DNA replication through G-quadruplex motifs is promoted by the *S. cerevisiae* Pif1 DNA helicase

Katrin Paeschke[1], **John A. Capra**[2], and **Virginia A. Zakian**[1,3]

[1] Department of Molecular Biology, Princeton University Princeton, NJ 08544-1014

[2] Gladstone Institutes, University of California, San Francisco, San Francisco, CA 94158

Abstract

G-quadruplex (G4) DNA structures are extremely stable four-stranded secondary structures held together by non-canonical G-G base pairs. Genome-wide chromatin immuno-precipitation was used to determine the *in vivo* binding sites of the multi-functional *S. cerevisiae* Pif1 DNA helicase, a potent unwinder of G4 structures *in vitro*. G4 motifs were a significant subset of the high confidence Pif1 binding sites. Replication slowed in the vicinity of these motifs, and they were prone to breakage in Pif1-deficient cells, while non G4 Pif1 binding sites did not show this behavior. Introducing many copies of G4 motifs caused slow growth in replication stressed Pif1 deficient cells that was relieved by spontaneous mutations that eliminated their ability to form G4 structures, bind Pif1, slow DNA replication and stimulate DNA breakage. These data suggest that G4 structures form *in vivo* and that they are resolved by Pif1 to prevent replication fork stalling and DNA breakage.

Introduction

G-quadruplex (G4) structures are four-stranded structures held together by non-canonical Hoogsteen G-G base pairs (Lipps and Rhodes, 2009). A stretch of single-stranded DNA can form an intra-molecular G4 structure if it contains four tracts of two or more guanines (G-tracts) separated by random sequence segments of variable length. Computational approaches have identified genomic regions with the ability to form G4 structures (called G4 motifs) (Huppert, 2008). There are more than 500 G4 motifs in the *S. cerevisiae* nuclear genome, a number that excludes the large number of G4 motifs in the repetitive telomeric and rDNAs (Capra et al., 2010; Hershman et al., 2008). Using both gel mobility shifts and circular dichroism, five of five tested non-telomeric *S. cerevisiae* G4 motifs readily formed G4 structures *in vitro* (Capra et al., 2010), demonstrating that the algorithm does indeed identify motifs that are able to form G4 structures. Of course, the ability of a G4 motif to form a G4 secondary structure *in vitro* does not mean that it forms this structure *in vivo*.

[3]Corresponding author: vzakian@princeton.edu; +1-609-258-6770.

Throughout this paper, we use G4 motif to refer to a sequence that has the potential to form a G4 structure.

Computational analyses suggest that G4 motifs have important functions. They are several times more abundant in human and *S. cerevisiae* than expected from the GC content of these genomes (Huppert, 2010) (JAC, unpublished results). In both organisms, G4 motifs are evolutionarily conserved. Moreover, in human and *S. cerevisiae*, the pattern of conservation suggests that the ability to form G4 structures is under evolutionary constraints (Capra et al., 2010; Nakken et al., 2009). In both yeast and mammals, G4 motifs are enriched in telomeric and rDNA, as well as at transcriptional regulatory sites and at preferred mitotic and meiotic double strand break (DSB) sites (Huppert, 2010). G4 structures also form in RNA, where their presence can affect RNA stability or translation (Huppert, 2008). Together, these data suggest that G4 structures regulate telomere metabolism, genomic transcription, meiotic DSB formation/processing, and/or mRNA transactions. However, if G4 structures do form *in vivo*, as these genome-wide analyses suggest, due to their high thermal stability, their presence is predicted to cause problems for replication (Lipps and Rhodes, 2009).

The *S. cerevisiae* helicase Pif1 is a member of a highly conserved 5′-3′ DNA helicase family found in nearly all eukaryotes (Bochman et al., 2010). Pif1 was first identified due to its positive role in maintenance of mitochondrial DNA (Foury and Dyck, 1985). However, there are two Pif1 isoforms, one targeted to mitochondria and one to nuclei (Schulz and Zakian, 1994; Zhou et al., 2000). Nuclear Pif1 is a multi-functional protein. Its best-studied role is its negative regulation of telomerase where it removes telomerase from DNA ends (Boule et al., 2005; Myung et al., 2001; Schulz and Zakian, 1994; Zhou et al., 2000). Pif1 also acts in a locus specific manner within the ribosomal DNA (rDNA) where it helps maintain the replication fork barrier (Ivessa et al., 2000). In addition, Pif1 has more global functions in chromosomal DNA replication as it cooperates with DNA polymerase and Dna2 in Okazaki fragment maturation. Finally, Pif1 has repair functions as it localizes to DNA damage foci and suppresses accumulation of toxic recombination intermediates that are generated by the Sgs1 helicase in *top3* cells (Bochman et al., 2010).

Both human (Sanders, 2010) and *S. cerevisiae* (Ribeyre et al., 2009) Pif1 helicases bind and unwind G4 structures *in vitro*. Although several other DNA helicases are also able to unwind G4 structures, including some whose mutation leads to inherited human diseases (WRN, BLM, and FANC-J) (Paeschke et al., 2010), the *S. cerevisiae* Pif1 DNA helicase is a particularly potent unwinder of G4 structures, as it unwinds G4 structures even under single cycle conditions. In side-by-side comparison, Pif1 is ~70 times more active on G4 structures than WRN, a human RecQ helicase (KP and VAZ, in preparation). Genetic evidence suggests that helicase mediated unwinding of G4 structures is important as mutations in the human FANC-J helicase or its *Caenorhabditis elegans* homolog Dog-1 result in instability of G-rich DNA (Cheung et al., 2002; London et al., 2008). Likewise a sequence from the human genome that can form G4 structures *in vitro* is prone to deletion when inserted into the genome of *pif1* mutant yeast cells (Ribeyre et al., 2009).

Although eukaryotic genomes are replete with G4 motifs, there is virtually no mechanistic data on their impact on replication and integrity of eukaryotic chromosomes. Here we

provide insight into these questions by exploring the consequences of the absence of *S. cerevisiae* Pif1 on G4 motifs *in vivo.* Using genome-wide approaches, we find that G4 motifs overlap a subset of the high confidence Pif1 binding sites. In the absence of Pif1, replication forks paused and breakage was more frequent near G4 motifs that bound Pif1 but not at non G4 Pif1 binding sites. When G4 motifs were present in high copy numbers in replication stressed *pif1* mutant cells, they caused slow growth, which was relieved by mutations that eliminated their ability to form G4 structures. Moreover, these naturally generated mutant G4 motifs no longer bound Pif1, no longer slowed DNA replication and no longer stimulated DNA breakage in *pif1* cells. Based on these findings, we propose that G4 structures form *in vivo*, and their presence can impede replication and cause chromosome breakage when they are not resolved by Pif1.

Results

System and methods used to identify Pif1 sites of action

To determine sites of Pif1 function, we carried out chromatin immuno-precipitation (ChIP) using a strain expressing epitope tagged Pif1. In previous experiments, Pif1 gave relatively low ChIP signals at telomeres and rDNA (Ivessa et al., 2000; Zhou et al., 2000). To improve the Pif1 ChIP signal, we tagged Pif1-K264A, a mutant version of Pif1 in which the invariant lysine in the Walker A box is mutated to alanine. Although Pif1-K264A is a null mutant *in vivo* and has no ATPase/helicase activity *in vitro*, it binds single-stranded DNA as well as wild-type (WT) Pif1 (Boule et al., 2005; Zhou et al., 2000). Because Pif1-K264A cannot unwind DNA, we reasoned that it would be trapped at its sites of binding and allow more accurate identification of Pif1 binding sites.

Asynchronously growing log phase cultures expressing Myc-tagged Pif1-K264A or WT Pif1 were processed for ChIP and DNA labeling and then hybridized to Agilent whole genome arrays. In all of the genome-wide ChIP experiments, we excluded rDNA and telomeres from the analysis as these multi-copy sequences are difficult to analyze with the same methods used for single copy DNA.

Pif1-K264A binding sites include G4 motifs

We identified 1123 regions in chromosomal DNA that were Pif1-K264A associated in asynchronous cells (see Table S1 for complete list of Pif1-K264A binding sites; Fig. 1A-D for the pattern of Pif1-K264A binding within four ~10 kb regions). Pif1-K264A showed a general preference for G-rich sequences. We define G-rich as a sequence with a GC content greater than the average GC-content of the *S. cerevisiae* genome (ie.,>38% GC). The Pif1-K264A binding sites had an average GC-content of 42.3%, which is significantly higher than expected (p < 0.001).

When considering whether Pif1-K264A binding (or later, high DNA Pol2 occupancy) overlapped a G4 site, we considered a site positive if it was within 500 bp of the peak of Pif1-K264A (or DNA Pol2) binding. This window was used due to the 500 bp average size of DNA fragments in the immuno-precipitate and the small size of G4 motifs, which span an average of only 54 bps. Using this metric, Pif1-K264A binding sites were significantly

associated with G4 motifs (q < 0.001) with 128 of the 1123 peaks (11%) overlapping a G4 motif. Thus, 25% (138 of 558) of the G4 motifs were Pif1-K264A associated (Fig. 1A,B). The number of G4 motifs and overlapping binding sites are not equal, because several binding sites overlapped more than one G4 motif. If no window was used, G4 motifs were still significantly Pif1-K264A associated (72 overlapping motifs, q < 0.001).

Other genomic features that were significantly Pif1-K264A associated included highly transcribed genes (8.5% of binding sites, 31% of highly transcribed genes), mitotic DSB sites (45% of binding sites, 40% of mitotic DSB sites), and meiotic DSB sites (38% of binding sites, 18% of meiotic DSBs). Many of these binding sites fall into more than one category. For example, 54% (q < 0.001) of the Pif1-K264A associated G4 motifs overlapped with preferred mitotic DSB sites. Although these binding sites are also of interest, we focus here on the functional significance of Pif1 binding to G4 motifs.

To confirm the associations of Pif1-K264A sites identified by the arrays, we performed standard ChIP followed by quantitative real-time PCR (ChIP/qPCR) on 16 sites that showed high Pif1-K264A binding in the arrays (Fig. 1E). Six of these sites were G4 motifs (Fig1E, black). (Table S2 has sequences of G4 motifs analyzed here). Five were G-rich non G4 sites (GR, white) and five were not G-rich (ie., <38% GC; NG, white). We also verified negative sites by using ChIP/qPCR on seven sites that did not bind Pif1-K264A in the array studies (nP, grey) and examined two sites that were expected to have high Pif1-K264A binding but were not analyzed on arrays, the right telomere of chromosome VI-R (Tel) and rDNA (rD) (controls; Fig. 1E, striped). Each of the 16 sites that showed high Pif1-K264A binding in the arrays had robust Pif1-K264A binding by qPCR with enrichments ranging from 4.3 to 7.3 (six G4 motifs) and 2.5 to 4.8 (ten GR and NG sites). As anticipated, binding to all non Pif1 binding regions was low (from 0.5 to 1.6 fold) while binding to the VI-R telomere and rDNA was particularly high (7.2 and 11.7, respectively; these high values are not due to the repetitive nature of the two sequences as signals were normalized to input DNA). Thus, results from qPCR analysis supported the genome-wide ChIP data.

We also carried out ChIP-microarray studies using WT Pif1 and identified 1584 binding sites (Table S1). As expected for a catalytically active helicase that can move away from its binding sites, Pif1 binding was more delocalized than for Pif1-K264A (Fig. 1A–D, WT Pif1 binding in grey). Nonetheless, the spots identified with the two proteins overlapped significantly (565 WT Pif1 sites overlapped Pif1-K264A binding sites; q < 0.001) and had similar associations with genomic features and G4 motifs (data not shown).

Pif1 associates with G4 motifs in late S phase

Next we determined the timing of Pif1 binding to G4 motifs. We synchronized cells expressing both Myc-tagged Pif1 and HA-tagged DNA Pol2, the catalytic subunit of the leading strand DNA polymerase and analyzed samples throughout the synchronous S phase for both DNA Pol2 (Fig. 2A,C) and Pif1 binding (Fig. 2B,D). A similar experiment was done using a strain expressing Myc-tagged Pif1-K264A with comparable results (Figure S1A). At each time point, we monitored Pif1 and DNA Pol2 binding to eight strong Pif1 binding sites: five at G4 motifs (Fig. 2A,B) and three at non G4 sites (Fig. 2C,D). We also

tested two sites with low Pif1 binding in the arrays that were 6 kb to the left and 4 kb to the right of Chr $XIII_{G4}$ (Fig. 2A,B).

For all sites, the time point with the highest DNA Pol2 binding was considered the time of semi-conservative replication for that site. By this criterion, four of the G4 motifs replicated in mid S phase (75 min after release from G1 arrest) while Chr IV_{G4} replicated earlier in S phase (55 min) (Fig. 2A). The sites 6 kb to the left and 4 kb to the right of the Chr $XIII_{G4}$ site replicated at 55 and 75–85 min, consistent with the replication fork moving left to right through the 10 kb region containing the Chr $XIII_{G4}$ motif (Fig. 2A). The three non G4 Pif1 binding sites replicated at ~50 (IX_{NG}) and ~55 min (XV_{GR} and $PGK1_{NG}$) (Fig. 2C).

The simplest prediction for a model in which Pif1 promotes fork progression by unwinding G4 structures is that Pif1 binding to G4 motifs occurs before or at about the same time as DNA replication of the motif. Yet, at all of the G4 sites, maximal Pif1 binding occurred at 85 min, after DNA replication of the site (Fig. 2B). Pif1 association with the G4 motifs in late S phase was high, seven (Chr IX_{G4} and XI_{G4}) to over 14 fold (Chr $XIII_{G4}$) enrichment. However, Pif1 binding to G4 motifs was also significant earlier in the cell cycle. For example, Pif1 binding at each of the five G4 motifs was significantly higher than its binding to both of the negative control sites (sites 6 kb and 4 kb to either side of $XIII_{G4}$; $p < 0.01$) at all time points except for the G4 motif on chromosome IV at 70 min (the difference was significant only when compared to the 6 kb control site). The level of enrichment varied from time point to time point and among the different G4 motifs but was never higher than three, except at the 85 min time point. Patterns of Pif1 binding to non G4 binding sites was similar to that seen at G4 motifs except that the difference in the levels of binding between the 85 min time point and early time points was not as dramatic (Fig. 2D). Thus, at both G4 and non G4 sites, Pif1 binding occurred throughout the cell cycle but was highest after semi-conservative replication of the site.

DNA Pol2 levels are higher at G4 motifs in the absence of Pif1

Given the stability of G4 structures, we anticipated that if these structures form *in vivo*, they would slow DNA replication. High occupancy by DNA Pol2 has been used to identify sites where replication forks move more slowly than elsewhere in the genome (Azvolinsky et al., 2009). In order to test whether Pif1 activity at specific sites affects their replication, we epitope tagged DNA Pol2 and determined if reduced Pif1 affects DNA Pol2 association genome-wide. We used the partial loss of function *pif1-m2* allele (rather than *pif1Δ*) because respiratory competent *pif1-m2* cells have a near WT growth rate (Schulz and Zakian, 1994). We reasoned that sites that have higher DNA Pol2 binding in *pif1-m2* compared to WT cells would be sites where replication is Pif1-dependent. As another control, we determined replication timing of Pol2 binding sites, including several G4 motifs, in synchronized *pif1-m2* cells and found that replication timing was not changed in the absence of Pif1 (compare Figure S1B to data for WT cells in Fig. 2A).

DNA Pol2 associated DNA was prepared from WT and *pif1-m2* asynchronous cells and hybridized to genome-wide microarrays. Using the same criteria as for Pif1-K264A, we found 1484 (WT) and 1398 (*pif1-m2*) sites with high DNA Pol2 occupancy (Table S3; Fig. 3A–D for patterns of DNA Pol2 binding in *pif1-m2* (black) and WT (grey) cells through four

10 kb regions). There was significant overlap between the high DNA Pol2 genome wide association data in WT and *pif1-m2* cells with 833 sites overlapping in the two strains (q < 0.001). That is, most replication pause sites were seen in both strains.

Despite this overall similarity, high DNA Pol2 binding was not associated with G4 motifs in WT cells (q = 0.97), while in *pif1-m2* cells, there was a strong overlap of G4 motifs with high DNA Pol2 binding (q < 0.001). Of the 1398 high DNA Pol2 occupancy sites in *pif1-m2* cells, 149 (11%) overlapped one or more G4 motifs. Thus, 158 of the 558 (28%) G4 motifs had high DNA Pol2 occupancy in *pif1-m2* cells. Moreover, those G4 motifs identified as having high DNA Pol2 binding in *pif1-m2* cells were significantly more likely than other G4 sites to have high Pif1-K264A binding. Of the 158 G4 motifs with high DNA Pol2 association in *pif1-m2* cells, 77 overlapped regions of strong Pif1-K264A binding (49%; q < 0.001). In contrast, most of the G4 motifs that did not have high DNA Pol2 association in *pif1-m2* cells also did not overlap a high Pif1-K264A site (15% or 61 of 394 such sites; q = 0.47) (Table S3). The chromosomal positions of G4 motifs with high Pif1-K264A binding, high DNA Pol2 occupancy in *pif1-m2* cells, and the overlap of these two classes is summarized in Fig. 7A.

In marked contrast to G4 motifs, DNA Pol2 occupancy at non G4 Pif1 binding sites was not sensitive to Pif1 as non G4 Pif1 binding sites were similarly likely to have high DNA Pol2 occupancy in WT (0.546) and *pif1-m2* (0.527) cells. That is, using high DNA Pol2 occupancy as a monitor of replication fork pausing, Pif1-K264A was often associated with sites of pausing, but pausing at most of these sites was not increased in the absence of Pif1, unless the site was also a G4 motif.

To validate the microarray data, we examined DNA Pol2 binding by ChIP/qPCR in both WT and *pif1-m2* cells at 16 high Pif1 binding sites (six G4 motifs, ten non G4 sites, of which five were G-rich and five were not) as well as at seven regions that did not bind Pif1 (Fig. 3E). DNA Pol2 occupancy was about twice as high at each of the six G4 motifs in *pif1-m2* (black) compared to WT (grey) cells, and these differences were highly significant (p < 0.001) (Fig 3E, fold enrichments above columns). Similarly, about two fold higher DNA Pol2 binding was seen at G4 motifs in synchronous *pif1-m2* versus synchronous WT cultures (Figure S1B). At the 17 non G4 sites, DNA Pol2 binding was not affected by the absence of Pif1, regardless of their GC content or their Pif1 binding status (Fig. 3E). These data argue that by the criterion of DNA Pol2 occupancy, absence of Pif1 result in replication fork slowing specifically near those G4 motifs to which Pif1 normally binds.

Replication fork progression is slowed at G4 motifs in HU grown *pif1-m2* cells

To confirm that replication slows at G4 motifs in the absence of Pif1we used 2D gels (Brewer and Fangman, 1987). We examined replication through three G4 motifs that had high Pif1-K264A binding and high DNA Pol2 occupancy in *pif1-m2* cells (Fig. 4A–C) and five non G4 sites (Fig. 4D,E). In experiments described later in the paper, we found that hydroxyurea (HU) affects growth of *pif1-m2* cells. Therefore, we examined replication in asynchronous WT and *pif1-m2* cells growing with or without low levels of HU. For the non G4 sites, only the HU data are shown, as the results were identical in both growth conditions.

In WT cells, there was no evidence for fork slowing at any of the G4 (Fig. 4A–C) or non G4 (Fig. 4E) sites, even when cells were grown in HU. Likewise, fork slowing was not seen at non G4 sites with or without HU in *pif1-m2* cells. Fork slowing was only evident in fragments with G4 motifs in HU grown *pif1-m2* cells (Fig. 4A–C). For each of the three fragments with a G4 motif, there was a 3–4 fold increase in the signal for forked replication intermediates in HU grown *pif1-m2* compared to HU grown WT cells (Fig. 4A–C). The increased numbers of replication intermediates were only seen for fragments with G4 motifs when the DNA had been crosslinked *in vivo* with psoralen, suggesting that these intermediates contained nicks and/or single-stranded gaps. Quite unexpectedly, the increase in replication intermediates was not limited to the site of the G4 motif but occurred throughout the arc of forked replication intermediates. At all three G4 loci, molecules similar to converging replication forks were seen in HU grown *pif1-m2* cells (indicated by arrows).

Regional, rather than site specific, replication fork slowing was also seen near the Chr XIII$_{G4}$ motif by the criterion of high DNA Pol2 occupancy (Fig. 2E). Consistent with the genome-wide analysis (Fig. 3A), DNA Pol2 levels were higher at the Chr XIII$_{G4}$ motif in *pif1-m2* versus WT cells even in the absence of HU (Fig. 2E, solid grey line), but the increase was more dramatic in HU grown cells (Fig. 2E). This high DNA Pol2 association extended for 1.5 kb to one side and 4 kb to the other side of the G4 motif. Consistent with the 2D gels, there was no increase in DNA Pol2 occupancy specifically near the Chr XIII$_{G4}$ motif in WT cells with or without HU (Fig. 2E, dotted lines; even in WT cells, the level of DNA Pol2 binding was higher for all DNA sequences in HU grown cells due to the negative effects of HU on fork progression). As a control, we established that HU did not change the time of Pif1 binding to its target sites (Figure S1C,D). Thus, by two different methods, we see regional slowing of forks in the vicinity of G4 motifs in *pif1-m2* but not WT cells, and this pausing is exacerbated by growth in HU.

G4 motifs increase recombination between direct repeats in Pif1 deficient cells

G4 motifs are significantly associated with sites of spontaneous mitotic DSBs (Capra et al., 2010). Here we show that replication forks slowed in the vicinity of G4 motifs in *pif1-m2* cells (Fig. 3,4). Breakage of forks paused near G4 motifs could explain this association between G4 sites and DSB sites. To test this possibility, we monitored the effects of G4 motifs on recombination by inserting G4 motifs between two repeats. It is well documented that DSBs between repeats increase the rate of recombination between them in a manner that deletes the intervening DNA. We constructed a strain with a partial duplication of *ADE2* (Fig. 5A). The region between the repeats contained the *URA3* gene and a site for inserting a sequence to determine its effects on recombination, which produces Ade$^+$ Ura$^-$ cells.

We inserted seven strong Pif1 binding sites, four G4 motifs (Chr I$_{G4}$, Chr IX$_{G4}$, Chr X$_{G4}$, Chr XIII$_{G4}$), and three non G4 Pif1 binding sites (Chr VII$_{NG}$, I$_{GR}$, XV$_{GR}$) as well as a G-rich site with no Pif1 binding (Chr XIV$_{nP}$) into the recombination substrate (Fig. 5A). Each sequence was inserted in both orientations relative to the direction of replication through the repeats (le, G-rich strand is template for leading strand synthesis; lg, G-rich strand is

template for lagging strand synthesis). Recombination rates were measured in WT and *pif1* cells using fluctuation analysis (Lea and Coulson, 1949) (Fig. 5B,C).

In both WT and *pif1* cells, the four non G4 inserts, regardless of orientation, yielded recombination rates similar to those in similar assays (0.7 to 3.4e-6 events per cell division; Freudenreich et al., 1998). Thus, recombination with any one of the four non G4 inserts was not Pif1-sensitive, even if the site was G-rich or a strong Pif1 binding site. In WT cells, inserts with G4 motifs had recombination rates in the same range as the control inserts (0.6 to 2.6 e-6; Fig. 5C). However, the recombination rate for each of the three G4 substrates was ~8 to 144 fold higher in *pif1* versus WT cells (Fig. 5C, last column). Recombination rates were not markedly orientation dependent. In support of the interpretation that recombination in this assay was initiated by DSBs, phosphorylated H2A, a regional marker for DSBs (Downs et al., 2000), was high near the G4 motifs that stimulated recombination (but not at other sites) in *pif1* (but not in WT) cells (Fig. 5D). These data make a strong argument that Pif1 suppresses DNA breakage specifically at those G4 motifs to which it binds, but not at non G4 Pif1 binding sites. Being G-rich or a non G4 Pif1 binding site was not sufficient to confer Pif1-dependent fragility.

When *pif1* cells contain many copies of G4 motifs, the motifs are often mutated in a manner that prevents formation of G4 structures

To determine if *pif1* cells are hypersensitive to G4 motifs, we cloned three G4 motifs in both orientations in the conditional high copy number plasmid FAT10 (FAT10-G4) (Fig. 6A; Chr IX_{G4}, X_{G4}, IV_{G4}). As in the recombination assay, in the leading orientation, the G-rich strand is the template for leading strand replication and in the lagging orientation it is the template for lagging strand replication. FAT10 carries a promoter defective version of *LEU2* (Runge and Zakian, 1989) (Fig. 6A). When cells are grown without leucine, FAT10 is present at 200–400 copies/cell. Empty vector (data not shown) and vector with an insert that did not bind Pif1-K264A in either the array or ChIP/qPCR analyses (Chr XIV_{nP}; FAT10 control) were used as controls. Each of the FAT10 plasmids was introduced into WT, *pif1-m2*, *pif1*, and *sgs1* cells and examined for their effects on growth in minus leucine medium with or without HU (Fig. 6B,C, and data not shown). Sgs1 is the sole yeast RecQ family helicase (*sgs1* cells were not tested on HU because they are HU-sensitive; Yamagata et al., 1998).

Regardless of the orientation of the insert or the strain, cells carrying the FAT10-G4 plasmids grew as well on minus leucine medium as cells with control plasmids (Fig. 6B,C left; data not shown). However, when *pif1-m2* or *pif1* cells were grown with HU, each of the FAT10-G4 plasmids, regardless of insert orientation, caused slow growth (Fig. 6B,C right). In contrast, FAT10-G4 plasmids had no detectable effects on growth of HU grown WT cells. Since replication forks paused in the vicinity of G4 motifs in *pif1* cells (Fig. 3,4), and HU alone causes replication fork stalling, we speculate that a large excess of G4 motifs in Pif1 deficient cells lowers the amount of HU that results in irreversible cell cycle arrest.

The slow growing *pif1-m2* or *pif1* cells from the HU minus leucine plates were streaked twice on plates without HU and then tested again for growth on HU minus leucine. After re-streaking, cells with FAT10-G4 grew as well on HU medium as cells with control plasmids.

To understand this change, plasmid DNA was isolated from individual colonies and sequenced. We also isolated plasmids from WT and mutant cells that were never grown in HU (Fig. 6D).

All of the G4 inserts from plasmids recovered from WT cells grown with or without HU were unchanged (0 of 420, 0.02% mutation frequency; Fig. 6D). Inserts in plasmids from *sgs1* cells also had no mutations (0/210, 0.04% mutation frequency). In contrast, many of the inserts in plasmids isolated from *pif1-m2* and *pif1* cells were mutated (Fig. 6D,E). Mutations were even seen in inserts from Pif1 deficient cells grown without HU (No HU: *pif1-m2*, 7/210, 3.5%; *pif1Delta;*, 23/210, 11% mutation frequency; Fig. 6D). However, mutations were more frequent in HU-grown cells (HU: *pif1-m2*; 27/210, 13%; *pif1Delta;*, 45/210, 21.4% mutation frequency; Fig. 6E). The mutation frequency was higher in the complete absence of Pif1 (*pif1-m2*: 34/420, 8.1%; *pif1Delta;* cells: 78/420, 18.5%). A higher fraction of the inserts were mutated when they were present on the lagging strand template (leading: 36/210, 17%; lagging: 66/210, 31%; data from *pif1-m2* and *pif1Delta;* cells were combined). This assay was the only one in this paper where the orientation of the G4 motif affected the magnitude of the effect. The high frequency of insert mutations was not due to a mutator phenotype in *pif1* cells as there were no mutations in control inserts from HU grown *pif1-m2* (30 inserts) or *pif1Delta;* (30 inserts) cells and *pif1* cells did not have higher mutation rates for structural genes like *CAN1* (KP and VAZ, unpublished results).

Most of the mutated inserts arose independently as the sequences of 78/102 (76%) were unique (Fig. 6F). Each mutated insert had multiple base pair changes per insert with an average of 5.4 (~8%) nucleotide changes per insert (see Fig. 6F for representative mutations in Chr IX$_{G4}$). Most mutations (~78%) were located within the G4 motif. Moreover, ~20% of the Gs within G tracts were mutated, a higher mutation frequency than for any other nucleotide, including Gs that were not in G-tracts (4% of these Gs were mutated). Even more dramatically, almost all mutated inserts (98/102) were no longer able to form G4 structures by the computational method used to identify the motifs (Capra et al., 2010).

These results can be explained if the addition of many G4 motifs in cells already under HU-induced replication stress provides selection for cells in which some of the inserts are mutated in a way that prevents their forming G4 structures. If Pif1 is acting at G4 structures, the mutated G4 inserts should no longer bind Pif1. If G4 structures are responsible for the replication defects and fragility, the mutated inserts should not affect replication nor cause chromosome breakage in *pif1* cells. The first prediction was tested using ChIP/qPCR on cells expressing Pif1-K264A-Myc and carrying FAT10 plasmids with the control insert (Fig. 6G, grey bar), G4 inserts (black bars) or mutant G4 inserts (striped bar) (Fig. 6G). Whereas the original IV$_{G4}$, IX$_{G4}$, and X$_{G4}$ inserts bound Pif1-K264A strongly (dark bars), the mutated inserts did not (grey bars) (light grey bar). The same mutated versions of IX$_{G4}$, and X$_{G4}$ inserts were inserted into the direct repeat assay. In *pif1* cells, the mutated inserts no longer slowed replication by the criterion of high DNA Pol2 binding (Fig. 5E), no longer stimulated direct repeat recombination (Fig. 6H), and were no longer prone to breakage by the criterion of H2A phosphorylation (Fig. 5D).

Discussion

The evidence for conservation and possible functions of G4 structures combined with the discovery of DNA helicases that unwind them has generated renewed interest in these non-canonical DNA secondary structures. Mutation of two FANC-J family helicases results in instability of endogenous G-rich sequences (Cheung et al., 2002; London et al., 2008). However, helicases function in RNA transcription, processing and translation, not just DNA mechanics, and the FANC-J studies do not allow the conclusion that these helicases act directly on G-rich or G4 structures. To our knowledge, this is the first demonstration of a DNA helicase having a direct role at G4 motifs *in vivo*. We also provide evidence for a mechanistic basis for instability of G4 motifs in the absence of a G4-resolving helicase.

G4 motifs were among the high confidence Pif1 binding sites (Fig. 1; q < 0.001), providing evidence that Pif1 acts directly on G4 motifs. However, consistent with its being a multi-functional helicase, G4 motifs were only a subset of Pif1 binding sites. In addition, only 25% of the G4 motifs were Pif1 associated. This measurement is an underestimate both because we used stringent criteria to identify binding sites and because we excluded the ~140 G4 motifs in telomeric DNA and the ~900 G4 motifs in rDNA, even though Pif1 bound extremely well to both (Fig. 1E). It is also possible that not all G4 motifs form G4 structures or that their formation is limited to a time (e.g. meiosis) that was not monitored in our experiments or that different helicases act on different sets of G4 motifs. The demonstration that those G4 motifs that were Pif1 associated were more likely than other G4 motifs or other non G4 Pif1 binding sites to impede replication (Fig. 3) and to stimulate DNA breakage (Fig. 5) in *pif1* cells makes a strong argument that their binding to Pif1 is biologically, as well as statistically, significant.

Unexpectedly, peak Pif1 binding to G4 motifs occurred after replication of the regions containing the motifs, although there was also significant binding to the motifs at earlier times in the cell cycle (Fig. 2). This binding pattern is consistent with the cell cycle regulated abundance of nuclear Pif1, which is present throughout the cell cycle but is maximally expressed in late S/G2 phase (Vega et al., 2007). Perhaps Pif1 unwinds G4 motifs throughout the cell cycle but binds and unwinds G4 motifs again after replication as a sort of failsafe mechanism that makes sure that the genome is free of G4 structures prior to chromosome condensation. Alternatively, other yeast helicases that act mainly earlier in the cell cycle could contribute to resolution of G4 structures. Another possibility is that G4 structures do not need to be resolved for forks to move past them (see below), but they must be resolved prior to mitosis.

By two criteria, DNA Pol2 occupancy (Fig. 2E) and 2D gels (Fig. 4), replication forks slowed in the vicinity of G4 motifs in *pif1-m2* cells. The genome-wide studies are particularly compelling as those G4 motifs that bound Pif1 were much more likely than non Pif1 binding G4 motifs to have high DNA Pol2 occupancy in *pif1-m2* cells (q < 0.001 versus q = 0.47). Likewise, four of four tested G4 motifs that were Pif1 associated in WT cells stimulated recombination in the direct repeat assay in *pif1* cells while three of three non G4 Pif1 binding sites did not (Fig. 5). Even more dramatically, when two G4 motifs were mutated so they no longer are predicted to form G4 structures but retained the same high GC

content they had prior to mutation, neither of the mutated motifs bound Pif1 (Fig. 6G) or increased DNA Pol2 occupancy (Fig. 5E), DNA breakage (Fig. 5D), or recombination in the absence of Pif1 (Fig. 6H). Thus, being Pif1 associated or being G-rich was not sufficient to affect replication or chromosome fragility in *pif1* mutant cells.

The second unanticipated result from this study is that the replication fork slowing near G4 motifs in *pif1-m2* cells was regional rather than site specific. This characteristic was evident from both 2D gels (Fig. 4) and DNA Pol2 levels (Fig. 2E). In contrast, stable protein complexes slow fork progression in a site-limited manner (Deshpande and Newlon, 1996; Greenfeder and Newlon, 1992; Ivessa et al., 2003). It was harder to detect fork slowing by 2D gels than by ChIP, perhaps because the latter is more sensitive. Alternatively, and we think more likely, replication intermediates in the vicinity of a G4 motif were difficult to isolate intact, as required for 2D gels but not for ChIP, as their recovery required *in vivo* psoralen cross-linking. This requirement suggests that while forks can ultimately bypass G4 structures, the DNA in the vicinity of the bypassed motifs is often damaged, containing nicks or gaps. Together these data indicate that forks slow in both their approach and movement away from a G4 motif, a behavior that suggests that unresolved G4 structures affect DNA topology/chromatin structure or generate torsional stress that acts over several kb.

In addition, Pif1 deficient, but not WT or *sgs1*, cells were sensitive to large numbers of G4 motifs under conditions where replication was impaired by HU (Fig. 6). The slow growth of Pif1 deficient cells in HU was eliminated by spontaneous mutations in the G4 motifs (Fig. 6). The majority of these mutations were located in the G-residues of the G-tracts, and, in virtually all cases, these mutations eliminated the predicted ability of the insert to form a G4 structure. Three of three tested spontaneous mutations lost Pif1 binding and their negative effects on chromosome integrity concomitant with losing the ability to form a G4 structure. The frequent mutations in G residues within G4 motifs must confer a selective advantage that makes it easier to maintain high plasmid copy number.

It seems unlikely that all of Pif1's effects are due to its unwinding G4 structures, especially as many of its binding sites were not G4 motifs. For example, 8% of its binding sites were at highly transcribed genes, most of which lack a G4 motif. Because Pif1 has the unusual property of being more active at displacing RNA than DNA from a DNA substrate (Boule and Zakian, 2007), it might act at these genes by removing RNA from highly transcribed regions. However, other Pif1 functions might be related to unwinding G4 structures. For example, Pif1's role in generating long flap Okazaki fragments might occur when a G4 structure forms in an extruded flap, and its resolution by Pif1 generates a longer than average length flap that requires Dna2 for processing (Fig. 7B). Pif1's critical but poorly understood role in mitochondrial DNA might also involve G4 structures, as remarkably, the very AT-rich yeast mitochondrial genome has a nearly tenfold higher density of G4 motifs than nuclear DNA (Capra et al., 2010).

We end with a speculative model to explain our data (Fig. 7B,C). We propose that growth in HU or reduced Pif1 increases the probability of G4 structure formation and/or persistence. If a G4 structure is present when the region containing it is replicating, forks slow as they

approach (and move away from) the structure but are usually able to bypass the G4 structure, just as they bypass other DNA lesions, leaving nicks or gaps behind. More rarely, forks arrest at G4 structures in *pif1-m2* cells (see converged forks; Fig. 4, arrows). If G4 structures are present at the end of S phase, Pif1 binds to and resolves them, and doing so suppresses breakage at these sites as chromosomes condense in preparation for mitosis. This model may be relevant to the function of human PIF, which also unwinds G4 structures *in vitro* (Sanders, 2010) since like yeast Pif1 (Vega et al., 2007), human PIF is cell cycle regulated with peak abundance in late S/G2 phase (Mateyak and Zakian, 2006).

Experimental Procedures

All experiments were done in YPH background (Sikorski and Hieter, 1989). Yeast strains and primers are listed (Table S4,5). ChIP was performed as described (Azvolinsky et al., 2006). Immunoprecipitated DNA was analyzed by either genome-wide microarray analysis or qPCR. For genome-wide analysis, immuno-precipitated DNA was amplified and labeled with minor modifications of instructions in the Agilent Yeast ChIP on chip protocol v9.2 (http://www.chem.agilent.com). For all genome-wide ChIP experiments, the signal in each spot was normalized to the hybridization signal obtained for the same spot using input DNA. Binding sites were identified from the median standardized array values using ChIPOTle 2.0 and a significance cutoff of 0.05 after the Benjamini-Hochberg multiple hypothesis correction (Buck et al., 2005). When testing multiple association hypotheses, we controlled the false discovery rate and report q-values for each test (Benjamini and Hochberg, 1995; Storey, 2003). G4 motifs were identified as described (Capra et al., 2010). The association of two sets of genome features was evaluated by comparing the observed overlap to the amount of overlap found when the query regions were randomly placed 1000 times across the genome (Capra et al., 2010). The genome annotations come from a variety of sources (Supplemental Methods; Capra et al., 2010). For 2D gels, DNA was crosslinked with psoralen (Gasser et al., 1996), and electrophoresis performed as in (Brewer and Fangman, 1987). Each 2D gel was carried out a minimum of three times on independent DNA preparations. The strain used for the direct repeat recombination experiments (YPH499 derivative called yBL3100) and the vectors for modifying it were generated by B. Lenzmeier. Test sequences from yeast DNA were cloned into two vectors, A2DRIV-A and A2DRIV-B, for integration on Chr VI between ORF yFR020W and yFR021W. They differ from each other only by the orientation of *ADE2*, allowing the test sequence to be integrated in either orientation into the yeast genome. *pif1* was created prior to each experiment. Recombination rates were determined by fluctuation analysis using the method of the median (Lea and Coulson, 1949). For FAT10 assays, fragments with G4 motifs or control sequence were PCR amplified from the natural chromosomal region and cloned into FAT10. For each experiment FAT10 was freshly transformed into otherwise isogenic WT, *pif1-m2*, *pif1* , and *sgs1* cells. Plasmid DNA was recovered from yeast cells, transformed into bacteria, and G4 inserts plus flanking DNA from the isolated plasmids were sequenced. Mutations in G4 motifs for FAT10_ChIP and recombination assay were created using PCR site directed mutagenesis (Agilent).

References

Azvolinsky A, Dunaway S, Torres J, Bessler J, Zakian VA. The *S. cerevisiae* Rrm3p DNA helicase moves with the replication fork and affects replication of all yeast chromosomes. Genes Dev. 2006; 20:3104–3116. [PubMed: 17114583]

Azvolinsky A, Giresi P, Lieb J, Zakian V. Highly transcribed RNA polymerase II genes are impediments to replication fork progression in Saccharomyces cerevisiae. Mol Cell. 2009; 34:722–734. [PubMed: 19560424]

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. JRoy Statist Soc Ser B. 1995; 57:289–300.

Bochman ML, Sabouri N, Zakian VA. Unwinding the functions of the Pif1 family helicases. DNA Repair (Amst). 2010; 9:237–249. [PubMed: 20097624]

Boule J, Vega L, Zakian V. The Yeast Pif1p helicase removes telomerase from DNA. Nature. 2005; 438:57–61. [PubMed: 16121131]

Boule JB, Zakian VA. The yeast Pif1p DNA helicase preferentially unwinds RNA DNA substrates. Nucleic Acids Res. 2007; 35:5809–5818. [PubMed: 17720711]

Brewer BJ, Fangman WL. The localization of replication origins on ARS plasmids in *S. cerevisiae*. Cell. 1987; 51:463–471. [PubMed: 2822257]

Buck MJ, Nobel AB, Lieb JD. ChIPOTle: a user-friendly tool for the analysis of ChIP-chip data. Genome Biol. 2005; 6:R97. [PubMed: 16277752]

Capra JA, Paeschke K, Singh M, Zakian VA. G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae. PLoS Comput Biol. 2010; 6:e1000861. [PubMed: 20676380]

Cheung I, Schertzer M, Rose A, Lansdorp PM. Disruption of dog-1 in Caenorhabditis elegans triggers deletions upstream of guanine-rich DNA. Nat Genet. 2002; 31:405–409. [PubMed: 12101400]

Deshpande AM, Newlon CS. DNA replication fork pause sites dependent on transcription. Science. 1996; 272:1030–1033. [PubMed: 8638128]

Downs JA, Lowndes NF, Jackson SP. A role for Saccharomyces cerevisiae histone H2A in DNA repair. Nature. 2000; 408:1001–1004. [PubMed: 11140636]

Foury F, Dyck EV. A PIF-dependent recombinogenic signal in the mitochondrial DNA of yeast. EMBO J. 1985; 4:3525–3530. [PubMed: 16453651]

Freudenreich CH, Kantrow SM, Zakian VA. Expansion and length-dependent fragility of CTG repeats in yeast. Science. 1998; 279:853–856. [PubMed: 9452383]

Gasser R, Koller T, Sogo JM. The stability of nucleosomes at the replication fork. J Mol Biol. 1996; 258:224–239. [PubMed: 8627621]

Greenfeder SA, Newlon CS. Replication forks pause at yeast centromeres. Mol Cell Biol. 1992; 12:4056–4066. [PubMed: 1508202]

Hershman SG, Chen Q, Lee JY, Kozak ML, Yue P, Wang LS, Johnson FB. Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. Nucleic Acids Res. 2008; 36:144–156. [PubMed: 17999996]

Huppert JL. Hunting G-quadruplexes. Biochimie. 2008; 90:1140–1148. [PubMed: 18294969]

Huppert JL. Structure, location and interactions of G-quadruplexes. FEBS J. 2010

Ivessa AS, Lenzmeier BA, Bessler JB, Goudsouzian LK, Schnakenberg SL, Zakian VA. The *Saccharomyces cerevisiae* helicase Rrm3p facilitates replication past nonhistone protein-DNA complexes. Mol Cell. 2003; 12:1525–1536. [PubMed: 14690605]

Ivessa AS, Zhou JQ, Zakian VA. The *Saccharomyces* Pif1p DNA helicase and the highly related Rrm3p have opposite effects on replication fork progression in ribosomal DNA. Cell. 2000; 100:479–489. [PubMed: 10693764]

Lea DE, Coulson CA. The distribution of the numbers of mutants in bacterial populations. J Genet. 1949; 49:264–285. [PubMed: 24536673]

Lipps HJ, Rhodes D. G-quadruplex structures: in vivo evidence and function. Trends Cell Biol. 2009; 19:414–422. [PubMed: 19589679]

London TB, Barber LJ, Mosedale G, Kelly GP, Balasubramanian S, Hickson ID, Boulton SJ, Hiom K. FANCJ is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts. J Biol Chem. 2008; 283:36132–36139. [PubMed: 18978354]

Mateyak M, Zakian V. Human PIF helicase is cell cycle regulated and associates with telomerase. Cell Cycle. 2006; 23:2796–2804. [PubMed: 17172855]

Myung K, Chen C, Kolodner RD. Multiple pathways cooperate in the suppression of genome instability in *Saccharomyces cerevisiae*. Nature. 2001; 411:1073–1076. [PubMed: 11429610]

Nakken S, Rognes T, Hovig E. The disruptive positions in human G-quadruplex motifs are less polymorphic and more conserved than their neutral counterparts. Nucleic Acids Res. 2009; 37:5749–5756. [PubMed: 19617376]

Paeschke K, McDonald KR, Zakian VA. Telomeres: Structures in need of unwinding. FEBS letter. 2010; 584:3769–3772.

Ribeyre C, Lopes J, Boulé J, Piazza P, Guédin A, Zakian V, Mergny JL, Nicolas A. The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences *in vivo*. PLoS Genet. 2009; 5:e1000475. [PubMed: 19424434]

Runge KW, Zakian VA. Introduction of extra telomeric DNA sequences into *Saccharomyces cerevisiae* results in telomere elongation. Mol Cell Biol. 1989; 9:1488–1497. [PubMed: 2657397]

Sanders CM. Human Pif1 helicase is a G-quadruplex DNA binding protein with G-quadruplex DNA unwinding activity. Biochem J. 2010

Schulz VP, Zakian VA. The *Saccharomyces PIF1* DNA helicase inhibits telomere elongation and *de novo* telomere formation. Cell. 1994; 76:145–155. [PubMed: 8287473]

Sikorski RS, Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics. 1989; 122:19–27. [PubMed: 2659436]

Storey JD. The positive false discovery rate: A Bayesian interpretation and the q-value. Annals of Statistics. 2003; 31:2013–2035.

Vega L, Phillips J, BRT, DPT, Onigbanjo M, Zakian V. Sensitivity of yeast strains with long G-tails to levels of telomere-bound telomerase. PLoS Genet. 2007; 3:1065–1075.

Yamagata K, Kato J, Shimamoto A, Goto M, Furuichi Y, Ikeda H. Bloom's and Werner's syndrome genes suppress hyperrecombination in yeast sgs1 mutant: implication for genomic instability in human diseases. Proc Natl Acad Sci USA. 1998; 95:8733–8738. [PubMed: 9671747]

Zhou JQ, Monson EM, Teng SC, Schulz VP, Zakian VA. The Pif1p helicase, a catalytic inhibitor of telomerase lengthening of yeast telomeres. Science. 2000; 289:771–774. [PubMed: 10926538]
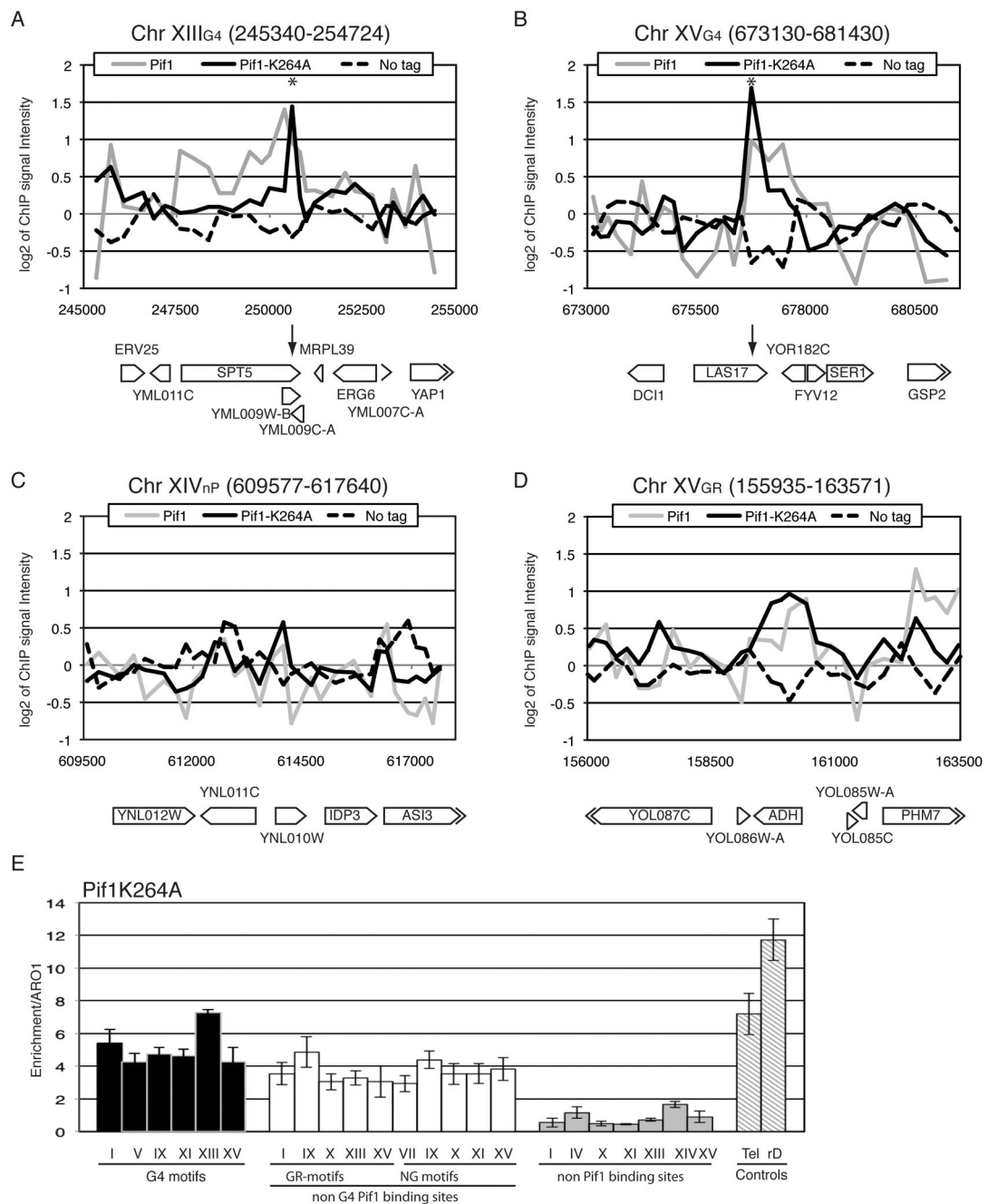
Figure 1. Pif1 binds G4 motifs *in vivo*
(A–D) Here and in subsequent figures, roman numerals indicate the chromosome on which the region is located. Log_2 values of Pif1-K264A (solid dark lines) and WT Pif1 (grey lines) binding compared to the no tag control (dotted line) within four 10 kb regions, two containing a Pif1 binding G4 motif (XIII_{G4}, A; XV_{G4}, B) and two G-rich but non G4 regions, one with no significant Pif1 binding (XIV_{nP}, C) and one with strong Pif1 binding (XV_{GR}, D). Here and in subsequent figures, asterisks mark locations of G4 motifs; arrows indicate protein binding sites; annotations corresponding to the coordinates are shown below the plot (related to Table S1). E) qPCR analysis of Pif1-K264A association at G4 motifs

(black), non G4 Pif1-K264A binding sites (both NG, not G-rich and GR, non G4 G-rich) (white), non Pif1 binding regions (nP, grey) and controls (VI-R telomere, Tel and rDNA, rD; striped). Here and in all subsequent Pif1 ChIP analysis, values were normalized to both input and the amount of *ARO1* DNA in the immuno-precipitate. Here and elsewhere error bars are +/− SD.
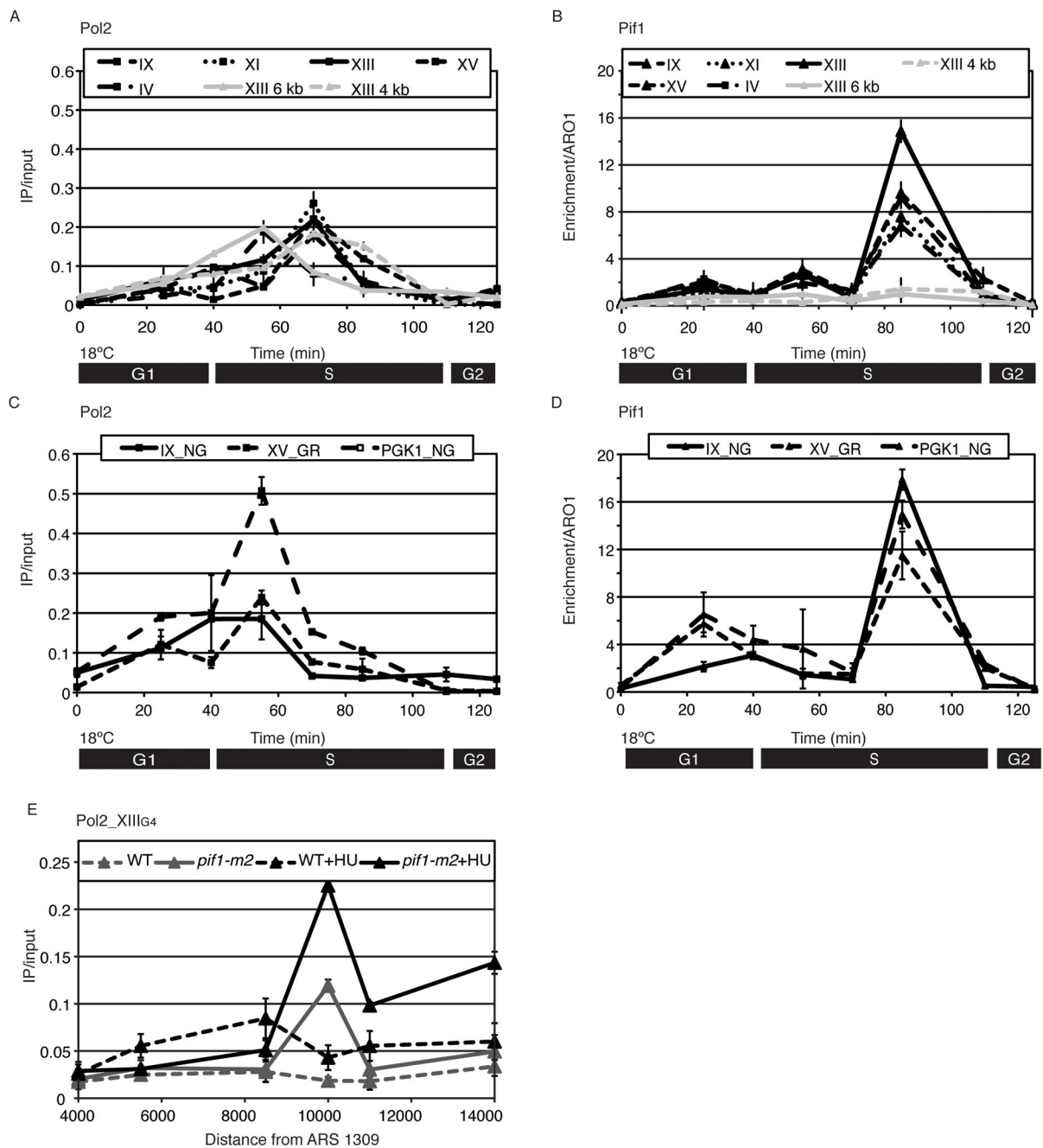
Figure 2. Pif1 binds G4 motifs after semi-conservative replication of the motif
(A–D) Cells expressing HA tagged DNA Pol2 (Pol2) and MYC tagged Pif1 (Pif1) were synchronized, and samples taken for ChIP at the indicated times. DNA in the immuno-precipitate (IP) was quantitated by qPCR. Here and in all subsequent DNA Pol2 ChIPs, the amount of DNA in the IP was normalized to input DNA (IP/input). Log2 values for DNA Pol2 (A,C) or Pif1 enrichment (B,D) at various Pif1 binding sites. Loci examined were G4 motifs (A, B, black), non G4 regions with no Pif1 binding (6 kb to the left and 4 kb to the right of Chr XIII$_{G4}$; A, B, grey) and non G4 regions with strong Pif1 binding (IX$_{NG}$, non G-rich, XV$_{GR}$ G-rich), and a highly transcribed gene (PGK1$_{NG}$) (C, D) (related to Figure S1)

(E) DNA Pol2 binding to the Chr XIII$_{G4}$ motif and its flanking regions was determined in asynchronous cells growing with or without HU in *pif1-m2* and WT cells. The G4 motifs is 10 kb from ARS1309.
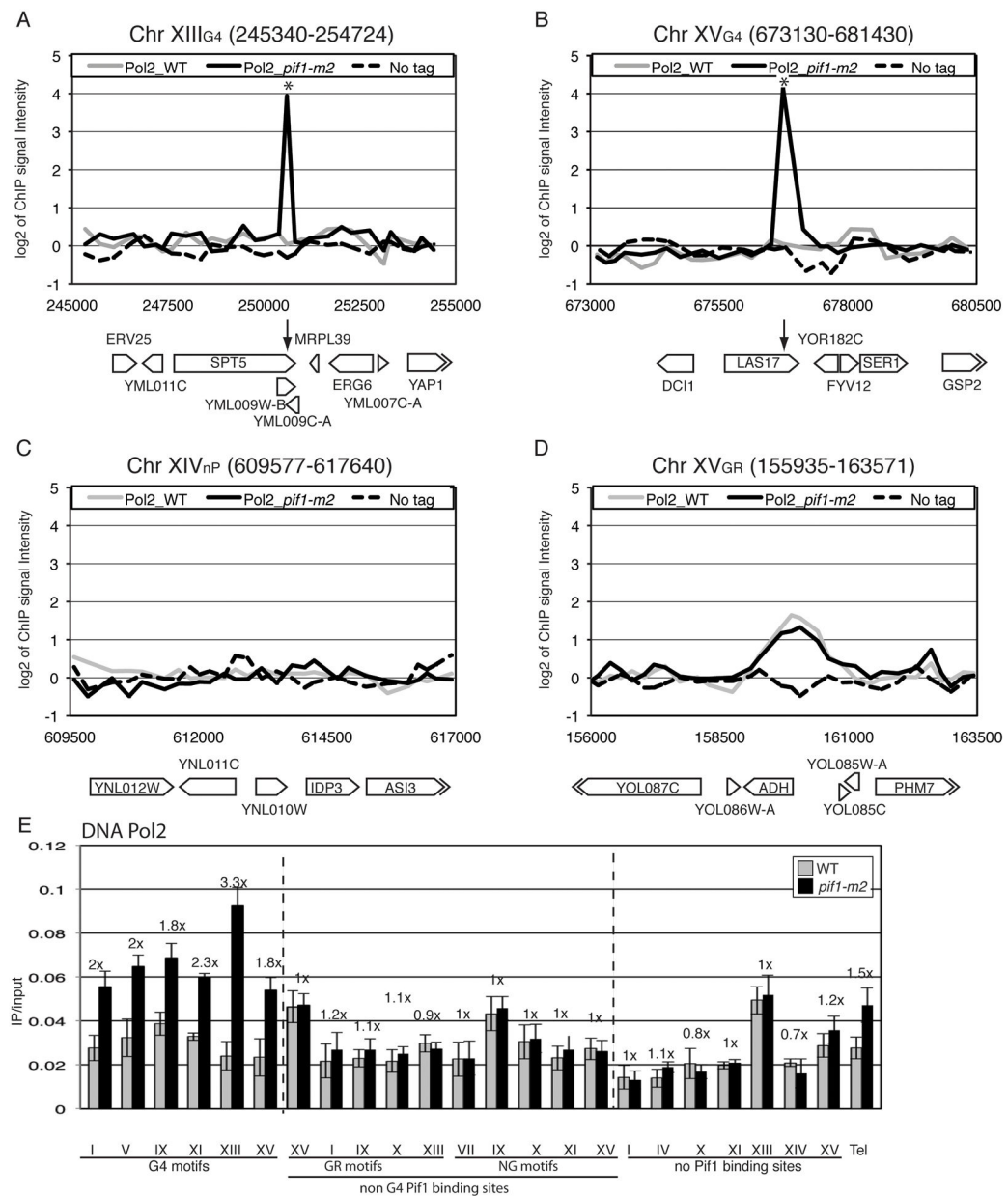
Figure 3. G4 motifs have increased DNA Pol2 occupancy in *pif1-m2* versus WT cells
Asynchronous WT and *pif1-m2* cells expressing DNA Pol2-MYC were analyzed by ChIP-microarray experiments (Fig. 1A–D). Log_2 values of DNA Pol2 binding in *pif1-m2* (solid dark lines) and WT cells (grey lines) compared to the no tag control (dotted line) within four 10 kb regions. (A,B) Two regions have G4 motifs ($XIII_{G4}$, A; XV_{G4}, B) and two are G-rich regions, one that did not (XIV_{nP}, C) and one that did bind Pif1 (XV_{GR}, D) (related to Table S3) (E) qPCR analysis of DNA Pol2-MYC association in WT (grey) and *pif1-m2* (black) cells at G4 motifs, non G4 Pif1-K264A binding sites (NG, not G-rich; GR, G-rich non G4), non Pif1 binding regions and the VI-R telomere (Tel). The numbers above the column indicate the fold DNA Pol2 enrichment in *pif1-m2* over WT cells at each locus.
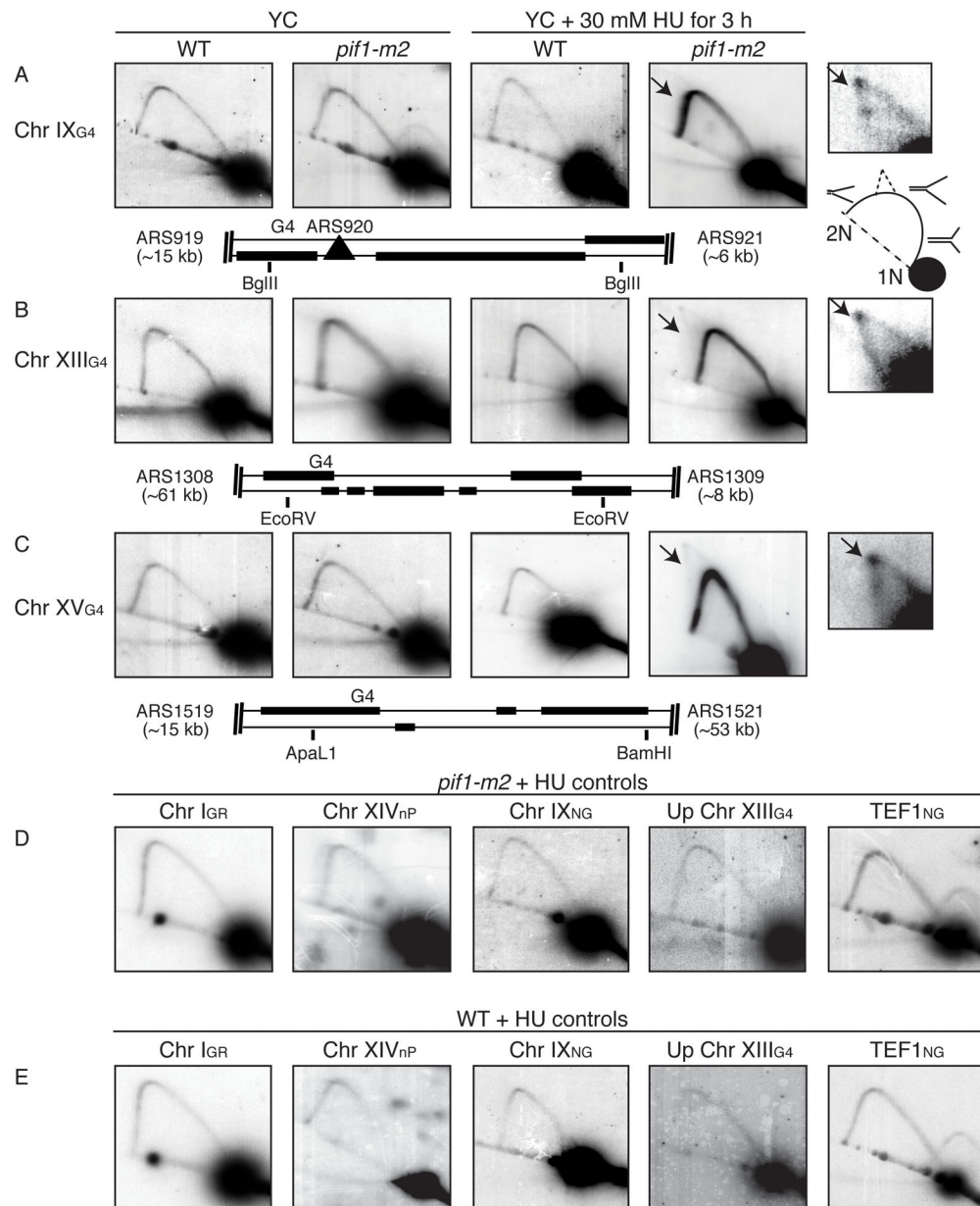
Figure 4. Two-dimensional (2D) gel analysis detects replication fork slowing in the vicinity of G4 motifs in HU grown *pif1-m2* cells

DNA was prepared from asynchronous WT or *pif1-m2* cells grown with or without 30 mM HU for 3 h, digested with restriction enzymes, and analyzed by 2D gels. Restriction enzymes were chosen so that the G4 motif is positioned ~75% of the way through the arc of forked replication intermediates. (A–C) Replication through three G4 motifs. (D, E) Replication through five non G4 loci. (Only DNA from HU grown cells are shown for samples in D and E.) Cartoon at the bottom of each panel indicates genomic location, distance to the nearest ARS, digestion site, and position of G4 motif. Media and strain conditions are indicated at top of figure. Arrows point to converging forks. The fraction of DNA in forked intermediates (FI) and in the 1N spot was determined using ImageQuant.

The average ratio of three independent experiments $\pm$ SD of FI/1N $_{pif1\text{-}m2}$ divided by FI/1N$_{WT}$ (no HU) was 1.0 ± 0.4 (Chr IX$_{G4}$), 1.3 ± 0.4 (Chr XIII$_{G4}$) and 1.1 ± 0.5 (Chr XV$_{G4}$); in HU grown cells, values were 3.1 ± 0.4 (IX$_{G4}$), 4.2 ± 0.7 (XIII$_{G4}$) and 3.2 ± 0.7 (XV$_{G4}$). D/E) 2D gel analysis of five control regions from genomic DNA from WT (E) and *pif1-m2* (D) cells in 30 mM HU: Chr I$_{GR}$ (strong Pif1 binding region), Chr XIV$_{nP}$ (G-rich region, no Pif1 binding), Chr IX$_{NG}$ (non G rich Pif1 binding site), Chr XIII$_{NG}$ (~ 10 kb upstream of Chr XIII$_{G4}$; up Chr XIII$_{G4}$, no Pif1 binding), and the highly transcribed TEF1 gene (no Pif1 binding).

A

ARS606 (~25 kb) ···· ──☐──█ ADE2 ──┤ URA3 ├─┤ CR ├─█──── ADE2 ────···· ARS607 (~5 kb)

ARS606 (~25 kb) ···· ──☐──█ ADE2 ──┤ URA3 ├─⭐─█──── ADE2 ────···· ARS607 (~5 kb)

Plate: FOA, Ade-

ARS606 (~25 kb)

───── ADE2 ─────···· ARS607 (~5 kb)

B

C

	WT	*pif1Δ*	*pif1Δ* / WT
I (le)_G4	1.17E -06	1.00E -05	8.61
I (lg)_G4	7.71E -07	7.44E -06	9.65
IX (le)_G4	6.10E -07	6.31E -06	10.35
IX (lg)_G4	9.13E -07	7.44E -06	8.15
XIII (le)_G4	2.64E -06	1.17E -04	44.31
XIII (lg)_G4	9.62E -07	1.39E -04	144.17
VII (le)_GN	2.86E -06	2.27E -06	0.80
VII (lg)_GN	3.41E -06	3.55E -06	1.04
XIV (le)_nP	7.92E -07	1.07E -06	1.35
XIV (lg)_nP	1.54E -06	1.40E -06	0.91
I (le)_GR	3.38E -06	3.65E -06	1.08
I(lg)_GR	1.88E -06	1.25E -06	0.66
XV (le)_GR	2.40E -06	2.59E -06	1.08
XV (lg)_GR	2.77E -06	2.72E -06	0.98
X_G4	1.30E -06	1.10E -05	8.48

D

E

Figure 5. G4 motifs enhance direct repeat recombination and are fragile sites in Pif1 deficient cells

G4 and non G4 Pif1 binding sites were integrated into a recombination cassette on Chromosome VI. (A) Schematic diagram of recombination assay. Star indicates cloning region. (B) Recombination rates for WT (grey) and *pif1Δ* (dark) cells. *pif1-m2* cells were not used because for unknown reasons, the direct repeat recombination substrate could not be recovered in this background. The inserts were 90 to 150 bps. All inserts were tested in both leading (le) and lagging (lg) strand orientation relative to the nearest ARS. (C) Average recombination rates from four independent experiments. Enrichment values are *pif1Δ* /WT for a given insert and orientation. (D) Binding of phosphorylated H2A was measured by

ChIP/qPCR at different recombination substrates in WT (grey) and *pif1* (dark) cells. Phosphorylated H2A levels were determined at two G4 motifs IX_{G4}, X_{G4}, two mutated G4 motifs IX_{G4mut} and X_{G4mut} as well as two non G4 Pif1 binding sites (I_{GR}, VII_{NG}). Levels of phosphorylated H2A at I_{GR} and VII_{NG} in WT and *pif1* ranged between 0.4–1.3. All phosphorylated H2A values were normalized to input (IP/input). (E) DNA Pol2 was tagged in recombination substrates and analyzed by ChIP/qPCR in WT (grey) and *pif1* (dark) cells. DNA Pol2 occupancy was analyzed at two different G4 motifs (IX_{G4} and X_{G4}) and two mutated G4 motifs IX_{G4mut} and X_{G4mut}.

A FAT10 6763 bp (Amp R, TRP1, 2-micron ORI, leu2-d, SphI)

B YC-LEU | YC-LEU+100 mM HU
WT/control, m2/control, WT/IX_le, m2/IX_le, WT/IX_lg, m2/IX_lg

C m2/control, WT/X_le, m2/X_le, WT/X_lg, m2/X_lg

D

	IV leading	IV lagging	IX leading	IX lagging	X leading	X lagging	Control
WT	0/35	0/35	0/35	0/35	0/35	0/35	0/30
pif1-m2	0/35	1/35	1/35	3/35	1/35	1/35	0/30
pif1Δ	3/35	2/35	3/35	4/35	4/35	7/35	0/30
sgs1	0/35	0/35	0/35	0/35	0/35	0/35	0/30

E

HU	IV leading	IV lagging	IX leading	IX lagging	X leading	X lagging	Control
WT	0/35	0/35	0/35	0/35	0/35	0/35	0/30
pif1-m2	2/35	4/35	1/35	6/35	5/35	9/35	0/30
pif1Δ	3/35	8/35	6/35	9/35	7/35	12/35	0/30

F

WT — Chr IX$_{G4}$
TTTGGAGGGTACGGTGGGTAATAAGGGAAGGTATCGGGATTGGGGTAGGCCATTAAGGGA

pif1-m2
TTTGGAGGGTACGGTATGTAATAAGGGAAGGTATCGAGATTGGGGTAGGCCATTAGTTCA
TTCGAAGCGTACGGTGGGTAATAAGGATAGGTATCCTGATTGGGGTAGGCCATTAAGGGA
TTTGGAGGGTACGGTGGGTAATAAGAGAAGGTATCGGGATTGCTGTAGGCCATTAAGGGA
TTTGGAGCCTAAAATGGGTAATAATAGAAGGTATAAGGATTGCTATAGGCCATTAAGGGA
TTTGGAGGGTACGTAGCGTAATAAGGGAAGGTATCGGGATTGGGGTAGGCCATATTGGGA
TTGGGAGGGTACGGTCGTAATAAGATAAGGTATCGGGATTGGGGTAGGCCATTAAGGGA
TTTGACGGGTACGGTGTGTAATAAGGGAAGGTATCAAGATTGGGGTAGGCCATTAAGGGA
G4$_{mut}$ GCGTAATAACGGAAGGTATCGGGATTCGCG

G Pif1K264A — Enrichment/ARO1
C, IV, IX, X (G4 motifs); IV, IX, X (G4 motifs mutated)

H Recombination rate — WT, pif1Δ
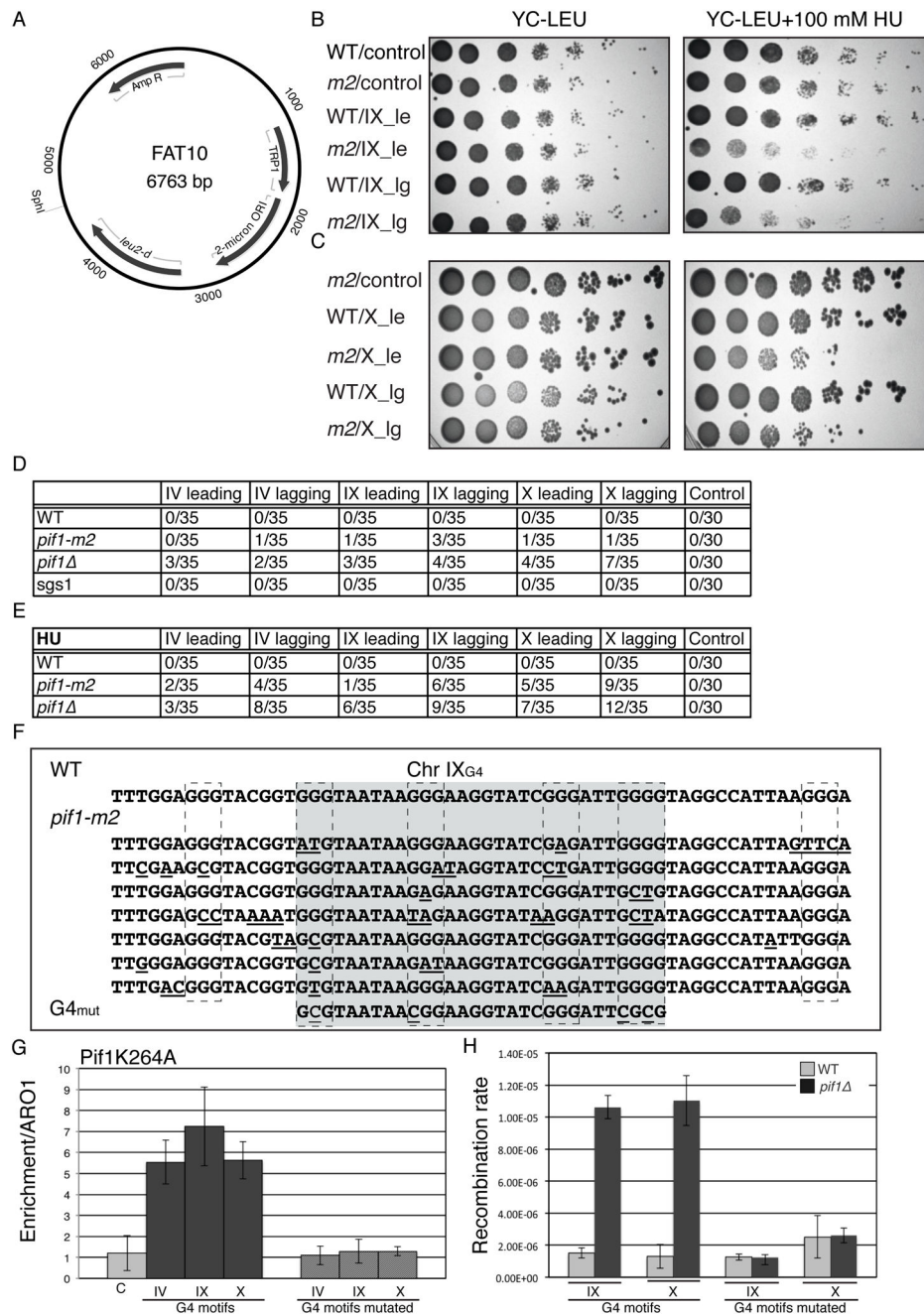IX, X (G4 motifs); IX, X (G4 motifs mutated)

Figure 6. G4 motifs are mutated in the absence of Pif1
(A) Map of FAT10 plasmid used to introduce extra copies of G4 motifs. (B, C) 1:10 dilutions of WT and *pif1-m2* (m2) cells carrying control plasmid (control) or plasmid with: Chr IX$_{G4}$ (B) or Chr X$_{G4}$ (C) on the leading (le) or lagging (lg) strand. Cells were plated on YC-LEU (left) and YC-LEU+100 mM HU (right). (D) Fraction of mutated inserts in various strains after plating on YC-LEU (D) or YC-LEU+HU (E). (F) Examples of mutations recovered from Chr IX$_{G4}$; mutated bases are underlined. Gray shaded area indicates G4 motif. Alternative G4 islands are indicated by a dashed box. (G) ChIP/qPCR analysis of Pif1-K264A to indicated inserts in FAT10. Association was examined at the control insert

(light grey) IV_{G4}, IX_{G4}, X_{G4} (dark bar) as well as at IV_{G4mut}, IX_{G4mut}, X_{G4mut} (grey bar). (H) Recombination rates for WT (grey) and *pif1* (dark) cells at two G4 motifs (IX_{G4}, X_{G4}) and the same motifs mutated (IX_{G4mut}, X_{G4mut}).
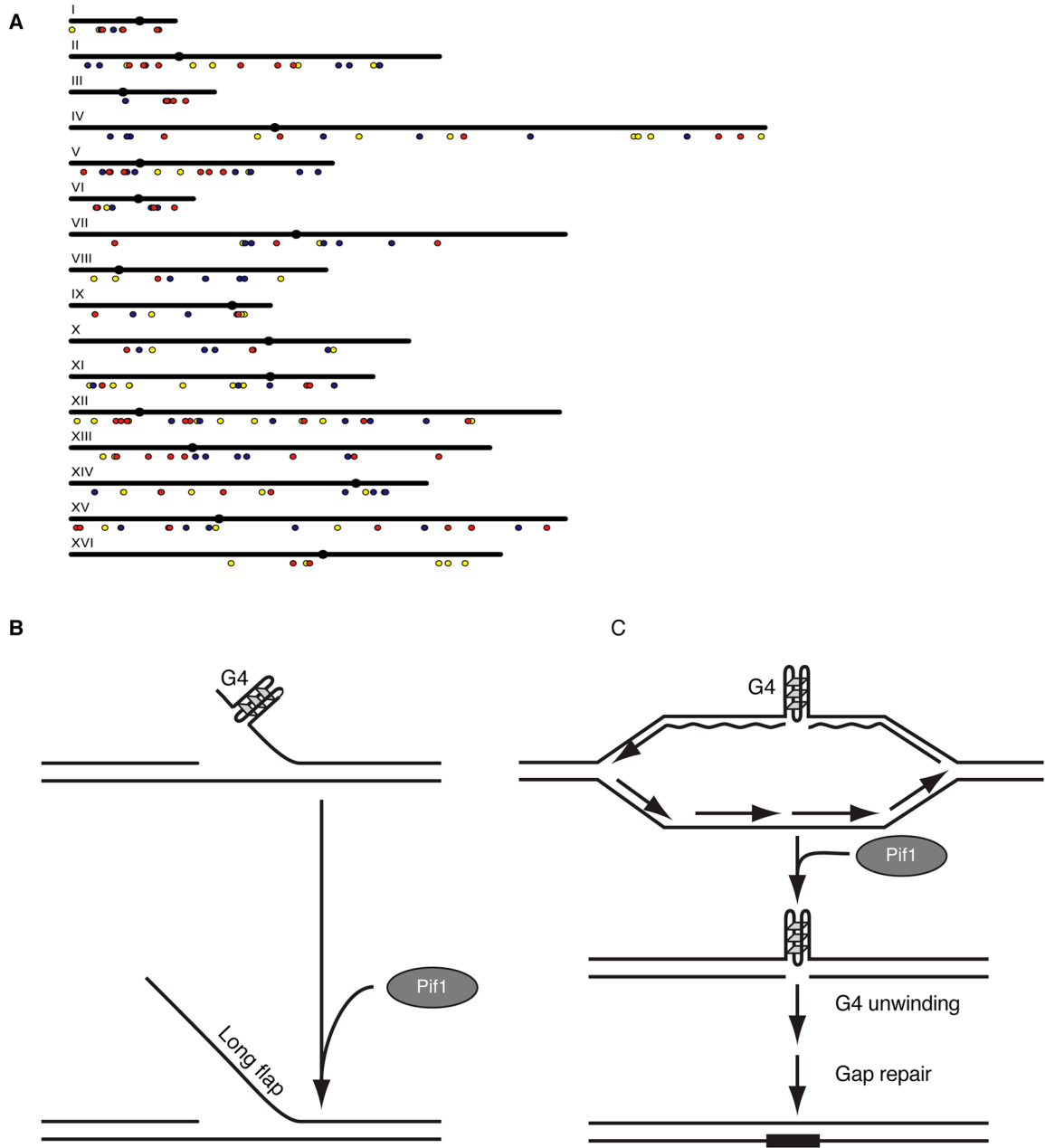
Figure 7. Model of Pif1 function at G4 structures
(A) Distribution of G4 motifs that bound Pif1-K264A (yellow), had high DNA Pol2 binding in *pif1-m2* cells (blue), or both (red) among the 16 yeast chromosomes (related to Table S2). Filled circles mark centromeres (B, C) Working model for Pif1 action at G4 structures on Okazaki fragments (B) and after replication of a region with a G4 structure (C). The figure shows structure formation on the leading strand template, but our data suggest that G4 structures form on both leading and lagging strand templates. Repair synthesis is denoted by thick line. See discussion for details.