Medical School

Clinical Medicine, School of Medicine and Population Health

University of Sheffield

The PIF1 DNA helicase in genome maintenance: from molecular mechanism to specific inhibitors as potential drugs

Francesca Romana Scommegna

A thesis submitted to the University of Sheffield for the degree of Doctor of Philosophy

December 2023

ACKNOWLEDGEMENT

I would like to thank my supervisor Cyril Sanders for his help with this project and his precious teaching. I also want to thank all the people in the project, especially Dr Francesca Pisani for being my second supervisor and Dr Silvia Onesti (Elettra Sincrotrone, Trieste) for hosting me in her lab for the secondment. I also want to thank Dr Didier Roche, Dr Vincent Rodeschini and especially Dr Mark Wever at the Edelris Keymical (Lyon, France) for the strong collaboration and work together.

ABSTRACT

Helicase enzymes use the energy from nucleotide triphosphate (NTP) hydrolysis [1] to resolve DNA secondary structures generating single-stranded DNA [2]. PIF1 is a 5'–3' helicase belonging to SF1B (Super Family 1B) helicase family, first identified in yeast and then in bacterial and human cells. Human PIF1 (hPIF1) plays roles in genome maintenance; *in vitro*, it has double-stranded DNA (dsDNA) and G quadruplex (G4) DNA unwinding activity and G4 DNA specific binding activity [2]. It recognizes and processes synthetic stalled DNA replication fork-like substrates reversing the replication strand to restart replication [3]. In intact cells, hPIF1 is required for replication fork stability; some tumour cells depend upon hPIF1 and siRNA-mediated knockdown causes their apoptosis, suggesting potential therapeutic applications of PIF1.

In this work, hPIF1 and unimolecular G4 DNA interaction have been investigated through an internally controlled 'pull-down' DNA binding assay. Several unimolecular G4 DNA substrates with different structures and conformations were tested; changes in apparent G4 DNA binding affinity during the ATP hydrolysis cycle were investigated by reproducing ATP hydrolysis steps with nucleotide cofactor analogues.

In G4 DNA unwinding assay hydrolyzable ATP is not essential while $ADP{\cdot}MgF_4^-$ is most efficient at stimulating unwinding. The data suggest that G4 DNA is recognised in the apo state, ATP hydrolysis leads to structural rearrangements and an increase in ssDNA binding affinity, presumably linked to translocation and G4 DNA unwinding. To identify hPIF1 residues involved in G4 DNA interaction, a model based on *To*PIF1- G4 DNA crystal structure was tested by biochemical assays. However preliminary data suggested that the binding model may not be universal.

A second branch of this work is focused on the discovery of hPIF1 inhibitors. A collaboration with Edelris Keymical (Lyon, France) followed three different approaches for novel small molecule inhibitors identification, resulting in new compounds able to inhibit hPIF1 helicase activity.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ALT	Alternative lengthening of telomere
Alt-EJ	Alternative end-joining
ATM	Ataxia-telangiectasia mutated kinase
ATR	ATM and Rad3-related protein
ATRIP	ATR-interactive
BER	Base excision repair
BLM	Bloom syndrome helicase
BP	Butylparaben
BPA	Bisphenol A
BSA	Bovine Serum Albumin
*Bs*Pif1	Bacteroids Pif1
CD	Circular Dichroism
CDK2	Cyclin dependent-kynase
CHK1	Checkpoint 1
CHK2	Checkpoint 2
DDR	DNA damage repair
DMS	Dimethyl sulphate
DMSO	Dimethyl Sulphoxide
DNA-PKcs	Phosphoinosidite 3 kinase(PI3K)-related protein kinase
DR	Direct repeats
DSB	Double strand breaks
dsDNA	Double strand DNA
DTT	dithiothreitol
EDTA	Ethylene diamine tetra acetic-acid
EtOH	Ethanol
GST	Gluthation S transferase
GRS	G4-recognizing surface
HD	Helicase domain

HEPES	4-(2-Hydroxyethyl)piperazine-1-ethsnesulfonic acid
His	Histidine
hPIF1	Human PIF1
HR	Homologous recombination
HRDC	Helicase and RNase C-terminal
IC_{50}	Inhibition concentration at 50%
IPTG	Isopropyl β-D-1 thiogalactopyranoside
IR	Ionizing radiation
kDa	kilo Dalton
LB	Lysogeny broth
MMR	Mismatch repair
MW	Molecular weight
NA	Nucleic acid
NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
nM	Nano molar
NMR	Nuclear magnetic resonance
NTP	Nucleotide triphosphate
OBD	Origin binding domain
OH·	Hydroxyl radicals
ORF	Open reading frame
pADRr	PARP1 ribosylation
PAGE	Polyacrylamide gel elctrophoresis
PARPi	PARP inhibitor
PBS	Phosphate buffrered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PMSF	Phenyl methyl sulphonyl fluoride
PNK	T4 polynucleotide kinase
PSFF	Pif1-family specific sequebce
RFB	Replication fork barrier

RMS	RHAU-specific motif
RPA	Replication protein A
*Sc*Pif1	*Saccharomyces cervisiae* Pif1
SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromatography
SF	Super-family
SL	Synthetic lethality
SSA	Single-strand annealing
STR	Short tandem repeats
SUMO	Small ubiquitin-like modified
TAE	Tris/Acetate/EDTA
TBE	Tris/Borate/EDTA
TCEP	Tris (2-carboxyethyl)phospine
TEMED	N,N,N',N'-tetramethylethane-1,2-diamine
*To*PIF1	Termus oshimai PIF1
WH	Winged helix
UV	Ulta violet
ZB	Zymo Broth

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Helicases

Helicases are nucleotide triphosphate (NTP)-dependent enzymes that translocate on nucleic acids (NA) and generate single-stranded NA from NA secondary structures. They were discovered in 1976 when the term “helicases” was applied to an ATP-dependent double-stranded DNA (dsDNA) unwinding activity. The first helicases were isolated from bacteria, then bacteriophages and plants [2].

The Falaschi laboratory at the International Centre for Genetic Engineering and Biotechnology in Trieste (Italy) isolated the first human helicase. In HeLa cells, they found four different molecular species but were able to characterise only one, named DNA helicase I, by measuring its ability to displace a radioactively labelled oligodeoxynucleotide annealed to bacteriophage M13 DNA [4]. A subsequent analysis of helicase enzymes defined an NTP-dependent DNA translocase activity required for unwinding DNA secondary structures and displacing bound proteins. Helicases are involved in many cellular events, if not all, that involve nucleic acids such as transcription, replication, recombination and DNA repair [2].

1.1.1 The classification of helicases

Six helicase superfamilies 1-6 (SF1-6) have been defined (Figure 1), with different properties and mechanisms of action. SF1&2 are more numerous and diverse, while SF3-6 are less well characterised. One of the structural similarities among all classes is the “core NTP*ase* domain” which includes conserved residues involved in NTP binding and hydrolysis, and an arginine finger that is involved in NTP hydrolysis and energy coupling to mechanical action (Figure 1). Other classifications are possible from a mechanistic point of view, discriminating these enzymes based on their specificity for DNA, RNA or DNA-RNA hybrids and polarity of translocation, 3'–5' or 5'–3' [5].

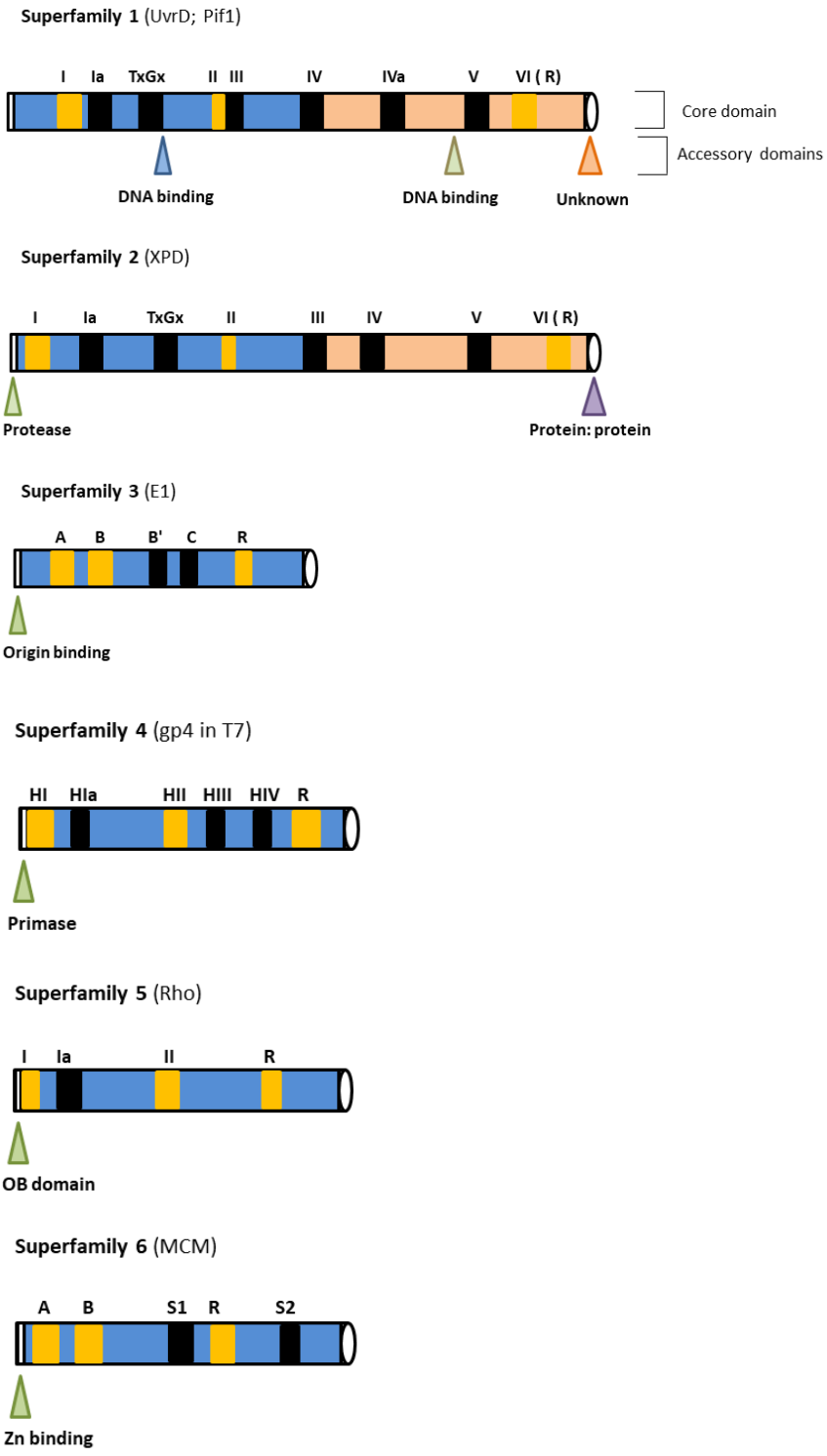


Figure 1. Helicase classification. Each superfamily is represented by one of its members (in brackets). In the core domain, the universal structural elements involved in NTP hydrolysis are represented in yellow; the accessory domains are represented with the arrows below the main structure; their position and functions change also in the same superfamily. In the core domain, the main motifs are indicated at the top of the structure; domain 6 contains the arginine finger indicated with R. (Figure obtained with Biorender and based on [6]).

Highly conserved amino acid sequences named “helicase motifs” are used in the classification. The motifs are usually clustered in a region of 200-700 aa, separated by low and highly conserved sequences, and they are usually involved in NTP binding and hydrolysis, coupling energy to structural events, or binding and unwinding of nucleic acids.

The *Q motif* is conserved among SF1 helicases that coordinate an adenine base and plays a role in orienting ATP for hydrolysis; *motif I* (also known as Walker A) is present in many nucleotide-binding proteins, forming a phosphate-binding loop, and is involved in binding Mg^{2+} and NTP; *motif II* (or Walker B) is involved in Mg^{2+} coordination and NTP hydrolysis; *motif III* is involved in DNA binding but its proximity to *motif II* suggests its involvement in transducing the energy of NTP hydrolysis to DNA movement; *motif IV* and *motif V* are respectively involved in the hydrolysis of the NTP and interaction with the nucleic acid phosphate backbone, *motif VI* is part of the NTP-binding cleft and is involved in coupling the helicase unwinding and NTP*ase* activities of the protein [6].

1.1.1.1 Superfamilies 1 and 2

The best-characterised superfamily is SF1. In this class two sub-families can be identified according to the polarity of ssDNA translocation: A and B. Examples of SF1A enzymes are the Rep and UvrD helicases in prokaryotes that translocate 3'–5'. Examples of SF1B helicase are RecD and Dda in bacteria and PIF1 in bacteria and eukaryotes that translocate 5'–3'. Mechanistic data indicates that the SF1 enzymes operate as monomers. However, SF1A Rep helicase undergoes a dimerization when binding ss or dsDNA. Also, *Sc*Pif1 binding to ssDNA induces dimerization [7]. From a structural point of view, SF1A&B enzymes have a core domain based on the RecA-fold [8]. Some have additional N- or C-terminal domains that interface with additional proteins [5] (Figure 2). Mechanistically speaking, differences between SF1A and SF1B are observed because of their opposite polarity but they seem to bind ssDNA in the same orientation [6]. In both classes, RecA-like domains 1A and 2A form a cleft where the ATP binds. In SF1A, domain 1A moves toward domain 2A after ATP binding translocating so in the 3'–5' direction; while in SF1B when ATP binds in the cleft,

domain 2A moves towards domain 1A providing 5'–3' translocation ([6], Figure 3). In all SF1 helicases, Rec-A like domains 1A and 2A are followed by accessory domains like a pin or a wedge whose role is to separate base pairs. The location of the pin changes according to the polarity of the helicase.

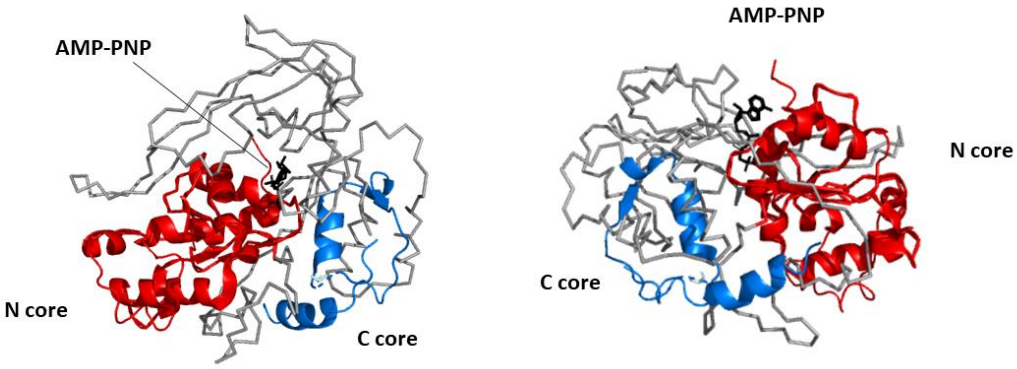


Figure 2. Core structure of an SF1A helicase. Structural representation of bacterial Pif1 core domain (pdb: 8bnx); the core is formed by tandem RecA-like folds. In the structure the N- (red) and the C-terminal (blue) RecA-like domains are represented with a NTP analogue (AMP·PNP) binding at the interface. In the N core domain are located motif 1 and 2 (Walker A motif: IFFTGSAGTGKS and Walker B motif: LVIDEIS) while motif 6 (YVALSR), with the arginine finger, is located in the C core domain. The structure is shown with two views.

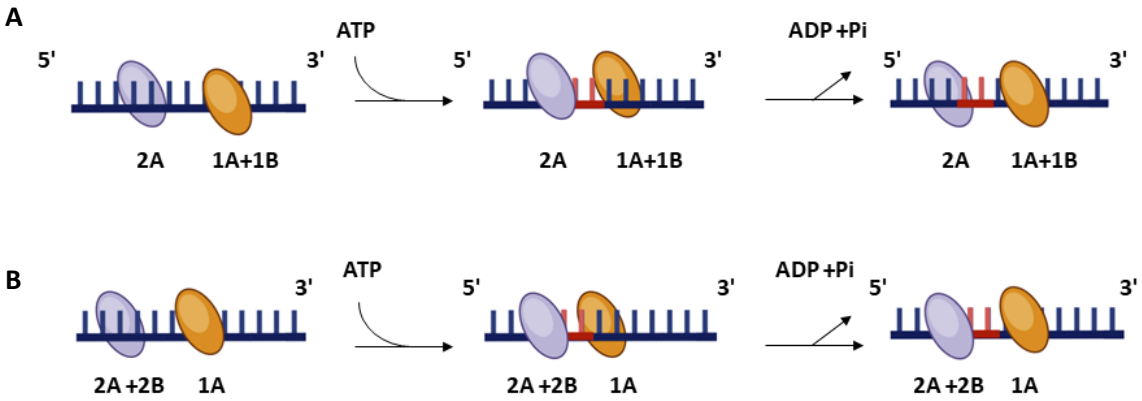


Figure 3. Translocation mechanism of SF1 helicases. (A) SF1A helicases translocate 3´—5´. Before ATP hydrolysis, the DNA is bound to domains 1A and 2A of the enzyme. When the ATP binds there is a closure of the cleft between the two domains that causes the DNA to slide across domains 1A/1B. After ATP hydrolysis the ssDNA is tightly bound to domain 1A and weakly to 2A and therefore the DNA is released. The domains change position, moving apart and pulling the DNA along the ssDNA binding channel. (B) SF1B translocation has 3´—5´ polarity. After ATP binding, the cleft closes between the 1A and 2A domains with a consequent sliding of domains 2A and 2B along the DNA backbone. The 1A domain still binds DNA tightly, anchoring the DNA while domains 2A and 2B slides along it. Once the conformational change has happened, 1A domain releases the clamp on the DNA backbone and the ATP hydrolysis starts allowing a complete relaxing of the cleft in an opened conformation. A consequence of this change of conformation is the sliding of the NA along the surface of domain 1A, moving away from 2A and 2B. Each round of ATP binding and hydrolysis causes the translocation of one base in 5´—3´ direction. (Figure obtained with Biorender and based on [7]).

SF2 enzymes are the largest helicase family. They can translocate on single or double-stranded NA, while most of the class, excluding the XPD family, display a 3'–5' polarity.

According to their mechanistic characteristics, it is possible to divide the SF1 and SF2 helicases into different categories: (i) Those able to unwind duplex NA substrates with a preferred polarity such as RecQ, Ski2-like and all the SF2 helicases; (ii) Those that locally destabilize NA duplexes without any specific direction of unwinding such as the DEAD-box family helicases and (iii) Those that can translocate on duplex substrates without strand displacements, such as RecG, Rig-I and Swi/Snf2 [9].

1.1.1.2 Superfamilies 3-6

SF3-6 helicases are hexameric and many are involved in the unwinding of dsDNA at the replication fork where up to hundreds of megabases can be unwound. Topological linking of the helicases to DNA by its threading through a central channel in the hexamer increases the enzyme processivity. Some hexameric helicases can self-assemble around DNA without additional cofactors [10] while closed-ring helicases require the help of ATP-dependent loading partners to access the template (*e.g.* bacterial DnaB which requires the clamp loader DnaC [11]). Loading cofactors are essential to avoid unregulated helicase loading and deleterious consequences for the cell *e.g.* DNA over-replication, leading to genome instability.

SF3-6 helicases have either a RecA (prokaryotic proteins) and translocate 5'–3', or an AAA^+-like (ATPase Associated with various cellular Activities, eukaryotic proteins) and translocate 3'–5', ATP*ase* core. Nucleotide-binding sites are located at the interface between monomers, where each subunit contributes different motifs for ATP binding and hydrolysis. From a structural perspective, the more complete example of a hexameric helicase is the SF3 helicase E1 from papillomavirus shown bound to the replication fork [12]. In E1, the AAA^+ motor pulls the replication fork junction against the helicase collar domain and one of the DNA strands (3') through the enzyme complex. At the entrance to the helicase domain collar the replication fork dsDNA is separated, with one strand remaining outside the helicase. Of the six N-terminal dsDNA-origin binding domains (OBD), two are in a fixed position, with one tracking and stabilising the path of the dsDNA to the replication fork unwinding point. A second OBD stabilises the bath of the 5' ssDNA strand that is excluded from the complex while passing across the top of the helicase collar domain [12].

Another important example of the behaviour of replicative hexameric helicases comes from the analysis of the SF6 helicase MCM, essential for replication initiation and elongation in eukaryotes. The MCM hexameric ring has a large central channel to accommodate dsDNA during initiation and ssDNA during processive DNA unwinding (elongation phase) [13]. Recently, the mechanism of coupled yeast MCM loading by ORC (origin replication complex) has been identified. Cryo-EM studies confirmed that two rings are loaded separately at the two inverted DNA sites in *ori*.

During this, the first MCM hexamer is loaded at a specific site with dsDNA within the central channel which promotes the engagement of ORC at the N-terminus of MCM. Afterwards, ORC can recruit the second MCM hexamer in the opposite orientation forming the double hexamer [14].

1.1.2 Mechanistic aspects of helicase function: active vs passive mechanism

Essential for the unwinding activity of helicases is the presence of a pin or wedge that splits the dsDNA during translocation, and whose location varies according to the helicase directionality. For SF1A helicases (3'–5' directionality) the pin is in domain 2A, at the junction of the duplex and ssDNA (to be ready for the strands separation) whereas in SF1B helicases (5'–3' directionality) it is in the 1B domain. Essential in the pin is the role of some residues in unwinding the duplex [15, 16].

There are different ways to describe the unwinding process and different 'units' to measure it: kinetic step size indicates the base pairs unwound prior to a rate-limiting kinetic step; physical step size, indicates the number of base pairs unwound simultaneously; chemical steps, refers to the number of base pairs unwound per ATP hydrolysed. On the other hand, translocation can be calculated as the time required for a helicase to reach the end of the specified length of ssDNA [17].

A way to classify a helicases' action is based on their ability to separate the duplex in an 'active' or 'passive' way [18]. In the passive model of unwinding, the helicase entraps ssDNA arising from the transient thermal unwinding of the upstream double-stranded nucleic acid (dsNA) end to progress fork unwinding. In contrast, in an active mechanism, the interaction of the enzyme with the dsNA acts to destabilise it, promoting the unwinding activity. As enzymes, helicases, act by lowering the activation energy barrier of the reaction they catalyze. It is possible to discriminate between active and passive mechanisms according to the magnitude of the Activation Energy (Ea) indicated here as 'B', which indicates the barrier that the enzyme must overcome to proceed with base separation. For an active enzyme, the velocity of unwinding (v_{un}) is close to the velocity of DNA translocation (v_{trans}), whilst for a passive enzyme $v_{un} < v_{trans}$ (Figure 4). When T is the temperature and with k_B

Boltzman's constant, then if B < T k_B the helicase is active and the barrier is negligible, otherwise if B > T k_B the helicase would be passive, and the unwinding reaction would be the rate-limiting step.

In practice, many parameters need to be considered to understand the behaviour of B, including a key role played by GC/AT content and the energy of base-pair formation. According to the theory above, studies have been made to discriminate between active and passive helicases. The RecQ helicase of prokaryotes (belonging to SF2) has been classified as an active helicase, while bacteriophage T4 gp41 helicase is considered a passive one [18]. However, most helicase unwinding has both active and passive components.

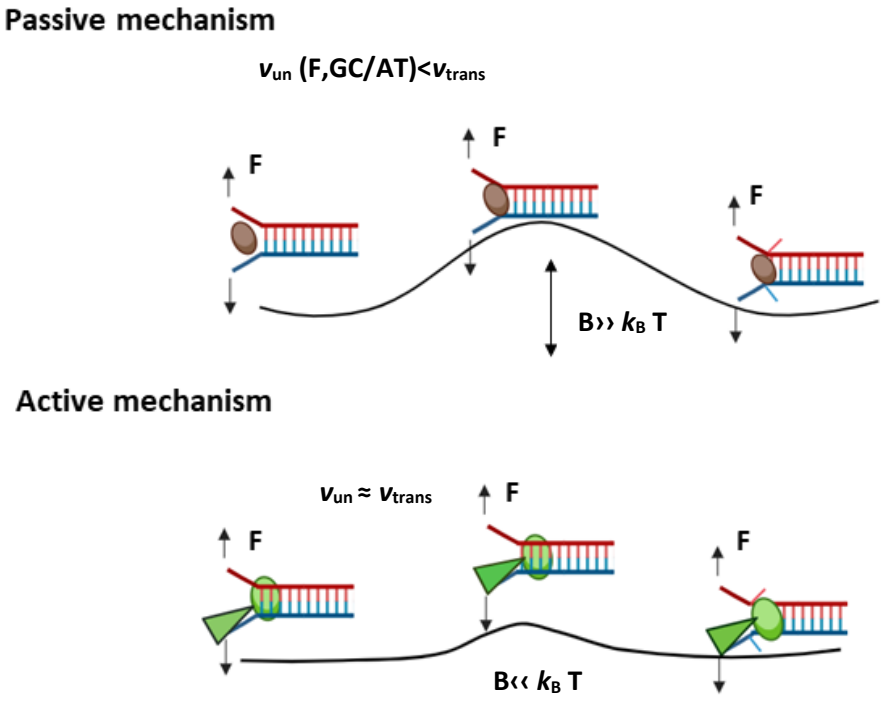


Figure 4. Passive and active mechanisms. In the passive mechanism, the enzyme (brown ball) must overcome a barrier B to proceed with the unwinding. In this case v_{un}<v_{trans} (v_{un}- helicase rate of unwinding of dsNA; v_{trans}- helicase rate of translocation on ssNA). Other parameters influencing the unwinding rate are the NA composition and the tension (F) at the fork junction. The active mechanism can destabilize the dsNA with the use of ATP (green triangle) and reduce the kinetic barrier to fork melting (B < k_BT). The velocity of unwinding is almost the same of the translocation v_{un}~ v_{trans}. Figure obtained with Biorender and based on [19].

1.2 DNA replication stress

DNA replication is a highly regulated, semi-conservative, process through which cells generate new genomic DNA material in the S phase of the cell cycle [19]. Despite such regulation, it is possible that exogenous or endogenous factors (see section 1.2.1.1) interfere with the replication mechanism causing replication stress and genome instability of both normal and cancer cells [20]. Replication stress is also considered to be a hallmark of cancer [21]. Several repair mechanisms exist that can be activated when the stress interferes with replication forks.

1.2.1 DNA damage, DNA repair pathway and genome stability

1.2.1.1 DNA damage

Cells are constantly exposed to different DNA-damaging agents that can be either endogenous or exogenous in origin. Damage includes oxidized bases, modified sugars, abasic sites, DNA strand breaks, DNA-DNA and DNA-protein crosslinks and thymine dimers. These events can cause mutagenesis, affect the replication fork, bring about the formation of non-B form DNA secondary structures or disrupt DNA methylation patterns, altering the correct function of the DNA. Helicases are central players in pathways that repair DNA damage.

(i) Endogenous DNA damage

During DNA replication, high-fidelity DNA synthesis is determined by biochemical and structural events that ensure the correct base is incorporated in the daughter strand. Sometimes, however, base substitutions and insertions occur, and additional replication errors can happen such as strand slippage events at repetitive sequences causing insertions and deletions of nucleotides. Also, topoisomerase enzymes can be a source of DNA errors during their catalytic cycle [22] and bases can be damaged, leading to substitution errors, through spontaneous events (e.g., cytosine, adenine and 5-methyl cytosine deamination) and reactive oxygen species.

(ii) Exogenous DNA damage

Exogenous DNA damaging agents include ionizing radiation (IR), UV and chemical agents. IR can damage DNA directly or indirectly by radiolysis of water to generate highly reactive hydroxyl radicals (OH•) resulting in single-strand DNA breaks. Single-stranded breaks can be converted to highly toxic double-strand breaks during replication [23]. Ultraviolet [24] solar radiation such as UV-C (190-290 nm), and to a lesser extent UV-B (290-320nm), can cause covalent linkages between two adjacent Pyrimidines, while UV-A (320-400 nm) can cause DNA adduct formation through photooxidation reactions. Chemical DNA damaging agents include natural toxins from microorganisms normally used as a defence response (*e.g*. aflatoxin from *Aspergillus sp.)* [25] and man-made chemicals in everyday products, such as butylparaben (BP) and bisphenol A (BPA) found in pharmaceuticals, cosmetic and food [22].

Extreme heat and cold or hypoxia can be considered as an environmental source of stress and have been identified as the cause of mutagenesis in human cells at trinucleotide repeats leading to neurodegenerative disorders.

Fortunately, eukaryotic cells have repair mechanisms and pathways able to fix the damaged DNA, avoiding permanent damage to the cells.

1.2.1.2 DNA damage repair pathways (DDR)

The actions of helicases are critical for genome stability [26]. Most helicases involved in DNA repair belong to SF1 and 2 (section 1.1.1) and have distinct functions within a pathway. Four major DNA repair pathways (Figure 5) have been identified in mammalian cells, nucleotide excision repair (NER), base excision repair (BER), mismatch repair (MMR) and double-stranded break (DSB) repair, sub-classified into homologous recombination (HR) and non-homologous end-joining (NHEJ). Other minor and alternative repair pathways have also been identified, such as the alternative end-joining (alt-EJ) and single-strand annealing (SSA) pathways.

When DNA damage is identified, the cell cycle is blocked, accurate repair is provided for, and abnormal cell apoptosis is promoted [27]. The DDR is characterised

by a highly regulated signalling pathway, mediated by proteins such as ATM (ataxia-telangiectasia mutated kinase), ATR (ATM and Rad3-related protein) and DNA-PKcs (phosphoinositide 3 kinase (PI3K)-related protein kinase). ATM is activated by DSBs and RPA (replication protein A); CHK1 and CHK2 (checkpoint kinase 1 and 2) are protein kinases that when targeted by ATM can inhibit the activity of CDK (cyclin dependent-kinase) through mechanisms supported by p53. When CDK activity is inhibited the cell cycle checkpoints are inhibited or arrested [28]. ATR is a DNA replication stress response kinase activated by genotoxic stress; it is recruited to DNA breakage sites through interactions with ATR-interacting protein (ATRIP). Once the breakage has been localised, a series of other factors are required to activate ATR and subsequently to activate CHK1 which arrests the cell cycle. DNA-PKcs is recruited by DSBs and involved in NHEJ steps. It suppresses DSB-induced and spontaneous HR directing the repair process towards NHEJ. Lastly, DNA-PKcs can contribute to DDR signalling in ATM-deficient conditions [27].

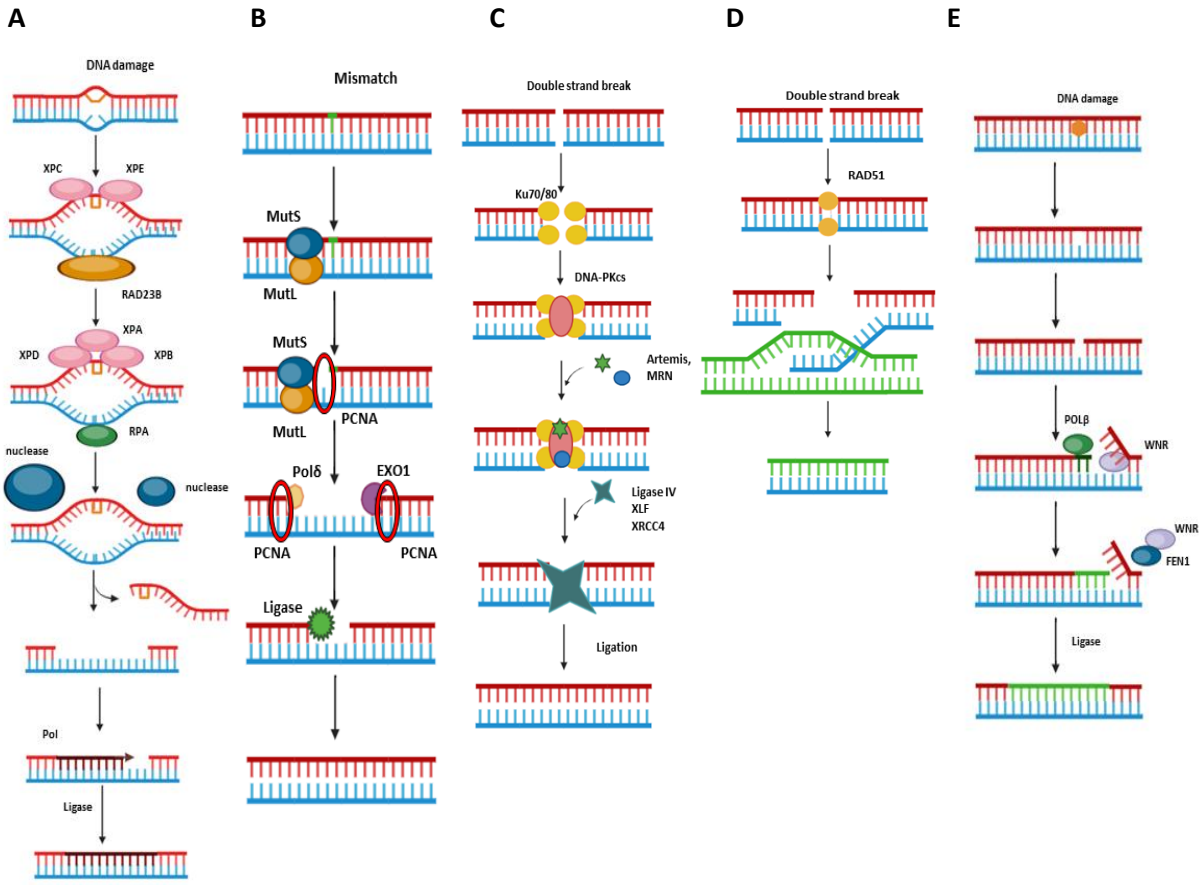


Figure 5. Schematic representation of the main DNA repair pathways and main players. (A) NER (Nucleotide Excision Repair) pathway (B) MMR (Mismatch Repair) pathway (C) NHEJ (Non-homologous DNA end joining) pathway (D) HR (Homologous Recombination) and (E) BER (Base Excision Repair) pathway. Figure realised with Biorender and based on [29], [32], [33], [34].

(i) Nucleotide excision repair (NER)

NER resolves helix-distorting lesions that interfere with base pairing. The DNA surrounding the lesion is cut and the damage is replaced using the DNA replication machinery. The NER pathway is mechanistically one of the most evolutionary conserved repair pathways, consisting of three steps: (i) damage recognition, (ii) damage verification and (iii) the excision of an ssDNA fragment. The prokaryotic NER pathway consists of a trimeric complex of UvrA, UvrB and UvrC (UvrABC) with the support of the helicase UvrD (helicase II): UvrA scans the DNA for damage, UvrB verify the damage, UvrC cleaves the DNA either side of the damage while UvrD is involved in the removal of the excised fragment [29].

The eukaryotic NER pathway is more complex, involving almost 30 proteins. The damage is first recognised by the XPC complex, which recruits the general transcription factor TFIIH to the lesion site. Two of the TFIIH subunits are the helicases XPB and XPD. XPD is a 5'–3' SF2A helicase containing an iron-sulphur (FeS) cluster. During the recruitment of TFIIH by XPC, the 3'–5' helicase XPB (a homolog of UvrD) directs TFIIH loading onto ssDNA, but only its ATP*ase* activity is necessary. Subsequently, XPD leads to a further unwinding of the DNA until it finds the damaged 8site. Additional repair factors are then recruited and DNA integrity is restored. [29]. Several hereditary diseases leading to cancer predisposition correlate to defects in the NER pathways, such as xeroderma pigmentosum (XP, *XPD* mutated) and trichothiodystrophy (TTD, *XPB* or *XPD* mutated) [30, 31].

(ii) Double-stand break (DSB) repair

DSBs are highly toxic and two pathways are involved in their resolution, HR (Homologous Recombination) and NHEJ (Non-Homologous End Joining). NHEJ mediates repair by directly ligating the ends of DSBs together, causing sometimes the deletion or mutation of DNA around the DSB site (error-prone). Classical NHEJ (C-NHEJ) usually occurs during G0/G1 and G2 of the cell cycle. HR is a conserved and mechanistically complex process that acts to restore the original DNA sequence to the site of damage by using a copy on a sister chromatid. One important role of HR is to restore stalled or broken replication forks. HR typically takes place in the mid-S

and mid-G2 phases of the cell cycle when DNA replication occurs, and sister templates are available. In HR, the recombinase RAD51 plays a key role in the homology search and single-strand DNA (ssDNA) invasion. Part of the DNA around the DSB is removed and, using the homologous sister chromatid as a template, a new DNA strand is restored at the DSB site. Many mammalian helicases from several sub-families (*e.g.* ReQ, FeS cluster, SNF2/SWI2-like) are involved in HR [32].

(iii) Mis-match repair (MMR)

MMR deals with dNTP misincorporation and insertion and deletion loops that can form during DNA replication. The mismatches in the DNA are recognised as lesions that are excised and then the correct DNA is re-synthesised in its place [33]. MMR in bacteria requires MutS, MutL MutH and helicase UvrD (helicase II), which has a role in displacing the aberrant DNA stretch.

Eukaryotic MMR is very similar to prokaryotic MMR. Many proteins are involved, including some homologous of the prokaryotic players: MutS is replaced by MutSα, MutSβ and MSH4-MSH5 that recognise mismatched bases; MutL is replaced by MutLα, MutLβ and MutLγ that nicks the DNA segment to be removed, to initiate repair synthesis. Whether a helicase is used in eukaryotic MMR has been long debated. However, it has been seen that the MMR pathway interacts with helicases belonging to the RecQ family or the FANCJ helicase (see below) [29].

(iv) Base excision repair (BER)

The BER pathway removes chemically damaged bases, but no helicase is used. In BER, the damaged base is recognised, removed and replaced by new DNA [33]. BER pathway is initiated by DNA glycosylases (monofunctional- generate a hydrolytic apurinic/ apyrimidinic site or bifunctional- with an additional lysate activity) recognition of a damaged base. The lesion recognition is followed by a short-nucleotide (SN-BER) or a long-patch (LP-BER) where various proteins are coordinated until the ligation is completed [34]. Inefficient repair would lead to the accumulation of toxic lesions and apoptosis.

1.2.1.3 Stabilising DNA replication forks

One important action of helicases is to stabilize stalled replication forks, prevent replication stress and maintain genomic stability. First, helicases or helicase complexes can act as checkpoints in the replication steps. For example, at the replication initiation phase, the MCM2-7 (minichromosome maintenance) helicase complex bound to the dsDNA origin of replication is activated by CDK2 (cyclin-dependent kinase 2) to allow replisome loading on the DNA. CDKs also prevent re-replication, stopping MCM2-7 from reloading on DNA until the next cell cycle [35]. In addition, several helicases are involved in replication fork remodelling such as the 3'–5' RecQ family helicases RECQ1, BLM (Bloom syndrome helicase), and WRN (Werner syndrome helicase), which play a central role in replication fork restoration and replication restart [36, 37]. Notably, the RecQ helicases are highly expressed in cancer, to promote replication fork progression under conditions of replication stress [38, 39].

RECQ1 can unwind the leading strand of stalled replication forks (fork reversal) and promote branch migration [40]. It is regulated by PARP1 (poly (ADP-ribose) polymerases), keeping it at the stalled replication fork until replication stress is relieved. PARPs are a family of nucleoproteins involved in different DDR pathways and in particular the PARP1 family influence the NER, HR, and BER pathways [41]. The WRN helicase also has a 3'–5' exonuclease activity, whose action is specific to restoring forks containing gaps in the leading strand [42]. BLM can resolve Holliday Junctions (HJ) and related DNA structures, rescuing stalled replication forks [43]. Also, the Bloom complex can catalyses replication fork regression, important for repairing stalled replication forks at DNA crosslink [44].

The Dna2 helicase, with the help of WRN/BLM ATP*ase* activity, can resect reversed forks and promote the production of 3' ssDNA overhangs. WRN and BLM have been shown to have pro- and anti-recombinogenic activity and instead of promoting HR, they can promote the restart of the fork by resolving HJs [45].

DDX11 is a 5'–3' SF2 helicase [46] that has a key role in the replication stress response caused by the formation of non-canonical B form DNA [47]. Most importantly,

the DDX11 helicase interacts with Timeless, a protein in the fork protection complex, and is involved in protecting fork integrity under conditions of replication stress [48].

We can assume then how important is the combined action of different proteins for the stability and the remodelling of RFs to determine the survival and genetic stability of a cell. Moreover, the action of some of these factors is taken advantage of in the development of anticancer therapy strategies, especially in pathways involving replication stress or critical DNA repair proteins [49].

1.3 Non-B form DNA structures and genome stability

Although the main DNA configuration in cells is the double-stranded B DNA form, more than 20 non-canonical secondary structures have been found, such as triple-helices, hairpins, cruciforms, G-quadruplex and slipped structures, that form under specific conditions. Non-B DNA structures have both physiological and pathological functions in the cell [50]. They play a vital role in biological processes such as replication and gene expression, but there is also evidence of their role in inducing genetic instability and consequently human diseases such as neurodegenerative disorders and cancer [51, 52].

Non-canonical B-form DNA structures can be generated during DNA replication when a long single-stranded region on the lagging strand forms [53]. Condition for the formation of non-canonical DNA structure can also occur during the DDR, for example DSB repair, where short inverted repeats allow the formation of a hairpin near the breakpoint [53]. Helicases have a critical role in regulating non-B DNA formation by unwinding such structures [52].

1.3.1 Non-B form DNA structures

In vivo, genomic DNA is packaged into chromatin and the “free” DNA between the histones usually maintains a B-conformation. However, the conditions that allow the release of the DNA from the histones, separation of the DNA duplex and generation of negative supercoiling, for example in transcription and replication, facilitate the formation of non-B DNA structure. Each type of non-B DNA structure

shows a different distribution across the genome: for example, G4 (G-quadruplex) DNA (Figure 6 E) and Z-DNA are enriched at GC-rich promoter regions; DR (direct repeats), H-DNA, and MR (mirror repeats) are instead enriched in low complexity repetitive sequences, while IR (inverted repeats) and STR (short tandem repeats) are distributed between gene-rich and gene-poor regions respectively [50].

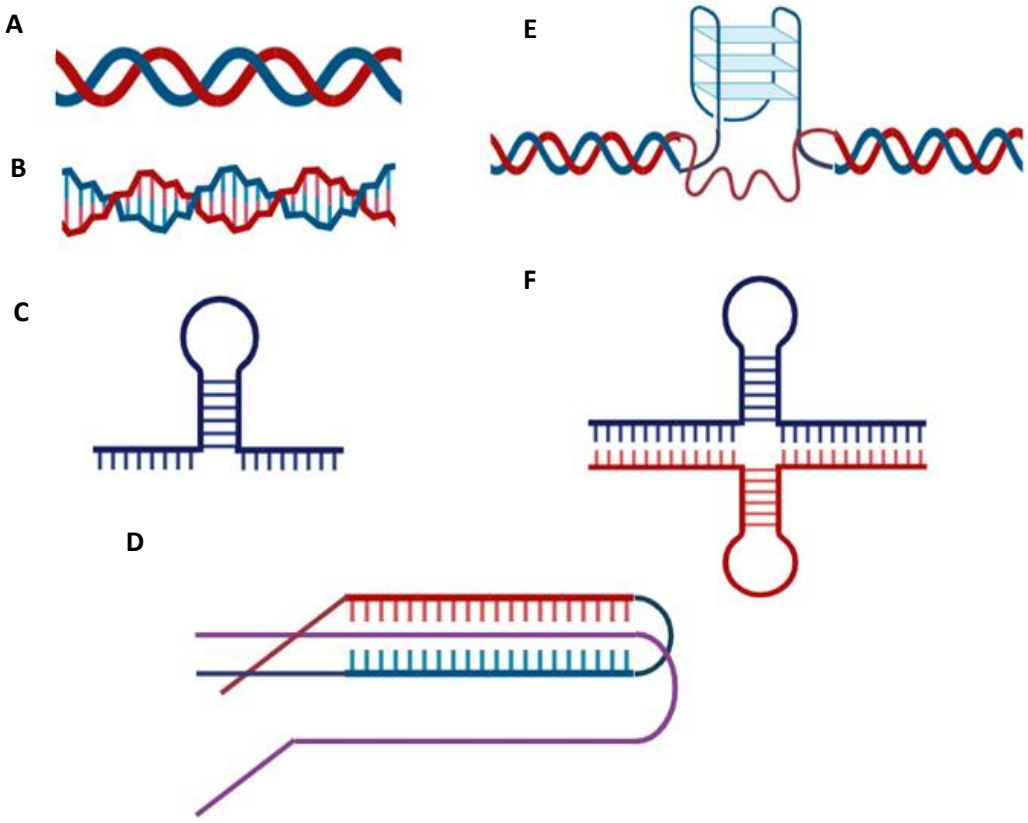


Figure 6. Cartoon representation of some non-canonical secondary structures in the human genome. (A) Canonical B-form DNA. (B) Left-handed structure caused by Z-DNA; (C) Hairpin DNA; (D) Triplex DNA structure; (E) G quadruplex structure. (F) Cruciform structure. Figure realised with Biorender and based on [50].

(i) Z-DNA

Z-DNA is so-called because of the perpendicular arrangement that forms in the sugar-phosphate backbones, giving rise to a zigzag pattern. It forms a left-handed double-helix with deep grooves that cannot be distinguished as major or minor. The negatively charged phosphate backbones are close, making the repulsion of phosphate groups more powerful and the DNA less stable than the B-form DNA.

Although considered to be a transient DNA conformation, Z-DNA may have a specific biological function in cells as it is, for example, located in the promoter regions of some oncogenes such as cMYC [51]. Z-DNA formation can result in DNA deletion and loss of control sequences in cancer-related genes leading to disease [54].

(ii) Triplex-DNA

DNA triplexes form between a Watson-Crick base paired duplex and a single DNA-stranded Hoogsteen base-paired in the duplex major groove. DNA triplexes are classified according to the nature of the third strand, *i.e.,* whether it is a separate molecule (intermolecular triplex) or part of the duplex (intramolecular triplex). Discrimination is also made between parallel or antiparallel triplexes: in a parallel triplex the third strand consists of a homo-pyrimidine sequence that binds the dsDNA in parallel to the complementary strand while in antiparallel triplexes, a purine-rich third strand binds to the dsDNA in an antiparallel configuration [55]. *In vitro*, DNA-triplexes are relatively unstable compared to duplex DNA [56].

H-DNA is a specific intra-molecular triplex, that forms when a mirror-repeat of homopurine-homopyrimidine sequences is present in the DNA. The energy provided by supercoiling allows localised unpairing of the dsDNA and one of the single strands then flips back and base pairs to form the triplex. Although generally poorly understood, this DNA conformation is abundant in eukaryotic cells and can affect gene expression by interrupting the activity of RNA polymerase. H-DNA forming sequence can stall replication fork progression, possibly causing dsDNA breaks. [57] [53].

(iii) i-motif

i-motifs consist of two parallel duplexes that interact in an antiparallel manner with each other and are held together by hemi-protonated cytosine-cytosine^{+} (C: C^{+}) base pairs. An intramolecular i-motif structure forms from four different C-tracts within the same strand, while inter-molecular i-motif forms after the association of two or four separate DNA strands, although four-stranded structures are unlikely to form *in vivo*.

In i-motifs, there are two narrow minor grooves between the antiparallel strands, and they can occur in two alternative types: face-to-face and back-to-back. The i-motif conformation and stability are influenced by the number of cytosine bases, pH, metal ions, and factors that may interact with the DNA [51].

The i-motif may form *in vivo* in cytosine-rich DNA that is particularly abundant near gene regulatory regions, especially those of oncogenes, and human telomeric DNA. They may form together with G-quadruplexes (in the opposing strand; see below), playing a complementary role in gene regulation. Some studies demonstrated a different presence of i-motifs in different stages of the cell cycle, and this suggests they play opposing roles in regulating gene expression [58]. The persistence of i-motif at telomeres leads to the uncapping and release of telomere-binding proteins and telomere dysfunction, which induces a DNA damage response, cycle arrest, senescence, and apoptosis. Dimeric i-motifs in centromeric regions play a role in the directing of centromere location and in providing architectural features of the centromere [59].

(iv) G quadruplex (G4) DNA

G-quadruplex (G4) DNA structures are characterised by cyclic hydrogen-bonded arrays of four guanine bases, linked through eight hydrogen bonds by Hoogsteen base-pairing and stabilised by monovalent cations (K^+ or Na^+) that confer high thermodynamic stability [51]. They form in single-stranded DNA or RNA at the sequence G≥3NxG≥3NxG≥3NxG≥3.

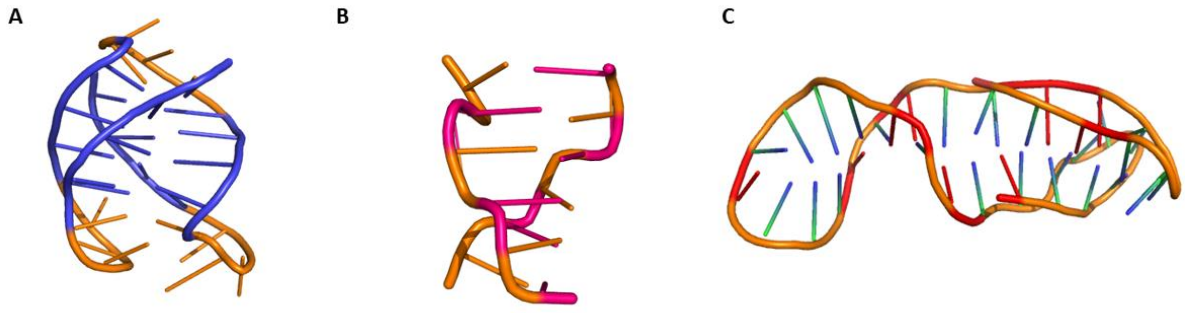


Figure 7. Molecular graphic representation of some non-canonical DNA secondary structures present in the human genome. (A) G4 DNA (pdb: 201d). dG bases in violet represent the quartets; dT bases in orange form the loop. (B) Z-DNA (pdb: 6aqt). dC in pink, dG in orange. (C) i-motif (pdb: 7o5e). dC in red are hemi-protonated and holding together the structure.

1.4 G4 DNA structure

G4 structures are polymorphic and able to adopt different conformations. G quadruplex DNA can be unimolecular or intermolecular [60] and, depending on the sequence and the length of the loops which influence the adopted structure, they are classified into four different groups: edgewise loops containing two adjacent antiparallel strands, diagonal loops connecting the two opposite antiparallel strands, double-chain-reversal loops connecting adjacent parallel strands and V-shaped loops connecting two corners of the G-tetrad core [61] [62] (Figure 8). The loops play an important role in structural stability and G4 folding [63, 64].

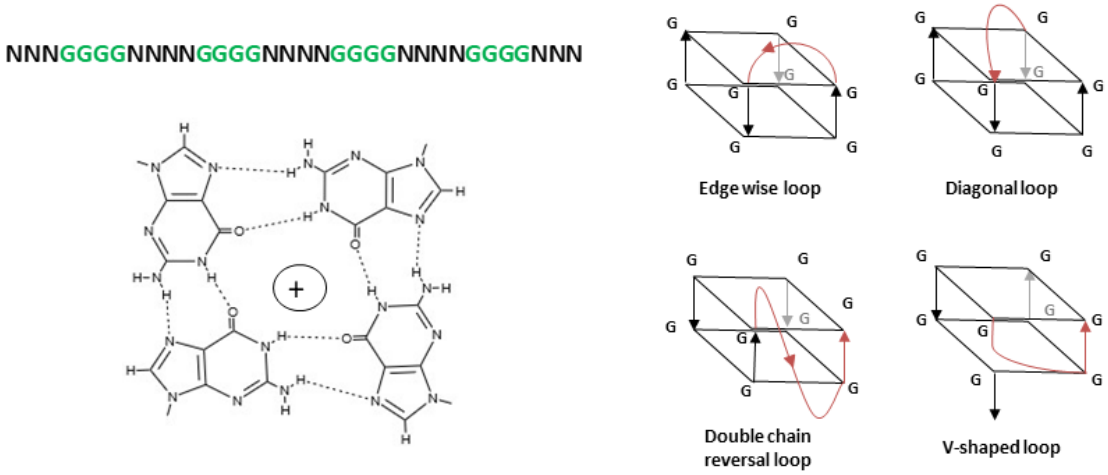


Figure 8. G-quadruplex structure and topologies. On the left, four guanines are linked through Hoogsteen base-pair and stabilised by a monovalent cation (K^+> Na^+ >Li^+) forming the G-tetrad; the consensus sequence is shown on the top of the structure, with the guanines forming the G tracks highlighted in green. On the right, different loop orientations. Figure based on [60].

Several techniques have been used to study G quadruplexes, such as electrophoresis, chromatography and mass spectroscopy, to retrieve knowledge about G4 molecular size [65] [66]; circular dichroism (CD) for information about G4 structures and possible conformations (parallel, antiparallel) and NMR spectroscopy along with X-ray crystallography to obtain atomic resolution structures of G4 DNA [67]. Specifically, the structural study of G-quadruplexes through NMR spectroscopy gives information for kinetics and dynamic studies, critical for the understanding of G4s as a therapeutic target [68].

1.4.1 Biophysics of G4 DNA structures

G-quadruplex structural polymorphism is influenced by factors such as pH, cations and other ligands. G4 structures are generally destabilised at low pH conditions (while the i-motif that could form at the same place in dsDNA is usually stable at pH 4-5 [69]). Some G4 conformations, however, can be possible at low and neutral pH as demonstrated by Yan *et al.*(2018), whose CD spectroscopy revealed a change of loop conformation for G4 DNA changing pH from 7 to 4.5 [70].

Another important factor influencing G4 structure is molecular crowding, defined as all the elements composing the intracellular environment (lipids, polysaccharides, nucleic acids) [71]. The human telomeric G4 structure is particularly sensitive to crowding elements and it is reported to be affected by the presence of PEG (polyethene glycol), used *in vitro* to reproduce the crowding *in vivo* [72] [73]. On the other hand, the *Oxytricha* telomere sequences usually assume an intermolecular anti-parallel conformation which rearranges to an intermolecular parallel under crowding conditions [74].

Finally, monovalent cations can affect variation in G4 structure. Cations are in the central channel formed by the quartet arrangement and can be positioned between the planes or along them. Na^+ and K^+ are the ions physiologically relevant, with K^+ inducing greater G4 stability compared to Na^+. A change of conformation in the glycosidic bonds of a human G4 telomeric structure has been observed in the presence of K^+ or Na^+ [75]. Divalent metal ions can also affect G4 structure and stability, but they are primarily found in RNA G4s, where they can stabilise the structure [76]. The ion in the central channel is dynamic and changing the ion concentration can change the coordination and eventually the stability of the G quadruplex. The binding's dynamicity is connected to the function of the structure [77].

1.4.2 G4 DNA folding

Understanding the folding dynamics of G4 structures is important to understand their biological functions. For example, the *Oxytricha nova* telomeric G quadruplex adopts a 'pre-folded' intermediate structure in solution (presumably a hairpin-like conformation) in the absence of cations, while their addition results in antiparallel G quadruplex formation [78, 79]. Frelih *et al*. (2020), analysing the G4 'pre-folded' state, found a series of structure's intermediates [80]; the hydrogen bond interactions mark energy barriers that need to be overcome to obtain the final fold. Among the most likely intermediate structures in G4 folding are hairpins and triplexes (Figure 9) [81]. All intermediate states, hairpins and triplexes, could lead to a G4 conformation (on-pathway) or be part of an alternative pathway leading to G4 alternatives (off-pathway) [82]. However, although folding mechanisms have been proposed, experimental evidence is hard to obtain.

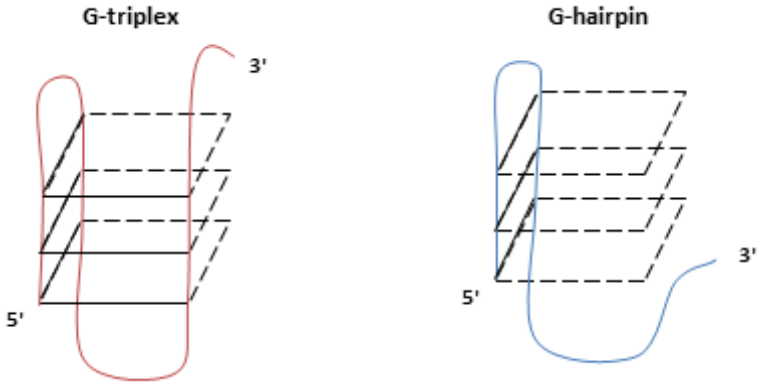


Figure 9. Proposed folding of G4 intermediates. G-triplex on the left; G-hairpin on the right. Those structures were observed through smFRET as human G4 telomeric intermediate structures. G-hairpin formation seems to be favoured in the presence of Na^+, while G-triplex intermediate seems to be happy both in the presence of Na^+ and K^+ [78].

1.4.3 G4 DNA in the human genome

In the human genome, G4 motifs are unevenly distributed: $G4^{lo}$ DNA regions are depleted of G4 DNA motifs and characteristic of most genes, while $G4^{hi}$ regions are enriched with G4 DNA motifs. This relative balance is related to gene function, where $G4^{lo}$ genes are associated with nucleic acid binding, nucleosome assembly, cell adhesion and regulation of cell division functions, while transcription factors, growth

factors and cytokines are G^{hi}-associated genes. Tumour suppressor genes are $G4^{lo}$ genes showing maximal genomic stability, while oncogenes are $G4^{hi}$ and therefore potentially prone to G4 DNA-induced instability. The regions of the genome that are energetically most favourable to the formation of G-quadruplexes are the telomeric regions and some promoters. G4 DNA structures are also implicated in the epigenetic regulation of human genes like in ATRX, a chromatin remodelling factor deficient in an X-linked human genetic disease. Patients with X fragile disease are characterised by alpha thalassemia and mental retardation; during transcription of X fragile genes, G4 RNA structures carry expansion of the trinucleotide (CGG) repeat responsible for X-Fragile disease [83].

1.4.4 G4 DNA in telomeric regions

Telomeres consist of tandem arrays of simple G-rich sequence repeats located at the ends of a linear chromosome, that protect them from the improper activation of DNA-damage-response pathways. Telomeric G-quadruplexes are resolved in the S phase of the cell cycle, allowing telomerase to access the G-overhang for telomere synthesis to take place. In vertebrates, many different proteins play a role in telomere maintenance including the RecQ family helicase WRN [84] [85] and RTEL (regulator of telomere length) that have clearly defined roles in unwinding secondary structures formed by G-rich DNA [86].

1.4.5 Human helicases and the resolution of G4 DNA

To form G4 DNA, the dsDNA needs to denature and become single-stranded. During transcription, potential G4 DNA forming sequences in the template strand are stabilised/neutralised as the co-transcriptional RNA/DNA hybrid forms. However, in the non-template strand, G4 DNA will form spontaneously while the G4-forming motif is un-paired. During replication, the helicase-dependent separation of the two strands of dsDNA can allow guanine-rich strands to fold into G-quadruplexes on both leading and lagging strands. However, specialised SF1 (PIF1) and SF2 helicases (e.g. BLM, FANCJ, WRN) act in their resolution, while the SF3-6 replicative helicases may be

blocked by G4s [87] [88]. WRN and PIF1 were reported to unfold G4 DNA on the lagging strand [89, 90] while BLM and FANCJ act on the leading strand G4s [91]. Any unresolved G4 DNA replication fork barriers that persist through mitosis can cause gaps that are converted to double strand breaks with consequent genomic instability. When helicase dependent G4 resolution pathways are deficient, there can be ramifications such as cell senescence and death or mutagenesis leading to carcinogenesis or neurological problems. This is underscored by the fact that certain human diseases are caused by the dysfunction of G4-resolving helicases *e.g.*, Bloom (BLM) and Werner (WRN) syndrome [83].

G4s are also key regulators of telomere lengthening. During alternative lengthening of telomere (ALT) G4s act as a barrier and need to be unfolded [92]. For this purpose, RECQ1 helicase plays a role, in unfolding the G quadruplex.

(i) Pif1 helicase

Importantly, among the G4 DNA processing helicases, in yeast, PIF1 is considered the most efficient in this task [93], and the human ortholog is the subject of this work and will be discussed later in detail.

(ii) Dna2 helicase-nuclease

The Dna2 protein is a ssDNA-specific endonuclease and 5'–3' helicase. Mammalian Dna2 plays a role in Okazaki fragment maturation, but some studies show Dna2 has a role in the DNA replication stress response and DNA repair by recognizing and resolving telomeric G quadruplex structures and cleaving DNA at G4 sites. Dna2 depletion causes telomere replication defects and treatment of cells with G4-stabilising ligands increases the frequency of these events [94].

(iii) The iron-sulphur (Fe-S) cluster helicases

Fe-S helicases contain an iron-sulphur cluster capable of accepting and donating electrons. The role of the Fe-S cluster in helicase action is not clear, but in *E. coli* DinG it at least has a structural role in configuring protein segments involved in DNA unwinding [95]. The human Fe-S helicase FANCJ unwinds G4 structures in a 5'–3' direction in an ATP-dependent way *in vitro*. This activity of FANCJ is reduced by the G4-DNA binding and stabilisation ligand telomestatin. In eukaryotic cells, FANCJ

unwinding of G4 DNA is fundamental for genomic stability [96]. Evidence suggests that a specific pocket in the FANCJ structure, not present in other G4 processing DNA helicases (PIF1, RTEL1 and RecQ family) recognises G4 DNA through two conserved lysine residues (K141/K142) in the G4 recognition peptide. The recognition peptide shares a similarity with the RHAU/DHX36 helicase Rhua18 G4 binding structure (see (v)).

(iv) The RecQ helicases

The RecQ helicases, found in bacteria and eukaryotes, share a 450 amino acid (aa) conserved helicase core domain. Of the five human RecQ proteins BLM, WRN and RECQ4, are linked to inherited diseases associated with cancer and premature ageing and involve genome stability. G4 DNA is a preferred target of these enzymes and a conserved C-terminal RCQ domain, required for processing replication forks, is responsible for binding G4 DNA [97] and may account for targeting specificity [58]. The RCQ has a Zn^{2+}- binding subdomain and a winged helix (WH) and it is known to be the major site of protein-DNA interactions [98]. In addition to the RCQ domain, a second DNA binding domain (HRDC) is involved in Holliday junction and ssDNA binding while also considered to be an auxiliary binding domain for G4 DNA and fork DNA structures [97]. No structural data show how RecQ helicases bind G4 DNA. However, a structure of *C. sakazakii* RecQ bound to unfolded G4 DNA shows the presence of a guanine-specific pocket that may be essential for G4 resolution, by capturing a G residue 'flipped' out from a stack [99].

(v) DHX36

RNA helicases are ubiquitous enzymes participating in RNA regulatory pathways and RNA metabolism and most belong to the SF1 and SF2. The largest families of human RNA helicases are the DEAD box (DDX, 40 proteins) and DEAH box (DHX, 15 proteins among which DHX36) [100]. G4 RNA binding and unwinding helicases will not be considered in depth here, other than DHX36 (RHAU) for the reasons outlined below.

DHX36 (also called RHAU) was originally identified as an RNA helicase [101]. However, it is an RNA and DNA helicase belonging to the DEAH box family involved in targeting (binding) and unfolding G4 RNA and DNA, specifically promoting the

translation of mRNA and resolving human telomeric G4 RNA structures [102]. DHX36 resolves G4 DNA structures in an ATP-dependent manner but there is no need for ATP in DHX36 G4 DNA binding [103]. Because of their dynamicity, obtaining a crystal structure of helicase complexes with G4 DNAs is difficult. In 2015, Phan's group resolved the 18 amino acids RHAU-specific motif (RMS) or DHX36-specific motif (DSM) bound to a 5' parallel G4 DNA through NMR and then obtained a crystal structure [104, 105]. Importantly though, a G4-bovine RHAU crystal structure provided the first detailed insights into protein-G4 NA binding and unwinding. In the RHAU-G4 DNA structure, the DSM adopts an α-helical fold upon DNA binding and, together with the OB-fold-like sub-domain, parallel G-quadruplexes is selectively bound. The binding is linked to structural changes in helicase core that induce G4 unfolding, one base at a time [106].

1.5 Targeting non-B form DNA as an anticancer strategy

Targeting non-B form DNA structures is an emerging anticancer strategy. The first suggestion that it could be feasible came with the knowledge that telomeres and the promoters of oncogenes and genes with roles in cell signalling (*e.g.*, the small GTPase KRAS) form G4 DNA structures. Targeting strategies aim to exploit G4 DNA sequences in the promoters of oncogenes such as *c-MYC*, stabilising them to down-regulate gene expression, or in telomeres where G4 stability can reduce the accessibility of telomerase that is often overexpressed in tumours and required for tumour survival. Globally, the stabilisation of G4 DNA structures in the genome could be used to induce a DNA damage response and cell death in tumours deficient in helicases required for their metabolism [107].

Since G-quadruplexes are discrete folded globular DNA structures, analogous to folded protein molecules, this offers advantages for ligand recognition and stable binding. For example, the large planar surface of a terminal G-quadruplex has led to the development of many small-molecule families based on heteroaromatic systems that complement the G-quartet. NMR studies have revealed structural diversity in G4 DNA folding and this diversity of G-quadruplexes structure can enhance ligand selectivity, for example by exploiting grooves and loops in the structure. *E.g.*,

considering KIT1 and KIT2 G-quadruplex structures in promoter regions of the KIT proto-oncogene; KIT1 presents a cleft, unique to its structure, suitable for ligand binding and high throughput *in silico* screening for drug discovery. Many small molecules are now available that stabilize G-quadruplexes that may be used as a platform to develop more specific agents [107].

1.6 Therapeutic potential of targeting the DNA damage response

As previously discussed in section 1.2.1.2, mammalian cells have developed DNA damage response (DDR) pathways to monitor the integrity of the genome [108]. Defects in DDR pathway led to genome instability associated with cancer [108]. However, cancer specific defects in DDR also represent an opportunity to selectively target tumours, through compensatory DDR pathways that are often preserved or augmented to avoid apoptotic cell death [109].

Cancer therapy aims to specifically kill cancer cells by directly acting on a molecular or cellular aspect of cell function. For example, cytotoxic cancer agents target the cell cycle because cancer cells grow faster than normal cells. Other approaches modulate growth signals promoting entry into the cell cycle, through hormonal manipulation, therapeutic antibodies and drugs that inhibit the signalling pathway. In order to target the cell cycle, a good strategy is to exploit the known effects of DNA-damaging drugs. Their enhanced cell killing is caused by the cells' attempts to replicate the damaged DNA, often causing replication forks to stall or collapse. If this occurs, they are recognised by the checkpoint machinery and the cell-cycle arrests, leading to either DNA repair or cell death through apoptosis or senescence. The toxicity of DNA damage-based cancer therapy can be further modulated by targeting DNA repair pathways. For example**,** a fundamental role in cellular resistance to DNA cross-linking agents is played by FA/BRCA pathway. One therapeutic strategy is the use of compounds that induce DNA inter-strand cross-links, but their effectiveness is reduced by elevation of FA gene expression. However, FANCM and FANCJ have been identified as useful targets for modulation by small molecules, to re-instate the toxicity of chemotherapy in cancer cells [67].

1.6.1 Synthetic lethality

DDR is essential for genome stability and cell survival; in cancer, the genome stability is lost while the DDR pathway keeps its role providing new therapeutic opportunity [110]. Initially, in fact, drugs that pharmacologically inhibit DNA repair were proposed as agents to enhance chemo sensitivity and radio sensitivity [111]. However, likely to be the most successful approach is exploiting synthetic lethality to selectively kill the cancer cells and spare the normal ones [27]. The initial illustration of the efficacy of this approach comes from the success of PARP inhibitors and therapeutic application in BRCA-mutant ovarian cancers to target tumour suppressor gene loss (see below) [112].

Synthetic lethality (SL) was first observed in the abnormal eye phenotypes of *Drosophila* caused by two genes *Bar* and *Glass*. These abnormal flies could only survive and reproduce without observing a combination of the two phenotypes. The two genes are involved in embryonal development and encode transcription factors expressed in the eye. Loss of both genes simultaneously is not compatible with life. This concept was further studied in yeasts and then in human diseases, facilitated by the development of RNAi technology. SL, therefore, refers to a genetic principle where the combination of two genetic perturbations is lethal while each individually is not. Cells have developed redundant pathways to maintain cellular homeostasis, to ensure that cells can survive a single gene/pathway loss because another gene/pathway can compensate for it. Starting from this concept, new drug treatments for cancer have been developed. In cancer therapy applications of SL two aspects are important: first, the genes that are synthetic lethal with oncogenic driver mutations are not necessarily mutated and second, drugs that do not act as single agents could be useful in combination with another drug synthetic lethal with the first one [113].

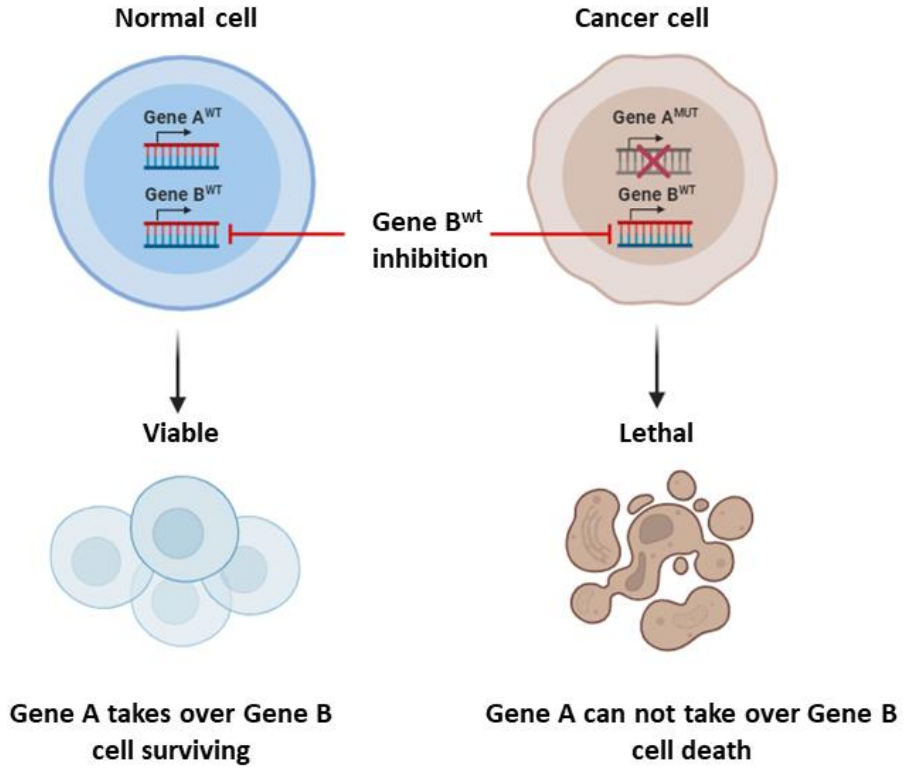


Figure 10. Synthetic lethality overview. In normal cells, the inhibition of gene B in isolation is not lethal to the cell as gene A can take over gene's B functions. The cells then will survive. In cancer cells, gene A is mutated and therefore its function is lost. It can not take over gene B functions anymore when gene B is inhibited, and the cell will go through apoptosis. (Figure realised with Biorender).

1.6.1.1 PARP inhibitors and synthetic lethality

A key example of the exploitation of the SL concept is the use of PARP inhibitors to treat patients lacking wild-type BRCA1 and BRAC2 genes. The actions of PARP proteins, of which there are 17 isoforms, include DNA damage recognition by binding to single-strand breaks [114] and the NAD^+-dependent synthesis of poly(ADP-ribose) chains on glutamate, aspartate and lysine residues of target proteins. PARP1 is the dominant class in the family and is involved in the BER, NER, HR and NHEJ repair pathways. At DNA lesions, PARP1 ribosylation (pADPr) modifies PARP itself, DNA repair proteins, histone H1 and other proteins including transcription factors to initiate the repair process [41].

PARP inhibitors (PARPi) work by competing with NAD^+ at PARP's active site, limiting PARP's ability to catalyse the formation of pADPr chains. Reduced pADPr prevents the recruitment of additional repair factors and the release of PARP from the

site of DNA damage, trapping it on damaged DNA further limiting the access of other repair proteins. This leads to the accumulation of SSB that are converted in DSB during DNA replication that is cell lethal. Usually, these lesions would be repaired by HR, but not in BRCA1^{-} or BRCA2^{-} deficient cells where HR is compromised, therefore indicating that the inhibitors are operating in a classic SL context. An advantage of exploiting the SL concept is that inhibitors can potentially be used as monotherapies where they are exclusively toxic to cancer cells with minimal side effects for the patients [115].

1.7 The PIF1 helicase

1.7.1 Discovery of the PIF1 helicase

The PIF1 gene was first identified in *Saccharomyces cerevisiae* as a gene with an essential role in the repair of mtDNA (mitochondrial DNA) after UV light and ethidium bromide treatments [116], while also affecting the recombination frequency between *rho*$^{+}$ and *rho*$^{-}$ allelic strains, indicating that it was a mitochondrial protein. A PIF1-enriched protein fraction revealed an ssDNA-dependent ATP*ase* activity and an associated DNA helicase activity, while a null-PIF1 mutant showed no activity in a similar purified fraction [117]. The gene was subsequently cloned and sequenced, to reveal the conserved SF1 helicases consensus motifs described in section 1.1.1.1; PIF1 shows homology with the *E. coli* UvrD helicase and the RAD protein of *S. cerevisiae* [118]. Through the screening of yeast mutants, PIF1 was re-discovered as a protein involved in telomere maintenance. PIF1 mutations cause an increase in the length and heterogeneity of telomeres, the loss of sub-telomeric genes and an increase in *de novo* telomere formation at broken chromosome ends [1, 119].

Analysis of the translated database identified a second PIF1-like gene in *S. cerevisiae*, RRM3, a gene previously known to repress recombination in ribosomal DNA [120]. Similarities and differences with PIF1 have been observed, in particular, Rrm3p and Pif1p both affect telomeres but differently as Rrm3p promotes telomere replication. However, the main role of Rrm3p is to promote replication through difficult-to-replicate sequences such as the Replication Fork Barrier (RFB) in ribosomal DNA, for example [121].

Further studies in *S. cerevisiae* demonstrated the involvement of PIF1 in conjunction with Dna2 in Okazaki fragment processing [122] and, more interestingly, a role in preventing genomic instability caused by sequences capable of forming G4 DNA [123]. More recently [124], PIF1 helicases have been shown to support DNA replication termination in yeast. Type I or II topoisomerases are critical for the removal of supercoils and torsional strain when two replication forks converge. However, when the final stretch of parental DNA becomes so short, supercoils cannot form and topoisomerase is excluded from the DNA. PIF1 is thought to help in overcoming the torsional strain at converging forks by unwinding the final stretch of parental dsDNA.

A screen in *S. pombe* identified a role for PIF1 in genomic DNA repair caused by sensitivity to alkylating agents [125], but notably, *S. pombe* does not have an RRM3-like gene (only one PIF1 protein). Another interesting aspect of the PIF1 family is that although first discovered in yeast and then human cells, a highly related protein is widely found in bacteria [126]. Biochemical analysis suggests that bacterial PIF is involved in resolving issues arising during DNA replication, recombination and repair that are in common with eukaryotes [127]. Many bacteria have linear chromosomes, which makes their replication problems closer to eukaryotes and correlates with PIF1's essential action at eukaryotic telomeres.

Human PIF1, as to be discussed in more detail, is a G4 DNA binding and processing enzyme, first indicated by its requirement for the stability of the human minisatellite CEB1 repeat when inserted into the yeast genome. In PIF1 deficient human cells, replication fork progression slows and chromosomal rearrangement increases near G4 DNA forming sequences [128].

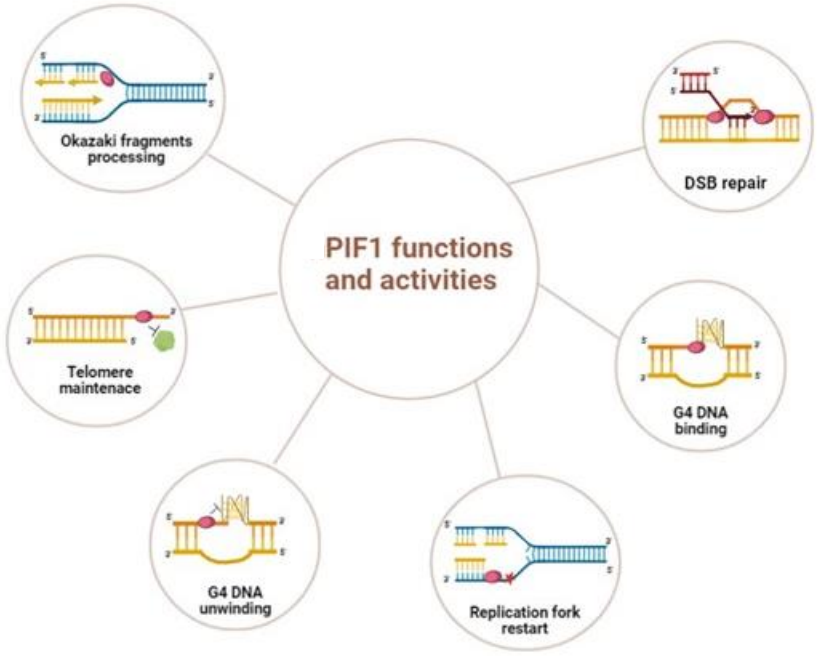


Figure 11. Summary of PIF1 functions and activities in bacteria, yeast, and human cells. Figure obtained using Biorender website and based on [121] [122] [123] and [127].

1.7.2 The structure of PIF1 proteins

PIF1 is an SF1B helicases that translocate in the 5'–3' direction using ATP [129], with the seven conserved motifs representative of the class (I, Ia, II, III, IV, V and VI) and three addition motifs (A, B and C) that PIF proteins share with RecD helicase family. Unlike RecD helicases, PIF1 has a unique PIF1-family specific sequence (PFSS) of 23 amino acids located between motifs I and III [126, 130]. Interestingly, *Bacteroides* PIF1 (*Bs*Pif1) can process a wide range of DNA substrates in a similar way to eukaryotic PIF1 [127], and its primary amino acid sequence is more similar to hPIF1 than *Sc*Pif1 (*Saccharomyces cervisiae*, yeast). Due to difficulties in obtaining purified hPIF1, the first crystal structures were obtained for *Bs*Pif1. *Bs*Pif1 has three structural domains, 1A and 2A with a RecA-like fold make the helicase core and the 2B accessory domain. Domains 1A and 2A are commonly seen in all SF1 and SF2 helicases [5] while 2B adopts a SRC-homology 3 (SH3)-like fold as in the structure of RecD2 [131]. The small 1B domain is inserted in the 1A domain, as also observed in other SF1B DNA helicases such as Dda and RecD2 [24, 131], where it acts as a “wedge” for separating the dsDNA strands [132]. Within 2B there is a rail-like β-hairpin which has its turn close to domain 1B, where it may assist in the unwinding process. A 2C

domain is characteristic of *Sc*Pif1 and other yeast, located in the domain 2B and absent in hPIF1 and *Bs*Pif1 structures. The conserved PFSS in domain 1A interacts structurally with the wedge and a short helix to stabilize the region for ssDNA binding, thus playing an indirect role in Pif1 enzymatic activity [132]. In *Bacteroides* and yeast, a mutational analysis of the PFSS motif suggests it is critical for ATP*ase* activity [133, 134].

High resolution hPIF1 core domain structures were obtained in 2019 [135] (Figure 12). They reveal the architecture typical of RecD2 helicases. In the RecA-like domains the Walker A (or motif I), the Walker B (motif II) and motif IV (arginine finger motif) are involved directly in NTP/Mg^{2+} binding and hydrolysis. The other conserved motifs are largely involved in energy transduction events coupling NTP*ase* to helicase activity [135]. Dehghani-Tafti *et al.* (2019) analysed ssDNA binding channel residues in hPIF1. These are highly conserved among species, and mutations result in a ~ 50-fold decrease in affinity when mutated, with corresponding impairment of DNA-dependent ATP*ase* and helicase activity. The wedge domain sequence is not conserved among the species. Eukaryotic PIF1 proteins have N-terminal and C-terminal accessory domains, while bacterial enzymes have only the helicase core. In hPIF1 (Figure 11) the N-terminal domain has non-specific DNA binding activity that improves the activity of the helicase core [135 136] (Figure A1, appendix). The role and activity of the C-terminal extension is unclear [135].

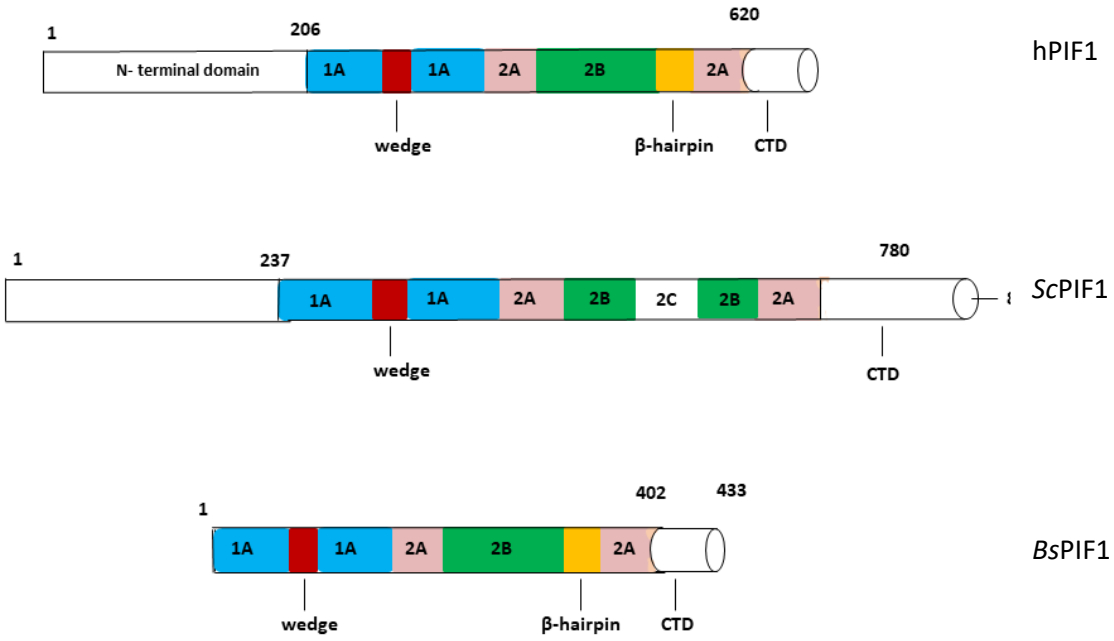


Figure 12. Stick diagram summarising the PIF1 proteins stuctures: hPIF1, *Sc*PIF1 and *Bs*PIF1. Core domain of each protein is coloured distinguishing the domains.

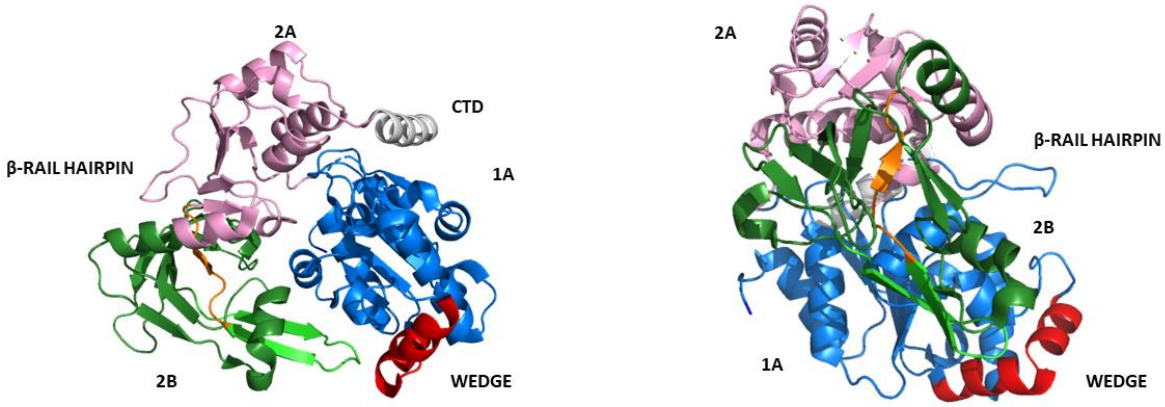


Figure 13. Apo hPIF1 crystal structure. (pdb: 6hpt) Cartoon representation of the hPIF1 core domain from two different views. Domain 1A in blue, domain 2A in pink and domain 2B in green. The C-term domain named CTD in white and the functional wedge domain in red are structurally part of the A domains while the rail β-harpin (orange) belongs to domain 2B. Figure realised with PyMol.

1.7.3 Biochemical activities of PIF1 proteins

*Sc*Pif1 has several biochemical activities: it preferentially unwinds fork DNA substrates with a 5' ssDNA overhanging, it unwinds DNA:RNA duplexes [137] and reduces telomerase processivity releasing it from telomeric DNA [138]. Most importantly, PIF1 proteins interact with G4 DNA structures. When nuclear *Sc*Pif1 is

inactive, rearrangements of G-rich human minisatellite CEB1 have been observed when inserted into the yeast genome [123]. Correlating with its role in processing of G4 DNA forming sequences *in vivo* [123], yeast PIF1 also unwinds G4 DNA *in vitro* [139 140]. Importantly, analysis of the yeast and human proteins also demonstrates specific binding to G4 DNA sequences *in vitro. Sc*Pif1 processes Okazaki fragments, making a flap during RNA removal by Dna2 [122] and when *Sc*Pif1 is overproduced, telomeres shorten [1], suggesting that PIF1 regulates telomerase activity. While the mechanism of PIF1 displacing telomerase from telomere ends is unclear, single-molecule experiments show the complex telomere-telomerase to be disrupted by the helicase and its efficiency depending on PIF1 concentration [141].

*Bs*Pif1 can unwind dsDNA, fork-like substrates and D-loops. It acts at stalled DNA replication forks and promotes Okazaki fragment maturation, lengthening the flap at the 5'-end of the fragment. Like *Sc*Pif1, bacteroid PIF1 unwinds partially duplex DNA with 5'-overhangs fork-like substrates but it is more efficient on DNA-DNA than DNA-RNA substrates [142]. An important feature of the PIF1 helicase family is the ability to unwind G4 DNA substrates, a characteristic confirmed for prokaryotic PIF proteins [132]. Like *Sc*Pif1, *Bs*Pif1 disrupts telomerase from telomere [142].

hPIF1 cellular and biochemical activities are conserved relative to *Sc*Pif1: it acts in telomere maintenance, processes Okazaki fragments [143, 144] and unwinds DNA structures resembling stalled DNA replication forks [145]. G4 DNA binding and unwinding activity is an intrinsic activity of helicase core [146] and hPIF1 also has a DNA strand annealing activity. Gu *et al.* (2008) localised this to the N-terminal accessory domain, while George *et al.* (2009) demonstrated that the core helicase domain had such a function. hPIF1 can also recognize and process stalled DNA replication forks, reversing the leading replication strand to generate ssDNA that could be used for replication restart [145]. Analysis *in vitro* has also shown that hPIF1 interacts with telomeric DNA and inhibits telomerase [147] demonstrating additional shared functions with yeast PIF1.

1.7.4 Mechanistic aspects of PIF1 protein

(i) *Bs*Pif1

The first structures of PIF1 were obtained with the *Bacteroides* protein. Several *Bs*Pif1 crystal structures were obtained, including apo, ADP, ADP·AlF_3^- and AMP·PNP bound. These revealed the ATP binding pocket between domains 1A and 2A. The absence of a Q motif (an exception in the Pif1 family) interacting with the adenine of ATP (Section 1.1.1) was noteworthy, but structural alignments showed a methionine (M10) residue in *Bs*Pif1 instead of the Q motif glutamine, interacting with adenine. The structural studies also reported residues coordinating the phosphate moiety: five amino acids in motif I coordinate the α- and β-phosphates while another five residues (two in motif I, one in motif III, one in motif IV and one in motif VI) act as 'a γ-phosphate sensor' [134]. *Bs*Pif1 was also crystallised with ADP·AlF_3^- and ssDNA, revealing the ssDNA binding channel across domains 1A and 2A. The structures also highlighted significant structural rearrangements occurring upon ssDNA binding: while domain 1A of the *Bs*Pif1-AMP·PNP structure can be superposed to the corresponding domain in *Bs*Pif1-ADP·AlF_3^--ssDNA, domains 2A and 2B make a significant movement. Domain 2A goes through a rotation of 22° while domain 2B reaches its final configuration after a 40° rotation and a displacement of 12.3 Å.

The structures also showed that three additional loops (1-3) in domain 2B undergo structural changes when the protein binds nucleotide and ssDNA. Among them, loop3 protrudes from the SH3 core domain and is part of a prominent β-hairpin; mutational analysis demonstrated its involvement in DNA binding and unwinding. The upcoming dsDNA faces the 2B domain which undergoes a closing movement upon nucleotide binding. Loop3 (top of the hairpin) closes on the 1B domain forming a tunnel where the ssDNA passes through in a 5'–3' direction during ATP hydrolysis. Once the dsDNA is firmly blocked by domains 2B and 1B, the force due to the translocation is sufficient for unwinding the dsDNA. It has been also hypothesised that once the translocation is over, domain 2B may move back to start another cycle [134].

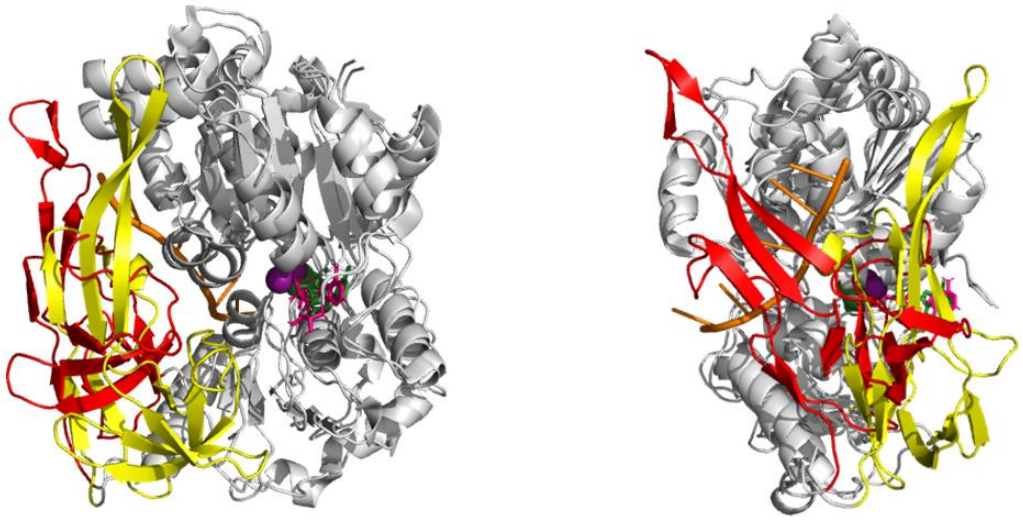


Figure 14. Alignment *Bs*Pif1 crystal structures complexed with AMP·PNP (pdb: 5fte) and ssDNA and ADP·AlF$_3$ (pdb:5ftb). The alignment shows the structural rearrangement of domain 2B, highlighted in red for AMP·PNP and yellow for ADP·AlF$_3$ structure. RMSD 2.6 Å.

A Bacteroides sp Pif1 crystal structure complexed with a symmetrical partially unwound dual-fork dsDNA (10 nts oligo-dT at both 3' and 5' ends) shows the need for two Pif1 molecules for binding and unwinding of a fork. When the first molecule of *Ba*Pif1 binds a 5' arm, the first base-pair of the fork starts to break while the *Ba*Pif1 molecule at the 3' arm acts as an accessory protein preventing the rehybridization of the two strands while also engaging the second base-pair for the next separation event. The communication of the two *Ba*Pif1 molecules happens through electrostatic interactions, regulating the unwinding activity [148].

(ii) *Sc*Pif1

*Sc*Pif1 crystal structures have also been obtained with and without nucleotides and ssDNA bound [149]. ADP·AlF$_4$ is bound in the cleft between domains 1A and 2A as observed in *Bacteroides* [150]. The *Sc*Pif1 Q motif is structurally implicated in adenine nucleotide binding and required for efficient ATP hydrolysis [149]. Lu *et al.* (2018) obtained structures with several ssDNA oligonucleotides including G-rich oligos that bind with higher affinity and subsequentially identified the protein residues specifically recognizing the Gs to be Glu724 (2A domain), Lys387 (1A domain) and Lys312 (1B domain). Further experiments confirm this *Sc*Pif1 nucleobase preference: the

dissociation constant ($K_{D,app}$) calculated for the G-rich ssDNA substrate is two folds lower than the one for a polyT substrate and in the presence of ATP the G-rich substrate $K_{D,app}$ is reduced while the poly T substrate one is enhanced. This observation suggests that the ATP may increase the binding with G rich sequences to *Sc*Pif1 and reduce the affinity for the polyT substrate [149]. Also, *Sc*Pif1 seems to interact with two short ssDNA substrates at the same time [151], suggesting the possibility of two binding sites for ssDNA and the possibility of binding long ssDNA molecules.

No crystal structure of *Sc*Pif1 in complex with G4 DNA has been obtained, but with a combination of low resolution SAXS and biochemical data, the authors hypothesise the presence of a 'clamp' for G4 binding. The 'clamp' has one arm in domain 2C and the second in the wedge domain, fixed at the entrance of the ssDNA binding channel [149]. However, most PIF proteins including hPIF1 do not have a 2C domain (this is largely a fungal-specific insertion in domain 2B) and the wedge residues implicated in G4 DNA binding by mutational analysis are not conserved [135].

Based on biochemical data, it has been suggested that G4 DNA can activate yeast Pif1 duplex unwinding activity by inducing its (Pif1) dimerization. The mechanism is thought to involve, first, the binding of two molecules of Pif1 to the G4 DNA motif followed by a protein dimerization event [3 140]. While a monomeric form of Pif1 can unwind RNA-DNA heteroduplexes and G4 structures, to unwind dsDNA the binding of multiple Pif1 molecules per DNA substrate is required. The inability of a Pif1 monomer to unwind dsDNA might be important for preventing it from unwinding dsDNA after it has displaced proteins, removed RNA strands or resolved G4 structures from within a ssDNA gap [152]. This activation mechanism by G4 could help in the rescue of replication forks on the leading strand stalled by G4 DNA formation.

smFRET experiments on a partial duplex DNA with a 3'-40 nt poly d(T) overhang, showed *Sc*Pif1 is recruited to the 3' ss-dsDNA junction and induces DNA looping by reeling in the ssDNA through ATP*ase*-dependent translocation activity. smFRET studies also demonstrated that *Sc*Pif1 acting as a monomer is unable to unwind dsDNA but multiple Pif1 molecules are needed to promote dsDNA unwinding [152]. A magnetic tweezer analysis of *Sc*Pif1 revealed that the protein can translocate for long

distances on ssDNA substrates [150], and with a high velocity when it does not need to separate the dsDNA.

A recent study investigated *Sc*Pif1-G4 DNA processing activity on a dsDNA hairpin substrate containing a G4 motif from the c-Myc promoter sequence. The study wanted to understand the mechanism at the base of G4 resolution by *Sc*Pif1, revealing two possible mechanisms. When ssDNA is available in the proximity of the enzyme, and upon collision with the G4 structure, *Sc*Pif1 shows a strand-switching behaviour where the secondary structure is resolved (after *Sc*Pif1 collides multiple times with the G4). In this case, multiple rounds of strand-switching mechanisms are needed. If the ssDNA is not available, *Sc*Pif1 pauses in front of the G4 structure until it removes it and afterwards resumes translocation. The switching behaviour can explain *Sc*Pif1*'s* preference for G4 structure conformations [153]. The switching model, however, does not support the dimerization of *Sc*Pif1, as previously supported by Zhang *et al.* (2016) in studies on a replication fork substrate with G4 structure at the free 5' single-stranded flap [154].

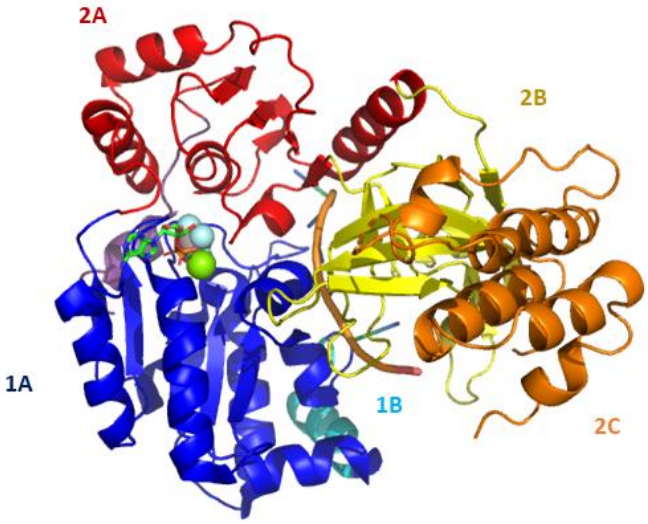


Figure 15. *Sc*Pif1 crystal structure complexed with ssDNA and ADP·AlF$_4$ (pdb: 5o6e). Domain 1A in blue, 1B in cyan, 2A in red, 2B in yellow, 2C in orange (unique of yeast Pif1) and CTD in purple. The ssDNA (TTTGGGTT) is located between domains 1A and 2B while the nucleotide is located between domains 1A and 2A. The G rich ssDNA is close to the 2C domain as well, suggesting its involvement in the G4 DNA binding.

(iii) *To*PIF1

Like *Bacteroides*, the bacteria *Thermus oshimai* has been used as a structural model for Pif1 proteins and is also comprised of a helicase core domain only. Many *To*PIF1 structures are available in the pdb database (https://www.rcsb.org/), with much overlap with the *Bacteroides* model. Notably, the only structure of apo Pif1 bound to ssDNA is available for *To*PIF1, which shows that dimerization of the protein is a structural consequence in *To*PIF1 ssDNA binding. Also, the dimerization happens with the 2B domain blocked in an open conformation (open ssDNA binding cleft) and it inhibits *To*PIF1 helicase activity [155].

Most importantly, however, thus far, the only crystal structure of a Pif1 protein bound to a G4 DNA substrate has been obtained with *Thermus oshimai* [156]. Domains 1B and 2B were identified as a G4-Recognizing Surface (GRS) along with specific residues that recognize the whole G4 DNA structure.

(iv) hPIF1

hPIF1 is, in general, less well structurally characterised compared to yeast and bacterial homologoues. hPIF1 helicase domain crystal structures show the domain architecture typical of the helicase family. Structures of apo and nucleotide bound enzyme have been obtained, allowing modelling of the structural events along the ATP*ase* chemical reaction coordinate. During ATP binding (AMP-PNP, ground state mimic), rearrangement of domains 1A and 2A is observed, with a positional change of ~4 Å of motif A causing a structural change of domain 2B. Also, domains 2A and 2B undergo a positional shift that prepares the protein for ATP hydrolysis [135]. In Dehghani *et al.* (2019), a post-catalytic model ($ADP{\cdot}ALF_4^-$) of the protein is very flexible, showing movements of the protein structure that breaks H bonds controlling the orientation of domains 2A and 2B, and their movement allowing the formation of the ssDNA binding cleft while the rail β-hairpin moves towards the wedge [135].

Unlike yeast and bacterial PIF1, a crystal structure of hPIF1 complexed with ssDNA is not available. A model however has been obtained by superposition of the *Sc*Pif1 and *Bs*Pif1 ssDNA-$ADP{\cdot}AlF_4^-$ structures and modelling domain movements,

which confirms a high degree of conservation of the ssDNA binding channel residues in PIF1 family helicases [135] (Figure 13).

Table 1 below summarises the important PIF1 information described above.

Protein	Crystal structures	pdb	Features
*Bs*Pif1	apo	5ftd	Absence of Q motif [133]
	ADP	5ftc	ssDNA channel across domains 1A and 2A [133]
	$ADP{\cdot}AlF_4$	5fhf	structural rearrangements occurring upon ssDNA binding [133]
	AMP·PNP	5ftb	β-hairpin involvement in the DNA unwinding and binding [133]
	ssDNA $ADP{\cdot}AlF_3^-$	5fte	two Pif1 molecule to unwind replication fork (*Ba*Pif1) [144]
*Sc*Pif1	G rich ssDNA $ADP{\cdot}ALF_4$	506e	First Pif1 protein identified [115]
	ssDNA $ADP{\cdot}ALF_4$	5o6b	Atypical 2C domain present [144]
			Q motif involved in adenosine binding [145]
			Clamp to bind G4 DNA with one arm in 2C domain and the second in the wedge [145]
			*Sc*Pif1 monomer to unwind DNA:RNA and G4 substrates [148]
			*Sc*Pif1 dimer to unwind dsDNA substrates [139]
			*Sc*Pif1 patrolling activity used to unwind R-loops [146]
			Two G4 DNA unwinding mechanisms [149]
*To*Pif1	Apo	6s3e	Dimerization after ssDNA binding [151]
	ssDNA $ADP{\cdot}MgF_4$	6s3i	G4 recognition surface identified (GRS) in domain 1B and 2B [152]
	ssDNA $ADP{\cdot}AlF_4$	6s3m	
	ssDNA apo	6s3p	
	G4-$ADP{\cdot}AlF_4$	7oar	
hPIF1	apo	6hpt	General protein architecture of the PIF family conserved [134]
	AMP·PNP	6hph	ssDNA binding channel conserved [134]
	$ADP{\cdot}AlF_4^-$	6hpu	Role for all conserved motifs accounted for, including A, B and C [134]

Table 1. Summary table of notable properties of Pif1 proteins. See main text for details.

1.7.5 Human PIF1 and oncogene-driven replication

siRNA–mediated hPIF1 depletion results in a combination of apoptosis, reduced cell survival, hypersensitivity to therapeutic DNA replication inhibitors and defective cell cycle progression in several cancer cell lines, independent of p53 status. hPIF1-depleted cells show elevated levels of cyclin E as well as p21 suggesting that hPIF1 down-regulation prevents cells from entering S-phase. Non-cancerous cells do not show such a response [129]. From these studies it was hypothesised that hPIF1 has an essential role in the restart of the replication fork when stalled after disruption by hydroxyurea (HU) treatment. Generally speaking, when the cell cycle has been stalled irreversibly by DNA damage, hPIF1 is required for replication recovery through its role in S-phase re-entry. Oncogene activation and the loss of control of the entry into the S-phase leads to chronic DNA replication stress (defined as stalling or slowing or replication fork progression [157]) in different types of tumour cells [158]. Further studies showed that hPIF1 is required to support replication fork progression in *ras* transformed cells [101], mimicking the conditions of replication stress experienced in the early stages of tumorigenesis [159].

Moreover, depletion of hPIF1 in the presence of a G4 DNA-stabilising ligand slows replication forks further with an increased chance of them stalling during the replication, suggesting the forks' movement can be disrupted by G4 DNA formation in GC–rich regions and that hPIF1 has a role in their resolution. The *in vitro* data showing G4 DNA binding and the unwinding of G4 DNA and stalled replication forks correlates well with cell-based studies supporting the notion that PIF1 supports replication fork progression and re-start, possibly through CHK1-dependent mechanisms [145, 146]. hPIF1 helps forks to progress by binding and unwinding non-canonical structures [128]. All the findings described above suggest a potential therapeutic application of PIF1, where inhibition could specifically induce apoptosis in tumour cells while insensitive non-tumour cells survive [129].

1.8 Hypothesis, aims and objectives

As summarised above and widely discussed in the literature, hPIF1 is a key helicase in maintaining genome stability, hypothesised to stabilize replication forks stalled [145] when they encounter G4 DNA structures. Moreover, it has been shown that siRNA mediated hPIF1 knocked down causes some tumour cells to undergo apoptosis while non-tumour cells survive [129], making PIF1 a potential cancer therapeutic target, through the synthetic lethality concept. This research work had two major aims connected to developing hPIF1 as a therapeutic target. First, I sought to better understand the mechanistic basis of hPIF1 G4 DNA binding and unwinding and second, to facilitate development of new hPIF1 inhibitors able to inhibit helicase activity.

The specific objectives of this thesis work were:

1. To understand the substrate preference of G4 DNA-hPIF1 interactions through an investigation of binding to both synthetic and physiological G4 DNA structures, with different sequences and conformations.

2. To determine the critical residues involved in G4 DNA binding through a site-directed mutagenesis and biochemical assay approach, starting from a comparison between hPIF1 and a structure of bacterial *To*PIF1 bound to G4 DNA.

3. To understand hPIF1 G4 DNA binding and unwinding and how they are influenced by nucleotide co-factors and analogues representing steps in the cycle of ATP hydrolysis steps.

4. To identify novel hPIF1 helicase activity inhibitors *in vitro*, based on hits (binders) identified in a previous thermal shift assay, and synthesised through new synthetic routs.

5. To assist in improving the inhibitory activity of EN300-02473, a known binder and inhibitor of hPIF1 previously identified by co-crystallisation screens with

the hPIF1 helicase domain, through chemical modification of its scaffold and *in vitro* assay of unwinding inhibition.

2 MATERIAL AND METHODS

2.1 Generation of protein expression constructs

The wild-type expression constructs used in this work were already present in the lab (GST FL hPIF1, GST hPIF1 HD206-END, GST hPIF1 HDΔC26 and His_{12}SUMO FL hPIF1); they were in pET11c backbone and made in home. More info about the construct is in Dr Dehghani-Tafti's PhD thesis, University of Sheffield (2018). Figure 15 represents the plasmid used for cloning GST hPIF1 proteins.

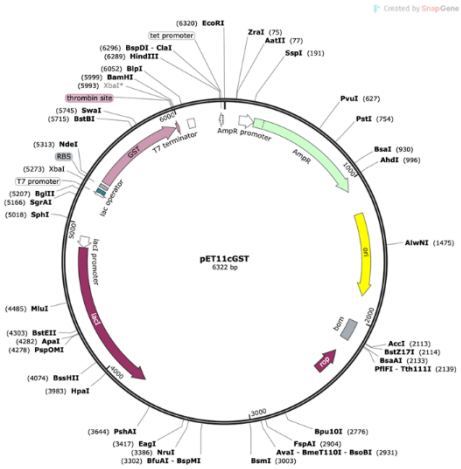


Figure 16. Plasmid used for GST proteins production. The coding sequence of hPIF1 ORFs was inserted into a pET11c backbone, specifically between the XbaI site and BamHI. The vector map is obtained using SnapGene.

2.1.1 DNA manipulation and cloning of hPIF1 mutants

Mutants were generated by overlapping primer extension. Table 2 shows the sequences of primers used. Full-length, codon-optimised, wild-type PIF1 was used as a template for the PCR. Table 2 also shows the flanking primers (top and bottom) used for cloning into the expression vectors, where the wild-type ORF was replaced by the mutated ORF. Restriction enzyme sites used for cloning are underlined.

2.1.2 PCR reaction

PCR reactions were performed with upstream flanking primers and reverse mutagenic primers (_B in Table 1) and downstream flanking primers and forward mutagenic primers (_T in Table 1) as pairs (1 µM each). PCR was performed in a 100 µL reaction containing 1X Pfu buffer, 0.2 mM dNTP, 2 mM $MgCl_2$, 1 ng/µL

template and 0.05 U/µL Pfu DNA polymerase. The PCR cycling conditions were the following: initial denaturation at 95°C for 3 min followed by 25 cycles of 95°C for 1 min, annealing at 55°C for 1 min and extension at 72°C for 2 min.

The PCR products were run on a 1% agarose gel (section 2.3.1), the DNA bands were identified using UV transillumination and excised and purified using a QIAGEN MinElution gel extraction kit, according to the manufacturer's instructions. A secondary PCR was set up, with the gel purification products as templates (~20 ng each) and SynP206 and PifBglII as primers (Table 2). The PCR conditions were as above, and the final products were purified with the QIAGEN QIAquick PCR Purification Kit according to the manufacturer's instructions. The purified products were digested for 2h at 37°C in the indicated buffer with restriction enzyme NdeI cutting in the sequence *5'-CATATG/3'-GTATAC* and BglII cutting in the sequence *5'-AGATCT/3'-TCTAGA).* Products were resolved on a 1% agarose gel and purified as before.

Primer	Sequence
K556A_T	**GCA**AGCCAGGGTATGACACT
K556A_B	TGTCATACCCTGGCT**TGC**ATGAATGCTCATTGC
H525E_T	**GCA**GCAGATCGTTGGACCGT
H525E_B	GTCCAACGATCTGC**TGC**AATAACTTCGGTAA
Q542A_T	**GCG**CAGCTGCCGCTGCAAC
Q542A_B	TGCAGCGGCAGCTG**CGC**ACGTGACAGCAGTTG
S540A_T	**GCA**CGTCAGCAGCTGCCGCT
S540A_B	GGCAGCTGCTGACG**TGC**CAGCAGTTGACCACC
R528A_T	**GCT**TGGACCGTTCAGGCAACC
R528A_B	CCTGAACGGTCCA**AGC**ATCTGCATGAATAAC
H555A_T	**GCA**AAAAGCCAGGGTAYGAC
H555A_B	CATACCCTGGCTTTTT**TGC**AATGCTCATTGCCCA
R290E_B	**GAA**CCTGGTGTGCGTCAGGGT
R290E_T	CTGACGCACACCAGG**TTC**CTGGGCCAGTGCAA
SynP206	5'-ATGCTCTAGACATATGCAGCTGTCTGAAGAACAG
PifBgl II	5'-CATGAGATCTTCACAGAATCGGATCCATATTT
SynPIF206	5'-ACGCAGATCTTCACAGGCTACGACCACGACG
SynPifBsa	5'-GTGGGTCTCGGATCCTCACAGAATCGGATCCATATTTTCTT

Table 2. List of forward (T) and reverse (B) primers used for hPIF1 mutagenesis. The codons modified are indicated in bold.

2.1.3 Agarose gel electrophoresis

DNA fragments and plasmids were analysed on 1% (w/v) agarose gels containing 0.5 µg/mL ethidium bromide for visualisation of double-stranded DNA. DNA samples were loaded on the gel with 1X DNA-loading dye (10X: 20% (w/v) Ficoll 400, 0.1 mM EDTA (pH 8.0), 1% (w/v) SDS, 0.25 (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol) and run at a constant voltage of 18 V/cm in 1X TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA).

2.1.4 DNA ligation

Of each PCR fragment obtained, 10 ng were ligated with 10 ng of pETHis$_{12}$SUMO vector digested with NdeI and BamHI (New England Biolabs). The DNA was combined and heated at 68°C for 3 min followed by 1 min on ice. 1 µL of 2X ligation reaction mix (2X buffer, 80 units of T4 DNA ligase (New England Biolabs) and 7.8% (w/v) PEG 8000) was added and reactions incubated at RT for 10 min and then on ice for 5 min.

2.1.5 Transformation of *E.coli*

Competent cells were generated by the method of Inoue et al 1990 [160]. The ligation products were used to transform 100 µL of DH5α competent cells; after 20 min incubation on ice the cells were heat shocked for 30 sec at 42°C, allowed to recover for 2 mins on ice and incubated for 1h at 37°C with 360 µL of outgrowth media (1X Lysogeny Broth (LB): 10 g/L tryptone, 5g/L yeast extract and 10 g/L NaCl), 0.4% (w/v) glucose, 25 mM KCl, 10 mM $MgCl_2$). The transformed cells were plated on LB agar plates (1X LB, agar (15 g/L), 100 µg/mL carbenicillin) and incubated overnight at 37°C. A single colony from each ligation plate was picked and inoculated for overnight growth at 37°C (LB, 100 µg/ml carbenicillin) and plasmid purification.

2.1.6 Plasmid purification

The overnight cultures (100 mL) were pelleted at 3 000 x *g* for 5 min at 4°C, thoroughly drained, and then re-suspended in 1.6 mL of solution I (25 mM Tris-Cl, 10 mM EDTA, 50 mM glucose, pH 8.0), and incubated for 5 min with 0.4 mL 20 mg/mL lysozyme[161]. 4 mL of solution II (0.2 M NaOH, 1%SDS) was added, and the reaction was incubated for 4 min, then 3 mL of solution III (3M KOAc (potassium acetate pH 5.5) was added, mixed gently, and left on ice for 10 min. The lysate was centrifuged at 25 000 x *g* for 10 min at 4°C; 8.5 mL of supernatant was removed and added to 0.7 vol. of isopropanol in a fresh tube. The tubes were spun at 25 000 x *g* for 10 min at 4°C. The pellets obtained were drained and dissolved in 0.5 mL of TE (10 mM Tris-Cl pH 8, 0.1 mM EDTA pH 8.0) for 10 min at 37°C. The solution was transferred to a 1.5 mL Eppendorf tube with 0.5 mL ice-cold 5 M LiCl, mixed and left for 5 mins on ice, then spun for 5 min at 21000 x *g* at 4°C. Isopropanol (1 mL) was mixed with the supernatant and spun at 21000 x *g* for 15 min at room temperature. The supernatant was removed, and the pellets dissolved in 0.2 mL 20 µg/mL RN*ase* in TE and incubated for 20 min at 37°C. 1.5 vol. of PEG/NaCl (1 vol. H_2O, 4 vol. 5M NaCl, 25 vol. 13% w/v PEG8000) was added and the tubes incubated for 20 min on ice. The tubes were spun at 21 000 x *g* for 20 min at 4°C, drained, and dissolved by warming at 37°C in 360 µL of TE. 40 µL of 5 M NH_4Ac was added and the solution extracted with 400 µL of phenol-chloroform-isoamyl alcohol (25:24:1). DNA was precipitated with 2.5 vol. of ethanol (EtOH), pellets washed with 0.75 mL of 80% EtOH and then absolute EtOH (3 min spins in a microfuge). Pellets were dried under vacuum and dissolved in 100 µL of deionised H_2O.

2.1.7 Sub-cloning into a GST-vector

All the mutated ORFs were re-cloned into a GST expression vector using XbaI and BamHI overhangs. PCR products were generated with SynPif206 and SynPifBsa as primers and processed as described in section 2.1.2.

2.1.8 DNA sequencing

The sequences of all mutants obtained were confirmed by Sanger DNA sequencing (Genewiz, sequence cover of 100 bp per read).

2.2 Protein production and analysis

2.2.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

To analyse the molecular weight and purity protein samples were run on a 10% SDS-PAGE gel: bis-acrylamide 29:1, 375 mM Tris-Cl pH 8.8, 0.1% (w/v) SDS with a 5% (w/v) stacking gel (5% acrylamide: bis-acrylamide 29:1, 125 mM Tris-Cl pH 6.8, 0.1% (w/v) SDS) and were set with 0.1% (w/v) APS and 0.1% (v/v) TEMED (tetramethylethylenediamine). Protein samples were dissolved in 1X Laemmli sample buffer (62.5 mM Tris-Cl pH 6.8, 10% (v/v) glycerol, 0.1% (w/v) bromophenol-blue, 2% (w/v) SDS, 2.5 % /v/v) β-mercaptoethanol) and were boiled (5 mins at 94°C) before loading. Gels were run at 8 V in SDS-PAGE running buffer (25 mM Tris-base, 0.1% (w/v) SDS, 0.2 M glycine) and stained with Coomassie-blue solution (1.2% Coomassie brilliant blue, 40% (v/v) methanol, 10% (v/v) acetic-acid) and de-stained with 7.5% methanol, 7.5% acetic-acid.

2.2.2 Protein quantification

All measurements were made in triplicate. Proteins were quantified with the Bradford (BioRad) assay according to the manufacturer’s instructions.

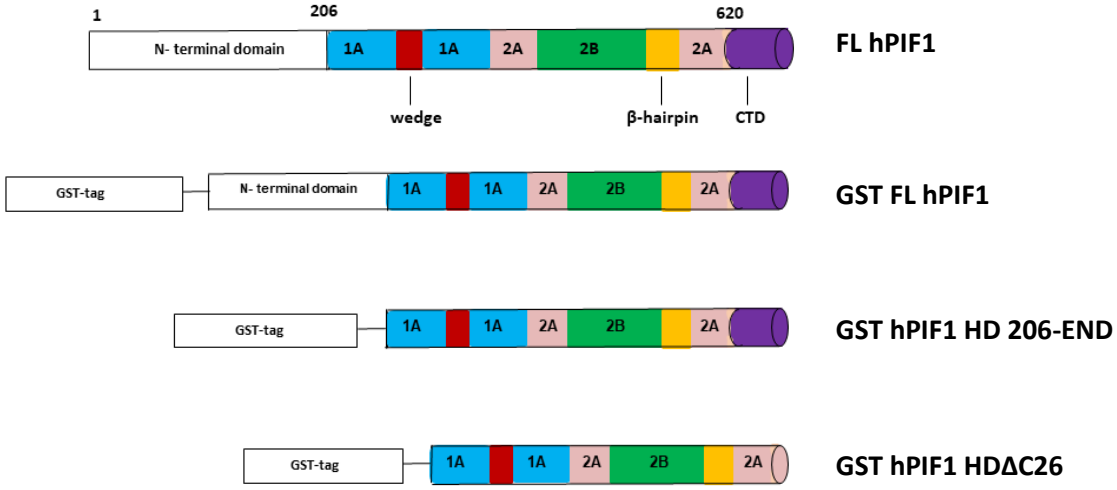


Figure 17. Summary of hPIF1 constructs used in this work. At the top FL hPIF1, tag-free protein used to exploit hPIF1 helicase activity. The remaining constructed illustrated are GST- tagged proteins: GST FL hPIF1, GST hPIF1 HD206-END and GST hPIF1 HDΔC26 (core domain).

2.3 FL hPIF1 purification

2.3.1 Protein expression

HisSUMO-tagged PIF1 expression construct (1-10 ng, His_{12}SUMO FL hPIF1), was used to transformed ArcticExpress™ BL21(DE3) cells (Novagen) as described above (section 2.3.1). The cells were then plated on Zymo Broth (ZB) agar plates (ZB: 10 g/L peptone, 5 g/L yeast extract and 10 g/L NaCl; agar 15 g/L, 0.4% (w/v) glucose, 25 mM M9 minimal salts, 2 mM $MgSO_4$) supplemented with carbenicillin 100 µg/ml and gentamicin 20 mg/ml for selection and incubated at 37°C overnight. pETGSTPIF1 expression constructs were used to transform BL21(DE3) cells and plated on ZB agar plates (carbenicillin 100 µg/mL).

The starter culture was set up (1X ZB (10 g/L peptone, 5 g/L yeast extract and 10 g/L NaCl), 0.4% (w/v) glucose, 25 mM M9 salt, 2 mM $MgSO_4$, "ZB media") with antibiotic selection (carbenicillin 100 µg/mL +/- gentamicin 20 mg/mL, as appropriate) and a single colony was picked, placed in 10 mL of broth and incubated at 37°C. After ~7h, 300-500 µL of culture was used to inoculate 1L of ZB media in 2L flasks (prepared as above, with antibiotics) and grown overnight at 22°C with shaking (200

rpm). Once the culture reached an A of 0.1, 1M IPTG (isopropyl β-D-1-thiogalactopyranoside) was added to each flask to a final concentration of 0.4 mM. Cells were incubated at 8°C in the incubator for 2 days. Cells were harvested by centrifugation at 2 200 *x g* for 18 minutes at 4°C, flash-frozen with liquid nitrogen and stored at -80°C.

2.3.2 Full-length hPIF1 purification

All the operations were at 4°C. Thawed cell paste (~40 g) was re-suspended in lysis buffers (50 mM HEPES, 10% sucrose, 2.5 mM TCEP, 1 mM PMSF (phenylmethanesulfonyl fluoride)) 0.1 M NaCl (low salt) in a volume of 1.5 mL per gramme of cells and then 1/50 vol of lysozyme [25], 50 mg/mL was added to the pellet and incubate for 30 minutes. 1.5 mL per gramme of cell paste of lysis buffer/1 M NaCl (high salt) was added, and the lysate was sonicated at 10-15 microns amplitude for 30-sec pulses in three cycles using a Soniprep 150 sonicator (MSE). The sonicated cell lysate was cleared at 20 000 x *g* for 30 minutes at 4°C. The supernatant was recovered and 0.125 volumes of 5 M NaCl was added followed by a 1/20 volume of 10% polyethyleneimine P (polymin P) pH 8.0). After 5 minutes on ice, the cell solution was cleared (20 000 x *g*, 5 min) and proteins precipitated by the addition of 0.226 g per mL $(NH_4)_2SO_4$ (40% saturation at 0°C, 20 minutes equilibration, 20 minutes centrifugation at 20 000 *x g*). The pellet was thoroughly drained and dissolved in His-buffer/20 mM imidazole (50 mM HEPES pH 7.5; 0.5 M NaCl; 20 mM imidazole; 10% sucrose; 1 mM TCEP (tris(2-carboxyethyl)phosphine) and the solution cleared again by centrifuging at 20 000 *x g* for 10 min.

2.3.3 His-trap affinity chromatography

The protein solution was loaded on an equilibrated 5 mL His-trap HP column (Cytiva), the column washed with 5 column volumes (col. vol.) of His-buffer (50 mM HEPES pH 7.5, 10% sucrose, 1 mM TCEP, 20 mM imidazole, 0.5 M NaCl) and the protein eluted on a 15 col. vol. gradient of His-buffer to 200 mM imidazole. The column was run at a flow rate of 2 mL/min and 3.5 mL fractions were collected. Protein extraction and chromatography were analysed by SDS-PAGE (section 2.1.9)

and peak fractions pooled. Protein concentration was determined by the Bradford (BioRad) assay, section 2.2.2).

2.3.4 SUMO-tag cleavage

The His_{12}SUMO-tag was cleaved with overnight dialysis at 4°C (dialysis buffer: 50 mM Tris- HCl pH 7.5, 300 mM NaCl, 10% glycerol; 0.5 mM TCEP) with SUMO Express Protease (Novagen) at 1U per mg of protein.

2.3.5 Second His-trap chromatography

The cleaved/dialysed protein solution was re-applied to a His-Trap column (Cytiva, 2 X 1 mL in line), previously equilibrated with 5 col. vol. of His trap buffer/5 mM imidazole (50 mM Tris-Cl pH 7.5, 300 mM NaCl, 5% sucrose, 2.5 mM TCEP), washed with 2 col. vol. of His trap buffer/5 mM imidazole then a gradient was created over 15 col. vol. to His trap buffer/55 mM imidazole for elution (1 mL/min and 1 mL volumes collected). The peak fractions were analysed and identified by 10% SDS-PAGE and the purest fractions were pooled and concentrated with a 5 000 kDa cut-off centrifugal filtration device (Vivaspin) at 6 000 x *g*, 4°C.

2.3.6 Gel filtration chromatography (SEC)

A Superdex 200 10/300 GL column was equilibrated with 20 mM HEPES pH 7.5, 300 mM NaCl, 1 mM EDTA, 2.5% sucrose, 1 mM TCEP. Protein (~ 0.5 mL) was loaded on the column and eluted at a flow rate of 0.5 mL/min and 0.5 mL fractions were collected. The protein fractions were analysed by 10% SDS-PAGE. Peak fractions were concentrated on a 10 kDa cut-off centrifugal filtration device (Vivaspin). Aliquoted, flash frozen in liquid N_2 and stored at -80°C.

The optimised scheme for the protein production using pETHis_{12}SUMO is summarised in Figure 18.

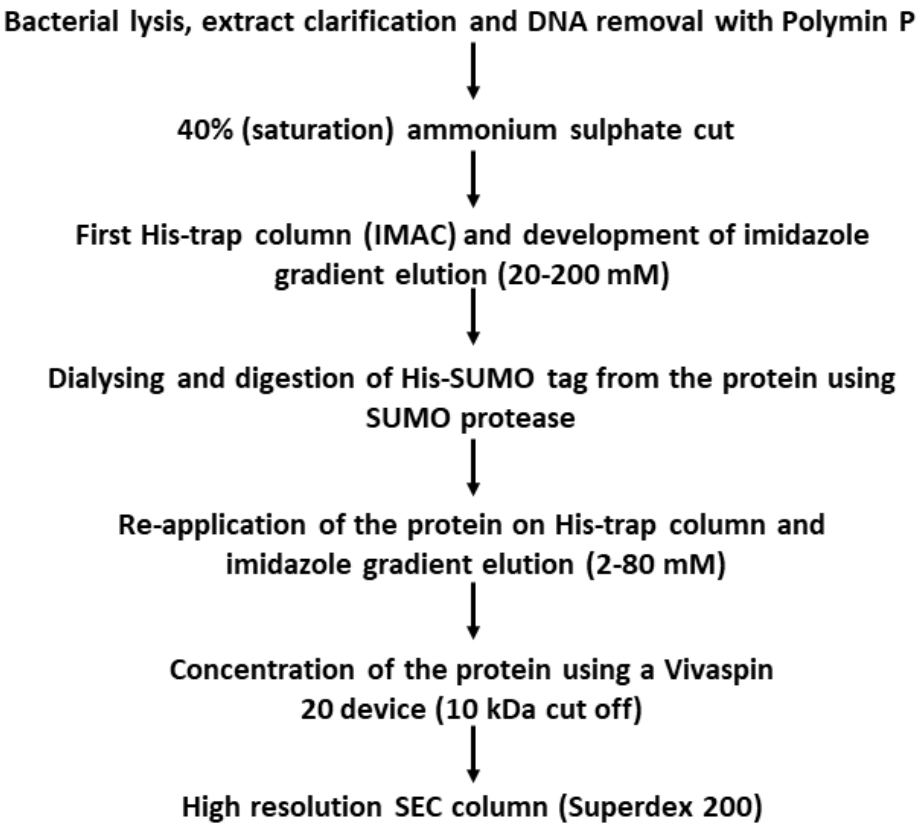


Figure 18. Summary of the FL hPIF1 purification steps.

2.4 Purification of GST proteins

2.4.1 pETGSTPIF1 expression

The products of expression using the GST tag were GST FL hPIF1, GST hPIF1 HDΔC26 and GST hPIF1 HD 206-END. The plasmids used pETGSTPIF1FL, pETGSTHD 206ΔC and pETGSTHD206-END were already available in the lab.

Transformation of the GST-proteins was performed in BL21(DE3) competent cells, with the same protocol used in section 2.3.1. A starter culture and 1 L pre-induction culture were established as above in section 2.3.1 (carbenicillin selection only). Once the culture reached an A of 0.1, they were induced with 1M IPTG. Expression was performed at 22°C in the incubator for 6-8 hours. Cells were harvested and stored as above.

2.4.2 GST FL hPIF1 purification

2.4.2.1 Cell lysis

The cell paste was thawed and re-suspended in 1.5 mL of lysis buffer/0.1 M NaCl per gramme of cells (lysis buffer: 50 mM Tris-Cl pH 7.5, 10 mM EDTA, 10% glycerol, 10 mM DTT, 1 M PMSF) and 1/50 vol. of 50 mg/mL lysozyme [25] was added to the cell paste. After 30 minutes, the same volume of lysis buffer/0.7 M NaCl was added to the lysate and sonicated at 10-15 microns amplitude for 30 second pulses in three cycles using a Soniprep 150 (MSE). The lysate was cleared at 25,000 x *g* for 30 minutes at 4°C and then 0.125 vols. 5 M NaCl was added. Polyethyleneamine P (10% w/v stock pH 8.0) was then added dropwise with mixing to 0.5% w/v. After the solution was equilibrated and cleared for 5 min at 25 000 x *g*, 0.291 g/mL of $(NH_4)_2SO_4$ was added to the supernatants and equilibrated for 20 min before spinning for 20 min at 25 000 x *g*. The protein pellet obtained was thoroughly drained and dissolved in 1 mL per gramme of lysis buffer/ 0.4M NaCl. Protein was bound overnight to equilibrated GSH-sepharose beads (Glutathione Sepharose™ 4 fast flow, Cytiva), 1mL beads per ~30 g of cells.

The next day, the glutathione beads were collected by centrifugation and washed in batch 2 times for 10 minutes with 10 volumes of lysis buffer (50 mM Tris-Cl pH 7.5, 10 mM EDTA, 10% (v/v) glycerol, 10 mM DTT, 1 mM PMSF)/1 M NaCl. The beads were then loaded on a drip column and washed first with 40 volumes of lysis buffer/1 M NaCl and then with 40 volumes lysis buffer/0.2 M NaCl.

2.4.2.2 Protein elution

GST-elution buffer (2-4 volumes; 25 mM Tris-Cl pH 8.0, 0.3 M NaCl, 1 mM EDTA, 10% (v/v) glycerol, 20 mM reduced glutathione, 5 mM DTT pH 8 at 4°C) was added to the column and the eluted solution (2-4 volumes) collected.

2.4.2.3 Ion-exchange chromatography

A Source S column buffer was prepared (20 mM Na phosphate pH 7.2, 2.5 mM DTT, 10% (v/v) glycerol, 0.5 mM EDTA, 0.5 mM PMSF) and the concentrated protein

was diluted with 2.5 vol. of S-buffer/no-salt and load on a 1 mL source S ion-exchange column equilibrated with 10 col. vol. S buffer/0.1 M NaCl. The column was washed with 2 col. vol. of 0.1 M NaCl buffer and a gradient was developed with S buffer to 0.6 M NaCl. The column was run at a flow rate of 1 mL/min and the fractions collected in 1 mL fraction volume. The peak fractions were identified and analysed by 10% SDS-PAGE. The peak fractions were concentrated to ~0.4 mL in a 30 kDa cut off centrifugal filtration device (Vivaspin) at 4°C.

2.4.2.4 Gel filtration chromatography (SEC)

A Superdex 200 10/300 GL column was equilibrated with 20 mM Tris-Cl pH 7.5, 300 mM NaCl, 1 mM EDTA, 5% (v/v) glycerol, 5 mM DTT and 0.1 mM PMSF. The SEC column was performed as in section 2.3.6

The most productive protocol for GST FL hPIF1 purification is summarised in Figure 19.

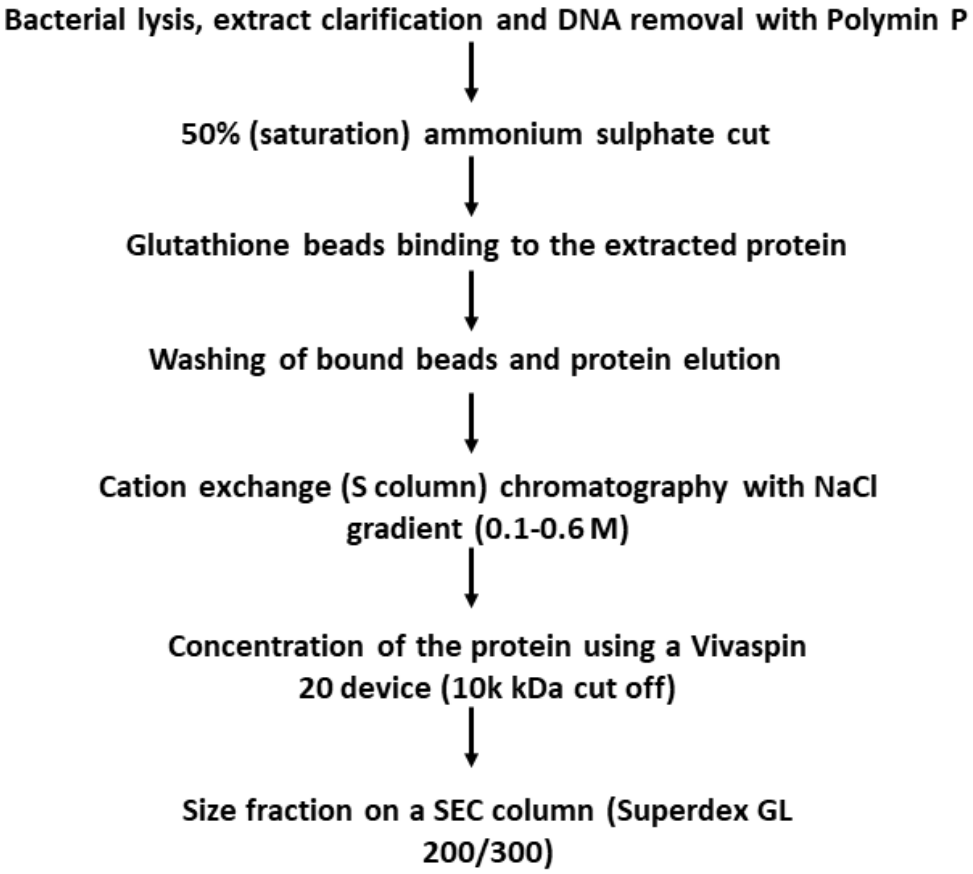


Figure 19. Summary of GST FL hPIF1 purification steps.

2.4.3 GST helicase domain purification

The protein extraction and elution from GST sepharose were performed as in sections 2.4.2.1 and 2.4.2.2. GST hPIF1 HD 206-END than GST hPIF1 HDΔC26 were purified according to the same protocol, using Q anion, heparin and gel filtration as described below.

2.4.3.1 Q anion exchange chromatography

A Q column buffer was prepared with 20 mM Tris pH 8, 2.5 mM DTT, 10% (v/v) glycerol, 0.5 mM EDTA, 1 mM PMSF. The protein previously eluted from the drip column was diluted in the same volume of 20 mM Tris pH 8.8 and 10% glycerol buffer, loaded on a 1mL Q anion exchange column previously equilibrated with 10 col vol of Q buffer/0.1 M NaCl. The column was washed with 2 col. vol. of Q buffer/0.1 M

NaCl buffer and a gradient developed with Q buffer to 0.7 M NaCl. The column was run at a flow rate of 1 mL/min and the fractions collected in 1 mL fraction volume. Column fractions were analysed by 10% SDS-PAGE. The protein in the flow through was purified further.

2.4.3.2 Heparin chromatography

The protein fractions collected from the Q anion column flow through were diluted in the same volume of Heparin buffer (20 mM NaPhosphate buffer pH 6.3, 2.5 mM DTT, 10% (v/v) glycerol, 0.5 mM EDTA, 0.5 mM PMSF). The Hi-Trap Heparin column was equilibrated with 10 col. vol. of Heparin buffer/0.1 M NaCl. The protein sample was applied to the column which was then washed with 2 col vol. of Heparin buffer/0.1 M NaCl and the protein eluted with a gradient developed with Heparin buffer 0.7 M NaCl (over 15 col. vols). The column was run at a flow rate of 1 mL/min and 1 mL fractions collected. The peak fractions were identified and analysed on a 10% SDS-PAGE; the selected fractions were concentrated on a 30 kDa cut off centrifugal filtration device (Vivaspin) to < 0.5 mL.

2.4.3.3 Gel filtration chromatography

Gel filtration and storage of the protein were performed as before. The optimised column purification steps are summarised in Figure 20.

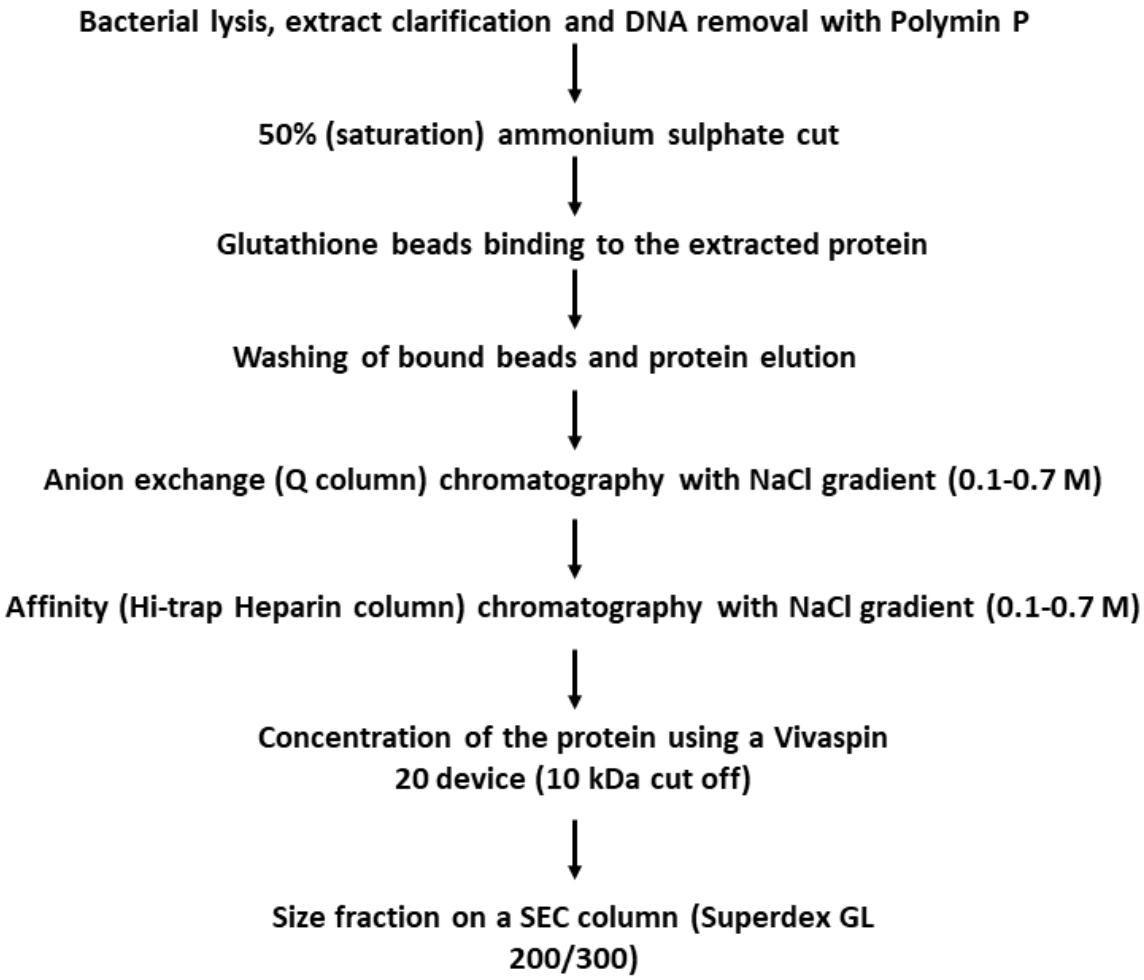


Figure 20. Summary of GST hPIF1 helicase domains purification steps.

2.5 Protein quantification

FL hPIF1, purified free from tags, was quantified by Abs_{280}, using the calculated molar extinction coefficients (ε FL-hPIF1: 38960) and the Beer-Lambert law ($A= \varepsilon c l$). The protein was denatured in the presence of 7 M guanidine hydrochloride and heated to 55°C for 5 mins before reading A_{280} in a quartz cuvette.

2.6 Radiolabelled substrates

2.6.1 Partially single- and double-stranded substrate preparation

A partially duplex and single-stranded substrate for helicase assays was generated from two partially complementary oligonucleotides 5'-GGGTACCGAGCTCGAATTCG (PSB) and 5'$(T)_{55}$-CGAATTCGAGCTCGGTACCC (PST55) giving a substrate with a 20 bp duplex and 55-base 5' T tail. First, the short PSB oligonucleotide was radiolabelled with ^{32}P using T4 polynucleotide kinase and γ [^{32}P]-ATP in a 5 µL reaction with 50 pmol PSB, 1.5 µL of γ [^{32}P]-ATP (6000 Ci/mmol), 0.5 µL PNK (1 U/L), 0.5 µL X10 PNK buffer (New England Biolabs), 2 µL H_2O and incubated at 37°C for 1h. After, the reaction was incubated for 5 minutes at 95°C followed by 1 min on ice. 0.5 µL of the reaction was added in 24.5 µL of stop buffer (10mM EDTA, 10% SDS) while the remaining 4.5 µL was added to 2 µL of X10 annealing buffer (100 mM Tris-Cl pH 8.0, 10 mM EDTA, 1 mM NaCl), 3 µL 80% (v/v) glycerol, 10 µL H_2O and 0.5 µL 100 µM PST55 in a final volume of 20 µl. DNA strands were annealed by heating the reaction in a boiling water for 5 mins followed by cooling down. The substrate was purified on an 8% polyacrylamide gel (19:1 acrylamide: bis-acrylamide) with 1 X TBE running buffer (89 mM Tris, 89 mM boric acid, 2 mM EDTA). The gel was exposed to film and the developed film was used as a template to cut the substrate band from the gel. The substrate was eluted in elution buffer (100 mM NaCl, 1 mM EDTA, 10 mM Tris-Cl pH 8.0) overnight (4^{o}C) after crushing the gel slice.

2.6.2 Unimolecular G4 DNA substrate preparation

The sequences of the G4 substrates and corresponding mutated sequences used are summarised in a Table 3. Substrates were labelled with ^{32}P using thermostable polynucleotide kinase and γ [^{32}P]-ATP in a 10 µL reaction with 10 µM oligonucleotide, 1.5-3 µL of γ [^{32}P]-ATP (6000 Ci/mmol), thermostable PNK (Prokazyme, 1 U/L), 1 X PNK buffer, 0.1 mg/mL BSA and 4.5% w/v of 13% PEG 8000 and incubated at 75°C for 3h. 0.5 µL of the labelled oligonucleotide was added to 24.5 µL of stop buffer (1% SDS, 10 mM EDTA) as reference for quantification (as section 2.6.1). The remaining 9.5 µL of labelled oligonucleotide was combined in a 50 µL reaction with 5 µL of 10X G4 buffer (1.5 M KCl, 100 mM Tris pH 8, 10 mM EDTA),

7.5 µL 80% (v/v) glycerol and 27 µL H_2O and heated in boiling water for 5 minutes followed by cooling down in the fridge with stirring until the temperature reached < 25°C. The substrates were purified on an 12% polyacrylamide gels (19:1), with 1X TBE running buffer (89 mM Tris, 89 mM boric acid, 2 mM EDTA) and 7.5 mM KCl. Substrates were recovered from the gel as described above (section 2.6.1) and eluted in 1X G4 buffer overnight after crushing the gel slices.

In the unwinding assay, the unimolecular G4 DNA substrate was synthesized starting from two partially complementary radiolabelled oligonucleotides: 5'-T_{55}-CGAATTCGAGCTCGGTACCC and 5'-GGGTACCGAGCTCGAATTCG-$(C)_{30}$. The oligonucleotides were annealed at 250 µM, 3h at 60°C; labelling and purification follow as above in section 2.6.1.

2.6.3 Substrates recovery and quantification

The substrates were recovered by filtration through a Costar Spin-X centrifugal filter (0.22 µm cellulose acetate membrane). The quantification was made by creating a grid on DE81 anion exchange paper and spotting 1 µL of counting control to determine specific activity of the ^{32}P-labelled DNA (sample in stop buffer, see above section 2.6.1) and 1 µl of recovered substrate each in three squares. The paper was washed 3 times in a 0.5 M Na-phosphate pH 7.0, then 30 seconds in 70% EtOH and the finally in absolute EtOH before exposure of the filter for ~5-10 minutes to a phosphor imaging screen. The screen was read on a Fujifilm FLA-3000 series phosphor imager and the data collected. The ‘specific activity’ of the known amount of spotted primer was calculated and then the concentration of the substrate in reference to this value.

2.7 Helicase assay for FL hPIF1

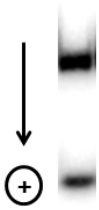
FL hPIF1 was diluted in protein dilution buffer (20 mM Tris pH 7.5, 150 mM NaCl, 2.5 mM DTT, 5% (v/v) glycerol, 1 mM PMSF, 1mg/mL BSA) first to a final concentration of 1 µM before establishing an appropriate dilution series in the same buffer (10X stocks for addition to the reaction). The optimal reaction buffer conditions

were determined to be 20 mM HEPES pH 7.5, 2.5 mM DTT, 5 mM ATP, 5.5 mM $MgCl_2$, 5 nM unlabelled PSB (see above for sequence, section 2.6.1) competitor, 1 mg/mL BSA, 0.1 nM substrate. Reactions were performed at 37°C for 15 minutes and stopped with 1/4th volume of 5 X stop buffer (120 mM EDTA, 0.25% (w/v) SDS, 60% (v/v) glycerol, 0.1% (w/v) bromophenol blue), 1 µM poly T55 oligonucleotide competitor). The strand displacement control was a boiled reaction (5 min at 95°C followed by incubation on ice for 1 min) with no protein added and the negative control received no protein. 20-40 µL of samples were loaded on a polyacrylamide gel (8% acrylamide (19:1 acrylamide to Bis-acrylamide), 0.25X TBE (89 mM Tris, 89 mM boric acid; 2mM EDTA), 0.05% (w/v) SDS gel and running buffer) and run at 8 V/cm until the bromophenol blue had migrated 4 cm. The gel was dried on 3 MM paper and exposed to a phosphor imaging screen.

2.7.1 Screening inhibitors compounds

All screens were performed using 10 nM FL hPIF1. The compounds were dissolved in DMSO and added to reactions to give a concentration from 1 to 4 mM according to the purpose of the assay later performed. Helicase assays were performed as described in section 2.6. IC_{50} measurements were determined using a doubling dilutions series of compound in DMSO. The reference reaction (positive control) contains FL hPIF1 and DMSO at the same concentration as the test reactions. The unwinding percentage was calculated by quantification of substrate and product obtained by phosphor imager analysis.

The unwinding activity of each sample was calculated as follows and represented in Figure 21: every reaction runs on an acrylamide gel and resolves in two bands. The first (N) represents the native substrate (partially double- and single-stranded DNA) and the second (D) denatured substrate (single-stranded DNA. The fraction dissociated (%) was calculated as in the formula (a), through which is possible to quantify the relative fraction unwound in the reaction, relative to the positive control (formula b). Lastly, we calculate the % of inhibition to quantify the power of the



$$\text{FRACTION DISSOCIATED (\%)} = \frac{D}{N+D} \quad (a)$$

RELATIVE UNWOUND (RU): fraction dissociated in the presence of specific inhibitor/ Positive control (DMSO+ protein)

$$\text{RU: } \frac{D_{fraction}}{N_{fraction} + D_{fraction}} \Big/ \frac{D_{pos\ cont}}{N_{pos\ cont} + D_{pos\ cont}} \quad (b)$$

$$\text{\% inhibition of inhibitor (I)} = 100\% - RU \quad (c)$$

Figure 21. Schematic representation of the calculation made to quantify the helicase assays. A detailed description is in the text above.

potential small molecule inhibitor considered.

2.7.2 IC_{50} determination

IC_{50} values were determined as described in section 2.6 for FL-hPIF1, in a compound titration with a variable starting concentration (from 4 mM to 1 mM, according to the purpose of the assay) in a doubling dilution series of 9 points and a protein concentration of 10 nM. The data were then analysed as described in section 2.7.1 and an IC_{50} curve was obtained in GraphPad prism.

2.8 Helicase assay on G4 substrates

For G4 DNA unwinding assays, proteins were diluted in a dilution buffer as described in section 2.7, but substituting NaCl with 150 mM KCl and adding 0.1% NP40-Alt. 200 µL reactions (20 mM HEPES pH 7.5, 2.5 mM DTT, 5 mM ATP, 5.5 mM $MgCl_2$, 5 nM unlabelled competitor DNA oligonucleotide (PSB), 1 mg/mL BSA, 0.02% NP40-Alt, 2.5 mM labelled DNA substrate) were incubated with the 0.2 nM G4 DNA substrate and protein for 15 minutes at RT and stopped with 50 µL sarkosyl stop buffer (0.1 M EDTA, 10 mM Tris pH 8, 0.25% w/v Sarkosyl, 50% (v/v) glycerol, 0.05 % w/v bromophenol blue). Boiled samples were performed as above (section 2.7) and 20 µL of reaction were loaded on an 8% acrylamide gel (19:1 acrylamide to Bis-acrylamide), 0.25X TBE (89 mM Tris, 89 mM boric acid; 2mM EDTA), 7.5 mM KCl, 0.0875% Sarkosyl. The gel was run at 8 V/cm until the bands reached 8 cm and afterwards it was dried on 3 MM paper and exposed to a phosphor imaging screen.

2.8.1 Modulation of cofactors on hPIF1 helicase activity on G4 DNA

The modulation of hPIF1 helicase activity by nucleotide cofactors was investigated in the helicase assay on a G4 DNA substrate. Specifically, 1.5 mM $MgCl_2$, 1mM ADP/1.5 mM Mg^{2+}, 1 mM ATP/1.5 mM Mg^{2+}, 1 mM AMP·PNP/1.5 mM Mg^{2+} and ADP·MgF_4^- formed in the presence of 1 mM ADP, 1.5 m Mg^{2+} and 35 mM NaF. The magnesium salt used was $MgCl_2$.

The modulation of hPIF1 binding activity by nucleotide cofactors was investigate in the Mcay assay, in the same conditions as above.

2.9 Bioinformatic analysis

The two PIF1 atomic models from human (pdb: 6hpt) and *Termus oshimai* (pdb: 6s3e) were aligned and analysed in PyMOL. To evaluate the level of conservation of the protein domains of interest, sequences were aligned with Clustal Omegna and ESPript 3.0 [162]. Sequences were retrieved from UniProt (https://www.uniprot.org/) and relevant eukaryotic, bacteria and yeast species were chosen including model

organisms and species with PIF1 crystal structures present in the PDB (https://www.rcsb.org/).

2.10 GST "McKay-type" assay

The GST "McKay" pulldown assay [40] is described in Figure 17. The GST-tagged protein is mixed with radio-labelled DNA substrates and incubated with glutathione-S-transferase beads. The bound DNA substrates are eluted in SDS stop buffer and analysed on a poly acrylamide gel.

2.10.1 Optimisation reaction conditions

The optimal conditions used for the assay are as follows: A 10 µl final volume of beads per reaction was used comprising of 5 µL of glutathione sepharose and 5 µL of Sephadex G15. The beads were blocked overnight in a blocking buffer (20 mM HEPES pH 7.2, 100 mM KCl and 1 mg/mL BSA) with 1 µL of 100 µM oligo 5'-$(T)_{55}$ per mL of buffer, 2.5 µg of plasmid competitor per mL of buffer and 2 mM $MgCl_2$. Blocked beads were then equilibrated in 10 vols wash/protein dilution buffer (20 mM HEPES pH 7.2, 100 mM KCl, 0.1 mg/mL BSA, 5% v/v glycerol, 0.1% NP-40 (Alt), 5 mM DTT).

GST hPIF1 proteins (GST FLhPIF1, GST hPIF1 HDΔC26 and GST hPIF1 HD206-END) were diluted to the required concentrations in wash/protein dilution buffer and used at 12.5 nM to 200 nM in reactions. The DNA substrates were diluted in a 1X G4 DNA buffer (150 mM KCl, 10 mM Tris pH 8, 1 mM EDTA) to a concentration of 10 nM. The optimal reaction buffer conditions were determined to be at 20 mM HEPES pH 7.5, 100 mM KCl, 1 mg/mL BSA, 5% v/v glycerol, 0.1% NP-40 Alt, 0.25 nM substrate and 0.125 ng/µL plasmid competitor.

The binding reaction (200 µL) was incubated with 10 µl of beads for 30 minutes at RT, with end-over-end mixing. The reactions were washed with 3X in 1mL wash/protein dilution buffer and after discarding the supernatant, the final traces of buffer were removed with a pipette and duck-billed tip (0.2 mm). Bound products were recovered for 20 minutes in 50 µl of McKay stop buffer ((10 mM Tris-Cl pH 8.0,

150 mM KCl, 2% Ficol, 11 mM EDTA, 0.1 % w/v SDS, 1 μM oligo $d(T)_{55}$, 0.0125 % w/v bromophenol blue and xylene cyanol). 20 μL of samples were loaded on a polyacrylamide gel (8% acrylamide (1:19), 0.25 XTBE (89 mM Tris, 89 mM boric acid; 2mM EDTA), with 7.5 mM KCl in gel and running buffer) and run at 8 V/cm until the bromophenol blue had migrated 8 cm. 25% input substrate mix was load on the gel along with 50% input of individual substrates as markers. The gel was dried on 3 MM paper and exposed to a phosphor imaging screen.

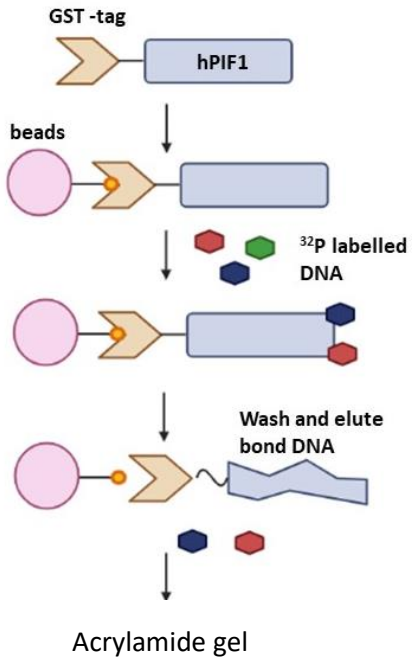


Figure 22. Schematic representation of GST "McKay-type" pulldown assay. The GST-tagged protein is incubated with glutathione beads (pink) and radiolabelled DNA substrates (ssDNA and G4 DNA, indicated with red, blue and green). After binding and washing, beads are recovered, and the DNA substrate is eluted with SDS stop buffer. The recovered products are analysed on an 8% polyacrylamide/TBE/KCl gel and visualized by phosphor imaging.

Substrate	Sequence
$T_1G_4T_4$	5'- TGGGGTTTTGGGGTTTTGGGGTTTTGGGG
M_$T_1G_4T_4$	5'- TAGGGTTTTGAGGTTTTGGAGTTTTGGGA
T_{50}	5'- TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTTTTTTTT-
T_{40}	5'- TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-
T_{30}	5'- TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-
T_{20}	5'- TTTTTTTTTTTTTTTTTTTT-
T_{10}	5'- TTTTTTTTTT-
$T_0G_4T_4$	5'- GGGGTTTTGGGGTTTTGGGGTTTTGGGG
M_ $T_0G_4T_4$	5'- AGGGTTTTGAGGTTTTGGAGTTTTGGGA
T_1_Myc 1245	5'- TGGGAGGGTTTTTAGGGTGGGGA
T_1_M_Myc 1245	5'- TGGGAAGGTTTTTAGAGTGGAGA
Telo 22	5' -AGGGTTAGGGTTAGGGTTAGGG
M_Telo 22	5' – AGGGTTAAGGTTAGAGTTAGGA
$T_1G_3T_4$	5'- TGGGTTTTGGGTTTTGGGTTTTGGGG
M_$T_1G_3T_4$	5'- TGGGTTTTAGGTTTTGAGTTTTGGA
$T_1G_4T_3$	5'- TGGGGTTTGGGGTTTGGGGTTTGGGG
M_$T_1G_4T_3$	5'- TAGGGTTTGAGGTTTGGAGTTTGGGA

Table 3. Names and sequences of substrates used in the "McKay-type" assay. The underlined nucleotides indicate where a base substitution has happened to obtain a 'mutated' form of the original DNA sequence.

2.11 DMS (dimethyl sulphate) footprinting

For radiolabelled G4 DNA substrates methylation [163]: in 200 µL of 50 mM cacodylate buffer, 1 mM EDTA, pH 7.0, was added 10 µL of 10% DMS (in EtOH), vortexed for 20 seconds and incubated at RT for 5 min. Reactions were terminated

with 50 µL of 1.5 M NaOAc pH 7.0 and 1 M 2-mercaptoethanol and DNA precipitate with 1.25 mL of chilled ethanol (15 minutes at 16 000 x *g*, 4ºC). The pellets were dissolved in 200 µL of 0.3 M NaOAc, 1 mM EDTA, pH 8 and precipitate with 2.5 vol. of EtOH. The new pellets so obtained were washed in 80% and then in 100% ethanol and vacuum dried.

For piperidine cleavage: the pellets of methylated DNA were dissolved in 100 µL of 1 M piperidine and incubated at 90ºC for 30 min. The contents were recovered to the bottom of cooled tubes by quick centrifugation and 1.3 mL of butanol was added, the tubes vortexed for 10 seconds, and spun at 16 000 x *g*. The pellets were dissolved in 100 µL of 1% SDS, transferred to a new tube, and precipitated again with butanol. Afterwards, the pellets were washed with 80% and 100% ethanol and dried before being dissolved in 50% formamide loading buffer (98% formamide, 1 mM EDTA, 0.025% (w/v) bromophenol blue). Reactions were run on a 12% acrylamide (19:1) denaturing PAGE gel (sequencing gel, 1X TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA, 8 M urea) and quantified by Phosphorimager analysis.

2.12 ATP*ase* assay

hPIF1 ATP*ase* activity was determined in 20 mM HEPES pH 7.2, 2 mM DTT, 5 mM ATP, 0.0125 µM (γ^{32} P) ATP (6000 Ci/mmol), 0.1 mM NP40-Alt, 0.1 mg/ml BSA, 45 mM KCl, 5.5 mM $MgCl_2$, 200 nM DNA. The protein was diluted in 200 mM Hepes pH 7.5, 150 mM KCl, 2.5 mM DTT, 5% (v/v) glycerol, 0.1 mM BSA, 0.1 mM PMSF. Reactions were incubated at 22°C in a time course up to 60 minutes and the ATP released measured by the charcoal binding method [164].

2.13 Figures

The pictures obtained in this work were obtained with Power Point and Biorender (https://www.biorender.com/) where icons and templates were selected to produce them. SnapGene was also used to create a plasmid representation. PoyMOL, Schrödinger (https://www.pymol.org) is the software used to obtain the molecular graphic pictures.

3 RESULTS

3.1 PIF1 constructs for biochemical analysis

Previous hPIF1 investigations in the Sanders lab focused on the analysis of truncated forms of hPIF1 encompassing the helicase domain (HD) or “helicase core” (Figure 23). A detailed description of the hPIF1 HD structure is given in the introduction section 1.7.4. Although it is now possible to produce the full-length protein (Saba Dehghani Tafti, PhD thesis 2019, University of Sheffield), an optimized protocol for producing milligram quantities of full-length PIF1 protein (FL hPIF1) free from expression tags was implemented here. This protein is used for helicase and ATP*ase* assays. In addition, GST fusion constructs of FL hPIF1 and the helicase core were produced for “GST pull-down/ McKay” experiments with radiolabeled DNA substrates

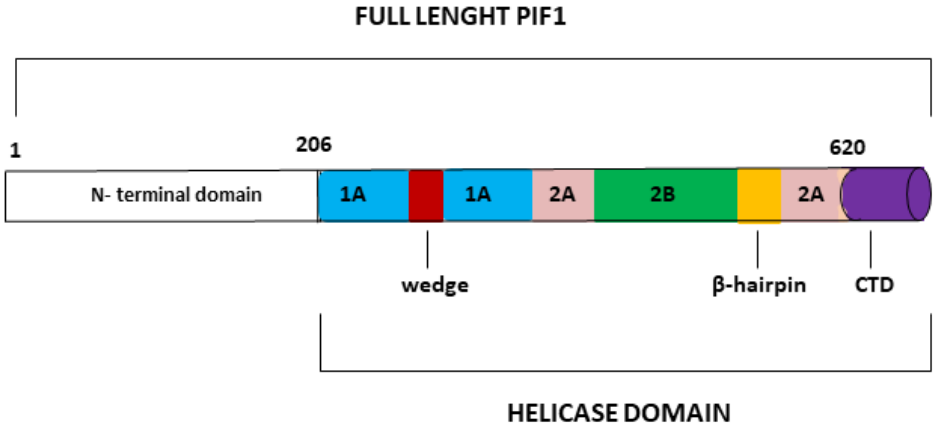


Figure 23. Cartoon of human PIF1. The picture shows the domains organisation with the helicase core domain (residues 206-620) shown in detail.

3.1.1 Production of full-length, tag-free, hPIF1 (FL hPIF1)

Previously, the expression of hPIF1 proteins at low temperatures was key to producing large quantities of helicase domain proteins from *E. coli* [135]. Trialing of various tags in the lab demonstrated that a His-SUMO tag produced the greatest yield of soluble protein at low-temperature expression conditions, 2 days at 8°C using *E. coli* BL21(DE3) ArcticExpress™ cells. First, a His_6 tag was used, but further optimization of purification procedures resulted in the adoption of a His_{12} tag for these studies.

From the FL hPIF1 expression, approximately 47 g wet weight of cells were harvested from 18 L of bacteria culture. The chromatograms with the SDS-PAGE analysis are shown in Figure 24. Note, the Arctic chaperone Cpn10/60 is a significant contaminant present after the first His-trap step (Figure 24 A). Complete cleavage of the SUMO tag was achieved without notable degradation of FL hPIF1 (Figure 24 B). Degradation had been a particular problem when other site-specific proteases had been used previously (TEV, thrombin), possibly due to the largely unstructured nature of the N-terminal domain of hPIF1 (Figure A1, appendix). The final SEC step resulted in a high yield of purified protein (Figure 24 C). ATP*ase* Walker A and B box mutants purified similarly demonstrated the absence of helicase and ATP*ase* activity (see section 3.6.6). The procedure yielded approximately 0.4 mg of FL hPIF1 per gramme of wet cells.

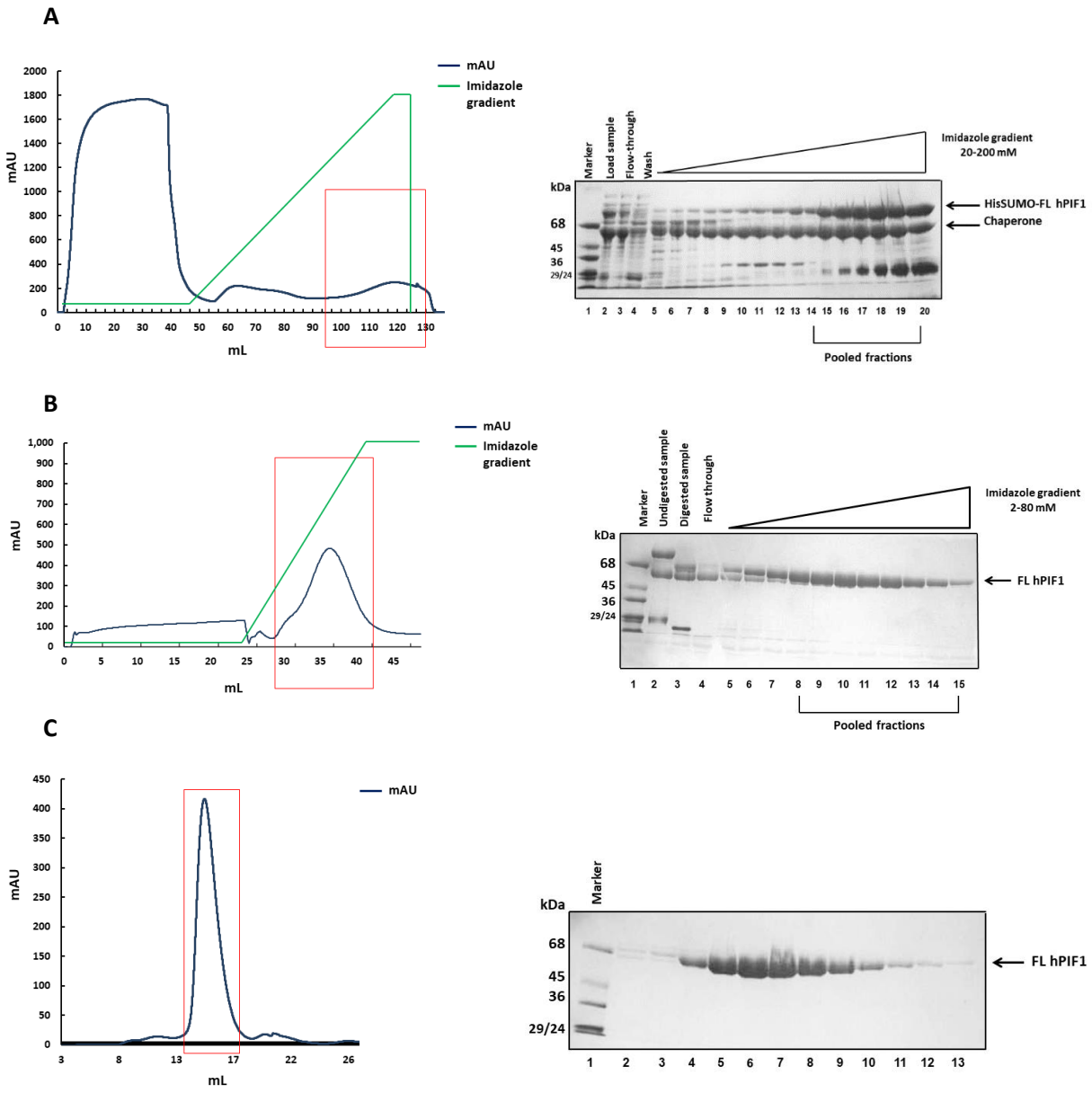


Figure 24. Chromatograms and SDS-PAGE analysis of FL hPIF1 purification. (A) First His-trap column. The imidazole gradient 20-200 mM is indicated in green. In the SDS-PAGE gel: lane 1, marker; lane 2, loaded sample; lane 3, flow-through; lane 4, wash; lane 5 to lane 20 eluted fractions. (B) Second His-trap (2-80 mM imidazole gradient in green) purification after digesting the His-SUMO tag off the protein. In the SDS-PAGE gel: lane 1, marker; lane 2, undigested sample; lane 3, digested sample; lane 4, flow through; from lane 5 to 15 eluted fraction. The expected size of FL hPIF1 is ~68 kDa, as indicated (C) SEC column. In the SDS-PAGE gel: lane 1, marker; from lane 2 to lane 13, eluted fractions.

3.1.2 GST FL hPIF1 purification

The GST hPIF1 proteins used in this work were GST FL hPIF1 (FL hPIF1), GST hPIF1 HDΔC26 (hPIF1 HD) and GST hPIF1 HD 206-END (hPIF1 HD-E) (Figure 25). The GST-hPIF1 expression construct using the wild-type hPIF1 sequence has been described previously [145], but never gave protein of high quality and purity following the removal of the GST tag with thrombin. However, it was considered that the quantity and quality would be adequate to perform pull-down assays, and it was also re-cloned here to have the *E. coli* expression optimised ORF using a pET15b plasmid, to obtain a higher yield of protein. All constructs were purified using the same procedure.

GST FL hPIF1 proteins were expressed in BL21(DE3) cells. Critical for optimal expression was an induction at an absorbance (A) of the culture of 0.1. Expression was performed at 22°C for 6-8 hours, in contrast to the low temperatures and extended incubations for the expression of proteins in ArcticExpress™ cells (see above). The harvest and the extraction of the protein were according to standard procedures, but the column purification steps were optimised. The chromatograms obtained for the protein purification steps and the SDS-PAGE analysis are shown in Figure 26. The procedure yielded approximately ~ 0.2 mg of GST FL hPIF1 per gramme of wet cells.

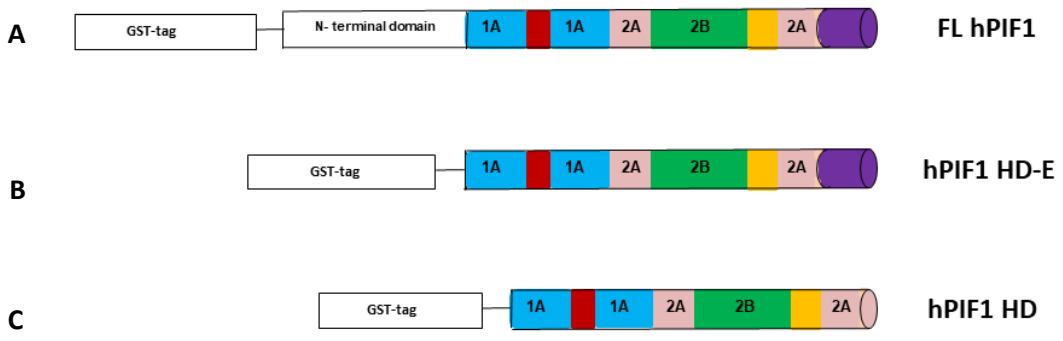


Figure 25. GST hPIF1 constructs used in this work. (A) GST FL hPIF1 construct; (B) GST hPIF1 HD206-END construct; (C) GST hPIF1 construct. On the side of each construct are indicated the abbreviations that will be used in this work.

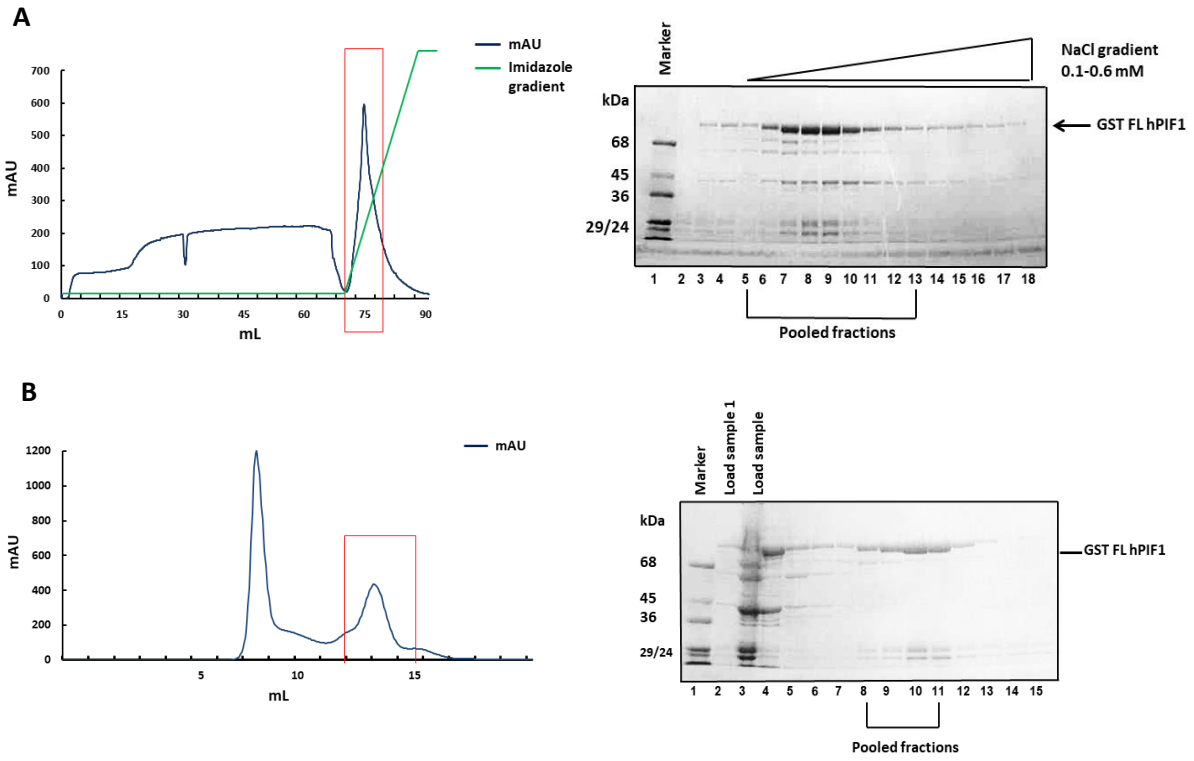


Figure 26. Chromatograms and SDS-PAGE analysis of GST FL hPIF1 purification. (A) Ion exchange chromatography after protein extraction and pull-down. NaCl gradient, green, 0-100% NaCl. In the SDS-PAGE: lane 1, marker; from lane 2 to lane 4 flow through; from lane 5 to lane 18 eluted fractions. The fractions of interest ("Pooled") are indicated below the gel. The GST FL hPIF1 is indicated on the side (~ 90 kDa). (B) SEC column chromatography (Superdex 200 10/300 GL). In the SDS-PAGE: lane 1, marker; lane 2, unconcentrated load protein; lane 3, concentrated load protein; lane 4 first peak sample (high MW fraction, ~7 ml); from lane 5 to lane 15, eluted fractions (~10-15 ml elution). The faction collected, concentrated, and stored as indicated as "pooled fractions".

3.1.3 GST hPIF1 helicase domain purification

Expression and extraction of the helicase domain proteins hPIF1 HD and hPIF1 HD-E were performed as described for FL hPIF1. The SDS-PAGE analysis is shown in Figure 27 for hPIF1 HD-E. The procedure yielded approximately ~0.28 mg of hPIF1 HD-E per gramme of wet cells and ~ 0.11 mg of hPIF1 HD per gramme of wet cells.

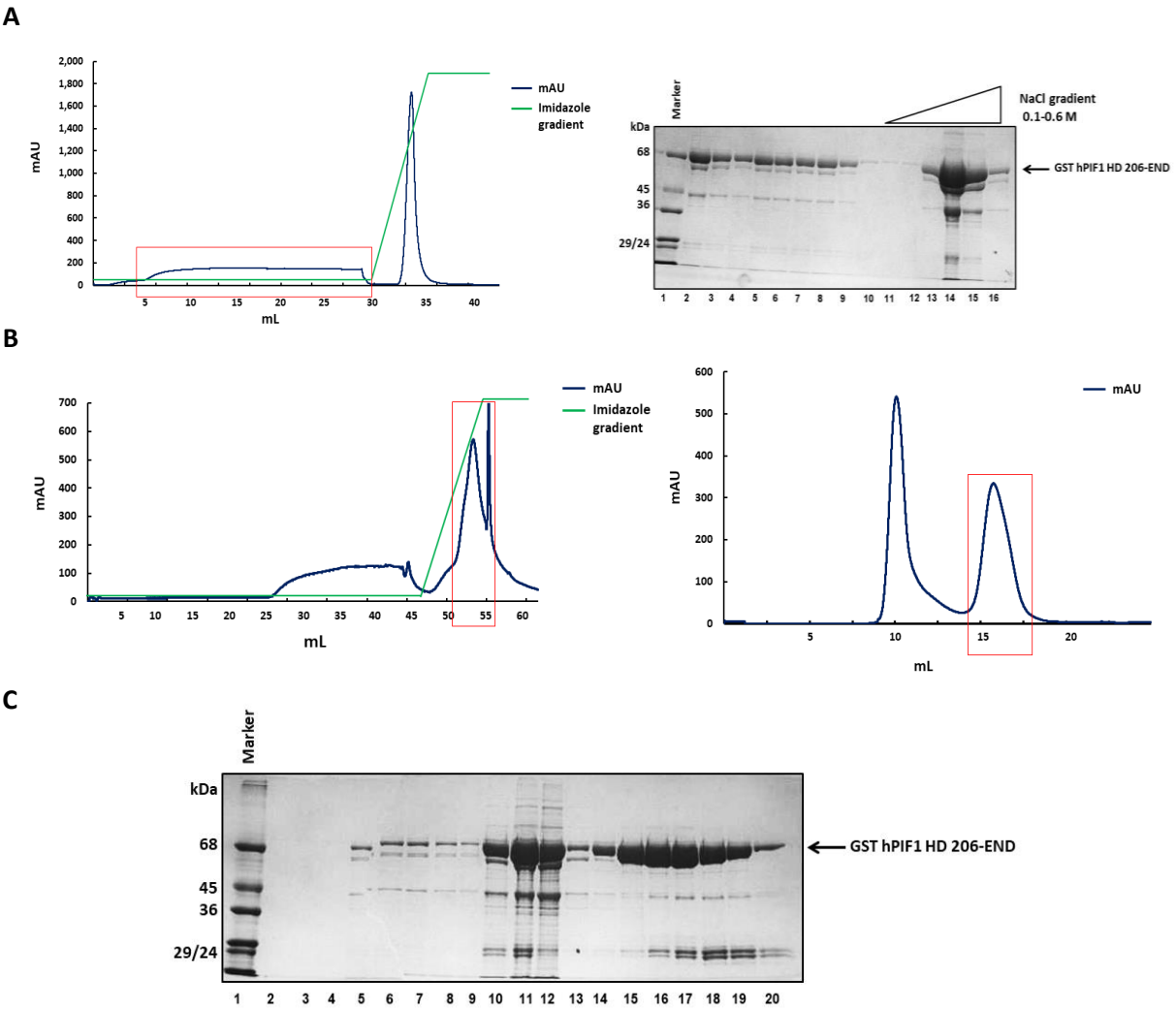


Figure 27. Chromatograms and SDS-PAGE of GST hPIF1 HD 206-END protein purification. (A) Anion exchange chromatography (Q column) after protein extraction and pulldown. Gradient is indicated in green (0-100% NaCl). In the SDS PAGE gel: lane 1, marker; lane 2, eluted protein from glutathione (GSH) sepharose column; lane 3, eluted protein from GSH sepharose column diluted in 20 mM Tris-Cl buffer pH 8.0; from lane 4 to lane 16 eluted fractions. Flow through fraction from lane 4 to lane 8 (highlighted in the red box in the chromatogram were collected for the next column step. (B) Heparin affinity chromatography and gel filtration chromatography; in both graphs the red box indicated the fractions collected for the next step in the purification. (C) In the SDS PAGE gel, lane 1, marker; from lane 2 to lane 9, flow through from the heparin column; lane 10, sample of the pooled fractions from the peak; lane 11, concentrated heparin peak applied to SEC; lane 12, high molecular weight material (discarded) in the void of the gel filtration (~ 8-10 ml elution volume, first peak in B, left graph), from lane 13 to lane 20, eluted fractions (red box).

3.1.4 Purification of GST hPIF1 mutants

A series of mutants of GST hPIF1 HD 206-END were expressed and purified, as above (section 3.1.3). However, their yield was much lower than wild type, ~ 0.02 mg per gramme of wet cells. The reasoning behind the selection of the mutants is reported in section 3.5, where they were employed in DNA binding assays. While for some proteins with amino acid substitutions, the expression and production were successful, for others (such as R290A) the yield obtained in the intermediate purification steps was so low it was not possible to proceed. Figure 28 shows 5 µg of the successfully purified variant proteins run on an SDS PAGE gel after determining their concentration.

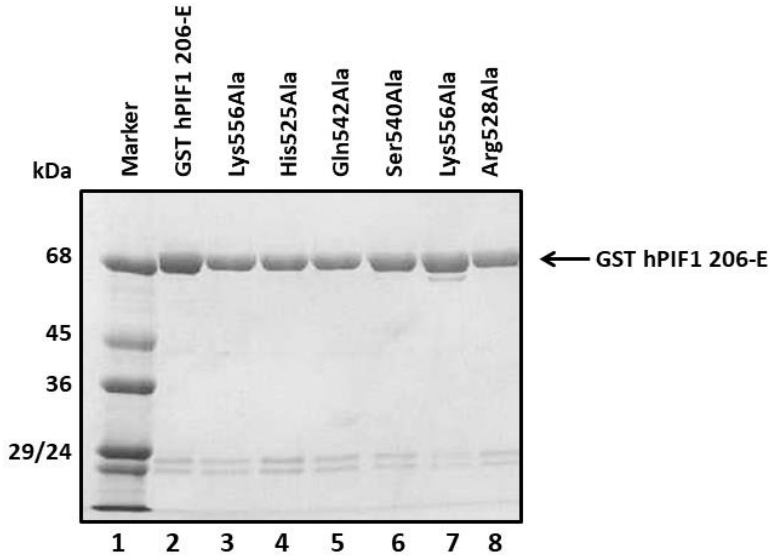


Figure 28. SDS-PAGE gel of GST hPIF1 mutants. The concentration and purity of variant proteins was check by loading 5 µg of protein on a SDS PAGE gel using the wt GST hPIF1 HD 206-END as control. Lane 1 marker; lane 2, GST hPIF1 HD 206-END; lane 3, Lys556Ala; lane 4, His525Ala; lane 5, Gln542Ala, lane 6, Ser540Ala, lane 7, Lys556Ala (express in BL21(DE3) Arctic cells); lane 8, Arg528Ala. All the variant proteins migrate at the same level (~ 68 kDa); the difference in the intensity of the bands could be due to variable levels of purity. In lane 7, the faint band underneath the main one could be contaminating Chaperon Cpn10/60. In the case of K556A, the use of Arctic cells trialled to potentially increase protein yield.

3.1.5 Protein quantification

Purified FL hPIF1 was quantified by Abs_{280} spectroscopy and employing the calculated molar extinction coefficient. Because of the reduced yield and purity, the hPIF1 HD-E fusion protein concentrations were determined using the Bradford protein assay (BioRad). The concentrations measured are reported in Table 4.

	Extinction coefficient (ε_M)	**Molar concentration**	**mg/mL**
FL-hPIF1	38960	33.1 µM	2.23 mg/mL
GST hPIF1 HD 206-END		73.4 µM	5.10 mg/mL
GST hPIF1 HDΔC26		37.7 µM	2.62 mg/mL
GST FL hPIF1		18.5 µM	1.7 mg/mL
GST-STOP		1200 µM	28.9 mg/mL
GST Lys556Ala		8.3 µM	0.6 mg/mL
GST His525Ala		6.2 µM	0.45 mg/mL
GST Gln542Ala		6.2 µM	0.45 mg/mL
GST Ser540Ala		12.1 µM	0.87 mg/mL
GST Arg528Ala		12.5 µM	0.9 mg/mL
GST Lys556Ala/His525Ala		23 µM	1.66 mg/mL

Table 4. Concentration of the purified proteins. The concentrations of the proteins are reported both in mg/mL and in µM. For FL-hPIF1 the molar extinction coefficient (Abs_{280}) is given.

3.2 Biochemical analysis of G4 DNA binding

3.2.1 Production of G4 DNA substrates

hPIF1 binds and unwinds G quadruplex DNA [146]. Previous analysis had been done using synthetic tetramolecular G4 DNA substrates, but the natural substrates of the enzyme are most likely to be unimolecular.

To investigate hPIF1 binding activity, several unimolecular G4 DNA structures were synthesized and end-labelled with ^{32}P as described in section 2.5.2. Substrate $T_0G_4T_4$ was employed as a substrate that is uniquely G4 in character. $T_0G_4T_4$ is based on the *Oxytricha nova* telomeric sequence [165] and forms 4 G tetrads from four dGGGG run with intervening dTTTT segments. A single-strand mutant control,

M_$T_0G_4T_4$, was also synthesized with an A instead of a G in each tetrad, which makes the DNA unable to form G quartets and fold into G4 DNA (see Table 8).

As shown in Figure 29 A, $T_0G_4T_4$ folded in the presence of 150 mM K^+ ions, is resolved as three species in an acrylamide gel, one major species (S1) accounting for ~65% of the folded DNA and two minor and poorly resolved faster moving forms (indicated as S2 and S3 in Figure 27 A, lane 2). The major folded form S1 was used for analysis. In the gel, other bands of high molecular weight (lower electrophoretic mobility) are visible in the folded $T_0G_4T_4$ lane (lane 2, Figure 29 A). These are likely DNA that has formed multimeric G4 DNA species. The control M_$T_0G_4T_4$ "mock hybridised" in parallel under the same conditions, runs predominantly as a single band of reduced mobility (lane 1) compared to $T_0G_4T_4$.

The unimolecular G4 DNA substrate considered as a 'reference substrate' in the analysis presented here is $T_1G_4T_4$, which is identical to $T_0G_4T_4$ except for the addition of a single 5' dT residue. Like $T_0G_4T_4$, $T_1G_4T_4$ folds into three species, a major species that was used as the reference and two minor faster migrating species (Figure 29 B, lane 3). As observed for $T_0G_4T_4$, unfolded material is also present in lane 3 (Figure 29 B), co-migrating with the mock hybridised mutant control M_$T_1G_4T_4$ (lane 4), and high molecular weight bands likely to be multimeric G4 DNA.

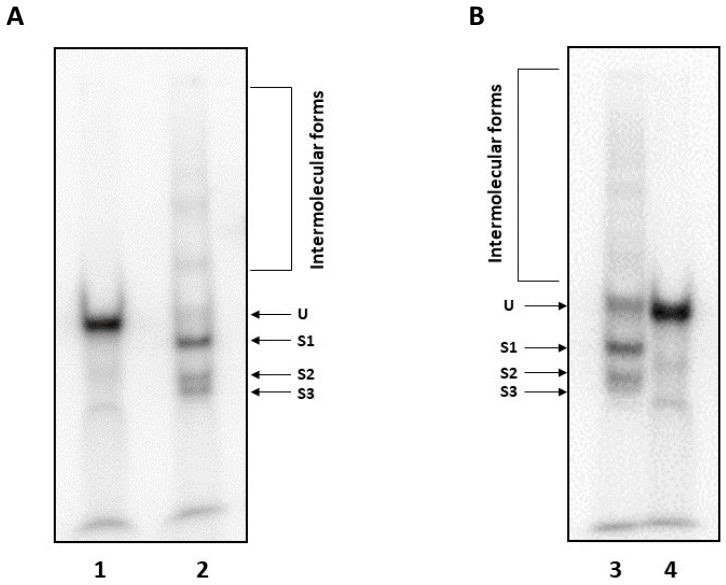


Figure 29. Formation of G4 substrates in 150 mM KCl. (A) Hybridised $T_0G_4T_0$ substrate and its mutant M_$T_0G_4T_4$: lane 1, the "mock hybridised" single stand mutant, M_$T_0G_4T_4$; lane 2, folded forms of $T_0G_4T_4$. The latter resolves as three main species (S1, S2 and S3) but high molecular weight bands are visible too. They likely include some unfolded material (U) migrating at the same level of the mutated species, and multimeric G4 DNA species. (B) $T_1G_4T_4$ substrate and its mutant M_$T_1G_4T_4$: lane 3, $T_1G_4T_4$ resolving as three main species (S1, S2 and S3) with unfolded material (U) and intermolecular G4 forms; lane 4, M_ $T_1G_4T_4$ single strand mutant.

Along with the G4 DNA substrate based on the *Oxytricha* telomeric sequence, two cellular G-quadruplexes, T_1Myc1245 and Telo22, and their single-stranded mutants (M_T_1Myc1245 and M_Telo22) were produced that are unable to fold (G–A substitutions in each tetrad, Figure 30. Table 8). T_1Myc1245 has been identified in the promoter of the Myc oncogene and the folded G4 DNA has three tetrads formed from 4 dGGG repeats. The second physiological G4 structure used is Telo22, formed from the human telomeric repeat, comprising four dGGG runs (forming three G tetrads) with intervening dTTA. Both substrates, T_1Myc1245 and Telo22 have a single ssDNA residue (T or A for Telo22) at the 5' end as in $T_1G_4T_4$. These substrates, end-labelled with ^{32}P, are shown in Figure 30. Telo22 shows a low specific activity of ^{32}P labelling compared to the other G4 substrates used. Also, hybridised Telo22 (lane 3) shows a significant band co-migrating with its mutated form (lane 4), but notably no appreciable multimeric high molecular weight forms formed, in contrast to all other G4 forming sequences used including the Myc sequence (lanes 1 and 2).

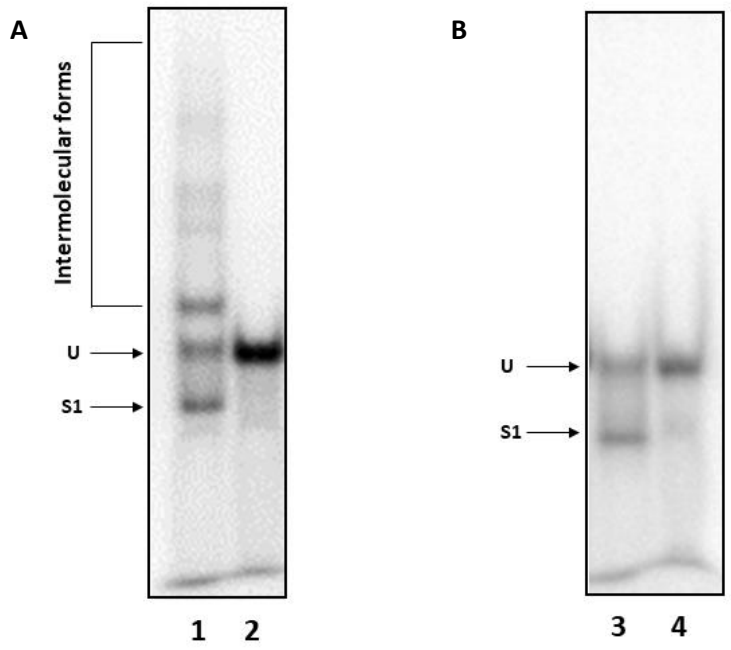


Figure 30. Formation of radiolabelled 'physiological' G4 substrates. (A) T_1Myc1245 (lane 1) resolved as one main folded species, S1, used in the analysis, with unfolded material (U) migrating at the same level of the mutant form and intermolecular material of high molecular weight; lane 2, M_T_1Myc1245, single stand mutant substrate. (B) Folded Telo22 (lane 3) resolving as a single folded form (S1) and unfolded material co-migrating at the level of M_Telo22 in lane 4.

A synthetic variant of the sequences described above was also produced, $T_1G_3T_4$ and $T_1G_4T_3$. $T_1G_3T_4$ contains four dGGG runs with intervening dTTTT segments and a T_1 5' tail residue. $T_1G_4T_3$ is made of four dGGGG runs with the intervention of dTTT segments and a T_1 5' tail. As observed in Figure 31, $T_1G_4T_3$ substrate resolves as two fast-migrating species (S1 and S2), both considered in my analysis.

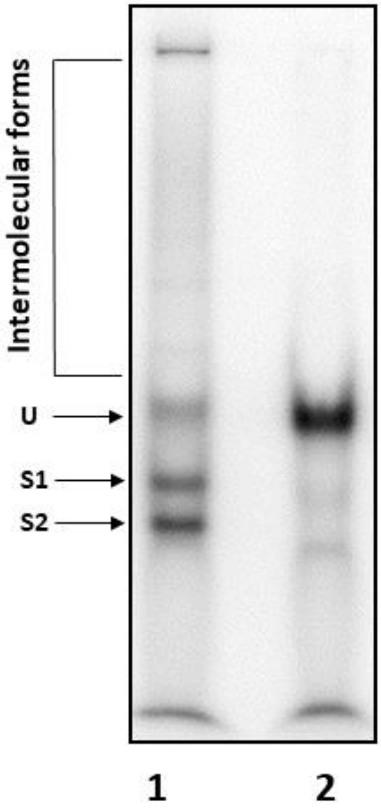


Figure 31. Formation of radiolabelled G4 substrates $T_1G_4T_3$. Lane 1, folded $T_1G_4T_3$ resolved as two main species (S1 and S2) with both considered for the analysis, unfolded material (U) migrating at the same level as the mutated specie and intermolecular G4 material. Lane 2, M_$T_1G_4T_3$ single strand mutant substrate.

To verify the folding of the G quadruplex DNA substrates, DMS footprinting was performed for ^{32}P end-labelled substrates $T_1G_4T_4$, T_1Myc1245 and Telo22, along with their mutant controls. DMS methylates guanine (and adenine) in ssDNA (or dsDNA) but reacts poorly with G residues that are part of a G quartet, *i.e.*, the G residues are protected by the G4 folding. Therefore, different patterns of reactivity between mutant and folded forms are expected to be observed after piperidine cleavage of reacted bases and the resolution of products on a denaturing gel. As with the mutant,

thermally denatured G4 DNA (boiled, low KCl concentration) should react with DMS and subsequent piperidine cleavage.

In Figure 32, N and B indicated native and boiled substrates respectively. Panel A represents the DMS footprint on the $T_1G_4T_4$ substrate and its mutant: lanes 1&2 indicates the footprints of the slower migrating species (S1, Figure 29 B), lanes 3&4 represents the two faster migrating species S2 and S3 combined together (Figure 29 B) and lanes 5&6 indicates the mutant M_$T_1G_4T_4$. Neither folded species (S1 or S2/3 pool) showed significant DMS-guanine reactivity in the G runs, compared to background cleavage products in intervening T residues, confirming quadruplex formation. Curiously, only the boiled substrate in the species 2/3 pool demonstrated a degree of enhanced G reactivity in the boiled control (Figure 32 lane 3 compared to 4). Enhanced DMS reactivity was observed for all G residues in the M_$T_1G_4T_4$ control (lane 5), but unexpectedly the extent of reactivity (mostly lower) was different in the boiled control (lane 6). However, it is possible that this could be partially explained by a low load of material (the full-length product appears less intense at the top of the gel). To an extent, these observations contrast qualitatively with previous and extensive footprint analysis obtained in the Sanders lab for the "G_4T_4" motif embedded within longer oligonucleotides (Appendix, Figure A2). However, they confirm G4 DNA formation, as expected.

Figure 32 panel B shows T_1Myc1245 footprint analysis: lane 1 is the footprint of the native (folded) species showing low reactivity to DMS, confirming G4 formation; its boiled form in lane 2, however, seems to have an unexpected low reactivity too (as for $T_1G_4T_4$, described above). Lanes 3 and 4 represent the mutant control, both lanes show similar behaviour, as expected, with an enhanced DMS-guanine reactivity.

Figure 32 panel C shows the Telo22 analysis: lane 1 is the native form demonstrating low DMS reactivity, as expected, but also the case with its boiled form (lane 2). Lanes 3 and 4 represent M_Telo22 in its native and boiled forms, with an enhanced guanine reactivity to DMS.

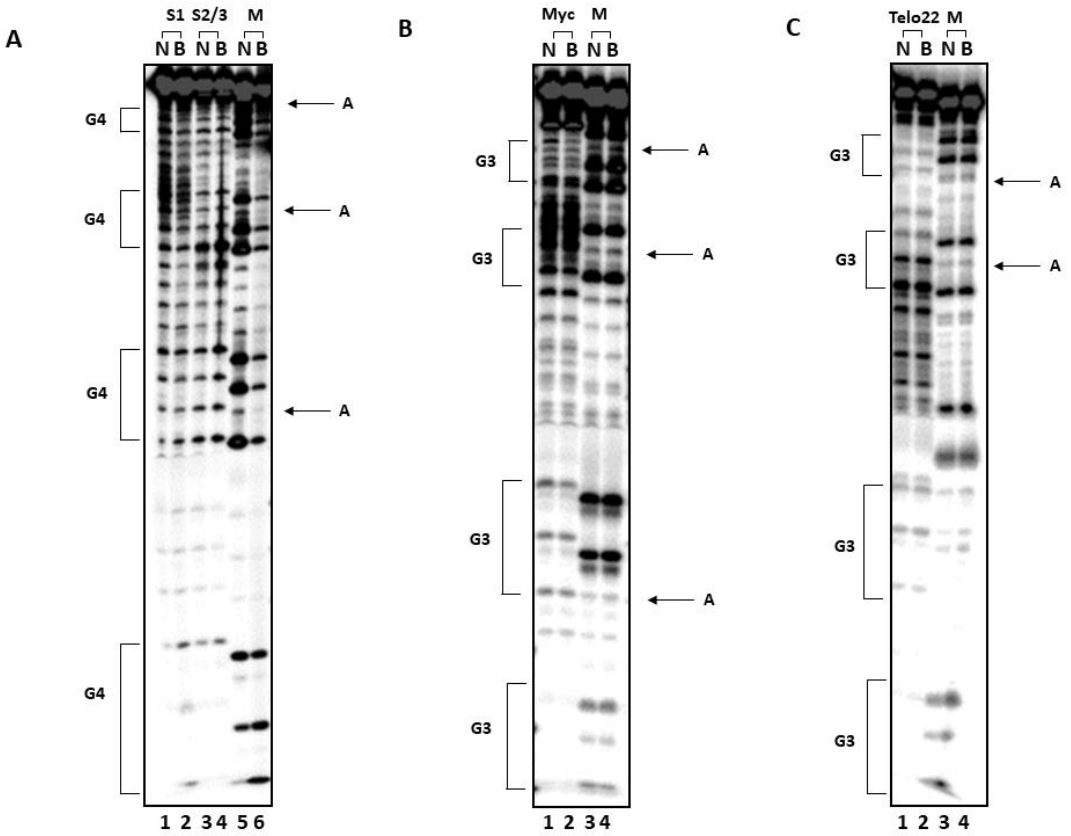


Figure 32. DMS footprint of the unimolecular G4 substrates and their mutants used in this work. Panel A: $T_1G_4T_4$ and M_$T_1G_4T_4$ DMS footprint. Lane 1, native (N) and lane 2, boiled (B) of the main species S1; lane 3&4, native (N) and boiled (B) of the minor species S2 and S3; lane 5, native (N) and lane 6, boiled (B) of the single strand mutant M_$T_1G_4T_4$. On the right side of the panel are indicated the 4 guanines for folding in stacks, on the left the As substituting the Gs in the mutant control are indicted with an arrow. Panel B: T_1Myc1245 and M_T_1Myc1245 DMS footprint. Lane 1, native (N) and lane 2, boiled (B) of T_1Myc1245; lane 3, native (N) and lane 4, boiled (B) of the single strand mutant M_T_1Myc1245. Panel C: Telo22 and M_Telo22 DMS footprint. Lane 1, native (N) and lane 2, boiled (B) of Telo22, lane 3 native (N) and lane 4 boiled (B) single strand mutant M_Telo22. Also, for panel B and C the Gs making the quadruplexes are indicated on the left side of the gel while the As substituting the Gs are indicated with an arrow on the right.

All the substrates, G4 and ssDNA, synthesized and used in the analysis of this work are reported in the table below (Table 5).

SUBSTRATE	SEQUENCE	STRUCTURE	ABBREVIATION
oligo d(T)$_{10}$	TTTTTTTTTT	unfolded	T10
oligo d(T)$_{20}$	TTTTTTTTTTTTTTTTTTTT	unfolded	T20
oligo d(T)$_{30}$	TTTTTTTTTTTTTTTTTTTT TTTTTTTTTT	unfolded	T30
oligo d(T)$_{40}$	TTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTT	unfolded	T40
oligo d(T)$_{50}$	TTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTT TTTTTTTTTT	unfolded	T50
dT(G$_4$T$_4$)$_3$G$_4$	5'-TGGGGTTTTGGGGTTTTGGGGTTTTGGGG	folded	$T_1G_4T_4$
dT(A$_1$G$_3$T$_4$)$_3$A$_1$G$_3$	5'-TAGGGTTTTGAGGTTTTGGAGTTTTGGGA	unfolded	$M\_T_1G_4T_4$
d(G$_4$T$_4$)$_3$G$_4$	5'-GGGGTTTTGGGGTTTTGGGGTTTTGGGGTTTT	folded	$T_0G_4T_4$
d(A$_1$G$_3$T$_4$)$_3$A$_1$G$_3$	5'-AGGGTTTTGAGGTTTTGGAGTTTTGGGATTTT	unfolded	$M\_T_0G_4T_4$
dT(G$_3$)$_4$A$_3$T$_6$G$_1$	5'-TGGGAGGGTTTTTAGGGTGGGGA	folded	T_1Myc1245
dT(G$_2$A$_1$)$_4$A$_3$T$_6$G$_1$	5'-TGGGAAGGTTTTTAGAGTGGAGA	unfolded	M_T_1Myc1245
d(A$_1$G$_3$T$_2$)$_3$A$_1$G$_3$	5'-AGGGTTAGGGTTAGGGTTAGGG	folded	Telo22
d(A$_2$G$_2$T$_2$)$_3$A$_2$G$_2$	5'- AGGGTTAAGGTTAGAGTTAGGA	unfolded	M_Telo22
dT(G$_3$T$_4$)$_3$G$_3$	5'-TGGGTTTTGGGTTTTGGGTTTTGGG	folded	$T_1G_3T_4$
dT(G$_4$T$_3$)$_3$G$_4$	5'-TGGGGTTTGGGGTTTGGGGTTTGGGG	unfolded	$T_1G_4T_3$

Table 5. Summary of the substrates used in the G4 binding analysis. In the table are indicated their chemical names, sequences, structures, and the abbreviations used in this thesis work. The As underlined indicates where a G has been substituted with an A.

3.2.2 Establishment of a GST 'pulldown' McKay-type assay to investigate hPIF1 DNA interactions

Previous attempts in the laboratory to establish a unimolecular G4 DNA binding assay based on the electrophoretic mobility shift assay (EMSA) have been unsuccessful. However, a "McKay-type binding assay was established and investigated further here. The assay employs a GST tag on hPIF1 as an affinity handle for a "pull-down", rather than an antibody interaction used in the original McKay assay [166]. The advantage of this assay is that ^{32}P labelled substrates (see Table 5 for substrates used) can be mixed in binding reactions and different substrates can compete freely for binding under solution conditions. After incubation, complexes are recovered on glutathione sepharose (GSH-sepharose), DNA released with SDS-containing buffer and analysed on a PAGE gel. The assay was established using 10 µL of beads for a 200 µL binding reaction, such that the concentration of immobilised glutathione on the beads is 800-fold greater than the concentration of GST hPIF1 typically used.

A primary issue encountered was background substrate binding in reactions without hPIF1. As summarised in Figure 33, the principal source of the background was identified as substrate binding to the GSH beads, and not the control GST protein (GST STOP) (compare lanes 15-17 with beads alone to 10-14 with GST STOP protein titration). Figure 33 also illustrates the general format used in the experimental scheme, as reported henceforth: lanes 1-3 are marker lanes for the individual substrates used (T50, $T_0G_4T_4$ and M_$T_0G_4T_4$); lane 4 shows 25% of the input mix of substrates. Note, that each substrate is present at molar equivalence, so differences in band intensity are reflective of the specific activity of the substrate (labelling efficiency). Lanes 5-17 show 50% of the products recovered with an increasing concentration of GST FL hPIF1, GST (from 0 nM to 100 nM protein) or beads alone, in the absence of nucleotide cofactors. GST FL hPIF1 shows a concentration-dependent increase in binding to substrate dT_{50} while binding to the short ssDNA substrate M_$T_0G_4T_4$ was only observed at the highest protein concentration tested (100 nM). The 'pure' G4 DNA substrate $T_0G_4T_4$ demonstrated GST FL hPIF1 concentration-dependent binding across the titration range, indicating

a binding affinity greater than dT_{50} (shown in a bar graph for the lowest protein concentration (12.5 nM), $T_1G_4T_4$, $M\_T_1G_4T_4$ and T50). Thus, this experiment demonstrates high-affinity and selective binding of hPIF1 to unimolecular G4 DNA. Optimisation to reduce the background was performed as follows (Figure 33).

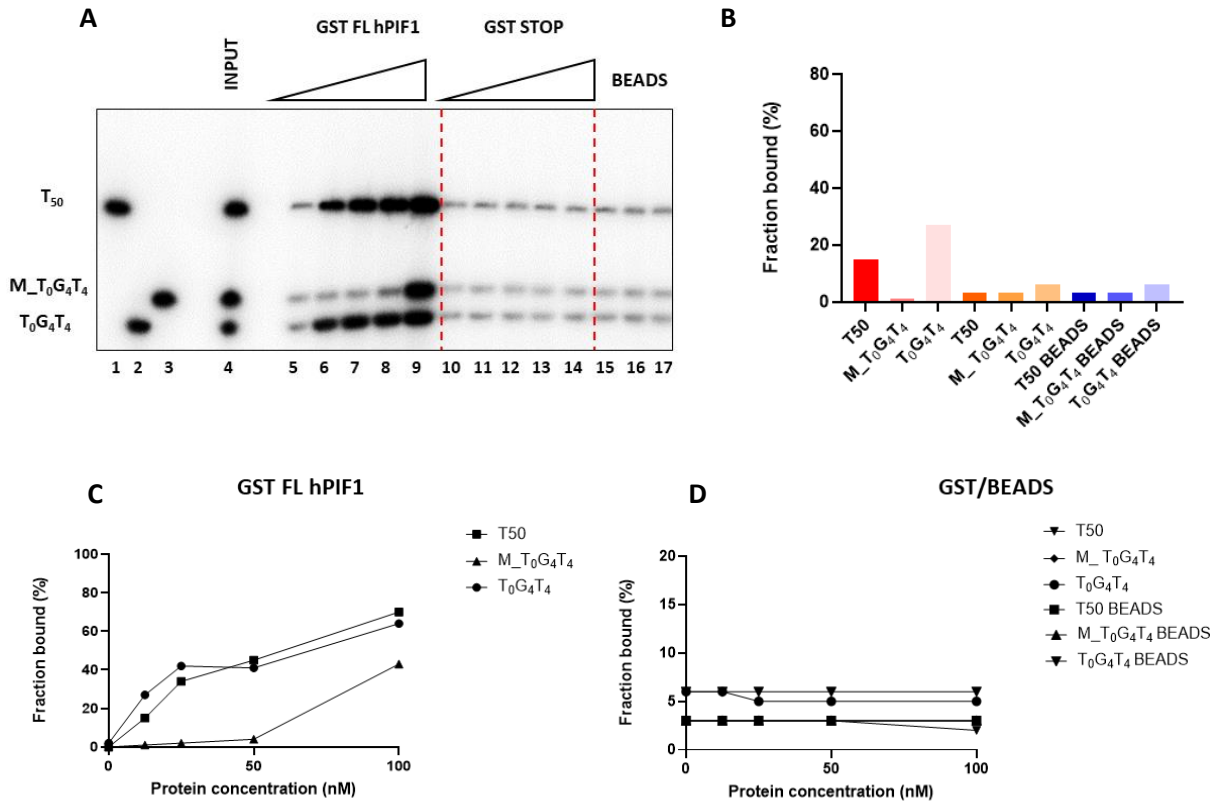


Figure 33. McKay binding assay with GST FL hPIF1, GST STOP and GSH beads alone. The substrates used were single strand T50, $T_1G_4T_4$ and $M\_T_1G_4T_4$ at a concentration of 0.25 nM final. (A) The protein concentration range used was 0 nM, 12.5, 25, 50, 100 nM, from lane 5 to lane 9 for FL and 10-14 for GST STOP protein, along with beads alone and no protein (3 repeats, lanes 15-17). (B) Data plot at 12.5 nM protein concentration. The data shows binding at low protein concentration, confirming the affinity of FL hPIF1 for both long ssDNA and G4 DNA. (C) Data plot of FL hPIF1 titration. Recovery of all substrates increases with increasing protein concentration. (D) Data plot of GST STOP titration and beads. As observed in the graph, increasing GST STOP concentration does not influence recovery of substrates, showing a low constant binding through the titration. The same level of recovery is observed for the beads alone, confirming them to be the main source of background in the experiments.

GSH-sepharose beads, responsible for binding the GST-tagged protein were combined in different ratios at a fixed volume (10 µL) with sephedex G15, superdex 75, superdex 200 and sephacryl 100 for pull-downs. A 10 µL total bead volume was considered to be the minimum required to give technical control over their recovery. Reactions contained 25 nM GST FL hPIF1 (from lane 5 to lane 17) and control GST STOP protein (from lane 18 to lane 30) with T50, $T_1G_4T_4$ and M_$T_1G_4T_4$ substrates (lanes 1, 2 and 3, Figure 34). GSH-sepharose was combined with the “inert” beads in a ratio of 25% and 50% to investigate the background contribution of the type of beads employed. Taken together with the results shown in Figure 33 above, the experiment shows that all inert beads except superdex 75 (lanes 22-24) have a lower background binding compared to GSH-sepharose alone when GST STOP is used in the assay (compared to lane 18, 100% GSH-sepharose, to lanes 19, 22, 25 and 28). The experiment also shows that an equivalent of 5 µL of GSH-sepharose is sufficient to recover the products of binding. To ensure a good level of binding and at the same time control the background the concentration of GSH-sepharose was used at 50%, while maintaining a constant volume of beads at 10 µL, by the addition of sephadex G15 (lanes 8 and 21).

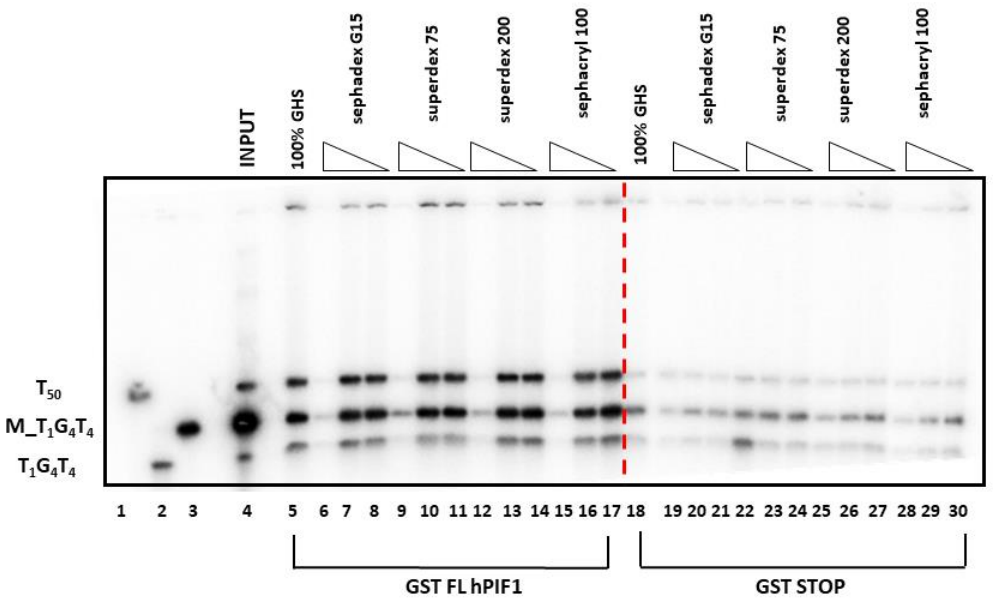


Figure 34. McKay assay background troubleshooting. Different combinations of beads were tested with 25 nM GST FL hPIF1 and GST-STOP protein on T50, $T_1G_4T_4$ and M_$T_1G_4T_4$ substrates at 0.25 nM. Lane 1-3, markers of the substrates used; lane 4, 50% input load on the gel; lanes 5 and 18, 100% GSH beads; each three lanes the percentage of GSH beads increases (0%, 50%, 75%), the missing percentage is compensated by other beads: lanes 6-8 and 19-21 decreasing concentration (100%, 50%, 25%) sephedex G15; lanes 9-11 and 22-24 superdex 75; lanes 12-14 and 25-27 superdex 200; lanes 15-17 and 28-30 sephacryl 100.

3.2.2.1 Optimization of KCl concentration for binding

K^+ ions are the dominant intracellular anion [167] while being optimal for G4 DNA stability. Given the salt-dependent stability of protein DNA interactions and the KCl requirement for G4 stability, KCl was explored over the range 55 mM, 75 mM, 95 mM, 115 mM and 135 mM KCl, in the McKay binding assay (Figure 35). Both FL hPIF1 and hPIF1 HD were employed at 50 nM concentration. In the assay, two long single-strand substrates (T40 and T50, lanes 1 and 2 were used as positive binding controls and a 'pure' G4 substrate ($T_0G_4T_4$) and its mutant M_$T_0G_4T_4$ (lanes 3 and 4). For GST FL hPIF1, if we accept that for technical reasons there has been a low recovery of product in lane 8 (95 mM KCl), then binding of G4 DNA is favoured at low salt concentrations (< 95 mM KCl) while binding to ssDNA increase with increasing KCl (lanes 6-10). For the helicase core domain (which lacks the non-specific N-terminal DNA binding domain intermediate KCl concentrations (75-95 mM) favour binding of both ssDNA and G4 DNA compared to higher and lower concentrations

(lane 13, compared to 11 and 12 and 14 and 15). The pattern emerging with the GST control protein (GST STOP) is that high salt concentrations promote background binding of ssDNA while lower KCl concentrations (except the lowest) favour G4 DNA binding. Since the aim was to find a compromise between the binding affinity of the protein to the various substrates and the background influenced by the salt concentration, a final concentration of 100 mM KCl (equivalent to the intracellular salt concentration [168]) was chosen for the analysis. This also approximates to the intracellular K^+ concentration [169].

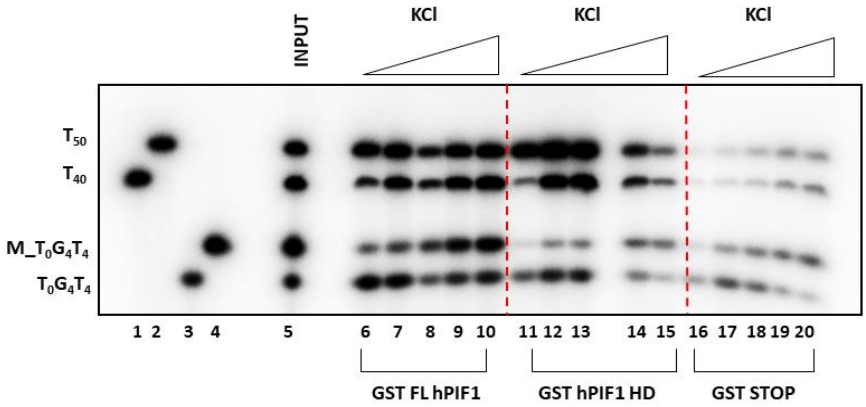


Figure 35. McKay assay exploring a KCl titration (55, 75, 95, 115 and 135 mM KCl). Lane 1-4 substrates markers; lane 5, 25% substrate input, lanes 6-10 KCl titration with 50 nM GST FL hPIF1; lanes 11-15 KCl titration with 50 nM GST HD ΔC206; lanes 16-20 KCl with 50 nM GST STOP protein.

3.3 Exploring hPIF1 binding activity

3.3.1 GST McKay pulldown assay with ssDNA substrates and nucleotide cofactors

ssDNA binding by PIF1 proteins has been well studied [135]. To support the use of the McKay assay in binding studies I asked if results obtained from previous gel-shift and biophysical binding studies could be recapitulated with this approach. Accordingly, the length dependence of hPIF1 ssDNA binding affinity and changes induced in the presence of nucleotide cofactors was assessed.

In the assay reported in Figure 36, ^{32}P labelled poly d(T) substrates increasing in length were analysed (T10, 20, 30, 40 and 50, marker in lanes 1-5, Figure 36 A), in

both apo and transition state analogue ($ADP \cdot MgF_4^-$) conditions, using FL hPIF1 and helicase core hPIF1 HD (titration 12.5 nM-100 nM). As observed previously [145], without a nucleotide cofactor, the extent of hPIF1 ssDNA binding increases with increasing poly T length (lanes 7-14, Figure 36 A). Also, the data indicate a higher affinity for ssDNA binding to FL hPIF1 compared to helicase core, likely attributable to the non-specific DNA binding activity of the N-terminal domain.

A significant increase in binding affinity is observed in the presence of $ADP \cdot MgF_4^-$, where binding is also observed for the shorter poly T substrates (T20) for FL hPIF1 (lanes 19-26). Interestingly, both the T10 and T20 substrate (lanes 1 and 2, Figure 36 A) show a very low binding affinity for helicase core compared to the other poly d(T) substrates, with or without nucleotide cofactor.

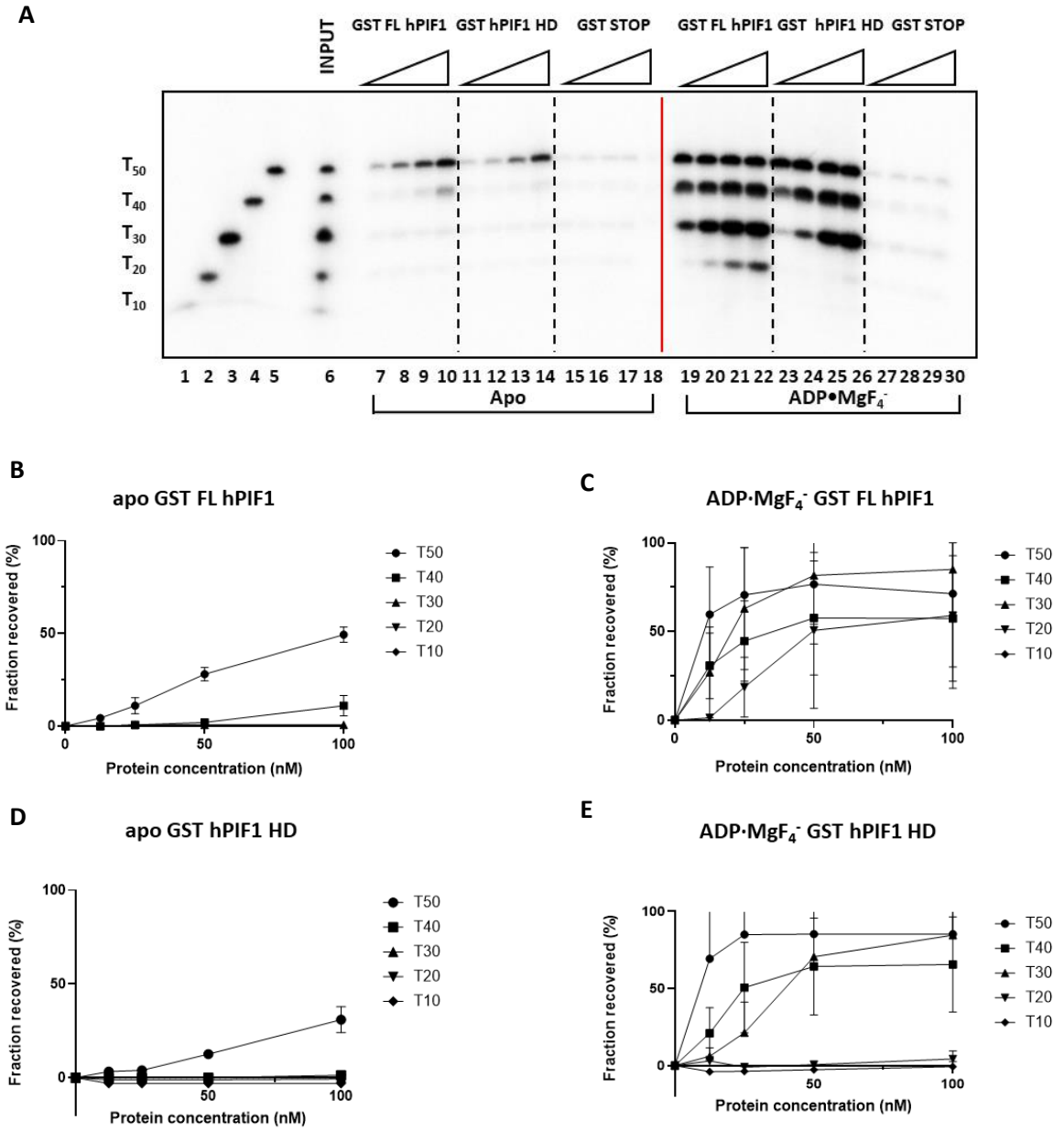


Figure 36. McKay assay investigating hPIF1 ssDNA length-dependent affinity and change of binding in the presence of ADP·MgF_4^-. (A) Acrylamide gel showing products of binding reactions increasing length poly T substrates. Lanes 1-5, markers T10 to T50; lane 6, 25% substrate input; lanes 7-10, GST FL hPIF1 titration (12.5, 25, 50 and 100 nM) apo conditions; lanes 11-14, GST hPIF1 HDΔC26 titration, apo conditions; lane 15-18, GST STOP titration apo conditions; lanes 19-22, GST FL hPIF1 titration with ADP·MgF_4^- (transition state analogue: 1 mM ADP, 1.5 mM $MgCl_2$, 35 mM NaF); lanes 23- 26 GST hPIF1 HDΔC26 titration with ADP·MgF_4^-; lanes 27-30 GST STOP with ADP·MgF_4^-. (B) data plot of apo condition GST FL hPIF1; (C) data plot of transition state analogue GST FL hPIF1. (D) data plot of apo conditions GST hPIF1 HDΔC26; (E) data plot of transition state analogue GST hPIF1 HDΔC26.

3.3.2 DNA binding by GST FL hPIF1, GST hPIF1 HD∆C26 and GST hPIF1 HD 206-END compared

Using the McKay assay, DNA binding by FL hPIF1 was compared to two truncated forms, hPIF1 HD and hPIF1 HD-E, from 12.5 nM to 200 nM. The investigation focused on the unimolecular G4 substrate $T_1G_4T_4$, its unfolded mutant M_$T_1G_4T_1$ and the ssDNA oligo $d(T)_{50}$ as a control for binding. As observed in Figure 37 A, the binding of all DNA substrates increased with increasing protein concentration when GST FL and GST HD-E were compared. In the graphed data we can see that 50% binding (substrate recovery) is already detected at 25 nM GST FL hPIF1 while for the truncate hPIF1 forms, the same level is reached at high protein concentration (100 nM) for G4 substrate. Furthermore, both FL and HD-E demonstrated a higher affinity for the G4 DNA substrate compared to the unfolded mutant control of the same length. A further comparison was conducted between the two helicase forms, hPIF1 HD and hPIF1 HD-E, over the same titration range (Figure 37 B), without substantial differences in the DNA binding being observed.

The data plot for the 50 nM protein shown in Figure 37 C further highlights the observations above: with FL hPIF1 protein the difference in recovery between the folded DNA and its mutated form is ~6 folds while with hPIF1 HD-E is ~14 folds and with the core domain (hPIF1 HD) is ~3 folds. The latter observations suggest that a further investigation on the role of the 21 amino acids at the C terminal end of hPIF1 requires further investigation.

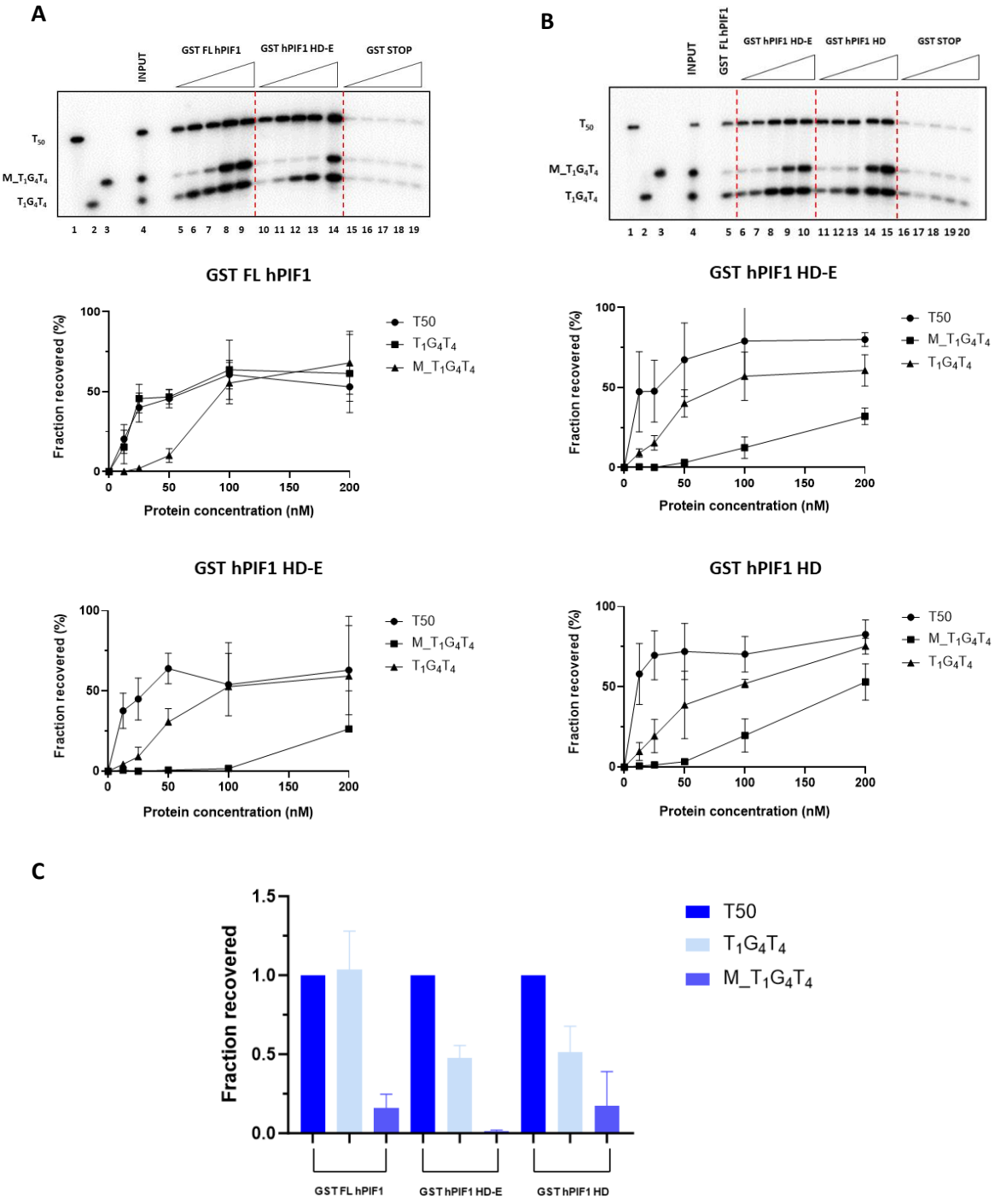


Figure 37. Binding of GST FL hPIF1 and truncation mutants to T50, $T_1G_4T_4$ and M_$T_1G_4T_4$ substrates. (A) GST FL hPIF1 and GST HD-END compared; in the acrylamide gel: lanes 1-3, marker controls; lane 4, 25% input; lanes 5-9 titration (12.5, 25, 50, 100 and 200 nM) GST FL hPIF1 protein; lanes 10-14 titration GST HD 206-END (HD-END, 12.5-200 nM); lanes 15-19, titration GST STOP control (12.5-200 nM). Data plot binding of FL hPIF1 and GST HD 206-END on T50, $T_1G_4T_4$ and M_$T_1G_4T_4$ substrates. (B) GST HD-END and GST HDΔC206 compared. In the acrylamide gel: lanes 1-3, marker controls; lane 4, 25% input; lane 5, 50 nM GST FL hPIF1 control; lanes 6-10, titration GST hPIF1 HD-END; lanes 11-15, titration GST hPIF1 HDΔC26; lanes 16-20, titration GST STOP control (protein range 12.5-200 nM). Data plot of GST hPIF1 HD 206-END and GST hPIF1 HDΔC26 on T50, $T_1G_4T_4$ and M_$T_1G_4T_4$ substrates. (C) Bar graph for 50 nM protein concentration pull-down, normalised to T50 recovery.

3.3.3 hPIF1 binding to G4 DNA with and without a 5' tail

The structure of *Thermus Oshimai* PIF1 bound to G4 DNA (see section 3.5) revealed residues binding the G4 quartets but also a key role of a 5' single-stranded tail residue in binding [170]. I asked whether the 5' tail plays the same key role in hPIF1 G4 DNA binding and synthesised G4 DNA with ($T_1G_4T_4$) and without ($T_0G_4T_4$) the 5' tail for analysis in binding assays.

First, the binding of the two truncated forms of the protein, hPIF1 HD-E and hPIF1 HD was investigated (12.5-200 nM) with oligo $d(T)_{50}$ as a binding control, $T_0G_4T_4$ and its mutant M_$T_0G_4T_4$ (lanes 1-3, Figure 38). As shown in Figure 38 and as previously observed (see section 3.4.2), both truncated hPIF1 forms bind $d(T)_{50}$ efficiently even at low protein concentrations, but the short single-strand substrate M_$T_0G_4T_4$ is bound only at high protein concentrations (100-200 nM, lanes 6-10 ad 11-15). Both GST hPIF1 proteins show an increase in binding to the G4 substrate with increasing protein concentration, and higher affinity binding compared to the unfolded control substrate M_$T_0G_4T_4$. The level of binding to the G4 DNA between the core domain and the construct with the additional C-terminal 21 residues (hPIF1 HD-E) shows small differences, but with an apparent slightly higher affinity for the core domain (Figure 38 B). Figure 38 C shows the binding at 100 nM protein concentration, with the data normalised to $d(T)_{50}$ where binding is saturated.

These observations demonstrate that a single-stranded “tail” residue is not required for hPIF1-G4 DNA binding and the C-terminal 21 residues of hPIF1 have little influence on G4 DNA binding.

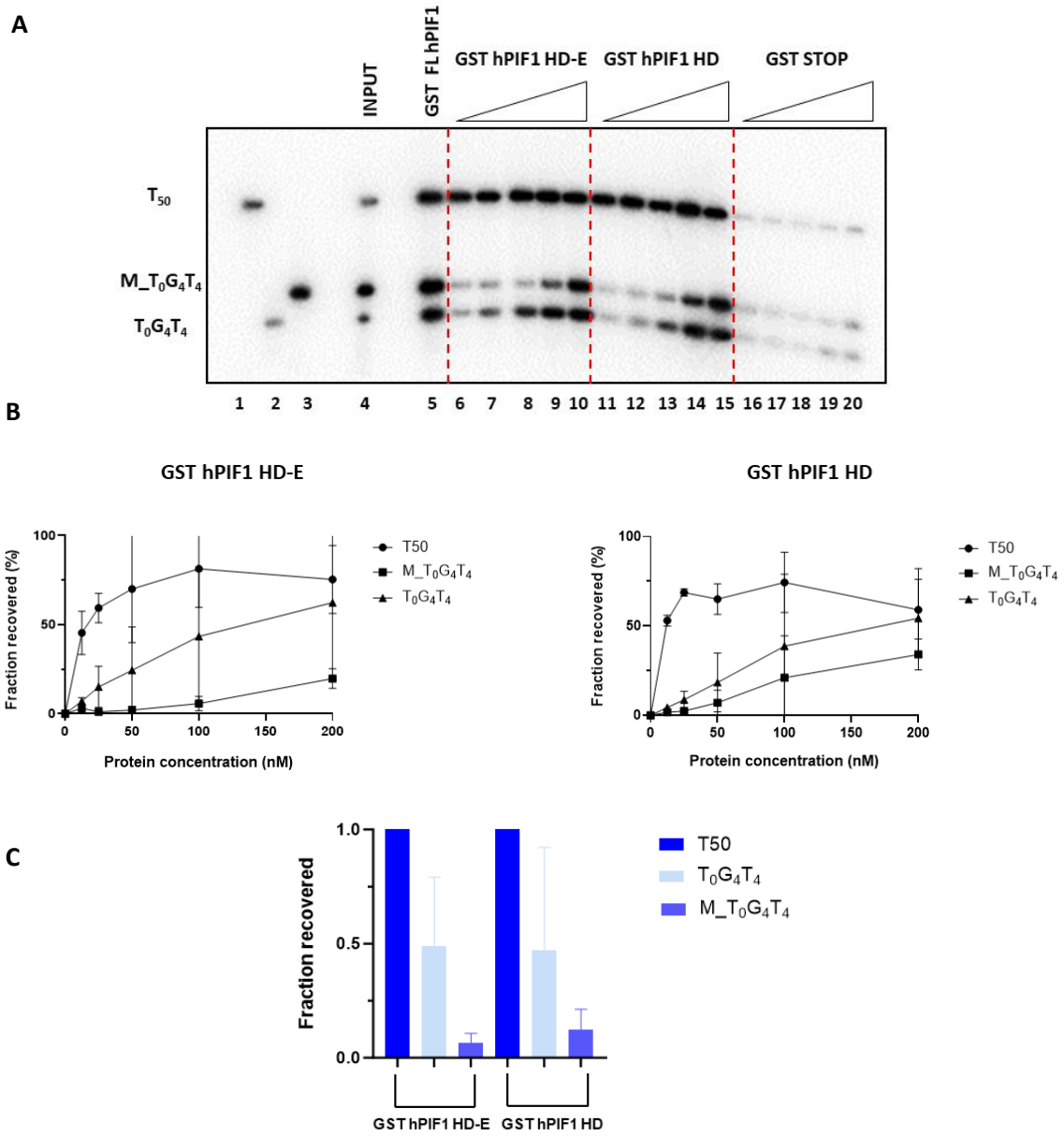


Figure 38. McKay assay investigating G4 DNA binding to two truncated hPIF1 forms (0.25 nM T50, $T_0G_4T_4$ and M_$T_0G_4T_4$ substrates). (A) Acrylamide gel showing the products of "pulldown" assays. Lanes 1-3 markers; lane 4, 25% input; lanes 6-10 GST hPIF1 HD-E titration (12.5, 25, 50, 100 and 200 nM); lanes 11-15 GST hPIF1 HD; lanes 16-20 GST STOP. In the marker lanes and input, the signal detected from $T_0G_4T_4$ is reduced compared to the other substrates due to its lower specific activity (B) and (C) data plot for the two truncated PIF1 forms (n=3 repeats), showing an increase in binding with increasing protein concentration. (D) Graph at 100 nM protein concentration with data normalized to T50 recovery.

3.3.4 Modulation of hPIF1-DNA binding affinity by nucleotide cofactors

hPIF1 affinity for single-stranded (ss) DNA increases in the presence of the transition state analogue ADP·MgF_4^- [135] compared to all other states (apo, product (ADP) and ground state mimic (AMP·PNP)). To explore how the affinity for G4 DNA changes during the ATP hydrolysis cycle (Figure 39), binding to substrates T50, $T_1G_4T_4$ and M_$T_1G_4T_4$ (lanes 1, 2 and 3 Figure 39) was tested with 50 nM hPIF1 HD-E and the GST STOP control under apo conditions and in the presence of in the following cofactors and singular components: $MgCl_2$; $MgCl_2$ and ADP (product); $MgCl_2$, ADP and NaF (transition state ADP·MgF_4^-); ADP; $MgCl_2$ and NaF; ADP and NaF; NaF.

Under apo conditions and with $MgCl_2$ (lanes 5 and 6), hPIF1 HD-E demonstrated high affinity for the long ssDNA substrate (oligo T50), lower affinity for the G4 DNA substrate and minimal binding to its non-folding control (M_$T_1G_4T_4$). Strikingly, however, the situation drastically changes in the presence of the transition state analogue (lane 8), with a loss of affinity for the G4 DNA but an increase for the short single stranded substrates. In the presence of ADP alone or combination with NaF or $MgCl_2$, a substantial affinity for all the substrates analysed is lost (lanes 7, 9, 11). No change in binding is determined by the presence of NaF alone or with $MgCl_2$ (lanes 10 and 12).

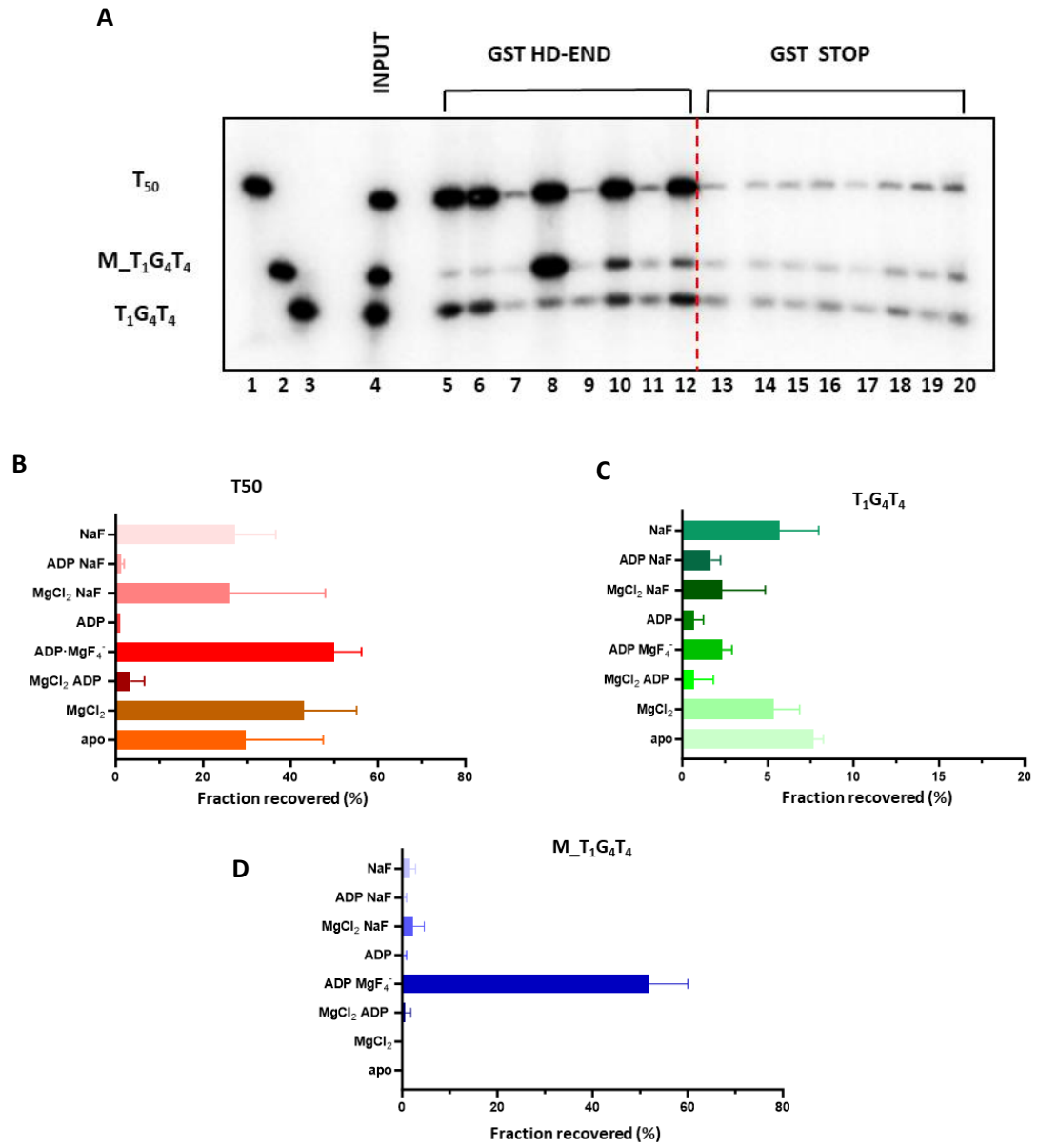


Figure 39. McKay assay investigating changes in hPIF1 binding in the presence of cofactors. (A) Representative acrylamide gel analysis of binding products recovered with of 50 nM GST hPIF1 HD-END or GST STOP with 0.25 nM T50, $T_1G_4T_4$ and M_$T_1G_4T_4$ substrates. Lanes 1-3, markers; lane 4, 25% input; lanes 5 and 13 apo (50% recovery); lanes 6 and 14 $MgCl_2$ (50% recovery); lanes 7 and 15 $MgCl_2$ + ADP (50% recovery); lanes 8 and 16 $MgCl_2$, ADP, NaF (50% recovery); lanes 9 and 17 ADP (50% recovery), lanes 10 and 18 $MgCl_2$ + NaF (50% recovery); lanes 11 and 19 ADP + NaF (50% recovery); lanes 12 and 20 NaF (50% recovery). (B) Data plot of T50 change of affinity in the presence of the various cofactors; (C) Data plot of $T_1G_4T_4$ change of affinity in the presence of the various cofactors; (D) Data plot of M_$T_1G_4T_4$ change of affinity in the presence of the various cofactors (n=3 repeats).

To extend this analysis, binding of T50, $T_1G_4T_4$ and M_$T_1G_4T_4$ (lanes 1-3, Figure 40) to GST FL hPIF1 and GST hPIF1 HD206-END (50 nM) in the absence or presence of each nucleotide cofactor (ATP/$MgCl_2$, AMP·PNP/$MgCl_2$, ADP/$MgCl_2$ and ADP·MgF_4^-) was tested. For the FL protein, there is a high affinity for the long ssDNA substrate and the G4 substrate and a low affinity for the short ssDNA substrate (M_$T_1G_4T_4$) under apo conditions and with $MgCl_2$ (lanes 5 and 6), as seen before (Figure 37). At the ground state (AMP·PNP, lane 7), there is a reduction (30-50%) in binding affinity for all substrates compared to apo/$MgCl_2$ conditions. As before (Figure 39), with the transition state analogue (ADP·MgF_4^-, lane 8) there is a large increase in binding for the short single-stand substrate and a loss of affinity for the G4 DNA. In the product state (ADP/$MgCl_2$, lane 9) FL hPIF1 shows an apparent loss of affinity for all substrates. In the presence of ATP (lane 10), the lowest affinity is observed for the single-stranded DNA substrates. Notably, however, there is an intermediate level of binding to the G4 DNA, greater than that observed with the transition state analogue, but less than that observed under apo/$MgCl_2$ or with AMP·PNP/$MgCl_2$, not observed with helicase domain (GST HD-E, lanes 11-15). For hPIF1 HD-E, binding under apo, ADP/$MgCl_2$ and ADP·MgF_4^- is as described in Figure 37. In addition, there is a near complete loss of affinity for all substrates in the presence of ATP/$MgCl_2$ and AMP·PNP/$MgCl_2$.

In summary, all substrates bind with low affinity to helicase core in the ground (AMP·PNP/$MgCl_2$) and product state (ADP/$MgCl_2$), and with ATP/$MgCl_2$. G4 DNA binds with high affinity only in the absence of nucleotide cofactors and the affinities for G4 DNA and short ssDNA substrates are modulated in opposing ways with and without the transition state mimic ADP·MgF_4^-. Furthermore, the data suggest that the presence of the N-terminal domain (residues 1-205) can promote binding to all DNA substrates under all conditions.

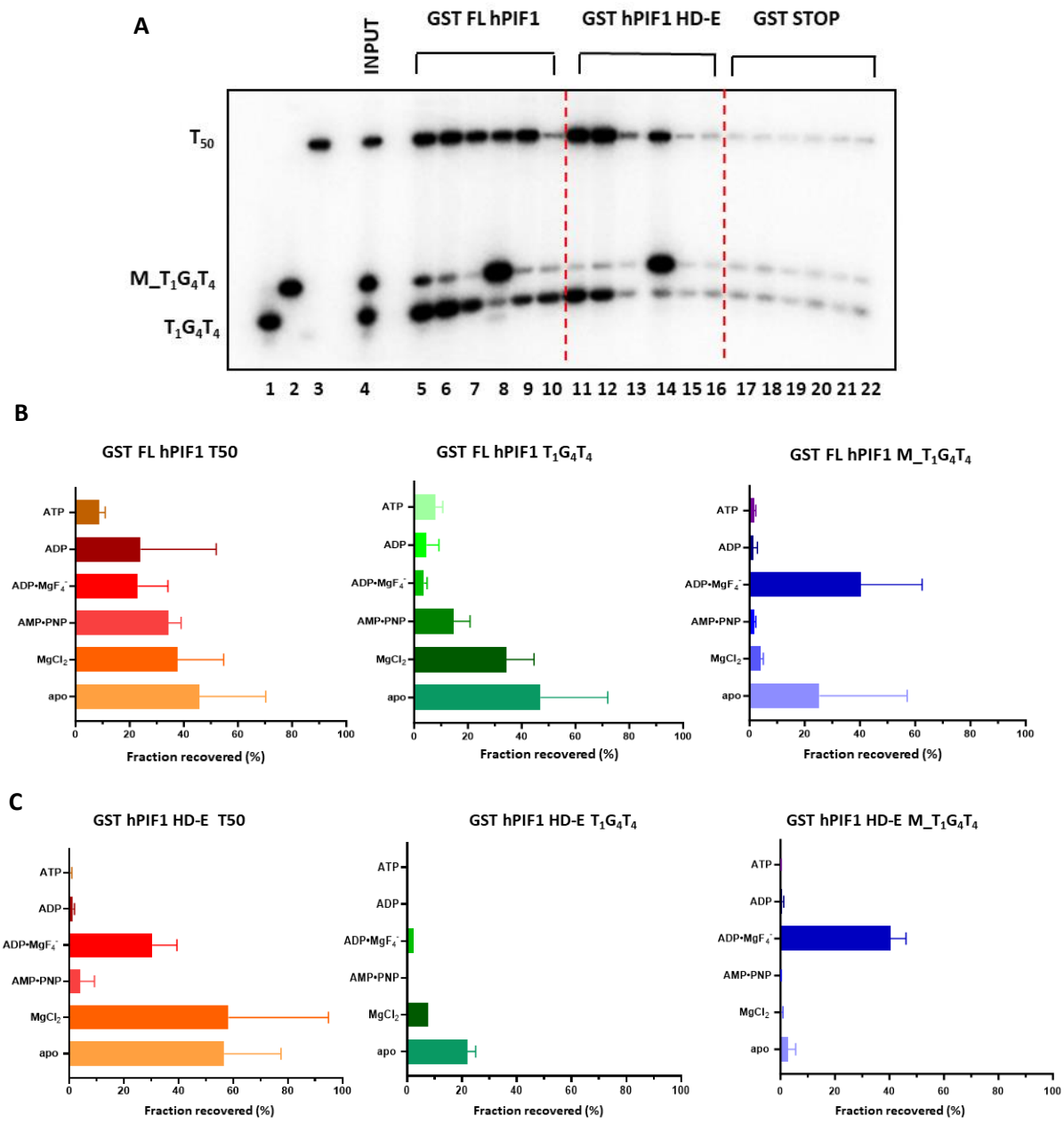


Figure 40. McKay assay investigating changes in hPIF1 DNA binding affinity in the presence of nucleotide cofactors (substrate $T_1G_4T_4$). (A) Acrylamide gel showing products recovered from binding to GST FL hPIF1, GST hPIF1 HD206-END and GST STOP (50 nM) with 0.25 nM T50, $T_1G_4T_4$ and M_$T_1G_4T_4$ substrates. Lanes 1-3, markers; lane 4, 25% input; lanes 5, 11 and 17, apo; lanes 6, 12 and 8, $MgCl_2$; lanes 7, 13 and 19, AMP·PNP; lanes 8, 14 and 20, ADP·MgF_4^-; lanes 9, 15 and 21, ADP/$MgCl_2$; lanes 10, 16 and 22 ATP/$MgCl_2$. (B) Data plot of GST FL hPIF1 binding to the three substrates. (C) Data plot of GST hPIF1 HD206-END binding to the three substrates (n=3 repeats).

To complement the analysis above, binding to a ‘tailless’ G4 substrate ($T_0G_4T_4$) was investigated together with T50 and M_$T_0G_4T_4$ (lanes 1, 2 and 3, Figure 41) in the absence and presence of nucleotide cofactors. Both FL hPIF1 and hPIF1 HD-E (50 nM) were compared. Experiments were performed and the data was analysed as in Figure 40. It can be concluded that there are no significant differences in hPIF1 binding to the tailless G4 substrate compared to the G4 substrate under all conditions.

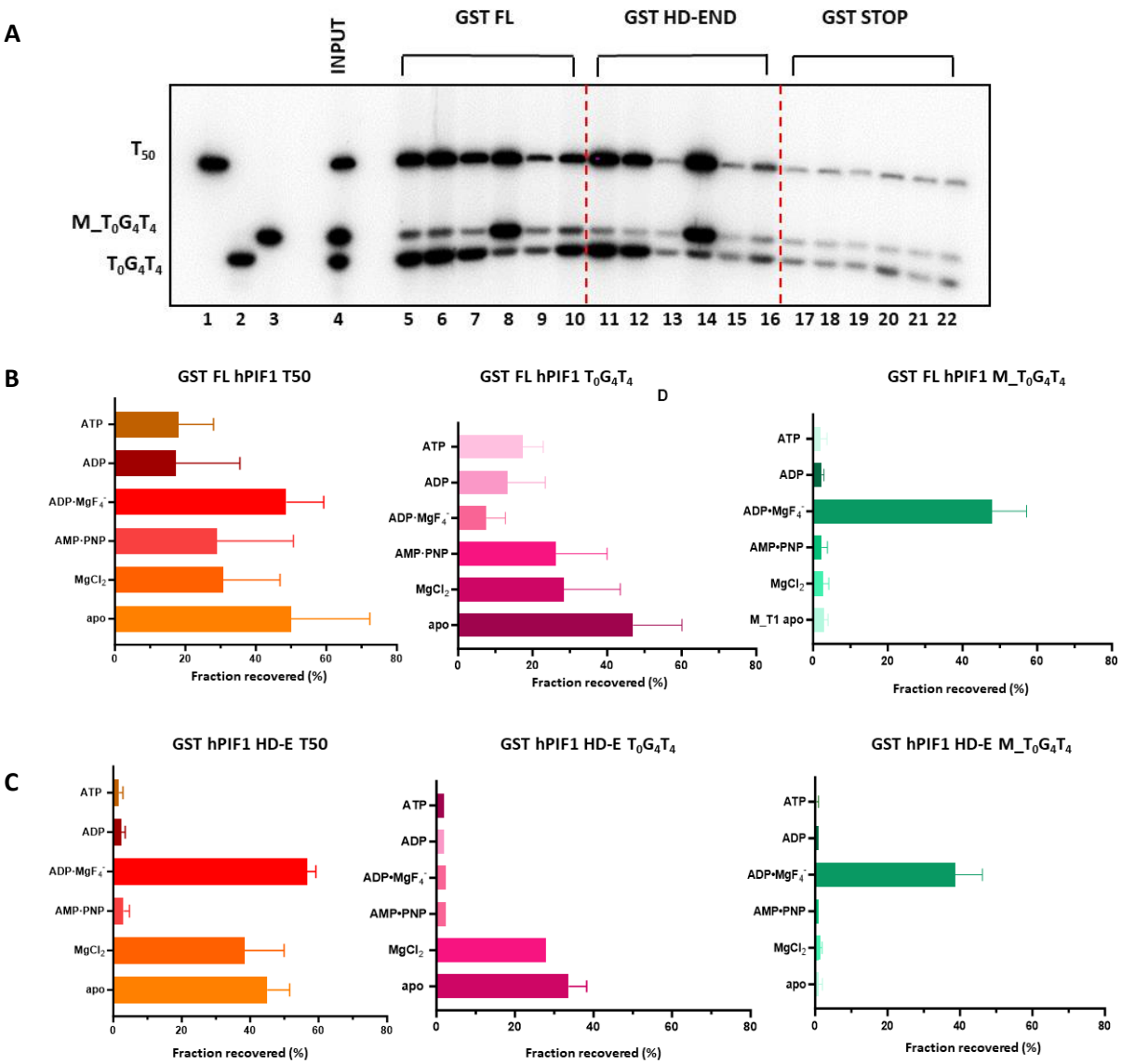


Figure 41. McKay assay investigating changes in hPIF1 DNA binding affinity in the presence of nucleotide cofactors (substrate $T_0G_4T_4$). (A) Acrylamide gel showing products recovered from binding to GST FL hPIF1, GST hPIF1 HD-END and GST STOP (50 nM) with 0.25 nM T50, $T_0G_4T_4$ and M_$T_0G_4T_4$ substrates. Lanes 1-3, markers; lane 4, 25% input; lanes 5, 11 and 17, apo (50% recovery); lanes 6, 12 and 8, $MgCl_2$ (50% recovery); lanes 7, 13 and 19, AMP·PNP (50% recovery); lanes 8, 14 and 20, ADP·MgF_4^- (50% recovery); lanes 9, 15 and 21, ADP/$MgCl_2$ (50% recovery); lanes 10, 16 and 22, ATP/$MgCl_2$ (50% recovery); (n = 3 repeats). (B) Data plot of GST hPIF1 HD-END binding to the three substrates. (C) Data plot of GST FL hPIF1 binding to the three substrates (n= 3 repeats).

3.3.5 Binding of hPIF1 to 'cellular' G4 substrates

The unimolecular substrates used in the analysis above can be considered as 'synthetic' test substrates, although conforming to the G4 DNA forming consensus sequence [145]. Two 'cellular' G4 structures were chosen for complementary analysis: substrate T_1Myc1245 consisting of the folded G4 core sequence found in the promoter region of the Myc oncogene [171] with a single 5' T extension and Telo22 from the human telomeres sequence, also with a 5' tail of 1 nt (adenine). Notably, these substrates have only three G4 stacks and different loop lengths and sequences (Figure 42).

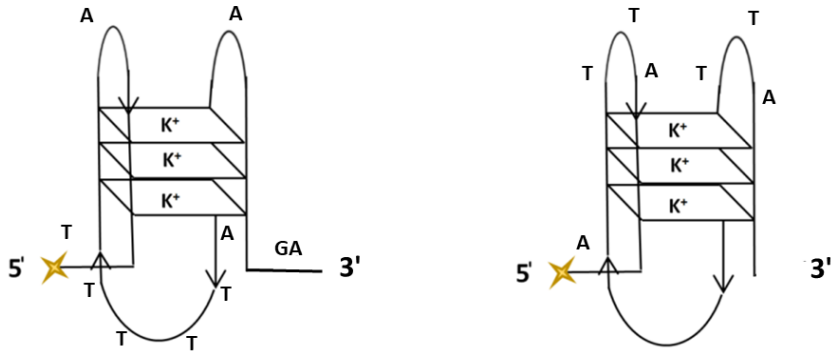


Figure 42. Graphic representation of T_1Myc1245 G4 structure (left) and Telo22 G4 structure (right). Both structures are radio-labelled at the 5' end, with the loop residues indicated on the side of the structure. The three G4 stacks are stabilized by a K^+ cation indicated in the middle of each stack.

First, substrate Telo22 was analysed in parallel with T50, $T_1G_4T_4$ and M_$T_1G_4T_4$ (0.25 nM; lanes 1-4, Figure 43) with FL hPIF1 and hPIF1 HD (12.5 nM to 200 nM) in the absence of nucleotide cofactors (apo conditions). While GST hPIF1 binding to the substrates previously used (T50, $T_1G_4T_4$ and M_$T_1G_4T_4$, lanes 6-10) was as previously observed (Figure 37), FL hPIF1 affinity for Telo22 was significantly reduced and detected only at the highest protein concentrations (100-200 nM, lanes 9 and 10 and (B) left). For the helicase domain construct (lanes 11-15 and (B) right), there is little detectable binding to Telo22 at all protein concentrations, further illustrating the impact of the non-specific binding activity of residues 1-205.

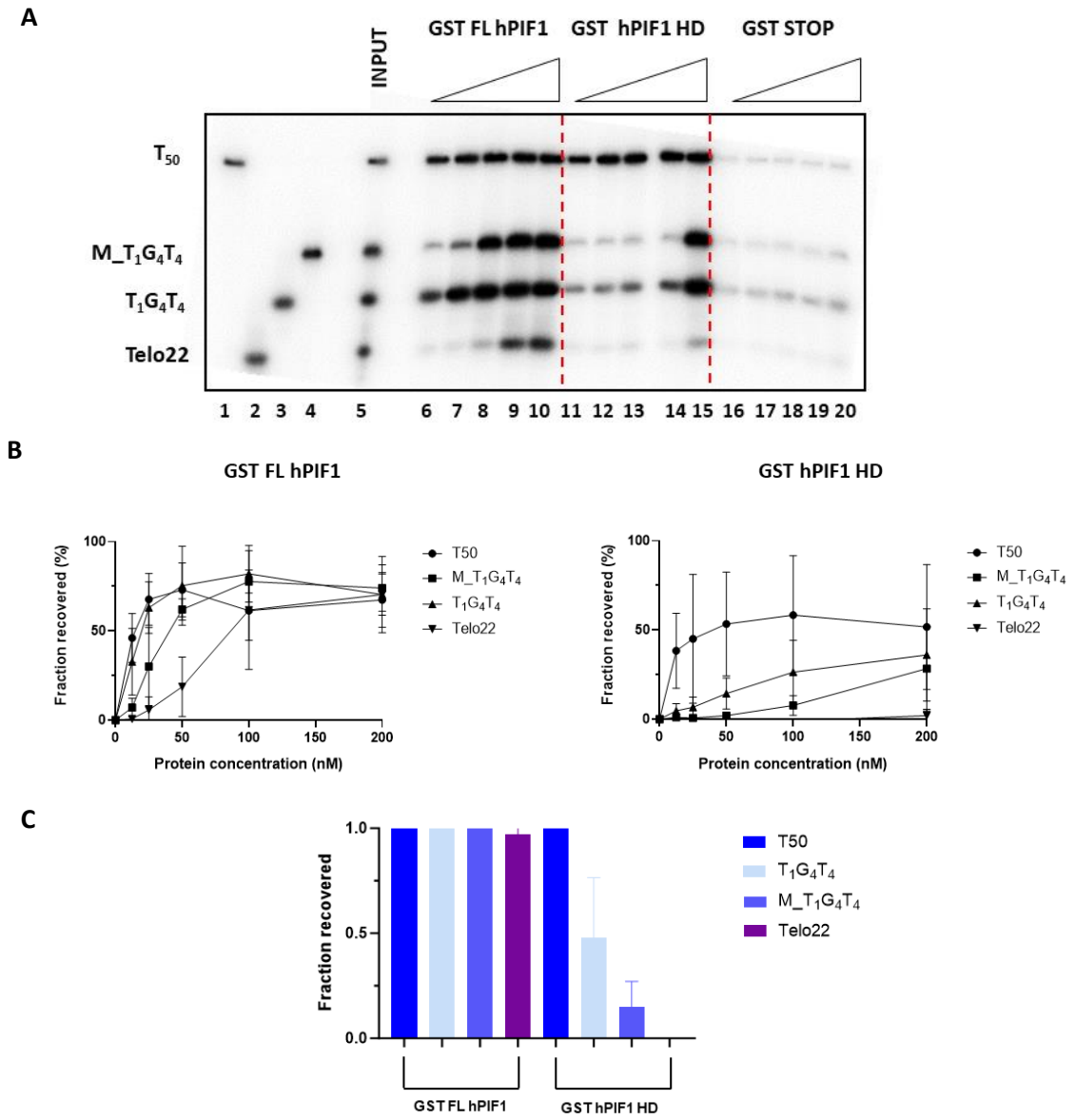


Figure 43. McKay assay investigating hPIF1 binding to the Telo22 substrate and standard test substrates. (A) Acrylamide gel analysis of reaction products recovered from binding assays with GST FL hPIF1, GST hPIF1 HD-END and GST STOP (12.5, 25, 50, 100 and 200 nM), 0.25 nM T50, $T_1G_4T_4$ and M_$T_1G_4T_4$ and Telo22 substrates. Lanes 1-4, markers; lane 5, 25% input; lanes 6-10, GST FL hPIF1 titration; lanes 11-15, GST hPIF1 HD206-END titration; lanes 16-20, GST STOP titration. (B) Data plots for the two hPIF1 proteins (n=3 repeats). (C) Data normalised to T50 at 100 nM protein concentration. (n=3 repeats)

Next, substrate T_1Myc1245 representing a G4 DNA forming sequence from the *Myc* oncogene promoter was analysed. Again, the investigation was conducted in parallel with T50, $T_1G_4T_4$ and its mutated form, M_ $T_1G_4T_4$ (0.25 nM; lanes 1-4 Figure 44). The analysis focused on the two truncated forms of hPIF1: hPIF1 HD-E and hPIF1 HD (12.5-200 nM). As with Telo22 (Figure 44), high hPIF1 binding to T_1Myc1245 G4 DNA was detected only at high protein concentrations (100-200 nM, lanes 11 and 16) with no difference between the two isoforms of the protein, while $T_1G_4T_4$ bound with higher affinity.

Also however, in this experiment a difference in DNA binding between the $T_1G_4T_4$ substrate and its mutant M_$T_1G_4T_4$ for the hPIF1 helicase domain proteins, with and without the C-terminal 21 amino acids, was observed. In Figure 44 C, hPIF1 HD-E shows a lower affinity for both substrates, suggesting the C-terminal residues have reduce affinity for DNA, but this could be due to differences in protein specific activity. However, as previously observed in Figure 37, there is a clear difference in *relative* affinity for the folded and unfolded ssDNA mutant for individual each protein. For hPIF1 HD-E at 100 nM, the the recovery of G4 DNA (Figure 44 C) is ~12 fold greater than for the unfolded mutant. However, for the core helicase domain protein, the relative difference in binding detected is less, at ~5 fold between substrates.

In summary, the observations described above demonstrate significantly reduced binding affinity of hPIF1 for both Telo22 and T_1Myc1245 compared to the synthetic test substrate $T_1G_4T_4$. These substrates differ in the number of G4 quartets (3 compared to 4), the length and sequence of the loop residues, and therefore also likely the folding conformation (see section 1.4). Next, the hypothesis that the loops and number of G4 quartets can influence hPIF1-G4 DNA binding was tested.

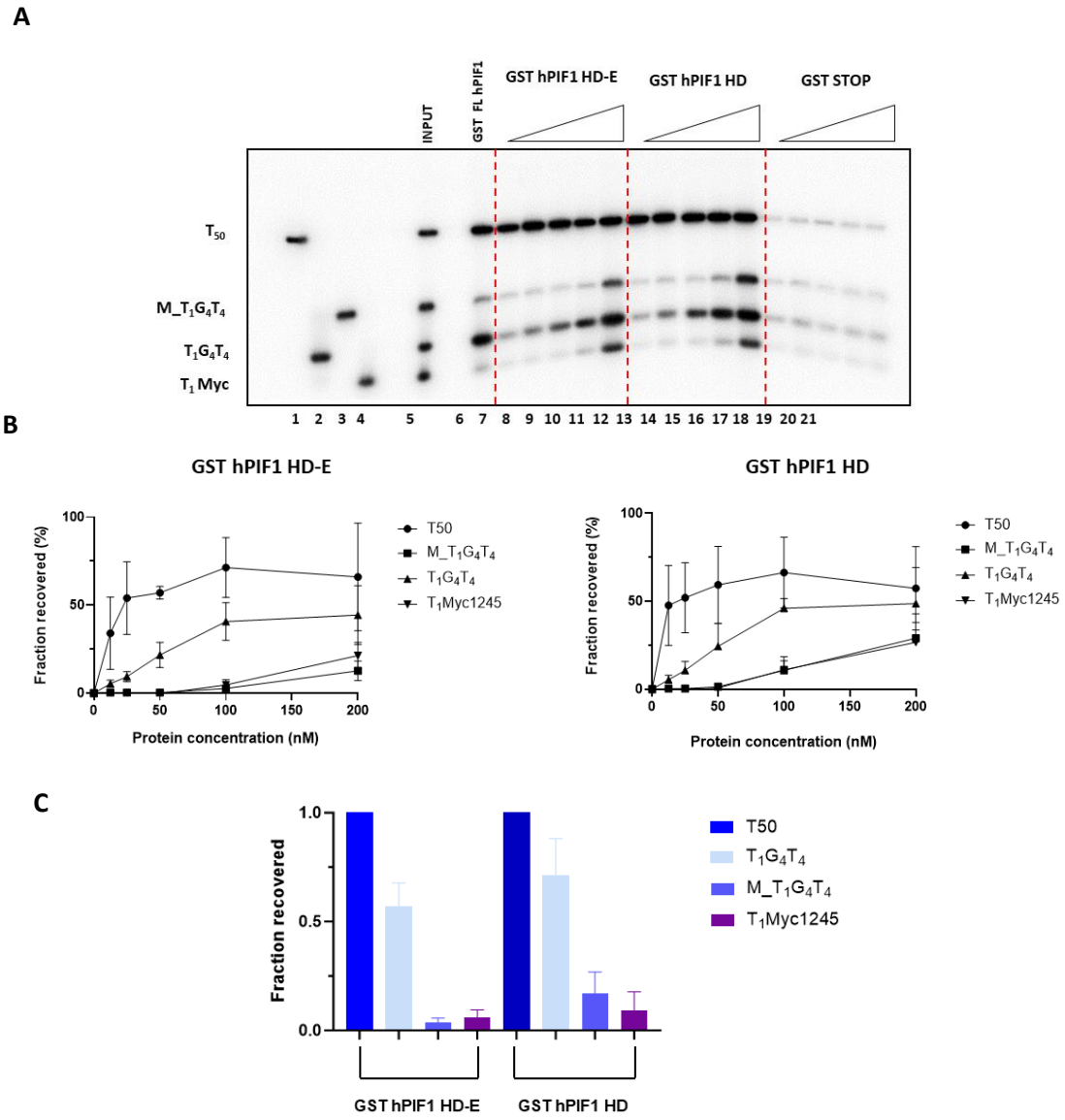


Figure 44. McKay assay investigating hPIF1 change binding to T_1Myc1245 and standard test substrate. (A) Acrylamide gel of recovered products from binding to GST hPIF1 HD-END, GST hPIF1 HDΔC and GST STOP (12.5, 25, 50, 100 and 200 nM) with 0.25 nM T50, $T_1G_4T_4$ and M_$T_1G_4T_4$ and T_1Myc1245 substrates. Lanes 1-4, markers; lane 5, 25% input; lane 6, 50 nM GST FL hPIF1 positive binding control; lanes 7-11, GST FL hPIF1 titration; lanes 12-16, GST hPIF1 HD206-END titration; lanes 17-21, GST STOP titration. (B) Data plots for the two HD proteins (n=3 repeats). (C) Data normalised to T50 at 100 nM protein concentration.

3.3.6 Influence of G4 stack number on hPIF1-G4 DNA binding

A synthetic G4 substrate $T_1G_3T_4$ with three G quartets and intervening $d(T)_4$ loops were analysed in 'McKay' binding assays as before, in parallel with T50, $T_1G_4T_4$ and its mutant M_ $T_1G_4T_4$ (0.25 nM; lanes 1-4, Figure 45), focusing on the truncated proteins hPIF1 HD-E and hPIF1 HD (12.5 nM to 200 nM). As observed in Figure 45, lane 4 and 5, $T_1G_3T_4$ appeared to be unstable and likely to be completely infolded (lanes 4 and 5). Despite this, no nucleic acid products related to $T_1G_3T_4$ were recovered in binding reactions with hPIF1 helicase domain proteins (lanes 7-16). Together with the results for telomeric and Myc G4 DNA substrates described above, these results suggest that substrates with four quartets bind to hPIF1 G4 with higher affinity than those with three.

In addition, the side-by-side comparison between hPIF1 HD-E and hPIF1 HD revealed an observation consistent with Figure 37 and here illustrated in the bar charts in (C). In each case, the ratio of G4 DNA to unfolded control ($T_1G_4T_4$ /M_$T_1G_4T_4$) is 2-3 fold greater for hPIF1 HD-END compared to HD, suggesting that the C-terminal 21 (predominantly acidic) residues may be a negative regulator of ssDNA binding. Moreover, significant is also the change of affinity between the folded form and its short ssDNA mutant: in the truncated form hPIF1 HD-E the affinity detected at 100 nM protein for G4 DNA (Figure 45 C) is ~12 fold bigger than its mutant. In the core domain protein, the difference in binding detected is less, with an affinity for the G4 substrate increased of ~5 folds.

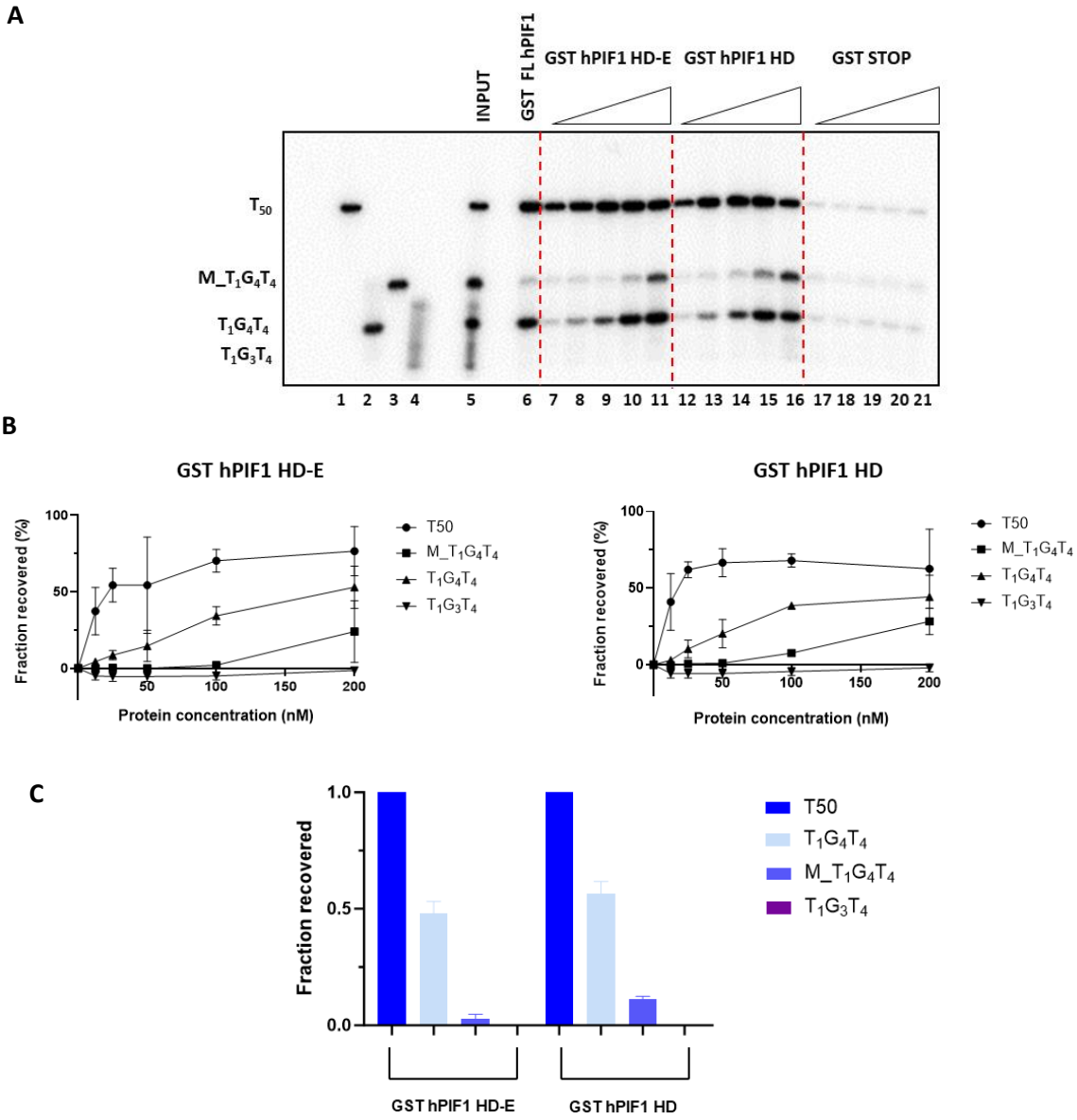


Figure 45. McKay assay investigating hPIF1 binding to $T_1G_3T_4$ and standard test substrate. (A) Acrylamide gel analysis of products recovered from GST hPIF1 HD-END, GST hPIF1ΔC26 and GST STOP binding reactions (12.5, 25, 50, 100 and 200 nM) with 0.25 nM T50, $T_1G_4T_4$ and $M\_T_1G_4T_4$ and $T_1G_3T_4$ substrates. Lanes 1-4, markers; lane 5, 25% input; lane 6, 50 nM FL hPIF1 positive binding control; lanes 7-11, hPIF1 HD-E titration; lanes 12-16, titration of hPIF1 HD; lanes 17-21, GST STOP titration. (B) Data plot for the two HD proteins (n=3 repeats). (C) Data normalised to T50 at 100 nM protein concentration.

3.3.7 Influence of loop length on hPIF1-G4 DNA binding

From the literature, we know that the loop sequence and length influence G4 DNA conformation (parallel or antiparallel-folded structures; see section 1.4) and also their structure stability [172]. Substrates $T_1G_4T_3$, $T_1G_4T_2$, $T_1G_4T_1$ were synthesised (in addition to $T_1G_4T_4$) but only a stable folded species could be obtained for $T_1G_4T_3$ (Dr. C. M. Sanders). $T_1G_4T_3$ resolved as two main species (Figure 29): S1 slow migrating and S2 fast migrating ('T3' slow and 'T3' fast, respectively). hPIF1 HD-E binding to the two species was analysed in parallel reactions rather than together due to difficulties in resolving them, along with T50 and M_$T_1G_4T_3$ (lanes 1-3 and 16-18, Figure 46). Interestingly, the two 'T3' substrates were recovered from hPIF1 HD-E binding reactions at significantly different extents across the titration range (lanes 6-10 compared to 21-25). Focusing on Figure 46 C (bar chart shown for 100 nM) we can appreciate 'T3 slow' binding being ~ 4.5 fold more than 'T3 fast' while, relatively to the unfolded control substrate (M_$T_1G_4T_3$), the extent of 'T3 slow' recovery is 5 folds higher whereas 'T3 fast' is the same. These data suggest that the G4 fold conformation can influence hPIF1-G4 DNA binding affinity.

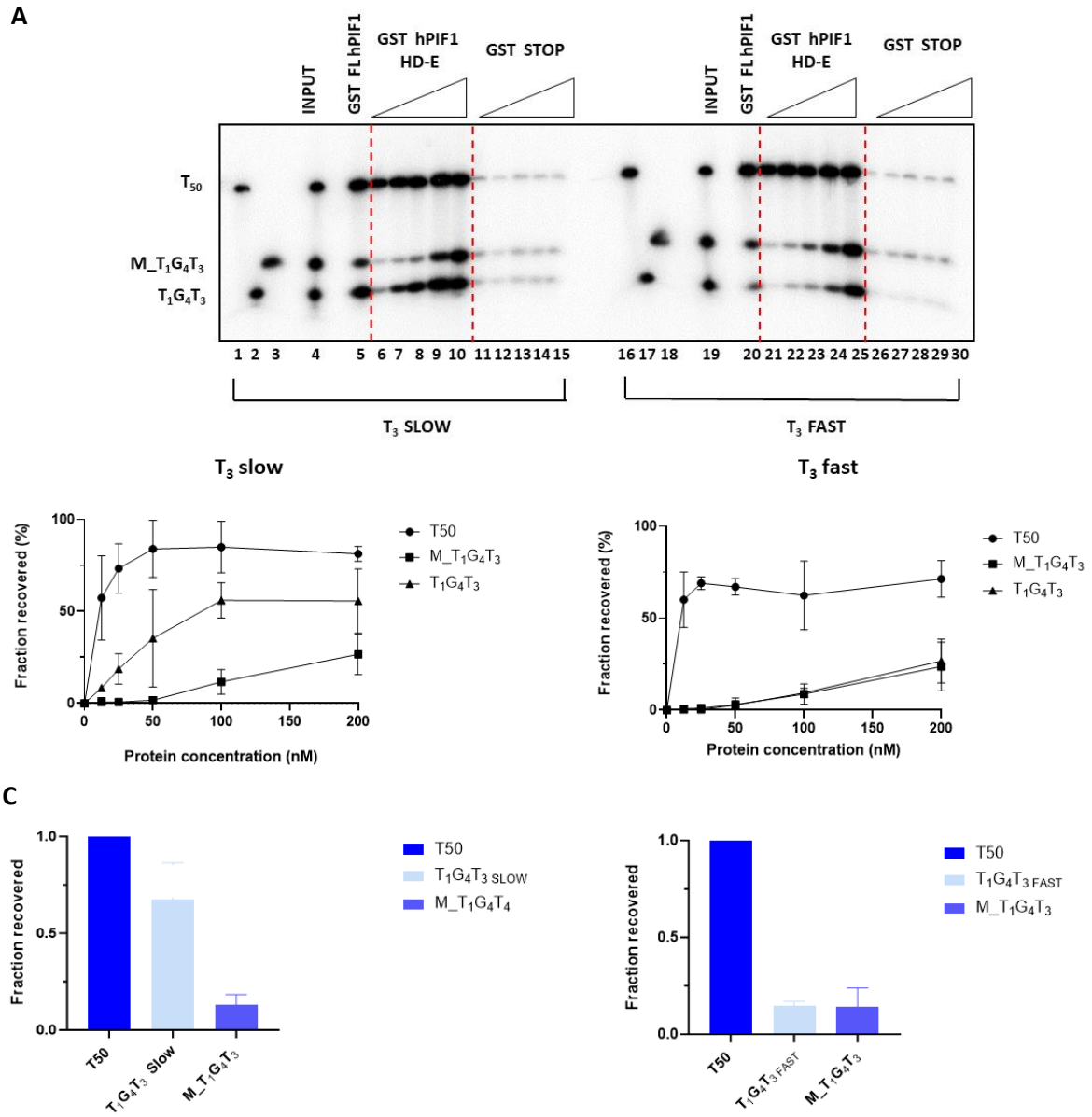


Figure 46. McKay assay investigating hPIF1 binding to $T_1G_4T_3$ and standard test substrates. (A) Acrylamide gel of products recovered from GST hPIF1 HD-END and GST STOP binding reactions (12.5, 25, 50, 100 and 200 nM) with 0.25 nM T50, M_$T_1G_4T_3$ and $T_1G_4T_3$ substrates. $T_1G_3T_4$ resolves as two species indicated as 'T3 slow' and 'T3 fast'. Lanes 1-3, markers; lane 4, 25% input; lane 5, 50 nM GST FL hPIF1 positive binding control; lanes 6-10, GST FL hPIF1 titration; lanes 11-15, GST STOP titration; lanes 16-18 markers; lane 19, 25% input; lanes 21-25, GST hPIF1 HD-END titration; lanes 26-30, GST STOP titration. (B) Data plots for the two substrates (n=3 repeats). (C) Binding data normalised to T50 at 100 nM protein concentration.

3.3.8 hPIF1 ATP*ase* activity and competitor challenge

hPIF1 is a DNA-dependent ATP*ase* [145]. To complement the substrate binding studies the stimulation of hPIF1 ATP*ase* activity by DNA substrates was investigated further. In particular, whether unimolecular G4 DNA would stimulate ATP*ase* activity is not known.

hPIF1 is not stable at 37°C and has a very low melting temperature (unpublished data). First, time course assays at 22°C and 37°C using a T50 ssDNA substrate were performed (we know hPIF1 ATP*ase* activity is stimulated by long ssDNA substrates [145]). Although an initial rate of ATP hydrolysis was greater at 37°C compared to 22°C, after 15 minutes at 37°C, hPIF1 loses all ATP*ase* activity reaching a plateau at 10 nmol ATP hydrolyzed (100 nmol per assay), whereas at 22°C ATP hydrolysis increases further with time (Figure 47 A).

Two hPIF1 ATP*ase* mutants were also analyzed: Glu307Gln (Walker B mutant, directly involved in ATP hydrolysis) and Lys234Ala (Walker A mutant, involved in ATP binding). Both mutant proteins do not show activity, as expected, when analyzed at 22°C along with the wild-type protein (Figure 47 B).

Two unimolecular G4 substrates, $T_0G_4T_4$ and $T_1G_4T_4$, previously analyzed in binding assays were assayed in parallel with T50 and their unfolded mutants M_$T_0G_4T_4$ and M_$T_1G_4T_4$ (Figure 47 C and D). The folded G4 DNA substrates stimulate very poorly ATP hydrolysis compared to T50 and their mutant ssDNA controls. However, the nature of this low-level G4 DNA stimulated activity is unclear and could be due to unfolded DNA.

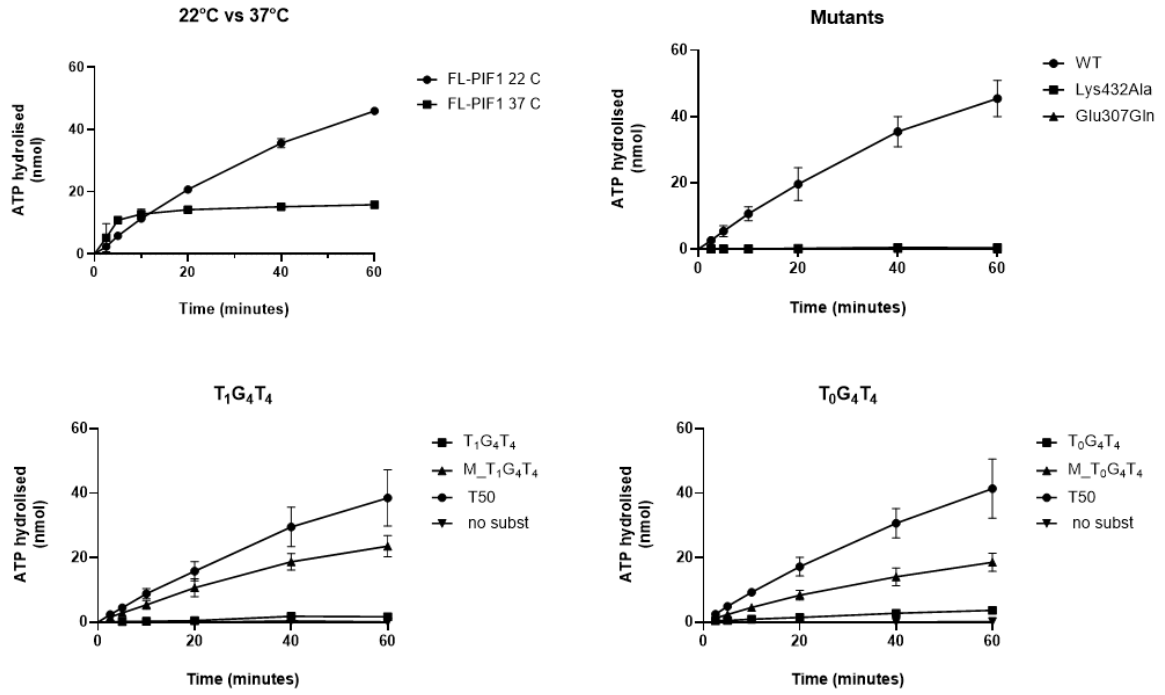


Figure 47. hPIF1 ATP*ase* activity (n=3 repeats). (A) hPIF1 ATP*ase* is temperature sensitive. (B) Two ATPase mutants, Glu307Gln and Lys234Met, were investigated showing no ATPase activity. (C&D) Analysis on G4 substrates, respectively $T_1G_4T_4$ and $T_0G_4T_4$. The G4 substrates stimulate a low level of hPIF1 ATP*ase* activity.

The unexpected activity of the ssDNA binding channel mutants observed later in section 3.4.5, suggests an overlap between the ssDNA and G4 DNA binding site. This was further investigated with a competitor challenge experiment. A McKay assay using 0.25 nM substrates T50, $T_1G_4T_4$ and its mutant M_$T_1G_4T_4$ was performed with 50 nM hPIF1 HD. The reaction was challenged with free $T_0G_4T_4$ and the single strand M_ $T_0G_4T_4$ competitors in a range from 0 nM to 500 nM (0.25, 1.25, 7.5, 25, 125, 500 nM), *i.e.,* to a 10-fold excess over protein. As shown in Figure 48, significant competition in binding to the radiolabeled substrate was only observed at 25 nM or greater unlabeled competitor (lanes 9 and 16), as expected. Surprisingly, the G4 DNA competitor (lanes 6-11) was an effective competitor for binding of all substrates, with a preference for M_$T_0G_4T_4$, but also T50, where competition binding was effectively complete at 500 nM (10-fold molar excess over protein). For the M_$T_0G_4T_4$ competitor (lanes 13-18), binding of labelled T50 was relatively unaffected, even at the highest concentrations of competitor (lane 18), consistent with the length-dependent affinity of ssDNA binding ([145]; Figure 36). However, unlabeled M_$T_0G_4T_4$

was an effective competitor for the homologous substrate, as expected, but also the G4 DNA binding at higher concentrations (lanes 16-18 and (B)). Although the experiment was subject to a high degree of experimental variability (Figure 48 B), these trends were born out in several repeats. The best interpretation of these results is that a high-affinity G4 DNA binding site overlaps with the ssDNA binding site.

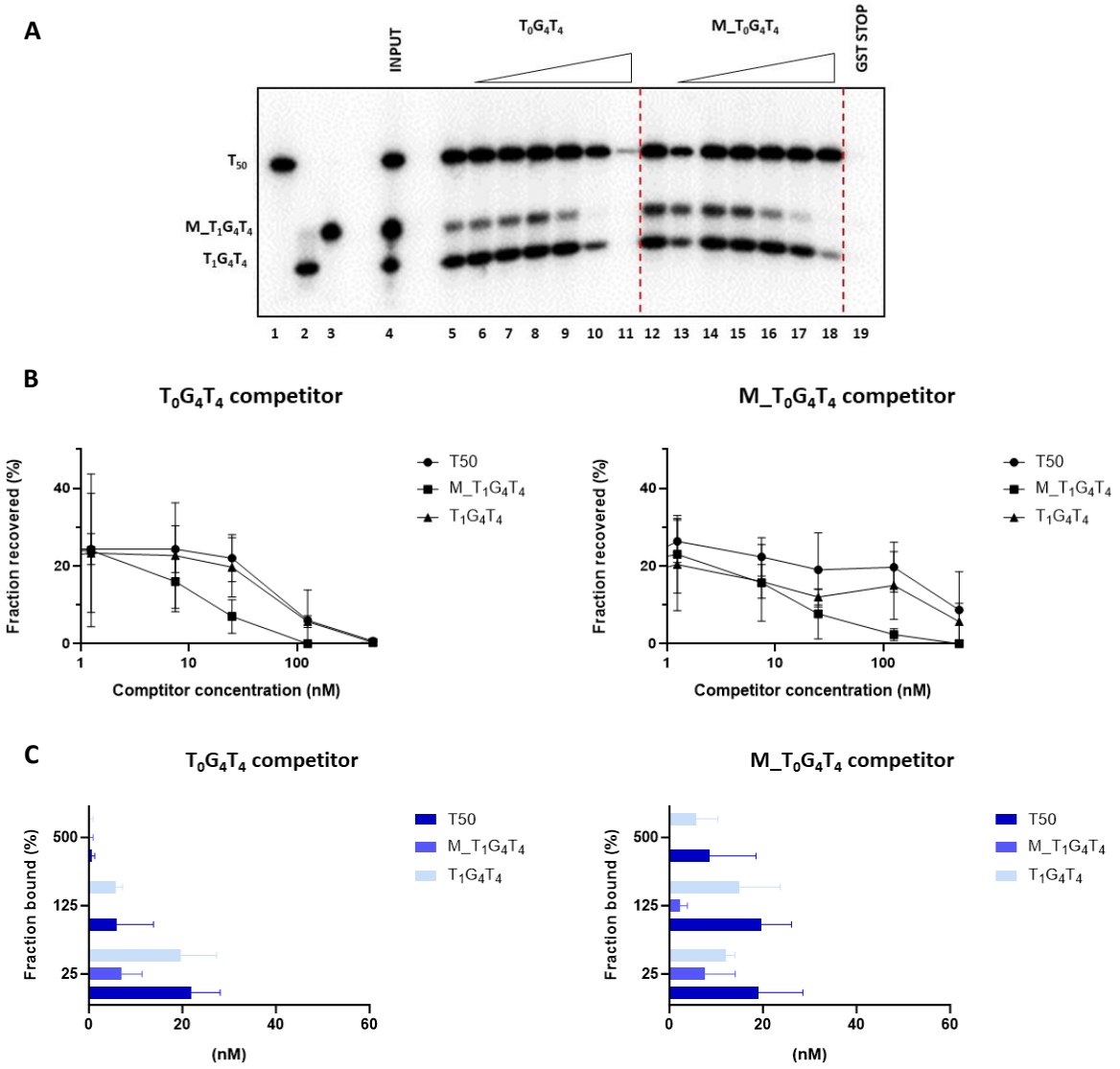


Figure 48. McKay assay investigating the effect of a competitor DNA in hPIF1 binding. (A) The acrylamide gel shows products recovered from binding reactions with 50 nM GST hPIF1 HDΔC206, ^{32}P labelled T50, $T_1G_4T_4$ and M_$T_1G_4T_4$ substrates (0.25 nM), challenged with $T_0G_4T_4$ and M_$T_0G_4T_4$ in an increasing concentration (0, 0.25, 1.25, 7.5, 25, 125, 500 nM). Lanes 1-3, markers; lane 4, 25% input; lanes 5-11, $T_0G_4T_4$ competitor (0-500 nM); lanes 12-18, M_$T_0G_4T_4$ competitor (0-500 nM); lane 18, 50 nM GST STOP with no competitor as background control. (B) Data plots for the DNA recovered (n= 3

3.3.9 Effect of the G4 DNA stabilizer Phen-DC_3 on hPIF1-G4 DNA binding

Phen-DC_3 is a ligand, highly selective for binding G4 DNA binding over ssDNA and dsDNA, able to inhibit G4 DNA unwinding by helicases [173]. It stabilises G4 DNA structures with different topologies and conformations [174], stacking externally on the 5' quartet of a parallel G4 DNA in human c-myc promoter [173]. Also, it is specifically known for its high affinity for the mini satellite CEB1 G4 DNA and its ability to inhibit its unwinding by the yeast Pif1 helicase [175]. It is also able to induce human telomeric DNA to fold into an antiparallel G4 quadruplex structure [173].

Here, I asked if Phen-DC_3 could affect the binding of hPIF1 to G4 DNA. The compound was added to binding reactions in the range 0 µM -0.5 µM with 50 nM GST FL hPIF1, hPIF1HD and GST STOP and 0.25 nM T50, $T_1G_4T_4$ and M_$T_1G_4T_4$ substrates, and the standard pull-down assay was performed (Figure 49). No changes in binding for the long ssDNA substrate T50 were observed in the presence of Phen-DC3, for either protein construct. However, the presence of Phen-DC_3 decreases the binding of both $T_1G_4T_4$ and M_$T_1G_4T_4$ substrates to hPIF1 full-length and core helicase domain. In Figure 49 C, a comparison between no stabilizer and 0.3 µM Phen-DC_3 shows how Phen-DC_3 affects hPIF1 binding to the G quadruplex, with a decrease of ~ 6 folds for the full-length protein and ~3 fold for the core domain. Loss of ssDNA affinity for is M_$T_1G_4T_4$ also observed, but not for T50 (note, for hPIF1 HD the affinity for T50 seems to be promoted by the addition of stabilizer but the experiment here shown is one repeat).

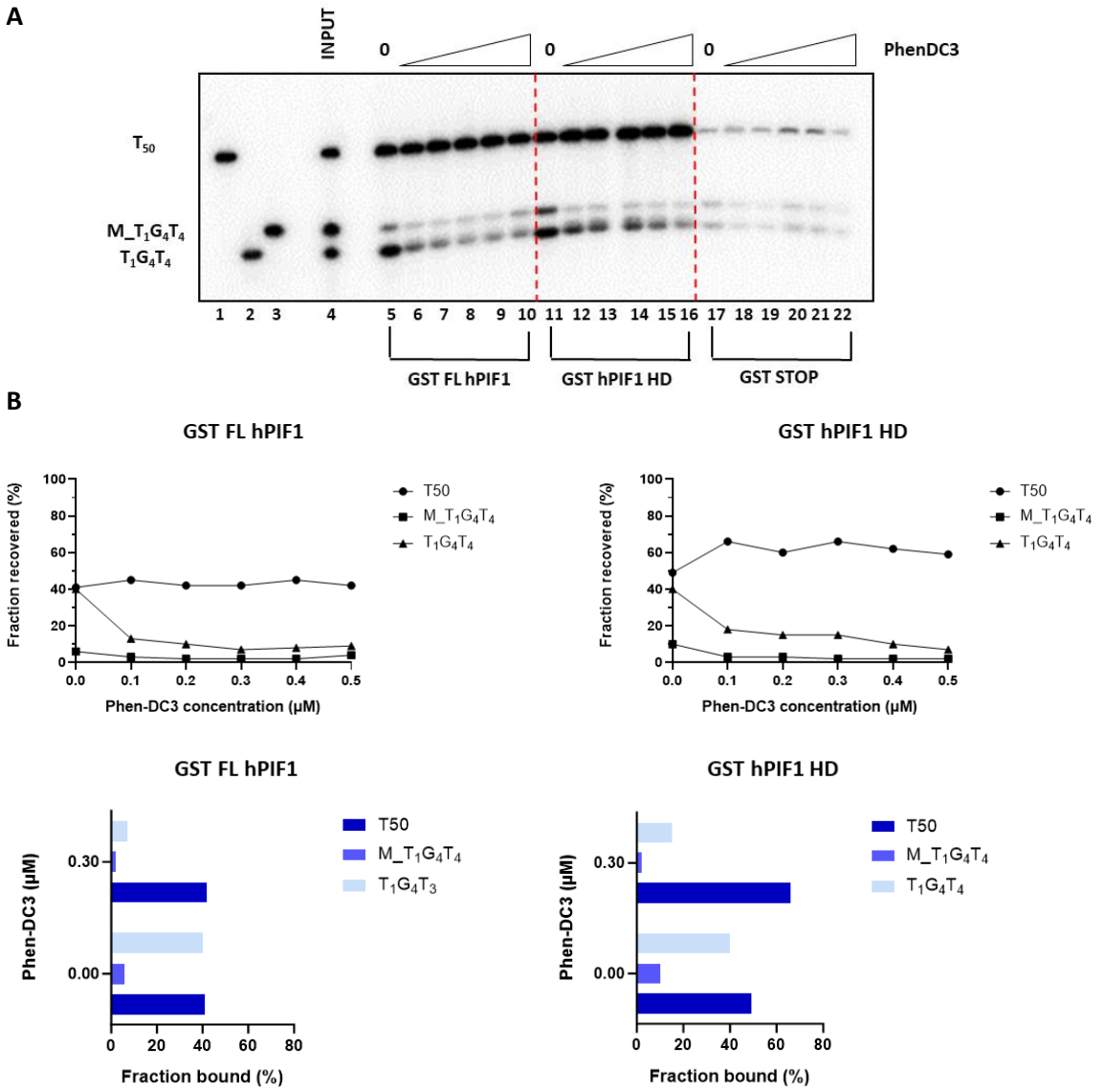


Figure 49. McKay assay investigation the effect of Phen-DC3 on hPIF1 DNA binding. (A) Acrylamide gel of reaction products recovered from binding fractions with 50 nM FL hPIF1, hPIF1 HD and GST STOP and T50, $T_0G_4T_4$ and M_$T_0G_4T_4$ substrates (0.25 nM), challenged with the presence of Phen-DC3 (0, 0.1, 0.2, 0.3, 0.4, 0.5 μM). Lanes 1-3, markers; lane 4, 25% input; lanes 5-10, reactions with increasing Phen-DC3 (0-0.5 μM) and 50 nM GST FL hPIF1; lanes 11-16, reactions with increasing Phen-DC3 (0-0.5 μM) and 50 nM GST hPIF1ΔC26; lanes 17-22, Phen-DC3 (0-0.5 μM) with 50 nM GST STOP. (B) Data plot of the analysed binding reactions (n=1 repeats). (C) Data plot of 0.3 μM and 0 PhenDC3 with full length and core domain protein.

3.4 Testing the *T. oshimai* PIF1 G4 DNA binding model with hPIF1

3.4.1 Construction mutant PIF1 ORFs and production of variant recombinant proteins for functional analysis

Based on a BioRxiv preprint describing a *T. oshimai*-G4 DNA-ADP·MgF_4^- crystal structure that became available during this study (August 2021), and published later (July 2022) [170], mutant hPIF1 ORFs for the generation of recombinant protein were constructed to test a model for hPIF1 G4 DNA binding. Figure 48 A shows the structure determined for *To*PIF1 bound to a G4 DNA substrate (pdb 7OAR), which identified domains 1B and 2B as part of a G4 DNA recognition surface, and the specific residues within that recognize the G4 DNA structure including individual tetrads. Figure 50 B shows the structure of the DNA substrate bound in 7OAR. The model proposed and schematized in Figure 50 C is called the 'sandwich model', with the G quadruplex structure between two PIF1 proteins. It was proposed that PIF1 on the 5' DNA end-segment is unwinding and bound to G4 DNA, while the other bound at the 3' DNA end segment mimics a post-catalytic binding state, bound to ssDNA only. Residues in *To*PIF1 that interact with the G4 DNA substrate are indicated with semi-transparent boxes in Figure 51. Notably, the authors (Dai *et al.*, 2022) highlighted the importance of a 5' ssDNA residue immediately adjacent to the first G of the G4 DNA for substrate recognition, thus potentially defining the minimal G4 DNA binding substrate as $N_1G{\geq}3N_XG{\geq}3N_XG{\geq}3N_XG{\geq}3$. Residues were selected for site-directed mutagenesis (SDM) following an analysis of available structures and a bioinformatics analysis of PIF1 sequences. Note, that atomic coordinates for the *To*PIF1-G4 DNA structure were not released until publication of the article in August 2022.

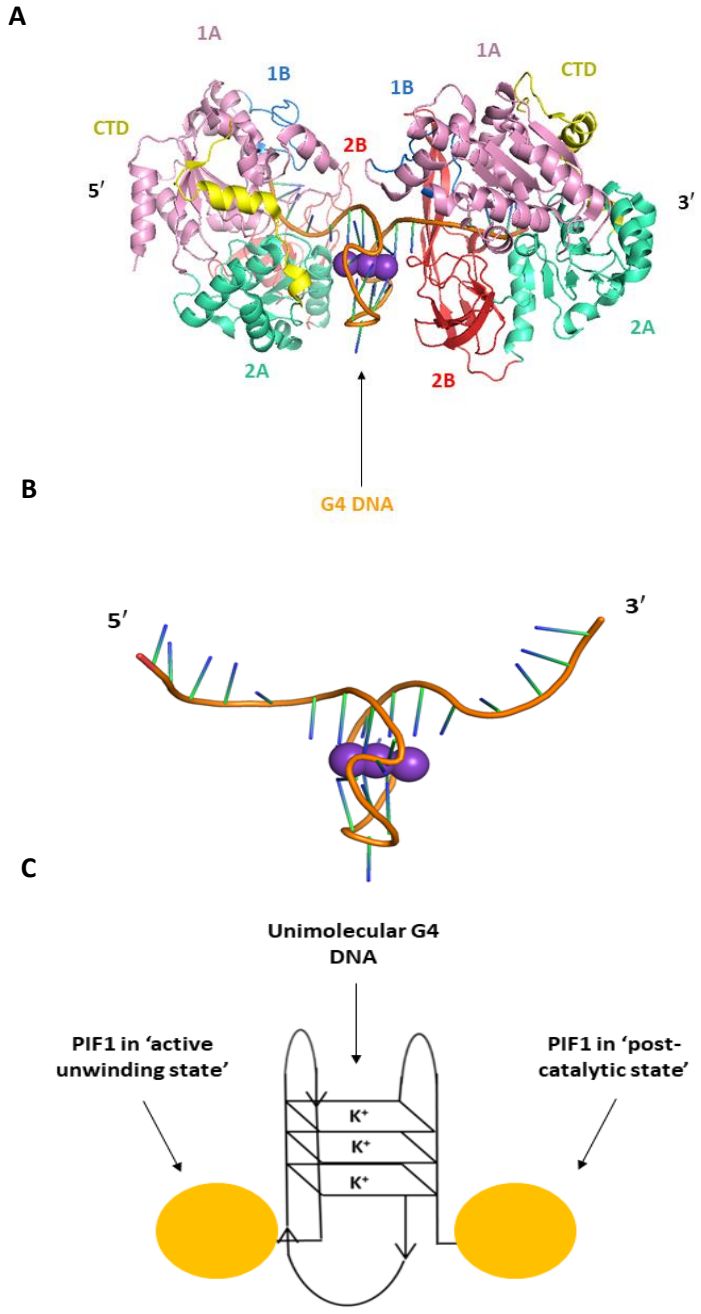
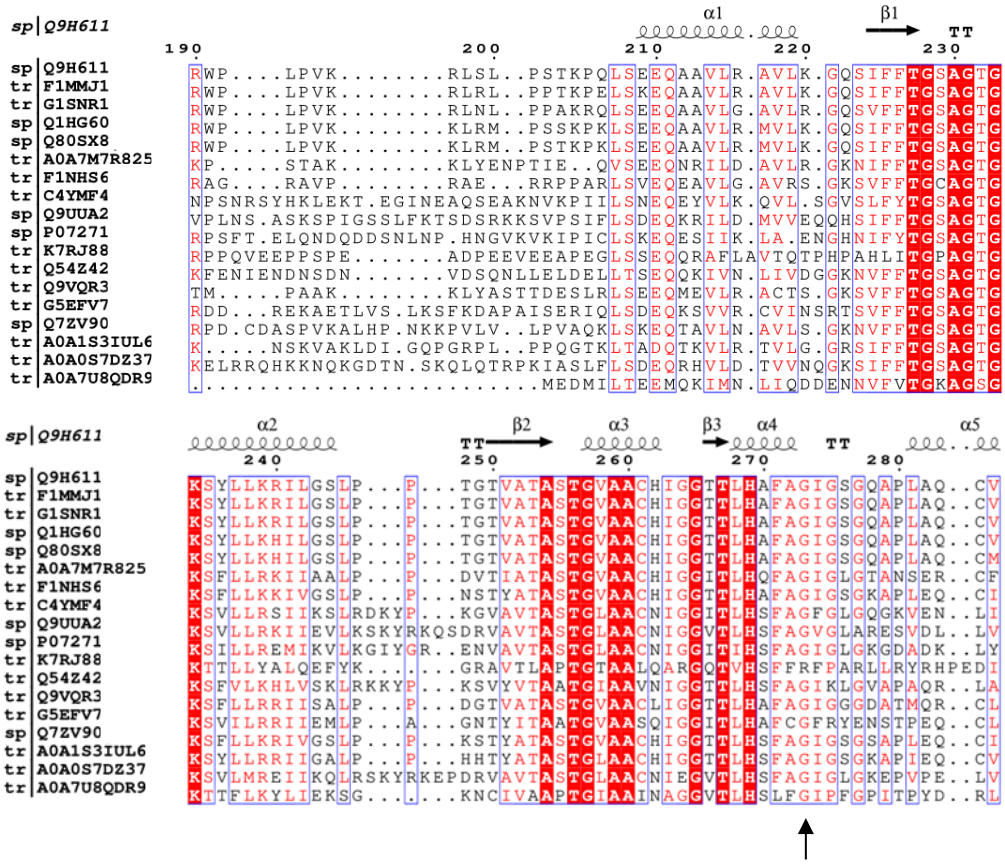


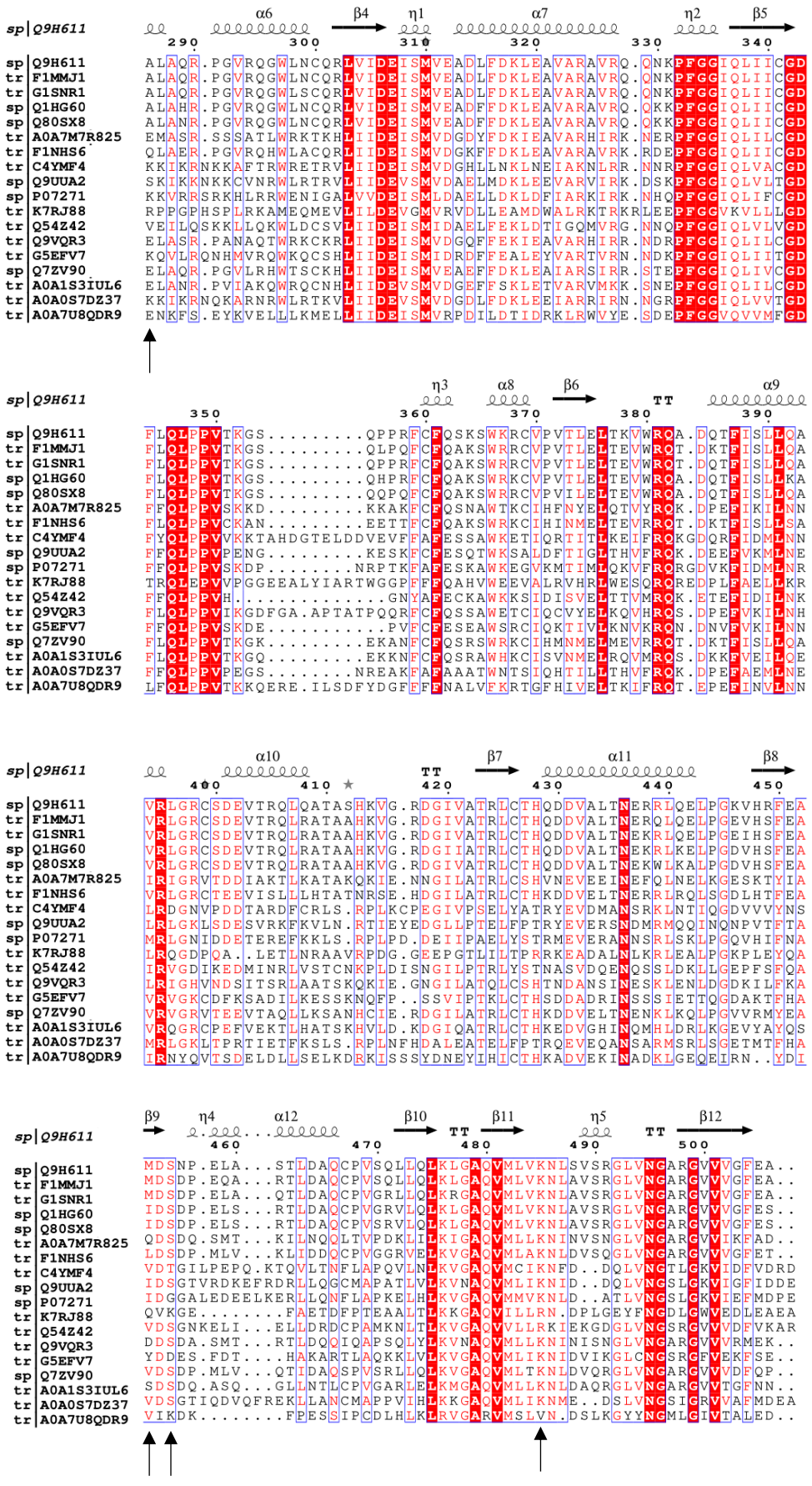
Figure 50. *To*PIF1 binding to G4 DNA and ADP•AlF_4^- (pdb: 7oar). (A) The overall structure of the *To*PIF1-G4 DNA-ADP•AlF_4 complex with the two PIF1 molecules binding one to the 5'- and one to the 3' ssDNA components of the G4 DNA substrate used. PIF1 domains are represented as follows: 1A in pink, 1B in blue, 2A in cyan, 2B in red, C terminal (CTD) in yellow and G4 DNA in orange with purple K^+ ions. Picture adapted from [7] using Pymol. (B) Representation of the G4 DNA bound in 7OAR. The structure is comprised of a 5' tail of 6 thymidine nucleotides a 3' tail of 8 thymidine nucleotides and a G4 DNA forming segment: TTTTTTGGGTGGGTGGGTGGGTTTTTTTT. The G quadruples are made of three stacks, and each is stabilised by a K^+ ion; the loops between the stacks are made of a T_1 residue. (C) cartoon diagram of the 'sandwich model'.

Figure 51. *To*PIF1 residues involved in G4 binding. The critical residues for G4 DNA binding identified in 7AOR are highlighted in three different semi-transparent colours according to their interaction with the G quadruplex DNA structure. (i) in red the residues interacting with the 5' tail residue only; (ii) in blue the residues interacting both with one of the tetrads and the first 5' tail residue adjacent to the G4 DNA; (iii) in green the residues interacting with one of the tetrads only. The *To*PIF1 domains are indicated in the sequence with coloured lines according to the key at the top.

3.4.2 Sequence alignment of PIF1 ORFs

Eighteen PIF1 sequences, with representatives from the animal kingdoms, were selected for multiple sequence alignment using Clustal Omega (https://www.ebi.ac.uk/jdispatcher/msa/clustalo). Included were several key model organisms (Figure 52). In Figure 52, residues interacting with the G4 DNA substrate, identified in the *To*PIF1-G4 DNA structure, are indicated with arrows below the sequence lines, Note, bacterial and most archaeal PIF1 proteins consist of helicase cores only and do not have N- and C- appendages associate with sequences in higher eukaryotes or helicase core insertions typical of yeast PIF1 (2C insertion, see Figure 14 Introduction).




sp|Q9H611
α6 β4 η1 α7 η2 β5
290 300 310 320 330 340
sp|Q9H611 ALAQR.PGVRQGWLNCQRLVIDEISMVEADLFDKLEAVARAVRQ.QNKPFGGIQLIICGD
tr|F1MMJ1 ALAQR.PGVRQGWLNCQRLVIDEISMVEADLFDKLEAVARAVRQ.QNKPFGGIQLIICGD
tr|G1SNR1 ALAQR.PGVRQGWLSCQRLVIDEISMVEADLFDKLEAVARAVRQ.QNKPFGGIQLIICGD
sp|Q1HG60 ALAHR.PGVRQGWLNCQRLVIDEISMVEADFFDKLEAVARAVRQ.QKKPFGGIQLIICGD
sp|Q80SX8 ALANR.PGVRQGWLNCQRLVIDEISMVEADFFDKLEAVARAVRQ.QKKPFGGIQLIICGD
tr|A0A7M7R825 EMASR.SSSATLWRKTKHLIIDEISMVDGDYFDKIEAVARHIRK.NERPFGGIQLILCGD
tr|F1NHS6 QLAER.PGVRQHWLACQRLIIDEISMVDGKFFDKLEAVARAVRK.RDEPFGGIQLIICGD
tr|C4YMF4 KKIKRNKKAFTRWRETRVLIIDEISMVDGHLLNKLNEIAKNLRR.NNRPFGGIQLVACGD
sp|Q9UUA2 SKIKKNKKCVNRWLRTRVLIIDEVSMVDAELMDKLEEVARVIRK.DSKPFGGIQLVLTGD
sp|P07271 KKVRRSRKHLRRWENIGALVVDEISMLDAELLDKLDFIARKIRK.NHQPFGGIQLIFCGD
tr|K7RJ88 RPPGPHSPLRKAMEQMEVLILDEVGMVRVDLLEAMDWALRKTRKRLEEPFGGVKVLLLGD
tr|Q54Z42 VEILQSKKLLQKWLDCSVLIIDEISMIDAELFEKLDTIGQMVRG.NNQPFGGIQLVLVGD
tr|Q9VQR3 ELASR.PANAQTWRKCKRLIIDEISMVDGQFFEKIEAVARHIRR.NDRPFGGIQLILCGD
tr|G5EFV7 KQVLRQNHMVRQWKQCSHLIIDEISMIDRDFFEALEYVARTVRN.NDKPFGGIQLIITGD
sp|Q7ZV90 ELAQR.PGVLRHWTSCKHLIIDEISMVEAEFFDKLEAIARSIRR.STEPFGGIQLIVCGD
tr|A0A1S3IUL6 ELANR.PVIAKQWRQCNHLIIDEVSMVDGEFFSKLETVARVMKK.SNEPFGGIQLILCGD
tr|A0A0S7DZ37 KKIKRNQKARNRWLRTKVLIIDEVSMVDGDLFDKLEEIARRIRN.NGRPFGGIQLVVTGD
tr|A0A7U8QDR9 ENKFS.EYKVELLLKMELLIIDEISMVRPDILDTIDRKLRWVYE.SDEPFGGVQVVMFGD
η3 α8 β6 TT α9
350 360 370 380 390
sp|Q9H611 FLQLPPVTKGS.....QPPRFCFQSKSWKRCVPVTLELTKVWRQA.DQTFISLLQA
tr|F1MMJ1 FLQLPPVTKGS.....QLPQFCFQAKSWRRCVPVTLELTEVWRQT.DKTFISLLQA
tr|G1SNR1 FLQLPPVTKGS.....QPPRFCFQAKSWRRCVPVTLELTEVWRQT.DQTFISLLQA
sp|Q1HG60 FLQLPPVTKGS.....QHPRFCFQAKSWRKCVPVTLELTEVWRQA.DQTFISLLKA
sp|Q80SX8 FLQLPPVTKGS.....QQPQFCFQAKSWRRCVPVILELTEVWRQA.DQTFISLLQA
tr|A0A7M7R825 FFQLPPVSKKD.....KKAKFCFQSNAWTKCIHFNYELQTVYRQK.DPEFIKILNN
tr|F1NHS6 FLQLPPVCKAN.....EETTFCFQAKSWRKCIHINMELTEVRRQT.DKTFISLLSA
tr|C4YMF4 FYQLPPVVKKTAHDGTELDDVEVFFAFESSAWKETIQRTITLKEIFRQKGDQRFIDMLNN
sp|Q9UUA2 FFQLPPVPENG.....KESKFCFESQTWKSALDFTIGLTHVFRQK.DEEFVKMLNE
sp|P07271 FFQLPPVSKDP.....NRPTKFAFESKAWKEGVKMTIMLQKVFRQRGDVKFIDMLNR
tr|K7RJ88 TRQLEPVVPGGEEALYIARTWGGPFFFQAHVWEEVALRVHRLWESQRREDPLFAELLKR
tr|Q54Z42 FFQLPPVH.....GNYAFECKAWKKSIDISVELTTVMRQK.ETEFIDILNK
tr|Q9VQR3 FLQLPPVTKGDFGA.APTATPQQRFCFQSSAWETCIQCVYELKQVHRQS.DPEFVKILNH
tr|G5EFV7 FFQLPPVSKDE.....PVFCFESEAWSRCIQKTIVLKNVKRQN.DNVFVKILNN
sp|Q7ZV90 FLQLPPVTKGK.....EKANFCFQSRSWRKCIHMNMELMEVRRQT.DKTFISLLQA
tr|A0A1S3IUL6 FLQLPPVTKGQ.....EKKNFCFQSRAWHKCISVNMELRQVMRQS.DKKFVEILQE
tr|A0A0S7DZ37 FFQLPPVPEGS.....NREAKFAFAAATWNTSIQHTILLTHVFRQK.DPEFAEMLNE
tr|A0A7U8QDR9 LFQLPPVTKKQERE.ILSDFYDGFFFFNALVFKRTGFHIVELTKIFRQT.EPEFINVLNN
α10 β7 TT α11 β8
400 410 420 430 440 450
sp|Q9H611 VRLGRCSDEVTRQLQATASHKVG.RDGIVATRLCTHQDDVALTNERRLQELPGKVHRFEA
tr|F1MMJ1 VRLGRCSDEVTRQLRATAAHKVG.RDGIVATRLCTHQDDVALTNERQLQELPGEVHSFEA
tr|G1SNR1 VRLGRCSDEVTRQLRATAAHKVG.RDGIVATRLCTHQDDVALTNEKRLQELPGEIHSFEA
sp|Q1HG60 VRLGRCSDEVTRQLRATAAHKVG.RDGIIATRLCTHQDDVALTNEKRLKELPGDVHSFEA
sp|Q80SX8 VRLGRCSDEVTRQLRATAAHKVG.RDGIVATRLCTHQDDVALTNEKWLKALPGDVHSFEA
tr|A0A7M7R825 IRIGRVTDDIAKTLKATAKQKIE.NNGILATRLCSHVNEVEEINEFQLNELKGESKTYIA
tr|F1NHS6 VRLGRCTEEVISLLLHTATNRSE.HDGILATRLCTHKDDVELTNERRLRQLSGDLHTFEA
tr|C4YMF4 LRDGNVPDDTARDFCRLS.RPLKCPEGIVPSELYATRYEVDMANSRKLNTIQGDVVVYNS
sp|Q9UUA2 LRLGKLSDESVRKFKVLN.RTIEYEDGLLPTELFPTRYEVERSNDMRMQQINQNPVTFTA
sp|P07271 MRLGNIDDETEREFKKLS.RPLPD.DEIIPAELYSTRMEVERANNSRLSKLPGQVHIFNA
tr|K7RJ88 LRQGDPQA..LETLNRAAVRPDG.GEEPGTLILTPRRKEADALNLKRLEALPGKPLEYQA
tr|Q54Z42 IRVGDIKEDMINRLVSTCNKPLDISNGILPTRLYSTNASVDQENQSSLDKLLGEPFSFQA
tr|Q9VQR3 LRIGHVNDSITSRLAATSKQKIE.GNGILATQLCSHTNDANSIESKLENLDGDKILFKA
tr|G5EFV7 VRVGKCDFKSADILKESSKNQFP..SSVIPTKLCTHSDDADRINSSSIETTQGDAKTFHA
sp|Q7ZV90 VRVGRVTEEVTAQLLKSANHCIE.RDGILATRLCTHKDDVELTNENKLKQLPGVVRMYEA
tr|A0A1S3IUL6 VRQGRCPEFVEKTLHATSKHVLD.KDGIQATRLCTHKEDVGHINQMHLDRLKGEVYAYQS
tr|A0A0S7DZ37 MRLGKLTPRTIETFKSLS.RPLNFHDALEATELFPTRQEVEQANSARMSRLSGETMTFHA
tr|A0A7U8QDR9 IRNYQVTSDELDLLSELKDRKISSSYDNEYIHICTHKADVEKINADKLGEQEIRN..YDI
β9 η4 α12 β10 TT β11 η5 TT β12
460 470 480 490 500
sp|Q9H611 MDSNP.ELA...STLDAQCPVSQLLQLKLGAQVMLVKNLSVSRGLVNGARGVVVGFEA..
tr|F1MMJ1 MDSDP.EQA...RTLDAQCPVSQLLQLKLGAQVMLVKNLAVSRGLVNGARGVVVGFEA..
tr|G1SNR1 MDSDP.ELA...QTLDAQCPVGRLLQLKRGAQVMLVKNLAVSRGLVNGARGVVVGFEA..
sp|Q1HG60 IDSDP.ELS...RTLDAQCPVGRVLQLKLGAQVMLVKNLAVSRGLVNGARGVVVGFES..
sp|Q80SX8 IDSDP.ELS...RTLDAQCPVSRVLQLKLGAQVMLVKNLAVSRGLVNGARGVVVGFES..
tr|A0A7M7R825 QDSDQ.SMT...KILNQQLTVPDKLILKIGAQVMLLKNINVSNGLVNGARGVVIKFAD..
tr|F1NHS6 LDSDP.MLV...KLIDDQCPVGGRVELKVGAQVMLAKNLDVSQGLVNGARGVVVGFET..
tr|C4YMF4 VDTGILPEPQ.KTQVLTNFLAPQVLNLKVGAQVMCIKNFD..DQLVNGTLGKVIDFVDRD
tr|Q9UUA2 IDSGTVRDKEFRDRLLQGCMAPATLVLKVNAQVMLIKNID..DQLVNGSLGKVIGFIDDE
sp|P07271 IDGGALEDEELKERLLQNFLAPKELHLKVGAQVMMVKNLD..ATLVNGSLGKVIEFMDPE
tr|K7RJ88 QVKGE.....FAETDFPTEAALTLKKGAQVILLRN.DPLGEYFNGDLGWVEDLEAEA
tr|Q54Z42 VDSGNKELI...ELLDRDCPAMKNLTLKVGAQVVLLRKIEKGDGLVNGSRGVVVDFVKAR
tr|Q9VQR3 DDSDA.SMT...RTLDQQIQAPSQLYLKVNAQVMLLKNINISNGLVNGARGVVVRMEK..
tr|G5EFV7 YDDES.FDT....HAKARTLAQKKLVLKVGAQVMLIKNIDVIKGLCNGSRGFVEKFSE..
sp|Q7ZV90 VDSDP.MLV...QTIDAQSPVSRLLQLKVGAQVMLTKNLDVQRGLVNGARGVVVDFQP..
tr|A0A1S3IUL6 SDSDQ.ASQ...GLLNTLCPVGARLELKMGAQVMLLKNLDAQRGLVNGARGVVTGFNN..
tr|A0A0S7DZ37 VDSGTIQDVQFREKLLANCMAPPVIHLKKGAQVMLIKNME..DSLVNGSIGRVVAFMDEA
tr|A0A7U8QDR9 VIKDK.....FPESSIPCDLHLKLRVGARVMSLVN.DSLKGYYNGMLGIVTALED..

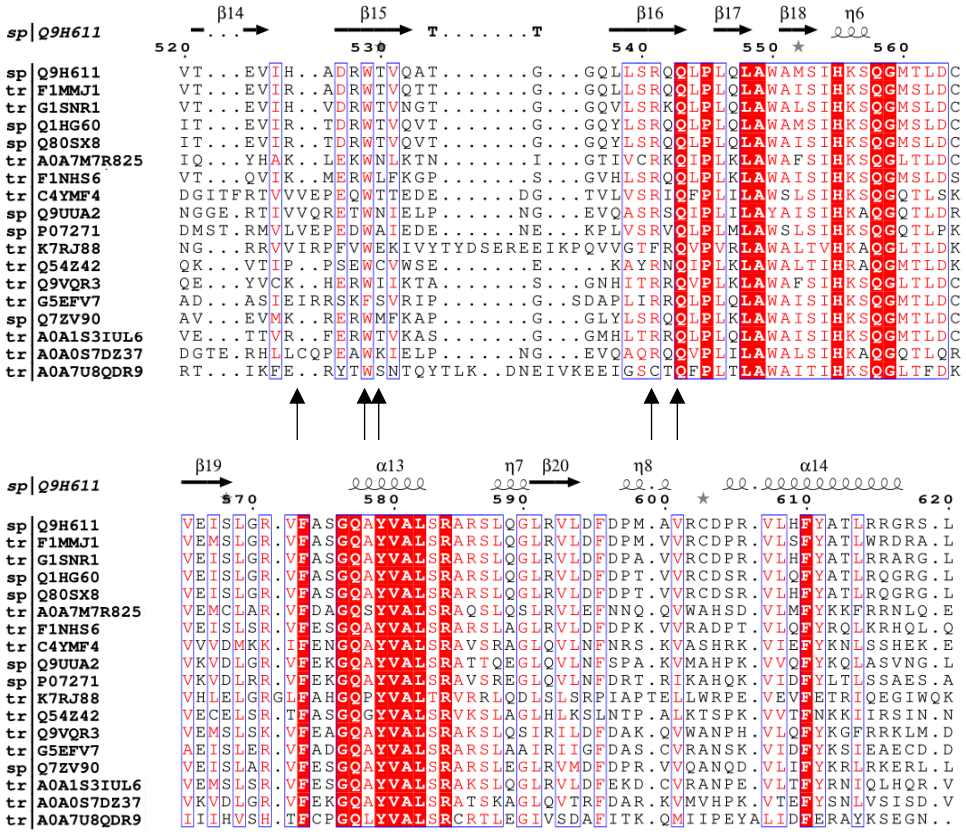


Figure 52. PIF1 sequence alignment. The first line, PIF1_HUMAN is the sequence of the human PIF1 protein (nuclear, isoform a) aligned against all other sequences (*i.e.* residues numbered are from hPIF1). Identical residues are boxed in red, similar residues are in red and non-conserved residues are in black. The blue framed regions have a global similarity of 0.7. A black arrow indicates the residues binding the G4 DNA substrate in the *To*PIF1 structure, identifying the human PIF1 residues considered in this analysis. The sequences of the model organism analysed were taken from the UniProt database (https://www.uniprot.org/). Code Q9H611 for *Homo sapiens* PIF1 protein; F1MMJ1 for *Bos taurus*; G1SNR1 for *Oryctolagus cuniculus*; Q1HG60 for *Rattus norvegicus*; Q80SX8 for *Mus musculus*; A0A7M7R852 for *Apis mellifer*; F1NHS6 *Gallus gallus*; C4YMF4 for *Candida albicans*; Q9UUA2 *Schizosaccharomyces pombe*; P07271 *Saccharomyces cerevisiae*; K7RJ88 for *Thermos oshimai*; Q54Z42 for *Dictyostelium discoideum*; Q9VQR3 for *Drosophila melanogaster*; G5EFV7 for *Caenorhabditis elegans*; Q7ZV90 for *Danio rerio*; A0A1S3IUL6 for *Lingual unguis*; A0A0S7DZ37 for *Aspergillus* lentulus; A0A7U8QDR9 for *Bacteroides fragilis*. Secondary structural elements of hPIF1 are indicated on top of the sequences as α, β and η; TT indicates the β turn. The *Thermus oshimai* sequence is indicated as "Thermus".Sequences were aligned with Clustal Omega and the final picture prepared with ESPript 3.0.

3.4.3 Defining potential G4 DNA binding residues in hPIF1 and a rationale for the choice of residues to mutate

All interactions observed between *To*PIF1, and G4 DNA are with the first unpaired 5' T residues (“5' T”) abutting the G4 DNA segment of the substrate or with the G quadruplex tetrads. *To*PIF1 R135 interacts with 5' T and the second (middle) tetrad of the G4 DNA at the entrance of the ssDNA binding channel between the Rec domains. The corresponding residue in humans is Gly273 and this Gly is conserved in vertebrates, invertebrates and yeast (Figure 52). It is unlikely that Gly273 can fulfill a similar role to Arg135, while no residue with similar properties is present at a similar position in the structure.

*To*PIF1 Arg150, interacting with the 3' most tetrad, is in the wedge (1B) domain of PIF1, previously implicated as a potential “plough-like” structure involved in pushing dsDNA apart [149]. This residue is not conserved (hPIF1 Ala286). Even though the wedge domain is a significant structural and functional feature of the PIF1 protein, it shows high sequence variability. However, the role of *To*PIF1 Arg150 could be fulfilled by Arg290 in human PIF1. Although not at a similar structural position (Figure 53) the residue shows a high level of conservation in vertebrate, invertebrate, and yeast PIF1 and the corresponding residue in yeast (*Saccharomyces*, Arg324) has been implicated previously in PIF1-G4 DNA interactions [149].

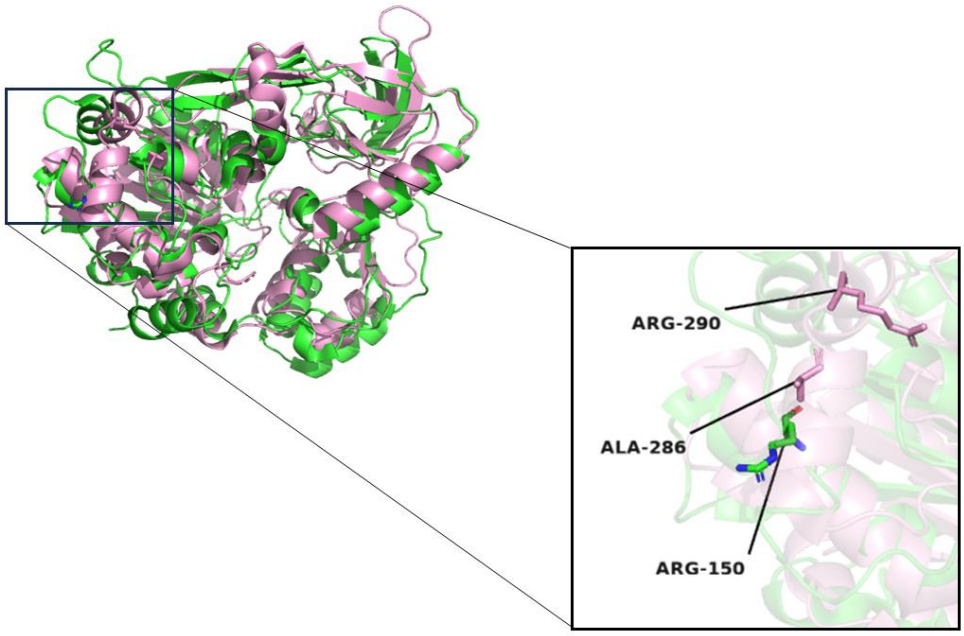


Figure 53. Structural alignment of hPIF1 (pink) and *To*PIF1 (green) with focus on Arg290/Ala286/Arg150, residue (wedge /1B domain). Apo form of the human (pdb: 6hpt) and *To* proteins (pdb: 6s3e) were aligned. ToPIF1 Arg150 interacts with the third (3') tetrad of the G quadruplex structure, it is not conserved but close to the residue equivalent to human Arg290. The enlarged box represents the close-up view of the residues Arg290 (in pink), Ala286 (in pink) and Arg150 (in green), highlighting the distance between the two residues.

Several residues from domain 2B interact with 5' T and G tetrads. Gln327 interacts with the 5' G tetrad PO_4 and is substituted with Met453 in humans (not conserved). Phe450 at an adjacent position could potentially have a function but in a different way, possibly through π-π stacking with bases. *To*PIF1 Lys329 is conserved in bacterial PIF1. The corresponding residue is serine or threonine in mammals and fungi, but glycine or aspartate in *S. cerevisiae* and *C. elegans* respectively. Analogous cationic-π interactions with G4 DNA are unlikely in hPIF1, but hydrogen bonding with DNA is in principle possible.

*To*PIF1 R355 is in motif B of domain 2B and conserved (lysine or arginine). The corresponding residue in humans is Lys485, and this residue has a role in hPIF1 ssDNA binding [135], consistent with the observed interaction, exclusively, with the 5' T residue of the G4 DNA structure.

Notably, the G4 DNA structure lies against the rail hairpin. In the rail hairpin, Arg392 interacts with the 5'-ssDNA residue and the 5' G tetrad via cationic π

interactions. In the *To*/human structural superposition (Figure 52), His525 is at the same position. While the residue is mostly Arg/Lys/His in all vertebrate and bacterial sequences aligned, in the yeast sequences however, Val or Leu is found at the corresponding position. His525 could substitute for *To* Arg392 and have a similar function in binding the G4 structure, but the residue is not conserved.

*To*PIF1 Phe394 is mostly substituted with glutamic acid, aspartic acid and occasionally serine or threonine in most other sequences (hPIF1 Asp527). However, adjacent Trp529 is at a similar structural position to *To* Phe394 and is conserved (with occasional substitutions to F). Trp529 could be involved in π-π stacking interactions with the ssDNA 5' T residue. *To*PIF1 Val395 interacts with the 5' T residue. Lysine or arginine is found at this position in mammals and is a candidate for electrostatic interactions with the nucleotide. Finally, Arg419 is substituted with Gln542 in humans. Glutamine at this position is conserved in mammals and often there is a polar residue present in other sequences (not conserved).

In summary, hPIF1 Lys485 is conserved and a candidate for interaction with the first 5' nucleotide abutting G4 DNA (5' N). Trp452 may also interact with 5' N, while His525 could, in principle, perform the role of *To*PIF1 Arg392 and interact with a G tetrad and 5' N. Figure 53 summarizes the information given above and highlights corresponding residues between human and *To*PIF1. Figure 55 shows a structural alignment of the two structures (human and *To*) with the residues involved in the G quadruplex binding represented with sticks. Table 6 is a comparison between residues involved in hPIF1 ssDNA binding analyzed in a previous study [135] and the residues identified as interacting with G4 DNA in the *To*PIF1 structure, highlighting the relative lack of conservation in the proposed G4 DNA interaction [170].

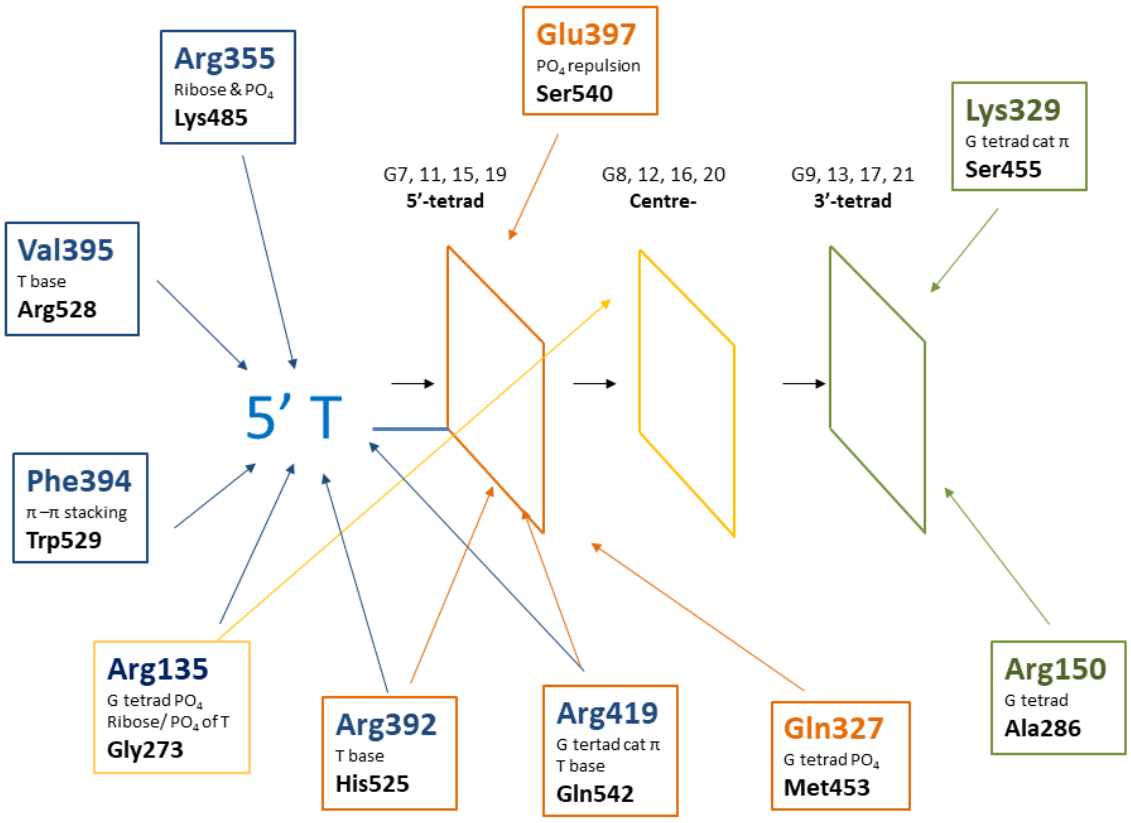


Figure 54. *To*PIF1 residues involved in G4 binding according to the Dai *et al.* study [7] (pdb 7OAR). The figure shows how key *To*PIF1 residues binds the G4 DNA and specifically whether they interact the single 5' T ss DNA residue (blue arrows), the 5' G tetrad (orange arrows), the middle tetrad (yellow arrows) or the 3' tetrads (green arrows). Corresponding human PIF1 residues are in black.

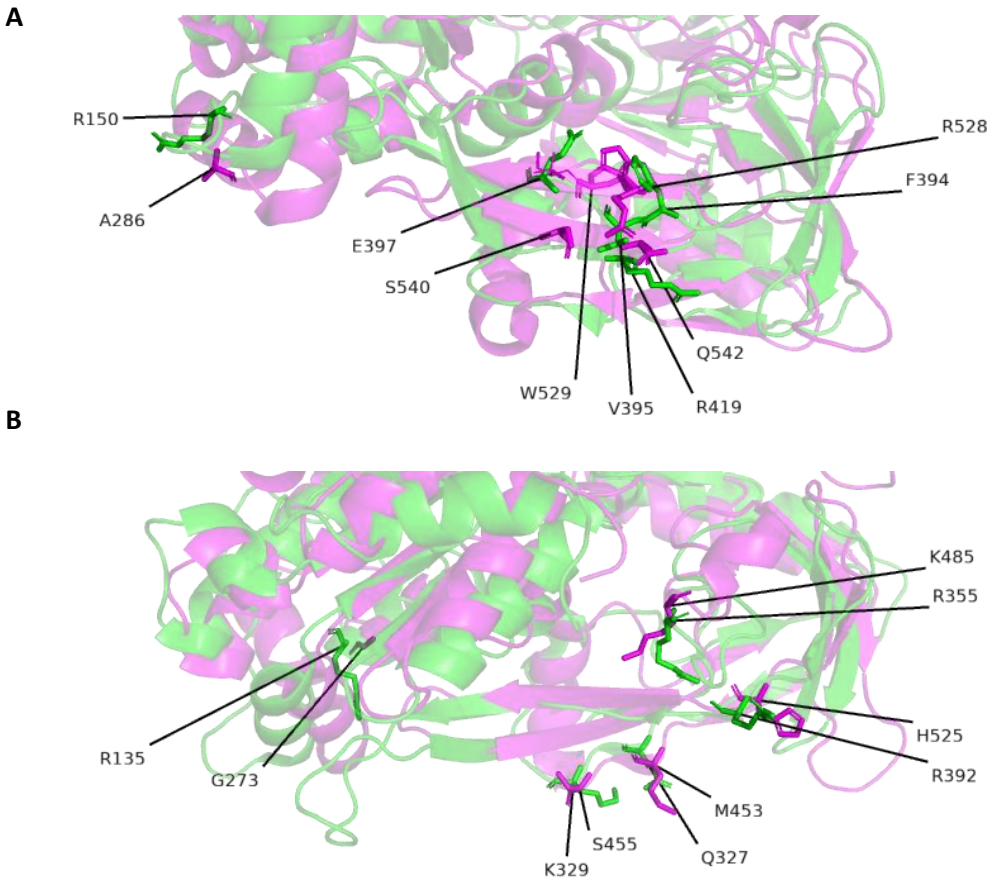


Figure 55. Structural alignment of hPIF1 (pdb: 6hpt, pink) and *To*PIF1 (pdb: 6s3e, green) structures in apo condition with a focus on the residues involved in the G4 binding. (A) Focus on residues in domains 2B and 1A. (B) Focus on residues in the rail β-hairpin and wedge.

hPIF1 residue	domain	Motif/ functional site	BsPIF1 residues	ScPIF1 residues	ToPIF1 residues
His269	1A	ssDNA binding channel	His68	His393	His131
Asn495	2B	ssDNA binding, C motif	Asn296	Asn533	Asn357
His555	2A	ssDNA binding channel	His361	His705	His432
Lys556	2A	ssDNA binding channel	Lys362	Lys706	Lys433
Phe573	2A	ssDNA binding channel	Phe379	Phe723	Phe451
Lys485	2B	B motif	Val287	Lys525	Arg355
Asn486	2B	ssDNA binding, B motif	Asn288	Asn526	Asn356

Table 6. Summary of the hPIF1 residues involved in ssDNA binding. Previous analysis identified the critical residues in hPIF1 involved in ssDNA binding [6]. In the table, the *Bacteroides*, yeast and *Thermus oshimai* ssDNA binding residues are reported, highlighting the high (near complete) degree of conservation between species, further confirmed in Fig alignment 35 for a wider range of species.

3.4.4 Construction of hPIF1 mutants and variant proteins

Table 7 summarizes the domain and motif location of the residues mutated for this study and discussed above. Two existing mutants in the ssDNA binding channel, Lys556A and Lys556E, previously characterized in the lab [135], were employed as a control since they are known to be defective in ssDNA binding and therefore predicted not to be defective in G4 DNA binding, allowing separation of function. All the ORFs were *E. coli* codon-optimized and constructed *de novo* by PCR-directed mutagenesis (section 2.1.2). The ORFs were first cloned in a pET His_{12}SUMO vector and subsequently sub-cloned into a pET-GST expression vector. Sequences were confirmed by Sanger DNA sequencing (Genewiz). All the proteins were purified from 18L of *E. coli* culture as described and discussed in section 3.1.4.

*To*PIF1 residue	hPIF1	human domain	human Motif/ functional site
Arg135	Gly273A	1A	ssDNA binding channel
Arg150	Ala286	1B	Separation wedge
Gln327	Met453A	2B	mechanical coupling of ATP hydrolysis
Arg541	Ser540A	2B	Rail β-hairpin
Lys329	Ser455A	2B	mechanical coupling of ATP hydrolysis
Arg355	Lys485E	2B	B motif
Arg392	His525A	2B	mechanical coupling of ATP hydrolysis
Phe394	Trp529A	1B	Rail β-hairpin
Val395	Arg528A	1B	Rail β-hairpin
Arg419	Gln542A	1B	Rail β-hairpin

Table 7. Summary of the *To*PIF1 residues involved in G4 DNA binding and their respective hPIF1 mutations constructed in this work. The domain and the functional site for each residue are summarized also. Note hPIF1 Ala286 is at a corresponding position to ToArg150, both in the wedge domain. However, the yeast equivalent to Arg290 has previously been implicated in G4 DNA binding and was selected for analysis here.

3.4.5 Testing hPIF1 mutants in G4 DNA binding assays

The *T. oshimai* data indicated that a set of residues distinct from those involved in ssDNA binding were responsible for G4 DNA interactions. The rationale for using the McKay assay was that mutants in the ssDNA binding channel could be assayed in parallel with mutants in the putative G4 DNA binding site with G4 and ssDNA substrates in the same binding reaction. Mutant Lys556A and a double mutant Lys556A/His555A were expected to be defective in ssDNA but not G4 DNA binding [135], while mutants in the G4 DNA binding site should show the converse. Binding reactions were performed with T50, $T_1G_4T_4$ and M_$T_1G_4T_4$ and hPIF1 HD-E.

The results obtained with these assays were unexpected. First, in Figure 56, for the ssDNA binding channel mutants Lys556A and Lys556A/His555A, where we hypothesised a decrease of affinity for T50 substrate but no change of affinity for the G4 DNA, the experiment however shows a decrease in affinity for both ssDNA and G4 DNA (lanes 10-19, Figure 56). These data are best illustrated in the graphs in Figure 54 B.

Furthermore, the mutants selected according to the *T. oshimai* model were tested (Figure 57). At first glance, the mutants worthy of attention seem to be Gln542A (lanes 15-17, Figure 57 A) and Arg528A (lanes 12-14, Figure 57 B) which show an apparent loss of affinity for the G4 substrate. Looking more carefully at the data reported in Figure 57 C (protein concentration 100 nM), it can be seen that there is a loss of affinity for the short ssDNA substrate M_$T_1G_4T_4$ also. Looking at the fraction bound to folded $T_1G_4T_4$ substrate compared to mutant M_$T_1G_4T_4$, we see a ratio of 5:1 for the wild type, 10:1 for Lys556A, 7:1 for His525A, 12:1 for Gln542A; 5:1 for Ser540A (as wild type) and 9:1 for Arg528A. Therefore, the ratios between the two substrates increases in the case of all mutants, suggesting predominantly an effect on the binding to M_$T_1G_4T_4$.

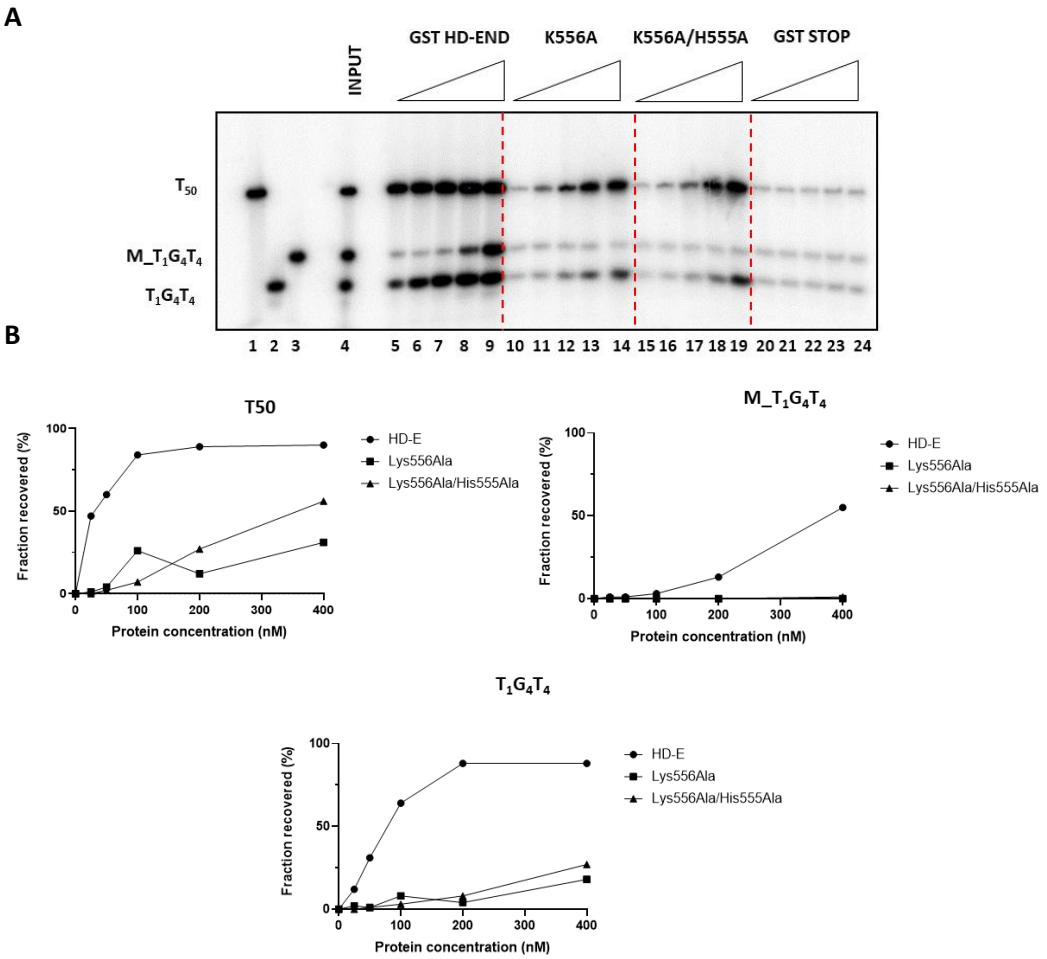


Figure 56. McKay assay with hPIF1 ssDNA binding channel mutants. The mutants were tested with 0.25 nM T50, $T_1G_4T_4$ and M_$T_1G_4T_4$ substrates, in a protein range 25, 50, 100, 200, 400 nM. Lanes 1-3 markers; lane 4, 25% input; lanes 5-9, GST HD-END titration; lanes 10-14, Lys556A titration; lanes 15-19, Lys556A/His555A titration; lanes 20-24, GST STOP titration. (B) Data plot of the wild type and mutants binding to substrates. n= 1 repeats.

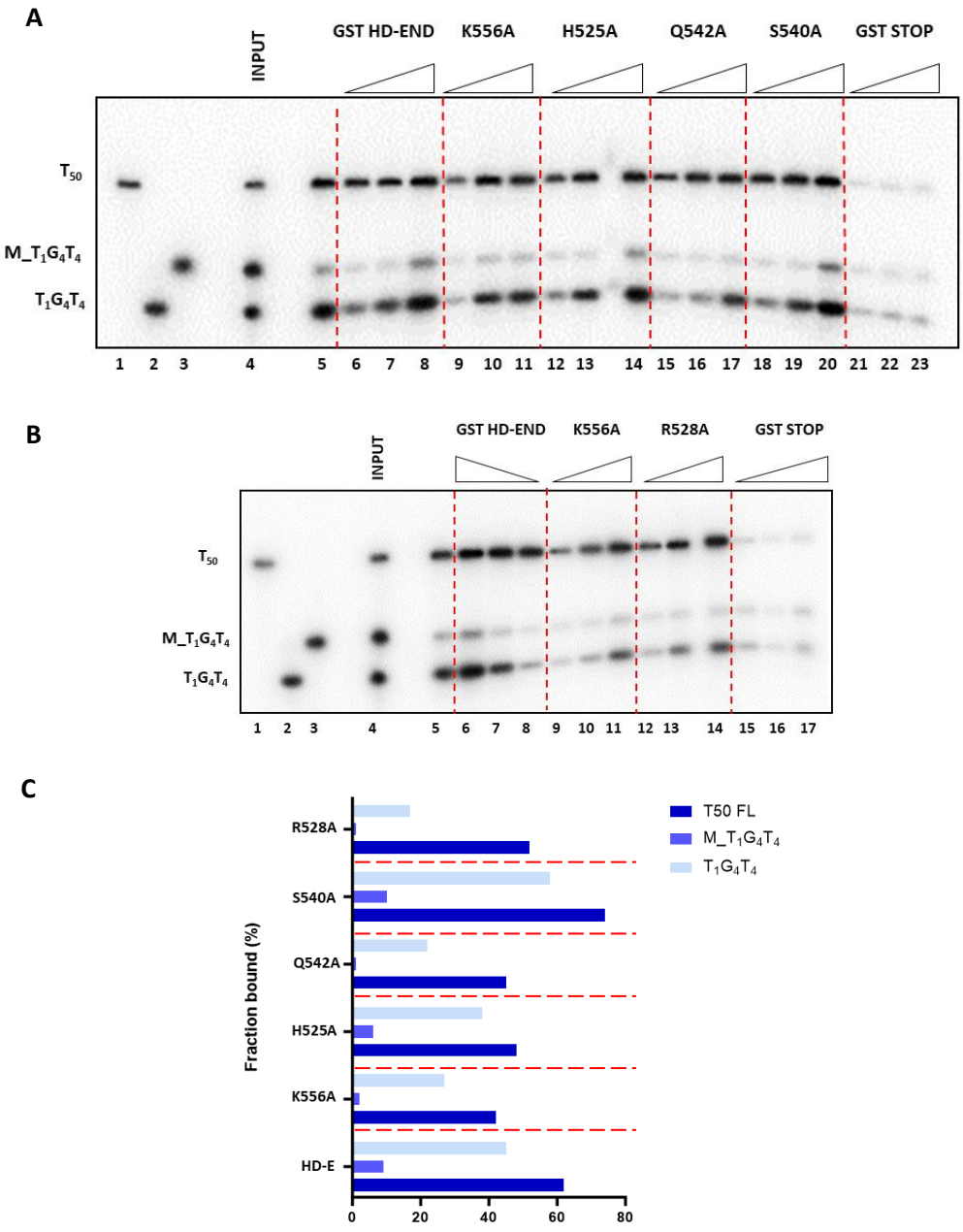


Figure 57. McKay assay with hPIF1 mutants to test a G4 DNA binding model (0.25 nM T50, $T_1G_4T_4$ and M_$T_1G_4T_4$ substrates and 25-100 nM protein). (A) Acrylamide gel of the mutants in the rail β-hairpin: lanes 1-3, markers; lane 4, 25% input; lane 5, 50 nM GST FL hPIF1 positive biding control; lanes 6-8, GST HD-END titration; lanes 9-11, K556A ssDNA binding channel control mutant titration; lanes 12-14, H525A titration; lanes 14-17, Q542A titration; lanes 18-20, S540A titration; lanes 21-23, GST STOP. (B) Acrylamide gel of the mutants in the 2B domain: lanes 1-3 markers; lane 4, 25% input; lane 5, 50 nM GST FL hPIF1; lanes 6-8, GST hPIF1 HD-END titration; lanes 9-11, K556A ssDNA binding channel control; lanes 12-14, R528A titration; lanes 15-17, GST STOP titration. (C) Data plot of the mutants at 100 nM concentration. Each bar represents a mutant while each colour indicates the fraction of substrate recovered.

3.5 Unwinding of unimolecular G4 substrates by hPIF1

Yeast PIF1 is considered to be the best optimized for G4 DNA resolution of all yeast helicases studied [176]. Also, in intact cells, replication forks slow in the presence of G4 DNA stabilizing ligands, and this is exacerbated when PIF1 is knocked down. Previously, hPIF1 helicase activity had been assayed with synthetic tetramolecuar G4 substrates [146]. Here, unimolecular G4 DNA unwinding was explored using a gel-based helicase assay.

Substrate PST55 is a partially single- and double-stranded DNA substrate used previously [145] and in this work (see later, section 3.6) to investigate hPIF1 dsDNA unwinding activity [146]. T55G4_ds20 is a unimolecular G4 substrate consisting of a 55 nucleotides 5' poly(T) tail, a G4 DNA segment G_4T_4 based on the O*xytricha* telomere and a 20 base pair component following. In the control substrate T55G4cont_ds20, the unimolecular G4 DNA is unable to fold because of a base substitution (A for a G) made in each stack, as for the substrates used in the McKay assays (Figure 58). Knowing how hPIF1 acts on PST55, I first compared activity on the G4 substrate using substrate PST55 as an unwinding control, and G4 and control substrates with both strands labelled.

In Figure 57, unwinding was compared with and without ATP/Mg^{2+}, with increasing hPIF1 concentration (3.125 -100 nM) and 0.2 nM substrate. hPIF1 dsDNA unwinding requires ATP/Mg^{2+} [129]. Lane 1 is thermally dissociated substrate T55G4cont_ds20, used as a mobility control to help identify unwound products. As observed in Figure 59 A lanes 18-25, a low level of substrate PST55 unwinding was observed without nucleotide cofactor. However, for substrate T55G4_ds20, there was very little displacement of the 20 base pair component of the substrate, but the appearance of a lower mobility product (see schematic in Figure 59 B) indicated unwinding of G4 DNA by hPIF1 in the absence of ATP/Mg^{2+} (lanes 2-9 and Figure 59 C). In the presence of ATP/Mg^{2+} up to a 2.5-fold increase in G4 DNA unwinding was observed (Figure 59 C). Moreover, the low level of nucleotide independent G4 DNA unwinding seems to stay constant with increasing protein concentration, while the fraction unwound in the presence of ATP/Mg^{2+}increases with the increasing protein concentration.

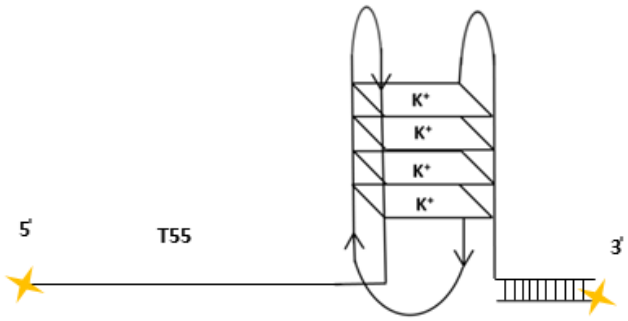


Figure 58. Schematic representation of the T55G4_ds20 substrate. The substrate comprises a 5' tail of 55 T residues and 20 base pairs dsDNA at the 3' end, with an intervening G4 DNA segment (GGGGTTTTGGGGTTTTGGGGTTTTGGGG. The substrate is labelled at the 5' end of the top and bottom strands. Yellow starts at the 5' ends indicates where the substrate has been labelled with γ^{32}P.

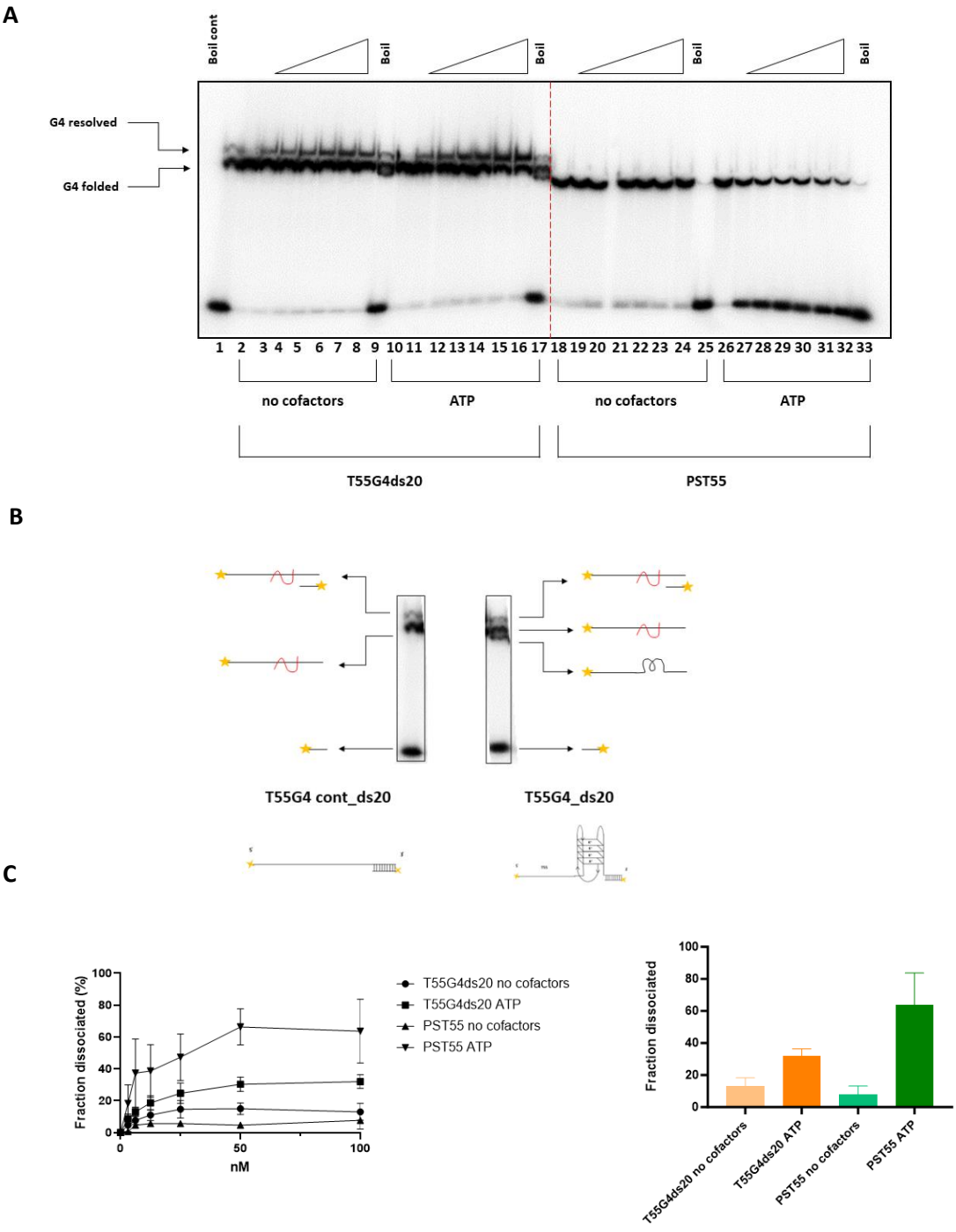


Figure 59. Unwinding of theT55G4_ds20 substrate. (A) Gel showing FL hPIF1 unwinding T55G4_ds20 (lanes 2-18) and PST55 (lanes 18-33) with increasing FL hPIF1 (3.125, 6.25, 12.5 25, 50 and 100 nM, 0.2 nM substrate), without and with of 5 mM ATP/Mg^{2+}. In the gel: lane 1, boil of T55g4contds2, lanes 2-9, increasing protein and no cofactor, T55G4_ds20; lanes 10-17, increasing protein with 5 mM ATP/Mg^{2+} and T55G4_ds20; lanes 18-25, increasing protein and no cofactor, with substrate PST55; lanes 26-33, increasing protein with 5 mM ATP/Mg^{2+} and substrate PST55. (B) Schematic explanation of T55G4_ds20 dissociation and its control T55G4cont_ds20. (C) Data plot of substrates unwinding (right); fraction dissociated at 100 nM protein concentration (left).

3.5.1 G4 unwinding in the presence of nucleotide analogues

Next, the effect of nucleotide analogues mimicking the steps in ATP hydrolysis on hPIF1 G4 DNA unwinding activity was investigated. As previously tested in the binding assay (section 3.3.4), five conditions were compared: apo (no cofactor), AMP·PNP, ADP·MgF_4^-, ADP and ATP (all 1 mM with 1.5 mM Mg^{2+}) along with high ATP/Mg^{2+} concentration (5 mM ATP and 5.5 mM Mg^{2+}), in a protein titration from 6.25 to 100 nM, using substrates T55G4_ds20 and its mutated control T55G4cont_ds20 (Figure 60). Again, a low level of hPIF1 G4 DNA unwinding was observed in the apo condition, but only in the presence of ADP·MgF_4^-, which mimics the transition state of ATP hydrolysis, is a high fraction of the substrate dissociated (complete dissociation compared to 10% dissociation without cofactor and 25% with ATP/Mg^{2+}). This is a surprising observation because it highlights that hPIF1 does not need hydrolysable ATP to display high-efficiency helicase activity on a G4 DNA substrate. Also, a difference between 1 mM and 5 mM ATP is observed with higher concentration of ATP favoring hPIF1 unwinding activity. AMP·PNP (non-hydrolysable ATP) on the other hands, shows a dissociation of ~30% (Figure 58 B) at high cofactors concentration comparable to the dissociation obtained at high ATP concentrations.

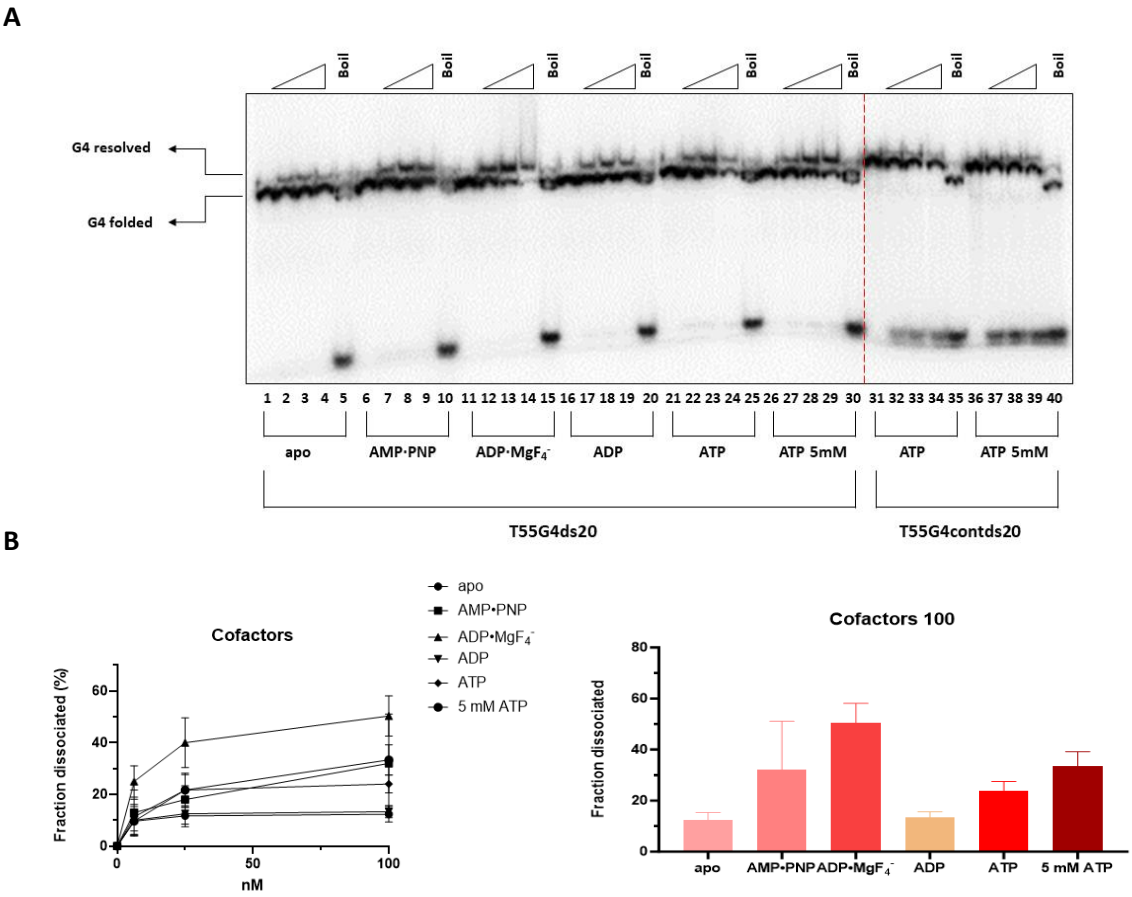


Figure 60. hPIF1 unwinding of G4 DNA in the presence of nucleotide cofactors (0, 6.25, 25, and 100 nM hPIF1 with 0.2 nM T55G4_ds20. (A) Representative acrylamide gel: lanes 1-4, in apo condition; lane 5, boil in apo; lanes 6-9, with AMP·PNP/Mg^{2+}; lane 10, boil; lanes 11-14, ADP·MgF_4^-: lane 15, boil with ADP·MgF_4^-; lanes 16-19, with ADP/Mg^{2+}; lane 20, boil with ADP/Mg^{2+}; lanes 21-24, with 1 mM ATP/Mg^{2+}; lane 25, boil with 1 mM ATP/Mg^{2+}; lanes 26-29, with 5 mM ATP Mg^{2+}; lane 30, boil with 5 mM ATP Mg^{2+}; lanes 31-34, with 1 mM ATP Mg^{2+}; lane 35, boil with 1 mM ATP Mg^{2+}; lanes 36-39, with 5 mM ATP; lane 40, boil with 5 mM ATP Mg^{2+}. (B) (on right) data plot of unwinding as a function of protein concentration with the different cofactors; (on left) bar graph of data at 100 nM concentration.

3.6 Discovery and analysis of hPIF1 inhibitors

3.6.1 Helicase assay of FL hPIF1

The radiometric helicase assay employs a synthetic partially single- and double-stranded substrate, ^{32}P radiolabelled on the 5' end of the short oligonucleotide that generates the dsDNA component of the substrate (Figure 61).

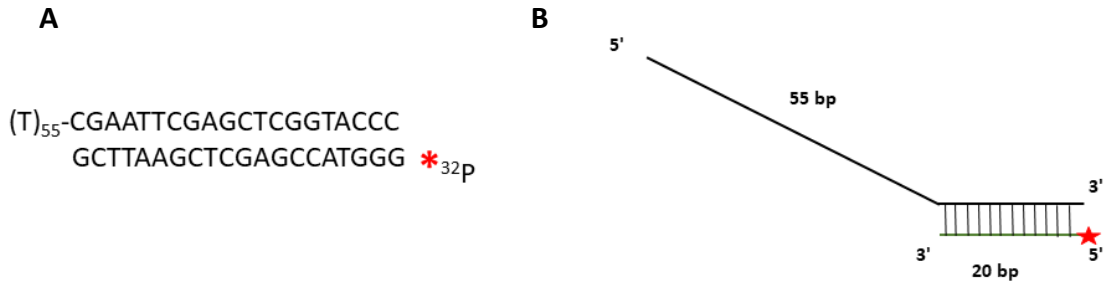


Figure 61. The partially single-and partially double-stranded DNA substrate use for radiometric helicase assay. (A) The sequences of the two strands. (B) Stick diagram of the substrate: the ssDNA 5´ tail is of 55 T residues, followed by a dsDNA portion of 20 base pairs.

While the general helicase assay method was established [146], further modifications to the reaction conditions were investigated to assess the range of options for downstream analysis. To investigate the activity of FL hPIF1 purified by a new route (described in section 3.1.1), the protein was first compared to an archival stock in the lab purified by a previously established protocol ("Stock FL hPIF1" in Figure 62). The assay was performed with 2% DMSO, which would be present in inhibitor screens.

In the electrophoresis assay of helicase reaction products (Figure 62 A), lanes 1 and 9 are a 'boil' control (no hPIF1) showing the mobility of the labelled single-stranded product following thermal denaturation; lanes 2 and 10, are controls showing the native substrate (no hPIF1) while in lanes 3-8 and 11-16 are protein titrations (from 0 nM to 20 nM) to evaluate hPIF1 helicase activity. The protein purified by the new route ("New FL-hPIF1" in Figure 62 A, right panel) appeared to have increased activity at low protein concentrations (<5 nM) compared to the archival stock, but comparable activity at higher protein concentration, demonstrating the successful purification of FL hPIF1 via a new route and buffer conditions (section 2.2.3). Note also, the results of ATP*ase* (Figure 47) assay demonstrate a complete lack of ATPase activity for hPIF1 walker A and B mutants purified by the same route. The screening of the two proteins also indicates that the maximum hPIF1 unwinding activity was obtained at 10 nM, suggesting an appropriate protein concentration to use for inhibitor screening.

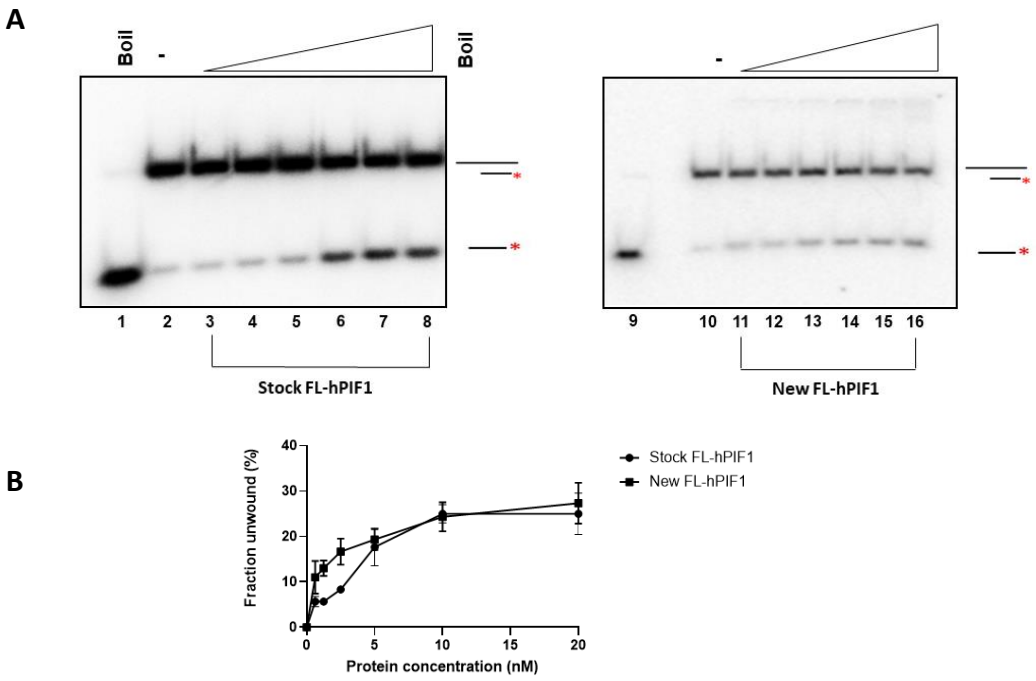


Figure 62. Comparison between the new purified FL hPIF1 and the archival lab FL hPIF1 stock. (A) Electrophoresis analysis of helicase assays comparing the two proteins (0.1 nM substrate): lanes 1 and 9, thermally denatured substrate: 'boil'; lanes 2 and 10, native substrate (no protein) (-); lanes 3-8, stock FL hPIF1 titration; lanes 11-16, new FL hPIF1 titration. (B) Quantification of reaction products (n=3 repeats). Both the proteins show a plateau of unwinding activity (~ 30%) at 10 nM.

An assessment was also made as to whether DTT (2 mM) could be replaced with TCEP (1 mM) as the reducing agent. No significant difference was observed and the original DTT was retained (data not shown). A time course analysis of the reactions was investigated by Dr Sanders. No increase in unwinding was observed after 15 minutes of incubation at 37°C.

Before proceeding with the screening of hPIF1 inhibitors, the tolerance of the protein to DMSO (dimethyl sulfoxide) was investigated (synthesized small molecules were dissolved in DMSO). In Figure 63 A, the DMSO range was from 0%-5%. FL hPIF1 DNA unwinding was sensitive to increasing DMSO (Figure 63 B). From the slope of the graph, it can be estimated that every 1% increase in DMSO results in ~3% decrease in unwinding activity. In future reactions, the percentage of DMSO did not exceed 2%.

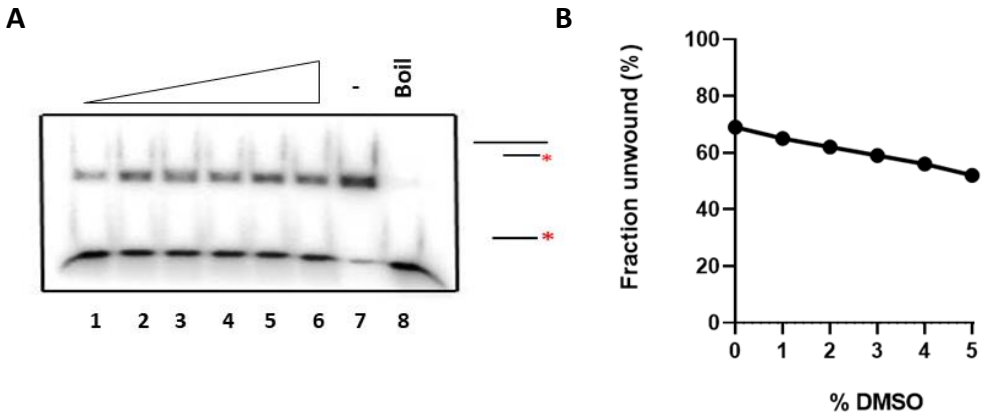


Figure 63. Effect of DMSO on hPIF1 helicase activity. (A) Electrophoresis of helicase reaction products lanes 1-6, 0%-5% DMSO in 1 % increments with 10 nM FL hPIF1 and 0.1 nM substrate, lane 7 native substrate without protein (-), lane 8 boil. (B) data plot showing a decrease in unwinding of ~ 15% when the percentage of DMSO is increased from 0% to 5%.

3.6.2 Analysis of EN300-163911 and its derivatives

Because of chemical novelty, potential for chemical modifications and its synthetic feasibility, EN300-163911 originally identified in DSF assay was selected for further development (Edelris). First, to evaluate the inhibitory activity on FL hPIF1 helicase action, an unwinding assay was performed with EN300-163911 from 4 mM - 1.95 µM in a doubling dilution series (2% DMSO), with a boil, a no-inhibitor (protein + 2% DMSO) and a negative (no protein + 2% DMSO) control. Figure 64 shows inhibitory activity and dose-response curve for EN300-163911. However, inhibition was only observed at relatively high concentrations of EN300-163911 > 2 mM and no IC_{50} value could be determined, although not unexpected for a fragment-like molecule.

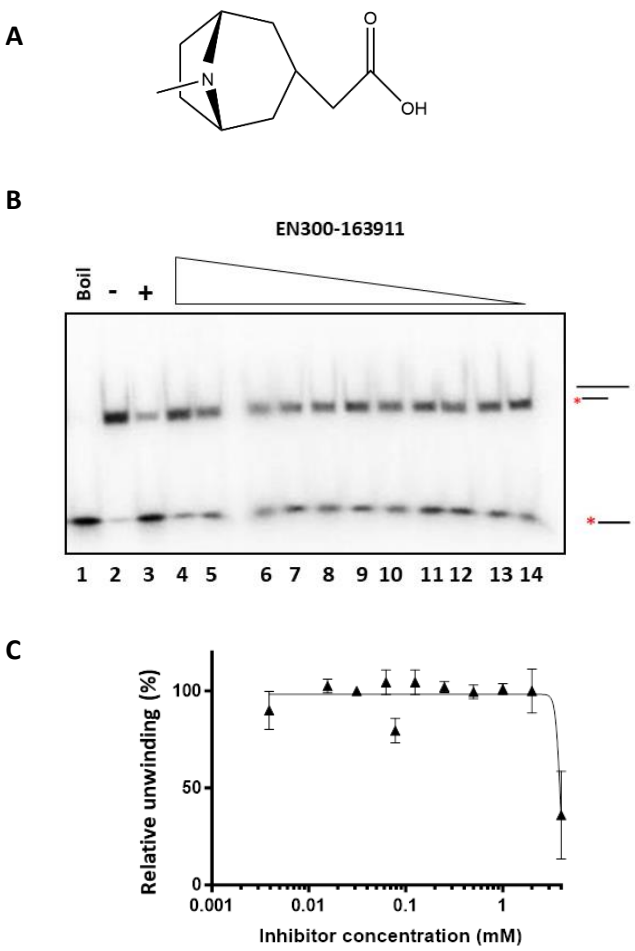


Figure 64. Analysis of EN300-163911 (183.25 Da) in the helicase assay (10 nM hPIF1, 0.1 nM substrate). (A) Chemical structure of EN300-163911. (B) Electrophoretic analysis of reaction products from helicase assays: lane 1, boil; lane 2, native substrate + 2% DMSO (-); lane 3, FL-hPIF1 + 2% DMSO (+); lanes 4-14, EN300-163911 doubling dilution series from 4 mM to 1.95 µM. (C) Data plot of EN300-0163911 activity (n=3 repeats); in the graph the % of fraction unwound relative to the no compound control is plotted against inhibitor concentration (mM). EN300-116391 shows inhibitory activity at high concentrations only (IC_{50} not determined).

A first set of EN300-163911 derivatives, "F-series compounds" (F1-F5), were synthesised (Edelris), dissolved in DMSO to obtain as stock concentration of 100 mM, and tested for inhibition at 1 mM (Figure 65). F1 appeared to show near completely inhibition of DNA unwinding, while F2 and F4 showed inhibition > 60%. However, in IC_{50} assays (doubling dilution series from 1 mM, Figure 65 B), fragments F2 and F4 showed similar behavior and poor inhibition while F1, previously determined to be the most active analogue in the single-point assay, shows a clear concentration-dependent inhibition of unwinding. An IC_{50} of 0.485 ± 0.04 mM was determined for F1 (Figure 65 C).

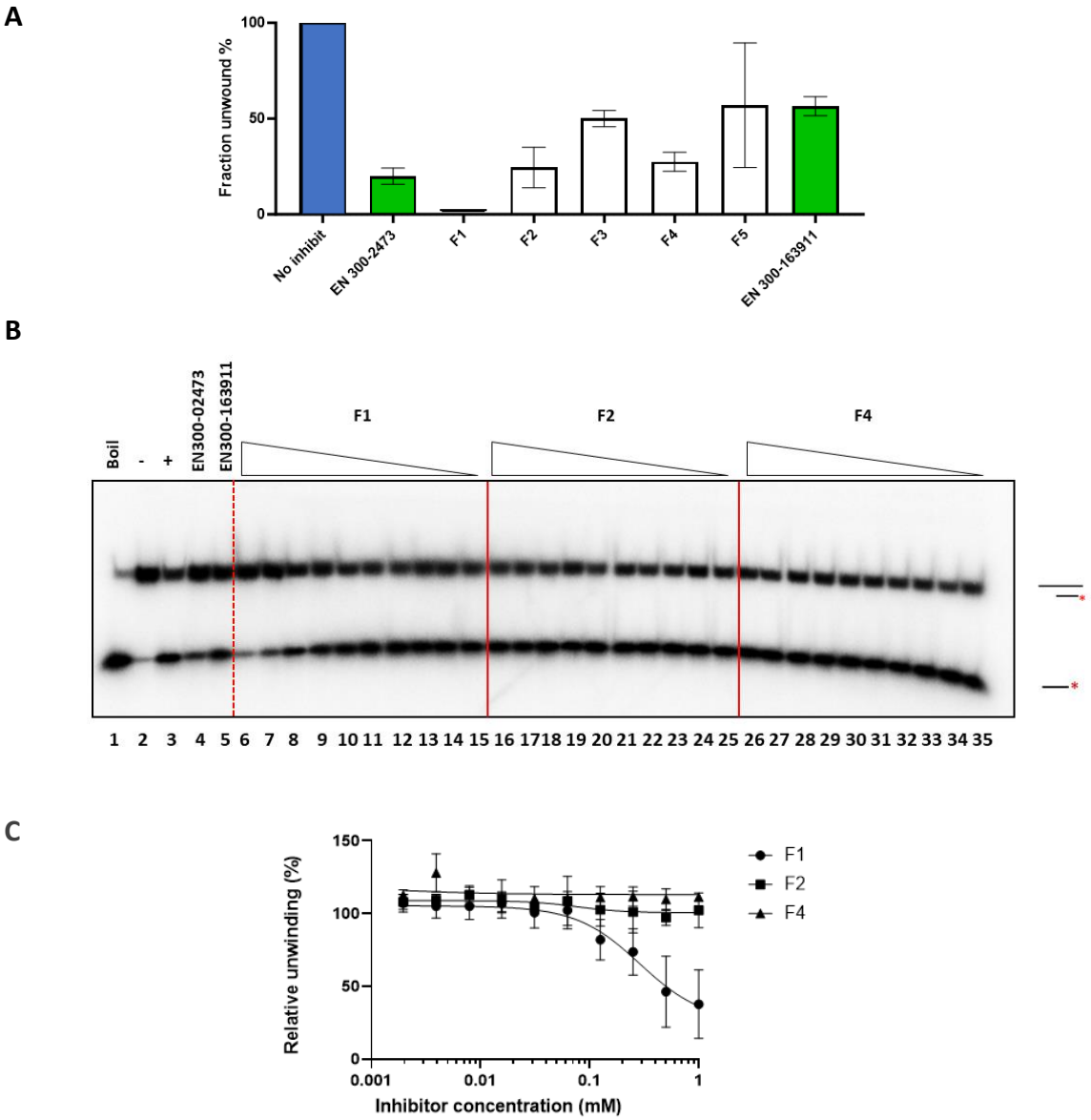


Figure 65. Screening of EN300-116391 analogues in the hPIF1 helicase assay (10 nM hPIF1, 0.1 nM substrate). (A) Data plot from single point (1 mM) screening analysis (n=3 repeats) relative to DMSO control (no inhibitor + hPIF1). F1 appears to completely inhibit FL hPIF1 dependent DNA unwinding, and F2 and F4 inhibition is > 60%. (B) Electrophoresis analysis of helicase reaction products for F1, F2, F4 titrations (doubling dilutions from 1 mM to 1.95 µM): lane 1, boil; lane 2, native substrate (-); lane 3, (+) + FL hPIF1; lane 4, EN300-02473 1 mM; lane 5, EN300-163911 (parent compound) 1 mM; lanes 6-15, F1 double dilution series; lanes 16-25, F2 doubling dilutions; lanes 26-35, F4 doubling dilutions. (C) data plot of F1, F2 and F5 IC_{50}s (n=3 repeats). A titration curve is obtained for F1 only with an IC_{50} value of 0.485 ± 0.04 mM determined.

Based on the analysis of the F-series compounds another set of 11 fused-heterocyclic derivatives (A1-10, B2) was synthesised and tested at a concentration of 2 mM (single point) for inhibition of helicase activity (Figure 66 A). A1, A3 and A4 show an apparent complete inhibition, while A7 and B2 appeared to have a similar inhibition of ~80%. A7 was excluded from further analysis after a poor activity in a 4-point titration (data not shown) as well as B2 which showed high variability in results, probably due to its partial solubility in DMSO.

IC_{50} values for inhibition of helicase activity were determined for A1, A3 and A4 (from 2 mM in a doubling dilution series, Figure 67) and the values obtained were, respectively 0.131 ± 0.007 mM, 0.084 ± 0.018 mM and 0.321 ± 0.032 mM.

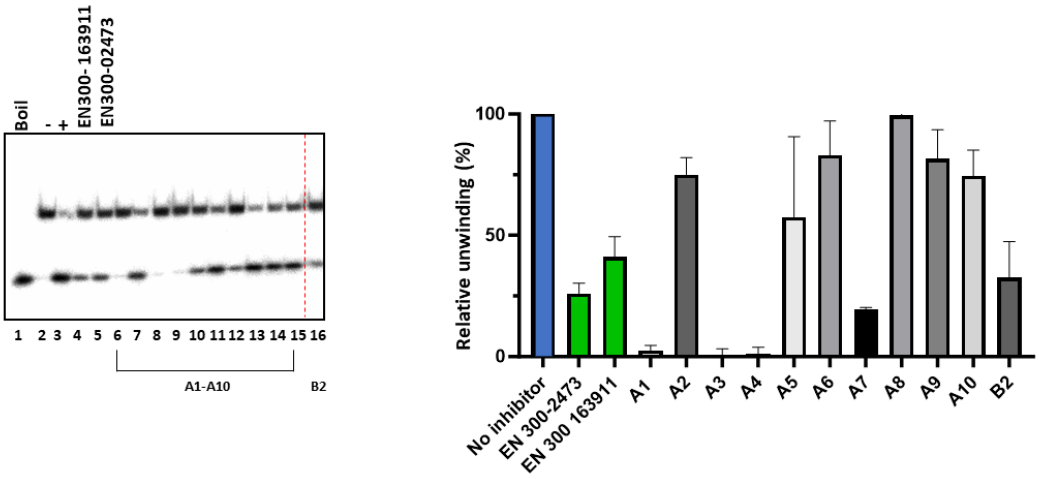


Figure 66. Screening of EN300-16391 derivatives at 2 mM by helicase assay (10 nM hPIF1, 0.1 nM substrate). (A) Electrophoresis analysis of reaction products. Lane 1, boil; lane 2, native substrate (-); lane 3, + hPIF1, no inhibitor (+); lane 4, 1 mM EN300-02473; lane 5, 1 mM EN300-163911; lanes 6-16, screening of compounds as indicated. (B) Data plot of the screening (n=3 repeats). A1, A3 and A4 show an almost complete inhibition of helicase activity; A7 and B2 ~ 80%, inhibition.

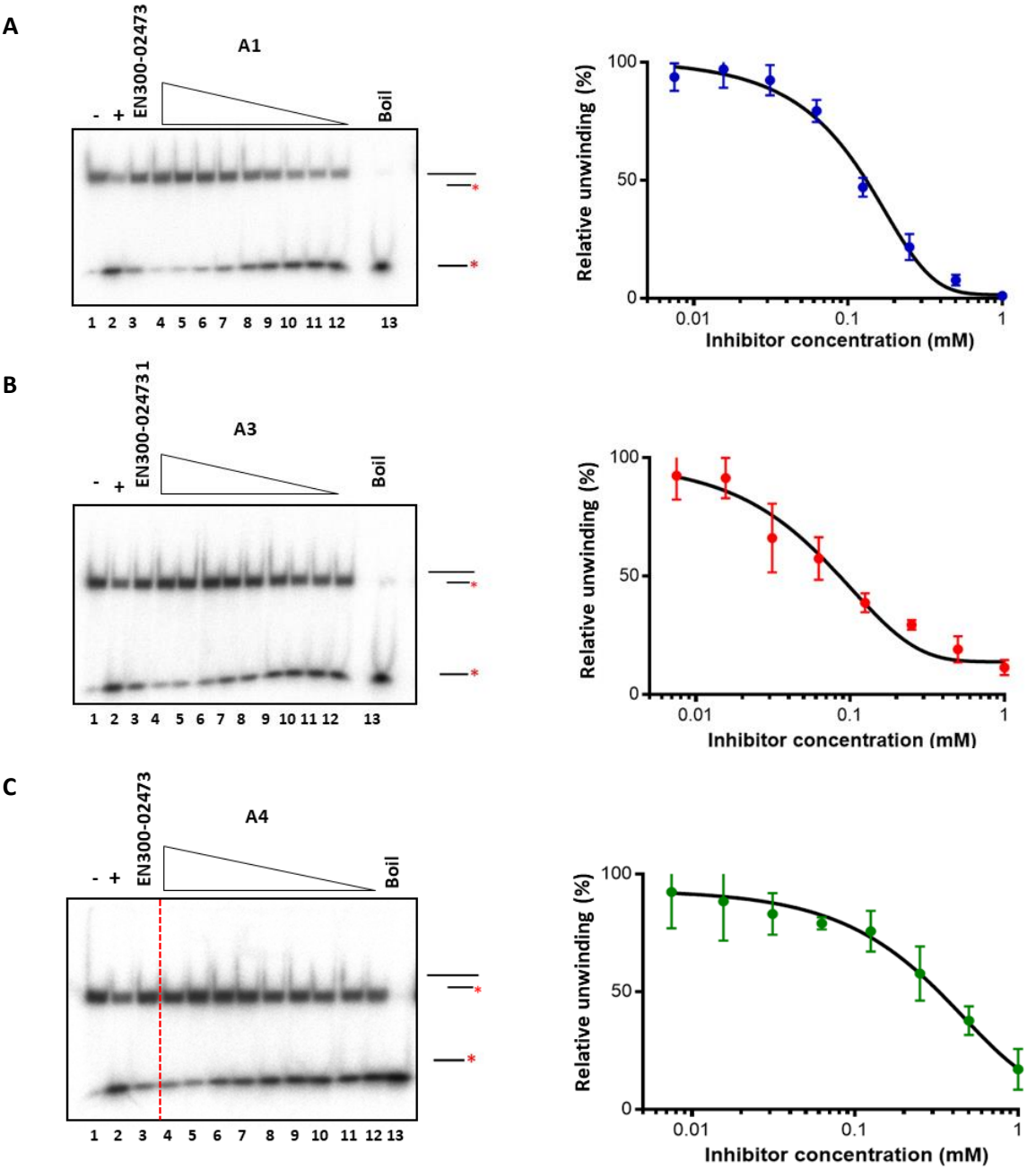


Figure 67. IC_{50} analysis of compounds A1, A3 and A4 (n=3 repeats; 10 nM hPIF1, 0.1 nM substrate). The data obtained from the electrophoresis of helicase reaction products (left) are plotted in the graph on the right. (A) A1, IC_{50} 0.131 ± 0.007 mM. (B) A3, IC_{50} 0.084 ± 0.018 mM; (C) A4, IC_{50} 0.131 ± 0.032 mM.

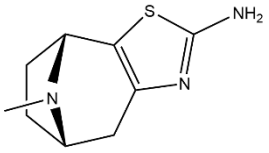
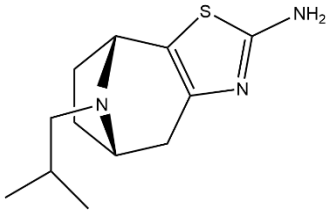
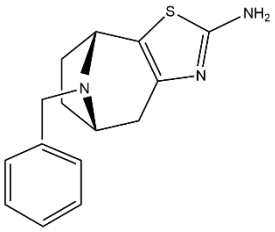
Name	Structure	IC_{50}
A1		**0.131 ± 0.007 mM**
A3		**0.084 ± 0.018 mM**
A4		**0.321 ± 0.032 mM**

Table 8. Structures and IC_{50} value of EN300-163911 derivatives A1, A3 and A4.

3.6.2.1 A3 derivative series

Compound A3 (reshipped by Edelris) was considered for the development of new derivatives, to introduce diversity in the compound scaffold while limiting the synthetic steps required. The alkyl-N-substituted derivatives were first screened at 1 mM (Figure 68 A), identifying AF1459 (~30% of inhibition), AF1463 (~60% of inhibition), AF1471 (~ 25% of inhibition) and AF1477 (~40% inhibition) as molecules of interest. All had higher inhibitory activity than their parent compound A3 (20% of inhibition). A 4-point titration of the four molecules (Figure 68 B revealed that AF1459 and AF1463 had the highest level of dose-responsive inhibition and IC_{50} values of ~0.800 ± 0.158 mM for AF1459 and ~0.271 ± 0.021 mM for AF1463 were then determined (Figure 69 A and B). Since chemical modification of this potential inhibitor through the alkyl substitutions did not bring any improvement, further investigation was not undertaken.

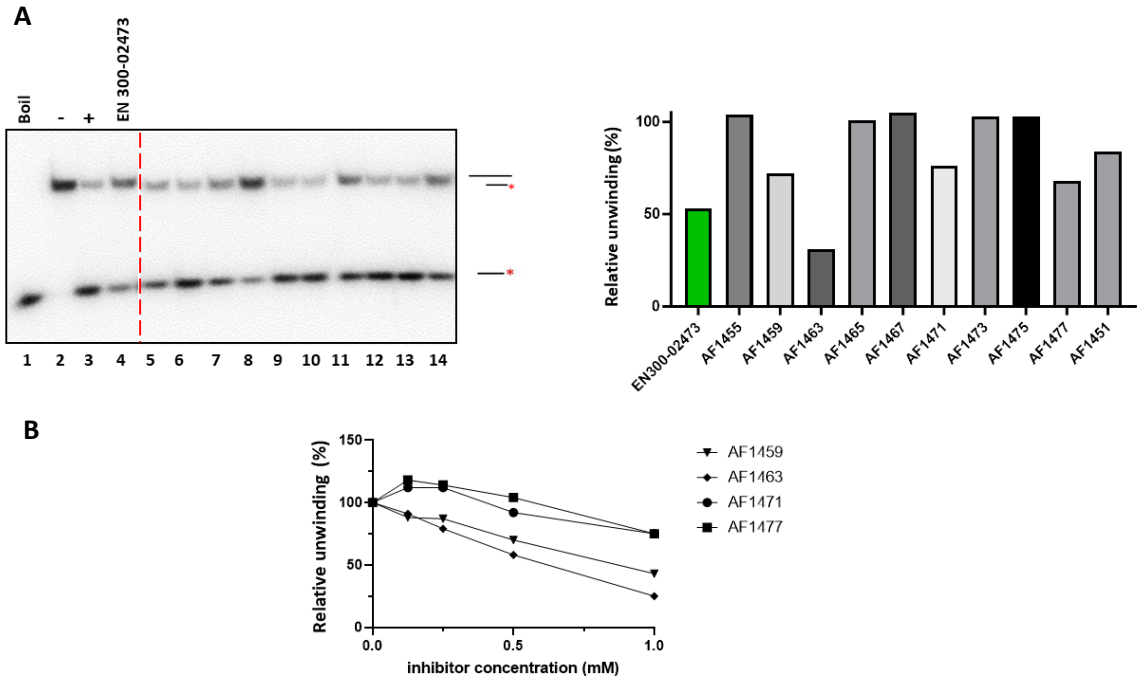


Figure 68. Screening of the A3 derivatives at 1 mM (10 nM hPIF1, 0.1 nM substrate). (A) The electrophoresis analysis of helicase assays with inhibitors: lane 1, boil; lane 2, native substrate (-); lane 3, + hPIF1 (+); lane 4, 1 mM EN300-02473 as positive control of inhibition; lanes 5-14, single point screen of the inhibitors: 5, AF1455; 6, AF1459; 7, AF1463; 8, AF1465; 9, AF1467; 10, AF1471; 11, AF1473; 12, AF1475; 13, AF1477, 14, AF1451. On the right is the data plot obtained from the screening (n=1 repeat) (B) Data plot of the 4-point screen of selected small molecules inhibitors (n=1 repeat).

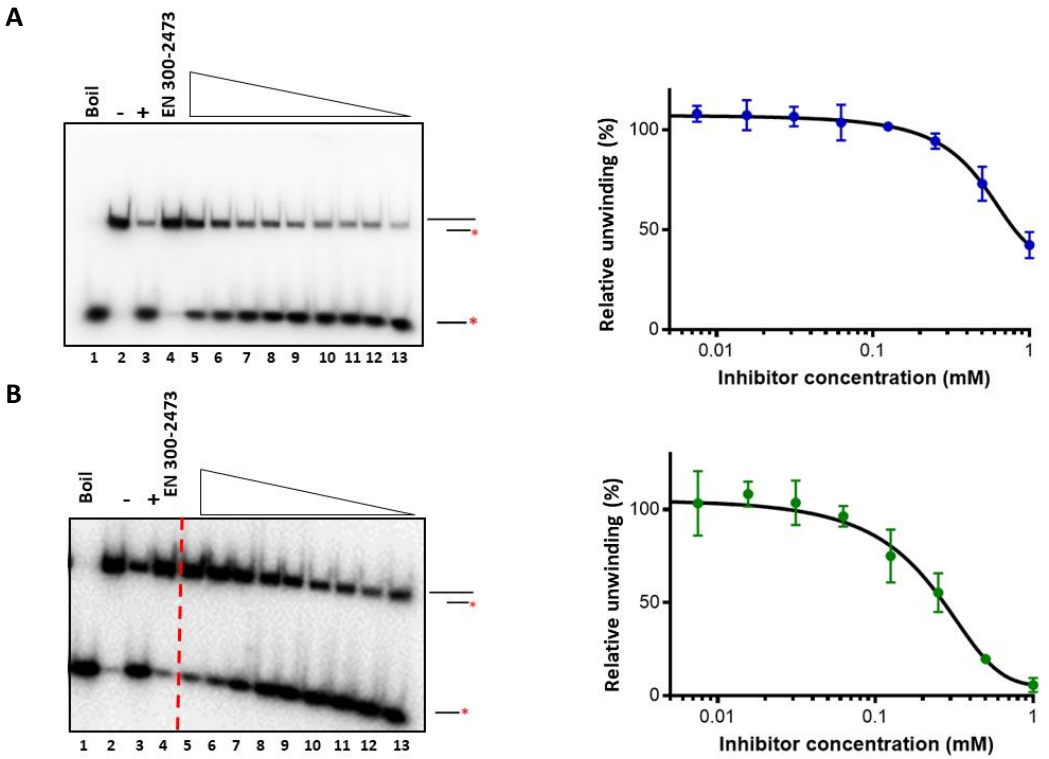


Figure 69. IC_{50} determination for compounds AF1459 and AF1463 (n=3 repeats; 10 nM hPIF1, 0.1 nM substrate). The data obtained from the electrophoresis analysis of reaction products (left) are plotted in the graph on the right. (A) IC_{50} AF1459: The IC_{50} value obtained is 0.819 ± 0.158 mM. (B) IC_{50} AF1463: The IC_{50} value obtained is 0.265 ± 0.021 mM.

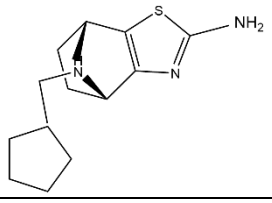
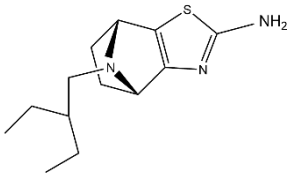
Name	Structure	IC_{50}
AF1495		0.800 ± 0.158 mM
AF1463		0.271 ± 0.021 mM

Table 9. Structure and IC_{50} of EN300-163911 derivatives AF1459 and AF1463.

3.6.3 Potential inhibitors identified by virtual screening of the Edelris chemical collection

A second approach adopted in the discovery of hPIF1 inhibitors was virtual screening. From the docking of Edelris chemical library (Dr. M. Wever) more than 300 entities were identified as potential binders, but the number was reduced when some cut-offs such as the average binding free energy and the virtual ligand efficiency were refined. 96 compounds were screened at 1.5 mM through the standard radiometric helicase assay. Seven compounds showed inhibition similar in magnitude to EN300-02476 (1.5 mM): A5, 53% inhibition; A6, 40% inhibition; B6, 88% inhibition; C4, 36% inhibition; C6, 65% inhibition; D5, 25% inhibition and H6 with 73% inhibition (see Figure 70). Also, ~15 additional compounds demonstrated inhibitory activity, although somewhat reduced compared to the control. The seven principal inhibitors selected share a similar chemical structure. Accordingly, compounds B6 and C6 that showed the highest levels of inhibition and have a similar structure were investigated further.

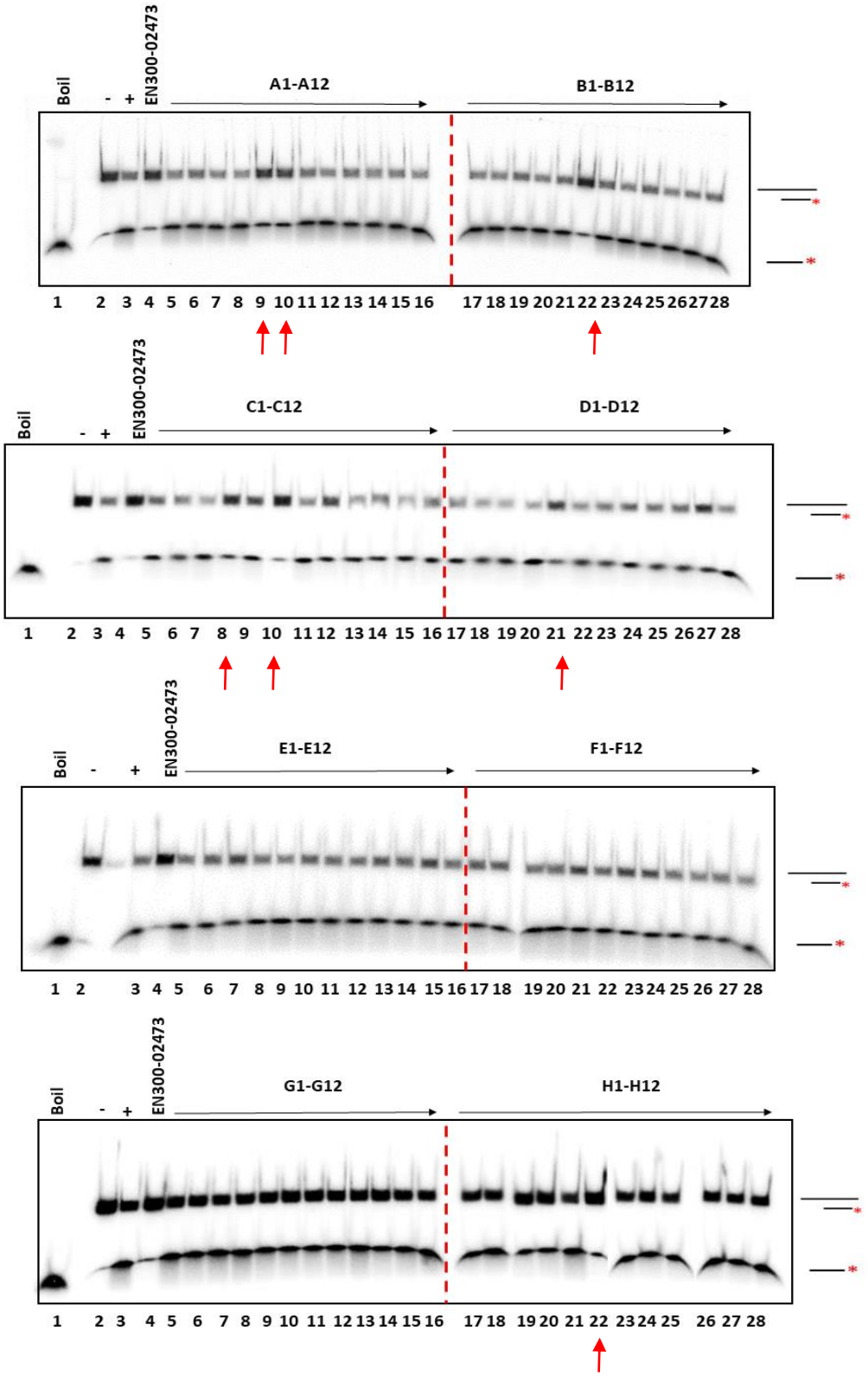


Figure 70. Biochemical screening at 1.5 mM of 96 compounds from the Edelris virtual screen (10 nM hPIF1, 0.1 nM substrate). The electrophoresis analysis of helicase reaction products is shown. In each panel: lane 1, boil; lane 2, native substrate (-); lane 3, hPIF1, no inhibitor (+); lane 4, EN300-02473 as a positive control for inhibition; lanes 5-28, single point screening of the hits. The newly synthesised small molecules of interest are A5, A6, B6, C4, C6, D5 and H6 (red arrow).

A helicase assay was performed to obtain an IC_{50} value for B6 and C6 (doubling dilution series from 1 mM, Figure 71). Although a dose response was obtained for each compound, IC_{50} values could not be determined from the data, while partial solubility was a compounding issue (the molecules are high molecular weight), even after processing in a sonicating water bath for 30 minutes.

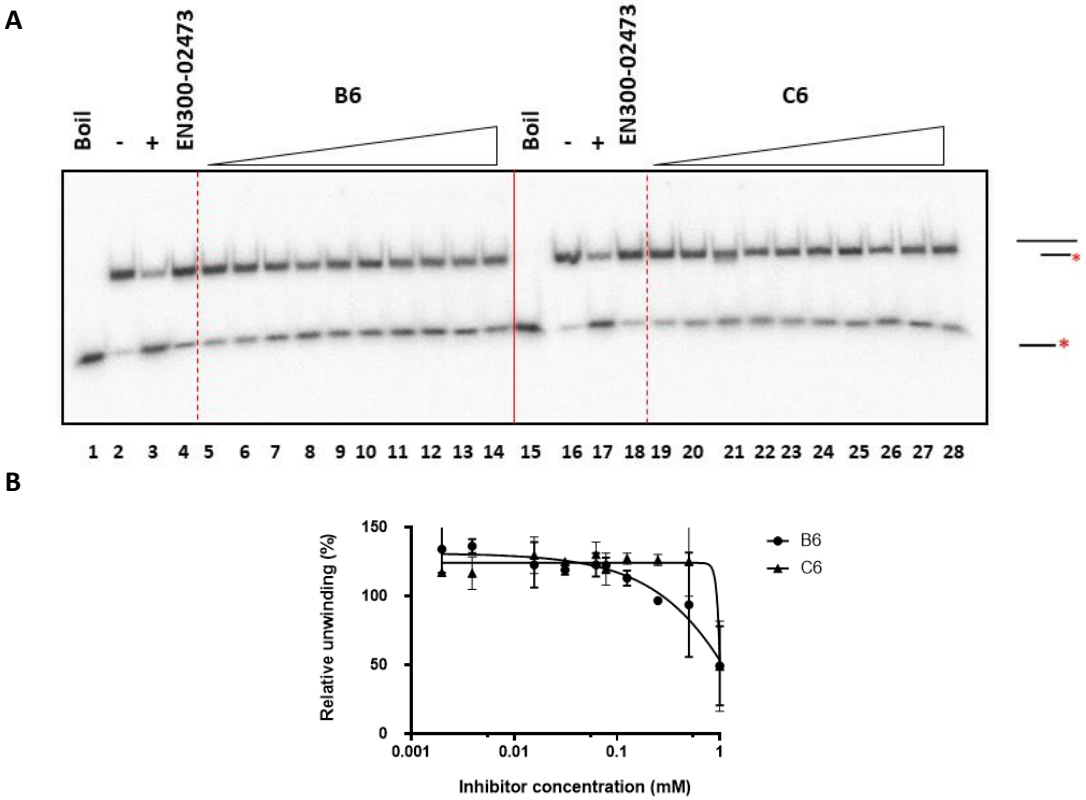


Figure 71. IC_{50} determination for compounds B6 and C6 (n=3 repeats; 10 nM hPIF1, 0.1 nM substrate). (A) Electrophoretic analysis of the reaction products of unwinding assays: lanes 1 and 15, boil; lanes 2 and 16, native substrate (-); lanes 3 and 17, hPF1 no inhibitor (+); lanes 4 and 18, 1 mM EN300-02473 inhibitor control; lanes 5-14 and 19-28, B6 and C6 doubling dilution series from 1 mM. (B) Data plot of the assay data; for both compounds a suitable curve for IC_{50} determination was not obtained.

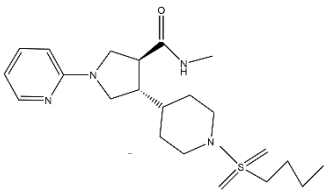
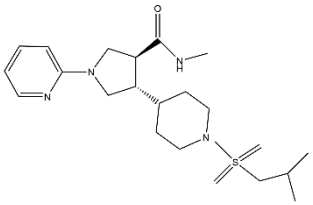
	Structure		
B6		**C6**	

Table 10. Structures of B6 and C6 small molecules.

Based on the activity observed for B6 and C6, derivatives were designed with improved features such as solubility. A first cluster of derivatives included piperidine substitutions placed in the aminothiazole pocket giving a stronger binding mode, were produced. After evaluating inhibitory activity at 1 mM (Figure 72), several hits were identified: AF1485 showed an inhibitory activity of ~ 80%, AF1486 showed ~ 70% while AF1487 shows the highest inhibitory activity of > 90%. Before proceeding with IC_{50} evaluation, these molecules were tested in a 4-point titration, where AF1485 demonstrated high variability so its analysis was stopped. IC_{50} values were determined for AF1486 and AF1487 (10-point doubling dilution series from 1 mM, Figure 73); the values obtained were 524 ± 0.209 µM for AF1486 and 364 ± 0.65 µM for AF1487, indicating an improvement on parent compounds B6 and C6. Although the work on B6 and C6 derivatives stopped at this point, the synthesis gave novel molecules with moderate activity as potential starting point for the synthesis of more analogues.

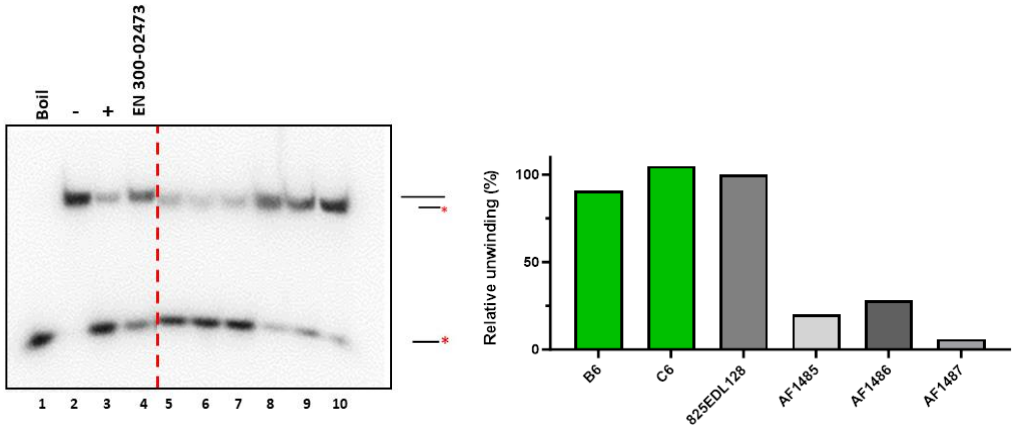


Figure 72. Screening of B6-C6 derivatives at 1 mM (n= 1; 10 nM hPIF1, 0.1 nM substrate). In the electrophoresis analysis of reaction products (left): lane 1, boil; lane 2, native substrate (-); lane 3, hPIF1 no inhibitor (+); lane 4, 1 mM EN300-02473 as positive control of inhibition; lanes 5-10, screening of the potential inhibitors. On the right is the data plot.

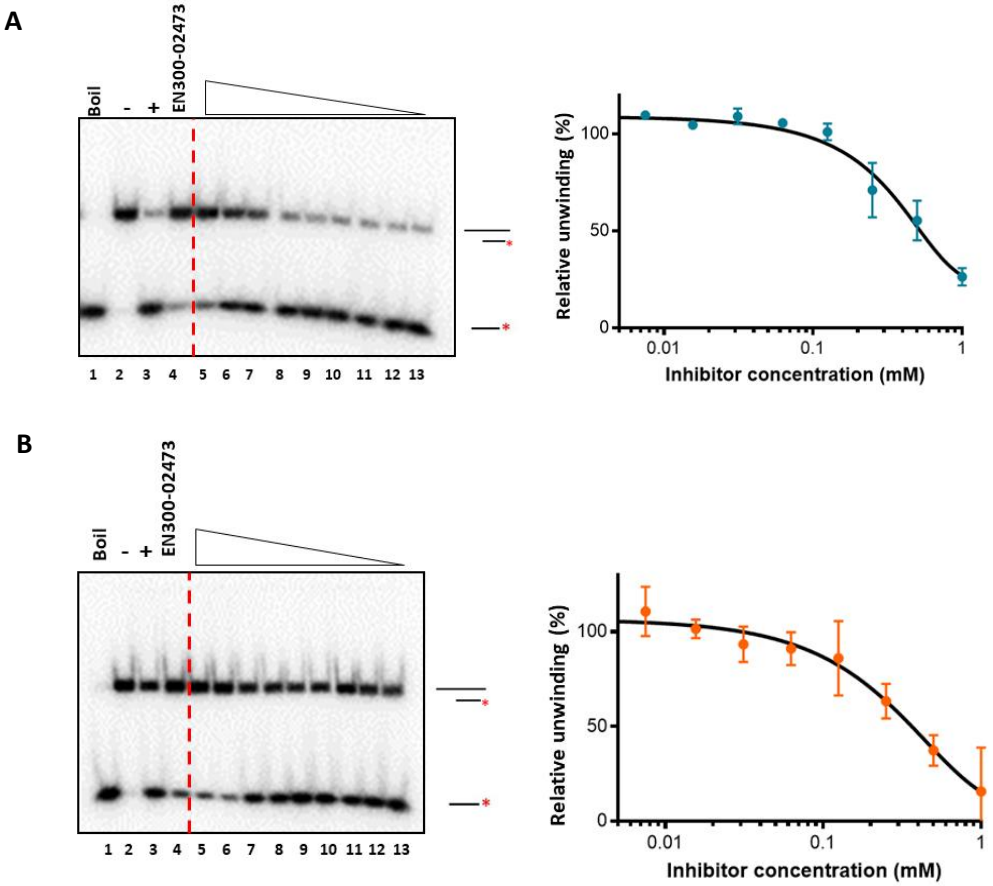


Figure 73. IC_{50} analysis of AF1486 and AF1487 (n=3 repeats; 10 nM hPIF1, 0.1 nM substrate). The data obtained with electrophoresis of reaction products (left) and data plotted in the graphs on the right. (A) AF1486, 0.498 ± 0.209 mM IC_{50} value obtained. (B) AF1487, 0.320 ± 0.065 mM IC_{50} value obtained.

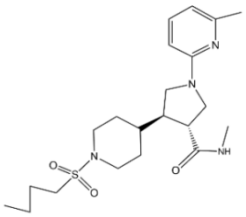
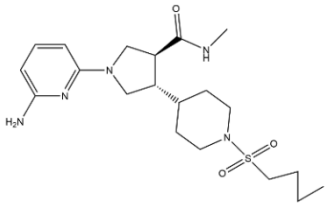
Name	Structure	IC_{50}
AF1486		0.498 ± 0.209 mM
AF1487		0.320 ± 0.065 mM

Table 11. Structures and IC_{50} values of B6 and C6 derivatives: 1. AF1486; 2. AF1487

3.6.4 Development and screening of EN300-02473 derivatives

3.6.4.1 Re-evaluation of inhibitor EN300-02473

EN300-02473 (MW 233.3 Da) was identified from a crystallographic fragment screen (XChem, Diamond Light Source, Oxford) and served as a scaffold for new derivatives. EN300-02473 activity was previously verified in a helicase assay, with an IC_{50} of 0.11 ± 0.02 mM determined for the compound sourced from Enamine. Moreover, at 233.3 Da it is fragment-like and suitable for chemical modification to improve its inhibitory activity [177].

Initially, EN300-02473 was resynthesize (Dr Wever, Edelris) and provided as AF1499. However, when tested in the helicase assay (Figure 74) alongside EN300-02473 purchased from Enamine, different IC_{50} values were determined: EN300-02473 revealed an IC_{50} of 0.061 ± 0.018 mM while for AF1499 it was not possible to obtain an IC_{50} value, although inhibitory activity was observed.

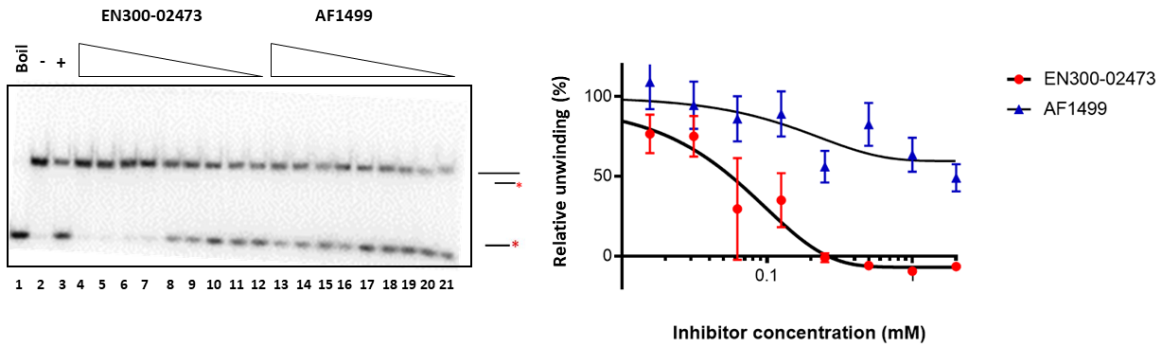


Figure 74. IC_{50} determination for EN300-02473 and AF1499 (10 nM hPIF1, 0.1 nM substrate). The electrophoresis data from helicase assays (left) and data plot (right). The IC_{50} value obtained for EN300-02473 was 0.061 ±0.018 mM (n=3 repeats); no IC_{50} was obtained for AF1499 (n=3 repeats).

New synthetic approaches were adopted at Edelris (see Table 12), providing a total of five preparations of EN300-02473, including EN300-02473 repurchased from Enamine. At Edelris, structures were confirmed to be the same by NMR (data not shown). As observed in Figure 75, none of the newly synthesized compounds reproduced the activity of EN300-02473 except for the new batch purchased from Enamine (AF1759). However, the data obtained would also indicate higher IC_{50} value for EN300-02473 (Enamine). In addition, there has also been a high degree of variability observed in measured IC_{50} values (~0.06-0.3 mM), even within experimentalist in the laboratory.

A comparison between EN300-02473 and the re-synthesised compound (Edelris) shows a difference in colour (near colourless solution for the new synthesis compared to dark brown for the Enamine compound) and the presence of an unidentified impurity was shown by NMR (See Figure A3 in Appendix). An IC_{50} value of 0.855 ± 0.135 mM was obtained for AF1757 (Figure 76, significantly lower than the one obtained from Enamine EN300-02473.

Table 12. Summary of EN300-02473 analogues along with the synthetic route used to obtain them. Descriptions and more information about the synthetic routes here used are in Dr Waver PhD Thesis, Universite Grenoble Alpes, 2023.

Compound	Synthetic route
AF1756	Thiouea route
AF1757	N-Boc-thiourea route
AF1758	Suzuki route
AF1759	Purchased by Edelris from ENAMINE
AF1770	Purchased by Edelris from ENAMINE and purified by preparative HPLC

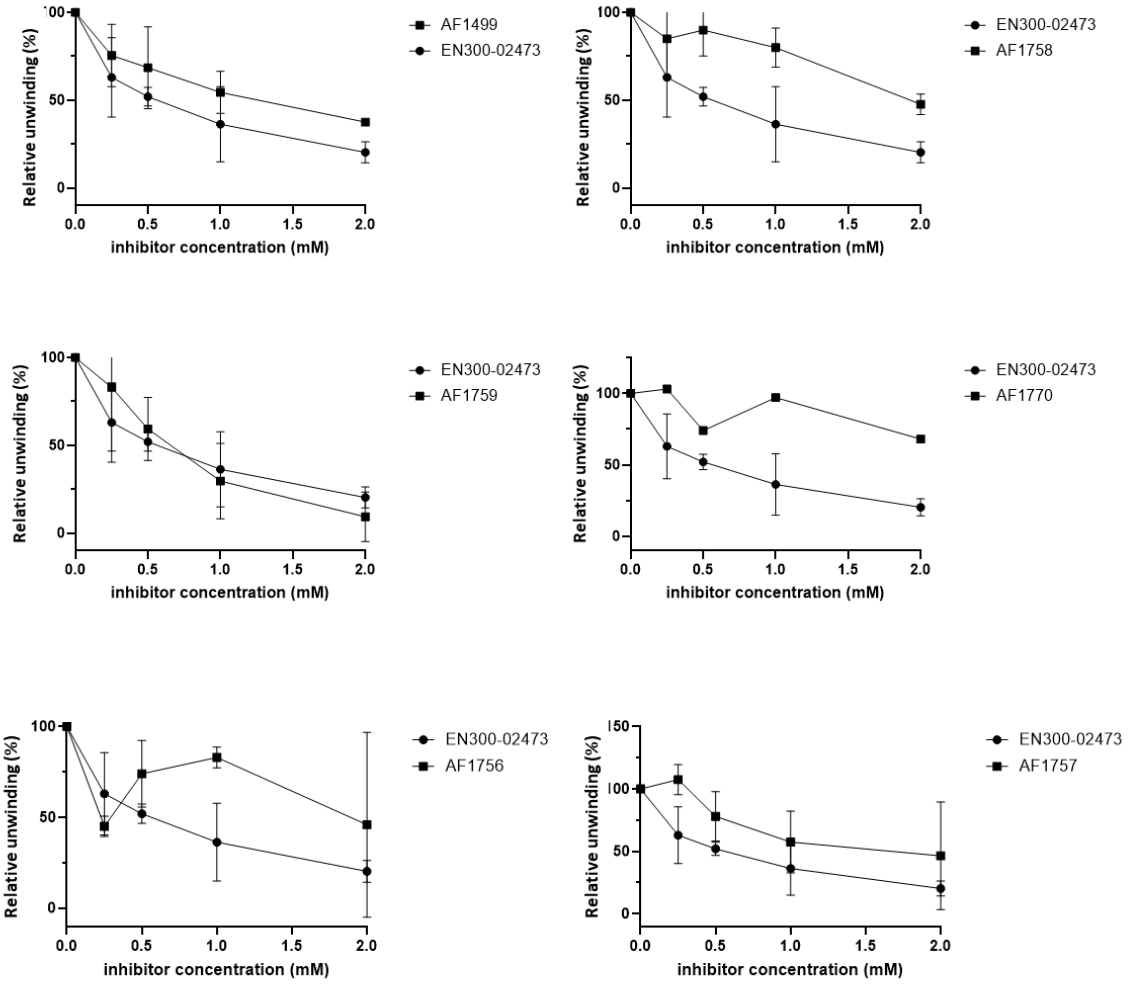


Figure 75. Analysis of EN300-02473 (Enamine) and re-synthesised batches (10 nM hPIF1, 0.1 nM substrate). The graphs report a 4-point screening of the analogous compounds. Only AF1759 re-purchased from Enamine showed activity like the original batch obtained (n=2 repeats, except AF1770, n = 1).

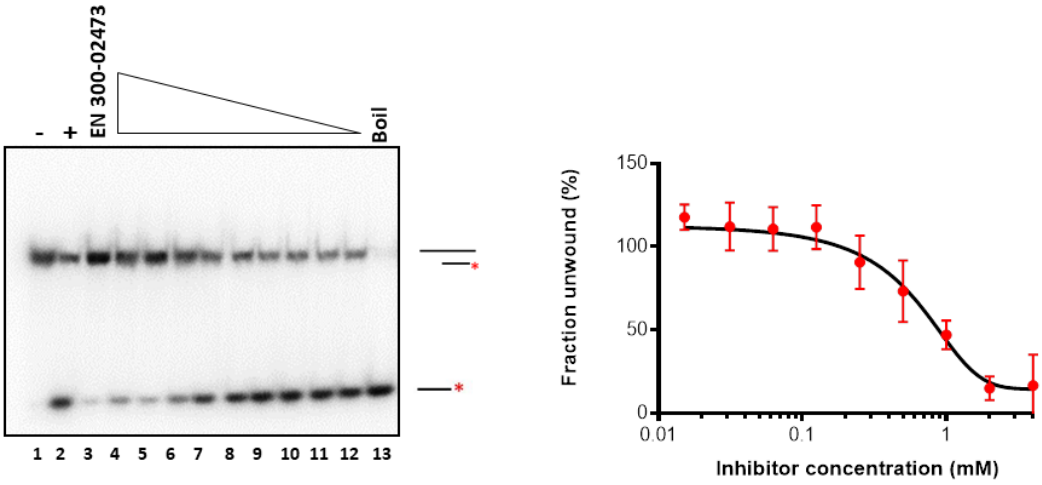


Figure 76. IC_{50} determination for AF1757 (10 nM hPIF1, 0.1 nM substrate). Electrophoresis analysis of helicase reaction products (left): lane 1, native substrate (-); lane 2, hPIF1 no inhibitor (+); lane 3, 1 mM EN300-02473 (Enamine); lanes 5-12, AF1757 doubling dilutions from 4 mM; lane 15, boil. From the data plot (right) the IC_{50} value obtained was 0.885 ± 0.135 mM (n=3 repeats).

3.6.4.2 Analysis of EN300-02473 derivatives

In an attempt to improve the potency of inhibition, the EN300-02473 scaffold (Figure 77) was modified chemically (Edelris) and compounds first screened at 1 mM in the helicase assay (Figure 77). Two series of potential inhibitors were provided: in 'series 1' (shipment BL21113, Figure 77) the scaffold was modified at the R^1 position in the 2-aminothiazole part. The most promising derivatives identified were AF1080, AF1082 and AF1093 (Figure 78, top) with an inhibition of ~ 90%, very similar to the parent compound EN300-02473 (1 mM, from Enamine). AF1085 and AF1087 showed an apparent complete inhibition. In the 'series 2' (shipment BL22113, Figure 78) modifications to the scaffold were made at the R^2 group of the benzene ring, (Figure 78, bottom) AF1753, AF1755 and AF1768 appear to have ~ 75% of inhibition, AF1762 >80% inhibition while AF17563 and AF1766 show an apparent complete inhibition. All molecules showing > 75% inhibition were screened in a 4-point titration (data not shown), identifying AF1763, AF1766 and AF1768 as suitable candidates for IC_{50} determination. IC_{50} values were determined (Figure 79) to be, respectively, 0.119 ± 0.046 mM, 0.101 ± 0.021 mM, and 0.163 ± 0.045 mM, as summarised in Table 13.

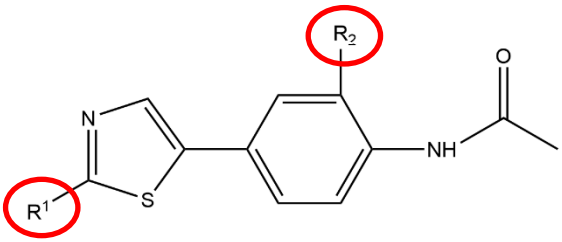


Figure 77. Structure of EN300-02473. The groups highlighted represents the exit routes for the synthesis of new derivatives.

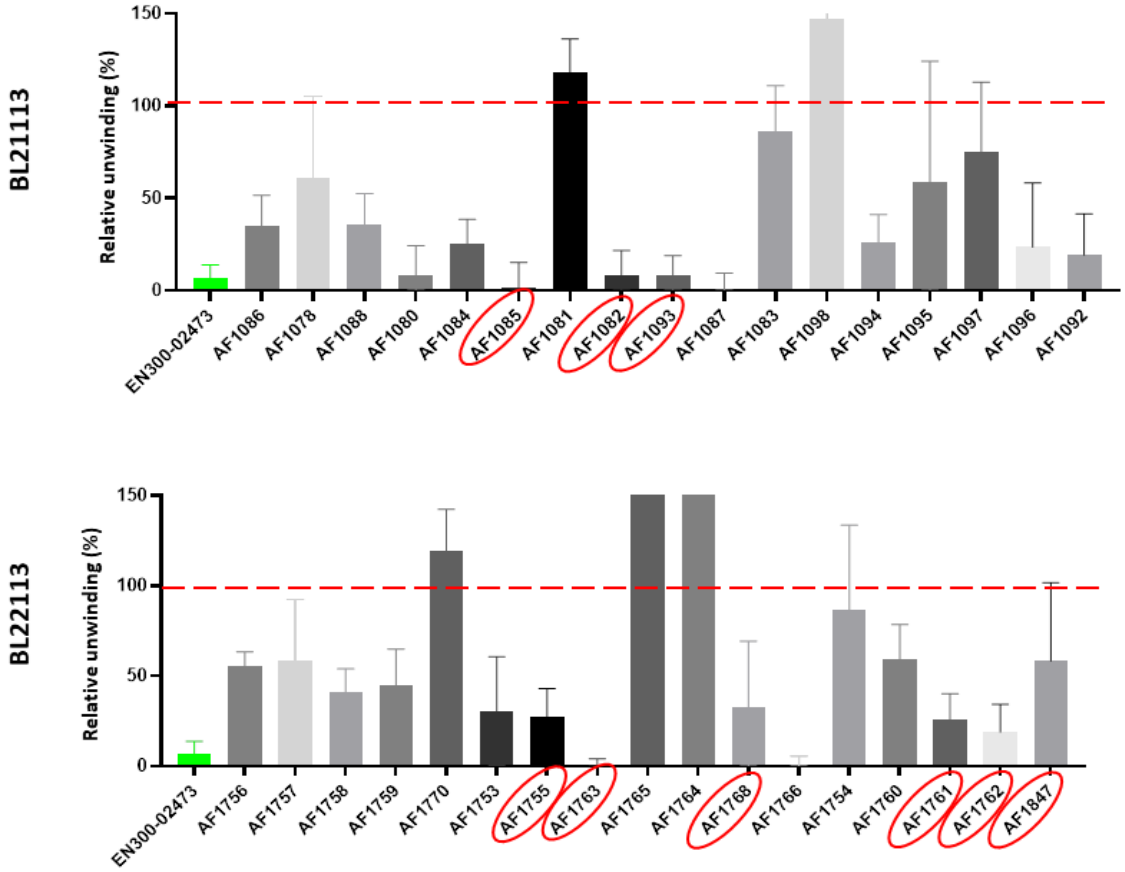


Figure 78. Screening EN300-02473 derivatives at 1 mM (10 nM PIF1 and 0.1 nM substrate PST55). Compounds highlighted in red show insolubility at high concentrations (≥ ~ 1 mM) in the aqueous reaction (n=3 repeats).

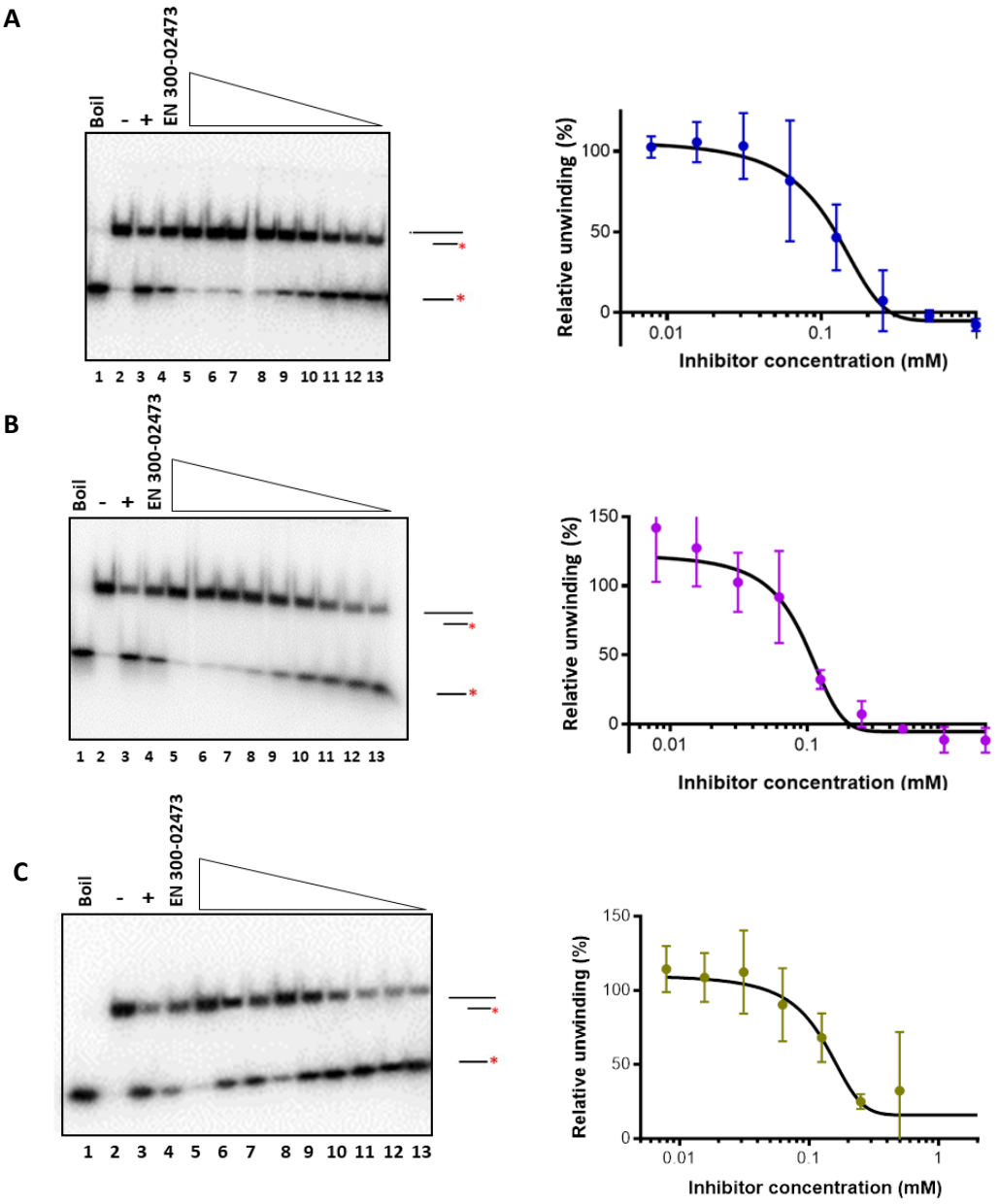


Figure 79. IC_{50} determination for EN300-02473 derivatives. Left, electrophoresis analysis of reaction products (n=3 repeats; 10 nM PIF1, 0.1 nM substrate): lane 1, boil; lane 2, native substrate (-); lane 3, hPIF1 no inhibitor (+); lane 4, 1 mM EN300-02473 (Enamine); lanes 5-13, doubling dilutions of compound from 1 mM. Data plots are shown on the right (A) AF1763, diluted from 1mM, IC_{50} 0.119 ± 0.046 mM. (B) IC_{50} of AF1766 diluted from 2 mM, 0.101 ± 0.021 mM. (C) AF1768 diluted from 1 mM IC_{50} 0.163 ± 0.045 mM.

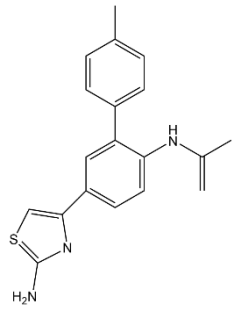
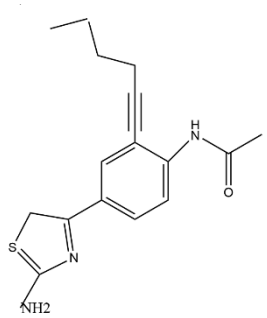
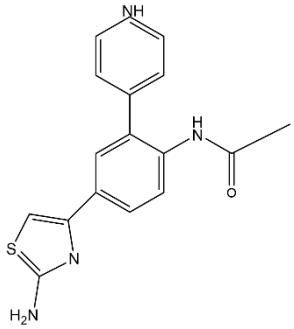
Name	Structure	IC_{50}
AF1763		0.119 ± 0.046 mM
AF1766		0.101 ± 0.021 mM
AF1768		0.163 ± 0.045 mM

Table 13. Structure and IC_{50} of EN300-02473 derivatives AF1763, AF1766 and AF1768.

A further series of compounds with modifications to the R^1 group of the EN300-02473 scaffold, series 3 from shipment BL22082, were generated and screened at 1 mM in the radiometric helicase assay (Figure 80), followed by a 4-point screen to selected molecules for IC_{50} evaluation. Only AF1718 (Table 14) was analysed and the IC_{50} value obtained was 0.587 ± 0.061 mM (Figure 81).

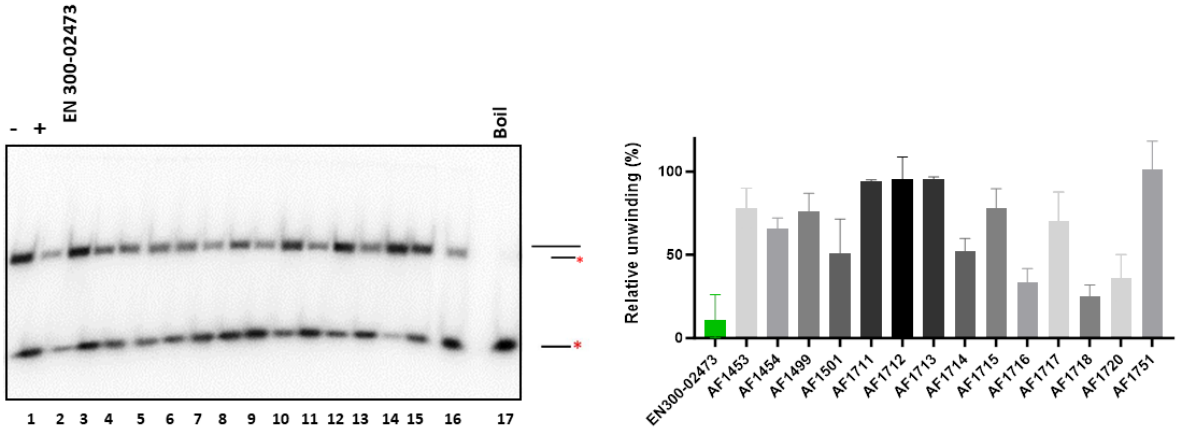


Figure 80. Single point screening of EN300-02473 derivatives (n=3 repeats; 10 nM hPIF1, 0.1 nM substrate). Compounds AF1716, AF1718 and AF1720 show inhibition of > 70%.

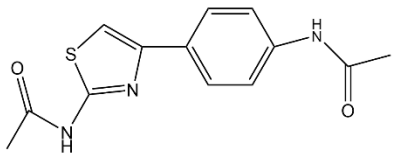
	Structure	IC_{50}
1		0.587 ± 0.061 mM

Table 14. Structure and IC_{50} of AF1718.

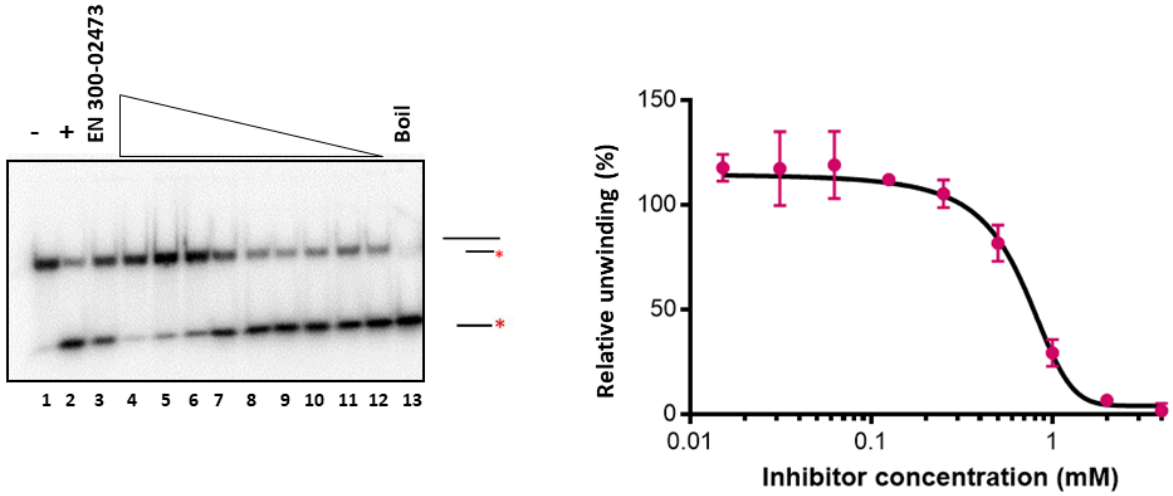


Figure 81. IC_{50} analysis of AF1718 (n=3 repeats; 10 nM hPIF1, 0.1 nM substrate). Left, electrophoresis analysis of reaction products: lane 1, negative (-) control; lane 2, hPIF1 no inhibitor (+); lane 3, 1 mM EN300-02473; lanes 4-12, doubling dilutions from 4 mM; lane 15, boil. From the data plot on the right an IC_{50} value of 0.587 ± 0.061 mM was obtained.

A final series ('series 4', shipment BL23036, 5 compounds) of five EN300-02473 derivatives was provided by Edelris with further modification in R^1. Compounds in series 4 were screened directly in a 4-point titration series (0.25-2 mM, Figure 82). IC_{50} values of 0.190 ± 0.08 mM and 0.420 ± 0.128 mM (Figure 83) were determined for AF1843 and AF1844 respectively.

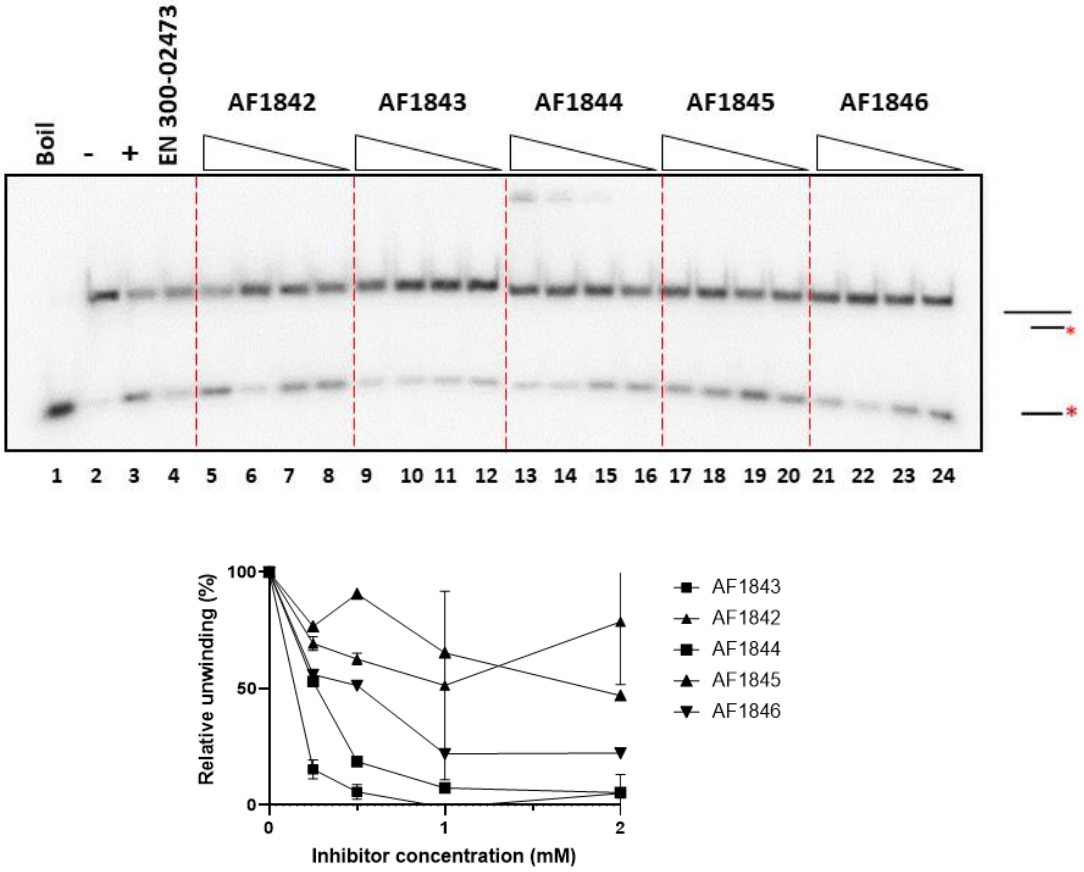


Figure 82. Four-point screen of EN300-02473 derivatives (2, 1, 0.5 and 0.25 mM; 10 nM PIF1, 0.1 nM substrate). In the electrophoresis: lane 1, boiled; lane 2, (-) native substrate; lane 3 (+) hPIF1, no inhibitor; lane 4, 1 mM EN300-02473; lanes 5-8, titration of AF1842; lanes 9-12, AF1843; lanes 13-16, AF1844, lanes 17-20, AF1845, and lanes 21-24, AF1846 (all n=3 repeats).

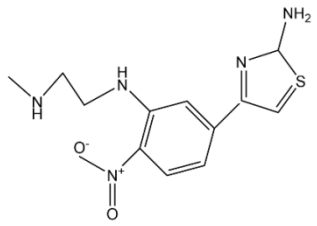
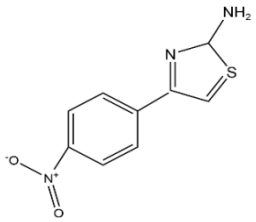
Name	Structure	IC_{50}
AF1843		0.190 ± 0.080 mM
AF1844		0.420 ± 0.128 mM

Table 15. Structure and IC_{50} of AF1843 and AF1844.

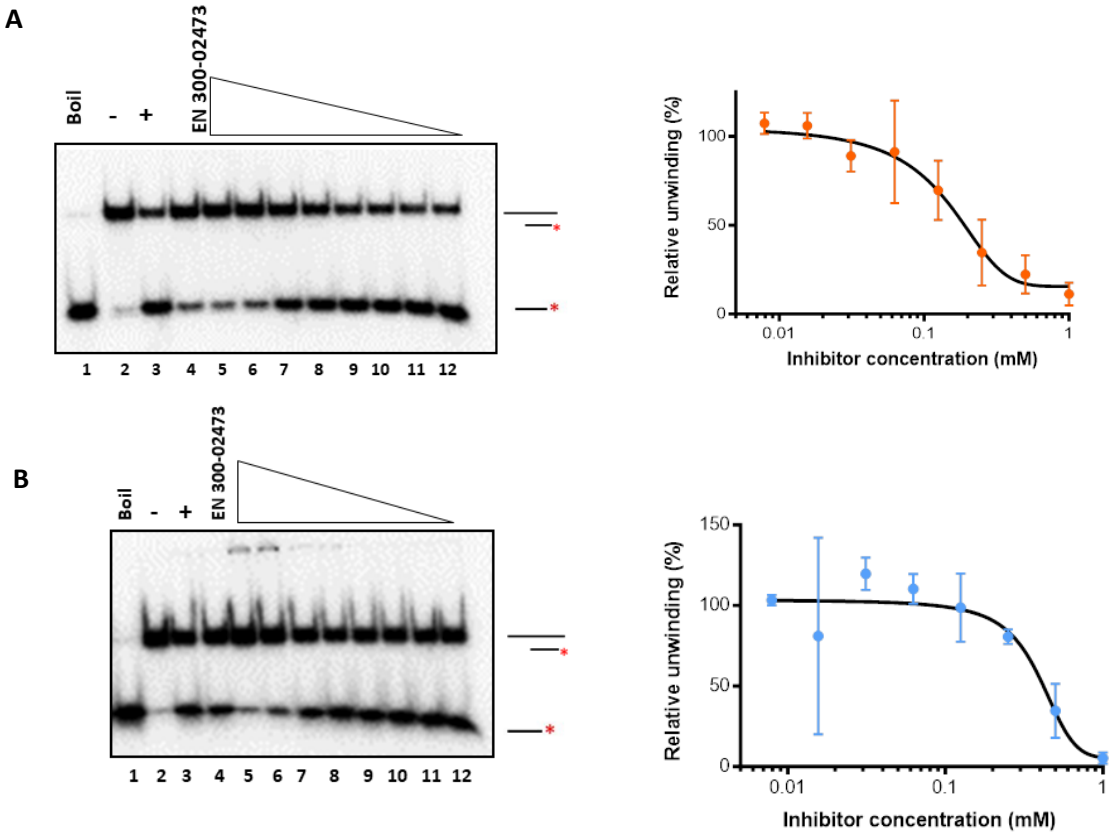


Figure 83. IC_{50} determination for (A) AF1843 and (B) AF1844 (n=3 repeats, 10 nM hPIF1, 0.1 nM substrate). In the analysis of reaction products by electrophoresis (left): lane 1, boil; lane 2 (-), native substrate; lane 3, (+) hPIF1 no inhibitor; lane 4, 1 mM EN300-02473; lanes 5-12, doubling dilutions from 1 mM. The data plot in the graphs on the right was used to determine IC_{50} values, 0.190 ± 0.080 mM for AF1843 and 0.420 ± 0.128 m for AF1844.

4 DISCUSSION

The important role of helicases in maintaining genome stability has made them attractive targets for treating cancer, where loss of function in tumours dependent on upregulated helicases to counter replication stress can initiate apoptotic cell death via the DNA damage response. Moreover, a full understanding of the role of helicases and their interacting proteins in DNA replication and repair during specific stages of the cell cycle is important for their development as targets and for identifying new targets for cancer therapy. To date, helicase MCM2-7 and the RecQ helicases (such as BLM and WRN), have been identified as potential targets [178] [179, 180] and drug discovery campaigns have been initiated. For example, an inhibitor named NSC617145 has been developed against the WRN helicase that can induce synthetic lethality in certain cell types [181], while for BLM an inhibitor known as ML216 has been developed as a potential anti-cancer drug causing anti-proliferative and DNA damage-inducing effects. In this case of ML216 however, the risk of developing secondary mutations in normal cells is high, as another repair pathway can be compromised [180, 182]. It has been suggested that hPIF1 is an attractive cancer therapy target with some potential advantages as such. When PIF1 in human cells is knocked down by siRNA and in combination with the action of some DNA damaging agents, the survival rate of some p53-deficient and p53-proficient cancer cell lines is significantly reduced [129, 183]. Studies have also demonstrated that PIF1 knockout mice show no abnormal phenotypes and survive [183], showing that PIF1 is a non-essential helicase. Taken together, these observations suggest the feasibility of a synthetic lethality-based therapeutic approach and that potential therapies involving PIF1 are unlikely to be toxic to normal cells. The only PIF1 inhibitor recently reported is called Tideglusib (TD), initially tested on *Bacteroides sp* Pif1 and afterwards on hPIF1 with successful results in *in vitro* unwinding assays. TD also displays a role (both *in vitro* and *in vivo*) as a neuroprotective agent mediated by PPAR-γ (Peroxisome proliferator-activated gamma) activation and it is an ATP-non-competitive inhibitor of GSK-3β [184]. However, as shown in Appendix, Figure 4A, the apparent inhibition on FL hPIF1 has not been reproduced in our hands, raising questions on the efficiency of the inhibitor.

Critical to hPIF1 role in genome stability are its interactions with G4 DNA. This has been well documented in a yeast model system [123], but less so in humans. However, knock-down of hPIF1 slows the progression of replication forks and this is

exacerbated in the presence of G4 DNA stabilising agents. Importantly, data in yeast and humans show that hPIF1 is a structure-specific G4 DNA-binding protein. Accordingly, it is important to understand the molecular basis of G4 DNA binding, as well as unwinding, since targeting the G4-binding activity could represent a potential route for specific therapeutic intervention. Here, an investigation of hPIF1 inhibition by potential small-molecule inhibitors was pursued. The investigations were based primarily on a previous drug discovery campaign initiated in Dr Sanders' lab and carried on in collaboration with Edelris Keymical, our partner in Lyone (France). Also, this work investigates the structural basis of hPIF1 binding to G4 DNA structures.

4.1 Production of purified recombinant hPIF1 mutants

In this work, several proteins were produced. His_{12}SUMO FL hPIF1 is the expression construct used for the purification of the full-length nuclear form of hPIF1. The expression uses the ArcticExpressTM BL21 (DE3) cells that co-express the GroEL/ES related chaperone (Cpn10/60) from the psychrophilic bacteria *O. Antarctica*, which allows a low induction temperature (8°C) necessary to reach a good yield of target protein. Moreover, working at low temperatures in both expression and purification is considered a fundamental condition for producing active enzymes. The purification of FL hPIF1 was performed in three column steps: the protein was first run on a His-trap column, the His-SUMO tag was removed with SUMO protease and the protein was then loaded onto a second His-trap column. Finally, the protein was run on an SEC column. The yield reached with this purification was around 0.4 mg per gramme wet weight of cells, considerably higher than that obtainable using a simple His6 tag construct used previously (S. Dehghani-Tafti, PhD thesis, UoS 2019).

4.1.1 Purification of GST fusion proteins

GST proteins were also produced in this work: GST FL hPIF1 (FL hPIF1), GST HD 206-END (hPIF1 HD-E), and GST HDΔC26 (hPIF1 HD). The choice of using a GST tag to investigate hPIF1-G4 interaction comes with the low non-specific binding exhibited by GST compared to His in pull-down assays. Moreover, the working conditions for the GST proteins expression is more advantageous, since GST

proteins can be expressed in the standard BL21(DE3) cells (6-8 hours induction), avoiding the low temperature and long expression times (2-3 days) needed with the Arctic Express™ (DE3) system. The mutants mentioned above in this work were successfully cloned or sub-cloned in a GST-tag vector to be expressed and purified as GST fusion proteins. The aim was to use a GST-pulldown “McKay” assay to investigate the binding of the mutants to different substrates of interest, both G4s and ssDNA, in an internally controlled way. The yield obtained was approximately 0.2 mg of protein per gramme of wet cells. However, the SDS-PAGE analysis showed the presence of low molecular weight contaminants, although these were deemed not detrimental to the intended use of the protein. Similar issues were encountered for the “wild-type” truncated forms hPIF1 HD-E and hPIF1 HD.

The production and purification of the GST-tagged helicase domain mutants (section 3.1.4) proved problematic, resulting in a significantly lower yield of protein compared to wild-type in all cases. In an attempt to improve protein production, PIF1 ORFs were cloned with a His-SUMO followed by a GST tag and the PIF1 ORF. The rationale for this was that His_{12}SUMO tagging of hPIF1 (and other proteins in the lab) had proven highly effective in the Arctic Express™ system, but GST tagged proteins were very poorly expressed. The only downside of the Arctic system is the presence of the highly over-expressed chaperone, which must be eliminated during the purification. In this case, the purification steps followed as for the His-SUMO FL hPIF1 previously discussed, with the elimination of the SUMO-tag in the process. A higher yield of protein was in this way obtained for the most problematic mutants (e.g. K556A), but with some residual chaperone (Cpn10/60) present.

4.2 Establishing an assay for hPIF1 unimolecular G4 DNA binding

Previous attempts in the lab to establish a gel-shift assay with uniquely unimolecular G4 DNA substrate had proven unsuccessful. Here, an alternative approach was taken (McKay assay), as will be discussed below.

4.2.1 Unimolecular G4 DNA substrates formation

As illustrated in section 1.4, G4 motifs are widely distributed in the human genome with different functions according to their location. However, in certain cases, they might be considered a problem. Specialised helicases can resolve these noncanonical secondary structures and avoid genome instability [87, 88]. hPIF1 can resolve G quadruplexes, which is not a particularly unique feature of helicases [176], but has also a more highly specialised ability to bind G4 structures. This property is shared with a sub-set of the helicases involved in the DDR such as WRN, BLM and FANCJ. Among the species, *S. cerevisiae* Pif1 has been demonstrated to be the most potent G4 unwinding helicase in the cell, able to unwind G4 DNA under single-cycle conditions [176].

A previous analysis from the lab had shown that hPIF1 can bind and resolve tetramolecular G4 DNA [146]. One of the aims of this work was to understand how hPIF1 binds and resolves unimolecular G4 DNA structures, the more physiologically relevant form. To do this, a series of radiolabelled unimolecular G4 DNA substrates were produced and folded in the presence of K^+ ions that dominate intracellularly [169] and result in more stable G4 structures compared to Na^+ [185]. A “reference” G4 substrate $T_0G_4T_4$ was produced on the basis of the *Oxytricha nova* telomeric structure, made of four G stacks and interconnecting T_4 segments with no ssDNA extensions at the 5' and 3' ends, thus representing a ‘pure’ G4 substrate. The rationale for this choice was that folding patterns for the substrate have been studied [165], and the T loops were considered to be relatively neutral in terms of sequence.

The ‘reference’ substrate was produced along with its ssDNA mutant ($M\_T_0G_4T_4$) where one of the Gs in each quartet is substituted with an A and therefore unable to fold into a stable G4 secondary structure. The ssDNA mutant is an important mobility control for the folded structure during its purification (Figure 29), as well as serving as a control for binding ssDNA of the same length and very similar sequence in assays. $T_0G_4T_4$ folded into three species that migrated at a faster rate compared to the unfolded control ($M\_T_0G_4T_4$), and some presumably unfolded residual material was also evident co-migrating with $M\_T_0G_4T_4$. One limitation of this study is that, due to the lack of specific instruments at our disposal, we do not know the folded conformation of these G4 species. Moreover, this is data that is not easy to

obtain. Techniques such as CD can differentiate between parallel and antiparallel G4 DNA composition, but they would require the preparation of substrates at a scale, which is not easy to achieve and given the limitations of the resolution of gel electrophoresis. Still, they can't determine precise structure. A similar discussion applies to the $T_1G_4T_4$ substrate that formed 3 folded species, although one dominated over the others, and substrate $T_1G_4T_3$. Notably, the physiological G4 DNAs, Telo22 and T_1Myc1245, resolved as one folded species but also resulted in a higher percentage of unfolded material, probably due to lower stability. $T_1G_3T_4$ folded into two species that, again, we are unable to describe structurally. Interestingly, the binding analysis showed a different level of hPIF1 binding to the two species.

The folded status of the G4 substrates produced was confirmed by DMS footprinting. Despite this being a well-established technique in the lab and always successfully performed with unimolecular G4 DNA using the G_4T_4 motif in the context of long oligonucleotides (Figure 3A), the analysis of the short G4 substrates used here was more problematic. The basis of this technique is the ability of DMS to react with the guanines when not hydrogen bonded, but not in folded G4 DNA. Reacted sites are then sites for piperidine cleavage. Accordingly, it would be expected to see reactivity only with the ssDNA mutants and the boiled (denatured) G4 reactions, and not with the folded G4 substrates. Looking at Figure 32, panel A, the situation is not clear: There is not a substantial difference in reactivity between native and boiled G4 DNA, but for M_ $T_1G_4T_0$ it is clearly possible to see the results of DMS-guanine reactivity and discriminate where the G–A substitutions are. It could be that the short length of the substrate (< 30 nts) facilitates rapid reannealing of the G4 state before significant DMS reactivity.

4.3 Establishment of a GST pulldown "McKay" assay

To probe hPIF1 G4 DNA binding activity, a GST 'McKay'-type pulldown assay was performed. The original assay was an immunoassay able to quantify the interaction of SV40 T antigen with different radiolabelled DNA sequences from the viral genome; the bound DNA was then separated from the free and analysed by agarose gel electrophoresis [166]. In the assay different DNA substrates (*e.g.,* both

ssDNA and G4 DNA) labelled at the 5'-end with ^{32}P γ-ATP can be mixed and freely compete for protein binding. The GST-tagged protein can bind to the glutathione beads providing a recovery handle for the radiolabelled substrate. After incubation and washing, the bound products are recovered and run on a polyacrylamide gel where it will be possible to discriminate the relative affinity for each substrate. In principle, the assay is an internally controlled system, minimizing the effects of experimental variation and useful for revealing small differences in affinity.

Despite the McKay assay protocol being already established and successfully employed in the past, optimisation is necessary for each protein used. One issue encountered here was substrate binding to the beads used (background), as explored in Figures 33 and 34. The background level was much lower when the assay had been performed in the past in the lab, and this may indicate batch variability between beads. Indeed, this was observed, and coloured beads designed to aid their visualisation during such pulldown experiments were particularly problematic (EZviewTM Re glutathione affinity gel, Sigma). To help minimise this matter GSH beads used were minimised and combined with relatively inert beads (sephadex G15) to maintain a suitable volume of beads for technical handling. Pre-blocking with non-specific protein and DNA was also helpful but did not completely eliminate the problem.

The KCl concentration used was explored. KCl is required for G4 stability and since DNA recognition is mostly determined by shape and charge [185], binding stringency would be influenced by its concentration (weaker binding with higher salt). A KCl concentration of 100 mM was chosen as ideal both because of lower background binding and being close to the intracellular K^+ concentration [186].

4.3.1 Exploring hPIF1 DNA binding with the McKay assay

The investigation started with verifying known DNA binding properties of hPIF1 with the McKay assay. Previous EMSA data for hPIF1 [145] showed that the hPIF1 helicase core displays a length-dependent binding affinity without nucleotides (higher affinity for longer ssDNA substrates > 35 nts). Also, there is a large increase in ssDNA binding affinity for small ssDNA substrates (~10 nucleotides) observed in the

presence of cofactor ADP·AlF_4^-, which correlates with large structural rearrangements in helicase core (section 1.7.4, [135]). The first pulldowns performed used poly d(T) substrates from 10 nts to 50 nts and confirmed the observations above: hPIF1 affinity is length dependent with an increase of binding affinity in presence of the transition state analogue (Figure 36).

FL hPIF1 shows a higher binding affinity for all the substrates tested (G4, short ssDNA and long ssDNA). We hypothesise a role played by the presence of the N-terminal domain, which has a non-specific ssDNA binding activity [145] [136]. Binding to truncated hPIF1 helicase domain constructs with and without a C-terminal peptide was also tested (GST hPIF1 HD 206-END and GST hPIF1 HDΔC206). Interestingly, when looking at Figure 37 C, we can see how in the core helicase domain missing the C-terminal (hPIF1 HD), the binding ratio between M_$T_1G_4T_4$ and $T_1G_4T_4$ substrate is 1:25 while in hPIF1 HD-E it is 1:3. This results suggests it is possible that the C-terminal domain 21 residues, not visible in the electron density maps of the HD core [135] but predominantly acidic, may act as a negative regulator of ssDNA binding.

To better understand hPIF1-G4 DNA binding a series of G4 DNA substrates was produced (section 3.3.3). A comparison between $T_0G_4T_4$ and $T_1G_4T_4$ (differing in the presence or absence of a 5' T residue on the core G4 fold) did not highlight any significant differences in binding, suggesting that the tail is not essential for hPIF1 binding to G4 structures, unlike what has been observed for G4 DNA unwinding [146], and supported by structural observations of *To*PIF1 bound to G4 DNA [156].

G4 DNA is a secondary structure unevenly distributed in the human genome, with a high concentration in telomeres and proto-oncogenes [187]. Therefore, two ‘physiological’ G quadruplexes were also considered in the analysis, Telo22 found in telomeres and T_1Myc1245 found in the Myc proto-oncogene promoter, synthesised with a 5' tail of one residue (as in $T_1G_4T_4$), and characterised by three G4 stacks. With the McKay assay, these two physiological G4 DNAs were analysed along with substrate $T_1G_4T_4$ adopted as a reference. Interestingly, the two ‘physiological’ G4 DNA substrates showed little or no binding to hPIF1 (both full length and helicase domains). These results suggest that hPIF1 may preferentially bind the more stable forms of G4 DNA with four or more G4 stacks. This observation was later confirmed by the analysis of substrate $T_1G_3T_4$, a similar substrate to the reference, but made of

three G4 stacks instead of four and showing no binding to hPIF1 (but also a certain instability in the assay, Figure 43). These results raise the question of whether the high G4 DNA concentration composed of three G4 stacks found in telomeres and the MYC proto-oncogene promoter is a target for hPIF1.

Several helicases are involved in telomere maintenance, including BLM,WRN, RTEL and PIF1, with a predominant role of BLM in cases of telomere dysfunction [188]. *Sc*Pif1 is involved in telomere DNA replication [189], inhibition of telomerase and its overexpression leads to a shortening of telomeres. I*n vitro* and *in vivo* data suggest that this is also the case in human cells, with hPIF1 is involved in telomere elongation affecting telomerase processivity [147]. However, our results raise the question of whether hPIF1 can be a major player in telomerase maintenance, since no binding to the telomer G4 was observed.

The investigation of the $T_1G_4T_3$ substrate aimed to understand the role of the loop length in binding (we failed to produce stable folded substrates with one or two T loops). Interestingly, when synthesized, the substrate resolved as two species in gel electrophoresis, and both were analysed in this work and showed a remarkably different level of binding to the hPIF1. These data highlight also that hPIF1 is exquisitely sensitive to the nature of the G4 fold, as well as G4 stack content. Thus, it appears that hPIF1, and by extension the other G4 DNA resolving helicases in the cell, may be specialised to deal with different classes of the polymorphic G4 DNA structures. It would have been very interesting to know what the conformation of the two species, but for reasons outlined above this is a non-trivial endeavour. Nonetheless, the $T_1G_4T_3$ investigation confirmed the importance of the T loops and fold in the G4 structures and its importance for binding, to the best of my knowledge, never previously reported in the literature for any G4 DNA binding protein.

All the binding data obtained here needs to be verified by K_D analysis to complement the investigation. A lack of time prohibited this here. However, it may be problematic as previous efforts to use microscale thermophoresis for PIF1 HD have been unsuccessful in the lab, with titration curves experiencing erratic fluctuation at high protein concentration without a stable binding plateau being reached. A likely cause is protein aggregation or multimerization, that would also affect other methods such as fluorescence anisotropy. As well, it would be important to validate the data

obtained in the McKay assay with the substrates used in the analysis singularly, to rule out potential interference (although the substrate concentrations used were very much lower than the protein concentration).

4.3.2 Changes in DNA binding affinity during the ATP*ase* cycle

One aim of this study was to understand how hPIF1 G4 DNA affinity changes during the ATP*ase* cycle. The objective was to use cofactors to reproduce steps in an idle cycle of ATP hydrolysis where AMP·PNP/Mg^{2+} (non-hydrolysable ATP) mimics the ground state, ADP·MgF_4^- is a transition state analogue and ADP/Mg^{2+} the product state. To obtain hPIF1 core domain structures bound to ssDNA ADP·AlF_3^- was used (*i.e*. crystallisation included aluminium chloride). However, a study by Jin *et al* (2017) investigating high-resolution structural data indicates that when Mg^{2+} is present, more often than not, the enzyme synthesises MgF_4^- instead of AlF_3^-. In fact, Mg is octahedral and forms complexes with 6 ligands while trifluoromagnesate is five-coordinate; also, MgF_4^- is isoelectric with PO_4^- making it ideal for mimicking the phosphoryl transfer [190].

It is known that there is a large increase in ssDNA binding affinity going between apo and transition state, as confirmed here (Figure 34). In the McKay assay combining oligo $d(T)_{50}$, $T_0G_4T_4$ / $T_1G_4T_4$ and their mutant controls (same length, 28/29 nt), in the apo condition the affinity of hPIF1 is higher for the long ssDNA substrate than for the G4 substrate, while little or no binding is observed for the short ssDNA, confirming again the previous data obtained with EMSA. Importantly, the situation completely changes with ADP·MgF_4^- where the affinity for the G4 substrate is completely lost and the affinity for all ssDNA substrates is vastly increased. Interestingly with ADP/Mg^{2+} affinity for all the substrates is reduced or lost.

Notably, unimolecular G4 DNA (substrates $T_1G_4T_4$ and $T_0G_4T_4$) does not stimulate ATP hydrolysis significantly (Figures 40 and 41), and the low level of stimulation observed in the assay could be due to the unfolded or transiently unfolded G4 DNA (see section 3.5).

4.3.3 Mutational analysis of a potential G4 DNA binding sites to hPIF1

A principal objective of this work was to investigate the interactions of the hPIF1 helicase with G4 DNA structures and identify domains and residues directly involved in the binding. During the study the first crystal structure of a PIF1 protein, from *Thermus oshimai,* complexed with G4 DNA became available. The structure at 2.58 Å resolution was obtained in the presence of $ADP{\cdot}AlF_4^-$, with the G4 structure made of three stacks, a 6 nt 5' tail and an 8 nt 3' tail. The authors concluded that the structure reveals the G4-Recognizing Surface (GRS) and the residues directly involved in the binding, and the model was tested biochemically through DNA binding and helicase assays of GRS mutants.

The bioinformatics analysis presented here did not reveal significant conservation in the GRS among a diverse selection of species including model organisms. Comparing several model organism sequences, very low genetic conservation (~10%) is observed and low conservation is observed in the structural analysis too. When comparing the nine residues involved in *To*PIF1-G4 DNA interaction to hPIF1, only a few residues show conservation: *To*PIF1 R355 is in motif B of domain 2B and its human corresponsive residue is K485. Arginine and lysine have very similar structures and are both positively charged so the idea of lysine fulfilling the arginine role is feasible. hPIF1 W529 structurally correspond to *To*PIF1 F394, with tryptophan and phenylalanine being both with aromatic rings, and possibly considered conserved. Finally, H525 in human could fulfil the role of *To*PIF1 R392 (arginine and histidine, both positively charged). In the mutants' analysis, R290 was also considered as a study with yeast Pif1 [149] revealed R290 (*Sc*Pif1 R325) to be considered conserved and implicated indirectly in G4 DNA binding.

GST-hPIF1 proteins with altered amino acids ("mutants") in the GRS were tested in the McKay DNA binding assay along with proteins with amino acid substitutions in the conserved ssDNA binding channel. The rationale was that for the ssDNA binding channel mutants, a loss of affinity for the ssDNA substrates was expected, as actually observed, but not loss of binding affinity for the G4 DNA substrate. However, the latter was not observed and the affinity for the G4 DNA substrates was also decreased. Among the putative GRS mutants, Q542A and R528A appeared to have reduced affinity for the G4 substrate, but on closer analysis the ratio of recovery of

$T_1G_4T_4$ compared to M_$T_1G_4T_4$ suggested a relative *decrease* of affinity for the short ssDNA substrate. This is not consistent with the initial hypothesis since Q542 and R528 are not located in the ssDNA binding channel. The robust wild-type level of binding to the long ssDNA substrate ($d(T)_{50}$) suggests that the proteins are folded and active. One possible explanation could be an overlap between the ssDNA and the G4 binding sites, which was supported by the trend observed in the competitor binding assay (Figure 46). However, it should be emphasised that the McKay DNA binding experiments were only performed once due to time constraints, although the internally controlled nature of the assays with multiple substrates gives a level of reliability to the results.

The loss of affinity for the G4 substrates in the case of the ssDNA binding channel mutants, could be linked to the possibility of partial unfolding of the G4 structure during their recovery from the reactions, and the formation of a partially single-stranded intermediate. Indeed, the G4 helicase assays (Figure 59) indicated that G4 unwinding could occur without ATP hydrolysis (Figure 60), but it should be emphasised that the substrates used in binding reactions and helicase assays are not equivalent. This idea is supported by findings identifying intermediate structures in the folding (and presumably also unfolding) of G4 DNA, as discussed in section 1.4.2 of the introduction. Also, a recent study on BRCA2-G4 DNA interactions suggests the formation of a G-triplex intermediate which has a short half-life and when recognized by the protein can prevent the formation of folded G4 DNA. More experimental work, however, is needed before drawing any solid conclusion.

Overall, however, the studies presented here suggest that the *To*PIF1 GRS data should be accepted with caution. First, the proposed GRS has very low conservation. This is particularly surprising given the very high degree of conservation in the ssDNA binding channels across the evolutionary spectrum (Table 6) [135], and the fact that G4 DNA binding is a defining characteristic of PIF proteins. Also, the *To*PIF1-G4 DNA structure was obtained in the presence of $ADP{\cdot}AlF_4^-$. Here, I observed the complete loss of G4 DNA binding affinity in the presence of transition state analogue as detected by the McKay pulldown assay. Furthermore, hPIF1 does not bind with appreciable affinity to G4 substrates with 3 G quadruplexes, as used for the *To*PIF1 analysis. Finally, a different model for G4 DNA binding has been proposed for *Sc*Pif1

involving the unique yeast 2C insertion and the relatively unconserved wedge domain [149]. Again, it would be surprising to see the same structural recognition problem solved in two entirely different ways within a structurally conserved protein family. The *To*PIF1-G4 structure may represent a low affinity, post catalytic, interaction.

Experiments with the G4 stabilizer Phen-DC3 showed inhibitory activity of the drug on hPIF1-G4 DNA binding. Previous studies on Phen-DC3 demonstrated its poor permeability into the nucleus, but with the help of cell-penetrating peptide, there is an increased localization of the drug in the nucleus, allowing Phen-DC3 to be used for studying G4 structures and its potential to target specific tissue or tumours, to induce a DNA damage response through G4 DNA stabilisation [191]. In the case of hPIF1, Phen-DC3 inhibited G4 DNA binding. However, the analysis also showed a decrease in binding to short ssDNA control (M_$T_1G_4T_4$), but not d(T)50. It is possible that Phen-DC3 was able to induce the control substrate to form a G4 fold, which was not recoverable by hPIF1 binding. Nonetheless, the ability of Phen-DC3 to prevent hPIF1-G4 DNA interactions could represent a new therapeutic route.

4.4 G4 DNA unwinding by hPIF1

As discussed in the introduction, because of their ability to generate single-stranded nucleic acids, helicases are involved in the resolution of non-canonical secondary structures such as G4 DNA, that can induce genome instability. It has been suggested that PIF1 could be one of the most important helicases in G4 DNA resolution, at least in yeast [139] [140]. As indicated by the PIF1 requirement for the stability of the human minisatellite CEB1 repeat in the yeast genome, yeast PIF1 acts as a G4 DNA binding and processing enzyme [123]. Experiments conducted *in vitro* on synthetic tetra-molecular G4 DNA substrates [146], demonstrated that hPIF1 requires a 5' ssDNA tail of sufficient length (optimum unwinding at 55 bases) to initiate efficient unwinding of simple partially single-stranded duplex (ss/dsDNA) substrates as well as G4 DNA [146] [145].

Investigation of hPIF1 helicase activity on G4 DNA was comparatively underexplored in this work, but despite so, some interesting observations were made.

We know that hPIF1 is able to unwind tetramolecular G4 DNA when this has a long ssDNA > 35 nts for helicase loading [146] and that it needs ATP to retrieve energy and display its helicase activity on dsDNA substrates (no unwinding observed without ATP/Mg^{2+}). The substrate used for the unwinding assay presented here has a 5' ssDNA tail of 55 nts labelled with ^{32}P, followed by a G4 structure (based on *Oxytricha* telomer T_4G_4, as for the binding assays) and a dsDNA component of 20 nts, with the 20 base oligonucleotide also 5'-end labelled. The reactions with protein show very minimal displacement of the 20 base oligonucleotide but formation of a slower mobility product indicative of a partially double- and single-strand DNA with the G4 resolved (Figure 59).

The first new observation is that hPIF1 can display a minimal level of helicase activity on the partially double- and single-stranded substrate PST55, even without ATP. However, the ATP-independent unwinding of the G4 DNA appears far more robust. This is not entirely surprising as, even if helicases have been defined as ATP-dependent enzymes, they can function even without using ATP. For example, [192] hepatitis C virus NS3 (SF2) can unwind dsDNA without ATP. However, G4 DNA unwinding without ATP has been more widely described: smFRET analysis demonstrated as RHAU and BLM helicases do not need ATP (or other nucleotides) binding or hydrolysis to unfold G4 DNA [193], differently from what is observed with WRN helicase which needs ATP to display its activity [194].

This findings on nucleotide-dependent hPIF1 G4 unwinding activity are further confirmed in Figure 60. Here, with cofactors that mimic steps in the ATP hydrolysis cycle, a complete unwinding of G4 DNA is observed in the presence of $ADPMgF_4^-$, while ADP and AMP-PNP support an intermediate level of unwinding, greater than no cofactor. It would have been interesting to evaluate the effect of nucleotide analogues on DNA unwinding using the partial double-single strand substrate PST55 or other G4 substrates, but lack of time limited the analysis.

To summarize, the investigation of hPIF1:G4 DNA interactions supports the idea that hPIF1 binds selectively to G4 DNA structures. Although further investigation is required, the results presented here do not support similarities with the proposed *To*PIF1-G4 DNA binding model. Indeed, the mutational analysis and competitor

challenge experiments open the possibility of an overlap between ssDNA and G4 DNA binding sites. Further investigation is needed to address this and understand the structural basis of G4 DNA binding by the human protein, which would be best achieved through direct structural observations. Additionally, while the unwinding assays show interesting dynamics in the hPIF1 processing of G4 DNA, further analysis of other radiolabelled substrates is required. However, the data do indicate that G4 DNA unwinding is coupled to the large structural movements that occur upon ssDNA binding and ATP hydrolysis and support a model of initial hPIF1-G4 DNA recognition in the apo state.

These results obtained from both unwinding and binding assays on G4 DNA could have implications for understanding how hPIF1 displays its helicase activity. The protein is likely to load on a ssDNA segment of >35 nt (binding observed in apo condition to long ssDNA, Figures 36 and 37), where affinity for G4 DNA is also high. When ATP/Mg^{2+} is bound, (AMP·PNP/Mg^{2+} ground state mimic) structural rearrangements in helicase core begin [135] and hPIF1 loses G4 DNA affinity and subsequent ATP hydrolysis results in large structural rearrangements, and an increase in ssDNA binding affinity presumably linked to translocation and DNA unwinding. Also, no G4 DNA binding is observed in the product state (ADP/Mg2+), as expected.

4.5 Helicase assay to probe for inhibitors of DNA unwinding

The second objective of this work is to exploit hPIF1 as a therapeutic target and develop inhibitors of hPIF1 helicase activity. Helicases and their interactive proteins in DNA replication and repair pathways are considered to be targets for cancer therapy [182]. Several helicases including BLM and WRN have been reported to be successfully targeted with small molecule inhibitors in cell-based models [179 180], while hPIF1 has also been considered to be of particular interest [129 183]. The second objective of this thesis work was to identify and develop small molecule inhibitors for hPIF1. So far, information on potential inhibitors is limited, and in the literature, only tideglusib (an FDA-approved drug) has recently been reported to inhibit PIF1 *in vitro* [184].

With this aim a drug discovery campaign targeting hPIF1 has been developed in the Sanders laboratory, and furthered here in collaboration with Edelris Keymical, Lyon (France), through the European Union Horizon 2020 Antihelix Project. Three approaches have been adopted to identify potential inhibitors/binders: (i) Direct binding of fragment-like molecules (< 200 Da) was screened using differential scanning fluorimetry (DSF/ thermal shift assay). This approach (Dr. Dehghani-Tafti, University of Sheffield, (2018)) identified 67 binders from the MRCT fragment library (n = 2200), among which EN300-163911 was selected by Dr. Mark Wever at Edelris for further development. (ii) A molecular docking of the Edelris compound collection was performed by Dr. Mark Wever to identify potential hPIF1 inhibitors. (iii) A 1.7 Å structure of the hPIF1 helicase domain (HD) bound to an EN300-02473 was obtained in a high-throughput co-crystallization screen (Diamond Light Source, Oxford, UK. Dehghani-Tafti and Dr. Ben Bax, Medicines Discovery Institute, Cardiff University) [135]. Chemical derivatives with potentially improved inhibitory activity were synthesised (Dr. Mark Wever, Edelris) and analysed here using a radiometric helicase inhibitor assay.

The medicinal chemistry and chemical synthesis aspect of the analysis was technically executed by Edelris, our partner in France. Here, a radiometric unwinding assay with a simple synthetic partially single- and double-stranded dsDNA as a general helicase substrate for FL hPIF1 was exploited.

Previously, all initial screens (enzymatic assays and protein thermal shift) for hPIF1 inhibitors had been conducted in the lab using hPIF1 HDΔC26. Once initial hits were obtained, a more focused evaluation (secondary screens) was performed using FL hPIF1. The reason for this was for ease and economy since the full-length enzyme was difficult to purify in sufficient quantity (although the purity and activity were excellent). First, the activity of hPIF1 purified as a His-SUMO tagged protein was evaluated. Previously, the protein was purified in the lab as a His-tagged protein but with a substantially lower yield and with greater effort than the newly established protocol. Here, as described in materials and methods, some changes to buffer conditions were also evaluated but, a side-by-side comparison with ‘archive’ protein (FL hPIF1) suggested that nothing significant would be gained from the changes. Furthermore, a titration of the FL hPIF1 protein was performed to determine the best

concentration to use for testing small molecule inhibitors. It was observed that maximum dissociation of the synthetic radiolabelled dsDNA was obtained between 5 nM and 10 nM hPIF1, ideal for the analysis of potential inhibitors of hPIF1. Evaluation of hPIF1's DMSO (solvent for small molecules) tolerance, showed a decrease of 3% of unwinding for each 1% increase in DMSO concentration in the reaction. A maximum of 2% of DMSO in the reaction was chosen for the analysis.

4.5.1 Inhibitor assay

Three approaches had previously been adopted in a drug discovery campaign to identify inhibitors/binders: (i) Direct binding of fragment-like molecules (< 200 Da) was screened using differential scanning fluorimetry (DSF/ thermal shift assay). With this approach (Dr Dehghani-Tafti, University of Sheffield, (2018)) identified 67 binders from the MRCT fragment library (n = 2200), among which EN300-163911 was selected by Dr Mark Wever at Edelris for further development. (ii) A molecular docking of the Edelris compound collection was performed by Dr Mark Wever to identify potential hPIF1 inhibitors. (iii) A 1.7 Å structure of the hPIF1 helicase domain (HD) bound to an EN300-02473 was obtained in a high-throughput co-crystallization screen (Diamond Light Source, Oxford, UK. Dehghani-Tafti and Dr Ben Bax, Medicines Discovery Institute, Cardiff University) [135]. To follow, several different routes were adopted here to expand on these findings and identify/develop inhibitors with improved potency.

4.5.2 EN300-163911 and its derivatives

EN300-163911 (183.25 Da), originally identified in a DSF screen, was selected at Edelris for further development for several reasons: EN300-163911's chemical structure was easy to modify and improve because of its potential for diversification but also starting material was cheap and easy to obtain. Also, the molecule had not been exploited before, according to the chemical literature and chemical databases such as Reaxys and SciFinder. Despite this, when inhibitory activity was explored, it

was not possible to determine an IC_{50} value, but a concentration-dependent inhibitory activity was detected at the upper concentrations tested (Figure 64). However, this is not to be unexpected for fragment (rule of three) sized molecules. Further development of its structure through chemical modification resulted in several derivatives and the evaluation of an IC_{50} for three of them, A1, A3 and A4 (section 3.6.2) was possible (84 to 321 μM). Structurally speaking, the thiazole amine replacement was crucial to improve the compound activity along with the tropane core. However, after those improvements, no additional improvement in activity through optimisation of the compounds' structure was achieved because larger chemical substitutions would not improve the inhibition. Therefore, a decision was taken with Edelris to stop the synthesis of further analogues and evaluate binding through structural biology/co-crystallisation in collaboration with Dr Silvia Onesti in Electra Sincrotrone, Trieste, (consortium member). These experiments are ongoing.

4.5.3 Potential inhibitors identified by virtual screening of the Edelris chemical collection

Docking of the Edelris chemical library collection identified more than 300 entities as potential binders/inhibitors. 96 compounds were selected for analysis, among which B6 and C6 resulted in a high level of inhibition (section 3.6.3). However, despite inhibition being observed in a single-point screening (1.5 mM), it was not possible to obtain an IC_{50} value from them. This was partially linked to the relative insolubility of the compounds under helicase reaction conditions, likely linked to the compound size and possibly to the presence of a sulphonamide-alkyl. Despite sonication to drive compounds into solution in DMSO stocks, no improvement in IC_{50} determination was obtained. Further improvements resulted in a series of derivatives and the selection of two compounds, AF1486 and AF1487, with improved inhibition and determination of IC_{50} values. Despite this, the molecular weight of these compounds is high (423.6 Da) and therefore further evaluation and modification of their structure is needed to make molecules suitable for progression. However, we can consider this route of screening as successful because a novel compound series of hPIF1 inhibitors was identified through a structure-based approach.

4.5.4 Development and screening of EN300-02473 derivatives

EN300-02473 was the compound most widely explored in this work. Previously identified in a co-crystal bound to hPIF1 HDΔC26, an IC_{50} value of <0.1 mM was determined (S. Dehghani-Tafti, PhD thesis, UoS 2019) for the compound purchased from Enamine. Also, structural analysis of the interaction between this inhibitor and hPIF1 suggested a clear rationale for its inhibitory activity, being sandwiched between the 2A and 2B domains of the enzyme and making close contact with a conserved Lysine (Lys556) in the ssDNA binding channel of the enzyme. Mutation of Lys556 completely inhibits helicase activity [135]. The results obtained from this investigation suggested that EN300-02473 specifically binds hPIF1 and it can be developed into a more potent inhibitor. Additionally, because of the reproducibility of the results and its known binding mode at the atomic level, EN300-02473 was adopted as a 'standard inhibitor' against which others in this work were evaluated. At Edelris, EN300-02473 was resynthesized through three different routes, while structure and purity were confirmed by NMR. Surprisingly, however, compared to the original compound purchased from Enamine, a striking difference in their inhibitory activities was observed, as well as physical differences in the colours of the compounds obtained. The difference in colours suggested the presence of a metal impurity, which was demonstrated to be palladium by ICP-MS (Inductively Coupled Plasma Mass Spectrometry) in the Enamine batch. For one of the re-synthesized EN300-02473 compounds it was possible to obtain an IC_{50} value but at a much higher value (0.885 ± 0.135 mM), suggesting that the faux inhibitory activity of the Enamine product may be due to a contaminant, but none the less, the compound is a genuine hPIF1 inhibitor.

The EN300-02473 structure has been modified and improved to obtain more potent molecules. In the analysis two groups went under separate modification, indicated as R1 and R2 (section 3.6.4.2). All the series of modifications brought successful improvement in the inhibitory activity of EN300-02473, evidencing six derivatives with improved IC_{50}. Looking at the successful inhibitors' structure, it is interesting to see how their IC_{50} value improves when the residue in position R1 or R2 is small, suggesting a better fitting of the inhibitor in the hPIF1 binding pocket. A

possibility in the future would be to combine the modification made on the two residues together to obtain more potent single molecules.

4.6 Conclusions and Future Work

This thesis work contributed to improving knowledge of the hPIF1 helicase. A biochemical assay able to analyse simultaneously binding of multiple DNA substrates was developed and demonstrated that hPIF1-G4 DNA binding is selective for G4 DNA conformation. Through this internally controlled 'pull-down' assay, changes in apparent G4 DNA binding affinity have been detected during the ATP hydrolysis cycle, that fit with a model for initial recognition of G4 DNA in the apo state followed by unwinding and ssDNA translocation. The study also demonstrates that hydrolysable ATP is not fundamental for hPIF1 to unwind G4 structures, but it is greatly stimulated by its presence. Also, the mutational analysis of a potential GRS first identified in bacterial PIF1 highlighted differences with bacterial model, opening the possibility that further investigations are required before the studies identifying the GRS can be considered conclusive. The work, also contributed to the development of new hPIF1 small molecule inhibitors, starting from novel routes never explored before, but also advanced understanding of the potential of EN300-02473, previously crystallised with the helicase core domain, as a route for drug development.

Further investigations should address the structural basis of G4 DNA binding specificity observed here, to give a more complete overview of hPIF1 activity on G4 structure. Also, more G4 substrates should be tested both for binding and in the unwinding assay to understand the mechanisms adopted by the helicase while resolving G4 DNA at stalled replication forks. One of the most important questions to address in the future concerns the true nature of the G4 DNA recognizing surface in hPIF1.

The data obtained with the inhibitor analysis, demonstrating a higher inhibitory effect with structurally smaller EN300-02473-like molecules, helps understanding of the binding pocket and mechanism of inhibition. Further development of the molecule should lead to molecules able to be tested in cell-based assays. Co-crystalisation of potential hits developed by new synthetic routes with hPIF1 should be a priority.

5 APPENDIX

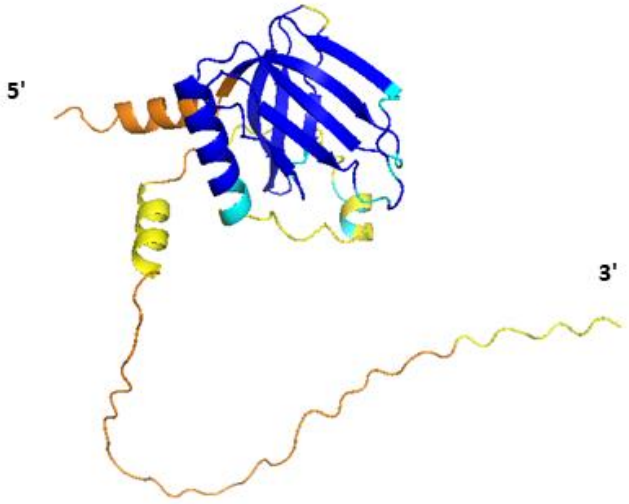


Figure A1. hPIF1 N-terminal AlphaFold prediction. The colours indicate the 'model confidence' of the prediction: orange very low; yellow low; cyan blue confident; blue very high.

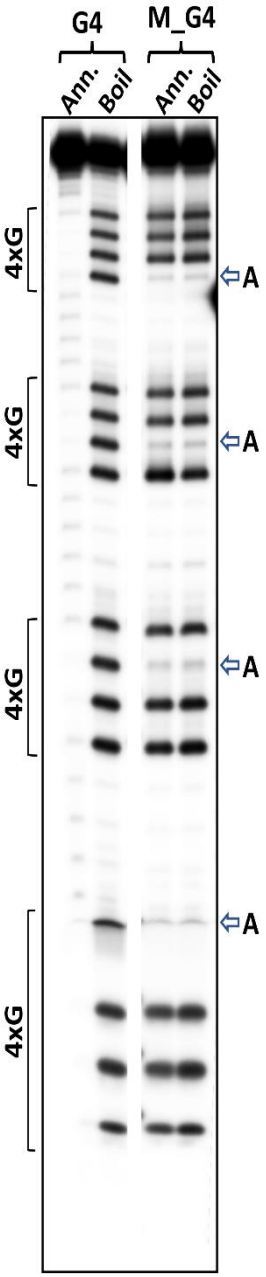


Figure A2. DMS footprint of unimolecular $T_{10}G_4T_4$ substrate and its mutant previously obtained in the lab. In the picture, the G quartet (G4) and its mutant (M_G4) are analysed in their native (ann.) and boiled (boiled) form. On the left side of the gel are indicated the G quartets (G) while on the right side, the adenines (A) substituted to a guanine (G) are indicated with an arrow. In the G4, the guanines are protected from reaction with DMS because of the folding of the quartet while in the mutated form, one G in the quartet is substituted with an A which does not allow the folding and therefore G reacts with DMS. Also, the boiled forms, confirm it.

Figure A3. Different batches of EN300-02473 analysed in this work. 1. Original stock compound used as a reference for the helicase assay; light brown colour in DMSO solution; 2. AF1756, synthesized at EDELRIS through thiourea route as a beige powder and nearly colourless in DMSO solution. The same aspect is observed for AF1757 and AF1758, respectively synthesise through N-boc-thiourea and Suzuki route. 3. AF1759, compound purchase from ENAMINE by Edelris, presented as brown powder turning dark brown in DMSO solution; 4. AF1770; Purchased from ENAMINE by Edelris and purified by HPLC. The compound shows a light beige colour when powder and nearly colourless when dissolved in DMSO.

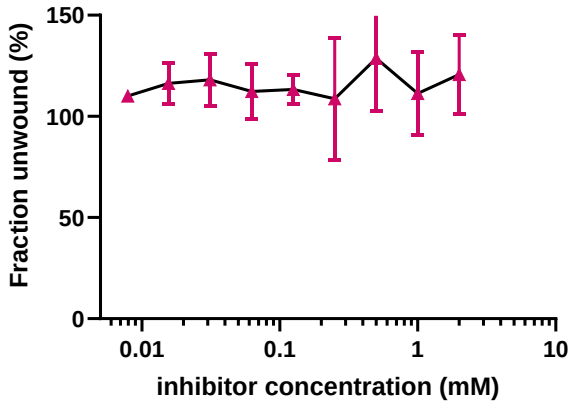


Figure 4A. IC_{50} determination of Tideglusib (TD). Data plot reported in the graph shows the attempt to determine an IC_{50} for TD, from 2 mM in a doubling dilution series (n=3 repeats), 10 nM FL hPIF1, 0.1 nM substrate. This drug was suggested to be a potent inhibitor on hPIF1 [176].

6 REFERENCES

[1] Zhou J-Q.; Monson E. K.; Ten S-C.; Schulz V. P.; Zakiant V. A. (2000) Pif1p helicase, a catalytic inhibitor of telomerase in yeast, *Science 289, 771-774.*

[2] Brosh R. M. Jr.; Matson S. W. (2020) History of DNA helicases, *Genes (Basel) 11,255.*

[3] Duan X. L.; Liu N. N., Yang Y. T.; Li H. H.; Li M.; Dou S. X.; Xi X. G. (2015) G-quadruplexes significantly stimulate Pif1 helicase-catalyzed duplex DNA unwinding, *Journal of Biological Chemistry 290, 7722-7735.*

[4] Tuteja N.; Tuteja R.; Rahman K.; Kang L-Y.; Falaschi A. (1990) A DNA helicase from human cells, *Nucleic Acids Research 18, 6785-6792*

[5] Singleton M. R.; Dillingham M. S.; Wigley D. B. (2007) Structure and mechanism of helicases and nucleic acid translocases, *Annual Review of Biochemistry 76*, *23-50.*

[6] Raney K. D.; Byrd A. K.; Aarattuthodiyil S. (2013) Structure and mechanisms of SF1 DNA helicases, *Advances in Experimental Medicine and Biology 767*, *17-46.*

[7] Barranco-Medina S.; Galletto R. (2010) DNA binding induces dimerization of *Saccharomyces cerevisiae* Pif1, *Biochemistry 49*, *8445-8454.*

[8] Pellegrini L..; Yu D. S.; Lo T.; Anand S.; Lee MY; Blundell T. L.; Venkitaraman A. R. (2002) Insights into DNA recombination from the structure of a RAD51–BRCA2 complex, *Nature 420, 287-293.*

[9] Fairman-Williams M. E.; Guenther U. P.; Jankowsky E. (2010) SF1 and SF2 helicases: family matters, *Current Opinion in Structural Biology 20*, *313-324.*

[10] Sedman J.; Steniund A. (1996) The initiator protein El binds to the bovine papillomavirus origin of replication as a trimeric ring-like structure, *The EMBO Journal 15, 5085-5092.*

[11] Kobori J. A.; Kornber A. (1982) The *Escherichia coli* DnaC gene product. Properties of the DnaB-DnaC protein complex, *The Journal Of Biological Chemistry 257, 13770-13775*.

[12] Javed A.; Major B.; Stead J. A.; Sanders C. M.; Orlova E. V. (2021) Unwinding of a DNA replication fork by a hexameric viral helicase, *Nature Communication 12*, *5535.*

[13] Bochman M. L.; Schwacha A. (2009) The MCM complex: unwinding the mechanism of a replicative helicase, *Microbiology and Molecular Biology Reviews 73*, *652-683.*

[14] Miller T. C. R.; Locke J.; Greiwe J. F.; Diffley J. F. X.; Costa A. (2019) Mechanism of head-to-head MCM double-hexamer formation revealed by cryo-EM, *Nature 575*, *704-710.*

[15] Saikrishnan K.; Powell B.; Cook N. J.; Webb M. R.; Wigley D. B. (2009) Mechanistic basis of 5'-3' translocation in SF1B helicases, *Cell 137*, *849-859.*

[16] Meir A.; Greene E. C. (2021) Srs2 and Pif1 as model systems for understanding SF1A and SF1B helicase structure and function, *Genes (Basel) 12*, *1319.*

[17] Spies M. (2013) DNA helicases and DNA motor proteins, *Advance in experimental Medicine and Biology, 767, 97-112.*

[18] Manosas M.; Xi X. G.; Bensimon D.; Croquette V. (2010) Active and passive mechanisms of helicases, *Nucleic Acids Research 38*, *5518-5526.*

[19] Bhowmick R.; Hickson I. D., Liu, Y. (2023) Completing genome replication outside of S-phase, *Molecular Cell 83*, *3596-3607.*

[20] Saxena S.; Zou L. (2022) Hallmarks of DNA replication stress, *Molecular Cell 82*, *2298-2314.*

[21] Bester A. C.; Roniger M.; Oren Y. S.; Im M. M.; Sarni D.; Chaoat M.; Bensimon A.; Zamir G.; Shewach D. S.; Kerem B. (2011) Nucleotide deficiency promotes genomic instability in early stages of cancer development, *Cell 145*, *435-446.*

[22] Chatterjee N.; Walker G. C. (2017) Mechanisms of DNA damage, repair, and mutagenesis, *Environmental and Molecular Mutagenesis 58*, *235-263.*

[23] Lliakis G. (1991) The role of DNA double strand breaks in ionizing radiation-induced killing of eukaryotic cells, *BioEssays: news and reviews in molecular, cellular and developmental biology 13*, *641-648.*

[24] He X.; Byrd A. K.; Yun M. K.; Pemble C. W.; Harrison D.; Yeruva L.; Dahl C.; Kreuzer K. N.; Raney K. D.; White S. W. (2012) The T4 phage SF1B helicase Dda is structurally optimized to perform DNA strand separation, *Structure 20*, *1189-1200.*

[25] Smela M. E.; Currier S. S., Bailey E. A.; Essigmaann J. M. (2001) The chemistry and Biology of aflatoxin B_1: from mutational spectrometry to carcinogenesis, *Carcinogenesis 22*, *535-545.*

[26] Tsegay P. S.; Lai Y.; Liu Y. (2019) Replication stress and consequential instability of the genome and epigenome, *Molecules 24*, *3870.*

[27] Cheng B.; Pan W.; Xing Y.; Xiao Y.; Chen J.; Xu, Z. (2022) Recent advances in DDR (DNA damage response) inhibitors for cancer therapy, *European Journal of Medicinal Chemistry 230*, *114109.*

[28] Alhmoud J. F.; Woolley J. F.; Al Moustafa A. E.; Malki M. I. (2020) DNA damage/repair management in cancers, *Cancers (Basel) 12*, *1050.*

[29] Kuper J., and Kisker C. (2013) DNA Helicases in NER, BER, and MMR, *Advances in Experimental Medicine and Biology 767*, *203-224.*

[30] Black J. O. (2016) Xeroderma Pigmentosum, *Head and Neck Pathology Vol. 10*, 139-144.

[31] Theil A. F.; Nonnekens J.; Wijgers N.; Vermeulen W.; Giglia-Mari G. (2011) Slowly progressing nucleotide excision repair in trichothiodystrophy group A patient fibroblasts, *Molecular and Cellular Biology 31*, *3630-3638.*

[32] Huselid E.; Bunting S. F. (2020) The regulation of homologous recombination by helicases, *Genes (Basel) 11*, *498.*

[33] Cervelli T.; Borghini A.; Galli A.; Andreassi, M. G. (2012) DNA damage and repair in atherosclerosis: current insights and future perspectives, *International Journal of Molecular Sciences 13*, *16929-16944.*

[34] Gohil D.; Sarker A. H.; Roy R. (2023) Base Excision Repair: mechanisms and impact in biology, disease, and medicine, *International Journal of Molecular Science 24, 14186.*

[35] Gaillard H.; Garcia-Muse T.; Aguilera A. (2015) Replication stress and cancer, *Nature Reviews Cancer 15*, *276-289.*

[36] Davies S. L.; North P. S.; Hickson I. D. (2007) Role for BLM in replication-fork restart and suppression of origin firing after replicative stress, *Nature Structural & Molecular Biology 14*, *677-679.*

[37] Davies S. L.; North P. S.; Hickson I. D. (2007) Role for BLM in replication-fork restart and suppression of origin firing after replicative stress, *Nature Structural & Molecular Biology 14, 677-679.*

[38] Popuri V.; Croteau D. L.; Brosh R. M. Jr.; Bohr V. A. (2012) RECQ1 is required for cellular resistance to replication stress and catalyzes strand exchange on stalled replication fork structures, *Cell Cycle 11*, *4252-4265.*

[39] Thangavel S.; Mendoza-Maldonado R.; Tissino E.; Sidorova J. M.; Yin J.; Wang W.; Monnat R. J. Jr.; Falaschi A.; Vindigni, A. (2010) Human RECQ1 and RECQ4 helicases

play distinct roles in DNA replication initiation, *Molecular and Cellular Biology 30*, *1382-1396.*

[40] Berti M.; Ray Chaudhuri A.; Thangavel S.; Gomathinayagam S.; Kenig S.; Vujanovic M.; Odreman F.; Glatter T.; Graziano S.; Mendoza-Maldonado R.; Marino F.; Lucic B.; Biasin V.; Gstaiger M.; Aebersold R.; Sidorova J. M.; Monnat R. J. Jr.; Lopes M.; Vindigni A. (2013) Human RECQ1 promotes restart of replication forks reversed by DNA topoisomerase I inhibition, *Nature Structural & Molecular Biology 20*, *347-354.*

[41] Javle M.; Curtin N. J. (2011) The role of PARP in DNA repair and its therapeutic exploitation, *British Journal of Cancer 105*, *1114-1122.*

[42] Atkinson J.; McGlynn P. (2009) Replication fork reversal and the maintenance of genome stability, *Nucleic Acids Research 37*, *3475-3492.*

[43] Wu L.; Hickson, I. D. (2006) DNA helicases required for homologous recombination and repair of damaged replication forks, *Annual Review of Genetics 40*, *279-306.*

[44] Ralf C.; Hickson I. D.; Wu, L. (2006) The Bloom's syndrome helicase can promote the regression of a model replication fork, *Journal of Biological Chemistry 281*, *22839-22846.*

[45] Neelsen, K. J.; Lopes M. (2015) Replication fork reversal in eukaryotes: from dead end to dynamic response, *Nature Reviews Molecular Cell Biology 16*, *207-220.*

[46] van Schie J. J. M.; Faramarz A.; Balk J. A.; Stewart G. S.; Cantelli E.; Oostra A. B.; Rooimans M. A.; Parish J. L.; de Almeida Esteves C.; Dumic K.; Barisic I.; Diderich K. E. M.; van Slegtenhorst M. A.; Mahtab M.; Pisani F. M.; Te Riele H.; Ameziane N.; Wolthuis R. M. F.; de Lange J. (2020) Warsaw Breakage Syndrome associated DDX11 helicase resolves G-quadruplex structures to support sister chromatid cohesion, *Nature Communication 11*, *4287.*

[47] Santos D.; Mahtab M.; Boavida A.; Pisani F. M. (2021) Role of the DDX11 DNA helicase in Warsaw Breakage Syndrome etiology, *International Journal of Molecular Sciences 22, 2308*.

[48] Cali F.; Bharti S. K.; Di Perna R.; Brosh R. M. Jr.; Pisani, F. M. (2016) Tim/Timeless, a member of the replication fork protection complex, operates with the Warsaw Breakage Syndrome DNA helicase DDX11 in the same fork recovery pathway, *Nucleic Acids Research 44*, *705-717.*

[49] Patel D. R.; Weiss, R. S. (2018) A tough row to hoe: when replication forks encounter DNA damage, *Biochemical Society Transactions 46*, *1643-1651.*

[50] Georgakopoulos-Soares I.; Morganella S.; Jain N.; Hemberg M.; Nik-Zainal S. (2018) Noncanonical secondary structures arising from non-B DNA motifs are determinants of mutagenesis, *Genome Research 28*, *1264-1271.*

[51] Du Y.; Zhou X. (2013) Targeting non-B-form DNA in living cells, The *Chemical Record 13*, *371-384.*

[52] Wang G.; Vasquez K. M. (2023) Dynamic alternative DNA structures in biology and disease, *Nature Review of Genetics 24*, *211-234.*

[53] Wang G.; Vasquez, K. M. (2014) Impact of alternative DNA structures on DNA damage, DNA repair, and genetic instability, *DNA Repair (Amst) 19*, *143-151.*

[54] Ravichandran S.; Subramani V. K.; Kim, K. K. (2019) Z-DNA in the genome: from structure to disease, *Biophysical Reviews 11*, *383-387.*

[55] Chan P. P.; Glazer P. M. (1997) Triplex DNA: fundamentals, advances, and potential applications for gene therapy, *Journal of Molecular Medicine 75, 267-282*.

[56] Lee J. E.; Taegon K.; Soo Yong K.; (2010) Detection of triple helix DNA formation of guanine-rich oligonucleotide in sodium ion abundant buffer by cross-checking FRET scheme, *Chemical Physics Letters 490*, *230-233.*

[57] Krasilnikova M. M.; Mirkin, S. M. (2004) Replication stalling at Friedreich's ataxia (GAA)n repeats in vivo, *Molecular and Cellular Biology 24*, *2286-2295.*

[58] Mendoza O.; Bourdoncle A.; Boule J. B; Brosh R. M. Jr.; Mergny, J. L. (2016) G-quadruplexes and helicases, *Nucleic Acids Research 44*, *1989-2006.*

[59] Abou Assi H.; Garavis M.; Gonzalez C.; Damha M. J. (2018) i-motif DNA: structural features and significance to cell biology, *Nucleic Acids Research 46*, *8038-8056.*

[60] Spiegel J.; Adhikari S.; Balasubramanian S. (2020) The structure and function of DNA G-quadruplexes, *Trends in Chemistry 2*, *123-136.*

[61] Phan A. T.; Kuryavyi V.; Patel D. J. (2006) DNA architecture: from G to Z, *Current Opinion in Structural Biology 16*, *288-298.*

[62] Enemark E. J.; Joshua-Tor L. (2006) Mechanism of DNA translocation in a replicative hexameric helicase, *Nature 442*, *270-275.*

[63] Risitano A.; Fox K. R. (2004) Influence of loop size on the stability of intramolecular DNA quadruplexes, *Nucleic Acids Research 32*, *2598-2606.*

[64] Hazel mP.; Huppet J; Balasubramanian S; Neidle S. (2004) Loop-Length-Dependent Folding of G-Quadruplexes, *JACS* 126, *16405-16415.*

[65] Rosu F.; Gabelica V.; Poncelet H.; De Pauw E. (2010) Tetramolecular G-quadruplex formation pathways studied by electrospray mass spectrometry, *Nucleic Acids Research 38*, *5217-5225.*

[66] Sen D.; Gilbert W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis, *Nature 334*, *364-366.*

[67] Adrian M.; Heddi B.; Phan A. T. (2012) NMR spectroscopy of G-quadruplexes, *Methods 57*, *11-24.*

[68] Lin, C. D., J.; Yang, D. (2019) NMR studies of G-quadruplex structures and G-quadruplex-interactive compounds, *Methods in Molecular Biology 2035*, *157-176.*

[69] Day H. A.; Pavlou P.; Waller Z. A. (2014) i-motif DNA: structure, stability and targeting with ligands, *Bioorganic & Medicinal Chemistry 22*, *4407-4418.*

[70] Ying Z. M.; Tu B.; Liu L.; Tang H.; Tang L. J.; Jiang, J. H. (2018) Spinach-based fluorescent light-up biosensors for multiplexed and label-free detection of microRNAs, *Chemical Communication 54*, *3010-3013.*

[71] Nishio M.; Tsukakoshu K.; Ikebukuro K. (2021) G-quadruplex: flexible conformational changes by cations, pH, crowding and its applications to biosensing, *Biosensor and Bioelectronics 178, 113030*.

[72] Miyoshi D.; Karimata H.; Sugimoto N. (2005) Drastic effect of a single base difference between human and tetrahymena telomere sequences on their structures under molecular crowding conditions, *Angewandte Chemie International Edition England 44*, *3740-3744.*

[73] Miyoshi D.; Nakao A.; Sugimoto N. (2002) Molecular crowding regulates the structural switch of the DNA G-quadruplex, *Biochemistry 41,15017-15024*.

[74] Kan Z-Y.; Yao Y.; Wang P.; Li X. H.; Hao Y. H.; Tan Z. (2006) Molecular crowding induces telomere G-quadruplex formation under salt-deficient conditions and enhances its competition with duplex formation, *Angewandte Chemie International Edition England 45*, *1629-1632.*

[75] Balagurumoorthy P.; Brahmachari S. K. (1994) Structure and stability of human telomeric sequence, *Journal of Biological Chemistry 269*, *21858-21869.*
[76] Balaratnam S.; Basu S. (2015) Divalent cation-aided identification of physico-chemical properties of metal ions that stabilize RNA G-quadruplexes, *Biopolymers 103*, *376-386.*
[77] Bhattacharyya D.; Mirihana Arachchilage G.; Basu S. (2016) Metal cations in G-quadruplex folding and stability, *Frontieris in Chemistry 4*, *38.*
[78] Smith F.; Feigon J. (1992) Quadruplex structure of *Oxytricha* telomeric DNA oligonucleotides, *Nature 356*, *164–168.*
[79] Neidle, S. (2016) Quadruplex nucleic acids as novel therapeutic targets, *Journal of Medicinal Chemistry 59, 5987-6011.*
[80] Frelih T.; Wang B.; Plavec J.; Sket P. (2020) Pre-folded structures govern folding pathways of human telomeric G-quadruplexes, *Nucleic Acids Research 48*, *2189-2197.*
[81] Hou X. M.; Fu Y. B.; Wu W. Q.; Wang L.; Teng F. Y.; Xie P.; Wang P. Y.; Xi X. G. (2017) Involvement of G-triplex and G-hairpin in the multi-pathway folding of human telomeric G-quadruplex, *Nucleic Acids Research 45*, *11401-11412.*
[82] Grun J. T.; Schwalbe H. (2022) Folding dynamics of polymorphic G-quadruplex structures, *Biopolymers 113 e23477*.
[83] Maizels N. (2012) G4 motifs in human genes, *Annual of NY Academy of Sciences 1267*, *53-60*.
[84] Mohaghegh P.; Karow J. K.; Brosh J. R. M. Jr; Bohr V. A.; Hickson I. D. (2001) The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases, *Nucleic Acids Research 29, 2843-2849*.
[85] Schulz V. P.; Zakian V. A.; Ogburn C. E.; McKay J.; Jarzebowicz M. A.; Edland S. D.; Martin G. M. (1996) Accelerated loss of telomeric repeats may not explain accelerated replicative decline of Werner Syndrome cells, *Human Genetics 97, 750-754*.
[86] Lipps H. J.; Rhodes D. (2009) G-quadruplex structures: *in vivo* evidence and function, *Trends in Cellular Biology 19*, *414-422.*
[87] Lohman T. M.; Tomko E. J.; Wu C. G. (2008) Non-hexameric DNA helicases and translocases: mechanisms and regulation, *Nature Review Molecular Cell Biology 9*, *391-401.*
[88] Lerner L. K.; Sale J. E. (2019) Replication of G-quadruplex DNA, *Genes (Basel) 10, 95*.
[89] Dahan D.; Tsirkas I.; Dovrat D.; Sparks M. A.; Singh S. P.; Galletto R.; Aharoni A. (2018) Pif1 is essential for efficient replisome progression through lagging strand G-quadruplex DNA secondary structures, *Nucleic Acids Research 46*, *11847-11857.*
[90] Sato K.; Martin-Pintado N.; Post H.; Altelaar M.; Knipschee P. (2021) Multistep mechanism of G-quadruplex resolution during DNA replication, *Science Advances 7, eabf8653*.
[91] Drosopoulos W. C.; Kosiyatrakul S. T.; Schildkraut C. L. (2015) BLM helicase facilitates telomere replication during leading strand synthesis of telomeres, *Journal of Cell Biology 210*, *191-208.*
[92] Clynes D.; Jelinska C.; Xella B.; Ayyub H.; Scott C.; Mitson M.; Taylor S.; Higgs D. R.; Gibbons R. J. (2015) Suppression of the alternative lengthening of telomere pathway by the chromatin remodelling factor ATRX, *Nature Communication 6*, *7538.*
[93] Paeschke K.; Bochman M. L.; Garcia P. D.; Cejka P.; Friedman K. L.; Kowalczykowski S. C.; Zakian V. A. (2013) Pif1 family helicases suppress genome instability at G-quadruplex motifs, *Nature 497*, *458-462.*

[94] Lin W.; Sampathi S.; Dai H.; Liu C.; Zhou M.; Hu J.; Huang Q.; Campbell J.; Shin-Ya K.; Zheng L.; Chai W.; Shen B. (2013) Mammalian DNA2 helicase/nuclease cleaves G-quadruplex DNA and is required for telomere integrity, *EMBO Journal 32*, *1425-1439*.

[95] Cheng K.; Wigley D. B. (2018) DNA translocation mechanism of an XPD family helicase, *Elife 7*, *e42400.*

[96] Cheung I.; Schertzer M.; Rose A.; Lansdorp P. M. (2002) Disruption of dog-1 in Caenorhabditis elegans triggers deletions upstream of guanine-rich DNA, *Nature Genetics 31*, *405-409.*

[97] Huber M. D.; Duquette M. L.; Shiels J. C.; Maizels N. (2006) A conserved G4 DNA binding domain in RecQ family helicases, *Journal of Molecular Biology 358*, *1071-1080.*

[98] Lee J. W.; Kusumoto R.; Doherty K. M.; Lin G. X.; Zeng W.; Cheng W. H.; von Kobbe C.; Brosh R. M. Jr.; Hu J. S.; Bohr V. A. (2005) Modulation of Werner Syndrome protein function by a single mutation in the conserved RecQ domain, *Journal of Biological Chemistry 280*, *39627-39636.*

[99] Voter A. F.; Qiu Y.; Tippana R.; Myong S.; Keck J. L. (2018) A guanine-flipping and sequestration mechanism for G-quadruplex unwinding by RecQ helicases, *Nature Communication 9*, *4201.*

[100] Bourgeois C. F.; Mortreux F.; Auboeuf D. (2016) The multiple functions of RNA helicases as drivers and regulators of gene expression, *Nature Reviews Molecular Cell Biology 17, 426-438.*

[101] Booy E. P.; McRae E. K.; Howard R.; Deo S. R.; Ariyo E. O.; Dzananovic E.M.; Meier M.; Stetefeld J.; McKenna S. A. (2016) RNA helicase associated with AU-rich element (RHAU/DHX36) interacts with the 3'-tail of the long non-coding RNA BC200 (BCYRN1), *Journal of Biological Chemistry 291*, *5355-5372*.

[102] Booy E. P.; Meier M.; Okun N.; Novakowski S. K.; Xiong S.; Stetefeld J.; McKenna S. A. (2011) The RNA helicase RHAU (DHX36) unwinds a G4-quadruplex in human telomerase RNA and promotes the formation of the P1 helix template boundary. *Nucleic Acid Research 40, 4110-4124.*

[103] Tippana R.; Chen M. C.; Demeshkina N. A.; Ferre-D'Amare, A. R.; Myong, S. (2019) RNA G-quadruplex is resolved by repetitive and ATP-dependent mechanism of DHX36, *Nature Communication 10*, *1855.*

[104] Heddi B.; Cheong V. V.; Martadinata H.; Phan A. T. (2015) Insights into G-quadruplex specific recognition by the DEAH-box helicase RHAU: Solution structure of a peptide-quadruplex complex, *PNAS USA 112*, *9608-9613.*

[105] Heddi B; Cheong V. V.; Schmitt E.; Mechulam Y.; Phan A. T. (2020) Recognition of different base tetrads by RHAU (DHX36): X-ray crystal structure of the G4 recognition motif bound to the 3'-end tetrad of a DNA G-quadruplex, *Journal of Structural Biology 209*, *107399.*

[106] Chen M. C.; Tippana R.; Demeshkina N. A.; Murat P.; Balasubramanian S.; Myong S.; Ferre-D'Amare A. R. (2018) Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36, *Nature 558*, *465-469.*

[107] Balasubramanian S.; Hurley L. H.; Neidle, S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy?, *Nature Review Drug Discovery 10*, *261-275.*

[108] Liang Y.; Lin S. Y.; Brunicardi F. C.; Goss J.; Li K. (2009) DNA damage response pathways in tumor suppression and cancer treatment, *World Journal of Surgery 33*, *661-666.*

[109] Curtin N. J. (2023) Targeting the DNA damage response for cancer therapy, *Biochemical Society Transactions 51*, *207-221*.

[110] Minchom A.; Aversa C.; Lopez J. (2018) Dancing with the DNA damage response: next-generation anti-cancer therapeutic strategies, *Therapeutic Advances in Medical Oncology 10*.

[111] Ding J.; Miao Z. H.; Meng L. H.; Geng M. Y. (2006) Emerging cancer therapeutic opportunities target DNA-repair systems, *Trends in Pharmacology Science 27*, *338-344.*

[112] Huang A.; Garraway L. A.; Ashworth A.; Weber B. (2020) Synthetic lethality as an engine for cancer drug target discovery, *Nature Review Drug Discovery 19*, *23-38.*

[113] Setton J.; Zinda M.; Riaz N.; Durocher D.; Zimmermann M.; Koehler M.; Reis-Filho J. S.; Powell S. N. (2021) Synthetic lethality in cancer therapeutics: the next generation, *Cancer Discovery 11*, *1626-1635.*

[114] Sparks M. A.; Singh S. P.; Burgers P. M.; Galletto, R. (2019) Complementary roles of Pif1 helicase and single stranded DNA binding proteins in stimulating DNA replication through G-quadruplexes, *Nucleic Acids Research 47*, *8595-8605.*

[115] Gavande N. S.; VanderVere-Carozza P. S.; Hinshaw H. D.; Jalal S. I.; Sears C. R.; Pawelczak K. S.; Turchi J. J. (2016) DNA repair targeted therapy: the past or future of cancer treatment?, *Pharmacology & Therapeutics 160*, *65-83.*

[116] Foury F.; Kolodynski J. (1983) pif mutation blocks recombination between mitochondrial p+ and p genomes having tandemly arrayed repeat units in *Saccharomyces cerevisiae*, *PNAS USA 80, 5345-5349.*

[117] Foury F.; Lahaye A. (1987) Cloning and sequencing of the PIF gene involved in repair and recombination of yeast mitochondrial DNA, *The EMBO Journal 6*, *1441 -1449.*

[118] Lahaye A.; Leterma S.; Foury F (1993) PIF1 DNA helicase from *Saccharomyces cerevisiae*. Biochemical characterization of the enzyme, *The Journal of Biological Chemistry 268, 26155-26161*.

[119] Schulz V. P.; Zakian V. A. (1994) The *Saccharomyces* PIF1 DNA helicase inhibits telomere elongation and *de novo* telomere formation, *Cell 76*, *145-155.*

[120] Keil R. L.; McWilliam A. D. (1993) A Gene with specific and global effects on recombination of sequences from tandemly repeated genes in *Saccharomyces cerevisiae, Genetics 135, 711-718.*

[121] Ivessa A. S.; Zhou J-Q; 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.*

[122] Budd M. E.; Reis C. C.; Smith S.; Myung K.; Campbell J. L. (2006) Evidence suggesting that Pif1 helicase functions in DNA replication with the Dna2 helicase/nuclease and DNA polymerase delta, *Molecular and Cellular Biology 26*, *2490-2500.*

[123] Ribeyre C.; Lopes J.; Boule J. B.; Piazza A.; Guedin A.; Zakian V. A.; Mergny J. L.; Nicolas A. (2009) The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo, *PLoS Genetics 5*, *e1000475.*

[124] Deegan T. D.; Baxter J.; Ortiz Bazan M. A.; Yeeles J. T. P.; Labib K. P. M. (2019) Pif1-family helicases support fork convergence during DNA replication termination in eukaryotes, *Molecular Cell 74*, *231-244.*

[125] Tanaka H.; Ryu G. H.; Seo Y. S.; Tanaka K.; Okayama H.; MacNeill S. A. 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*.

[126] Bochman M. L.; Judge C. P.; Zakian V. A. (2011) The Pif1 family in prokaryotes: what are our helicases doing in your bacteria?, *Molecular Biology of the Cell 22*, *1955-1959*.

[127] Liu N. N.; Duan X. L.; Ai X.; Yang Y. T.; Li M.; Dou S. X.; Rety S.; Deprez E.; Xi,X. G. (2015) The *Bacteroides sp. 3_1_23* Pif1 protein is a multifunctional helicase, *Nucleic Acids Research 43*, *8942-8954.*

[128] Gagou M. E.; Ganesh A.; Phear G.; Robinson D.; Peterman E.; Cox A.; Meuth M. (2014) Human PIF1 helicase supports DNA replication and cell growth under oncogenic-stress, *Oncotarget 5, 11381-11398*.

[129] Gagou M. E.; Ganesh A.; Thompson R.; Phear G.; Sanders C.; Meuth M. (2011) Suppression of apoptosis by PIF1 helicase in human tumor cells, *Cancer Research 71*, *4998-5008.*

[130] Bochman M. L.; Sabouri N.; Zakian, V. A. (2010) Unwinding the functions of the Pif1 family helicases, *DNA Repair (Amst) 9*, *237-249.*

[131] Saikrishnan K.; Griffiths S. P.; Cook N.; Court R.; Wigley D. B. (2008) DNA binding to RecD: role of the 1B domain in SF1B helicase activity, *EMBO Journal 27*, *2222-2229.*

[132] Zhou X.; Ren W.; Bharath S. R.; Tang X.; He Y.; Chen C.; Liu Z.; Li D.; Song, H. (2016) Structural and functional insights into the unwinding mechanism of *Bacteroides sp* Pif1, *Cell Reports 14*, *2030-2039.*

[133] Geronimo C. L.; Singh S. P.; Galletto R.; 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 ATP*ase* activity, *Nucleic Acids Research 46*, *8357-8370.*

[134] 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.; Xi X. G. (2016) Crystal structures of the *Bs*Pif1 helicase reveal that a major movement of the 2B SH3 domain is required for DNA unwinding, *Nucleic Acids Research 44*, *2949-2961.*

[135] Dehghani-Tafti S., Levdikov V., Antson A. A.; Bax B.; Sanders C. M. (2019) Structural and functional analysis of the nucleotide and DNA binding activities of the human PIF1 helicase, *Nucleic Acids Research 47*, *3208-3222.*

[136] Gu, Y., Masuda, Y., and Kamiya, K. (2008) Biochemical analysis of human PIF1 helicase and functions of its N-terminal domain, *Nucleic Acids Res 36*, *6295-6308.*

[137] Byrd A. K.; Raney K. D. (2017) Structure and function of Pif1 helicase, *Biochemical Society Transaction 45*, *1159-1171*.

[138] Boule J. B.; Vega L. R.; Zakian V. A. (2005) The yeast Pif1p helicase removes telomerase from telomeric DNA, *Nature 438*, *57-61.*

[139] Huber M.; Lee D.; Maizeld N. (2002) G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition, *Nucleic Acids Research 30, 3954-3961*.

[140] Lopes J.; Piazza A.; Bermejo R.; Kriegsman B.; Colosio A.; Teulade-Fichou M. P.; Foiani M.; Nicolas A. (2011) G-quadruplex-induced instability during leading-strand replication, *EMBO Journal 30*, *4033-4046*.

[141] Li J. R.; Yu T. Y.; Chien I. C.; Lu C. Y.; Lin J. J.; Li H. W. (2014) Pif1 regulates telomere length by preferentially removing telomerase from long telomere ends, *Nucleic Acids Research 42*, *8527-8536.*

[142] Liu N. N.; Duan X. L.; Ai X.; Yang Y. T.; Li M.; Dou S. X.; Rety S.; Deprez E.; Xi X. G. (2015) The *Bacteroides sp. 3_1_23* Pif1 protein is a multifunctional helicase, *Nucleic Acids Res 43*, *8942-8954.*

[143] Mateyak M. K.; Zakian V. A. (2006) Human PIF helicase is cell cycle regulated and associates with telomerase, *Cell Cycle 5*, *2796-2804*.

[144] Rossi M. L.; Pike J. E.; Wang W.; Burgers P. M. J.; Campbell J. L.; Bambara R. A. (2008) Pif1 helicase directs eukaryotic Okazaki fragments toward the two-nuclease cleavage pathway for primer removal, *Journal of Biological Chemistry 283*, *27483-27493*.

[145] George T.; Wen Q.; Griffiths R.; Ganesh A.; Meuth M.; Sanders C. M. (2009) Human Pif1 helicase unwinds synthetic DNA structures resembling stalled DNA replication forks, *Nucleic Acids Research 37*, *6491-6502.*

[146] Sanders C. M. (2010) Human Pif1 helicase is a G-quadruplex DNA-binding protein with G-quadruplex DNA-unwinding activity, *Biochemical Journal 430*, *119-128.*

[147] Zhang, D. H.; Zhou B.; Huang Y.; Xu L. X.; Zhou J. Q. (2006) The human Pif1 helicase, a potential *Escherichia coli* RecD homologue, inhibits telomerase activity, *Nucleic Acids Research 34*, *1393-1404.*

[148] Su N.; Byrd A. K.; Bharath S. R.; Yang O.; Jia Y.; Tang X.; Ha T.; Raney K. D.; Song H. (2019) Structural basis for DNA unwinding at forked dsDNA by two coordinating Pif1 helicases, *Nature Communication 10*, *5375.*

[149] 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.; Xi X. G. (2018) Insights into the structural and mechanistic basis of multifunctional *S. cerevisiae* Pif1p helicase, *Nucleic Acids Research 46*, *1486-1500.*

[150] Mustafi M.; Kwon Y.; Sung P.; Greene E. C. (2023) Single-molecule visualization of Pif1 helicase translocation on single-stranded DNA, *Journal of Biological Chemistry 299*, *104817.*

[151] Ramanagoudr-Bhojappa R.; Byrd A. K.; Dahl C.; Raney K. D. (2014) Yeast Pif1 accelerates annealing of complementary DNA strands, *Biochemistry 53*, *7659-7669.*

[152] Zhou R.; Zhang J.; Bochman M. L.; Zakian V. A.; Ha T. (2014) Periodic DNA patrolling underlies diverse functions of Pif1 on R-loops and G-rich DNA, *Elife 3*, *e02190.*

[153] Valle-Orero J.; Rieu M.; Tran P. L. T.; Joubert A.; Raj S.; Allemand J. F.; Croquette V.; Boule J. B. (2022) Strand switching mechanism of Pif1 helicase induced by its collision with a G-quadruplex embedded in dsDNA, *Nucleic Acids Research 50*, *8767-8778.*

[154] Zhang B.; Wu W Q.; Liu N. N.; Duan X. L.; Li M.; Dou S. X.; Hou X. M.; Xi X. G. (2016) G-quadruplex and G-rich sequence stimulate Pif1p-catalyzed downstream duplex DNA unwinding through reducing waiting time at ss/dsDNA junction, *Nucleic Acids Research 44*, *8385-8394.*

[155] Dai Y. X.; Chen W. F.; Liu N. N.; Teng F. Y.; Guo H. L.; Hou X. M.; Dou S. X.; Rety S.; Xi X. G. (2021) Structural and functional studies of SF1B Pif1 from Thermus oshimai reveal dimerization-induced helicase inhibition, *Nucleic Acids Research 49*, *4129-4143.*

[156] Dai Y. X.; Guo H. L.; Liu N. N.; Chen W. F.; Ai X.; Li H. H.; Sun B.; Hou X. M.; Rety S.; Xi X. G. (2022) Structural mechanism underpinning *Thermus oshimai* Pif1-mediated G-quadruplex unfolding, *EMBO Reports 23*, *e53874*.

[157] Mazouzi A.; Velimezi G.; Loizou J. I. (2014) DNA replication stress: causes, resolution and disease, *Experimental Cell Research 329*, *85-93.*

[158] Bartkova J.; Rezaei N.; Liontos M.; Karakaidos P.; Kletsas D.; Issaeva N.; Vassiliou L. V.; Kolettas E.; Niforou K.; Zoumpourlis V. C.; Takaoka M.; Nakagawa H.; Tort F.; Fugger K.; Johansson F.; Sehested M.; Andersen C. L.; Dyrskjot L.; Orntoft T.; Lukas J.; Kittas C.; Helleday T.; Halazonetis T. D.; Bartek J.; Gorgoulis V. G. (2006) Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints, *Nature 444*, *633-637.*

[159] Di Micco R.; Fumagalli M.; Cicalese A.; Piccinin S.; Gasparini P.; Luise C.; Schurra C.; Garre M.; Nuciforo P. G.; Bensimon A.; Maestro R.; Pelicci P. G.; d'Adda di Fagagna F. (2006) Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication, *Nature 444*, *638-642.*

[160] Inoue H.; Nojima H., Okayama H. (1990) High efficiency transformation of *Escherichia coli* with plasmids, *Gene 96, 23-8.*

[161] Bimboim H. C.; Doly J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA, *Nucleic Acids Research 7, 1513-1523.*

[162] Robert X.; P. Gouet (2014) Deciphering key features in protein structures with the new ENDscript server, *Nucleic Acids Research*.

[163] Maxam A. M.; Gilbert W. (1977) A new method for sequencing DNA. *PNAS USA*

[164] Iggo R. D.; Lane D. P. (1989) Nuclear protein p68 is an RNA-dependent ATPase, *EMBO Journal 8*, 1827-1831.

[165] Horvath M. P.; Schweiker V.; Bevilacqua J.; Ruggles J.; Schultz S. (1998) Crystal structure of the *Oxytricha nova* telomere end binding protein complexed with single strand DNA, *Cell Press 42,320-324*.

[166] McKay R. D. G. (1981) Binding of Sivian Virus 40 T antigen-related protein to DNA, *Journal of Molecular Biology 145*, *471-488.*

[167] Korolev N. (2021) How potassium came to be the dominant biological cation: of metabolism, chemiosmosis, and cation selectivity since the beginnings of life, *BioEssays: news and reviews in molecular, cellular and developmental biology 43*, *e2000108.*

[168] Malnic G; Muto S.; Giebisch G. (2008) Regulation of potassium excretion, In *The Kidney* (Press, A., Ed.), *1301-1347.*

[169] Rastegar A. (1990) Serum potassium*, Vol. Clinical methods: The history, physical, and laboratory examinations, 3rd edition.*

[170] Dai Y-X; Guo H-L.; Liu N-N.; Chen W-F; Ai X.; Li H-H.; Sun B.; Hou X-M; Rety S.; Xi X-G. (2021) Structural mechanism underpinning *Thermus oshimai* Pif1-mediated G-quadruplex unfolding, *EMBO Journal 23, e53874*.

[171] Wang W.; Hu S.; Gu Y.; Yan Y.; Stovall D. B.; Li D.; Sui G. (2020) Human MYC G-quadruplex: from discovery to a cancer therapeutic target, *Biochimica et Biophysica Acta Reviews on Cancer 1874*, *188410.*

[172] Hazel P.; Huppert J.; Balasubramanian S.; Neidle S. (2004) Loop-length-dependent folding of G-Quadruplexes, *JACS 126, 16405-16415*.

[173] Ghosh A.; Trajkovski M.; Teulade-Fichou M. P.; Gabelica V.; Plavec J. (2022) Phen-DC(3) induces refolding of human telomeric DNA into a chair-type antiparallel G-quadruplex through ligand intercalation, *Angewandte Chemie International Edition England 61*, *e202207384.*

[174] De Cian A.; DeLemons E.; Mergny J.-L.; Teulade-Fichou M.-P.; Monchaud D. (2007) Highly efficient G-quadruplex recognition by bisquinolinium compounds, *JACS129, 1856-1857.*

[175] Piazza A.; Boule J. B.; Lopes J.; Mingo K.; Largy E.; Teulade-Fichou M. P.; Nicolas A. (2010) Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in *Saccharomyces cerevisiae*, *Nucleic Acids Research 38*, *4337-4348.*

[176] Paeschke K.; Capra J. A.; Zakian V. A. (2011) DNA replication through G-quadruplex motifs is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase, *Cell 145*, *678-691.*

[177] Lipinski C. A.; Lombardo F.; Dominy B. W.; Feeney P. J. (1996) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advance Drug Delivery Reviews 46, 3-26.*

[178] Byun T. S.; Pacek M.; Yee M. C.; Walter J. C.; Cimprich K. A. (2005) Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint, *Genes and Development 19*, *1040-1052.*

[179] Sallmyr A.; Tomkinson A. E.; Rassool F. V. (2008) Up-regulation of WRN and DNA ligase IIIalpha in chronic myeloid leukemia: consequences for the repair of DNA double-strand breaks, *Blood 112*, *1413-1423.*

[180] Rosenthal A. S.; Dexheimer T. S.; Gileadi O.; Nguyen G. H.; Chu W. K.; Hickson I. D.; Jadhav A.; Simeonov A.; Maloney D. J. (2013) Synthesis and SAR studies of 5-(pyridin-4-yl)-1,3,4-thiadiazol-2-amine derivatives as potent inhibitors of Bloom helicase, *Bioorganic and Medicinal Chemistry Letters 23*, *5660-5666.*

[181] Moles R.; Bai X. T.; Chaib-Mezrag H.; Nicot C. (2016) WRN-targeted therapy using inhibitors NSC 19630 and NSC 617145 induce apoptosis in HTLV-1-transformed adult T-cell leukemia cells, *Journal of Hematology & Oncology 9*, *121.*

[182] Datta A.; Brosh R. M. Jr. (2018) New Insights Into DNA Helicases as Druggable Targets for Cancer Therapy, *Frontieris in Molecular Biosciences 5*, *59.*

[183] Snow B. E.; Mateyak M.; Paderova J.; Wakeham A.; Iorio C.; Zakian V.; Squire J.; Harrington L. (2007) Murine Pif1 interacts with telomerase and is dispensable for telomere function in vivo, *Molecular and Cellular Biology 27*, *1017-1026.*

[184] Zhou X.; Pan Y.; Qu Y.; Ke X. (2022) Tideglusib inhibits Pif1 helicase of *Bacteroides sp*. via an irreversible and cys-380-dependent mechanism, *ACS Omega 7*, *31289-31298.*

[185] Bardin C.; Leroy J. L. (2008) The formation pathway of tetramolecular G-quadruplexes, *Nucleic Acids Research 36*, *477-488.*

[186] Zacchia M.; Abategiovanni M. L.; Stratigis S.; Capasso G. (2016) Potassium: from physiology to clinical implications, *Kidney Disease (Basel) 2*, *72-79.*

[187] Hershman S. G.; Chen Q.; Lee J. Y.; Kozak M. L.; Yue P.; Wang L. S.; Johnson F. B. (2008) Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*, *Nucleic Acids Research 36*, *144-156*.

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

[189] Zhou J.-Q.; Monson E. K.; Teng S.-C.; Schulz V. P. (2000) Pif1p helicase, a catalytic inhibitor of telomerase in yeast, *Science 289, 771-774*.

[190] Jin Y.; Molt R. W. Jr.; Blackburn G. M. (2017) Metal fluorides: tools for structural and computational analysis of phosphoryl transfer enzymes, *Topics in Current Chemistry 375*, *36.*

[191] Prasad B.; Doimo M.; Andreasson M.; L'Hote V.; Chorell E.; Wanrooij S. (2022) A complementary chemical probe approach towards customized studies of G-quadruplex DNA structures in live cells, *Chemical Science Journal 13*, *2347-2354.*

[192] Reynolds K. A.; Cameron C. E.; Raney, K. D. (2015) Melting of duplex DNA in the absence of ATP by the NS3 helicase domain through specific interaction with a single-strand/double-strand junction, *Biochemistry 54*, *4248-4258.*

[193] Budhathoki J. B.; Ray S.; Urban V.; Janscak P.; Yodh J. G.; Balci H. (2014) RecQ-core of BLM unfolds telomeric G-quadruplex in the absence of ATP, *Nucleic Acids Research 42*, *11528-11545.*

[194] Tippana R.; Hwang H.; Opresko P. L.; Bohr V. A.; Myong S. (2016) Single-molecule imaging reveals a common mechanism shared by G-quadruplex-resolving helicases, *PNAS USA 113*, *8448-8453.*