Supplementary Information

Strand-switching mechanism of Pif1 helicase induced by its collision with a G-quadruplex embedded in dsDNA.

Jessica Valle-Orero\textsuperscript{1,2,3,*}, Martin Rieu\textsuperscript{1,2}, Phong Lan Thao Tran\textsuperscript{4}, Alexandra Joubert\textsuperscript{4}, Saurabh Raj\textsuperscript{1,5}, Jean-François Allemand\textsuperscript{1,2}, Vincent Croquette\textsuperscript{1,2,6*}, Jean-Baptiste Boulé\textsuperscript{4,*}

\textsuperscript{1} Laboratoire de physique de L’École Normale Supérieure de Paris, CNRS, ENS, Université PSL, Sorbonne Université, Université de Paris, 75005 Paris, France.
\textsuperscript{2} Institut de Biologie de l’École Normale Supérieure de Paris (IBENS), École Normale Supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France.
\textsuperscript{3} The American University of Paris, 75007 Paris, France.
\textsuperscript{4} Structure et Instabilité des Génomes, Museum National d’Histoire Naturelle, INSERM, CNRS, Alliance Sorbonne Université, 75005 Paris, France.
\textsuperscript{5} Kusuma School of Biological Sciences, Indian Institute of Technology, Delhi, 110016, India.
\textsuperscript{6} ESPCI Paris, Université PSL, 75005 Paris, France.

*Corresponding authors: jean-baptiste.boule@mnhn.fr, vincent.croquette@phys.ens.fr, jessica.valle-orero@phys.ens.fr
**SUPPLEMENTARY TABLES**

| Oligonucleotide | Sequence |
|-----------------|----------|
| **LagG4**       | 5’TCTTTCTTCTTCTATAACCTCCACCAGTGTCCTTTTGGTGCTCTGGTGCTCCTTCGAATTTTTTTTGCAACTGCACAGATTGA CATAGCATGATAAGGGAGGTGGGAGGGGAGGGGAA GGATCGTACGTAGCATCGCTTGCACTGACTGGCCTCTCAGTGCAAGCGATGCTGTACGTACGACCTTCCCCAACCCCCCCCTCATGCTATGTCAATCGTGAC AGTTGCTTTTTTACCGGCGCTATTAGCTTTCCATACCAGCTGGCAACATCCATCATGATCCGCTACTCCCA-3’ |
| **LeadG4**      | 5’TCTTTCTTCTTCTATAACCTCCACCAGTGTCCTTTTGGTGCTCTGGTGCTCCTTCGAATTTTTTTTGCAACTGTCACGATTGA CATAGCATGATACCTTCCCCACCCTCCCCACCCTCCCCTAT CGTACGTACGATCGCTTGCACTGACTGGCCTCTCAGTGCAAGCGATGCTGTACGTACGACCTTCCCCAACCCCCCCCTCATGCTATGTCAATCGTGAC AGTTGCTTTTTTACCGGCGCTATTAGCTTTCCATACCAGCTGGCAACATCCATCATGATCCGCTACTCCCA-3’ |
| **Oli7**        | GCCCGCGC |
| **RNA oligo**   | GCA-UGA-UAC-CCU-CCC-CAC-CCU-CCC-CAC-CCU-CCC- |
| **LNA oligo**   | GCA-TGA-TAC-CTT-CCC-CAC-CCT-CCC-C56-668-666-6 For: 5 = lnA; 6 = lnC; 8 = lnT |

**Supplementary Table S1. Sequences of the hairpins used in this study.** The colored bases correspond to the following assemblies: blue for G4 motif, grey for the complementary sequence of the G4 motif, red for the loop, brown for the region complementary to OliBiotin, green for the region complementary to OliDBCO. Sequences of different Oligonucleotides in the measurements: Oli7 is the 7-base oligonucleotide complementary to the loop; RNA and LNA oligonucleotides are complementary to the G4 motif in the LeadG4 substrate.
| Name            | Sequence                                                                 |
|-----------------|--------------------------------------------------------------------------|
| OliDBCO         | ATT CGA AGA GCA CCA GAA AGA CCA AAA GAC ACG GTG AAG GAT TAG ACAGAA GAA GAC 3’DBC0 |
| OliBiotin       | DualBiotin TGG GAG TAG CGG ATG ATG GAT GTT GCC AGC TGG TAT GGAAGC TAA TAG CGC CGG T |
| OliLoop         | gcttGCACTGAGAGcgccggcCTCTCAGTGC                                         |
| Oligo5LeadG4    | GTC TTC TTC TGT CTA ATC CCT CAC CGT GTC TTT TGG TCT TTC TGG TGC TCTTCG AAT TTT TTT TGC AAC TGT CAC GAT TGA CAT AGC ATG ATA AGG GGAGGG TGG GGA GGG TGG GGA AGG ATC GTA CGT ACA GCA TC |
| Oligo3LeadG4    | aag cGA TGC TGT ACG TAC GAT CCT TCC CCA CCC TCC CCA CCC TCC CCT TATCAT GCT ATG TCA ATC GTG ACA GTT GCT TTT TTT ACC GGC GCT ATT AGC TTCCAT ACC AGC TGG CAA CAT CCA TCA TGA TCC GCT ACT CCC A |
| Oligo5LagG4     | GTC TTC TTC TGT CTA ATC CCT CAC CGT GTC TTT TGG TCT TTC TGG TGC TCTTCG AAT TTT TTT TGC AAC TGT CAC GAT TGA CAT AGC ATG ATA CCT TCC CCACCC TCC CCA CCC TCC CCT ATC GTA CGT ACA GCA TC |
| Oligo3LagG4     | aag cGA TGC TGT ACG TAC GAT AGG GGA GGG TGG GGA GGG TGG GGA AGGTAT CAT GCT ATG TCA ATC GTG ACA GTT GCT TTT TTT ACC GGC GCT ATT AGCTTC CAT ACC AGC TGG CAA CAT CCA TCA TGA TCC GCT ACT CCC A |

**Supplementary Table S2.** Sequences of oligonucleotides used to construct the hairpins of our assays. In the OliLoop the first 4 lowercase bases are the overhangs for the ligation; the second set of bases in lowercase are the ones composing the loop. On the other oligonucleotides, the lowercase bases correspond to overhangs.
Supplementary Figure S1. Hairpin Construct. Hairpin construct of 87 base pairs that incorporates the 26 bp c-Myc G4 oncogene promoter sequence (c-Myc Pu27) in the middle of the hairpin (blue). Two assays were designed, one containing the sequence in the strand before the loop (LagG4), and another after the loop (LeadG4). There are 7 nucleotides available for Pif1 to attach to the hairpin (Pif1 loading). The 5’-end of the hairpin is complementary to a 58-base 3’-DBCO modified oligonucleotide (OliDBCO-brown), and the 3’end to a 57-base oligonucleotide (OliBiotin - green). The loop is colored in red.

Supplementary Figure S2. Relationship between the total time spent at the G4 position (resolving time) and the number of strand switches. The linear relationship shows that the larger the resolving time, the larger the number of strand-switching events. This implies that strand-switching occurs with a constant rate (but with individual times distributed exponentially given this rate), which does not depend on a particular value of the resolving time, and thus strand-switching and resolution of the G4 are independent.
**Supplementary Figure S3. G4 lifetime.** A) The closure of the hairpin, when the force is reduced from 19 to 7 pN, is impaired by the presence of the G4 structure, observed as a long pause at a length of ~70 nm. When the force is increased and maintained at 11 pN, the blockage is continuous until the G4 structure unfold spontaneously, causing the fully closure of the hairpin (at 13500 s). B) Comparison of the lifetime of the G4 under the effects of the helicase (in log scale). If the helicase does not intervene, the G4 has an average lifetime of about 27400 ± 6300 s. Whilst, if Pif1 collides with the G4 structure embedded in the hairpin, the structure is resolved within an average time of 26.6 ± 6.7 s.
Supplementary Figure S4. Pif1 dynamics on a LagG4 assay. A) In the absence of G4 structure. When G4 is not formed in the dsDNA hairpin, Pif1 translocates through it without being stalled by the structure, and thus no blockage at the G4 position is observed. B) and C) In the presence of G4. Here are representative traces of the hairpin extension in which G4 has formed, as seen by a blockage at 7 pN (first cycles). When the force is kept constant at 11 pN (time > 200 s for A) and >3750 s for B)), and Pif1 binds to the hairpin, it translocates through the lagging strand until it gets stalled by the G4 structure, as seen by a blockage at around 60 nm. The helicase resolves the structure within tens of seconds, resuming translocation. Later, another helicase binds and translocates through the whole hairpin without being blocked by the G4, indicating that it was resolved in the previous event.
Supplementary Figure S5. Probability of resolving G4 for LeadG4 and LagG4 assays. G4 resolving probability is assessed by determining the proportion of traces that do not show a blockage on the second Pif1 event. No blockage was observed in 97% and in 87% of the traces for LagG4 and LeadG4 substrates respectively.

Supplementary Figure S6. Probability of Pif1 being stalled at the G4 position in the LeadG4 assay. Helicase stalling probability. When G4 is present (G4 structure bar) within the cycles (as observed by a blockage at 7 pN) Pif1 gets blocked at the G4 position while translocating onto the leading strand 85% of the time. However, if G4 was not present (no G4 bar), a blockage was only observed in 5% of the traces.
Supplementary Figure S7. **A)-B)** Representative traces of the collision of Pif1 with G4 in the LeadG4 assay under 1 mM ATP and 50 µM ATP respectively. **C)** Comparison of the unzipping/rezipping speeds during strand switching under both ATP concentrations. Speed values are the average of all the measured transitions given in bp/s, and the errors are standard deviations.
Supplementary Figure S8. Pif1 dynamics during its interaction with a DNA:LNA heteroduplex.  **A)** Representative trace of Pif1. A characteristic force-extension trace showing the blockage at 7 pN due to the binding of the LNA oligo to its complementary hairpin sequence on the leading strand. At 11 pN, Pif1 is loaded into the solution. It takes several helicases to interact with the complex until the oligo is removed, as observed by the two burst of unzipping/zipping events, tens of seconds apart from each other. **B)** Hairpin Sketch. The sketch shows how Pif1 interacts with the oligo within a DNA hairpin context (1), and successfully removes it (2). **C)** DNA sequence. DNA sequence highlighting the region where the LNA oligo, composed of 9 LNA (green) and 25 DNA bases (turquoise), is complementary to the G4 sequence (dark blue), and thus will form a hybrid complex.
Supplementary Figure S9. Pif1 dynamics during its interaction with a DNA:RNA heteroduplex.  
A) Representative trace of Pif1. A characteristic force-extension trace showing the blockage (at 7 pN) by the RNA oligo when binds to its complementary hairpin sequence on the leading strand. At 11 pN, Pif1 is loaded into the solution, and after one translocation through the hairpin it is able to remove the RNA oligo.  
B) Hairpin Sketch. The sketches show how Pif1 interacts with the oligo in the leading strand of the DNA hairpin (1), and successfully removes it (2).  
C) DNA sequence. DNA sequence highlighting the region where the RNA oligo is complementary to the G4 sequence (dark blue) within the hairpin and thus will form a hybrid complex (purple).
Supplementary Figure S10. Pif1 interacting with β^A-ori structure formed on the leading strand. A) and B) show characteristic traces of the interaction between Pif1 and G4 formed from β^A-ori sequence. Strand switching at the G4 position is observed in both cases.
Supplementary Figure S11. Comparison of the stalling time in the LeadG4 complex at two different Pif1 concentrations: 6 and 60 nM. Here, the stalling time in this substrate was analyzed separately for both concentrations, to show that the differences are insignificant. Again, the data shows a single exponential characterized by a time constant $\tau'_{G4} (6 \text{ nM}) = 0.58 \pm 0.06 \text{ s (blue)}$ and $\tau'_{G4} (60 \text{ nM}) = 0.51 \pm 0.07 \text{ s (red)}$. Errors were derived from the fitting procedure using IgorPro.
Supplementary Figure S12. Study of the dynamics of Pif1 at a concentration of 60 nM while interacting with a DNA hairpin containing an embedded G4 structure. A) Representative trace of Pif1 when it interacts with a LagG4 assay. A characteristic blockage at the G4 position is observed on the first helicase event on the lagging strand. Translocation is resumed and the hairpin closes. Due to the high concentration the rate of a helicase binding to the hairpin increases, and the frequency of events is highly increased. B) Representative trace of Pif1 when interacts with a LeadG4 assay. A characteristic blockage and strand switching are observed at the G4 position, when the structure forms on the lagging strand. Once Pif1 resolves the structure and the hairpin recloses, multiple helicase events are subsequently observed within a time window of 100s. C) Distribution of resolving time of G4 at 60nM Pif1 concentration. The resolving time shows a single exponential characterized by a time constant $\tau_{\text{LagG4}}$ (60 nM) = 14.6 ± 3.1 s and $\tau_{\text{LeadG4}}$ (60 nM) = 5.90 ± 1.05 s.
Supplementary Figure S13. Comparison of patrolling and strand switching. If Pif1 starts patrolling at the loop, and sequesters the length between the loop and the G4 structure in the leading strand, we would observe a shortening of the hairpin of $\Delta L = 32$ nt (which is the distance between loop-G). In contrast, when Pif1 undergoes strand switching between the loop and the G4 structure, we observe a shortening of the hairpin by $2\Delta L = 62$ nt. This distance corresponds to the annealing of the base pairs between the loop and the G4 (or partial rezipping of the hairpin), resulting in double the decrease in length.
SUPPLEMENTARY METHODS

Conversion procedure to transform nm into bp: shown for LeadG4 substrate

Pif1 translocates along the hairpin from the lagging strand to the loop causing its total opening at a constant force of 11 pN. From our traces we measure the total extension of the hairpin $\Delta L_{\text{TOTAL}}$ in units of nm. We then computed a histogram of these extensions and fit a Gaussian function, obtaining a mean total extension of 93.0 nm with a standard deviation of 6.1 nm (Supplementary Figure S12). This length corresponds to a translocation of around 90 base pairs. Thus, the ratio $90 \text{ bp}/93 \text{ nm}$ provides our conversion factor used to compute the position of Pif1 throughout the hairpin from nm to bp.

Histogram of the total extension in nm units. Histogram of the total extension of our LeadG4 assay at 11 pN (blue points), and the Gaussian fit (red) defining the peak at 93 nm and a spread of 6.1 nm.

We also measured the stalling position (G4 extension) in nm (step 5 in Figure 2A&B), and apply the conversion factor ($90/93$) to obtained the G4 extensions in bp units. And again, we computed a histogram for those G4 extensions and fit it with a Gaussian (Supplementary Figure S13). This corresponds to the average position in the hairpin at which Pif1 is stalled by the G4 structure (stalling position). In this assay, Pif1 encounters the G4 during the rezipping of the hairpin, at a stage where $49.9 \pm 4.8$ base pairs (data corresponds to mean and standard deviation) remain still opened (hairpin extension).

Histogram of the G4 position in base pair units. Histogram of the converted extensions of the G4 positions (blue points), and the Gaussian fit (red) defining the peak at 49.9 bp and a spread of 4.8 bp.
At the constant force of 11 pN, Pif1 translocates through the lagging strand (unzipping) all the way to the loop and into the leading strand (rezipping). We then measured separately the unzipping \(v_{unz}\) and rezipping \(v_z\) velocities of about 122 and 101 transitions, respectively, by fitting the position of the bead with a linear function. Our results showed a mean and standard deviation of \(v_{unz} = 97.8 \pm 7.6 \text{ nm/s}\) and \(v_z = 116.8 \pm 14.1 \text{ nm/s}\). We used our conversion factor determined before to provide the speed in bp/s. From the spread on the histograms of the total extension we determined an error in the conversion factor of about 6 to 8%, which probably arises from the error in the force and on the zero position. Thus, this % was added to the error on the speeds before converting them to bp/s.

This method was applied to all the speeds measured on our other substrates: \textit{LagG4}, \textit{LNA}:DNA, and \textit{RNA}:DNA. Our values used for the conversion and the final velocities in both nm/s and bp/s are shown in Supplementary Table S3.

### Comparing translocation speeds for the different substrates.

Pif1 unwinds the double-stranded DNA along the lagging strand (\textit{LagG4} assay) at a constant speed of \(v_{unz1} = 108.37 \pm 18.19 \text{ bp/s}\) before interacting with the G4 structure, and resumes translocation after being stalled by the G4 at a speed of \(v_{unz2} = 95 \pm 20.13 \text{ bp/s}\).

We also measured both the unzipping and rezipping velocities during strand-switching transitions on the \textit{LeadG4} substrate, and obtained very similar values within one standard deviation, of \(v_{unz} = 94.64 \pm 9.98 \text{ bp/s}\) and \(v_z = 113.05 \pm 16.07 \text{ bp/s}\). Hence, these measurements imply that both events are the result of the same helicase process (Supplementary Table S3).

We also analyzed the unzipping and rezipping speeds of Pif1 during strand switching in the presence of LNA and RNA oligonucleotides, and found for LNA a translocation speed of \(v_{unz} = 108.22 \pm 25.22 \text{ bp/s}\) and \(v_z = 108.58 \pm 20.77 \text{ bp/s}\), and for RNA of \(v_{unz} = 113.39 \pm 21.91 \text{ bp/s}\) and \(v_z = 86.81 \pm 32.66 \text{ bp/s}\) for unzipping and rezipping respectively. These velocities are very similar, within one standard deviation, to the values measured on our \textit{LeadG4} assay.

| Substrate | Total Ext (nm) | Conversion bp/nm | Stalling position (bp) | Speed Unzipping \(v_{unz}\) in nm/s | Speed Unzipping \(v_{unz}\) in bp/s | Speed zipping \(v_z\) in nm/s | Speed zipping \(v_z\) in bp/s |
|-----------|----------------|------------------|------------------------|-------------------------------------|------------------------------------|-------------------------------|-------------------------------|
| LagG4     | 98.06 ± 8.2    | 90/98.1          | 46.3 ± 3.3             | 118.1 ± 17.4                       | 108.4 ± 18.2                       | NA                            | NA                            |
| LeadG4    | 93.01 ± 6.11   | 90/93.01         | 49.9 ± 4.8             | 97.8 ± 7.6                         | 94.6 ± 10.0                        | 116.8 ± 14.1                   | 113.1 ± 16.1                   |
| LNA:DNA   | 98.71 ± 7.30   | 90/98.71         | 54.3 ± 5.6             | 118.7 ± 23.6                       | 108.2 ± 25.2                       | 119.1 ± 18.8                   | 108.6 ± 20.8                   |
| RNA:DNA   | 94.49 ± 8.05   | 90/94.49         | 50.7 ± 3.9             | 119.1 ± 19.4                       | 113.4 ± 21.9                       | 91.1 ± 31.7                    | 86.8 ± 32.7                    |

**Supplementary Table S3. Translocation speeds.** Measurements on the total extension of the molecule (in nm) when the helicase unwinds the full hairpin (i.e. 90 base pairs). This extension was used to obtain a conversion factor to transform all Pif1 translocation speeds from nm/s to bp/s, for all of our substrates. In the \textit{LagG4} substrate we obtained two unzipping speeds: one before Pif1 interacts with G4, and a second one when it resumes translocation after resolving the G4 (indicated with a *). The errors on these measurements correspond to the standard deviation of the distribution.