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

Nicholas M. Andis, Christopher W. Sausen, Ashna Alladin, and Matthew L. Bochman

Materials included in the Supporting Information:

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Cloning TePif1-DENQ – Oligonucleotides MB1209 and MB1210 were used for site-directed mutagenesis of plasmid pMB427 to create the ATPase- and helicase-dead D106N, E107Q allele of the gene encoding TePif1. This generated expression vector pMB481. The sequence of this and all other vectors was verified by DNA sequencing (ACGT, Inc., Wheeling, IL, USA).

Cloning TePif1-ΔWYL – The sequencing encoding the first 441 amino acids of TePif1 was PCR-amplified using oligonucleotides MB1035 and MB1124 and cloned into the *Bam*HI and *Xho*I sites of pMB131 [1] to generate the expression vector pNMA1.

Cloning TePif1-4x – The sequence encoding the 4x point mutant (R470A, C494A, R501A, R504A) was synthesized as a gBlock by IDT (Coralville, IA, USA). The full gene encoding TePif1-4x was cloned in two parts: the 5′ section was PCR amplified from pMB427 using oligonucleotides MB1035 and MB1123, and the 3′ section was PCR amplified from the gBlock using oligonucleotides MB1122 and MB1036. These pieces were then fused using isothermal DNA assembly (IDA) [2], the full-length gene was PCR-amplified using oligonucleotides MB1035 and MB1036, and the product was cloned into the *Bam*HI and *Xho*I sites of pMB131 [1] to generate the expression vector pNMA2.

Cloning the WYL domain – Using pMB427 as a template, the WYL domain and plasmid backbone were PCR-amplified using oligonucleotides MB1219 and MB1220. The reaction was treated with *Dpn*I for 2 h at 37°C, and the expression plasmid (pMB479) was assembled by intramolecular IDA of the PCR product.

Cloning ToPif1 – A synthetic gBlock DNA encoding the ToPif1 protein sequence (Accession no. WP_016329679.1) was codon optimized for expression in *E. coli* and synthesized by IDT. The DNA was PCR-amplified using oligonucleotides MB1033 and MB1034 (Table S1) and cloned into the *Bam*HI and *Xho*I sites of pMB131 [1] to generate the expression vector pNMA3.

Cloning TyPif1 – A synthetic gBlock DNA encoding the TyPif1 protein sequence (Accession no. WP_012546127.1) was codon optimized for expression in *E. coli* and synthesized by IDT. The DNA was PCR-amplified using oligonucleotides MB985 and MB986 (Table S1) and cloned into the *Bam*HI and *Xho*I sites of pMB131 [1] to generate the expression vector pMB415.

DNA substrate preparation – DNA substrates for gel shifts and helicase assays were made by 5′-end labeling the oligonucleotides indicated in Supplemental Table S1 with T4 polynucleotide kinase (PNK) and γ[^{32}P]-ATP. Labeled oligonucleotides were separated from free label using illustra ProbeQuant G-50 micro columns (GE Healthcare) following the manufacturer's instructions. Oligonucleotides were annealed by incubating complementary or partially complementary oligonucleotides overnight at 37°C in Annealing Buffer (20 mM Tris–HCl [pH 8], 4% glycerol, 0.1 mM EDTA, 40 μg/ml BSA, 10 mM DTT and 10 mM MgOAc) [3]. The blunt dsDNA substrate was created by annealing oligonucleotides MB733 and MB822, the 5 ′-tailed substrate was made by annealing oligonucleotides MB1167 and MB820, the 3 ′-tailed substrate was made by annealing oligonucleotides MB820 and MB1168, and the fork substrate was made by annealing oligonucleotides MB1167 and MB1168.

SUPPLEMENTAL TABLES

TABLE S1. Oligonucleotides used in this study.

Name	Sequence (5 ′-3 ′)
MB459	TTTTT
MB460	TTTTTTTTTT
MB461	TTTTTTTTTTTTTTT
MB462	TTTTTTTTTTTTTTTTTTT
MB463	TTTTTTTTTTTTTTTTTTTTTTTT
MB464	TTTTTTTTTTTTTTTTTTTTTTTTTTTT
MB547	GACGTCATAGACGATTACATTGCTAGGACATGCTGTCTAGAGACTATCGC
MB548	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
MB549	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCC
MB550	GGGGGGGGGGGGGGGGGGGAGGGGGGGGGGGGGGGGGAGGGGGGGG
MB551	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
MB733	ACCGTTGTGCAACTGAGTGGACAACGTGTCACTCACATAGCGTTC
MB734	GAACGCTATGTGAGTGACACCAACAGGTGAGTCAACGTGTTGCCA
MB820	GAACGCTATGTGAGTGACAC
MB821	GTGTCACTCACATAGCGTTC
MB822	GAACGCTATGTGAGTGACACGTTGTCCACTCAGTTGCACAACGGT
MB985	TGGAAGTTTTGTTCCAAGGTCCAGGATCCATGATCGAACTGGATCTGAATGAAGAATTTG
MB986	TCTCAGTGGTGGTGGTGGTGGTGCTCGAGATCAATAATGCGGACATCAATGATCTTCTCG
MB1033	CGGGATTCATGAGTCATCTGCTGCTGTTCGTAACC
MB1034	CGATCGATCTCGAGCAGTGAAGGCCATCCATGGCTTTTCTG
MB1035	CGGGATCCATGTCCAATAACAAACTGTCCATGAATGAGATCG
MB1036	CGATCGATCTCGAGTTTAACCACTTCACGGTCCTCAATGAGTTC
MB1122	GCGTTTAGCGTGGGCGGTTACCATTCATAA
MB1123	TTATGAATGGTAACCGCCCACGCTAAACGC
MB1124	CGATCGATCTCGAGCTCGGACATACTTAATTTTTTTTCGCTGTTTTTGC
MB1167	TTTTTTTTTTTTTTTTTTTTTTTTTGTGTCACTCACATAGCGTTC
MB1168	GAACGCTATGTGAGTGACACTTTTTTTTTTTTTTTTTTTTTTTT
MB1209	TACCGCGAAATTGATACTATTATCATTAACCAGATTTCGATGGTCCGTGCCGACTTGTTA
MB1210	TAACAAGTCGGCACGGACCATCGAAATCTGGTTAATGATAATAGTATCAATTTCGCGGTA
MB1219	CAAGGTCCAGGATCCAAAATGGAACTGATTAATCGGGCGATCGAAC
MB1220	AATCAGTTCCATTTTGGATCCTGGACCTTGGAACAAAACTTCC
TP-G4	TGGACCAGACCTAGCAGCTATGGGGGAGCTGGGGAAGGTGGGAATGTGA
IV-G4	GGGGAGGGGAAGGGGAGGGG

SUPPLEMENTAL FIGURES

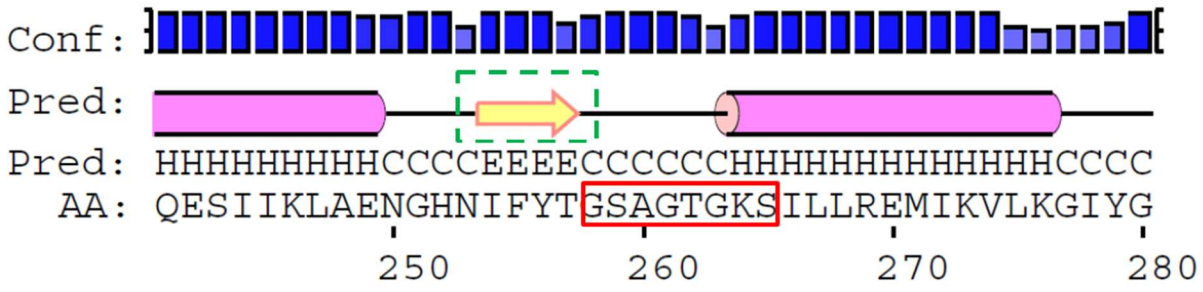

Figure S1. There is a predicted β-sheet upstream of the ScPif1 Walker A box. The secondary structure of ScPif1 was predicted using the PSIPRED Protein Sequence Analysis Workbench (http://bioinf.cs.ucl.ac.uk/psipred/). A portion of the predicted secondary structure (aa 241-280) is shown here, with the Walker A box sequence denoted by a red outline and the predicted β-sheet marked by the green dashed box.

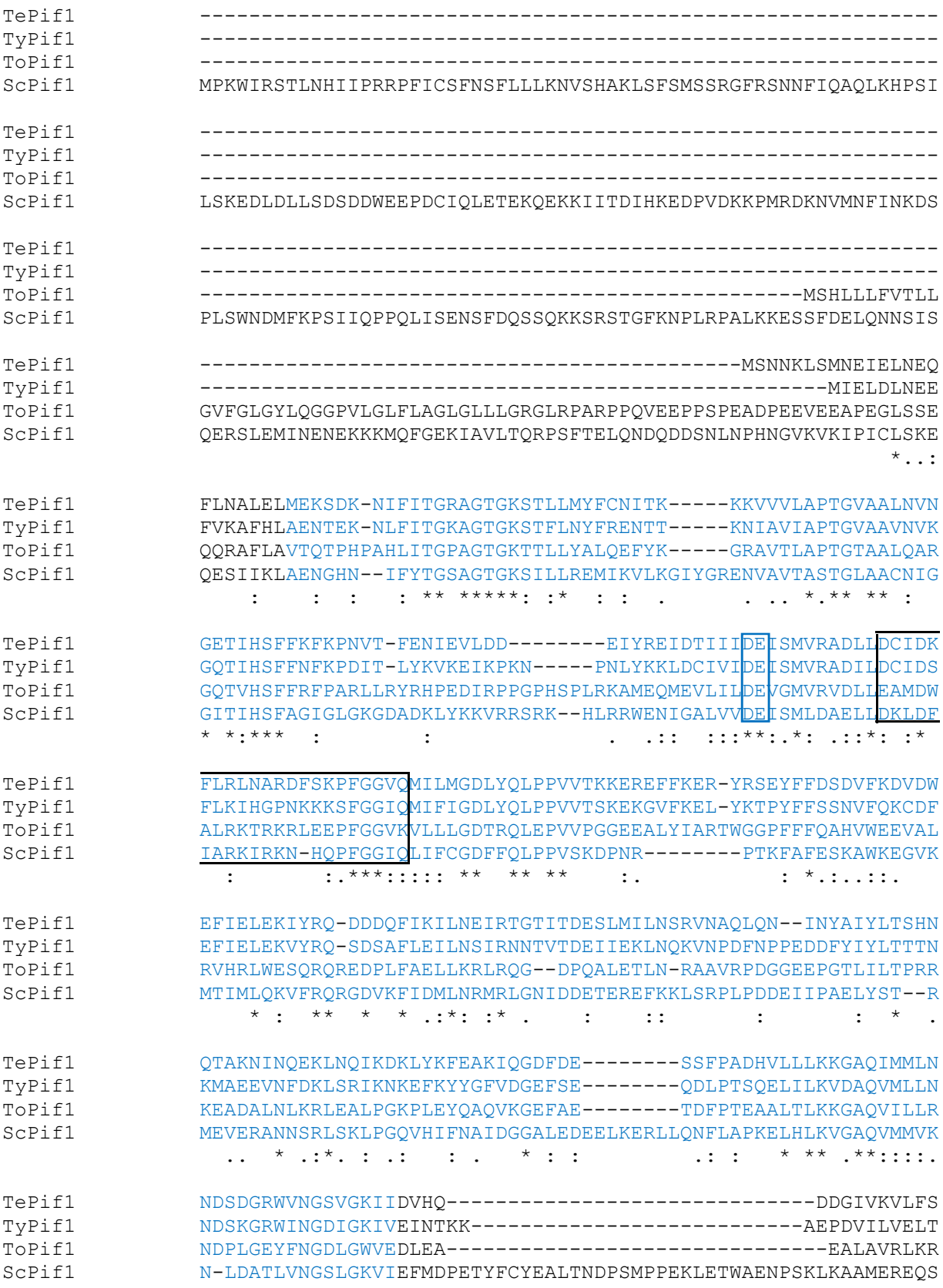
TePif1 ----------
TyPif1 ----------
ToPif1 ----------
ScPif1 MPKWIRSTLNHIIPRRPFICSFNSFLLLKNVSHAKLSFSMSSRGFRSNNFIQAQLKHPSI
TePif1 ----------
TyPif1 ----------
ToPif1 ----------
ScPif1 LSKEDLDLLSDSDDWEEPDCIQLETEKQEKKIITDIHKEDPVDKKPMRDKNVMNFINKDS
TePif1 ----------
TyPif1 ----------
ToPif1 ----------MSHLLLFVTLL
ScPif1 PLSWNDMFKPSIIQPPQLISENSFDQSSQKKSRSTGFKNPLRPALKKESSFDELQNNSIS
TePif1 ----------MSNNKLSMNEIELNEQ
TyPif1 ----------MIELDLNEE
ToPif1 GVFGLGYLQGGPVLGLFLAGLGLLLGRGLRPARPPQVEEPPSPEADPEEVEEAPEGLSSE
ScPif1 QERSLEMINENEKKKMQFGEKIAVLTQRPSFTELQNDQDDSNLNPHNGVKVKIPICLSKE
TePif1 FLNALELMEKSDK-NIFITGRAGTGKSTLLMYFCNITK-----KKVVVLAPTGVAALNVN
TyPif1 FVKAFHLAENTEK-NLFITGKAGTGKSTFLNYFRENTT-----KNIAVIAPTGVAAVNVK
ToPif1 QQRAFLAVTQTPHPAHLITGPAGTGKTTLLYALQEFYK-----GRAVTLAPTGTAALQAR
ScPif1 QESIIKLAENGHN--IFYTGSAGTGKSILLREMIKVLKGIYGRENVAVTASTGLAACNIG
TePif1 GETIHSFFKFKPNVT-FENIEVLDD--------EIYREIDTIIIDEISMVRADLLDCIDK
TyPif1 GQTIHSFFNFKPDIT-LYKVKEIKPKN-----PNLYKKLDCIVIDEISMVRADILDCIDS
ToPif1 GQTVHSFFRFPARLLRYRHPEDIRPPGPHSPLRKAMEQMEVLILDEVGMVRVDLLEAMDW
ScPif1 GITIHSFAGIGLGKGDADKLYKKVRRSRK--HLRRWENIGALVVDEISMLDAELLDKLDF
TePif1 FLRLNARDFSKPFGGVQMILMGDLYQLPPVVTKKEREFFKER-YRSEYFFDSDVFKDVDW
TyPif1 FLKIHGPNKKKSFGGIQMIFIGDLYQLPPVVTSKEKGVFKEL-YKTPYFFSSNVFQKCDF
ToPif1 ALRKTRKRLEEPFGGVKVLLLGDTRQLEPVVPGGEEALYIARTWGGPFFFQAHVWEEVAL
ScPif1 IARKIRKN-HQPFGGIQLIFCGDFFQLPPVSKDPNR--------PTKFAFESKAWKEGVK
TePif1 EFIELEKIYRQ-DDDQFIKILNEIRTGTITDESLMILNSRVNAQLQN--INYAIYLTSHN
TyPif1 EFIELEKVYRQ-SDSAFLEILNSIRNNTVTDEIIEKLNQKVNPDFNPPEDDFYIYLTTTN
ToPif1 RVHRLWESQRQREDPLFAELLKRLRQG--DPQALETLN-RAAVRPDGGEEPGTLILTPRR
ScPif1 MTIMLQKVFRQRGDVKFIDMLNRMRLGNIDDETEREFKKLSRPLPDDEIIPAELYST--R
TePif1 QTAKNINQEKLNQIKDKLYKFEAKIQGDFDE--------SSFPADHVLLLKKGAQIMMLN
TyPif1 KMAEEVNFDKLSRIKNKEFKYYGFVDGEFSE--------QDLPTSQELILKVDAQVMLLN
ToPif1 KEADALNLKRLEALPGKPLEYQAQVKGEFAE--------TDFPTEAALTLKKGAQVILLR
ScPif1 MEVERANNSRLSKLPGQVHIFNAIDGGALEDEELKERLLQNFLAPKELHLKVGAQVMMVK
TePif1 NDSDGRWVNGSVGKIIDVHQ------------------------------DDGIVKVLFS
TyPif1 NDSKGRWINGDIGKIVEINTKK--------------------------AEPDVILVELT
ToPif1 NDPLGEYFNGDLGWVEDLEA--------------------------------EALAVRLKR
ScPif1 N-LDATLVNGSLGKVIEFMDPETYFCYEALTNDPSMPPEKLETWAENPSKLKAAMEREQS

Figure S2. CLUSTAL alignment of the *Thermotoga elfii* (TePif1), *Thermodesulfovibrio yellowstonii* (TyPif1), *Thermus oshimai* (ToPif1), and *Saccharomyces cerevisiae* (ScPif1) PIF1 helicases. The helicase domain is colored blue, and the WYL domain in TePif1 and TyPif1 is green. The PIF1 family signature sequence[4] is outlined in black and spans two rows of the alignment. The catalytic residues in the Walker B box are outlined with a blue rectangle, and the WYL domain residues predicted to bind ligands[5] are outlined with green rectangles.

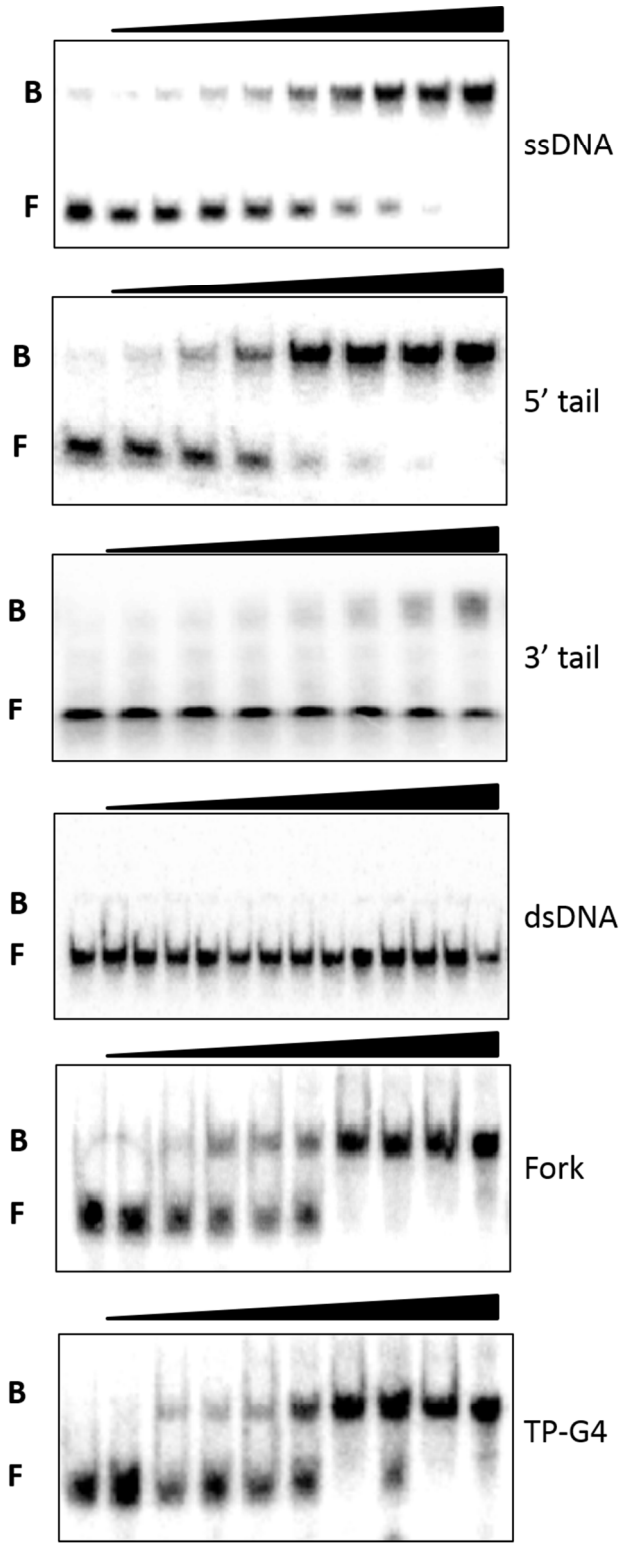

FIGURE S3. Representative gel shifts for the substrates shown in Figure 3. The positions of the free substrate (F) and bound substrate (B) are shown. The first lane in each gel image is the no-protein control. The black triangles denote the increasing concentration of TePif1 that was titrated into the binding reactions. The ssDNA substrate was oligo MB733 (45 nt long), the 5′ tail substrate was made by annealing oligos MB733 and MB820 (25-nt 5′ tail, 20-bp duplex), the 3′ tail substrate was made by annealing oligos MB821 and MB734 (25-nt 3′ tail, 20-bp duplex), the dsDNA substrate was made by annealing oligos MB733 and MB822 (45-bp, blunt ends), the fork substrate was made by annealing oligos MB733 and MB734 (25-nt 5′ and 3′ tails, 20-bp duplex), and the TP-G4 substrate was folded using oligo TP-G4 (see Table S1).

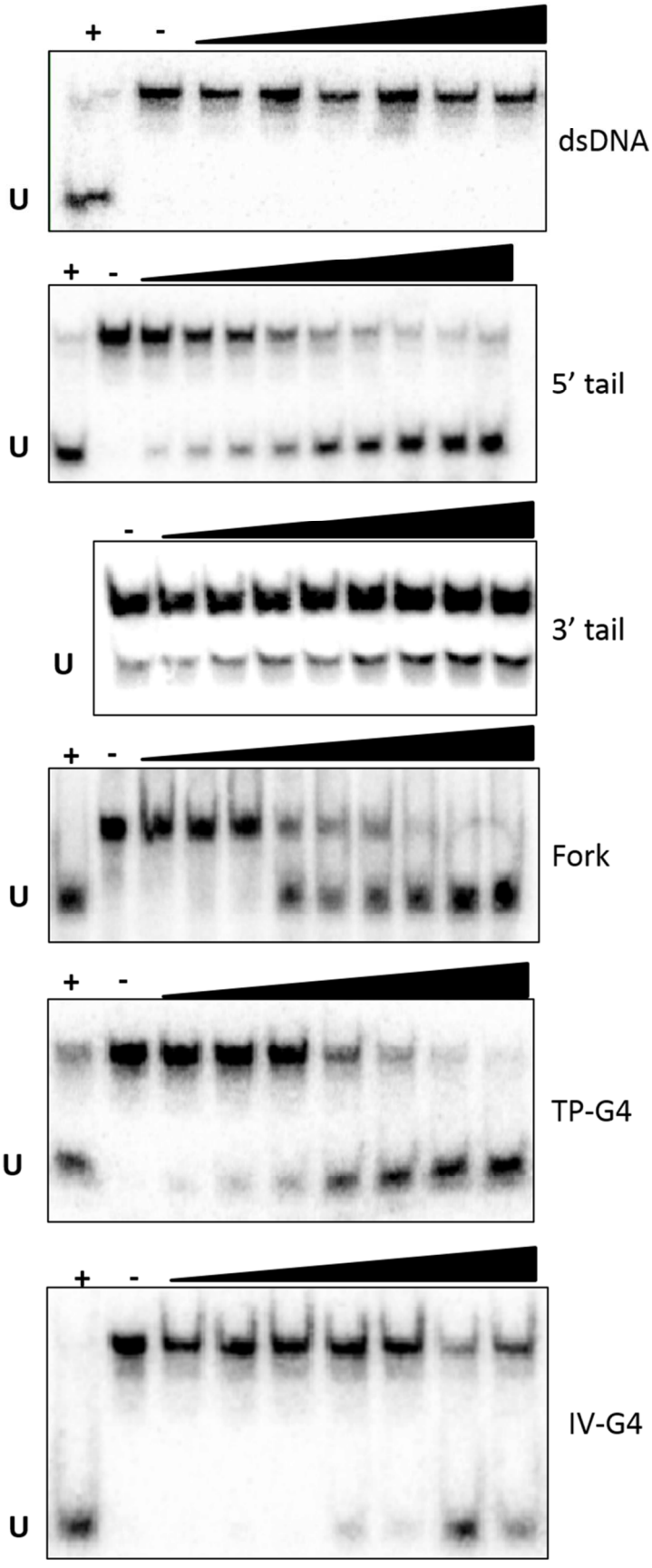

FIGURE S4. Representative helicase assays for the substrates shown in Figure 5. The position of the unwound substrate is marked with a U. In lanes marked with a +, the substrate was boiled before running it on the gel. In lanes marked with a -, the native substrate was run on the gel. The black triangles denote the increasing concentration of TePif1 that was titrated into the unwinding reactions. The dsDNA, 5′ tail, 3′ tail, fork, and TP-G4 substrates were identical to those used in Figures 3 and S3. The IV-G4 substrate was formed by *in vitro* folding of the IV-G4 oligo (Table S1).

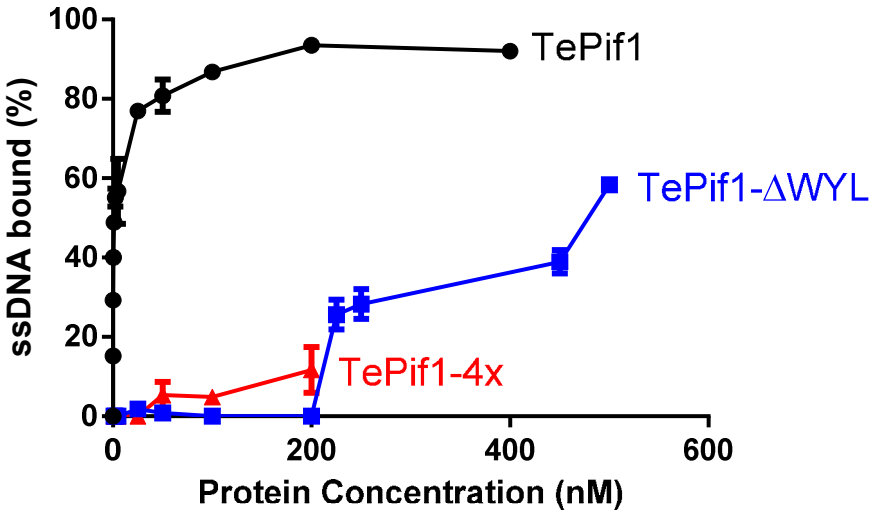

FIGURE S5. Deletion and mutation of the TePif1 WYL domain inhibits ssDNA binding relative to the wild-type protein.

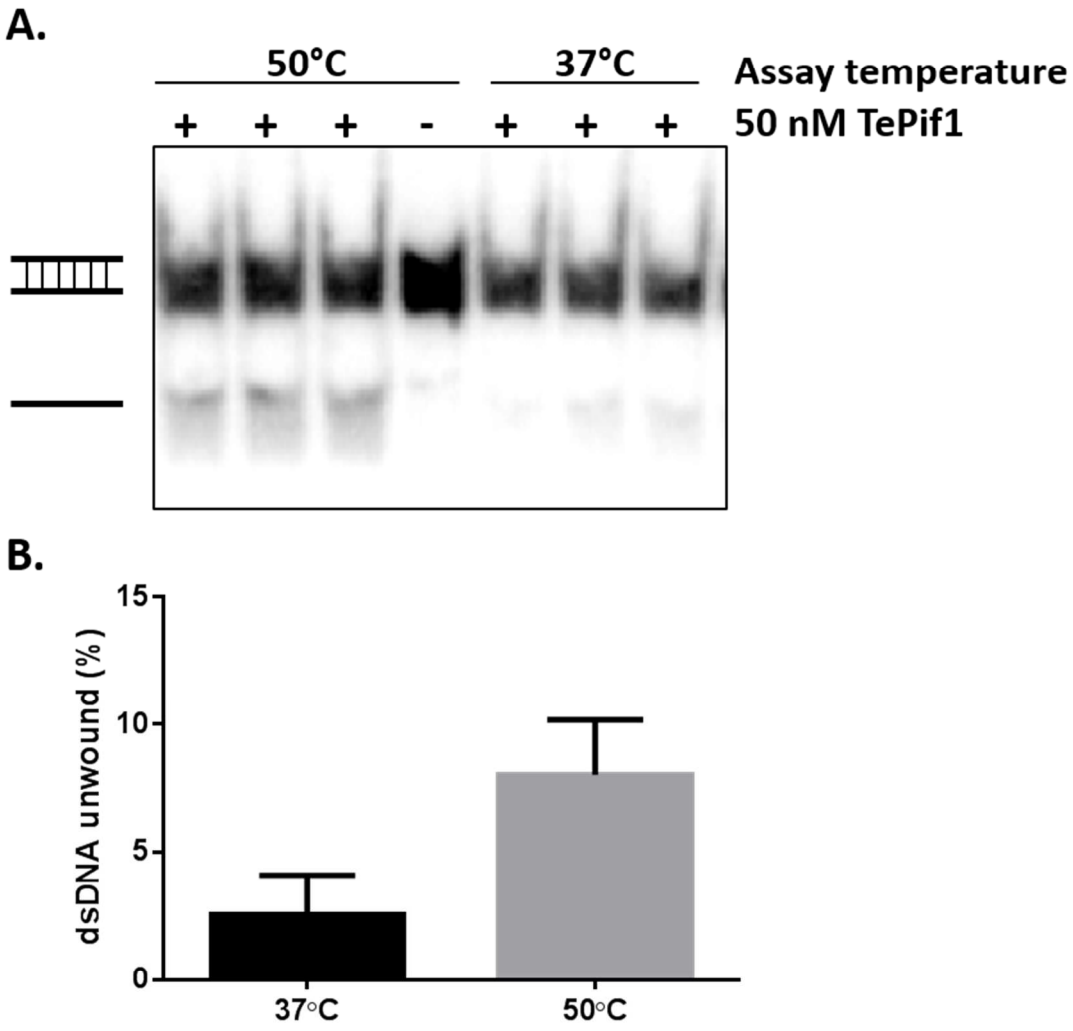

FIGURE S6. Increased reaction temperature leads to increased unwinding of blunt dsDNA. *A*, Helicase gel showing unwinding of the dsDNA substrate at 37 and 50°C. Triplicate reactions containing 50 nM TePif1 and 1 nM radiolabeled dsDNA are shown. The no-protein control (-) was performed at 50°C to also show that higher temperature incubation does not lead to appreciable thermal denaturation of the substrate. *B*, Quantification of the results from A.

SUPPLEMENTAL REFERENCES

[1] 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.

[2] Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases, *Nature methods 6*, 343-345.

[3] Kanter, D. M., and Kaplan, D. L. (2011) Sld2 binds to origin single-stranded DNA and stimulates DNA annealing, *Nucleic Acids Res 39*, 2580-2592.

[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] 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.