The WYL domain of the *Thermotoga elfii* PIF1 helicase is an accessory single-stranded DNA binding module

Nicholas M. Andis[1], Christopher W. Sausen[1], Ashna Alladin[1,2], and Matthew L. Bochman[1]

[1] Molecular and Cellular Biochemistry Department, Indiana University, Bloomington, Indiana 47405
[2] Present address: EMBL Heidelberg, Meyerhofstrasse 1, 69117 Heidelberg, Germany

Running title: The TePif1 WYL domain binds ssDNA

To whom correspondence should be addressed: Matthew L. Bochman, Molecular and Cellular Biochemistry Department, Indiana University, 212 South Hawthorne Drive, Simon Hall MSB1 room 405B, Bloomington, Indiana 47405, Telephone: (812) 856-2095; FAX (812) 856-5710; E-mail: bochman@indiana.edu

Keywords: DNA helicase, ATPase, DNA binding protein, DNA-protein interaction, genomic instability

---

ABSTRACT

In the yeast *Saccharomyces cerevisiae*, the PIF1 family helicases Pif1 and Rrm3 aid in the maintenance of nuclear and mitochondrial genome stability. Despite great progress in understanding the roles of these enzymes as a whole, the functions of the N- and C-terminal domains that flank their central helicase core remain unclear. Here, we characterized the biochemical activities of the *Thermotoga elfii* PIF1 helicase (TePif1), which contains a C-terminal WYL domain. As is typical of helicases from thermophilic organisms, recombinant TePif1 was amenable to over-expression and purification, thermostable *in vitro*, and displayed activities similar to its better-studied eukaryotic homologs. We also found that the WYL domain was necessary for high affinity single-stranded DNA (ssDNA) binding and impacted both ATPase and helicase activities. Our results indicate that TePif1 and other bacterial PIF1 helicases may act as accessory helicases at replication forks to sense ssDNA levels behind the replicative helicase. Further, our findings predict that the domains of unknown function found in TePif1 homologs like *S. cerevisiae* Pif1 and Rrm3 may contain motifs needed for ssDNA binding.

DNA helicases are motor proteins that separate double-stranded DNA (dsDNA) into single-stranded DNA (ssDNA) templates to facilitate cellular processes that are necessary to maintain genome integrity, such as DNA replication, recombination, and repair (1). As such, these enzymes are abundant and conserved throughout evolution, with many being essential for life (2). Indeed, mutations in human DNA helicases are often linked to diseases characterized by genomic instability and a predisposition to cancers (3).

The 5′-3′-directed Superfamily Ib PIF1 helicases exemplify all of these traits (4). Members of this protein family are found in bacteria and eukaryotes (4,5), and in the case of the fission yeast *Schizosaccharomyces pombe*, the single PIF1 helicase Pfh1 is essential for viability (6). There are also multiple lines of evidence suggesting that human PIF1 (hPIF1) may act as a tumor suppressor (7-11), including the fact that mutation of a conserved residue in the PIF1 family signature sequence (4) is linked to breast cancer (12).

The *Saccharomyces cerevisiae* genome encodes two PIF1 family helicases: Pif1 and Rrm3. Recently, *S. cerevisiae* Rrm3 was shown to physically associate with Orc5 (13), a subunit of the hexameric origin recognition complex (Orc1-6) that demarcates all DNA replication origins in eukaryotes. Orc5 binding requires a short motif in Rrm3's large N-terminal domain, which otherwise is of unknown function and is predicted to be natively disordered (4). Rrm3/Orc5 association modulates DNA replication during replication stress and does not require the ATPase/helicase activity of Rrm3. Intriguingly, bioinformatics analyses have also recently suggested that a PIF1

helicase/Orc5 interaction is also important in amoebae (14). It appears that a gene fusion event occurred in *Dictyostelium fasciculatum* between the genes encoding a PIF1 family helicase and the Orc3 subunit, creating a single protein (Pif1-Orc3). Orc3 and Orc5 interact in *S. cerevisiae* (15), so it is tempting to speculate that the *D. fasciculatum* Pif1-Orc3 and Orc5 proteins recapitulate the important Rrm3/Orc5 interaction found in yeast. The amoeboid work also found PIF1 helicases with accessory domains containing motifs implicated in ubiquitination (Ubox and CUE-like motifs) and single-stranded nucleic acid binding (RNase H1-like RNA binding and CCHC zinc finger motifs) (14).

What other accessory domains are found in PIF1 helicases, and how might these domains impact helicase function or be predictive of novel roles for these enzymes as with *D. fasciculatum* Pif1-Orc3 and *S. cerevisiae* Rrm3/Orc5 above? Here, we report that several species of thermophilic bacteria encode PIF1 family helicases that contain C-terminal WYL domains, which adopt Sm-like SH3 β-barrel folds and are predicted ligand binding modules (16). We found that *Thermotoga elfii* Pif1 (TePif1) is thermostable in solution and displays ATPase, DNA binding, and DNA unwinding activities similar to other members of this helicase family. Mutational analysis indicated that the WYL domain of TePif1 inhibits ATPase activity but stimulates helicase activity. Further, TePif1 lacking its WYL domain or containing mutations that neutralize its predicted positively charged ligand binding channel had reduced ssDNA binding affinity. The recombinant WYL domain in isolation bound ssDNA, suggesting that it functions as an accessory ssDNA binding interface in the context of the full-length TePif1. Thus, the uncharacterized C-terminal domains of PIF1 helicases such as hPIF1 may contain cryptic ssDNA binding modules that are important for function.

RESULTS

Generation of recombinant PIF1 helicases from thermophilic bacteria – A major impediment to working with eukaryotic PIF1 helicases is the difficulty in over-expressing and purifying sufficient quantities of recombinant protein for biochemistry. For instance, *S. cerevisiae* Pif1 is the most studied PIF1 family helicase in part because the full-length enzyme can be routinely over-expressed and purified from *Escherichia coli* (4). However, far less is known about *S. cerevisiae* Rrm3 because the full-length protein is exceptionally difficult to purify (17,18). To overcome these issues, we and others have previously turned to PIF1 helicases in bacteria, which have proven to be very tractable for *in vitro* experimentation (19-21). This is similar to the use of homologous proteins from thermophilic archaea to study difficult-to-purify replication factors from eukaryotes (22).

Lacking PIF1 homologs in hyperthermophilic archaea (MLB, observations), we instead sought to characterize PIF1 helicases from thermophilic bacteria in an effort to identify enzymes that are robust and amenable to biochemistry and structural biology. By querying the NCBI Protein database with the *S. cerevisiae* Pif1 sequence, we found that multiple thermophilic bacteria encode PIF1 family helicases, including *T. elfii* (TePif1), *Thermus oshimai* (ToPif1), and *Thermodesulfovibrio yellowstonii* (TyPif1) (Fig. 1A). These proteins contain an N-terminal helicase domain that includes the characteristic PIF1 family signature sequence motif (4,5) (Fig. 1 A and B). Additionally, they all include a more C-terminally located UvrD_C_2 domain, which is a domain that adopts a AAA-like fold and is found near the C-terminus of many helicases (23).

We also found that a variety of PIF1 helicases from thermophilic bacteria contain a WYL domain at their extreme C-terminus (Fig. 1A and B). Such domains are often found in proteins associated with CRISPR-Cas systems, those encoded in SOS DNA repair-associated operons, and a host of other enzymes (reviewed in (16)). In some proteins with tandem WYL domains, it is hypothesized that they may enable oligomerization. Alternatively, based on predicted structural homology with Sm-like SH3 β-barrels, they may function as ligand binding domains.

To begin to characterize PIF1 helicases from thermophilic bacteria and determine the function of the WYL domain, we generated recombinant Te-, To-, and TyPif1 proteins (Fig. 1C and data not shown) by over-expression in *E. coli*. Although all preparations were enzymatically active *in vitro* (Fig. 2-5 and data not shown), wild-

type and mutant versions of TePif1 expressed to the highest levels, and thus, TePif1 is the focus of the work herein.

TePif1 is a DNA-stimulated ATPase – We first investigated the ATPase activity of TePif1. As *T. elfii* grows between 50 and 72°C, with an optimal temperature of 66°C (24), we sought to determine the range of temperatures over which recombinant TePif1 was active *in vitro*. As shown in Figure 2A, significant ATP hydrolysis was observed between 4 and 70°C, with maximum activity at 20-30°C. These data suggest that TePif1 could be a suitable enzyme for use in helicase-dependent amplification (HDA) or other isothermal molecular biology applications. Regardless, because TePif1 was highly active at 37°C, we used this temperature for convenience in the remainder of our biochemical assays.

Like most helicases, the ATPase activity of TePif1 was greatly stimulated by the presence of ssDNA (Fig. 2B, $k_{1/2(ATP)}$ = 2.96 ± 0.13 nM). By including ssDNA substrates of various lengths, we also found that the stimulation of TePif1 ATPase activity was length dependent, with a maximal stimulation of ~6-fold in the presence of ≥ 20 nt poly(dT) ssDNA (Fig. 2C). However, ssDNA as short as 5 nt was still able to stimulate ATPase activity ~2.7-fold, which is consistent with *S. cerevisiae* Pif1 being able to bind to very short lengths of ssDNA (25). In contrast to ssDNA, dsDNA did not significantly stimulate TePif1 ATPase activity (Fig. 2B).

Because the *S. cerevisiae* Pif1 also preferentially binds to and unwinds G-rich DNA *in vitro*, we assessed the effects of ssDNA sequence on TePif1 activity. Using equimolar concentrations of poly(dA), (dC), (dG), (dT), or random sequence 50mer oligonucleotides, we found that all ssDNAs stimulated TePif1 ATPase activity relative to control reactions lacking ssDNA (all $p < 0.001$ *vs.* no ssDNA) (Fig. 2D). Further, poly(dT) ssDNA preferentially stimulated activity ($p < 0.01$ *vs.* all other ssDNAs) (Fig. 2D). There were no significant differences in the rates of ATP hydrolysis in the presence of the other ssDNAs.

To ensure that the observed ATP hydrolysis was due to TePif1 and not a contaminating enzyme from *E. coli*, we also generated a Walker B box mutant of TePif1 (TePif1-DENQ) by mutating the acidic D106 and E107 residues to their uncharged amide counterparts (D106N, E107Q) (Fig. 1B). The Walker B box is an ATPase motif necessary for hydrolysis (26), and indeed, TePif1-DENQ was unable to hydrolyze ATP (Fig. 2B).

TePif1 is active on substrates that mimic replication forks – It is unclear what role(s) bacterial PIF1 helicases play *in vivo* (5), but *in vitro* substrate preference can be indicative of *in vivo* function. For instance, *S. cerevisiae* Pif1 is highly active on G-quadruplex substrates *in vitro* and suppresses genomic instability at sites capable of forming such structures *in vivo* (19). To examine the types of DNA substrates that TePif1 might preferentially act on, we assayed for DNA binding and unwinding of a series of simple substrates *in vitro*. As predicted by the stimulation of ATPase activity in Figure 2, TePif1 bound ssDNA but displayed minimal binding to blunt-ended dsDNA (Fig. 3A). However, substrates containing 5′, 3′, or both ssDNA extensions on duplex DNA were appreciably (> 20%) bound by TePif1. The enzyme preferentially bound ssDNA, 5′-tailed, and fork substrates, as indicated by the high affinity apparent binding constants (ssDNA K_d = 0.93 ± 0.11 nM, 5′-tail K_d = 4.53 ± 0.39 nM, fork K_d = 4.40 ± 0.26 nM) to all three substrates (Fig. 3B), though there were no significant differences between these three binding affinities ($p > 0.2$).

We next assessed *in vitro* helicase activity. Based on the poor binding to blunt dsDNA (Fig. 3A), it was unsurprising that TePif1 displayed no significant unwinding of this substrate (Fig. 4A). Consistent with the 5′-3′ directionality of PIF1 helicases (4), TePif1 displayed only basal levels of unwinding for the 3′-tailed substrate but was a robust helicase on the 5′-tailed substrate. Similarly, the fork substrate was unwound to completion (Fig. 4A). To compare unwinding of the 5′-tail and fork substrates, we next assayed for helicase activity as a function of helicase concentration. TePif1 unwound both substrates with similar activity (fork K_M = 80 ± 7 pM, 5′-tail K_M = 156 ± 9 pM) (Fig. 4B). In addition, the kinetics of unwinding were also comparable ($t_{1/2} \approx$ 1 min) (Fig. 4C). Thus, substrates that resemble replication forks or broken forks retaining 5′ ssDNA are preferred TePif1 substrates *in vitro*. As with ATP hydrolysis, the TePif1-DENQ mutant was unable to unwind DNA (data not shown).

The WYL domain impacts ATPase, DNA binding, and helicase activity – We next focused on the role of the TePif1 WYL domain. It is hypothesized that WYL domains may mediate oligomerization, so we sought to determine the oligomeric state of TePif1 in solution. In our preparative gel filtration chromatography, the peak of TePif1 eluted with a calculated molecular weight (MW) of 67.1 kDa (68. 5 kDa MW expected) (Fig. 5A), suggesting that recombinant TePif1 is monomeric. As further evidence that the WYL domain does not mediate oligomerization, recombinant TePif1 lacking this domain (TePif1ΔWYL) or containing mutations (R470A, C494A, R501A, and R504A) in the putative ligand binding channel (TePif1-4x; Fig. 1B) (16) also eluted as monomers from gel filtration: TePif1ΔWYL calculated MW = 52.1 kDa, expected MW = 58.8 kDa; TePif1-4x calculated MW = 63.7 kDa, expected MW = 68.2 kDa (Fig. 5A).

If not oligomerization, then what is the function of the TePif1 WYL domain? To address this question, we compared the biochemical activities of the TePif1ΔWYL and TePif1-4x mutant proteins with wild-type. As shown in Figure 5B, ablation of the WYL domain by deletion or mutation stimulated ATP hydrolysis 3.5-fold and 5.7-fold (respectively) relative to wild-type TePif1 (all in the presence of ssDNA). When these same protein preparations were used in helicase assays, however, DNA unwinding was inhibited by WYL domain ablation (Fig. 5C). The K_M for TePif1-4x helicase activity (0.77 ± 0.10 nM) was approximately 10-fold greater than wild-type TePif1, and the apparent K_M for TePif1ΔWYL helicase activity (16.4 ± 1.1 nM) was nearly 200-fold greater than wild-type.

Based on our biochemical observations, one factor that affects both ATPase and helicase activity is ssDNA. Thus, we hypothesized that the TePif1 domain may be a ssDNA binding module. When we assessed the ssDNA binding activity of TePif1ΔWYL, it was completely eliminated at 50 nM (Fig. 5D) and only reached levels significantly above background at > 200 nM protein (Supplemental Figure S1). This suggested that ssDNA is a ligand bound by the WYL domain.

To further investigate this possibility, we used the TePif1-4x mutant. The R470A, C494A, R501A, and R504A mutations eliminate positive charge in the putative ligand binding channel in the WYL domain, and we predicted that ssDNA binding would thus also be impaired in this mutant. As shown in Figure 5D, this hypothesis proved correct; 50 nM TePif1-4x bound only ~5% of the ssDNA substrate, an approximately 15-fold reduction relative to the wild-type protein.

Finally, we over-expressed and purified the TePif1 WYL domain in isolation (Fig. 1B and C) to test for its ability to bind to ssDNA. In gel shift assays, the WYL domain did bind to a poly(dT) 50mer substrate (Fig. 5D), and mutation of the four residues in the putative binding channel in the WYL domain (4xWYL mutant) reduced ssDNA binding. This corresponds with the wild-type TePif1, TePif1ΔWYL, and TePif1-4x data, demonstrating that the TePif1 WYL domain is an accessory ssDNA binding module. The implications of these findings are discussed below.

DISCUSSION

Eukaryotic PIF1 family helicases, especially those in metazoa, generally contain a centrally located helicase domain flanked by N- and C-terminal domains that are often large and of unknown function (4). In contrast, many bacterial PIF1 helicases are smaller proteins comprised of only the helicase domain (4,5). However, it has recently been appreciated that PIF1s from a variety of prokaryotic and eukaryotic organisms contain fusions of the PIF1 helicase core to domains with known or predicted enzymatic activities ((14) and MLB, observations). Here, we investigated the function of the C-terminal WYL domain that is found in many PIF1 helicases from thermophilic bacteria. In the context of TePif1, we found that the WYL domain is an accessory ssDNA binding module important for regulating the biochemical activities of the helicase

The WYL domain provides another DNA binding interface for TePif1 – The data in Figure 5 indicate that the TePif1 WYL domain binds ssDNA. However, this is not the only DNA binding interface in TePif1. Although the TePif1-ΔWYL and TePif1-4x mutant proteins bound ssDNA poorly at a concentration of 50 nM, increasing the helicase concentration to 200-500 nM revealed ssDNA binding of up to ~60% (Fig. S1). Thus, the WYL domain appears to be needed for high affinity ssDNA binding, but other motifs

in TePif1 can bind to ssDNA with lower affinity. This could explain why helicase activity decreased for the TePif1ΔWYL and TePif1-4x proteins (Fig. 5C). A lower affinity for the substrate and greater off-rate due to the lack of a high affinity ssDNA binding interface would both decrease TePif1 DNA unwinding.

Perhaps the helicase domain enables binding to dsDNA-ssDNA junctions, while the WYL domain is exclusively involved in interactions with ssDNA. A similar phenomenon has been observed for *S. cerevisiae* Pif1 using single molecule experiments, where the helicase does not translocate along ssDNA but localizes to a dsDNA-ssDNA junction and reiteratively reels in ssDNA (27). *S. cerevisiae* Pif1 lacks a WYL domain, but it does contain N- and C-terminal domains of unknown function (4) that could harbor a cryptic ssDNA binding interface.

It should be noted that some helicases that are highly homologous to TePif1, such as ToPif1 (35% identity, 56% similarity), lack a WYL domain (Fig. 1A). Thus, they may display a different spectrum of DNA substrate preferences. In the case of ToPif1, however, its domain organization is slightly different from that of TePif1 and TyPif1, which have well defined helicase, UvrD_C_2, and WYL domains (Fig. 1A). In ToPif1, the UvrD_C_2 domain is larger than in TePif1 and TyPif1, and it is predicted to overlap with the C-terminal portion of the helicase domain. Perhaps this unique domain arrangement generates a ssDNA binding surface that can function in a similar manner to the WYL domain found in TePif1.

What is the in vivo *role of TePif1?* – To date, it is still unclear what role(s) PIF1 family helicases play in bacteria, even in species like *T. elfii* that encode just a single family member (5). The only indications that we do have come from biochemical data generated using recombinant proteins. Aside from an ability to bind and unwind G-quadruplex DNA structures that is conserved across PIF1 helicases from multiple bacterial phyla (19), DNA structures that mimic simple replication forks or broken forks with 5′ ssDNA tails appear to be preferred substrates for TePif1 (Fig. 3 and 4) and the Pif1 helicase from *Bacteroides sp. 3_1_23* (28). By analogy to the predicted role of *S. cerevisiae* Rrm3 as an accessory replicative helicase needed for efficient replication fork progression past nucleoprotein barriers (4,29-31), TePif1 and other bacterial PIF1 helicases may serve a similar function during replication (Fig. 6). If the TePif1 helicase domain does interact with the dsDNA-ssDNA junction as postulated above, perhaps its WYL domain acts as a sensor for ssDNA accumulating behind replication forks. We found that alterations to the TePif1 WYL domain affect helicase activity (Fig. 5C), so this could be a mechanism to enhance or inhibit DNA unwinding at the replication fork as protein barriers or DNA lesions are encountered (Fig. 6). It will be informative to determine if bacterial PIF1 helicases have the ability to disrupt stable nucleoprotein complexes and/or physically interact with replisome components.

TePif1 is active across a wide range of temperatures – T. elfii can be cultured at temperatures ranging from 50 to 72°C, but fails to grow at 45 and 75°C (24). Recombinant TePif1 displayed ATPase activity at an even wider range of temperatures (4-70°C), but maximal activity occurred at temperatures (20-30°C) (Fig. 2A) well below the optimal growth temperature of *T. elfii*. The reason for this phenomenon is unclear, but it could be due to lack of a partner protein that stabilizes or stimulates TePif1 activity *in vivo*. We could not assay for high-temperature helicase activity because the short DNA substrates used in our assays have a T_m of 53.4°C, but we predict that temperature would affect helicase activity in a similar manner to ATP hydrolysis. This broad range of temperatures at which TePif1 is active is similar to the appreciable ATPase activity from 37-65°C of the *Picrophilus torridus* MCM helicase (32). *P. torridus* is a thermophilic archaeon that grows between 45 and 65°C, with a growth optimum of 60°C (33). The MCM helicases from the thermophilic *Methanobacterium thermoautotrophicum* and hyperthermophilic *Sulfolobus islandicus* display similarly thermostable enzymatic activities (34,35).

We also found that TePif1 is stable and remains active for DNA unwinding even when stored at room temperature for at least 1 week (data not shown). This fact and the above data suggest that TePif1 may be a valuable enzyme for isothermal molecular biology applications, such as HDA (36). Ideally, the helicase used in HDA can unwind blunt dsDNA, and though TePif1 largely

lacks this ability at 37°C, increasing the reaction temperature may increase helicase activity on blunt substrates. The addition of the polymerase in the HDA reaction may also stimulate TePif1 activity, a phenomenon that has been observed with the bacteriophage T7 DNA polymerase and replicative helicase combination (37). As PIF1 family helicases are also able to unwind a variety of non-canonical DNA secondary structures (4,19,38), TePif1 may be a valuable accessory helicase in traditional HDA reactions that can be inhibited by such structures. Indeed, our ongoing work focuses on the ability of TePif1 to unwind structured DNA and its performance in HDA.

In conclusion, we biochemically characterized the first PIF1 family helicase from a thermophilic bacterium, demonstrating that TePif1 has similar activities to its better studied bacterial and eukaryotic homologs. Further, we also found that the WYL domain in TePif1 functions to bind ssDNA and regulate ATPase and helicase activity. Our data combined with similar investigations of PIF1 helicases fused to domains with other enzymatic activities (14) will shed light on the *in vivo* functions of PIF1 family proteins in diverse organisms and may suggest roles for the uncharacterized non-helicase domains of eukaryotic PIF1s.

EXPERIMENTAL PROCEDURES

Nucleotides, oligonucleotides, and other reagents – ATP was purchased from GE Healthcare (Little Chalfont, UK) or DOT Scientific (Burton, MI, USA). α-^{32}P-ATP and γ-^{32}P-ATP were purchased from PerkinElmer (Waltham, MA, USA). The oligonucleotides used in this work were synthesized by IDT (Coralville, IA, USA) and are listed in Supplemental Table S1. All enzymes were from New England Biolabs (Ipswich, MA, USA) unless otherwise noted.

Protein cloning, expression, and purification – A synthetic gBlock DNA encoding the TePif1 protein sequence (Accession no. WP_028843943) was codon optimized for expression in *E. coli* and synthesized by IDT. The DNA was PCR-amplified using oligonucleotides MB1035 and MB1036 (Table S1) and cloned into the *Bam*HI and *Xho*I sites of pMB131 (19) to generate the expression vector pMB427. The sequence of this and all other vectors was verified by DNA sequencing (ACGT, Inc., Wheeling, IL, USA). Details of the TePif1-DENQ, TePif1-ΔWYL, TePif1-4x, WYL, ToPif1, and TyPif1 clonings can be found in the Supplemental Data.

Expression plasmids were transformed into chemically competent Rosetta 2(DE3) pLysS cells and selected for at 37°C on LB medium supplemented with 100 μg/mL ampicillin and 34 μg/mL chloramphenicol. Fresh transformants were used to inoculate one or more 5-mL LB cultures supplemented with antibiotics and incubated at 30°C for ~6 h with aeration. These starter cultures were then diluted 1:100 in ZYP-5052 autoinduction medium containing 1x trace metals mix (39), 100 μg/mL ampicillin, and 34 μg/mL chloramphenicol and incubated at 22°C with agitation to OD_{600} >3 (15-18 h). Cells were harvested by centrifugation for 10 min at 5,500 x g and 4°C. Cell pellets were weighed and frozen at -80°C prior to lysis or for long-term storage.

Frozen cell pellets were thawed at room temperature by stirring in 4 mL/g cell pellet resuspension buffer (25 mM Na-HEPES (pH 7.5), 5% (v/v) glycerol, 300 mM NaOAc, 5 mM MgOAc, and 0.05% Tween-20) supplemented with 1x protease inhibitor cocktail (Sigma), and 20 μg/mL DNase I. Cells were lysed by six passes through a Cell Cracker operated at >1000 psi. All subsequent steps were performed at 4°C. The soluble fraction was clarified by centrifugation for 30 min at 33,000 x g followed by filtering the supernatant through a 0.22-μm membrane. This mixture was then applied to a Strep-Tactin Sepharose column (IBA) pre-equilibrated in resuspension buffer using an ÄKTA Pure (GE Healthcare Life Sciences). The column was washed with 20 column volumes (CVs) of resuspension buffer, 10 CVs of resuspension buffer supplemented with 5 mM ATP, and 10 CVs of resuspension buffer. Protein was eluted with 15 CVs of resuspension buffer supplemented with 2.5 mM desthiobiotin (IBA). Column fractions were examined on 8% SDS-PAGE gels run at 20 V/cm and stained with Coomassie Brilliant Blue R-250 (BioRad). Peak fractions were pooled, concentrated with Amicon Ultra-4 30K centrifugal filters, and loaded onto a HiPrep 16/60 Sephacryl S-200 HR column (GE Healthcare Life Sciences) pre-equilibrated in gel filtration buffer (25 mM Na-HEPES (pH 7.5), 5% glycerol, 300 mM NaCl, 5 mM $MgCl_2$, and 0.05% Tween-20). The protein was eluted with 1.5 CVs gel filtration buffer, and

fractions were analyzed by SDS-PAGE as above. Peak fractions were pooled, snap-frozen with liquid nitrogen, and stored at -80°C.

ATPase assays – ATP hydrolysis was measured via two methods. The data in Figure 2A were generated by thin-layer chromatography (TLC) using the method described in (40) with some modifications. Briefly, reactions were prepared in helicase buffer (25 mM Na-HEPES [pH 8.0], 5% glycerol, 50 mM NaOAc [pH 7.6], 150 µM NaCl, 7.5 mM MgOAc, and 0.01% Tween-20) and contained 50 nM TePif1, 1 µM oligonucleotide MB551 (Table S1), and 6 µCi of α[^{32}P]-ATP. The reactions were incubated for 15 min at temperatures ranging from 4-90°C and stopped by the addition of 2% SDS and 20 mM EDTA. ATP was then separated from ADP by TLC in a buffer composed of 1 M formic acid and 400 mM LiCl for 1 h. The ratio of ATP to ADP was quantified by phosphoimaging using a Typhoon FLA 9500 Variable Mode Imager and ImageQuant 5.2 software. Background levels of autohydrolysis were subtracted at each temperature tested, and the data were normalized to the highest ATPase activity in each experiment to facilitate comparison of the relative ATPase activity across the temperature range examined.

The ATPase data in Figures 2B-D and Figure 5B were generated using a NADH-coupled ATPase assay as described in (41). Briefly, reactions were performed in ATPase buffer (25 mM Na-HEPES [pH 8.0], 5% glycerol, 50 mM NaOAc [pH 7.5], 150 µM NaCl and 0.01% NP-40 substitute) supplemented with 5 mM ATP (pH 7.0), 5 mM $MgCl_2$, 0.5 mM phospho(enol)pyruvic acid, 0.4 mM NADH, 5 U/mL pyruvate kinase, 8 U/mL lactate dehydrogenase, 50 nM helicase (unless otherwise stated), and 1 µM DNA substrate. Absorbance at 340 nm was read at 37°C in 96-well plates using a BioTek Synergy H1 microplate reader. Absorbance readings were converted to ATP turnover based on NADH concentration. It was assumed that 1 µM NADH oxidized is proportional to 1 µM ATP hydrolyzed.

DNA binding assay – DNA binding was measured via gel shifts. The proteins were incubated at the indicated concentrations with 1 nM radiolabeled DNA for 30 min at 37°C in resuspension buffer. Protein–DNA complexes were separated from unbound DNA on 8% 19:1 acrylamide:bis-acrylamide gels in TBE buffer at 10 V/cm. Gels were dried under vacuum and imaged using a Typhoon 9210 Variable Mode Imager. DNA binding was quantified using ImageQuant 5.2 software. All radiolabeled DNA substrates were prepared using T4 polynucleotide kinase and γ-^{32}P-ATP by standard methods. See the Supplemental Materials for additional details.

Helicase assay – DNA unwinding was measured by incubating the indicated concentrations of helicase with 5 mM ATP and 1 nM radiolabeled DNA in resuspension buffer. Reactions were incubated at 37°C for 30 min and stopped with the addition of 1× Stop-Load dye (5% glycerol, 20 mM EDTA, 0.05% SDS and 0.25% bromophenol blue). Unwound DNA was separated on 8% 19:1 acrylamide:bis-acrylamide gels in TBE buffer at 10 V/cm and imaged as for the DNA binding assays. Unwinding time courses were performed under the same conditions but with a single helicase concentration of 50 nM. Reactions were incubated at 37°C for times ranging from 0 to 30 min and stopped as above.

Statistical analyses – All data were analyzed and plotted using GraphPad Prism 6 (GraphPad Software, Inc). The plotted values are averages, and the error bars were calculated as the standard deviation from three or more independent experiments. *P*-values were determined by analysis of variance (ANOVA). We defined statistical significance as $p < 0.05$.

Acknowledgements: We thank Cody Rogers and David Nickens for carefully reading this manuscript, Garrett Booher, Matan Cohen, and Ankon Paul for contributing early results, and members of the Bochman and van Kessel labs for feedback.

Conflict of interest: The authors declare that they have no conflicts of interest with the contents of this article.

Author contributions: NMA conducted most of the experiments, analyzed the results, and helped to write the paper. CWS performed the experiments in Figure 2A, analyzed the results, and helped to write the paper. AA initiated this project and performed the protein purification experiments and enzymatic assays. MLB conceived the project, performed the experiments in Figure 5E, analyzed the results, and wrote the paper with NMA and CWS.

REFERENCES

1. Abdelhaleem, M. (2010) Helicases: an overview. *Methods Mol Biol* **587**, 1-12
2. Eki, T. (2010) Genome-Wide Survey and Comparative Study of Helicase Superfamily Members in Sequenced Genomes. in *Advances in Genetics Research* (Urbano, K. V. ed.), Nova Science Publishers, Inc. pp 168-203
3. Uchiumi, F., Seki, M., and Furuichi, Y. (2015) Helicases and human diseases. *Front Genet* **6**, 39
4. Bochman, M. L., Sabouri, N., and Zakian, V. A. (2010) Unwinding the functions of the Pif1 family helicases. *DNA Repair (Amst)* **9**, 237-249
5. Bochman, M. L., Judge, C. P., and 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
6. Zhou, J.-Q., Qi, H., Schulz, V., Mateyak, M., Monson, E., and Zakian, V. (2002) *Schizosaccharomyces pombe pfh1+* encodes an essential 5' to 3' DNA helicase that is a member of the *PIF1* sub-family of DNA helicases. *Mol. Biol. Cell* **13**, 2180-2191
7. Gagou, M. E., Ganesh, A., Phear, G., Robinson, D., Petermann, E., Cox, A., and Meuth, M. (2014) Human PIF1 helicase supports DNA replication and cell growth under oncogenic-stress. *Oncotarget* **5**, 11381-11398
8. Gagou, M. E., Ganesh, A., Thompson, R., Phear, G., Sanders, C., and Meuth, M. (2011) Suppression of apoptosis by PIF1 helicase in human tumor cells. *Cancer Res* **71**, 4998-5008
9. Aswad, L., Yenamandra, S. P., Ow, G. S., Grinchuk, O., Ivshina, A. V., and Kuznetsov, V. A. (2015) Genome and transcriptome delineation of two major oncogenic pathways governing invasive ductal breast cancer development. *Oncotarget* **6**, 36652-36674
10. Li, C. W., and Chen, B. S. (2016) Investigating core genetic-and-epigenetic cell cycle networks for stemness and carcinogenic mechanisms, and cancer drug design using big database mining and genome-wide next-generation sequencing data. *Cell Cycle* **15**, 2593-2607
11. Gala, M. K., Mizukami, Y., Le, L. P., Moriichi, K., Austin, T., Yamamoto, M., Lauwers, G. Y., Bardeesy, N., and Chung, D. C. (2014) Germline mutations in oncogene-induced senescence pathways are associated with multiple sessile serrated adenomas. *Gastroenterology* **146**, 520-529
12. Chisholm, K. M., Aubert, S. D., Freese, K. P., Zakian, V. A., King, M. C., and Welcsh, P. L. (2012) A genomewide screen for suppressors of Alu-mediated rearrangements reveals a role for PIF1. *PLoS One* **7**, e30748
13. Syed, S., Desler, C., Rasmussen, L. J., and Schmidt, K. H. (2016) A Novel Rrm3 Function in Restricting DNA Replication via an Orc5-Binding Domain Is Genetically Separable from Rrm3 Function as an ATPase/Helicase in Facilitating Fork Progression. *PLoS Genet* **12**, e1006451
14. Harman, A., and Manna, S. (2016) Identification of Pif1 helicases with novel accessory domains in various amoebae. *Mol Phylogenet Evol* **103**, 64-74
15. Matsuda, K., Makise, M., Sueyasu, Y., Takehara, M., Asano, T., and Mizushima, T. (2007) Yeast two-hybrid analysis of the origin recognition complex of Saccharomyces cerevisiae: interaction between subunits and identification of binding proteins. *FEMS Yeast Res* **7**, 1263-1269
16. Makarova, K. S., Anantharaman, V., Grishin, N. V., Koonin, E. V., and Aravind, L. (2014) CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems. *Front Genet* **5**, 102
17. Ivessa, A. S., Zhou, J.-Q., Schulz, V. P., Monson, E. M., and Zakian, V. A. (2002) *Saccharomyces* Rrm3p, a 5' to 3' DNA helicase that promotes replication fork progression through telomeric and sub-telomeric DNA. *Genes Dev.* **16**, 1383-1396
18. Wilson, M. A., Kwon, Y., Xu, Y., Chung, W. H., Chi, P., Niu, H., Mayle, R., Chen, X., Malkova, A., Sung, P., and Ira, G. (2013) Pif1 helicase and Poldelta promote recombination-coupled DNA synthesis via bubble migration. *Nature* **502**, 393-396
19. Paeschke, K., Bochman, M. L., Garcia, P. D., Cejka, P., Friedman, K. L., Kowalczykowski, S. C., and Zakian, V. A. (2013) Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature* **497**, 458-462

20. 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., and Xi, X. G. (2016) Crystal structures of the BsPif1 helicase reveal that a major movement of the 2B SH3 domain is required for DNA unwinding. *Nucleic Acids Res*
21. Zhou, X., Ren, W., Bharath, S. R., Tang, X., He, Y., Chen, C., Liu, Z., Li, D., and Song, H. (2016) Structural and Functional Insights into the Unwinding Mechanism of Bacteroides sp Pif1. *Cell reports* **14**, 2030-2039
22. Kelman, L. M., and Kelman, Z. (2014) Archaeal DNA replication. *Annu Rev Genet* **48**, 71-97
23. Conserved Protein Domain Family *UvrD_C_2*. National Center for Biotechnology Information, https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=316096
24. Ravot, G., Magot, M., Fardeau, M. L., Patel, B. K., Prensier, G., Egan, A., Garcia, J. L., and Ollivier, B. (1995) Thermotoga elfii sp. nov., a novel thermophilic bacterium from an African oil-producing well. *Int J Syst Bacteriol* **45**, 308-314
25. Barranco-Medina, S., and Galletto, R. (2010) DNA binding induces dimerization of Saccharomyces cerevisiae Pif1. *Biochemistry* **49**, 8445-8454
26. Hanson, P. I., and Whiteheart, S. W. (2005) AAA+ proteins: have engine, will work. *Nat Rev Mol Cell Biol* **6**, 519-529
27. Zhou, R., Zhang, J., Bochman, M. L., Zakian, V. A., and Ha, T. (2014) Periodic DNA patrolling underlies diverse functions of Pif1 on R-loops and G-rich DNA. *eLife* **3**, e02190
28. Liu, N. N., Duan, X. L., Ai, X., Yang, Y. T., Li, M., Dou, S. X., Rety, S., Deprez, E., and Xi, X. G. (2015) The Bacteroides sp. 3_1_23 Pif1 protein is a multifunctional helicase. *Nucleic Acids Res* **43**, 8942-8954
29. Bruning, J. G., Howard, J. L., and McGlynn, P. (2014) Accessory Replicative Helicases and the Replication of Protein-Bound DNA. *J Mol Biol*
30. Ivessa, A. S., Lenzmeier, B. A., Bessler, J. B., Goudsouzian, L. K., Schnakenberg, S. L., and Zakian, V. A. (2003) The *Saccharomyces cerevisiae* helicase Rrm3p facilitates replication past nonhistone protein-DNA complexes. *Mol. Cell* **12**, 1525-1536
31. Azvolinsky, A., Dunaway, S., Torres, J., Bessler, J., and Zakian, V. A. (2006) The *S. cerevisiae* Rrm3p DNA helicase moves with the replication fork and affects replication of all yeast chromosomes. *Genes Dev* **20**, 3104-3116
32. Goswami, K., Arora, J., and Saha, S. (2015) Characterization of the MCM homohexamer from the thermoacidophilic euryarchaeon Picrophilus torridus. *Sci Rep* **5**, 9057
33. Schleper, C., Puehler, G., Holz, I., Gambacorta, A., Janekovic, D., Santarius, U., Klenk, H. P., and Zillig, W. (1995) Picrophilus gen. nov., fam. nov.: a novel aerobic, heterotrophic, thermoacidophilic genus and family comprising archaea capable of growth around pH 0. *J Bacteriol* **177**, 7050-7059
34. Shechter, D. F., Ying, C. Y., and Gautier, J. (2000) The intrinsic DNA helicase activity of Methanobacterium thermoautotrophicum delta H minichromosome maintenance protein. *J Biol Chem* **275**, 15049-15059
35. Xia, Y., Niu, Y., Cui, J., Fu, Y., Chen, X. S., Lou, H., and Cao, Q. (2015) The Helicase Activity of Hyperthermophilic Archaeal MCM is Enhanced at High Temperatures by Lysine Methylation. *Frontiers in microbiology* **6**, 1247
36. Jeong, Y. J., Park, K., and Kim, D. E. (2009) Isothermal DNA amplification in vitro: the helicase-dependent amplification system. *Cell Mol Life Sci* **66**, 3325-3336
37. Stano, N. M., Jeong, Y. J., Donmez, I., Tummalapalli, P., Levin, M. K., and Patel, S. S. (2005) DNA synthesis provides the driving force to accelerate DNA unwinding by a helicase. *Nature* **435**, 370-373
38. Bochman, M. L., Paeschke, K., and Zakian, V. A. (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet*
39. Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**, 207-234

40. Bochman, M. L., Bell, S. P., and Schwacha, A. (2008) Subunit organization of Mcm2-7 and the unequal role of active sites in ATP hydrolysis and viability. *Mol Cell Biol* **28**, 5865-5873
41. Rogers, C. M., Wang, J. C.-Y., Noguchi, H., Imasaki, T., Takagi, Y., and Bochman, M. L. (2017) Yeast Hrq1 shares structural and functional homology with the disease-linked human RecQ4 helicase. *Nucleic Acids Res*
42. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948

FOOTNOTES

This work in this paper was supported by funds from the College of Arts And Sciences, Indiana University, a Collaboration in Translational Research Pilot Grant from the Indiana Clinical and Translational Sciences Institute, and the American Cancer Society (RSG-16-180-01-DMC).

The abbreviations used are: dsDNA, double-stranded DNA; ssDNA, single-stranded DNA; hPIF1, human PIF1; TePif1, *Thermotoga elfii* Pif1; ToPif1, *Thermus oshimai* Pif1; TyPif1, *Thermodesulfovibrio yellowstonii* Pif1; HDA, helicase dependent amplification; MW, molecular weight; TePif1ΔWYL, TePif1 lacking its C-terminal WYL domain; TePif1-4x, TePif1 containing R470A, C494A, R501A, and R504A mutations; CV, column volume; TLC, thin-layer chromatography

FIGURE LEGENDS

FIGURE 1. PIF1 family helicases in thermophilic bacteria. *A*, Sequence alignment of TePif1, TyPif1, and ToPif1. The sequences of the three proteins were aligned using ClustalW (42), and the BOXSHADE program in the Biology WorkBench suite (http://workbench.sdsc.edu) was used to color-code conserved residues. The completely conserved residues are in green, conserved similarities are in cyan, and identical residues are yellow. The PIF1 family helicase domain is underlined in black, and the PIF1 family signature sequence motif is denoted by the black box. The WYL domain is underlined in green. TePif1 residues that were mutated in this work are marked with asterisks. *B*, Domain schematics of the proteins that were the focus of this work. The PIF1 family helicase domain (PIF1) is blue, the UvrD_C_2 domain (DC2) is red, and the WYL domain is green. The size of the proteins in amino acids is shown. *C*, SDS-PAGE of purified recombinant protein preparations. Fractions from the TePif1 (left) and WYL (right) purifications were separated on 8% and 12% acrylamide gels, respectively, and stained with Coomassie blue. The positions of the MW markers are shown.

FIGURE 2. TePif1 ATPase activity is preferentially stimulated by long tracts of poly(dT) ssDNA. *A*, TePif1 is an active ATPase at a wide range of temperatures. TLC was used to measure TePif1 ATPase activity in the presence of 1 μM poly(dT) 50mer ssDNA at the indicated temperatures. The data are normalized to the highest level of ATP hydrolysis measured in each experiment. *B*, TePif1 ATPase activity is greatly stimulated by the presence of ssDNA. The coupled ATPase assay was used to measure ATP hydrolysis by TePif1 at the indicated concentrations in the absence (- DNA) or presence (+DNA) of 1 μM poly(dT) 50mer oligonucleotide. *C.*, The effect of ssDNA length on TePif1 ATPase activity. ATPase hydrolysis by 50 nM TePif1 was measured in the presence of poly(dT) oligonucleotides of the indicated lengths. The slopes (rates) of the hydrolysis curves are plotted *vs.* ssDNA length. *D*, The effect of ssDNA sequence on TePif1 ATPase activity. The random-sequence, poly(dA), poly(dC), poly(dG), and poly(dT) oligonucleotides were all 50 nt long. In this and all other figures, the data points are averages from ≥ 3 independent experiments, and the error bars represent the standard deviation.

FIGURE 3. TePif1 preferentially binds ssDNA with high affinity. *A*, Binding of 50 nM TePif1 to the five substrates shown. *B*, DNA binding as a function of helicase concentration. TePif1 binds tightly to ssDNA and duplex DNA with 5′ (5′ tail) or 5′ and 3′ ssDNA extensions (Fork).

FIGURE 4. TePif1 unwinds DNA substrates with 5′ ssDNA tails. *A*, DNA unwinding by 50 nM TePif1. *B*, Helicase activity as a function of helicase concentration. *C*, The kinetics of fork and 5′-tailed substrate unwinding by 50 nM TePif1.

FIGURE 5. The WYL domain binds ssDNA. *A*, gel filtration chromatography trace of wild-type TePif1 and the calculated MW of TePif1 (black), TePif1ΔWYL (blue), and TePif1-4x (red) based on the standard

curve (*inset*). *B*, TePif1 ATPase activity is stimulated by ablation of the WYL domain. ATP hydrolysis was measured using the coupled assay for the indicated concentrations of TePif1, TePif1ΔWYL, and TePif1-4x. *C*, Helicase activity is inhibited by ablation of the WYL domain. *D*, The TePif1 WYL domain binds ssDNA. Binding by 50 nM wild-type TePif1, truncated protein lacking the WYL domain (TePif1ΔWYL), full-length protein containing R470A, C494A, R501A, and R504A mutations (TePif1-4x), the isolated wild-type WYL domain, and the quadruple R470A, C494A, R501A, and R504A WYL mutant (WYL-4x) is shown.

FIGURE 6. Hypothetical model for TePif1 activity at replication forks. During replication, TePif1 may travel with the replisome, as has been suggested for *S. cerevisiae* Rrm3 (31). Perhaps it uses its 5′-3′ unwinding activity (red arrow) to act as an accessory helicase to aid the hexameric 5′-3′ replicative helicase in fork progression past replication barriers. In the face of genotoxic stress, the replicative polymerases and helicase can become uncoupled, generating excess ssDNA behind the helicase. TePif1 may use its WYL domain as a sensor for this ssDNA and putatively slow the progress of the replicative helicase until the full replisome can be reestablished (*i.e.*, protective fork stalling). In the absence of TePif1, the replicative helicase continues to unwind the genome, generating long tracts of ssDNA that are fragile and can lead to fork breakage.

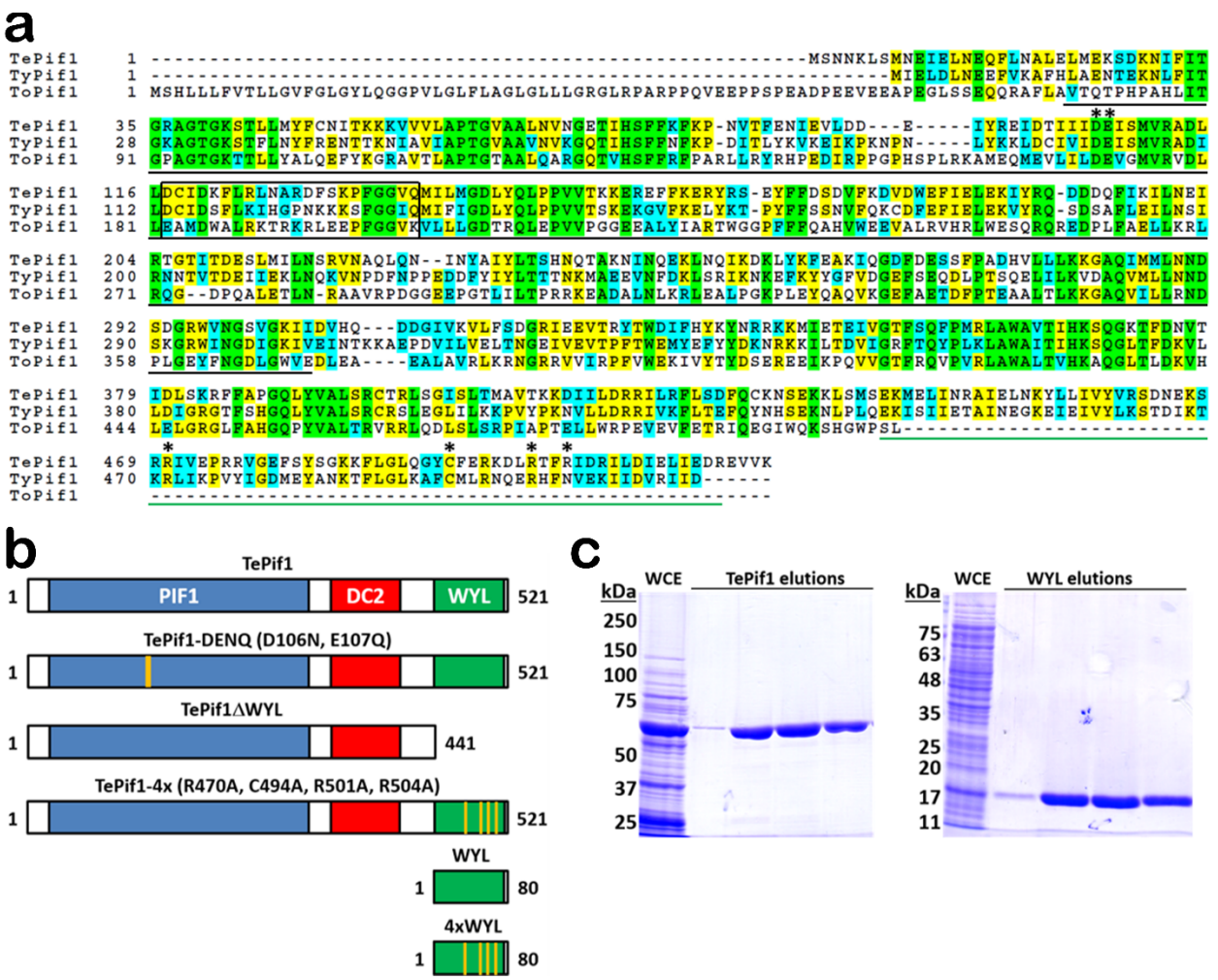

FIGURE 1

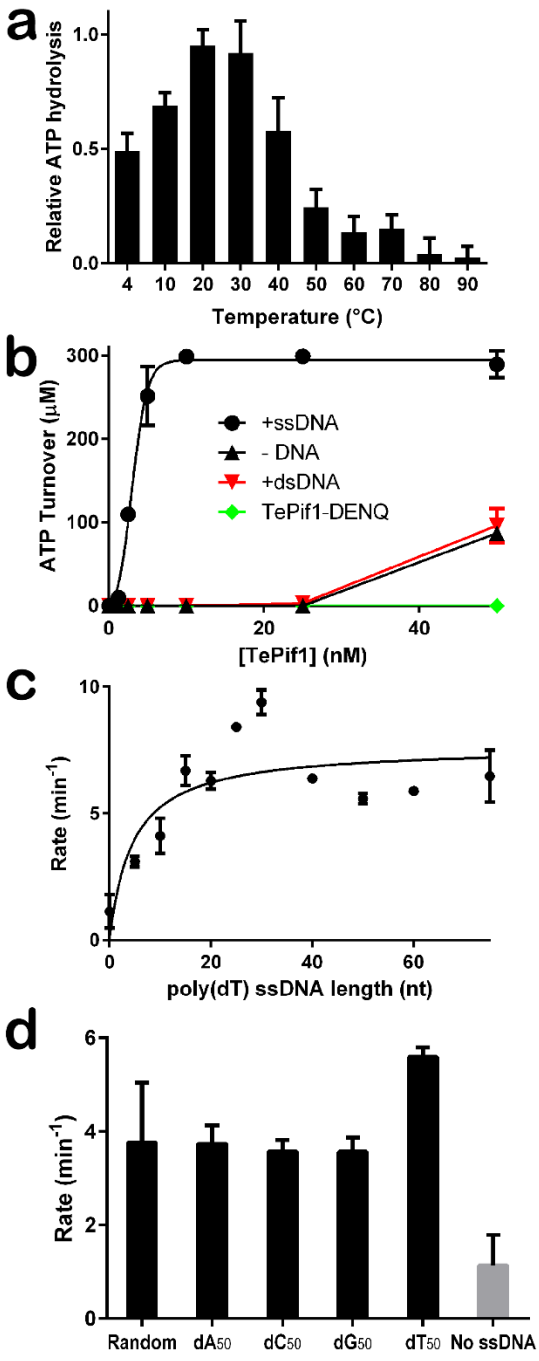

FIGURE 2

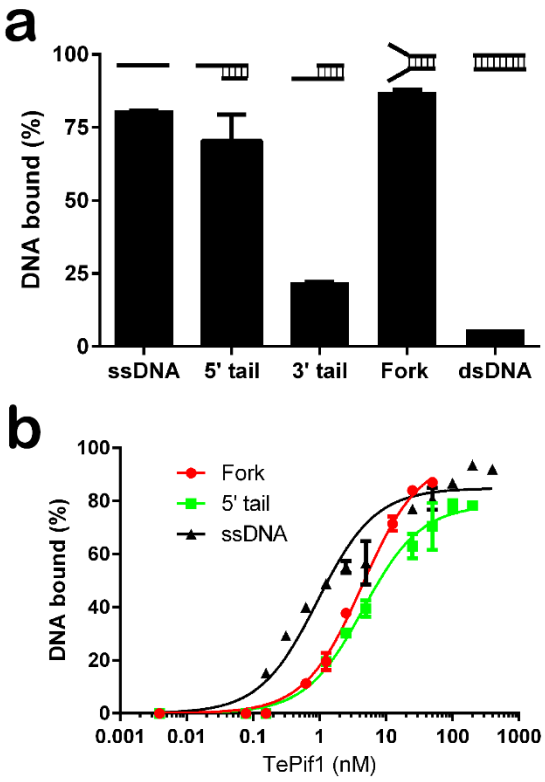

FIGURE 3

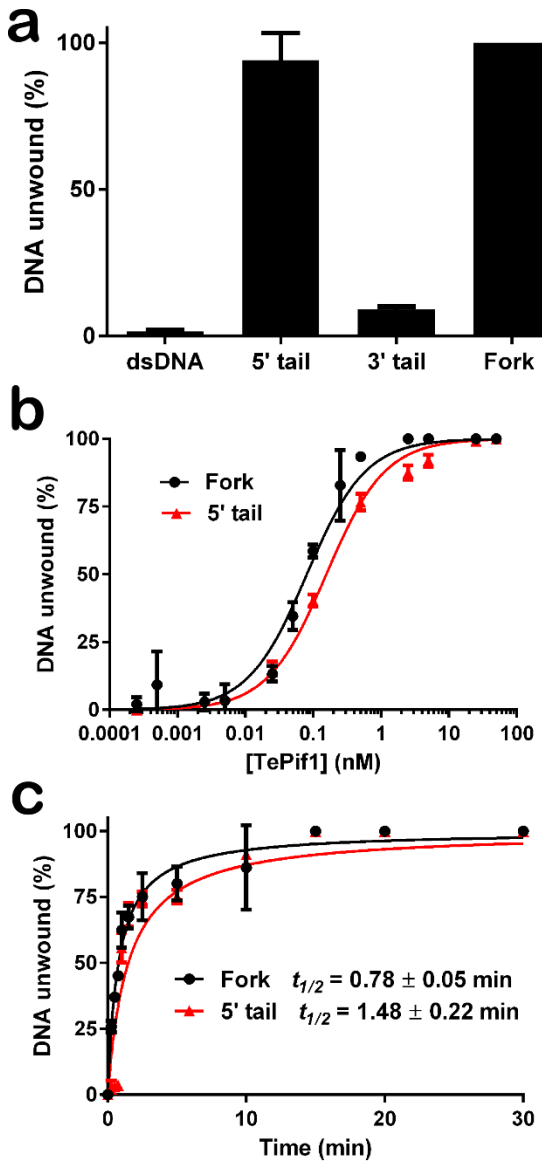

FIGURE 4

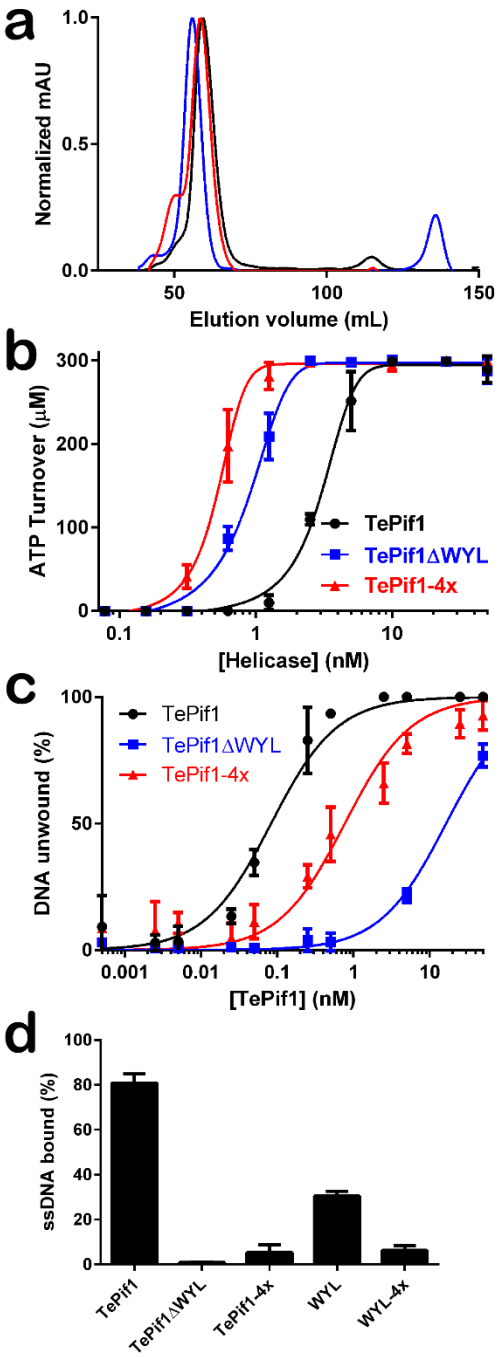

FIGURE 5

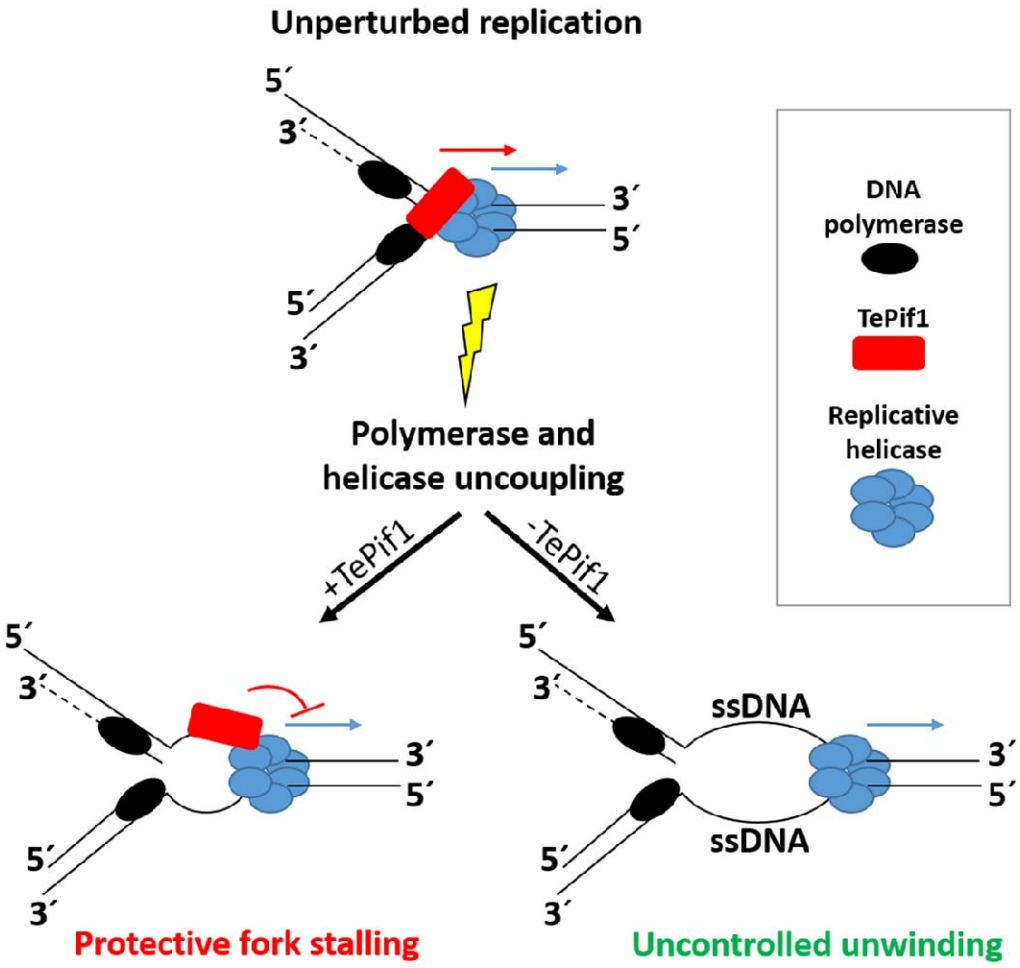

FIGURE 6