Short loop length and high thermal stability determine genomic instability induced by G-quadruplex-forming minisatellites

Aurèle Piazza1,7, Michael Adrian2,§, Frédéric Samazan1,§, Brahim Heddi2, Florian Hamon3, Alexandre Serero1, Judith Lopes1,‡, Marie-Paule Teulade-Fichou3, Anh Tuân Phan2* & Alain Nicolas1,**

Abstract

G-quadruplexes (G4) are polymorphic four-stranded structures formed by certain G-rich nucleic acids, with various biological roles. However, structural features dictating their formation and/or function in vivo are unknown. In S. cerevisiae, the pathological persistency of G4 within the CEB1 minisatellite induces its rearrangement during leading-strand replication. We now show that several other G4-forming sequences remain stable. Extensive mutagenesis of the CEB25 minisatellite motif reveals that only variants with very short (≤ 4 nt) G4 loops preferentially containing pyrimidine bases trigger genomic instability. Parallel biophysical analyses demonstrate that shortening loop length does not change the monomorphic G4 structure of CEB25 variants but drastically increases its thermal stability, in correlation with the in vivo instability. Finally, bioinformatics analyses reveal that the threat for genomic stability posed by G4 bearing short pyrimidine loops is conserved in C. elegans and humans. This work provides a framework explanation for the heterogeneous instability behavior of G4-forming sequences in vivo, highlights the importance of structure thermal stability, and questions the prevailing assumption that G4 structures with short or longer loops are as likely to form in vivo.

Keywords genomic instability; G-quadruplex; minisatellite; Phen-DC5; Pif1
Subject Categories DNA Replication, Repair & Recombination

Introduction

G-quadruplexes (G4) are four-stranded structures formed by certain G-rich DNA or RNA sequences consisting in the stacking of multiple ‘G-quartets’ (a planar arrangement of four guanines (Gellert et al., 1962)) coordinated by cations (Williamson et al., 1989). Intramolecular G4-forming sequences typically contain four tracts of consecutive guanines separated by relatively short-loop regions of the form GxNxGxNxGxNxGx, where N can be any nucleotide, and x is usually 7 or less (Huppert & Balasubramanian, 2005; Guedin et al., 2010). Biophysical and structural studies revealed an impressive diversity of G4 conformations depending on the number of G-quartets, the length of the loops, and their sequences as well as different strand orientation (Burge et al., 2006) and handedness (Chung et al., 2015). However, how this conformational diversity and the thermodynamic properties of these transient secondary structures modulate their cellular functions remains poorly understood.

Compelling evidence implicates G4 motifs in various biological processes (reviewed in Maizels & Gray, 2013), including regulation of transcription (Siddiqui-Jain et al., 2002; Law et al., 2011), telomere capping (Paeschke et al., 2005, 2008), replication initiation at certain origins (Valton et al., 2014; Foulk et al., 2015), programmed genome rearrangements (Cahoon & Seifert, 2009), and accidental genomic instability (Kruisselbrink et al., 2008; Ribeyre et al., 2009; Piazza et al., 2010, 2012; Lopes et al., 2011) as well as RNA maturation, translation, and transport (Wieland & Hartig, 2007; Decorriade et al., 2011; Subramanian et al., 2011). During replication, the formation of intramolecular G4 is likely facilitated by the occurrence of single-strand DNA regions, but the determinants that affect the folding and stability of G4 in vivo remain to be elucidated. In vitro, several helicases unwind G4 that are strong impediments to the replicative polymerase progression (Woodford et al., 1994). Consequently, the formation and persistence of G4 in helicase defective cells or upon G4 stabilization with G4 ligands are highly suspected to drive the
genomic instability of G4-prone genomic regions (Cheung et al., 2002; Kruisielbrink et al., 2008; Rodriguez et al., 2012; Vannier et al., 2012; Koole et al., 2014).

In previous studies, we examined the genomic instability of the G4-forming human minisatellite CEB1 in mitotically growing S. cerevisiae cells. In the absence of Pif1, an evolutionary conserved G4 unwinding helicase (Ribeyre et al., 2009; Sanders, 2010; Paeschke et al., 2013), frequent expansion/contraction of the CEB1 tandem array is observed (Ribeyre et al., 2009). This instability depends on the ability of the CEB1 motif (39 nt) to form G4 in vitro and was not observed with the G-mutated array (Ribeyre et al., 2009; Lopes et al., 2011; Piazza et al., 2012). Consistently, treatment of wild-type (WT) cells with the Phen-DC3 G4-ligand (De Cian et al., 2004; Ambrus et al., 2005; Monchaud et al., 2008; Piazza et al., 2010) phenocopies the Pif1 deletion in vivo (Piazza et al., 2010; Lopes et al., 2011). Physical analysis of replication intermediates by 2D-gel revealed that G4 specifically perturbs the leading-strand replication, thus yielding CEB1 internal rearrangements in an orientation-dependent manner (Lopes et al., 2011). Here, we use this sensitive assay to characterize the G4 determinants dictating genomic instability in yeast. We assayed several validated G4-forming sequences and found that some but not all minisatellites exhibit instability. We identified the molecular determinants driving the in vivo instability by extensive mutagenesis of the stable human CEB25 minisatellite motif and parallel biophysical characterization of the resulting G4 structure by UV, CD and NMR spectroscopy. The CEB25 G4 structure has been recently solved by NMR (Amrane et al., 2012); it adopts an all-parallel strand arrangement connected by propeller loops, the first and third loop being a single T residue and the central loop being 9 nt long. Each motif in a CEB25 tandem array adopts the same monomorphic structure, leading to the possibility to form a homogeneous ‘pearl-necklace’ G4 structure in minisatellites (Amrane et al., 2012). Here, we show that only variants with shortened loops (≤ 4 nt) and a maximal total loop length of 5 nt containing pyrimidines exhibit minisatellite instability. Shortening the loop does not alter the monomorphic core structure of the CEB25 G4 but drastically increases its thermal stability, in correlation with its in vivo behavior. Finally, we performed a bioinformatics analysis of single-nucleotide loop G4 motifs in various model organisms. This enabled us to severely narrow the fraction of potential G4 motifs in the S. cerevisiae, S. pombe, C. elegans, and human genomes that might be ‘at risk’ to trigger genome instability. Strikingly, short pyrimidine loops are clearly under-represented compared to purine loops, but are strongly enriched for DNA damage upon treatment of human cells with the G4-ligand pyridostatin (Rodriguez et al., 2012). This study highlights the conserved threat for genomic stability posed specifically by highly stable G4 structures and alters the prevailing assumptions that G4 structures with short or longer loops are as likely to form in vivo and/or exert phenotypes.

Results

Heterogeneous behavior of chromosomally integrated G4-forming minisatellites

Here, we assayed the rearrangement frequency (also referred to as ‘instability’) of various synthetic minisatellites comprising natural G4 motifs and variant sequences (Supplementary Table S1, Materials and Methods). All arrays were chromosomally inserted near the ARS305 replication origin (Materials and Methods), and oriented so that the G-rich strand is template for the leading-strand replication machinery (‘Orientation I’ in Fig 1A in Lopes et al (2011)) (Supplementary Table S2; Materials and Methods). This is our most sensitive and best characterized location for the study of G4-induced rearrangements (Lopes et al., 2011).

In WT cells, a CEB1-WT array is rather stable (4 rearrangements/159 colonies) but undergoes frequent rearrangements upon addition of 10 μM Phen-DC₃ or in the absence of Pif1 (23/192 and 39/66; P-value vs. WT cells = 9.52 × 10⁻⁴ and 2 × 10⁻²¹, respectively) (Fig 1B, Supplementary Table S3). In contrast, CEB25-WT remained stable in both contexts (0/192 and 1/192, respectively), not significantly different from WT cells (0/192) (Fig 1B, Table 1). Thus, conditions that induced expansion-contraction of CEB1 exert no effect on CEB25. This is not due to an intrinsic inability of CEB25 to rearrange since, like CEB1, it exhibits expansion and contraction in the rad27Δ mutant (data not shown).

To investigate the behavior of other G4-prone sequences, we constructed three other minisatellite arrays each containing 18 identical G4 motifs. The G4-prone sequences were separated from one another by a non-G4 sequence spacer in order to prevent inter-motif G4 formation (Fig 1A; spacer italicized in gray; full array information in Supplementary Table S1). We chose the well-characterized CEB25 motif forms a unique G4 structure utilizing an isolated guanine propeller loops (Phan et al., 2007; Todd et al., 2011). Intriguingly, we found that the c-Myc allele exhibited significant destabilization upon Phen-DC₃ treatment and Pif1 deletion (17/96 and 12/23, P-value vs. untreated WT cells = 4.56 × 10⁻⁶ and 1.3 × 10⁻¹⁰, respectively), while the c-Kit and Bcl2-MBR alleles remained stable in the same conditions (Fig 1B, Supplementary Table S3). Thus, c-Myc behaves like CEB1-WT, while c-Kit and Bcl2-MBR behave like CEB25-WT. Hence, despite being able to form G4 in vitro, only a subset of G4-forming sequences exhibit genomic instability in the same yeast assay.

The 9-nt central loop of the CEB25 G4 is required and sufficient to stabilize the array in vivo

The sharp differences in the behavior of the G4-prone sequences prompted us to investigate the underlying molecular basis, using the CEB25 G4 as a model. To achieve this, we assayed the instability of CEB25 allele variants bearing modified G4 motifs (listed in Table 1, full allele information in Supplementary Table S1) and performed biophysical analyses of the G4 variants, presented afterward.

A striking structural feature of the CEB25-WT G4 motif is the presence of a long central loop of 9 nt (Fig 2A). To address whether this loop account for the stable in vivo behavior of CEB25-WT
Determination of G4 motif, which was confirmed with an independent allele bearing a different 9-nt-long loop (Supplementary Fig S1, Supplementary Table S4). Thus, these CEB1 loop size variants behave like CEB1-WT.

Conversely, we substituted the central single-nucleotide adenine loop within the G4 motif of CEB1 by the 9-nt central loop of CEB25. Strikingly, the CEB1-loopCEB25 allele remained fully stable in both Phen-DC3-treated WT cells and in the pif1 Δ mutant (0/192 and 0/188, respectively; P-values vs. CEB1-WT < 2.2 × 10^{-16}) (Fig 2B, Supplementary Table S3). The abolishment of the CEB1 instability was confirmed with a second allele bearing a different 9-nt-long loop (Supplementary Fig S1, Supplementary Table S4). Thus, these CEB1 loop size variants behave like CEB25-WT. Altogether, these results demonstrate that a single long loop within the G4 motif, although not affecting the ability to adopt a G4 structure in vitro, is required and sufficient to stabilize the minisatellite in vivo.

(Also referred to as L191, with the numbers indicating the sizes of three loops), we first replaced it by a single thymine residue to yield the CEB25-L111(T) variant (Fig 2A). Whereas the CEB25-L111(T) array is stable in WT cells (0/96 rearrangements), it became unstable upon addition of Phen-DC3 (42/192) or deletion of pif1 (21/32) (Fig 2A, Table 1). These instabilities are the highest ever measured in our experimental system, especially for such short minisatellites (13 motifs). These results were confirmed with an independent strain bearing a shorter CEB25-L111(T)-8m; it is also highly destabilized in the presence of Phen-DC3 or in the absence of Pif1 (10/94 and 17/38, respectively). Thus, the variant CEB25-L111(T) behaves like CEB1.

Figure 1. Heterogeneous instability phenotype of different G4-forming tandem repeats in WT cells treated or not with Phen-DC3, and in pif1 Δ cells.

A Motif sequence of different G4-forming tandem repeats. G4 motif is underlined. G-tracts are shown in bold. The c-Myc, c-Kit, and Bcl2-MBR G4-forming sequences have been separated by the neutral CEB1 spacer (in gray) to prevent the formation of irrelevant G4 conformations resulting from the tandem organization. Details about the minisatellite size, number of motifs, and GC content are provided in Supplementary Table S1.

B Southern blot analysis of the G4-forming minisatellites CEB1-WT (26 motifs; WT: ORT7131, pif1 Δ: ORT7137), CEB25-WT (13 motifs; WT: ORT7167, pif1 Δ: ORT7175), c-Myc (18 motifs; WT: ORT7358, pif1 Δ: ORT7345-8), c-Kit (18 motifs; WT: ORT7359, pif1 Δ: ORT7346), and Bcl2-MBR (18 motifs; WT: ORT7357, pif1 Δ: ORT7344) in WT cells treated for 8 generations with DMSO (control) or the G4-ligand Phen-DC3 (10 μM), and in pif1 Δ cells. The number of colonies analyzed per lane and the total rearrangement frequencies are indicated. Each blot may not show all the colonies analyzed to obtain the final rearrangement frequency. DNA was digested with EcoRI that cuts within 20 nt at each side of the minisatellite, and membranes have been hybridized with the appropriate probe. The same molecular ladder (Lambda DNA digested by HindIII/EcoRI) is run in the first lane of each blot. Frequencies and statistical comparison are reported in Supplementary Table S3.
Table 1. Sequence and in vivo instability of CEB25 allele variants in different contexts, and thermal stability of their associated G4.

<table>
<thead>
<tr>
<th>Allele</th>
<th>Motif sequence</th>
<th>Genomic instability (%)</th>
<th>Thermal stability (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>WT cells</td>
<td>WT cells + Phen-DCj</td>
</tr>
<tr>
<td>CEB25-WT (L191)</td>
<td>AAGGCTGGCTGTAAGTGCTGGGCTGGT</td>
<td>0/192</td>
<td>0/192</td>
</tr>
<tr>
<td>CEB25-L171</td>
<td>AAGGCTGGCTAAGTGCTGGGCTGGT</td>
<td>0/96</td>
<td>0/192</td>
</tr>
<tr>
<td>CEB25-L151</td>
<td>AAGGCTGGCTAAGTGCTGGGCTGGT</td>
<td>1/192 (0.5%)</td>
<td>0/96</td>
</tr>
<tr>
<td>CEB25-L142(TTT)</td>
<td>AAGGCTGGCTTTTTGGGCTGGT</td>
<td>0/96</td>
<td>4/192 (2.1%)</td>
</tr>
<tr>
<td>CEB25-L132(TGT)</td>
<td>AAGGCTGGCTTGTGGGCTGGT</td>
<td>1/192 (0.5%)</td>
<td>0/96</td>
</tr>
<tr>
<td>CEB25-L131(TTT)</td>
<td>AAGGCTGGCTTTTTGGGCTGGT</td>
<td>3/96 (3%)</td>
<td>3/192 (1.6%)</td>
</tr>
<tr>
<td>CEB25-L131(TTT)</td>
<td>AAGGCTGGCTTTTTGGGCTGGT</td>
<td>0/96</td>
<td>4/192 (2.1%)</td>
</tr>
<tr>
<td>CEB25-L113(TTT)</td>
<td>AAGGCTGGCTTTTTGGGCTGGT</td>
<td>0/96</td>
<td>2/192 (1%)</td>
</tr>
<tr>
<td>CEB25-L211(TT)</td>
<td>AAGGCTGGCTTTTTGGGCTGGT</td>
<td>1/384 (0.3%)</td>
<td>66/572 (11.5%)***</td>
</tr>
<tr>
<td>CEB25-L211(TT)</td>
<td>AAGGCTGGCTTTTTGGGCTGGT</td>
<td>0/192</td>
<td>63/380 (16.5%)***</td>
</tr>
<tr>
<td>CEB25-L212(TT)</td>
<td>AAGGCTGGCTTTTTGGGCTGGT</td>
<td>0/192</td>
<td>13/380 (1.4%)***</td>
</tr>
<tr>
<td>CEB25-L211(AA)</td>
<td>AAGGCTGGCTTTTTGGGCTGGT</td>
<td>0/96</td>
<td>15/188 (7.9%)***</td>
</tr>
<tr>
<td>CEB25-L221(TT)</td>
<td>AAGGCTGGCTTTTTGGGCTGGT</td>
<td>0/96</td>
<td>18/192 (9.4%)***</td>
</tr>
<tr>
<td>CEB25-L212(TT)</td>
<td>AAGGCTGGCTTTTTGGGCTGGT</td>
<td>0/96</td>
<td>7/180 (3.9%)**</td>
</tr>
<tr>
<td>CEB25-L212(TT)</td>
<td>AAGGCTGGCTTTTTGGGCTGGT</td>
<td>1/95 (1%)</td>
<td>11/176 (6.3%)**</td>
</tr>
<tr>
<td>CEB25-L222(TT)</td>
<td>AAGGCTGGCTTTTTGGGCTGGT</td>
<td>ND</td>
<td>2/144 (1.4%)</td>
</tr>
<tr>
<td>CEB25-L222(AA)</td>
<td>AAGGCTGGCTTTTTGGGCTGGT</td>
<td>0/96</td>
<td>0/192</td>
</tr>
<tr>
<td>CEB25-L111(T)</td>
<td>AAGGCTGGCTTTTTGGGCTGGT</td>
<td>0/96</td>
<td>42/192 (21.9%)***</td>
</tr>
<tr>
<td>CEB25-L111(T)-G30A</td>
<td>AAGGCTGGCTTTTTGGGCTGGT</td>
<td>0/92</td>
<td>22/96 (22.9%)***</td>
</tr>
<tr>
<td>CEB25-L111(T)-G12T</td>
<td>AAGGCTGGCTTTTTGGGCTGGT</td>
<td>0/96</td>
<td>0/184</td>
</tr>
<tr>
<td>CEB25-L111(C)</td>
<td>AAGGCTGGCTTTTTGGGCTGGT</td>
<td>1/96 (1%)</td>
<td>42/192 (21.9%)***</td>
</tr>
<tr>
<td>CEB25-L111(A)</td>
<td>AAGGCTGGCTTTTTGGGCTGGT</td>
<td>0/96</td>
<td>2/192 (1%)</td>
</tr>
</tbody>
</table>

Underlined: oligo used for Tm measurement (full G4 thermal stability data, see Supplementary Table S4). Bold: modifications relatively to CEB25-WT. All the alleles used to measure genomic instability contain 13 motifs. ND: not determined. *P-value versus WT cells < 0.05. **P-value versus CEB25-WT allele < 0.05.

Mutagenesis of the unstable CEB25-L111 variant

The unstable behavior of CEB25-L111(T) strongly suggests that persistent G4s are formed in vivo. To confirm that CEB25-L111(T) instability depends on G4 folding, we constructed the CEB25-L111(T)-G12T array (Table 1) bearing a single G→T substitution in one of the four G-triplets involved in CEB25 G4 formation in vitro. As expected, this single-point mutation abolished the minisatellite
instability in both Phen-DC₃-treated WT cells and in pifΔ cells (Table 1, and Supplementary Fig S2). Consistently, the single-point mutation of another G-triplet (G30A) not involved in CEB25 G4 formation (Amrane et al, 2012) had no effect on the rearrangement frequencies: The resulting CEB25-L111(T)-G30A allele was stable, but strikingly, the CEB25 variants were stable in Phen-DC₃-treated WT cells (Table 1, and Supplementary Fig S2). These results demonstrate that unlike the natural CEB1 minisatellite sequence (Piazza et al, 2010, 2012), the destabilization of the variant CEB25-L111(T) minisatellite depends on its G4 motif.

**Total loop length and position requirements for CEB25 instability in vivo**

Next, we investigated the granularity of the loop length effect on CEB25 instability. First, we shortened the 9-nt central loop of CEB25 from the S' end to yield the CEB25-L171, CEB25-L151, and CEB25-L131(TGT) variants. Remarkably, these constructs were stable upon Phen-DC₃ treatment and in the absence of Pif1 (Fig 3A, Table 1). Then, we built the CEB25-L121(TT), CEB25-L131(TTT), and CEB25-L141(TTTT) variants homogenized to bear only T in the central loop. Upon treatment of WT cells with Phen-DC₃, the CEB25-L141(TTTT) and CEB25-L131(TTT) variants, like CEB25-L131(TGT), were stable, but strikingly, the CEB25-L121(TT) variant was destabilized (63/380, P-value vs. WT cells = 1.2 × 10⁻¹²), suggesting that the CEB25 variants become significantly unstable when the central loop is less than 3 nt in length (Fig 3B). Consistently, in pifΔ cells, CEB25-L121(TT) was also unstable (26/52, P-value vs. WT cells = 2.7 × 10⁻¹⁵) (Table 1 and Fig 3A). However, in contrast to the stable CEB25-L131(TGT) variant, CEB25-L131(TTTT) was clearly destabilized (16/78, P-value vs. WT cells = 1 × 10⁻⁶), quantitatively slightly less than the CEB25-L121(TT) and much less than the CEB25-L111(T) variant. As well, the CEB25-L141(TTTT) was slightly unstable (12/94, P-value vs. WT cells = 1.5 × 10⁻³). These results indicate that the threshold of instability of the CEB25 variants in the presence of Phen-DC₃ and in the absence of Pif1 is ≤ 2 and 4 nt in length, respectively, but also depends on the nucleotide composition (see below, Fig 3B). This threshold difference between the two conditions might reflect the higher sensitivity of the mutant situation.

Next, we asked whether the position of the longer loop within the G4 motif would affect the instability of CEB25. For this purpose,
we built the CEB25-L311(TTT) and CEB25-L113(TTT) variants in which the 3-nt loop has been moved in the first and third position, respectively. In Phen-DC3-treated cells and pif1Δ cells, both constructs were stable (Fig 3C, Table 1). Similarly, we moved the 2-nt loop in first or third position in CEB25-L211(TT) and CEB25-L112(TT), respectively. Both alleles exhibited a significant increase of instability upon Phen-DC3 treatment (66/572 and 13/380, P-values vs. WT cells = 1.6 × 10−14 and 6.1 × 10−3, respectively) or in the absence of Pif1 (13/77 and 13/91, P-values vs. WT cells = 3.7 × 10−10 and 2.1 × 10−7) (Fig 3C, Table 1). Thus, a single 2-nt loop located at any position within the G4 motif limits but does not preclude CEB25 instability. Quantitatively, bearing the 2- and 3-nt loops in lateral positions is more innocuous for the stability of the array than in the central position (Fig 3C).
Moreover, we examined the impact of the combinatorial presence of several loops of variable length. The addition of a second 2-nt loop (TT) in the CEB25-L212(TT) variants did not abolish CEB25 instability, but decreased it on average 2- to 3-fold compared to the variants bearing only one 2-nt loop in both Phen-DCs and pif1A context, respectively (Fig 3D, Table 1). However, the CEB25-L222(TT) variant bearing three 2 nt loops became stable in these conditions (Fig 3D, Table 1). Hence, each 2-nt loop contributes to a decrease in the destabilizing potential of the G4 motif (Fig 3D).

Altogether, the above experiments uncovered a drastic decrease of the CEB25 G4-dependent instability with an incremental increase of a single loop from 1 to 3 nt and outlined the subtle combinatorial burden of each loop, above which CEB25 remains stable.

All variant sequences form intra-molecular parallel G4 resembling native CEB25

To rationalize the observations above, we investigated the conformational and thermodynamic properties of CEB25 variant oligonucleotides (sequences underlined in Table 1), including: (i) several mutants to probe the effect of central loop shortening by replacing loop sequence with poly-thymine, that is, CEB25-L111(T), CEB25-L121(TT), CEB25-L131(TTT), and CEB25-L141(TTTT) or by truncating natural loop residues from the 5’ side, that is, CEB25-L131(TGT), CEB25-L151, and CEB25-L171 (folds later shown in Fig 4E); (ii) two mutants to assess positional consequence of 3-nt propeller loop within the structure, that is, CEB25-L311(TTT) and CEB25-L113(TTT); (iii) five mutants to address the position and number of 2-nt loops, that is, CEB25-L211(TT), CEB25-L122(TT), CEB25-L221(TT), CEB25-L212(TT), CEB25-L122(TT), and CEB25-L222(TT) (Fig 4F-I); (iv) two mutants to measure the stability of all 1-nt loops with all C or A residues, that is, CEB25-L111(C) and L111(A); and (v) one mutant bearing a mutated G-tract, that is, L111(T)-G12T.

CEB25 (CEB25) forms a parallel-stranded three-layered G4 with three propeller loops of 1, 9, and 1 nt, respectively (Fig 4D) (Amrane et al, 2012). The in vitro formation of a single three-layered G4 structure for all variant sequences was confirmed by NMR spectra showing twelve major imino proton peaks (four for each G-tetrad layer) at ~10-12 ppm (Fig 4A; Supplementary Fig S3) (Adrian et al, 2012). Thermal difference UV absorption spectra (TDS) and CD spectroscopy were used to support G4 formation (Mergny et al, 2005) and to identify their strand orientations (Gray et al, 2008), respectively (Fig 4B and C; Supplementary Fig S4). When dissolved in 1 mM KPi buffer, TDS of each mutant generally showed typical pattern of a G4 structure with a negative minimum at 295 nm and two positive maxima at 240 and 275 nm (Fig 4B; Supplementary Fig S4) (Mergny et al, 2005). Concurrently, CD spectrum of each mutant displayed a positive maximum at 260 nm and a negative minimum at 240 nm, characteristic of a parallel-stranded G-quadruplex (Fig 4C; Supplementary Fig S4) (Gray et al, 2008). The stoichiometry of G4 was deduced based on solvent-exchange protection pattern of its imino proton peaks. For each of the mutants, there were four peaks left after one hour exposure in D2O solvent (which are associated with one well-protected middle G-tetrad layer within a three-layered G4) (Fig 4A; Supplementary Fig S3), thus implying monomeric nature of folded G4. Supported by NMR, UV, and CD data, all variant sequences form intra-molecular parallel-stranded G4 structures, similar to that of native CEB25. Thus, the differential behavior of the CEB25 variants in vivo cannot be explained by conformational change of the G4.

Thermal stability of variant sequences is dependent on loop sizes

The thermal stability of CEB25 and variant G4s was measured from the melting temperatures (Tm) in heating/cooling experiments performed by UV and CD spectroscopy (Table 1, Supplementary Table S4). Parallel G4 containing all 1-nt propeller loops of a pyrimidine residue is known to be extremely stable in physiological salt condition at ~100 mM K+ (Rachwal et al, 2007a; Guedin et al, 2010). Indeed, the melting temperature of L111(T) was above 80°C and could not be accurately determined even at relatively low concentration of potassium cations in 5–20 mM KPi buffer. For this reason, all sequences were dissolved in 1 mM KPi buffer to yield melting temperatures within the sensitive temperature region of CD or UV heating/cooling experiments.

Compared with the native CEB25-L191 sequence that was characterized with a Tm of 55.1°C, a drastic increase of TmUV to 73.4°C was recorded for CEB25-L111(T) (Fig 5A, Table 1) (Rachwal et al, 2007a). The CEB25-L212(TT), CEB25-L131(TTT) and L141(TTTT) sequences were found to have TmUV of 67.9°C, 63.3°C, and 63.9°C, respectively (Fig 5A, Table 1). Adding one thymine to a 1- and 2-nt poly-thymine central loop monotonically decreases melting temperatures by ΔTmUV (1 > 2 nt) = −5.5°C and ΔTmUV (2 > 3 nt) = −4.6°C, respectively. Interestingly, the 4-nt poly-thymine central loop in the L141(TTTT) marginally stabilizes the structure relative to the L131(TTT), that is, ΔTmUV (3 > 4 nt) = 0.6°C. It may result from inter-residue interaction within the longer loop in L141(TTTT). These results were confirmed independently by CD spectroscopy (Fig 5A, Supplementary Table S4).

Other variants CEB25-L131(TTT), L151, and L171 conserved the 3’ end sequence of natural loop, whose two residues (GT) were found to render base pairing interaction with flanking residues at 5’ end of the strand (Amrane et al, 2012). Containing 3-nt central loop of TGT sequence, L131(TGT) has slightly lower TmUV of 61.9°C compared to those of L131(TTT) (ΔTmUV = −1.4°C). Addition of two purine residues to construct a 5-nt central loop of AGTGT sequence such as in L151 lowered TmUV to 59.7°C. Elongation of the central loop to 7 nt with TAAGTGT sequence in L171 produced TmUV of 61.0°C, comparable to those of L151 (Table 2, Fig 5A). Notably, at least one Watson-Crick base pair presumably between A2 and T16 similar to that observed in CEB25-L191 was formed in L171 (Supplementary Fig S5). Indeed, additional hydrogen bond interactions from base pair formation have been shown to raise the thermal stability of CEB25 (Amrane et al, 2012).

The all-thymine loop position (of 2 or 3 nt length) within the G4 barely affects the thermal stability of the structure (Fig 5B, Table 1). The inclusion of double 2-nt thymine loops at different positions such as in L221(TT), L212(TT), and L122 (TT) similarly lowered TmUV to 61.1°C, 62.1°C, and 61.5°C, respectively (Fig 5B, Table 1 and Supplementary Table S4). Dramatic reduction of
thermal stability was observed in all 2-nt thymine loop L222(TT) with TmUV of 54.9°C. As loop position only moderately affects thermal stability, the difference in melting temperatures of L222(TT) and L121(TT) (ΔTmUV = -13.0°C) can be attributed to additive effect of 2-nt thymine loops at three loop positions (Fig 5B, Table 1).
Phen-DC₃ similarly binds and stabilizes CEB25 G4 variants bearing different loop length

Phen-DC₃ exhibits a high affinity and an exceptional selectivity for G4 over dsDNA (De Cian et al., 2007) but poorly discriminates between different G4 conformations (Largy et al., 2011). The recently published NMR structure of the ligand in a 1:1 complex with the c-Myc Pu24'T G4 provides the basis for this universal G4 recognition (Chung et al., 2014). Using the FRET melting method on oligonucleotides [L191, L1311(TTT), L121(TT), and L1111(TT)] labeled with fluorescein and tetramethylrhodamine at the 3’ and 5’ ends, respectively, we verified that Phen-DC₃ binds and stabilized similarly CEB25 G4 variants bearing different central loop length: While the thermal stabilities of the G4 formed by the labeled oligonucleotides are very close to the values measured by UV and CD spectroscopy, addition of 1 molar equivalent of Phen-DC₃ resulted in a stabilization (ΔTₘ) of 9.6°C [for L191 and L1111(TT)] to 13°C [L121(TT)] and 14°C [L1311(TTT)] (Supplementary Table S4). This

Figure 5. CEB25 variant instability correlates to thermal stability of its associated G4.

A, B Thermal stability dependence on loop length and position as measured by UV and CD spectroscopy. All melting temperatures (Tₘ) were obtained in 1 mM KPi buffer at ~4 μM DNA strand concentrations. (A) Thermal stability of CEB25 G4 variants is inversely correlated to the central loop length (P-values obtained using the Spearman correlation test). Other loops are single thymine. The two TₘUV values for a central loop of 3 nt correspond to L1311(TTT) and L121(TT), respectively. (B) Effect of the position of a single 2- or 3-nt-long loop and permutation of two or three 2-nt-long loops on the thermal stability of CEB25 G4 variants. All loop residues are thymine.

C In vivo instability of CEB25 allele variants plotted as a function of the melting temperature of the corresponding G4 measured by UV spectroscopy, in WT cells treated with Phen-DC₃ (left panel) and in pifΔA cells (right panel). P-values and correlation coefficients were obtained using a two-tailed Spearman correlation test.

D Sequence effect of single loop residue substitutions on the thermal stability of the CEB25-L111 G4. Melting temperatures (Tₘ) were obtained in 1 mM KPi buffer at ~4 μM DNA strand concentrations.

E Sequence effect of three 1-nt-long loops on the CEB25 instability in WT cells treated with Phen-DC₃ (left panel, strains ANT1903, ANT1953, and ANT1936) and in pifΔA cells (right panel, strains ANT1917, ANT1974, and ANT1980). Analysis was done as in Fig 1B.
similar increase in stability indicates that Phen-DC₃ binds and stabilizes G₄ bearing different central loop length to similar extents. Since Phen-DC₃ inhibits G₄ unwinding by Pif1 in vitro (Piazza et al., 2010), this similar recognition of G₄ variants by Phen-DC₃ is consistent with the treatment of WT cells that quantitatively phenocopies the absence of Pif1 (Supplementary Fig S6).

**CEB25 variant instability is correlated with thermal stability of their G₄**

The above results reveal a striking correlation between the G₄ thermal stability and the *in vivo* genomic instability of the cognate minisatellite (Fig 5C). However, the total loop length (and hence the overall volume of the structure or the amount of ssDNA in the loops) could be a confounding factor since it is also negatively correlated to the thermal stability (*P* < 5 × 10⁻⁵). To address whether the G₄ thermal stability dictates minisatellite instability *in vivo* independently of the loop length, we substituted all the single thymine residues in the CEB25-L111(T) allele by either cytosine or adenine to yield the CEB25-L111(C) and CEB25-L111(A) sequences, respectively (Fig 5E). While T-to-C substitutions in CEB25-L111(C) had no effect on the thermal stability of the structure (*Tm* = 74.7°C), thymine-to-adenine substitutions in all 1-nt loop structure in CEB25-L111(A) plummeted its melting temperature to 56.5°C [Δ*Tm* = −16.9°C relative to CEB25-L111(T) values] (Fig 5D, Table 1). It highlights the tremendous destabilization effect of purine residue inclusion into G₄ short loops (Rachwal et al., 2007b; Guedin et al., 2008). Strikingly, while the CEB25-L111(C) allele exhibited genomic instability levels very similar to those observed for CEB25-L111(T) in Phen-DC₃-treated WT cells (42/192 in both cases) and in pif1Δ cells (26/39 vs. 21/32), T-to-A substitutions in CEB25-L111(A) abolished the instability in WT-treated cells (2/192) and drastically decreased it in a pif1Δ mutant (7/107) (*P*-value vs. CEB25-L111(T) = 1.1 × 10⁻¹⁴ and 1.9 × 10⁻¹¹, respectively) (Fig 5E, Table 1). To further test the effect of the loop base composition, we also generated the CEB25-L121(AA) variant containing AA in the central loop. This 2-nt substitution decreased the *Tm* of the structure by 2.1°C compared to L121(TT) (Table 1). Consistently, this variant was unstable in both the WT Phen-DC₃- treated cells and in the absence of Pif1, but two- to fourfold less than CEB25-L121(TT) (15/188 vs. 63/380 (*P* = 4.4 × 10⁻³) and 6/45 vs. 26/52 (*P* = 1.8 × 10⁻⁴), respectively) (Table 1). Consistently with CEB25-L222(TT) being stable in any conditions, the CEB25-L222(AA) allele exhibited no instability (Table 1). We conclude that the base composition of the loop is another determinant that affects G₄-dependent CEB25 instability. The lower thermal stability of the G₄ folds containing A instead of C or T residues strongly suggests that G₄ thermal stability, but not the overall volume or amount of ssDNA in loops, is a direct determinant of the sequence instability *in vivo*.

**Single pyrimidine loop G₄ motifs are particularly at risk for genomic stability in other eukaryotic genomes**

Our study in *S. cerevisiae* points to G₄ motifs bearing short pyrimidine (C or T) loops as being at higher risk for genomic stability than those bearing short purine loops. It prompted us to examine the diversity of the potential G₄ motifs in other organisms. We determined single-nucleotide loop G₄ motifs (hereafter referred to as G₄L1 motifs, listed in Supplementary Table S5) and studied their base composition (Supplementary Table S6) in the *S. cerevisiae*, *S. pombe*, *C. elegans*, and human genomes.

The classical consensus ([G₃₋₅N₁₋₇G₃₋₅N₁₋₇G₃₋₅N₁₋₇G₃₋₅]₃) used to mine genomes for G₄-prone sequences (Huppert & Balasubramanian, 2005; Todd et al., 2005) identifies 27 and 30 motifs in the *S. cerevisiae* and the *S. pombe* genomes, respectively (Supplementary Fig S7A and B, Supplementary Table S5). Among those, only 3 and 2, respectively, bear single-nucleotide loops only, that all contain the most innocuous purine loops (Supplementary Fig S7B). Consequently, both yeasts genomes are devoid of the most detrimental G₄L1 motifs. The *C. elegans* genome contains 2,226 G₄-prone sequences, among which 1,172 match the G₄L1 motif (Fig 6A). Strikingly, the peculiarly high prevalence of mono-G-runs in the *C. elegans* genome accounts for 98% (1,153/1,172) of these motifs (95.6% perfect and 19.7% imperfect (e.g., bearing a single interrupting nucleotide)). Poly-G sequence G15 has been shown to form a propeller-type parallel G-quadruplex containing three single-residue guanine loops (Sengar et al., 2014). Overall, the *C. elegans* genome contains only 10 G₄L1 motifs bearing ≥2 pyrimidines, two of which are in essential genes (Fig 6A, Supplementary Table S5). This is 117-fold less than purine-rich monoG G₄L1 motifs. In the human genome, among the 376,000 G₄ motifs identified (Huppert & Balasubramanian, 2005; Todd et al., 2005), 18,153 are G₄L1 motifs. With the same base probabilities (mean human genome GC content of 41%), G₄L1 motifs containing only A loops are 11.1-fold more prevalent than those bearing only T loops, and G-containing motifs are 4-fold more prevalent than those bearing only C loops (Fig 6B). The trend is the same for G₄L1 motifs bearing non-identical loops 3.7-fold more G₄ motifs containing purine loops only over those bearing pyrimidine loops only (Fig 6B). This depletion is more pronounced in the repeated regions (Supplementary Fig S7C). In conclusion, the more detrimental pyrimidine-containing G₄L1 motifs are either absent (*S. cerevisiae* and *S. pombe*) or strongly under-represented compared to the purine-containing ones (*C. elegans* and human).

Then, we tested our prediction that pyrimidine-containing G₄L1 motifs would be more prevalent at sites of damage or rearrangement than purine-containing ones or than G₄ motifs bearing longer loops. First, we mapped the location of the 100–200-bp deletions that arise in *C. elegans* animals deficient for the dog-1 (Deletion Of G-rich DNA-1) helicase, ortholog of the G₄-unwinding FANC-J helicase (Kruisselbrink et al., 2008). The authors identified a total of 69 deletions (among which 65 were non-recurrent), all present at G₄ motifs. The majority (62/65) fell at G₄L1 motifs: 61 at perfect or almost perfect mono-G-runs (61/1,153, 5.3%) and one at a non mono-G motif (1/19, 5.3%) (Fig 7A). The three remaining deletions occurred at G₄ motifs that had two single-nt loops and one loop ≥1 nt (3/1,054, 0.3%) (Supplementary Fig S7D). Thus, the G₄L1 motifs are 18.6-fold more often affected by deletions than G₄ motifs bearing a single loop ≥1 nt (*P*-value = 1.8 × 10⁻¹⁴, two-tailed Fisher’s exact test), consistent with our findings in yeast.

The second G4-related study that we re-analyzed concerns the location of the DNA damage signaling marker phospho-γH2AX in human cells treated with the G₄-ligand pyridostatin (Rodriguez et al., 2012). Precisely, we mined the G₄L1 motifs loop composition...
in the 1,214 genes (proto-oncogenes and tumor suppressor genes) analyzed for the presence of γH2AX ChIP-Seq peaks (290 γH2AX-positive and 924 γH2AX-negative genes, see Materials and Methods, Supplementary Table S7). In agreement with our prediction, pyrimidine-containing G4L1 motifs are strongly enriched in the γH2AX-positive versus γH2AX-negative genes, while purine-containing G4L1 motifs are not (Fig 6D). This is true for G4L1 with both identical loops (7.3- and 3.6-fold for T- and C-containing loops \( P = 1.33 \times 10^{-9} \) and 0.033, respectively) vs. 1.1- and 1.6-fold for A- and G-containing loops \( P = 0.37 \) and 0.24, respectively) and non-identical loops (9.1- vs. 0.8-fold for pyrimidine- versus purine-containing loops, \( P = 1.58 \times 10^{-4} \) and 0.81, respectively) (Fig 6D). Conversely, γH2AX-positive genes were strongly enriched over γH2AX-negative genes for pyrimidine-containing (4.9-fold increase, \( P = 1.13 \times 10^{-8} \)) but not purine-containing (1.3-fold, \( P = 0.18 \)) G4L1 motifs (Fig 6E).

Thus, our analysis of the prevalence and loop composition of single-nt loop G4 motifs in these eukaryotic genomes and their association with DNA damage and genome rearrangement phenotypes in \( C\). \( e\)\( la\)\( g\)\( a\)\( s\) and human cells upon inhibition of G4 unwinding show that the rules dictating the instability of a G4 motifs determined in our model yeast system can be generalized to other evolutionary distant organisms.

**Discussion**

The G4 loops modulate minisatellite instability

In this study, we sought to decipher the heterogeneous instability phenotype of several G4-forming arrays in yeast. Like CEB1, we
found that the c-Myc tandem array was frequently rearranged but not the CEB25-WT, c-Kit, and Bcl2-MBR sequences that also form G4 in vitro (Phan et al., 2004, 2007; Ambrus et al., 2005; Todd et al., 2007; Kumar & Maiti, 2008; Nambari et al., 2011; Ambrane et al., 2012; Wei et al., 2012). The molecular determinants of this behavioral discrepancy reside in the G4 loops. Extensive mutagenesis of CEB25 G4 motif uncovered four determinants (detailed in Fig 7) that dictate sequence instability in vivo: (i) the length of a single loop that connects the G-stretches, (ii) the position of the longest loop, (iii) the total number of nucleotide in the loops, and (iv) the base composition of the loops.

The CEB25 rules are consistent with the unstable behavior of the CEB1 and c-Myc G4 sequences that exclusively contain loops of 1 or 2 nt (Phan et al., 2004; Ambrus et al., 2005; Adrian et al., 2014) and the stability of the c-Kit and Bcl2-MBR sequences that have two loops ≥ 4 nt (Fig 1A) (Phan et al., 2007; Todd et al., 2007; Nambari et al., 2011; Wei et al., 2012). Hence, our extensive mutagenesis study narrows the fraction of destabilizing G4-forming sequence to those matching the following consensus: G_\text{N}_xG_\text{N}_yG_\text{N}_zG_\text{w}, where N are preferentially pyrimidines, x, y ≤ 2, y ≤ 4, and x + y + z ≤ 7 nt.

Our biophysical studies demonstrated that all the CEB25 variant sequences having loops of 1 to 9 nt retained a single major intramolecular parallel G4 conformation (Fig 4). Thus, their distinct and continuous in vivo behavior cannot be explained by a drastic conformational change in the structure of the G4. Rather, we uncovered that their thermodynamic stability greatly differed (varying over 25°C, for the CEB25 variants in 1 mM K⁺) in a trend inversely correlated with the loop length (Guedin et al., 2010) (Fig 5A). Overall, our CEB25 in vitro data regarding the loop length and sequence, as well as the effect of Phen-DC₃ binding on G4 stability, are consistent with previous observations on other G4 (Rachwal et al., 2007a; Guedin et al., 2008, 2010; Agrawal et al., 2013; Tippana et al., 2014). Thus, we conclude that the G4 thermodynamic stability is a key determinant for their formation and persistence in vivo and thereof of their capacity to trigger the genomic instability of the arrays during replication by acting as a stable roadblock for the replicative polymerase (Lopes et al., 2011).

Notably, most of the CEB25 variants bearing a single loop of 3 nt remain stable in vivo, even though their associated G4 Tₘ are slightly higher than those of the unstable CEB25 variants bearing two loops of 2 nt (Fig 5B, and compare orange and blue instabilities in Fig 5C). This observation may suggest the existence of additional in vivo factors ensuring the genomic stability of the underlying sequence when a loop ≥ 3 nt is present. We envision that a G4-induced phenotype (in our case genomic instability) can be regulated by subtle changes in the G4 loops, either smoothly when acting on the structure stability below a certain loop length (≤ 2 nt), or more sharply when it exceeds this threshold. This can make G4 both versatile switches and fine-tunable regulators of discrete processes at an evolutionary time scale.

Narrowing the fraction of G4 motifs ‘at risk’ for genomic stability

The present study strongly suggests that the threat posed by short-loop G4 to genomic stability is a recurrent feature and is not limited to tandem repeats. In yeasts, the rarity of G4L1 could be explained by an evolutionary counter-selection. In contrast, the remaining presence of robust G4 motifs containing short loops in the C. elegans and human genomes (Huppert & Balasubramanian, 2005) suggests their beneficial role in other essential processes such as the regulation of gene expression. Perhaps to be evolutionary maintained, they preferentially require specialized binding or unwinding proteins to temper their potential to generate damage and rearrangements during replication. Differently, the presence of G4L1 motifs in tandem arrays aggravates the risk of instability (Lopes et al., 2011; Piazza et al., 2012). Likewise G4-forming
microsatellites of the form \((GGGN)_{>8}\) (related to our CEB25-L111 series) are particularly under-represented in the human genome, and the decreasing number of \((GGGA)_{>8} \geq (GGGT)_{>8} \geq (GGGC)_{>8}\) (599, 4 and zero occurrences, respectively) (Bacolla et al., 2008) correlates with the decreasing level of G4-induced instability in our yeast system (Fig 5E). This under-representation of \((GGGN)_{>8}\) sequences suggests that, at the evolutionary time scale, tandem arrays of such structures are prone to rearrange even in cells proficient for their unwinding, and drift toward shorter arrays with greater stability.

Notably, the telomeric sequence of almost all eukaryotes is tandem repeats, up to several kb in length, bearing the conserved ability to form G4 in vitro (Tran et al., 2011) but composed of a G-triplet accompanied by 2–4 other nucleotides, never single nucleotides. In light of our study, this conserved ability to form telomeric G4 of moderate stability (Tran et al., 2011) provides a useful compromise between the requirement for the structure in the biology of telomeres (as documented for ciliates, reviewed in Lipps & Rhodes, 2009) and the threat it may pose for the stability of the array. It might explain why, despite a considerable enrichment for G4 motifs at telomeres, G4 ligands such as pyridostatin did not induce a high level of damage at telomeres compared to interstitial clusters of G4 motifs (Rodriguez et al., 2012). On the contrary, sequences forming highly stable G4 are mostly present in a non-repeated fashion (Huppert & Balasubramanian, 2005; Bacolla et al., 2008), likely limiting their propensity to induce genome rearrangements.

Remarkably, G4-induced instability could in some instances be positively selected, as it may be exploited as a rudimentary inducer of genetic diversity: For example, the only short-loop G4 (identical to the one in CEB25-L121(TT)) in the genome of the bacteria *Neisseria gonorrhoeae* is located in the promoter of the pilin expression locus *pilE* and stimulates its recombination on polymorphic *pIL* pseudogenes, thus promoting antigenic variation (Cahoon & Seifert, 2009).

Having delineated the fraction of G4 motifs that are the most ‘at risk’ to trigger genome instability raises the question of how robust is our overall capacity to predict the existence of G4 structures from genomic sequences. Mostly based on biophysical studies on G4 structures formed by oligonucleotide in vitro, the G4 consensus motif of the form \(G_{x}C_{y}G_{z}C_{y}G_{z}\), where \(x\), \(y\), and \(z\) define the loop length, has largely been used in G4 prediction algorithms (Hazel et al., 2004; Huppert & Balasubramanian, 2005; Todd et al., 2005; Rachwal et al., 2007a; Kumar & Maiti, 2008; Guedin et al., 2010). A reasonable compromise between sensitivity and robustness consisted in restricting each loop to 7 nt (Huppert & Balasubramanian, 2005; Todd et al., 2005; Guedin et al., 2010), which identifies only 27 potentially G4-forming sequences in the *S. cerevisiae* genome.

Differently, Capra et al., by relaxing the loop length constraint to 25 nt each, identified 552 and 446 potential G4 sequences in the *S. cerevisiae* (Capra et al., 2010; Paeschke et al., 2011) and *S. pombe* (Sabouri et al., 2014) genomes, respectively. On the opposite side, the present data, allowing a maximum loop length of 3 nt, would call for only four G4 motifs in each yeast, all being isolated sequences bearing the most innocuous purine loops (Supplementary Fig S7). Thus, how many *S. cerevisiae* and *S. pombe* sequences really form a G4 able to create a replication impediment remains uncertain, but likely very few. If so, the enrichment of Pif1/Pfh1 binding at numerous potential G4 sequences defined with loops of 25 nt (i.e., 138 and 90) in the *S. cerevisiae* and *S. pombe* genomes, respectively (Paeschke et al., 2011; Sabouri et al., 2014), would suggest that other prominent factors than G4-forming capacity are at play. Along the same line, the genome-wide mapping of fragile sites in yeast cells exhibiting reduced levels of Polz revealed no association with potential G4 motifs (Song et al., 2014) with loops \(\leq 7\) nt or \(\leq 12\) nt each. However, a significant association was found using up to 25 nt as loop length, even when the sequences from the more stringent datasets were removed (Song et al., 2014), suggesting that a non-G4 confounding factor causes fragility.

In conclusion, we described the heterogeneous behavior of G4-forming sequences in yeast and identified their underlying structural and biophysical specificities. G4 loops, in correlation with the thermodynamic stability of the structure, appear as the main determinants. We also highlighted the risk of assuming the reliance of a phenotype on G4 structures solely based on the ability of a sequence to adopt such structure in vitro or be called by a relaxed bioinformatics prediction. Our efforts strongly advocate for more analytical G4 prediction algorithms and a thorough validation of the G4-dependent phenotype by combining, for example, mutagenesis of the G4 motif and enhancement of the phenotype with specific G4-stabilizing molecules.

## Materials and Methods

### Media

Liquid synthetic complete (SC) and solid yeast–peptone–dextrose (YPD) media have been prepared according to standard protocols (Tresco & Lundblad, 2001). SC media containing Phen-DC at 10 μM have been prepared as described previously (Piazza et al., 2010).

### Strains

Relevant genotypes of the *Saccharomyces cerevisiae* strains used in this study are listed in Supplementary Table S2. Strains with minisatellites inserted near ARS305 were derived from SY2209 (W303 RAD5+ background) (Fachinetti et al., 2010) by regular lithium acetate transformation, as described in Lopes et al. (2011). Briefly, minisatellites have been inserted near ARS305, in the intergenic region between YCL048w and YCL049c (precisely at chrII:41801-41840, yielding a small deletion of 39 bp), by replacement of a *URA3-hphMX* cassette in the strain ORT6143-13 (WT) or ORT7178-5 (pif1Δ). The minisatellite is oriented on the chromosome in order to have its G-rich strand on the Crick molecule (e.g., template for the leading machinery of forks emanating from ARS305, see orientation 1 in Fig IA in Lopes et al. (2011)). Correct integration and minisatellite size are verified by Southern blot. Alternatively, the PIF1 gene was deleted by transformation of a *pif1::His3* cassette after integration of the minisatellite. Correct *PIF1* deletion is verified by Southern blot using a probe external to the transforming fragment. The presence of the parental minisatellite size is also verified by Southern blot in the transformant.

### Minisatellite synthesis

The *CEB1-WT* (CEB1-WT-1.0 in Piazza et al. (2012)), *CEB1-loop-CEB25*, *CEB1-loopCEB25-m*, and *CEB25-WT* (CEB25-WT-0.7 in
Piazza et al (2012)) minisatellites have been synthesized using homemade PCR-based method described in Ribeyre et al (2009). Other minisatellites have been synthesized by GenScript. Minisatellites size, sequence, and GC content are listed in Supplementary Table S1.

**Measurement of minisatellite instability**

Minisatellite instability during vegetative growth has been measured as previously described in WT cells and pif1Δ cells (Ribeyre et al, 2009), and Phen-DC3-treated WT cells (Lopes et al, 2011). Briefly, untreated WT cells and pif1Δ cells from a fresh patch of cells made from a single colony bearing the parental allele size (checked by Southern blot) are diluted in 5 mL of YPD (2 × 10^5 cells/ml), grown for 8 generations at 30°C with shaking, and spread as single colonies on YPD plates. The instability measurement in these cells thus corresponds to the rearrangement frequency after 45–50 generations. To measure minisatellite instability upon Phen-DC3 treatment, WT cells from a fresh patch on YPD were grown for 8 generations at 30°C in liquid SC containing Phen-DC3 at 10 μM (Lopes et al, 2011). Isolated colonies or pools of colonies are analyzed by Southern blot using the EcoRI digestion that cut at each side of the minisatellite. The membranes are hybridized with a probe corresponding to the minisatellite of interest. The signals are detected with a Typhoon PhosphorImager (Molecular Dynamics). The elimination of potential early clonal events (that occurred early during the colony growth before liquid culture) has been performed as described in Lopes et al (2011). In mutant strains with very high minisatellite instability (for example, CEB25-L111(T) in the pif1Δ mutant), the probability of obtaining two independent rearrangements of the same size is high. Therefore, the removal of rearrangements of the same size (suspected early clonal events) leads to an underestimation of the real rearrangement frequency. To more accurately determine the minisatellite instability in these highly unstable trains, the rearrangement frequency has been determined with fewer colonies (12–24) but on a higher number of independent clones.

**DNA oligonucleotide preparation**

DNA oligonucleotides (sequences see Table 1 or Supplementary Table S4) were chemically synthesized on an ABI 394 DNA/RNA synthesizer. Oligonucleotides were purified and dialyzed successively against potassium chloride solution and water. Oligonucleotides were dissolved both in 1 mM potassium phosphate buffer (pH 7) and in 20 mM potassium phosphate buffer containing 70 mM potassium chloride (pH 7). DNA concentration was expressed in strand molarity using a nearest-neighbor approximation for the absorption coefficients of the unfolded species (Cantor et al, 1970).

**Thermal difference spectra**

Thermal difference spectra (TDS) were obtained by taking the difference between the absorbance spectra from unfolded and folded oligonucleotides that were, respectively, recorded much above (90°C) and below (20°C) its melting temperature (T_m). TDS provide specific signatures of different structural conformations (Mergny et al, 2005). The DNA oligonucleotides at approximately 4 μM strand concentrations were prepared in 1 mM potassium phosphate buffer (pH 7). Spectra were recorded between 220 and 320 nm on a JASCO V-650 UV/Vis spectrophotometer using 1-cm pathlength quartz cuvettes. For each experiment, an average of three scans was taken, and the data were zero-corrected at 320 nm.

**Circular dichroism**

Circular dichroism (CD) spectra were recorded on a JASCO-810 spectropolarimeter using 1-cm pathlength quartz cuvettes. The DNA oligonucleotides at approximately 4 μM strand concentration were prepared in 1 mM potassium phosphate buffer (pH 7). For each experiment, an average of three scans was taken, the spectrum of the buffer was subtracted, and the data were zero-corrected at 320 nm.

**UV/CD melting experiments**

The thermal stability of G4 structures formed by oligonucleotides was characterized in heating/cooling experiments by recording the UV absorbance at 295 nm and the CD ellipticity at 260 nm as a function of temperature (Mergny et al, 1998) using a JASCO V-650 UV/Vis spectrophotometer and a JASCO-810 spectropolarimeter, respectively. UV/CD melting experiments were conducted as previously described in Mergny and Lacroix (2003) at constant DNA strand concentrations of approximately 4 μM in 1 mM potassium phosphate buffer (pH 7). The heating and cooling rates were 0.2°C/min. Experiments were performed with 1-cm pathlength quartz cuvettes.

**NMR spectroscopy**

NMR experiments were performed on 600 MHz Bruker spectrometers at 25°C. The strand concentration of the NMR samples was typically 0.2–0.6 mM both in 1 mM potassium phosphate buffer (pH 7) and 20 mM potassium phosphate buffer containing 70 mM potassium chloride (pH 7). NMR spectra were zero-referenced to resonance of DSS compound.

**FRET melting**

Stabilization of compounds with quadruplex structure via FRET melting assay was performed in a 1.4-ml quartz cell in a fluorescence Cary Eclipse spectrophotometer with a 4-position Peltier effect thermostated cell holder. FRET melting assay was carried out with oligonucleotides equipped with FRET partners at each extremity: fluorescein/FAM molecule at 5’ end and tetramethylrhodamine (TAMRA) at 3’ end. G4-DNA oligonucleotides were prepared by heating the corresponding sequence at 90°C for 5 min in a 10 mM lithium cacodylate buffer (pH 7.4) with 1 mM KCl/99 mM LiCl, and cooling in ice for 30 min to favor the intramolecular folding by kinetic trapping. After addition of Phen-DC3 (0.2 μM), the final volume is 800 μl. Measurements were made with excitation at 492 nm and detection at 516 nm while heating at 25°C for 5 min and then from 25°C to 95°C at a 1°C/min rate.

**Bioinformatics analyses of G4L1 motifs**

The G4L1 motifs in the C. elegans (assembly 235, accessed from Ensembl on 01/30/2015) and human (GRCh38, accessed from the
USCS Web site on 01/20/2015) genomes were determined using custom scripts (available upon request) under R 2.13.1 (R Development Core Team, 2011). To avoid bias induced by the high prevalence in the human genome of tandem repeats of the form (GGGN)$_{8}$ (that match two or more G4L1 motifs), especially (GGGA)$_{8}$ (539 occurrences) (Bacolla et al., 2008), we distinguished G4L1 motifs belonging to unique regions versus repeated regions of the genome (3,542 and 14,438 in human, respectively, and 1,713 overlapping the junction of the two regions) (Supplementary Fig S7C). Repeated sequences were determined by UCSC with RepeatMasker and Tandem Repeats Finder with periodicities ≥ 12 bp and soft-masked in the GRCh38 genome assembly. We only counted non-overlapping identical G4L1 motifs, and we did not merge identical overlapping motifs (for example, the (GGGA)$_{7}$GGG sequence will be scored as two consecutive G4L1 motifs, not a single merged one nor five partially overlapping ones). However, overlapping motifs with different loop sequences are both scored (for example, GGGAGGGAGGTTGGAGGG will count for two G4L1 motifs, one with loops A-A-T and one with loops A-T-A). Overlaps are indicated for each G4L1 motif (Supplementary Table S5). The G4L1 motif loop composition of the C. elegans and human genomes is provided in Supplementary Table S6. Overall, in both C. elegans and human, a minor fraction of G4L1 motifs (16%) is considered overlapping (190/1,172 in the C. elegans genome and 2,859/18,153 in the human genome). Most of these overlaps occur in tandem repeats: In the C. elegans and human genomes, respectively, all (190/190) and 87% (2,485/2,859) of the overlapping sequences fell in the repeated portion of the genome, or at junctions between unique and repeated regions. In C. elegans, 188/190 are monoG-runs.

We also provide in Supplementary Table S5 the lists of G4 motifs with individual loops of 1–7 nt, which have been downloaded from QuadDB (now offline) (Wong et al., 2010) on 04/03/2012 (S. cerevisiae assembly 62) and 03/08/2012 (C. elegans assembly 180 and H. sapiens GRCh36). We determined the S. pombe G4 motifs using QGRS mapper (Kikin et al., 2006) using the assembly 294 on 01/31/2015.

Re-analysis of deletion breakpoint location in dog-1-deficient C. elegans

We used the list of G4 motifs and monoG-runs found at deletion breakpoints provided in Table S1 in the original study by Kruisell-brink et al. (2008). We manually determined the smaller possible G4 motif in each sequence. Non-monoG-run G4 motif sequences are presented in Supplementary Fig S7D. A two-tailed Fisher’s exact test was used to compare the proportion of affected monoG-runs and consensus G4 motif.

Re-analysis of pyridostatin-induced γH2AX signal

Phospho-γH2AX ChiP-Seq data following pyridostatin treatment of SV40-infected MRC-5 fibroblast cells have been obtained from Rodriguez et al. (2012). The study focused on a subset of 1,224 genes (482 proto-oncogenes and 742 tumor suppressors) for which a qualitative H2AX score was attributed (‘yes(**),’ ‘yes(*),’ ‘yes,’ ‘yes/no’, and ‘no’; Supplementary Dataset 2 in Rodriguez et al. (2012)). Using gene names of Supplementary Dataset 2 and custom scripts, we could retrieve the GRCh37 coordinates and G4 motif content from Supplementary Dataset 3 in Rodriguez et al. (2012). Next, these coordinates were lifted-over to the GRCh38 release using the online Ensembl lift-over tool, and duplicated entries were manually curated to obtain a final list of 1,214 genes (479 proto-oncogenes and 735 tumor suppressors) and their associated coordinates, density of G4 motifs (or PQS for potential quadruplex sequence, loops 1–7 nt), and H2AX score (Supplementary Table S7). We then measured the intersection between G4L1 motifs of different loop composition and H2AX-positive (‘yes’ to ‘yes(**),’ score 1–3) and H2AX-negative (‘no’ and ‘yes/no’, score 0) genes, in order to determine (i) the enrichment for certain G4L1 motifs in γH2AX-positive vs. γH2AX-negative genes (Fig 6D) and (ii) the enrichment for H2AX signal in genes containing G4L1 motifs bearing certain loops (Fig 6E). For simplicity, we considered only G4L1 motifs bearing either 3 purine loops or 3 pyrimidine loops. In each case, enrichment was normalized to the total size of the genes. Proportion of G4L1 motifs or of G4L1 motif-containing genes in the γH2AX-positive and γH2AX-negative classes were compared using a two-tailed Fisher’s exact test.

Statistical analysis

Rearrangement frequencies have been compared using a two-tailed Fisher’s exact test. Correlations between Tm and in vivo instability, loop size and Tm, and between instabilities were determined using a two-tailed Spearman non-parametric correlation test. Statistical cutoff has been set to 0.05. All statistical tests have been performed using R2.13.1 (R Development Core Team, 2011) or GraphPad Prism 4.03.

Supplementary information for this article is available online: http://emboj.embopress.org

Acknowledgements

We thank past and present members of our laboratories for helpful discussions. AP received fellowships from the Ministère de l’Education Nationale, de la Recherche, et de la Technologie (MENRT), and from the Association pour la Recherche sur le Cancer (ARC). AS was supported by a postdoctoral fellowship from the Fondation pour la Recherche Médicale (FRM). MA was supported by the Youcef Jameel scholarship. This research was supported by Grant ANR-12-BSV6-0002 (to AN and MPTF), Singapore Ministry of Education Academic Research Fund Tier 3 (MOE2012-T3-1-001) and Nanyang Technological University grants (to ATP) and the Singapore-France Merlion grant (to ATP and AN).

Author contributions

AP, MA, FS, AS, JL, FH, ATP, and AN designed the experiments. AP, MA, FS, BH, FH, AS, and JL performed the experiments. AP, MA, FS, BH, FH, AS, JL, MPTF, AH, ATP, and AN analyzed the data. AP performed the bioinformatics analyses. AP, MA, and AN wrote the manuscript with contributions by BH, FH, ATP, and MPTF.

Conflict of interest

The authors declare that they have no conflict of interest.

References


Chirality 20: 431–440


Determinants of G4-induced genomic instability

Aurèle Piazza et al


