UMEÅ UNIVERSITY

Biochemical analysis of Pfh1, the essential Pif1 family helicase in *Schizosaccharomyces pombe*

Jani Basha Mohammad

Department of Medical Biochemistry and Biophysics
Umeå 2019

Dissertation for PhD
ISBN: 978-91-7855-020-3
ISSN: 0346-6612
Possible series: 2011
Cover photo: © Jani Basha Mohammad
Electronic version available at: http://umu.diva-portal.org/
Printed by: KBC service center
Umeå, Sweden 2019

Today whatever you are is because of your past, so be good and do good; that decides your future.

My dad

List of abbreviations

G4: G-quadruplex

Sc: *Saccharomyces cerevisiae*

Sp: *Schizosaccharomyces pombe*

BaPif1: Bacteroides sp 2-1-16 Pif1

BsPif1: Bacteroides spp Pif1

hPif1: Human Pif1

hPif1-HD: Human Pif1-helicase domain

ssDNA: Single-stranded DNA

dsDNA: Double-stranded DNA

rDNA: Ribosomal DNA

tDNA: Transfer DNA

nPfh1: Nuclear Pfh1

SM: Signature motif

SF: Superfamily

CD: Circular dichroism

THT: Thioflavin T

ChIP: Chromatin immunoprecipitation

RPA: Replication protein A

Pol α: DNA Polymerase α

Pol δ: DNA Polymerase δ

Pol ε: DNA Polymerase ε

BLM: Bloom syndrome

SPR: Surface Plasmon Resonance

CMG: Cdc45–Mcm–GINS

RFB: Replication fork barrier

EMSA: Electrophoretic mobility shift assay

Table of Contents

Abstract

DNA stores the genetic information of all living organisms, and this information needs to be copied accurately and passed on to each daughter cell when a cell divides. However, the DNA replication machinery often meets obstacles in the genome that cause fork pausing and might result in DNA damage. DNA helicases are motor proteins that unwind duplex DNA structures using the energy from ATP hydrolysis. Helicases can also assist in replication fork progression by resolving obstacles that arise at hard-to-replicate sites such as tightly DNA-bound proteins, R-loops, and DNA secondary structures like G-quadruplexes (G4s). In this thesis, I focused on *Schizosaccharomyces pombe* DNA helicase Pfh1, which is localized in both the nucleus and the mitochondria and belongs to the evolutionary conserved Pif1 helicases. Pfh1 is an accessory replicative helicase, and the goal in this thesis was to gain a better mechanistic understanding of the role of nuclear Pfh1 (nPfh1). Our first aim was to elucidate the role of nPfh1 at G-quadruplex (G4) DNA. Aim two was to understand the function of nPfh1's signature motif. Aim three was to characterize the role of nPfh1 in strand annealing.

Some G-rich sequences can form a four-stranded DNA structure called G4 DNA, and the *S. pombe* genome contains about 450 bioinformatically predicted G4 structures. We selected two of these sequences, one located in the ribosomal DNA region and one located in the telomeric DNA region, and showed that they form inter- and intramolecular G4 structures, respectively. Next, we established a method to express and purify recombinant nPfh1 and demonstrated that nPfh1 binds to and unwinds these structures. In addition, Pfh1 bound to both the ribosomal and telomeric DNA regions *in vivo*, suggesting that Pfh1 can bind and unwind G4 structures *in vivo*. The purified nPfh1 also unwound RNA/DNA more efficiently than DNA/DNA structures, suggesting that nPfh1 has the ability to unwind R-loops *in vivo*. nPfh1 also showed protein displacement activity, suggesting that it can remove tightly bound proteins from DNA. All of these properties of nPfh1 suggest that it is important for fork progression and for preserving genome integrity.

Furthermore, nPfh1 stimulated strand annealing, and this activity did not require ATP hydrolysis. The strand-annealing activity was higher for complementary DNA/DNA compared to

RNA/DNA substrates and did not require a DNA overhang. Furthermore, by analysing Pfh1 truncated variants we demonstrated that the N-terminus region of Pfh1 was mainly responsible for the strand-annealing activity, however the C-terminus region also possessed some strand-annealing activity. Point mutations in the Pif1 signature motif (SM) have been shown to be associated with an increased risk of breast cancer in humans and with inviable *S. pombe* cells. We purified several SM variants and found that the unwinding and protein displacement activities of nPfh1 were dependent on the SM, but not the strand-annealing activity, suggesting that the SM is important for functions that require ATP hydrolysis.

In conclusion, in this thesis we identified nPfh1 as a potent G4 unwinder, and this is the only G4 unwinder identified in *S. pombe* to date. We also provided detailed mechanistic insights into nPfh1 and its different domains, and this has enhanced our understanding of Pfh1's role in maintaining genome integrity.

List of Papers

I. **G-rich telomeric and ribosomal DNA sequences from the fission yeast genome form stable G-quadruplex DNA structures in vitro and are unwound by the Pfh1 DNA helicase**

Marcus Wallgren, Jani B. Mohammad, Kok-Phen Yan, Parham Pourbozorgi Langroudi, Mahsa Ebrahimi, Nasim Sabouri

Nucleic Acids Research, Volume 44, Issue 13, 27 July 2016, Pages 6213–6231

II. **The Pif1 signature motif of Pfh1 is necessary for both protein displacement and helicase unwinding activities, but is dispensable for strand-annealing activity**

Jani B. Mohammad, Marcus Wallgren, Nasim Sabouri

Nucleic Acids Research, Volume 46, Issue 16, 19 September 2018, Pages 8516–8531

III. **Biochemical analysis of the strand annealing activity of the *S. pombe* Pif1 helicase**

Jani B. Mohammad, Nasim Sabouri

Manuscript

Introduction

The hunt for the genetic molecule reached its pinnacle in 1953 with the discovery of the DNA structure by James Watson and Francis Crick (1). DNA is written in a four-letter "alphabet", and it is the primary component of the chromosomes of all living organisms and stores all of the genetic information. The two strands of the DNA helix run in opposite directions, with one running in the 3'–5' direction and the other running in the 5'–3' direction. Within the helix, nitrogen bases form a ladder-like structure stabilized by hydrogen bonding. The DNA base number and composition varies from one species to another, and the human genome consists of about 3 billion base pairs arranged into 23 chromosome pairs.

DNA replication

Cell division requires the accurate replication of the organism's genetic information. During DNA replication, the leading strand is synthesized in a continuous manner along with the progressing replication fork, whereas the lagging strand is synthesized in the opposite direction and is replicated discontinuously. DNA replication initiation starts at the origin of replication (ORC). Initially, the MCM helicase forms a preinitiation complex involving Cdc6 and Cdt1 as loading factors. The ORC-Cdc6 complex recruits two MCM-Cdt1 complexes that surround the double-stranded DNA (dsDNA) (2-4). The kinase activity of CDK and DDK drives the assembly of this inactive hexameric replisome. This assembly then interacts tightly with DNA polymerase ε (Pol ε) to initiate replication (5-7). The CMG helicase unwinds the dsDNA generating single-stranded DNA (ssDNA) that is then coated with replication protein A (RPA). The DNA polymerase α (Pol α)-DNA primase complex lays down RNA primers for subsequent extension by Pol ε for leading strand synthesis and by Pol δ for lagging strand synthesis (8,9). The short, discontinuous sections of newly synthesized DNA on the lagging strand are known as Okazaki fragments. To make a continuous DNA strand on the lagging strand, the RNA primer is removed and the strands are ligated by DNA ligase 1 through a process called strand displacement. To carry out the ligation reaction, PCNA and DNA ligase 1 make extensive contact with the DNA by encircling the dsDNA (10). When bound to the polymerase core, the PCNA ring slides along

the DNA with the polymerase making it a highly processive enzyme capable of extending entire DNA chains without dissociating.

Hard-to-replicate sites

DNA replication proceeds bidirectionally from the origin of replication. However, replication might become unidirectional when replication forks meet regions that are hard to replicate or replication fork barriers (RFBs). In this thesis, I used *Schizosaccharomyces pombe* as my model organism. In *S. pombe,* ribosomal DNA (rDNA) and the mating type loci are two genomic locations where replication become unidirectional due to RFBs (11-13). Swi1/Swi3 is a non-nucleosomal complex that binds the rDNA and mating type loci and functions as natural RFBs (12,13). In *S. pombe,* rDNA is located in two clusters on both sides of chromosome III, and both clusters contain about 100–150 copies of rDNA repeats (11). DNA replication often encounters other natural obstacles along the genome, which if not removed might lead to replication pausing and DNA damage. For instance, in *S. pombe* replication is slowed down at highly transcribed RNA polymerase II and III genes and at telomeres (14,15). Slower fork progression might be due to tightly bound proteins at these sites and/or due to R-loop formations (16).

Moreover, DNA can form stable non-canonical secondary structures like the four-stranded G-quadruplex (G4) structure. G4 structures are thermodynamically stable, and if they are not resolved by helicases they can create obstacles to DNA and RNA metabolism (17). These structures will be described in more depth later in this thesis.

Helicases

DNA helicases were initially described as enzymes that use chemical energy to unwind duplex DNA. The first DNA helicase was purified and characterized in *Escherichia coli* in 1976 and was named TraI (Helicase I) (18). Helicases are motor proteins that translocate along nucleic acid strands using ATP as an energy source, and they play a vital role in nucleic acid metabolism, including DNA replication and repair, transcription, and translation (19). Helicases unwind dsDNA and secondary structures like G4s by hydrolyzing ATP, and they can also remove proteins that are tightly bound to nucleic acids and to chromatin (20-23). DNA

undergoes various harmful modifications due to exposure to UV light, toxic chemicals, and environmental stress, and this can lead to DNA damage. Damaged DNA is repaired by different DNA repair machineries that also involve DNA helicases (24).

Helicases are also involved in rewinding and annealing complementary strands of polynucleic acids in the presence or absence of nucleoside triphosphates (25). Some helicases involved in telomere metabolism, for instance, are known to have strand-annealing activity (26-29). In 2004, it was reported that RECQ5β is a DNA helicase with both DNA unwinding and DNA annealing activity (30). RECQ5β is the longer isoform of the RECQ5 helicase that is localized in the nucleus, and its annealing activity has been mapped to the C-terminus portion (aa 411–991) (30). The BLM helicase is another RECQ family helicase that shows strand-annealing activity (31,32), and in 2010, it was shown that the N-terminal region of the BLM/Sgs1 subdomain possesses strand-annealing activity *in vitro* (33). The WRN helicase is yet another RECQ family helicase with strand-annealing activity, and this activity has been was mapped to the C-terminal region (aa 1072–1150) (34,35).

Based on their shared sequence motifs, helicases are divided into six superfamilies (SFs) (19). The classification in SFs is based on amino acid sequences and the existence of helicase motifs. The majority of helicases fall into SF1 and SF2. SFs are also classified according to their translocation and unwinding polarity (19,36). Helicases that move along ssDNA are of Type A, while those moving along dsDNA are of Type B. Helicases with a 3'–5' directionality are of Type α, while those with a 5'–3' directionality are of Type β.

Superfamilies 1 and 2

SF1 is further divided into Type 1A and 1B (19,37). UvrD and Pcr A are members of SF1A and translocate with a 3'–5' directionality (19), while RecD, Pif1, and Dda are members of SF1B and translocate with a 5'–3' directionality (19). SF1 and SF2 contain at least seven conserved amino acid motifs that are very similar regarding organization, sequence, and secondary structures (38-40) and are named Q, I, Ia, II, III, IV, VI, and VI. The Q motif is conserved in more or less all SF1 helicases, and this motif coordinates the adenine base and is less conserved among those helicase families that do not exhibit ATP specificity (41). Motif I is also called the Walker A

motif, and the consensus sequence AxxGxGKT is associated with nucleotide binding proteins and with creating a phosphate binding loop (42). The motif Ia consensus sequence is TxxAA, which is involved in ssDNA binding and in transferring the energy from the ATP binding site to the DNA binding site (41,43). In SF1, motif Ia and motif III have been proposed to play a significant role in defining the translocation polarity (44). Motif II (Walker B), with the consensus sequence DExx, is involved in NTP hydrolysis (41,42,45), and the D and E residues coordinate the ATP-associated Mg^{2+} to activate the water molecule (46,47). Motif III, with the consensus sequence GDxxQLPP, is involved in DNA binding through the stacking of hydrogen bonds with the bases (41,45). The motif IV consensus sequence varies among families within SF1, and it supplies the stacking platform for the adenine as well as directs contact with the γ phosphate, which suggests that it might be directly involved in the hydrolysis of the NTP (43,48). Motif V interacts with the sugar-phosphate backbone of DNA (49). Motif VI has the consensus sequence VA(L/Y)TRA(K/R) and has been suggested to be involved in the ATPase and helicase activities (41,48).

SF2 is the largest and most diverse helicase superfamily, and these proteins are involved in various cellular processes (38). Translocation of SF2 motor cores along a nucleic acid mesh or grid is achieved by the concerted action of helicase domain 1 and 2 moieties (50,51). The SF2 helicases are characterized by the presence of 11 conserved motifs, Q, I, Ia, Ib, II, III, IV, Iva, V, Va, and VI. Motifs I and II contain the Walker A and B boxes, and these are the most conserved motifs across the SF (37). Motif IV is involved in the hydrolysis of NTPs (50), and motifs Ia, Ib, IV, Iva, and V make extensive contacts with the phosphodiester backbone of the DNA mesh (51). SF2 encompasses DEAD-box RNA helicases and some other RecQ-like and Snf2-like enzymes, and these are mostly type A helicases (38,41).

Superfamilies 3–6

SF3–6 encompass ring-forming helicases (37). SF3 belongs to the type A helicases, and SF3 helicases are found in small DNA viruses and are involved in various stages of the viral replication cycle (52). SF4 helicases are type B helicases from multiple systems including proteins from bacteriophages (T7 gp4 and SPP1 GP40), the bacterial helicase DnaB, and the eukaryotic mitochondrial helicase TWINKLE (19). SF5 includes the Rho helicases that act on

RNA substrates (19), and the essential unwinding character resembles that of SF4, which has 5'–3' polarity. SF6 helicases share a similar AAA+ α/β core and include the MCM, RuvB, RuvA, and RuvC helicases (19). The MCM protein is the replicative helicase in both archaea and eukaryotes.

Helicases and human diseases

Defects in helicases have been associated with various diseases such as premature aging, cancer, and neurodegenerative diseases (24,37,53). Defects in mitochondrial DNA helicases have been shown to lead to neuromuscular degenerative diseases (54,55), and the G4-binding helicases WRN, FANCJ, BLM, and hPif1 are associated with genetic diseases (56-58). Xeroderma pigmentosum is a rare autosomal recessive disorder caused by mutations in the Fe-S cluster of the SF2 XPD helicase.

Several diseases are linked to the RecQ helicases. The human genome encodes five RecQ helicases, whereas *E. coli*, *S. cerevisiae*, and *S. pombe* encode only one RecQ helicase (59). Variants of the RecQ-like DNA helicase family member REQL4 can lead to three distinct genetic disorders, namely Rothmund-Thomson syndrome, Baller-Gerold syndrome, and RAPADILINO (60). Bloom syndrome (BLM) is a rare autosomal recessive disorder connected to mutations in the BLM helicase in which cells exhibit increased numbers of chromatid gaps, breaks, and rearrangements in the chromosomes and accumulate large numbers of mutations at all loci, including coding and non-coding sequences of repetitive DNA (61). Werner syndrome is a rare autosomal recessive disorder linked to mutations in the WRN helicase, where cells exhibit reduced lifespan and increased genetic instability (62). The other two human RECQ helicases, RECQ1 and RECQ5, are not yet linked to any human diseases, but it seems that they might play a role in cancer predisposition or chromosomal instability (63). From here on out, I will focus on the Pif1 helicases, which have been the subject of my thesis.

Pif1 family helicases

Pif1 helicases are found in either single or multiple copies in different organisms, and the number of Pif1 helicases varies in prokaryotes and eukaryotes. For instance, humans and *S.*

pombe (Sp) encode a single homolog, whereas *Saccharomyces cerevisiae* (Sc) encodes two isoforms, ScPif1 and ScRrm3, and *Trypanosoma brucei* (Tb) has eight copies (64). All Pif1 helicases have highly conserved helicase motifs, but the C-terminus and N-terminus can differ significantly. Pif1 helicases have a 23 amino acid SM that is present between motifs II and III (64). This SM is found only in Pif1 helicases, and not in any other helicases that belong to SF1B (64). Pif1 helicases are found in both prokaryotes and eukaryotes, but below I focus mainly on eukaryotic Pif1 helicases.

ScPif1 and ScRrm3 helicases

Three decades ago, Pif1 (Petite integration frequency 1) helicase was first isolated in *S. cerevisiae* (hereafter called ScPif1) and was characterized as a mitochondrial helicase (65,66). Later on, Pif1 homologs were identified in both prokaryotes and eukaryotes (64,67). In a screen for rDNA recombination mutations, another Pif1 homolog was detected in *S. cerevisiae* and named Rrm3 (hereafter called ScRrm3) (64). Since then, ScPif1 and ScRrm3 have been the most well studied eukaryotic Pif1 helicases and have been shown to play a significant role in *S. cerevisiae* genome maintenance. ScPif1 has been shown to be recruited to the replisome only when it is required, whereas ScRrm3 is part of the replisome complex itself and thus moves along with the replication fork (68,69). ScRrm3 facilitates replication at many hard-to-replicate regions, including transfer DNA and 5S ribosomal DNA (5S rDNA) genes, RFBs on the ribosomal DNA (rDNA), and inactive origins (14,70-72). ScPif1 blocks fork progression at the RFBs in ribosomal DNA (72), and ScPif1 also plays a role in the processing of Okazaki fragments and in promoting break-induced replication repair of dsDNA breaks(73-76). ScPif1 also suppresses genome instability at predicted G4 structures to promote replication fork movement at these sites (69,77) and maintains the mitochondrial DNA (65,78-80).

By comparing its unwinding activity on different substrates, ScPif1 has been shown to be more efficient on RNA/DNA hybrids than on DNA/DNA substrates, suggesting that ScPif1 can unwind R-loops *in vivo* (81,82). ScPif1 also potently unwinds G4 structures (64,77). These activities have not been tested for ScRrm3 because full-length recombinant ScRrm3 has been very difficult to express and purify. Both helicases are involved in telomere maintenance. In the absence of ScPif1 and ScRrm3, the telomeres become more extended, and overexpression of

ScPif1 leads to shorter telomeres (83), showing that ScPif1 is a negative regulator of telomere length. It has also been shown that telomerase is removed from DNA ends by ScPif1 (84) and that lack of ScRrm3 leads to replication fork pausing at telomeres (71). Apart from its helicase activity, ScPif1 also shows strand-annealing activity and removes proteins from DNA *in vitro* (85,86).

SpPif1 helicase - Pfh1

S. pombe is a commonly studied model organism because many of its chromosomal features resemble those of human cells (87). Pfh1 (Pif1 homolog 1) is the *S. pombe* Pif1 helicase and is encoded by an essential gene and has functions in both nuclear and mitochondrial genome maintenance (88), although the role of Pfh1 in the mitochondrial genome is relatively unexplored. Knowledge of the roles and functions of Pif1 helicase came prior to this thesis mainly from ScPif1 because other eukaryotic Pif1 helicases are insoluble and/or unstable and thus are difficult to purify. Similar to ScRrm3, Pfh1 interacts with the replisome complex and moves along with the replisome during replication, suggesting that Pfh1 is a component of the replisome(16). Pfh1 interacts with the major components of the replisome complex, including DNA polymerase ε, MCM helicase, RPA, and PCNA, to promote replication and to suppress DNA damage at hard-to-replicate sites (16). Chromatin immunoprecipitation combined with sequencing (ChIP-seq) showed that Pfh1 binds to 20% of predicted G4 structures. Furthermore, fork progression slows down at predicted G4 structures, and an increased level of DNA damage has been detected in Pfh1-depleted cells, suggesting that Pfh1 is needed at these sites to promote replication fork progression (89). Pfh1 is also needed at other hard-to-replicate regions. By 2D gel analysis and ChIP-seq, it has been shown that replication at highly transcribed regions in RNA polymerase II and RNA polymerase III genes is dependent on Pfh1(90). Furthermore, a temperature-sensitive pfh1 mutant strain of *S. pombe* shows DNA damage at locations of transcribed RNA polymerase III genes, further supporting that idea that Pfh1 is needed to promote DNA replication at these sites(16). In further support of this, it has been shown that Pfh1 depletion leads to the accumulation of converged forks at replication termination sites such as RFBs (90,91).

The L319P point mutation in the hPif1 SM is found in many families with increased risk of breast cancer (58), and the corresponding mutation Pfh1L430P in *S. pombe* results in inviable cells, suggesting that this site is essential for the functions of Pfh1 (58).

Pfh1 is also enriched at telomeres *in vivo*, and overexpression of Pfh1 leads to telomere lengthening, suggesting that Pfh1 is a positive regulator of telomere length(92). Moreover, Pfh1 is also needed during Okazaki fragment maturation (93).

Mammalian Pif1 helicases

Human Pif1 (hPif1) belongs to the helicase 1B superfamily. The genomes of higher eukaryotes like humans and mice only encode one Pif1 family helicase, and hPif1 and mPif1 share nearly 84% sequence similarity across the whole open reading frame (94). In comparison, hPif1 and *S. pombe* Pfh1 share 36% sequence identity in the conserved helicase domain (95). Expression and purification of full-length hPif1 is a particularly difficult task, but a study of recombinant hPif1 showed that is has 5'–3' helicase activity that can unwind both DNA/DNA and RNA/DNA substrates (96). Similar to other Pif1 family members, hPif1 also localizes in both the mitochondria and nucleus (94,97). hPif1 is found in highly proliferating cells (27,59,94), and depletion of hPif1 in cancer cells by si-RNA results in apoptosis, reduced survival, hypersensitivity to therapeutic DNA replication inhibitors, and defective cell cycle progression. These defects are p53 independent, and non-cancerous cells do not show a similar response (98). During cancer development, hPif1 plays an essential role in overcoming the replication stress induced by oncogenes, and suppression of hPif1 results in slower migration of replication forks and increased replication fork arrest under normal cell cycle conditions (99). The binding efficiency of hPif1 to telomeric DNA is 100-fold higher compared to its binding efficiency to random DNA sequences, and overexpression of hPif1 shortens the telomere length (96). As mentioned earlier, the Pif1L319P point mutation in the SM is associated with human breast cancer (58). hPif1 shows protein displacement activity by facilitating fork progression at telomeric regions by removing DNA-bound proteins (100), and hPif1 also possesses ATP-independent strand-annealing activity (101). The purified recombinant hPif1 protein shows annealing activity in its N-terminus portion (aa 1–180) and helicase domain (101,102).

Structure of Pif1

There are four available crystal structures of Pif1 in different organisms, including *Bacteroides spp* Pif1 (BsPif1) (103), *Bacteroides sp 2-1-16* (104), *S. cerevisiae* Pif1P (ScPif1p) (105), and the hPif1 helicase domain (hPif1-HD) (104). Unlike other SF1B helicases, the Pif1 helicase also contains two RecA-like domains referred to as domain 1A and domain 2A in their catalytic core where the helicase motifs are clustered. These two domains are distinguished by a cleft where ATP binding and hydrolysis take place. The 1B domain of BsPif1 forms an ordered loop, whereas in BaPif1 and hPif1-HD the 1B domain is formed by a loop and an α-helix. The 1B domain is proposed to act like a wedge that separates the incoming dsDNA (106). In the BsPif1 and BaPif1 structures, residues of motifs I, III, IV, and V make contact with the ATP (103,104). The SM is located between motifs II and III (64) and is composed of an α-helix and a turn located in such a way as to approach the DNA binding site, and this is same in all three structures of Pif1 helicases (103,104).

G4 structures

In vitro

G4 DNA is formed when four guanines participate in Hoogsteen hydrogen bonding to form a planar arrangement termed a G-quartet, and these G-quartets stack together to form a G4 structure (107,108). G4 structures are stabilized by either monovalent cations like Na+, K+, and Li+ or by stabilizing ligands like $TMPyP_4$ and Phen-DC_3 (109). G4 structures can be intramolecular or intermolecular and can have various topologies like parallel, antiparallel, and hybrid structures (110). Intermolecular G4 structures are dependent on their loop length and the nature of the cation in the central channel. RNA and hybrid RNA/DNA sequences can also form stable G4 structures (111-113). The structural diversity of G4 structures is greater for DNA compared to RNA, and although G4 DNA structures are thermally stable, a distinguishing feature of RNA G4 structures is their greater stability compared to DNA G4 structures (112,114,115).

In vivo

In humans about 700,000 G4 motifs have been found in the genome using high-throughput sequencing methods (116). In *S. pombe* an algorithm searching for the consensus motif ($G_{\geq 3}\,N_{1\text{-}25})_3\,G_{\geq 3}$ found nearly 446 G4 motifs in the genome (89,117). In humans, G4-forming regions are not exclusively found in the nuclear genome, but are also found in the mitochondrial genome (118,119).

Increasing evidence indicates that intramolecular G4 structures are biologically relevant with numerous reports suggesting that G4 structures are present in living cells (120). Many biological processes are potentially influenced by G4 structures (110,121), and it is suggested that changes in some cellular processes might be due to misregulation of G4 structures leading to genome instability and subsequent pathological states (17,56). G4 structure formation might play beneficial roles, such as in telomere capping and the regulation of replication initiation or transcription, as well as harmful roles such as arresting the progression of the replication fork (122). G4 structures might have different biological roles, but these structures have to be resolved during replication and transcription because unresolved G4 structures might act as barriers that affect the progression of replication, transcription, and the reannealing of DNA duplexes.

G4 motifs are not randomly dispersed in the genomes of various organisms, but are enriched at transcriptional start sites and promoters, meiotic and mitotic double-strand break sites, replication hot spots, and telomeres (123,124). In humans, nearly 3,000 genes are predicted to be able to potentially form G4 structures in the 5'UTR region of their mRNA, which might repress translation of these genes (112,125). The predicted G4-forming motifs are particularly abundant in RNA transcripts of telomeres and TERRA (126). In *S. pombe*, G4 motifs are enriched in telomeres and rDNA regions (89).

The formation of G4s *in vivo* can be studied by using anti-G4 antibodies such as BG4, Sty49, and D1 (127-129). These studies have identified the presence of G4 structures in the telomeres of the ciliate *Stylonychia lemnae* (128) and have shown the cell cycle dependence of such structures in mammalian cells (130).

The formation of G4 structures *in vivo* is predicted to be favored by molecular crowding and superhelical stress as well as by specific G4-binding proteins (121,131). Various G4-

interacting proteins have been identified in different studies, and among these are several helicases that specifically unwind G4 structures *in vitro* (132). In the late 1990s, several reports showed that dsDNA helicases (SV40, BLM, SgsI) can unwind G4 structures *in vitro* (133-136).

Methods for studying G4 structure unfolding by helicases

Folded ssDNA with a predicted G4 structure sequence can form a mono, bi, or tetramolecular G4 structure, and these can be detected by their different migration patterns in native polyacrylamide gels. For instance, ssDNA migrates faster than the dimeric or tetrameric substrates in native gel electrophoresis, while monomeric (intramolecular) G4 structures migrate faster than ssDNA. Gel-based assays are one of the main biochemical techniques used to study DNA and RNA unwinding by helicases, and these assays are also commonly used to examine G4 unwinding by different helicases. In my thesis, many of the oligonucleotides that I used for the gel electrophoresis experiments were 5' end-labeled with ^{32}P so that I could visualize the results by phosphorimaging.

Surface plasmon resonance is another method used to study how helicases resolve G4s. This real-time method was employed, for instance, to study SV40 T-ag helicase activity at sites of G4 DNA. Binding of the SV40 T-ag helicase to the DNA substrate requires ATP, suggesting that the helicase activity is ATP dependent (137).

Single-molecule studies using fluorescence resonance energy transfer are another method for studying the unwinding of intramolecular G4 structures (138). This method was used, for instance, when studying the BLM helicase. Under physiological salt concentrations, the BLM helicase is active on intramolecular G4 in the absence of ATP. However, the same group also reported that the unwinding activity of BLM helicase is ATP dependent. To explain this apparent discrepancy, they suggested that the BLM helicase binds to the ssDNA region of the G4 and that this eventually leads to unwinding of the G4 structure (139).

In addition to these *in vitro* assays, *in vivo* assays are also used to detect G4 structures, for example, the use of G4-specific antibodies (130,140). Cells that are FANCJ helicase deficient show stronger nuclear staining with the anti-G4 antibody compared to cells producing normal levels of FANCJ, suggesting that the FANCJ helicase resolves G4s (140). Cells treated with G4

ligands – such as pyridostatin – that stabilize the G4 structures suffer from elevated levels of DNA damage, and this damage can be monitored by an antibody that recognizes phosphorylated histone 2AX (also called γH2AX) (141).

ChIP is also a common method to determine if DNA helicases function at G4 DNA. For example, both ScPif1 and Pfh1 are enriched at G4 motifs in vivo, suggesting that they might be enriched there for the purpose of unwinding G4s (69,89).

Aim

The aim of this thesis was to understand how Pfh1 maintains genome integrity.

Specific aims:

1. To determine the role of nPfh1 at G-quadruplex DNA
2. To understand the function of nPfh1's signature motif
3. To characterize the role of nPfh1 in strand annealing

Summary of my studies

Paper I: G-rich telomeric and ribosomal DNA sequences from the fission yeast genome form stable G-quadruplex DNA structures in vitro and are unwound by the Pfh1 DNA helicase

In this paper, we conducted several *in vitro* and *in vivo* experiments to explore the formation of G4 structures in *S. pombe*. Out of 446 predicted G4 structures (89), we selected two G4 motifs, one from the rDNA and the other from the telomeric DNA, to study in more detail. These sites were selected because the density of G4 motifs at these sites is significantly higher than in other regions of the *S. pombe* genome (89). Furthermore, G4 motifs at both of these genomic locations are evolutionarily conserved, suggesting that these G4 structures play important regulatory functions at these sites (89,117).

Not all *in silico*-predicted G4 structures can actually form a G4 structure. To determine whether a certain G-rich sequence can form a G4 structure, it is important to use different methods and conditions to examine G4 formation. Therefore, we used various conditions to determine if these two predicted G4 structures can indeed adopt a stable G4 structure. First, we used circular dichroism (CD) analysis, which is a commonly used method to examine the topology of folded oligonucleotides. A CD spectrum with a positive peak at 264 nm and a negative peak at 245 nm is typical of parallel G4 structures. If the positive absorption maximum is at 295 nm and the minimum at 265 nm, this is typical of anti-parallel G4 structures, whereas positive peaks at both 265 nm and 295 nm are indicative of a hybrid structure or a mixture of parallel and anti-parallel G4 structures (142-146). Using CD analysis, we demonstrated that both the rDNA and telomere DNA sequences from *S. pombe* form stable G4 structures. Reports of other examined G4 structures show that the stability of the G4 structures increases with K+ ions compared to Na+ ions, and this stabilization is further increased as the concentration of salt increases from 100 mM to 1 M (110). In our experiments, the rDNA G4 adopted a parallel G4 structure in the presence of both Na+ and K+ ions, and in the presence of K+ the rDNA G4 was highly thermally stable. The telomeric G4 DNA showed a broader positive CD peak in the presence of both Na+ and K+ ions, suggesting that it adopts a hybrid structure.

We also used native gel electrophoresis to examine the molecularity of the G4 structures. These assays showed that rDNA G4 forms an intermolecular G4 structure because the folded rDNA G4

oligonucleotides migrated slower in the gel than the unfolded rDNA oligonucleotides. Because the rDNA G4 sequence is a repetitive sequence in the rDNA loci region, these sequences might also form intermolecular G4 structures *in vivo*. We found that the telomeric G4 DNA formed an intramolecular structure because it migrated faster through the gel compared to unfolded oligonucleotides. Human telomeric DNA can adopt several different G4 topologies, but many of them also adopt intramolecular G4 structures, showing that both the human and *S. pombe* telomeric G4 structures form intramolecular G4 structures (147-151).

To further confirm our findings, we used the thioflavin T (THT) assay. THT is a small molecule that becomes fluorescent when it binds to G4 DNA, but not when binding to ssDNA (152). We observed an intense signal for telomeric and rDNA G4 oligonucleotides but no signal for ssDNA. Mutated oligonucleotides, where either a single G in one G-tract or all four G-tracts of the rDNA or telomere DNA were mutated, showed reduced fluorescence enhancement when compared to the non-mutated G4 oligonucleotides. These results were consistent with our CD results, where the mutated oligonucleotides demonstrated lower CD signal than the non-mutated oligonucleotides.

Next, we examined whether Pfh1 is enriched at rDNA and telomeric G4 DNA regions *in vivo* by using ChIP combined with quantitative PCR. In these experiments, we used both asynchronous and synchronous cells. Pfh1 was enriched at both the rDNA and telomeric regions, and the binding occupancy was higher for asynchronous cells compared to G2-arrested synchronous cells, suggesting that binding of Pfh1 occurs outside the G2 phase, probably during S-phase. In fact, Pfh1-depleted cells have increased replication pausing at predicted G4 structures (90), supporting our data showing that Pfh1 is enriched at these G4 sites during S-phase.

To determine if Pfh1 can bind to these structures *in vitro*, the nuclear isoform of Pfh1 was expressed and purified recombinantly. To improve the solubility of the protein (93,153,154), we examined several epitope tags in *E. coli*. His_6-Trx-nPfh1-Flag was our best candidate, where Trx was used as a carrier protein and gave increased solubility of Pfh1 and the His_6 and Flag tags improved the purity of the enzyme. To test the binding affinity of Pfh1, we performed EMSA experiments. EMSA is a common affinity electrophoresis method to study the interactions between proteins and nucleic acids. For these experiments we purified the nuclear isoform of Pfh1 as well as a helicase-dead mutant having an alanine substitution at the conserved lysine 338

(nPfh1-K338A) in the Walker A motif. nPfh1 variants displayed high binding to rDNA G4 folded in the presence of either Na+ or K+, but no or very poor binding to telomere sequences folded in K+. Next, we used a helicase assay to show that nPfh1 could unwind both types of G4 structures. nPfh1 unwound the rDNA G4 structures folded in Na+ very efficiently, indicating that Pfh1 is very efficient in unwinding G4 structures and in this case is more similar to ScPif1 than to bacterial Pif1 helicases (77). hPif1 also unwinds G4 structures *in vitro* (57). Together, these data suggest that resolving G4 structures is a common function for Pif1 helicases.

We also used a G4-stabilizing ligand, $PhenDC_3$, to examine how efficiently nPfh1 unwinds G4 structures. We found that nPfh1 was a very potent G4 unwinder because it could unwind these stabilized structures as well. The unwinding activity was higher for rDNA structures folded in Na+ and $PhenDC_3$ compared to those folded in K+ and $PhenDC_3$. However, there was no difference in unwinding activity for telomere structures folded in K+ and $PhenDC_3$ compared to telomere DNA folded in K+ without $PhenDC_3$.

We conclude that both the rDNA and telomeric DNA can form stable G4 structures *in vitro*. *In vivo,* the Pfh1 occupancy is higher at rDNA and telomere DNA regions. The rDNA and telomere sequences can form G4 structures and are further stabilized by ligands like $PhenDC_3$, and nPfh1 can bind and unwind these G4 structures even when stabilized by $PhenDC_3$. Thus, Pfh1 is another G4 unwinder of the evolutionarily conserved Pif1 family helicases and is the first G4 unwinder identified in *S. pombe*.

Paper II: The Pif1 signature motif of Pfh1 is necessary for both protein displacement and helicase unwinding activities, but is dispensable for strand-annealing activity

In this paper, we studied the role of the Pif1 SM *in vitro*, which is important to study because it is a characteristic motif for the Pif1 helicases and is associated with familial breast cancer (58,64). The Pif1 SM in Pfh1 is located at residues 428–450. To determine the role of the SM, we created several constructs, including nPfh1(WT), nPfh1-K338A (KA), nPfh1-L430P (LP), nPfh1-L430V (LV), nPfh1-L430A (LA), and nPfh1-Δ21(Δ21), and expressed and purified them following the purification protocols in paper 1 in this thesis. The LP variant is the corresponding variant that is found in several families with breast cancer and that also leads to inviability in *S. pombe* cells (58). The Δ21 variant lacks the first 21 amino acid residues in the SM. To understand the importance of the SM, we first used WT and KA as our controls in all experiments and then compared those data with the SM variants.

First, we examined the helicase unwinding activity of WT and KA on several different DNA and RNA substrates. In these experiments, WT unwound RNA/DNA substrates very efficiently compared to DNA/DNA substrates, which is similar to ScPif1 (81,82). These results might indicate that nPfh1 has a role in removing R-Loops or in displacing RNA bound to DNA *in vivo.* To determine the role of nPfh1 on replication forks, we tested two different substrates that resemble a replication fork, a Y-fork and an Okazaki fragment. nPfh1 very efficiently unwound the Okazaki fragment structure compared to the Y-fork substrate, and this suggests that nPfh1 might play a role in removing RNA/DNA flaps during Okazaki fragment maturation. These findings also agree with earlier observations made with recombinant NusA-Pfh1 helicase (93). Pfh1's unwinding efficiency from the highest efficiency to lowest was rDNA G4, Okazaki fragment, Y-fork, RNA/DNA, and DNA/DNA. KA was used as a control for our purification, and we did not detect any unwinding by this helicase-dead variant, indicating that the observed results are due to the intrinsic Pfh1 activity and not due to impurities from our protein preparations.

Next, we examined the binding pattern of nPfh1 for these substrates. nPfh1 and nPfh1-KA bound equally well to all of the substrates, showing that nPfh1-KA can bind to but cannot unwind the examined structures. nPfh1 showed higher binding affinity towards rDNA G4 and Okazaki fragments compared to the Y-fork substrate, and this binding affinity reflects nPfh1's helicase

activity. Furthermore, nPfh1 showed no or very poor binding efficiency towards DNA/DNA and RNA/DNA substrates, suggesting that the higher unwinding activity of RNA/DNA substrates is not due to higher affinity for the RNA/DNA structure, but is more likely due to the greater unwinding activity of nPfh. nPfh1 bound to ssDNA and ssRNA equally well, and this suggests that the binding efficiency depends on the structure of the substrate rather than on its chemical identity.

Some helicases can remove proteins from ssDNA. Pfh1, ScPif1, and ScRrm3 are all suggested to have this ability because they all promote fork progression at sites bound by stable nonnucleosomal protein complexes (70-72,90,91). ScPif1 can remove proteins from DNA *in vitro* (86,155,156), and to determine if Pfh1 also has a protein-displacement activity we performed a protein displacement assay using a biotin-labeled ssDNA bound to streptavidin. The binding between biotin and streptavidin is strong, but nPfh1 was able to displace the bound streptavidin from the biotin-labeled ssDNA. These results show that Pfh1 can remove proteins from DNA, suggesting that Pfh1 also performs this task *in vivo* at sights of tightly bound proteins in order to promote fork progression.

From the literature, we know that some helicases not only have unwinding activity, but also have a strand-annealing activity where they rewind two complementary oligonucleotide strands. To determine if Pfh1 has strand-annealing activity, we performed annealing assays by mixing partially complementary single-stranded oligonucleotides with nPfh1. We found that similar to ScPif1 and hPif1 (85,102) nPfh1 could also accelerate the annealing of complementary oligonucleotides, showing another evolutionarily conserved function of Pif1 helicases. Together, all the above-described functions of nPfh1 suggest that it has several different functions that help the cell to maintain the integrity of its genome.

In the available crystal structures of Pif1 helicases, the SM is folded into an α-helix, a loop, and part of a β sheet (103,104). To determine the role of the SM, we again performed a helicase assay and tested the LV, LA, LP, and Δ21 variants with the same substrates used for WT nPfh1. Of all the nPfh1 variants, LV was more active and could unwind all of the substrates that we tested. The unwinding activity was somewhat weaker when compared to WT nPfh1. In addition, WT and LV showed the same level of ATPase activity. These results indicate that the low helicase activity for LV might be due to something other than a lack of fully functional ATP

hydrolysis. The LP and Δ21 variants did not show any helicase activity, suggesting that the SM is important for the unwinding activities of Pfh1.

We then tested the binding preference of these nPfh1 variants. All of the SM variants showed binding to rDNA G4 and Okazaki fragments, and LV also showed binding towards Y-forks and ssDNA. None of the nPfh1 variants showed binding towards 5'–3' DNA/DNA, 5'–3' RNA/DNA, or ssRNA or ssDNA oligonucleotides.

Residue I118 in BaPif1 corresponds to L430 in *S. pombe*. Because LP did not bind to ssDNA nor did it unwind the partial duplex structure, it behaved similarly to recombinant BaPif1 I118P (104). BaPif1 I118A and SpPfh1 L430V behave similarly to WT in terms of binding and unwinding activities. The BaPif1 crystal structures show that I118 interacts with the part of the 1A domain that contacts the ssDNA, and it is proposed that the proline substitution at this position creates a kink in the α-helix that destabilizes the interaction with the 1A domain (104). This might be the reason why LP is unable to bind ssDNA or to unwind any of the substrates used in this study.

The SM variants did not show any protein displacement activity, suggesting that the SM is essential not only for unwinding, but also for protein displacement activity. Our *in vitro* findings correspond well to an *in vivo* study of the functions of ScPif1's SM (157) that showed, for instance, that ScPif1's SM is needed for functions in mtDNA maintenance and Okazaki fragment maturation, both of which are sites where Pfh1 also acts in an ATP-dependent manner (157).

Finally, all of the nPfh1 variants showed annealing activity. Surprisingly, LV showed better annealing activity compared to WT. These results suggest that the SM is not needed for the annealing activity of Pfh1, which is the only activity that we have tested for nPfh1 that does not require ATP hydrolysis.

In conclusion, we have demonstrated that nPfh1 has unwinding, binding, protein displacement, and annealing activities. The breast cancer-linked variant, nPfh1-LP, showed binding activity, although only for certain structures, and it showed annealing activity, but it lacked helicase and protein displacement activities. Because the corresponding mutation is linked to cancer in humans, perhaps these altered properties found in nPfh1-LP are also found in hPif1L319P, which could explain the genome instability seen in these cancer patients.

Paper III: Biochemical analysis of the strand-annealing activity of the *S. pombe* Pif1 helicase

The strand-annealing activity of Pfh1 was discovered by us and described for the first time in paper 2 in this thesis (158), and in paper 3 we characterized the role of Pfh1's strand-annealing activity. To perform the strand-annealing experiments, we mixed two complementary/partially complementary oligonucleotides and examined both the spontaneous formation of duplexes and the formation of duplexes in the presence of Pfh1 by native gels. We used DNA or RNA oligonucleotides that formed different structures with no or with different sizes of overhangs. We also expressed and purified several new Pfh1 constructs, namely N^{22-308} (containing residues 22–308 and lacking helicase motifs I and VI), N^{22-352} (containing residues 22–352 and including the Walker A motif and helicase motif I and lacking the helicase motif VI), $N^{22-352KA}$ (containing the same motifs but with the K338A point mutation), and $C^{309-805}$ (containing residues 309–805 and including helicase motifs I and VI and lacking the N-terminus). These constructs were used to determine if there is a specific domain that is responsible for Pfh1's strand-annealing activity.

Although two complementary oligonucleotides can spontaneously form duplex DNA, this process is quite slow. The presence of Pfh1 significantly accelerated the formation of duplex substrates from complementary oligonucleotides. The nPfh1-KA variant is mutated in the conserved K in the Walker A motif and cannot bind ATP. We found that nPfh1 and nPfh1-KA showed similar strand-annealing activity on all substrates tested, showing that ATP binding is not required for strand-annealing activity. Hence, we could exclude ATP from the reactions and thereby avoid nPfh1's unwinding activity and only focus on the strand-annealing activity.

Pfh1 cannot unwind a blunt-end substrate and it requires an overhang to bind to in order to initiate unwinding. Using a blunt-end substrate, we found that the presence of ATP did not inhibit strand annealing by Pfh1, suggesting that ATP binding does not regulate the switch from unwinding to rewinding by Pfh1. The presence of ATP does not inhibit the strand-annealing activity of ScPif1 either (85), suggesting that the regulation of unwinding and rewinding is conserved between the two Pif1 helicases. However we observed reduced annealing activity in the presence of ATP for N^{22-352}, suggesting that ATP binding inhibits this variant. Because we showed that this variant cannot unwind or hydrolyze ATP, perhaps the binding of ATP causes structural changes in the N^{22-352} variant that lead to strand-annealing inhibition.

We found that nPfh1 can accelerate the annealing of many different substrates, even those without an overhang, as well as RNA/DNA oligonucleotides. However, the annealing activity on complementary DNA/DNA oligonucleotides was much higher than on RNA/DNA. The N-terminus $N^{22\text{-}352}$ and $N^{22\text{-}352KA}$ nPfh1 variants showed greater annealing activity compared to full-length nPfh1. For instance, both N-terminus $N^{22\text{-}352}$ variants showed almost 100% annealing activity on both substrates that had two overhangs – the duplex DNA/DNA and Y-fork substrates – suggesting that overhangs on the substrates promote strand-annealing activity. Overall, we found that the N-terminus $N^{22\text{-}352}$ variants showed greater annealing activity compared to the full-length nPfh1 variants on all tested substrates, suggesting that full-length nPfh1 partly suppresses the strand-annealing activity.

We also tested the annealing activity of two other nPfh1 variants, $N^{22\text{-}308}$ and $C^{309\text{-}805}$. The $C^{309\text{-}805}$ variant showed similar levels of strand-annealing activity as the full-length nPfh1 in creating the duplex DNA/DNA substrate, but this activity was two-fold lower compared to the $N^{22\text{-}352}$ variants. The $N^{22\text{-}308}$ variant showed a very low level of annealing activity when compared to all other nPfh1 variants. These results suggest that the additional 44 amino acid residues present in the $N^{22\text{-}352}$ variant are important for nPfh1's annealing activity. Furthermore, not only the N-terminus region could accelerate strand annealing, but the C-terminus could as well, although less efficiently. This is similar to hPif1 that also possesses strand-annealing activity in both its N- and C-terminus regions (101).

We also used EMSA to determine if the strand-annealing activity of the different nPfh1 variants reflects their binding efficiency to single-stranded oligonucleotides or to the annealed substrates. Full-length nPfh1 effectively bound the substrates that had a longer overhang or that had two overhangs. Binding of the $C^{309\text{-}805}$ variant was very similar to full-length nPfh1 on almost all substrates. However, although the $N^{22\text{-}352}$ variants showed higher annealing activity on these substrates compared to full-length nPfh1, $N^{22\text{-}352}$ could not efficiently bind to any of these substrates nor to their corresponding single-stranded oligonucleotides. Surprisingly, the $N^{22\text{-}352}$ variants efficiently bound to the rDNA G4 substrate, suggesting that the $N^{22\text{-}352}$ variants can only effectively bind certain types of substrates. Therefore, the strand-annealing activity does not correlate well with the binding preferences of the different nPfh1 variants for different substrates.

These results suggest that there are likely other properties of nPfh1, and not just binding preference, that regulate the strand-annealing activity.

In conclusion, this study suggests that nPfh1 has annealing activity and that this annealing activity is greater for the truncated $N^{22\text{-}352}$ variant. The truncated $C^{309\text{-}805}$ variant also shows annealing activity, but it is less efficient compared to $N^{22\text{-}352}$. Therefore, similar to hPif1, both the N- and C-terminus regions of nPfh1 possess strand-annealing activity. The annealing activity was not dependent on ATP, suggesting that binding of ATP is not the regulating factor that controls the switch between unwinding and annealing activities.

Conclusions

In this thesis, I performed in-depth biochemical studies of Pfh1 to better understand what roles Pfh1 might play in the nuclear genome. One of the biggest challenges in my studies was to establish a protocol for expressing and purifying Pfh1, along with the different Pfh1 variants that I have been studying, because Pfh1 has proven to be very difficult to purify (93,159,160).The protocol I developed allowed me to purify not only the full-length nPfh1, but also many different Pfh1 variants, and this allowed me to perform many detailed biochemical studies of Pfh1 that were not previously possible.

Identifying which G4 motifs form G4 structures *in vivo* and which proteins are involved in G4 biology is very important because G4 structures are associated with many diseases such as BLM, xeroderma pigmentosum, and WRN (56-58). Increased knowledge about G4 biology also leads to a better understanding of which biological mechanisms G4 structures are involved in and what activities these structures regulate in the cell. In *S. pombe,* the rDNA and telomere DNA contain predicted and evolutionarily conserved G4 motifs that we showed form stable G4 structures, and we also showed that Pfh1 binds to both structures *in vivo* and *in vitro* and that it unwinds them *in vitro*. The formation of G4 structures in the telomeres is perhaps important for the protection of the telomeres and for the regulation of telomerase in *S. pombe* as has been suggested for other organisms (161,162). The formation of the G4 structure in the rDNA might be important for regulating the transcription of the rDNA genes.

Pfh1 is so far the only G4 unwinder identified in *S. pombe*, but there are most likely other helicases in *S. pombe* that are also able to unwind G4 structures. Furthermore, our work showed that Pfh1 unwound the intermolecular rDNA G4 structure more efficiently than the telomeric intramolecular G4 structure. Therefore, perhaps different helicases are specialized for unwinding different types of G4 structures. In searching the literature for G4 unwinding homologs (163), the Werner syndrome 3'–5' helicases Rqh1 and Hrq1 and the FANM 5'–3' helicase Fml1 are all good candidates for helicases in *S. pombe* that might be interesting to test for G4 unwinding.

During replication, the replication machinery often meets obstacles along the genome, including replication-transcription conflicts, R-loops, and non-canonical structures like G4s. If these obstacles are not resolved, they can lead to DNA damage. In paper 2, we showed that Pfh1 can resolve RNA/DNA structures more efficiently than DNA/DNA structures, and therefore it

might play a role in resolving R-loops *in vivo*. In the same study, we also showed that Pfh1 can remove streptavidin from DNA. This protein-displacement activity is probably important during replication-transcription conflicts to remove proteins that hinder the progression of the replication fork. This property is most likely also important for removing RFBs at rDNA and mating type loci. In fact, Pfh1 is not needed at RFBs in Swi1-deleted cells, suggesting that Pfh1 removes Swi1 from the DNA (90). Pfh1 variants with defects in the SM showed that the SM is important for both unwinding and protein displacement activities; however, because Pfh1 cells with LP (58) or LA (data not shown) mutations are not viable, we could not test these phenotypes *in vivo*. However, in an accompanying paper (157) with paper 2 in this thesis that was also published in *Nucleic Acids Research*, the SM of ScPif1 was tested *in vivo*. They found that the SM of ScPif1 is needed for all activities that require ATP hydrolysis, which is consistent with our *in vitro* work on Pfh1. Together these two studies were the first determination of the function of the SM in eukaryotes.

We found that the SM was not essential for Pfh1's strand-annealing activity. We also showed that Pfh1 had the ability to accelerate strand-annealing on many substrates and that similar to hPif1 both the N-terminal region and the helicase region had annealing activities. The strand-annealing property of Pfh1 might be important for preventing G4 structures from re-forming after Pfh1 has resolved them. Alternatively, strand annealing might be essential for telomere maintenance because Pfh1 binds to telomeres *in vivo* (92). There are many suggestions for how annealing helicases regulate their unwinding and rewinding activities. One report suggests, for instance, that multimerization might influence the strand-annealing activity because the hRECQ1 helicase shows annealing activity as a trimer and unwinding activity as a dimer or monomer (164). Likewise, perhaps multimerization of Pfh1 might promote strand annealing as well. However, more experiments are needed to determine what factors control the switch between unwinding and rewinding of Pfh1.

Finally, I have shown in this thesis that nPfh1 has many different functional properties, including strand annealing and protein-displacement activities. All of these properties play important roles in maintaining genome integrity, and most of these properties are likely also shared by hPif1 and are important for preserving genome stability in human cells.

Acknowledgments

This journey would not have been possible without the support of my family, friends, and teachers. To my family, thank you for encouraging me in pursuing my dreams. I am especially grateful to my parents who have supported me emotionally and financially. I always knew that you believed in me and wanted the best for me. Thank you to my parents **Sileman** and **Aashabi** for guiding me as a person.

First I would like thank to my supervisor **Nasim Sabouri** for giving me this opportunity and for designing an exciting PhD project. You are a fantastic teacher and have inspired me in many ways. I am so grateful for your support and endless patience throughout these years. I appreciate all the advice you have given me and your encouragement for dealing with scientific problems that has helped me to build my future.

Huge thanks to my co-supervisor **Erik Johansson**

My initial lab career started with **Rosario Garcia Gill** and **Karina Persson,** I appreciate all your advice and support, and I am very thankful for both of you.

Marcus Wallgren, you are my one of the best buddies I have had in Umeå. Whenever you're in the lab my workload becomes half because you are my walking notebook. I have missed your company, but I congratulate you on your studies. **Parham Pourbozorgi Langroudi**, we started our PhD career together and I have enjoyed your company. **Jan Jamroskovic,** I have really liked your company, especially during conferences. **Kok-Phen Yan,** thank you for the conversations and scientific support. I hope you have a wonderful life. **Ikenna Obi,** you do an exceptional job keeping the lab up and running. **Marco Deiana,** you are a very nice person. I am very thankful to the current and past members of the **Sabouri** lab.

Thanks to all the members of **Sjoerd Wanrooij** and **Paulina Wanrooij's** groups for their valuable conversations and advice during lab meetings.

Thanks to my officemates, **Kok-Phen, Jan Jamroskovic, Göran Bylund, Selma Dahmane, Parham Pourbozorgi Langroudi,** and **Marco Deiana** for wonderful conversations.

Thanks to **Ingrid Råberg, Jenny Fossen, Clas Wikström, Anna Sjöström**, and **Matilda Djärf** for helping me in administrative problems.

Thanks to my co-assistant teaching pals **Mahsa, Phong, Khalil, Venki, Farah, Parham, Josephin**, **Tohidul, Lars,** and **Thomas**.

I really appreciate my friends here in Umeå, **Kumar Appari, Venki-Bindu, Karthik-Gowthami, Pramod-Neetha, Lalitha-Surya, Abhay-Srilatha, Madhavi-Mohan, Syam-Reshma, Ravi-Vidhya, Harsha, Chitanya, Yougandhar, Ravishaw, Valentina, Soumya Bharth, Hareesha, Soumya, Jagan, Iftikar, Vivek, Rajesh, Sushma, Vimal, Praveen,** and **Sonali**. I am very thankful for your making my life in Umeå memorable and full of joy.

A big hug to all the players in Badminton Group.

Finally, I am very blessed to have my family, my partner **Afroz Banu**, my world **Sameeha Suhana** and **Farheen**, and my sister **Amerun-Ibrahim**, my in-laws **Sirajuddin-Arifa** Without your support and patience, my life would be incomplete.

I am very thankful and grateful for your love and affection. This thesis would not have possible without your support.

References

1. Watson, J.D. and Crick, F.H.C. (1953) Molecular Structure of Nucleic Acids - a Structure for Deoxyribose Nucleic Acid. *Nature*, **171**, 737-738.
2. Bell, S.P. and Labib, K. (2016) Chromosome Duplication in Saccharomyces cerevisiae. *Genetics*, **203**, 1027-1067.
3. Deegan, T.D. and Diffley, J.F. (2016) MCM: one ring to rule them all. *Curr Opin Struct Biol*, **37**, 145-151.
4. Ticau, S., Friedman, L.J., Ivica, N.A., Gelles, J. and Bell, S.P. (2015) Single-molecule studies of origin licensing reveal mechanisms ensuring bidirectional helicase loading. *Cell*, **161**, 513-525.
5. Takayama, Y., Kamimura, Y., Okawa, M., Muramatsu, S., Sugino, A. and Araki, H. (2003) GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. *Genes Dev*, **17**, 1153-1165.
6. Muramatsu, S., Hirai, K., Tak, Y.S., Kamimura, Y. and Araki, H. (2010) CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol (epsilon}, and GINS in budding yeast. *Genes Dev*, **24**, 602-612.
7. Sengupta, S., van Deursen, F., de Piccoli, G. and Labib, K. (2013) Dpb2 integrates the leading-strand DNA polymerase into the eukaryotic replisome. *Curr Biol*, **23**, 543-552.
8. Pursell, Z.F., Isoz, I., Lundstrom, E.B., Johansson, E. and Kunkel, T.A. (2007) Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science*, **317**, 127-130.
9. Jin, Y.H., Obert, R., Burgers, P.M., Kunkel, T.A., Resnick, M.A. and Gordenin, D.A. (2001) The 3'-->5' exonuclease of DNA polymerase delta can substitute for the 5' flap endonuclease Rad27/Fen1 in processing Okazaki fragments and preventing genome instability. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 5122-5127.
10. Pascal, J.M., O'Brien, P.J., Tomkinson, A.E. and Ellenberger, T. (2004) Human DNA ligase I completely encircles and partially unwinds nicked DNA. *Nature*, **432**, 473-478.
11. Sanchez, J.A., Kim, S.M. and Huberman, J.A. (1998) Ribosomal DNA replication in the fission yeast, Schizosaccharomyces pombe. *Exp Cell Res*, **238**, 220-230.
12. Dalgaard, J.Z. and Klar, A.J. (2000) swi1 and swi3 perform imprinting, pausing, and termination of DNA replication in S. pombe. *Cell*, **102**, 745-751.
13. Krings, G. and Bastia, D. (2004) swi1- and swi3-dependent and independent replication fork arrest at the ribosomal DNA of Schizosaccharomyces pombe. *Proc Natl Acad Sci U S A*, **101**, 14085-14090.
14. Azvolinsky, A., Giresi, P.G., Lieb, J.D. and Zakian, V.A. (2009) Highly transcribed RNA polymerase II genes are impediments to replication fork progression in Saccharomyces cerevisiae. *Molecular cell*, **34**, 722-734.
15. Deshpande, A.M. and Newlon, C.S. (1996) DNA replication fork pause sites dependent on transcription. *Science*, **272**, 1030-1033.
16. McDonald, K.R., Guise, A.J., Pourbozorgi-Langroudi, P., Cristea, I.M., Zakian, V.A., Capra, J.A. and Sabouri, N. (2016) Pfh1 Is an Accessory Replicative Helicase that Interacts with the Replisome to Facilitate Fork Progression and Preserve Genome Integrity. *PLoS genetics*, **12**.
17. Cea, V., Cipolla, L. and Sabbioneda, S. (2015) Replication of Structured DNA and its implication in epigenetic stability. *Front Genet*, **6**, 209.
18. Abdelmonem, M., Durwald, H. and Hoffmannberling, H. (1976) Enzymic Unwinding of DNA .2. Chain Separation by an Atp-Dependent DNA Unwinding Enzyme. *Eur J Biochem*, **65**, 441-449.
19. Singleton, M.R., Dillingham, M.S. and Wigley, D.B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem*, **76**, 23-50.
20. Patel, S.S. and Donmez, I. (2006) Mechanisms of helicases. *The Journal of biological chemistry*, **281**, 18265-18268.
21. Dillingham, M.S. (2011) Superfamily I helicases as modular components of DNA-processing machines. *Biochem Soc Trans*, **39**, 413-423.

22. Bernstein, K.A., Gangloff, S. and Rothstein, R. (2010) The RecQ DNA helicases in DNA repair. *Annu Rev Genet*, **44**, 393-417.
23. Wu, Y., Suhasini, A.N. and Brosh, R.M., Jr. (2009) Welcome the family of FANCJ-like helicases to the block of genome stability maintenance proteins. *Cell Mol Life Sci*, **66**, 1209-1222.
24. Brosh, R.M., Jr. (2013) DNA helicases involved in DNA repair and their roles in cancer. *Nat Rev Cancer*, **13**, 542-558.
25. Wu, Y. (2012) Unwinding and rewinding: double faces of helicase? *J Nucleic Acids*, **2012**, 140601.
26. Masuda-Sasa, T., Polaczek, P. and Campbell, J.L. (2006) Single strand annealing and ATP-independent strand exchange activities of yeast and human DNA2: possible role in Okazaki fragment maturation. *J Biol Chem*, **281**, 38555-38564.
27. Opresko, P.L., Otterlei, M., Graakjaer, J., Bruheim, P., Dawut, L., Kolvraa, S., May, A., Seidman, M.M. and Bohr, V.A. (2004) The Werner syndrome helicase and exonuclease cooperate to resolve telomeric D loops in a manner regulated by TRF1 and TRF2. *Mol Cell*, **14**, 763-774.
28. Opresko, P.L., Mason, P.A., Podell, E.R., Lei, M., Hickson, I.D., Cech, T.R. and Bohr, V.A. (2005) POT1 stimulates RecQ helicases WRN and BLM to unwind telomeric DNA substrates. *J Biol Chem*, **280**, 32069-32080.
29. Ghosh, A.K., Rossi, M.L., Singh, D.K., Dunn, C., Ramamoorthy, M., Croteau, D.L., Liu, Y. and Bohr, V.A. (2012) RECQL4, the protein mutated in Rothmund-Thomson syndrome, functions in telomere maintenance. *J Biol Chem*, **287**, 196-209.
30. Garcia, P.L., Liu, Y., Jiricny, J., West, S.C. and Janscak, P. (2004) Human RECQ5beta, a protein with DNA helicase and strand-annealing activities in a single polypeptide. *The EMBO journal*, **23**, 2882-2891.
31. Li, D., Liu, H., Jiao, L., Chang, D.Z., Beinart, G., Wolff, R.A., Evans, D.B., Hassan, M.M. and Abbruzzese, J.L. (2006) Significant effect of homologous recombination DNA repair gene polymorphisms on pancreatic cancer survival. *Cancer Res*, **66**, 3323-3330.
32. Li, D., Frazier, M., Evans, D.B., Hess, K.R., Crane, C.H., Jiao, L. and Abbruzzese, J.L. (2006) Single nucleotide polymorphisms of RecQ1, RAD54L, and ATM genes are associated with reduced survival of pancreatic cancer. *J Clin Oncol*, **24**, 1720-1728.
33. Chen, C.F. and Brill, S.J. (2010) An essential DNA strand-exchange activity is conserved in the divergent N-termini of BLM orthologs. *The EMBO journal*, **29**, 1713-1725.
34. Machwe, A., Xiao, L., Groden, J., Matson, S.W. and Orren, D.K. (2005) RecQ family members combine strand pairing and unwinding activities to catalyze strand exchange. *The Journal of biological chemistry*, **280**, 23397-23407.
35. Muftuoglu, M., Kulikowicz, T., Beck, G., Lee, J.W., Piotrowski, J. and Bohr, V.A. (2008) Intrinsic ssDNA annealing activity in the C-terminal region of WRN. *Biochemistry-Us*, **47**, 10247-10254.
36. Gorbalenya, A.E. and Koonin, E.V. (1993) Helicases - Amino-Acid-Sequence Comparisons and Structure-Function-Relationships. *Curr Opin Struc Biol*, **3**, 419-429.
37. Fairman-Williams, M.E., Guenther, U.P. and Jankowsky, E. (2010) SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol*, **20**, 313-324.
38. Singleton, M.R., Dillingham, M.S. and Wigley, D.B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem*, **76**, 23-50.
39. Enernark, E.J. and Joshua-Tor, L. (2008) On helicases and other motor proteins. *Curr Opin Struc Biol*, **18**, 243-257.
40. Lohman, T.M., Tomko, E.J. and Wu, C.G. (2008) Non-hexameric DNA helicases and translocases: mechanisms and regulation. *Nat Rev Mol Cell Bio*, **9**, 391-401.
41. Fairman-Williams, M.E., Guenther, U.P. and Jankowsky, E. (2010) SF1 and SF2 helicases: family matters. *Curr Opin Struc Biol*, **20**, 313-324.

42. Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P. and Blinov, V.M. (1989) 2 Related Superfamilies of Putative Helicases Involved in Replication, Recombination, Repair and Expression of DNA and Rna Genomes. *Nucleic Acids Res*, **17**, 4713-4730.
43. Velankar, S.S., Soultanas, P., Dillingham, M.S., Subramanya, H.S. and Wigley, D.B. (1999) Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell*, **97**, 75-84.
44. Saikrishnan, K., Powell, B., Cook, N.J., Webb, M.R. and Wigley, D.B. (2009) Mechanistic Basis of 5 '-3 ' Translocation in SF1B Helicases. *Cell*, **137**, 849-859.
45. Hall, M.C. and Matson, S.W. (1999) Helicase motifs: the engine that powers DNA unwinding. *Mol Microbiol*, **34**, 867-877.
46. Story, R.M. and Steitz, T.A. (1992) Structure of the Reca Protein-Adp Complex. *Nature*, **355**, 374-376.
47. Pause, A. and Sonenberg, N. (1992) Mutational Analysis of a Dead Box Rna Helicase - the Mammalian Translation Initiation-Factor Eif-4a. *Embo J*, **11**, 2643-2654.
48. Korolev, S., Hsieh, J., Gauss, G.H., Lohman, T.M. and Waksman, G. (1997) Major domain swiveling revealed by the crystal structures of complexes of E-coli Rep helicase bound to single-stranded DNA and ADP. *Cell*, **90**, 635-647.
49. GravesWoodward, K.L. and Weller, S.K. (1996) Replacement of Gly(815) in helicase motif V alters the single-stranded DNA-dependent ATPase activity of the herpes simplex virus type 1 helicase-primase. *J Biol Chem*, **271**, 13629-13635.
50. Lohman, T.M. and Bjornson, K.P. (1996) Mechanisms of helicase-catalyzed DNA unwinding. *Annu Rev Biochem*, **65**, 169-214.
51. Singleton, M.R., Scaife, S. and Wigley, D.B. (2001) Structural analysis of DNA replication fork reversal by RecG. *Cell*, **107**, 79-89.
52. Iyer, L.M., Aravind, L. and Koonin, E.V. (2001) Common origin of four diverse families of large eukaryotic DNA viruses. *J Virol*, **75**, 11720-11734.
53. Bochman, M.L. (2014) Roles of DNA helicases in the maintenance of genome integrity. *Mol Cell Oncol*, **1**, e963429.
54. Brandon, M., Baldi, P. and Wallace, D.C. (2006) Mitochondrial mutations in cancer. *Oncogene*, **25**, 4647-4662.
55. Ohta, S. (2003) A multi-functional organelle mitochondrion is involved in cell death, proliferation and disease. *Curr Med Chem*, **10**, 2485-2494.
56. Maizels, N. (2015) G4-associated human diseases. *EMBO Rep*, **16**, 910-922.
57. Sanders, C.M. (2010) Human Pif1 helicase is a G-quadruplex DNA-binding protein with G-quadruplex DNA-unwinding activity. *The Biochemical journal*, **430**, 119-128.
58. Chisholm, K.M., Aubert, S.D., Freese, K.P., Zakian, V.A., King, M.C. and Welcsh, P.L. (2012) A genomewide screen for suppressors of Alu-mediated rearrangements reveals a role for PIF1. *Plos One*, **7**, e30748.
59. Chakraverty, R.K. and Hickson, I.D. (1999) Defending genome integrity during DNA replication: a proposed role for RecQ family helicases. *Bioessays*, **21**, 286-294.
60. Nakayama, H., Nakayama, K., Nakayama, R., Irino, N., Nakayama, Y. and Hanawalt, P.C. (1984) Isolation and Genetic-Characterization of a Thymineless Death-Resistant Mutant of Escherichia-Coli-K12 - Identification of a New Mutation (Recq1) That Blocks the Recf Recombination Pathway. *Mol Gen Genet*, **195**, 474-480.
61. Shahrabani-Gargir, L., Shomrat, R., Yaron, Y., Orr-Urtreger, A., Groden, J. and Legum, C. (1998) High frequency of a common Bloom syndrome Ashkenazi mutation among Jews of Polish origin. *Genet Test*, **2**, 293-296.
62. Fukuchi, K., Martin, G.M. and Monnat, R.J. (1989) Mutator Phenotype of Werner Syndrome Is Characterized by Extensive Deletions. *P Natl Acad Sci USA*, **86**, 5893-5897.
63. Sharma, S., Doherty, K.M. and Brosh, R.M., Jr. (2006) Mechanisms of RecQ helicases in pathways of DNA metabolism and maintenance of genomic stability. *Biochem J*, **398**, 319-337.

64. Bochman, M.L., Sabouri, N. and Zakian, V.A. (2010) Unwinding the functions of the Pif1 family helicases. *DNA repair*, **9**, 237-249.
65. Foury, F. and Kolodynski, J. (1983) pif mutation blocks recombination between mitochondrial rho+ and rho- genomes having tandemly arrayed repeat units in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America*, **80**, 5345-5349.
66. Lahaye, A., Stahl, H., Thines-Sempoux, D. and Foury, F. (1991) PIF1: a DNA helicase in yeast mitochondria. *Embo J*, **10**, 997-1007.
67. Bochman, M.L., Judge, C.P. and Zakian, V.A. (2011) The Pif1 family in prokaryotes: what are our helicases doing in your bacteria? *Mol Biol Cell*, **22**, 1955-1959.
68. Azvolinsky, A., Dunaway, S., Torres, J.Z., Bessler, J.B. and Zakian, V.A. (2006) The S. cerevisiae Rrm3p DNA helicase moves with the replication fork and affects replication of all yeast chromosomes. *Genes Dev*, **20**, 3104-3116.
69. Paeschke, K., Capra, J.A. and Zakian, V.A. (2011) DNA replication through G-quadruplex motifs is promoted by the Saccharomyces cerevisiae Pif1 DNA helicase. *Cell*, **145**, 678-691.
70. Ivessa, A.S., Lenzmeier, B.A., Bessler, J.B., Goudsouzian, L.K., Schnakenberg, S.L. and Zakian, V.A. (2003) The Saccharomyces cerevisiae helicase Rrm3p facilitates replication past nonhistone protein-DNA complexes. *Molecular cell*, **12**, 1525-1536.
71. Ivessa, A.S., Zhou, J.Q., Schulz, V.P., Monson, E.K. and Zakian, V.A. (2002) Saccharomyces Rrm3p, a 5' to 3' DNA helicase that promotes replication fork progression through telomeric and subtelomeric DNA. *Genes & development*, **16**, 1383-1396.
72. Ivessa, A.S., Zhou, J.Q. and Zakian, V.A. (2000) The Saccharomyces Pif1p DNA helicase and the highly related Rrm3p have opposite effects on replication fork progression in ribosomal DNA. *Cell*, **100**, 479-489.
73. Pike, J.E., Burgers, P.M., Campbell, J.L. and Bambara, R.A. (2009) Pif1 helicase lengthens some Okazaki fragment flaps necessitating Dna2 nuclease/helicase action in the two-nuclease processing pathway. *The Journal of biological chemistry*, **284**, 25170-25180.
74. Rossi, M.L., Pike, J.E., Wang, W., Burgers, P.M., Campbell, J.L. and Bambara, R.A. (2008) Pif1 helicase directs eukaryotic Okazaki fragments toward the two-nuclease cleavage pathway for primer removal. *The Journal of biological chemistry*, **283**, 27483-27493.
75. Saini, N., Ramakrishnan, S., Elango, R., Ayyar, S., Zhang, Y., Deem, A., Ira, G., Haber, J.E., Lobachev, K.S. and Malkova, A. (2013) Migrating bubble during break-induced replication drives conservative DNA synthesis. *Nature*, **502**, 389-392.
76. Wilson, M.A., Kwon, Y., Xu, Y., Chung, W.H., Chi, P., Niu, H., Mayle, R., Chen, X., Malkova, A., Sung, P. *et al.* (2013) Pif1 helicase and Poldelta promote recombination-coupled DNA synthesis via bubble migration. *Nature*, **502**, 393-396.
77. Paeschke, K., Bochman, M.L., Garcia, P.D., Cejka, P., Friedman, K.L., Kowalczykowski, S.C. and Zakian, V.A. (2013) Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature*, **497**, 458-462.
78. O'Rourke, T.W., Doudican, N.A., Zhang, H., Eaton, J.S., Doetsch, P.W. and Shadel, G.S. (2005) Differential involvement of the related DNA helicases Pif1p and Rrm3p in mtDNA point mutagenesis and stability. *Gene*, **354**, 86-92.
79. Cheng, X., Dunaway, S. and Ivessa, A.S. (2007) The role of Pif1p, a DNA helicase in Saccharomyces cerevisiae, in maintaining mitochondrial DNA. *Mitochondrion*, **7**, 211-222.
80. Cheng, X., Qin, Y. and Ivessa, A.S. (2009) Loss of mitochondrial DNA under genotoxic stress conditions in the absence of the yeast DNA helicase Pif1p occurs independently of the DNA helicase Rrm3p. *Mol Genet Genomics*, **281**, 635-645.
81. Chib, S., Byrd, A.K. and Raney, K.D. (2016) Yeast Helicase Pif1 Unwinds RNA:DNA Hybrids with Higher Processivity than DNA:DNA Duplexes. *The Journal of biological chemistry*, **291**, 5889-5901.
82. Boule, J.B. and Zakian, V.A. (2007) The yeast Pif1p DNA helicase preferentially unwinds RNA DNA substrates. *Nucleic acids research*, **35**, 5809-5818.

83. Zhou, J., Monson, E.K., Teng, S.C., Schulz, V.P. and Zakian, V.A. (2000) Pif1p helicase, a catalytic inhibitor of telomerase in yeast. *Science*, **289**, 771-774.
84. Boule, J.B., Vega, L.R. and Zakian, V.A. (2005) The yeast Pif1p helicase removes telomerase from telomeric DNA. *Nature*, **438**, 57-61.
85. Ramanagoudr-Bhojappa, R., Byrd, A.K., Dahl, C. and Raney, K.D. (2014) Yeast Pif1 accelerates annealing of complementary DNA strands. *Biochemistry-Us*, **53**, 7659-7669.
86. Koc, K.N., Singh, S.P., Stodola, J.L., Burgers, P.M. and Galletto, R. (2016) Pif1 removes a Rap1-dependent barrier to the strand displacement activity of DNA polymerase delta. *Nucleic acids research*, **44**, 3811-3819.
87. Hoffman, C.S., Wood, V. and Fantes, P.A. (2015) An Ancient Yeast for Young Geneticists: A Primer on the Schizosaccharomyces pombe Model System. *Genetics*, **201**, 403-423.
88. Pinter, S.F., Aubert, S.D. and Zakian, V.A. (2008) The Schizosaccharomyces pombe Pfh1p DNA helicase is essential for the maintenance of nuclear and mitochondrial DNA. *Molecular and cellular biology*, **28**, 6594-6608.
89. Sabouri, N., Capra, J.A. and Zakian, V.A. (2014) The essential Schizosaccharomyces pombe Pfh1 DNA helicase promotes fork movement past G-quadruplex motifs to prevent DNA damage. *BMC Biol*, **12**, 101.
90. Sabouri, N., McDonald, K.R., Webb, C.J., Cristea, I.M. and Zakian, V.A. (2012) DNA replication through hard-to-replicate sites, including both highly transcribed RNA Pol II and Pol III genes, requires the S. pombe Pfh1 helicase. *Genes & development*, **26**, 581-593.
91. Steinacher, R., Osman, F., Dalgaard, J.Z., Lorenz, A. and Whitby, M.C. (2012) The DNA helicase Pfh1 promotes fork merging at replication termination sites to ensure genome stability. *Genes & development*, **26**, 594-602.
92. McDonald, K.R., Sabouri, N., Webb, C.J. and Zakian, V.A. (2014) The Pif1 family helicase Pfh1 facilitates telomere replication and has an RPA-dependent role during telomere lengthening. *DNA repair*, **24**, 80-86.
93. Ryu, G.H., Tanaka, H., Kim, D.H., Kim, J.H., Bae, S.H., Kwon, Y.N., Rhee, J.S., MacNeill, S.A. and Seo, Y.S. (2004) Genetic and biochemical analyses of Pfh1 DNA helicase function in fission yeast. *Nucleic acids research*, **32**, 4205-4216.
94. Mateyak, M.K. and Zakian, V.A. (2006) Human PIF helicase is cell cycle regulated and associates with telomerase. *Cell Cycle*, **5**, 2796-2804.
95. Zhou, J.Q., Monson, E.K., Teng, S.C., Schulz, V.P. and Zakian, V.A. (2000) Pif1p helicase, a catalytic inhibitor of telomerase in yeast. *Science*, **289**, 771-774.
96. Zhang, D.H., Zhou, B., Huang, Y., Xu, L.X. and Zhou, J.Q. (2006) The human Pif1 helicase, a potential Escherichia coli RecD homologue, inhibits telomerase activity. *Nucleic Acids Res*, **34**, 1393-1404.
97. Futami, K., Shimamoto, A. and Furuichi, Y. (2007) Mitochondrial and nuclear localization of human Pif1 helicase. *Biological & pharmaceutical bulletin*, **30**, 1685-1692.
98. Gagou, M.E., Ganesh, A., Thompson, R., Phear, G., Sanders, C. and Meuth, M. (2011) Suppression of apoptosis by PIF1 helicase in human tumor cells. *Cancer Res*, **71**, 4998-5008.
99. Gagou, M.E., Ganesh, A., Phear, G., Robinson, D., Petermann, E., Cox, A. and Meuth, M. (2014) Human PIF1 helicase supports DNA replication and cell growth under oncogenic-stress. *Oncotarget*, **5**, 11381-11398.
100. Anand, R.P., Shah, K.A., Niu, H., Sung, P., Mirkin, S.M. and Freudenreich, C.H. (2012) Overcoming natural replication barriers: differential helicase requirements. *Nucleic Acids Res*, **40**, 1091-1105.
101. George, T., Wen, Q., Griffiths, R., Ganesh, A., Meuth, M. and Sanders, C.M. (2009) Human Pif1 helicase unwinds synthetic DNA structures resembling stalled DNA replication forks. *Nucleic acids research*, **37**, 6491-6502.
102. Gu, Y., Masuda, Y. and Kamiya, K. (2008) Biochemical analysis of human PIF1 helicase and functions of its N-terminal domain. *Nucleic acids research*, **36**, 6295-6308.

103. Chen, W.F., Dai, Y.X., Duan, X.L., Liu, N.N., Shi, W., Li, N., Li, M., Dou, S.X., Dong, Y.H., Rety, S. *et al.* (2016) Crystal structures of the BsPif1 helicase reveal that a major movement of the 2B SH3 domain is required for DNA unwinding. *Nucleic acids research*, **44**, 2949-2961.
104. Zhou, X., Ren, W., Bharath, S.R., Tang, X., He, Y., Chen, C., Liu, Z., Li, D. and Song, H. (2016) Structural and Functional Insights into the Unwinding Mechanism of Bacteroides sp Pif1. *Cell reports*, **14**, 2030-2039.
105. Lu, K.Y., Chen, W.F., Rety, S., Liu, N.N., Wu, W.Q., Dai, Y.X., Li, D., Ma, H.Y., Dou, S.X. and Xi, X.G. (2018) Insights into the structural and mechanistic basis of multifunctional S-cerevisiae Pif1p helicase. *Nucleic acids research*, **46**, 1486-1500.
106. He, X., Byrd, A.K., Yun, M.K., Pemble, C.W.t., Harrison, D., Yeruva, L., Dahl, C., Kreuzer, K.N., Raney, K.D. and White, S.W. (2012) The T4 phage SF1B helicase Dda is structurally optimized to perform DNA strand separation. *Structure*, **20**, 1189-1200.
107. Gellert, M., Lipsett, M.N. and Davies, D.R. (1962) Helix formation by guanylic acid. *Proc Natl Acad Sci U S A*, **48**, 2013-2018.
108. Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K. and Neidle, S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res*, **34**, 5402-5415.
109. Lin, W., Sampathi, S., Dai, H., Liu, C., Zhou, M., Hu, J., Huang, Q., Campbell, J., Shin-Ya, K., Zheng, L. *et al.* (2013) Mammalian DNA2 helicase/nuclease cleaves G-quadruplex DNA and is required for telomere integrity. *The EMBO journal*, **32**, 1425-1439.
110. Bochman, M.L., Paeschke, K. and Zakian, V.A. (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet*, **13**, 770-780.
111. Zhang, J.Y., Zheng, K.W., Xiao, S., Hao, Y.H. and Tan, Z. (2014) Mechanism and Manipulation of DNA:RNA Hybrid G-Quadruplex Formation in Transcription of G-Rich DNA. *J Am Chem Soc*, **136**, 1381-1390.
112. Bugaut, A. and Balasubramanian, S. (2012) 5 '-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res*, **40**, 4727-4741.
113. Zheng, K.W., Wu, R.Y., He, Y.D., Xiao, S., Zhang, J.Y., Liu, J.Q., Hao, Y.H. and Tan, Z. (2014) A competitive formation of DNA: RNA hybrid G-quadruplex is responsible to the mitochondrial transcription termination at the DNA replication priming site. *Nucleic Acids Res*, **42**, 10832-10844.
114. Millevoi, S., Moine, H. and Vagner, S. (2012) G-quadruplexes in RNA biology. *Wires Rna*, **3**, 495-507.
115. Agarwala, P., Pandey, S. and Maiti, S. (2015) The tale of RNA G-quadruplex. *Org Biomol Chem*, **13**, 5570-5585.
116. Chambers, V.S., Marsico, G., Boutell, J.M., Di Antonio, M., Smith, G.P. and Balasubramanian, S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nature biotechnology*, **33**, 877-881.
117. Capra, J.A., Paeschke, K., Singh, M. and Zakian, V.A. (2010) G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae. *PLoS Comput Biol*, **6**, e1000861.
118. Bharti, S.K., Sommers, J.A., Zhou, J., Kaplan, D.L., Spelbrink, J.N., Mergny, J.L. and Brosh, R.M., Jr. (2014) DNA sequences proximal to human mitochondrial DNA deletion breakpoints prevalent in human disease form G-quadruplexes, a class of DNA structures inefficiently unwound by the mitochondrial replicative Twinkle helicase. *J Biol Chem*, **289**, 29975-29993.
119. Damas, J., Carneiro, J., Goncalves, J., Stewart, J.B., Samuels, D.C., Amorim, A. and Pereira, F. (2012) Mitochondrial DNA deletions are associated with non-B DNA conformations. *Nucleic Acids Res*, **40**, 7606-7621.
120. Murat, P. and Balasubramanian, S. (2014) Existence and consequences of G-quadruplex structures in DNA. *Curr Opin Genet Dev*, **25**, 22-29.
121. Lipps, H.J. and Rhodes, D. (2009) G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol*, **19**, 414-422.

122. Bochman, M.L., Paeschke, K. and Zakian, V.A. (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet*, **13**, 770-780.
123. Murat, P. and Balasubramanian, S. (2014) Existence and consequences of G-quadruplex structures in DNA. *Curr Opin Genet Dev*, **25**, 22-29.
124. Huppert, J.L. (2010) Structure, location and interactions of G-quadruplexes. *Febs J*, **277**, 3452-3458.
125. Beaudoin, J.D. and Perreault, J.P. (2013) Exploring mRNA 3'-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening. *Nucleic Acids Res*, **41**, 5898-5911.
126. Xu, Y., Suzuki, Y., Ito, K. and Komiyama, M. (2010) Telomeric repeat-containing RNA structure in living cells. *Proc Natl Acad Sci U S A*, **107**, 14579-14584.
127. Biffi, G., Di Antonio, M., Tannahill, D. and Balasubramanian, S. (2014) Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat Chem*, **6**, 75-80.
128. Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H.J. and Pluckthun, A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc Natl Acad Sci U S A*, **98**, 8572-8577.
129. Liu, H.Y., Zhao, Q., Zhang, T.P., Wu, Y., Xiong, Y.X., Wang, S.K., Ge, Y.L., He, J.H., Lv, P., Ou, T.M. *et al.* (2016) Conformation Selective Antibody Enables Genome Profiling and Leads to Discovery of Parallel G-Quadruplex in Human Telomeres. *Cell Chem Biol*, **23**, 1261-1270.
130. Biffi, G., Tannahill, D., McCafferty, J. and Balasubramanian, S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem*, **5**, 182-186.
131. Miyoshi, D., Karimata, H. and Sugimoto, N. (2006) Hydration regulates thermodynamics of G-quadruplex formation under molecular crowding conditions. *J Am Chem Soc*, **128**, 7957-7963.
132. Brazda, V., Haronikova, L., Liao, J.C. and Fojta, M. (2014) DNA and RNA quadruplex-binding proteins. *Int J Mol Sci*, **15**, 17493-17517.
133. Baran, N., Pucshansky, L., Marco, Y., Benjamin, S. and Manor, H. (1997) The SV40 large T-antigen helicase can unwind four stranded DNA structures linked by G-quartets. *Nucleic Acids Res*, **25**, 297-303.
134. Harrington, C., Lan, Y. and Akman, S.A. (1997) The identification and characterization of a G4-DNA resolvase activity. *Journal of Biological Chemistry*, **272**, 24631-24636.
135. Sun, H., Karow, J.K., Hickson, I.D. and Maizels, N. (1998) The Bloom's syndrome helicase unwinds G4 DNA. *Journal of Biological Chemistry*, **273**, 27587-27592.
136. Sun, H., Bennett, R.J. and Maizels, N. (1999) The Saccharomyces cerevisiae Sgs1 helicase efficiently unwinds G-G paired DNAs. *Nucleic Acids Res*, **27**, 1978-1984.
137. Plyler, J., Jasheway, K., Tuesuwan, B., Karr, J., Brennan, J.S., Kerwin, S.M. and David, W.M. (2009) Real-time investigation of SV40 large T-antigen helicase activity using surface plasmon resonance. *Cell Biochem Biophys*, **53**, 43-52.
138. Budhathoki, J.B., Ray, S., Urban, V., Janscak, P., Yodh, J.G. and Balci, H. (2014) RecQ-core of BLM unfolds telomeric G-quadruplex in the absence of ATP. *Nucleic Acids Res*, **42**, 11528-11545.
139. Budhathoki, J.B., Stafford, E.J., Yodh, J.G. and Balci, H. (2015) ATP-dependent G-quadruplex unfolding by Bloom helicase exhibits low processivity. *Nucleic Acids Res*, **43**, 5961-5970.
140. Henderson, A., Wu, Y., Huang, Y.C., Chavez, E.A., Platt, J., Johnson, F.B., Brosh, R.M., Jr., Sen, D. and Lansdorp, P.M. (2014) Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res*, **42**, 860-869.
141. Rodriguez, R., Miller, K.M., Forment, J.V., Bradshaw, C.R., Nikan, M., Britton, S., Oelschlaegel, T., Xhemalce, B., Balasubramanian, S. and Jackson, S.P. (2012) Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat Chem Biol*, **8**, 301-310.
142. Randazzo, A., Spada, G.P. and da Silva, M.W. (2013) Circular dichroism of quadruplex structures. *Top Curr Chem*, **330**, 67-86.

143. Vorlickova, M., Kejnovska, I., Sagi, J., Renciuk, D., Bednarova, K., Motlova, J. and Kypr, J. (2012) Circular dichroism and guanine quadruplexes. *Methods*, **57**, 64-75.
144. Karsisiotis, A.I., Hessari, N.M., Novellino, E., Spada, G.P., Randazzo, A. and Webba da Silva, M. (2011) Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. *Angew Chem Int Ed Engl*, **50**, 10645-10648.
145. Masiero, S., Trotta, R., Pieraccini, S., De Tito, S., Perone, R., Randazzo, A. and Spada, G.P. (2010) A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures. *Org Biomol Chem*, **8**, 2683-2692.
146. Paramasivan, S., Rujan, I. and Bolton, P.H. (2007) Circular dichroism of quadruplex DNAs: applications to structure, cation effects and ligand binding. *Methods*, **43**, 324-331.
147. Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R.A. and Yang, D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res*, **34**, 2723-2735.
148. He, Y., Neumann, R.D. and Panyutin, I.G. (2004) Intramolecular quadruplex conformation of human telomeric DNA assessed with 125I-radioprobing. *Nucleic Acids Res*, **32**, 5359-5367.
149. Phan, A.T., Luu, K.N. and Patel, D.J. (2006) Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K+ solution. *Nucleic Acids Res*, **34**, 5715-5719.
150. Lane, A.N. (2012) The stability of intramolecular DNA G-quadruplexes compared with other macromolecules. *Biochimie*, **94**, 277-286.
151. Williamson, J.R. (1994) G-quartet structures in telomeric DNA. *Annu Rev Biophys Biomol Struct*, **23**, 703-730.
152. Renaud de la Faverie, A., Guedin, A., Bedrat, A., Yatsunyk, L.A. and Mergny, J.L. (2014) Thioflavin T as a fluorescence light-up probe for G4 formation. *Nucleic Acids Res*, **42**, e65.
153. Kang, H.Y., Choi, E., Bae, S.H., Lee, K.H., Gim, B.S., Kim, H.D., Park, C., MacNeill, S.A. and Seo, Y.S. (2000) Genetic analyses of Schizosaccharomyces pombe dna2(+) reveal that dna2 plays an essential role in Okazaki fragment metabolism. *Genetics*, **155**, 1055-1067.
154. Bae, S.H., Kim, D.W., Kim, J., Kim, J.H., Kim, D.H., Kim, H.D., Kang, H.Y. and Seo, Y.S. (2002) Coupling of DNA helicase and endonuclease activities of yeast Dna2 facilitates Okazaki fragment processing. *The Journal of biological chemistry*, **277**, 26632-26641.
155. Galletto, R. and Tomko, E.J. (2013) Translocation of Saccharomyces cerevisiae Pif1 helicase monomers on single-stranded DNA. *Nucleic Acids Res*, **41**, 4613-4627.
156. Ramanagoudr-Bhojappa, R., Chib, S., Byrd, A.K., Aarattuthodiyil, S., Pandey, M., Patel, S.S. and Raney, K.D. (2013) Yeast Pif1 helicase exhibits a one-base-pair stepping mechanism for unwinding duplex DNA. *The Journal of biological chemistry*, **288**, 16185-16195.
157. Geronimo, C.L., Singh, S.P., Galletto, R. and Zakian, V.A. (2018) The signature motif of the Saccharomyces cerevisiae Pif1 DNA helicase is essential in vivo for mitochondrial and nuclear functions and in vitro for ATPase activity. *Nucleic acids research*, **46**, 8357-8370.
158. Mohammad, J.B., Wallgren, M. and Sabouri, N. (2018) The Pif1 signature motif of Pfh1 is necessary for both protein displacement and helicase unwinding activities, but is dispensable for strand-annealing activity. *Nucleic Acids Res*, **46**, 8516-8531.
159. Tanaka, H., Ryu, G.H., Seo, Y.S., Tanaka, K., Okayama, H., MacNeill, S.A. and Yuasa, Y. (2002) The fission yeast pfh1(+) gene encodes an essential 5' to 3' DNA helicase required for the completion of S-phase. *Nucleic acids research*, **30**, 4728-4739.
160. Zhou, J.Q., Qi, H., Schulz, V.P., Mateyak, M.K., Monson, E.K. and Zakian, V.A. (2002) Schizosaccharomyces pombe pfh1+ encodes an essential 5' to 3' DNA helicase that is a member of the PIF1 subfamily of DNA helicases. *Mol Biol Cell*, **13**, 2180-2191.
161. Smith, J.S., Chen, Q., Yatsunyk, L.A., Nicoludis, J.M., Garcia, M.S., Kranaster, R., Balasubramanian, S., Monchaud, D., Teulade-Fichou, M.P., Abramowitz, L. *et al.* (2011) Rudimentary G-quadruplex-based telomere capping in Saccharomyces cerevisiae. *Nat Struct Mol Biol*, **18**, 478-485.

162. Zahler, A.M., Williamson, J.R., Cech, T.R. and Prescott, D.M. (1991) Inhibition of telomerase by G-quartet DNA structures. *Nature*, **350**, 718-720.
163. Mendoza, O., Bourdoncle, A., Boule, J.B., Brosh, R.M., Jr. and Mergny, J.L. (2016) G-quadruplexes and helicases. *Nucleic Acids Res*, **44**, 1989-2006.
164. Muzzolini, L., Beuron, F., Patwardhan, A., Popuri, V., Cui, S., Niccolini, B., Rappas, M., Freemont, P.S. and Vindigni, A. (2007) Different quaternary structures of human RECQ1 are associated with its dual enzymatic activity. *PLoS Biol*, **5**, e20.