



**HAL**  
open science

# Rôle des G-quadruplexes dans l'instabilité génomique chez *Saccharomyces cerevisiae*

Aurele Piazza

► **To cite this version:**

Aurele Piazza. Rôle des G-quadruplexes dans l'instabilité génomique chez *Saccharomyces cerevisiae*. Biochimie, Biologie Moléculaire. Université Pierre et Marie Curie - Paris VI, 2012. Français. NNT : 2012PAO66445 . tel-00833453

**HAL Id: tel-00833453**

**<https://theses.hal.science/tel-00833453>**

Submitted on 12 Jun 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**THESE DE DOCTORAT DE  
L'UNIVERSITE PIERRE ET MARIE CURIE**

Spécialité :

**Biologie moléculaire**

École doctorale :

**Complexité du Vivant**

Présentée par :

**M. Aurèle Piazza**

Pour obtenir le grade de

**DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE**

**"Rôle des G-quadruplexes dans l'instabilité génomique chez**

***Saccharomyces cerevisiae*"**

soutenue le 28 Septembre 2012

devant le jury composé de :

M. Pierre Netter	Président du jury
M. Alain Nicolas	Directeur de thèse
M. Philippe Pasero	Rapporteur
M. David Shore	Rapporteur
M Benoit Arcangioli	Examineur
M Anh-Tuan Phan	Examineur



## RESUME

Les G-quadruplexes sont des structures à quatre brins formées spontanément par certains acides nucléiques riches en guanines dans des conditions physiologiques *in vitro* et qui présentent une grande variété de conformations. Ma thèse a consisté à montrer qu'ils se formaient *in vivo* et à étudier les conséquences pathologiques de leur persistance.

Le système expérimental utilisé au cours de ma thèse repose sur la mesure de la stabilité de séquences répétées en tandem humaines de type minisatellite chez *S. cerevisiae* en croissance mitotique. Le motif unitaire du minisatellite CEB1 forme un G-quadruplexe, efficacement déroulé par l'hélicase Pif1 *in vitro*. Dans un mutant déficient pour l'hélicase Pif1, CEB1 est fréquemment réarrangé (expansion ou contraction). J'ai dans un premier temps participé à montrer que cette instabilité de CEB1 dépendait de ses motifs G-quadruplexes, par mutagenèse dirigée. J'ai confirmé ce résultat par une seconde approche, en traitant des cellules sauvages avec un ligand spécifique des G-quadruplexes. Ce ligand induit spécifiquement l'instabilité du minisatellite CEB1, mais n'a pas d'effet sur le minisatellite CEB1 muté. Ces outils expérimentaux m'ont ensuite permis, en collaboration avec J. Lopes, de montrer que l'instabilité du minisatellite survient durant la réplication, lorsque les G-quadruplexes sont formés sur la matrice du brin à synthèse continue. Cette capacité à former des G-quadruplexes, entre autres propriétés des minisatellites riches en GC, induit la formation de grands réarrangements chromosomiques. Le dernier volet de ma thèse a consisté en l'étude de la relation structure-fonction des G-quadruplexes à l'instabilité génomique.

## ABSTRACT

G-quadruplexes are topologically diverse four-stranded structures formed by certain G-rich nucleic acids in physiological conditions *in vitro*. During my thesis, I studied their *in vivo* formation, and the pathological consequences of their persistency.

The experimental system used during my thesis rely on the measurement of the stability of human tandemly-repeated sequences called "minisatellite" in mitotically growing *S. cerevisiae*. The unitary motif of CEB1 form a G-quadruplex *in vitro*, efficiently unwound by the purified Pif1 helicase. In a Pif1-deficient strain, CEB1 is frequently rearranged (expansion and contraction). At first, we showed that this CEB1 instability was dependent on its G-quadruplex-forming motifs, by directed mutagenesis. I confirmed this result by treating cells with a specific G-quadruplex ligand: it specifically induces the instability of CEB1, but not its G-mutated counterpart. In collaboration with J. Lopes, I showed that the CEB1 instability occurred during replication, upon formation of G-quadruplexes in the leading strand template. The ability to form G-quadruplexes, among other GC-rich minisatellite features, induces gross chromosomal rearrangements. The last aspect of my thesis consisted in the study of the structure-function relationship of G-quadruplexes and genomic instability.

**Titre:** Rôle des G-quadruplexes dans l'instabilité génomique chez *Saccharomyces cerevisiae*

**Title:** Role of G-quadruplexes in genomic instability in *Saccharomyces cerevisiae*

**Mots clés:** G-quadruplexe, Pif1, instabilité génomique, réplication, minisatellite

**keywords:** G-quadruplex, Pif1, genomic instability, replication, minisatellite

**Adresse du laboratoire:**

Institut Curie - Centre de Recherche

Pavillon Trouillet, UMR3244, 5<sup>e</sup> étage Laboratoire Nicolas

26 rue d'Ulm

75248 Paris Cedex 05

France

# Remerciements

J'adresse mes sincères remerciements aux membres de mon jury de thèse qui ont pris le temps de lire et d'évaluer ce travail: mes rapporteurs Philippe Pasero et David Shore, mes examinateurs Benoit Arcangioli et Anh-Tuan Phan, ainsi que le président du jury Pierre Netter.

\*\*\*

Cette thèse, aboutissement de 4 années de travaux, n'est évidemment pas de mon seul fait. J'aimerais remercier ici ceux qui, directement ou indirectement, ont contribué à sa réalisation.

Je tiens en premier lieu à remercier Alain Nicolas pour m'avoir accueilli dans son laboratoire et fournit les conditions matérielles de la réalisation de ce travail. Matérielle, mais pas uniquement. Alain Nicolas m'a transmis son goût pour une science sûre, solide, pérenne. Une science certes plus lente, plus austère, dont la satisfaction ne tient pas aux fragiles effets d'annonce mais à l'assurance de paver un chemin sûr pour son propre futur et celui des autres.

Je voudrais tout particulièrement remercier Judith Lopes, de qui j'ai énormément appris pendant ces années au laboratoire. Quantifier l'importance de Judith dans le succès de ma thèse, en plus d'être impossible, serait inutile. L'évidence est qu'elle y a très largement contribué, que ce soit par la formation complète et rigoureuse qu'elle m'a prodiguée ou comme l'excellente collaboratrice qu'elle a été sur nos projets communs.

Je tiens aussi à remercier Jean-Baptiste Boulé, dont la passion communicative pour la science a continuellement irrigué la mienne. Plus encore, il en a épuré les appareils factices pour m'en révéler la substantifique moelle, le sens, qui au-delà des calculs alimentaires doit guider tout scientifique.

Cette thèse a fait l'objet de nombreuses collaborations avec d'excellents scientifiques exerçant dans différents domaines de la biologie que je tiens à remercier chaleureusement. Concernant l'étude biophysique des G-quadruplexes, j'ai une pensée particulière pour Jean-

Louis Mergny qui a participé aux études initiales de la formation des G-quadruplexes de CEB1, et dont l'intelligence et la clarté ont guidé le laboratoire dans ce qui devint mon sujet de thèse. Plus récemment, Anh-Tuan Phan et son excellent étudiant Michael Adrian ont déterminé les impressionnantes structures tri-dimensionnelles formées par les G-quadruplexes de CEB1 et CEB25, conférant à mon projet de thèse une dimension supplémentaire, passionnante: celle de l'étude de la relation structure-fonction des G-quadruplexes à l'instabilité génomique. Concernant l'étude *in vivo*, je ne pourrais jamais remercier assez Marie-Paule Teulade-Fichou, David Monchaud, et Eric Largy pour avoir développé et m'avoir fourni cet extraordinaire outils pour ma thèse que fut le ligand de G-quadruplexe Phen-DC<sub>3</sub>, ainsi que pour la qualité de notre si fructueuse collaboration. J'aimerais également remercier Katie Mingo et Barry Kriegsman, deux excellents étudiants américains qui ont participé à mes projets lors de leur stage d'été en France, rendu possible grâce au programme d'échange entre l'institut Curie et le MIT.

Les pages vont manquer à l'exhaustivité des remerciements que je tiens à adresser aux nombreuses autres personnes qui ont enrichies ma vie personnelle et scientifique durant ces années mémorables. Les membres du laboratoire tout d'abord, qui de proches collègues sont pour certains devenus des amis : Nicolas Robine dont j'ai beaucoup apprécié la finesse d'esprit, Ana-Maria Léon-Ortiz pour nos passionnantes conversations scientifiques et pour l'ADIC, l'inaltérable entrain d'Isabelle Morin, la rigueur allemande de Laurent Lacroix, le paternalisme hilarant d'Alexandre Serero, Gaël Millot et les progrès considérables qu'il m'a fait faire en statistiques, Sophie Loeillet pour sa disponibilité et son aide inconditionnelle, Valérie Borde pour son esprit critique, Claire Jubin pour son éclairage bio-informatique, et plus récemment les rayons de soleil que sont Roberta Sarno, Raphaëlle Laureau, Lou Fourrière, Lucie Carrière et Frédéric Samazan. Un grand merci également à deux travailleurs de l'ombre, mais non moins importants au bon déroulement de cette thèse: Marie-Francoise Liachenko pour sa diligence à traiter les multiples affaires administratives me concernant, et Lélia Soter pour sa fiabilité dans la préparation des milieux de culture, allégeant grandement ma charge de travail. Enfin je voudrais remercier Bozidarka Zaric, Joséphine Galipon, Adeline Berger, Lóránt Székvölgyi, Roberta Sciurano, Pierre Thouvenot et Solange Tanwo Nounga avec qui j'ai certes entretenu des relations

moins étroites en raison du peu de temps passé ensemble au laboratoire ou de la distance entre nos sujets de recherche respectifs, mais qui ont également contribué à l'atmosphère positive du laboratoire. J'aimerais étendre ma gratitude aux membres de l'UMR3244, avec une pensée particulière pour mes amis Hervé Técher, Suk-Min Jang et Ming Fang, ainsi que Arturo Londono et Benoit Le Tallec dont j'ai particulièrement apprécié l'enthousiasme et les passionnantes conversations que nous avons pu avoir.

Je remercie également l'ADIC (Sandrine Morlot, Maxime Deforêt, Darius Koester, Ana Joaquina-Jimenez, Sophie Colombo, Florent Ubelmann et les autres) qui prouve que l'on peut allier productivité ambitieuse et désintéressée à d'excellents moments, ainsi que Jacques Camonis et Andreï Zynoviev pour l'organisation du Babel Tower Club.

Je voudrais également remercier le Ministère de l'Enseignement Supérieur et de la Recherche et l'Agence pour la Recherche sur le Cancer" pour le financement de ma thèse, la Société Française de Génétique et l'Institut Curie pour l'aide financière qu'ils m'ont accordé pour les deux congrès auxquels j'ai participé aux Etats-Unis, et finalement l'Institut Curie pour l'excellent environnement scientifique qu'il propose.

Enfin, je suis profondément reconnaissant à Patrick Ferreira qui par son talent a suscité dès le lycée la passion que j'entretiens pour la biologie et confortée dans le supérieur par deux professeurs exceptionnels : Roland Douchet en classe prépa, et Pierre Netter à l'université.

\*\*\*

Plus personnellement maintenant, j'aimerais remercier ma famille et mes amis.

Mes parents qui m'ont toujours encouragé et soutenu moralement et financièrement dans la réalisation de mes projets d'étude. Au-delà des études, je les remercie pour toujours avoir mis leur intelligence à développer en moi le sens de l'essentiel, élément déterminant de mon épanouissement tant personnel que professionnel.

Ma sœur qui, de ses yeux brillant d'admiration, m'a aussi investi d'une mission de réussite qu'il serait idiot de négliger.



Mes amis qui ont été un inaltérable support moral, une source d'inspiration et de motivation. Des amis excellents sans lesquels j'aurais sans doute été bien différent.

Enfin, je suis extrêmement reconnaissant à Noémie pour sa compréhension, son soutien, et sa façon de rendre la vie belle.



# Table des matières

Abréviations	4
Liste des figures	5
Liste des tables	6
Avant-propos	7
<b>I. <u>Introduction</u></b>	<b>8</b>
<b>I.A. <u>Les G-quadruplexes</u></b>	<b>8</b>
<b>I.A.1. <u>Propriétés physico-chimiques des G-quadruplexes</u></b>	<b>8</b>
I.A.1.a. <u>Guanine, G-quartet, et G-quadruplexe</u>	8
I.A.1.b. <u>Diversité structurale des G-quadruplexes</u>	10
I.A.1.c. <u>Modalités de formation d'un G-quadruplexe dans l'ADN double-brin</u>	20
<b>I.A.2. <u>Les ligands des G-quadruplexes</u></b>	<b>27</b>
I.A.2.a. <u>Modalités d'interaction</u>	27
I.A.2.b. <u>Sélectivité, affinité, et spécificité</u>	27
I.A.2.c. <u>Le test FID</u>	29
I.A.2.d. <u>Exemples de ligands utilisés dans des études <i>in vivo</i></u>	29
<b>I.A.3. <u>Existence des G-quadruplexes <i>in vivo</i></u></b>	<b>32</b>
I.A.3.a. <u>Les séquences G-quadruplexes putatives des génomes</u>	32
I.A.3.b. <u>Rôles biologiques des G-quadruplexes</u>	35
I.A.3.c. <u>Conclusion</u>	45
<b>I.A.4. <u>Les G-quadruplexes comme cibles thérapeutiques potentielles dans le traitement des cancers</u></b>	<b>45</b>
I.A.4.a. <u>Inhibition de l'expression des oncogènes</u>	45
I.A.4.b. <u>Perturbation du métabolisme des télomères</u>	46
I.A.4.c. <u>Les effets secondaires et le challenge de la spécificité</u>	48
<b>I.A.5. <u>G-quadruplexes et instabilité génomique</u></b>	<b>50</b>
I.A.5.a. <u>Blocage des polymérases par un G-quadruplexe <i>in vitro</i></u>	50
I.A.5.b. <u>Des hélicases déroulent les G-quadruplexes <i>in vitro</i></u>	50
I.A.5.c. <u>Autres activités de déroulement des G-quadruplexes</u>	56
<b>I.B. <u>La réplication, un mécanisme à risque pour la stabilité du génome</u></b>	<b>57</b>
<b>I.B.1. <u>La réplication chez <i>S. cerevisiae</i></u></b>	<b>57</b>
I.B.1.a. <u>Généralités</u>	57
I.B.1.b. <u>ADN simple brin à la fourche</u>	61
I.B.1.c. <u>Quels couplages à la fourche ?</u>	63
<b>I.B.2. <u>Les séquences capables de former des structures secondaires de l'ADN sont « à risque » lors de la réplication</u></b>	<b>68</b>
<b>I.B.3. <u>Mécanismes qui limitent l'instabilité génomique lors de la réplication</u></b>	<b>69</b>
I.B.3.a. <u>Prévention de la formation des structures secondaires : RPA et la limitation de l'accumulation d'ADN simple brin à la fourche</u>	69
I.B.3.b. <u>Réparation des fourches bloquées ou cassées</u>	71
<b>I.B.4. <u>Les réarrangements génomiques</u></b>	<b>81</b>
I.B.4.a. <u>Instabilité des séquences répétées en tandem</u>	82
I.B.4.b. <u>Grands réarrangements chromosomiques (GCR) médiés par des séquences répétées dispersées</u>	82
<b>I.C. <u>Etude de la formation des G-quadruplexes et des mécanismes assurant la stabilité des séquences sous-jacentes chez <i>S. cerevisiae</i></u></b>	<b>86</b>
<b>I.C.1. <u>Contexte et but de la thèse</u></b>	<b>86</b>
I.C.1.a. <u>Etude de l'instabilité des séquences répétées de type minisatellite chez</u>	

la levure <i>S. cerevisiae</i>	86
I.C.1.b. <u>L'hélicase Pif1</u>	90
I.C.1.c. <u>Un rôle des G-quadruplexes dans l'instabilité des séquences minisatellites chez <i>S. cerevisiae</i>?</u>	92
<b>II. Résultats</b>	<b>94</b>
<b>II.A. Article 1 : <u>The Yeast Pif1 Helicase Prevents Genomic Instability Caused by G-Quadruplex-Forming CEB1 Sequences <i>In Vivo</i></u></b>	<b>94</b>
II.A.1. Contexte	94
II.A.2. Article	94
<b>II.B. Article 2 : <u>Genetic Instability Triggered by G-Quadruplex Interacting Phen-DC Compounds in <i>Saccharomyces cerevisiae</i></u></b>	<b>95</b>
II.B.1. Contexte	95
II.B.2. Article	95
II.B.3. Résultat supplémentaire : l'instabilité de CEB1 est soumise aux concentrations respectives de Pif1 et de Phen-DC <sub>3</sub> <i>in vivo</i>	96
<b>II.C. Article 3 : <u>G-Quadruplex-Induced Instability During Leading-Strand Replication</u></b>	<b>98</b>
II.C.1. Contexte	98
II.C.2. Article	98
II.C.3. Résultats supplémentaires	99
II.C.3.a. <u>Nature de la lésion induite par CEB1</u>	99
II.C.3.b. <u>Mutants de polymérase translésionnelles</u>	99
II.C.3.c. <u>Mutants d'hélicases</u>	105
II.C.3.d. <u>Mutants de réplication et du checkpoint intra-S</u>	107
<b>II.D. Article 4 : <u>Stimulation of Gross Chromosomal Rearrangements by the Human CEB1 and CEB25 Minisatellites in <i>S. cerevisiae</i> Depends on G-quadruplexes or Cdc13</u></b>	<b>111</b>
II.D.1. Contexte	111
II.D.2. Article	111
II.D.3. Résultats supplémentaires	112
II.D.3.a. <u>L'allèle naturel CEB1-1.8 induit des GCR à des taux plus faibles que l'allèle artificiel</u>	112
II.D.3.b. <u>Phen-DC<sub>3</sub> stimule la capacité de la répétition télomérique humaine à induire des GCR</u>	112
II.D.3.c. <u>Effet d'autres ligands de G-quadruplexes</u>	114
<b>II.E. Structure-function Relationship of G-Quadruplexes to Genomic Instability</b>	<b>117</b>
<b>II.E.1. Loop Length Affects the Instability of G-Quadruplex-Forming Sequences in <i>S. cerevisiae</i></b>	<b>117</b>
II.E.1.a. <u>CEB25 is stable in the absence of Pif1 or upon treatment with Phen-DC<sub>3</sub></u>	117
II.E.1.b. <u>The central loop of the CEB25 G-quadruplex stabilizes the minisatellite <i>in vivo</i></u>	120
<b>II.E.2. Multiple G-quadruplexes Formed by CEB1 Contributes to its Instability</b>	<b>124</b>
II.E.2.a. <u>The CEB1 motif adopt several G-quadruplex conformations</u>	124
II.E.2.b. <u>Point mutagenesis of CEB1 reveal G-runs involved in its instability</u>	126
II.E.2.c. <u>Does several G-quadruplexes cooperate to induce the instability of CEB1?</u>	132
<b>III. Discussion et perspectives</b>	<b>135</b>
<b>III.A. Formation des G-quadruplexes à la fourche de réplication</b>	<b>135</b>
III.A.1. Formation du G-quadruplexe entre l'hélicase réplicative et Pole	135
III.A.2. Découplage entre machinerie de synthèse continue et hélicase réplicative	136

III.A.3. Coopération entre motifs et aggravation des problèmes de réplication	138
III.A.4. Les G-quadruplexes se forment-ils sur la matrice du brin à synthèse discontinue ?	141
III.A.5. Implications pour l'instabilité génomique causée par d'autres structures secondaires	142
III.B. <u>Reprise de la synthèse et formation de réarrangements génomiques</u>	144
III.C. <u>Rôle de Pif1 dans la réplication des séquences G4P et le maintien de la stabilité du génome</u>	145
III.C.1. Pif1 promeut la réplication des séquences G4P endogènes de <i>S. cerevisiae</i>	145
III.C.2. Le rôle de déroulement des G-quadruplexes par Pif1 est-il conservé chez les eucaryotes ?	146
III.C.3. Pif1 est-elle la seule hélicase impliquée dans le maintien des séquences G4 chez <i>S. cerevisiae</i> ?	148
III.D. <u>Structure des G-quadruplexes et spécificité des hélicases pour le maintien des séquences sous-jacentes</u>	150
III.D.1. Différences d'instabilité entre les allèles naturels et artificiels de CEB1	150
III.D.2. Pourquoi un G-quadruplexe contenant une longue boucle n'induit-il l'instabilité du minisatellite?	151
III.D.3. Versatilité évolutive et physiologique des G-quadruplexes	153
Conclusions	155
Références	156

## **Abréviations**

ADN = acide désoxyribonucléique

ADNdb = ADN double-brin (= duplex)

ADNsb = ADN simple-brin

AFM = Atomic Force Microscopy

ARN = acide ribonucléique

BIR = Break-Induced Replication

CDB = cassure double-brin

FRET = Förster resonance energy transfer

G = guanine

G4P (séquence G4P) = séquence G-quadruplexe putative

GCR = Gross Chromosomal Rearrangement

RH = Recombinaison homologue

dHJ = double Jonction de Holliday

kb = kilobase

M = molaire (mol/L)

MMR = Mismatch Repair

ORF = Open Reading Frame (cadre ouvert de lecture)

pb = paire de bases

PCNA = Proliferating cell nuclear antigen

PCR = Polymerase Chain Reaction

RMN = résonance magnétique nucléaire

RNR = Ribonucléotide réductase

TSS = Transcription Start Site (site d'initiation de la transcription)

UTR = Untranslated region (région non traduite)

UV = ultraviolet

YPD = Yeast peptone dextrose

## Liste des Figures

	Page
<b>Figure 1</b> Appariements canoniques et non-canoniques des nucléotides: Guanine, G-quartet et G-quadruplexe	9
<b>Figure 2</b> Conformations des G-quadruplexes (i): les brins	12
<b>Figure 3</b> Conformations des G-quadruplexes (ii): les boucles	15
<b>Figure 4</b> G-quadruplexe en tandem: exemple de CEB25 et collier de perle	17
<b>Figure 5</b> G-quadruplexes exotiques	19
<b>Figure 6</b> La superhélicité négative promeut le désappariement du duplex et la formation du G-quadruplexe	23
<b>Figure 7</b> La G-loop est une structure formée co-transcriptionnellement contenant un hybride ARN/ADN et un ou plusieurs G-quadruplexes sur le brin non-transcrit	25
<b>Figure 8</b> Facteurs qui affectent potentiellement l'équilibre ADNdb - G-quadruplexe <i>in vivo</i>	26
<b>Figure 9</b> Ligands des G-quadruplexes: mode d'interaction	28
<b>Figure 10</b> Mesure de l'affinité et de la sélectivité des ligands de G-quadruplexe par le test FID	30
<b>Figure 11</b> Les G-quadruplexes dans le génome de <i>S. cerevisiae</i>	35
<b>Figure 12</b> Les télomères et la télomérase	38
<b>Figure 13</b> G-quadruplexes aux télomères du cilié <i>Styloichia lemnae</i>	40
<b>Figure 14</b> Modèle d'inhibition de la transcription par le G-quadruplexe formé dans le promoteur de <i>c-Myc</i>	42
<b>Figure 15</b> Un G-quadruplexe comme cible de cassures pour la variation antigénique du pili chez <i>N. gonorrhoeae</i>	44
<b>Figure 16</b> Modes d'action putatifs des ligands de G-quadruplexe responsables de l'inhibition de croissance des cellules tumorales	49
<b>Figure 17</b> Les hélicases de la famille RecQ sont conservées et déroulent les G-quadruplexes	53
<b>Figure 18</b> La réplication est initiée à partir de régions définies chez <i>S. cerevisiae</i>	58
<b>Figure 19</b> Le réplisome chez <i>S. cerevisiae</i>	60
<b>Figure 20</b> ADN simple brin à la fourche de réplication	62
<b>Figure 21</b> Mrc1 participe au couplage entre synthèse et déroulement du duplex en cas d'inhibition généralisée de la synthèse d'ADN	64
<b>Figure 22</b> Absence de couplage des deux machineries de synthèse, chez les bactéries et <i>S. cerevisiae</i>	67
<b>Figure 23</b> Détection de pause de réplication par gel bidimensionnel : exemple des séquences capables de former des structures secondaires chez <i>S. cerevisiae</i>	70
<b>Figure 24</b> Ré-initiation de la synthèse en cas de blocage d'une polymérase sur une matrice endommagée	73
<b>Figure 25</b> Les polymérases translésionnelles permettent la réplication d'une matrice endommagée	75
<b>Figure 26</b> Mécanismes de recombinaison homologue	77
<b>Figure 27</b> Mécanismes de template-switch	80
<b>Figure 28</b> Réarrangement des séquences répétées en tandem par recombinaison homologue	83
<b>Figure 29</b> Grands réarrangements chromosomiques causés par des séquences répétées	85
<b>Figure 30</b> Localisation et séquences des minisatellites humains CEB1 et CEB25	87
<b>Figure 31</b> L'hélicase Pif1 de <i>S. cerevisiae</i>	91
<b>Figure 32</b> L'instabilité de CEB-WT-1.7 en présence de Phen-DC3 est sensible à la concentration intra-cellulaire de Pif1	97
<b>Figure 33</b> Aucune CDB n'est détectée au niveau de CEB1 en absence de Pif1 par gel d'électrophorèse en champ pulsé	100
<b>Figure 34</b> Instabilité de CEB1 en absence de la polymérase translésionnelle Rev1	102
<b>Figure 35</b> Instabilité de CEB1 en absence de l'hélicase Sgs1	106
<b>Figure 36</b> La protéine Atr1 n'a pas d'effet sur l'instabilité de CEB1 dans les mutants <i>pif1Δ</i> et <i>pif1Δrrm3Δ</i>	108
<b>Figure 37</b> Rad53 est phosphorylé dans les mutants <i>rrm3Δ</i> et <i>pif1Δrrm3Δ</i> , mais pas dans le mutant <i>pif1Δ</i>	110
<b>Figure 38</b> L'allèle naturel CEB1-1.8 stimule moins le taux de GCR que l'allèle artificiel de même taille dans une souche WT	113
<b>Figure 39</b> Effet de divers ligands de G-quadruplexe sur la croissance et le taux de GCR d'une souche WT de <i>S. cerevisiae</i> .	115
<b>Figure 40</b> CEB25 instability at the <i>ARG4</i> locus in a WT strain treated by Phen-DC3, or in <i>pif1Δ</i> and <i>rad27Δ</i> mutants	118
<b>Figure 41</b> The shortening of the 9-nt loop of the CEB25 G-quadruplex induces its instability	121
<b>Figure 42</b> Insertion of the 9-nt loop of CEB25 within the G4P sequence of CEB25 stabilizes the minisatellite	123
<b>Figure 43</b> The CEB1 motif adopts several G-quadruplex conformations <i>in vitro</i>	125
<b>Figure 44</b> Instability of the CEB1 alleles mutated for their G-triplets	128
<b>Figure 45</b> Instability of the CEB1 alleles mutated for the G-doublet involved in form 1	129
<b>Figure 46</b> Instability of the CEB1 alleles mutated for the 6 contiguous G	131
<b>Figure 47</b> Summary of the behavior of the CEB1 mutant alleles <i>in vivo</i>	133
<b>Figure 48</b> L'accumulation de la matrice du brin à synthèse continue sous forme simple-brin favorise la formation de G-quadruplexe. (A) Effet des facteurs agissant en cis (richesse GC) et en trans (concentration en dNTP) sur la vite	137
<b>Figure 49</b> Mécanismes putatifs de collaboration entre motifs en mesure d'expliquer l'effet de taille sur l'instabilité de CEB1	139
<b>Figure 50</b> Pif1 est enrichi au niveau de certaines séquences G4P, et limite l'accumulation de Pol2 spécifiquement à ces sites	146
<b>Figure 51</b> Le ligand de G-quadruplexe pyridostatin et hPif1 colocalisent dans des cellules humaines	148
<b>Figure 52</b> Pif1 est redondant avec d'autres hélicases sur les G-quadruplexes à boucle longue	152

## Liste des Tables

<b>Table 1</b>	Activités connues des hélicases ADN sur des substrats G-quadruplexes intra- et inter-moléculaires	<b>p. 37</b>
<b>Table 2</b>	Minisatellites humains dont l'instabilité a été étudiée chez <i>S. cerevisiae</i> durant ma thèse	<b>p. 56</b>
<b>Table 3</b>	Instabilité de CEB1-1.8 inséré à proximité de l' <i>ARS305</i> dans différents mutants	<b>p. 66</b>
<b>Table 4</b>	Instabilité de CEB1-1.8 inséré à proximité d' <i>ARS305</i> après traitement par Phen-DC3	<b>p. 70</b>
<b>Table 5</b>	Instabilité d'allèles naturels et artificiels de CEB25 insérés à <i>ARG4</i> dans des souches WT, <i>pif1Δ</i> , et <i>rad27Δ</i> .	<b>p. 74</b>
<b>Table 6</b>	Instabilité d'allèles de CEB25 aux tailles de boucle variables inséré à proximité de l' <i>ARS305</i> dans l'orientation dans des souches WT ± Phen-DC3, et <i>pif1Δ</i>	<b>p. 75</b>



## Avant-propos

Les acides nucléiques sont une classe de macro-molécules essentielles aux organismes vivants, à leur reproduction et à leur évolution. Dans les cellules, l'ADN adopte la structure en double-hélice décrite par Watson et Crick. Toutefois, la chimie des bases autorise d'autres interactions que l'appariement canonique, et en conséquence d'autres conformations d'acides nucléiques. Le G-quadruplexe est l'une de ces structures alternatives, formée par l'appariement de quatre brins riches en guanines. Bien que caractérisée *in vitro* depuis plusieurs décennies, ce n'est que dans les cinq dernières années que la communauté de biologistes s'est intéressée à leur formation et leurs rôles *in vivo*. Mon arrivée au laboratoire coïncide avec cet éveil, et avec la nécessité de mettre en place des systèmes expérimentaux et des méthodes pour étudier la formation de ces structures et leurs rôles dans les cellules.

Cette structure apparaît encore comme exotique pour de nombreux biologistes moléculaires et généticiens. Aussi vais-je dresser, dans une première partie de mon introduction, un portrait exhaustif des propriétés biophysiques de cette famille de structures, ainsi que des démonstrations les plus pertinentes de leur existence et leurs rôles associés *in vivo*. A ces rôles « positifs » des G-quadruplexes, dans différents processus biologiques et organismes, requièrent des processus de maintien des séquences sous-jacentes au sein du génome: mon travail de thèse révèle qu'un défaut de déroulement de ces structures perturbe la réplication et conduit à l'instabilité génomique de ces séquences chez *S. cerevisiae*.

# I. INTRODUCTION





# I. Introduction

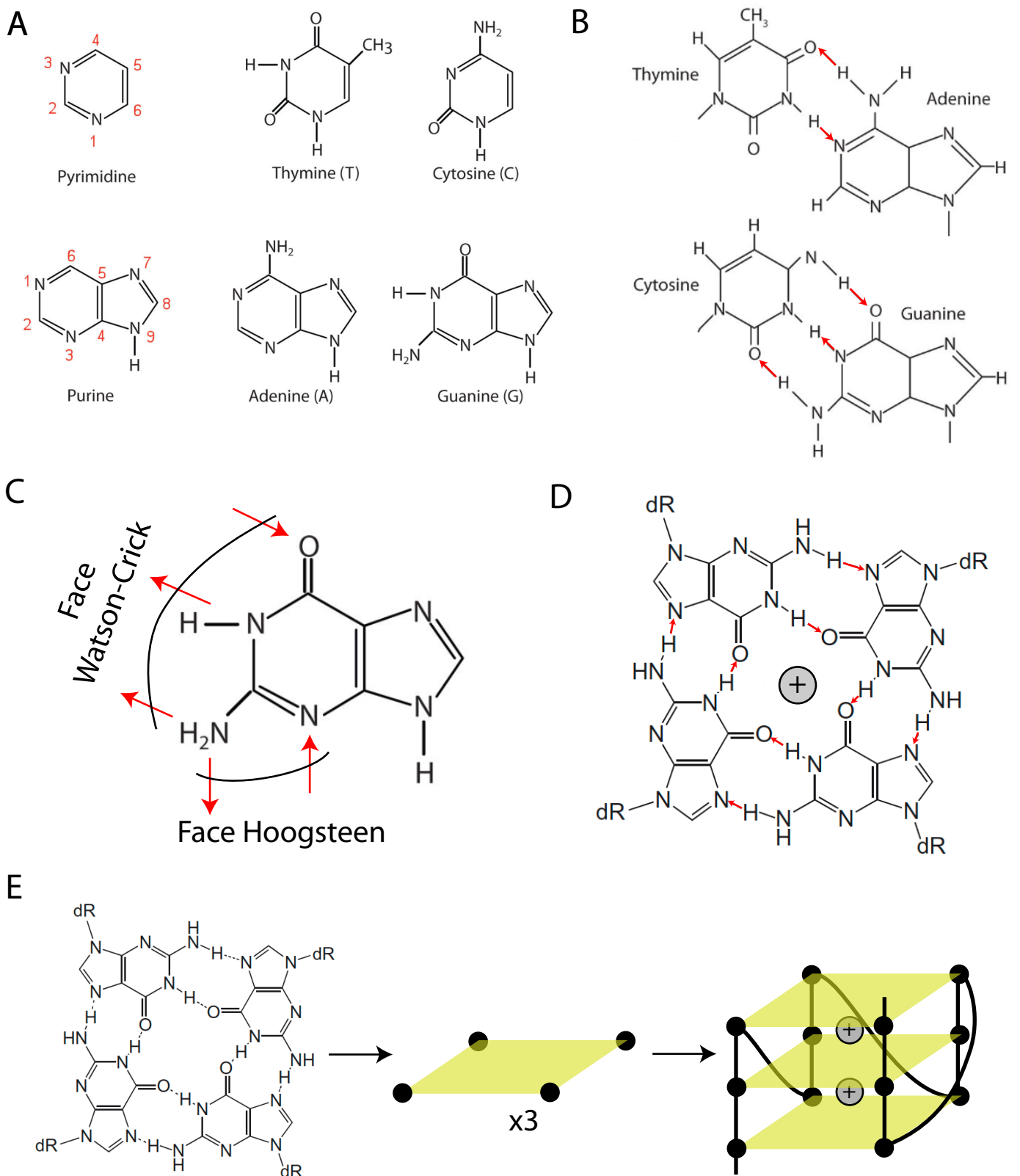
## I.A. Les G-quadruplexes

### I.A.1. Propriétés physico-chimiques des G-quadruplexes

#### I.A.1.a. Guanine, G-quartet, et G-quadruplexe

Les nucléotides sont les éléments unitaires des acides nucléiques. Ils sont composés d'un groupement phosphate, d'un sucre, et d'une base. La nature du sucre différencie l'ADN (désoxyribose) de l'ARN (ribose). La base, quant à elle, différencie les nucléotides : adénine (A), cytosine (C), guanine (G), thymine (T, dans l'ADN) et uracile (U, dans l'ARN)(**Figure 1A**). L'appariement canonique des nucléotides (A:T et C :G)(**Figure 1B**) fait intervenir des liaisons hydrogènes sur la face « Watson-Crick » (**Figure 1C**). Dans l'ADN, cet appariement canonique des bases, en plus de permettre la formation de la structure protectrice en double hélice, constitue intrinsèquement une possibilité de copie de la molécule, essentiel pour l'expression et la transmission de l'information génétique (Watson and Crick 1953).

Les nucléotides peuvent former des associations non-canoniques en utilisant leur face dite « Hoogsteen » (**Figure 1C**). C'est le cas des guanines qui, en s'associant via deux liaisons hydrogènes entre leurs faces « Hoogsteen » et « Watson-Crick », autorise la formation d'un plateau de 4 guanines appelé « G-quartet » (**Figure 1D**) (Gellert et al. 1962). Dans ce cas, le nombre de liaisons H impliquées s'élève à 8, contrairement aux 2 (pour A et T) ou 3 (pour G et C) liaisons H observées entre les nucléotides dans la double hélice. L'empilement d'au moins 2 G-quartets suffit à former la structure à 4 brins appelée « G-quadruplexe » (**Figure 1E**) (Arnott et al. 1974; Zimmerman et al. 1975). L'empilement ininterrompu de G-quartets suppose la présence dans la séquence primaire d'ADN de guanines successives : 2 pour la formation d'un G-quadruplexe à 2 G-quartets, 3 pour 3 G-quartets etc... Les plateaux de bases y sont espacés d'environ 3.2A, approximativement comme dans la double-hélice B (3.3A). Le squelette phosphate relie les guanines de G-quartets successifs, résultant en une structure hélicoïdale à 4 brins d'un pas de 30°. La taille des sillons varie entre 10 et 12 A suivant la géométrie *syn* ou *anti* adoptée par la base (voir **chapitre I.A.1.b.ii**). Ces G-quadruplexes sont exclusivement constitués d'ADN ou d'ARN : aucun G-quadruplexe mixte ARN/ADN n'a



été décrit. Qualitativement, ce qui est vrai pour les G-quadruplexes ADN en termes de stabilité l'est aussi pour les G-quadruplexes ARN (Sacca et al. 2005). Ma thèse portant sur l'étude *in vivo* des G-quadruplexes ADN, c'est à eux que je me référerai dorénavant, sauf indication contraire.

L'empilement des G-quartets au sein du G-quadruplexe requiert la présence de cations (Williamson et al. 1989; Laughlan et al. 1994). Le cation s'intercale entre 2 G-quartets au niveau de la cavité centrale, chargée négativement : ce cation « coordonne » deux G-quartets (**Figure 1D** et **1E**). Un G-quadruplexe à  $n$  G-quartets contient  $n-1$  cations. Le diamètre et la charge du cation sont des déterminants importants de la stabilité du G-quadruplexe, ainsi que de sa structure. Le Potassium ( $K^+$ ) est le meilleur coordinateur monovalent, suivi du Sodium ( $Na^+$ ) (Mergny et al. 1998). La capacité d'autres cations monovalents à coordonner les G-quartets a été examinée, et la qualité de cette coordination s'effectue suivant l'ordre  $K^+ \gg Na^+ > Rb^+ > NH_4^+ > Cs^+ \gg Li^+$  (Williamson et al. 1989; Venczel and Sen 1993). Le Lithium ( $Li^+$ ), en raison de son trop faible diamètre, est neutre vis-à-vis des G-quadruplexes : il ne permet pas leur formation ni ne les déstabilise une fois formés en présence de  $K^+$  ou  $Na^+$ , même à forte concentration (Mergny et al. 1998; Neidle and Balasubramanian 2006). Il est utilisé comme contrôle négatif pour l'étude des G-quadruplexes *in vitro*. Des cations divalents comme le Strontium ( $Sr^{2+}$ ), le Calcium ( $Ca^{2+}$ ) et le Magnésium ( $Mg^{2+}$ ) permettent également la formation des G-quadruplexes à faible concentration ( $\approx 1$  mM) (Blume et al. 1997; Wlodarczyk et al. 2005). Dans les cellules, les concentrations physiologiques de cations monovalents et divalents les plus abondants sont  $[K^+] \approx 140-160$  mM,  $[Na^+] \approx 5-15$  mM,  $[Mg^{2+}] \approx 0.5$  mM, et  $[Ca^{2+}] \approx 0.2$   $\mu$ M. Les autres cations sont présents à des concentrations inférieures au nanomolaire. Ainsi, des conditions physiologiques de sels et de pH sont propices à la formation des G-quadruplexes à partir d'ADN simple brin (Mergny et al. 1998). C'est en présence de concentrations physiologiques de  $Na^+$  et  $K^+$  que sont réalisées les expériences d'étude des G-quadruplexes *in vitro* (Huppert 2010).

#### I.A.1.b. Diversité structurale des G-quadruplexes

Les G-quartets empilés et reliés entre eux par le squelette phosphate constituent le « noyau » invariant de toute structure G-quadruplexe. Une propriété remarquable des G-quadruplexes réside dans les multiples conformations qu'ils peuvent adopter : la

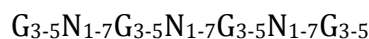
séquence primaire d'ADN et le nombre de molécules impliquées participent à cette diversité structurale et affectent la cinétique de formation et la stabilité du G-quadruplexe. Alors que la cinétique de formation se traduit par la constante d'association ( $k_{on}$ ) en  $s^{-1}$ , la stabilité thermique s'exprime en °C, le  $T_m$  étant la température à laquelle la moitié des molécules sont sous forme G-quadruplexe à l'équilibre.

#### *I.A.1.b.i. Nombre de brins*

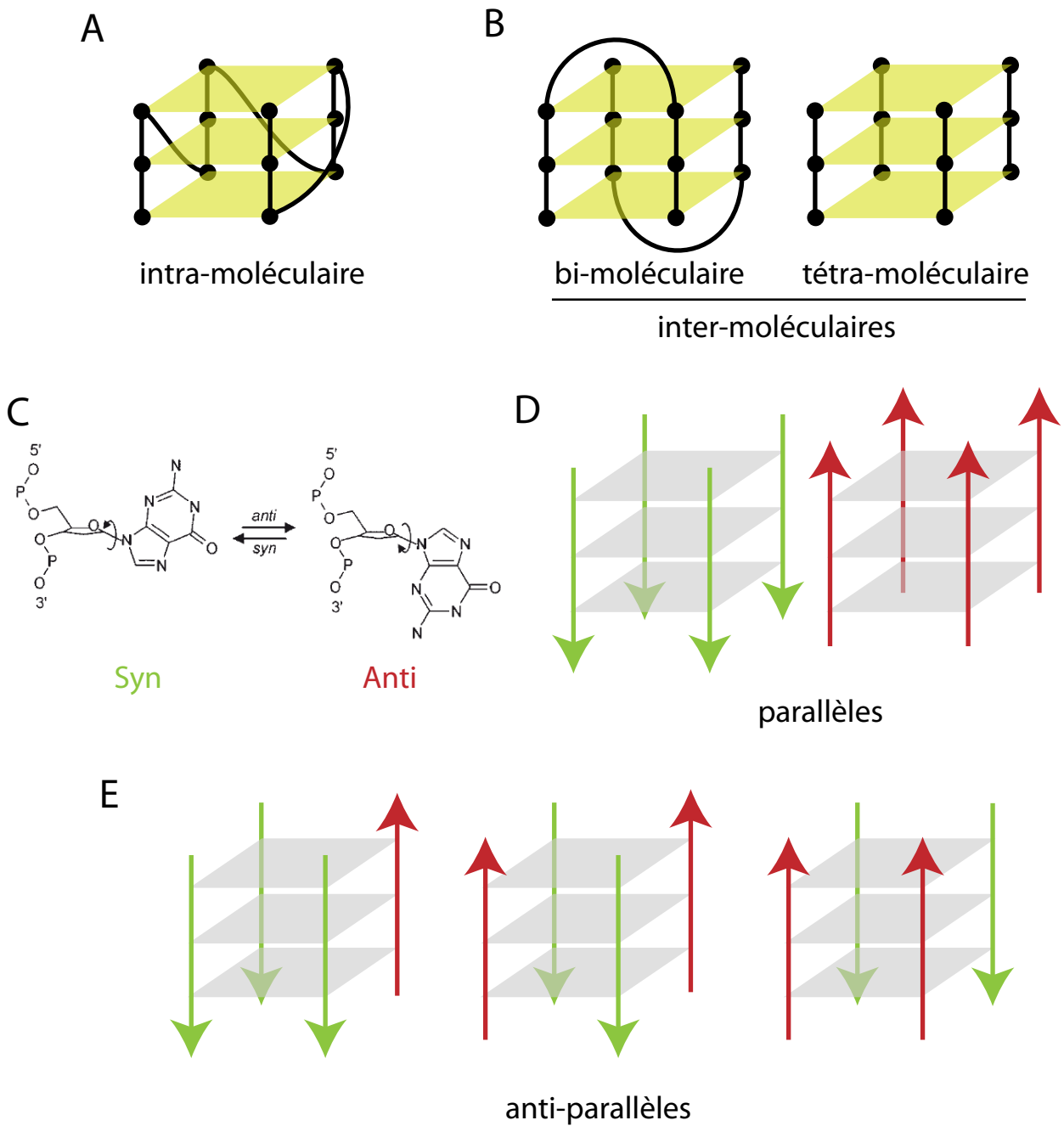
Le G-quadruplexe peut être formé par l'association de guanines provenant d'une ou de plusieurs molécules, avec des cinétiques de formation et une stabilité thermodynamique associées très différentes.

La formation d'un G-quadruplexe intra-moléculaire requiert une séquence d'ADN contenant au moins 4 blocs d'au moins 2 guanines séparés par un nombre variable de nucléotides. Ceci implique la formation de trois boucles simples brins au sein du G-quadruplexe (**Figure 2A**). Le nombre de guanines dans chaque bloc ainsi que la taille des boucles sont cruciaux pour la stabilité de la structure.

La proximité des guanines au sein de la molécule autorise l'association rapide ( $k_{on} < 10^{-2} - 10^{-3} s^{-1}$ ) et indépendante de la concentration d'oligonucléotides *in vitro* (Bouquin G4 Chapitre 2). Plus la taille des boucles est petite, plus la stabilité augmente, pour atteindre 80°C (Guedin et al. 2010a). Inversement, plus la taille des boucles augmente, plus la vitesse d'association et la stabilité thermique diminuent (**Chapitre I.A.1.b.iv**). Ces mesures de stabilité thermodynamique ont conduit à la formulation d'une séquence consensus susceptible de former un G-quadruplexe intra-moléculaire utilisé dans la recherche de séquences G-quadruplexes putatives (G4P) des génomes (voir **Chapitre I.A.3.a.i**):



Dans un G-quadruplexe inter-moléculaire, les brins d'ADN contiguës ne sont pas reliés covalentement, et peuvent faire intervenir deux ou quatre brins indépendants (G-quadruplexes respectivement bi- et tétra-moléculaires)(**Figure 2B**). Ces G-quadruplexes inter-moléculaires sont beaucoup plus stables que les G-quadruplexes intra-moléculaires : leur cinétique d'association, très lente, dépend de la concentration



**Figure 2 : Conformation des G-quadruplexes: les brins.** Représentation schématique d'un G-quadruplexe intra-moléculaire (A) et de deux G-quadruplexes inter-moléculaires (B). (C) Conformation syn (vert) ou anti (rouge) de la guanine vis-à-vis du sucre (Simonsson, Chem. Biol. 2001). (D) G-quadruplexes parallèles. (E) Exemple de G-quadruplexes anti-parallèles. Les guanines impliquées dans la structure présentent des conformations différentes (au moins sur un des 4 brins). (D et E) La couleur des brins indique la conformation des guanines (syn ou anti).



d'oligonucléotides ( $k_{on} \approx 10^6 - 10^8 \text{ s}^{-1}$ ) (Wyatt et al. 1996; Mergny et al. 2005; De Cian and Mergny 2007). En revanche, une fois formés, ils sont très résistants à la dénaturation thermique : aucune dénaturation du G-quadruplexe formé par l'association de  $TG_3T$  n'est détectée après plusieurs minutes à  $100^\circ\text{C}$ , même en présence de hautes concentration (7 M) d'urée (Shida et al. 1991). Ces cinétiques d'association longues et leur stabilité extrême posent la question de leur relevance biologique. Au **Chapitre I.A.3.b.i** sera abordé l'exemple du G-quadruplexe bi-moléculaire formé par l'extrémité télomérique  $(T_4G_4)_2$  chez le cilié *Stylonichia lemnae* au cours du cycle cellulaire.

#### *I.A.1.b.ii. Nombre de G-quartets*

Au sein du G-quartet, chaque guanine se trouve associée à ses voisines par 4 liaisons H (**Figure 1C**). Ainsi, l'association des brins entre eux augmente proportionnellement avec le nombre de G-quartets, avec pour conséquence d'augmenter la stabilité globale du G-quadruplexe. De plus, l'empilement des bases (« base stacking ») est un facteur stabilisateur des plateaux de bases, dans la double hélice comme dans le G-quadruplexe. Aussi, même si des G-quadruplexes se forment avec seulement 2 G-quartets, l'augmentation du nombre de G-quartets augmentera la stabilité d'ensemble de la structure en augmentant le « base stacking », notamment pour les G-quartets internes. Ceci peut conduire à l'empilement de deux G-quadruplexes indépendants via l'un de leur G-quartet externe (**Figure 5** et résultats **Chapitre II.E.2.a**) (Mukundan et al. 2011).

Cet effet du nombre de G-quartets sur la cinétique de formation et la stabilité a été étudié pour des G-quadruplexes inter-moléculaires formés par des oligonucléotides  $TG_{3-7}T$  (Mergny et al. 2005) et sur des G-quadruplexes intra-moléculaires  $(G_{3-7}T)_4$  (Rachwal et al. 2007a). L'augmentation du nombre de quartets conduit à une augmentation de la stabilité de la structure, ainsi que de la vitesse d'association des brins dans le cas des G-quadruplexes inter-moléculaires, chaque G supplémentaire augmentant la vitesse d'association d'un facteur 10 (Mergny et al. 2005).

#### *I.A.1.b.iii. Orientation des brins*

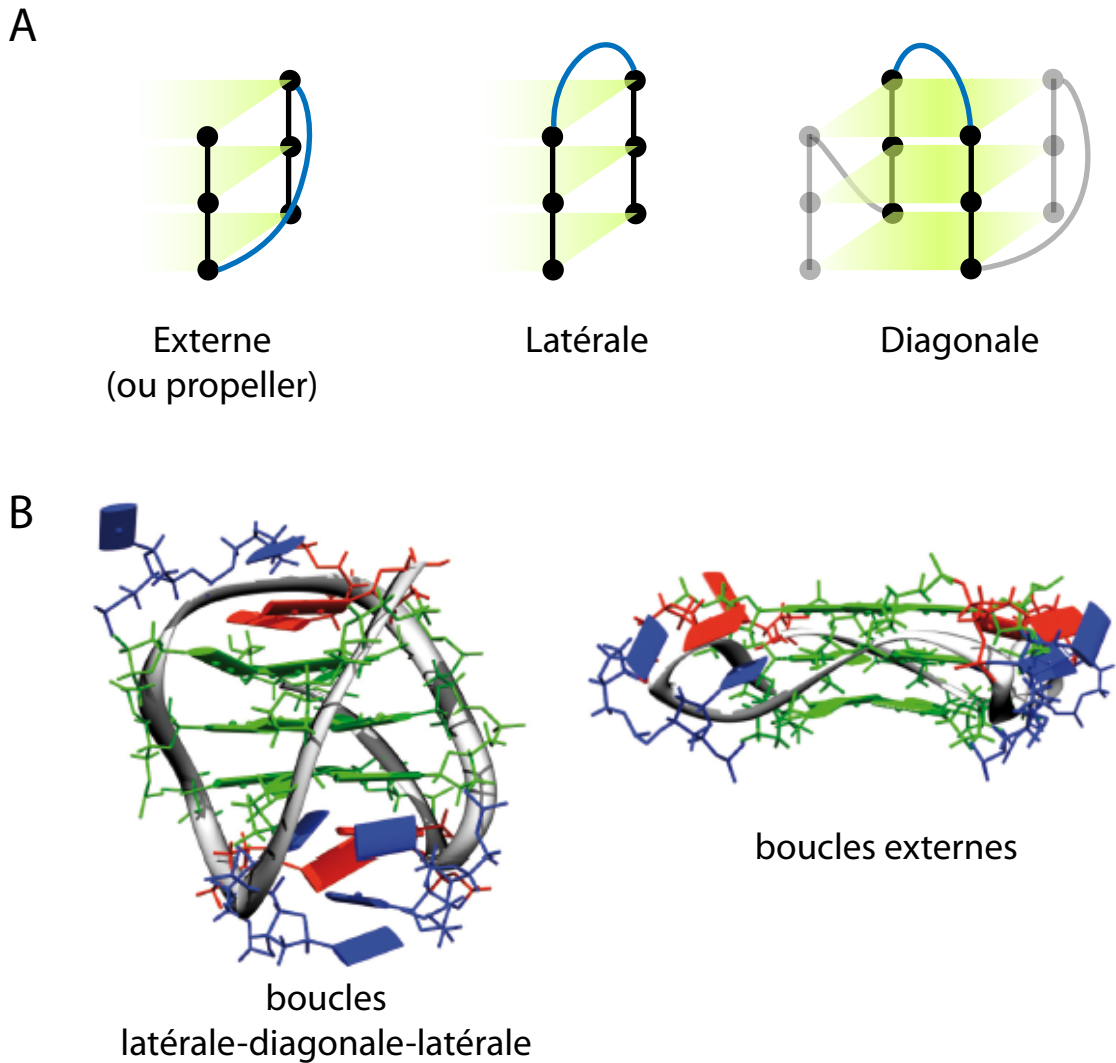
Au sein d'un G-quartet, les guanines peuvent adopter deux géométries en fonction de l'orientation de la base vis-à-vis du sucre : *syn* ou *anti* (**Figure 2C**). Lorsque toutes les guanines d'un G-quartet ont la même géométrie, la polarité des brins est identique et la structure parallèle (**Figure 2D**). Lorsqu'au moins un brin présente des

guanines avec une géométrie différente des autres brins (*syn-anti-anti-anti*, *syn-syn-anti-anti*, ou *syn-syn-syn-anti*), la structure est dite anti-parallèle (**Figure 2E**). Les G-quadruplexes parallèles sont généralement plus stables que les G-quadruplexes anti-parallèles, qu'ils soient intra-moléculaires ou inter-moléculaires (Chapitre 2, (Neidle and Balasubramanian 2006)).

#### *I.A.1.b.iv. Taille et séquence des boucles*

Les boucles sont un facteur déterminant de la diversité conformationnelle et de la stabilité des G-quadruplexes. Elles sont de trois types: latérales, diagonales, et externes, schématisées **Figure 3A**. Les trois boucles au sein du G-quadruplexe peuvent être de même type ou différentes (exemples **Figure 3B**). Le type de boucle a des implications évidentes sur l'orientation des brins : une boucle externe reliera deux brins de même orientation, alors qu'une boucle diagonale ou latérale reliera deux brins d'orientations opposées. La taille des boucles participe à la fois à la vitesse de formation, à la stabilité de la structure, et à l'orientation des brins : les boucles courtes sont généralement externes et favorisent donc une conformation parallèle. En revanche, étant moins contraintes, les longues boucles peuvent être de n'importe quel type et donnent lieu à des G-quadruplexes anti-parallèles (Rachwal et al. 2007a).

L'effet de la taille des boucles sur la stabilité des G-quadruplexes a été l'objet de nombreuses études (Risitano and Fox 2004; Rachwal et al. 2007a; Guedin et al. 2008; Kumar et al. 2008; Guedin et al. 2010a). L'équipe de J.-L. Mergny a étudié l'impact de la taille de la boucle sur la stabilité thermique des G-quadruplexes sur une séquence artificielle  $TG_3TG_3T_nG_3TG_3$ , en faisant varier  $n$  de 1 à 30 (Guedin et al. 2010a). La stabilité du G-quadruplexe diminue quand la taille de la boucle augmente, mais des G-quadruplexes stables se forment même avec une boucle de 21 nt (Bourdoncle et al. 2006; Guedin et al. 2010a). La stabilité n'est pas reliée à la somme totale de nucléotides dans les boucles, puisque le  $T_m$  de G-quadruplexes dont la somme est 17-18 nt est de 62,5°C avec l'arrangement de taille 1-15-1, 49.5°C avec 3-12-3, et 46.5°C avec 7-3-7 en présence de potassium (Rachwal et al. 2007a; Kumar et al. 2008). Ainsi, cet effet déstabilisateur des boucles longues est d'autant plus dramatique que la taille des autres boucles augmente (Rachwal et al. 2007a). Cette étude a des implications concernant la recherche de motifs capables de former des G-quadruplexes dans les génomes (voir **Chapitre I.A.3.a.i**).



**Figure 3: Conformation des G-quadruplexes: les boucles.** (A) Types de boucles possibles au sein des G-quadruplexes. Les brins reliés par des boucles externes sont dans la même orientation, alors que les brins reliés par une boucle latérale ou diagonale sont dans des orientations opposées. En conséquence, un G-quadruplexe comportant une boucle autre que externe est de type anti-parallèle. (B) Deux conformations adoptées par la séquence télomérique humaine (oligonucléotide:  $(GGGTTA)_3GGG$ ). La conformation de gauche a été obtenue par RMN en solution ( $Na^+$ ) (Yang & Patel, Structure 1993) alors que la conformation de droite a été obtenue par cristallographie ( $K^+$ ) (Parkinson Nature 2002). Si le coeur central reste le même (à la différence des angles glycosidiques entre base et sucre), l'arrangement des boucles leur confère des structures tridimensionnelles très différentes. Dans des situations d'encombrement stérique (PEG), la structure obtenue est très similaire à la structure de droite (tout parallèle, boucles externes) (Heddi & Phan, JACS, 2011). Les guanines sont représentées en vert.

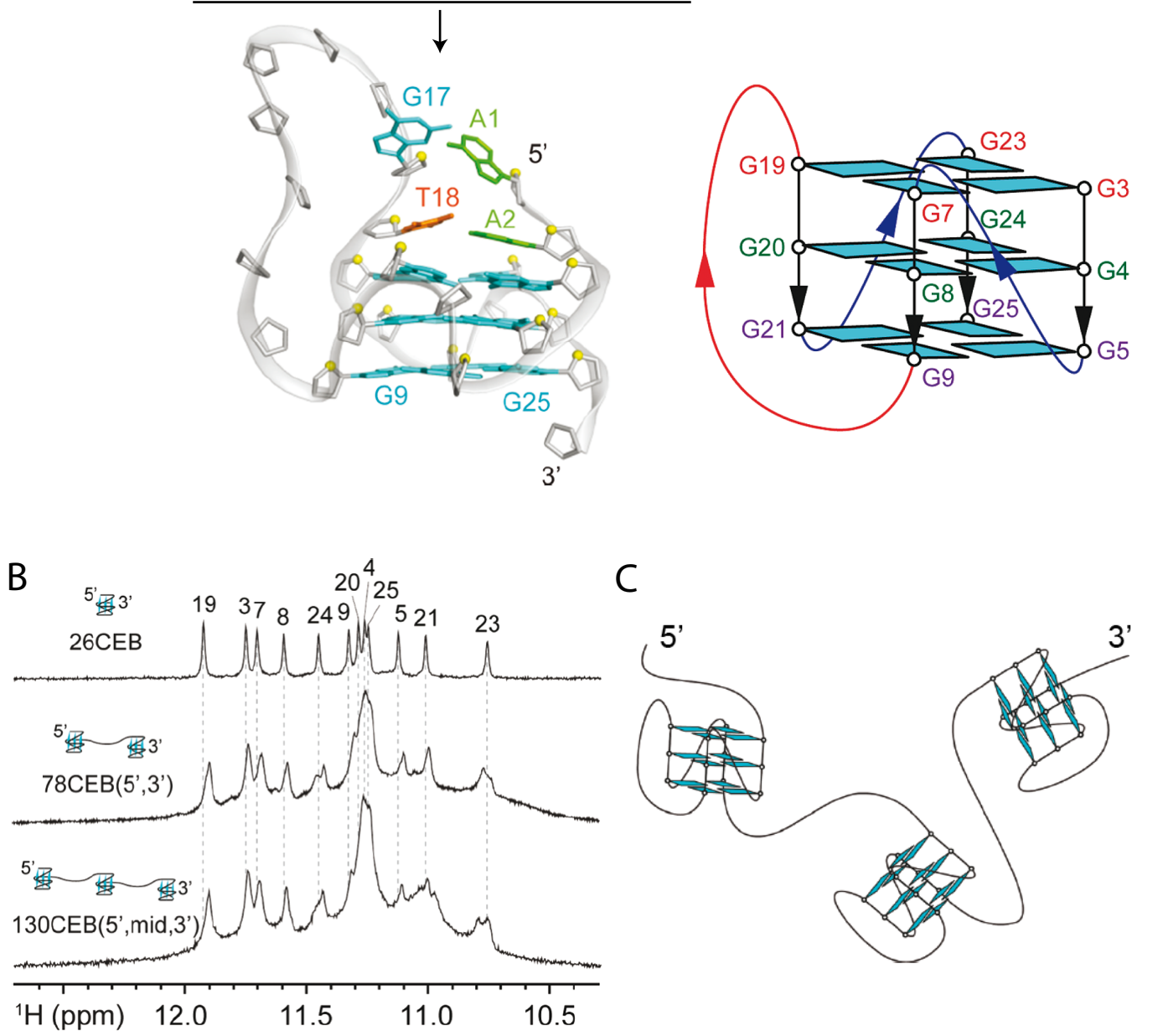
La séquence nucléotidique au sein de la boucle participe également à la stabilité du G-quadruplexe (Rachwal et al. 2007b; Guedin et al. 2008; Amrane et al. 2012). Un exemple stabilisateur provient de l'étude du minisatellite humain CEB25, objet de deux études durant ma thèse. Chaque motif forme un G-quadruplexe parallèle dont la structure a été résolue par RMN (**Figure 4A**) (Amrane et al. 2012). Toutes les boucles sont externes. La boucle centrale de 9 nt interagit avec la séquence flanquante via 2 de ses nucléotides (**Figure 4A**). Cette interaction, qui fait intervenir une association de bases canonique  $A_2:T_{18}$  et non-canonique  $A_1:G_{17}$  augmente la stabilité de la structure : la substitution des nucléotides de la boucle ou de la séquence flanquante résulte en une diminution de la stabilité thermique d'environ 5°C. Par ailleurs, une étude systématique de l'effet des 3 nucléotides A,T, et C au sein de boucles d'un nucléotide a également été réalisée (Guedin et al. 2008). La présence d'adénine dans les boucles latérales (1 et 3) est toujours défavorable, quelle que soient la nature du cation ( $K^+$  ou  $Na^+$ ) et la taille de la boucle centrale (3, 6, ou 9 nucléotides) en comparaison de thymine ou de cytosine. Le différentiel de  $T_m$  atteint entre la combinaison la moins stable (boucles 1 et 3 = A) et la plus stable (boucles 1 et 3 = C) est d'environ 10°C. La raison de cet effet déstabilisateur de l'adénine, ou stabilisateur des pyrimidines, n'est pas connue.

#### *I.A.1.b.v. Mise en tandem des séquences G4P*

Plusieurs séquences naturelles ont le potentiel pour former des G-quadruplexes en tandem. C'est le cas de nombreux minisatellites humains (voir **Article 4**), des séquences télomériques de nombreux organismes (Tran et al. 2010), ou de toutes séquences présentant un multiple de 4 triplets de G séparés par quelques nucléotides.

L'effet de la présence de plusieurs séquences G4P consécutives et recouvrantes au sein de longs ADNsb télomériques humains  $(TTAGGG)_n$  ( $n > 4$ ) affecte la stabilité des G-quadruplexes : l'ajout de répétitions aux 4 initialement requises pour former un G-quadruplexe réduit sa stabilité thermique (Yu et al. 2006b; Yu et al. 2006a; Petraccone et al. 2011). Cette déstabilisation est le résultat de l'augmentation de la taille des boucles, qui incluent les répétitions surnuméraires (Yue et al. 2011). L'étude d'oligonucléotide avec 8, 12, ou 16 répétitions TTAGGG montrent que la molécule (qui apparaît comme un collier de perle par AFM) forme rarement le nombre maximal de G-quadruplexe théoriquement possible (Wang et al. 2011a).

A CEB25: AAGGGTGGGTGTAAGTGTGGGTGGGTGTGAGTGTGGGTGTGGAGGTAGATGT



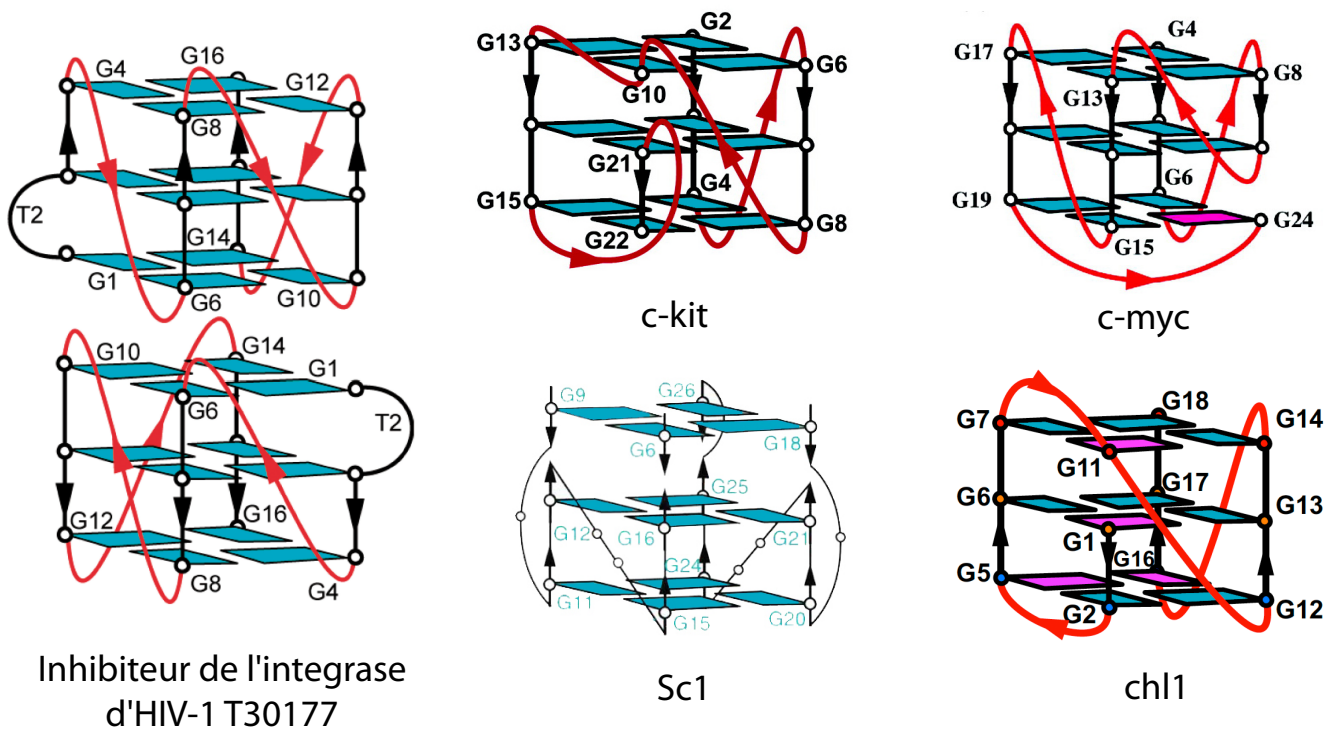
**Figure 4 : G-quadruplexe en tandem: exemple de CEB25 et collier de perle.** (A) Motif de CEB25, orienté 5'-3'. La séquence oligonucléotidique utilisée pour la détermination de la structure par RMN (26CEB) est soulignée, et correspond à un demi motif de CEB25. Structure représentée sous forme de ruban (Amrane et al. JACS 2012). Les nucléotides A1 et A2 interagissent avec les nucléotides de la boucle G17 et T18 respectivement. L'interaction A1-G17 est non-canonique. Une représentation simplifiée montre l'arrangement des boucles, toutes externes, et l'orientation des brins, parallèle. (B) Spectre imino-protons formé par l'oligonucléotide 26CEB seul (haut), un motif entier de CEB25 fusionnée au motif 26CEB suivant (milieu), et deux motifs de CEB25 suivit du demi-motif 26CEB (bas) (Amrane et al. JACS 2012). L'intensité des pics augmente mais leur position n'est pas affectée, ce qui suggère que les G-quadruplexes se forment indépendamment les uns des autres le long de la molécule. (C) Modèle du "collier de perle" de G-quadruplexes au sein d'une séquence minisatellite (Amrane et al. JACS 2012).

L'analyse de deux séquences G4P artificielles séparées par un linker de taille variable (TGGG)<sub>7</sub>N<sub>1-7</sub>(TGGG)<sub>7</sub> suggère quant à elle que les G-quadruplexes formés au sein de la même molécule se comportent indépendamment l'un de l'autre, probablement en raison de la distance importante entre eux (Payet and Huppert 2012). Plus récemment, l'analyse par AFM et RMN de trois motifs consécutifs du minisatellite humain CEB25 révèle une structure régulière de type « collier de perle », chaque motif formant un G-quadruplexe indépendamment des motifs voisins (**Figure 4B et 4C**).

Ainsi, lorsque plus d'une séquence G4P est présente au sein d'une molécule, la proximité de ces séquences est déterminantes pour leur comportement *in vitro* : si une région simple-brin de taille suffisante les sépare, les G-quadruplexes se formeront indépendamment sur la molécule (cas de CEB25). En revanche, si la taille de la région simple brin est de taille équivalente à la taille des boucles au sein de chaque séquence G4P (cas du télomère humain), ces triplets de G supplémentaires pourront être inclus au sein des boucles, conduisant à la formation de G-quadruplexes moins stables et moins nombreux. Cette formation de G-quadruplexes en tandem a des implications mécanistiques concernant l'instabilité des séquences minisatellites au cours de la réplication, qui seront discutées au **Chapitre III.A.3**.

#### *I.A.1.b.vi. G-quadruplexes atypiques*

En plus de la variété des conformations autorisées par les combinaisons de taille et séquence des boucles, des G-quadruplexes ne répondant pas à la séquence G4P consensus ont été caractérisés au cours des dernières années. Ces G-quadruplexes « exotiques » présentent pour la plupart des brins interrompus (**Figure 5**)(Kuryavyi and Patel 2010). Dans le cas d'un ARN (sc1), la structuration de l'ARNm (hairpins) amène à proximité des G éloignés sur la molécule et favorise la formation du G-quadruplexe (Phan et al. 2011). Dans ce cas, la jonction duplex-quadruplexe est reconnue spécifiquement par un peptide de la protéine FMRP, suggérant que ce mode de reconnaissance est à l'œuvre pour des ARN naturels, et constitue une cible privilégiée pour la reconnaissance spécifique par de petites molécules (voir **Chapitre I.A.2**). Certains possèdent également la propriété étonnante de se dimériser spontanément, comme le G-quadruplexe formé par l'inhibiteur d'HIV1 T30177 (**Figure 5**) (Do et al. 2011; Mukundan et al. 2011). Ces études soulignent la capacité des séquences riches en



**Figure 5: G-quadruplexes exotiques.** Représentations simplifiées des structures tridimensionnelles des G-quadruplexes formés par l'inhibiteur de l'intégrase d'HIV1 (T30177) (Mukundan et al. Nucl. Acids Res. 2011), du promoteur de c-kit (Phan et al. JACS 2007) et de c-Myc (Phan et al. Nature Chem. Biol. 2005), de l'ARN sc1 (Phan et al. Nature Str. Mol. Biol. 2011), et de l'intron du gène chl1 (Kuryavyi et al. Structure 2010). Les flèches indiquent l'orientation des brins 5'-3'. Aucune séquence à l'origine de ces G-quadruplexes ne contient 4 triplets de guanines: toutes les structures ont un brin interrompu. De plus, l'inhibiteur de l'intégrase d'HIV-1 forme un dimère symétrique. Les bases représentées en bleu ont une conformation syn, et les bases en violet une conformation anti, relativement au sucre.

G à former des G-quadruplexes, même si elles ne répondent pas exactement au consensus G4P intra-moléculaire (**Chapitre I.A.1.b.i**).

#### *I.A.1.b.vii. Conclusion*

En résumé, les G-quadruplexes apparaissent comme une famille de structures aux conformations très diverses qui se forment dans des conditions physiologiques *in vitro*. L'empilement central des G-quartets est le dénominateur structural autour duquel s'épanouissent des combinaisons variées d'arrangements de boucles et de brins. Mais si la séquence primaire influence la topologie de la structure, les déterminants structuraux ne s'arrêtent pas là : plusieurs formes peuvent coexister pour une même séquence, ou s'inter-converter en fonction des conditions expérimentales. C'est le cas des G-quadruplexes formés par la séquence télomérique humaine, le promoteur du gène c-Myc (Qin and Hurley 2008) et, comme nous le verrons au **Chapitre II.E.2**, du minisatellite CEB1.

#### **I.A.1.c. Modalités de formation d'un G-quadruplexe dans l'ADNdb**

Les études *in vitro* présentées jusqu'à maintenant ont été menées sur des ADN simples-brins, en absence du compétiteur naturel que constitue le brin complémentaire. Cette partie vise à introduire le problème central que constitue cette situation physiologique de compétition.

##### *I.A.1.c.i. La compétition avec l'ADN double brin*

Un G-quadruplexe se forme-t-il en présence de son brin complémentaire ? Autrement dit : quelle est la stabilité relative d'un ADNdb et d'un G-quadruplexe pour une séquence donnée ? La forme prépondérante adoptée par la séquence télomérique humaine (TTAGGG)<sub>n</sub> en présence du brin complémentaire est le duplex (Phan and Mergny 2002). De manière similaire, les profils thermodynamiques des formes duplexes et G-quadruplexes de 36 séquences humaines a montré que la forme duplex est plus stable que la forme G-quadruplexe, mais à des degrés variables (Kumar et al. 2008). Des expériences de molécules uniques ont montré que l'ajout du brin complémentaire en présence du G-quadruplexe télomérique pré-formé conduit à son déroulement rapide, dans la minute, à 37°C (Ying et al. 2003). Différemment, une étude de molécule unique a



montré que le G-quadruplexe du promoteur de c-kit se forme au sein d'un long substrat ADNdb dans environ 20-30% des molécules observées (Shirude et al. 2007).

Ainsi, même si des différences peuvent survenir d'une séquence à l'autre, la forme la plus stable et prépondérante à l'équilibre lorsque le brin riche en G est mis en présence de son complémentaire est la double-hélice classique (Lane 2012). Quelles conditions permettent alors une transition de l'ADNdb vers le G-quadruplexe ?

#### *I.A.1.c.ii. L'encombrement stérique*

Les conditions employées dans les tests *in vitro* sont loin de rendre compte des conditions intracellulaires, notamment en ce qui concerne l'encombrement stérique. En effet, le noyau des cellules contient une grande variété de molécules (acides nucléiques, protéines, polysaccharides, ...) pour une concentration massique d'environ 400 g/L (Miyoshi and Sugimoto 2008). *In vitro*, cet encombrement est simulé par des polymères tels que le PEG (polyéthylèneglycol). La présence de PEG affecte la stabilité de la double hélice (Miyoshi and Sugimoto 2008). Cet effet est particulièrement frappant en ce qui concerne l'équilibre entre ADNdb et G-quadruplexe : en présence de PEG, le duplex télomérique humain est déstabilisé au profit de la forme G-quadruplexe intramoléculaire (Miyoshi et al. 2004; Yu et al. 2006b; Zheng et al. 2010). De même, des concentrations croissantes de PEG diminuent la stabilité du duplex et augmente la stabilité du G-quadruplexe du promoteur de *NRAS* (Zheng et al. 2010). Le PEG permet également la formation et la stabilisation d'un G-quadruplexe au sein d'un long duplex après que celui-ci ait été ouvert par chauffage ou transcription du brin opposé (Zheng et al. 2010). Bien que la relevance du PEG pour simuler l'encombrement stérique produit par une variété d'agents au sein de la cellule puisse être discutée, ces conditions expérimentales se rapprochent des conditions physiologiques en ce qui concerne les volumes d'exclusion (Miyoshi and Sugimoto 2008) et tendent à relativiser les écarts de stabilité entre ADNdb et G-quadruplexe, qui diffèrent de 7 à 35°C suivant les séquences (Phan and Mergny 2002; Kumar et al. 2008). De plus, cet encombrement stérique, en diminuant l'hydratation de la structure restreint le nombre de conformations que peut adopter une même séquence : en présence de PEG, la séquence télomérique humaine arbore une seule conformation (parallèle, boucles latérales) au lieu d'un mélange hétérogène de 4 conformations différentes (**Figure 3B**) (Heddi and Phan 2011).

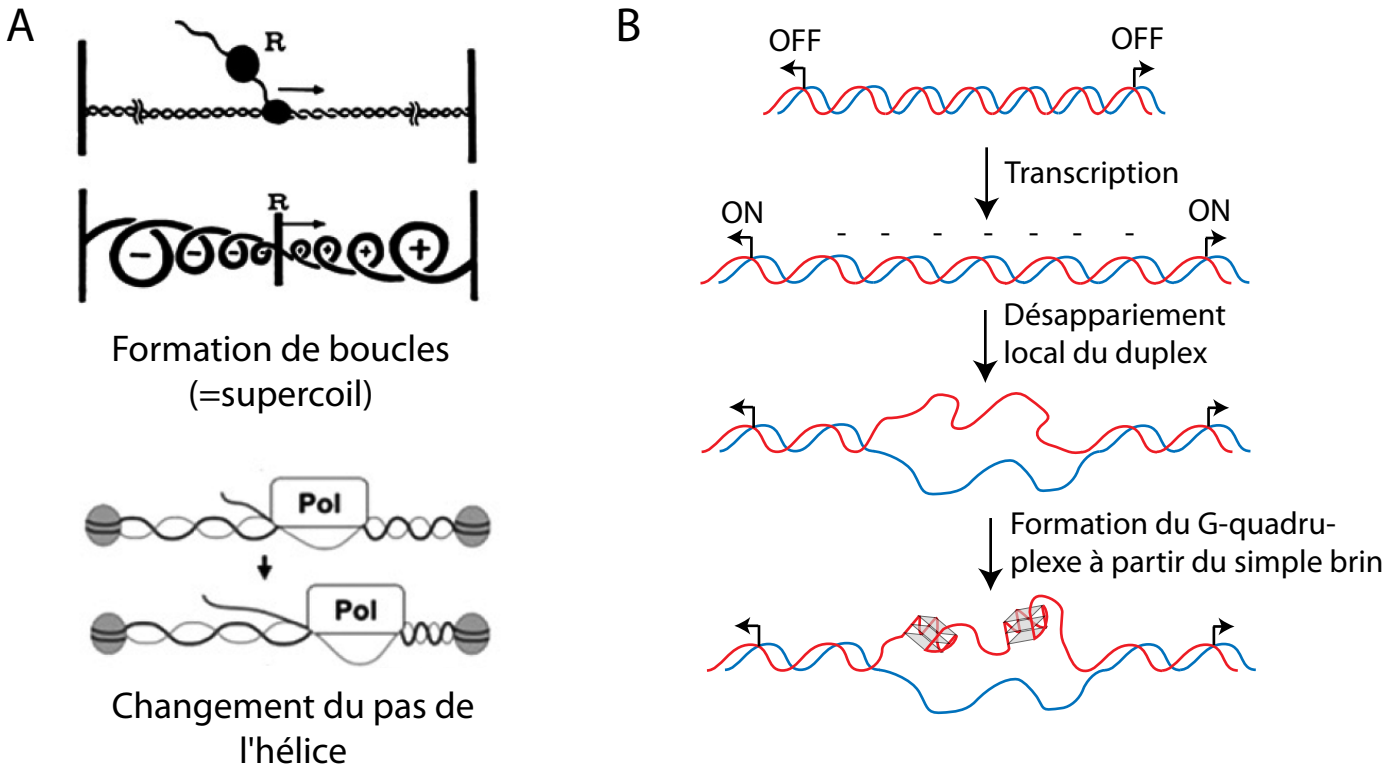
### I.A.1.c.iii. La superhélicité

La superhélicité correspond au nombre de tours qu'un brin d'ADN réalise autour de son brin complémentaire au sein de l'ADNdb (normalement un tous les 10 pb dans l'ADN-B non contraint). L'ADN est généralement sous-enroulé (superhélicité négative) *in vivo* (Zechiedrich et al. 2000). Des processus tels que l'ouverture du duplex ou la progression d'une machinerie de déroulement directionnelle au sein du duplex génèrent des variations locales de superhélicité, qui peuvent être relâchées par des topoisomérases, ou se propager dans la molécule sous forme de boucles (**Figure 6A**) (Koster et al. 2010; Roca 2011). Lorsque le duplex est grand et que les contraintes exercées sont importantes et non-relâchées survient un changement du nombre de « twist » au sein du duplex : les plateaux de base se rapprochent et la stabilité du duplex augmente lorsque la superhélicité est positive (Randall et al. 2009). En revanche, lorsque la superhélicité est négative, la stabilité du duplex diminue jusqu'à sa dénaturation, comme observé entre deux promoteurs actifs divergents (**Figure 6B**) (Kouzine et al. 2008). Cette déstabilisation du duplex fait basculer l'équilibre vers la forme G-quadruplexe du promoteur du gène c-Myc cloné dans un plasmide sous-enroulé (Sun and Hurley 2009). Ainsi, les variations locales de la superhélicité générées par la transcription ou la réplication sont susceptibles de modifier l'équilibre entre ADNdb et G-quadruplexe.

### I.A.1.c.iv. La G-loop

L'ADNdb est en compétition avec d'autres structures, notamment les hybrides ARN/ADN. La formation de tels hybrides laisse la molécule d'ADN déplacée sous forme simple brin. La structure résultante, observée par microscopie électronique, a été appelée R-loop (Thomas et al. 1976). Ainsi, en déplaçant l'équilibre duplex ADN vers hybride ARN/ADN, la transcription est, en théorie, à même d'autoriser la formation du G-quadruplexe sur l'ADNsb déplacé.

Le laboratoire de N. Maizels a démontré l'existence de G-quadruplexes formés sur le simple-brin d'une R-loop, par microscopie électronique d'ADN transcrit *in vitro* (**Figure 7A**) ou extrait d'*E. coli* (Duquette et al. 2004). Cette structure, appelée « G-loop », et a été observée en présence de plusieurs séquences G4P (c-Myc, Bcl-6, et RhoH, des régions de switch des chaînes lourdes d'Ig S $\mu$  et S $\gamma$ 3, ou de la répétition télomérique



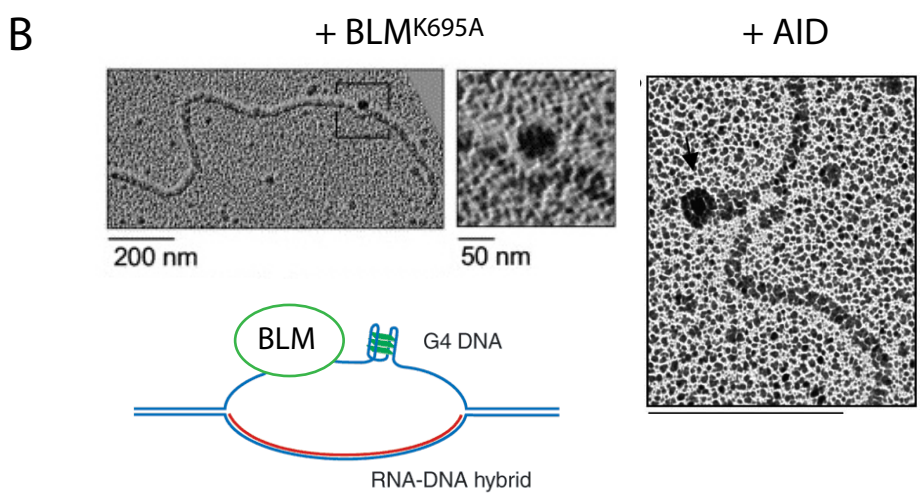
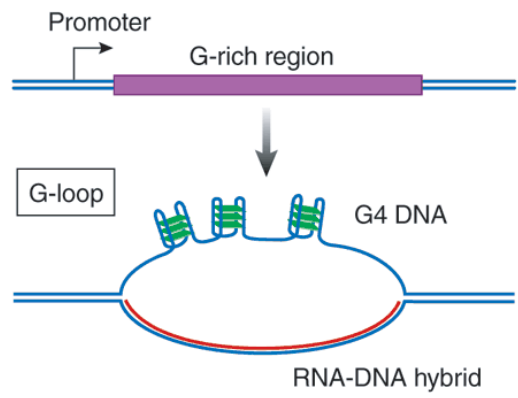
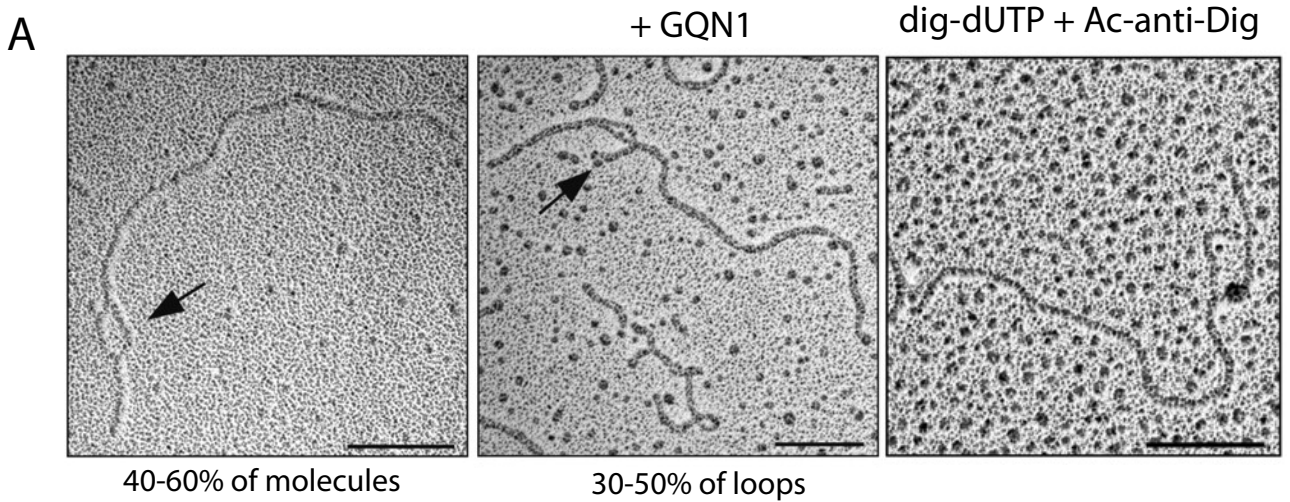
**Figure 6: La superhélicité négative promeut le désappariement du duplex et la formation du G-quadruplexe.** (A) Accumulation de contraintes positives en aval et négatives en amont d'une activité fixe de déroulement du duplex. Ces contraintes peuvent être absorbées sous forme de grandes boucles (haut), sans quoi elles modifient le pas de l'hélice (bas). Images tirées de (Lavelle Biochimie 2007). (B) Modèle de la formation de G-quadruplexes dans des régions localement sujettes à l'accumulation de contraintes négatives (exemple de deux promoteurs divergents actifs). La formation de boucles étant limitée par la taille de la région, les deux brins du duplex vont finir par se désapparer. La forme G-quadruplexe peut se former à partir du simple brin généré.

humaine) (Duquette et al. 2004; Duquette et al. 2005; Duquette et al. 2007). Aucune G-loop n'est formée en absence de transcription ou lorsque le brin riche en G est le brin transcrit (Duquette et al. 2004). Lorsque du digoxygénin-dUTP est ajouté lors de la réaction de transcription, un des deux brins de la G-loop est reconnu par un anticorps anti-dig couplé à une bille d'or (**Figure 7A**). L'ajout de RNaseA lors de la réaction de transcription n'affecte pas la fréquence de G-loop, alors qu'un traitement à la RNaseH l'abolit (Duquette et al. 2004). Ces résultats démontrent que la G-loop contient un hybride ARN/ADN formé co-transcriptionnellement. La présence de G-quadruplexes sur le brin opposé a été testée par digestion des plasmides avec la nucléase de G-quadruplexe GQN-1 (Sun et al. 2001). En présence de GQN-1, 30-50% des G-loops sont clivées (**Figure 7A**). L'ajout d'un excès 1000 fois supérieur d'ADNsb n'affecte pas l'efficacité du clivage, alors qu'un excès de 10 fois de G-quadruplexe l'abolit (Duquette et al. 2004), indiquant que c'est bien une structure G-quadruplexe qui est clivée par GQN-1 au niveau de la G-loop. De plus, des protéines de très haute affinité pour les G-quadruplexes, comme la nucléolin (Kd=0.5 nM) (Hanakahi et al. 1999) ou BLM (Sun et al. 1998), ciblent un des deux brins de la G-loop (**Figure 7B**). L'analyse de plasmides extraits d'une souche d'*E. coli* mutée pour l'hélicase RecQ et la RNaseH : 13% d'entre eux présentent une G-loop, dont 91% sont clivées après traitement par GQN-1 (Duquette et al. 2004). Ce résultat constitue la première preuve de l'existence des G-quadruplexes *in vivo*.

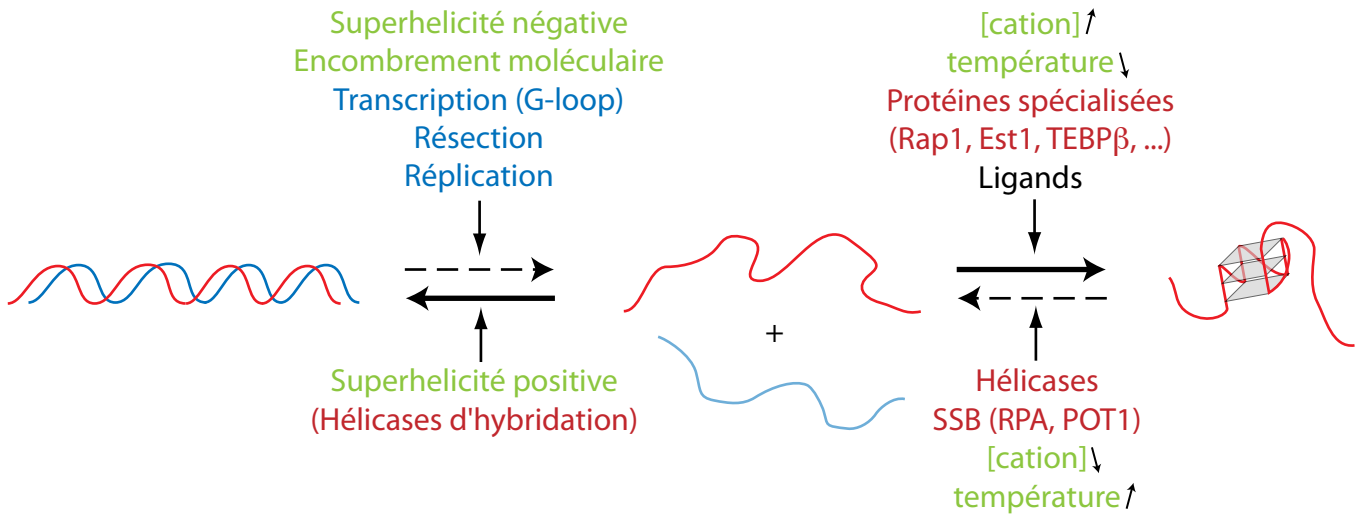
Ces résultats montrent que la G-loop est formée d'un hybride ARN/ADN et d'un simple-brin au sein duquel se sont formés des G-quadruplexes. Cette structure pourrait se former au niveau des gènes hautement transcrits comme le rDNA et ainsi permettre de maintenir ouvert le duplex, ou au niveau de la région S des chaînes lourdes d'immunoglobulines pour favoriser le ciblage par AID (Duquette et al. 2005).

#### *I.A.1.c.v. Conclusion*

Si l'équilibre mesuré *in vitro* entre ADNdb et G-quadruplexe est sensiblement en faveur de la forme duplex, le contexte cellulaire et les processus qui affectent localement sa stabilité sont à même de faire basculer l'équilibre en faveur de la forme G-quadruplexe. Une liste de ces activités biologiques, médiées ou non par des protéines, est présentée **Figure 8**. Ces structures étant très stables une fois formées, notamment les G-quadruplexes inter-moléculaires, des mécanismes à même de redéplacer



**Figure 7: La G-loop est une structure formée co-transcriptionnellement contenant un hybride ARN/ADN et un ou plusieurs G-quadruplexes sur le brin non-transcrit.** (A) Plasmides contenant un promoteur fort T7 et une séquence capable de former des G-quadruplexes après induction de la transcription *in vitro*, visualisés par microscopie électronique (gauche). Le pourcentage de molécules présentant cette structure est indiqué. En absence de transcription, aucune G-loop n'est observée. Le traitement des échantillons par la nucléase de G-quadruplexe GQN1 conduit à la formation de G-loops coupées uniquement sur l'un des deux brins (91% des G-loops)(milieu). La G-loop contient également un hybride ARN/ADN, détecté par incorporation de Dig-dUTP + Ac anti-Dig couplé à une bille d'or (droite). (Duquette et al. Genes & Dev. 2004). Barre = 500 nm. Schéma tiré de (Maizels, Nat. Str. Mol. Biol. 2006). (B) Un des brins de la G-loop est reconnu par plusieurs protéines présentant de l'affinité pour les G-quadruplexes *in vitro*: BLM<sup>K695A</sup> (helicase-dead, gauche) et AID (droite), ainsi que MutS $\alpha$  (non montré) (Huber et al. J. Mol. Biol. 2006; Duquette et al. Oncogene 2005).



**Figure 8: Facteurs qui affectent potentiellement l'équilibre ADNdb - G-quadruplexe *in vivo*.** Dans la cellule, l'équilibre entre ADNdb et G-quadruplexe est soumis aux propriétés physico-chimiques externes (cation, température, encombrement moléculaire) ou internes (superhelicité) à la molécule d'ADN (vert), aux activités positive ou négatives de protéines spécialisées (rouge), et à des processus biologiques générant activement de l'ADNsb (bleu). Certaines protéines pourraient directement catalyser la transition ADNdb -> G-quadruplexe.

l'équilibre vers les formes simple-brin (puis double-brins) semblent requis (**Figure 8**). Le rôle central des hélicases dans ce processus est abordé au **Chapitre I.A.5.b**.

## **I.A.2. Les ligands de G-quadruplexes**

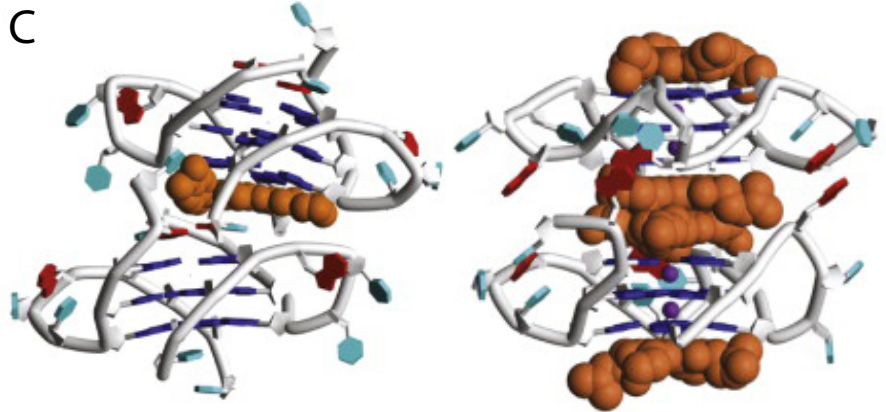
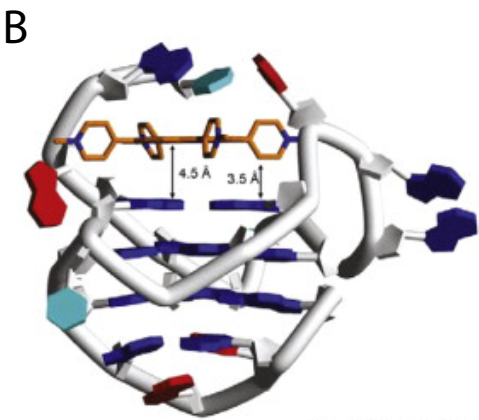
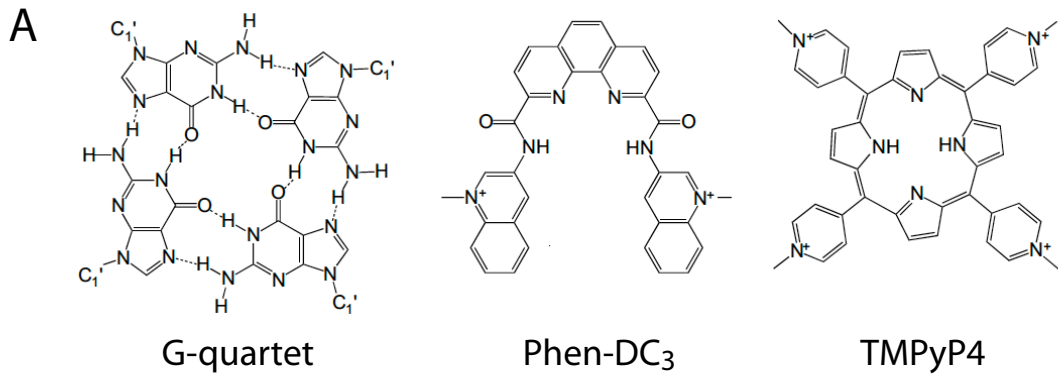
Le G-quadruplexe est une famille de structures aux topologies variées dont le G-quartet constitue l'unité essentielle et invariable. C'est cette partie invariable unique qui est la cible de la quasi-totalité des molécules appelées ligands de G-quadruplexe (De Cian et al. 2008b).

### I.A.2.a. Modalités d'interactions

La plupart des ligands de G-quadruplexe ciblent la partie invariante de la structure : le G-quartet. Aussi sont-ils généralement plans, et leur dimension recouvre celle d'un G-quartet (**Figure 9A**). Les structures obtenues par cristallographie de rayon X et par RMN montrent que le ligand se lie sur la face exposée d'un G-quartet externe (**Figure 9B**). La liaison se fait par interaction des orbitales électroniques  $\pi$  des groupements aromatiques du ligand et des guanines du G-quartet (**Figure 9B**). Cette interaction s'apparente à l'empilement naturel des G-quartets entre eux au sein du G-quadruplexe. Or, le « base stacking » conduit à une augmentation de la stabilité de la structure en limitant l'accessibilité du solvant. Ainsi, le ligand de G-quadruplexe n'est pas neutre vis-à-vis de la structure qu'il reconnaît ; il augmente sa stabilité, avec pour conséquence de déplacer l'équilibre thermodynamique vers la forme structurée. Dans certains cas même, il pourrait catalyser la formation du G-quadruplexe à partir d'ADN simple brin (Han et al. 1999a; De Cian and Mergny 2007). En plus d'augmenter la stabilité du complexe en comparaison du G-quadruplexe seul, ce mode d'interaction sur la face exposée des G-quartets externes du G-quadruplexe a des conséquences sur l'accessibilité des protéines aux nucléotides de la structure. Ceci est probablement à l'origine de l'inhibition qu'exercent nombre de ligands sur le déroulement des G-quadruplexes par les hélicases (voir **Article 2**).

### I.A.2.b. Sélectivité, affinité, et spécificité

Pour être utilisés dans les cellules et affecter spécifiquement les processus impliquant les G-quadruplexes, les ligands doivent être très sélectifs. Ceci est mesuré par des expériences de fixation du ligand au G-quadruplexe en présence d'un excès de



**Figure 9 : Ligands des G-quadruplexes: mode d'intercalation.** (A) Comparaison de la taille du G-quartet à deux ligands de G-quadruplexes. Tiré de (de Cian et al., Biochimie 2008). (B) Structure du G-quadruplexe formé par le promoteur de c-myc (G en bleu) avec le ligand TMPyP4 (orange), sur le G-quartet externe le plus accessible (Phan et al. Nat. Chem. Biol. 2005). (C) Exemple de dimérisation de deux G-quadruplexes télomériques humains par un ligand de G-quadruplexe. Les G-quadruplexes sont représentés en blanc, et les ligands en structure globulaire orange. Concernant le BRACO-19 (gauche), le ligand interagit sur le G-quartet inférieur du G-quadruplexe supérieur, et avec la boucle du G-quadruplexe inférieur, et non le G-quartet (Campbell et al. JACS 2008). A droite, deux ligands dérivés de naphthalène diimide permettent la dimérisation des G-quadruplexes. Les G-quartets externes sont également liés par le ligand (Parkinson et al., J. Mol. Biol 2008). Les images en (B) et (C) sont tirées de (Neidle Curr. Op. in Str. Biol. 2009).



compétiteur double-brins. Une molécule dont la sélectivité excède 20 est considérée comme un ligand de G-quadruplexe d'intérêt. Si la sélectivité vis-à-vis de l'ADNdb est évidemment un pré-requis, l'affinité l'est également pour espérer une application biologique et médicale, avec des  $K_d$  généralement inférieurs à 0,5  $\mu\text{M}$ . Étant donné le grand nombre de G-quadruplexes potentiellement formés dans les génomes, le ciblage spécifique requiert des molécules capables de discriminer différents G-quadruplexes. Cette recherche de spécificité, qui doit idéalement se faire sans perte de sélectivité ou d'affinité pour le G-quadruplexe d'intérêt, passe par le ciblage des boucles ou des sillons du G-quadruplexe, en plus du G-quartet, en ajoutant des branchements, des « bras », à la molécule plane centrale.

#### I.A.2.c. Le test FID

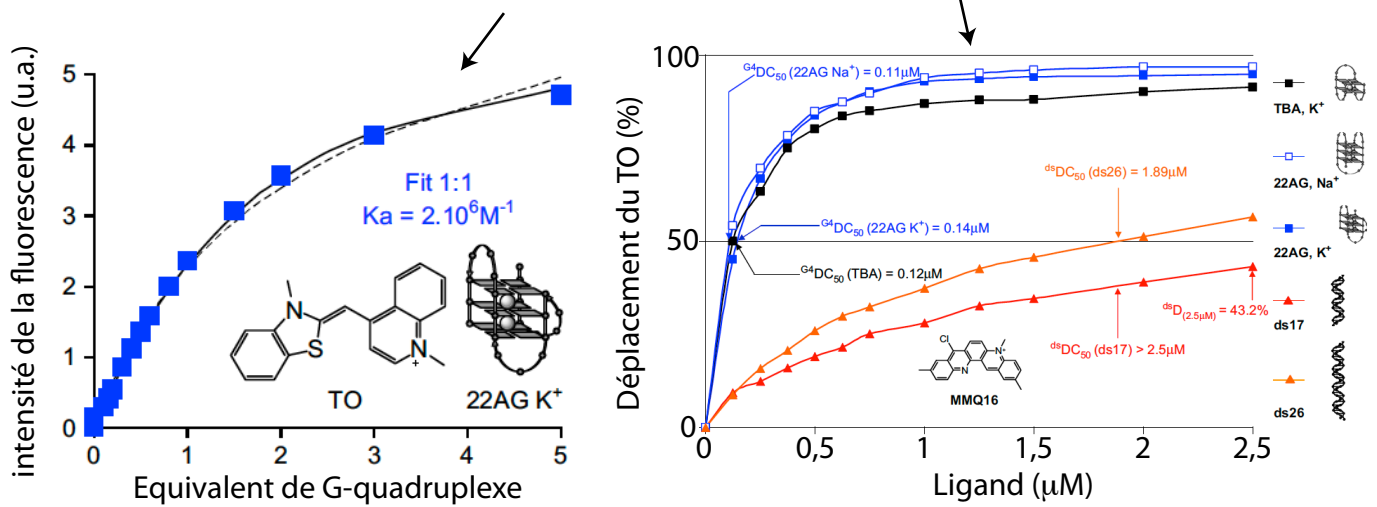
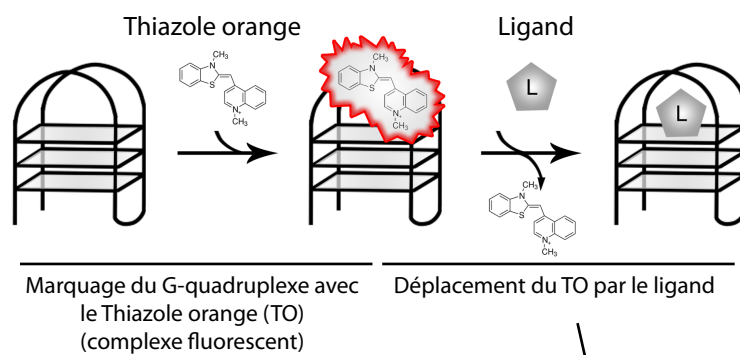
Comme le complexe G-quadruplexe/ligand est plus stable que le G-quadruplexe seul, la fixation peut être mesurée par l'analyse de la stabilité de la structure par FRET (De Cian et al. 2007b) ou spectroscopie UV (Rachwal and Fox 2007). Cependant, ces tests ne permettent pas de définir une constante d'affinité (Guedin et al. 2010b). Une autre méthode consiste à mesurer le déplacement d'un autre ligand de G-quadruplexe : le thiazole orange (TO). Une fois fixé sur un G-quartet, le TO devient plan et fluorescent. Son déplacement par le ligand d'intérêt est mesuré par perte de fluorescence (FID, pour Fluorescence Displacement Assay, schématisé **Figure 10A**) (Monchaud et al. 2008). Cette méthode permet de déterminer une constante d'affinité du ligand pour le G-quadruplexe. La sélectivité est déterminée en présence de concentrations croissantes d'ADNdb compétiteur (**Figure 10C**) (Monchaud et al. 2008).

#### I.A.2.d. Exemples de ligands utilisés dans des études *in vivo*

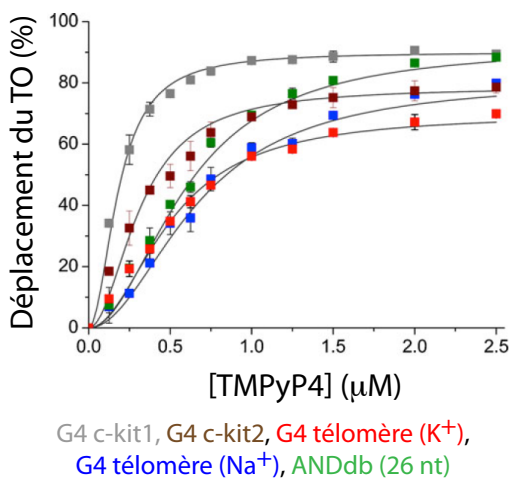
##### *I.A.2.d.i. Les porphyrines*

Les porphyrines TMPyP4 (5,10,15,20-Tetra-(N-Methyl-4-Pyridyl)Porphyrine) (**Figure 9A**) (Wheelhouse et al. 1998) et NMM (N-Methyl-Mesoporphyrin IX) (**Article 2, Figure 1**) (Ren and Chaires 1999) comptent parmi les premiers ligands de G-quadruplexes décrits. Ces molécules sont largement utilisées pour interférer avec des processus impliquant un G-quadruplexe, *in vitro* (Sun et al. 1999; Wu and Maizels 2001; Huber et al. 2002) et *in vivo* (Siddiqui-Jain et al. 2002; Hershman et al. 2008; Cahoon and Seifert 2009; Decorsiere et al. 2011; Vannier et al. 2012). Malheureusement, la faible

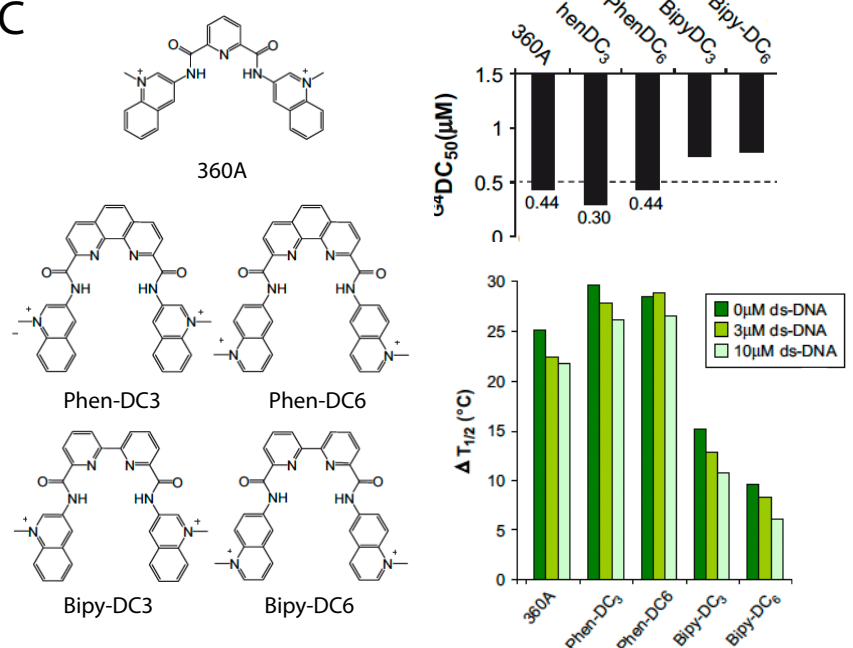
A



B



C



**Figure 10: Mesure de l'affinité et de la sélectivité de ligands de G-quadruplexe par le test FID. (A)** Principe du test FID (Fluorescence Displacement Assay) (Article 2, Figure 1). L'exemple de l'association du TO au G-quadruplexe formé par la répétition télomérique humaine AGGG(TTAGGG)<sub>3</sub> (22AG) (à gauche), et son déplacement par un ligand (MMQ<sub>18</sub>, à droite) est présenté. Une molécule de TO se fixe sur le G-quadruplexe. La DC<sub>50</sub>, exprimé en μM, est une mesure de l'affinité du ligand pour la structure. Le MMQ<sub>18</sub> a une meilleure affinité pour les G-quadruplexes que pour les ADNdb (Monchaud *et al.* Biochimie 2008). (B) Déplacement du TO par TMPyP4 sur différents G-quadruplexes et un ADNdb (Largy *et al.* Anal. Bioanal. Chem. 2011). (C) Affinité et sélectivité de bisquinoliniums (et MMQ<sub>18</sub>) pour les G-quadruplexes, déterminés par FID et expériences de dénaturation, respectivement. La sélectivité est mesurée en présence de concentrations variables d'ADNdb compétiteur. MMQ<sub>18</sub> est peu sélectif car un excès de 300 à 1000 fois d'ADNdb diminue fortement la stabilité du G-quadruplexe. En revanche, les bisquinoliniums sont peu, voire pas affectés par cet excès d'ADNdb compétiteur, soulignant leur sélectivité pour les G-quadruplexes (Monchaud *et al.* Biochimie 2008).

affinité du NMM pour les G-quadruplexes (Ren and Chaires 1999) et la spécificité au mieux 2 à 4 fois supérieure de TMPyP4 pour les G-quadruplexes en comparaison du duplex (**Figure 10B**) (Dixon et al. 2005; Parkinson et al. 2007; Largy et al. 2011) font de ces composés des ligands de G-quadruplexes médiocres, et proscrivent leur utilisation *in vivo*.

#### *I.A.2.d.ii. Les bisquinoliniums*

La famille des bisquinoliniums, développée par l'équipe de MP Teulade-Fichou, compte une dizaine de ligands dont les composés Phen-DC<sub>3</sub> font partie (**Figure 10C**) (De Cian et al. 2007a; De Cian et al. 2008a; Monchaud et al. 2008; Lacroix et al. 2010). L'affinité de ces composés pour le G-quadruplexe formé par le télomère humain est élevée, avec une DC<sub>50</sub>  $\approx$  0.3  $\mu$ M (**Figure 10C**) (Monchaud et al. 2008; Largy et al. 2011). Cette affinité élevée est associée avec un fort pouvoir stabilisateur sur le G-quadruplexe qui se traduit par une augmentation de la résistance à la dénaturation thermique ( $\Delta T_m$ , en °C) d'environ 25°C en présence de 360A et 30°C en présence Phen-DC<sub>3</sub> (**Figure 10C**, droite) (De Cian et al. 2007a; Monchaud et al. 2008). La sélectivité est mesurée par diminution du  $\Delta T_m$  en présence d'un excès de compétiteur double-brin: les Phen-DC et 360A sont très sélectifs, car en présence d'un excès d'ADNsb le  $\Delta T_m$  diminue peu (**Figure 10C**).

En plus de stabiliser le G-quadruplexe, le 360A stimule la formation de G-quadruplexes inter-moléculaires: les constantes de formation très faibles et dépendantes de la concentration d'ADNsb ( $k_{on} \approx 10^6-10^8 \text{ M}^{-3} \cdot \text{s}^{-1}$  à 100  $\mu$ M d'oligonucléotides) augmentent de plusieurs ordres de grandeur en présence de ce composé ( $k_{on} \approx 10^{12}-10^{13} \text{ M}^{-3} \cdot \text{s}^{-1}$ ) (Mergny et al. 2005; De Cian and Mergny 2007). Ainsi, 360A déplace l'équilibre entre ADNsb et G-quadruplexe vers la forme G-quadruplexe à la fois en favorisant la formation ( $k_{on}$  augmente), et prévenant la dénaturation du G-quadruplexe ( $k_{off}$  diminue).

Les composés Phen-DC n'ont pas de spécificité particulière: l'affinité est très similaire pour de nombreux G-quadruplexes, dont celui formé par CEB1 (voir **Article 2**) (Lacroix et al. 2010; Largy et al. 2011). Ces composés sont les plus affins et sélectifs pour les G-quadruplexes actuellement connus (Lacroix et al. 2010).

### I.A.3. Existence des G-quadruplexes *in vivo*

La capacité de certaines séquences riches en G à former spontanément des structures G-quadruplexes intra-moléculaires dans des conditions physiologiques *in vitro* suggère qu'elle se forment *in vivo* et pose la question de leurs potentiels rôles biologiques. L'analyse bioinformatique d'enrichissement, de localisation, et de conservation des séquences G-quadruplexes putatives (G4P) des génomes a suggéré des rôles pour ces structures, dont certaines ont fait l'objet de validations expérimentales.

#### I.A.3.a. Les séquences G-quadruplexes putatives des génomes

##### *I.A.3.a.i. Les algorithmes de recherche et leurs limitations*

Les algorithmes les plus simples scannent pour la présence de 4 triplets de G dans une fenêtre glissante de 100 pb (Eddy and Maizels 2006). Cet algorithme pose une contrainte sur la somme de la taille des 3 boucles :

$$G_3N_xG_3N_yG_3N_zG_3 \text{ avec } x+y+z \leq 88$$

Ce genre d'algorithme est toujours utilisé (Hershman et al. 2008), quoiqu'avec des fenêtres plus restreintes, le paramètre original d'une fenêtre de 100 pb étant bien trop permissif (voir taille des boucles, **Chapitre I.A.1.b.iv**). Un second type d'algorithme, plus contraignant, pose une limite à la taille de chaque boucle (Huppert and Balasubramanian 2005; Todd et al. 2005) :

$$G_{3-5}N_{1-7}G_{3-5}N_{1-7}G_{3-5}N_{1-7}G_{3-5}$$

Ce consensus permet la sélection de séquences capables de former des G-quadruplexes dont le  $T_m$  est généralement supérieur à 41°C. C'est ce consensus qui a été utilisé par le logiciel d'analyse de génome comme Quadparser (Huppert and Balasubramanian 2005; Todd et al. 2005), utilisable sur [www.quadruplex.org](http://www.quadruplex.org) (Wong et al. 2010). Cet algorithme couramment utilisé a permis l'identification des séquences G4P du génome humain (Huppert and Balasubramanian 2005; Todd et al. 2005). Il est par certains côtés trop laxiste car il sélectionne des séquences possédant 3 boucles de 5 à 7 nt, formant des G-quadruplexes peu stables *in vitro* (Rachwal et al. 2007c), et par certains côtés trop contraignant car il élimine toute séquence possédant une boucle supérieure à 7 nt, alors que des G-quadruplexes stables se forment avec des boucles

pouvant aller jusqu'à 25 nt, pour peu que les autres boucles soient courtes (**Chapitre I.A.1.b.iv**) (Guedin et al. 2010a).

Une troisième génération d'algorithme, dit « d'apprentissage », utilise les données de stabilité thermique de nombreux G-quadruplexes *in vitro* pour prédire, par similarité, la capacité d'une séquence à former un G-quadruplexe (Stegle et al. 2009). Suffisamment renseigné, cet algorithme pourrait se substituer à la recherche de motifs G4P par utilisation d'une séquence consensus.

#### *I.A.3.a.ii. Distribution des séquences G-quadruplexes dans le génome humain*

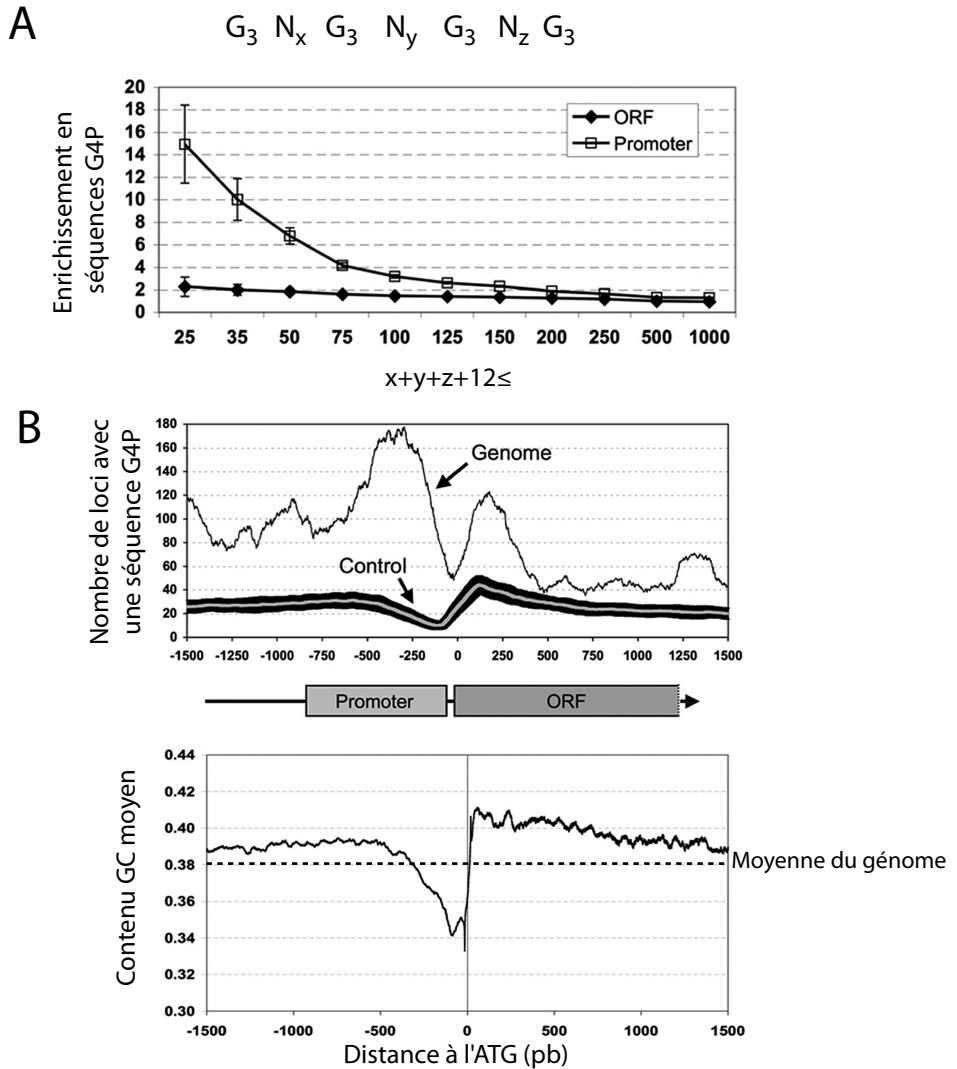
Certaines séquences capables de former des G-quadruplexes ont été identifiées à certains loci comme les télomères, le rDNA, et les régions de Switch des Ig avant l'avènement de la bioinformatique (Henderson et al. 1987; Sen and Gilbert 1988; Williamson et al. 1989). L'analyse bioinformatique a permis la cartographie exhaustive de séquences G-quadruplexes putatives (G4P) dans le génome humain: 376000 séquences G4P ont été identifiées (Huppert and Balasubramanian 2005; Todd et al. 2005). C'est moins que le nombre de G4P attendu par le hasard, indiquant que ces séquences sont sous-représentées dans le génome (Huppert and Balasubramanian 2005). Cette sous-représentation ne concerne pas tous les loci : à l'exception des exons (Eddy and Maizels 2009), les gènes sont enrichis en séquences G4P au niveau de leur promoteur, du 1<sup>er</sup> intron, et des régions 5' et 3'UTR (Huppert and Balasubramanian 2007; Du et al. 2008; Eddy and Maizels 2008; Huppert et al. 2008; Eddy and Maizels 2009). En effet, plus de 40% des gènes humains contiennent au moins une séquence G-quadruplexe dans leur promoteur (jusqu'à 1 kb du TSS) (Huppert and Balasubramanian 2007; Zhang et al. 2008). Les protooncogènes semblent particulièrement enrichis pour ces séquences G4P, contrairement aux gènes suppresseurs de tumeurs (Eddy and Maizels 2006). La formation de G-quadruplexes intra-moléculaires a été vérifiée *in vitro* pour plusieurs de ces séquences (Kumar and Maiti 2008; Qin and Hurley 2008) notamment celles présentes dans les promoteurs des oncogènes : c-Myc (Siddiqui-Jain et al. 2002), VEGF (Sun et al. 2005), PDGF-A (Qin et al. 2007), HIF-1 $\alpha$  (De Armond et al. 2005), KRAS (Cogoi and Xodo 2006), c-Kit (Rankin et al. 2005; Phan et al. 2007; Shirude et al. 2007), c-Myb (Broxson et al. 2011), Bcl-2 (Dai et al. 2006), et Ret (Guo et al. 2007).

Presque la moitié des gènes humains contiennent au moins une séquence G4P dans leur premier intron (Eddy and Maizels 2008). Cet enrichissement a lieu principalement dans la partie 5' de l'intron 1, très proche du TSS, et majoritairement sur le brin non-transcrit (Eddy and Maizels 2008). Il est intrigant de constater que les gènes sujets à des pauses de transcription juste en aval du TSS (Gilmour 2009) sont enrichis en séquences G4P dans leur premier intron, suggérant un rôle potentiel des G-quadruplexes dans cet arrêt et la maturation de ces ARNm (Reinhold et al. 2010; Eddy et al. 2011).

Concernant les régions UTR, environ 4000 (12%) et 5000 (15%) gènes humains possèdent au moins une séquence G4P dans leur 5'UTR et leur 3'UTR, respectivement (Huppert et al. 2008). Concernant le 5'UTR, cela représente une densité de G4P par kb cinq fois supérieure à celle observée pour le reste du génome (Beaudoin and Perreault 2010) et environ deux fois supérieure pour les 3'UTR (Huppert et al. 2008). Les séquences G4P dans l'ARN sont sous représentées dans les UTR, à l'exception des 50 premiers nucléotides du 5'UTR (Huppert et al. 2008). Cet enrichissement a suggéré un contrôle de l'expression génétique au niveau post-transcriptionnel pour certaines classes de gènes qui a été démontré expérimentalement depuis. Les rôles divers des G-quadruplexes dans le métabolisme des ARNs ne seront pas abordé dans ma thèse, mais ont fait l'objet de revues récentes (Bugaut and Balasubramanian 2012; Millevoi et al. 2012).

#### I.A.3.a.iii. *Distribution des séquences G-quadruplexes dans le génome de S. cerevisiae*

Une première étude utilisant un algorithme à fenêtre glissante de taille raisonnable ( $\leq 50$  nt) a identifié 854 séquences G4P chez *S. cerevisiae* (Hershman et al. 2008). Ces séquences sont enrichies dans les promoteurs, alors même que les promoteurs ne présentent pas une richesse GC particulière (**Figure 11B**). Cet enrichissement est maximal avec la taille de fenêtre la plus petite (25 nt,  $n=54$ ), et va décroissant quand la taille de fenêtre augmente (et conjointement le nombre de faux positifs) (**Figure 11A**). Ainsi, l'enrichissement diminue quand la stringence diminue. Ceci suggère que seules les séquences détectées avec un algorithme stringent (boucles courtes, haute stabilité *in vitro*) ont un rôle potentiel de contrôle transcriptionnel. De manière similaire à ce qui est observé avec la fenêtre de 25 nt, notre étude des



**Figure 11: Les G-quadruplexes dans le génome de *S. cerevisiae*.** (A) Enrichissement des séquences G4P dans les promoteurs (0 à -850 pb) et les ORF en fonction de la stringence de l'algorithme de recherche (taille totale de la séquence G4P). L'enrichissement est calculé en divisant la proportion de séquences G4P observées sur l'enrichissement moyen observé sur 1000 génomes de *S. cerevisiae* générés aléatoirement, construits en respectant les biais de fréquence des bases à chaque position de l'ATG (voir (B)). L'enrichissement de séquences G4P est fort dans le promoteur, et diminue quand la stringence de l'algorithme diminue. (B) Distribution des séquences G4P (taille inférieure à 50 pb) autour de l'ATG dans le génome humain. Le pic de séquences G4P dans le promoteur ne peut pas être simplement expliqué par une augmentation du pourcentage GC à ce niveau.

séquences G4P chez *S. cerevisiae* identifie 50 séquences G4P à l'aide d'un algorithme qui impose des contraintes sur la taille de chaque boucle ( $G_{3-5}N_{1-9}G_{3-5}N_{1-9}G_{3-5}N_{1-9}G_{3-5}$ )(**Article 3, Table S3**).

Contrairement aux séquences télomériques de la majorité des organismes séquencés à l'heure actuelle, les télomères de *S. cerevisiae* ne contiennent pas de séquence consensus capable de former systématiquement de G-quadruplexes intra-moléculaires stables (Tran Mergny NAR 2010). En revanche, la présence de triplets de G, dispersés au sein des  $\approx 300$  nt que compte chaque télomère, autorise des associations inter-moléculaires, similaires à celles observées chez *Stylochoia lemnae* (Tran et al. 2010). En faveur de cette hypothèse, les protéines télomériques de levure Rap1 et Est1 (qui lient respectivement l'ADNdb et l'ADNsb télomérique) catalysent la formation de G-quadruplexes inter-moléculaires *in vitro* (Giraldo and Rhodes 1994; Giraldo et al. 1994; Zhang et al. 2010).

Le faible nombre de séquences G4P identifiées dans le génome de *S. cerevisiae* suggère un rôle marginal des G-quadruplexes intra-moléculaires chez cet organisme. En revanche, la présence de protéines capables de catalyser la formation de G-quadruplexes inter-moléculaires, très stables une fois formés (**Chapitre I.A.1.b.i**), suggère l'existence de mécanismes à même de les dérouler efficacement.

### I.A.3.b. Rôles biologiques des G-quadruplexes

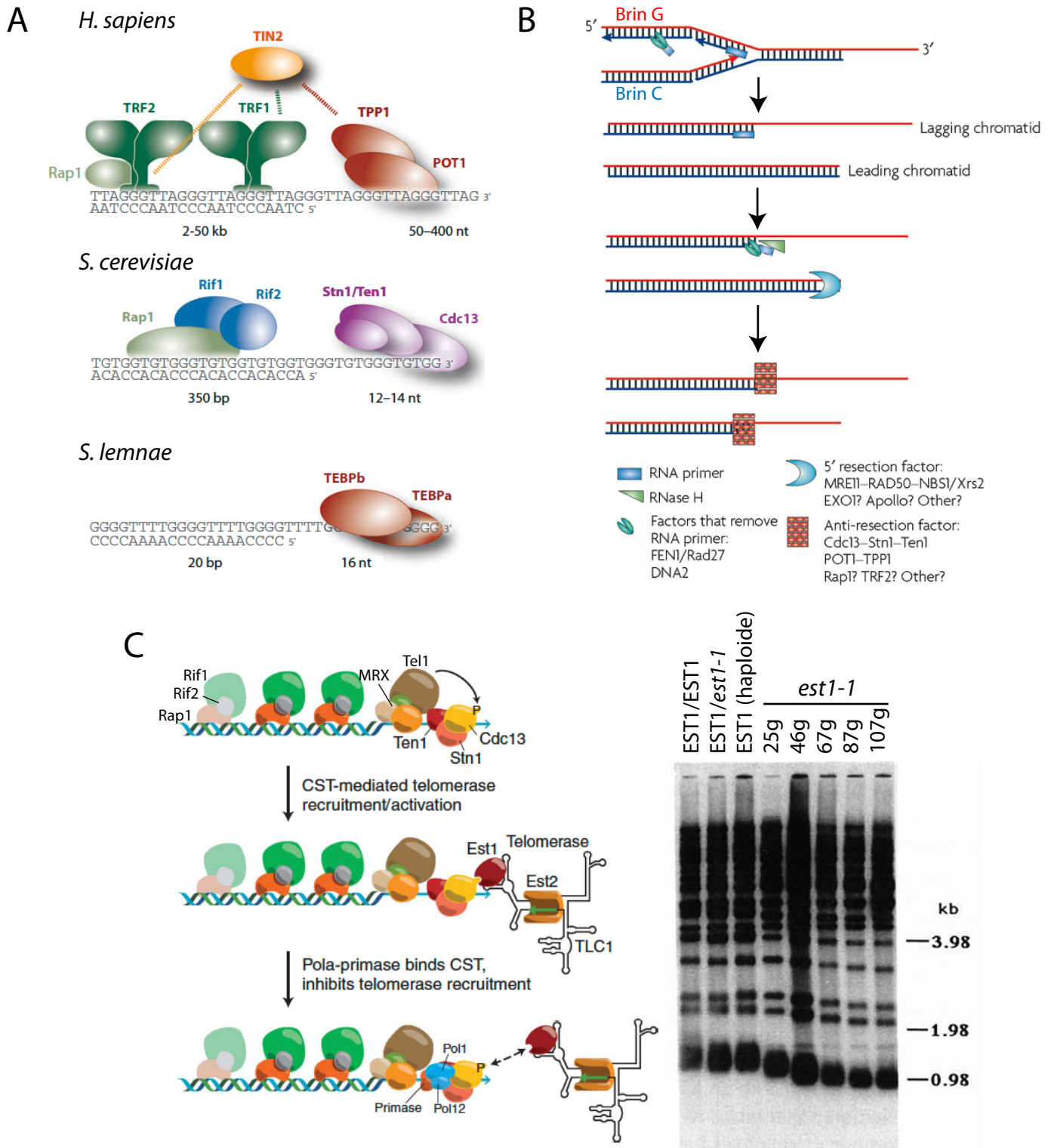
Plus de 25 ans après leur découverte, les G-quadruplexes étaient toujours considérés comme des « structures in search of functions » (Simonsson 2001). L'analyse bioinformatique des génomes suggère des rôles biologiques variés pour ces structures, mais ne constitue pas une démonstration expérimentale. L'objectif de cette partie est de présenter de manière exhaustive les preuves expérimentales les plus solides concernant l'existence des G-quadruplexes *in vivo* et leurs différents rôles biologiques. Ces démonstrations reposent soit sur la détection directe des structures *in vivo*, soit par la combinaison d'une approche de mutagénèse dirigée du motif G-quadruplexe et d'interférence chimique avec un ligand spécifique de G-quadruplexes.

#### *I.A.3.b.i. Capping télomérique chez *Stylochoia lemnae**



Les chromosomes eucaryotes sont linéaires et possèdent à leurs extrémités une structure nucléoprotéique appelée télomère (**Figure 12A**). La séquence télomérique consiste en la répétition d'un motif nucléotidique de 5-8 pb généré par la télomérase. Alors que la séquence télomérique assure l'intégrité des parties internes du chromosome en tamponnant pour un certain nombre de générations le « end replication problem » (**Figure 12B et 12C**), la structure multi-protéique (shelterin chez l'homme) cache l'extrémité télomérique aux machineries de détection des cassures doubles-brins (CDB) et régule l'activité de la télomérase (Palm and de Lange 2008). La réplication incomplète à l'origine de ce raccourcissement progressif des télomères, ainsi que des activités nucléases, contribuent à la formation d'une extrémité 3' débordante appelée « overhang » qui peut atteindre 400 nt chez l'homme (**Figure 12A**). La particularité de cette séquence télomérique est d'être riche en GC, avec un biais de brin fort qui place les G sur l'overhang. L'analyse *in vitro* de 15 séquences télomériques eucaryotes montre que la quasi-totalité forme spontanément des G-quadruplexes intra-moléculaires (Tran et al. 2010). C'est notamment le cas de la séquence télomérique humaine TTAGGG (un motif télomérique conservé chez les vertébrés (Moyzis et al. 1988; Meyne et al. 1989)), très étudiée *in vitro* et dont les quatre conformations ont été résolues par cristallographie aux rayons X (Parkinson et al. 2002) et RMN (**Figure 3B**) (Wang and Patel 1993; Dai et al. 2008; Heddi and Phan 2011).

Le premier exemple de formation des G-quadruplexes au télomère *in vivo*, et sans doute le mieux démontré, provient de l'étude du cilié *Stylonichia lemnae*, qui possède approximativement  $10^8$  chromosomes. L'ADN extrait du micronoyau et analysé par microscopie électronique montre des micro-chromosomes associés par leurs extrémités (**Figure 13A**) (Lipps 1980). Ces associations requièrent la présence d'ADN simple-brin puisque la digestion à la nucléase S1 les abolit. L'overhang télomérique de ces ciliés consiste en 16 nt T<sub>4</sub>G<sub>4</sub>T<sub>4</sub>G<sub>4</sub> (**Figure 12A**) qui s'associent entre eux *in vitro* sous forme de G-quadruplexes bi-moléculaires, parallèle ou anti-parallèle (**Figure 13B**) (Haider et al. 2002). Deux anticorps spécifiques et affins ont pu être obtenus : l'un est dirigé contre la forme parallèle (Sty3, K<sub>d</sub>=0.1 nM) et l'autre sans préférence de conformation (Sty49, K<sub>d</sub>≈2-4 nM) (Schaffitzel et al. 2001). Ces anticorps permettent la détection directe par immunohistochimie des G-quadruplexes télomériques de ces ciliés sur cellules fixées. Alors que l'incubation de cellules de *Stylonichia* fixées avec l'anticorps Sty3 ne produit aucun signal, l'anticorps Sty49 allume quasiment l'intégralité du macronoyau, indiquant



**Figure 12: Les télomères et la télomérase.** (A) Structure du télomère chez l'homme, *S. cerevisiae* et *S. lemnae*. Tiré de (Palm & de Lange, Ann. Rev. Genet. 2008). (B) "End replication problem" et activités nucléases responsables de la diminution progressive de la taille des télomères. Tiré de (Gilson & Géli Nat. Str. Mol. Biol. 2007). (C) Modèle d'action de la télomérase chez *S. cerevisiae*, tiré de (Shore & Bianchi EMBO J. 2008). Tel1 (marron) phosphoryle Cdc13, qui recrute la télomérase via sa sous-unité Est1. La télomérase rétro-transcrit alors une région de sa sous-unité Tlc1 (flèche verte) à l'extrémité 3' du télomère (flèche bleue), une réaction catalysée par Est2. En absence d'activité télomérase dans des cellules eucaryotes en croissance (cellules somatiques de vertébrés) la taille des télomères diminue progressivement. L'exemple d'un mutant de télomérase (*est1-1*) de *S. cerevisiae* est présenté à droite: le raccourcissement des télomères est suivi au cours des générations (25 à 107) par Southern blot (Lundblad & Zostak Cell 1989).

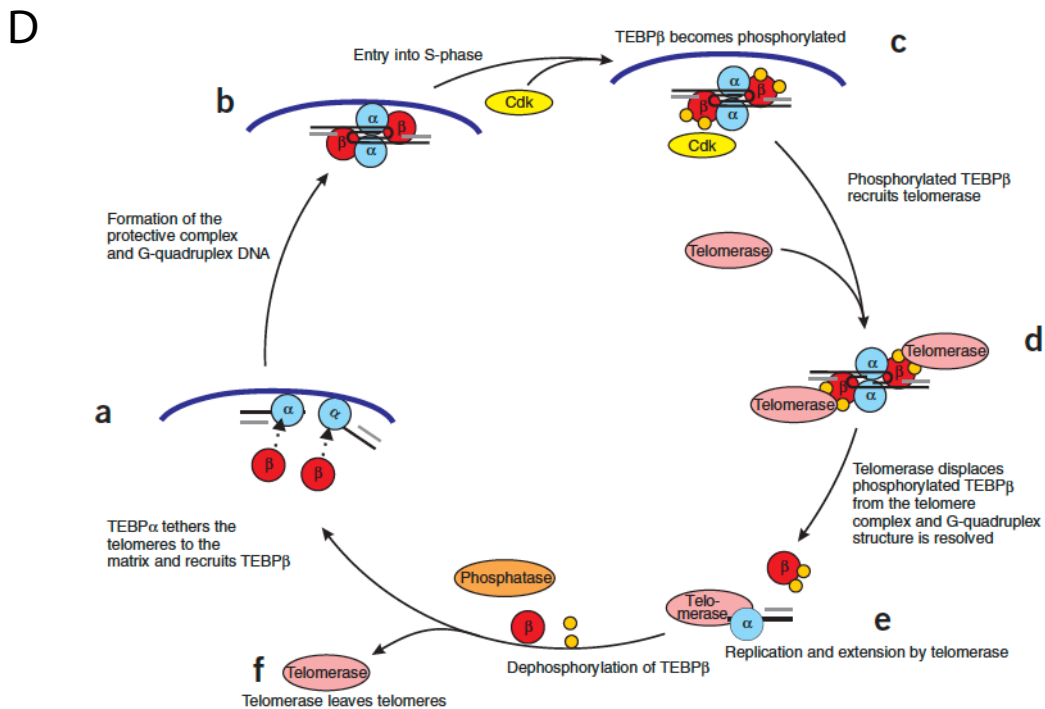
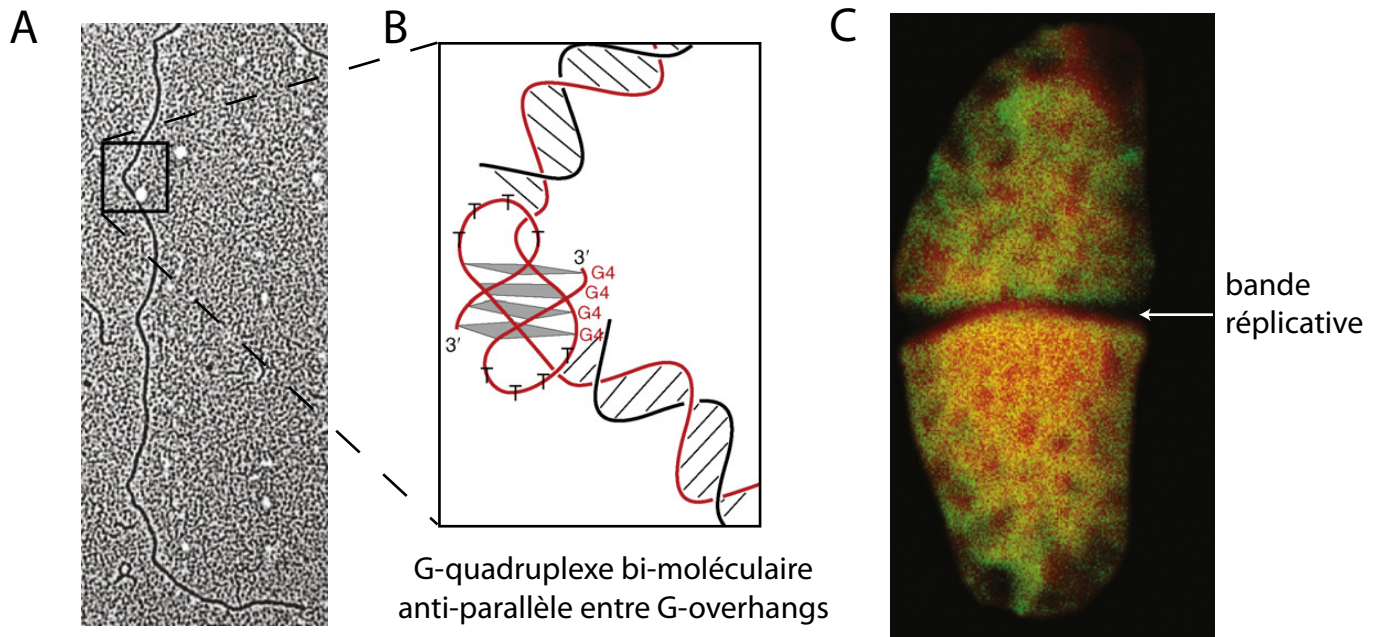
que c'est la forme anti-parallèle qui est formée *in vivo* (**Figure 13C**) (Schaffitzel et al. 2001). Seule la bande répliquative (la région où se concentre l'ADN en cours de réplication) n'est pas marquée, ce qui montre que l'anticorps révèle des G-quadruplexes sans induire leur formation *in vivo*. Ce marquage du noyau par Sty49, est perdu en cas de traitement par la nucléase S1, ce qui suggère que l'association entre chromosomes, observée par microscopie électronique, est bien due à la formation de G-quadruplexes bi-moléculaires *in vivo*.

Le formidable outil que constitue cet anticorps a permis d'étudier le rôle des protéines télomériques TEBP $\alpha$  et TEBP $\beta$  dans la formation des G-quadruplexes télomériques au cours du cycle cellulaire. TEBP $\beta$  catalyse la formation du G-quadruplexe bi-moléculaire *in vitro* et *in vivo* d'une façon dépendante de la présence de TEBP $\alpha$  (Paeschke et al. 2005). La phosphorylation de TEBP $\beta$  en phase S permet le recrutement de la télomérase et le déroulement du G-quadruplexe au niveau de la bande répliquative, restreignant ainsi l'activité télomérase à la phase S (**Figure 13D**) (Paeschke et al. 2008).

La détection directe des G-quadruplexes télomériques n'a pu être reproduit chez l'homme (ni chez aucun autre organisme), notamment en raison du faible nombre de chromosomes et de l'impossibilité de développer un anticorps spécifique d'une ou de plusieurs conformations de ce G-quadruplexe (Dai et al. 2008). Aussi, l'existence de G-quadruplexes aux télomères d'autres organismes provient d'arguments indirects (traitement par des ligands de G-quadruplexe ou inhibition d'hélicases déroulant activement les G-quadruplexes *in vitro*). Ces expériences révèlent plus un rôle « négatif » d'induction de lésions pour les G-quadruplexes lorsqu'ils ne sont pas correctement déroulés qu'une fonction « positive » de protection ou de régulation de la taille des télomères (**Chapitre I.A.4.b**). Cette dernière fonction est suggérée par l'activité inhibitrice sur la télomérase d'un G-quadruplexe en position distale de l'overhang (Zahler et al. 1991). Le rôle physiologique des G-quadruplexes aux télomères d'autres organismes reste donc à découvrir (Smith et al. 2011).

#### *I.A.3.b.ii. Régulation de l'expression génétique*

L'enrichissement des séquences G4P dans différentes régions régulatrices des gènes (**Chapitre I.A.3**) suggère un rôle des G-quadruplexes dans le contrôle de leur expression, au niveau transcriptionnel et post-transcriptionnel chez l'homme (Huppert

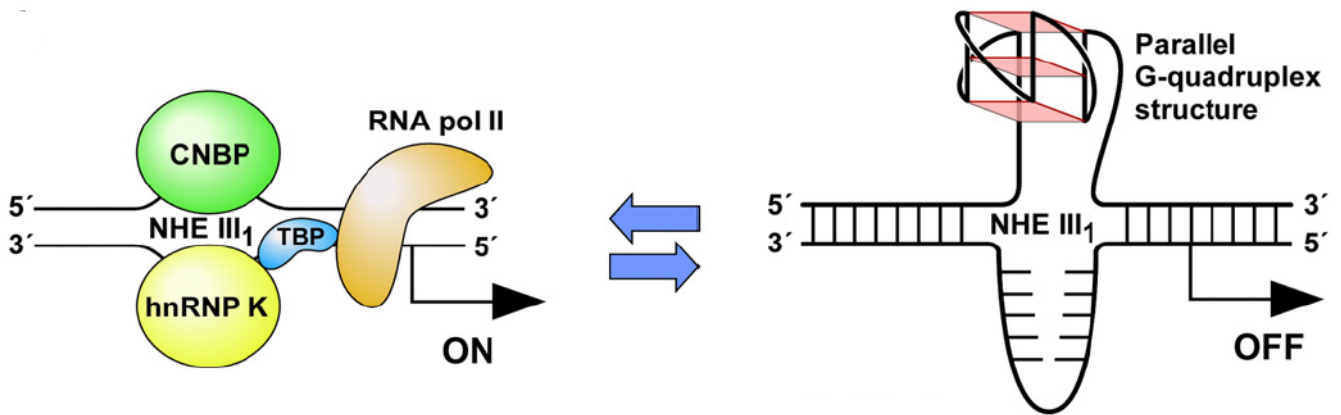


**Figure 13 : G-quadruplexes aux télomères du cilié *Stylonichia lemnae*.** (A) Visualisation en microscopie électronique des nano-chromosomes du cilié *Stylonichia lemnae* extraits du micronoyau (Lipps & Rhodes. Trends in Cell. Biol. 2009). (B) Représentation schématique du G-quadruplexe bi-moléculaire formé par les overhangs des télomères de deux nano-chromosomes (Lipps & Rhodes. Trends in Cell. Biol. 2009). (C) Immunohistochimie du macro-noyau de *Stylonichia lemnae* en présence de l'anticorps Sty49 (jaune). L'ADN est en rouge. La bande réplivative n'est pas marquée par l'anticorps Sty49 alors qu'un anticorps dirigé contre l'ADNdb allume cette bande comme le reste du noyau. L'absence de marquage n'est donc pas due à un problème d'accessibilité de l'anticorps (Lipps & Rhodes. Trends in Cell. Biol. 2009). (D) Modèle de régulation de la formation du G-quadruplexe bi-moléculaire au niveau des télomères des nano-chromosomes de *Stylonichia lemnae*. a) TEBP $\alpha$  recrute TEBP $\beta$  au niveau de l'overhang télomérique (T<sub>4</sub>G<sub>4</sub>)<sub>2</sub>. b) TEBP $\beta$  catalyse la formation du G-quadruplexe. c) En phase S, une Cdk phosphoryle TEBP $\beta$  qui se trouve délocalisé au niveau de la bande réplivative. d) TEBP $\beta$  phosphorylé recrute la télomérase et se fait déplacer par celle-ci. e) Le G-quadruplexe est déroulé. Le télomère est 40 fois répliqué et allongé par la télomérase. f) en fin de phase S, la télomérase est déplacée. Figure tirée de (Paeschke et al. Nat. Str. Mol. Biol. 2008).

and Balasubramanian 2007; Du et al. 2008; Eddy and Maizels 2008; Huppert et al. 2008; Eddy and Maizels 2009). Ma thèse portant sur les G-quadruplexes formés à partir d'ADN, dans la délicate situation de compétition avec l'ADN complémentaire, je n'aborderai pas les démonstrations des rôles post-transcriptionnels des G-quadruplexes dans le transport (Subramanian et al. 2011), la protection et la maturation (Decorsiere et al. 2011) et la régulation de la traduction (Wieland and Hartig 2007) des ARN, par ailleurs passionnants.

La présence d'au moins une séquence G4P dans le promoteur de 40% des gènes humains suggère un rôle de ces structures dans le contrôle transcriptionnel. Cet effet a été étudié par l'utilisation de systèmes rapporteurs fluorescents (luciférase ou GFP) ou par RT-qPCR chez plusieurs organismes (lignées humaines et bactérie) et avec plusieurs séquences G4P (Halder et al. 2012a). L'expression du gène rapporteur est mesurée en présence de la séquence sauvage et d'une séquence mutée incapable de former un G-quadruplexe *in vitro*, ou en présence et en absence d'un ligand stabilisateur de G-quadruplexe. Ces expériences ont pour la première fois été menées sur le G-quadruplexe du promoteur du gène c-Myc (Siddiqui-Jain et al. 2002) : la substitution d'un nucléotide crucial pour la formation du G-quadruplexe résulte en une activation transcriptionnelle d'un facteur  $\approx 3$ . À l'inverse, la stabilisation du G-quadruplexe avec le ligand TmPYP4 réprime la transcription par un facteur  $\approx 2$ . Ces résultats démontrent le rôle inhibiteur exercé sur la transcription d'un gène par un G-quadruplexe présent dans son promoteur (Siddiqui-Jain et al. 2002).

Cette inhibition transcriptionnelle par un G-quadruplexe a été montrée par mutagénèse ou interférence chimique sur d'autres gènes : TK1 (Basundra et al. 2010), c-kit (Bejugam et al. 2007), et k-ras (Cogoi and Xodo 2006). Le mécanisme par lequel s'opère l'inhibition de la transcription n'est pas connu. La délocalisation rapide (6-12 h) de trois facteurs de transcription activateurs (Sp1, CNBP et hnRNP-k) associée à une diminution de l'expression de c-Myc dans des cellules RAJI traitées par le ligand de G-quadruplexe GQC-05 suggère que la formation du G-quadruplexe supprimerait la cible duplex de facteurs de transcription activateurs (**Figure 14**) (Todd and Neidle 2008; Brown et al. 2011). Alternativement, le G-quadruplexe pourrait permettre le recrutement de protéines inhibitrices de la transcription, créant une boucle de rétrocontrôle négative simple comme l'illustre l'exemple du gène de l'insuline. En effet,

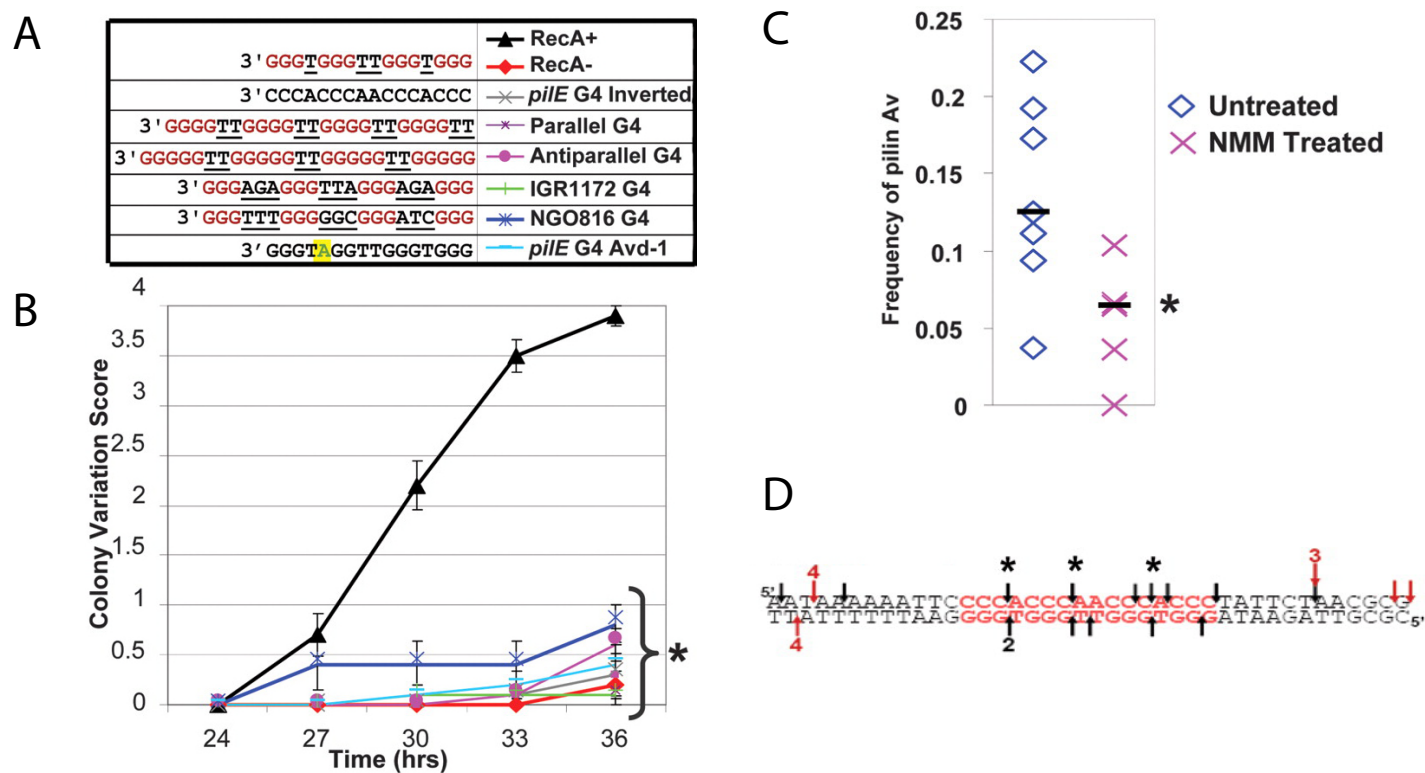


**Figure 14: Modèle d'inhibition de la transcription par le G-quadruplexe formé dans le promoteur de c-Myc.** CNBP et hnRNP-K sont deux facteurs de transcription activateurs qui lient l'ADNs, présent au niveau du site NHEIII (Nuclease Hypersensitive Element III). La formation du G-quadruplexe empêcherait la fixation de CNBP. Les auteurs émettent également l'hypothèse qu'un i-motif (structure secondaire formée par les cytosines protonées à pH acide) se formerait sur le brin complémentaire et empêcherait également la fixation d'hnRNP-K. Image tirée de (Qin & Hurley Biochimie 2008).

le minisatellite ILPR, présent entre les promoteurs des gènes de l'insuline et de l'insuline-like Growth Factor 2 (IGF-2), forme une structure G-quadruplexe particulière qui est spécifiquement reconnu par le produit de ces deux gènes. L'affinité de l'insuline et d'IGF-2 pour ce G-quadruplexe est très élevée, avec des valeurs de  $K_d$  atteignant  $10^{-13}$  M (Palumbo et al. 2008; Wang et al. 2009; Xiao et al. 2009; Xiao and McGown 2009). Des mutations ponctuelles, notamment des nucléotides des boucles qui interagissent entre elles et participent à l'originalité de cette structure, cause une diminution de l'affinité de plusieurs ordre de grandeur ( $K_d=10^{-8}$  M) (Xiao et al. 2009).

I.A.3.b.iii. *Ciblage par des nucléases et variation antigénique chez la bactérie Neisseria gonorrhoeae*

*Neisseria gonorrhoeae* est l'agent pathogène responsable de la gonorrhée. Il possède trois systèmes générateurs de diversité des antigènes exposés à sa surface qui lui permettent d'échapper à la détection par le système immunitaire. L'un de ces systèmes opère par conversion génique du gène *pilE* (qui code pour un pili exposé à la surface de la cellule) en utilisant les multiples copies transcriptionnellement éteintes et polymorphiques du même gène présentes ailleurs dans le génome. Un crible génétique a permis d'identifier une région cruciale du promoteur de *pilE* pour la variation antigénique du pili. Cette région est une séquence qui forme un G-quadruplexe *in vitro* (**Figure 15A**) (Cahoon and Seifert 2009). Son rôle fonctionnel dans la stimulation antigénique du pili a été démontré par une mutagenèse extensive du motif G4P et par interférence chimique avec un ligand de G-quadruplexe. Les mutations réalisées sont l'inversion de la séquence (inverted), deux séquences contenant chacune une mutation d'une guanine essentielle à la formation du G-quadruplexe (Avd-1 et -2), le remplacement par une séquence qui forme un G-quadruplexe anti-parallèle à 5 G-quartets (Antiparallel G4) ou un G-quadruplexe à 4 G-quartets parallèle (Parallel G4), ou l'introduction de 2 autres séquences G4P à 3 G-quartets et 3 boucles de 3 nucléotides présentes dans le génome de *N. gonorrhoeae* (IGR1172 et NG0816 G4)(**Figure 15A**). Contrairement à la séquence endogène, aucune de ces séquences, G4P ou G-mutée, ne permet la variation antigénique du pili (**Figure 15B**). Ainsi, la variation antigénique requiert spécifiquement le G-quadruplexe endogène dans la bonne orientation. Le traitement des cellules avec le ligand de G-quadruplexe NMM inhibe par un facteur 2 la fréquence de variation antigénique du pili (**Figure 15D**), ce qui suggère que la formation de la lésion



**Figure 15: Un G-quadruplexe comme cible de cassures pour la variation antigénique du pili chez *N. gonorrhoeae*.** (A) Séquences utilisées pour tester la variation antigénique du pili. Les G impliquées dans la formation d'un G-quadruplexe *in vitro* sont indiquées en rouge. (B) Variation du pili au cours du temps. Seule la séquence sauvage permet la variation du pili, d'une façon RecA-dépendante. (C) Le ligand de G-quadruplexe NMM inhibe la variation du pili. (D) Position des cassures simples et doubles-brins (\*) détectées par "break-site mapping". Les flèches noires indiquent la position des cassures dans une souche sauvage, et les flèches rouges dans une souche avec la séquence mutée Avd-1 (séquence, voir (A)). Tiré de (Cahoon & Seifert Science 2009).



recombinogène est inhibée lorsqu'un ligand est lié au G-quadruplexe. Les auteurs ont cartographié la position des cassures simple et double-brin dans et autour de la séquence G-quadruplexe (**Figure 15C**). La séquence G-quadruplexe intacte est requise pour la formation de cassure en son sein (aucune cassure n'est détectée au niveau de la séquence G4P avec la séquence mutée Avd-1) (**Figure 15C**). En résumé, le G-quadruplexe formé dans le promoteur du gène *pilE*, est la cible de cassures simples et doubles-brins dont la réparation par recombinaison conduit à la conversion du gène sur les copies polymorphiques ectopiques et silencieuses, stimulant ainsi la variation antigénique du pili. La nucléase responsable du ciblage spécifique de cette structure et de la formation des cassures simples et doubles-brins reste à déterminer.

Chez les mammifères, un rôle similaire de ciblage des cassures au niveau d'un G-quadruplexe a été proposé pour l'initiation de la recombinaison des chaînes lourdes d'immunoglobulines dans les lymphocytes, ainsi que dans la formation de translocations pathologiques entre les gènes d'immunoglobulines et le promoteur de c-Myc ou d'autres oncogènes, fréquemment trouvés dans les lymphomes (Duquette et al. 2005; Duquette et al. 2007).

### I.A.3.c. Conclusion

L'analyse bio-informatique des génomes et les validations expérimentales par mutagénèse et interférence chimique ont apporté des preuves solides de la formation des G-quadruplexes *in vivo*. Ces structures d'ADN arborent des rôles divers dans plusieurs processus biologiques essentiels, dont certains sont spécifiques de la conformation particulière d'un G-quadruplexe. C'est le cas du G-quadruplexe responsable de l'induction de lésions dans le promoteur du gène *pilE* chez *N. gonorrhoeae* dont la fonction ne peut être assurée par d'autres séquences G4P. Étant donnée la grande variété de conformations adoptées par les G-quadruplexes, cette observation incite à l'étude de la relation structure-fonction des G-quadruplexes à leurs rôles biologiques (voir notre étude **Chapitre II.E**).

## **I.A.4. Les G-quadruplexes comme cibles thérapeutiques potentielles dans le traitement des cancers**

### I.A.4.a. Inhibition de l'expression des oncogènes

L'analyse bioinformatique du génome humain a révélé un enrichissement des séquences G4P dans le promoteur des protooncogènes, contrairement aux gènes suppresseurs de tumeurs (Eddy and Maizels 2006). Les G-quadruplexes dans le promoteur de plusieurs de ces oncogènes inhibent leur expression, et cette inhibition est stimulée par le traitement avec un ligand de G-quadruplexe (**Chapitre I.A.3.b.ii**) (Cogoi and Xodo 2006; Mikami-Terao et al. 2008; Palumbo et al. 2008; Qin et al. 2010; McLuckie et al. 2011; Membrino et al. 2011). Ainsi, la stabilisation des structures G-quadruplexes apparaît comme un moyen efficace de causer une diminution globale de l'expression des oncogènes. Comme les structures tridimensionnelles des G-quadruplexes de nombreux oncogènes commencent à être disponibles, la mise au point de ligands spécifiques d'une structure particulière semble accessible, permettant ainsi de raffiner le ciblage du traitement (Brooks et al. 2010; Brown et al. 2011).

#### I.A.4.b. Perturbation du métabolisme des télomères

L'érosion progressive des télomères au cours des générations (voir **Figure 12C**), lorsqu'elle n'est pas compensée par un mécanisme d'élongation, limite la capacité proliférative des cellules et constitue un barrage à la tumorigénèse. En effet, la capacité proliférative illimitée des cellules tumorales dépend de la réactivation d'un mécanisme d'élongation des télomères qui, dans 80-85% des tumeurs, survient par surexpression de la télomérase (**Figure 12C**) (Kim et al. 1994; Palm and de Lange 2008). Ceci fait de la télomérase une cible thérapeutique privilégiée pour le traitement du cancer (De Cian et al. 2008b). La découverte de l'effet inhibiteur d'un G-quadruplexe formé dans le substrat de la télomérase sur son activité *in vitro* (Zahler et al. 1991) a permis d'envisager des stratégies d'inhibition de la télomérase indirectes, en altérant la structure de son substrat. La stabilisation d'un G-quadruplexe formée par le télomère humain par un ligand de G-quadruplexe *in vitro* inhibe l'élongation d'un oligonucléotide télomérique par la télomérase (Sun et al. 1997 ; Wang et al. 2011b). Ces expériences pionnières ont considérablement stimulé l'intérêt pour les G-quadruplexes comme cible thérapeutique et le développement de petites molécules capables de les stabiliser spécifiquement.

Ces molécules se localisent aux télomères sur chromosomes humains en métaphase (Granotier et al. 2005; Pennarun et al. 2005; Arnoult et al. 2008) et exercent des effets antiprolifératifs préférentiellement sur des cellules tumorales en culture (De Cian et al. 2008b; Gunaratnam et al. 2009). Cette cytotoxicité est corrélée à une

diminution de la taille totale des télomères (Riou et al. 2002; Burger et al. 2005; Gomez et al. 2006a; Gomez et al. 2006b; Brassart et al. 2007). Cependant, plusieurs observations suggèrent que l'effet exercé par les ligands de G-quadruplexes aux télomères soit plus l'induction des lésions que l'inhibition de leur extension par la télomérase: (i) l'entrée en sénescence induite par certains ligands (7-10 jours) est trop rapide pour être due à l'érosion progressive des télomères en absence d'un mécanisme d'élongation ( $\geq 40-50$  jours) (Riou et al. 2002; Tauchi et al. 2003; Pennarun et al. 2005; Gomez et al. 2006b; De Cian et al. 2008b); (ii) la surexpression de la télomérase ne limite ni leur cytotoxicité ni la diminution de la taille des télomères (De Cian et al. 2008b); (iii) une diminution de la taille des télomères n'est pas observée avec tous les ligands, alors qu'ils exercent des effets antiprolifératifs (Gowan et al. 2002); (iv) cette diminution de la taille des télomères a aussi lieu dans des cellules ALT, qui maintiennent leurs télomères en absence de télomérase (Riou et al. 2002; Incles et al. 2004); et (v) de nombreux signes de dommages de l'ADN sont visibles dans les cellules traitées, qu'elles soient positives ou non pour la télomérase (Riou et al. 2002; Incles et al. 2004; Gomez et al. 2006a; Salvati et al. 2007; Rizzo et al. 2009).

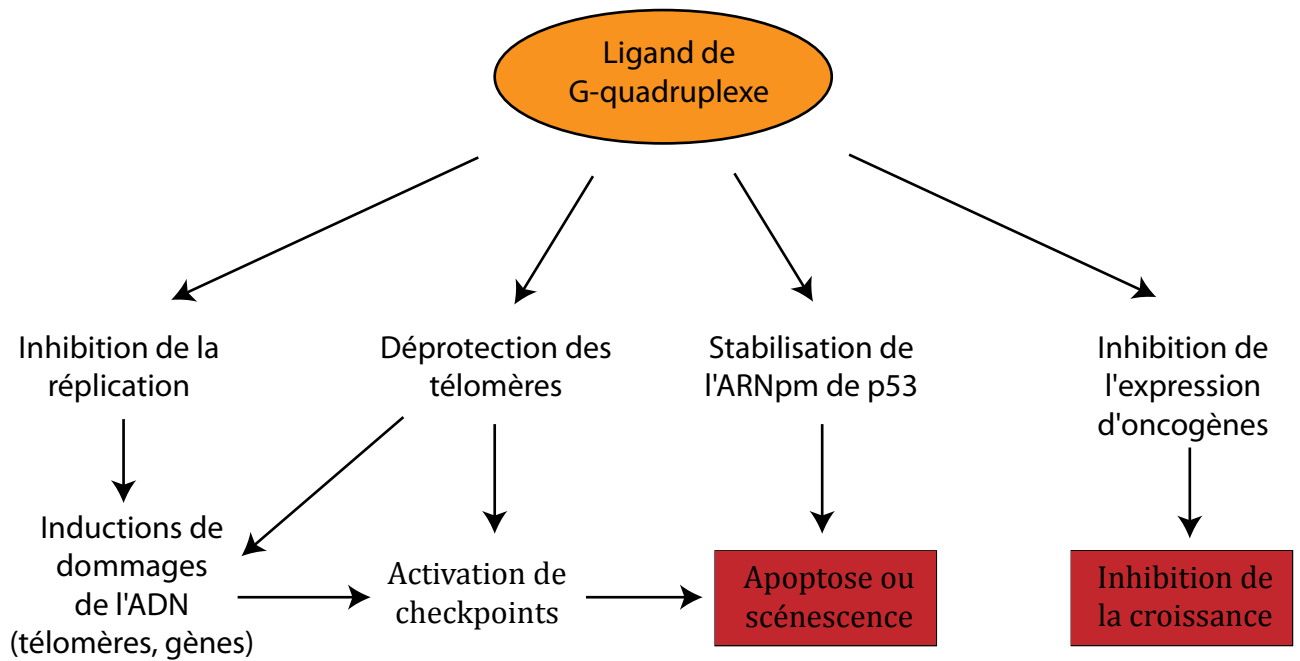
Cette induction de dommages peut avoir deux origines: la déprotection de l'extrémité télomérique et des défauts de réplication. La déprotection télomérique est suggérée par la rapide délocalisation de POT1, TRF2, TIN2 et d'hTERT des extrémités télomériques en présence de ligand, qui peut être contrecarrée par la surexpression de POT1 (Gomez et al. 2006b; Salvati et al. 2007). La formation de lésions durant la réplication est suggérée par la présence de doublets télomériques, de foyers  $\gamma$ H2AX (Gomez et al. 2006b; Phatak et al. 2007), de l'activation des kinases du checkpoint intra-S ATM et Chk1 (Tauchi et al. 2003), et d'une incorporation diminuée de BrdU aux télomères dans les cellules traitées (Rizzo et al. 2009). De plus, la sévérité de l'effet d'un ligand est corrélée avec la taille des télomères de la lignée traitée, suggérant une activité de ces ligands tout au long des télomères, et pas seulement à leur extrémité (Riou et al. 2002).

En fonction du type cellulaire, de la nature du ligand et de la concentration employée, la réponse cellulaire est soit l'entrée en sénescence, soit l'apoptose. Ces différences ne sont pas très bien comprises, étant donné l'hétérogénéité de la littérature en terme de ligands et de lignées cellulaires testées (De Cian et al. 2008b). Il n'est pas

non plus à exclure que ces ligands induisent des dommages durant la réplication ailleurs qu'aux télomères, comme révélé récemment par des expériences de ChIP-Seq anti- $\gamma$ H2AX en présence du ligand pyridostatin, au niveau du corps des gènes riches en séquences G4P (Rodriguez et al. 2012).

#### I.A.4.c. Les effets secondaires et le challenge de la spécificité

Les différents mécanismes par lesquels les ligands de G-quadruplexes sont susceptibles d'inhiber la croissance des cellules tumorales sont résumés **Figure 16**. Pourtant, étant donné le grand nombre de cibles potentielles dans le génome et leur implication dans des processus biologiques généraux, les effets toxiques sur des cellules saines semblent inévitables. Bien que les approches réductionnistes au niveau de nombreux gènes aient permis de valider rigoureusement l'effet de la formation d'un G-quadruplexe dans le promoteur d'un gène sur son expression (**Chapitre I.A.4.a**), il est frappant de noter qu'en dix ans, seulement deux études transcritomiques dans des cellules humaine ont été publiées (Mikami-Terao et al. 2008; Halder et al. 2012b). Etant donné la simplicité expérimentale de ce type d'étude, il est raisonnable de suspecter que ces expériences ont été réalisées mais n'ont pas donné de résultats facilement interprétables. Après 48h de traitement, TmPYP4 stimule l'expression de 33 gènes, et inhibe l'expression de 54 gènes diminue (Mikami-Terao et al. 2008). Les auteurs de cette première étude ne discutent pas une éventuelle corrélation avec la présence de séquences G4P dans le promoteur ou à proximité du TSS des gènes affectés. La seconde étude, publiée très récemment, concerne l'effet des bisquinoliniums Phen-DC<sub>3</sub> et 360A, affins et spécifique des G-quadruplexes, mais peu sélectifs (**Chapitre I.A.2.d.ii**) (De Cian et al. 2007a; Monchaud et al. 2008). Le traitement de cellules HeLa conduit à un changement massif de l'expression des gènes : respectivement 2212 et 526 gènes sont affectés (Halder et al. 2012b). Bien que les gènes contenant un motif G4P à proximité de leur TSS soient plus fréquemment affectés que les autres, cette étude révèle surtout l'effet pléiotropique de ces ligands, malgré leur haute spécificité envers les G-quadruplexes. Ainsi, le développement de ligands de plus en plus spécifiques d'une structure G-quadruplexe précise, par exemple en ciblant une conformation ou une séquence de boucle en plus du G-quartet externe, pourrait permettre de circonscrire l'action du ligand à un sous-groupe de G-quadruplexes *in vivo*, et ainsi limiter les effets cytotoxiques dus au ciblage de G-quadruplexes non désirés (Sissi et al. 2011). De ce



**Figure 16: Modes d'action putatifs des ligands de G-quadruplexe potentiellement responsables de l'inhibition de croissance des cellules tumorales.** La formation de dommages de l'ADN, consécutives à l'inhibition du déroulement des G-quadruplexes par des hélicases lors de la réplication (voir **Article 3**) ou aux cassures générées par la fusion de télomères déprotégés en mitose. Ces lésions ( $\gamma$ H2AX) participent à activer les checkpoints (ATM - Chk2 - p21, ou ATR - Chk1 - p53 - p21) qui dirigent la cellule vers la sénescence ou l'apoptose, dans la mesure où ces voies sont encore fonctionnelles dans la lignée cellulaire étudiée. La stimulation de la maturation de l'ARNpm de p53 en stabilisant le G-quadruplexe présent dans sa région 3'UTR participe probablement à la mise en place effective de ces voies d'apoptose et de sénescence. Par ailleurs, l'inhibition de l'expression d'oncogènes fréquemment surexprimés dans les tumeurs (c-Myc, c-Kit, hTERT, ...) sont susceptibles de limiter la prolifération des cellules.

point de vue, le développement des systèmes expérimentaux permettant de cribler la sélectivité des ligands dans un contexte cellulaire apparaît crucial (voir résultats **Chapitre II.D.3.c**).

### **I.A.5. G-quadruplexes et instabilité génomique**

#### **I.A.5.a. Blocage des polymérases par un G-quadruplexe *in vitro***

La pause spontanée d'une ADN polymérase à une séquence G4P dans le promoteur du gène de la beta-globine de poulet *in vitro* a été rapportée par Woodford *et al.* (Woodford et al. 1994). Cette pause, dépendante de la présence de potassium et levée après modification des G impliqués dans un triplet de guanines, a conduit les auteurs à proposer qu'une structure G-quadruplexe se formait dans le brin matrice et bloquait la progression de l'ADN polymérase et donc la synthèse en aval du G-quadruplexe. Cette observation, initialement réalisée avec de nombreuses polymérases ADN procaryotes, a été généralisée aux polymérases eucaryotes (Kamath-Loeb et al. 2001) ainsi qu'à des polymérases ARN et des transcriptases inverses (Tornaletti et al. 2008; Hagihara et al. 2010). Concernant les substrats, de très nombreuses autres séquences capables de former des G-quadruplexes *in vitro* (Han et al. 1999b; Broxson et al. 2011), et notamment des séquences répétées en tandem de type micro- et minisatellite (Weitzmann et al. 1997) induisent des pauses de synthèse par ces polymérases. Un G-quadruplexe est donc une matrice corrompue pour les polymérases. Cette inhibition a des implications techniques et biologiques. D'un point de vue technique, elle a été mise à profit comme test *in vitro* (« Polymerase stop assay ») de la capacité d'une séquence à former un G-quadruplexe, et de petites molécules à les stabiliser, au sein de longs ADN simples ou doubles-brins (Han et al. 1999a; Sun and Hurley 2010). Par ailleurs, ces séquences sont enclines à introduire des biais importants dans des techniques courantes comme l'amplification par PCR (Boan et al. 2004), et par conséquent dans toutes les technologies qui y font appel. Du point de vue biologique, la présence de telles séquences dans les génomes constituent des obstacles potentiellement nombreux pour les polymérases, durant la réplication, la transcription, ou la recombinaison. Étant donné la stabilité thermique de ces structures, des protéines capables de catalyser leur déroulement semblent requises.

#### **I.A.5.b. Des hélicases déroulent les G-quadruplexes *in vitro***

### *I.A.5.b.i. Généralités sur les hélicases*

Une hélicase est une translocase, c'est à dire une protéine qui se déplace activement sur un acide nucléique simple brin. Cette activité de translocation lui permet de déplacer un éventuel brin complémentaire. Une fois chargées sur une portion d'acide nucléique simple brin (généralement de plus de 6 nt), les hélicases se déplacent de manière orientée : soit dans le sens 5'-3', soit dans le sens 3'-5', une propriété qui restreint probablement le spectre de leurs substrats potentiels *in vivo* (voir discussion, **Chapitre III.D**). Des expériences récentes de molécule unique ont permis de dégager deux types d'hélicases suivant leur façon d'opérer le déroulement des acides nucléiques: des hélicases « actives » et des hélicases « passives » (ou « opportunistes »). Les hélicases « actives » séparent les deux brins d'un duplex ADN à vitesse constante, quelle que soit la composition en base ou les forces exercées sur la molécule, alors que les hélicases « passives » sont très sensibles à la nature du substrat (peu actives sur un fragment riche en GC) et aux forces exercées (très actives quand une force de séparation facilite l'ouverture du duplex) (Manosas et al. 2010). Etant donné leur stabilité, les G-quadruplexes ne sont probablement que substrat des hélicases de type « actif ». Les activités connues d'hélicases pour des substrats G-quadruplexes intra et/ou inter-moléculaires sont résumées **Table 1**.

### *I.A.5.b.ii. Les hélicases de la famille RecQ*

Les hélicases de la famille RecQ sont conservées chez les procaryotes et les eucaryotes (**Figure 17A**) et sont impliquées dans la réparation de l'ADN (exemple des protéines partenaires de Sgs1 **Figure 17B**). Elles possèdent toutes une directionnalité 3'-5', et le mode de déroulement de l'ADN double-brin, étudié avec l'hélicase bactérienne RecQ, est de type « actif » (Manosas et al. 2010).

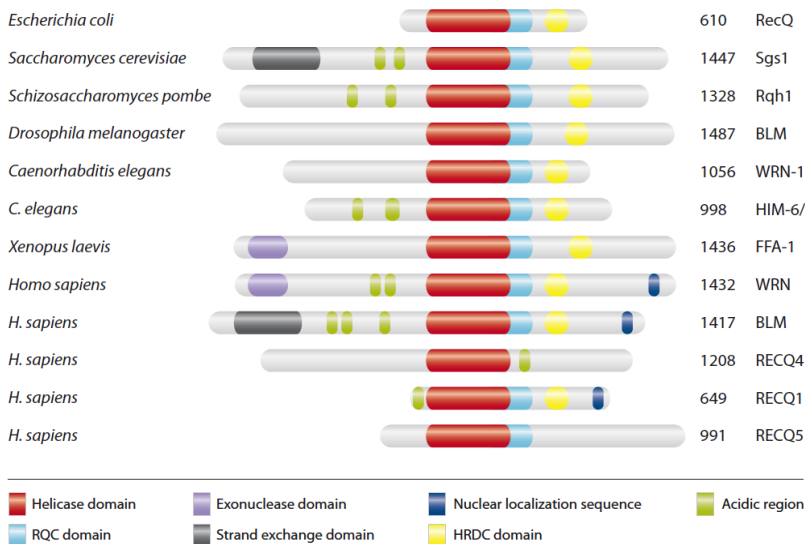
La capacité de ces hélicases à dérouler des G-quadruplexes a principalement été testée sur des substrats tétra-moléculaires très stables (**Chapitre I.A.1.b.i**) : RecQ, Sgs1, BLM, et WRN, mais pas RecQ1, déroulent ces substrats au moins dix fois plus efficacement que de l'ADN double brin *in vitro* (**Figure 17C et 17D**) (Sun et al. 1998; Sun et al. 1999; Li et al. 2001; Mohaghegh et al. 2001; Wu and Maizels 2001; Huber et al. 2002; Popuri et al. 2008). Le domaine RQC, qui jouxte le domaine hélicase, est spécifique de cette famille d'hélicase (**Figure 17A**). Il lie à haute affinité ( $K_d=3$  nM) un G-

**Table 1: Activités connues des hélicases ADN sur des substrats G-quadruplexes intra- et inter-moléculaires**

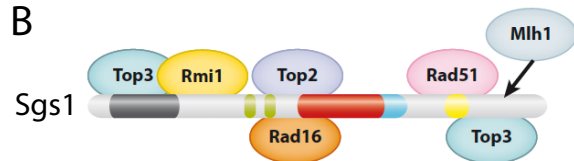
Organisme	Hélicase	Famille	Directionnalité	G-quadruplexe		Référence
				Inter-moléculaire	Intra-moléculaire	
<i>E. coli</i>	RecQ	RecQ	3'-5'	++	ND	Wu Maizels NAR 2001; Fry JBC 2001
	RecBCD	SF1	5'-3'	-	ND	
	UvrD	Rep	3'-5'	+/-	ND	Fry JBC 2001
<i>S. cerevisiae</i>	Sgs1	RecQ	3'-5'	++	ND	Han Biochemistry 2000; Huber NAR 2002; Sun Maizels 1999
	Pif1	Pif1-like	5'-3'	++	ND	Ribeyre PLoS Genetics 2009
	Dna2		5'-3'	+	ND	Masudaa-Sasa JBC 2008
<i>H. sapiens</i>	WRN	RecQ	3'-5'	++	ND	Mohaghegh NAR 2001; Fry JBC 2001; Li Biochemistry 2001
	BLM	RecQ	3'-5'	++	+	Liu JACS 2010; Sun JBC 1998; Huber NAR 2002; Huber JMB 2006; Kamath-Loeb PLoS ONE 2011; Li Biochemistry 2001; Wang NAR 2011
	RecQ1	RecQ	3'-5'	-	-	Popuri JBC 2008
	ChlR1		5'-3'	+ / ++	ND	Wu JBC 2011
	FANCI		5'-3'	++	ND	Wu MCB 2008
	Pif1	Pif1-like	5'-3'	++	ND	Sanders Biochem. J. 2010
	Dna2		5'-3'	+	ND	Masudaa-Sasa JBC 2008
G4R1	DEAH	3'-5'	+++	+++	Giri NAR 2011; Cearcy JBC 2008; Vaughn JBC 2005	



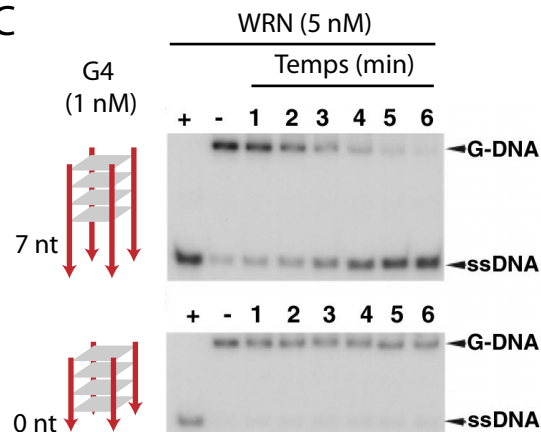
**A**



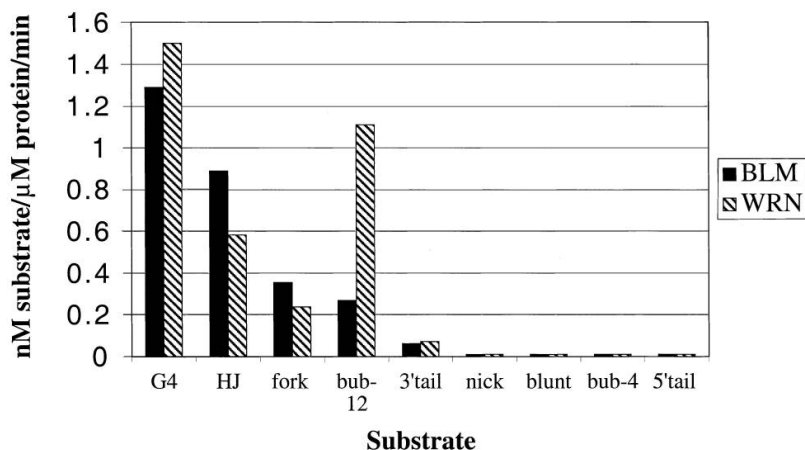
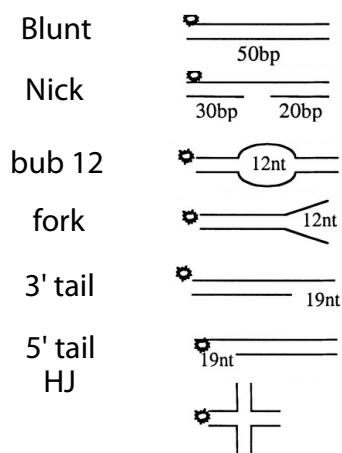
**B**



**C**



**D**



**Figure 17: Les hélicases de la famille RecQ sont conservées et déroulent les G-quadruplexes.**

(A) Structure et domaines conservés dans les hélicases de la famille RecQ. A l'exception de RECQ4, le domaine RQC, unique à cette famille d'hélicase, est conservé. Elles présentent en plus de nombreux autres domaines, comme l'HRDC. Tiré de (Bernstein et al. Ann. Rev. Genet. 2010). (B) Les domaines conservés des hélicases RecQ sont impliqués dans des interactions physiques avec de nombreuses protéines de la réparation de l'ADN, notamment de la recombinaison homologue (Rad51, Top3, Rmi1). L'exemple de Sgs1 est présenté ici. Tiré de (Bernstein et al. Ann. Rev. Genet. 2010). (C) L'hélicase WRN (5 nM) déroule un substrat G-quadruplexe tétra-moléculaire (G-DNA, 1 nM), uniquement lorsque celui-ci possède une extrémité simple-brin 3' pour permettre le chargement de l'hélicase. Tiré de (Mohaghegh et al. Nucl. Acids Res. 2001). (D) Efficacité de déroulement de différents substrats par les hélicases WRN et BLM, exprimé en nM de substrat déroulé par μM de protéine et par minute. La taille du duplex à dérouler dans tous les substrats (à l'exception du G-quadruplexe, G4) est de 50 pb. Le G-quadruplexe tétra-moléculaire est le substrat le plus efficacement déroulé par les deux protéines. Tiré de (Mohaghegh et al. Nucl. Acids Res. 2001).

quadruplexe inter-moléculaire, 3 fois supérieure à un substrat fourché, et 10 fois supérieure à une jonction de Holliday, l'autre substrat de prédilection de ces hélicases (Huber et al. 2006). Le fait que ces hélicases requièrent au moins 6 nt d'ADNsb en 3' du G-quadruplexe pour se charger et le dérouler (**Figure 17C**), suggère cependant qu'elles n'opèrent pas le déroulement du G-quadruplexe en se fixant directement à la structure. Le domaine RQC permettrait donc une reconnaissance spécifique du substrat G-quadruplexe, qui faciliterait le chargement de l'hélicase sur l'ADNsb flanquant (au moins 6 nt, **Figure 17C**) et subséquemment le déroulement du G-quadruplexe par le domaine hélicase (Mohaghegh et al. 2001; Wang et al. 2011b). J'aimerais tout de même émettre quelques réserves quant à l'efficacité préférentielle des hélicases RecQ sur les G-quadruplexes, qui repose sur la comparaison directe des paramètres cinétiques et d'affinité avec ceux obtenus avec d'autres substrats, et par des expériences de compétition avec un substrat « froid ». Ces paramètres cinétiques et d'affinités sont probablement affectés par les différences de nature entre G-quadruplexe et ADNdb : concernant les paramètres cinétiques, la distance à parcourir par l'hélicase (qui équivaut au nombre de plateaux de bases à rompre) pour dérouler l'intégralité d'un duplex est 5 à 7 fois supérieure à ce qui est requis pour dérouler un G-quadruplexe (seulement 3 ou 4 nt de long)(**Figure 17D**). Concernant l'affinité, un substrat G-quadruplexe tétramoléculaire dispose de 4 sites simple-brin permettant le chargement de l'hélicase, contre seulement un sur les autres substrats (**Figure 17C et 17D**). Ces deux différences entre les G-quadruplexes et les autres substrats conduit probablement à une surestimation systématique de l'activité de ces hélicases pour un substrat G-quadruplexe, et à la conclusion erronée d'une spécificité d'action particulière de ces hélicases pour ce type de substrat. Récemment, des expériences de molécule unique par FRET étudiant le déroulement de substrats G-quadruplexes intra-moléculaires et des substrats double-brin équivalents par BLM ont permis de définir plus rigoureusement les paramètres cinétiques de la réaction (Liu et al. 2010b). Ils relativisent les différences d'activité particulièrement élevées de cette hélicase entre G-quadruplexe et ADNdb précédemment mesurés (Liu et al. 2010b).

De manière intéressante, l'inhibition de la polymérase répliquative humaine pol $\delta$  par un G-quadruplexe intra-moléculaire formé par le microsatellite (CGG) $_n$  (Fry and Loeb 1994; Usdin and Woodford 1995) est spécifiquement levée en présence de WRN

(Kamath-Loeb et al. 2001). En revanche, WRN est incapable de lever l'inhibition de pol $\alpha$  et pol $\epsilon$  sur le même substrat, suggérant une coopération spécifique entre WRN et pol $\delta$ . L'hélicase RecQ d'*E.coli* peut se substituer à WRN. A l'inverse, UvrD est incapable de lever l'inhibition de pol $\delta$  à des concentrations qui permettent le déroulement du G-quadruplexe. Ces résultats suggèrent une interaction spécifique et une action coordonnée entre hélicases de la famille RecQ et la polymérase répliquative pol $\delta$  pour la réplication des séquences G4P (Kamath-Loeb et al. 2001).

#### I.A.5.b.iii. Autres hélicases

D'autres hélicases n'appartenant pas à la famille RecQ sont également capables de dérouler efficacement des substrats G-quadruplexe *in vitro*. Ces hélicases sont Pif1 et Dna2 chez *S. cerevisiae*, et ChlR1, FANCI, hPif1, et G4R1 chez l'homme. Un résumé qualitatif des activités de ces hélicases sur le déroulement des G-quadruplexes *in vitro* est présenté **Table 1**. L'étude de l'activité de Pif1 pour ces substrats *in vitro* a fait l'objet de deux études (J.-B. Boulé) présentées dans les **articles 1 et 2**.

Le cas de l'hélicase 3'-5' ADN et ARN G4R1 (également connue sous les noms de RHAU et DHX36) est particulièrement intéressant. Cette hélicase a initialement été identifiée par spectrométrie de masse de la fraction du lysat de cellule HeLa contenant l'essentiel de l'activité de déroulement de G-quadruplexes inter-moléculaires (Harrington et al. 1997; Vaughn et al. 2005). Cette fraction ne contient ni WRN, ni BLM, ni hPif1, ni FANCI, suggérant que ces hélicases ne sont pas responsables de la résolution des G-quadruplexes inter-moléculaires, au moins dans ces expériences sur lysat cellulaire (Vaughn et al. 2005; Creacy et al. 2008). La protéine G4R1 purifiée lie avec une très haute affinité les G-quadruplexes inter-moléculaires ADN et ARN ( $K_d = 77 \pm 6$  pM et  $39 \pm 9$  pM, respectivement) (Creacy et al. 2008), et plus encore les G-quadruplexes intra-moléculaires ADN ( $K_d = 2 - 6$  pM en fonction du G-quadruplexe testé) (Giri et al. 2011). Cette affinité de G4R1 est spécifique aux G-quadruplexes : à 300 pM, aucune fixation n'est détectée sur un oligonucléotide simple brin de même taille (Giri et al. 2011). Ces affinités sont supérieures de plusieurs ordres de grandeur à celles mesurées pour les autres hélicases humaines. Ainsi, G4R1 apparaît comme l'hélicase la plus efficace sur les G-quadruplexes intra-et inter-moléculaires *in vitro* et dans un extrait cellulaire.

#### I.A.5.b.iv. Instabilité des séquences G4P dans des mutants d'hélicases

Ces expériences *in vitro* suggèrent un rôle de déroulement des G-quadruplexes pour certaines hélicases afin d'assurer la réplication ou la transcription des séquences sous-jacentes *in vivo*. La fragilité des télomères humains en absence des hélicases WRN et BLM, et plus récemment de RTEL-1, suggère que la persistance pathologique des G-quadruplexes en absence d'hélicases spécialisées est responsable de leur propension à casser (Sfeir et al. 2009; Vannier et al. 2012). En faveur de cette hypothèse, des expériences menées sur des animaux *C. elegans* révèlent des délétions uniquement au niveau d'homopolymères de G dans le mutant *dog-1* (pour *Deletion of Guanine-rich DNA*) (Cheung et al. 2002; Youds et al. 2006; Kruisselbrink et al. 2008; Chung et al. 2011). La fréquence des délétions est nulle quand l'homopolymère est inférieur à 14 nt, et augmente avec le nombre de G dans l'homopolymère (Cheung et al. 2002; Kruisselbrink et al. 2008). Les séquences G4P sont également capables d'induire des délétions, quoique moins fréquemment : l'analyse par CGH montre que 62 des 525 homopolymères de G et 3 des 1742 séquences G4P ont subies une délétion. Cette approche n'a pas détectée de délétion au niveau d'autres régions du génome (séquences palindromiques, homopolymères A/T, séquences uniques) (Kruisselbrink et al. 2008). Ces résultats montrent que les régions riches en G et qui ont la capacité de former des G-quadruplexes subissent des réarrangement fréquents en absence de *dog-1*. Les auteurs proposent un rôle direct de *dog-1* dans le déroulement des G-quadruplexes *in vivo*. Cependant, la protéine *dog-1* n'ayant pu être purifiée, le seul support biochimique provient de la similarité de séquence entre DOG-1 et l'hélicase FANCI, qui déroule les G-quadruplexes *in vitro*. De plus, le faible nombre de séquences G4P non-homopolymère G<sub>n</sub> concernées par les délétions, et l'absence d'expériences de mutagenèse des triplets de G ou d'interférence chimique avec un ligand de G-quadruplexe ne permettent pas de conclure que le rôle de *dog-1* est bien de dérouler les G-quadruplexes du génome de *C.elegans* pour assurer la stabilité des séquences sous-jacentes *in vivo* : elle pourrait être impliquée dans la réparation fidèle des lésions induites par ailleurs par ces G-quadruplexes.

#### I.A.5.c. Autres activités de déroulement des G-quadruplexes *in vitro*

Les hélicases ne sont pas les seules protéines à promouvoir le déroulement des G-quadruplexes. La protéine télomérique humaine POT1, qui lie l'overhang télomérique riche (TTAGGG)<sub>n</sub>, promeut le déroulement du G-quadruplexe télomérique humain *in*

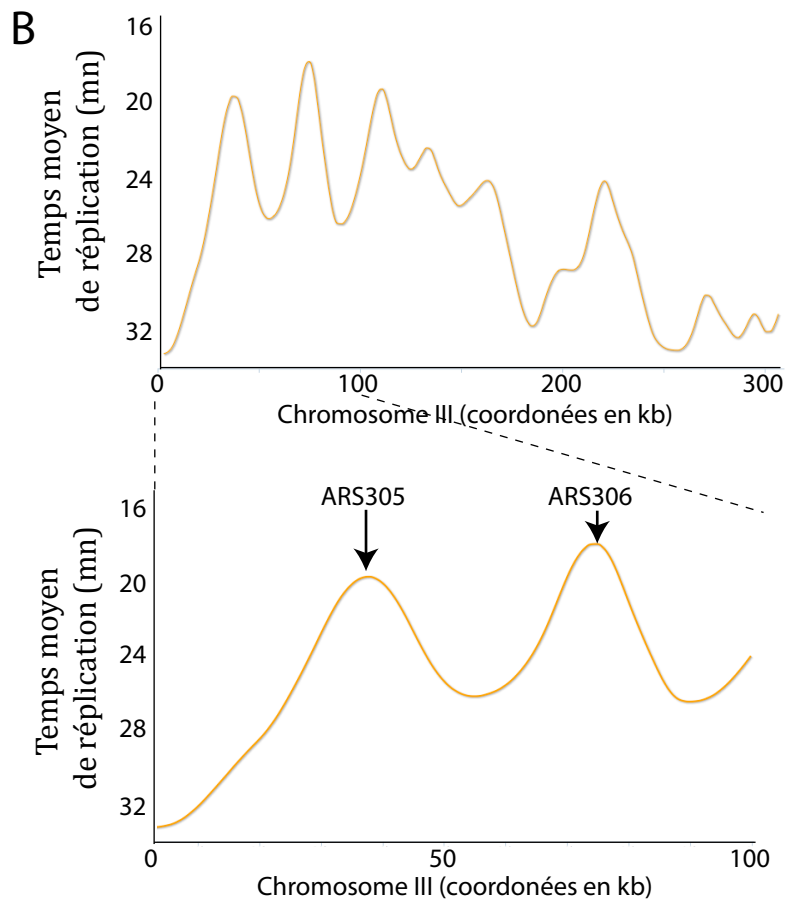
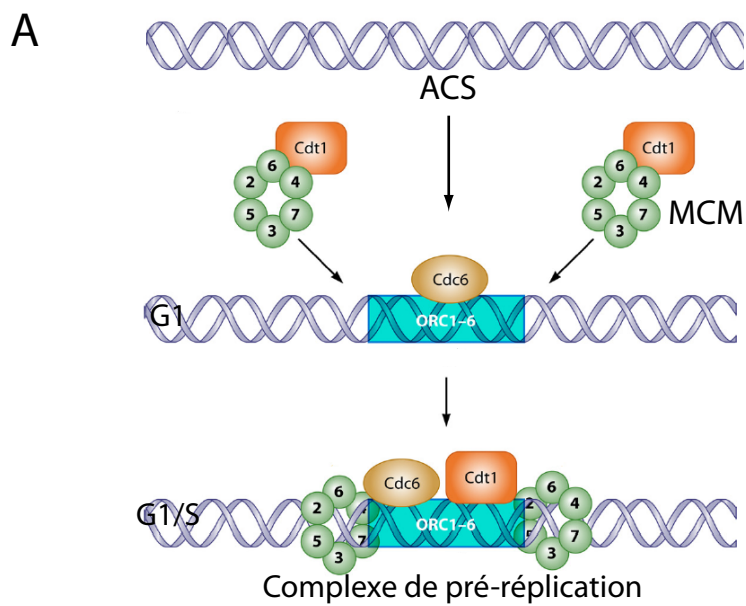
*vitro* (Zaug et al. 2005; Wang et al. 2011a). La nature de ce déroulement est sujette à controverse. Une première étude, menée en présence de la séquence G4P minimale GGG(TTAGGG)<sub>3</sub> concluait au simple déplacement de l'équilibre en piégeant la forme la forme simple-brin, l'effet de POT1 pouvant être mimé par l'ajout d'un ADNsb complémentaire du brin riche en G (Zaug et al. 2005). La seconde étude, plus récente, menée en molécule unique par AFM sur des substrats plus longs (et donc physiologiquement plus pertinents) suggère un rôle actif de déroulement des G-quadruplexes télomériques par POT1 (Wang Opresko JBC 2011). Dans cette dernière étude, POT1 déroule plus efficacement les G-quadruplexes qu'un oligonucléotide complémentaire. Les auteurs proposent que la présence d'ADNsb flanquant un G-quadruplexe dans de longs overhangs aide la fixation de POT1 et promeut le déroulement du G-quadruplexe (Wang et al. 2011a). De manière similaire à POT1, hRPA semble également capable de promouvoir le déroulement d'un G-quadruplexe télomérique : la cinétique de déroulement d'un G-quadruplexe formé par le même oligonucléotide que dans la première étude menée avec POT1 (GGG(TTAGGG)<sub>3</sub>), est  $\approx 20$  fois supérieure avec hRPA qu'avec l'oligonucléotide riche en C (Salas et al. 2006). Ainsi, contrairement à POT1, hRPA1 ne requiert pas d'ADNsb flanquant pour promouvoir le déroulement du G-quadruplexe.

## **I.B. La réplication, un mécanisme à risque pour la stabilité du génome**

### **I.B.1. La réplication chez *S. cerevisiae***

#### **I.B.1.a. Généralités**

La transmission du matériel génétique d'une génération à l'autre requiert la copie de l'intégralité du génome, qui a lieu lors de la phase S du cycle cellulaire. Chez la levure *S. cerevisiae*, la réplication est initiée à partir d'un nombre relativement bien défini d'origines, les ARS (pour « Autonomously Replicating Sequence »). Ces ARS contiennent une séquence consensus (ACS, pour « ARS Consensus Sequence ») sur lesquelles sont préalablement assemblés en G1 les complexes de pré-réplication. (**Figure 18A**). L'activation séquentielle et préférentielle de certains complexes de pré-réplication est à l'origine d'un programme temporel orchestré et robuste à l'échelle de la population cellulaire (**Figure 18B**) (Raghuraman et al. 2001; Yabuki et al. 2002; Feng et al. 2006;



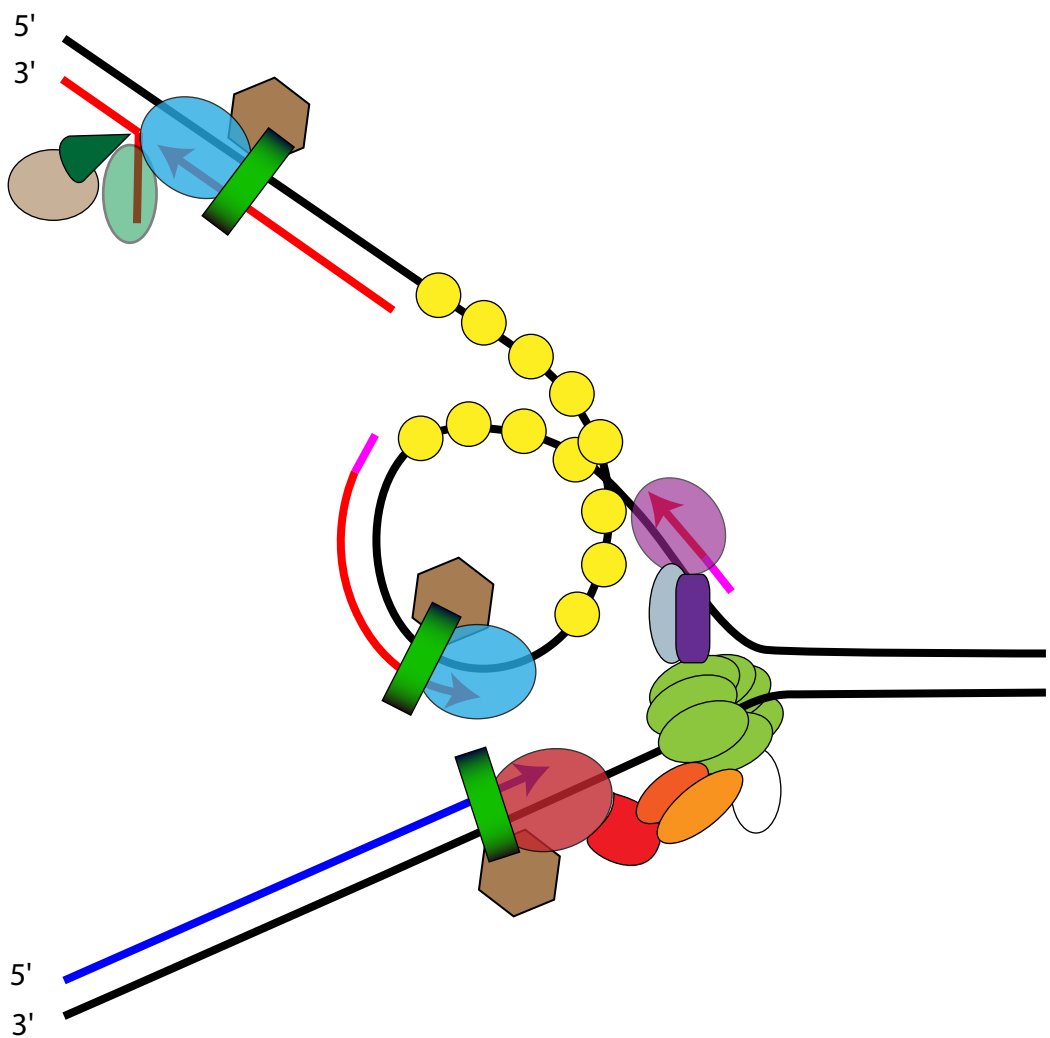
**Figure 18 : La réplication est initiée à partir de régions définies chez *S. cerevisiae*.** (A) Le complexe de pré-réplication est assemblée sur les ARS à l'issue de la mitose. Tiré de (Bochman & Schwacha, MMBR 2004). (B) Profil de réplication du chromosome III. Le temps moyen à partir duquel la région est répliquée est indiqué sur l'axe des ordonnées, obtenu à partir des données de couverture sur puces à ADN (Feng et al. Genes Cell 2002). Un détail des cent premiers kilobases du chromosome III, contenant les ARS305 et 306 (**Article 3**), est montré. Les régions répliquées précocément (sommet des pics) correspondent aux ARS actives (flèches). Graphiques obtenus avec OriDB (<http://cerevisiae.oridb.org/>).

McCune et al. 2008). La réplication procède bi-directionnellement à partir d'une origine activée. Ainsi, les régions flanquant les origines actives sont répliquées précocement et « directionnellement », alors que les régions équidistantes entre deux ARS actives, qui constituent les zones de terminaison de la réplication, sont répliquées plus tardivement et « bi-directionnellement » (**Figure 18B**) (Pavlov et al. 2002; Nick McElhinny et al. 2008; Fachinetti et al. 2010).

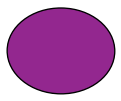
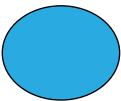


La phase d'élongation fait intervenir l'action concertée des protéines du « réplisome » (**Figure 19**), dont l'action consiste en l'ouverture de l'ADNdb parental et la copie des deux ADNsb ainsi générés. En raison du déplacement orienté 5'-3' universel des polymérases, un brin est synthétisé de manière continue (leading strand) et l'autre est synthétisé de manière discontinue (lagging strand), par fragments d'environ 200 pb (également appelés fragments d'Okazaki) subséquentement ligués entre eux. Trois machineries sont donc à l'œuvre au sein du réplisome : une machinerie de déroulement de l'ADN parental, une machinerie qui synthétise le brin continu, et une machinerie qui synthétise le brin discontinu (**Figure 19**).

La machinerie de déroulement consiste en un complexe hélicase hétéro-hexamérique, formé des protéines MCM2-7 chez les eucaryotes (Bochman and Schwacha 2009) qui enserme un des deux brins parentaux (Enemark and Joshua-Tor 2006). Chez les procaryotes ou les bactériophages, la directionnalité de l'hélicase répliquative étant 5'-3', elle est présente sur l'ADNsb matrice de la machinerie de synthèse discontinue (Marians 2008) alors que l'hélicase répliquative eucaryote, de directionnalité opposée, enserme le brin matrice de la machinerie de synthèse continue (Kaplan et al. 2003; Bochman and Schwacha 2008; Bochman and Schwacha 2009; Fu et al. 2011).






Les deux ADN simples-brins ainsi générés sont utilisés comme matrice par les ADN polymérases. Le brin continu est synthétisé par Pol $\epsilon$ , et le brin discontinu par Pol $\delta$  qui allonge de courts primers ARN/ADN générés par la Pol $\alpha$  (Nick McElhinny et al. 2008). Les mutants catalytiques de Pol $\epsilon$  ou Pol $\delta$  étant viables et accessoires pour la réplication, il est probable qu'elles puissent se substituer l'une à l'autre (Dua et al. 1999; Kesti et al. 1999). Deux Pol $\delta$  agissent probablement de concert sur le brin discontinu, comme suggéré par des expériences de microscopie à haute résolution effectuée chez la







Polymérase:

-  Pol $\alpha$
-  Pol $\delta$
-  Pol $\epsilon$
-  PCNA




Couplage:

-  Mrc1
-  Tof1
-  Csm3
-  Mcm10
-  Ctf4

Hélicase:

-  MCM2-7
-  Cdc45
- Autres:
-  RPA
-  RFC

Maturation du fragment d'Okazaki:

-  Rad27
-  Dna2
-  Lig1

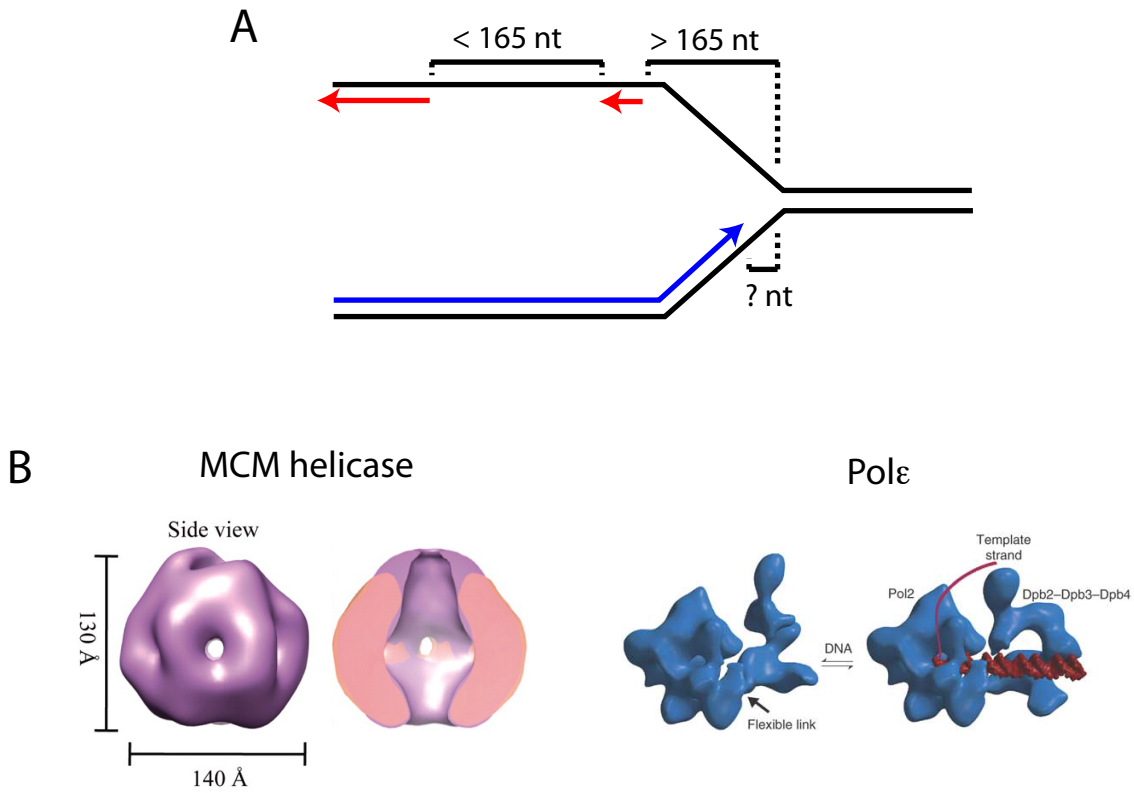
**Figure 19: Le réplisome chez *S. cerevisiae*.** Le brin synthétisé de manière discontinue est représenté en rouge. Le brin synthétisé de manière continue est représenté en bleu. PCNA est un anneau homo-trimérique de la sous-unité Pol30. RPA est un complexe hétéro-trimérique constitué des protéines Rfa1-3. De nombreuses protéines ont été omises par souci de clarté: Pol32 au niveau de Pol $\delta$ , les topo-isomérases 1 et 2 en aval et en amont de la fourche, la RNaseH et Pif1 au niveau du fragment d'Okazaki, les nucléosomes et le complexe d'assemblage de la chromatine.



bactérie *E. coli* (Georgescu et al. 2012; Lia et al. 2012). Lorsque la Pol $\delta$  rencontre le fragment d'Okazaki présent en aval, elle déplace le primer ADN/ARN, générant ainsi une structure « flap » clivée à sa base par l'endonucléase Rad27 (Liu et al. 2004) éventuellement aidée de l'hélicase et endonucléase Dna2 lorsque le flap excède quelques nucléotides (Pike et al. 2009). La réaction de ligation des deux fragments d'Okazaki est réalisée par Lig1 (Levin et al. 1997).

#### I.B.1.b. ADN simple-brin à la fourche

Les molécules d'ADNsb parental générées par l'hélicase répliquative sont engagées par deux machineries différentes, qui ne requièrent pas la présence de la même quantité d'ADNsb. La taille du fragment d'Okazaki, a récemment été réévaluée à  $\approx 165$  nt après clivage du flap généré lorsque la Pol $\delta$  rencontre le primer ARN/ADN du fragment d'Okazaki en aval, au maximum de 30 nt (Smith and Whitehouse 2012). C'est donc au moins 165 nt d'ADNsb qui doivent être présents au niveau de la matrice du brin à synthèse discontinue chez *S. cerevisiae* (**Figure 20A**). Conformément à cette prédiction, la taille de l'ADNsb présent seulement sur l'un des brins à la fourche est estimée à  $\approx 200$  nt par microscopie électronique (Sogo et al. 2002; Lopes et al. 2006b). La taille de l'ADNsb généré pour la machinerie de synthèse continue est considéré comme minime, bien qu'elle n'ait pu être précisément établie par microscopie électronique (Sogo et al. 2002; Lopes et al. 2006b). En effet, en plus de l'a priori que le simple-brin présent sur ces intermédiaires de réplication est systématiquement présent sur le brin à synthèse discontinue, en absence d'une synthèse d'ADN jusqu'au site de déroulement, rien n'empêche l'ADN parental simple-brin de se réapparier une fois les protéines présentes à la fourche digérées durant le protocole d'extraction. Même si la distance entre la polymérase et le complexe MCM2-7 n'est pas connu, les structures à basse résolution déterminées pour Pol $\epsilon$  et le complexe MCM arché suggère qu'une taille substantielle d'ADNsb est présente entre le site de déroulement du duplex parental et le site de synthèse, au moins au sein de ces protéines (Figure 24B) (Asturias et al. 2006; Sanchez-Berrondo et al. 2012). L'existence éventuelle d'une quantité non-négligeable d'ADNsb au niveau de la machinerie de synthèse continue a des implications concernant le lieu de formation des structures secondaires de l'ADN lors de la réplication (**Chapitre III.A.1**) (Mirkin and Mirkin 2007; Richard et al. 2008).



**Figure 20: ADN simple brin à la fourche de réplication.** (A) Représentation schématique d'une fourche de réplication et de la taille des régions d'ADNsb attendue sur chaque brin. (B) Structure du complexe MCM de *Bacillus cereus* déterminée à partir de données de cryo-microscopie électronique (Sanchez-Berrondo *et al.* Nucl. Acids Res. 2012). Ce complexe MCM présente de fortes similarités de séquences avec le complexe MCM2-7 eucaryote. A droite, la structure de Polε de *S. cerevisiae* déterminée par cryo-miscroscopie électronique et modélisée en cours de synthèse (Asturias *et al.* Nat. Str. Mol. Biol. 2006). L'ADNsb matrice parvient à la polymérase par un canal.

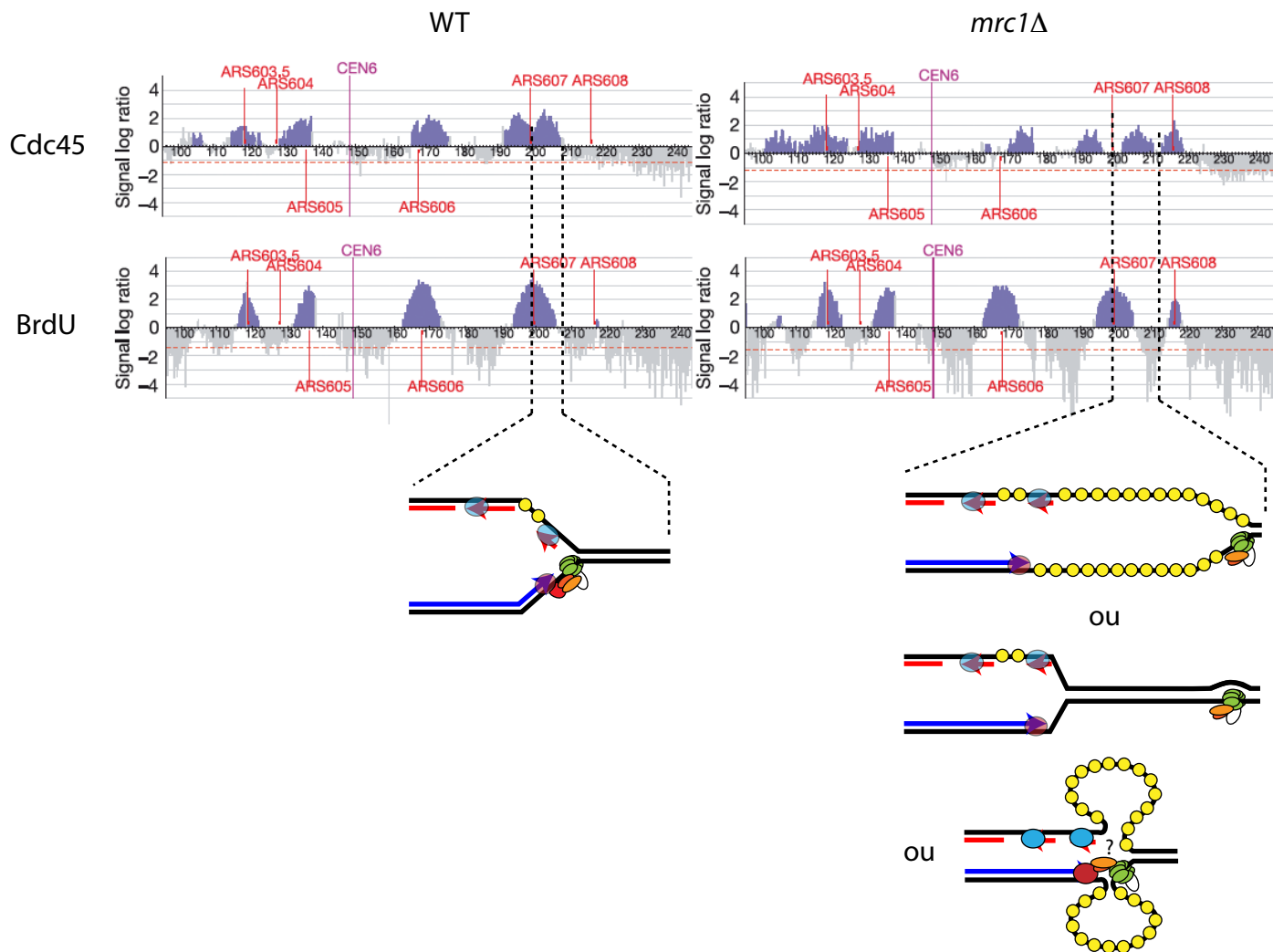
### I.B.1.c. Quels couplages à la fourche ?

#### *I.B.1.c.i. Couplage physique et fonctionnel entre les machineries de déroulement et de synthèse*

L'existence de couplages, physiques ou fonctionnels, entre hélicase répllicative et polymérase a été rapportée. Concernant le couplage fonctionnel, des expériences de répllication en molécule unique sur la machinerie de répllication du bactériophage T4 révèlent que l'hélicase répllicative seule est « passive » (**Chapitre I.A.5.b.i**) et a tendance à « glisser » fréquemment dans le sens opposé de son moteur moléculaire conduisant à des vitesses moyennes de déroulement faible ( $\approx 100$  nt/s). En absence d'hélicase, la polymérase recule et dégrade l'ADN néosynthétisé. En revanche, l'action simultanée de l'hélicase et de la polymérase conduit à une synthèse rapide (300 nt/s) et processive : l'hélicase lève la contrainte exercée par le duplex sur la polymérase tandis que la présence de la polymérase empêche le glissement de l'hélicase. De plus, la polymérase déstabilise les premières bases du duplex, conférant au duo polymérase-hélicase une activité de déroulement « active » (Manosas et al. 2012). Cette action ne requiert pas d'association physique entre les deux enzymes (Delagoutte and von Hippel 2001; Ishmael et al. 2003). Ces expériences révèlent donc un couplage fonctionnel entre hélicase et polymérase à la fourche pour le déroulement et la synthèse rapide et processive du brin continu.

Un couplage physique existe entre l'hélicase MCM2-7 et la Pol $\epsilon$ . Il est réalisé par le complexe hétérotrimérique formé par les protéines Mrc1, Tof1, et Csm3 (**Figures 19 et 21**) (Nedelcheva et al. 2005; Lou et al. 2008; Bando et al. 2009) qui voyagent avec la fourche de répllication (Katou et al. 2003; Bando et al. 2009). Concernant la machinerie de synthèse discontinue, Ctf4 et Mcm10 réalisent le couplage physique entre Pol $\alpha$  et l'hélicase répllicative (**Figure 19**) (Gambus et al. 2009; Lee et al. 2010).

L'absence de Mrc1 n'affecte pas l'association de Tof1 et Csm3 entre eux et avec la fourche de répllication, mais Tof1 et Csm3 sont requis pour s'associer à Mrc1 (Bando JBC 2009). Mrc1 interagit physiquement avec Pol $\epsilon$  (Lou et al. 2008), et le complexe interagit physiquement avec plusieurs sous-unités de l'hélicase MCM2-7 et Cdc45 (Katou et al. 2003; Nedelcheva et al. 2005; Bando et al. 2009). Le complexe entier est requis pour maintenir à proximité sur la molécule d'ADN la machinerie d'ouverture du duplex et les



**Figure 21: Mrc1 participe au couplage entre synthèse et déroulement du duplex en cas d'inhibition généralisée de la synthèse d'ADN.** ChIP-chip sur des cellules synchronisée 60' après l'entrée en phase S. Les cellules sont cultivées en présence de 200 mM d'hydroxyurée (HU) et de BrdU. L'anticorps utilisé est indiqué à gauche: le lieu de synthèse est révélé par l'anticorps anti-BrdU, et la position de l'hélicase est révélée avec l'anticorps anti-Cdc45. L'expérience est réalisée dans des cellules WT, et *mrc1Δ*. En absence de Mrc1, Cdc45 ne colocalise plus avec le site de synthèse. Légendes de la fourche, voir Figure 19.

machineries de synthèse en présence d'inhibiteur de polymérisation : les protéines de déroulement et de synthèse sont découplées dans des cellules déficientes pour l'une de ces protéines lorsqu'elles sont traitées à l'hydroxyurée (HU) (**Figure 21**) (Katou et al. 2003; Bando et al. 2009). Le mécanisme conduisant à l'inhibition de l'activité du complexe MCM2-7 par le complexe Mrc1-Tof1-Csm3 n'est pas connu (Nedelcheva-Velva et al. 2006; Bochman and Schwacha 2009), mais son maintien à une fourche bloquée et le couplage hélicase-polymérase dépendent de l'activation du checkpoint intra-S (Rad53) (Lopes et al. 2001; Sogo et al. 2002; Cobb et al. 2004). En plus de son rôle de couplage, Mrc1 est un des médiateurs de l'activation du checkpoint intra-S en cas d'accumulation d'ADNsb à la fourche, quasi-indépendamment de Tof1 et Csm3 (Osborn and Elledge 2003; Tourriere et al. 2005; Crabbe et al. 2010). Ainsi, en cas d'inhibition des machineries de synthèse, Mrc1 participe à l'activation du checkpoint qui en retour stabilise le réplisome et promeut le couplage MCM2-7-polymérase via Mrc1-Tof1-Csm3 (**Figure 21**).

Ces données concernant les couplages entre hélicase et polymérase ont été obtenus dans des conditions d'inhibition généralisée de la synthèse (HU), qui conduisent à l'activation du checkpoint intra-S (phosphorylation de Rad53). Qu'en est-il d'une inhibition ponctuelle d'une polymérase à une seule fourche, par exemple sur une matrice endommagée ? A ma connaissance, de telles données concernant l'existence d'un couplage au niveau d'une fourche individuelle suite à l'inhibition de l'une des machineries de synthèse n'est pas connu. Le couplage entre hélicase et polymérase requérant un checkpoint actif (Sogo et al. 2002), et le checkpoint intra-S nécessitant l'accumulation d'ADNsb à de nombreuses fourches (Shimada et al. 2002; Tercero et al. 2003; Byun et al. 2005), il est probable que le couplage entre hélicase répliquative et machinerie de synthèse ne soit pas effectif en cas d'inhibition de la synthèse à une fourche unique (voir discussion **Chapitre III.A.2**).

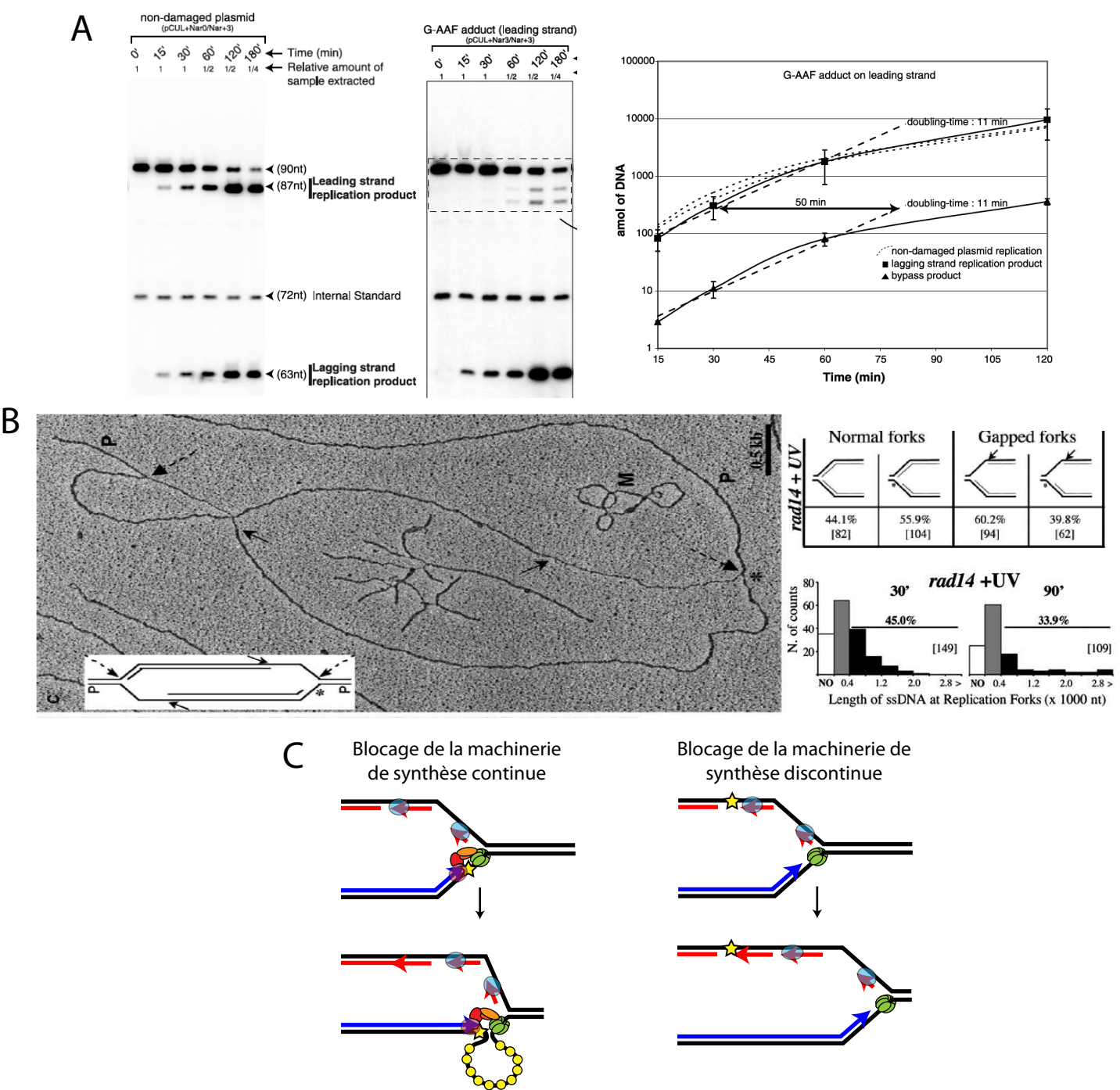
*I.B.1.c.ii. Absence de couplage entre machineries de synthèses continue et discontinue*

Si les polymérase des machineries continue et discontinue sont associées à l'hélicase répliquative, est-il besoin d'un mécanisme supplémentaire de couplage entre les deux machineries de synthèse? La question se pose si les deux machineries ne synthétisent pas l'ADN à la même vitesse, ce qui causerait une accumulation de l'ADN

simple brin parental au niveau de la machinerie la plus lente. Des études biochimiques sur fourches bactériennes reconstituées montrent qu'*in vitro* la vitesse de synthèse est la même entre machinerie à synthèse continue et discontinue (McInerney and O'Donnell 2004). Ceci est également vrai *in vivo* : la synthèse des brins continus et discontinus se fait à vitesse égale (Pages and Fuchs 2003). De manière intéressante, l'introduction d'un nucléotide modifié sur la matrice du brin à synthèse continue inhibe la synthèse sur ce brin, mais pas la synthèse discontinue. Ce résultat démontre qu'aucun mécanisme de couplage entre les deux machineries de synthèse n'existe chez la bactérie (**Figure 22A**) (Pages and Fuchs 2003). De plus, l'introduction de cette lésion sur la matrice opposée n'affecte la progression d'aucune des deux machineries, indiquant l'existence d'un mécanisme de tolérance à la lésion spécifique à la machinerie de synthèse discontinue (Pages and Fuchs 2003).

Cette absence de couplage semble conservée chez *S. cerevisiae*. En effet, l'analyse par microscopie électronique d'ADN extrait de levures incapables de réparer les dimères de thymine (souche *rad14Δ* traitée aux UV) révèle la formation de l'accumulation de longs ADNsb sur un des deux brins de la fourche (supposément sur le brin à synthèse continue)(**Figure 22B**) (Lopes et al. 2006b). Ceci suggère que l'arrêt de synthèse sur un brin n'affecte pas la synthèse sur le brin opposé et que l'arrêt d'une machinerie n'est pas suffisante pour arrêter la progression de l'hélicase, sur des distances qui peuvent atteindre 3 kb (**Figure 22B**) (Lopes et al. 2006b). L'absence d'ADNsb sur le brin à synthèse discontinue suggère quant à lui que, si une polymérase se trouve bloquée sur une lésion, la réplication peut continuer par réinitiation de la synthèse sur le fragment d'Okazaki suivant. Les ADN polymérases translésionnelles (TLS) se chargeront de contourner la lésion et permettre la reprise de la synthèse en aval pour finir la réplication des petits fragments d'ADNsb qui persistent entre fragments d'Okazaki, en phase S ou en G2 (McInerney and O'Donnell 2004; Lopes et al. 2006b). Cette réplication découplée de la fourche de réplication a lieu lors d'un processus appelé post-replicative repair (PRR), détaillé au (**Chapitre I.B.3.b.ii**). En absence des polymérases TLS, ces fragments simples-brins internes persistent, spécifiquement sur le brin à synthèse discontinue (**Figure 22B**) (Lopes et al. 2006b).

Ainsi, des découplages entre machineries de synthèses continue et discontinue sont observés chez la bactérie et *S. cerevisiae*: une polymérase bloquée, principalement



**Figure 22: Absence de couplage des deux machineries de synthèse, chez les bactéries et *S. cerevisiae*.** (A) La réplication chez *E. coli* procède à vitesse égale sur les deux brins, mais l'ajout d'une base modifiée sur la matrice de la machinerie de synthèse continue bloque la réplication sur ce brin, mais pas sur la machinerie opposée: à gauche, produits de réplication des machineries de synthèse continue (haut) et discontinue (bas) d'un plasmide non-modifié. Au centre, produits de réplication d'un plasmide contenant une base modifiée sur la matrice du brin continue: la synthèse de ce brin est bloquée et des produits de contournement sont observés (encadrés). A droite, cinétique de la synthèse des brins continus et discontinus sur plasmide intact (pointillé) ou modifié (pleine). Tiré de (Pagès & Fuchs, Science, 2003). (B) Accumulation d'ADNs asymétrique à la fourche dans une souche de *S. cerevisiae* déficiente pour la réparation des dimères de T, après traitement aux UV. La nature du brin affecté (qui s'accumule sous forme simple brin) est déduite grâce à la présence à la fourche d'une petite région simple brin caractéristique des fragments d'Okazaki. Tiré de (Lopes et al. Mol. Cell. 2006). (C) Modèle hypothétique à même d'expliquer cette accumulation d'ADNs spécifiquement sur la matrice du brin à synthèse continue: un blocage sur ce brin n'affecte pas la progression de la machinerie de synthèse discontinue. Un blocage d'une polymérase de la machinerie de synthèse discontinue, en raison de sa capacité intrinsèque de ré-initiation, n'entraîne pas l'accumulation d'ADNs, ni le ralentissement d'aucune machinerie de synthèse.

celle en charge du brin à synthèse continue, voit s'accumuler entre elle et l'hélicase des fragments d'ADNs<sub>b</sub> qui peuvent mesurer plusieurs kilobases (**Figure 22C**). Cette absence de couplage a des implications concernant l'instabilité des séquences capables de former des structures secondaires, notamment des séquences capables de former des G-quadruplexes que j'ai étudiées au cours de ma thèse, et qui seront discutées au **Chapitre III.A**.

## **I.B.2. Les séquences capables de former des structures secondaires de l'ADN sont « à risque » lors de la réplication**

De nombreux agents physico-chimiques, souvent classés suivant leur origine exogène ou endogène, endommagent l'ADN (Aguilera and Gomez-Gonzalez 2008). Alors que la plupart de ces agents endommageants agissent en *trans*, des régions du génome induisent des lésions en *cis*. En effet, certaines séquences peuvent conduire au repliement intramoléculaire de l'ADN via l'appariement (canonique ou non) de ses nucléotides. Le G-quadruplexe est une de ces « structures secondaires » (**Chapitre I.A**). Les séquences répétées inversées (palindromiques), même séparées par quelques nucléotides, peuvent former des structures de type « épingle à cheveux » (hairpin) ou cruciformes (Pearson et al. 1998). Les séquences riches en purines ou pyrimidines forment des triples hélices (triplex)(Mariappan et al. 1999). Ces structures secondaires sont un obstacle à la progression des ADN polymérases *in vitro* (**Chapitre I.A.5.a**) (Baran et al. 1991; Ohshima et al. 1996; Gacy et al. 1998), et de nombreux indices indiquent qu'ils le sont probablement *in vivo*. En effet, les répétitions microsatellites capables de former des « hairpins » ou des cruciformes, tels que CGG/GCC, CAG/CTG (Leach 1994; Pearson et al. 1998), ou des triplex, tel que GAA/CCT (Mariappan et al. 1999), sont instables dans plusieurs organismes modèles et présents au niveau de sites fragiles dans les cellules humaines (Krasilnikova and Mirkin 2004; Pollard et al. 2004; Mirkin and Mirkin 2007; Kim et al. 2008; Richard et al. 2008; Kerrest et al. 2009). Les séquences palindromiques induisent des événements de recombinaison entre séquences flanquantes (Gordenin et al. 1993; Lobachev et al. 1998). Ainsi, la capacité de ces structures à inhiber la progression des ADN polymérases *in vitro* corrèle avec leur instabilité (formation d'expansion/contraction) ou leur capacité à induire des réarrangements des régions flanquantes *in vivo*. De plus, ces séquences induisent généralement des pauses de réplication, visibles par gel bidimensionnel (lecture d'un gel



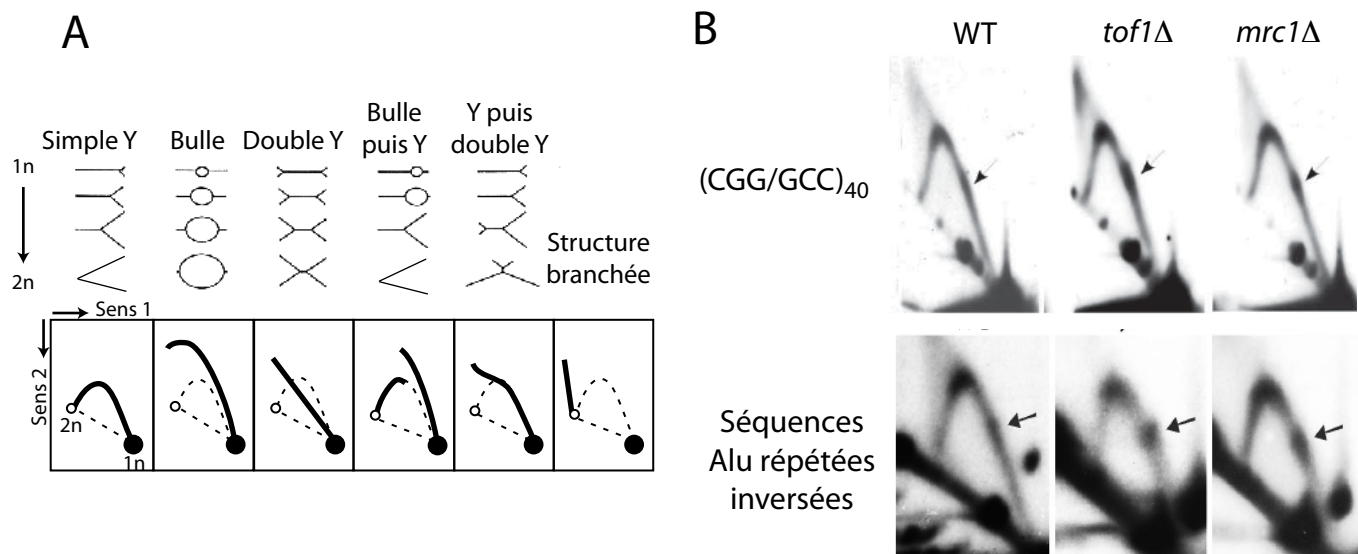
bidimensionnel, **Figure 23A**) au cours de la phase S (**Figure 23B**) (Krasilnikova and Mirkin 2004; Zhang and Freudenreich 2007; Voineagu et al. 2008; Kerrest et al. 2009; Voineagu et al. 2009; Liu et al. 2010a; Anand et al. 2012). Ces réarrangements, ainsi que la formation de structures branchées visibles en gel 2D, requièrent généralement l'action des protéines de la recombinaison homologue (**Chapitre I.B.4.a, Figures 26 et 28**) (Freudenreich et al. 1998; Kerrest et al. 2009). Ces observations, faites avec plusieurs séquences et chez plusieurs organismes, suggèrent que ces structures causent un défaut de réplication, et que ce défaut de réplication conduit à leur instabilité somatique par recombinaison homologue. Ces pauses de réplifications et cette instabilité peuvent être exacerbées dans certains mutants (exemples de *mrc1Δ* et *tof1Δ* **Figure 23B**), suggérant que des mécanismes à même de prévenir la formation de ces structures, de les dérouler une fois formées, et/ou de redémarrer la réplication en cas de blocage prolongé de la polymérase sur une structure persistante, sont requis pour assurer la réplication fidèle de ces séquences.

### **I.B.3. Mécanismes qui limitent l'instabilité génomique**

Une augmentation de l'instabilité génomique peut résulter d'une augmentation du taux des lésions de l'ADN et/ou d'un défaut dans la réparation fidèle de ces lésions. La prévention de la formation des structures secondaires ou leur déroulement en cas de formation est un premier rempart à la formation de lésions. Si toutefois une structure se forme et ne peut être résolue, d'autres mécanismes sont déployés pour résoudre le blocage de la synthèse qui en résulte.

#### **I.B.3.a. Prévention de la formation des structures secondaires**

De l'ADNsb est obligatoirement présent sur la matrice du brin à synthèse discontinue (**Chapitre I.B.1.b**). En effet, la Pol $\alpha$  qui initie la synthèse du fragment d'Okazaki crée un primer ADN/ARN environ tous les 165 nt, et de ce fait requiert la présence d'un ADN simple-brin d'au moins 165 nt (Smith and Whitehouse 2012). Une diminution de la concentration intra-cellulaire de Pol $\alpha$  conduit à une augmentation de l'instabilité génomique préférentiellement au niveau d'une grande séquence palindromique, supposément associée à une augmentation de la taille des fragments d'Okazaki, et donc la taille de l'ADN simple brin généré entre eux (Lemoine et al. 2005). De manière similaire, des mutations ponctuelles de Pol $\alpha$  causent une augmentation des



**Figure 23: Détection de pause de réplication par gel bidimensionnel : exemple des séquences capables de former des structures secondaires chez *S. cerevisiae*.** (A) Lecture d'un gel 2D. L'ADN est cross-linké, migré à voltage faible dans un gel d'agarose peu concentré (séparation des molécules par leur taille), puis à fort voltage dans un agarose concentré. La première migration sépare les molécules principalement selon leur taille. La deuxième migration sépare les molécules selon leur taille et le degré de leur branchement: plus la molécule sera branchée, plus elle sera retardée. Les profils possibles sont représentés au-dessous des structures. (B) Exemple de pauses de réplication visibles le long d'un arc Y, au niveau d'un microsatellite (CGG/GCC)<sub>40</sub> ou de séquences Alu répétées inversées, dans des souches de levure WT ou mutantes pour *TOF1* ou *MRC1* (Voineagu *et al.* PNAS 2008, Nat. Str. Mol. Biol. 2009). Cette pause se traduit par la formation d'un spot le long de l'arc Y, qui correspond à une espèce fourchée dont la durée de vie augmente relativement à la durée de vie des autres structures fourchées présentes tout au long de l'arc. L'intensité du spot, normalisé sur l'intensité du reste de l'arc Y, permet de réaliser une étude comparative de l'intensité de la pause entre différentes souches. Cependant, l'intensité de cette pause étant fonction du nombre de molécules présentant une pause à ce site, et de la durée de la pause, il convient de rester prudent quant à l'interprétation de la pause: deux spots de même intensité peuvent présenter des fréquences et des durées de pause très différentes. Ainsi, il n'est pas possible de savoir si les mutants *tof1*Δ et *mrc1*Δ causent une augmentation de la fréquence de pause, ou si pour un nombre équivalent de fourches bloquées, la durée de la pause augmente.

réarrangements entre séquences répétées inversées, qu'elles soient longues (1.5 kb) (Gordenin et al. 1993) ou courtes (30 et 80 pb) (Ruskin and Fink 1993). Cette situation d'accumulation d'ADN simple-brin peut également être obtenue par inhibition des polymérase, soit en limitant la disponibilité des précurseurs nucléotidiques (dNTP), soit en inhibant la réaction de polymérisation. La diminution de la concentration de dNTP peut être obtenue transitoirement par traitement des cellules avec l'hydroxyurée (HU), un inhibiteur de la ribonucléotide réductase (RNR) qui catalyse la réaction NDP → dNDP. L'accumulation de simple-brin aux fourches dans des cellules traitées à l'HU (Feng et al. 2006) est associée à une augmentation de l'instabilité de microsatellites formant des structures secondaires de type « hairpin » (Freudenreich et al. 1998 ; Yang et al. 2003). Ces résultats suggèrent que la taille de l'ADNsb normalement présent à la fourche limite la formation des structures secondaires. De manière complémentaire, le complexe hétéro-trimérique RPA (Rfa1-3) qui lie à très haute affinité l'ADNsb est en compétition avec la formation de structures secondaires et déplace l'équilibre vers la forme simple brin (**Chapitre I.A.5.c**). A ces mécanismes de prévention s'ajoutent les mécanismes de déroulement actifs de ces structures par les hélicases (**Chapitre I.A.5.b**) (Dhar and Lahue 2008; Anand et al. 2012).

#### I.B.3.b. Réparation des fourches bloquées ou cassées

Si ces mécanismes de prévention et de déroulement sont insuffisants, des mécanismes spécialisés sont mis en jeu pour redémarrer de la fourche ou la polymérase bloquée. Une fourche bloquée peut également être cassée, laissant une cassure asymétrique (une seule extrémité 3') dont la réparation restaure une fourche de réplication fonctionnelle.

Les mécanismes impliqués dans le redémarrage ou la réparation d'une fourche de réplication compromise, leurs modalités d'action, et leur fidélité vont faire l'objet de cette partie. Il faut néanmoins noter que dans la grande majorité des études citées, le bloc ou la lésion initiale sont rarement des structures secondaires, mais des dommages au sein de la matrice (site abasique, dimère de thymine, etc...) ou des blocages physiques de l'hélicase répliquative par un complexe protéine-ADN (Ivessa et al. 2002; Ivessa et al. 2003; Lambert et al. 2010; Dubarry et al. 2011). Cependant, l'inhibition des ADN polymérase répliquatives en présence d'un substrat impropre à la copie, quelle que soit sa nature, suggère une certaine unité dans la réponse cellulaire à ces événements de

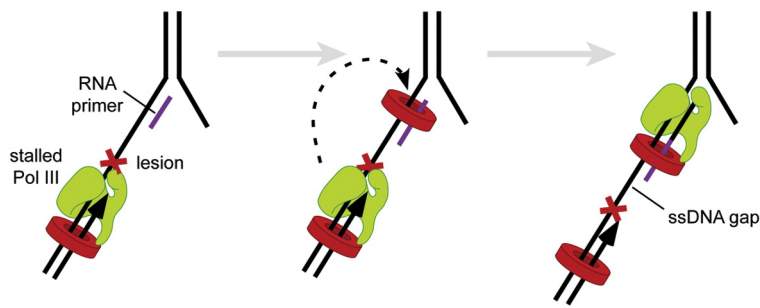
blocage : c'est moins la multitude des couples « lésion particulière-polymérase bloquée » que la conséquence du blocage de la polymérase (accumulation d'ADNsb, cassure asymétrique) qui est détectée par la cellule (Walter et al. 1998; Zou and Elledge 2003; Zou et al. 2003; Byun et al. 2005). Ainsi, les mécanismes de réponse et de résolution d'un blocage de synthèse décrits en présence d'une altération chimique de la matrice sont probablement les mêmes qu'en présence d'une structure secondaire.

#### *1.B.3.b.i. Ré-initiation de la synthèse en aval de la lésion*

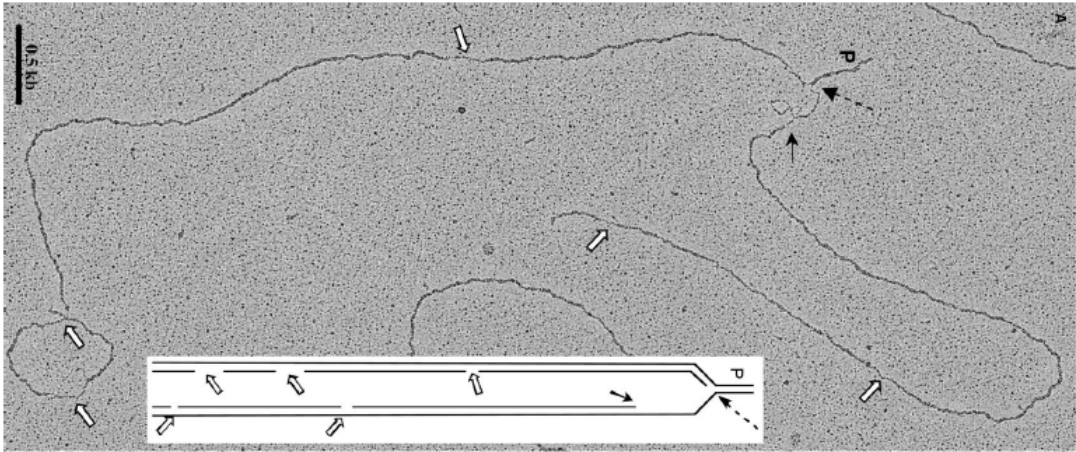
Lorsque le brin matrice d'une machinerie de synthèse est endommagé, une façon pour la cellule de poursuivre la réplication et d'éviter d'accumuler de l'ADNsb entre hélicase et polymérase consiste à ré-initier la synthèse en aval du site lésé. Cette ré-initiation (« repriming »), intrinsèquement présente pour la machinerie de synthèse discontinue (Pages and Fuchs 2003), existe également sur la machinerie de synthèse continue chez la bactérie (**Figure 24A**) (Heller and Marians 2006), la levure (Lopes et al. 2006b), et l'humain (Lehmann 1972; Cordeiro-Stone et al. 1999; Elvers et al. 2011).

Ce mécanisme de contournement du bloc de synthèse (généralement des dimères de thymine induits par exposition à des rayonnements UV) autorise la poursuite de la réplication mais a pour conséquence de laisser en amont de la fourche des régions d'ADNsb aux sites de lésion (Pages and Fuchs 2003; Karras and Jentsch 2010; Elvers et al. 2011), visibles par microscopie électronique en absence de synthèse translésionnelle et de NER (**Figure 24B**) (Lopes et al. 2006b). Ces régions seront répliquées subséquemment de manière découplée de la réplication, en S/G2, lors du « Post-Replication Repair » (PRR) (Lopes et al. 2006b; Daigaku et al. 2010; Karras and Jentsch 2010; Diamant et al. 2012), sans quoi la persistance de ce « gap » simple-brin pourrait conduire à la rupture d'une fourche de réplication à la phase S suivante (**Figure 24C**). Deux grands mécanismes existent pour accomplir leur réplication, l'un qualifié d'« error prone » qui fait intervenir des ADN polymérases peu fidèles, l'autre d'« error free » car il utilise l'information intacte d'une autre molécule. Le choix entre ces mécanismes de tolérance des dommages dépend des modifications post-traductionnelles de PCNA (Ulrich and Walden 2010): la mono-ubiquitination de sa lysine 164 (par Rad6/Rad18) promeut le bypass de la lésion par les polymérases translésionnelles, alors que sa poly-ubiquitination par Rad5 conduit à l'utilisation d'une autre copie de la séquence à répliquer (**Figure 24D**) (Carlile et al. 2009; Minca and Kowalski 2010).

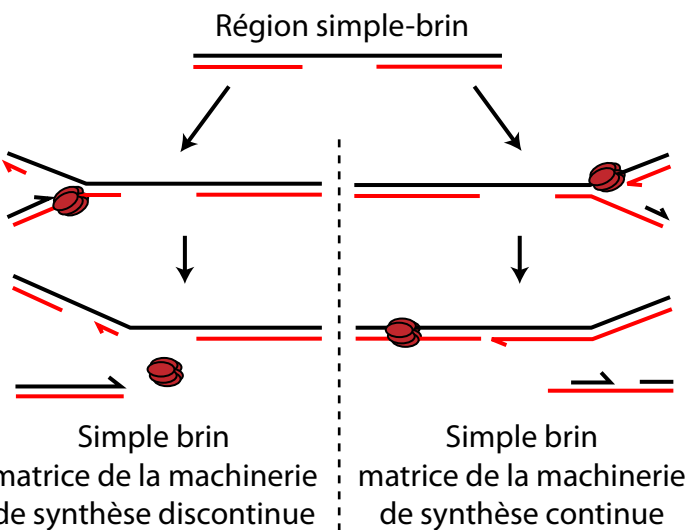
A



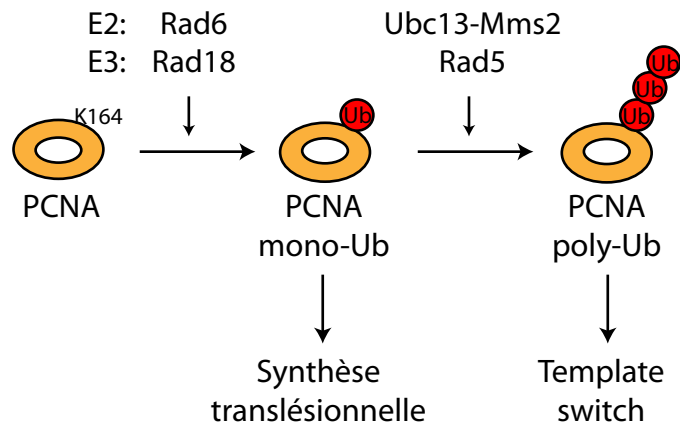
B



C



D



**Figure 24: Ré-initiation de la synthèse en cas de blocage d'une polymérase sur une matrice endommagée.** (A) Ré-initiation de la synthèse continue en aval d'une lésion. La ré-initiation nécessite la synthèse d'un primer ARN. Tiré de (Yao & O'Donnell, Curr. Op. Cell. Biol. 2009). (B) Fourche de réplication issue d'une souche de *S. cerevisiae* déficiente pour la synthèse translésionnelle et le mismatch repair après traitement aux UV, observée microscopie électronique (Lopes *et al.* Mol. Cell. 2006). (C) Modèle de cassure de fourche au niveau d'une région non-répliquée de l'ADN parental. (D) Les lésions et régions simple-brins laissée en amont de la fourche sont répliquées lors du PRR (post replication repair). La voie employée pour réaliser ce "gap-filling" dépend de l'état d'ubiquitination du clamp répliatif homo-trimérique PCNA (Pol30) sur sa lysine 164 (K164). La mono-ubiquitination par les E2 et E3 ubiquitin-ligases Rad6-Rad18 dirige vers un contournement de la lésion par des polymérases translésionnelles. La poly-ubiquitination subséquente par Ubc13-Mms2-Rad5 conduit à la réparation/réplication par template-switch.

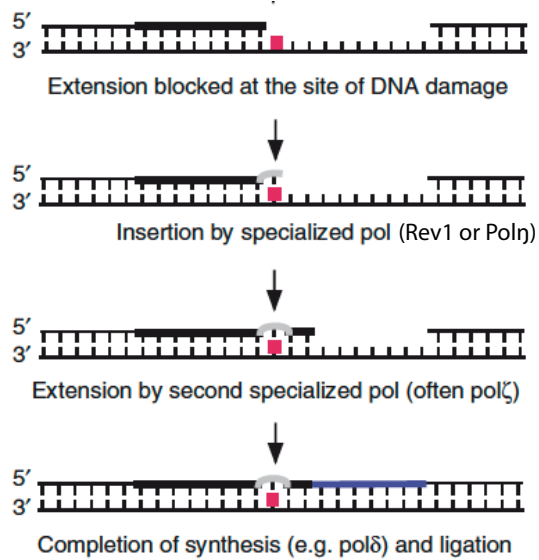
### *I.B.3.b.ii. La synthèse translésionnelle*

Ce « gap-filling », qui nécessite d'accommoder un nucléotide en absence d'un complémentaire identifiable par les polymérases répliquatives classiques, requiert l'action d'ADN polymérases spécialisées. Ces polymérases ne réparent pas la lésion mais permettent de la tolérer et d'empêcher sa conversion en une lésion plus toxique encore (la rupture d'une fourche de réplication au cycle ultérieur), au détriment du maintien de l'information génétique d'un ou de quelques nucléotides.

Les ADN polymérases translésionnelles appartiennent aux familles conservées des polymérases Y, qui incluent Pol $\eta$  (Rad30) et Rev1 chez *S. cerevisiae* et des polymérases B, représenté chez *S. cerevisiae* par Pol $\zeta$  (Rev3/Rev7). Ces enzymes sont peu fidèles, peu processives, et n'ont pas d'activité exonucléase 3'-5'. Malgré leurs similarités, elles présentent des activités biochimiques très différentes suivant la nature de la lésion à répliquer, ou la position de l'extrémité 3'-OH à étendre vis-à-vis de cette lésion (incorporation d'un nucléotide en face de la lésion, ou extension à partir d'un nucléotide mésapparié) (Prakash et al. 2005; Shachar et al. 2009). Ainsi, Rev1 ou Pol $\eta$  apposent le premier nucléotide en face de la lésion, et Pol $\zeta$ , plus fidèle, étend la molécule à partir de ce nucléotide mésapparié (**Figure 25**) (Haracska et al. 2001). Une fois la lésion passée, la polymérase répliquative (Pol $\delta$ ) se substitue à Pol $\zeta$  et finit la réplication de l'ADNsb restant jusqu'à l'ADN néo-synthétisé en aval (**Figure 25**). Le chargement spécifique et séquentiel de ces polymérases translésionnelles sur l'ADN endommagé nécessite la mono-ubiquitination de PCNA (K164) par Rad6/Rad18 (**Figure 24D**) (Prakash et al. 2005; Wood et al. 2007).

### *I.B.3.b.iii. Mécanismes utilisant une séquence homologue à la séquence lésée*

Une autre stratégie utilisée par la cellule pour répliquer une matrice endommagée consiste à utiliser l'information provenant d'une autre molécule intacte, impliquant un changement de matrice pour la polymérase, le « template-switch ». Comme la synthèse translésionnelle, il est mis en jeu au niveau de gaps simples-brins générés par un évènement de réinitiation lors du PRR (machinerie de synthèse continue et discontinue), ou à la fourche au niveau d'une polymérase bloquée en absence de réinitiation en aval (machinerie de synthèse continue uniquement). En raison de la



**Figure 25: Les polymérase translésionnelles permettent la réplication d'une matrice endommagée.** Adapté de (Takata *et al.* EMBO J. 2009).

similarité des transactions de brins et des protéines impliquées avec la recombinaison homologue (RH), je vais brièvement introduire ce mécanisme avant d'aborder les spécificités propres au template-switch.

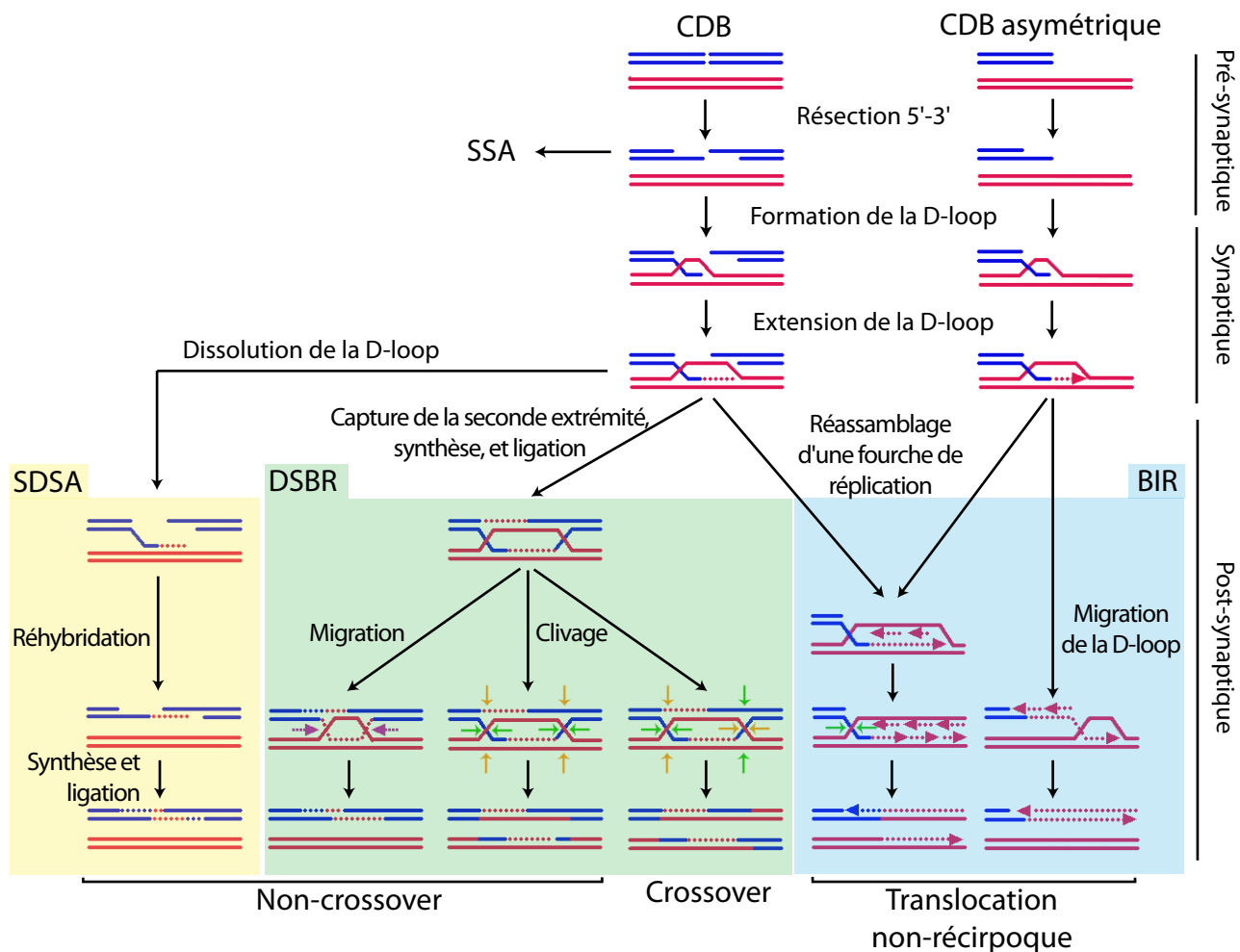
### - **La recombinaison homologue**

Une cassure double-brin (CDB) de l'ADN peut être réparée par deux grands mécanismes : la recombinaison homologue ou la jonction d'extrémités non-homologues (NHEJ, pour « Non-Homologous End Joining »). La recombinaison homologue est un mécanisme de réparation fidèle car il utilise l'information présente sur une autre molécule pour restaurer la linéarité du duplex lésé sans en altérer la séquence. C'est le principal mécanisme de réparation des CDBs chez *S. cerevisiae* en phase S et G2 (Barlow et al. 2008). La RH fait intervenir une succession de réactions, pour la plupart réversibles, lors de trois phases : pré-synaptique, synaptique, et post-synaptique (**Figure 26**).

Lors de l'étape pré-synaptique, chaque extrémité de la CDB est reconnue par le complexe hétéro-trimérique MRX (Mre11-Rad50-Xrs2), qui les maintient à proximité l'une de l'autre (Hopfner et al. 2002; Borde et al. 2004). En collaboration avec l'endonucléase Sae2, MRX initie la résection 5'-3' des extrémités de la cassure sur quelques centaines de nucléotides (Clerici et al. 2005). Cette résection peut au besoin être étendue sur plusieurs kilobases par Sgs1-Dna2 et Exo1 (Zhu et al. 2008; Cejka et al. 2010a). L'ADNsb ainsi généré de part et d'autre de la cassure est couvert par RPA. Rad52 catalyse alors le remplacement de RPA par Rad51 (Pâques and Haber 1999). Le nucléofilament de Rad51 est un intermédiaire essentiel de la recombinaison homologue et ses dérivés : il réalise la recherche d'une séquence homologue dans le génome et son invasion.

Lors de l'étape synaptique, le nucléofilament réalise la recherche d'homologie. Cette recherche procède par échantillonnage multiple de l'ADNdb donneur, dont l'efficacité augmente avec la taille du nucléofilament (Forget and Kowalczykowski 2012). L'augmentation de la mobilité de l'ADNdb à proximité du site de cassure favorise probablement la rencontre des molécules homologues (Mine-Hattab and Rothstein 2012). Une fois trouvée, le nucléofilament envahit l'ADNdb donneur. La D-loop (pour « Displacement loop ») ainsi formée est donc constituée du duplex formé par l'extrémité





**Figure 26: Mécanismes de recombinaison homologue.** Les étapes synaptiques et pré-synaptiques sont communes à toutes les voies de recombinaison homologue. Si des régions d'homologie sont présentes de part et d'autre de la CDB, un mécanisme de Single-Strand Annealing (SSA) peut intervenir, qui est présenté Figure 28. Les protéines requises à chaque étape et spécifiques à chaque voie sont détaillées dans le texte. Concernant le DSBR, les flèches roses indiquent une migration des jonctions de Holliday. Leur rencontre crée un hémicaténane qui, résolu par Top3, conduit à une réparation sans crossover. Les jonctions de Holliday peuvent également être résolues par des endonucléases (flèches vertes et oranges). Lorsque les deux jonctions sont résolues de manière identique (cas de gauche, flèches vertes), l'issue est un non-crossover, alors que quand les jonctions sont résolues de différentes manières (couple orange ou couple vert) l'issue de la réparation est un crossover.

résectée et la molécule donneuse complémentaire entouré du nucléofilament de Rad51, et du brin déplacé, qui peut éventuellement être capturé par l'autre extrémité de la cassure. Rad54 dégage l'extrémité 3' du brin envahissant pour permettre le chargement d'une ADN polymérase ( $Pol\delta$ ) (Heyer et al. 2006) qui va utiliser la matrice complémentaire pour restaurer la séquence présente au point de cassure, et ainsi créer une région complémentaire entre les deux extrémités de la cassure. Si la molécule donneuse provient de la chromatide sœur, la séquence néosynthétisée est le parfait complémentaire de l'autre extrémité de la cassure, et aucune conversion génique ne peut avoir lieu. En revanche, l'invasion du chromosome homologue dans une cellule diploïde ou d'une séquence ectopique peut conduire à la copie d'une séquence légèrement divergente qui, lors de l'étape post-synaptique, conduira à la formation d'un mésappariement au sein du duplex réparé et éventuellement à une conversion génique par action du MMR.

L'étape post-synaptique est l'étape la plus variable de la RH, qui conduit à distinguer 3 sous-familles aux issues variables : le SDSA (« Synthesis-Dependent Strand Annealing ») (Pâques and Haber 1999), le DSBR (« Double-Strand Break Repair ») (Szostak 1986), et le BIR (« Break-Induced Replication ») (Bosco and Haber 1998; Malkova et al. 2005) (**Figure 26**).

Dans le modèle du SDSA, la dissolution de la D-loop permet la réhybridation des deux extrémités simples-brins. La linéarité des deux brins peut alors être recouverte par simple copie de l'ADNsb généré par la résection et ligation à l'extrémité 5' présente en aval (**Figure 26**). Ce mécanisme, le plus courant dans les cellules en croissance mitotiques, ne conduit pas à la formation de crossover (Nassif et al. 1994; Bzymek et al. 2011).

Dans le modèle du DSBR, la deuxième extrémité simple-brin capture le brin déplacé de la D-loop. Une série de ligations des brins néosynthétisés conduit à la formation d'une double jonction de Holliday (dHJ) qui doit être résolue pour séparer les deux duplex. La façon de la résoudre peut conduire ou non à un crossing-over. La migration des jonctions l'une vers l'autre grâce au complexe hétéro-trimérique STR (Sgs1-Top3-Rmi1) aboutit à la formation d'un hémicaténane dont la résolution par Top3 conduit à un non-crossover (**Figure 26**) (Bussen et al. 2007; Cejka et al. 2010b ; Ashton et al. 2011). Alternativement, les jonctions de Holliday sont la cible d'endonucléases

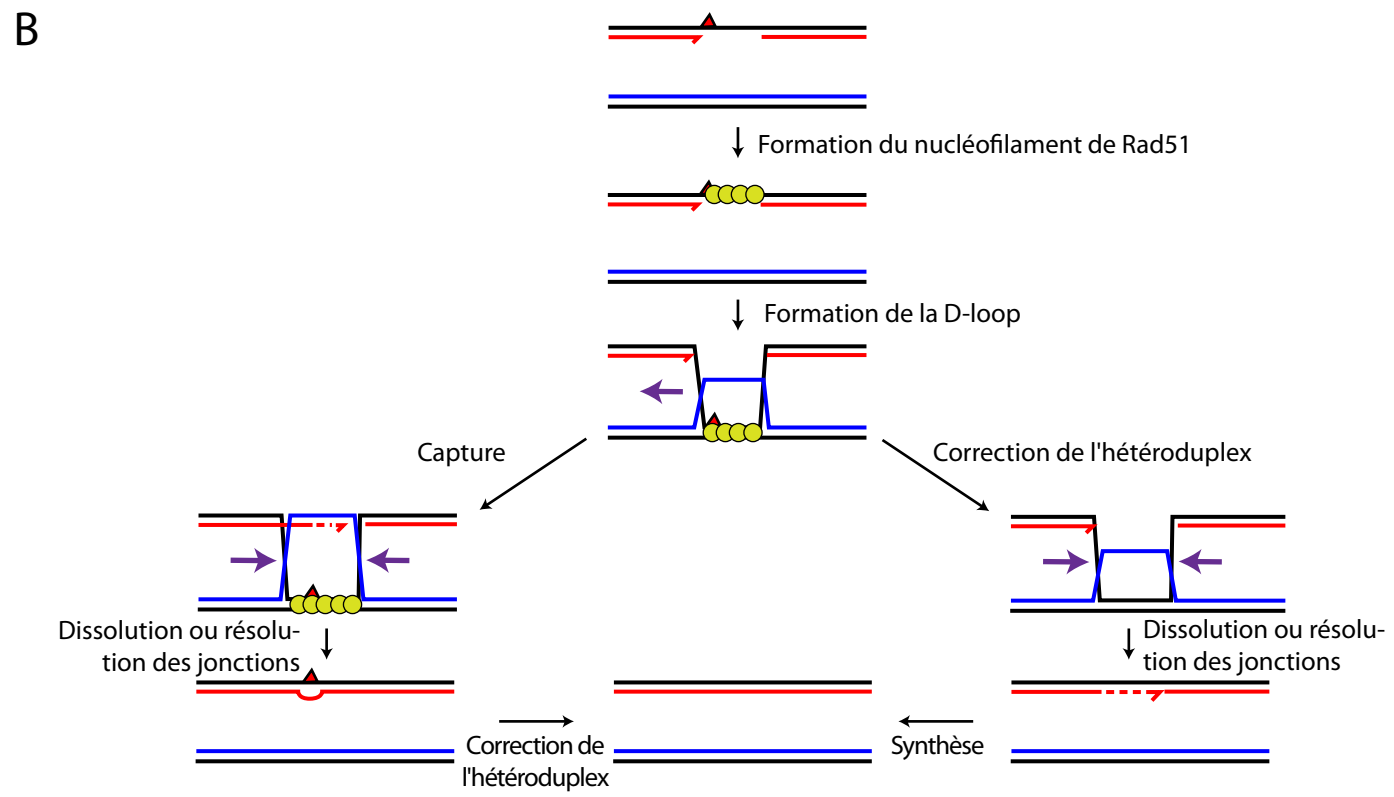
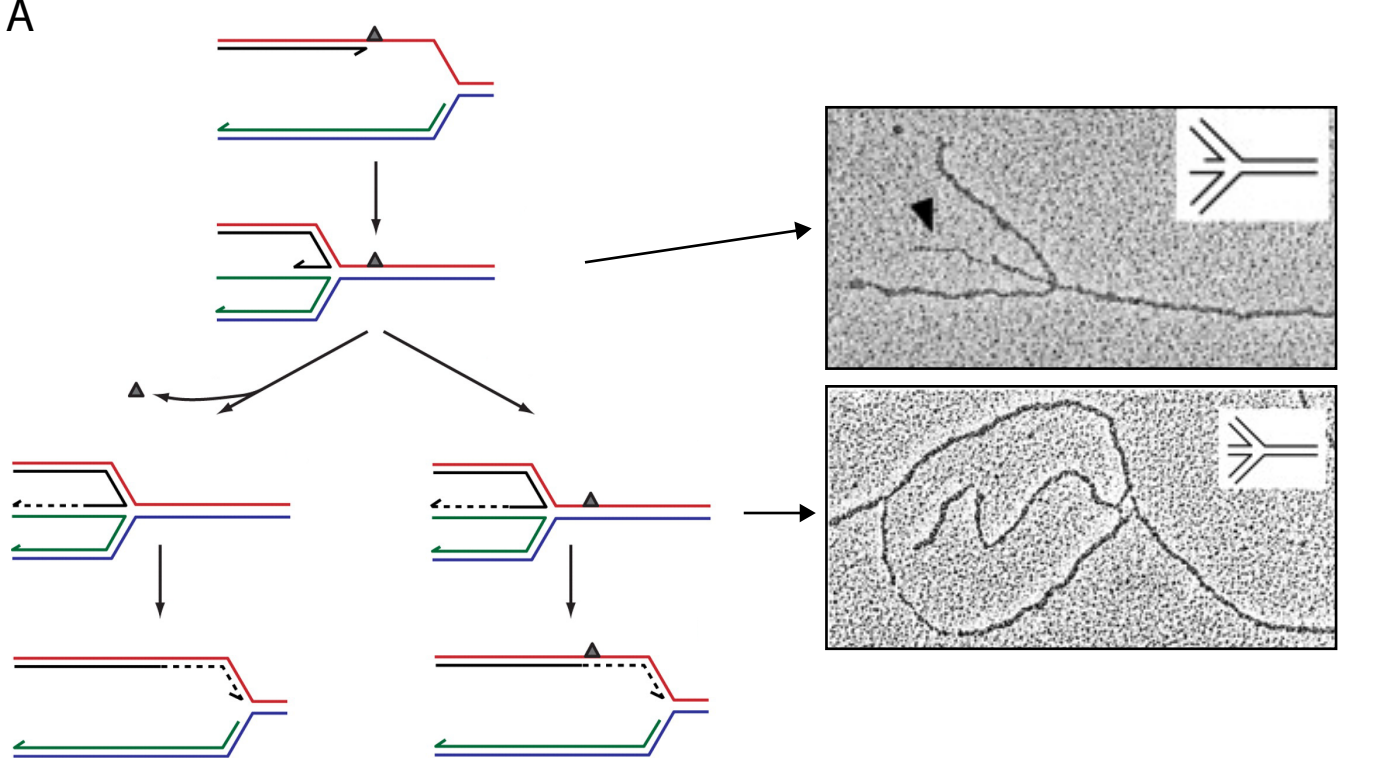
spécifiques et redondantes : Mus81-Mms4, Slx4-Slx1, et Yen1 (Ip et al. 2008; Ho et al. 2010; Schwartz and Heyer 2011). Suivant les brins clivés au niveau des jonctions, la résolution de la dHJ conduira soit à un crossover soit à un non-crossover (**Figure 26**) (Pâques and Haber 1999; Ho et al. 2010).

Lors du BIR, contrairement au SDSA, le brin envahissant n'est pas dissocié de la D-loop, et la synthèse se poursuit jusqu'à l'extrémité du chromosome où une séquences télomérique est copiée et stabilise le chromosome. Cette copie sur une longue distance se fait soit par migration de la D-loop ou de la jonction de Holliday, associée à la mise en place d'une synthèse discontinue sur le brin néo-synthétisé déplacé, soit par résolution de la jonction de Holliday qui permet l'établissement d'une fourche de réplication (**Figure 26**). Ce mode de réparation se solde par la formation d'une translocation non-réciproque associée à (i) une perte d'hétérozygotie si la réparation s'est faite sur le chromosome homologue, (ii) une perte d'hétérozygotie et un changement du nombre de copie si la région envahit est non-allélique (-1 pour l'extrémité distale du chromosome ; +1 pour la région copiée), (iii) la formation éventuelle de chromosomes dicentriques ; et dans tous les cas est associé à une forte mutagénèse ponctuelle (Deem et al. 2011). En présence d'une CDB « classique », le BIR est défavorisé par rapport au SDSA (Malkova et al. 2005; Lydeard et al. 2010; Marrero and Symington 2010). En revanche, il apparaît comme le seul mécanisme à même de réparer une CDB qui n'a qu'une seule extrémité, comme un télomère déprotégé ou une fourche de réplication cassée (Malkova et al. 2005). Dans ce dernier cas, le BIR est une façon de restaurer une fourche et poursuivre la réplication, au prix d'une mutagénèse accrue (**Figure 26**) (Deem et al. 2011).

### - **Le template-switch**

Comme la RH, le template-switch copie la séquence d'une molécule d'ADN autre que son complémentaire initial. Plusieurs mécanismes de template-switch semblent exister, suivant que la lésion se trouve au niveau d'une fourche bloquée ou au niveau d'une région d'ADNsb laissée en amont de la fourche (**Figure 27**) (Branzei 2011).

Dans le cas d'une réparation à la fourche, deux mécanismes de template-switch sont envisageables : une régression de fourche ou une invasion de la chromatide sœur (synthétisée de manière discontinue). La régression de fourche conduit à la formation d'une structure de type « chicken foot », qui ressemble à une jonction de Holliday



**Figure 27 : Mécanismes de "template-switch".** (A) La régression de fourche permet la réplication d'un ADN matriciel endommagé de deux façons: en utilisant l'ADN néo-synthétisé de la chromatide soeur comme matriciel (gauche), ou en réparant la matriciel lésée une fois réhybridée avec son complémentaire (droite). Des images de microscopie électronique de tels intermédiaires, observées dans des souches de levure déficientes pour Rad53 traitées à l'hydroxyurée sont présentées à droite (Sogo *et al.* Science 2002). (B) Modèle de template-switch sur une région simple brin interstitielle.

(Higgins et al. 1976; Atkinson and McGlynn 2009). La reprise de la synthèse continue se fait alors sur l'ADN néo-synthétisé de la chromatide opposée plutôt que sur l'ADN parental lésé (**Figure 27A**). Alternativement et non-exclusivement, l'ADN duplex parental reformé en aval de la jonction permet la correction de la base lésée. La migration de la « chicken foot » dans le sens opposé restaure une fourche fonctionnelle ayant répliqué fidèlement la lésion (**Figure 27A**). La régression de fourche a principalement été étudiée chez *E. coli*, mais des évidences génétiques et physiques suggèrent l'existence d'un tel mécanisme chez *S. cerevisiae* (Zou and Rothstein 1997; Sogo et al. 2002).

Le template-switch par invasion de la chromatide sœur, contrairement à la régression de fourche, requiert la protéine Rad51 et permet la réplication d'une matrice simple-brin endommagée à la fourche, ou laissée en amont après un événement de réinitiation de synthèse (**Figure 27B**) (Ulrich 2011). Dans les deux cas, le template-switch comporte obligatoirement une étape de recherche d'homologie, d'invasion de l'ADNdb homologue, et de copie de la séquence donneuse. Contrairement à la RH, le nucléo-filament Rad51 n'a pas une extrémité 3' libre et forme une D-loop close, flanquée par deux jonctions de Holliday (**Figure 27B**). Cette D-loop close permet de résoudre le blocage de synthèse de deux façons : en déplaçant un ADNsb qui peut être capturé et utilisé comme matrice « saine » par l'extrémité 3' à allonger, et en autorisant la correction de la base altérée au sein du duplex de la D-loop, initialement irréparable car présente au sein de l'ADNsb (**Figure 27B**). La façon dont la dHJ est résolue importe peu si le template-switch s'est fait entre chromatides sœurs, ce qui semble être généralement le cas (Ulrich 2011).

#### **I.B.4. Les réarrangements génomiques entre séquences répétées**

Les réarrangements génomiques consistent en l'établissement d'une continuité entre séquences normalement éloignées dans le génome. Ces réarrangements peuvent être des insertions, des délétions, des inversions et des translocations. Ils résultent de l'action impropre soit du NHEJ entre les extrémités de deux CDBs, soit de la RH et ses dérivés au niveau d'une CDB ou d'une fourche de réplication compromise. Notamment, les séquences répétées du génome constituent un risque pour les mécanismes qui utilisent une séquence homologue à la séquence lésée : la recombinaison entre copies

ectopiques, aussi appelée recombinaison non-allélique, est à l'origine de nombreux réarrangements. Aussi, la présence de séquences répétées en tandem ou dispersées dans le génome sont un vecteur d'instabilité génomique (Putnam et al. 2009; Chan and Kolodner 2011).

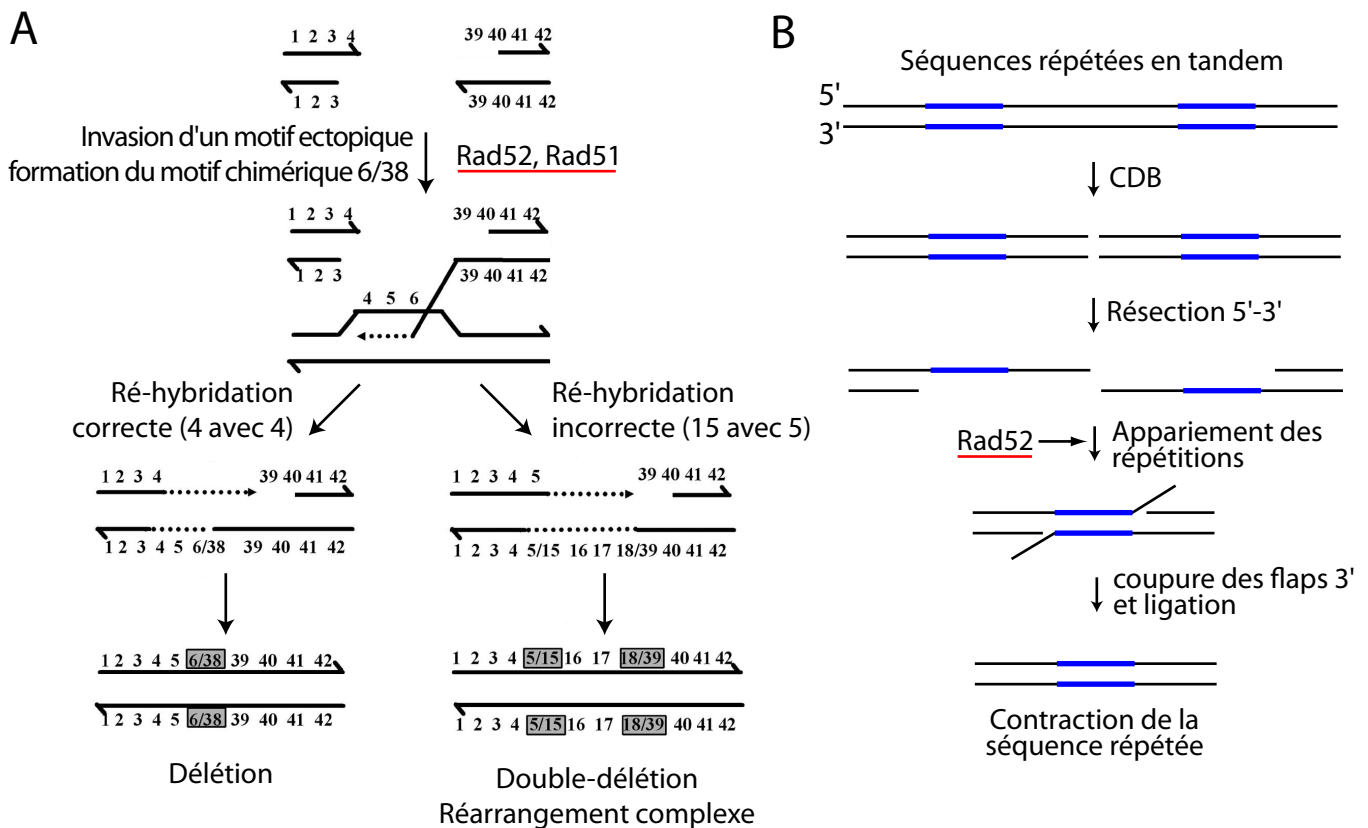
#### I.B.4.a. Instabilité des séquences répétées en tandem

Les étapes de recherche d'homologie (formation de la D-loop) ou de ré-hybridation, inhérentes à la RH et ses dérivés, sont l'occasion d'appariements incorrects entre motifs au sein de séquences répétées en tandem (exemple du SDSA **Figure 28A**). Ces évènements de mésappariements sont à l'origine des réarrangements de la séquence répétée réceptrice, puisque la synthèse initiée à partir de ces motifs mésappariés va conduire à la formation de jonctions chimériques et rompre la succession parentale des motifs. Suivant la position de l'extrémité envahissante et du motif envahi, le réarrangement sera une expansion ou une contraction. Dans le cas d'une réparation par DSBR (**Figure 26**), si la résolution de l'intermédiaire de recombinaison se fait par crossing-over, la copie donneuse sera également réarrangée. L'analyse par pedigree de la formation de réarrangements complexes suggère que de multiples cycles d'invasion/copie/dissociation surviennent en une seule génération (Lopes et al. 2006a).

Alternativement, deux motifs présents dans les simples-brins générés par résection de part et d'autre d'une CDB peuvent s'hybrider. La digestion des extrémités simples-brins 3' et la ligation des molécules conduit à une contraction de la séquence répétée. Ce mécanisme d'hybridation de simple-brin (SSA, single-strand annealing) dépend de la protéine Rad52, mais pas des protéines des phases synaptiques et post-synaptiques de la RH (Rad51, Rad54, ...). Il est schématisé **Figure 28B**.

#### I.B.4.b. Grands réarrangements chromosomiques (GCR) médiés par des séquences répétées dispersées

Les réarrangements entre copies distantes dans le génome sont qualifiés de « grands » réarrangements chromosomiques (GCR, pour Gross Chromosomal Rearrangement) et excluent l'instabilité des séquences répétées en tandem (expansion/contraction des micro- et minisatellites). Ces GCR peuvent être simples, comme une délétion, duplication, inversion, ou translocation et survenir lors d'un seul



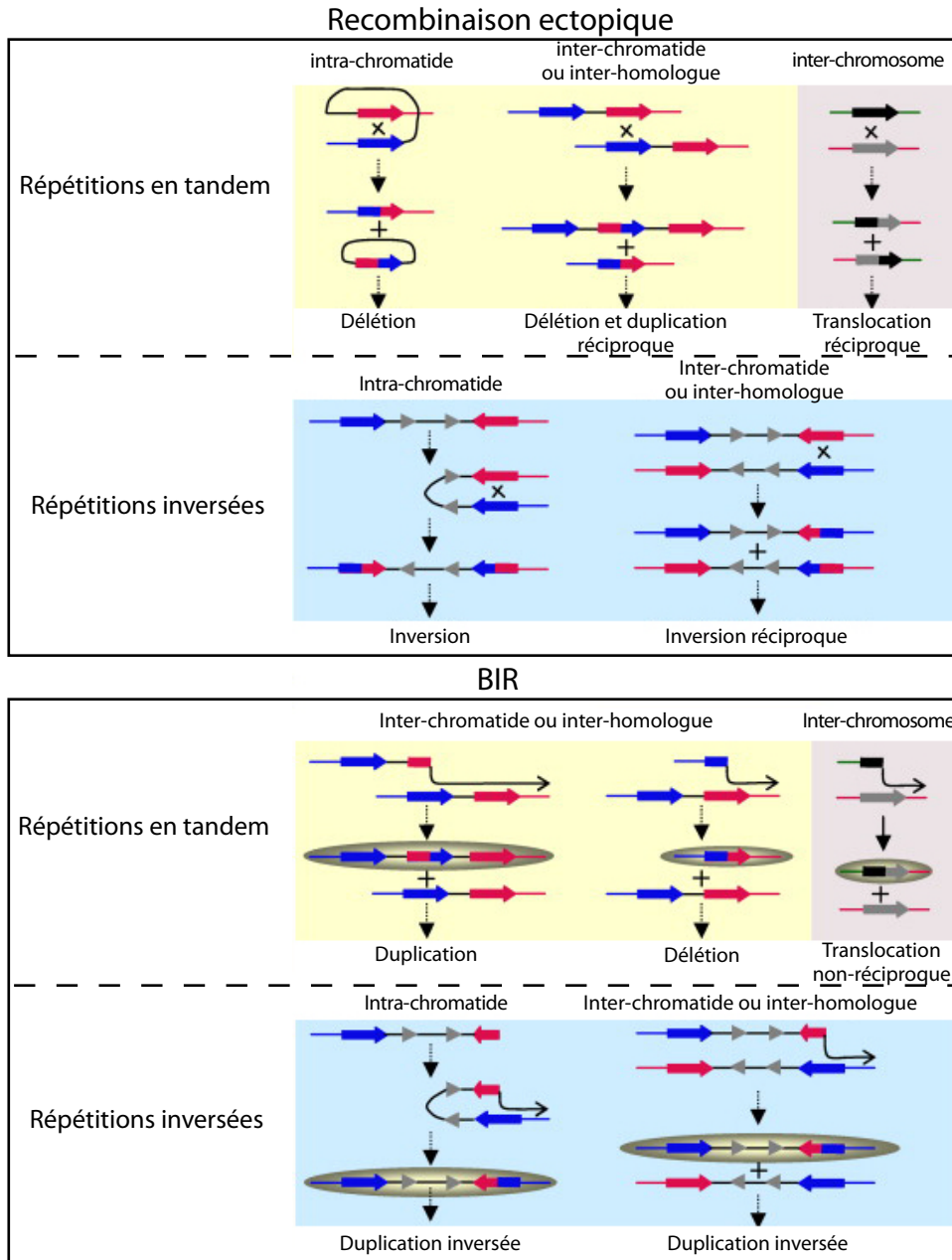
**Figure 28: Réarrangement des séquences répétées en tandem par recombinaison homologue.** (A) Exemples de la formation de réarrangements d'une séquence répétée en tandem par SDSA. Ce mécanisme requiert Rad52 et Rad51 pour l'étape d'invasion de brin. A gauche est montré un évènement de délétion simple, qui ne comporte qu'un motif chimérique (6/38) consécutif à une seule étape de mésappariement. Ce mécanisme peut également conduire à la formation d'expansions simples si le motif envahissant est inférieur au motif envahi. A droite, les mésappariements de motifs consécutifs conduisent à la formation de plusieurs motifs chimériques et à des réarrangements plus complexes (double-délétion, double-expansion, mixture). Adapté de (Lopes *et al.* Mol. Cell. Biol. 2006). (B) Modèle du SSA (Single-Strand Annealing) consécutif à la formation d'une CDB entre deux séquences répétées. Rad52 catalyse le déplacement de RPA, qui autorise l'hybridation entre répétitions. Ce mécanisme ne permet que la formation de contractions.

événement de recombinaison ectopique. Certains sont complexes, car ils contiennent une mixture de réarrangements simples, et sont probablement dus à une succession de recombinaisons ectopiques. C'est ce qui se produit lorsqu'un premier réarrangement conduit à la formation d'un chromosome dicentrique, qui casse en mitose jusqu'à la formation d'un réarrangement à même de le stabiliser sous forme mono-centromérique. Ainsi, la diversité des réarrangements médiés par des copies ectopiques à la séquence lésée dépend (i) de la phase du cycle (la phase G2 ajoute la possibilité de réarrangement sur la chromatide sœur), (ii) de la position de la copie ectopique dans le génome (même bras de chromosome, bras opposé, ou autre chromosome), et (iii) de son orientation, vis-à-vis de la position et l'orientation de la séquence réceptrice et des centromères.

Les mécanismes qui génèrent ces réarrangements ont principalement été étudiés à l'aide de systèmes génétiques chez les levures. À l'exception du SDSA, les sous-voies de la RH sont toutes aptes à produire des GCR (**Figure 26**). Le SSA n'est à l'origine que de délétions (**Figure 28B**). La voie du DSBR conduit à une délétion, une duplication, ou une inversion lorsque la copie ectopique provient du même chromosome que la séquence lésée, soit à la formation d'une translocation réciproque lorsqu'elle provient d'un autre chromosome (**Figure 29**). La voie du BIR est à l'origine de translocations non-réciproques (**Figure 29**). Chez *S. cerevisiae*, certaines translocations non-réciproques Rad51- et Rad52-dépendantes ne présentant qu'une homologie très limitée (< 70 pb) et imparfaite au point de jonction, ont suggéré le mécanisme de « micro-homology-mediated BIR » (MMBIR) (Bosco and Haber 1998; Chen and Kolodner 1999). La délétion terminale d'un chromosome peut être due à un événement de BIR sur un télomère, ou à l'addition *de novo* d'un télomère par la télomérase (Kramer and Haber 1993; Putnam et al. 2004; Pennaneach et al. 2006). Cette dernière réaction est inhibée par l'hélicase Pif1 (voir **Chapitre I.C.1.b**) (Schulz and Zakian 1994; Myung et al. 2001; Makovets and Blackburn 2009).

L'élucidation des voies impliquées dans la formation de ces réarrangements chez les organismes modèles a fourni les clés conceptuelles pour expliquer les réarrangements parfois complexes observés dans des cellules cancéreuses (Zhang et al. 2009; Chen et al. 2010).





**Figure 29: Grands réarrangements chromosomiques causés par des séquences répétées.** Différents réarrangements sont présentés, suivant la position de la copie ectopique (même chromatide, chromatide soeur, chromosome homologue, ou autre chromosome), et son orientation (en tandem ou inversée) vis-à-vis de la copie qui effectue la recherche d'homologie (en rouge). Les réarrangements causés par un mécanisme de recombinaison homologue résolu par crossing-over (haut) et par un mécanisme de BIR (bas) sont présentés. Suivant la position et l'orientation des répétitions sur le chromosome, certaines translocations ou inversions peuvent conduire à la formation de chromosomes dicentriques ou acentriques (non-montré). Adapté de (Chen *et al.* Seminars in Cancer Biol. 2010).

## **I.C. Étude de la formation des G-quadruplexes et des mécanismes assurant la stabilité des séquences sous-jacentes chez *S. cerevisiae***

### **I.C.1. Contexte et but de la thèse**

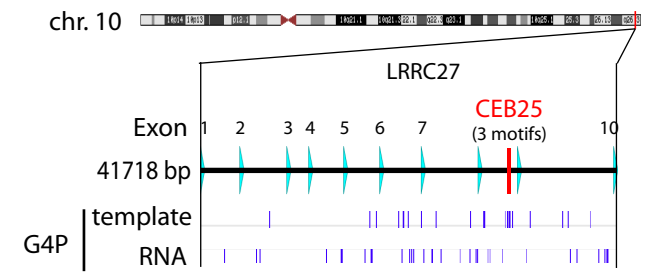
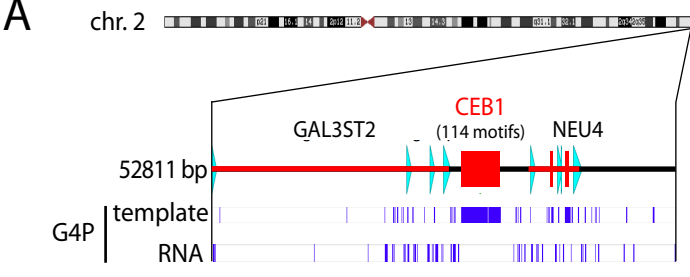
#### **I.C.1.a. Étude de l'instabilité des séquences répétées de type minisatellite chez la levure *S. cerevisiae***

##### *I.C.1.a.i. Les minisatellites CEB1 et CEB25*

Les génomes eucaryotes contiennent des séquences répétées en tandem. Ces séquences sont classées suivant la taille de leur motif unitaire en microsatellite (1-9 pb), minisatellite (10-100 pb), et satellite (>100 pb). L'analyse bio-informatique du génome humain (hg19) avec TRDB identifie ≈675000 de ces séquences, dont 53% sont des minisatellites. Ces minisatellites sont enrichis aux extrémités chromosomiques (**Article 4 Figure 6**) (Amarger et al. 1998). CEB1 (locus D2S90) et CEB25 (locus XXX) sont les deux minisatellites subtélomériques humains que nous avons étudiés. CEB1 est localisé dans la région promotrice du gène NEU4 (Figure), et CEB25 dans l'intron 9 du gène LRRC27 (**Figure 30A**). Ils sont dit « hypervariables », c'est-à-dire que leur taille varie fréquemment d'un individu à l'autre, par changement du nombre de motifs. De fait, CEB1 est extrêmement polymorphique dans la population, avec des tailles d'allèles allant de 0,2 à >7 kb (Buard and Vergnaud 1994; Buard et al. 1998). Cette instabilité survient dans la lignée germinale mâle, initialement estimée en analyse de pedigree à 13% par spermatozoïde, et qui est positivement corrélée à la taille parentale du minisatellite (Buard et al. 1998). De manière similaire, la taille des allèles de CEB25 est très variable dans la population, et comprise entre 0,5 et 20 kb (Amrane et al. 2012).

Le polymorphisme des allèles CEB1 et CEB25 implique aussi la séquence des motifs unitaires, dont les caractéristiques sont résumées **Table 2**. La longueur des motifs de l'allèle CEB1-1.8 (42 motifs, 1,8 kb) est comprise entre 37 et 43 nt (**Figure 30B**) (Lopes et al. 2006a). La séquence de l'allèle CEB25-1.7 (31 motifs, 1,7 kb) révèle également la présence de polymorphismes de séquence et de taille entre motifs (**Figure 30C**) (Amrane et al. 2012).

##### *I.C.1.a.ii. Étude de la stabilité du minisatellite CEB1 chez *S. cerevisiae**



**B** CEB1-1.8

```

1 AGTCGGTCCCTG CAG CGTCCGGTGGGAGGGAGGGGG
2 AGTCGGGTCCCTGGAGGCGTCCGGTGGAGGGAGGGGG
3 AGTCGGGTCCCTGGAGGCGTCCGGTGGGAGGGAGGGGG
4 AGTCGGGTCCCTGGAGGCGTCCGGTGGGAGGGAGGGGG
5 AGTCGGGTCCCTGGAGGCGTCCGGTGGGAGGGAGGGGG
6 AGTCGGGTCCCTGGAGGCGTCCGGTGGGAGGGAGGGGG
7 AGTCGGGTCCCTGGAG CCGTCCGGTGGGAGGGAGGGGG
8 AGTCGGGTCCCTGGAG CCGTCCGGTGGGAGGGAGGGGG
9 AGTCGGGTCCCTGGAGGCGTCCGGTGGGAGGGAGGGGG
10 AGTCGGGTCCCTGGAG CCGTCCGGTGGGAGGGAGGGGG
11 AGT GGGTCCCTGGAGGCGTCCGGTGGGAGGGAGGGGG
12 AGTCGGGTCCCTGGAGGCGTCCGGTGGAGGGAGGGGG
13 AGTCGGGTCCCTGGAGGCGTCCGGTGGGAGGGAGGGGG
14 AGTCGGGTCCCTGGAGGCGTCCGGTGGGAGGGAGGGGG
15 AGTCGGGTCCCTGGAG CCGTCCGGTGGGAGGGAGGGGG
16 AGTCGGGTCCCTGGAG CCGTCCGGTGGAGGGAGGGGG
17 AGTCGGGTCCCTGGAGGCGTCCGGTGGGAGGGAGGGGG
18 AGTCGGGTCCCTGGAGGCGTCCGGTGGGAGGGAGGGGG
19 AGTCGGGTCCCTGGAG CCGTCCGGTGGGAGGGAGGGGG
20 AGTCGGGTCCCTGGAGGCGTCCGGTGGGAGGGAGGGGG
21 AGTCGGGTCCCTGGAG CCGTCCGGTGGGAGGGAGGGGG
22 AGTCGGGTCCCTGGAGGCGTCCGGTGGGAGGGAGGGGG
23 AGT GGGTCC TGGAG CCGTCCGGTGGGAGGGAGGGGG
24 AGTCGGGTCCCT CAG CCGTCCGGTGGGAGGGAGGGGG
25 AGTCGGGTCCCTGGAG CCGTCCGGTGGAGGGAGGGGG
26 AGTCGGGTCCCTGGAGGCGTCCGGTGGGAGGGAGGGGG
27 AGTCGGGTCCCTGGAGGCGTCCGGTGGGAGGGAGGGGG
28 AGTCGGGTCCCTGGAG CCGTCCGGTGGGAGGGAGGGGG
29 AGT GGGTCC TGGAG CCGTCCGGTGGGAGGGAGGGGG
30 AGTCGGGTCCCT CAG CCGTCCGGTGGAGGGAGGGGG
31 AGTCGGGTCCCTGGAG CCGTCCGGTGGGAGGGAGGGGG
32 AGTCGGGTCCCTGGAGGCGTCCGGTGGGAGGGAGGGGG
33 AGTCGGGTCCCTGGAGGCGTCCGGTGGGAGGGAGGGGG
34 AGTCGGGTCCCTGGAGGCGTCCGGTGGGAGGGAGGGGG
35 AGTCGGGTCCCTGGAG CCGTCCGGTGGGAGGGAGGGGG
36 AGTCGGGTCCCT CAG CCGTCCGGTGGAGGGAGGGGG
37 AGTCGGGTCCCTGGAG CCGTCCGGTGGAGGGAGGGGG
38 AGTCGGGTCCCTGGAGGCGTCCGGTGGGAGGGAGGGGG
39 AGTCGGGTCCCTGGAGGCGTCCGGTGGGAGGGAGGGGG
40 AGTCGGGTCCCTGGAGGCGTCCGGTGGAGGGAGGGGG
41 AGTCGGGTCCCTGGAG CCGTCCGGTGGAGGGAGGGGG
42 A TCGGGTCCCTGGAG CCGTCCGGTGGGAGGGAGGGGG
GGGAGGGAGGGAGGGGG

```

**C** CEB25-0.8 (AL096806)

```

TGAGTGTAAAGTGTGGGTGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
TGGGTGTGAGTGTGGGTGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
TGGGTGTGAGTGTGGGCTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
TGCATGAATCTGATGTGCAGGTAATGTAAGGGTGGGTGTAAGTGTGGG
CGGGTGAATTTGTGCACTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
TGGGTGTAATGTGGGTGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
TGGGTGTGAGTGTGGGAGTGGAGGTAGATGTAAGCACTAGTGTAAATGTGGG
-----TGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
TGGGTGTGAGTGTGGGTGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
TGGGTGTGAGTGTGGGTGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
TGGGTGTAATTTGTGGCGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
TGGGTGTAATTTGTGGCGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
TGGGTGTGAGTGTGGGCTGGAGGTAGATGTAAGCACTAGTGTAAATGTGGG
-----TGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
TGGGTGTGAGTGTGGGCTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
TGGGTGTGAGTGTGGGTGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
TGGGTGTAAGCGGGTGTGGAGGTAGATGTAAGTGTGGGTGTGCTCATGGG

```

CEB25-1.7

```

1 TGAGTGTAAAGTGTGGGCTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
2 TGGGTGTAATTTGTGGCGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
3 TGGGTGTGAGTGTGGGTGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
4 TGGGTGTAATTTGTGGCGTGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
5 TGGGTGTGAGTGTGGGCTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
6 TGGGTGTGAGTGTGGGCTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
7 TGGGTGTGAGTGTGGGTGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
8 TGGGTGTAATTTGTGGCGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
9 TGGGTGTAATTTGTGGCGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
10 TGGGTGTGAGTGTGGGCTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
11 TGGGTGTAATTTGTGGCGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
12 TGGGTGTGAGTGTGGGCTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
13 TGGGTGTAATTTGTGGGTGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
14 TGGGTGTGAGTGTGGGTGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
15 TGGGTGTAATTTGTGGGTGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
16 TGGGTGTGAGTGTGGGTGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
17 TGGGTGTAAGTGTGGGTGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
18 TGGGTGTAATTTGTGGCGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
19 TGGGTGTGAGTGTGGGTGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
20 TGGGTGTAATTTGTGGGTGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
21 TGGGTGTAAGTGTGGGTGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
22 TGGGTGTGAGTGTGGGTGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
23 TGGGTGTGAGTGTGGGTGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
24 TGGGTGTAAGTGTGGGTGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
25 -----CTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
26 TGGGTGTAATTTGTGGCGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
27 TGGGTGTGAGTGTGGGTGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
28 -----TGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
29 TGGGTGTAAGCGGGTGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
30 TGGGTGTGAGTGTGGGTGTGGAGGTAGATGTAAGGGTGGGTGTAAGTGTGGG
31 TGGGTGTAAGCGGGTGTGGAGGTAGATGTAAGTGTGGGTGTGCTCATGGG

```

**Figure 30: Localisation et séquences des minisatellites humains CEB1 et CEB25.** (A) Localisation chromosomique et environnement local des minisatellites CEB1 et CEB25 (UCSC, hg19). Les séquences G4P de la région ont été déterminées avec QGRS Mapper avec les paramètres suivants: taille totale = 40 nt; taille de boucle < 21 nt, taille du run de G = 3 nt. La position des minisatellites a été déterminée avec Tandem Repeat Finder avec les paramètres de recherche par défaut. (B) Séquence de l'allèle CEB1-1.8 (Lopes et al. Mol. Cell. Biol. 2006). Le brin G, orienté de 3'-5' est représenté. Les polymorphismes d'un nucléotide sont indiqués en couleur. Le motif est polymorphe en taille de part les variations du nombre de guanine à l'extrémité du motif. (C) Séquences d'un allèle de CEB25 de 0.8 kb (CEB25-0.8, AL096806) et d'un allèle de 1.7 kb (CEB25-1.7) (Amrane et al. JACS 2012). Le brin G est orienté 5'-3'. La région capable de former un G-quadruplexe *in vitro* est soulignée.

Il y a plusieurs années, le laboratoire d'Alain Nicolas a utilisé la levure *S. cerevisiae* pour comprendre le mécanisme d'instabilité méiotique de CEB1 (Debrauwere et al. 1999), puis son comportement dans les cellules en croissance mitotique. Pour cela, le minisatellite a été inséré dans le génome au niveau d'un hotspot de recombinaison méiotique, le promoteur du gène *ARG4*. Il s'est ensuite intéressé aux facteurs cellulaires assurant sa stabilité dans des cellules en croissance mitotique à ce locus. La procédure permettant de déterminer la fréquence de réarrangements d'un minisatellite, ainsi que les différents sites d'insertions de CEB1 dans le génome étudiés au cours de ma thèse sont détaillés dans le chapitre de **Méthodes**.

Dans les cellules diploïdes et haploïdes sauvages, l'allèle naturel CEB1-1.8 manifeste une faible fréquence de réarrangement (respectivement 2/1990 et 5/1824 colonies analysées)(**Article 1**) (Lopes et al. 2002). La stabilité du minisatellite au cours des générations a été mesurée dans différents mutants, notamment le mutant *rad27Δ*. Rad27 est une endonucléase conservée (homologue humain : FEN-1) impliquée dans la maturation de la structure flap générée par le déplacement de l'extrémité 5' du fragment d'Okazaki sur le brin à synthèse discontinue au cours de la réplication. En l'absence de la protéine Rad27, CEB1-1.8 est fréquemment réarrangé (42 et 48 % respectivement dans des cellules diploïdes et haploïdes) (Lopes et al. 2002; Maleki et al. 2002; Lopes et al. 2006a). L'étude des mutants *rad51Δ* et *rad52Δ* montre que l'instabilité dépend de la voie de la recombinaison homologue (Lopes et al. 2006a) et n'est pas spécifique à CEB1 puisque les 11 séquences répétées en tandem testées sont déstabilisées, quelle que soit la taille du motif unitaire (de 1 à 192 nt) (**Article 1**). Le modèle d'instabilité propose qu'un défaut de maturation des fragments d'Okazaki conduirait à la formation de lésions recombinogènes (ADNsb ou CDB) qui, lorsqu'elles surviennent au niveau de motifs répétés, conduiraient à son réarrangement (**Chapitre I.B.4.a**).

L'équipe s'est également intéressée au rôle de Dna2, une protéine essentielle impliquée dans la maturation des fragments d'Okazaki (Pike et al. 2009). Le mutant thermosensible *dna2-1* induit une faible instabilité de CEB1 à température semi-permissive ( $\approx 1,5\%$ , 19/1293) (Lopes et al. 2002). La létalité du mutant *dna2Δ* est supprimée par délétion du gène *PIF1* (Budd et al. 2006). CEB1 est déstabilisé dans le double-mutant *pif1Δdna2Δ* (4,7%, 18/384), mais ceci au même niveau que dans le

**Table 2: Minisatellites humains dont l'instabilité a été étudiée chez *S. cerevisiae* durant ma thèse**

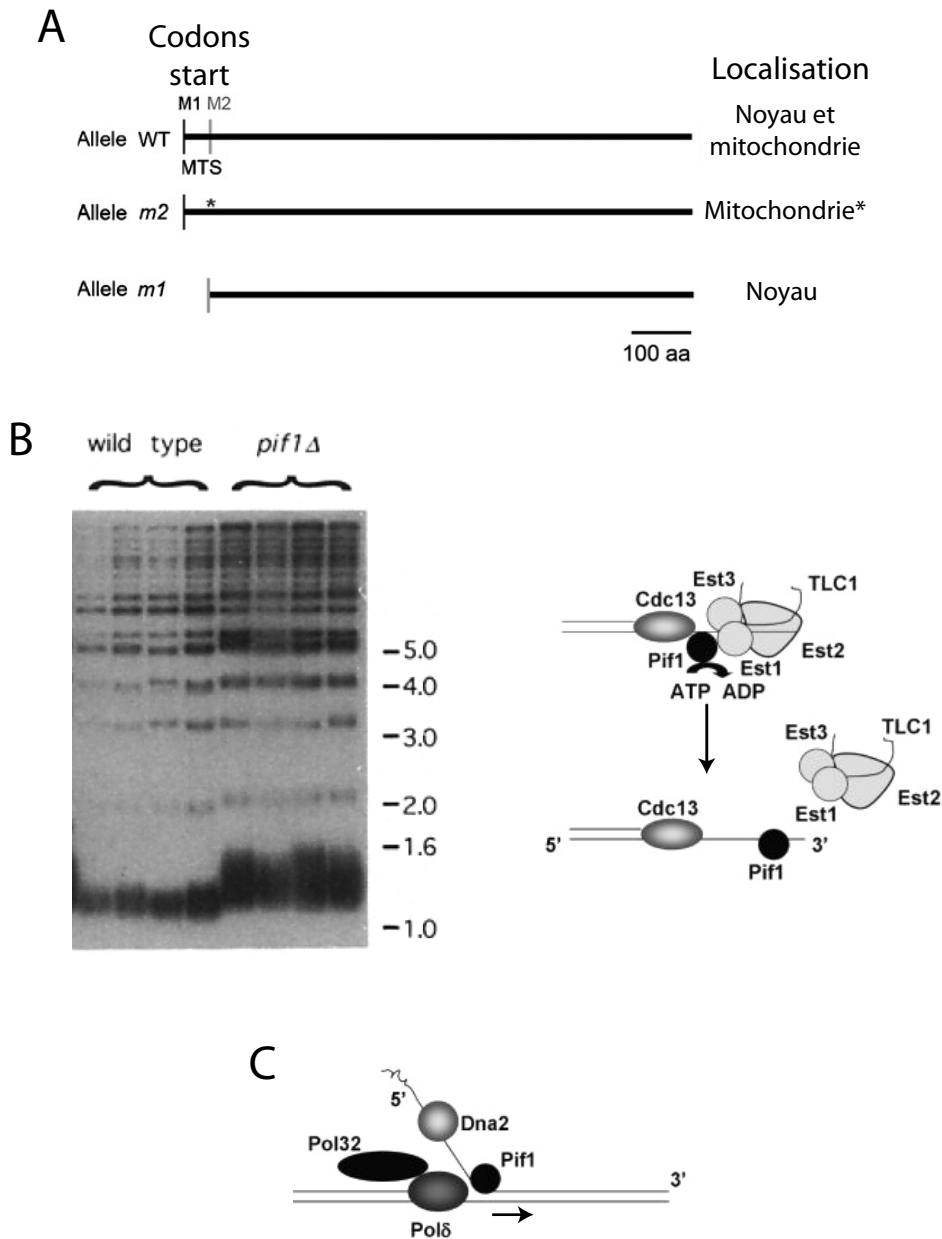
Minisatellite	Localisation	Distance au télomère	Taille du motif	%GC	GC-bias	Séquence du brin riche en G (5'-3')
CEB1	2q37.3	0.4 Mb	39 bp	77%	77%	GGGGGAGGGAGGGTGGCCTGCCGAGGTCCTGGGCTGA
CEB25	10q26.3	1.4 Mb	52 bp	56%	100%	AAGGGTGGGTGTAAGTGTGGGTGGGTGTGAGTGTGGGTGTGGAGGTAGATGT
HRAS1	11p15.5	0.5 Mb	28 bp	68%	74%	CCCTGGAGAGAAGGGGAGTGTGGCGT

simple-mutant *pif1Δ* (6%, 40/672)(**Article 1**). Ce résultat est à l'origine de l'étude du rôle de Pif1 dans l'instabilité de CEB1 développé lors de ma thèse.

#### I.C.1.b. L'hélicase Pif1

Pif1 (Petite integration frequency 1) est une hélicase ADN appartenant à la superfamille d'hélicase SF1, et membre fondateur d'une sous-famille conservée chez les eucaryotes (Foury and Kolodynski 1983; Foury and Lahaye 1987; Bochman et al. 2010). D'un point de vue biochimique, le coefficient de sédimentation de la protéine suggère qu'elle est sous forme monomérique en solution. Elle déroule les acides nucléiques avec une polarité 5'-3', mais est peu active sur des duplex ADN/ADN (**Article 1**) (Lahaye et al. 1993). Cette efficacité diminue avec la taille du substrat et augmente avec la concentration de protéine, suggérant un mode d'action distributif (Lahaye et al. 1993). En revanche l'efficacité de déroulement augmente significativement si le substrat présente un flap 3' ou si le brin à dérouler est un hybride ARN/ADN (Lahaye et al. 1993; Boule and Zakian 2007). Ces activités dépendent de la présence du cofacteur  $Mg^{2+}$ , d'ATP, et du domaine hélicase fonctionnel de Pif1, inactivé dans les mutants Pif1-K264A (ne lie pas l'ATP) et Pif1-K264R (n'hydrolyse pas l'ATP) (**Article 1**) (Boule et al. 2005; Boule and Zakian 2007).

La protéine Pif1 est impliquée dans plusieurs processus biologiques. En raison de la présence d'un codon alternatif d'initiation de la traduction, le gène *PIF1* code pour une forme mitochondriale (premier ATG) et une forme nucléaire (deuxième ATG)(**Figure 31A**) (Foury and Lahaye 1987; Lahaye et al. 1991; Rose et al. 1992; Schulz and Zakian 1994; Zhou et al. 2000 ; Futami et al. 2007). De ce fait, le mutant *pif1Δ* présente des phénotypes mitochondriaux et nucléaires qui peuvent être découplés par mutation ponctuelle de l'un ou l'autre des codons d'initiations de la traduction (Schulz and Zakian 1994). Toutefois, la mutation *pif1-m2* n'est pas équivalente à la délétion du gène, probablement par complémentation partielle des fonctions nucléaires par la forme mitochondriale (**Articles 1 et 3**) (Myung et al. 2001). La forme mitochondriale de Pif1 lie l'ADN mitochondrial, prévient la formation de CDB ou promeut leur réparation par recombinaison homologue, et limite l'apparition de mutations (O'Rourke et al. 2002; O'Rourke et al. 2005; Cheng et al. 2007). En son absence, les fonctions mitochondriales (respiration, régulation de l'homéostasie du Fer et du Zinc) sont perdues après 24-48h



**Figure 31: L'hélicase Pif1 de *S. cerevisiae*.** (A) Représentation schématique des deux isoformes de Pif1. L'ARNm contient deux codons d'initiation de la traduction. La forme mitochondriale est traduite à partir du premier AUG. Elle contient un signal d'adressage à la mitochondrie (MTS), clivé à l'issue de l'importation. La forme nucléaire est traduite à partir du second AUG. Ne possédant pas de MTS, elle ne peut pas être adressée à la mitochondrie. \* le mutant *pif1-m1* limitant certains phénotypes nucléaires, une fuite de la forme mitochondriale vers le noyau est soupçonnée. Tiré de (Bochman *et al.* DNA Rep. 2010). (B) Pif1 inhibe l'action de la télomérase au télomère. À gauche, télo-blot de 4 souches WT et 4 souches *pif1*Δ. Les bandes qui migrent à des poids moléculaires à 4 kb sont des télomères. Tiré de (Schulz & Zakian, Cell 1994). À droite, modèle d'action de Pif1 au télomère. Tiré de (Boulé & Zakian, Nucl. Acids Res. 2006). (C) Modèle d'action de Pif1 dans l'élongation du flap d'un fragment d'Okazaki durant la réplication. Pif1 est chargé sur le flap et, en se déplaçant de 5' vers 3', déplace le flap 5' conjointement à Polδ. Tiré de (Boulé & Zakian, Nucl. Acids Res. 2006).

de culture liquide (Foury and Lahaye 1987; O'Rourke et al. 2002; O'Rourke et al. 2005; Cheng et al. 2007; Guirola et al. 2010).

La forme nucléaire, stimule la pause de la fourche de réplication au niveau du site RFB du rDNA (Ivessa et al. 2000), et inhibe l'action de la télomérase au télomère (Schulz and Zakian 1994; Boule et al. 2005) et au niveau de CDB (Schulz and Zakian 1994; Myung et al. 2001; Makovets and Blackburn 2009; Zhang and Durocher 2010) : en son absence ou dans un mutant catalytique de Pif1, la taille moyenne des télomères augmente d'environ 100 pb (**Figure 31B**), et les réarrangements chromosomiques par addition *de novo* d'un télomère par la télomérase augmente  $\approx 1000$  fois en comparaison d'une souche sauvage (Myung et al. 2001). L'inhibition de la télomérase par Pif1 à une CDB (mais pas du télomère) requiert la phosphorylation de Pif1 par les kinases du checkpoint intra-S Rad53 et Mec1 (Makovets and Blackburn 2009). L'interaction génétique entre *PIF1* et plusieurs gènes codants pour des protéines impliquées dans la maturation des fragments d'Okazaki (dont *DNA2*) (Budd et al. 2006; Chang et al. 2009), ainsi que des expériences de reconstitution de la machinerie de maturation des fragments d'Okazaki *in vitro* suggèrent que Pif1 participe, conjointement à Pol $\delta$ , à la formation de longs flaps subséquentement clivés par Dna2 (**Figure 31C**), mais aucune expérience *in vivo* n'a pour l'heure corroboré ce rôle (Pike et al. 2009).

#### I.C.1.c. Un rôle des G-quadruplexes dans l'instabilité des séquences minisatellites chez *S. cerevisiae*?

##### *I.C.1.c.i. CEB1 est le seul minisatellite instable en absence de Pif1*

Étant donné la possible implication de Pif1 dans la maturation des fragments d'Okazaki, l'équipe chercha à comprendre si la déstabilisation de CEB1 dans le mutant *pif1 $\Delta$*  procédait de manière similaire à celle observée dans un mutant *rad27 $\Delta$* . De manière surprenante, la délétion de *PIF1* n'induit pas de réarrangement des 11 séquences répétées en tandem toutes instables dans le mutant *rad27 $\Delta$*  (**Article 1 Table 1**) : l'absence de Pif1 déstabilise donc spécifiquement CEB1. Par ailleurs, aucun autre mutant hélicase testé (*sgs1 $\Delta$* , *rrm3 $\Delta$* , *mph1 $\Delta$* , *srs2 $\Delta$* ) n'induit l'instabilité de CEB1 (**Article 1 Table 1**). Quelle propriété spécifique à CEB1 requiert l'action spécifique de Pif1 pour assurer sa stabilité?

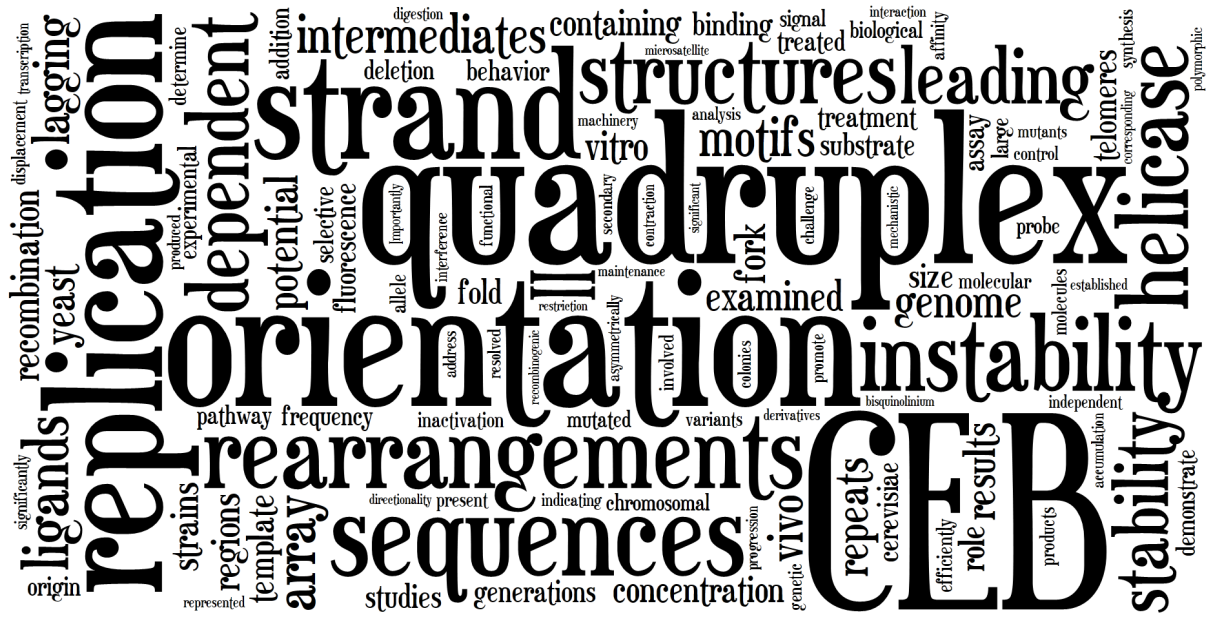
##### *I.C.1.c.ii. CEB1 forme des G-quadruplexes déroulés par Pif1 in vitro*



Le motif CEB1 contient plusieurs blocs de guanines contiguës (**Table 2**). Cette particularité du motif de CEB1 par rapport aux autres séquences répétées en tandem testée lui confère le potentiel de former une structure G-quadruplexe. Une collaboration initiée avec l'équipe de biophysiciens dirigée par JL. Mergny permet de montrer que le motif de CEB1 forme effectivement un G-quadruplexe stable *in vitro* (**Article 1 Figure 2**). La capacité de Pif1 à dérouler un substrat G-quadruplexe formé par le motif de CEB1 a été testée *in vitro* (voir **Article 1 Figure 2**). Ce G-quadruplexe inter-moléculaire (mixture bi et tétra-moléculaire) est extrêmement résistant à la dénaturation thermique. A stœchiométrie 1:1 entre Pif1 et son substrat, le déroulement du G-quadruplexe est >80% après 5 minutes d'incubation, alors que dans ces conditions seulement 25% du duplex ADN/ADN (20 nt) est déroulé (**Article 1 Figure 2**). L'analyse cinétique montre que le déroulement du G-quadruplexe est rapide, puisque pour une stœchiométrie 5:1 entre Pif1 et son substrat, >75% du G-quadruplexe est déroulé en 1 mn, alors que seulement 50% du duplex ADN/ADN est déroulé après 15 mn. Ces études montrent que Pif1 est capable de dérouler une structure G-quadruplexe très stable plus efficacement qu'un court duplex ADN/ADN. Cette capacité à dérouler les G-quadruplexes est conservée puisque l'hélicase Pif1 humaine présente également une activité de déroulement d'un substrat G-quadruplexe inter-moléculaire supérieure à celle observée sur un duplex ADN/ADN (Sanders 2010).

Les G-quadruplexes formés dans CEB1 et non-déroulés en absence de Pif1 sont-ils à responsables de l'instabilité spécifiquement observée avec ce minisatellite dans le mutant *pif1Δ* ?

## II. RÉSULTATS





**Article 1 :**

**The Yeast Pif1 Helicase Prevents Genomic Instability  
Caused by G-Quadruplex-Forming CEB1 Sequences *In Vivo***



## II. Résultats

### II.A. Article 1 : The Yeast Pif1 Helicase Prevents Genomic Instability Caused by G-Quadruplex-Forming CEB1 Sequences *In Vivo*

#### II.A.1. Contexte

Les évidences *in vitro* de la formation d'un G-quadruplexe par le motif de CEB1 et la capacité de Pif1 à dérouler cette structure a conduit à l'hypothèse que les G-quadruplexes formés dans CEB1 et non-déroulés en absence de Pif1 étaient responsable de son instabilité *in vivo*. L'objectif de mon stage de 2ème année de master a été de tester cette hypothèse, par mutagenèse dirigée des G impliquées dans la formation du G-quadruplexe. Mon travail a donc consisté, en collaboration avec J. Lopes, à synthétiser *in vitro* des minisatellites artificiels et à mesurer leur instabilité dans des souches WT et *pif1Δ*.



# The Yeast Pif1 Helicase Prevents Genomic Instability Caused by G-Quadruplex-Forming CEB1 Sequences In Vivo

Cyril Ribeyre<sup>1,9</sup>, Judith Lopes<sup>1,9</sup>, Jean-Baptiste Boulé<sup>1,2</sup>, Aurèle Piazza<sup>1</sup>, Aurore Guédin<sup>3</sup>, Virginia A. Zakian<sup>2</sup>, Jean-Louis Mergny<sup>3</sup>, Alain Nicolas<sup>1\*</sup>

**1** Recombinaison et Instabilité Génétique, Institut Curie Centre de Recherche, CNRS UMR3244, Université Pierre et Marie Curie, Paris, France, **2** Department of Molecular Biology, Princeton University, Princeton, New Jersey, United States of America, **3** Laboratoire de Biophysique, Museum National d'Histoire Naturelle USM 503, INSERM U565, CNRS UMR5153, Paris, France

## Abstract

In budding yeast, the Pif1 DNA helicase is involved in the maintenance of both nuclear and mitochondrial genomes, but its role in these processes is still poorly understood. Here, we provide evidence for a new Pif1 function by demonstrating that its absence promotes genetic instability of alleles of the G-rich human minisatellite CEB1 inserted in the *Saccharomyces cerevisiae* genome, but not of other tandem repeats. Inactivation of other DNA helicases, including Sgs1, had no effect on CEB1 stability. In vitro, we show that CEB1 repeats formed stable G-quadruplex (G4) secondary structures and the Pif1 protein unwinds these structures more efficiently than regular B-DNA. Finally, synthetic CEB1 arrays in which we mutated the potential G4-forming sequences were no longer destabilized in *pif1Δ* cells. Hence, we conclude that CEB1 instability in *pif1Δ* cells depends on the potential to form G-quadruplex structures, suggesting that Pif1 could play a role in the metabolism of G4-forming sequences.

**Citation:** Ribeyre C, Lopes J, Boulé J-B, Piazza A, Guédin A, et al. (2009) The Yeast Pif1 Helicase Prevents Genomic Instability Caused by G-Quadruplex-Forming CEB1 Sequences In Vivo. *PLoS Genet* 5(5): e1000475. doi:10.1371/journal.pgen.1000475

**Editor:** Orna Cohen-Fix, National Institute of Diabetes and Digestive and Kidney Diseases, United States of America

**Received:** January 12, 2009; **Accepted:** April 8, 2009; **Published:** May 8, 2009

**Copyright:** © 2009 Ribeyre et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This study was supported by La Ligue Nationale contre le Cancer "Equipe Labellisée LIGUE 2007" and in part by an E.U. FP6 "MolCancerMed" (LSHC-CT-2004-502943) grant. CR was supported by a graduate student fellowship from the MNERT and the Association pour la Recherche contre le Cancer (ARC). JBB was supported by a post-doctoral fellowship from the ARC. AP was supported by a graduate student fellowship from the MNERT. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: alain.nicolas@curie.fr

✉ Current address: Department of Molecular Biology, University of Geneva, Geneva, Switzerland

9 These authors contributed equally to this work.

## Introduction

At the chromosomal level, in addition to coding regions and epigenetic modifications, the biological information also resides in DNA secondary structures, but this layer remains to be further deciphered. Biophysical and structural studies have long established that in vitro DNA can adopt diverse structures different from the canonical Watson-Crick conformations [1]. However, for a long time, the hypothesis that these structures occur in the native chromosomal context, as an integral part of the functional architecture of a chromosome, has been regarded with a certain skepticism. One example of such a non canonical DNA structure is the G-quadruplex, also named G-tetraplex or G4 DNA. These structures form in vitro in guanine-rich sequences that contain four tracts of at least three guanines separated by other bases, and are stabilized by G-quartets that form between four DNA strands [2]. Under physiological conditions, long runs of G4-forming sequences promote the formation of highly stable structures that can form spontaneously in vitro and, once formed, are very resistant to thermal denaturation. It is also important to consider that sequences that form G4-DNA slowly in vitro may be more prone to fold in vivo owing to the action of proteins that promote and/or

stabilize their formation, such as the beta subunit of the ciliate *Oxytricha* telomere binding protein complex [3,4].

Evidence for in vivo formation of G4 DNA has emerged in recent years. Notably, G4 DNA has been observed by electron microscopy from transcribed human G-rich DNA arrays in bacteria [5] and has been detected at the end of the ciliate *Oxytricha* telomeres by immunochemistry [6,7]. As a complementary approach, genome-wide bioinformatic analyses have identified regions that have the potential to form G4 DNA within evolutionary diverse model systems, from bacteria to human. For example, in the human genome, more than 300,000 distinct sites have the potential to form G4 DNA [8,9]. These sequences are highly over-represented in the promoter regions of diverse organisms, including human [10], yeast [11] and bacteria [12]. In addition, potential G4-forming sequences are found in G-rich arrays such as telomeres, rDNA or G-rich micro- and minisatellites. Hence, it has been suggested that their presence might affect transcriptional or post-transcriptional events when the G4 forming sequence is within the transcribed region [11,13]. G4 DNA has also been proposed to participate in telomere capping, DNA replication and recombination [14]. However, it remains to be determined how and to what extent these secondary structures



## Author Summary

Changes in the primary DNA sequence are a major source of pathologies and cancers. The hereditary information also resides in secondary DNA structures, a layer of genetic information that remains poorly understood. Biophysical and structural studies have long established that, in vitro, the DNA molecule can adopt diverse structures different from the canonical Watson-Crick conformations. However, for a long time their existence in vivo has been regarded with a certain skepticism and their functional role elusive. One example is the G-quadruplex structure, which involves G-quartets that form between four DNA strands. Here, using in vitro and in vivo assays in the yeast *S. cerevisiae*, we reveal the unexpected role of the Pif1 helicase in maintaining the stability of the human CEB1 G-rich tandem repeat array. By site-directed mutagenesis, we show that the genomic instability of CEB1 repeats in absence of Pif1 and is directly dependent on the ability of CEB1 to form G-quadruplex structures. We show that Pif1 is very efficient in vitro in processing G-quadruplex structures formed by CEB1. We propose that Pif1 maintains CEB1 repeats by its ability to resolve G-quadruplex structures, thus providing circumstantial evidence of their formation in vivo.

affect these processes and how they are maintained through DNA replication despite causing a structural impairment to the various nucleic acid processing enzymes.

It is clear that DNA goes through a single strand configuration locally during processes like DNA replication, transcription or repair, and many models argue that this single stranded stage favors G4 DNA formation [14]. In vitro, several DNA helicases, such as the human BLM, WRN, FANCD1 and the *S. cerevisiae* Sgs1, can unwind G4 structures. They preferentially unwind G4 DNA over partially duplex DNA, forked DNA or Holliday junction substrates, and their helicase activity is inhibited in presence of G4 DNA ligands [15–18]. In *Caenorhabditis elegans* the FANCD1 homolog *dog-1* is involved in the maintenance of G-rich regions by preventing intrinsic instability and loss of these regions [19,20]. However, considering that different G-rich sequences can adopt very diverse secondary structures, and that in numerous instances genes encoding helicases are not essential, the questions of how many and which class of helicases are indeed able to process efficiently these secondary structures formed in guanine-rich regions in a given organism remains to be addressed. Also, until now, very few in vivo systems exist to study the involvement of helicases in processing these structures and assay artificially designed variant substrates.

In the present study, which was aimed at characterizing the mechanism(s) of rearrangement of tandem DNA repeats, we uncover an unexpected function of the Pif1 helicase with regards to processing G4 structures. Pif1 is a member of a conserved family of 5'-3' DNA helicases, with distant homology to the RecD bacterial helicase. The *S. cerevisiae* Pif1 protein is important both for maintenance of mitochondrial DNA [21,22] and as a negative regulator of telomerase-mediated telomere lengthening [23,24]. Here we report that Pif1 also affects stability of the G-rich CEB1 minisatellite when it is inserted into a yeast chromosome. In contrast, mutations in other helicases, including the *S. cerevisiae* RecQ homologue Sgs1, had no effect on CEB1 stability. In vitro, CEB1 formed G4 structures that were efficiently unwound by Pif1. Finally, mutation of the CEB1 repeats such that they were no longer able to form G4 structures made them insensitive to Pif1. Thus we demonstrated that one of the functions of the Pif1

helicase is to process G4 structures. As sequences with the ability to form G4 DNA are found throughout the yeast genome, beyond acting on intrinsically unstable repeats, we propose that the processing of G4 structures by Pif1 may facilitate DNA replication, transcription and/or repair.

## Results

### The DNA Helicase Pif1 Actively Destabilizes CEB1 during Vegetative Growth

We previously developed yeast strains to study the genetic instability of a natural 1.8 kb allele of the human minisatellite CEB1 inserted in the *S. cerevisiae* genome (Figure 1A). This allele (called CEB1-1.8) is composed of a tandem array of 42 polymorphic repeats of sizes varying between 36 and 43 base pairs (bp) [25] (Figure S1). In our standard assay, which measures the frequency of allele size variation after growth for seven generations at 30°C, approximately 0.3% of wild-type (WT) cells exhibit a change in CEB1 size (contractions and expansions). Using this system, we reported that CEB1-1.8 was strongly destabilized in the absence of the Rad27/FEN1 endonuclease (42% instability) [26].

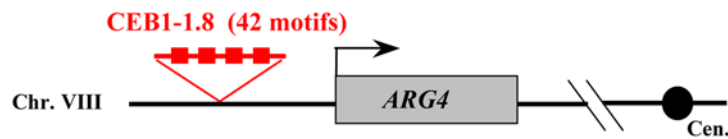
Recently, it was reported that the lethality caused by inactivation of the essential helicase/endonuclease Dna2, which participates with Rad27 in the maturation of Okazaki fragments, could be rescued by inactivation of the DNA helicase Pif1 [27]. These results prompted us to test if Pif1 also had an effect on the maintenance of CEB1 arrays in our system. Remarkably, in the absence of Pif1 (*pif1Δ*), the frequency of rearrangement by contractions or expansions of the parental allele increased 20-fold compared to WT cells (6% instability; Table 1, Figure 1B). As a control, a *pif1Δ* CEB1-1.8 strain containing a multicopy plasmid that expressed the WT *PIF1* gene under the control of the *PIF1* promoter did not exhibit CEB1 instability. Together, these results demonstrate that the absence of Pif1 destabilizes the CEB1-1.8 minisatellite at a rate of ~1% per cell per generation. CEB1 instability was not specific to tracts inserted at the *ARG4* locus as CEB1-1.8 inserted at the *ADP1* locus in chromosome III was stable in the presence of Pif1 but was rearranged in its absence (3.6% instability; 7/192). The difference in stability between the two chromosomal locations is not statistically significant (Fisher's Exact test,  $p = 0.28$ ).

To determine if the helicase activity of Pif1 was required to stabilize the CEB1-1.8 allele, we examined the stability of CEB1-1.8 in strains carrying the *pif1-K264A* or *pif1-K264R* mutations, which inactivate Pif1 ATPase/helicase activity [28]. In both mutants, the frequency of CEB1 rearrangement was increased approximately 10-fold over the WT level (3.2%; Table 1, Figure 1B). Thus, the helicase activity of Pif1 has a role in the stabilization of the CEB1 repeats during vegetative growth. Compared to the *pif1Δ* mutant, the frequency of size variants was approximately two-fold lower in both of the helicase-inactive mutants. This suggests that while ATPase/helicase activity is totally inactive in helicase-dead *pif1-K264A* mutant (see below, Figure 2F), the *pif1-K264A* polypeptide which retains wild type level of DNA binding [24], may act within a complex of proteins sufficient to partially protect CEB1 repeats from damage or recombinational repair.

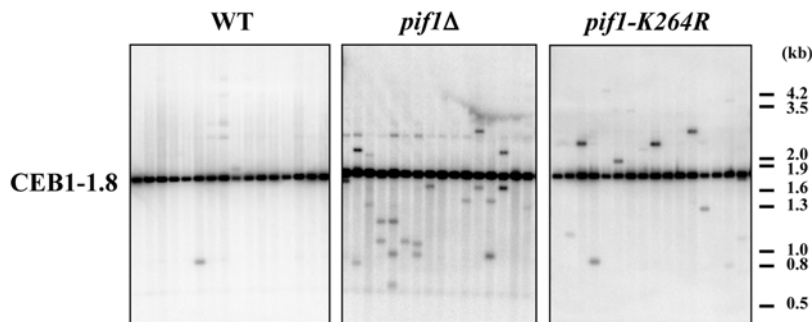
### CEB1-1.8 Rearrangements in *pif1Δ* Cells Are Often Complex and Depend on the Rad51- and Rad52-Dependent Homologous Recombination Pathway

To characterize the internal structures of CEB1-1.8 rearrangements obtained in the *pif1Δ* cells, we sequenced nine CEB1 contractions and compared them to the parental motif. As shown in Figure 1C and Figure S1, the sequenced contractions from

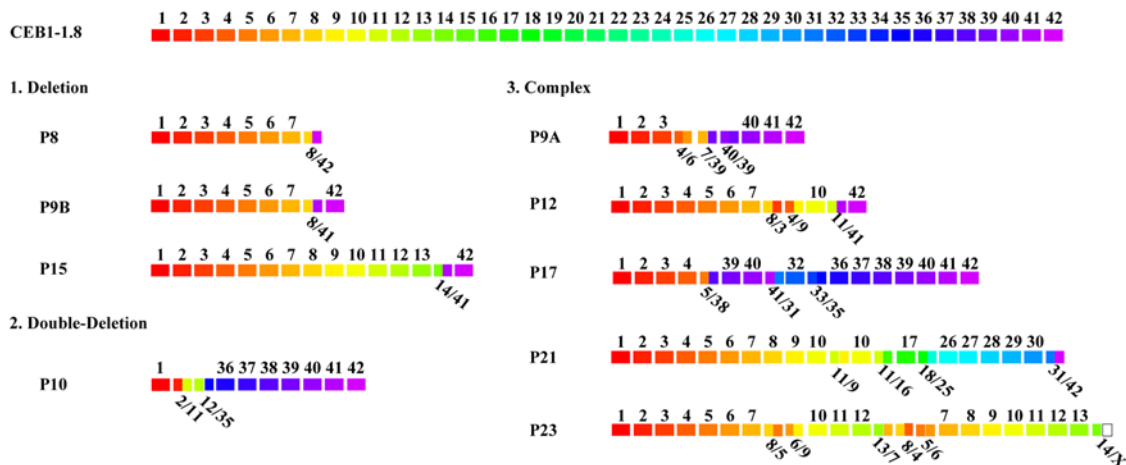
A



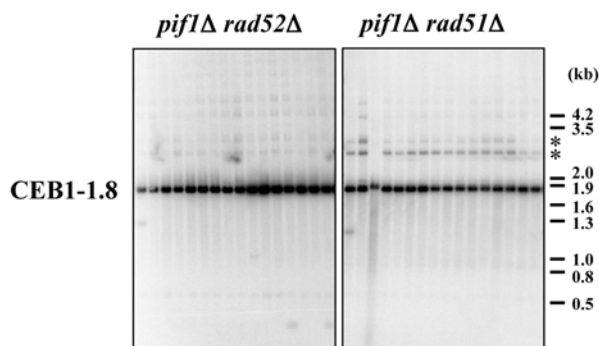
B



C



D



**Figure 1. The CEB1 minisatellite is unstable in *pif1Δ* deficient cells.** (A) Structure of the genomic locus containing CEB1-1.8. (B) Southern blot analysis of CEB1-1.8 instability in haploid strains: WT (ORT2914), *pif1Δ* (ORT4843), *pif1-K264R* (ORT5083-4E). Each lane contains DNA extracted from pools of 12 independent colonies digested by *AluI* and hybridized with a CEB1-0.6 probe. (C) Structure of CEB1-1.8 rearrangements obtained in *pif1Δ* haploids. Each of the 42 CEB1-1.8 repeats is represented by a colored box and numbered (top). Nine rearrangements were sequenced and classified in three categories (1 to 3). The name of each rearranged allele is at the left. Hybrid repeats are represented by the two colors corresponding to the fused repeats. The white box in P23 indicates a motif that cannot be attributed to a specific parental motif. (D) CEB1 instability is Rad52 and Rad51 dependent. Southern blot analysis of CEB1-1.8 instability in *pif1Δ rad52Δ* (ORD7565-2C) and *pif1Δ rad51Δ* (ORD7574-9B) haploid strains. Same legends as in (B). Additional bands marked with an asterisk are presumably due to partial digestion by *AluI*. doi:10.1371/journal.pgen.1000475.g001

**Table 1.** Instability of CEB1-1.8 in haploid strains.

Strain	Genotype	Number of rearrangements/total (%)	Fold increase vs. WT	p value vs. WT*
ORT2914	WT	5/1824 (0.3)	1	-
ORT4841	<i>pif1Δ</i>	40/672 (6.0)	20	<0.01
ORT4843	<i>pif1Δ</i>	11/192 (5.8)	19	<0.01
ORD7569	<i>pif1Δ/pif1Δ</i>	12/192 (6.3)	21	<0.01
ORT5083-4E	<i>pif1-K264R</i>	18/576 (3.2)	10	<0.01
ORT5087-5E	<i>pif1-K264A</i>	12/384 (3.2)	10	<0.01
ORT5084-2C	<i>pif1-m1</i>	1/192 (0.5)	2	NS
ORT5085-1C	<i>pif1-m2</i>	1/384 (0.3)	1	NS
ORT4848	<i>pif1Δ dna2Δ</i>	18/384 (4.7)	16	<0.01
ORT4880	<i>rrm3Δ</i>	0/336 (0)	<1	NS
ORD9304-9A	<i>pif1Δ rrm3Δ</i>	15/384 (3.9)	13	<0.01
ORT4849	<i>sgs1Δ</i>	0/192 (0)	<1	NS
ORD9922-4B	<i>pif1Δ sgs1Δ</i>	21/363 (5.8)	19	<0.01
ORT4885	<i>mph1Δ</i>	0/336 (0)	<1	NS
ORT4840	<i>srs2Δ</i>	2/276 (0.7)	2.3	NS
ORD6786-4A	<i>rif1Δ</i>	0/368 (0)	<1	NS
ORD7565-2C	<i>pif1Δ rad52Δ</i>	2/384 (0.5)	2	NS
ORD7574-9B	<i>pif1Δ rad51Δ</i>	1/384 (0.3)	1	NS
ORD7574-11C	<i>rad51Δ</i>	0/272 (0)	<1	NS

\*Fisher-test.

NS: non significant.

doi:10.1371/journal.pgen.1000475.t001

*pif1Δ* cells were all different from each other. Three were simple deletions, one was a double deletion and five were complex events.

To determine whether or not the destabilization of CEB1-1.8 in *pif1Δ* cells was dependent on homologous recombination, we tested the stability of CEB1-1.8 in *pif1Δ rad52Δ* and *pif1Δ rad51Δ* double-mutants. In both strains, rearrangement of CEB1-1.8 occurred at close to WT levels, strongly reduced compared to *pif1Δ* cells (Figure 1D and Table 1). We conclude that the molecular events leading to CEB1 rearrangement are repaired by homologous recombination, similar to what is seen in the absence of Rad27 [25].

### CEB1 Destabilization Is Not a Secondary Effect of Telomere or Mitochondrial Defect in *pif1Δ* Cells

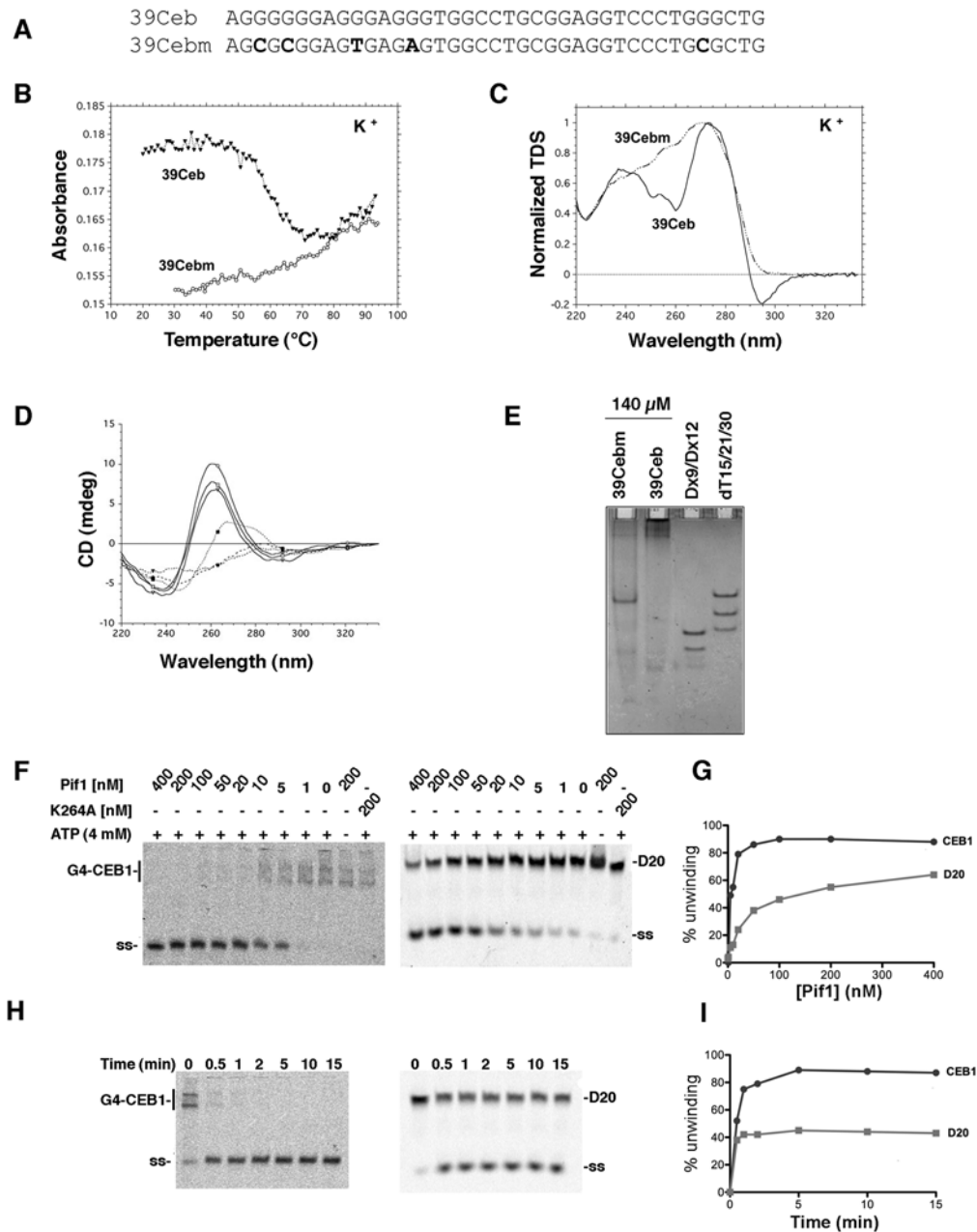
To determine if the effects of *pif1Δ* on CEB1 stability are a secondary consequence of the increased telomere length or mitochondrial DNA depletion that are characteristic of *pif1Δ* cells, we examined CEB1-1.8 stability in mutants that affect either telomere length or maintenance of mitochondrial DNA. The deletion of the *RIF1* gene results in telomere lengthening [29], a phenotype likely due to the enhanced access of telomerase to the telomere [30]. *RIF1* inactivation did not destabilize CEB1-1.8 (Table 1), indicating that long telomeres are not sufficient to destabilize CEB1-1.8 repeats.

Pif1 is present as two isoforms, one targeted to the nucleus and one to mitochondria. The *pif1-m1* mutation prevents the synthesis of the mitochondrial isoform, resulting in mitochondrial deficiency but leaving nuclear Pif1 functions intact. In *pif1-m2* cells, only the mitochondrial form is detected by western analysis [28], and this strain has normal mitochondrial function and long telomeres. However, telomere lengthening and de novo telomere addition are not as elevated in *pif1-m2* cells as in a *pif1Δ* strain suggesting that some nuclear function is retained in the *pif1-m2* allele [23]. As

expected, CEB1-1.8 was not destabilized (1/192) in *pif1-m1* cells (Table 1). Surprisingly, CEB1 was also stable in *pif1-m2* cells (1/384) (Table 1), a result that can be explained if *pif1-m2* cells retain sufficient nuclear Pif1 to carry out its role in maintaining CEB1 stability. To test if a low level of the Pif1-m2 polypeptide could be active in the nucleus, we examined complementation of the *pif1-m2* telomere phenotype by over expressing the *pif1-m2* protein from its own promoter on a multi-copy 2 μ plasmid in *pif1Δ* cells. Telomeres were shorter in the strain over-expressing the *pif1-m2* construct than in the control *pif1Δ* cells (data not shown). These results support our interpretation that in *pif1-m2* cells, there is sufficient nuclear Pif1 protein to stabilize CEB1, although it is insufficient to sustain normal length telomeres. A similar observation was recently reported in the fission yeast *S. pombe*. As in budding yeast, the Pif1 homolog Pfh1p is present as a mitochondrial and a nuclear isoforms. However, expression of the mitochondrial-only isoform is able to complement pfh1p nuclear defects, even though the protein is not detectable in the nucleus at the protein level by western blot [31].

### CEB1-1.8 Is Not Destabilized by Mutations in Other Helicases

We investigated if the inactivation of other helicases would also affect CEB1-1.8 stability. We previously showed that in a *dna2-1* strain, CEB1-1.8 was modestly destabilized (1.8% instability) [26]. The viability of the *DNA2* deletion in combination with the deletion of *PIF1* [27] allowed us to examine the behavior of CEB1 in the complete absence of *DNA2*. As indicated in Table 1, in the *pif1Δ dna2Δ* CEB1-1.8 strain, the frequency of CEB1 size variation was estimated at 4.7%, a value significantly higher than in wild-type cells ( $p < 0.01$ , Fisher's Exact Test), but not different than in the *pif1Δ* single mutant ( $p = 0.48$ , Fisher's Exact Test). This result



**Figure 2. Evidence for G4 structures formation by CEB1 minisatellite sequences.** (A) Oligonucleotide sequences. The 39Ceb oligonucleotide mimicks one full CEB1 repeat. 39Cebm is a control sequence with five base substitutions (shown in bold) (B) Melting profiles. Absorbance at 295 nm vs temperature plots for 39Ceb (triangles) and 39Cebm (circles) each at 3  $\mu$ M strand concentration. Melting experiments are performed in 10 mM lithium pH 7.2 cacodylate buffer supplemented with 0.1 M KCl. (C) Thermal difference spectra. Thermal difference spectra result from the difference between the absorbance recorded at  $79 \pm 2^\circ\text{C}$  and at  $40 \pm 2^\circ\text{C}$  in a 10 mM lithium pH 7.2 cacodylate buffer supplemented with 0.1 M KCl. Thermal difference spectra are normalized ( $\text{TDS}_{\text{norm}} = \text{TDS}/\text{max}(\text{TDS})$ ) over the 220–335 nm wavelength range. Full line: 39Ceb; dotted line: 39Cebm. (D) Circular dichroism spectra. Oligonucleotides were prepared at 140  $\mu$ M strand concentration and annealed in 1 M NaCl as in helicase experiments, then immediately diluted to 3  $\mu$ M in a 10 mM lithium cacodylate 1 M NaCl, pH 7.2 buffer. Full line: 39Ceb; dotted line: 39Cebm. Spectra were recorded at three different temperatures:  $25^\circ\text{C}$  (squares),  $65^\circ\text{C}$  (circles) and  $90^\circ\text{C}$  (triangles). (E) Behavior of the 39Ceb and 39Cebm sequences on a non-denaturing gel. Oligonucleotides were prepared at 140  $\mu$ M strand concentration, annealed in 1 M NaCl as in helicase experiments and loaded on a non-denaturing 15% acrylamide gel supplemented with 20 mM NaCl and run at  $21^\circ\text{C}$ . Migration markers are 1: double-stranded DNA (9 and 12 bp) and 2: (dT)<sub>15</sub>, (dT)<sub>21</sub> and (dT)<sub>30</sub> oligomers. (F) In vitro unwinding of 2 nM G-quadruplex DNA (G4-CEB1, left) versus 2 nM double stranded DNA oligonucleotide substrate (D20, right) in presence of decreasing amount of Pif1 or helicase-dead *pif1-K264A* for 15 minutes at  $35^\circ\text{C}$  in presence or absence of ATP. (G) Quantifications of the gels shown in F. (H) Kinetics of unwinding of G-quadruplex DNA (left) versus double stranded DNA oligonucleotide D20 (right) in presence of 100 nM Pif1. (I) Quantifications of the gels shown in H. doi:10.1371/journal.pgen.1000475.g002

indicates that the complete absence of Dna2 neither suppresses nor enhances the effects of Pif1 inactivation.

Next, we examined the inactivation of Rrm3, a 5'-3' DNA helicase that is closely related to Pif1 [32]. As shown in Table 1, deletion of the *RRM3* gene did not destabilize CEB1-1.8. Moreover, the frequency of rearrangement of CEB1 was not statistically different in the *pif1Δ rrm3Δ* (3.9%) and *pif1Δ* (6.0%) cells ( $p = 0.2$ , Fisher's Exact Test).

We tested three additional helicases with well characterized roles in genome stability for the effects on CEB1-1.8 stability. We examined the RecQ homolog Sgs1 helicase involved in multiple aspects of DNA recombination and repair [33–38], Srs2, a 3' to 5' helicase that disassembles abortive recombination intermediates [39], and the Mph1 helicase that plays a role in DNA repair [40]. Inactivation of these helicases did not destabilize the CEB1-1.8 array, and the inactivation of both Pif1 and Sgs1 helicases (*pif1Δ sgs1Δ* strain), induced the same CEB1 instability as the *pif1Δ* strain (Table 1). We conclude that the role of Pif1 in

stabilizing CEB1-1.8 is specific for Pif1, rather than a general function of DNA helicases involved in DNA repair or recombination.

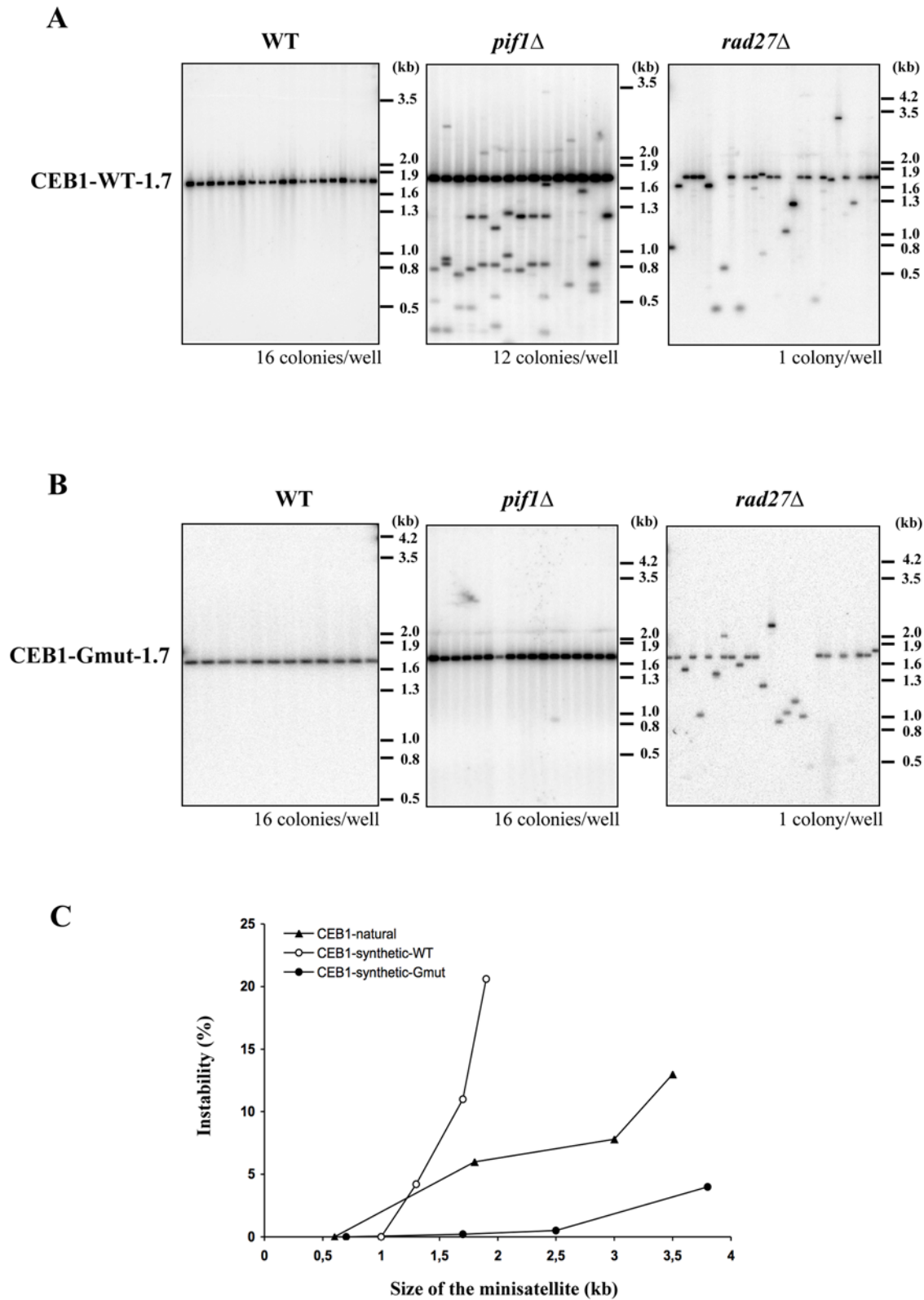
### All Tandem Repeated Sequences Are Destabilized in *rad27Δ* cells But Only CEB1 Is Destabilized in the Absence of Pif1

We examined CEB1 alleles of various sizes, a shorter allele CEB1-0.6 (14 repeats) and two longer alleles, CEB1-3.0 (65 repeats) and CEB1-3.5 (75 repeats). The two longer alleles were destabilized in *pif1Δ* cells, with instability increasing with the size of the array (Table 2, Figure 3C). For comparison, we performed similar studies in the *rad27Δ* cells. In all cases CEB1 rearrangements occurred at a lower frequency in the *pif1Δ* cells than in *rad27Δ* cells [25]. In the case of CEB1-1.8, for which the largest sample of cells was examined, its instability was approximately 5-fold higher in *rad27Δ* than in *pif1Δ* cells.

**Table 2.** Instability of various tandem repeated sequences in *pif1Δ* and *rad27Δ* cells.

Name of the tandemly repeated sequence	Size of the motif (bp)	Number of repeats	Strain	Genotype	Number of rearrangements/total (%)
CEB1-0.6	39–42	14	ORD7557-10B	<i>pif1Δ</i>	0/276 (0)
CEB1-1.8	39–42	42	ORT4841	<i>pif1Δ</i>	40/672 (6.0)
CEB1-3.0	39–42	65	ORD7598-12C	<i>pif1Δ</i>	15/192 (7.8)
CEB1-3.5	39–42	75	ORT4841-4E1	<i>pif1Δ</i>	25/192 (13.0)
<i>DAN4</i>	18	30	ORD7568	<i>pif1Δ/pif1Δ*</i>	0/192 (0)
			ORD6708	<i>rad27Δ/rad27Δ*</i>	3/52 (5.7)
<i>FLO1</i>	135	17	ORT4843	<i>pif1Δ</i>	0/192 (0)
			ORD6713-8D	<i>rad27Δ</i>	8/304 (2.7)
<i>HKR1</i>	42	21	ORT4841	<i>pif1Δ</i>	0/384 (0)
			ORD6713-8D	<i>rad27Δ</i>	11/158 (7)
<i>NUM1</i>	192	10	ORT4841	<i>pif1Δ</i>	0/384 (0)
			ORD6713-8D	<i>rad27Δ</i>	5/304 (1.6)
<i>hRAS1</i>	28	75	AND1228-2A	<i>pif1Δ</i>	0/384 (0)
			AND1228-8C	<i>rad27Δ</i>	28/55 (51)
Name of the plasmid with tandemly repeated sequence (motif)	Size of the motif (bp)	Number of repeats	Strain	Genotype	Fold increase vs. WT (frequency $10^{-5}$ )
pMD28 (G)	1	18	ORT5604	WT	1 (0.2)
			ORT5600	<i>pif1Δ</i>	2
			ORT6009	<i>rad27Δ</i>	86400
pBK1 (GAGT)	4	16	ORT5606	WT	1 (0.4)
			ORT4896	<i>pif1Δ</i>	1.5
			ORT6014	<i>rad27Δ</i>	31200
pBK3 (CAACG)	5	15	ORT5614	WT	1 (1.8)
			ORT4894	<i>pif1Δ</i>	1
			ORT6013	<i>rad27Δ</i>	7730
pBK10 (CAATCGGT)	8	10	ORT5602	WT	1 (0.7)
			ORT4892	<i>pif1Δ</i>	1
			ORT6011	<i>rad27Δ</i>	17250
pEAS20 (CAACGCAATGCGTTGGATCT)	20	3	ORT5608	WT	1 (0.8)
			ORT4898	<i>pif1Δ</i>	3
			ORT6016	<i>rad27Δ</i>	17800

\*The repeated sequence is homozygous.  
doi:10.1371/journal.pgen.1000475.t002



**Figure 3. CEB1 instability in *pif1* $\Delta$  cells depends on its potential to form G4 DNA.** Comparison of synthetic-CEB1-WT-1.7 (A) and synthetic-CEB1-Gmut-1.7 (B) instability in WT, *pif1* $\Delta$  and *rad27* $\Delta$  haploid strains by Southern blot analysis. In order to increase the number of independent colonies analyzed in *pif1* $\Delta$  and WT strains, colonies are pooled for DNA extraction and the number of colonies analyzed per well is indicated under each gel. DNA is digested by *ApaI/SpeI* and hybridized with a CEB1-WT or CEB1-Gmut probe. When several rearranged minisatellites migrate at the same size they are considered as clonal and are counted only one time. The frequency of instability for each synthetic minisatellite is reported in Table 3. (C) Frequency of natural and synthetic CEB1 minisatellites according to the size of the alleles in *pif1* $\Delta$  cells. doi:10.1371/journal.pgen.1000475.g003

Next, we examined the instability of four natural yeast minisatellites that are normally found in the coding regions of the *DAN4*, *FLO1*, *HKR1* and *NUM1* genes [41]. This set represents a large variety of motifs in term of size (18 to 192 bp) and repeat units (10–30). All of these motifs were altered in *rad27Δ* but not in *pif1Δ* cells (Table 2). Likewise, the GC-rich hRAS1 human minisatellite [42] was not altered when propagated in *pif1Δ* cells (0/384 colonies).

Finally, using the plasmid assay developed by Kokoska *et al.* (1998), we compared the behavior of four microsatellite sequences composed of 1, 4, 5 and 8 nucleotide motifs and a triplication of a 20 nucleotides motif in wild-type, *pif1Δ* and *rad27Δ* haploid cells (see Table 2 for sequence of motifs). As previously reported [43], the rearrangement frequencies in the wild-type strain were on the order of  $10^{-5}$ – $10^{-6}$  and were stimulated more than 10,000 fold in *rad27Δ* cells (Table 2). However, no significant increase in instability was detected in *pif1Δ* cells. Thus, in contrast to the strong and ubiquitous effects of Rad27 on minisatellite and microsatellite stability [41,43,44], the absence of Pif1 destabilized only the CEB1 arrays.

### The CEB1 Repeat Forms G4 Structures In Vitro

DNA oligonucleotides containing at least four successive runs of three or more guanines have been shown to fold into intramolecular G4 DNA in presence of physiological concentrations of monovalent cations [45]. Examination of the CEB1 repeat sequence revealed the presence of 3 to 5 triplets of guanines localized on the same strand in each repeat of the CEB1-1.8 allele (Figure 2A and Figure S1). It suggests that this minisatellite may form G4 structures, even if its primary sequence does not fit perfectly the  $d(G_{3+n}N_{1-7})_4$  consensus used for most bioinformatic analyses. To test this hypothesis, we examined in vitro the formation of secondary structures using a single-stranded oligonucleotide that mimicked a complete CEB1 repeat (39Ceb) or a control sequence in which five of the guanines had been mutated (39Cebm) (Figure 2A). Four complementary assays were performed to detect the formation of G4 structures:

First, 39Ceb and 39Cebm oligos were incubated in presence of 100 mM NaCl or KCl in conditions that favor G4 DNA formation. We measured the absorbance at 295 nm of 39Ceb and 39Cebm oligos at increasing temperatures. Indeed, an inverted transition corresponding to a conformational change associated with the temperature increase was observed with the 39Ceb oligo at a melting temperature ( $T_m$ ) of  $\approx 48^\circ\text{C}$  in NaCl and  $55^\circ\text{C}$  in KCl, while no clear transition was seen with the 39Cebm sequence (Figure 2B and Table S2). Truncated versions of this motif were also analyzed (Table S2). Second, thermal differential spectra (TDS), which measure the difference between UV absorbance spectra of the oligonucleotide measured at a temperature above  $T_m$  (unfolded state) and below  $T_m$  (folded state), provides a clear signature for each type of nucleic acid structures including G4 DNA [46]. We measured the TDS in  $K^+$  buffer for 39Ceb and 39Cebm. As shown in Figure 2C, 39Ceb exhibits the typical pattern of a G4 structure with two positive maxima at 240 and 275 nm and a negative minimum around 295 nm [46–48] while 39Cebm exhibited a different signature, which does not correspond to quadruplexes.

Third, we measured the circular dichroism (CD) spectra of the two oligonucleotides under experimental conditions that mimic the helicase assays (see below; briefly oligonucleotides were incubated at 140  $\mu\text{M}$  strand concentration for 48 hours in 1 M NaCl). A positive maxima around 260 nm and a negative minimum around 240 nm was observed in the CD spectra of 39Ceb, an observation in agreement with the formation of parallel G4 structures (Figure 2D) [49,50]. In contrast 39Cebm did not

exhibit a CD spectra characteristic of any G4 structure found so far. Furthermore, when prepared under these conditions, the quadruplexes were extremely stable, as shown by temperature-independent CD profiles between  $25^\circ\text{C}$  and  $90^\circ\text{C}$ . This demonstrates that these structures are extremely heat resistant (no melting transition was observed by absorbance at 295 nm when the sample was prepared with this protocol; data not shown).

Finally, 39Ceb and 39Cebm oligonucleotides were analyzed by polyacrylamide gel electrophoresis under native conditions where G4 structures are expected to show different mobility compared to unstructured oligonucleotides. No migration anomaly was found for 39Ceb when incubated in 100 mM LiCl, which does not stabilize G4 secondary structures [51] (data not shown). When 39Ceb is incubated in a sodium buffer at high strand concentration (Figure 2E; conditions identical as for helicase experiments, see below), bands of very low mobility were clearly visible. Intermolecular G4 structure formation was revealed by slow migrating bands as compared to the migration pattern of 39Cebm mutated control (Figure 2E). These higher order species likely correspond to bimolecular, tetramolecular (or higher) G4 structures. These experiments were repeated at lower strand concentration (50 nM or 4  $\mu\text{M}$ ), both in sodium and potassium. As expected for multimers (dimers, tetramers or species of even higher stoichiometry), concentration-dependent profiles were obtained (Figure S2).

In conclusion, in all assays, the oligonucleotides containing the G-strand of the CEB1 motif exhibited the hallmarks of G4 structure formation in vitro whereas the 39Cebm control sequence did not. Depending on buffer conditions, strand concentration and incubation protocol, a variety of different quadruplex structures could be obtained with this sequence, arguing for the possible formation of multiple quadruplexes in vivo.

### Pif1 Protein Unwinds G4 CEB1 DNA In Vitro

If CEB1 also forms G4 DNA in vivo, Pif1 might inhibit CEB1 rearrangements by unwinding these structures. The prediction of this model is that Pif1 should be able to unwind these structures. To test this prediction oligonucleotides containing one CEB1 repeat were incubated in vitro using conditions that favor the formation of intermolecular G4 structures (see Materials and Methods). The G4-DNA substrate was first incubated in the presence of decreasing amount of purified recombinant Pif1. Upon 15 minutes incubation at  $35^\circ\text{C}$ , 5 nM Pif1 was enough to unwind 50% of the 20 fmol (2 nM) G4-DNA, while at least 20 times more Pif1 was necessary to unwind 20 fmol (2 nM) of a double-stranded oligonucleotide substrate (Figure 2F, G). The unwinding of both substrates required Pif1 helicase activity as no unwinding is observed in absence of ATP, or when the substrate is incubated in presence of saturating amount of the *pif1-K264A* helicase-dead mutant (Figure 2F). The rate of G4-DNA unwinding was also faster than unwinding of the double-stranded DNA substrate (Figure 2H, I). Indeed, 100 nM Pif1 was able to unwind 20 fmol (2 nM) of G4-DNA substrate in less than 5 minutes, while the enzyme was only able to unwind about 40% of the double-stranded substrate over the entire time course. These results demonstrate that Pif1 is more efficient at unwinding G4-DNA structures than regular double-stranded DNA.

### Synthetic CEB1 Alleles Without G4 Prone Sequence Are Stable in *pif1Δ* Cells

The in vitro experiments demonstrating the propensity of the CEB1 repeat to form G4 structures and the ability of Pif1 to unwind these structures led us to consider that Pif1 might unwind G4 structures in CEB1 in vivo. If this model is correct, mutations in

CEB1 that eliminate its ability to form G4 structures might render it insensitive to Pif1. For these experiments, we developed a method combining both in vitro and in vivo steps to construct long (>1 kb) synthetic CEB1 alleles (see Text S1). We generated two categories of synthetic CEB1 arrays based on two different repeat units. The first category, named synthetic-CEB1-WT, was based on the repetition of the most common motif of the natural polymorphic CEB1-1.8 allele (Figure S3, A, D). The second category, named CEB1-Gmut, was made from oligonucleotides in which 5 dispersed G bases were changed to either C, A or T in order to disrupt the original 5 G-triplets on the G-rich strand (Figure S3, A, E). In vitro analysis of the secondary structures of CEB1-Gmut oligonucleotides demonstrated that, as expected, they were unable to form G4 structures (39Cebm, Figure 2 and Table S2).

The rearrangement frequency of the synthetic-CEB1-WT arrays (1.0, 1.3, 1.7, 1.9 and 2.3 kb long) and of the synthetic-CEB1-Gmut arrays (0.7, 1.7, 2.5 and 3.8 kb long) in WT, *pif1Δ* and *rad27Δ* cells is reported in Table 3 and summarized in Figure 3. As observed for the natural CEB1 alleles, the rearrangement frequency of the synthetic-CEB1-WT arrays was low in WT cells and increased in a size dependent manner in both *pif1Δ* and *rad27Δ* cells. In all cases, the frequency of instability for similarly sized alleles was higher in the synthetic-CEB1-WT arrays than in the natural CEB1 alleles. We attribute this difference to the greatly reduced polymorphism of the synthetic allele. However, the most striking result was that mutations in G4 prone motifs strongly decreased the frequency of their rearrangement in *pif1Δ* cells. We observed only one rearrangement of the CEB1-Gmut-1.7 allele among the 383 colonies analyzed (0.2%) while the synthetic-CEB1-WT-1.7 allele was rearranged in 38/343 *pif1Δ* colonies (11%) (Figure 3 and Table 3). Similarly, the large synthetic-CEB1-Gmut-3.8 array, which contains approximately 97 repeats, yielded only a few rearrangements in the *pif1Δ* and WT strains (4% and 2%, respectively; this difference was not statistically different,  $p = 0.18$ , Fisher's Exact Test). In contrast, CEB1-Gmut arrays

rearranged in *rad27Δ* cells and the frequency of rearrangement increased in a size dependent-manner (Table 3). Thus, the synthetic and natural CEB1 alleles behaved similarly while the artificial CEB1 arrays containing mutation of G4-prone sequences were stabilized in *pif1Δ* but not in *rad27Δ* cells. These results strongly support our proposal that formation of G4 structures within the CEB1 array is responsible for their instability in vivo and that this secondary structure is processed by the Pif1 helicase.

## Discussion

In the present study, we provide new insights into the biochemical and biological functions of the evolutionary conserved Pif1 helicase. Our main findings are: (i) inactivation of Pif1 increased the frequency of rearrangement of the G-rich CEB1-1.8 tandem array, (ii) this increased rearrangement was specific for Pif1 as mutation of other helicases did not affect the stability of CEB1 and other repeats were stable in *pif1Δ* cells, (iii) the G-rich strand of the CEB1 repeat unit formed G-quadruplex structures in vitro, (iv) Pif1 readily unwound the CEB1 G4 structures in vitro and, (v) mutation of the G4-forming motifs stabilized CEB1 in *pif1Δ* cells. Destabilization of CEB1 in *pif1Δ* cells was not an indirect consequence of other *pif1Δ* phenotypes such as respiratory deficiency or long telomeres. Thus, the experiments reported here uncover a new activity for the Pif1 helicase, the ability to process G4 secondary structures, and suggest that this activity contributes to genome stability by preventing the rearrangement of G4 forming repeats in vivo.

## Mechanism of CEB1 Repeats Instability

In previous studies, we reported that human CEB1 repeats inserted into the yeast genome are highly unstable in absence of the Rad27 endonuclease and slightly unstable in a *dna2-1<sub>ts</sub>* mutant [25,26]. Since Rad27 and Dna2 are involved in the processing of flap structures during Okazaki fragment maturation [52], we concluded that CEB1 instability was likely due to the accumula-

**Table 3.** Instability of synthetic minisatellites in WT, *pif1Δ* and *rad27Δ* cells.

Minisatellite	Number of repeats	Strain	Genotype	Number of rearrangements/total (%)
CEB1-WT-2.3	58	AND1207-9B	WT	6/384 (1.5)
CEB1-WT-1.7	44	AND1212-10D	WT	2/708 (0.3)
CEB1-WT-1.0	26	AND1213-1D	WT	0/192 (0)
CEB1-WT-1.9	48	AND1202-13D-P14C3	<i>pif1Δ</i>	32/154 (20.6)
CEB1-WT-1.7	44	AND1202-11A	<i>pif1Δ</i>	38/343 (11)
CEB1-WT-1.3	33	AND1202-11A-L8C12	<i>pif1Δ</i>	8/189 (4.2)
CEB1-WT-1.0	26	ORT6108-4	<i>pif1Δ</i>	0/192 (0)
CEB1-WT-1.7	44	AND1218-1A	<i>rad27Δ</i>	15/50 (30)
CEB1-WT-1.0	26	ORT6110-1	<i>rad27Δ</i>	32/576 (5.5)
CEB1-Gmut-3.8	97	AND1206-5D	WT	8/384 (2)
CEB1-Gmut-1.7	42	AND1227-5C	WT	0/192 (0)
CEB1-Gmut-3.8	97	AND1206-4C	<i>pif1Δ</i>	8/192 (4)
CEB1-Gmut-2.5	64	AND1206-4C-1B6-1E1	<i>pif1Δ</i>	1/192 (0.5)
CEB1-Gmut-1.7	42	AND1206-4C-D11P2	<i>pif1Δ</i>	1/383 (0.2)
CEB1-Gmut-0.7	19	ORT6107-1	<i>pif1Δ</i>	0/192 (0)
CEB1-Gmut-3.8	97	AND1206-4B	<i>rad27Δ</i>	35/51 (68.6)
CEB1-Gmut-1.7	42	AND1226-18B/-17C	<i>rad27Δ</i>	52/98 (53)
CEB1-Gmut-0.7	19	ORT6109	<i>rad27Δ</i>	23/552 (4.1)

doi:10.1371/journal.pgen.1000475.t003



tion of unresolved flap structures during replication. We proposed that these intermediates would form recombinogenic structures that are repaired by homology-dependent strand displacement and annealing (SDSA) [53].

Here we show that inactivation of Pif1 also resulted in CEB1 instability. As in *rad27Δ* cells, the CEB1 rearrangements in *pif1Δ* cells had a high frequency of complex events (Figure 1C; [25]). In addition, in both mutants, CEB1 rearrangements depended on Rad52/Rad51-dependent homologous recombination (Table 1). These similarities suggest that the repair of the lesion leading to CEB1 rearrangement in the absence of either Pif1 or Rad27 occurs by SDSA, although the recombinogenic lesion may be different (for example a single-strand gap or a double-strand break). In *pif1Δ* and *rad27Δ* cells, the frequency of rearrangements increased with the size of the allele (Figure 3C; [25]). In *rad27Δ* cells, this increased instability may reflect the increased probability that longer arrays are more likely to contain more than one improperly processed flap. Similarly, in *pif1Δ* cells, long CEB1 minisatellites could form G4 structures with a higher probability, especially if quadruplexes involve G-tracts from adjacent repeats. Alternatively, lesions in small alleles could be rare or more often resected into the non-repeated flanking sequences, leading to the preferential restoration of the parental sequence by homologous recombination in G2 cells using the intact sister chromatid as a template [53].

#### CEB1 Repeats Are Unstable in *pif1Δ* Cells Only if They Are Able to Form G4 Structures

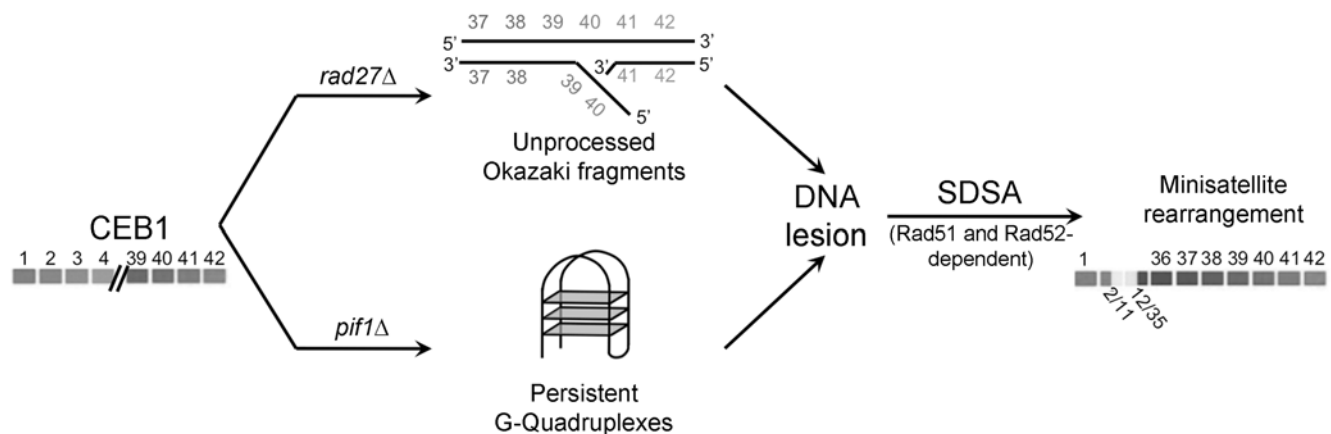
Whereas all micro- and minisatellites sequences tested are unstable in *rad27Δ* cells ([43,44] this study), only CEB1 was unstable in *pif1Δ* cells (Table 2). The CEB1 sequence is G/C rich (72%) with a high strand bias (23 G and 7 C per repeat of 39 bases). However, the instability of CEB1 in *pif1Δ* cells can not be attributed solely to its G/C rich sequence as the human hRAS1 minisatellite, which is also G rich (68%) with a strong bias (14 G and 5 C per repeat of 28 bases), was stable in the absence of Pif1. Each CEB1 repeat contains putative G4 signature motifs. Our biophysical analyses of CEB1 and hRAS oligonucleotides showed that the CEB1 motif readily formed G4 structures in vitro while hRAS1 did not (Figure 2 and Table S2). Moreover, synthetic CEB1 minisatellites in which the runs of guanine were mutated to disrupt their ability to form G4 structures were no longer unstable in *pif1Δ* cells. We propose that the recombinogenic lesions formed in the absence of Pif1 are unresolved

intra- or inter-motifs G4 structures. Thus, while CEB1 alleles are unstable in both *pif1Δ* and *rad27Δ* cells, the events that initiate instability, unprocessed Okazaki fragments (in *rad27Δ* cells) or persistent G4 structures (in *pif1Δ* cells) are different (Figure 4). As a result, all tandem arrays are unstable in the absence of Rad27, including the synthetic G4-mutated CEB1 alleles, while only CEB1 was unstable in *pif1Δ* cells.

#### In Vivo Roles of Pif1

What do our results suggest about the role(s) of Pif1 in the cell? Owing to the alternative use of a translation start site, *PIF1* generates two isoforms, one with mitochondrial and one with nuclear functions. Several observations indicate that Pif1 is involved in the maintenance of mitochondrial DNA. Specifically, Pif1 increases the frequency of recombination between  $\rho^+$  and certain  $\rho^-$  tandemly repeated mitochondrial genomes [21]. The loss of Pif1 is thought to trigger mtDNA breakage in specific regions, leading the authors to propose that Pif1 recognizes a specific but uncharacterized DNA topology [22,54]. Although the  $\sim 75$  kb *S. cerevisiae* mitochondrial genome is AT-rich, it contains numerous G-rich stretches. We speculate that in the absence of mitochondrial Pif1, breaks occur due to defective processing of G4 structures and these breaks are repaired by recombination. Alternatively, G4 DNA can create a structural target for factors involved in DNA recombination.

In the nucleus, Pif1 affects telomere length through direct inhibition of telomerase [23,28] the specialized reverse transcriptase that lengthens telomeres in most eukaryotes. In vivo and in vitro data suggest that telomerase inhibition is achieved by direct displacement of telomerase from a DNA end [24]. Since Pif1 exhibits a marked preference for RNA-DNA hybrid unwinding in vitro [55], Pif1 is proposed to inhibit telomerase by unwinding the RNA-DNA hybrid formed between the telomerase RNA, TLC1, and the telomeric DNA end. Pif1-mediated removal of telomerase from DNA ends can explain the effects of *pif1* mutations on both telomere length and de novo telomere addition [23,56] as well as its inhibition of gross chromosomal rearrangements [57]. Human Pif1 (hPIF) may have similar functions as ectopic expression of hPIF causes telomere shortening and decreased telomerase processivity in vitro [58]. In addition, hPIF co-immunoprecipitates with telomerase subunits and telomerase activity [59]. Importantly for the present study, most telomeric DNA sequences, including yeast and human telomeric DNA, can form G4 structures in vitro.



**Figure 4. Proposed model for CEB1 rearrangements in *rad27Δ* and *pif1Δ* mutants.** In the absence of Rad27, accumulation of unresolved flap structures inside CEB1 during replication generates recombinogenic structures that are repaired by homology-dependent strand displacement and annealing (SDSA). While in the absence of Pif1, the persistence of unprocessed G-quadruplex secondary structures in CEB1, during replication, transcription or other processes, initiates DNA lesions that are also repaired by SDSA, leading to minisatellite rearrangements. doi:10.1371/journal.pgen.1000475.g004

Moreover, G4 structures have been detected at ciliate telomeres in vivo [6]. In budding yeast, no evidence of the presence of G4 structures in the telomeric single stranded region has yet been reported, but proteins that bind or process G4 DNA in vitro are nevertheless present at yeast telomeres. In particular, in vitro studies have shown that the telomere binding protein Rap1 binds double-stranded telomeric DNA and promote the formation of G-quadruplex structures [60]. It is not known if this reaction occurs in vivo, but it is tempting to speculate that the formation of G4 DNA is necessary to promote the assembly of functional telomere. Alternatively or in addition to its ability to inhibit telomerase directly, Pif1 could counteract the formation of G4 structures in telomeric DNA, thus antagonizing the formation of proper telomere architecture. Consistent with this hypothesis, it has been shown that Pif1 overexpression compromises the viability of yeast strains with compromised telomere end protection [61].

Several studies suggest that Pif1 also has non-telomeric roles in replication and repair of nuclear DNA. First, in the rDNA, Pif1 helps maintain the replication fork barrier during replication [32]. Second, Pif1 is recruited to Rad52 DNA repair foci after gamma irradiation [62]. Third, lack of Pif1 suppresses the lethality of a Dna2 deletion, a helicase/endonuclease involved in the processing of Okazaki fragments by removing long 5' flaps. Although the role of Pif1 in Okazaki fragment maturation is unclear, it is proposed to act by extending the flaps created by the lagging strand replicative polymerase at the junction of two consecutive Okazaki fragments [27]. Like Pif1, Dna2 is involved in telomere maintenance [63] and is able to process G4 DNA in vitro [64]. Thus, the two enzymes may act in concert to remove toxic intermediates, including G4-DNA, which could arise during lagging strand replication and, if not appropriately processed, promote formation of recombinogenic DNA lesions, such as double strand breaks.

Finally, considering that in addition to G4-unwinding, Pif1 more efficiently unwinds RNA/DNA hybrids than DNA/DNA substrates [55], it is also to be envisaged that Pif1 plays a more general role in yeast cells when potential G4 structure can form, for example, during transcription.

### Multiplicity and Specificity of G4-Processing Helicases

Budding yeast as well as all the other organisms encodes a large number of helicases. Current estimate in *S. cerevisiae* is approximately 120. This multiplicity raises the question of their specific substrate(s) and function(s), an issue which remains often unresolved and controversial. In *S. cerevisiae*, the RecQ homolog Sgs1 helicase was proposed to resolve G4 DNA, a conclusion primarily based on its ability, and more generally of members of the RecQ family, to resolve G4 DNA structures in vitro [16]. Compelling evidence for the involvement of Sgs1 in G4 DNA metabolism in vivo finally came from the survey of global gene expression analysis in absence of Sgs1 [11]. The authors found that the set of genes which expression level is affected in *sgs1* mutant is biased towards genes that contain potential G4 forming sequences in their ORFs. To our surprise, the deletion of *SGS1* had no effect on CEB1 stability (Table 1). The lack of in vivo redundancy between Sgs1 and Pif1 in this novel assay is interesting and allows several hypotheses. First, it is possible that Sgs1 and Pif1 do not recognize the same set of G4 structures. G4 forming sequences can give rise to secondary structures exhibiting very diverse sizes, topologies (parallel or anti-parallel) and arrangements (intra- or inter-molecular) [65], and these structures may be recognized or processed differently depending on helicase. Second, Sgs1 may not recognize the G4 substrates generated by CEB1 in vivo due to the polarity of the single strand region flanking the G4-DNA structure (Pif1 is a 5'-3' helicase while Sgs1 has a 3'-5'

polarity). Third, it is likely that the numerous repeats in CEB1 that contain G4 forming sequences lead to the formation of highly stable structures in vivo that only some helicases are able to unwind. Finally, in the absence of more direct evidences for Sgs1 involvement in G4 DNA unwinding in vivo, there is also a possibility that Sgs1 plays a minor role in maintaining G4 DNA forming sequences. In multicellular organisms, the relationships between genomic instability, G-quadruplex structures and helicases functions have also been suspected. Studies in human cells deficient for the Werner, Bloom and RTEL helicases showed defects in telomere maintenance in vivo while G4 DNA is highly suspected to form at mammalian telomeres [66,67] and a recent study reports the correlation between genomic stability and G4 DNA unwinding by the human FANCDJ helicase [18]. Similarly, in *Caenorhabditis elegans*, the disruption of the RTEL homolog DOG-1 triggers deletions of polyguanine tracts matching the G4 DNA signature [20].

Finally, it should be mentioned that the inactivation of the potential Pif1 homolog in mice has no detectable phenotype, in particular regarding change in telomere length homeostasis [68]. In light of our present study, the stability of other repeated potentially G4 forming sequences in mice and mammalian cells should be examined. Also, taking advantage of the present yeast system allowing to test natural and synthetic substrates, we anticipate that further studies of *pif1Δ* cells will allow to uncover the multiple roles of this evolutionary conserved helicase, facilitate the characterization of G4 structures in vivo and finally enhance our understanding of the dynamics of G4 formation and function in vivo.

## Materials and Methods

### Yeast Strains

The relevant genotypes and sources of haploid and diploid *S. cerevisiae* strains (S288C background) used in this study are indicated in Table S1.

### Identification of Minisatellite Rearrangements

Examination of CEB1 instability during vegetative growth was done as previously described [25]. Individual colonies or colonies pools were analyzed by Southern blot depending on the rearrangement frequency (for rearrangement frequency >20%, individual colonies were privileged). Southern blots were performed using *AluI* digestion for natural CEB1 minisatellites and *ApaI/SpeI* for synthetic minisatellites and the corresponding membranes were hybridized with the radiolabeled CEB1-0.6 and CEB1-synthetic probes, respectively. For the analysis of the yeast minisatellite instability (*DAN4*, *FLO1*, *HKR1* and *NUM1*), Southern blots were performed using *AluI* digestion (which does not cut in these repeats). Membranes were hybridized with the radiolabeled purified PCR product of the corresponding minisatellite (primer sequences available under request). For the analysis of the human *hRAS1* minisatellite instability, Southern blots were performed using *ApaI/SpeI* digestion and *hRAS1* probe obtained from the p37Y8 plasmid (gift from D. Kirkpatrick). Detection of signals was done with a Storm PhosphorImager (Molecular Dynamics). For pools of genomic DNA from 12 or 16 colonies/wells, rearrangement is counted when the intensity of the rearranged minisatellite, quantified with ImageQuant software, corresponds to 1/12 or 1/16 of the total amount of signals measured in the lane. When several rearranged minisatellites migrate at the same size they are considered as clonal and are counted only once.

## Sequencing of CEB1 Alleles

The internal structure of rearranged alleles was determined by DNA sequencing as described previously [25].

## Analysis of G-Quadruplex Secondary Structure

Oligonucleotides were synthesized by Eurogentec (Belgium). Concentrations of all oligodeoxynucleotides were estimated using extinction coefficients provided by the manufacturer and calculated with a nearest neighbor model [69] under low salt conditions at 60°C in order to destabilize quadruplex formation. The sequences studied are shown in Table S2. Oligonucleotides chosen for non denaturing gel electrophoresis were first purified under denaturing conditions.

Melting experiments were conducted as previously described [70]. Denaturation was followed by recording the absorbance at 240 or 295 nm [47,71]. Melting experiments were typically performed at a concentration of 4  $\mu$ M per strand. Thermal difference spectra (TDS) were obtained by difference between the absorbance spectra from unfolded and folded oligonucleotides that were respectively recorded much above and below its melting temperature ( $T_m$ ).

Circular dichroism (CD) spectra were recorded on a JASCO-810 spectropolarimeter using a 1 cm path length quartz cuvette in a reaction volume of 580  $\mu$ l. Oligonucleotides were either *i*) prepared as a 4  $\mu$ M solution in 10 mM lithium cacodylate pH 7.2, 100 mM NaCl or KCl buffer and annealed by heating to 90°C for 2 min, followed by cooling to 20°C or *ii*) preincubated for 48 hours at higher strand concentration (140  $\mu$ M) in a 10 mM lithium cacodylate pH 7.2, 1 M NaCl buffer. Scans were performed at 25°C to 90°C over a wavelength range of 220–335 nm with a scanning speed of 500 nm/min, a response time of 1 s, 1 nm pitch and 1 nm bandwidth.

Formation of G4-DNA was confirmed by non-denaturing PAGE. In this case, oligonucleotides were either directly observed by UV shadow (when incubated at high strand concentration) or 5' labeled with T4 polynucleotide kinase. Prior to the incubation, the DNA samples were heated at 90°C for 10 min and slowly cooled (2 h) to room temperature (or 60°C for 48 hours). Oligonucleotides were first treated with 50 mM LiOH (to unfold quadruplexes) for 10 minutes followed by HCl neutralization. Samples were incubated at 10 nM or 4  $\mu$ M strand concentration in Tris-HCl 10 mM pH 7.5 buffer with 100–1000 mM Li<sup>+</sup> or K<sup>+</sup>. 10% sucrose was added just before loading. Oligothymidylate markers (dT<sub>15</sub>, dT<sub>21</sub>, or dT<sub>30</sub>) or double-stranded markers (DX<sub>9</sub>: 5'-d-GCGATACGG+5'-d-CCGATACGC DX<sub>12</sub>: 5'-d-GCGTGACTTCGG+5'-d-CCGAAGTCAC-GC) were also loaded on the gel.

## Analysis of G-Quadruplex Unwinding by Pif1 In Vitro

Recombinant Pif1 was purified to homogeneity by affinity chromatography as described [55]. A Cy5-labeled oligonucleotide containing a 5' poly(dA) tail followed by a CEB1 repeat (5'-Cy5-AAAAAAAAAAGGGGAGGGAGGGTGGCCTGCGGAGG-TCCCTGGGCTG) was synthesized by Eurogentec (Belgium). For formation of the G-quadruplex, a solution of CEB1 oligo at 140  $\mu$ M in 1 M NaCl was denatured 5 min at 100°C, then incubated at 65°C for 48 hours to promote formation of G4 intermolecular structures [72]. The double-stranded DNA control was made by annealing a 5'-Cy5-labeled 20 mer oligonucleotide to a 40 mer oligonucleotide, leaving a 20 nucleotide-long 5' single-stranded DNA overhang. Briefly, 10  $\mu$ M of each oligonucleotide were mixed in a buffer containing 10 mM Tris pH 8.0 and 5 mM Mg<sup>2+</sup>. The mixture was denatured 5 minutes at 95°C and slowly let to cool to room temperature. The double-stranded DNA

substrate was further purified from non annealed single-stranded DNA on a MiniQ anion exchange column.

Helicase assays were carried out by incubating indicated amounts of Pif1 and 2 nM nucleic acid substrate at 35°C. Standard reaction buffer was 20 mM Tris pH 7.5, 50 mM NaCl, 100  $\mu$ g/ml bovine serum albumin, 2 mM DTT, 5 mM Mg<sup>2+</sup> and 4 mM ATP. For kinetic studies, reactions were started by addition of ATP in presence of 100 nM Pif1 and 2 nM substrate. 10  $\mu$ l aliquots were withdrawn at indicated times and the reactions stopped by addition of 2  $\mu$ l deproteinizing/loading buffer (6% Ficoll, 50 mM EDTA pH 8.0, 2.5  $\mu$ g/ $\mu$ l Proteinase K) and incubated further 15 minutes at 35°C. Reaction products were loaded on a 10% polyacrylamide non-denaturing gel and resolved by electrophoresis at 4°C and 10 V/cm in TBE 1 $\times$  buffer. Gels were dried and scanned with a storm PhosphorImager (Molecular Dynamics) and quantified using ImageQuant software (GE Healthcare).

## Statistical Analysis

Fisher exact test was performed using R software [73].

## Supporting Information

**Figure S1** Sequences of the G-strand of CEB1-1.8 parental allele and of nine rearrangements obtained in the *pif1* $\Delta$  haploid strain (ORT4841). Polymorphic DNA bases are highlighted. The numbers at right in parentheses indicate the corresponding repeat in the parental CEB1-1.8 allele. Two numbers separated by dash represent hybrid repeats. Junction regions, which are delimited by polymorphisms of CEB1-1.8 derived from repeats involved in the deletions/duplications, are shaded in grey. X indicates a repeat of unknown origin or which cannot be attributed to a specific repeat in the parental CEB1-1.8 allele.

Found at: doi:10.1371/journal.pgen.1000475.s001 (1.66 MB PDF)

**Figure S2** Behavior of the 39Ceb and 39Cebm sequences on a non-denaturing gel. Two strand concentrations were tested: radiolabeled only (around 50 nM) or supplemented with 4  $\mu$ M of cold oligonucleotide. Samples were treated with 50 mM LiOH to unfold quadruplexes, reannealed in 1 M NaCl buffer (top) or KCl (bottom) for 2 hours and loaded on a non-denaturing 15% acrylamide gel and run at 26°C. Migration markers are double-stranded DNA (9 and 12 bp) and (dT)<sub>15</sub>, (dT)<sub>21</sub> and (dT)<sub>30</sub> oligomers.

Found at: doi:10.1371/journal.pgen.1000475.s002 (0.11 MB PDF)

**Figure S3** Synthesis of artificial CEB1 minisatellites by PCR. (A) Nucleotide sequence of the CEB1-WT and CEB1-Gmut motifs. Repeats of at least three consecutive guanines are highlighted in grey in the CEB1-WT motif. Point mutations interrupting the G-triplets in the CEB1-Gmut motif are underlined. (B) Schematic representation of CEB1-concatemers synthesized by PCR. Two complementary oligonucleotides for CEB1-Gmut are represented (up and low), each composed of two identical CEB1-Gmut motifs (see Text S1 for sequences). After the first cycle of denaturation and annealing, the oligonucleotides can perfectly anneal along the two motifs and no elongation is possible (left), or they can shift and only one motif is annealed and the second motif is used as DNA template for elongation (right) resulting in addition of one motif at the end of the cycle. (C) After 30 cycles, DNA is deposited in agarose gel and the smear corresponds to a population of CEB1-concatemers of various sizes. White square indicates the part of the gel that will be cut in order to extract DNA and clone it in pGEM-T Easy vector. Sequences of the synthetic minisatellites, CEB1-WT-1.0 (D) and CEB1-Gmut-1.7 (E), with 26 and 42 repeats respectively. The sequence of the parental motif (CEB1-WT or

CEB1-Gmut) used for the synthesis is indicated above the sequence of the synthetic minisatellite. Mutations and small deletions introduced during the concatemer synthesis are highlighted in red and in grey, respectively.

Found at: doi:10.1371/journal.pgen.1000475.s003 (0.51 MB PDF)

**Table S1** List of strains used in this study.

Found at: doi:10.1371/journal.pgen.1000475.s004 (0.10 MB PDF)

**Table S2** Sequence of the oligonucleotides used and their respective melting temperatures.

Found at: doi:10.1371/journal.pgen.1000475.s005 (0.08 MB PDF)

**Text S1** Supplementary material and methods.

Found at: doi:10.1371/journal.pgen.1000475.s006 (0.13 MB PDF)

## References

- Mirkin SM (2008) Discovery of alternative DNA structures: a heroic decade (1979–1989). *Front Biosci* 13: 1064–1071.
- Sen D, Gilbert W (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* 334: 364–366.
- Fang G, Cech TR (1993) The beta subunit of *Oxytricha* telomere-binding protein promotes G-quartet formation by telomeric DNA. *Cell* 74: 875–885.
- Fry M (2007) Tetraplex DNA and its interacting proteins. *Front Biosci* 12: 4336–4351.
- Duquette ML, Handa P, Vincent JA, Taylor AF, Maizels N (2004) Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev* 18: 1618–1629.
- Schaffitzel C, Berger I, Postberg J, Hanes J, Lipps HJ, et al. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylomychia lemnae* macronuclei. *Proc Natl Acad Sci U S A* 98: 8572–8577.
- Paeschke K, Simonsson T, Postberg J, Rhodes D, Lipps HJ (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat Struct Mol Biol* 12: 847–854.
- Huppert JL, Balasubramanian S (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res* 33: 2908–2916.
- Todd AK, Johnston M, Neidle S (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res* 33: 2901–2907.
- Huppert JL, Balasubramanian S (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res* 35: 406–413.
- Hershman SG, Chen Q, Lee JY, Kozak ML, Yue P, et al. (2008) Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. *Nucleic Acids Res* 36: 144–156.
- Rawal P, Kummaraasetti VB, Ravindran J, Kumar N, Halder K, et al. (2006) Genome-wide prediction of G4 DNA as regulatory motifs: role in *Escherichia coli* global regulation. *Genome Res* 16: 644–655.
- Du Z, Zhao Y, Li N (2008) Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. *Genome Res* 18: 233–241.
- Maizels N (2006) Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat Struct Mol Biol* 13: 1055–1059.
- Fry M, Loeb LA (1999) Human Werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)<sub>n</sub>. *J Biol Chem* 274: 12797–12802.
- Huber MD, Lee DC, Maizels N (2002) G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. *Nucleic Acids Res* 30: 3954–3961.
- Mohaghegh P, Karow JK, Brosh RM Jr, Bohr VA, Hickson ID (2001) The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucl Acids Res* 29: 2843–2849.
- Wu Y, Shin-ya K, Brosh RM Jr (2008) FANCD1 helicase defective in Fanconi anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Mol Cell Biol* 28: 4116–4128.
- Cheung I, Schertzer M, Rose A, Lansdorff PM (2002) Disruption of dog-1 in *Caenorhabditis elegans* triggers deletions upstream of guanine-rich DNA. *Nat Genet* 31: 405–409.
- Kruisselbrink E, Guryev V, Brouwer K, Pontier DB, Cuppen E, et al. (2008) Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCD1-defective *C. elegans*. *Curr Biol* 18: 900–905.
- Foury F, Kolodny J (1983) *pif* mutation blocks recombination between mitochondrial rho+ and rho- genomes having tandemly arrayed repeat units in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 80: 5345–5349.
- Cheng X, Dunaway S, Ivesa AS (2007) The role of Pif1p, a DNA helicase in *Saccharomyces cerevisiae*, in maintaining mitochondrial DNA. *Mitochondrion* 7: 211–222.
- Schulz VP, Zakian VA (1994) The *saccharomyces* PIF1 DNA helicase inhibits telomere elongation and de novo telomere formation. *Cell* 76: 145–155.
- Boulé JB, Vega LR, Zakian VA (2005) The yeast Pif1p helicase removes telomerase from telomeric DNA. *Nature* 438: 57–61.
- Lopes J, Ribeyre C, Nicolas A (2006) Complex minisatellite rearrangements generated in the total or partial absence of Rad27/hFEN1 activity occur in a single generation and are Rad51 and Rad52 dependent. *Mol Cell Biol* 26: 6675–6689.
- Lopes J, Debrauwere H, Buard J, Nicolas A (2002) Instability of the human minisatellite CEB1 in *rad27A* and *dna2-1* replication-deficient yeast cells. *EMBO J* 21: 3201–3211.
- Budd ME, Reis CC, Smith S, Myung K, Campbell JL (2006) Evidence Suggesting that Pif1 Helicase Functions in DNA Replication with the Dna2 Helicase/Nuclease and DNA Polymerase  $\delta$ . *Mol Cell Biol* 26: 2490–2500.
- Zhou J, Monson EK, Teng S, Schulz VP, Zakian VA (2000) Pif1p helicase, a catalytic inhibitor of telomerase in yeast. *Science* 289: 771–774.
- Hardy CF, Sussel L, Shore D (1992) A RAPI-interacting protein involved in transcriptional silencing and telomere length regulation. *Genes Dev* 6: 801–814.
- Teng SC, Chang J, McCowan B, Zakian VA (2000) Telomerase-independent lengthening of yeast telomeres occurs by an abrupt Rad50p-dependent, Rif-inhibited recombinational process. *Mol Cell* 6: 947–952.
- Pinter SF, Aubert SD, Zakian VA (2008) The *Schizosaccharomyces pombe* Pfh1p DNA helicase is essential for the maintenance of nuclear and mitochondrial DNA. *Mol Cell Biol* 28: 6594–6608.
- Ivesa AS, Zhou JQ, Zakian VA (2000) The *Saccharomyces* Pif1p DNA helicase and the highly related Rrm3p have opposite effects on replication fork progression in ribosomal DNA. *Cell* 100: 479–489.
- Watt PM, Louis EJ, Borts RH, Hickson ID (1995) Sgs1: a eukaryotic homolog of *E. coli* RecQ that interacts with topoisomerase II in vivo and is required for faithful chromosome segregation. *Cell* 81: 253–260.
- Gangloff S, McDonald JP, Bendixen C, Arthur L, Rothstein R (1994) The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase. *Mol Cell Biol* 14: 8391–8398.
- Mimitou EP, Symington LS (2008) Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature* 455: 770–774.
- Oh SD, Lao JP, Taylor AF, Smith GR, Hunter N (2008) RecQ helicase, Sgs1, and XPF family endonuclease, Mus81-Mms4, resolve aberrant joint molecules during meiotic recombination. *Mol Cell* 31: 324–336.
- Watt PM, Hickson ID (1996) Failure to unwind causes cancer. *Genome stability. Curr Biol* 6: 265–267.
- Zhu Z, Chung WH, Shim EY, Lee SE, Ira G (2008) Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. *Cell* 134: 981–994.
- Veaute X, Jeusset J, Soustelle C, Kowalczykowski SC, Le Cam E, et al. (2003) The Srs2 helicase prevents recombination by disrupting Rad51 nucleoprotein filaments. *Nature* 423: 309–312.
- Prakash R, Krejci L, Van Komen S, Anke Schurer K, Kramer W, et al. (2005) *Saccharomyces cerevisiae* MPH1 Gene, Required for Homologous Recombination-mediated Mutation Avoidance, Encodes a 3' to 5' DNA Helicase. *J Biol Chem* 280: 7854–7860.
- Verstrepen KJ, Jansen A, Lewitter F, Fink GR (2005) Intragenic tandem repeats generate functional variability. *Nat Genet* 37: 986–990.
- Jauert PA, Edmiston SN, Conway K, Kirkpatrick DT (2002) *RAD1* controls the meiotic expansion of the human HRAS1 minisatellite in *Saccharomyces cerevisiae*. *Mol Cell Biol* 22: 953–964.
- Kokoska RJ, Stefanovic L, Tran HT, Resnick MA, Gordenin DA, et al. (1998) Destabilization of yeast micro- and minisatellite DNA sequences by mutations affecting a nuclease involved in Okazaki fragment processing (*rad27*) and DNA polymerase delta (*pol3-t*). *Mol Cell Biol* 18: 2779–2788.
- Freudenreich CH, Kantrow SM, Zakian VA (1998) Expansion and length-dependent fragility of CTG repeats in yeast. *Science* 279: 853–856.
- Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S (2006) Quadruplex DNA: sequence, topology and structure. *Nucl Acids Res* 34: 5402–5415.

## Acknowledgments

We thank members of our laboratory, J.A. Londoño Vallejo (Institut Curie, Paris), R. Rothstein, L. Lacroix, F. Leroy and C. Thurat (MNHN, Paris) for helpful discussions; Ph. Bois, M. Cardon and G. Millot for technical advice, materials and helpful discussions; and H. Debrauwère, F. Foury, V. Géli, E. Gilson, D. Kirkpatrick, S. Loeliet, V. Lundblad, K. Myung, T. Petes, C. Soustelle and D. Stillman for providing strains and plasmids.

## Author Contributions

Conceived and designed the experiments: CR JL JBB JLM AN. Performed the experiments: CR JL JBB AP AG JLM. Analyzed the data: CR JL JBB AP VAZ JLM AN. Wrote the paper: CR JL JBB VAZ JLM AN.

46. Mergny JL, Li J, Lacroix L, Amrane S, Chaires JB (2005) Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res* 33: e138.
47. Mergny JL, Phan AT, Lacroix L (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett* 435: 74–78.
48. Petraccone L, Erra E, Randazzo A, Giancola C (2006) Energetic aspects of locked nucleic acids quadruplex association and dissociation. *Biopolymers* 83: 584–594.
49. Paramasivan S, Rujan I, Bolton PH (2007) Circular dichroism of quadruplex DNAs: applications to structure, cation effects and ligand binding. *Methods* 43: 324–331.
50. Gray DM, Wen JD, Gray CW, Repges R, Repges C, et al. (2008) Measured and calculated CD spectra of G-quartets stacked with the same or opposite polarities. *Chirality* 20: 431–440.
51. Simonsson T (2001) G-quadruplex DNA structures—variations on a theme. *Biol Chem* 382: 621–628.
52. Bambara RA, Murante RS, Henricksen LA (1997) Enzymes and reactions at the eukaryotic DNA replication fork. *J Biol Chem* 272: 4647–4650.
53. Pâques F, Haber JE (1999) Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* 63: 349–404.
54. Foury F, Dyck EV (1985) A PIF-dependent recombinogenic signal in the mitochondrial DNA of yeast. *EMBO J* 4: 3525–3530.
55. Boulé JB, Zakian VA (2007) The yeast Pif1p DNA helicase preferentially unwinds RNA DNA substrates. *Nucleic Acids Res* 35: 5809–5818.
56. Mangahas JL, Alexander MK, Sandell LL, Zakian VA (2001) Repair of chromosome ends after telomere loss in *Saccharomyces*. *Mol Biol Cell* 12: 4078–4089.
57. Myung K, Chen C, Kolodner RD (2001) Multiple pathways cooperate in the suppression of genome instability in *Saccharomyces cerevisiae*. *Nature* 411: 1073–1076.
58. Zhang DH, Zhou B, Huang Y, Xu LX, Zhou JQ (2006) The human Pif1 helicase, a potential *Escherichia coli* RecD homologue, inhibits telomerase activity. *Nucleic Acids Res* 34: 1393–1404.
59. Mateyak MK, Zakian VA (2006) Human PIF helicase is cell cycle regulated and associates with telomerase. *Cell Cycle* 5: 2796–2804.
60. Giraldo R, Rhodes D (1994) The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. *EMBO J* 13: 2411–2420.
61. Vega LR, Phillips JA, Thornton BR, Benanti JA, Onigbanjo MT, et al. (2007) Sensitivity of yeast strains with long G-tails to levels of telomere-bound telomerase. *PLoS Genet* 3: 1065–1075.
62. Wagner M, Price G, Rothstein R (2006) The absence of Top3 reveals an interaction between the Sgs1 and Pif1 DNA helicases in *Saccharomyces cerevisiae*. *Genetics* 174: 555–573.
63. Choe W, Budd M, Imamura O, Hoopes L, Campbell JL (2002) Dynamic localization of an Okazaki fragment processing protein suggests a novel role in telomere replication. *Mol Cell Biol* 22: 4202–4217.
64. Masuda-Sasa T, Polaczek P, Peng XP, Chen L, Campbell JL (2008) Processing of G4 DNA by Dna2 Helicase/nuclease and RPA provides insights into the mechanism of Dna2/RPA substrate recognition. *J Biol Chem* 283: 24359–24373.
65. Bates P, Mergny JL, Yang D (2007) Quartets in G-major. The First International Meeting on Quadruplex DNA. *EMBO Rep* 8: 1003–1010.
66. Crabbe L, Verdun RE, Haggblom CI, Karlseder J (2004) Defective telomere lagging strand synthesis in cells lacking WRN helicase activity. *Science* 306: 1951–1953.
67. Ding H, Schertzer M, Wu X, Gertsenstein M, Selig S, et al. (2004) Regulation of murine telomere length by Rtel: an essential gene encoding a helicase-like protein. *Cell* 117: 873–886.
68. Snow BE, Mateyak M, Paderova J, Wakeham A, Iorio C, et al. (2007) Murine Pif1 interacts with telomerase and is dispensable for telomere function in vivo. *Mol Cell Biol* 27: 1017–1026.
69. Cantor CR, Warshaw MM, Shapiro H (1970) Oligonucleotide interactions. 3. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers* 9: 1059–1077.
70. Mergny JL, Lacroix L (2003) Analysis of thermal melting curves. *Oligonucleotides* 13: 515–537.
71. Sacca B, Lacroix L, Mergny JL (2005) The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. *Nucleic Acids Res* 33: 1182–1192.
72. Bachrati CZ, Hickson ID (2006) Analysis of the DNA unwinding activity of RecQ family helicases. *Methods Enzymol* 409: 86–100.
73. Ihaka R, Gentleman R (1996) R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics* 5: 299–314.

## SUPPLEMENTAL DATA

### Supplemental Materials and Methods

**Yeast strains and plasmids.** *S. cerevisiae* strains were grown as previously described [1]. The relevant genotypes and sources of haploid and diploid *S. cerevisiae* strains (S288C background) used in this study are indicated in supplemental Table S1. The CEB1-0.6, CEB1-1.8 and CEB1-3.0 alleles were inserted on chromosome VIII in the *ARG4* promoter as described previously [1,2]. The CEB1-3.5 allele is an expansion of CEB1-1.8 isolated from ORT4841.

The *KanMX4* disrupted strains were constructed by PCR amplification of the *geneX::KanMX4* cassette with flanking regions from the BY strains of the EUROSCARF deletants collection. After transformation of the yeast strain ORT2914 with the purified PCR product, G418<sup>R</sup> strains were verified for correct chromosomal integration by Southern blot. *pif1::HIS3* strains were built using plasmid M4754 (provided by David Stillman) which allows replacement of *KanMX4* by *HIS3* [3]. The *rad51::LEU2* and the *rad52::LEU2* disruptions originated from ORT2615 and ORT2916 respectively [4] were introduced by genetic crosses. To construct *DNA2* deletion, the *HygMX4* cassette was amplified from pAG32 [5] with primers containing homology with flanking regions of *DNA2*: DNA2ApMJ696A (5'-GAGTACTCATTGTGCAAGCAAACACTGACAATTGAAGAGATCGTCAGGATCCCCGGGTTAATTAA GGC-3') and DNA2BpMJ696B (5'-AGCTCGTTTTTCGACACTGGTG CTGTGATAGCTTTCCTGTTATGGAGAAGCTCTTCTTATCCCCCTGTC-3'). The PCR product was used to transform ORT4841 resulting in ORT4848.

The strains ORT4841, ORT2914 and ORD6713-8D were all transformed with replicative plasmids pBK3, pBK1, pMD28, pEAS20 and pBK10 [6]. These plasmids contain various

repetitive tracts in frame with the *URA3* marker for examining micro- and minisatellites instability.

The *hRAS1* minisatellite was amplified by PCR from the p37Y8 plasmid (gift from D. Kirkpatrick) using Ras 5' and Ras 3' primers and Taq polymerase (Invitrogen) as previously described [7]. The PCR fragment was cloned in pGEM-T Easy vector (Promega) and the *ApaI-SpeI* fragment containing *hRAS1* was inserted in pJL82 plasmid derived from pUC19. pJL82 contains a 3.1 kb *HindIII-PstI* fragment from *DED81-ARG4* region in which an unrelated polylinker DNA sequence, polyI, replaces a part of *ARG4* promoter [2]. *arg-* strains (WT, *pif1Δ* and *rad27Δ*) were transformed by electroporation with *HindIII-PstI* fragment containing minisatellite and selected for arginine prototrophy. Correct integration was verified by Southern blot analysis of Arg<sup>+</sup> transformants using *EcoRV/PvuII* digestion and a *DED81* probe.

In order to create *PIF1* point mutations, a *PIF1* fragment was amplified from yeast genomic DNA using PIF1-223-*XhoI* (5'-TCATGCTCGAGACATTAAGAAAGGCGCGTCT-3') and PIF1+1023-*XbaI* (5'-GAATCTCTAGAATCGACAACCAAAGCACCAA-3') primers. This fragment was digested by *XbaI/XhoI* (highlighted in primers) and inserted into integrative vector pRS306 containing *URA3* marker at *XbaI/XhoI* sites to create pJL69. The pJL69 vector was next used to create *PIF1* point mutations by site directed mutagenesis using QuikChange Site-Directed Mutagenesis (Stratagene). The following primers were used (mutated bases are highlighted): *pif1-K264R* was created using PIF1-K264R-up (5'-GAGTGCCGGTACCGGTAGATCCATTCTTTTACGTG-3') and PIF1-K264R-low (5'-CACGTAAAAGAATGGATCTACCGGTACCGGCACTC-3'), *pif1-m1* using PIF1-M1-up (5'-CTTGTATCAATCAATTTTGGGCCCAAAGTGGATAAGATC-3') and PIF1-M1-low (5'-GATCTTATCCACTTTGGGGCCCAAATGATTGATACAAG-3'), *pif1-m2* using PIF1-M2-up (5'-CCAAACTATCGTTTTCTGCCTCGAGTCGTGGTTTCAGGTC-3') and PIF1-M2-low (5'-GACCTGAAACCACGACTCGAGGCAGAAAACGATAGTTTGG-3'). After verification by

sequencing of the presence of the mutation, a *XbaI/XhoI* fragment (*pif1-K264R* and *pif1-m1*) or a *XbaI/BseRI* fragment (*pif1-m2*) was cloned into pRS306 plasmid at the same sites (*XbaI/XhoI* for *pif1-K264R* and *pif1-m1*; *XbaI/BseRI* for *pif1-m2*). This creates pJL71 (*pif1-K264R*), pJL72 (*pif1-m1*) and pJL73 (*pif1-m2*).

For *pif1-m2* overexpression, the entire ORF containing *pif1-m2* mutation plus the 5' and 3' noncoding region was amplified from yeast genomic DNA (from strain ORT5085-1C) with primers PIF1A (5'-AATGGCAAGTTTGCCGAA-3') and PIF1B (5'-TGCTTCCTGTCAGCTTGGTT-3'); next this PCR product was cloned in pCR2.1 vector (Invitrogen) which creates pJL75. A *SacI/NotI* fragment from pJL75 containing the entire PCR product was introduced in a 2 $\mu$  replicative plasmid pRS426 containing the *URA3* marker to create pJL76. This plasmid was transformed in the yeast strain ORT4841 and the resulting transformants (ORT5086) were selected for uracil prototrophy.

The *PfIMI/BseRI* fragment from pVS102PKA [8], which carries *pif1-K264A* mutation, was inserted at *PfIMI/BseRI* sites in pJL75, replacing *pif1-m2* mutation by *pif1-K264A* mutation, to create pJL78. The *SacI/NotI* fragment, containing the entire ORF with *pif1-K264A* mutation plus the 5' and 3' noncoding region from pJL78 was inserted at *SacI/NotI* in pRS306 plasmid, which creates pJL79. The plasmids pJL71 (*pif1-K264R*), pJL72 (*pif1-m1*), pJL73 (*pif1-m2*) and pJL79 (*pif1-K264A*) were linearized with *Bsu36I*, *AflIII*, *AflIII* and *BstEII* respectively, and introduced by transformation pop-in pop-out at the *PIF1* locus in the yeast strain ORT2914. Correct integration of the plasmids was verified by Southern blot analysis.

**Instability of tandem repeated sequences on plasmids.** For plasmids pBK3, pBK1, pMD28, pEAS20 and pBK10, alteration in tract length results in an out of frame insertion which can be selected on medium containing 5-Fluoro-Orotic-Acid (5-FOA) as previously described [6].



**Synthetic minisatellites.** The synthetic minisatellites were generated by PCR using complementary primers: 39x2-up (5' TCAGCCCAGGGACCTCCGCAGGCCACCCTCCCTCCCCCTCAGCCCAGGGACCTCCGCAGGCCACCCTCCCTCCCCC 3') and 39x2-low (5' GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGAGGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA 3') for the CEB1-WT minisatellites and the primers 39x2mut-up (5' TCAGCGCAGGGACCTCCGCAGGCCACTCTCACTCCCGGCTCAGCGCAGGGACCTCCGCAGGCCACTCTACTCCCGC 3') and 39x2mut-low (5' GCGGGAGTGAGAGTGGCCTGCGGAGGTCCCTGCGGCTGAGCGGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGGCTGA 3') for the CEB1-Gmut minisatellites. Mutations interrupting each triplet of guanines are underlined. Principle of the minisatellite synthesis is represented in Figure S3 B. PCR reaction was done with a final concentration of 0.37  $\mu$ M for each oligonucleotide, the PFU Ultra polymerase (1U) (Stratagene) in a total volume of 15  $\mu$ l using the buffer described in Jeffrey *et al.* [9]. Reaction conditions were: 3 min at 95°C for one cycle, and 30 s at 95°C, 30 s at 64°C and 30 s at 72°C +2 s/cycle for 30 cycles and the reaction was terminated by 10 min at 72°C. At the end of the PCR, 0.2 $\mu$ l of recombinant Taq polymerase (Invitrogen) was added during 6 min at 72°C for A-tailing procedure. To increase the quantity of amplified DNA, four identical PCR reactions were performed in parallel. These PCR products were pooled, precipitated and electrophoresed in 0.8% agarose gel. DNA fragments around 1 kb were extracted from agarose gel with NucleoSpin Extract II kit (Macherey–Nalgen), cloned in pGEM-T Easy vector (Promega) and sequenced with the Big Dye Terminator version 3.1 kit (Perkin Elmer) and 20% Betaine (Sigma) (Supplemental Figure S3). Synthetic CEB1 minisatellites were inserted on chromosome VIII in the *ARG4* promoter as described for *hRAS1* minisatellite.

## Supplemental References

1. Lopes J, Ribeyre C, Nicolas A (2006) Complex minisatellite rearrangements generated in the total or partial absence of Rad27/hFEN1 activity occur in a single generation and are Rad51 and Rad52 dependent. *Mol Cell Biol* 26: 6675-6689.
2. Debrauwere H, Buard J, Tessier J, Aubert D, Vergnaud G, et al. (1999) Meiotic instability of human minisatellite CEB1 in yeast requires DNA double-strand breaks. *Nat Genet* 23: 367-371.
3. Voth WP, Jiang YW, Stillman DJ (2003) New 'marker swap' plasmids for converting selectable markers on budding yeast gene disruptions and plasmids. *Yeast* 20: 985-993.
4. Soustelle C, Vedel M, Kolodner R, Nicolas A (2002) Replication protein A is required for meiotic recombination in *Saccharomyces cerevisiae*. *Genetics* 161: 535-547.
5. Goldstein AL, McCusker JH (1999) Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. *Yeast* 15: 1541-1553.
6. Kokoska RJ, Stefanovic L, Tran HT, Resnick MA, Gordenin DA, et al. (1998) Destabilization of yeast micro- and minisatellite DNA sequences by mutations affecting a nuclease involved in Okazaki fragment processing (rad27) and DNA polymerase delta (pol3-t). *Mol Cell Biol* 18: 2779-2788.
7. Jauert PA, Edmiston SN, Conway K, Kirkpatrick DT (2002) *RAD1* controls the meiotic expansion of the human HRAS1 minisatellite in *Saccharomyces cerevisiae*. *Mol Cell Biol* 22: 953-964.
8. Zhou J, Monson EK, Teng S, Schulz VP, Zakian VA (2000) Pif1p helicase, a catalytic inhibitor of telomerase in yeast. *Science* 289: 771-774.
9. Jeffreys AJ, Neumann R, Wilson V (1990) Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. *Cell* 60: 473-485.

## Supplemental Figure Legends

**Figure S1: Sequences of the G-strand of CEB1-1.8 parental allele [1] and of nine rearrangements obtained in the *pif1Δ* haploid strain (ORT4841).** Polymorphic DNA bases are highlighted. The numbers at right in parentheses indicate the corresponding repeat in the parental CEB1-1.8 allele. Two numbers separated by dash represent hybrid repeats. Junction regions, which are delimited by polymorphisms of CEB1-1.8 derived from repeats involved in the deletions/duplications, are shaded in grey. X indicates a repeat of unknown origin or which cannot be attributed to a specific repeat in the parental CEB1-1.8 allele.

**Figure S2: Behavior of the 39Ceb and 39Cebm sequences on a non-denaturing gel.** Two strand concentrations were tested: radiolabeled only (around 50 nM) or supplemented with 4  $\mu$ M of cold oligonucleotide. Samples were treated with 50 mM LiOH to unfold quadruplexes, reannealed in 1 M NaCl buffer (top) or KCl (bottom) for 2 hours and loaded on a non-denaturing 15% acrylamide gel and run at 26°C. Migration markers are double-stranded DNA (9 and 12 bp) and (dT)<sub>15</sub>, (dT)<sub>21</sub> and (dT)<sub>30</sub> oligomers.

**Figure S3: Synthesis of artificial CEB1 minisatellites by PCR.** (A) Nucleotide sequence of the CEB1-WT and CEB1-Gmut motifs. Repeats of at least three consecutive guanines are highlighted in grey in the CEB1-WT motif. Point mutations interrupting the G-triplets in the CEB1-Gmut motif are underlined. (B) Schematic representation of CEB1-concatemers synthesized by PCR. Two complementary oligonucleotides for CEB1-Gmut are represented (up and low), each composed of two identical CEB1-Gmut motifs (see “Supplemental Experimental Procedure” for sequences). After the first cycle of denaturation and annealing, the oligonucleotides can perfectly anneal along the two motifs and no elongation is possible (left), or they can shift and only one motif is annealed and the second motif is used as DNA template for elongation (right) resulting in addition of one motif at the end of the cycle. (C) After 30 cycles, DNA is deposited in agarose gel and the smear corresponds to a population of CEB1-concatemers of various sizes. White square indicates the part of the gel that will be cut in order to extract DNA and clone it in pGEM-T Easy vector. Sequences of the synthetic minisatellites, CEB1-WT-1.0 (D) and CEB1-Gmut-1.7 (E), with 26 and 42 repeats respectively. The sequence of the parental motif (CEB1-WT or CEB1-Gmut) used for the synthesis is indicated above the sequence of the synthetic minisatellite. Mutations and small

deletions introduced during the concatemer synthesis are highlighted in red and in grey, respectively.

**TABLE S1. Genotypes of yeast strains used in this study**

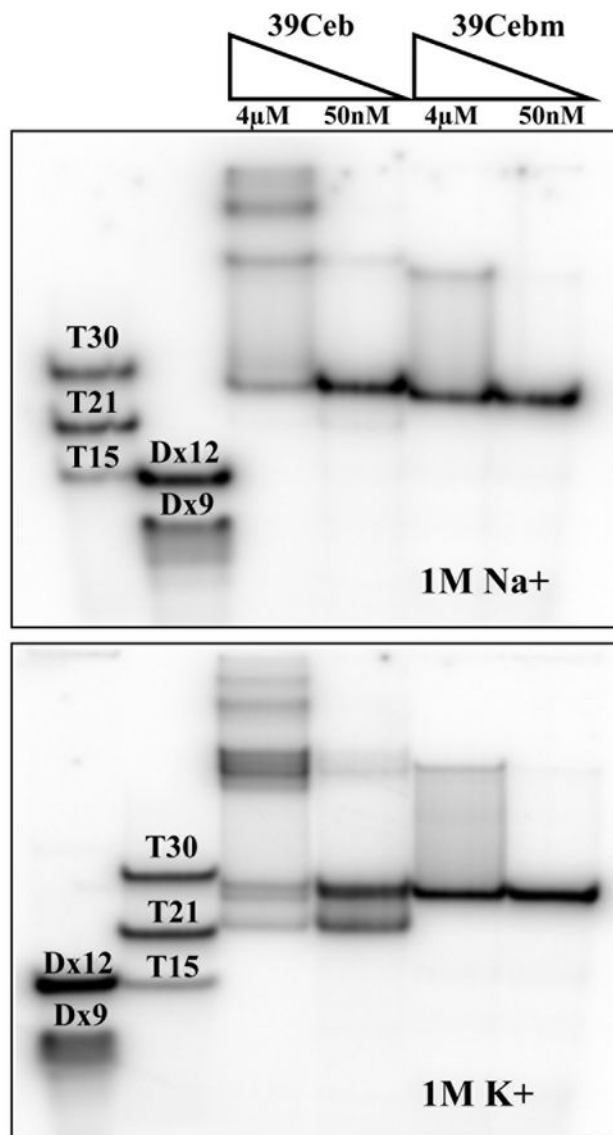
Strain	Relevant genotype	Source
ORT2914	<i>MAT<math>\alpha</math></i> CEB1-1.8 <i>ARG4</i>	Lopes <i>et al.</i> 2006
ORT4840	<i>MAT<math>\alpha</math></i> <i>srs2::KanMX4</i> CEB1-1.8 <i>ARG4</i>	This study
ORT4849	<i>MAT<math>\alpha</math></i> <i>sgs1::KanMX4</i> CEB1-1.8 <i>ARG4</i>	This study
ORT4880	<i>MAT<math>\alpha</math></i> <i>rrm3::KanMX4</i> CEB1-1.8 <i>ARG4</i>	This study
ORT4885	<i>MAT<math>\alpha</math></i> <i>mph1::KanMX4</i> CEB1-1.8 <i>ARG4</i>	This study
ORT5084-2C	<i>MAT<math>\alpha</math></i> <i>pif1-m1</i> CEB1-1.8 <i>ARG4</i>	This study
ORT5085-1C	<i>MAT<math>\alpha</math></i> <i>pif1-m2</i> CEB1-1.8 <i>ARG4</i>	This study
ORT5083-4E	<i>MAT<math>\alpha</math></i> <i>pif1-K264R</i> CEB1-1.8 <i>ARG4</i>	This study
ORT5087-5E	<i>MAT<math>\alpha</math></i> <i>pif1-K264A</i> CEB1-1.8 <i>ARG4</i>	This study
ORT4841	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> CEB1-1.8 <i>ARG4</i>	This study
ORT4843	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> CEB1-1.8 <i>ARG4</i>	This study
ORD7557-10B	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> CEB1-0.6 <i>ARG4</i>	This study
ORD7598-12C	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> CEB1-3.0 <i>ARG4</i>	This study
ORT4841-4E1	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> CEB1-3.5 <i>ARG4</i>	This study
ORD7569	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> CEB1-1.8 <i>ARG4</i> / <i>pif1::KanMX4</i> CEB1-1.8 <i>ARG4</i>	This study
ORT5086	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> CEB1-1.8 <i>ARG4</i> + pJL76	This study
ORT4848	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> <i>dna2::HygMX4</i> CEB1-1.8 <i>ARG4</i>	This study
ORT5604	<i>MAT<math>\alpha</math></i> CEB1-1.8 <i>ARG4</i> + pMD28	This study
ORT5606	<i>MAT<math>\alpha</math></i> CEB1-1.8 <i>ARG4</i> + pBK1	This study
ORT5614	<i>MAT<math>\alpha</math></i> CEB1-1.8 <i>ARG4</i> + pBK3	This study
ORT5602	<i>MAT<math>\alpha</math></i> CEB1-1.8 <i>ARG4</i> + pBK10	This study
ORT5608	<i>MAT<math>\alpha</math></i> CEB1-1.8 <i>ARG4</i> + pEAS20	This study
ORT5600	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> CEB1-1.8 <i>ARG4</i> + pMD28	This study
ORT4896	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> CEB1-1.8 <i>ARG4</i> + pBK1	This study
ORT4894	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> CEB1-1.8 <i>ARG4</i> + pBK3	This study
ORT4892	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> CEB1-1.8 <i>ARG4</i> + pBK10	This study
ORT4898	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> CEB1-1.8 <i>ARG4</i> + pEAS20	This study
AND1228-2A	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> hRAS1-2.1 <i>ARG4</i>	This study
AND1228-8C	<i>MAT<math>\alpha</math></i> <i>rad27::HIS3</i> hRAS1-2.1 <i>ARG4</i>	This study
ORT6009	<i>MAT<math>\alpha</math></i> <i>rad27::HIS3</i> CEB1-1.8 <i>ARG4</i> + pMD28	This study
ORT6014	<i>MAT<math>\alpha</math></i> <i>rad27::HIS3</i> CEB1-1.8 <i>ARG4</i> + pBK1	This study
ORT6013	<i>MAT<math>\alpha</math></i> <i>rad27::HIS3</i> CEB1-1.8 <i>ARG4</i> + pBK3	This study
ORT6011	<i>MAT<math>\alpha</math></i> <i>rad27::HIS3</i> CEB1-1.8 <i>ARG4</i> + pBK10	This study
ORT6016	<i>MAT<math>\alpha</math></i> <i>rad27::HIS3</i> CEB1-1.8 <i>ARG4</i> + pEAS20	This study
ORD7574-11C	<i>MAT<math>\alpha</math></i> <i>rad51::LEU2</i> CEB1-1.8 <i>ARG4</i>	This study
ORD6786-4A	<i>MAT<math>\alpha</math></i> <i>rif1::KanMX4</i> CEB1-1.8 <i>ARG4</i>	This study
ORD6713-8D	<i>MAT<math>\alpha</math></i> <i>rad27::HIS3</i> CEB1-1.8 <i>ARG4</i>	Lopes <i>et al.</i> 2006
ORD7574-9B	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> <i>rad51::LEU2</i> CEB1-1.8 <i>ARG4</i>	This study
ORD7565-2C	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> <i>rad52::LEU2</i> CEB1-1.8 <i>ARG4</i>	This study
ORD9922-4B	<i>MAT<math>\alpha</math></i> <i>pif1::HIS3</i> <i>sgs1::KanMX4</i> CEB1-1.8 <i>ARG4</i>	This study
ORD9304-9A	<i>MAT<math>\alpha</math></i> <i>pif1::HIS3</i> <i>rrm3::KanMX4</i> CEB1-1.8 <i>ARG4</i>	This study
ORD6708	<i>MAT<math>\alpha</math></i> <i>rad27::HIS3</i> / <i>rad27::HIS3</i> CEB1-1.8 <i>ARG4</i> / CEB1-0.6 <i>arg4bg</i>	Lopes <i>et al.</i> 2002
AND1207-9B	<i>MAT<math>\alpha</math></i> CEB1-WT-2.3 <i>ARG4</i>	This study
AND1212-10D	<i>MAT<math>\alpha</math></i> CEB1-WT-1.7 <i>ARG4</i>	This study
AND1213-1D	<i>MAT<math>\alpha</math></i> CEB1-WT-1.0 <i>ARG4</i>	This study
AND1202-13D-P14C3	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> CEB1-WT-1.9 <i>ARG4</i>	This study
AND1202-11A	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> CEB1-WT-1.7 <i>ARG4</i>	This study
AND1202-11A-L8C12	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> CEB1-WT-1.3 <i>ARG4</i>	This study
ORT6108-4	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> CEB1-WT-1.0 <i>ARG4</i>	This study
AND1218-1A	<i>MAT<math>\alpha</math></i> <i>rad27::HIS3</i> CEB1-WT-1.7 <i>ARG4</i>	This study
ORT6110-1	<i>MAT<math>\alpha</math></i> <i>rad27::HIS3</i> CEB1-WT-1.0 <i>ARG4</i>	This study
AND1206-5D	<i>MAT<math>\alpha</math></i> CEB1-Gmut-3.8 <i>ARG4</i>	This study
AND1227-5C	<i>MAT<math>\alpha</math></i> CEB1-Gmut-1.7 <i>ARG4</i>	This study
AND1206-4C	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> CEB1-Gmut-3.8 <i>ARG4</i>	This study
AND1206-4C-1B6-1E1	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> CEB1-Gmut-2.5 <i>ARG4</i>	This study
AND1206-4C-D11P2	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> CEB1-Gmut-1.7 <i>ARG4</i>	This study
ORT6107-1	<i>MAT<math>\alpha</math></i> <i>pif1::KanMX4</i> CEB1-Gmut-0.7 <i>ARG4</i>	This study
AND1206-4B	<i>MAT<math>\alpha</math></i> <i>rad27::HIS3</i> CEB1-Gmut-3.8 <i>ARG4</i>	This study
AND1226-18B/-17C	<i>MAT<math>\alpha</math></i> <i>rad27::HIS3</i> CEB1-Gmut-1.7 <i>ARG4</i>	This study
ORT6109	<i>MAT<math>\alpha</math></i> <i>rad27::HIS3</i> CEB1-Gmut-0.7 <i>ARG4</i>	This study

**Table S2 : Sequence of the oligonucleotides used and their respective melting temperatures**

Name <sup>a</sup>	Sequence (5' => 3') <sup>b</sup>	Tm <sup>c</sup> Na <sup>+</sup> (°C)	Tm <sup>c</sup> K <sup>+</sup> (°C)
24Ceb	TGAGGGGGGAGGGAGGGAGGGTGG	<b>42</b>	<b>69</b>
29Ceb	AGGGGGGAGGGAGGGTGGCCTGCGGAGGT	<b>38</b>	<b>54</b>
39Ceb	AGGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTG	<b>48</b>	<b>55</b>
24Cebm	TGAG <u>CGCGGAGT</u> GAG <u>AG</u> TGGCCTG	<i>dx</i>	<i>dx</i>
39Cebm	AG <u>CGCGGAGT</u> GAG <u>AG</u> TGGCCTGCGGAGGTCCCTG <u>CG</u> GCTG	<i>dx</i>	<i>dx</i>
28hRAS1	GGCGTCCCCTGGACAGAA <b>GGGG</b> AGTGT	<i>dx</i>	<i>dx</i>
28hRAS2	GGCATTCCCTGGACAGAA <b>GGG</b> CAAGTGT	<i>dx</i>	<i>dx</i>
28hRAS3	GGCGTCCCCTGGAGAGAA <b>GGG</b> CCAGTGT	<i>dx</i>	<i>dx</i>

<sup>a</sup> 24Ceb, 29Ceb and 39Ceb oligonucleotides correspond to sequences found in CEB1 minisatellite repeats. 24Cebm and 39Cebm are mutant sequences in which guanines blocks have been disrupted. 28hRAS1, 28hRAS2 and 28hRAS3 are control sequences found in *hRAS1* GC rich minisatellite. <sup>b</sup> Blocks of 3 or more guanines are shown in red. Mutations that interrupt guanine blocks appear in bold/underlined. <sup>c</sup> Melting temperature, in °C (average of 2-4 independent experiments) determined in a 10 mM lithium pH 7.2 cacodylate buffer supplemented with 0.1M NaCl or KCl (last column). Values in bold were determined at 295 nm and could be attributed to quadruplex thermal denaturation. *dx*: no transition at 295 nm but evidence for duplex formation, with a melting temperature >37°C.

**Figure S2**

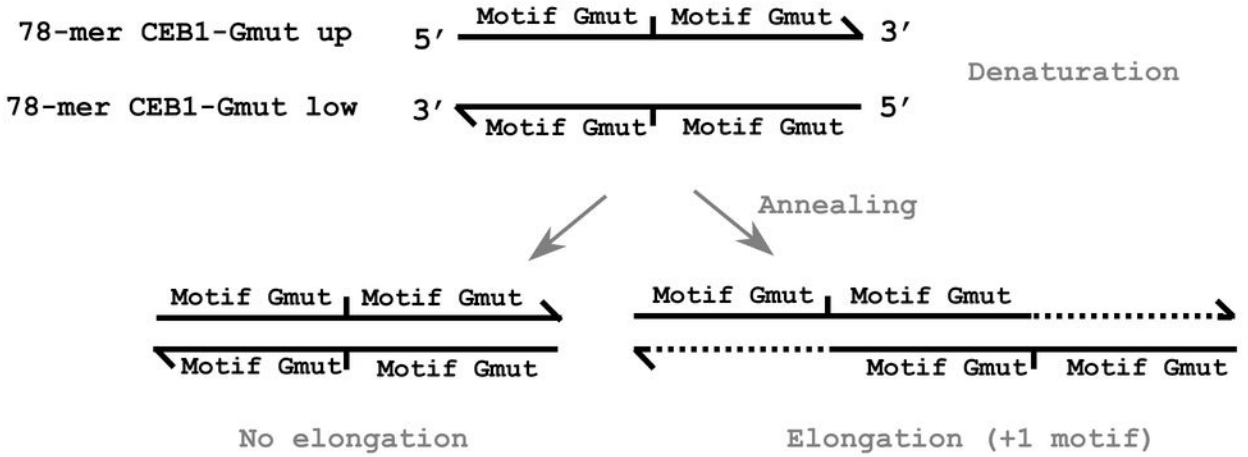


# Figure S3

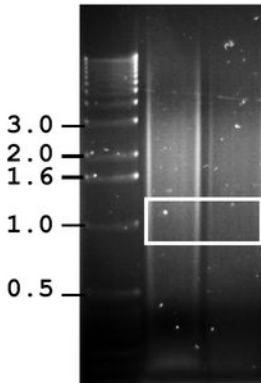
**A**

Motif CEB1-WT 5' **GGGGGGAGGGAGGG**TGGCCTGCGGAGGTCCCT**GGG**CTGA 3'  
 Motif CEB1-Gmut 5' **GCGCGGAG**T**GAGAG**TGGCCTGCGGAGGTCCCT**GCG**CTGA 3'

**B**



**C**



**D**

## Synthetic CEB1-WT-1.0

Motif CEB1-WT:

5' GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA 3'

GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	1
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	2
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	3
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	4
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	5
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	6
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	7
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	8
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	9
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	10
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	11
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	12
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	13
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	14
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	15
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	16
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	17
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	18
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	19
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	20
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	21
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	22
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	23
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	24
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	25
GGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	26



# Figure S3

## E Synthetic CEB1-Gmut-1.7

Motif CEB1-Gmut:

```
5' GCGCGGAGTGAGGTGGCCTGCGGAGGTCCTGCGTGA 3'

GCGC-----GTGGCCTGCGGAGGTCGCTGCGCTGA      1
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCCTGCGCTGA      2
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCCTGCGCTGA      3
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA      4
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA      5
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA      6
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA      7
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA      8
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA      9
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCCTGCGCTGA     10
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA     11
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCCTGCGCTGA     12
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA     13
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCCTGCGCTGA     14
GCGCGGAGTGAGAGTGG---GCGGAG---CCCTGCGCTGA     15
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA     16
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCCTGCGCTGA     17
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCCTGCGCTGA     18
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCCTGCGCTGA     19
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA     20
GCGCAGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA     21
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA     22
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA     23
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA     24
GCGCGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA     25
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA     26
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA     27
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCCTGCGCTGA     28
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA     29
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCCTGCGCTGA     30
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA     31
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA     32
GCGCGGAGTGAGAGTGGCCTGCGGAGGTC---GCG---TGA     33
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA     34
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA     35
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA     36
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA     37
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCCTGCGCTGA     38
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCGCTGA     39
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCGCTGCG---TGA     40
GCGCGGAGTGAGAGTGGCCTG-----CCCTGCGCTGA     41
GCGCGGAGTGAGAGTGGCCTGCGGAGGTCCTGCGCTGA     42
```

**Article 2 :**

**Genetic Instability Triggered by G-Quadruplex  
Interacting Phen-DC Compounds in *Saccharomyces  
cerevisiae***



## **II.B. Article 2 : Genetic Instability Triggered by G-Quadruplex Interacting Phen-DC Compounds in *Saccharomyces cerevisiae***

### **II.B.1. Contexte**

L'approche par mutagenèse combinée à l'activité de déroulement des G-quadruplexes par Pif1 *in vitro* suggère fortement que ce sont les G-quadruplexes formés par CEB1 et non déroulés en absence de Pif1 qui sont responsables de son instabilité *in vivo*. Cependant, cette approche n'est pas une démonstration formelle : les mutations effectuées indiquent juste que les cinq guanines mutées, ou certaines des cinq guanines mutées, sont nécessaires à l'instabilité de CEB1. Ces mutations pourraient éventuellement affecter une autre propriété de CEB1 que sa capacité à former un G-quadruplexe, comme par exemple sa capacité à recruter une protéine inconnue. Pour démontrer plus formellement que c'est bien le G-quadruplexe qui est à l'origine de l'instabilité de CEB1, nous décidons d'utiliser une seconde approche, d'interférence chimique à l'aide de ligands de G-quadruplexe sur des cellules sauvages. Pour cela nous initiions une collaboration avec l'équipe de MP teulade-Fichou à l'Institut Curie qui synthétise des ligands très affins et sélectifs des G-quadruplexes : Phen-DC<sub>3</sub> et Phen-DC<sub>6</sub>. Réciproquement, notre système expérimental permet d'étudier d'éventuels effets aspécifiques de ces ligands, en présence de l'allèle CEB1 G-muté.



# Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in *Saccharomyces cerevisiae*

Aurèle Piazza<sup>1</sup>, Jean-Baptiste Boulé<sup>1</sup>, Judith Lopes<sup>1</sup>, Katie Mingo<sup>1,2</sup>, Eric Largy<sup>3</sup>, Marie-Paule Teulade-Fichou<sup>3</sup> and Alain Nicolas<sup>1,\*</sup>

<sup>1</sup>Recombinaison et Instabilité Génétique, Institut Curie Centre de Recherche, CNRS UMR3244, Université Pierre et Marie Curie, 26 rue d'Ulm, 75248 Paris Cedex 05, France, <sup>2</sup>Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA and <sup>3</sup>Institut Curie Centre de Recherche, CNRS UMR176, Université Paris XI, Bât. 110, 91405 Orsay, France

Received January 28, 2010; Revised and Accepted February 16, 2010

## ABSTRACT

G-quadruplexes are nucleic acid secondary structures for which many biological roles have been proposed but whose existence *in vivo* has remained elusive. To assess their formation, highly specific G-quadruplex ligands are needed. Here, we tested Phen-DC<sub>3</sub> and Phen-DC<sub>6</sub>, two recently released ligands of the bisquinolinium class. *In vitro*, both compounds exhibit high affinity for the G4 formed by the human minisatellite CEB1 and inhibit efficiently their unwinding by the yeast Pif1 helicase. *In vivo*, both compounds rapidly induced recombination-dependent rearrangements of CEB1 inserted in the *Saccharomyces cerevisiae* genome, but did not affect the stability of other tandem repeats lacking G-quadruplex forming sequences. The rearrangements yielded simple-deletion, double-deletion or complex reshuffling of the polymorphic motif units, mimicking the phenotype of the Pif1 inactivation. Treatment of Pif1-deficient cells with the Phen-DC compounds further increased CEB1 instability, revealing additional G4 formation per cell. In sharp contrast, the commonly used *N*-methyl-mesoporphyrin IX G-quadruplex ligand did not affect CEB1 stability. Altogether, these results demonstrate that the Phen-DC bisquinolinium compounds are potent molecular tools for probing the formation of G-quadruplexes *in vivo*, interfere with their processing and elucidate their biological roles.

## INTRODUCTION

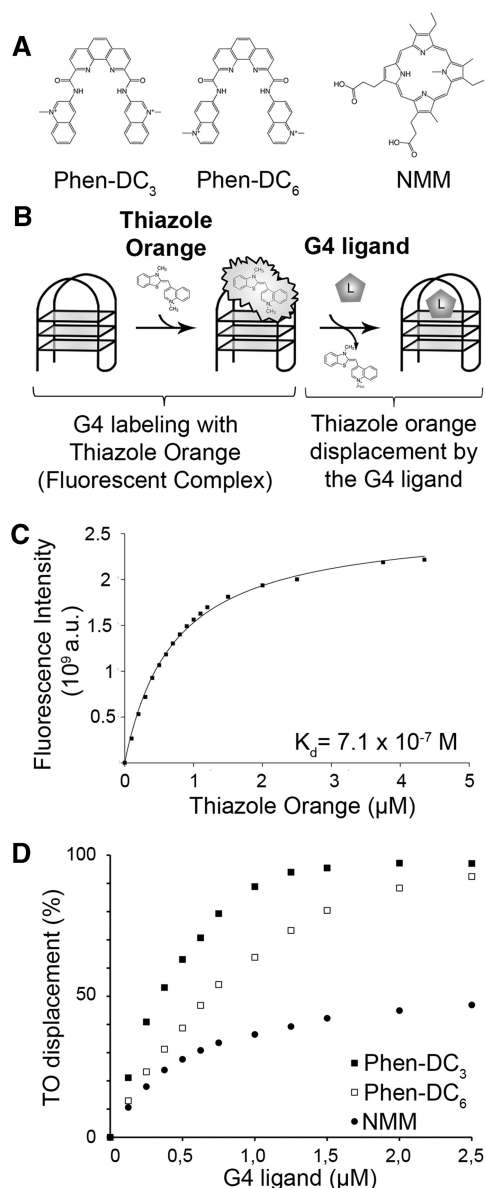
G-quadruplexes (G4) are four-stranded nucleic acid helical structures that spontaneously form *in vitro* within certain G-rich sequences. The unitary motif of this structure is composed of four guanines stabilized by non-canonical H-bonding in a coplanar arrangement (called a G-quartet) (1). The uninterrupted stacking of at least three G-quartets stabilized by monovalent cations (Na<sup>+</sup> or K<sup>+</sup>) is sufficient to form a G4. Quadruplexes can result from intra-molecular folding of one DNA strand containing four triplets of G separated by few bases or from intermolecular association of several strands and adopt a large variety of conformations, depending on the size and sequence of intervening loops, and have been documented by numerous structural studies (2).

Evidences concerning the *in vivo* formation of G4 and their involvement in several biological pathways remain limited but are starting to emerge (3–5). Formation of G4 in transcribed human G-rich DNA arrays in bacteria was visualized by electron microscopy (5). In ciliates, the formation of G4 was detected by immunochemistry (6). In this organism, the formation of G4 is triggered by specific telomere end-binding proteins which in turn regulates telomere protection from degradation and cell-cycle-dependent accessibility to telomerase (6,7). Using genetic approaches, the formation of G4 was shown to participate in the instability of the human CEB1 minisatellite inserted on a yeast chromosome (8), in a gene conversion pathway resulting in pilin antigenic variation in the bacteria *Neisseria gonorrhoeae* (9), and in the instability of guanine-rich regions in the *Caenorhabditis elegans* genome in absence of the *dog-1*

\*To whom correspondence should be addressed. Tel: +33 0 1 56 24 65 20; Fax: +33 0 1 56 24 66 44; Email: alain.nicolas@curie.fr

helicase (10). Complementarily, computational studies have provided a wealth of information concerning the occurrence and location within the genome of sequences having potential to form intra-molecular G4, as inferred from their primary DNA sequence (11). These potential G4-forming sequences are statistically over-represented at several loci, including telomeres of most eukaryotic organisms, at the rDNA loci in yeast (12) and human (13), and are significantly enriched in promoters sequences in human (14), yeast (12) and in *C. elegans* (15). These studies suggest that G4-DNA structures could exert a regulatory effect in *cis* on gene expression, either by recruiting factors at promoters or helping maintain a chromatin organization to favor or repress transcription. Some of these hypotheses have started to be submitted to experimental challenge, but it is still unclear how much of these potential G4 really form *in vivo* and how they specifically affect, for example, replication, transcription or recombination of genomic regions surrounding them.

To address the biological roles of G4, another approach of general use is the stabilization of these structures *in vivo* using specific ligands. The presence of G4-forming sequences at human telomeres and the fact that the first generation of G4-binding ligands were able to inhibit telomerase (16) have largely contributed to make G4 the archetypal higher order nucleic acid structure for the design of selectively targeting ligands in the presence of duplex DNA. Among the large number of ligands produced to date (17), the porphyrin derivatives *N*-methyl-mesoporphyrin IX (NMM, Figure 1A) and TmPyP4, and the perylene dimide derivative PIPER (18) are commercially available and thus have been often used for biochemical and biological studies. However, these two porphyrins are not optimal tools for probing quadruplex functions *in vivo* since NMM is selective for G4 over duplex DNA but is a relatively low-affinity ligand (19) and TmPyP4 has a high affinity but a poor selectivity for quadruplex DNA (20,21). Equally, the propensity of PIPER to aggregate in aqueous media (22) and its binding to duplex DNA renders its biological use questionable. The natural product telomestatin, which fulfills both requirements of selectivity and affinity, was promising and thus has been used to probe quadruplex structures *in vivo* and *in vitro* (23,24). However, telomestatin suffers several disadvantages such as poor water-solubility, chemical instability, and at present is only accessible by an arduous multi-step synthetic pathway (25) that makes large-scale use difficult. In this context, we recently developed the bisquinolinium family of compounds that includes Phen-DC<sub>3</sub> and Phen-DC<sub>6</sub> [Figure 1A, referred as compounds 2a and 2b respectively in ref. (26)], two promising molecules that emerged for their strong quadruplex stabilizing ability and an exquisite selectivity for quadruplex over duplex DNA. Indeed, the G4 recognition properties of Phen-DC<sub>3</sub> and Phen-DC<sub>6</sub> rival or surpass those of the best G4-binders such as Braco-19, telomestatin and their pyridine analogues (360A, 307A), which all exhibit a high selectivity for G4 and an affinity in the nanomolar range (27–29). In addition, the chemical stability and ease of preparation of our Phen-DC compounds are important advantages (27).



**Figure 1.** Fluorescent intercalator displacement of the G-quadruplex formed by the CEB1 motif (G4-CEB1) with Phen-DC<sub>3</sub>, Phen-DC<sub>6</sub> and NMM. (A) Phen-DC<sub>3</sub>, Phen-DC<sub>6</sub> and NMM structures. (B) Schematic representation of the FID assay. (C) TO association curve with G4-CEB1. (D) TO displacement by (filled square) Phen-DC<sub>3</sub>, (open square) Phen-DC<sub>6</sub> or (filled circle) NMM in presence of K<sup>+</sup>.

Although the Phen-DC compounds are very promising candidates to foster biological studies involving potential G4-forming sequences, their *in vivo* efficiency remains to be established (30). Therefore, we conducted *in vitro* and *in vivo* studies in the yeast *Saccharomyces cerevisiae* to assay the biological activity of the Phen-DC ligands with respect to the formation and the processing of G4, based on the ability of the human CEB1 minisatellite inserted in the yeast genome to efficiently form G4 structures and the ability of the yeast Pif1 helicase to resolve these structures (8). In budding yeast, the inactivation of Pif1, but not of other helicases, destabilizes the CEB1 tandem array and yields recombination-dependent size variants (8).

We show that *in vitro* both Phen-DC compounds bind the G4 structure formed by CEB1 with high affinity, and inhibit G4 unwinding by Pif1 efficiently and selectively. In WT cells, Phen-DC compounds specifically induce the instability of G4-prone CEB1 arrays, a phenomenon further increased in Pif1-deficient mutants. These results demonstrate that the Phen-DC compounds are exquisite molecular tools to probe the *in vivo* formation of G4, to interfere with their processing, and thereof provide mechanistic insights.

## MATERIALS AND METHODS

### G-quadruplex fluorescence intercalator displacement assay

The fluorescence intercalator displacement (FID) assay was performed in lithium cacodylate buffer solution (10 mM LiAsO<sub>2</sub>Me<sub>2</sub>) containing NaCl or KCl (100 mM) and adjusted to pH 7.2 with HCl. The fluorescence spectra were recorded with a HORIBA Jobin-Yvon Fluoromax-3<sup>®</sup> spectrofluorimeter in the wavelength range from 510 to 750 nm, in a 3-ml quartz cell (path length 1 cm), using the following experimental parameters: excitation wavelength, 501 nm; increment, 1 nm; optical slit widths, 3.0/3.0 nm; integration time, 0.1 s. The fluorescence spectrum of the buffer solution was recorded and was systematically subtracted from all following spectra. The solutions of DNA (0.25 μM) were mixed with thiazole orange (TO) (final concentration 0.5 μM) and reference fluorescence spectra were recorded. Increasing concentrations (from 0 to 10 molar equivalents, i.e. from 0 to 2.5 μM) of ligands to be tested were then added. After a 3-min shaking and equilibration period, a fluorescence spectrum was recorded for each ligand addition step. The degree of TO displacement (percentage of ligand-induced decrease of TO fluorescence) is calculated from the fluorescence area *FA* as:

$${}^{\text{TO}}\text{D}(\%) = 100 - (FA/FA_0) \times 100$$

where *FA*<sub>0</sub> is the fluorescence area of the reference spectrum. The percentage of displacement is then plotted as a function of the concentration of added compound.

### Helicase assays

Helicase assays were carried out by incubating 100 nM of *S. cerevisiae* Pif1 and 2 nM of Cy5-labeled G4 DNA or 1 nM forked DNA substrate at 35°C, in a buffer containing 20 mM Tris pH 7.5, 50 mM NaCl, 100 μg/ml bovine serum albumin, 2 mM DTT, 5 mM Mg<sup>2+</sup> and 4 mM ATP, and the indicated amount of the G4 binding ligands Phen-DC<sub>3</sub>, Phen-DC<sub>6</sub> or NMM. Reactions were pre-incubated at 35°C for 10 min and started by addition of 4 mM ATP. Reactions (10 μl) with G4-DNA substrates were stopped by addition of 2 μl stop buffer (17% Ficoll, 50 mM EDTA, 3 mg/ml Proteinase K) and further incubated at 35°C for 10 min. Reaction products were loaded on a 8% polyacrylamide non-denaturing gel and resolved by electrophoresis at 4°C and 10V/cm in TBE 1X buffer containing 10 mM NaCl (for G4-DNA formed in NaCl) or 10 mM KCl buffer (for G4-DNA formed in KCl). Reactions containing the forked DNA substrate

(10 μl) were stopped by addition of 2 μl of 17% Ficoll, 50 mM EDTA and 150 nM unlabelled fd20 oligonucleotide. Reaction products were loaded on a 10% polyacrylamide non-denaturing gel and resolved by electrophoresis at 4°C and 10V/cm in TBE 1× buffer. Gels were scanned with a Storm PhosphorImager (Molecular Dynamics) at 635 nm and quantified using ImageQuant software (GE Healthcare). Two independent lots of NMM were purchased from Frontier Scientific and assayed with similar outcome.

### Yeast strains and CEB1 minisatellites

The genotypes of the *S. cerevisiae* strains (S288C background) used here are reported in the Supplementary Table S1. The *rad51Δ* strain (AND1239-5C) containing CEB1-WT-1.7 was obtained after sporulation of the diploid obtained by crossing AND1202-11A (CEB1-WT-1.7 *pif1::KanMX*) with the BY4742 *rad51::KanMX* strain (31). The nucleotide sequences of the natural human CEB1-1.8 (see Supplementary Figure S6) and the synthetic CEB1-WT-1.7 and CEB1-Gmut-1.7 minisatellites, each containing 42 motifs, have been previously reported (8,32). The CEB1 arrays were inserted on the yeast chromosome VIII, upstream of the *ARG4* promoter (32).

### Induction of minisatellite instability

Cells taken from a fresh patch were suspended at a density of  $2 \times 10^5$  cells/ml into 5 ml of rich Yeast–Peptone–Dextrose (YPD) medium containing the indicated concentration of Phen-DC<sub>3</sub> and Phen-DC<sub>6</sub> and 1% DMSO, or in YPD containing only 1% DMSO (control treatment). Cells were grown for eight generations at 30°C with agitation, plated as individual colonies on YPD plates and grown at 30°C. For the induction in synthetic complete (SC) medium with NMM and Phen-DC<sub>3</sub>, the final concentration of DMSO reached in the liquid culture and in the control treatment was 0.4%. For the NMM treatment in SC medium, cells were grown overnight in presence of the drug, diluted to  $2 \times 10^5$  cells/ml in the same media and grown for an additional eight generations at 30°C, thus undergoing approximately four additional generations in presence of the drug compared to the control condition or the Phen-DC<sub>3</sub> treatment.

### Analysis of minisatellite instability

Colonies grown from treated cultures were inoculated in 96-well megaplaque in YPD for 24–48 h at 30°C. Pools of 4–16 colonies were made right before DNA extraction. DNA was digested with ApaI and SpeI to probe for the CEB1 minisatellite, or with AluI to probe for the *DAN4* and *FLO1* minisatellites (8). Digestion products were migrated in a 0.8% agarose–TBE 1× gel and analyzed by Southern blot using a radiolabeled probe corresponding to the minisatellite of interest. Blots were analyzed using a Storm PhosphorImager (Molecular Dynamics) and quantified using ImageQuant software (GE healthcare). To account for secondary rearrangements (occurring after plating) in the analysis, each visible band was quantified and normalized to the mean of the



intensity of the parental bands. Bands with signal intensity two times lower than expected for a primary rearrangement were removed from the counting. In numerous instances pools of colonies were depooled for verification and sequencing.

Additional experimental protocols are reported in Supplementary Material and Methods.

## RESULTS

### Phen-DC compounds strongly bind to the G-quadruplex form of CEB1

Our screening and initial analysis of a large series of potential G4-binding ligands was performed with two quadruplex-forming oligonucleotides: 22AG [5'-AG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>-3'] mimicking a human telomeric sequence and TBA [5'-G<sub>2</sub>T<sub>2</sub>G<sub>2</sub>TGTG<sub>2</sub>T<sub>2</sub>G<sub>2</sub>-3'], the thrombin binding aptamer sequence (27). The bisquinolinium-dicarboxamide derivatives Phen-DC<sub>3</sub> and Phen-DC<sub>6</sub> (illustrated in Figure 1A) were among the most selective G4 ligands.

Since G4 can fold differently depending on primary nucleotide sequence (33), we wished to evaluate the capacity of the Phen-DC compounds to bind to the quadruplex conformation adopted by the more complex human minisatellite CEB1 motif used in the present study. To this end, we folded a single-stranded oligonucleotide representing the prevalent 39nt CEB1 repeat motif (G4-CEB1) (8) and assayed its interaction with the Phen-DC and NMM compounds using the G4-FID assay (34), whose principle is illustrated in Figure 1B. This method is based on the competitive displacement of the fluorescent intercalator probe TO by a putative G4 ligand. TO being highly fluorescent when bound to DNA and virtually non-fluorescent when free in solution, the decrease of TO fluorescence as function of increasing ligand concentration thus serves to evaluate binding affinity of the ligand for G-quadruplex DNA (expressed as DC<sub>50</sub>: concentration required to decrease the fluorescence signal of 50%). As shown in Figure 1C, fluorimetric titration of TO by the G4-CEB1 substrate showed a strong fluorescence increase, indicating that TO efficiently bound to the G4-CEB1 conformation. The titration curve adopted a regular shape that could be fitted with a 1/1 stoichiometry model, giving a  $K_d$  value in the micromolar range ( $K_d = 7.1 \times 10^{-7}$  M). This behavior as well as the  $K_d$  value are similar to results obtained with the TBA and the human telomeric G4-DNA (27), thereby indicating that, in our experimental conditions, G4-CEB1 accommodates the probe at a single site corresponding to one of the two external G-quartets. The displacement of the TO by Phen-DC compounds was then examined in K<sup>+</sup> (Figure 1D) and Na<sup>+</sup> (Supplementary Figure S1) conditions. Both Phen-DC<sub>3</sub> and Phen-DC<sub>6</sub> isomers display a strong ability to displace TO ( $DC_{50} = 0.4-0.5 \mu\text{M}$ ) with complete displacement (100% quenching) being reached at a low ligand/quadruplex ratio (5 and 10 for Phen-DC<sub>3</sub> and Phen-DC<sub>6</sub>, respectively). Considering previous results obtained with the telomeric quadruplex and the low concentration of G4-DNA

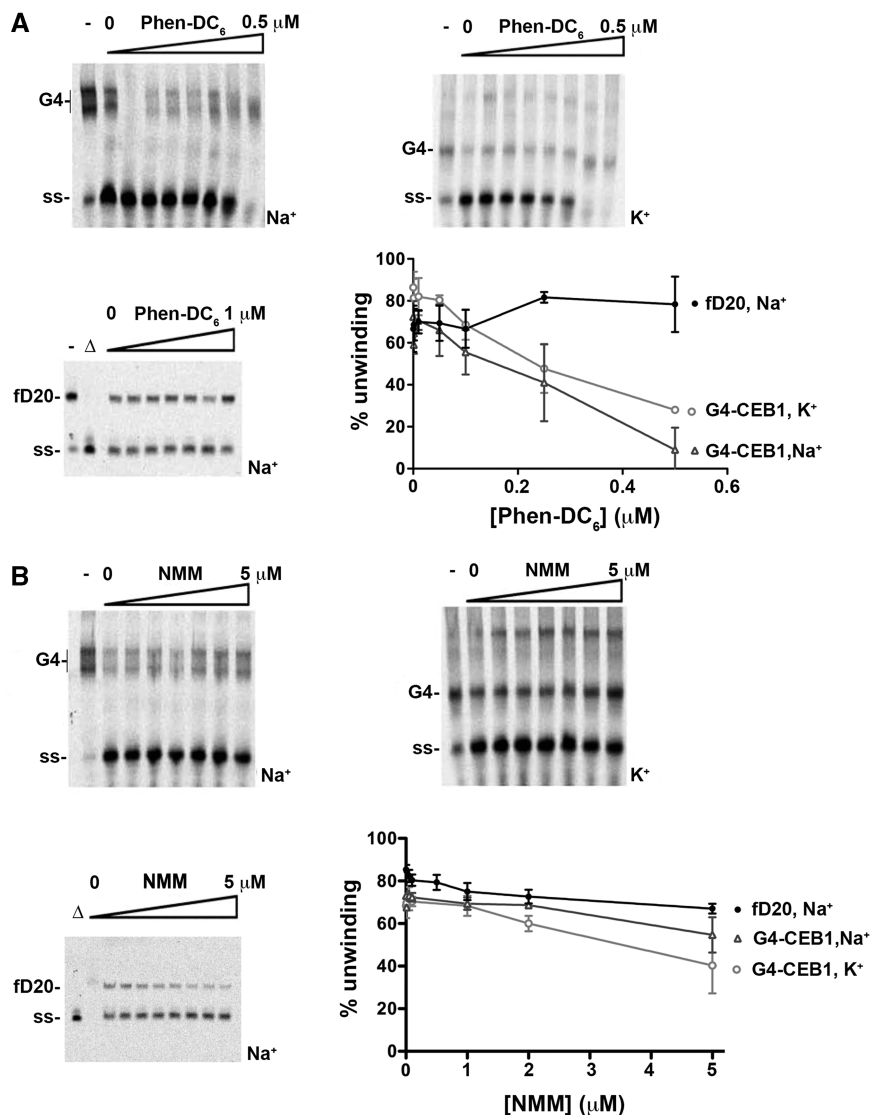
(0.25  $\mu\text{M}$ ) used in the test, these data unambiguously reveal a strong interaction between Phen-DC compounds and G4-CEB1 (Figure 1D, Supplementary Figure S1). In sharp contrast, the NMM was much less efficient in displacing TO since saturation was not reached even at high ligand/quadruplex ratio (>40). This result indicates that NMM exhibits a much lower affinity for G4-CEB1 than our Phen-DC compounds (Figure 1D).

### Phen-DC compounds specifically inhibit G4-DNA unwinding by the Pif1 helicase *in vitro*

Recently, we showed that the yeast Pif1 helicase efficiently unwinds the G4-CEB1 substrate (8). To investigate the ability of the two Phen-DC derivatives to interfere with such G4 unwinding enzymatic activities, we tested their effect on G4-CEB1 unwinding by Pif1 *in vitro*. The G4-CEB1 substrates were made in two conditions: i.e. in presence of either Na<sup>+</sup> or K<sup>+</sup> as a monovalent cation (see Supplementary Materials and Methods) (8). This led to intermolecular G4-CEB1 structures with different gel migration properties, likely corresponding to a different folding of the G4-DNA. As shown in Figure 2A, Pif1 efficiently unwound the G4-CEB1 substrates formed in either conditions but was rapidly inhibited by low concentrations of Phen-DC<sub>6</sub> [ $k_{i(\text{K}^+)} = 250$  nM;  $k_{i(\text{Na}^+)} = 300$  nM]. Phen-DC<sub>3</sub> showed comparable activity (Supplementary Figure S2).

For comparison, we also tested the inhibition of G4-CEB1 unwinding by NMM, reported to be a potent inhibitor of G4-DNA unwinding by the *Escherichia coli* RecQ helicase and its eukaryotic homologues Sgs1 and BLM helicases (35). Strikingly, NMM was fairly inefficient in inhibiting G4 DNA unwinding by Pif1. At 5  $\mu\text{M}$  concentration, the unwinding of G4-CEB1 by Pif1 was inhibited by ~50% (Figure 2B). Therefore, Phen-DC compounds are more effective than NMM to inhibit G4-CEB1 unwinding by Pif1, a result consistent with the difference in their respective binding ability as evaluated by the G4-FID assay.

Next, to address the specificity of Phen-DC compounds for G4-DNA, we tested their effect on the unwinding of a forked 20-mer double-stranded DNA (dsDNA) substrate (fD20). Phen-DC<sub>6</sub> had no effect on fD20 unwinding by Pif1 at concentrations up to 1  $\mu\text{M}$  concentration (Figure 2A). We could not test higher concentrations of the drug in this assay because higher concentrations of Phen-DC<sub>6</sub> resulted in non-specific aggregation of DNA substrates that were not resolved in the polyacrylamide gel, regardless of the substrate used (G4-DNA or dsDNA). This effect is reminiscent of DNA compaction by polyamines, a situation frequently encountered with cationic ligands (36,37). In comparison, NMM also exhibited specificity for G4-DNA substrates since unwinding of the forked substrate was not inhibited by high concentrations of NMM (5  $\mu\text{M}$ ) (Figure 2B). In conclusion, contrary to NMM, the Phen-DC compounds are active at low concentrations as potent and specific inhibitors of G4-CEB1 unwinding by Pif1 helicase *in vitro*.

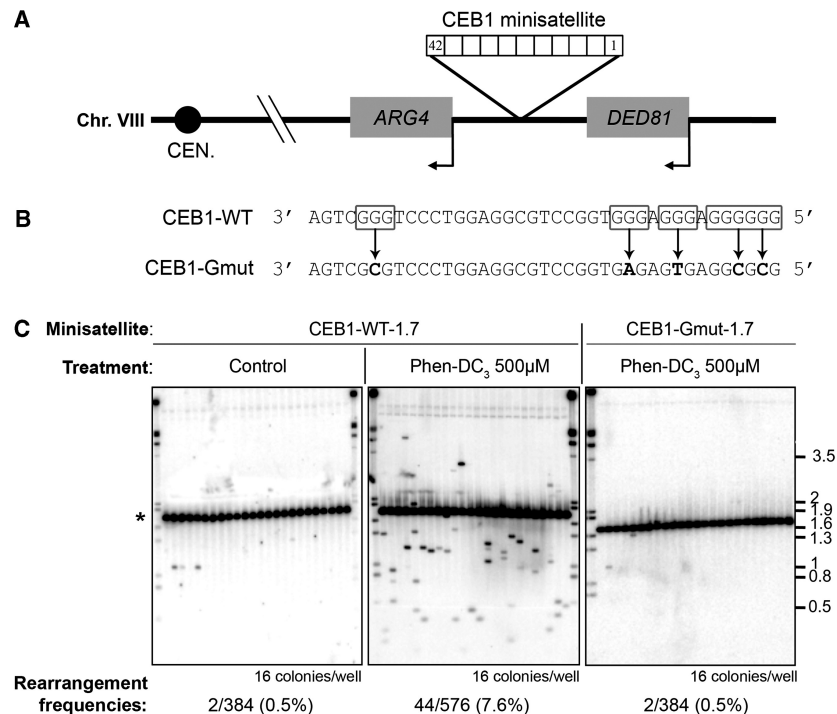


**Figure 2.** Inhibition of G4-CEB1 unwinding by Phen-DC<sub>6</sub> and NMM *in vitro*. (A) Pif1 unwinding of G4-CEB1 or forked DNA substrate (fD20) in presence of increasing concentrations of Phen-DC<sub>6</sub>. The G4-CEB1 substrates were formed in presence of Na<sup>+</sup> or K<sup>+</sup> as indicated. Quantifications (mean ± SD) from three independent experiments are shown. Open triangle indicates boiled substrate. Dash indicates absence of Pif1. Position of G4-CEB1 (G4) and unwound substrate (ss) are indicated. (B) Same as in (A), using NMM as a G4 ligand.

### The Phen-DC compounds induce genetic instability of the G4-prone CEB1 minisatellite

To test the effect of the Phen-DC compounds *in vivo* and their specificity towards G4-DNA substrates, we examined their capacity to trigger the instability of the G4-prone human CEB1 minisatellite inserted in the yeast genome (38) (Figure 3A). In WT yeast cells, the synthetic 1.7 kb CEB1 array (CEB1-WT-1.7) composed of 42 wild-type motifs of 39 nt (Figure 3B) is rather stable (8). It is highly destabilized in the absence of the Rad27/FEN1 endonuclease or of the Pif1 helicase, but in the latter case, only when the CEB1 repeats contains G4-forming sequences (8). Mutation of the G4-forming motifs in the synthetic CEB1 array (CEB1-Gmut-1.7, Figure 3B) stabilizes CEB1 in *pif1Δ* cells but not in *rad27Δ* cells (8). Thus, this system provides a well-defined genetic assay to monitor the formation of G4 structures *in vivo*.

We first treated WT cells carrying the synthetic CEB1-WT-1.7. The frequency of CEB1 instability was measured by the appearance of size variants (expansions or contractions) visualized by Southern blot analysis (Figure 3C). Treatment of cells grown in rich media (YPD) with increasing doses of Phen-DC<sub>3</sub> or Phen-DC<sub>6</sub> (up to 500 μM and 200 μM, respectively) had no effect on cell growth (Supplementary Figure S3A). In untreated cells or upon 1% DMSO control treatment (Phen-DC compounds are solubilized in pure DMSO), CEB1 rearrangements were rare: 2/708 and 2/384 respectively (Table 1). In contrast, upon treatment with 200 μM and 500 μM of Phen-DC<sub>3</sub>, the frequency of size variants increased to 4.2% (16/384 rearranged colonies) and 7.6% (44/576), respectively (Figure 3C, Table 1). This dose-response was statistically significant ( $P = 0.029$ ). Similarly, treatment with 200 μM Phen-DC<sub>6</sub> also



**Figure 3.** Phen-DC compounds trigger G-quadruplex-dependent CEB1 instability. (A) Schematic representation of the CEB1 minisatellite insertion in the chromosome VIII. (B) Sequences of CEB1-WT and CEB1-Gmut motifs. G-runs involved in G-quadruplex formation are boxed. Point mutations introduced in CEB1-Gmut motif and preventing G-quadruplex formation *in vitro* are shown in bold. (C) Southern blot analysis of WT cells carrying CEB1-WT-1.7 (strain AND1212-10D) or CEB1-Gmut-1.7 (AND1227-5C) after control (1% DMSO) or 500 μM Phen-DC<sub>3</sub> treatment. The number of colonies analyzed and the rearrangement frequencies are indicated below each gel. Size markers (kb) are indicated on the right. The position of the parental minisatellite alleles (42 repeats, 1.7 kb) is indicated by an asterisk.

**Table 1.** Frequencies of CEB1 rearrangements in WT cells

Strain:	AND1212-10D		AND1227-5C	
	CEB1-WT-1.7		CEB1-Gmut-1.7	
Minisatellite:	Rearrangt. Freq.	<i>P</i> -value versus control	Rearrangt. Freq.	<i>P</i> -value CEB1-WT-1.7 versus CEB1-Gmut-1.7
YPD media				
Untreated	2/708 (0.3%) <sup>a</sup>	NS	0/192 <sup>a</sup>	NS
Treatment				
Control	2/384 (0.5%)	NA	2/384 (0.5%)	NS
Phen-DC <sub>6</sub> 200 μM	18/384 (4.7%)	3.4e <sup>-4</sup>	1/384 (0.2%)	6.4e <sup>-5</sup>
Phen-DC <sub>3</sub> 200 μM	16/384 (4.2%)	1.2e <sup>-3</sup>	0/384	2.6e <sup>-5</sup>
Phen-DC <sub>3</sub> 500 μM	44/576 (7.6%)	9.1e <sup>-8</sup>	2/384 (0.5%)	9.1e <sup>-8</sup>
NMM 200 μM	0/384	NS	3/368 (0.8%)	NS
NMM 500 μM	0/192	NS	ND	NA
SC media				
Treatment				
Control	0/192	NA	ND	NA
Phen-DC <sub>3</sub> 2 μM	2/384 (0.5%)	NS	ND	NA
Phen-DC <sub>3</sub> 20 μM	27/384 (7%)	2.4e <sup>-5</sup>	3/384 (0.8%)	5.8e <sup>-6</sup>
NMM 20 μM	0/384	NS	ND	NA

<sup>a</sup>Data from ref. (8).

Rearrangement frequencies were compared using a two-tailed Fischer's exact test. NA, not applicable; ND, not determined; NS, not significant.

stimulated significantly the formation of CEB1-WT-1.7 rearrangements to yield a frequency of 4.7% (18/384) (Table 1), indicating that both compounds, structurally close to each other, behave similarly and yielded variants of diverse size. Lower concentrations of

Phen-DC were active in different media (see below). Again, treatment with 200 μM or 500 μM NMM did not stimulate CEB1-WT-1.7 instability (Table 1, Supplementary Figure S4). We concluded that the two Phen-DC isomers are equally able to enter the cells and

to trigger a biological response in the nucleus, namely to induce an 8-15 fold stimulation of CEB1 rearrangements.

Then, to test if the CEB1 instability triggered by the Phen-DC compounds relied specifically on the formation and/or stabilization of G4 structures, we carried out similar experiments with cells bearing a CEB1-Gmut-1.7 array. In contrast to the CEB1-WT-1.7 array, the CEB1-Gmut-1.7 array remained stable upon treatment with up to 500  $\mu\text{M}$  Phen-DC<sub>3</sub> or 200  $\mu\text{M}$  Phen-DC<sub>6</sub> (Figure 3C, Table 1). In these conditions and in both control experiments (untreated or incubated with 1% DMSO), the frequencies of CEB1 rearrangements are in the range of 0.5% or less, and are not significantly different from each other. Also, the treatment with 200  $\mu\text{M}$  NMM had no effect (Table 1). To further test the specificity of the Phen-DC compounds, we also examined the behavior of two natural yeast minisatellites DAN4 and FLO1, which are devoided of potential G4-forming sequences. Neither minisatellite was destabilized upon treatment with Phen-DC<sub>3</sub> or Phen-DC<sub>6</sub> (0/384 rearrangement, at 200  $\mu\text{M}$  in each case). Thus, CEB1 instability observed in WT cells upon treatment with Phen-DC compounds depends on the presence of G4-forming sequences, mimicking what is observed in *pif1* $\Delta$  cells (8).

Although consistent with its weak activity in our *in vitro* assays (Figure 2), the lack of effect of NMM on CEB1 stability is rather surprising since a lower dose of NMM (8  $\mu\text{M}$ ) was reported to significantly alter transcription of yeast genes bearing G4-forming sequences in their promoter regions (12). An important difference between this study and ours is that we used rich media (YPD) instead of SC media. Therefore, we re-examined the dose effect of NMM and Phen-DC compounds in cells grown in SC media. We found that NMM had a mild effect on growth at 40  $\mu\text{M}$ . In contrast, 20  $\mu\text{M}$  Phen-DC<sub>3</sub> or 8  $\mu\text{M}$  Phen-DC<sub>6</sub> was sufficient to reduce growth up to 50% (Supplementary Figure S3B). These effects were observed whether or not the strains carried the CEB1 minisatellite (Supplementary Figure S3B). Consequently, although we preferred to pursue our study using YPD media and concentrations of drugs not affecting cell growth, we nevertheless examined CEB1 instability upon drug treatment in SC media. We observed that the treatment of cells with 20  $\mu\text{M}$  NMM did not destabilize CEB1-WT-1.7 (0/384 rearrangement) (Table 1, Supplementary Figure S4). At 2  $\mu\text{M}$  Phen-DC<sub>3</sub>, CEB1-WT-1.7 remained stable. In contrast, upon treatment with 20  $\mu\text{M}$  Phen-DC<sub>3</sub>, we observed 7% instability (27/384) of CEB1-WT-1.7 and no instability of the CEB1-Gmut-1.7 allele (3/384 rearrangements,  $P = 5.8 \times 10^{-6}$ ) (Table 1, Supplementary Figure S5). Altogether, these results demonstrate that the Phen-DC molecules are active in both rich and synthetic complete medium.

#### The Phen-DC compounds increase CEB1 instability in Pif1-deficient strains

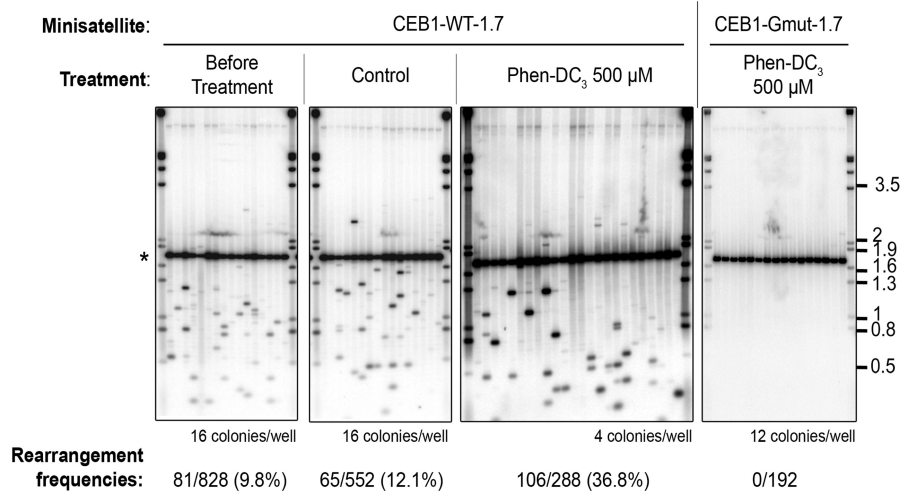
The occurrence of G4-dependent CEB1 rearrangements in treated wild-type cells revealed that some G4 stabilized by

the Phen-DC compounds escape from the unwinding activity of Pif1. This is consistent with the observations that these compounds inhibit G4-CEB1 unwinding by the Pif1 helicase *in vitro* (Figure 2), but whether or not Pif1 remove all or a fraction of the G4 that form in CEB1 remained to be evaluated. To address this question, we examined the frequency of CEB1-WT-1.7, CEB1-1.8 and CEB1-Gmut-1.7 rearrangements triggered by Phen-DC<sub>3</sub> or Phen-DC<sub>6</sub> in the absence of Pif1 (*pif1* $\Delta$  cells) and in a strain carrying a helicase-dead allele of *PIF1* [*pif1-K264A* (39)].

In our assays, the background of CEB1 rearrangements in WT cells is low (<1%) and roughly negligible compared to the rapid and strong effect of the Phen-DC compounds (Table 1), but this is not the case in *PIF1* mutant cells, in which the G4-prone CEB1 minisatellite is substantially unstable [see ref. (8) and present study]. Thus, to take into account CEB1 rearrangements which may pre-exists before the treatment, we estimated the effect of the G4-ligands in two steps. First, we measured the frequency of pre-existing rearrangements in the starting cell patches and then, as in WT cells, divided the assay culture in two samples: one being treated by the Phen-DC compounds and the other incubated in 1% DMSO alone (control treatment). Thus, upon subtraction of the pre-existing events (see Supplementary Materials and Methods for details on the procedure used for correction), we could estimate the frequency of rearrangements generated during the drug treatment only (eight generations). The *pif1* $\Delta$  cultures treated with 200  $\mu\text{M}$  Phen-DC<sub>6</sub> or 500  $\mu\text{M}$  Phen-DC<sub>3</sub> exhibited a high frequency of CEB1-WT-1.7 rearrangements, 20.9% and 36.8%, respectively, instead of 12.1% in the control treatment ( $P$ -values versus control treatment are  $4.1 \times 10^{-4}$  and  $2.7 \times 10^{-16}$ , respectively) (Figure 4, Table 2). Similarly, the frequency of CEB1-1.8 rearrangements raised from 5.2% to 12.2% ( $P = 1.1 \times 10^{-3}$ ) and 2.7% to 7.3%, ( $P = 5.5 \times 10^{-3}$ ) in the *pif1* $\Delta$  and *pif1-K264A* cells, respectively. Thus, taking into account the pre-existing events, estimated at 9.8% (CEB1-WT-1.7) and 2.1% (CEB1-1.8) in *pif1* $\Delta$  cell cultures, and at 1.3% (CEB1-1.8) in the *pif1-K264A* cell cultures, the treatments yielded a 3- to 12-fold increase of CEB1 rearrangements. Finally, as in WT cells, the Phen-DC compounds did not stimulate instability of the CEB1-Gmut-1.7 minisatellite (Figure 4). Altogether, the above results demonstrate that the Phen-DC compounds are able to stimulate CEB1 rearrangements in the presence or absence of Pif1, and in this latter case, uncover the *in vivo* formation of a higher number of G4 than in WT cells.

#### Sequence of CEB1 rearrangements induced by the Phen-DC compounds in WT and Pif1 deficient cells

To determine the nature of the rearrangements induced upon Phen-DC treatment in WT and *pif1* $\Delta$  cells, we used the naturally polymorphic CEB1-1.8 minisatellite in which the presence of multiple base substitutions along the 42 motifs allows a precise analysis of the contribution of the parental motifs in the sequenced size variants (32). As for the synthetic CEB1-WT-1.7 minisatellite, the treatment of WT cells grown in YPD with 500  $\mu\text{M}$  Phen-DC<sub>3</sub>



**Figure 4.** Treatment with Phen-DC<sub>3</sub> increases the instability of CEB1-WT-1.7 in *pif1Δ* cells. Southern blot analysis of CEB1-WT-1.7 (AND1202-11A) or CEB1-Gmut-1.7 (AND1206-4C-D11P2) instability in *pif1Δ* cells, before and after control (1% DMSO) or 500 μM Phen-DC<sub>3</sub> treatment. Other legends are as in Figure 3.

**Table 2.** Frequency of CEB1 rearrangements in *pif1Δ* cells

Strain	AND1202-11A CEB1-WT-1.7	AND1206-4C-D11P2 CEB1-Gmut-1.7	ORT4841 CEB1-1.8	ORT5087-5E CEB1-1.8
Mutation	<i>pif1Δ</i>	<i>pif1Δ</i>	<i>pif1Δ</i>	<i>pif1-K264A</i>
Before treatment	81/828 (9.8%)	1/383 (0.2%) <sup>a</sup>	4/192 (2.1%)	5/375 (1.3%)
Treatment				
Control	67/552 (12.1%)	ND	20/384 (5.2%)	10/375 (2.7%)
Phen-DC <sub>6</sub> 200 μM	77/368 (20.9%)	0/368	ND	ND
Phen-DC <sub>3</sub> 500 μM	106/288 (36.8%)	0/192	41/336 (12.2%)	26/358 (7.3%)

<sup>a</sup>Data from ref. (8).  
ND: not determined.

stimulated the rearrangements of the CEB1-1.8 allele ~14-fold (29/768 versus 5/1824 in the untreated cells,  $P = 1.8e^{-11}$ ) and yielded variants of diverse size. Thus, the presence of base polymorphisms located outside the triplets of G involved in CEB1 quadruplex formation does not interfere with the drug target. The sequences of 14 and 13 CEB1-1.8 contractions obtained upon treatment of WT and *pif1Δ* cells are reported in Supplementary Figure S6. They were different from each other. In both strains, simple deletions with one chimerical motif, double deletion with two chimerical motifs, and complex events with multiple chimerical motifs and various internal reshuffling of the parental polymorphic markers occurred (Figure 5A and B), suggesting that the underlying mechanisms of rearrangements were similar in WT and *pif1Δ* cells. No *de novo* mutagenic events were observed. Thus, the nature of the rearrangements obtained after treatment with Phen-DC<sub>3</sub> is diverse and reflects the same general pattern of events as in the untreated *pif1Δ* (8) and *rad27Δ* (32) mutants.

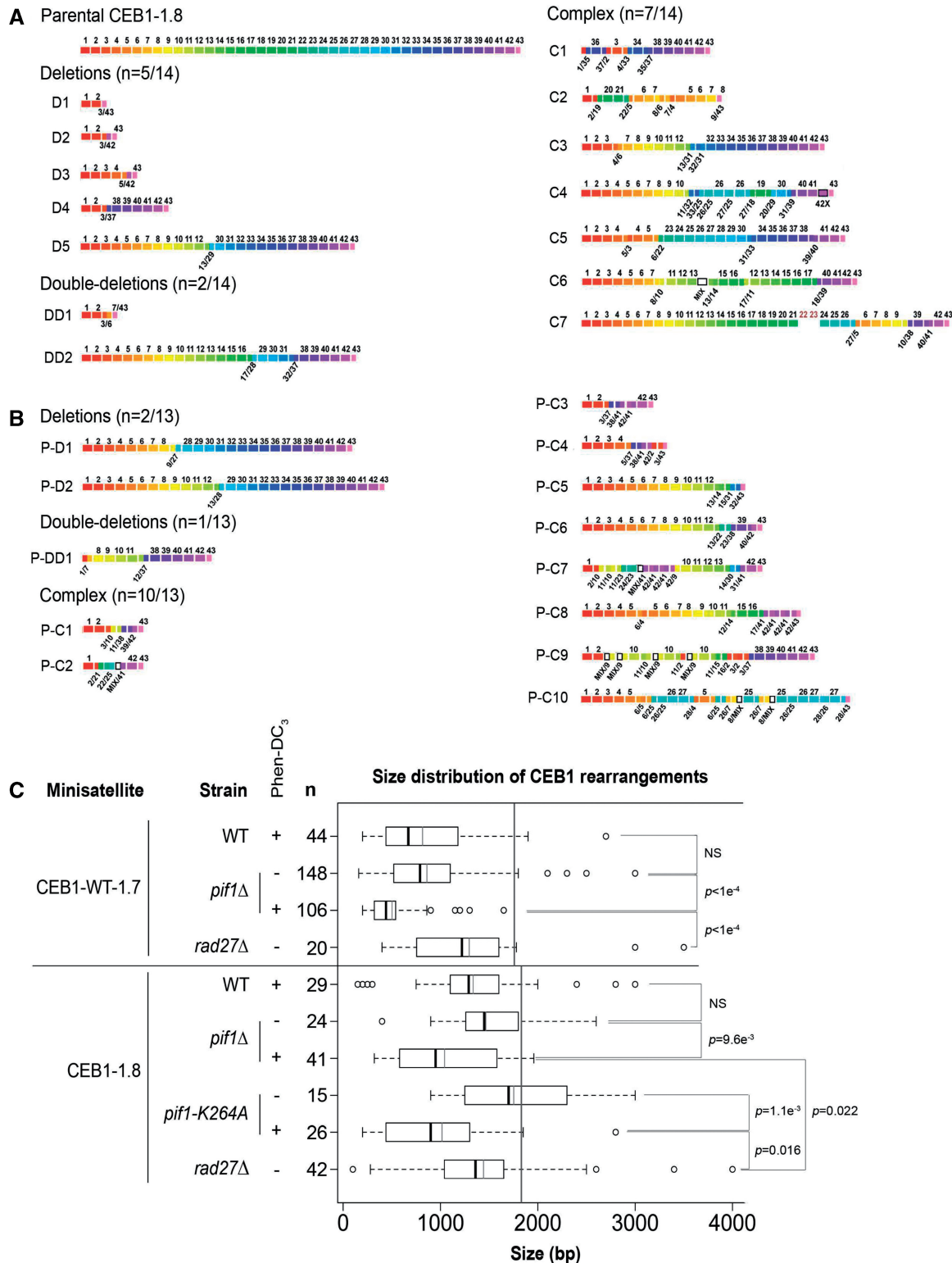
#### Phen-DC<sub>3</sub> treatment of Pif1-deficient cells yields unusually short CEB1 rearrangements

In comparison to the rearrangements observed in the absence of Pif1 alone, we noted that the rearrangements produced upon treatment of Pif1-deficient cells yielded a

larger number of CEB1 fragments of small size, corresponding to short variants of 15 or less repeat units (Figure 4). A comparative analysis of the size of the CEB1-WT-1.7 and CEB1-1.8 rearrangements obtained in *pif1Δ* or *pif1-K264A* cells with or without treatment is reported in Figure 5C. Strikingly, in all cases, despite the existence of pre-existing rearrangements, the mean size of the rearrangements was shorter in Phen-DC<sub>3</sub>-treated versus untreated Pif1-deficient cells ( $P \leq 9.6e^{-3}$ ). We also examined the size of the CEB1 variants produced in the *rad27Δ* cells. This cell exhibits a high level of instability for CEB1-WT-1.7 (30%) (8) and CEB1-1.8 (41/158, 25.9%) that is comparable to the instability measured in *pif1Δ*-treated cells (Table 2). The mean size of the CEB1 rearrangements in *rad27Δ* is much higher than in both *pif1Δ* and *pif1-K264A* treated cells ( $P \leq 2.2e^{-2}$ ) (Figure 5C). Hence, the increase in short alleles upon treatment of Pif1-deficient cells is not only a consequence of a high level of instability, but rather a specific effect of Phen-DC<sub>3</sub> treatment.

#### CEB1 rearrangements induced by Phen-DC compounds depend on the homologous recombination pathway

In previous studies, we demonstrated that in *rad27Δ* and *pif1Δ* cells the formation of CEB1 rearrangements depends on the process of homologous recombination,



**Figure 5.** Sequence and size of CEB1 rearrangements produced upon treatment with Phen-DC<sub>3</sub>. (A) Schematic representation of the natural CEB1-1.8 allele and of 14 rearrangements produced upon treatment of WT cells (ORT2914) with 500 μM Phen-DC<sub>3</sub>. Nucleotide sequences are reported in Supplementary Figure S6. Chimerical motifs combining the polymorphisms of more than three parental motifs (called ‘Mix’ in Supplementary Figure S6) are shown as a white box. Sequences of the two red motifs in C7 have not been determined. (B) Schematic representation of 13 rearrangements produced upon treatment of *pif1*Δ cells (ORT4841) with 500 μM Phen-DC<sub>3</sub>. (C) Size distribution of CEB1-WT-1.7 and CEB1-1.8 rearrangements in WT (AND1212-10D and ORT2914), *pif1*Δ (AND1202-11A and ORT4841), *pif1-K264A* (ORT5087-5E) or *rad27*Δ (AND1218-1A and ORD6713-8D) cells, with or without treatment with 500 μM Phen-DC<sub>3</sub>. Datasets are from this study, except the size distribution of rearranged CEB1-WT-1.7 alleles in *rad27*Δ cells, obtained from ref. (8) (strain AND1218-1A). The *n* indicates the number of variants analyzed. The vertical gray lines indicate the size of the parental alleles. For each distribution, dark bar: median; gray bar: distance between the first and third quartiles; whiskers: extreme values; white circles: outliers. *P*-values were obtained by comparing the samples with a Mann–Wilcoxon statistical test.

since it was abolished in the absence of the *RAD51* or *RAD52* genes (8,32). Similarly, we examined the behaviour of the CEB1-WT-1.7 array in a *rad51Δ* strain upon treatment with 200 μM Phen-DC<sub>3</sub> (Supplementary Figure S7). Compared to the WT strain (4.2% rearrangements, Table 1), we observed no rearrangements (0/384) in the absence of Rad51 ( $P = 2.6e^{-5}$ ). Therefore, Phen-DC treatment also leads to CEB1 rearrangements through the homologous recombination pathway.

## DISCUSSION

In this study, we investigated the *in vitro* and *in vivo* activities of two very potent G4 ligands: Phen-DC<sub>3</sub> and Phen-DC<sub>6</sub>. Based on *in vitro* biophysical and biochemical criteria, they are best ranked among current G4 ligands (26,27,29,30,40) and these promises clearly extend to their biological activity. We reported that the Phen-DC compounds: (i) exhibit a strong affinity for preformed G4-CEB1 structures *in vitro*, (ii) inhibit the unwinding by Pif1 of the resulting G4-ligand structure(s), (iii) manifest biological activity in intact yeast cells, (iv) destabilize the G4-prone but not the G4-mutated CEB1 array in WT cells, (v) further destabilize CEB1 in Pif1 deficient cells, (vi) induce recombination-dependent CEB1 size variants (expansions and contractions) similar to the spontaneous rearrangements produced in the absence of Pif1, and (vii) unexpectedly yield short size CEB1 variants. Thus, as hoped, the two Phen-DC compounds behave as potent biological drugs targeting G4 structures and interfering with their processing. To choose between the Phen-DC<sub>3</sub> and Phen-DC<sub>6</sub> molecules, the differences are subtle. Both are chemically stable, easy to prepare in large amounts (27), and qualitatively their *in vitro* and *in vivo* effects are similar as expected from their closely related chemical structures. The advantage of Phen-DC<sub>3</sub> is a slightly higher solubility in buffered solutions, allowing the use of higher drug concentrations in culture media.

To be optimal, a ligand should exhibit high selectivity and affinity for its substrate. Convincingly, the two Phen-DC derivatives fulfill both requirements. They bind with high affinity to various types of G-quadruplexes (27,40) and to the human minisatellite G4-CEB1 (Figure 1). This binding behavior is attributed to the crescent shape and the size of the Phen-DC scaffold that fit with the dimensions of a G-quartet, both features being favorable to an optimal  $\pi$ -overlap between the two aromatic surfaces (26).

We extensively compared the effects of the Phen-DC compounds to NMM, a molecule previously reported to bind G4 (19,41,42), to inhibit their unwinding by RecQ (43), Sgs1, and BLM (35) helicases *in vitro*, and to exert G4-related biological effects in budding yeast (12) and *N. gonorrhoeae* (9). Thus, we expected to observe an effect of NMM in our assays. The Phen-DC compounds act at nanomolar concentrations ( $K_i = 250$  nM in  $K^+$ ) as specific inhibitors of G4-CEB1 unwinding by Pif1 *in vitro* (Figure 2), which is a very active helicase on these substrates (8). In contrast, although our *in vitro*

experiments confirmed that NMM is a selective G4 ligand (as assessed by FID), it is inefficient in inhibiting the Pif1 helicase over a wide range of concentrations (Figure 2) and does not induce any CEB1 instability in yeast (Table 1, Supplementary Figure S4). The low affinity of NMM for G4-CEB1 structures can explain the discrepancies observed between treatment with NMM and Phen-DCs. An alternative and not mutually exclusive interpretation is that these molecules could recognize different G4 topologies and/or creates distinct G4-ligand structures, which would be differently processed by Pif1 or other endogenous helicases. Along these lines, the examination of the sequence of CEB1 motifs as well as their organization in a direct tandem array (placing all G triplets on the same strand) suggest that the CEB1 minisatellite has the potential to form topologically distinct intra- and inter-motifs G4 structures with heterogeneous loop sizes (11). Structurally, these two quolinium rings, which display a pronounced electron-acceptor character that contributes to the increased stacking interactions with guanines, have the potential to recognize various types of G4 *via* stacking on the external quartets terminating the quadruplex core. Thus, we envisage that the Phen-DC compounds may have a large spectrum of G4-structure targets (27,29,30), whereas NMM might be more specific of certain G4 structures (41).

The instability observed in treated wild type cells and its increase when Pif1 activity is compromised (revealing a higher number of G4) suggest that Pif1 greatly limits the effect of these G4 ligands *in vivo*. Mechanistically, this can be due to the unwinding of only a fraction of G4-CEB1 by Pif1 *in vivo*. In addition to the thermodynamic shift toward the folded state in presence of the ligand, which could trigger the formation and/or stabilization of G4 structures that would otherwise not stably form naturally, the persistency of G4 can be due to the inhibition of other potential G4-processing helicases. In *S. cerevisiae*, Sgs1 and Dna2 were reported to exhibit a G4-unwinding activity *in vitro* (44,45) but their *in vivo* activity on G4 substrates have not been demonstrated. We know that the absence of Sgs1 does not stimulate CEB1 instability, and that the deletion of the *DNA2* gene in a *pif1Δ* strain has the same effect as the inactivation of Pif1 alone (8). Further studies will be undertaken to decipher redundant and overlapping G4 processing pathways in wild-type, single and multi-mutant contexts.

The structure of the CEB1 rearrangements produced in Phen-DC treated and untreated *pif1Δ* and *rad27Δ* cells are of similar types, including simple deletion, double deletions and a majority of complex events. They are interpreted to be the consequence of the repair of an initiating lesion through the homologous recombination pathway, in which multiple cycles of misaligned synthesis-dependent strand annealing (SDSA) (46) reactions generate the various reshuffling of the parental minisatellite (8,32) [see illustration Figure 6 in ref. (32)]. However, we noted that the CEB1-WT-1.7 rearrangements obtained in the *pif1Δ*-treated cells were unusually short, suggesting that the Phen-DC treatment was not exactly mimicking the untreated situations. To confront this unexpected

observation, we made a comparative analysis of the size of the CEB1-WT-1.7 and CEB1-1.8 rearrangements obtained in the *pif1Δ* and the *pif1-K264A* mutants with or without treatment. Strikingly, in all pairwise comparisons of strains with the same genotype, the rearranged variants were shorter in the treated sample (Figure 5C). Why does the treatment of Pif1-deficient cells yield extensive contractions? An attractive possibility is that the concomitant formation (47) and the persistency of several G4 on the same strand may cause multiple distant lesions favoring large internal deletions. Mechanistically, these mutagenic events may be reminiscent of the deletions of G4-prone sequences reported in *C. elegans* mutated for the FANC-J helicase homolog *dog-1* (10). The lesion initiating the CEB1 rearrangements could be a single-strand nick occurring within the G4-prone region as observed in *N. gonorrhoeae* (9), or a single-strand gap or a double-strand break resulting from the persistency of unprocessed intra-molecular G4 accumulating in CEB1 during replication, transcription or other biological processes generating single-stranded DNA.

To conclude, the present study demonstrates that the Phen-DC<sub>3</sub> and Phen-DC<sub>6</sub> bisquinolinium compounds are reliable and potent small molecules to specifically target and trigger a G-quadruplex-dependent biological process, namely tandem-repeat instability. The independent and combinatorial use of genetic and chemical interfering approaches provides new perspectives to probe the G4-forming sequences in yeast and other organisms (11).

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank members of our laboratories, A. Londono-Vallejo, J.-L. Mergny and E. Blackburn for helpful discussions; D. Monchaud for preparing the Phen-DC compounds; G. Millot for advices on statistical analyses.

## FUNDING

Institut Curie; the Centre National de la Recherche Scientifique; and La Ligue Nationale contre le Cancer 'Equipe Labellisée LIGUE 2007' (to A.N.); by an E.U.; FP6 'MolCancerMed' (LSHC-CT-2004-502943 to M.-P., T.-F.); graduate student fellowship from the Ministère de l'Éducation Nationale, de la Recherche et de la Technologie (to A.P.); a post-doctoral fellowship from the Association pour la Recherche sur le Cancer (to J.B.B.); Massachusetts Institute of Technology - Institut Curie exchange program fellowship (to K.M.); a BDI graduate student fellowship from the Institut Curie and the Centre National de la Recherche Scientifique (to E.L.). Funding for open access charge: Institut Curie.

*Conflict of interest statement.* None declared.

## REFERENCES

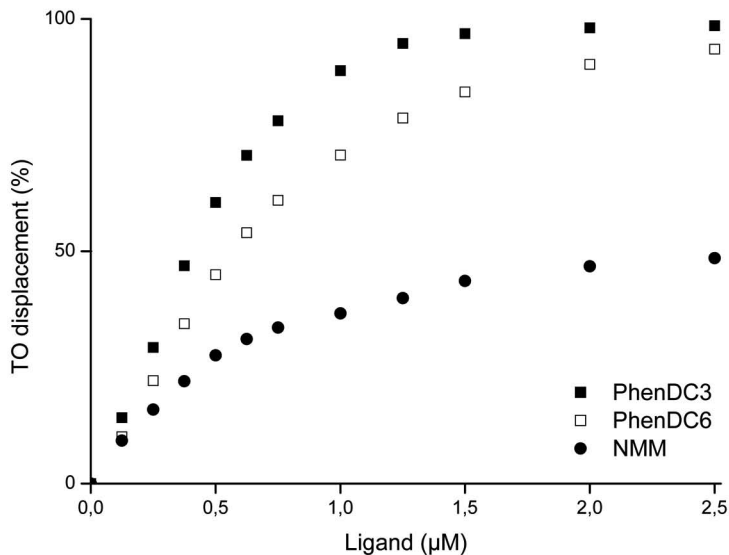
- Gellert, M., Lipsett, M.N. and Davies, D.R. (1962) Helix formation by guanylic acid. *Proc. Natl Acad. Sci. USA*, **48**, 2013–2018.
- Phan, A.T., Kuryavyi, V., Luu, K.N. and Patel, D.J. (2006) Structural diversity of G-quadruplex scaffolds. In Neidle, S. and Balasubramanian, S. (eds), *Quadruplex Nucleic Acids*. Ch. 3, RSC publishing, Cambridge, UK, pp. 81–99.
- Lipps, H.J. and Rhodes, D. (2009) G-quadruplex structures: *In vivo* evidence and function. *Trends Cell. Biol.*, **19**, 414–422.
- Maizels, N. (2006) Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat. Struct. Mol. Biol.*, **13**, 1055–1059.
- Duquette, M.L., Handa, P., Vincent, J.A., Taylor, A.F. and Maizels, N. (2004) Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.*, **18**, 1618–1629.
- Paeschke, K., Juraneck, S., Simonsson, T., Hempel, A., Rhodes, D. and Lipps, H.J. (2008) Telomerase recruitment by the telomere end binding protein- $\beta$  facilitates G-quadruplex DNA unfolding in ciliates. *Nat. Struct. Mol. Biol.*, **15**, 598–604.
- Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D. and Lipps, H.J. (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures *in vivo*. *Nat. Struct. Mol. Biol.*, **12**, 847–854.
- Ribeyre, C., Lopes, J., Boule, J.B., Piazza, A., Guedin, A., Zakian, V.A., Mergny, J.L. and Nicolas, A. (2009) The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences *in vivo*. *PLoS Genet.*, **5**, e1000475.
- Cahoon, L.A. and Seifert, H.S. (2009) An alternative DNA structure is necessary for pilin antigenic variation in *Neisseria gonorrhoeae*. *Science*, **325**, 764–767.
- Kruisselbrink, E., Guryev, V., Brouwer, K., Pontier, D.B., Cuppen, E. and Tijsterman, M. (2008) Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCD1-defective *C. elegans*. *Curr. Biol.*, **18**, 900–905.
- Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K. and Neidle, S. (2006) Quadruplex DNA: Sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.
- Hershman, S.G., Chen, Q., Lee, J.Y., Kozak, M.L., Yue, P., Wang, L.S. and Johnson, F.B. (2008) Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **36**, 144–156.
- Hanakahi, L.A., Sun, H. and Maizels, N. (1999) High affinity interactions of nucleolin with G-G-paired rDNA. *J. Biol. Chem.*, **274**, 15908–15912.
- Huppert, J.L. and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 406–413.
- Wong, H.M. and Huppert, J.L. (2009) Stable G-quadruplexes are found outside nucleosome-bound regions. *Mol. Biosyst.*, **5**, 1713–1719.
- Sun, D., Thompson, B., Cathers, B.E., Salazar, M., Kerwin, S.M., Trent, J.O., Jenkins, T.C., Neidle, S. and Hurley, L.H. (1997) Inhibition of human telomerase by a G-quadruplex-interactive compound. *J. Med. Chem.*, **40**, 2113–2116.
- De Cian, A., Lacroix, L., Douarre, C., Temime-Smaali, N., Trentesaux, C., Riou, J.-F. and Mergny, J.-L. (2008) Targeting telomeres and telomerase. *Biochimie*, **90**, 131–155.
- Fedoroff, O.Y., Salazar, M., Han, H., Chemeris, V.V., Kerwin, S.M. and Hurley, L.H. (1998) NMR-based model of a telomerase-inhibiting compound bound to G-quadruplex DNA. *Biochemistry*, **37**, 12367–12374.
- Ren, J. and Chaires, J.B. (1999) Sequence and structural selectivity of nucleic acid binding ligands. *Biochemistry*, **38**, 16067–16075.
- Parkinson, G.N., Ghosh, R. and Neidle, S. (2007) Structural basis for binding of porphyrin to human telomeres. *Biochemistry*, **46**, 2390–2397.
- Ping, W., Lige, R., Hanping, H., Feng, L., Xiang, Z. and Zheng, T. (2006) A phenol quaternary ammonium porphyrin as a potent telomerase inhibitor by selective interaction with quadruplex DNA. *ChemBioChem*, **7**, 1155–1159.



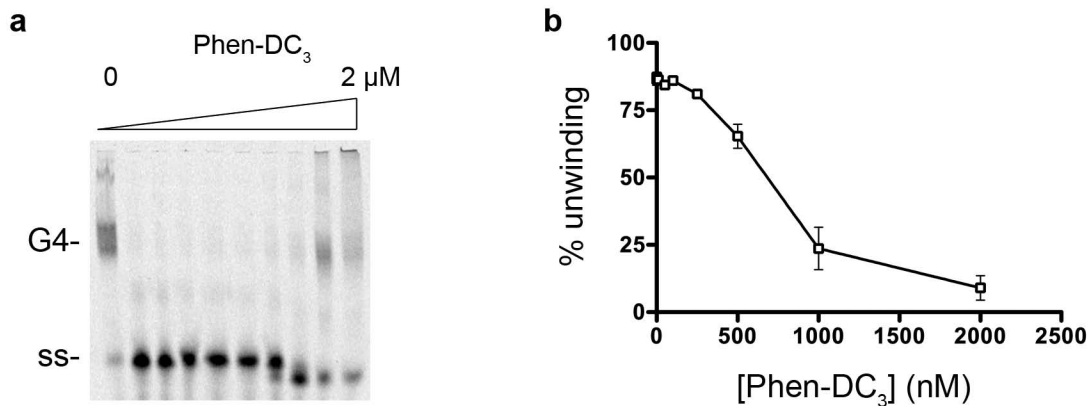
22. Kern, J.T., Thomas, P.W. and Kerwin, S.M. (2002) The relationship between ligand aggregation and G-quadruplex DNA selectivity in a series of 3,4,9,10-perylenetetracarboxylic acid diimides. *Biochemistry*, **41**, 11379–11389.
23. De Cian, A., Guittat, L., Shin-ya, K., Riou, J.-F. and Mergny, J.-L. (2005) Affinity and selectivity of G4 ligands measured by FRET. *Nucleic Acids Symp. Ser.*, **49**, 235–236.
24. Temime-Smaali, N., Guittat, L., Sidibe, A., Shin-ya, K., Trentesaux, C. and Riou, J.-F.O. (2009) The G-quadruplex ligand telomestatin impairs binding of topoisomerase III to G-quadruplex-forming oligonucleotides and uncaps telomeres in alt cells. *PLOS ONE*, **4**, e6919.
25. Doi, T., Yoshida, M., Shin-ya, K. and Takahashi, T. (2006) Total synthesis of (r)-telomestatin. *Org. Lett.*, **8**, 4165–4167.
26. De Cian, A., DeLemos, E., Mergny, J.L., Teulade-Fichou, M.P. and Monchaud, D. (2007) Highly efficient G-quadruplex recognition by bisquinolinium compounds. *J. Am. Chem. Soc.*, **129**, 1856–1857.
27. Monchaud, D., Allain, C., Bertrand, H., Smargiasso, N., Rosu, F., Gabelica, V., De Cian, A., Mergny, J.L. and Teulade-Fichou, M.P. (2008) Ligands playing musical chairs with G-quadruplex DNA: A rapid and simple displacement assay for identifying selective G-quadruplex binders. *Biochimie*, **90**, 1207–1223.
28. De Cian, A., Cristofari, G., Reichenbach, P., De Lemos, E., Monchaud, D., Teulade-Fichou, M.P., Shin-Ya, K., Lacroix, L., Lingner, J. and Mergny, J.L. (2007) Reevaluation of telomerase inhibition by quadruplex ligands and their mechanisms of action. *Proc. Natl Acad. Sci. USA*, **104**, 17347–17352.
29. De Cian, A., Grellier, P., Mouray, E., Depoix, D., Bertrand, H., Monchaud, D., Teulade-Fichou, M.P., Mergny, J.L. and Alberti, P. (2008) Plasmodium telomeric sequences: structure, stability and quadruplex targeting by small compounds. *ChemBiochem.*, **9**, 2730–2739.
30. Henn, A., Joachimi, A., Goncalves, D.P., Monchaud, D., Teulade-Fichou, M.P., Sanders, J.K. and Hartig, J.S. (2008) Inhibition of dicing of guanosine-rich shRNAs by quadruplex-binding compounds. *ChemBiochem.*, **9**, 2722–2729.
31. Winzler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H. *et al.* (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science*, **285**, 901–906.
32. Lopes, J., Ribeyre, C. and Nicolas, A. (2006) Complex minisatellite rearrangements generated in the total or partial absence of Rad27/hFEN1 activity occur in a single generation and are Rad51 and Rad52 dependent. *Mol. Cell. Biol.*, **26**, 6675–6689.
33. Neidle, S. (2009) The structures of quadruplex nucleic acids and their drug complexes. *Curr. Op. Struct. Biol.*, **19**, 239–250.
34. Monchaud, D., Allain, C. and Teulade-Fichou, M.-P. (2006) Development of a fluorescent intercalator displacement assay (G4-FID) for establishing quadruplex-DNA affinity and selectivity of putative ligands. *Bioorg. Med. Chem. Lett.*, **16**, 4842–4845.
35. Huber, M.D., Lee, D.C. and Maizels, N. (2002) G4 DNA unwinding by BLM and Sgs1p: Substrate specificity and substrate-specific inhibition. *Nucleic Acids Res.*, **30**, 3954–3961.
36. Raspaud, E., Olvera de la Cruz, M., Sikorav, J.L. and Livolant, F. (1998) Precipitation of DNA by polyamines: A polyelectrolyte behavior. *Biophys. J.*, **74**, 381–393.
37. Slama-Schwok, A., Peronnet, F., Hantz-Brachet, E., Taillandier, E., Teulade-Fichou, M.P., Vigneron, J.P., Best-Belpomme, M. and Lehn, J.M. (1997) A macrocyclic bis-acridine shifts the equilibrium from duplexes towards DNA hairpins. *Nucleic Acids Res.*, **25**, 2574–2581.
38. Lopes, J., Debrauwere, H., Buard, J. and Nicolas, A. (2002) Instability of the human minisatellite CEB1 in *rad27A* and *dna2-1* replication-deficient yeast cells. *EMBO J.*, **21**, 3201–3211.
39. Zhou, J., Monson, E.K., Teng, S., Schulz, V.P. and Zakian, V.A. (2000) Pif1p helicase, a catalytic inhibitor of telomerase in yeast. *Science*, **289**, 771–774.
40. De Cian, A. (2007) Identification and characterization of G-quadruplex ligands: to target telomeres and/or telomerase?. Université Pierre et Marie Curie - Paris VI, Paris.
41. Arthanari, H., Basu, S., Kawano, T.L. and Bolton, P.H. (1998) Fluorescent dyes specific for quadruplex DNA. *Nucleic Acids Res.*, **26**, 3724–3728.
42. Li, Y., Geyer, C.R. and Sen, D. (1996) Recognition of anionic porphyrins by DNA aptamers. *Biochemistry*, **35**, 6911–6922.
43. Wu, X. and Maizels, N. (2001) Substrate-specific inhibition of RecQ helicase. *Nucleic Acids Res.*, **29**, 1765–1771.
44. Sun, H., Bennett, R.J. and Maizels, N. (1999) The *Saccharomyces cerevisiae* Sgs1 helicase efficiently unwinds G-G-paired DNAs. *Nucleic Acids Res.*, **27**, 1978–1984.
45. Masuda-Sasa, T., Polaczek, P., Peng, X.P., Chen, L. and Campbell, J.L. (2008) Processing of G4 DNA by Dna2 helicase/nuclease and RPA provides insights into the mechanism of Dna2/RPA substrate recognition. *J. Biol. Chem.*, **36**, 24359–24373.
46. Pâques, F. and Haber, J.E. (1999) Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.*, **63**, 349–404.
47. Renciuik, D., Kejnovska, I., Skolakovska, P., Bednarova, K., Motlova, J. and Vorlickova, M. (2009) Arrangements of human telomere DNA quadruplex in physiologically relevant K<sup>+</sup> solutions. *Nucleic Acids Res.*, **37**, 6625–6634.

## **SUPPLEMENTARY DATA**

## **SUPPLEMENTARY FIGURES**

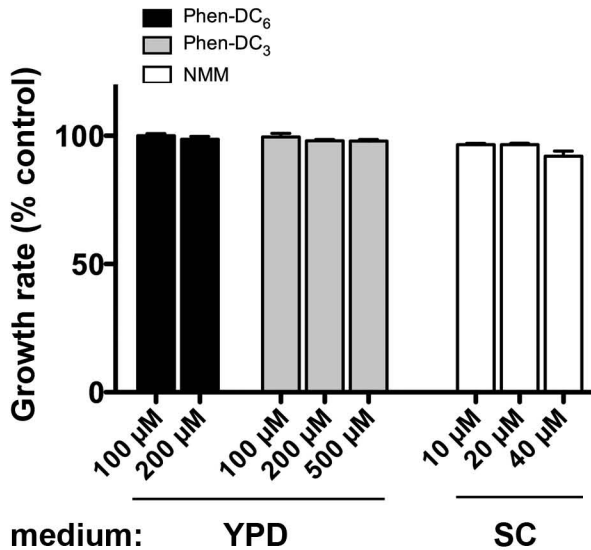


**Supplementary Figure 1:** Association of Phen-DC<sub>3</sub>, Phen-DC<sub>6</sub>, and NMM to G4-CEB1. Thiazole Orange (TO) displacement by increasing concentrations of (■) Phen-DC<sub>3</sub>, (□) Phen-DC<sub>6</sub> or (●) NMM in presence of Na<sup>+</sup>.

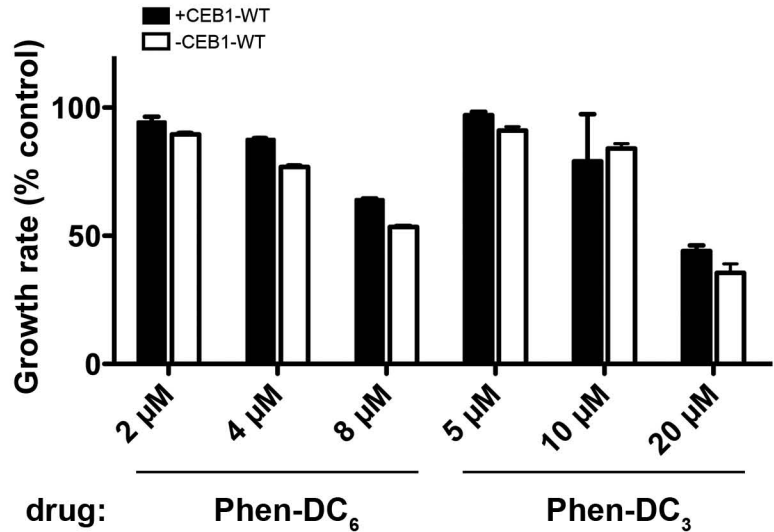


**Supplementary Figure 2:** Inhibition of G4-CEB1 unwinding by Phen-DC<sub>3</sub> *in vitro*. (a) Pif1 unwinding of G4-CEB1 in presence of increasing concentrations of Phen-DC<sub>3</sub>. The G4-CEB1 substrates were formed in presence of Na<sup>+</sup>. Position of G4-CEB1 (G4) and unwound substrate (ss) are indicated. (b) Quantifications (mean  $\pm$  SD) from three independent experiments.

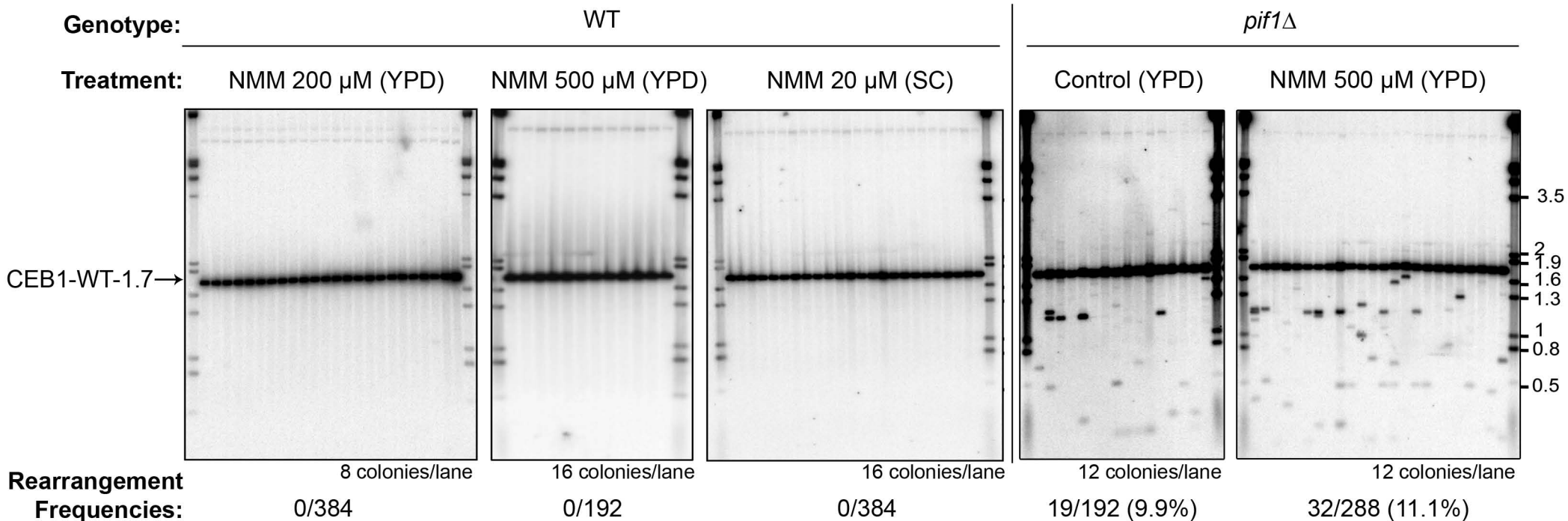
a



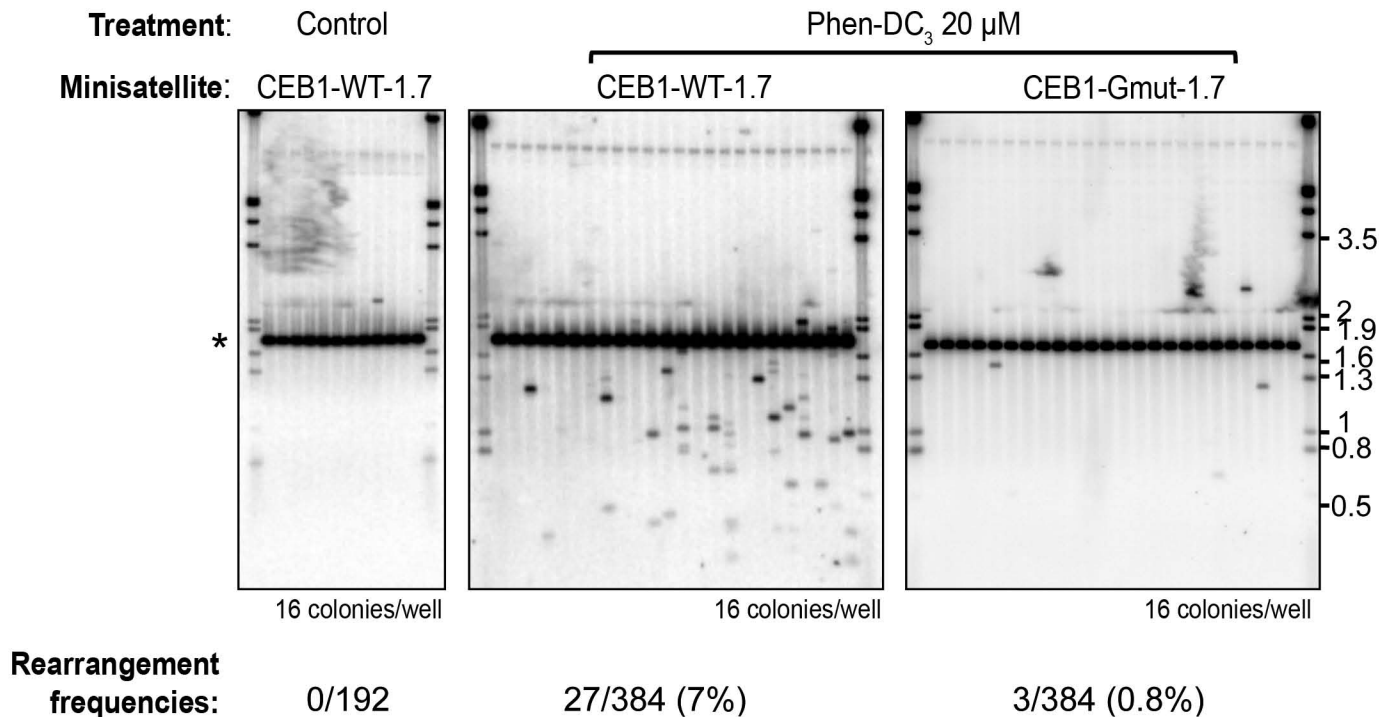
b



**Supplementary Figure 3:** Effect of Phen-DC<sub>6</sub>, Phen-DC<sub>3</sub>, and NMM on growth of WT cells. **(a)** Mean of the exponential growth rates (doubling time during exponential growth) of WT cells (AND1212-10D) in presence of indicated concentrations of Phen-DC<sub>6</sub>, Phen-DC<sub>3</sub> or NMM, in YPD or SC medium, expressed as a percentage of exponential growth rate in control conditions (1% DMSO). Error bars show standard deviation of 2 independent experiments. **(b)** Mean of the exponential growth rates of WT cells carrying (AND1212-10D) or not (AND1222-1D) CEB1-WT-1.7 in SC medium with the indicated concentrations of Phen-DC<sub>6</sub> (n=2 experiments) or Phen-DC<sub>3</sub> (n=4), expressed as a percentage of exponential growth rate in control conditions (1% DMSO). Error bars show standard deviation.



**Supplementary Figure 4:** Southern blot analysis of CEB1-WT-1.7 instability in wild-type and *pif1*Δ cells treated with NMM. WT cells carrying CEB1-WT-1.7 (AND1212-10D) were grown for 8 generations in YPD or SC in presence of the indicated concentrations of NMM. For the treatment in SC, a preculture of 4 generations was made with NMM before the 8 generations. *pif1*Δ cells carrying CEB1-WT-1.7 (AND1202-11A) were grown in YPD for 8 generations with or without 500 μM NMM. The number of colonies analyzed and the rearrangement frequencies are indicated below each gel. Size markers (kb) are indicated on the right. An arrow indicates the position of the parental CEB1-WT-1.7 allele.

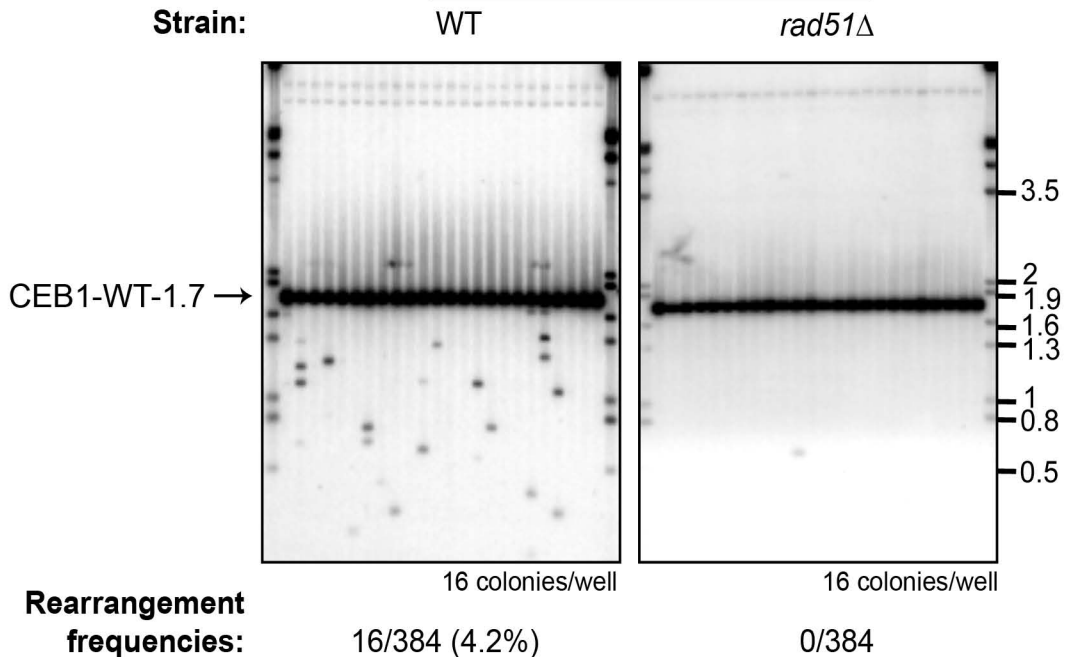


**Supplementary Figure 5:** Southern blot analysis of CEB1-WT-1.7 and CEB1-Gmut-1.7 instability upon treatment with 20 μM Phen-DC<sub>3</sub> in SC medium. WT cells carrying CEB1-WT-1.7 (AND1212-10D) or CEB1-Gmut-1.7 (AND1227-5C) were grown in presence of 0.4% DMSO (control treatment) or 20 μM Phen-DC<sub>3</sub> as indicated. The number of colonies analyzed and the rearrangement frequencies are indicated below each southern blot. Size markers (kb) are indicated on the right. The position of the parental minisatellite alleles (42 repeats, 1.7kb) is indicated by an asterisk.

**Supplementary Figure 6:** Sequences of the parental minisatellite CEB1-1.8 and of rearrangements obtained upon 500  $\mu$ M Phen-DC<sub>3</sub> treatment of WT (ORT2914) and *pif1* $\Delta$  (ORT4841) cells. The structure of the rearrangements is schematized in **Fig. 5a** and **b**, respectively. G-strand is shown, in the 3'-5' orientation. Polymorphisms between motifs are highlighted: substitution of the more common base by a A is shown in green; by a C in pink; by a G in blue; by a T in red; and base deletion are shown in yellow. Chimerical motifs combining the polymorphisms of more than 3 parental motifs are called "Mix". The Motif 42X in rearrangement C4 correspond to the sequence of the motif 42, with two additional repetitions of the terminal AGGG repeats highlighted in purple. The motifs 23 and 24 indicated by a star in C7 have not been sequenced, but the size determination of the band by Southern blot and PCR is sufficiently accurate to estimate that 2 motifs are missing.



Phen-DC<sub>3</sub> 200 μM



**Supplementary Figure 7:** CEB1 rearrangements are Rad51-dependent. Southern blot analysis of CEB1-WT-1.7 instability in WT (AND1212-10D) and *rad51*Δ cells (AND1239-5C) upon 200 μM Phen-DC<sub>3</sub> treatment. Size markers (kb) are indicated on the right. The rearrangement frequencies and the number of colonies analyzed in each lane are given below each gel. The position of the parental CEB1-WT-1.7 allele (42 repeats, 1.7kb long) is indicated by an arrow.

**Piazza et al.**

**Supplementary Table 1: Relevant Genotypes of yeast strains**

<b>Strain</b>	<b>Relevant genotype</b>	<b>Source</b>
ORT2914	<i>MATa</i> CEB1-1.8	Ref. (1)
ORD6713-8D	<i>MATa rad27::HIS3</i> CEB1-1.8	Ref. (1)
ORT4841	<i>MATa pif1::KanMX4</i> CEB1-1.8	Ref. (2)
ORT5087-5E	<i>MATa pif1-K264A</i> CEB1-1.8	Ref. (2)
AND1212-10D	<i>MATa</i> CEB1-WT-1.7	Ref. (2)
AND1227-5C	<i>MATa</i> CEB1-Gmut-1.7	Ref. (2)
AND1202-11A	<i>MATa pif1::KanMX4</i> CEB1-WT-1.7	Ref. (2)
AND1206-4C-D11P2	<i>MATa pif1::KanMX4</i> CEB1-Gmut-1.7	Ref. (2)
AND1218-1A	<i>MATa rad27::HIS3</i> CEB1-WT-1.7	Ref. (2)
AND1222-1D	<i>MATa arg4Δ2060</i>	This study
AND1239-5C	<i>MATa rad51::KanMX4</i> CEB1-WT-1.7	This study

## SUPPLEMENTARY MATERIALS AND METHODS

**Proteins.** The recombinant nuclear form of the *S. cerevisiae* Pif1 protein was expressed and purified as previously described (3).

**Oligonucleotide substrates.** G4-DNA substrates used in helicase assays were formed *in vitro* by incubating 150 $\mu$ M of Cy5-labeled A<sub>9</sub>39Ceb oligonucleotide (5'-Cy5-AAAAAAAAAAGGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTG) in presence of 1 M NaCl or 1 M KCl at 96°C for 5 minutes, followed by a 48 hours incubation at 65°C. The forked DNA substrate was formed by annealing of 10  $\mu$ M fD20 oligo (5'-Cy5-AAAAAAAAAACGCCATGGTGATCCGAGTGC) with an equimolar amount of Tmp1 oligonucleotide (5'-CACTGGCCGTCTTACGGTCGGCACTCGGATCACCATGGCG) in the presence of 10 mM Tris pH 8.0 and 10 mM MgCl<sub>2</sub>. The oligonucleotides mixture was incubated 3 minutes at 96°C in a dry bath, which was let to slowly cool to room temperature over an hour. The annealed forked substrate (fD20) was then resolved by electrophoresis on a native polyacrylamide gel and gel purified.

**Yeast Media preparation.** Yeast Peptone Dextrose (YPD) medium and Synthetic Complete (SC) medium were prepared according to standard protocols (4). Phen-DC<sub>3</sub> and Phen-DC<sub>6</sub> were solubilized in pure Dimethyl Sulfoxide (100% DMSO) at the maximal concentration of 20 mM for Phen-DC<sub>6</sub> and 50 mM for Phen-DC<sub>3</sub>. Above those concentrations drugs are no more soluble according to Beer-Lambert law. NMM (N-Methyl Mesoporphyrin IX, Frontier Biosciences) was diluted in water to a final concentration of 2 or 10 mM. Right before usage, NMM, Phen-DC<sub>3</sub>,

Phen-DC<sub>6</sub>, or DMSO was diluted in SC or YPD medium and media were filter-sterilized. Phen-DC<sub>3</sub> and Phen-DC<sub>6</sub> were used in order not to exceed a final concentration of 1% DMSO in the culture.

**Measurement of cell growth and viability.** Growth experiments were carried out with a WT strains bearing or not CEB1-WT-1.7 (AND1212-10D and AND1222-1D, respectively) in SC or YPD media. Freshly plated colonies were inoculated in liquid culture at an optical density at 600 nm (OD<sub>600</sub>) of 0.6 to 0.7 in liquid SC or YPD and grown overnight at 30°C. The culture was then diluted at an OD<sub>600</sub> of 0.3 in 5 ml of pre-warmed SC or YPD containing the desired Phen-DC<sub>3</sub>, Phen-DC<sub>6</sub> or DMSO concentration. Growth measurement with NMM was performed only in SC medium. Cells growth was monitored by spectroscopy, and expressed as a percentage of the number of generation made during exponential growth (between 2 and 8 hours) compared to the control treatment (water for NMM, and DMSO 1% or 0.4% for Phen-DC<sub>3</sub> and Phen-DC<sub>6</sub>, in YPD or SC respectively).

**Sequencing of CEB1 rearrangements.** WT (ORT2914) and *pif1*Δ (ORT4841) strains carrying the natural CEB1-1.8 allele of 42 polymorphic repeats plus a truncated one (see Supplementary Figure 6 for sequence) were used to determine the nature of rearrangements produced upon treatment with 500 μM Phen-DC<sub>3</sub>. Induction of minisatellite instability was carried out as described in **Material and Methods**. Colonies carrying a size contraction of CEB1 were identified by analyses of pooled colonies by Southern blot, followed by single colony isolation. PCR amplification of rearranged minisatellites and sequencing methods were performed as

previously described (1). The rearrangements were interpreted as the more parsimonious reshuffling of the parental motifs, based on the analysis of the polymorphic nucleotides.

**Correction of instability measured after the treatment of *pif1*Δ cells.** Due to the inherent instability of the CEB1 minisatellite in the Pif1-deficient strains before treatment, a subpopulation of cells may contain rearranged minisatellites. Thus, the effect of the treatment was measured as follow: *pif1*Δ cells were taken from a colony patch, diluted in water, plated as individual cells, and grown at 30°C. Rearrangement frequency in those colonies was determined by Southern blot, as described above. The pre-existing rearranged fraction ( $F_{p\text{-rea}}$ ) of CEB1-WT-1.7 in a starting patch of *pif1*Δ (AND1202-11A) and of CEB1-1.8 in *pif1*Δ (ORT4841) and *pif1*-*K264A* cells (ORT5087-5E) was estimated by Southern blot (see Table 2). We subtracted this fraction of cells to the total number of cells analyzed ( $N_{\text{tot}}$ ) after treatment for 8 generations, as well as the number of cells bearing a rearranged minisatellite ( $N_{\text{rea}}$ ) counted in the Southern blot. The corrected rearrangement frequency ( $F_{\text{cor}}$ ) is then expressed as:

$$F_{\text{cor}} = \frac{N_{\text{rea}} - F_{p\text{-rea}}N_{\text{tot}}}{(1 - F_{p\text{-rea}})N_{\text{tot}}}$$

For example, the calculation for *pif1*Δ cells bearing CEB1-WT-1.7 (AND1202-11A) treated by 500 μM of Phen-DC<sub>3</sub> for 8 generations and for which  $F_{\text{rea}}=9.8\%$  (Table 2) is:

$$\frac{106 - 0.098 \times 288}{(1 - 0.098) \times 288} = \frac{78}{260} = 0.3$$

**Statistical analysis.** The rearrangement frequencies measured on the Southern blots were compared using the two-tailed Fischer's Exact test, with an alpha cut-off of 0.05. The

distribution of the size of the rearrangements was compared using a two-tailed Mann-Whitney-Wilcoxon statistical test, with an alpha cut-off of 0.05.

## **SUPPLEMENTARY REFERENCES**

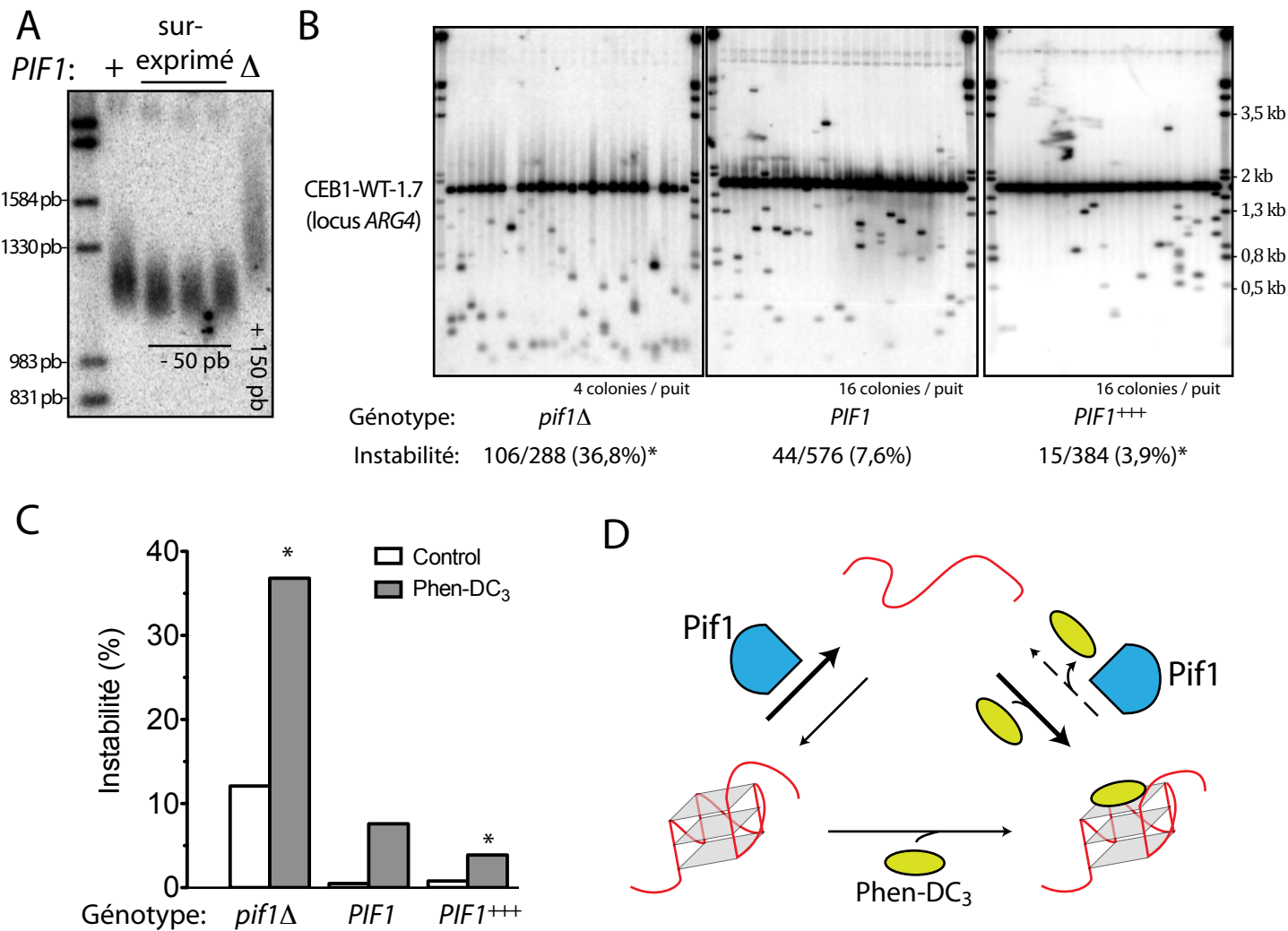
1. Lopes, J., Ribeyre, C. and Nicolas, A. (2006) Complex minisatellite rearrangements generated in the total or partial absence of Rad27/hFEN1 activity occur in a single generation and are Rad51 and Rad52 dependent. *Mol. Cell. Biol.*, **26**, 6675-6689.
2. Ribeyre, C., Lopes, J., Boule, J.B., Piazza, A., Guedin, A., Zakian, V.A., Mergny, J.L. and Nicolas, A. (2009) The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo. *PLoS Genet.*, **5**, e1000475.
3. Boulé, J.B. and Zakian, V.A. (2007) The yeast Pif1p DNA helicase preferentially unwinds RNA DNA substrates. *Nucleic Acids Res.*, **35**, 5809-5818.
4. Treco, D.A. and Lundblad, V. (2001) Preparation of yeast media. *Curr. Prot. Mol. Biol.*, **Chapter 13**, Unit 13-11.



### **II.B.3. Résultat supplémentaire : l'instabilité de CEB1 est soumise aux concentrations respectives de Pif1 et de Phen-DC<sub>3</sub> *in vivo***

Phen-DC<sub>3</sub> induit une instabilité accrue de CEB1 en absence de l'hélicase Pif1 : lors du traitement, 7,6% des cellules WT ont réarrangées CEB1 quand l'induction d'instabilité dans une souche *pif1Δ* est de ≈30% (**Article 2, Figure 3 et 4**). Ce résultat suggère que dans une souche WT, Pif1 contribue à limiter l'instabilité de CEB1 induite par le ligand. Pour le confirmer, j'ai mesurer l'effet du traitement par Phen-DC<sub>3</sub> sur une souche sur-exprimant la protéine Pif1. La souche sauvage est transformée avec un plasmide multi-copie (2μ, *leu2-cs*, 100-200 copies par cellules) contenant *PIF1* sous contrôle de son promoteur endogène (YEp13PIF1, (Foury and Lahaye 1987)). La sur-expression de Pif1 est vérifiée par le fait que les télomères sont plus courts que dans une souche sauvage (**Figure 32A**)(Zhou et al. 2000). Les cellules ont été traitées par Phen-DC<sub>3</sub> pendant 8 générations en milieu YPD (pas de pression de sélection sur le plasmide) et ont été étalées sur milieu sélectif (SC -Leucine) . Dans ces conditions, la fréquence de réarrangements est de 3.9% (15/384). C'est significativement moins que les 7,6% (44/576) observés dans les cellules sauvages (p-value = 0.017, test exact de Fisher, two-tailed)(**Figure 32B et 32C**). Ainsi, l'instabilité de CEB1 dépend des concentrations respectives de Phen-DC<sub>3</sub> (**Article 2, Table 1**) et de Pif1. Ceci suggère un modèle dans lequel Phen-DC<sub>3</sub> et Pif1 agissent en sens opposé sur le même équilibre, entre ADNsb et G-quadruplexe au niveau de CEB1 *in vivo* (**Figure 32D**).





**Figure 32: L'instabilité de CEB1 en présence de Phen-DC<sub>3</sub> est sensible à la concentration intracellulaire de Pif1.** (A) Analyse de la taille des télomères de souches *PIF1* (AND1212-10D), *pif1 $\Delta$  (AND1202-11A), et surexprimant *PIF1* (ORT6522) par Southern blot. Les télomères sont  $\approx 50$  pb plus courts dans une souche sur-exprimant *PIF1*, et  $\approx 150$  pb plus long dans une souche *pif1 $\Delta$ . Digestion = XhoI. Sonde = télomérique (pCT300). (B) Analyse de l'instabilité de CEB1-WT-1.7 en présence de Phen-DC<sub>3</sub> 500  $\mu$ M dans des souches WT, *pif1 $\Delta$ , et sur-exprimant Pif1. Digestion = ApaI/SceI. Sonde = CEB1. (C) Quantification de l'instabilité de CEB1-WT-1.7 dans des souches WT (*PIF1*), *pif1 $\Delta$ , et sur-exprimant *PIF1*, en présence ou en absence de Phen-DC<sub>3</sub> 500  $\mu$ M. (B et C) \* *p*-value vs. *PIF1*<sup>+</sup> < 0,05, test exact de Fisher, two-tailed. (D) Modèle d'interaction entre Pif1 et Phen-DC<sub>3</sub> au niveau des G-quadruplexes formés par CEB1. A gauche, l'équilibre est déplacé vers la forme simple brin par Pif1. Ainsi Pif1 élimine des cibles potentielles pour Phen-DC<sub>3</sub> (bas). Les expériences menées *in vitro* montre que Phen-DC<sub>3</sub> déplace l'équilibre vers la forme G-quadruplexe de deux façons: en promouvant la formation du G-quadruplexe (de Cian et al. 2007) et en inhibant le déroulement du G-quadruplexe par Pif1 (Article 2, Figure 2). Les équilibres sont sensibles à la concentration de Pif1, et/ou à la concentration de Phen-DC<sub>3</sub>.****

**Article 3 :**

**G-Quadruplex-Induced Instability During  
Leading-Strand Replication**



## **II.C. Article 3 : G-Quadruplex-Induced Instability During Leading-Strand Replication**

### **II.C.3. Contexte**

En l'absence de Pif1 ou en présence de ligands de G-quadruplexe, la persistance de G-quadruplexes au sein de CEB1 est responsable de son instabilité. Cette instabilité dépend d'une voie de recombinaison homologue fonctionnelle, suggérant la formation d'une lésion recombino-gène causée par le G-quadruplexe. Cependant, le moment de formation du G-quadruplexe, de même que le mécanisme conduisant à la formation de cette lésion, restaient inconnus.

Le G-quadruplexe intra-moléculaire se forme spontanément dans des conditions physiologiques en absence de brin complémentaire (**Chapitre I.A.1.a**). Aussi les processus qui génèrent activement de l'ADNsb sont propices à leur formation. C'est le cas de la réplication : la taille d'ADNsb générée sur la matrice de la machinerie de synthèse discontinue est d'au moins 165 nt, et de taille inconnue sur le brin matrice opposé. En plus d'être favorable à leur formation, ce processus pourrait également y être sensible : une fois formés au sein de leur matrice, les G-quadruplexes inhibent la progression des ADN polymérasés (**Chapitre I.A.5.a**). Cette inhibition pourrait constituer l'évènement initial à l'origine de la formation de la lésion recombino-gène (long ADNsb, ou fourche cassée)(**Chapitre I.B.3.b.iii**).

En utilisant les outils mis en place au préalable (mutagénèse de CEB1 et ligand de G-quadruplexe), nous avons étudié l'effet de la formation de G-quadruplexes sur la dynamique de réplication au niveau de CEB1 par gel bidimensionnel. Un avantage considérable de notre système expérimental, au contraire des séquences formant d'autres structures secondaires étudiées jusque-là, est que le brin sur lequel se forme la structure est connu. Ainsi, si l'instabilité de CEB1 dépend de son orientation vis-à-vis de la réplication, le brin de la matrice et de la machinerie de synthèse affecté peut être déterminé sans ambiguïté. Pour étudier cette éventuelle instabilité spécifique d'un brin, nous avons inséré CEB1 dans une région asymétrique du point de vue de la réplication, à proximité d'une origine de réplication précoce, l'ARS305. Cette étude a été réalisée en collaboration avec J. Lopes.



# G-quadruplex-induced instability during leading-strand replication

Judith Lopes<sup>1,5</sup>, Aurèle Piazza<sup>1,5</sup>, Rodrigo Bermejo<sup>2</sup>, Barry Kriegsman<sup>1</sup>, Arianna Colosio<sup>2</sup>, Marie-Paule Teulade-Fichou<sup>3</sup>, Marco Foiani<sup>2,4</sup> and Alain Nicolas<sup>1,\*</sup>

<sup>1</sup>Recombinaison et Instabilité Génétique, Institut Curie Centre de Recherche, CNRS UMR3244, Université Pierre et Marie Curie, Paris Cedex 05, France, <sup>2</sup>Istituto F.I.R.C. di Oncologia Molecolare, Milano, Italy, <sup>3</sup>Institut Curie Centre de Recherche, CNRS UMR176, Université Paris XI, Orsay, France and <sup>4</sup>Dipartimento di Genetica e di Biologia dei Microorganismi, Università degli Studi di Milano, Milano, Italy

**G-quadruplexes are four-stranded nucleic acid structures whose biological functions remain poorly understood. In the yeast *S. cerevisiae*, we report that G-quadruplexes form and, if not properly processed, pose a specific challenge to replication. We show that the G-quadruplex-prone CEB1 tandem array is tolerated when inserted near ARS305 replication origin in wild-type cells but is very frequently destabilized upon treatment with the potent Phn-DC<sub>3</sub> G-quadruplex ligand, or in the absence of the G-quadruplex-unwinding Pif1 helicase, only when the G-rich strand is the template of leading-strand replication. The orientation-dependent instability is associated with the formation of Rad51–Rad52-dependent X-shaped intermediates during replication detected by two-dimensional (2D) gels, and relies on the presence of intact G-quadruplex motifs in CEB1 and on the activity of ARS305. The asymmetrical behaviour of G-quadruplex prone sequences during replication has implications for their evolutionary dynamics within genomes, including the maintenance of G-rich telomeres.**

*The EMBO Journal* (2011) 30, 4033–4046. doi:10.1038/emboj.2011.316; Published online 26 August 2011

**Subject Categories:** genome stability & dynamics

**Keywords:** G-quadruplex; genomic instability; Pif1; recombination; replication

## Introduction

The maintenance of genome stability is crucial for normal cell growth, as revealed by the numerous chromosomal rearrangements that are a hallmark of cancer cells (Stephens *et al*, 2011). Mechanistically, genome stability is supported by the evolution of complex repair and surveillance cellular mechanisms monitoring faithful replication and lowering mutagenic DNA damages, induced upon exposure to genotoxic agents or

spontaneously generated inside cells (Aguilera and Gomez-Gonzalez, 2008). However, from an evolutionary perspective, genetic variation is a source of diversification within species, generating subtle or abrupt phenotypic variations suited for immediate adaptation and Darwinian selection to act on.

A long-standing and puzzling problem underlying the molecular mechanisms controlling the fluidity of genomes is the abundant presence in all organisms of nucleotide sequences ‘at risks’ of rearrangements (Durkin and Glover, 2007). Whether or not these sequences are detrimental or evolutionary advantageous and how they are propagated and tolerated within genomes are recurrent questions. At one end of the scale length are heterochromatin domains, such as centromeres and telomeres regions, which contain long highly repetitive sequences at risk of gross chromosomal rearrangements. On the other end of the scale length are microsatellite arrays composed of simple repeating DNA sequences of 1–10 nucleotides, which are prone to small length changes associated with human diseases. Another class of puzzling tandem repeats are minisatellite sequences which are arrays of 10–100 nucleotide motifs present in all genomes and also prone to contraction or expansion (Richard *et al*, 2008; present study). In the human genome, highly polymorphic G-C rich minisatellites (like *hCEB1* studied here) are overrepresented in subtelomeric regions (Vergnaud and Denoeud, 2000) but minisatellites are also found in interstitial regions, near or within coding regions (Richard and Dujon, 2006). Importantly, due to their extended motif length, minisatellite arrays have great potential to carry functional information: they can be the substrate of binding proteins (Law *et al*, 2011) or encode polypeptides (Richard and Dujon, 2006). Thus, the addition or deletion of minisatellite repeat units in individuals of a population is a potent mechanism to create genetic polymorphisms.

The molecular mechanisms, which drive the maintenance and variation of minisatellites, remain poorly understood. It is well established that the repetitiveness of microsatellite sequences facilitates polymerase slippage and allows out-of-register homologous recombination. Repetitiveness is thought to provide a high local concentration of homologous templates favouring maintenance. Sequence composition also matters; within repeat motifs, the embedded sequence-directed determinants allowing the formation of transient non-Watson–Crick DNA structure are additional challenges. The formation of secondary structures, when DNA is single stranded, can trigger fork pausing and instability. Stable DNA secondary structures may act as structural impediments to polymerase progression and their processing yield recombinogenic intermediates (Mirkin and Mirkin, 2007). This is a well-described challenge for the subset of microsatellites that can form hairpin secondary structures and hence facilitate intramolecular slippage of the replicative DNA polymerases (Samadashwily *et al*, 1997).

Other nucleic acid secondary structures may also challenge replication (Mirkin and Mirkin, 2007). Among these are

\*Corresponding author. Recombinaison et Instabilité Génétique, Institut Curie Centre de Recherche, CNRS UMR3244, Université Pierre et Marie Curie, 26 rue d’Ulm, 75248 Paris Cedex 05, France.

Tel.: +33 015 624 6520; Fax: +33 015 624 6644;

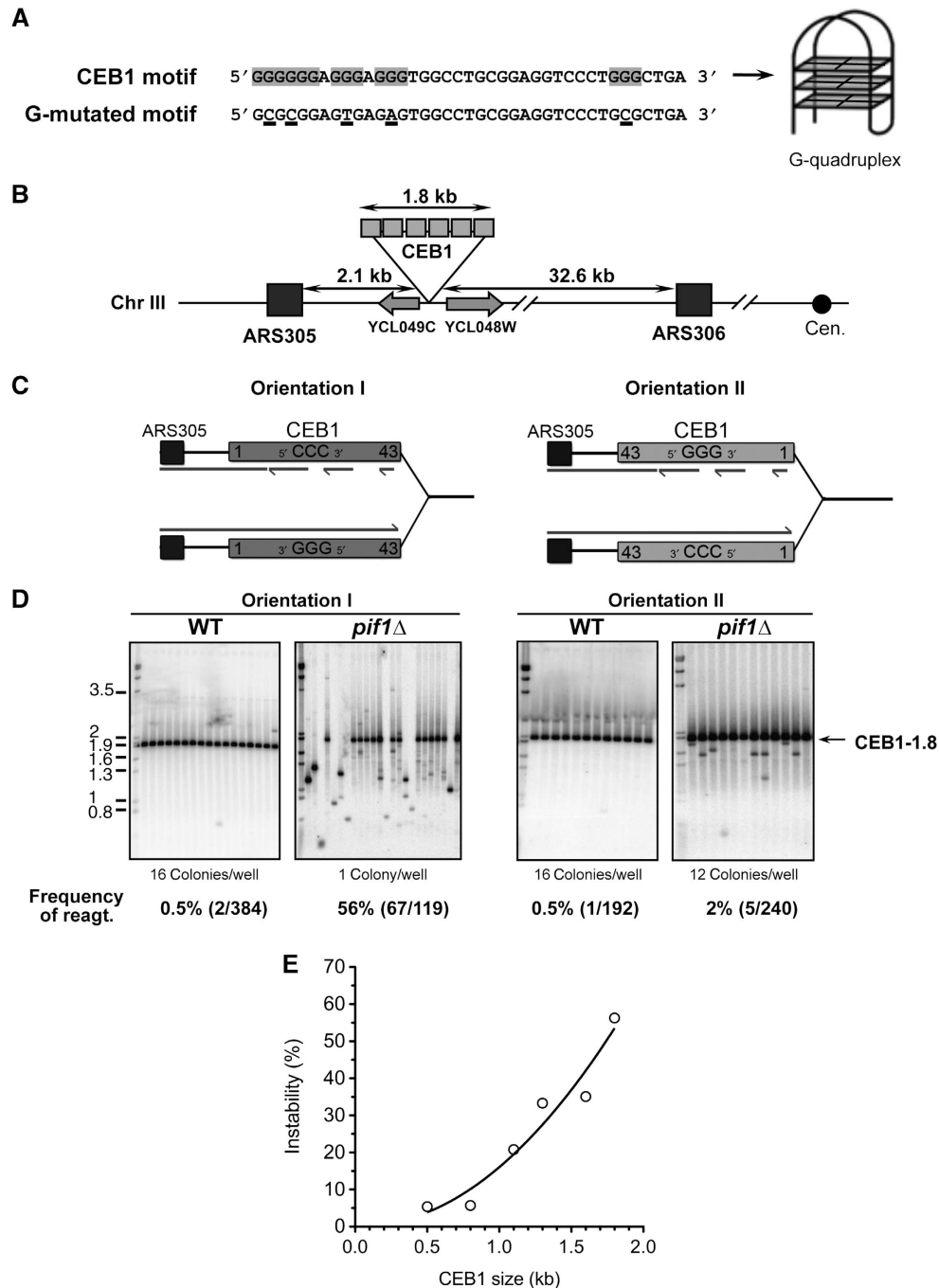
E-mail: alain.nicolas@curie.fr

<sup>5</sup>These authors contributed equally to this work

Received: 28 June 2011; accepted: 9 August 2011; published online: 26 August 2011

guanine-rich repeats, which have the capacity to adopt stable G-quadruplex structures. G-quadruplexes are composed of quartets of coplanar guanines (G-quartets) stabilized by non-

canonical Hoogsteen hydrogen bonding. Then, the stacking of three G-quartets stabilized by monovalent cations ( $\text{Na}^+$  or  $\text{K}^+$ ) is sufficient to form a G-quadruplex (illustrated in



**Figure 1** Orientation-dependent destabilization of CEB1 inserted near *ARS305* in *pif1*Δ cells. (A) Nucleotide sequence of CEB1 motifs shown in the 5'-3' orientation. CEB1 motif: more common motif in the natural CEB1-1.8 minisatellite (Lopes *et al*, 2006a). The G-tracts involved in the G-quadruplex formation (Ribeyre *et al*, 2009) are highlighted. G-mutated CEB1 motif: sequence of the synthetic CEB1 G-mutated motif assembled in the artificial CEB1-Gmut-1.7 tandem array (Ribeyre *et al*, 2009). The mutated nucleotides are underlined. (B) Map of the chromosome III region in which the CEB1 arrays were inserted (see Supplementary Methods). (C) Mode of replication of CEB1 by forks emanating from *ARS305* upon insertion in orientation I (left) or II (right). In orientation I, the G-quadruplex-forming strand is the template of leading-strand replication. In orientation II, the G-quadruplex-forming strand is the template of lagging-strand replication. (D) Behaviour of CEB1-1.8 in orientations I (left) and II (right) in WT (ORT6119-4 and ORT6135-36, respectively) and in *pif1*Δ (ORT6146-1 and ORT6136-8, respectively) cells visualized by Southern blot analysis. The number of colonies analysed per lane is indicated. The molecular ladder (in kb) allows for the size determination of the DNA fragments at a resolution of  $\pm 1$  motif (40 bp). Genomic DNAs were digested with *ApaI/XhoI* (orientation I for WT and *pif1*Δ cells), *ApaI/NcoI* (orientation II for WT cells), and *ApaI/SacII* (orientation II for *pif1*Δ cells). Membranes were hybridized with radiolabelled CEB1 probe. The total rearrangement frequencies (%), and in parenthesis the absolute number of rearrangements over the total number of colonies analysed, are reported below the blots. Statistical analyses (*P*-values) of rearrangement frequencies between strains are reported in Table I. (E) Size effect of CEB1 in orientation I on its rearrangement frequency in *pif1*Δ cells. A non-linear, second order fit has been applied ( $R^2 = 0.9133$ ). Strain names and numerical data are reported in Supplementary Tables S1 and S2, respectively.

Figure 1A). They can adopt a large variety of conformations depending on the size and sequence of intervening loops and, accordingly, G-quadruplexes will differ in stability and molecular properties (Burge *et al*, 2006), and their targeting thus require specific cognate ligands (De Cian *et al*, 2008). Experimental evidence arguing for the *in vivo* role of G-quadruplexes, and hence their formation have proven difficult and remains limited and heterogeneous (Maizels, 2006). G-quadruplexes regulate telomerase accessibility at the telomeres of the *Styloichia lemnae* protist, as directly observed throughout the cell cycle by immunostaining (Paeschke *et al*, 2008). Also, recent genetic and molecular studies argue that G-quadruplexes cap compromised telomeres (Sfeir *et al*, 2009; Smith *et al*, 2011). Beside their involvement in telomere homeostasis, G-quadruplexes also appear to promote genetic instability, as suggested by the presence of G-quadruplex motifs in the vicinity of chromosomal breakpoints in humans (Nambiar *et al*, 2010; De and Michor, 2011) and in *C. elegans* animal defective for *dog-1*, the homologue of the G-quadruplex-unwinding helicase FANCF (Kruisselbrink *et al*, 2008). This proposal has recently been subjected to experimental challenges. Site-directed mutational analysis and G-quadruplexes stabilization by specific ligands uncovered a direct role of G-quadruplexes in regulation of gene conversion of the pilin locus in *Neisseria gonorrhoeae* (Cahoon and Seifert, 2009) and in the instability of the hCEB1 minisatellite in *S. cerevisiae* (Ribeyre *et al*, 2009; Piazza *et al*, 2010; present study).

The next critical issue is how G-quadruplexes are processed? *In vitro* several DNA helicases, such as the mammalian BLM and WRN RecQ orthologues, PIF1, and FANCF unwind G-quadruplex structures (Sanders, 2010; Wu and Brosh, 2010). In *S. cerevisiae*, the Sgs1 (RecQ orthologue) and Pif1 helicases which belong to different helicase subfamilies and work in opposite directionality (3'-5' and 5'-3', respectively) unwind G-quadruplexes *in vitro* (Sun *et al*, 1999; Ribeyre *et al*, 2009). *In vivo*, our previous studies (and present data) clearly demonstrate that Pif1 is involved in the processing of the G-quadruplexes formed by CEB1 (Ribeyre *et al*, 2009; Piazza *et al*, 2010). Consistently, it has been recently reported that potential G-quadruplex-forming sequences are significantly enriched among the numerous Pif1 and Pif1-helicase dead (*pif1-K264A*) chromatin binding regions (1584 and 1153 regions, respectively), with an estimate that ~25% (138/558) of the potential G-quadruplex-forming sequences of the *S. cerevisiae* genome might be processed by Pif1 (Paeschke *et al*, 2011). Why, at last, Pif1 binds to so many chromosomal regions and certainly not only at G-quadruplex sites remains to be elucidated.

Another issue is to determine in which circumstance G-quadruplex forms *in vivo*. The formation of single-stranded DNA being most favourable circumstances, the folding of DNA G-quadruplex is most likely occurring during transcription and replication (Maizels, 2006). Strong evidences that G-quadruplex forms during replication remained to be brought. In this direction, Paeschke *et al* (2011) recently reported that in *pif1-m2* cells (proficient for the Pif1 mitochondrial function) but not in wild-type cells, a small fraction (11%) of the DNA Pol2 occupancy sites overlap regions containing potential G-quadruplex-forming sequences and that replication fork progression is slowed in hydroxyurea grown *pif1-m2* cells in a regional but not in G-quadruplex site-specific manner. Here, complementary to the Paeschke *et al* (2011)

study, we took advantage of the clearest demonstration that the minisatellite CEB1 can form G-quadruplex in *S. cerevisiae* cells (Ribeyre *et al*, 2009; Piazza *et al*, 2010) and specifically address whether G-quadruplexes affect replication and, if they do so, have different effects upon formation in the leading or in the lagging-strand template. The temporal program of replication, which initiates from a limited number of well-defined origins along the chromosomes being well established (Raghuraman *et al*, 2001), our key experimental strategy was to insert CEB1 in both orientations in an intergenic region near the early-firing replication origin *ARS305* and monitor its stability. This method was previously developed to determine the leading- and lagging-strand polymerases using a reporter strand-specific mutagenic assay (Nick McElhinny *et al*, 2008). Importantly, the selected chromosomal context further allows to demonstrate the role of replication by modulating the activities of the replication origins. Thus, the G-quadruplex-driven tandem-repeat instability assay in *S. cerevisiae* enabled us to demonstrate that G-quadruplexes frequently form on the leading strand in wild-type cells treated with the specific G-quadruplex-ligand Phen-DC<sub>3</sub> and in the absence of the Pif1 helicase.

## Results

### Experimental system and rationale

To determine the role of G-quadruplexes on yeast replication, we made use of the G-rich minisatellite CEB1 which is a sensitive reporter of G-quadruplexes formation and processing by monitoring its change of length during mitotic growth (Ribeyre *et al*, 2009; Piazza *et al*, 2010). Increase (expansion) or decrease (contraction) of the total number of CEB1 motifs (Figure 1A) can be visualized by Southern blot analysis of DNA extracted from individual or pooled cell colonies (Lopes *et al*, 2006a) and allow for the determination of a rearrangement frequency, also referred as 'instability'.

To examine a potential instability bias upon replication by the leading or the lagging machinery, we inserted the natural CEB1-1.8 allele containing 42 slightly polymorphic 39 nt motifs arranged as direct repeats (for full sequence, see Lopes *et al*, 2006a) in an asymmetrical position relatively to the dynamic of replication between *ARS305* and *ARS306* (Supplementary Figure S1; Pavlov *et al*, 2002; Fachinetti *et al*, 2010). Specifically, CEB1 was inserted in both orientations (I and II) in an intergenic region 2.1 kb to the right of *ARS305* and 32.6 kb away from *ARS306*, the closest downstream origin on the left arm of chromosome III (Figure 1B; Supplementary Methods). As to determine the DNA polymerases involved in leading- and lagging-strand replication, the rationale of this experimental design is that CEB1 will primarily be replicated by the fork emanating from *ARS305*, so that the leading- and lagging-strand templates can be unambiguously assigned (Nick McElhinny *et al*, 2008). In orientation I, the CEB1 G-quadruplex-forming strand will be the leading-strand template and becomes the lagging-strand template in orientation II (Figure 1C). The intergenic insertion of CEB1 was chosen to avoid transcriptional interference and all strains were grown in rich media to avoid eventual selective pressure. Furthermore, to avoid any risk of background heterogeneity, all strains were derived by transformation of the SY2209 haploid strain, previously used to examine replication dynamics (Fachinetti *et al*, 2010).



### Orientation- and size-dependent instability of CEB1 near ARS305 in Pif1-deficient cells

To examine the role of Pif1 in the stability of CEB1 inserted near ARS305, we measured the frequency of CEB1-1.8 size variants in WT and *pif1*Δ cells carrying CEB1 in orientation I or II. In WT cells, CEB1-1.8 is rather stable in both orientations, showing 0.5% (2/384) and 0.5% (1/192) of rearrangements, respectively (Figure 1D; Table I). In *pif1*Δ cells, the orientation greatly matters. In orientation II, the frequency of CEB1-1.8 rearrangements (2.0%) is not significantly different than in WT (0.5%) (*P*-value *pif1*Δ orientation II versus WT orientation II = 0.23). In sharp contrast, it is extremely unstable in orientation I. The instability reaches 56.3% (67/119) (*P*-value *pif1*Δ orientation I versus WT orientation I =  $1.24e^{-47}$  and *P*-value *pif1*Δ orientation I versus *pif1*Δ orientation II =  $1.69e^{-33}$ ) and yields a large diversity of size variants (Figure 1D; Table I), which ascertains the independent origin of these events. Importantly, when instability reaches such high levels, our measures are likely underestimated. To eliminate potential clonal events, similar size variants were counted only once (see Materials and methods), although sequencing showed that they could be different (Supplementary Figure S5). Thus, in *pif1*Δ cells, CEB1 in orientation I is at least 28-fold more unstable than in orientation II and than in either orientation in WT cells.

Formally, the lack of CEB1 size variants in orientation II could be explained by the induction of massive cell lethality. To test this possibility, we examined cell viability in the WT and *pif1*Δ cells carrying CEB1 in either orientation. We measured cell viability in a single generation assay by separating pairs of mother and daughter cells under the microscope and monitoring colony growth (Supplementary Figure S2A). In both wild-type and *pif1*Δ cells, the insertion of CEB1 near ARS305 in either orientation has no effect on cell viability: over 90% of budded mother cells (a landmark of entry into S-phase) give rise to viable colonies (Supplementary Figure S2B). Therefore, induction of cell death in orientation II is not the source of the orientation effect of CEB1.

Next, we investigated the relationship between array size and stability. As illustrated in Figure 1E, the overall frequencies vary by at least one order of magnitude, from 5.4% for the shortest allele (11 repeats) to 56.3% for the longest alleles (CEB1-1.8, 43 repeats) (Supplementary Table S2). Thus, the rearrangement frequency increases steadily with the number of CEB1 motifs (*P* =  $2.78e^{-3}$ , Spearman correlation test), apparently without threshold and in a more-than-linear re-

lationship despite the underestimate of the rearrangement frequencies for the longer alleles (Figure 1E).

In sum, the above data demonstrate that Pif1 is required to stabilize CEB1 near ARS305 in an orientation-dependent manner with rearrangement frequencies tightly correlated to the number of repeats. With respect to the mode of replication, the extreme instability corresponds to the orientation in which the CEB1 G-rich strand is the template of leading-strand synthesis.

### The helicase activity of Pif1 is required to stabilize CEB1

To determine if the helicase activity of Pif1 was required to stabilize CEB1-1.8, we examined strains carrying the *pif1*-K264A mutation that inactivates the Pif1 ATPase/helicase activity (Schulz and Zakian, 1994). We observed that CEB1 is very unstable in orientation I (27/76, 35.5%) but not in orientation II (6/192, 3.1%) (*P* =  $9.77e^{-12}$ ; Table I). Thus, the helicase activity of Pif1 has a role in the stabilization of CEB1 in the G-leading orientation. Compared with *pif1*Δ cells, the frequency of size variant in orientation I was ~2-fold lower in *pif1*-K264A cells (*P* =  $5.3e^{-3}$ ). This may reflect the fact that the *pif1*-K264A mutant retains wild-type level of DNA binding *in vitro* (Boulé *et al*, 2005), which could, to a certain extent, prevent G-quadruplex formation in CEB1 (Ribeyre *et al*, 2009). Thus, the absence of Pif1 or loss of its helicase activity results in a CEB1 instability induction that is 28- and 11-fold, respectively, higher in orientation I than in orientation II.

### Orientation-independent destabilization of CEB1 near ARS305 in rad27Δ/FEN1 cells

In a previous study, we found that the absence of Rad27/FEN1, a 5'-3' Flap endonuclease involved in the processing of Okazaki fragments formed during lagging-strand synthesis (Kunkel and Burgers, 2008) destabilizes all types of tandem repeats (microsatellite and minisatellite), including CEB1-1.8 inserted near ARG4 (Lopes *et al*, 2002, 2006a; Richard *et al*, 2008). Compared with WT cells, the inactivation of Rad27 increases the frequency of CEB1 rearrangements in both orientations (Table I), and the difference between the orientations I (18%) and II (11%) is not significant (*P* = 0.27). These results demonstrate that the biased behaviour of CEB1 near ARS305 is specific to the inactivation of Pif1 and, in the occurrence, not a general feature shared by another mutation perturbing replication.

**Table I** Rearrangement frequencies of CEB1 arrays placed near ARS305 in both orientations in WT and mutant strains

Minisatellite	Genotype	Orientation I	Orientation II	<i>P</i> -value orientation I versus II
CEB1-1.8	WT	2/384 (0.5%)	1/192 (0.5%)	NS
	<i>pif1</i> Δ	67/119 (56.3%)*	5/240 (2%)	$1.69e^{-33}$
	<i>pif1</i> -K264A	27/76 (35.5%)*†	6/192 (3.1%)	$9.77e^{-12}$
	<i>pif1</i> Δ <i>rad51</i> Δ	6/191 (3.1%)†	0/192	0.03
	<i>pif1</i> Δ <i>rad52</i> Δ	1/192 (0.5%)†	1/192 (0.5%)	NS
	<i>pif1</i> Δ <i>rad54</i> Δ	17/167 (10.2%)*†	10/191 (5.2%)*	NS
	<i>sgs1</i> Δ	2/192 (1%)†	2/192 (1%)	NS
	<i>rrm3</i> Δ	0/192†	0/192	NS
	<i>rad27</i> Δ	14/81 (17.3%)*†	9/83 (10.8%)*†	NS
CEB1-Gmut-1.7	<i>pif1</i> Δ	0/384	ND	NA

NA, not applicable; ND, stability not determined; NS, *P*-value orientation I versus orientation II not significant.

\**P*-value single or double mutants versus WT <0.05.

†*P*-value *pif1*-K264A, single or double mutants versus *pif1*Δ <0.05.

**Other helicases, as well as Rev1 polymerase, are not required to ensure CEB1 stability**

To examine whether this orientation-dependent instability of CEB1 near ARS305 is specific to the absence of Pif1, and not mutations in other helicases, we assayed the stability of CEB1-1.8 in both orientations in strains deleted for the *RRM3* or *SGS1* helicase genes. Sgs1 was reported to unwind G-quadruplexes *in vitro* (Sun *et al*, 1999), and Rrm3 is, like Pif1, a member of the SF1 helicase superfamily, reported to travel with the replication fork and to promote its progression at pause sites (Ivessa *et al*, 2003; Azvolinsky *et al*, 2006). In the *sgs1Δ* or *rrm3Δ* cells, CEB1 remained stable in either orientation (Table I). Hence, the destabilization of CEB1 in orientation I is specific to the inactivation of the Pif1 helicase. Additionally, it shows that neither Pif1, nor Sgs1, nor Rrm3 seems required to ensure CEB1 stability when the G-strand is template for the lagging-strand synthesis (see Discussion). A recent study suggested that in DT40 chicken cells, REV1, a polymerase of the Y-family, promotes the replication of G-quadruplex-forming sequences during leading-strand synthesis (Sarkies *et al*, 2010). To address the eventual role of the orthologous *REV1* gene in yeast, we constructed and assayed the instability of CEB1 near ARS305 in *rev1Δ* cells in the presence or absence of Phen-DC<sub>3</sub> treatment. *rev1Δ* cells behaved like wild type (Table II). Thus, further studies will be required to generalize the role of Rev1 on potentially forming G-quadruplex sequences.

**G-quadruplex-dependent instability of CEB1 near ARS305 in WT and pif1Δ cells**

Does the orientation-dependent instability of CEB1 near ARS305 depend on its capacity to form G-quadruplexes

*in vivo*? To address this issue and to demonstrate the *in vivo* formation of G-quadruplexes, we conducted two independent experimental approaches.

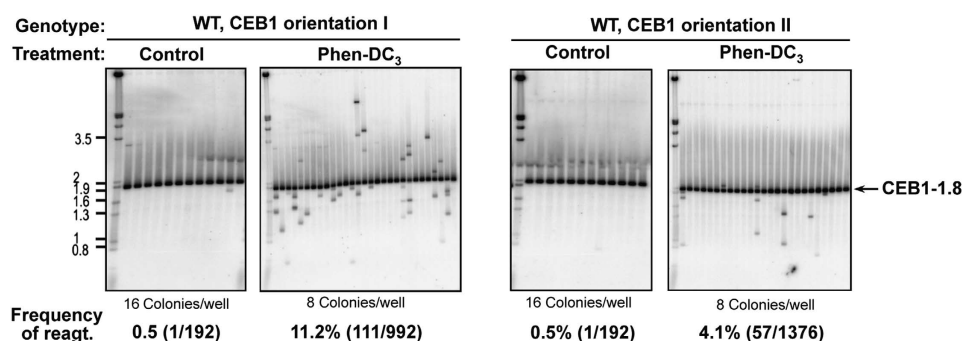
First, we asked whether or not CEB1 near ARS305 could be destabilized in WT cells upon treatment with a G-quadruplex ligand, in an orientation-dependent manner. We previously showed that the potent G-quadruplex-stabilizing compound Phen-DC<sub>3</sub> (Monchaud *et al*, 2008) efficiently inhibits G-quadruplex unwinding by Pif1 *in vitro*, and specifically triggers instability of G-quadruplex-prone CEB1 alleles inserted near the ARG4 locus (Piazza *et al*, 2010). Thus to establish that G-quadruplexes impede the replication machinery, we examined the stability of CEB1-1.8 inserted in either orientation near ARS305 in WT cells grown for eight generations with or without Phen-DC<sub>3</sub> at 10 μM. Remarkably, compared with the control samples (with 0.02% DMSO, as in the treated sample), which showed a single rearrangement out of 192 colonies (0.5%), Phen-DC<sub>3</sub> treatment triggered instability of CEB1 in both orientations but to different extents: in orientation I, we observed 111/992 rearrangements (11.2%, *P*-value versus control = 4.5e<sup>-8</sup>) and in orientation II, 57/1376 rearrangements (4.1%, *P*-value versus control = 7.1e<sup>-3</sup>) (Figure 2; Table II). The difference between both orientations is highly significant (*P*-value = 6.9e<sup>-11</sup>). These data strongly suggest that the instability of CEB1 near ARS305 depends on the *in vivo* formation of G-quadruplexes promoted or stabilized by the Phen-DC<sub>3</sub> compound.

As a second, complementary experimental approach, we examined the behaviour of a version of CEB1 mutated for its G-quadruplex-forming sequences (sequence of the G-mutated motif is shown in Figure 1A; Ribeyre *et al*, 2009). For this purpose, we inserted our synthetically built G-mutated CEB1

**Table II** Rearrangement frequencies of CEB1-1.8 in WT and mutant strains treated or not with Phen-DC<sub>3</sub>

CEB1-1.8 Genotype\treatment	Orientation I			Orientation II		
	Control	Phen-DC <sub>3</sub>	<i>P</i> -value (control versus Phen-DC <sub>3</sub> )	Control	Phen-DC <sub>3</sub>	<i>P</i> -value (control versus Phen-DC <sub>3</sub> )
WT	1/192 (0.5%)	111/992 (11.2%)	4.5e <sup>-8</sup>	1/192 (0.5%)	57/1376 (4.1%)**	7.1e <sup>-3</sup>
<i>rad51Δ</i>	2/176 (1.1%)	14/576 (2.9%)*	NS	2/384 (0.5%)	16/564 (2.8%)	0.013
<i>rad52Δ</i>	1/192 (0.5%)	1/154 (0.6%)*	NS	0/176	0/192*	NS
<i>rad54Δ</i>	0/192	1/384 (0.3%)*	NS	5/192 (2.6%)	8/368 (2.2%)	NS
<i>pif1Δ</i>	ND	ND	NA	3/191 (1.6%)	47/480 (9.8%)*	6.7e <sup>-5</sup>
<i>rev1Δ</i>	0/192	18/192 (9.4%)	5.0e <sup>-6</sup>	0/192	4/192 (2.1%)	NS

NA, not applicable; ND, not determined; NS, not significant. \**P*-value versus WT <0.05. \*\**P*-value versus orientation I <0.05.



**Figure 2** Orientation-dependent induction of CEB1 instability upon treatment of WT cells with the G-quadruplex-ligand Phen-DC<sub>3</sub>. Southern blot analyses of CEB1-1.8 behaviour in ORT6119-4 (orientation I) and ORT6135-36 (orientation II) cells treated or not with 10 μM of Phen-DC<sub>3</sub>. Genomic DNAs were digested with *ApaI/XhoI* (orientation I) and *ApaI/NcoI* (orientation II). Other legends as in Figure 1D.

allele of 42 motifs, named CEB1-Gmut-1.7 (for the full sequence of the minisatellite, see Ribeyre *et al*, 2009), near *ARS305*, at the same location as the natural CEB1-1.8 allele. Mutation of these motifs prevents G-quadruplex formation *in vitro* and stabilizes CEB1 inserted at the *ARG4* locus (Ribeyre *et al*, 2009; Piazza *et al*, 2010). The CEB1-Gmut-1.7 array, inserted near *ARS305* in orientation I in a *pif1* $\Delta$  mutant is stable (0/384), indicating that the high instability (56.3%) observed with CEB1-1.8 in this context relies on its G-quadruplex-forming potential (Table I).

Consistently and independently, the above data obtained from *pif1* $\Delta$  and Phen-DC<sub>3</sub>-treated WT cells strongly suggest that G-quadruplexes formed within CEB1 account for its differential instability behaviour in orientations I and II near the *ARS305* replication origin.

**The orientation-dependent behaviour of CEB1 results from the activity of ARS305 and can be reversed**

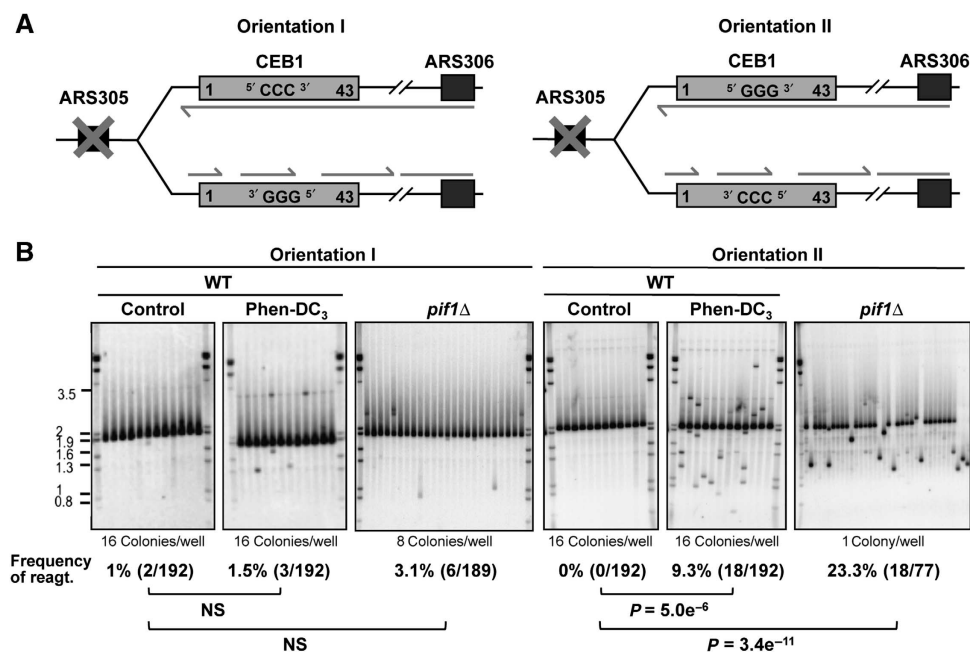
To rule out the possibility that the orientation-dependent behaviour of CEB1 could depend on other *cis*-acting elements than the replication directionality, we deleted the *ARS305* region (Figure 3A; Supplementary Methods). In this context, it is expected that CEB1 will be mostly replicated from the closest but nevertheless distant *ARS306* origin, and little from the distant and poorly active *ARS300-304* origins (see Supplementary Figure S1; Poloumienko *et al*, 2001; Pavlov *et al*, 2002). It implies that in *ARS305*-deleted cells, the G-strand of CEB1-1.8 in orientation I will be the template of lagging-strand replication, and *vice versa* orientation II will be the template of leading replication (Figure 3A). If only the mode of replication across CEB1 matters, the prediction is that the orientation bias will be switched. Indeed, the deletion of *ARS305* profoundly modifies the behaviour of CEB1:

it is still destabilized, but in the opposite orientation. In the absence of Pif1, we observed 6/189 (3.1%) rearrangements in the CEB1 orientation I and 18/77 (23.3%) rearrangements in orientation II (Figure 3B). The difference in orientation is statistically significant ( $P = 3.17e^{-6}$ ). To further establish this remarkable change of behaviour and the dependency on G-quadruplexes formation, we also examined the behaviour of CEB1-1.8 in WT cells treated or not by Phen-DC<sub>3</sub>. Consistently, the Phen-DC<sub>3</sub> treatment stimulates the instability of the CEB1-1.8 array placed in orientation II (18/192, 9.3%,  $P$ -value versus control =  $5e^{-6}$ ) but not significantly in orientation I (3/192, 1.5%) (Figure 3B). Again, the difference in orientation is statistically significant ( $P = 1.1e^{-3}$ ).

Together, the above data strongly support the interpretation that in the absence of Pif1, or upon stabilization of G-quadruplexes by Phen-DC<sub>3</sub> in wild-type cells, G-quadruplexes form within CEB1 when the G-rich strand is the template for leading replication and hence destabilize CEB1.

**G-quadruplex-dependent accumulation of X-shaped replication intermediates**

To determine whether the presence of CEB1 near *ARS305* and its orientation-dependent behaviour perturb the progression of replication, we arrested cells in G1 with  $\alpha$ -factor, released them to progress synchronously into S-phase, and then examined DNA content doubling by FACS analysis as well as replication intermediates by two-dimensional (2D)-gel electrophoresis (see Materials and methods). Origin firing generates ‘bubble arc’ containing forks proceeding bidirectionally and ‘large Y’ molecules resulting from forks migrating asymmetrically outside of the origin fragment. Other intermediates such as X-shaped and ‘cone’ signals detected in connections with various processes such as DNA recom-



**Figure 3** The deletion of *ARS305* reverses the orientation-dependent behaviour of CEB1. (A) Mode of CEB1 replication upon deletion of *ARS305*. CEB1 is replicated by forks emanating from *ARS306*. In orientation I, the G-quadruplex-forming strand is the template of lagging-strand replication whereas in orientation II, it is the template of leading-strand replication. This is the opposite of Figure 1C where CEB1 was replicated by the nearby *ARS305*. (B) Southern blot analysis of CEB1-1.8 behaviour placed in orientation I or II in WT cells (ORT6165-2 and ORT6166-52, respectively) treated or not with 10  $\mu$ M of Phen-DC<sub>3</sub>, and in *pif1* $\Delta$  cells (ORT6169-2 and ORT6170-8, respectively). Genomic DNA was digested with *ApaI/XhoI* (orientation I) and *ApaI/NcoI* (orientation II). Other legends as in Figure 1D.

bination (Ivessa *et al*, 2003; Bzymek *et al*, 2010) can also be visualized (Figure 4A).

In WT cells synchronously released from the G1 block, DNA doubling occurs within  $\sim 45'$  as monitored by FACS (Supplementary Figure S3), and replicative Y-shaped molecules generated by forks emanating asymmetrically from *ARS305* accumulate within 10–30' without detectable difference between the two strains carrying CEB1 in either orientation (Figure 4A). CEB1 is asymmetrically placed in the *HaeII*–*SacII* or *HaeII*–*ApaI* restriction digests products (see Figure 4A), implying that the typical increase of signal within the descending part of the Y arc (corresponding to long and mostly replicated Y molecules) may mask fork arrest within CEB1. To specifically address this possibility, we used a different restriction digestion (*ApaI*–*SpeI*) to probe CEB1 replication intermediates in a different position along the Y arc (Supplementary Figure S4). Signal intensity along the ascending Y arc was regular and remained constant after having fully replicated CEB1 in WT cells, indicating that the CEB1 array *per se* does not impair replication fork progression (Supplementary Figure S4).

In *pif1* $\Delta$  mutants, Y replication intermediates appear with appropriate kinetics across CEB1 placed in either orientation (Figure 4A). Strikingly, a signal of X-shaped intermediates (see arrow in Figure 4A, right panel) accumulates in the strain carrying CEB1 in orientation I. Along the time course, the proportion of this X-spike (expressed as the X/Y signal ratio) increases between 20' and 45', to reach 0.8 at 45' (Figure 4C). In sharp contrast, the X-spike is barely observed in the strain carrying CEB1 in orientation II, and never detected in WT strains, even upon longer gel exposure. Quantitatively, the X/Y signal ratio in *pif1* $\Delta$  cells is at least 10-fold higher in orientation I than in orientation II. These data demonstrate that *pif1* $\Delta$  cells accumulate aberrant intermediates migrating as X-shaped molecules during replication depending on the CEB1 orientation. Formation of X-structures correlates with very high frequency of CEB1 rearrangements. Interestingly, the formation of these X-shaped intermediates in *pif1* $\Delta$  cells requires intact G-triplets within CEB1, since such structures are not observed upon replication of the CEB1-Gmut-1.7 minisatellite in orientation I (Figure 4B and C; Supplementary Figure S3B).

Using a different restriction strategy (*ApaI*–*SpeI*), we probed CEB1 replication intermediates presenting forks at CEB1 in the ascending part of the Y arc. The signal intensity was stronger in the ascending than in the descending part of the Y arc in *pif1* $\Delta$  cells (Supplementary Figure S4). This was not the case with the CEB1-Gmut-1.7 allele. These observations suggest that replication slows down at CEB1 in the absence of Pif1 depends on the presence of intact G-triplets (Supplementary Figure S4).

Finally, we also examined the replication intermediates in WT cells treated with Phen-DC<sub>3</sub>. We observed that G1-arrested cells released into the Phen-DC<sub>3</sub>-containing media exhibited a significant delay in S-phase onset, compared with the control cells (equivalent DMSO concentration of 0.02%) that seemed unaffected, revealed both by the FACS profiles (Supplementary Figure S3C) and by the appearance of the replication intermediates on 2D gel (Figure 4D). This effect of Phen-DC<sub>3</sub> on the S-phase entry is independent of the presence and the orientation of CEB1. Remarkably, like in the absence of Pif1, X-shaped structures accumulate in

the Phen-DC<sub>3</sub>-treated WT cells during the replication of CEB1 in orientation I, but not in the opposite orientation (Figure 4E; Supplementary Figure S3D). Thus, the appearance of X-shaped intermediates dependent on the CEB1 orientation in circumstances that impair G-quadruplex processing, strongly suggests that G-quadruplexes form during leading-strand replication and, if not resolved, trigger CEB1 rearrangements.

### **The formation of X-shaped molecules and CEB1 rearrangements depend on the homologous recombination pathway**

On 2D gels, molecules migrating in the X-spike have been related to recombination intermediates such as double Holliday junctions (dHJ) (Bzymek *et al*, 2010) or convergent forks at sites of replication termination (Lopes *et al*, 2003). Among these two possibilities, the formation of recombination intermediates is the only one that require the activity of the Rad51 and Rad52 proteins, which play a central role in the repair of DSBs and the restart of stalled replication fork by homologous recombination (Pâques and Haber, 1999). Therefore, to distinguish between these possibilities, we deleted the *RAD51* or *RAD52* gene in *pif1* $\Delta$  cells carrying CEB1 in orientation I and then examined the replication intermediates. In contrast to *pif1* $\Delta$  cells, reexamined in parallel, X-shaped structures were not detected in the *pif1* $\Delta$  *rad51* $\Delta$  and *pif1* $\Delta$  *rad52* $\Delta$  double-mutant strains (Figure 4F and G). We, thus, concluded that X-shaped intermediates arising in *pif1* $\Delta$  cells during the replication of CEB1 in orientation I are recombination intermediates.

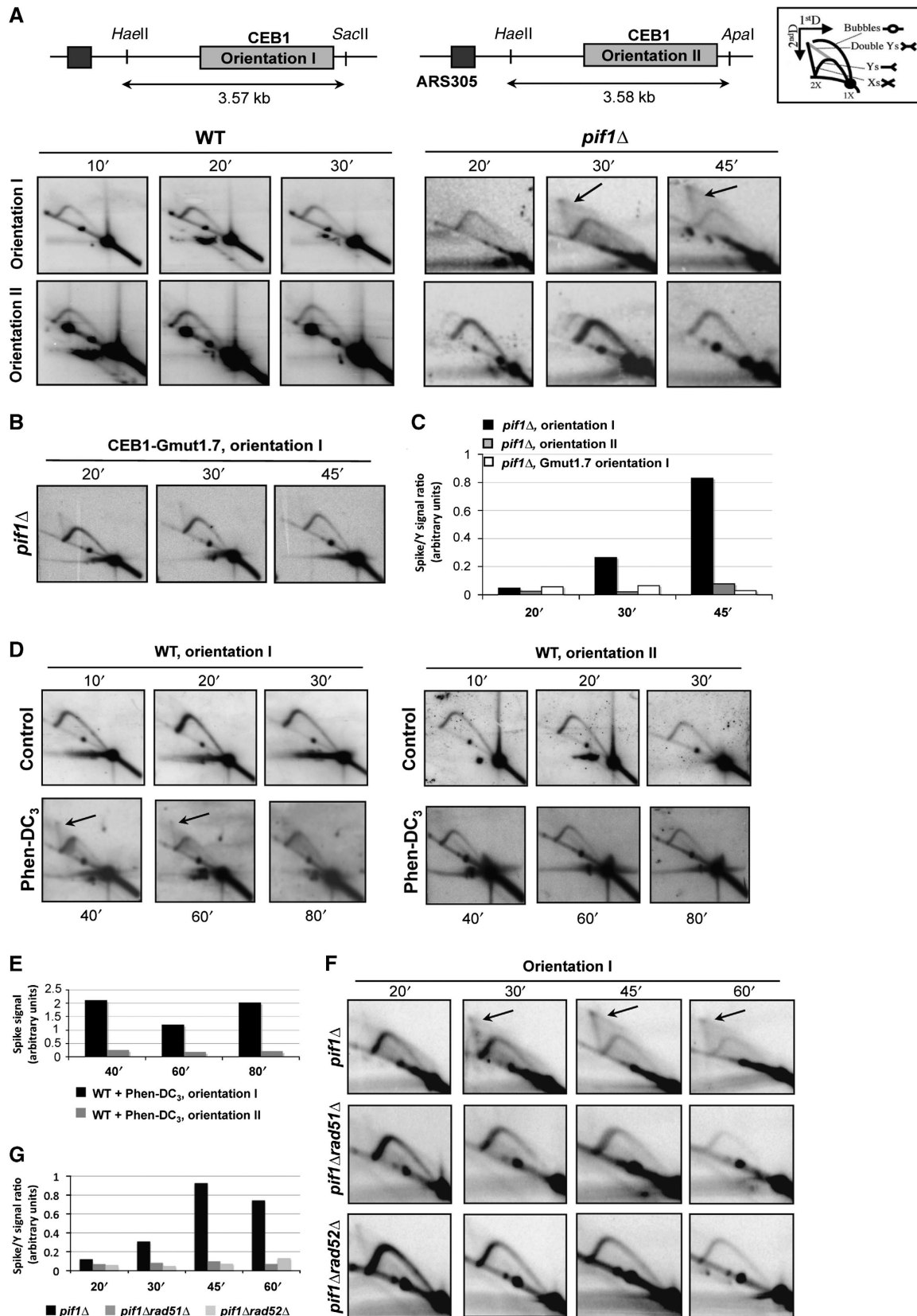
### **The formation of the CEB1 rearrangements also depends on the homologous recombination pathway**

The above results raise the prediction that the occurrence of CEB1-1.8 rearrangements near *ARS305* is also dependent on the activity of the homologous recombination pathway. To test this prediction, we examined the instability of CEB1 in strains deleted for the *RAD51*, *RAD52*, or *RAD54* genes (Supplementary Table S1). The Rad54 protein is also involved in homologous recombination at synapsis and post-synapsis stages (Heyer *et al*, 2006). The behaviour of CEB1 in these various strain backgrounds is reported in Table I. In the untreated *rad51* $\Delta$ , *rad52* $\Delta$ , and *rad54* $\Delta$  strains, we observed no or few rearrangements, and the treatment with Phen-DC<sub>3</sub> yielded no or few instability (Table II). Accordingly, the CEB1 instability in orientation I was abolished in the *pif1* $\Delta$  *rad52* $\Delta$  strain, and strongly decreased in the *pif1* $\Delta$  *rad51* $\Delta$  and *pif1* $\Delta$  *rad54* $\Delta$  double mutants compared with the *pif1* $\Delta$  strain ( $P$ -values =  $7.8e^{-34}$ ,  $7.4e^{-28}$ , and  $1.9e^{-17}$ , respectively; Table I). Thus, the remaining Rad51- and Rad54-independent rearrangements are likely produced by break-induced replication or single-strand annealing mechanisms (Pâques and Haber, 1999; Malkova *et al*, 2005). In orientation II, in all cases the frequencies of rearrangements were low (Table I).

Again, we examined cell viability. The absence of Rad51 has no detectable effect with CEB1 in both orientations (Supplementary Figure S2B). However, the double-mutant *pif1* $\Delta$  *rad51* $\Delta$  bearing CEB1 in orientation I exhibited a markedly reduced viability (67%) compared with *pif1* $\Delta$  and *rad51* $\Delta$  single mutants ( $P = 7.5e^{-6}$  and  $5.9e^{-3}$ , respectively), whereas cells bearing CEB1 in orientation II did not

(Supplementary Figure S2B). This result suggests that lesions produced in CEB1 in orientation I in the absence of Pif1 require the homologous recombination pathway to be repaired, otherwise leading to cell death or senescence.

We concluded that the production of CEB1 rearrangements in orientation I generated in the absence of Pif1 or upon treatment of WT cells by Phen-DC<sub>3</sub> and the correlated formation of the X-shaped structures during replication are fully



dependent on the activity of the homologous recombination pathway. Its role is likely to allow the processing of unresolved G-quadruplex structures forming during leading-strand replication, albeit with the consequence to generate CEB1 size variations by a process of template switch or synthesis-dependent strand annealing (SDSA) (Figure 5).

### Structure of the CEB1 rearrangements produced upon Phen-DC<sub>3</sub> treatment in either orientation

To further explore the behaviour of CEB1 near *ARS305* with respect to its orientation, we also asked whether the size variants obtained in the Phen-DC<sub>3</sub>-treated cells carrying the polymorphic CEB1-1.8 allele in either orientation were of different types. To do so, we first examined the size distribution of the variants. Compared with the parental CEB1-1.8 array containing 42 repeats, the size distribution of the variants, ranging from 3 to  $\approx 200$  repeats (median orientation I: 38 and median orientation II: 39), is not significantly different in both orientations and includes a majority of contraction events (78% in orientation I and 65% in orientation II, not significantly different). Having screened a large number of cells, we could also determine the sequence of 12 contraction events in each orientation, and compared them with the parental array, and with each other. No *de novo* mutation was found. As schematically represented in Supplementary Figure S5, all sequenced contractions retained their tandem array structure with full-length CEB1 motif, and are different from each other even if they contain the same number of motifs (for examples: D4 and DD1 (21 motifs) in orientation I or D4, D5, and C3 (13 motifs) in orientation II). As previously observed in *rad27Δ*, *pif1Δ*, and wild-type cells carrying CEB1 near *ARG4* and treated by Phen-DC<sub>3</sub> (Piazza *et al*, 2010), we recovered three categories of rearrangements (simple and double deletions, as well as complex reshuffling of the parental motifs) in similar proportions. Qualitatively, looking at the average size of the rearrangements and the motif shuffling, we did not detect notable difference

between the events produced in orientations I and II. This analysis gives no evidence that the rearrangements produced by the Rad51/Rad52 pathway are differently initiated in orientations I and II.

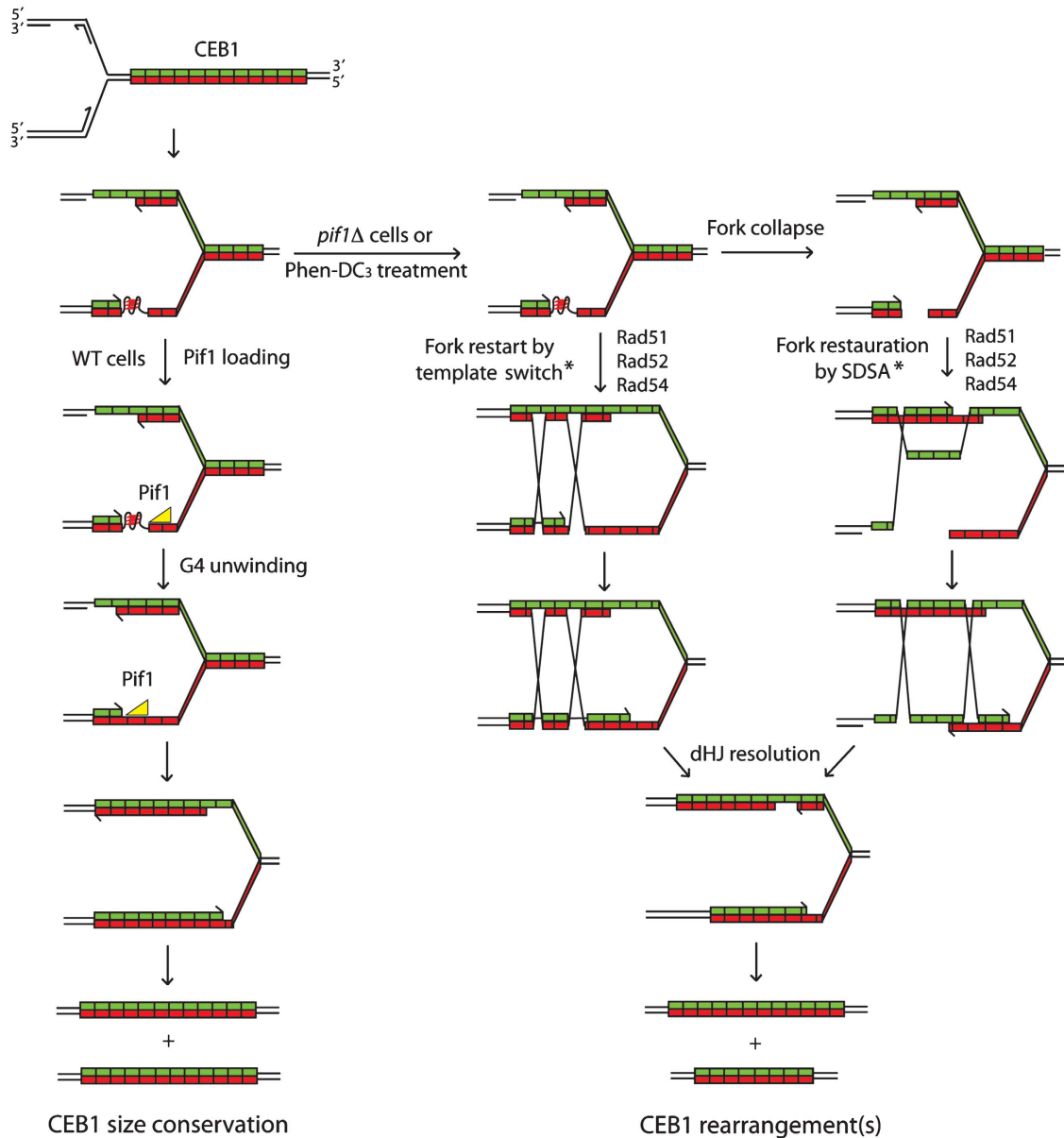
## Discussion

Using complementary genetic, chemical, and physical analyses, we provide unprecedented insights into the *in vivo* consequence of DNA G-quadruplex secondary structures formation on leading-strand replication and minisatellite instability. We show that (i) Yeast cells deficient for G-quadruplex processing, either by interaction with the Phen-DC<sub>3</sub> G-quadruplex ligand or by inactivation for the Pif1 helicase, frequently rearrange the reporter G-quadruplex-prone CEB1 minisatellite inserted in close proximity to the early-firing *ARS305* in an unexpected orientation-dependent manner. (ii) The instability is dependent on the G-quadruplex-forming sequence of CEB1 since site-directed mutations of the G runs abolish the instability. (iii) The instability is almost exclusively observed in orientation in which the G-quadruplex-forming sequence is the template of the leading-strand synthesis (see Figure 1). (iv) This orientation effect relies on the *cis*-effect of *ARS305*. Deletion of *ARS305* not only modifies the orientation bias but also reverses it because, in the absence of *ARS305*, the same region becomes replicated in the opposite direction from *ARS306*. (v) The orientation-dependent behaviour correlates with the formation of abnormal X-shaped intermediates during replication in CEB1. Formation of these intermediates, as well as the rearrangements, is dependent on the activity of the Rad51/Rad52 homologous recombination pathway.

### Model for G-quadruplex formation and its interference with leading-strand synthesis

In which biological circumstances G-quadruplexes fold *in vivo* remain unknown. G-quadruplexes have been implicated in

**Figure 4** Recombination-intermediate formation during CEB1 replication depends on the CEB1 orientation and on the formation of G-quadruplexes. **(A)** Time-course analysis by 2D-gel electrophoresis of replication intermediates formed in CEB1 arrays in orientations I and II during S-phase in synchronized WT and *pif1Δ* cells. The WT ORT6119-4 (orientation I), ORT6135-36 (orientation II), and the *pif1Δ* ORT6136-8 (orientation II) cells carry the CEB1-1.8 array. Due to the very high instability of CEB1-1.8 orientation I in *pif1Δ* cells which yields clonal rearrangements in the cell population (see Table I), the *pif1Δ* strain ORT6123-1-5 used here carries the CEB1-1.6 array in orientation I. Cells were arrested in G1 by  $\alpha$ -factor treatment, released into a synchronous S-phase, and aliquots examined by FACS to follow progression into the S-phase (see Supplementary Figure S4A). Genomic DNA was prepared from cells collected at the indicated times (min). Methods for 2D-gel analyses are described in Materials and methods. Restriction strategies and expected DNA fragment sizes for CEB1 in orientations I (*HaeII/SacII*) and II (*HaeII/ApaI*) are shown in the top panel. Membranes were hybridized with a radiolabelled CEB1-0.6 probe (Lopes *et al*, 2006a). A schematic representation of replication intermediates is presented in the right panel. Arrows on gel images outline abnormal X-shaped molecules (spike). The low intensity second Y arc in the *pif1Δ* strain with CEB1 orientation II corresponds to partially digested DNA. **(B)** Time-course analysis by 2D-gel electrophoresis of DNA intermediates formed in CEB1-Gmut-1.7 in orientation I during S-phase in a *pif1Δ* strain (ORT6157-1). Genomic DNA was digested with *HaeII/ApaI*. DNA fragments are revealed using a radiolabelled CEB1-Gmut probe (Ribeyre *et al*, 2009). Other legends as in **(A)**. FACS analyses reported in Supplementary Figure S4B. **(C)** Relative time-course quantification of the spike signal normalized to the Y arc signal observed in experiments in **(A)** and **(B)**. The spike signal is abundant in *pif1Δ* cells carrying the natural G-quadruplex-forming sequence as template of leading-strand replication (orientation I). **(D)** Time-course analysis by 2D-gel electrophoresis of DNA intermediates formed in CEB1 arrays in orientations I (left panel) and II (right panel) in synchronized WT cells, treated or not with 10  $\mu$ M Phen-DC<sub>3</sub>. In the Phen-DC<sub>3</sub>-treated samples, the shift from the 1C to 2C peak is slightly delayed compared with the control sample incubated in the same final concentration of DMSO (0.02%) (FACS, see Supplementary Figure S4C). Thus, control and treated samples were examined at different times in order to compare cells with similar progression into S-phase. Other legends as in **(A)**. **(E)** Time-course quantifications of spike signal observed in Phen-DC<sub>3</sub>-treated samples in the experiments of **(D)**. **(F)** Time-course analysis by 2D-gel electrophoresis of replication intermediates in synchronized *pif1Δ* (ORT6123-1-5) and *pif1Δ rad51Δ* (ORT6139-5) strains carrying CEB1-1.6 in orientation I, and *pif1Δ rad52Δ* (ORT6153-2) cells carrying CEB1-1.8 in orientation I. Other legends as in **(A)**. The spike signal observed in *pif1Δ* cells **(A)** is confirmed in this duplicate experiment and is no more observed upon inactivation of Rad51 and Rad52. FACS analyses reported in Supplementary Figure S4D. **(G)** Relative time-course quantification of the spike signal normalized to the Y arc signal observed in experiments of **(F)**. The abundant spike signal in *pif1Δ* cells carrying the natural G-quadruplex-forming sequence as template of leading-strand replication (orientation I) depends on the presence of the Rad51 and Rad52 proteins.



**Figure 5** A model for the replication of the CEB1 sequence upon formation of a G-quadruplex in the leading-strand template. CEB1 is schematically represented as a green (C-strand) and red (G-strand) array. The replication fork moves from left to right, implying that the G-rich strand of CEB1 is the template for leading-strand replication. Each motif has a certain probability to form a G-quadruplex, which will in turn stuck the replicative polymerase (Pol $\epsilon$ ). To be noted, in contrast to hairpins, G-quadruplexes fold only in the G-rich strand without an equivalent counterpart in the opposite strand (i-motif forms only at unphysiological acidic pH; Gehring *et al*, 1993). In a WT context (left part), Pif1 loads in the downstream single-stranded DNA and, while translocating in the 5'-3' orientation, unwind the G-quadruplex. Note that, due to the presence of Pol $\epsilon$  upstream of the G-quadruplex, the loading and thus the activity of potential 3'-5' helicases (like Sgs1) will require further processing (removal of the polymerase and 3'-5' resection). Thus, in this context, the G-quadruplex is a substrate only for 5'-3' helicases. Once the G-quadruplex is unwound, the replication can proceed normally. If the G-quadruplex persists, in Pif1-deficient cells or upon G-quadruplex stabilization by Phen-DC<sub>3</sub> (right part), fork is restarted and the G-quadruplex is bypassed or removed by a template-switch mechanism. Alternatively, an accidental single-strand break of the leading-strand template or the G-quadruplex cleavage can convert the stalled fork into a broken fork. Fork will then be restored by a synthesis-dependent strand annealing (SDSA) mechanism. Both of these SDSA and template-switch mechanisms involve a Rad52-dependent Rad51 nucleofilament formation followed by a Rad54-dependent dsDNA invasion step during which CEB1 motifs can be misaligned (denoted by an \*). The subsequent extension and reannealing of a misaligned invading strand will lead to the formation of the CEB1 rearrangements. Depending on the resolution of the double Holliday junction (dHJ), one or both alleles will be rearranged.

the regulation of DNA- and RNA-related processes such as telomere maintenance, genome rearrangements, transcription, splicing, and translation (Maizels, 2006; Krusselbrink *et al*, 2008; Paeschke *et al*, 2008; Cahoon and Seifert, 2009; Reinhold *et al*, 2010). Theoretically, the folding of a G-quadruplex DNA may initiate either from duplex or from single-

stranded DNA. A large number of natural proteins have been found to specifically interact with G-quadruplex structures (Wu and Brosh, 2010). They can play a sequence-specific role in destabilizing a target duplex DNA and/or acts to shift the equilibrium in favour of the G-quadruplex-folded state. Alternatively, the intramolecular folding of G-quadruplexes

on single-stranded DNA either during transcription or during replication might be easier without the competition from the complementary strand which in duplex would favour standard Watson–Crick base pairing. With this view, our mechanistic interpretation of the formation of G-quadruplexes and its specific interference with leading-strand replication, up to the ultimate formation of the rearrangement products, is illustrated in Figure 5. Globally, it addresses three major questions: how the G-quadruplex may form during S-phase, how defective G-quadruplex resolution leads to a recombinogenic lesion, and finally how the repair of these recombinogenic lesions generates the variety of rearrangements depicted in Supplementary Figure S5.

Along this path, the first issue is where single-stranded DNA might form during replication to promote the formation G-quadruplexes? During normal replication, the leading strand is synthesized continuously and the lagging strand copied in short ~200 bp Okazaki fragments which are then joined together after removal of their 5' ends by the Rad27/FEN1 Flap endonuclease (Kunkel and Burgers, 2008). During the unwinding of the DNA duplex by the MCM complex at the replication fork, single-stranded DNA may transiently forms between the leading Pole polymerase and the fork and/or between the lagging Okazaki fragments which are RNA primed and successively elongated by the Pol $\alpha$  and Pol $\delta$  polymerases (Kunkel and Burgers, 2008). In the case of CEB1, the sequence context may compromise the ability of the replicative polymerases to remain processive. We, thus, propose that the repetitiveness of the array would provide aggravating opportunities to further slow down replication and ultimately create the more drastic G-quadruplex obstacle. It is known that G-quadruplexes can hinder DNA polymerases progression *in vitro* (Weitzmann *et al*, 1997). Replication pause site can be visualized in 2D gel as a discrete spot signal along the Y arc that can be localized using different restriction digests. Here, in most 2D gels (Figure 4), we observed that the distribution of signal along the Y arc was homogenous, suggesting the lack of a specific site for replication pausing within CEB1. However, to be noticed, using the most asymmetric *ApaI/SpeI* restriction digestion that places CEB1 at the distal end of the Y arc, we noted a slight accumulation of replication intermediates within the hereof compacted CEB1 region (see Supplementary Figure S4). Hence, whether or not the wild-type CEB1 sequence itself perturbs the dynamic of Pole elongation remains to be further explored. Of concern, the sensitivity of 2D gel may not be sufficient to detect low frequency pausing events distributed along the CEB1 array. Differently, the coupling between the replicative helicase and polymerase as well as the coupling between the leading- and lagging-strand synthesis may also be impaired, leaving additional opportunities to create single-stranded gap regions or more complex branched molecules. Thus, as illustrated in Figure 5, we propose that intramolecular G-quadruplex folds on the leading strand between CEB1 and the fork duplex.

The next critical issue is how G-quadruplexes are processed? *In vitro* several DNA helicases, such as the mammalian BLM and WRN RecQ orthologues, PIF1, FANCI unwind G-quadruplex structures (Sanders, 2010; Wu and Brosh, 2010). In *S. cerevisiae*, the Sgs1 and Pif1 helicases unwind G-quadruplex *in vitro* (Sun *et al*, 1999; Ribeyre *et al*, 2009).

The present data demonstrate the *in vivo* role of Pif1 to ensure CEB1 stability during replication but we found no effect of inactivating the Sgs1 or Rrm3 helicases. It suggests that Sgs1 or Rrm3 may have no role in unwinding G-quadruplexes *in vivo*, act on different G-quadruplex substrates that remain to be uncovered, or show functional redundancies with each other or with the other helicases of *S. cerevisiae*. Our current mechanistic model is illustrated in Figure 5. Consistent with the recent report of Pif1 binding to chromosomal regions containing G-quadruplex potentially forming sequences (Paeschke *et al*, 2011), we propose that Pif1 unwinds the G-quadruplexes that form during leading-strand replication either by its direct recruitment to the G-quadruplex structure or indirectly, as being part of the 'replisome or fork-movement machinery' in order to resolve such impediments. The 5'–3' directionality of Pif1 would play an important role in this process (Figure 5). In contrast, in the absence of Pif1 or in Phen-DC<sub>3</sub>-treated wild-type cells, the G-quadruplexes formed by CEB1 will remain unprocessed and channelled into the recombinational-repair pathway, otherwise leading to cell death.

The present observations that factors of the homologous recombination pathways involved in the nucleofilament formation (Rad51 and Rad52) and strand invasion (Rad51 and Rad54) are required for the production of the accumulation of X-shaped intermediates, in correlation with the production of complex CEB1 rearrangements, demonstrate that unresolved G-quadruplexes formed in the leading-strand template lead to a recombinogenic substrate. The 2D-gel properties of the identified repair intermediates suggest that they represent X-shaped structures typical of unresolved joint molecules likely forming between the sister chromatids, but whether it arises from an initiating nick, a single-strand gap, or a double-strand break remains to be determined. As illustrated in Figure 5, out-of-register Rad51-dependent strand invasion can occur within the array and multiple cycles of invasion and destabilization events can yield the large variety of CEB1 rearrangements (Supplementary Figure S5). In Phen-DC<sub>3</sub>-treated WT cells as well as in *pif1 $\Delta$*  cells, the frequency of rearrangements increased with the size of the CEB1 array (Figure 1; Ribeyre *et al*, 2009) likely reflecting the increased probability of longer arrays to form one or several G-quadruplex structures. Alternatively, lesions in small alleles may be more frequently repaired in proper register or be resected into the non-repeated flanking regions, leading to the preferential restoration of the parental sequence.

### **Does G-quadruplexes form in the lagging-strand template?**

The stability of CEB1 inserted near *ARS305* in orientation II (G-strand being the lagging template) was surprising for two reasons. First, single-stranded DNA inherently forms on the lagging strand between elongating Okazaki fragments. Second, Pif1 involvement on the lagging-strand replication has been suggested by the genetic interactions of *PIF1* with *DNA2*, *POL32*, *CDC9*, *POL3*, and *RAD27* (Budd *et al*, 2006; Chang *et al*, 2009) and investigated *in vitro* on reconstituted Okazaki fragment processing machineries (Pike *et al*, 2009). Several interpretations can be invoked. It may indicate that G-quadruplexes do not fold on the template strand during lagging-strand synthesis. Alternatively, if G-quadruplexes



form, it may reveal that other helicases than Pif1 are specialized to resolve them on the lagging strand. So far, we have tested deletion of the *SGS1*, *RRM3*, *MPH1*, and *DNA2* (in combination with the Pif1 deletion) genes and found no effect of inactivating these candidate helicases (Ribeyre *et al*, 2009; Table I). Differently, it may also reveal the higher tolerance of the lagging-strand replication machinery to bypass the block by priming DNA synthesis downstream, although downstream priming on the leading strand is also possible (Heller and Marians, 2006).

To further address the possibility that CEB1-G-quadruplex may be differently processed on the leading and lagging strand, we also examined the behaviour of CEB1 in WT cells treated with Phen-DC<sub>3</sub> in either orientation. Again, the instability of CEB1 was significantly more pronounced in orientation I than in II, but quantitatively the bias was less pronounced in WT-treated cells than in the absence of Pif1 (~3–8-fold versus 28-fold, respectively; see Tables I and II and Figure 3) and importantly, it results from a slight increase of instability in orientation II compared with untreated WT cells. Mechanistically, it raises the possibility that CEB1-G-quadruplexes are processed by Pif1 in that strand, but in a dispensable manner due to the presence of other redundant G-quadruplex-processing activities.

### Implication on the localization of G-quadruplex-forming sequences within genome

The 'at risk' destabilization of CEB1 according to the direction of replication raises the question where the potential G-quadruplex-forming sequences are located within the yeast genome relative to the position of the origins of replication. To cross these features, we stringently mined the sequence of the yeast genome with Quadparser (Huppert and Balasubramanian, 2005) for G-quadruplexes with loops size ≤9 nt and then examined the position of the 50 best potential G-quadruplex-forming sequences relatively to the experimentally validated closest early- and middle-firing 5' and 3' ARS (Fachinetti *et al*, 2010). The methods and the source of information are described in Supplementary Methods. The resulting information is reported in Supplementary Table S3. If to be maintained, the G-quadruplex-forming sequences are preferentially located on the lagging-strand template, we expected to find them closer to the ARS in 5' than the ARS in 3'. Clearly, no bias towards a particular location on one side of the interorigin distance exists ( $P=0.19$  Mann-Whitney-Wilcoxon test). Thus, we conclude that in the *S. cerevisiae* genome the location of single G-quadruplex potentially forming sequences is not correlated with the replication directionality. Another evolutionary feature may reside in the arrangement of G-quadruplex-forming sequences in arrays like in CEB1. Indeed, whereas no negative selective pressure seems to apply for single G-quadruplex-forming sequences in the human genome (except in exonic coding sequence) (Huppert and Balasubramanian, 2005), G-quadruplex-prone tandem arrays are among the most unstable (Weitzmann *et al*, 1997), and the number of motifs in these arrays inversely correlates with the stability of the G-quadruplexes determined *in vitro* in nine vertebrate genomes (Bacolla *et al*, 2008). This correlation is rationalized in *S. cerevisiae* by the fact that except the subtelomeres and telomere repeats, interstitial tandem arrays of potential G-quadruplex-forming sequences are absent and that the majority of rearrangements

produced upon the Phen-DC<sub>3</sub> treatment in WT cells, and even more in *pif1Δ*-treated cells are extensive contractions (Piazza *et al*, 2010), all tending to eliminate potentially unstable G-quadruplex arrays, unless counteracted by functional selective pressure. Interestingly, along this line, despite the potential drawback of the G-quadruplex-forming repeats which we report here, it is remarkable that telomeric repeats in most eukaryotic phyla are TTAGGG or other potentially forming G-quadruplex sequences (Tran *et al*, 2010) and that both the subtelomeric repeats and the single-stranded overhang are oriented such that the G-rich strand is replicated by the lagging-strand machinery (Makovets *et al*, 2004; Gilson and Geli, 2007). Together with the utility of these G-quadruplex sequences to be able to form potentially useful stable structures in normal or pathological situations (Smith *et al*, 2011) such predominant evolution of telomeric repeats may result from the need to protect chromosome ends from damageable contractions and overall being replicated by the safer lagging-strand machinery is an adaptation to the challenge of leading-strand DNA replication.

## Materials and methods

### Media and strains

Media were prepared as previously described (Piazza *et al*, 2010). Strain was derived by transformation of the SY2209 *S. cerevisiae* strain of the W303 background (Fachinetti *et al*, 2010) and listed in Supplementary Table S1. Methods of strain construction are described in Supplementary Methods.

### Measurement of CEB1 instability

CEB1 instability during vegetative growth was measured by Southern blot analysis as previously described (Ribeyre *et al*, 2009; Piazza *et al*, 2010). Briefly, cells were suspended at a density of  $2 \times 10^5$  cells/ml into 5 ml of rich YPD medium, grown for eight generations at 30°C with agitation, and then plated as individual colonies on YPD media incubated at 30°C. Individual colonies or colony pools were analysed by Southern blot using the following genomic digestions: *ApaI/XhoI* for CEB1-1.8 (orientation I), *ApaI/NcoI* or *ApaI/SacII* for CEB1-1.8 (orientation II), and *ApaI/SacII* for CEB1-Gmut-1.7. The membranes were hybridized with radiolabelled CEB1-0.6 or CEB1 Gmut probes. Signals were detected with a Typhon PhosphorImager (Molecular Dynamics). To measure CEB1 instability upon Phen-DC<sub>3</sub> treatment, cells were grown for eight generations in liquid SC media containing 10 μM of Phen-DC<sub>3</sub> or, as a control, grown in the equivalent concentration of DMSO (0.02%) (Piazza *et al*, 2010). For pools of genomic DNA of 8, 12, or 16 colonies/well, a rearrangement is counted if the intensity of the rearranged minisatellite, quantified with ImageQuant software and normalized to its size, corresponds to 1/8 or 1/12 or 1/16 of the total amount of signals measured in the lane. Rearrangements migrating at the same size are considered as early clonal events, counted only once, and the remaining counts removed of the total number of colonies analysed. Sequencing of several of these same-sized alleles allowed us to confirm their clonal origin in most cases. However, in mutant strains with very high CEB1 instability (i.e., *pif1Δ* with CEB1 orientation I), the probability of obtaining two independent rearrangements of the same size is quite high. Consequently, the removal of all rearranged alleles of the same size may lead to an underestimation of the rearrangement frequency.

### Statistical analysis

The rearrangement frequencies measured on Southern blots have been compared using the two-tailed Fischer's exact test. The correlation between the instability and the size of CEB1 has been tested using the Spearman correlation test. Size distributions of the rearrangements have been compared using the two-tailed Mann-Whitney-Wilcoxon median comparison test. All the statistical tests have been performed using R 2.12.0. An  $\alpha$  cutoff of 0.05 has been applied for each test.

### Sequencing of CEB1 rearrangements

The sequencing of a sample of CEB1 contractions was performed as previously described (Lopes *et al*, 2006a). Colonies that carry a CEB1 contraction visualized by Southern blot were isolated, genomic DNA extracted and the CEB1 region was PCR amplified and sequenced using flanking primers. Compared with the sequence of the progenitor polymorphic CEB1-1.8 allele, the origin of the motifs was defined by the most parsimonious interpretation to describe the reshuffling of the parental motifs (see Supplementary Figure S5).

### Physical analyses of replication intermediates by 2D gels

After an O/N preculture at 25°C in YPD or SC medium, cells were diluted to 10<sup>7</sup> cells/ml in fresh medium and grown for 2 h at 25°C. Cell synchronization was performed on diluted cells (10<sup>7</sup> cells/ml) by adding  $\alpha$  factor (Euromedex) to a final concentration of 3  $\mu$ g/ml for about 130 min at 25°C. In the case of Phen-DC<sub>3</sub> or DMSO regimen, 10  $\mu$ M of Phen-DC<sub>3</sub> or the equivalent concentration of DMSO was added 15 min before the end of  $\alpha$  factor synchronization. The release was performed at 30°C in YPD, or in SC medium containing either Phen-DC<sub>3</sub> at 10  $\mu$ M, or 0.02% DMSO. At each time point, 1 ml of cells was taken for FACS analysis ( $\approx$ 10<sup>7</sup> cells) and 200 ml cultures were arrested by addition of 2 ml of 2% sodium azide for DNA extraction. Before DNA extraction, cells were washed with cold water and treated with Psoralen (Sigma). In all, 5 ml of cold water was added to the pellet and cells were transferred in culture dishes that contain six wells, in ice. In all, 300  $\mu$ l of Psoralen solution (0.2 mg/ml in ethanol 100%) was added and cells were incubated for 5 min in the dark. Cells were irradiated for 10 min with UV light of 365 nm. The light source was mounted at a distance of 6–7 cm above the culture dish, always in ice. The procedure is repeated three times with a total irradiation time of 40 min. Cells were then transferred in a 50-ml Falcon tube in ice, and washed with cold water. DNA extraction was performed as described in 'Method 3' DNA Extraction Procedures by Lopes *et al* (2003). 2D-gel electrophoresis was carried out as originally described by Brewer and Fangman (1987). DNA was transferred to GeneScreen Plus<sup>®</sup> Hybridization Transfer Membrane (Perkin-Elmer) in 10  $\times$  SCC by capillarity blotting. Quantification was done using ImageQuant 5.2 (Molecular Dynamics) as previously described (Lopes *et al*, 2003).

## References

- Aguilera A, Gomez-Gonzalez B (2008) Genome instability: a mechanistic view of its causes and consequences. *Nat Rev Genet* **9**: 204–217
- Azvolinsky A, Dunaway S, Torres JZ, Bessler JB, Zakian VA (2006) The *S. cerevisiae* Rrm3p DNA helicase moves with the replication fork and affects replication of all yeast chromosomes. *Genes Dev* **20**: 3104–3116
- Bacolla A, Larson JE, Collins JR, Li J, Milosavljevic A, Stenson PD, Cooper DN, Wells RD (2008) Abundance and length of simple repeats in vertebrate genomes are determined by their structural properties. *Genome Res* **18**: 1545–1553
- Boulé JB, Vega LR, Zakian VA (2005) The yeast Pif1p helicase removes telomerase from telomeric DNA. *Nature* **438**: 57–61
- Brewer BJ, Fangman WL (1987) The localization of replication origins on ARS plasmids in *S. cerevisiae*. *Cell* **51**: 463–471
- Budd ME, Reis CC, Smith S, Myung K, Campbell JL (2006) Evidence suggesting that Pif1 Helicase functions in DNA replication with the Dna2 Helicase/Nuclease and DNA Polymerase  $\delta$ . *Mol Cell Biol* **26**: 2490–2500
- Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res* **34**: 5402–5415
- Bzymek M, Thayer NH, Oh SD, Kleckner N, Hunter N (2010) Double Holliday junctions are intermediates of DNA break repair. *Nature* **464**: 937–941
- Cahoon LA, Seifert HS (2009) An alternative DNA structure is necessary for pilin antigenic variation in *Neisseria gonorrhoeae*. *Science* **325**: 764–767
- Chang M, Luke B, Kraft C, Li Z, Peter M, Lingner J, Rothstein R (2009) Telomerase is essential to alleviate pif1-induced replication stress at telomeres. *Genetics* **183**: 779–791
- De S, Michor F (2011) DNA secondary structures and epigenetic determinants of cancer genome evolution. *Nat Struct Mol Biol* **18**: 950–955
- De Cian A, Lacroix L, Douarre C, Temime-Smaali N, Trentesaux C, Riou JF, Mergny JL (2008) Targeting telomeres and telomerase. *Biochimie* **90**: 131–155
- Durkin SG, Glover TW (2007) Chromosome fragile sites. *Annu Rev Genet* **41**: 169–192
- Fachinetti D, Bermejo R, Cocito A, Minardi S, Katou Y, Kanoh Y, Shirahige K, Azvolinsky A, Zakian VA, Foiani M (2010) Replication termination at eukaryotic chromosomes is mediated by Top2 and occurs at genomic loci containing pausing elements. *Mol Cell* **39**: 595–605
- Gehring K, Leroy JL, Gueron M (1993) A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature* **363**: 561–565
- Gilson E, Geli V (2007) How telomeres are replicated. *Nat Rev Mol Cell Biol* **8**: 825–838
- Heller RC, Marians KJ (2006) Replication fork reactivation downstream of a blocked nascent leading strand. *Nature* **439**: 557–562
- Heyer WD, Li X, Rolfsmeier M, Zhang XP (2006) Rad54: the Swiss Army knife of homologous recombination? *Nucleic Acids Res* **34**: 4115–4125
- Huppert JL, Balasubramanian S (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res* **33**: 2908–2916
- Ivessa AS, Lenzmeier BA, Bessler JB, Goudsouzian LK, Schnakenberg SL, Zakian VA (2003) The *Saccharomyces cerevisiae* helicase Rrm3p facilitates replication past nonhistone protein-DNA complexes. *Mol Cell* **12**: 1525–1536
- Kruisselbrink E, Guryev V, Brouwer K, Pontier DB, Cuppen E, Tijsterman M (2008) Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCD1-defective *C. elegans*. *Curr Biol* **18**: 900–905
- Kunkel TA, Burgers PM (2008) Dividing the workload at a eukaryotic replication fork. *Trends Cell Biol* **18**: 521–527

### Cell viability assay

To examine cell viability, cells taken from a fresh patch on YPD plate were suspended to a final concentration of 5  $\times$  10<sup>4</sup> cells/ml in 5 ml of YPD, and grown O/N at 30°C. The exponentially growing culture was sonicated (Biolock Scientific sonicator, Vibra Cell 75041) for 5 s, and cells spread on YPD plate at room temperature. Using a micro-dissector (Singer Instruments), budded cells were isolated on a grid and regularly examined to separate mother and daughter cells. After 3 days of plate incubation at 30°C, the cells forming colonies were examined to determine the number of mother–daughter pairs giving rise to 2, 1, or 0 outgrowth.

### Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

## Acknowledgements

We thank Wolf-Dietrich Heyer and Jean-Baptiste Boulé for critical reading of the manuscript; Gael Millot for advises on statistical analysis; Valérie Borde, Thelma Capra, Rodney Rothstein, Arturo Londono-Vallejo, and members of our laboratories for helpful discussions. The AN team is supported by the Ligue Nationale contre le Cancer (LNCC; Equipe Labellisée EL2007.LNCC/AN and EL2010.LNCC/AN). AP is supported by a graduate student fellowship from the 'Ministère de l'Éducation Nationale, de la Recherche et de la Technologie'. BK was supported by the 2010 Massachusetts Institute of Technology—Institut Curie exchange program fellowship.

*Author contributions:* JL, AP, RB, MF, and AN designed the experiments. JL and AP performed the genetic experiments. JL, RB, and AC performed the 2D-gel analyses. JL, AP, RB, BK, AC, MF, and AN analysed the data. MPTF provided the Phen-DC3 molecule. JL, AP, RB, MF, and AN wrote the manuscript.

## Conflict of interest

The authors declare that they have no conflict of interest.

- Law MJ, Lower KM, Voon HP, Hughes JR, Garrick D, Viprakasit V, Mitson M, De Gobbi M, Marra M, Morris A, Abbott A, Wilder SP, Taylor S, Santos GM, Cross J, Ayyub H, Jones S, Ragoussis J, Rhodes D, Dunham I *et al* (2011) ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell* **143**: 367–378
- Lopes J, Debrauwere H, Buard J, Nicolas A (2002) Instability of the human minisatellite CEB1 in *rad27Δ* and *dna2-1* replication-deficient yeast cells. *EMBO J* **21**: 3201–3211
- Lopes J, Ribeyre C, Nicolas A (2006a) Complex minisatellite rearrangements generated in the total or partial absence of Rad27/hFEN1 activity occur in a single generation and are Rad51 and Rad52 dependent. *Mol Cell Biol* **26**: 6675–6689
- Lopes M, Cotta-Ramusino C, Liberi G, Foiani M (2003) Branch migrating sister chromatid junctions form at replication origins through Rad51/Rad52-independent mechanisms. *Mol Cell* **12**: 1499–1510
- Lopes M, Foiani M, Sogo JM (2006b) Multiple mechanisms control chromosome integrity after replication fork uncoupling and restart at irreparable UV lesions. *Mol Cell* **21**: 15–27
- Maizels N (2006) Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat Struct Mol Biol* **13**: 1055–1059
- Makovets S, Herskowitz I, Blackburn EH (2004) Anatomy and dynamics of DNA replication fork movement in yeast telomeric regions. *Mol Cell Biol* **24**: 4019–4031
- Malkova A, Naylor ML, Yamaguchi M, Ira G, Haber JE (2005) RAD51-dependent break-induced replication differs in kinetics and checkpoint responses from RAD51-mediated gene conversion. *Mol Cell Biol* **25**: 933–944
- Mirkin EV, Mirkin SM (2007) Replication fork stalling at natural impediments. *Microbiol Mol Biol Rev* **71**: 13–35
- Monchaud D, Allain C, Bertrand H, Smargiasso N, Rosu F, Gabelica V, De Cian A, Mergny JL, Teulade-Fichou MP (2008) Ligands playing musical chairs with G-quadruplex DNA: a rapid and simple displacement assay for identifying selective G-quadruplex binders. *Biochimie* **90**: 1207–1223
- Nambiar M, Goldsmith G, Moorthy BT, Lieber MR, Joshi MV, Choudhary B, Hosur RV, Raghavan SC (2010) Formation of a G-quadruplex at the BCL2 major breakpoint region of the t(14;18) translocation in follicular lymphoma. *Nucleic Acids Res* **39**: 936–948
- Nick McElhinny SA, Gordenin DA, Stith CM, Burgers PM, Kunkel TA (2008) Division of labour at the eukaryotic replication fork. *Mol Cell* **30**: 137–144
- Paeschke K, Capra JK, Zakian VA (2011) DNA replication through G-Quadruplex motifs is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase. *Cell* **145**: 678–691
- Paeschke K, Juranek S, Simonsson T, Hempel A, Rhodes D, Lipps HJ (2008) Telomerase recruitment by the telomere end binding protein-beta facilitates G-quadruplex DNA unfolding in ciliates. *Nat Struct Mol Biol* **15**: 598–604
- Pâques F, Haber JE (1999) Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* **63**: 349–404
- Pavlov YI, Newlon CS, Kunkel TA (2002) Yeast origins establish a strand bias for replicational mutagenesis. *Mol Cell* **10**: 207–213
- Piazza A, Boule JB, Lopes J, Mingo K, Largy E, Teulade-Fichou MP, Nicolas A (2010) Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **38**: 4337–4348
- Pike JE, Burgers PM, Campbell JL, Bambara RA (2009) Pif1 helicase lengthens some Okazaki fragment flaps necessitating Dna2 nuclease/helicase action in the two-nuclease processing pathway. *J Biol Chem* **284**: 25170–25180
- Poloumienko A, Dershowitz A, De J, Newlon CS (2001) Completion of replication map of *Saccharomyces cerevisiae* chromosome III. *Mol Biol Cell* **12**: 3317–3327
- Raguraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, Conway A, Lockhart DJ, Davis RW, Brewer BJ, Fangman WL (2001) Replication dynamics of the yeast genome. *Science* **294**: 115–121
- Reinhold WC, Mergny JL, Liu H, Ryan M, Pfister TD, Kinders R, Parchment R, Doroshov J, Weinstein JN, Pommier Y (2010) Exon array analyses across the NCI-60 reveal potential regulation of TOP1 by transcription pausing at guanosine quartets in the first intron. *Cancer Res* **70**: 2191–2203
- Ribeyre C, Lopes J, Boule JB, Piazza A, Guedin A, Zakian VA, Mergny JL, Nicolas A (2009) The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences *in vivo*. *PLoS Genet* **5**: e1000475
- Richard GF, Dujon B (2006) Molecular evolution of minisatellites in hemiascomycetous yeasts. *Mol Biol Evol* **23**: 189–202
- Richard GF, Kerrest A, Dujon B (2008) Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol Mol Biol Rev* **72**: 686–727
- Samadashwily GM, Raca G, Mirkin SM (1997) Trinucleotide repeats affect DNA replication *in vivo*. *Nat Genet* **17**: 298–304
- Sanders CM (2010) Human Pif1 helicase is a G-quadruplex DNA-binding protein with G-quadruplex DNA-unwinding activity. *Biochem J* **430**: 119–128
- Sarkies P, Reams C, Simpson LJ, Sale JE (2010) Epigenetic instability due to defective replication of structured DNA. *Mol Cell* **40**: 703–713
- Schulz VP, Zakian VA (1994) The *saccharomyces* PIF1 DNA helicase inhibits telomere elongation and *de novo* telomere formation. *Cell* **76**: 145–155
- Sfeir A, Kosiyatrakul ST, Hockemeyer D, MacRae SL, Karlseder J, Schildkraut CL, de Lange T (2009) Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell* **138**: 90–103
- Smith JS, Chen Q, Yatsunyk LA, Nicoludis JM, Garcia MS, Kranaster R, Balasubramanian S, Monchaud D, Teulade-Fichou MP, Abramowitz L, Schultz DC, Johnson FB (2011) Rudimentary G-quadruplex-based telomere capping in *Saccharomyces cerevisiae*. *Nat Struct Mol Biol* **18**: 478–485
- Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, Pleasance ED, Lau KW, Beare D, Stebbings LA, McLaren S, Lin ML, McBride DJ, Varela I, Nik-Zainal S, Leroy C, Jia M, Menzies A, Butler AP, Teague JW *et al* (2011) Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**: 27–40
- Sun H, Bennett RJ, Maizels N (1999) The *Saccharomyces cerevisiae* Sgs1 helicase efficiently unwinds G-G paired DNAs. *Nucleic Acids Res* **27**: 1978–1984
- Tran PL, Mergny JL, Alberti P (2010) Stability of telomeric G-quadruplexes. *Nucleic Acids Res* **39**: 3282–3294
- Vergnaud G, Denoeud F (2000) Minisatellites: mutability and genome architecture. *Genome Res* **10**: 899–907
- Weitzmann MN, Woodford KJ, Usdin K (1997) DNA secondary structures and the evolution of hypervariable tandem arrays. *J Biol Chem* **272**: 9517–9523
- Wu Y, Brosh Jr RM (2010) G-quadruplex nucleic acids and human disease. *FEBS J* **277**: 3470–3488

## Supplemental information

### Supplementary Figure Legends

#### Supplementary Table 1. Genotypes of Yeast Strains Used in This Study

All strains are MATa *ura3-1, his3-11,15, leu2-3,112, trp1-1, RAD5*. \* CEB1 placed in orientation I.  
\*\* CEB1 placed in orientation II (illustrated in Figure 1C).

#### Supplementary Table 2. Rearrangement Frequencies of CEB1 Alleles of Various Sizes in Orientation I in *pif1Δ* Cells

#### Supplementary Table 3. Localization of G-Quadruplex-Forming Sequences Relatively to Surrounding Early-Firing Replication Origins.

G-quadruplex-forming (G4F) sequences ( $G_3N_{1-9}G_3N_{1-9}G_3$ ) list have been determined using Quadparser (Huppert and Balasubramanian, 2005) and manually curated to eliminate telomeric G-quadruplex-forming sequences. Strand of the G-quadruplex-forming sequence is indicated (+ is Watson and – is Crick strand). Coordinates of the closest early-firing origins (Fachinetti et al., 2010) in 5' and in 3', as well as the distance to each G4F sequence, are indicated.

#### Supplementary Figure 1. ARS Location and Replication Dynamics in the 120 kb Region Surrounding the *ARS305* Region of Chromosome III

The CEB1 insertion site, near *ARS305* is shown. Replication profiles were obtained from the DNA Replication Origin Database website (<http://www.oridb.org>) (Nieduszynski et al., 2007). Black arrows indicate active origins. Grey arrows indicate inactive or poorly active ARS.

#### Supplementary Figure 2. Viability of WT, *pif1Δ*, *rad51Δ* and *pif1Δ rad51Δ* Cells with CEB1-1.8 in the Orientation I and II

(A) At t=0, budded cells from an exponentially growing culture are put on a grid by micromanipulation on a YPD plate. After 1-3h, mother and daughter cells are separated and put in the adjacent coordinate on the grid (for more details, see Supplementary Methods). Cells that failed to separate after 3 hours on plate were removed of the analysis. The number of colonies formed by each pairs of cells (0, 1, or 2 colonies) is counted after 3 days of growth at 30°C, and the results are reported in (B). (B) Percentage of separated pairs of cells that gave rise to 0, 1, or 2 colonies. The strains are the same as in Table 1. The control *pif1Δ* strain without CEB1 insert is ORT6133-2 (see Supplementary Table 1). Number of pairs analyzed, as well as the percentage of viable cells, are indicated. Statistical significance was calculated by comparing the distributions of pairs in the 0, 1, and 2 colonies categories with the Fischer exact test ( $\alpha$  cutoff of 0.05).

### **Supplementary Figure 3. Cell Cycle Progression Analysis by FACS After $\alpha$ -Factor Synchronisation**

(A) The WT ORT6119-4 (orientation I), ORT6135-36 (orientation II) cells and the *pif1* $\Delta$  ORT6123-1-5 (orientation I), ORT6136-8 (orientation II) cells were arrested in G1 by  $\alpha$ -factor treatment and release into fresh medium. Samples were collected at the indicated time points (minutes) for FACS analysis and 2G-gels electrophoresis (Figure 4A).

(B) Cell cycle progression of the synchronized *pif1* $\Delta$  strain (ORT6157-1) with the CEB1-Gmut-1.7 in orientation I. Other legends as in (A). The 2D-gel electrophoresis is presented in Figure 4B.

(C) Cell cycle progression of the synchronized WT ORT6119-4 (orientation I) and ORT6135-36 (orientation II) cells treated or not with 10  $\mu$ M Phen-DC<sub>3</sub>. In the Phen-DC<sub>3</sub>-treated samples, the shift from the 1C to 2C peak is slightly delayed compared to the control sample. Other legends as in (A). The 2D-gel electrophoresis is presented in Figure 4D.

(D) Cell cycle progression of the synchronized *pif1* $\Delta$  (ORT6123-1-5) and *pif1* $\Delta$  *rad51* $\Delta$  (ORT6139-5) and *pif1* $\Delta$  *rad52* $\Delta$  (ORT6153-2) cells carrying CEB1 in the orientation I. Other legends as in (A). The 2D-gel electrophoresis is presented in Figure 4F.

### **Supplementary Figure 4. Recombination-intermediates Formation during S-phase Seems Associated With Fork Pausing**

(A) Time course analysis by 2D-gel electrophoresis of DNA intermediates formed in CEB1 arrays in orientation I during S-phase in synchronized WT (ORT6119-4) and *pif1* $\Delta$  (ORT6123-1-5) cells. Restriction digests and expected DNA fragment sizes for CEB1 in the orientation I (HaeII/SacII and ApaI/SpeI) are shown in the top panel. Progressive doubling of the signal intensity is expected to occur in the descending part of the Y arc (large Y) with the HaeII/SacII digest in WT and *pif1* $\Delta$  cells (see also Figure 4A). With the ApaI/SpeI digest in *pif1* $\Delta$  cells, the accumulation of signal in the ascending part of the Y arc (small Y), associated with the decrease of the large Y arc signal, suggests fork pausing inside the CEB1-G-quadruplex repeats. Blue lines outline CEB1 position in the Y arc. Other legends as in Figure 4A.

(B) Time course analysis by 2D-gel electrophoresis of DNA intermediates formed in CEB1-Gmut-1.7 in orientation I during S-phase in a *pif1* $\Delta$  strain (ORT6157-1). Genomic DNA were digested with HaeII/ApaI or SpeI. Contrary to the *pif1* $\Delta$  cells with CEB1 G-quadruplex-prone sequence, the signal of replication intermediates remains in the descending part of the Y arc even with the SpeI digest, suggesting that the absence of G-quadruplex-forming sequence abolishes the replication fork pausing. Other legends as in Figure 4B.

(C) Relative time course quantification of the large and small Y arc signals observed in experiments A and B. Left panel: Ratio of large Y / small Y with CEB1 located in the descending part of the Y arc (HaeII/SacII and HaeII/ApaI digests). Right panel: Ratio of large Y / small Y with CEB1 located in the ascending part of the Y arc (ApaI/SpeI and SpeI digests).

### **Supplementary Figure 5. Sequence Analysis of the CEB1 Rearrangements Induced by Phen-DC<sub>3</sub> in WT cells**

Schematic representation of the polymorphic motif structure of the parental CEB1-1.8 allele and of 24 contractions obtained in WT strains carrying CEB1-1.8 array in the orientation I and II (strains ORT6119-4 and ORT6135-36, respectively) treated with 10  $\mu$ M of Phen-DC<sub>3</sub>. Among the 43 motifs in the natural CEB1-1.8 allele, 25 can be distinguished by one or several polymorphisms (Lopes et al., 2006) that have been used to determine the more likely structure of the rearrangements. The rearrangements fell into 3 categories: simple-deletion, double-deletion, and complex reshuffling of the parental sequence. The three classes of rearrangements are generated independently of CEB1 orientation. White motif in II-C1 contains polymorphisms of at more than 2 parental motifs.

### **Supplementary methods**

#### **Yeast strain construction**

All yeast strains used in this study were built by transformation of the parental SY2209 *Saccharomyces cerevisiae* strain of the W303 background (Fachinetti et al., 2010). Their genotypes are reported in Supplementary Table 1. The introduction of the CEB1 arrays near the *ARS305* locus was performed in several steps. First, the *URA3* marker was inserted on Chr. III in the intergenic region between the divergent *YCL048w*-*YCL049c* genes, between positions 41801-41840 (Cherry et al., 1997). This insertion was obtained by lithium-acetate transformation of the SY2209 strain with a PCR product constructed to provide sufficient flanking homology: i.e. containing an upstream fragment of 385 bp (from 41416 to 41801 on Chr. III), the *URA3* gene amplified from pRS416 and a downstream fragment of 412 bp (from 41840 to 42252 on Chr. III). It generated strain ORT5094-1. Then, the Hygromycin resistant marker (*HphMX4*) (Goldstein and McCusker, 1999) from pMJ696 was inserted in ORT5094-1 between positions 41841 and 41900 on Chr. III by long flanking homology transformation to yield ORT6143-13. Then, this strain carrying the *URA3* and *HphMX4* genes in close proximity was used to introduce the CEB1 minisatellite arrays (CEB1-1.8 orientation I, CEB1-1.8 orientation II or CEB1-Gmut-1.7) using PCR products containing the desired CEB1 allele and appropriate Chr. III flanking regions to substitute the markers and introduce CEB1. Sequentially, the transformants of strain ORT6143-13 were selected on 5FOA media (selecting *ura-* strains) and then screened for Hygromycin sensitivity. The CEB1-containing PCR products were constructed as follows. The flanking upstream fragment of 385 bp (from 41416 to 41801 on Chr. III) was amplified from yeast genomic DNA with a couple of primers, each containing a terminal *Apal* restriction site. The PCR product digested by *Apal* was inserted into pJL80 (pUC19 derived plasmid containing the CEB1-1.8 orientation I) to create plasmid pJL115. Then, the downstream fragment of 412 bp (from 41840 to

42252 on Chr. III) was amplified with two primers to add terminal XhoI and AflII sites, and inserted into pJL115 to create pJL116. The Bsu36I/AflII fragment from pJL116 carrying the CEB1-1.8 minisatellite (orientation I) flanked by the Chr. III regions was used for transformation of the ORT6143-13 strain to produce ORT6119-4 in which the CEB1 allele is cleanly inserted, without additional marker that may generate transcriptional interference with replication. Homologous integration was verified by Southern blot analysis of the transformants. Similarly, an upstream 385 bp XhoI/SacII and downstream 412 bp Apal/AflII chr. III fragments were successively integrated into pJL85 (pUC19 derivated plasmid with CEB1-1.8 orientation II) to create plasmid pJL122. The XhoI/AflII fragment from pJL122, was used to transform ORT6143-13 resulting in ORT6135-36. To insert the CEB1-Gmut-1.7 array near *ARS305*, it was amplified from plasmid pPA58-4 (pGEM<sup>®</sup>-T Easy vector (Promega) containing CEB1-Gmut-1.7 minisatellite) with primers adding terminal SacII and Apal restriction sites and cloned in pJL122 to replace the CEB1-1.8 minisatellite by the CEB1-Gmut-1.7. The resulting plasmid, pJL125, was digested by XhoI/AflII and used to transform ORT6143-13 resulting in ORT6582-4. All transformants were verified by Southern blot analysis of genomic DNA allowing for the confirmation of the markers loss and their replacement by the expected CEB1 array. Inactivation of the *PIF1*, *RAD54*, *REV1*, *RRM3*, and *SGS1* genes with the selectable *KanMX4* marker (G418 resistance) were obtained by transformation of the ORT6119-4 and ORT6135-36 strains with PCR products amplified from the corresponding strains of the EUROSCARF deletants collection (Winzeler et al., 1999). Similarly, the *pif1::HIS3*, *rad27::HIS3*, *rad51::LEU2* and *rad52::LEU2* disruptions were obtained by transformation with PCR products from strains ORD9922-4B, ORD6713-8D, ORD7574-9B and ORD7565-2C, respectively (Ribeyre et al., 2009). Double mutant strains were similarly constructed by transformation of single mutant strains. ORT6146-1 and ORT6136-8 were both transformed by the PCR products *rad51::LEU2* and *rad52::LEU2*. The *pif1-K264A* point mutation was introduced at the *PIF1* locus by pop-in and pop-out transformation (Scherer and Davis, 1979) of ORT6119-4 and ORT6135-36, as previously described (Ribeyre et al., 2009). To build the deletion of *ARS305*, plasmid pMJ593 (Dershowitz et al., 2007) containing the *ARS305* deletion (1100 bp deletion of BamHI-ClaI fragment, from 38606 to 39706 on Chr. III) plus the flanking regions (3' end of *YCL51W* and the 5' end of *YCL48W*), was linearized by HpaI and integrated by pop-in pop-out transformation of ORT6119-4 and ORT6135-36. For all transformants, integration at the homologous locus was verified by Southern blot analyses of genomic DNA using appropriate restriction digest and probes flanking the targeted locus.

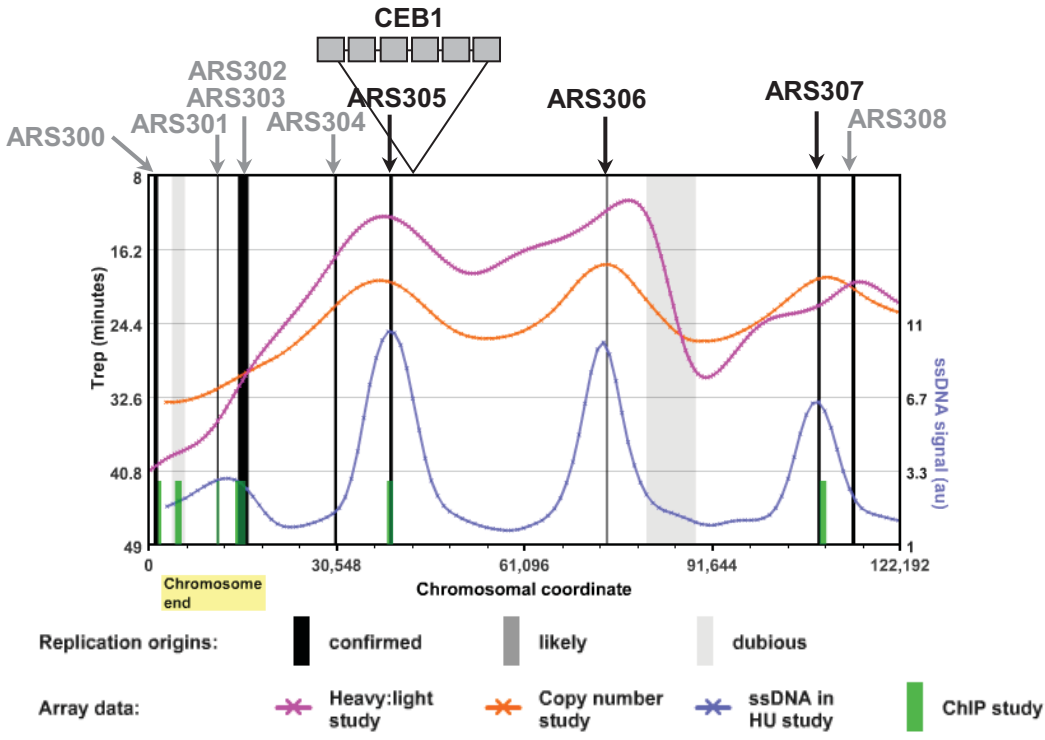
### Supplementary references

Cherry, J.M., Ball, C., Weng, S., Juvik, G., Schmidt, R., Adler, C., Dunn, B., Dwight, S., Riles, L., Mortimer, R.K. and Botstein, D. (1997) Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature*, 387, 67-73.

- Dershowitz, A., Snyder, M., Sbia, M., Skurnick, J.H., Ong, L.Y. and Newlon, C.S. (2007) Linear derivatives of *Saccharomyces cerevisiae* chromosome III can be maintained in the absence of autonomously replicating sequence elements. *Mol Cell Biol*, 27, 4652-4663.
- Goldstein, A.L. and McCusker, J.H. (1999) Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. *Yeast*, 15, 1541-1553.
- Nieduszynski, C.A., Hiraga, S., Ak, P., Benham, C.J. and Donaldson, A.D. (2007) OriDB: a DNA replication origin database. *Nucl. Acids Res.*, 35, D40-46.
- Scherer, S. and Davis, R.W. (1979) Replacement of chromosome segments with altered DNA sequences constructed in vitro. *Proc Natl Acad Sci U S A*, 76, 4951-4955.
- Winzler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H., Chu, A.M., Connelly, C., Davis, K., Dietrich, F., Dow, S.W., El Bakkoury, M., Foury, F., Friend, S.H., Gentalen, E., Giaever, G., Hegemann, J.H., Jones, T., Laub, M., Liao, H., Liebundguth, N., Lockhart, D.J., Lucau-Danila, A., Lussier, M., M'Rabet, N., Menard, P., Mittmann, M., Pai, C., Rebischung, C., Revuelta, J.L., Riles, L., Roberts, C.J., Ross-MacDonald, P., Scherens, B., Snyder, M., Sookhai-Mahadeo, S., Storms, R.K., Veronneau, S., Voet, M., Volckaert, G., Ward, T.R., Wysocki, R., Yen, G.S., Yu, K., Zimmermann, K., Philippsen, P., Johnston, M. and Davis, R.W. (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science*, 285, 901-906.

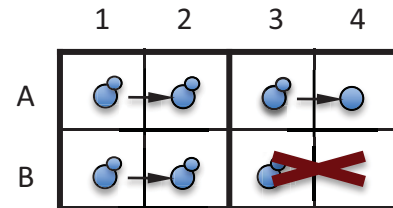
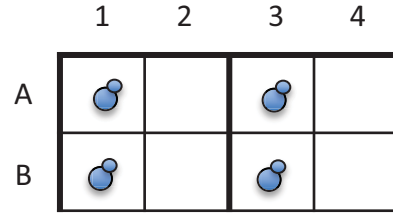
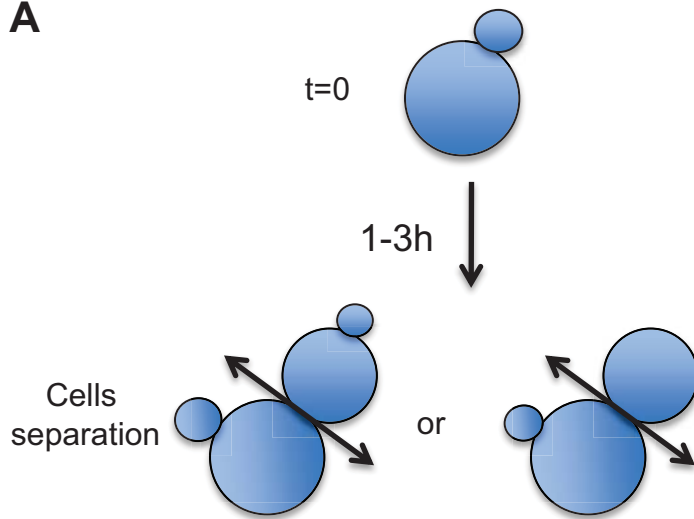


# Supplementary Figure 1



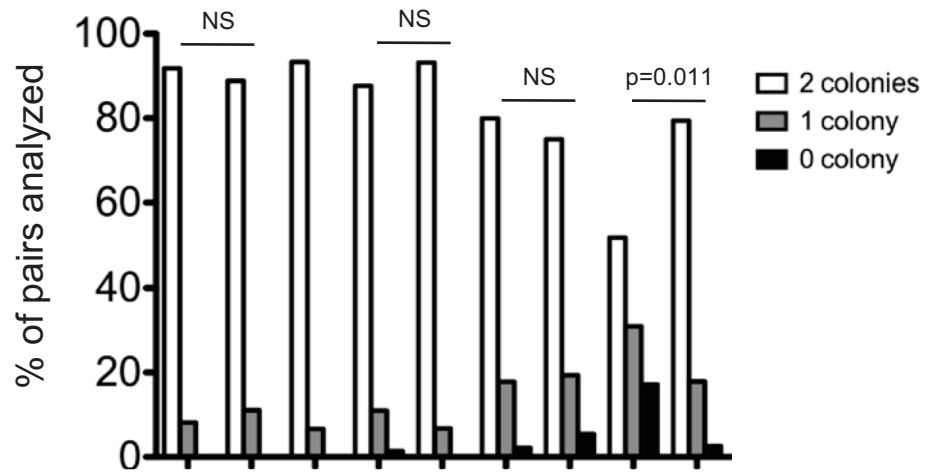
## Supplementary Figure 2

**A**



Number of colonies formed per separated pairs is counted after 3 days of growth at 30°C.

**B**



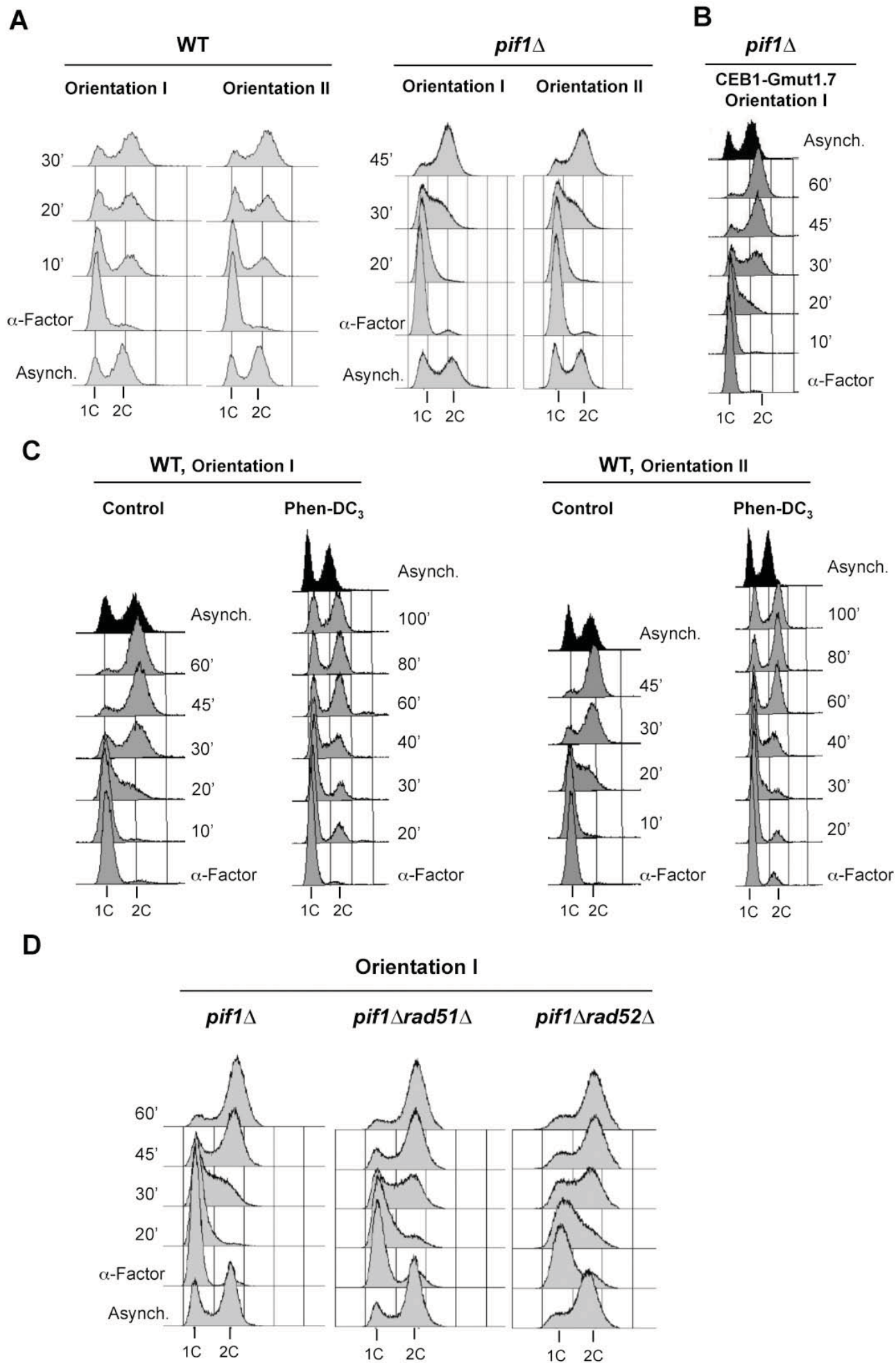
Pairs analyzed: 49 27 30 73 44 45 36 58 39

Viable cells (%): 96 94 97 93 97 89 85 67 88

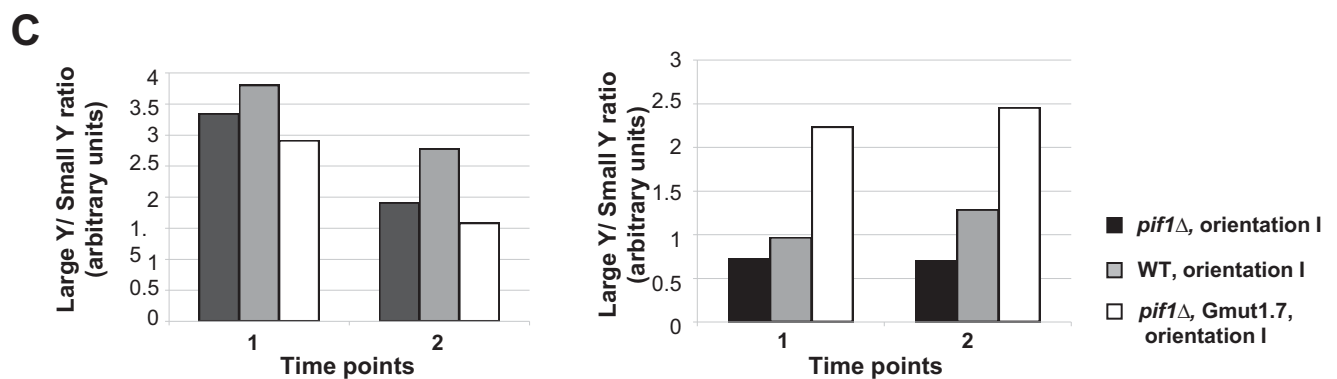
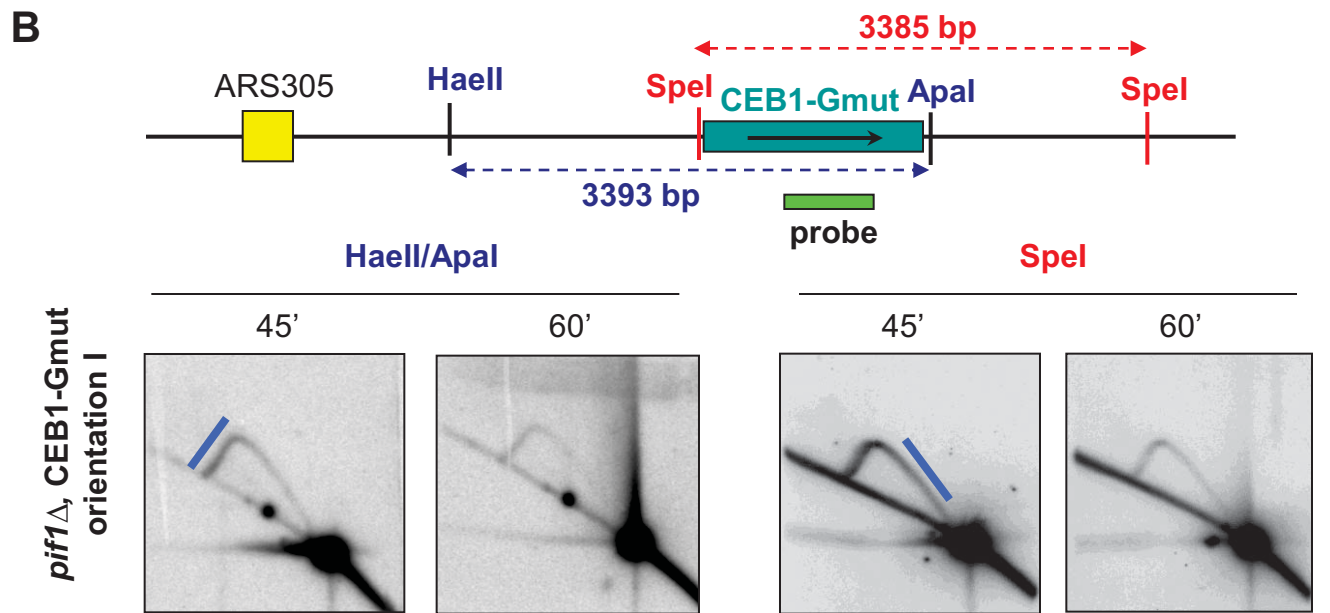
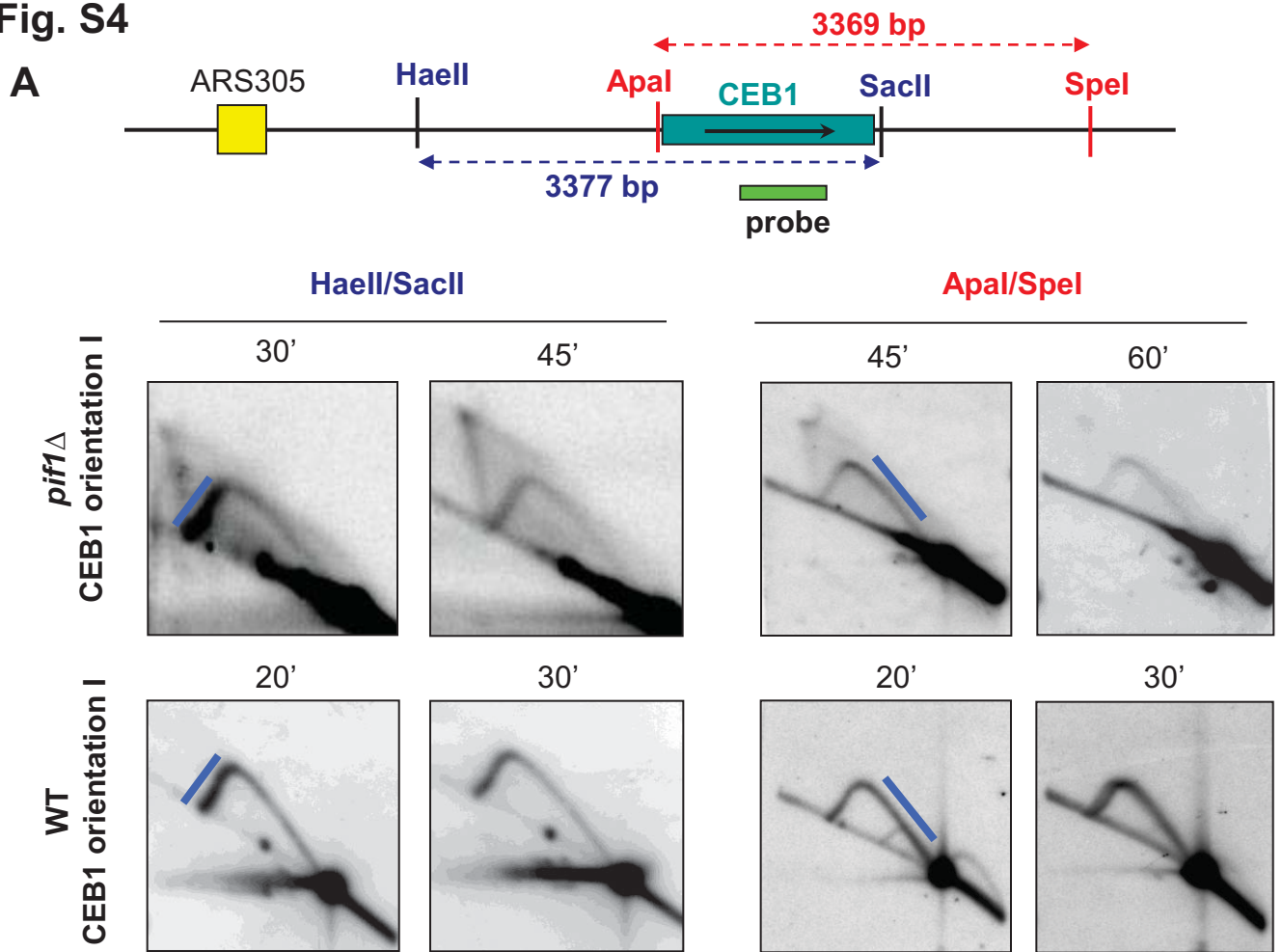
CEB1 orientation: I II - I II I II I II

Genotype: WT *pif1*Δ *rad51*Δ *pif1*Δ *rad51*Δ

p-values (Fischer exact test)	Orientation I	Orientation II
<i>rad51</i> Δ vs. <i>pif1</i> Δ <i>rad51</i> Δ	5.9e <sup>-3</sup>	0.82
<i>pif1</i> Δ vs. <i>pif1</i> Δ <i>rad51</i> Δ	7.5e <sup>-6</sup>	0.13

**Fig. S3**

**Fig. S4**



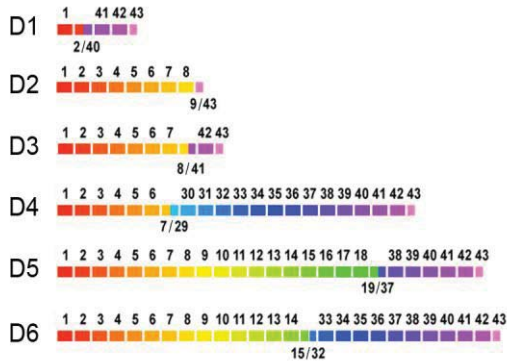
# Supplementary Figure 5

## Parental CEB1-1.8



### Orientation I

#### Deletions (n=6/12)



#### Double-deletions (n=3/12)

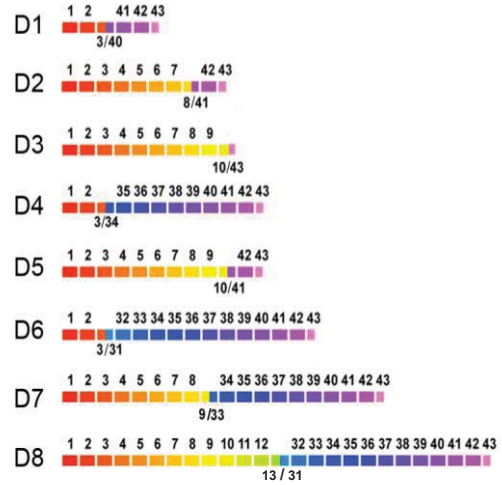


#### Complex (n=3/12)

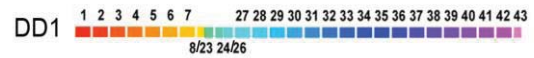


### Orientation II

#### Deletions (n=8/12)



#### Double-deletions (n=1/12)



#### Complex (n=3/12)



Supplementary Table 1

<b>Strain</b>	<b>Relevant genotype</b>
SY2209	MAT a, <i>ura3-1, his3-11,15, leu2-3,112, trp1-1, RAD5</i>
ORT5094-1	WT, <i>URA3</i> - ARS305
ORT6143-13	WT, <i>URA3-hphMX</i> - ARS305
ORT6119-4	WT, CEB1-1.8 (I)* - ARS305
ORT6135-36	WT, CEB1-1.8 (II)** - ARS305
ORT6165-2	WT, CEB1-1.8 (I) - <i>ARS305Δ</i>
ORT6166-52	WT, CEB1-1.8 (II) - <i>ARS305Δ</i>
ORT6582-4	WT, CEB1-Gmut-1.7 (I) - ARS305
ORT6133-2	<i>pif1::KMX</i>
ORT6146-1	<i>pif1::KMX</i> , CEB1-1.8 (I) - ARS305
ORT6136-8	<i>pif1::KMX</i> , CEB1-1.8 (II) - ARS305
ORT6123-1-5	<i>pif1::KMX</i> , CEB1-1.6 (I) - ARS305
ORT6148-3	<i>pif1::KMX</i> , CEB1-1.3 (I) - ARS305
ORT6148-2	<i>pif1::KMX</i> , CEB1-1.1 (I) - ARS305
ORT6148-4	<i>pif1::KMX</i> , CEB1-0.8 (I) - ARS305
ORT6148-1	<i>pif1::KMX</i> , CEB1-0.5 (I) - ARS305
ORT6169-2	<i>pif1::KMX</i> , CEB1-1.8 (I) - <i>ARS305Δ</i>
ORT6170-8	<i>pif1::KMX</i> , CEB1-1.8 (II) - <i>ARS305Δ</i>
ORT6157-1	<i>pif1::KMX</i> , CEB1-Gmut-1.7 (I) - ARS305
ORT6167-3	<i>pif1-K264A</i> , CEB1-1.8 (I) - ARS305
ORT6168-18	<i>pif1-K264A</i> , CEB1-1.8 (II) - ARS305
ORT6127-2	<i>rad27::HIS3</i> , CEB1-1.8 (I) - ARS305
ORT6142-2	<i>rad27::HIS3</i> , CEB1-1.8 (II) - ARS305
ORT6145-4	<i>rad51::LEU2</i> , CEB1-1.8 (I) - ARS305
ORT6144-1	<i>rad51::LEU2</i> , CEB1-1.8 (II) - ARS305
ORT6581-3	<i>rad51::LEU2, pif1::KMX</i> , CEB1-1.8 (I) - ARS305
ORT6150-2	<i>rad51::LEU2, pif1::KMX</i> , CEB1-1.8 (II) - ARS305
ORT6139-5	<i>rad51::LEU2, pif1::KMX</i> , CEB1-1.6 (I) - ARS305
ORT6151-2	<i>rad52::LEU2</i> , CEB1-1.8 (I) - ARS305

Supplementary Table 1

ORT6152-3	<i>rad52::LEU2</i> , CEB1-1.8 (II) - ARS305
ORT6153-2	<i>rad52::LEU2, pif1::KMX</i> , CEB1-1.8 (I) - ARS305
ORT6154-5	<i>rad52::LEU2, pif1::KMX</i> , CEB1-1.8 (II) - ARS305
ORT7107-1	<i>rad54::KMX</i> , CEB1-1.8 (I) - ARS305
ORT7108-1	<i>rad54::KMX</i> , CEB1-1.8 (II) - ARS305
ORT7115-10	<i>rad54::KMX, pif1::HIS3</i> , CEB1-1.8 (I) - ARS305
ORT7116-1	<i>rad54::KMX, pif1::HIS3</i> , CEB1-1.8 (II) - ARS305
ORT7111-1	<i>rev1::KMX</i> , CEB1-1.8 (I) - ARS305
ORT7112-7	<i>rev1::KMX</i> , CEB1-1.8 (II) - ARS305
ORT6574-2	<i>rrm3::KMX</i> , CEB1-1.8 (I) - ARS305
ORT6575-1	<i>rrm3::KMX</i> , CEB1-1.8 (II) - ARS305
ORT6599-1	<i>sgs1::KMX</i> , CEB1-1.8 (I) - ARS305
ORT7100-1	<i>sgs1::KMX</i> , CEB1-1.8 (II) - ARS305

---

**Supplementary Table 2**

<b>Strain</b>	<b>CEB1 size (kb)</b>	<b>Absolute rearrangement frequency</b>
ORT6148-1	0.5	5/92 (5.4%)
ORT6148-4	0.8	4/70 (5.7%)
ORT6148-2	1.1	16/77 (20.8%)
ORT6148-3	1.3	24/72 (33.3%)
ORT6123-1-5	1.6	27/77 (35.1%)
ORT6146-1	1.8	67/119 (56.3%)



**Supplementary Table 3**  
**G4F sequences**

Chr	Start	End	Sequence (G= strand + ; C= strand -)	strand	ARS 5' of G4F sequence				ARS 3' of G4F sequence				interorigin distanc	Distance in % of the inter-origin to the 5' ARS	Machinery in charge of the G4 motif replication
					name	Start	End	distance to G4	name	Start	End	distance to G4			
2	128566	128594	GGGCTGGGCATTATAGGGAGTTTGAAGGG	1	ARS202	63186	63421	65276.5	ARS207.5	198193	198434	69733.5	135010	48.34938153	lagging
2	490618	490648	CCCCCTGAGAGCCCTGCCCTGACCTGCC	-1	ARS216	486661	486909	3848	ARS220	622625	622894	132126.5	135974.5	2.829942379	leading
2	680201	680224	GGGGTAGTAGGGCAGGGAAAGGGG	1	ARS220	622625	622894	57453	no ARS						lagging
3	50757	50780	CCCCAACCCCATCAGTGCCC	-1	ARS305	39158	39706	11336.5	ARS306	74457	74677	23798.5	35135	32.26554718	leading
3	214677	214705	GGGCCAGGGTCCATGGCGGGTCCGGGG	1	ARS310	166494	167340	47774	ARS315	224807	225053	10239	58013	82.35050764	leading
4	258692	258723	GGGGGACAATGGGTGAGTTGTCGGGTAAGGG	1	ARS410	253789	254038	4794	ARS413	329564	329813	70981	75775	6.326624876	lagging
4	411435	411465	CCCACACCCGGGCACCCAAAGTCCACCC	-1	ARS414	408070	408312	3259	ARS415	435056	435388	23772	27031	12.05652769	leading
4	636358	636398	GGGAAAGGGCGTCCGCAAGGGCAAGGGCAAGG	1	ARS419	567490	567737	68764.5	ARS421	639859	640108	3605.5	72370	95.01796324	leading
4	866789	866809	CCCGAGCCCGTCCCGCGCCC	-1	ARS423	753159	753391	113524	ARS428	913780	914029	47105.5	160629.5	70.67444025	lagging
4	1242538	1242557	GGGGAGGGGAAGGGGAGGGG	1	ARS432.5	1165998	1166221	76438	no ARS						lagging
4	1281391	1281420	CCCTTGTCCTCAATACATCCCTGCTGGCCC	-1	ARS432.5	1165998	1166221	115296	no ARS						leading
4	1330164	1330182	CCCCATCCCTCCCTCCC	-1	ARS432.5	1165998	1166221	164063.5	no ARS						leading
4	1359723	1359756	CCGATCCCATATGCCATCCCAAGATCATGCC	-1	ARS432.5	1165998	1166221	193630	no ARS						leading
4	1525262	1525294	GGGTACGAGTGGGGAGGTAGGGTAATGGAGGG	1	ARS432.5	1165998	1166221	359168.5	no ARS						lagging
4	1525362	1525417	GGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTG	1	ARS432.5	1165998	1166221	359280	no ARS						lagging
5	327200	327222	CCCCAACTTCTCCCTCCCTCCC	-1	ARS514	287504	287750	39584	ARS516	353504	353751	26416.5	66000.5	59.97530322	lagging
5	522885	522910	CCCTATTCCCGCCCATACACCCC	-1	ARS520	498417	499343	24017.5	ARS523	569020	570085	46655	70672.5	33.984223	leading
5	545172	545196	GGGAAAAAGGGGGGGGGGGGGGG	1	ARS520	498417	499343	46304	ARS523	569020	570085	24368.5	70672.5	65.51911988	leading
6	190598	190629	GGGATTCAATGGGGGAAAGTGGGAGCGGG	1	ARS606	167606	168041	22790	ARS607	199382	199493	8824	31614	72.0883153	leading
6	255485	255534	GGGGCACACGTGCGGGAGTTTCAAAGGGCAGAA	1	ARS607	199382	199493	56072	no ARS						lagging
7	197275	197302	CCCCTCCCCCTCCTCCGAAGCAACCC	-1	ARS707	163180	163447	33975	ARS710	203917	204159	6749.5	40724.5	83.42643863	lagging
7	302728	302749	CCCCGCCCTGTCCCTTTGCC	-1	ARS714	285951	286246	16640	ARS717	388658	388892	86036.5	102676.5	16.20623999	leading
7	467441	467467	CCCCAGCAAACACCCCTCCTCCC	-1	ARS718	421093	421342	46236.5	ARS719	484932	485160	17592	63828.5	72.43864418	lagging
7	726593	726618	CCCATACCTATATACACCCATACCC	-1	ARS728	715273	715556	11191	ARS729	777967	778216	51486	62677	17.85503454	leading
7	851433	851471	CCCATCGGAAAGCCCATATCCCGCCATATTCATC	-1	ARS731	834492	834736	16838	ARS731.5	888350	888599	37022.5	53860.5	31.26224227	leading
8	289308	289332	GGGGGGGGTGGGAAAAAAGGGG	1	ARS807	133347	133591	155851	ARS820	447619	447853	158416	314267	49.59190752	lagging
8	382904	382931	CCCTCCCCCTGGCGGAACCCATCATCCC	-1	ARS807	133347	133591	249448.5	ARS820	447619	447853	64818.5	314267	79.37470368	lagging
9	356347	356385	GGGTACGGTGGGTAATAAGGGAAGGTATCGGGAT	1	ARS919	341853	342096	14391.5	ARS920	357156	357393	908.5	15300	94.0620915	leading
10	623184	623209	GGGGTGGTGGCCCTAAGGGAGCGGG	1	ARS1019	612542	612975	10438	ARS1021	683328	683817	60376	70814	14.74002316	lagging
10	639703	639721	CCCCCCCCCCCCCCCC	-1	ARS1019	612542	612975	26953.5	ARS1021	683328	683817	43860.5	70814	38.06238879	leading
11	321974	322001	GGGGAGGGGAAAGAAAGGGGAAGGGG	1	ARS1106.7	213080	213385	108755	ARS1109	329322	329571	7459	116214	93.5816683	leading
12	169109	169143	GGGTATTATAATGGGTTCTGTGAGGGCAAAGAGGG	1	ARS1209	156646	156883	12361.5	ARS1211	231179	231422	62174.5	74536	16.58460341	lagging
12	421491	421507	GGGGAGGGGGGGGGGG	1	ARS1215	412668	412897	8716.5	ARS1216	450485	450726	29106.5	37823	23.04550141	lagging
12	442819	442839	CCCCCCCCACAACCTACCCC	-1	ARS1215	412668	412897	30046.5	ARS1216	450485	450726	7776.5	37823	79.43975888	lagging
12	759130	759160	GGGGGCACACCGGGTAGGGGTCCGGAGGG	1	ARS1223	744942	745179	14084.5	ARS1226	794020	794269	34999.5	49084	28.69468666	lagging
12	906560	906587	CCCCAGGACCTCTCACCCCAACCC	-1	ARS1226	794020	794269	112429	no ARS						leading
12	954518	954553	GGGGGGAGTGGAGAGGGCTGTGAGGGTCAAGTC	1	ARS1226	794020	794269	160391	no ARS						lagging
13	70381	70408	GGGAGATAGGAAGGGGCGGGGTCCGGG	1	ARS1303	31687	31935	38583.5	ARS1305	94216	94463	23945	62528.5	61.70546231	leading
13	409843	409869	CCCCAGCCCGGCCCGGTCCACCCC	-1	ARS1312	370976	371221	38757.5	ARS1320	535595	535843	125863	164620.5	23.54354409	leading
13	482827	482852	GGGAGATAAGAGGGAGCAGGGTGGGG	1	ARS1312	370976	371221	111741	ARS1320	535595	535843	52879.5	164620.5	67.87793744	leading
14	51913	51933	GGGAAGGGTGGAGGAGCTGGG	1	no ARS				ARS1407	89528	89802	37742			leading
14	661656	661674	GGGAGGAAGGGCAGAGGG	1	ARS1426	635660	635901	25884.5	no ARS						lagging
15	236337	236363	CCAGGCGCAGGCCAAGCCCAAGCCC	-1	ARS1510	166974	167220	69253	ARS1511	277529	277778	41303.5	110556.5	62.6403694	lagging
15	494609	494639	CCCCTTTGATATCCACGCCCCACGCCCC	-1	ARS1514	489645	490129	4737	ARS1516	566409	566643	71902	76639	6.18092616	leading
15	780553	780574	GGGATAGGGCTCTATGGGTGGG	1	ARS1516	566409	566643	214037.5	ARS1526	874190	874434	93748.5	307786	69.54101226	leading
15	831998	832026	GGGATTCGGCAAATCGATGGGCGCTGGG	1	ARS1516	566409	566643	265486	ARS1526	874190	874434	42300	307786	86.25668484	leading
15	1059283	1059307	CCCCCCTCGCCCCGACGCGACCC	-1	ARS1529	981454	981690	77723	no ARS						leading
16	307880	307898	GGGGGGAGGGGAGTTGGG	1	ARS1614	289483	289704	18295.5	ARS1618	384536	384784	76771	95066.5	19.24494959	lagging
16	490095	490121	GGGCTTGGCTTGGGCTTGGGCTGGG	1	ARS1619	418132	418359	71862.5	ARS1621	511619	511940	21671.5	93534	76.83035046	leading
16	811918	811958	CCCACGCCAGCCAGTCTAATACCCCAATTGGT/	-1	ARS1626.5	776921	777152	34901.5	ARS1627	819153	819393	7335	42236.5	82.6335042	lagging

lagging= 27  
leading= 23

### II.C.3. Résultats supplémentaires

Afin de mieux comprendre les mécanismes d'instabilité sur le brin à synthèse continue, et étudier l'éventuelle formation de G-quadruplexes sur la matrice du brin discontinue, j'ai mesuré l'instabilité de CEB1 dans différents mutants d'hélicases, de réplication, et du PRR. J'ai également cherché à déterminer la nature moléculaire de la lésion recombino-gène induite par les G-quadruplexes durant la réplication. Les résultats présentés ici sont encore préliminaires.

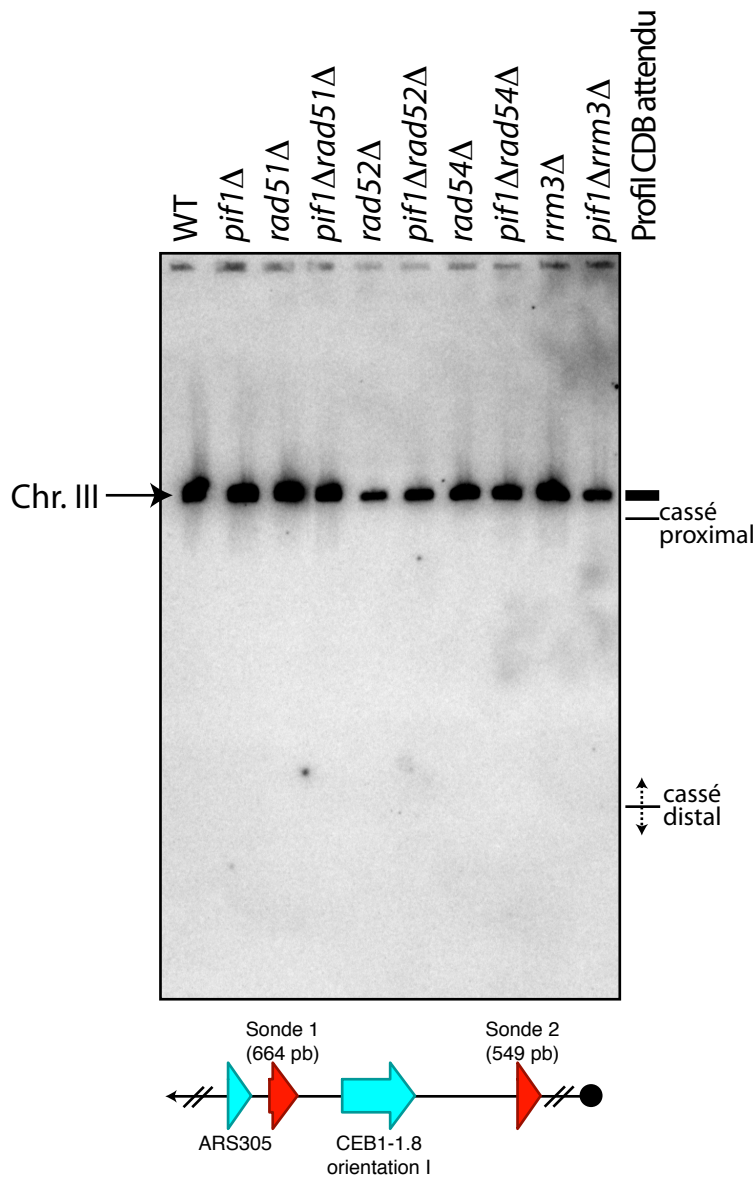
#### II.C.3.a. Nature de la lésion induite par CEB1

La voie de la recombinaison homologue est requise pour la formation du spike en gel 2D au niveau de CEB1 lors de la phase S et pour la formation de réarrangements de CEB1. La nature de la lésion peut donc être une CDB, une fourche cassée (CDB asymétrique), ou une région d'ADNsb. Des arguments génétiques et physiques sont en défaveur de la formation d'une cassure de la molécule d'ADN.

Du point de vue génétique, la dépendance totale ou quasi-totale à Rad51 et Rad54 suggère qu'un mécanisme de SSA (Rad51- et Rad54-indépendant mais Rad52-dépendant (Pâques and Haber 1999)) n'est pas à l'œuvre au sein de CEB1, alors qu'une CDB au sein d'une séquence répétée en tandem est un substrat propice à l'accomplissement d'un tel mécanisme (**Figure 28B**).

J'ai examiné la formation éventuelle d'une CDB au niveau de CEB1 par électrophorèse en champ pulsé du chromosome III dans les mutants *rad51Δ*, *rad52Δ*, ou *rad54Δ*, combinés ou non à la délétion de *PIF1*. Aucun chromosome cassé n'est détecté dans ces mutants, pourtant déficients pour la réparation d'une CDB (**Figure 33**). Étant donnée la fréquence de ces lésions par génération en absence de Pif1 ( $\approx 5\%$ ) et l'intensité des intermédiaires de recombinaison visibles sur cellules synchronisées en gel 2D, ces chromosomes cassés auraient du être détectés. Aussi est-il probable que la lésion recombino-gène ne soit pas une cassure de fourche ou une CDB, mais plutôt, par élimination, une région d'ADNsb à la fourche ou laissée en amont de la fourche.

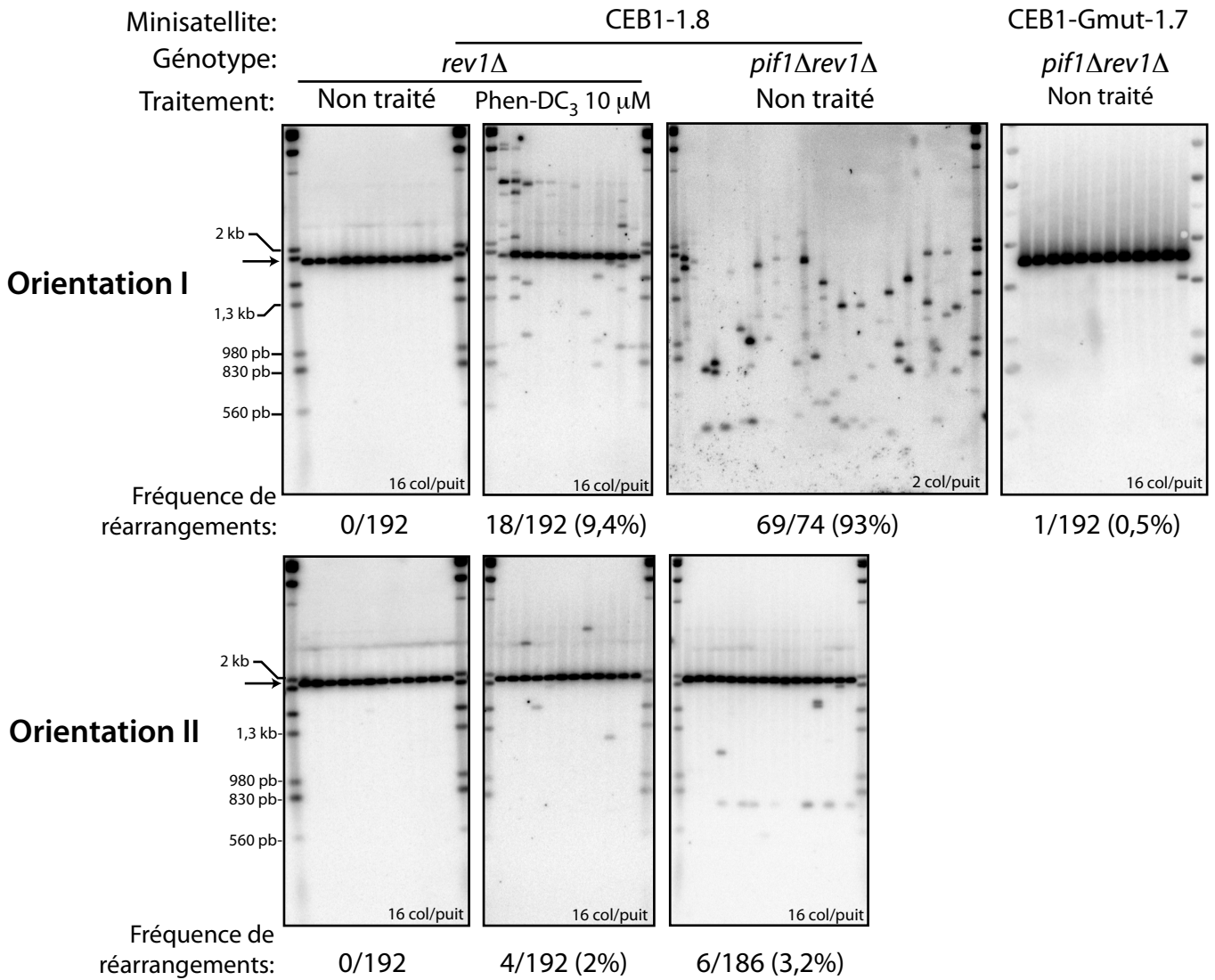
#### II.C.3.b. Mutants de polymérase translésionnelles



**Figure 33: Aucune CDB n'est détectée au niveau de CEB1 en absence de Pif1 par gel d'électrophorèse en champ pulsé.** Analyse de souches WT (ORT6119-4), *pif1*Δ (ORT6146-2), *rad51*Δ (ORT6145-4), *pif1*Δ*rad51*Δ (ORT6581-3), *rad52*Δ (ORT6151-2), *pif1*Δ*rad52*Δ (ORT6153-2), *rad54*Δ (ORT7107-2), *pif1*Δ*rad54*Δ (ORT7115-10), *rrm3*Δ (ORT6574-2), et *pif1*Δ*rrm3*Δ (ORT6595-2) contenant CEB1-1.8 dans l'orientation I à proximité de l'ARS305 par gel d'électrophorèse en champ pulsé. Deux sondes situées de part et d'autres de CEB1 ont été utilisées. La sonde 1 est située à 1036 pb en position distale de CEB1, et la sonde 2 est située à 2404 pb en position proximale de CEB1. La taille du chromosome III ainsi que la taille des deux fragments cassés sont indiquées: le fragment distal (sonde 1) mesure 41 kb, et le fragment proximal (sonde 2) mesure 275 kb.

Si l'intermédiaire recombino-gène est une région d'ADNsb qui persiste à la fourche ou derrière elle, les mécanismes dits « error-prone » (translèsion) et « error-free » (template-switch) du PRR sont susceptibles d'intervenir pour permettre sa réplication. Cependant, seul le template-switch est à même de générer des réarrangements des séquences répétées en tandem (**Chapitre I.B.4.a**). Pour étudier si la voie « error prone » est en compétition avec la voie « error free » au niveau de CEB1, j'ai étudié l'effet de mutants de polymérases translésionnelles Rev1 et Polζ (Rev3/7) sur l'instabilité de CEB1 dans les orientations I et II. Rev1 est impliquée dans l'accommodation du premier nucléotide face à une lésion et Polζ prolonge la synthèse à partir de ce premier nucléotide incorporé (**Chapitre I.B.3.ii, Figure 25**). Rev1 est une polymérase translésionnelle spécifique des matrices riches en G : elle incorpore préférentiellement des C, même en présence d'un site abasique (Nelson et al. 1996; Haracska et al. 2002), et est requise pour la réplication fidèle des guanines endommagées par un traitement au 4-NQO (Washington et al. 2004; Wiltout and Walker 2011). Une publication récente suggère un rôle de cette polymérase dans la réplication d'une séquence G4P dans des cellules de poulet DT40, lorsque le G-quadruplexe se forme sur le brin matrice de la machinerie de synthèse continue (Sarkies et al. 2010). Cependant, ni la mutagénèse du motif G4P, ni le traitement des cellules avec un ligand de G-quadruplexe n'ont été réalisées, ce qui ne permet pas de conclure que le phénotype observé dépend bien de la présence du G-quadruplexe.

CEB1 est stable dans les deux orientations dans les mutants *rev3Δ* et *rev1Δ*, et dans les double mutants *pif1Δrev3Δ* et *pif1Δrev1Δ* avec CEB1 dans l'orientation II (**Table 3**). Ainsi, l'absence d'instabilité de CEB1 dans l'orientation II en absence de Pif1 n'est pas imputable à la réplication/réparation des G-quadruplexes qui s'y forment par les polymérases Polζ et Rev1 au détriment du template-switch. Dans l'orientation I, CEB1 est instable dans le double-mutant *pif1Δrev3Δ*, mais en présence d'un nombre trop important d'évènements clonaux, la fréquence de réarrangement n'a pu être précisément déterminée (**Table 3**). Dans le mutant *pif1Δrev1Δ*, en revanche, l'instabilité de CEB1 est très importante (93%, 69/74, p-value vs. *pif1Δ* =  $1,3e^{-8}$ ) (**Figure 34**). Cette instabilité dépend de la capacité de CEB1 à former des G-quadruplexes puisque l'allèle CEB1-Gmut-1.7 est stable dans la souche *pif1Δrev1Δ* (0,5%, 1/192) (**Figure 34**). En revanche, l'instabilité de CEB1 dans une souche *rev1Δ* traitée par Phen-DC<sub>3</sub> n'est pas



**Figure 34: Instabilité de CEB1 en absence de la polymérase translésionnelle Rev1.** (A) Analyse de l'instabilité de CEB1-1.8 inséré à proximité de l'ARS305 dans les orientations I (haut) et II (bas) par Southern Blot, dans les souches *rev1Δ* (respectivement ORT7111-1 et ORT7112-7) traitées ou non par Phen-DC3, et dans les souches *pif1Δrev1Δ* (respectivement ORT7117-8 et ORT7118-1). L'instabilité de CEB1-Gmut-1.7 (orientation I) dans une souche *pif1Δrev1Δ* (ORT7198-1) a également été testée. Digestions: Apal/XhoI (orientation I) et Apal/NcoI (orientation II) pour CEB1-1.8; SacII/ApaI pour CEB1-Gmut-1.7. Sondes: CEB1 et CEB1-Gmut. Le nombre de colonies analysées dans chaque puits et la fréquence de réarrangements totale est indiquée. La flèche indique la position du minisatellite parental.

**Table 3: Instabilité de CEB1-1.8 inséré à proximité de l'ARS305 dans différents mutants**

Genotype	Orientation I	Orientation II	p-value I vs. II
WT	2/384 (0.5%)	1/192 (0.5%)	NS
<b>Helicases</b>			
<i>pif1Δ</i>	67/119 (56.3%)*	14/430 (3.2%)	4.8e <sup>-39</sup>
<i>pif1Δ + scPIF1</i>	1/93 (1.1%) <sup>°</sup>	ND	NA
<i>pif1-K264A</i>	27/76 (35%)* <sup>°</sup>	6/192 (3.1%)	3.8e <sup>-11</sup>
<i>pif1-m2</i>	7/183 (3.8%) <sup>°</sup>	5/188 (2.6%)	NS
<i>sgs1Δ</i>	2/192 (1%)	2/192 (1%)	NS
<i>rrm3Δ</i>	0/192	0/192	NS
<i>chl1Δ</i>	0/192	0/192	NS
<i>pif1Δsgs1Δ</i>	>95% <sup>°</sup>	9/190 (4.7%)	NA
<i>pif1Δrrm3Δ</i>	5/79 (6.3%) <sup>°</sup>	12/384 (3.1%)	NS
<i>pif1Δchl1Δ</i>	20/48 (41.7%)	6/192 (3.1%)	2.4e <sup>-11</sup>
<b>Réplisome</b>			
<i>rad27Δ</i>	14/81 (17.3%)*	9/83 (10.8%)* <sup>σ</sup>	NS
<i>mrc1Δ</i>	1/192 (0.5%)	2/192 (1%)	NS
<i>tof1Δ</i>	0/192	1/192 (0.5%)	NS
<i>pif1Δmrc1Δ</i>	24/51 (47%)	11/191 (5.7%)	2.9e <sup>-11</sup>
<i>pif1Δtof1Δ</i>	51/139 (36.7%) <sup>°</sup>	17/171 (9.9%) <sup>°</sup>	1.6e <sup>-8</sup>
<b>Recombinaison homologue</b>			
<i>rad51Δ</i>	2/176 (1.1%)	2/384 (0.5%)	NS
<i>rad52Δ</i>	1/192 (0.5%)	0/176	NS
<i>rad54Δ</i>	0/192	5/192 (2.6%)	NS
<i>pif1Δ rad51Δ</i>	6/191 (3.1%) <sup>σ</sup>	0/192	0.03
<i>pif1Δ rad52Δ</i>	1/192 (0.5%) <sup>σ</sup>	1/192 (0.5%)	NS
<i>pif1Δ rad54Δ</i>	17/167 (10.2%)* <sup>σ</sup>	10/191 (5.2%)*	NS
<b>Synthèse translésionnelle</b>			
<i>pol32Δ</i>	1/192 (0.5%)	2/192 (1%)	NS
<i>rev1Δ</i>	0/192	0/192	NS
<i>rev3Δ</i>	0/192	2/192 (1%)	NS
<i>pif1Δpol32Δ</i>	29/61 (47.5%)	14/189 (7.4%) <sup>°</sup>	2.7e <sup>-11</sup>
<i>pif1Δrev1Δ</i>	69/74 (93%)	6/186 (3.2%)	4.3e <sup>-39</sup>
<i>pif1Δrev3Δ</i>	10/31 (32%) clonaux	4/192 (2%)	4.5e <sup>-7</sup>

\*= p-value  $x\Delta$  vs. WT < 0.05

<sup>°</sup>= p-value *pif1ΔxΔ* or other *PIF1* mutants vs. *pif1Δ* < 0.05

**Table 4: Instabilité de CEB1-1.8 inséré dans l'orientation I à proximité d'ARS305 après traitement par Phen-DC<sub>3</sub> 10 µM dans différents mutants**

Genotype	Orientation I			Orientation II		
	Untreated	Phen-DC <sub>3</sub> 10 µM	p-value	Untreated	Phen-DC <sub>3</sub> 10 µM	p-value
WT	2/384 (0.5%)	111/992 (11.2%)*	4,5e <sup>-8</sup>	1/192 (0.5%)	57/1376 (4.1%)	7,1e <sup>-3</sup>
<i>pif1Δ</i>	67/119 (56.3%)*	ND	NA	14/430 (3.2%)*	47/480 (9.8%)*	8,9e <sup>-5</sup>
<b>Helicases</b>						
<i>sgs1Δ</i>	2/192 (1%)	43/192 (22.4%)*	5,2e <sup>-12</sup>	2/192 (1%)	ND	NA
<i>rrm3Δ</i>	0/192	39/384 (10%)	1,4e <sup>-7</sup>	0/192	12/384 (3.1%)	0,01
<b>Replication proteins</b>						
<i>tof1Δ</i>	0/192	45/384 (11.7%)	6,5e <sup>-9</sup>	1/192 (0.5%)	30/384 (7.8%)*	5,3e <sup>-5</sup>
<b>Homologous recombination and PRR proteins</b>						
<i>rad51Δ</i>	2/176 (1.1%)	14/576 (2.4%)*	NS	2/384 (0.5%)	16/564 (2.8%)	0,013
<i>rad52Δ</i>	1/192 (0.5%)	1/154 (0.6%)*	NS	0/176	0/192*	NS
<i>rad54Δ</i>	0/192	1/384 (0.3%)*	NS	5/192 (2.6%)	8/368 (2.2%)	NS
<b>Translesion polymerases</b>						
<i>pol32Δ</i>	1/192 (0.5%)	32/384 (8.3%)	3e <sup>-5</sup>	2/192 (1%)	20/384 (5.2%)	0,011
<i>rev1Δ</i>	0/192	18/192 (9.4%)	5e <sup>-6</sup>	0/192	4/192 (2%)	NS

\*= p-value *xΔ* vs. WT < 0.05

significativement différente de ce qui est observé avec la souche WT, dans les deux orientations (**Table 4**). Ces résultats préliminaires suggèrent que Rev1 est impliquée dans la réplication des G-quadruplexes de CEB1, et que Phen-DC<sub>3</sub> inhibe cette activité *in vivo*.

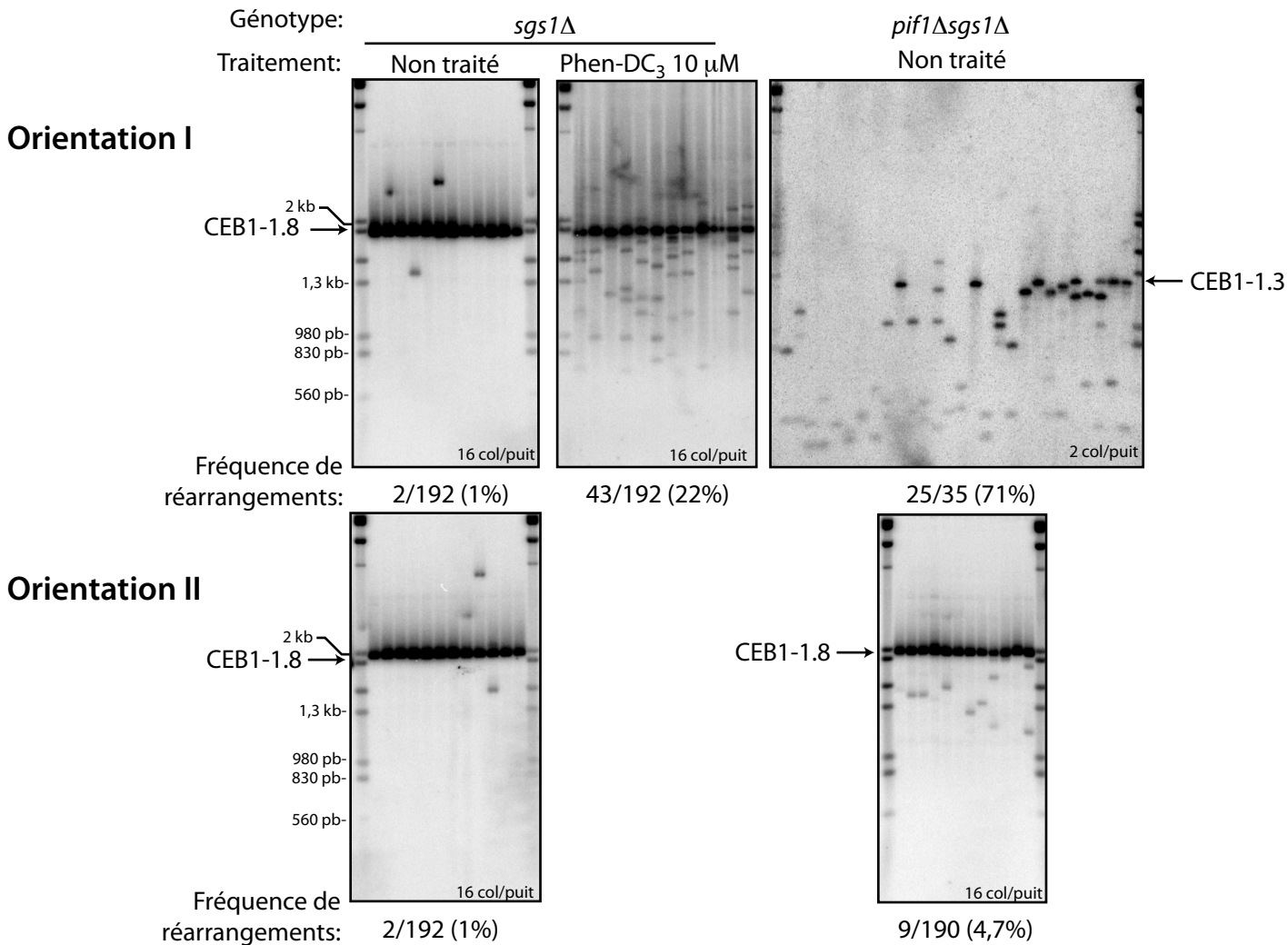
### II.C.3.c. Mutants d'hélicases

La faible instabilité de CEB1 dans l'orientation II est-elle due à des activités de déroulement redondantes à Pif1 sur ce brin ? J'ai testé l'effet de la délétion de trois hélicases candidates (Sgs1, Rrm3, Chl1), basées sur leur activité *in vitro* sur les G-quadruplexes (Sgs1) (Sun et al. 1999), l'homologie avec Pif1 (Rrm3) ou des hélicases déroulant les G-quadruplexes dans d'autres organismes (Chl1) (Wu et al. 2012). Aucun de ces simples mutants (*sgs1Δ*, *rrm3Δ*, et *chl1Δ*) ne présente une instabilité de CEB1 dans les orientations I et II accrue en comparaison de la souche WT (**Table 3**). En combinaison avec la délétion de *PIF1*, ou en présence du ligand Phen-DC<sub>3</sub>, aucun de ces mutants n'a d'effet significatif sur l'instabilité de CEB1 dans l'orientation II (**Table 3 et 4**). En revanche, ces mutants affectent différemment l'instabilité de CEB1 dans l'orientation I.

L'instabilité dans le double mutant *pif1Δchl1Δ* n'est pas significativement différente de celle observée dans le mutant *pif1Δ*, suggérant que Chl1 ne joue aucun rôle dans la stabilisation de CEB1. La situation est différente dans les autres mutants.

Dans le double mutant *pif1Δsgs1Δ*, l'instabilité de CEB1-1.8 dans l'orientation I est quasiment de 100%. Pour estimer plus précisément l'augmentation d'instabilité entre les mutants *pif1Δ* et *pif1Δsgs1Δ*, un allèle plus court de CEB1 (CEB1-1.3) a été examiné. La fréquence de réarrangement de cet allèle s'accroît de 33% (24/72) en absence de Pif1 à 71% (25/35) dans le double-mutant *pif1Δsgs1Δ* (p-value = 3,6e<sup>-4</sup>). De manière similaire à la délétion de *PIF1*, le traitement par Phen-DC<sub>3</sub> conduit à une instabilité accrue CEB1-1.8 dans le mutant *sgs1Δ* (22,4%, 43/192) en comparaison d'une souche WT (11,2%, 111/992, p-value = 2,4e<sup>-6</sup>) (**Figure 35, Table 4**). Ainsi, Sgs1 limite l'instabilité de CEB1 lorsque les G-quadruplexes se forment sur le brin à synthèse continue. Savoir si cette limitation de l'instabilité est due à son activité de déroulement des G-quadruplexes ou à son activité de résolution des intermédiaires de recombinaison





**Figure 35: Instabilité de CEB1 en absence de l'hélicase Sgs1.** Analyse de l'instabilité de CEB1 inséré à proximité de l'ARS305 dans les orientations I (haut) et II (bas) par Southern blot, dans des souches *sgs1Δ* (respectivement ORT6599-1 et ORT7100-1) traitées ou non par Phen-DC3, et dans des souches *pif1Δsgs1Δ* (respectivement ORT7105-10-4 et ORT7106-2). L'allèle testé est CEB1-1.8, à l'exception de la souche ORT7105-10-4, qui contient un allèle plus court d'environ 1,3 kb (CEB1-1.3). Digestions: XhoI/ApaI (orientation I) et NcoI/ApaI (orientation II).

en non-crossover au sein du complexe STR (Sgs1-Top3-Rmi1) (Gangloff et al. 1994; Chang et al. 2005; Mullen et al. 2005; Ashton et al. 2011) reste à déterminer.

Une souche *rrm3Δ* traitée par Phen-DC<sub>3</sub> présente une instabilité de CEB1 de 10% (39/384) dans l'orientation I et 3.1% (12/384) dans l'orientation II, similaire à ce qui est observé dans une souche WT traitée (**Table 3**). En revanche, l'instabilité de CEB1 dans l'orientation I est significativement diminuée dans le double-mutant *pif1Δrrm3Δ* (6,3%, 5/79) en comparaison du mutant *pif1Δ* (67/119, p-value = 5,1e<sup>-14</sup>)(**Table 3**). Ce résultat peut s'expliquer par l'action potentielle de Rrm3 à deux niveaux : soit en amont de Pif1 pour permettre la formation d'un G-quadruplexe par CEB1, soit en aval pour permettre la réparation de la lésion induite par le G-quadruplexe. Dans ce dernier cas, la suppression de l'instabilité devrait être associée à une augmentation de la létalité, comme observé dans le mutant *pif1Δrad51Δ* (**Article 3, Figure S2**). Pourtant, aucune perte de viabilité n'est observée sur 18 couples mère-fille *pif1Δrrm3Δ* analysés. Dans ce cas, quel rôle Rrm3 joue-t-elle en amont de Pif1?

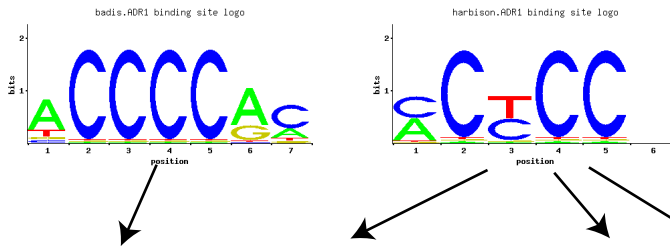
Une activité de Rrm3 pourrait être d'éliminer un complexe protéine-ADN (Ivessa et al. 2000; Ivessa et al. 2003; Makovets et al. 2004) au sein de CEB1 et ainsi permettre la formation de G-quadruplexes. Une protéine candidate est le facteur de transcription Adr1 qui possède des sites de fixation consensus au sein de CEB1 (**Figure 36A**). Cependant son inactivation n'a pas d'effet : CEB1 est stable dans le triple mutant *pif1Δrrm3Δadr1Δ*, pas différent de ce qui est observé dans le double mutant *pif1Δrrm3Δ* (**Figure 36B**). Ainsi, le rôle de Rrm3 reste à déterminer. L'hypothèse alternative que nous souhaitons tester au laboratoire serait une rôle indirect de Rrm3 sur la concentration intracellulaire en dNTP. En effet, l'activation du checkpoint intra-S (Mec1 et Rad53 phosphorylé) dans le mutant *rrm3Δ* et *rrm3Δpif1Δ* (**Figure 37A**) aboutit à la dégradation de Sml1 (un inhibiteur de la RNR)(Zhao et al. 2001; O'Rourke et al. 2005; Hashash et al. 2011), et à une augmentation de la vitesse de synthèse (Poli et al. 2012). L'étude des mutants *sml1Δ* et *pif1Δsml1Δ* est en cours.

#### II.C.3.d. Mutants de couplage et du checkpoint intra-S : Tof1 et Mrc1

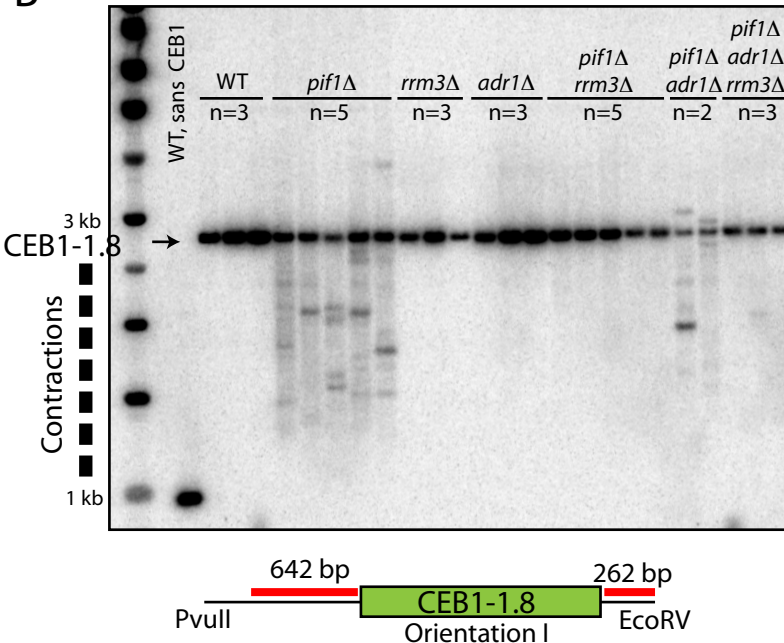
Mrc1 et Tof1 permettent le couplage des MCM2-7 à Polε en cas d'inhibition généralisée de la réplication, limitant potentiellement l'accumulation d'ADNsb à la fourche (**Chapitre I.B.1.c.i**). J'ai voulu examiner si le couplage entre hélicase et

A

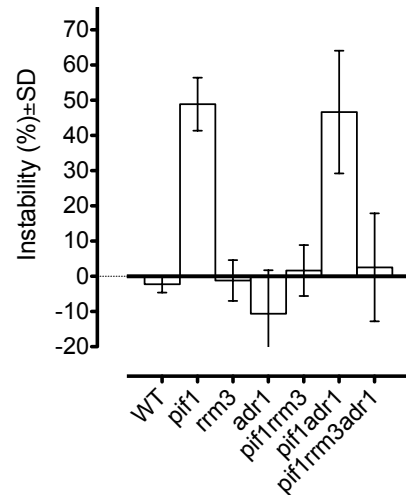
Sites consensus de fixation d'Adr1

Motif de CEB1: TCAGCCAGGGA**CCTCC**GCAGGCCAC**CCTCCCTCC**CCCC

B



Quantification des smears

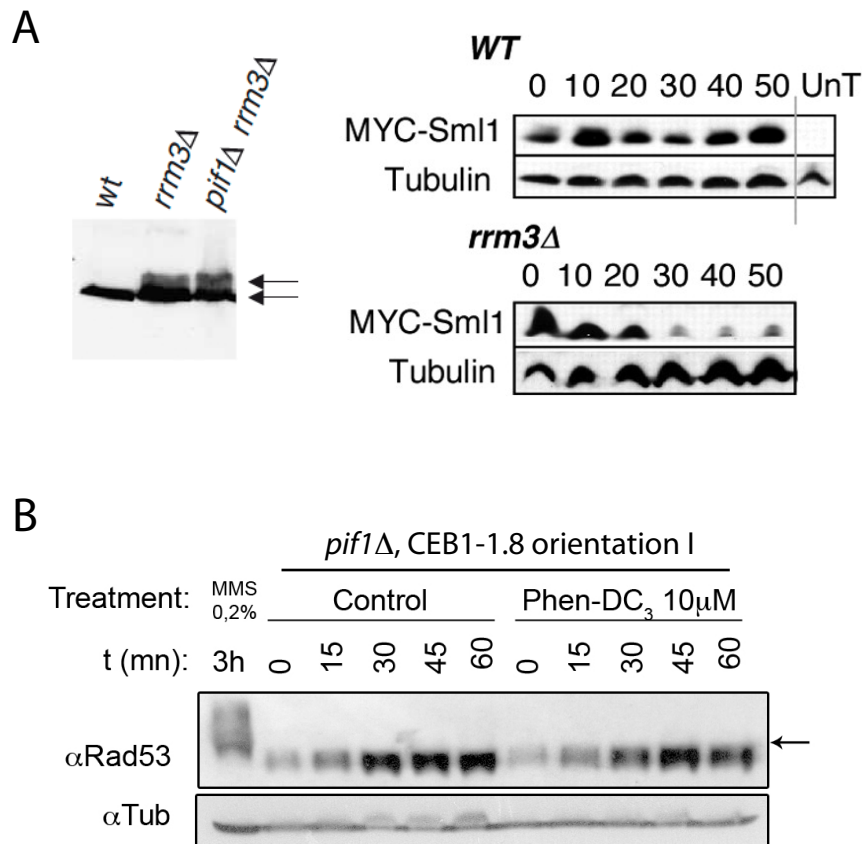


**Figure 36: La protéine Adr1 n'a pas d'effet sur l'instabilité de CEB1 dans les mutants *pif1Δ* et *pif1Δrrm3Δ*.** (A) Deux motifs trouvés significativement enrichis aux sites de fixation d'Adr1 in vivo (source, SGD). La présence de ces sites, à un nucléotide près, dans le motif de CEB1 est indiqué. (B) Analyse de la taille de CEB1 dans une population de levures par Southern blot.

La digestion employée laisse 953 pb d'ADN flanquant, de part et d'autre de CEB1, contre lesquelles est dirigée la sonde (rouge). La sonde n'est pas dirigée contre CEB1 pour éviter que l'intensité du signal varie avec la taille de l'allèle. Le fragment contenant CEB1-1.8 mesure 2758 pb (indiqué sur le blot). Des souches WT (ORT6119), *pif1Δ* (ORT6146), *rrm3Δ*(ORT6574), *adr1Δ* (ORT7168), *pif1Δrrm3Δ* (ORT6595 et AND1257-4D/5C), *pif1Δadr1Δ* (AND1257-4C/15A) et *pif1Δrrm3Δadr1Δ* (AND1257-6A/7B/8D) contenant CEB1-1.8 dans l'orientation I, ainsi qu'une souche sans CEB1 (SY2209), ont été analysées. n= nombre de clones indépendants analysés pour chaque génotype. L'instabilité du minisatellite dans la population est mesuré par quantification du signal du smear divisé par le signal total mesuré dans le puit. La moyenne des clones et l'écart type est présenté à droite.

polymérase réalisé par Mrc1 et Tof1 prévenait la formation de G-quadruplexes au sein de CEB1, et donc son instabilité. Dans les simples mutant *mrc1Δ* et *tof1Δ*, CEB1-1.8 est stable dans les deux orientations (**Table 3**). Dans le double-mutant *pif1Δmrc1Δ*, l'instabilité de CEB1-1.8 dans les orientations I et II (respectivement 47 et 5,7%) n'est pas significativement différente de ce qui est mesuré dans le mutant *pif1Δ* (respectivement 56,3 et 3,2%)(**Table 3**). Le double-mutant *pif1Δtof1Δ* présente en revanche une instabilité de CEB1 légèrement diminuée dans l'orientation I (36,7%) en comparaison du mutant *pif1Δ* (p-value = 0,013), associée à une augmentation significative de la fréquence de réarrangements dans l'orientation II (9,9%, p-value vs. *pif1Δ* =  $1,7e^{-3}$ ) (**Table 3**). Un changement dans la dynamique de réplication de CEB1 pourrait expliquer cette diminution de l'instabilité dans l'orientation I et cette augmentation dans l'orientation II. En faveur de cette interprétation, le déclenchement d'une ARS cryptique entre *ARS305* et *ARS306* a été observé dans les mutants *mrc1Δ* et *tof1Δ* traités à l'HU (Crabbe et al. 2010). Cette observation doit être vérifiée dans la souche *pif1Δtof1Δ* non-traitée. Cette diminution de l'asymétrie de l'instabilité mis à part, les délétions de *MRC1* et *TOF1* n'induisent pas d'instabilité supplémentaire de CEB1 en absence de Pif1. Ceci indique que les protéines de couplage Mrc1 et Tof1 ne participent pas à la stabilisation de CEB1 en absence de Pif1.

Je me suis également intéressé à une éventuelle activation du checkpoint intra-S en suivant au cours de la phase S l'état de phosphorylation de Rad53. Aucune phosphorylation de Rad53 n'est détectée dans une souche *pif1Δ* contenant CEB1 dans l'orientation I, et ce même lorsque les cellules sont traitée par Phen-DC<sub>3</sub> (**Figure 37B**). Ceci peut-être dû à une activation trop faible (une seule fourche affectée) dans une fraction minoritaire de la population pour que la phosphorylation de Rad53 soit détectable par Western blot. Aussi le rôle d'activation du checkpoint par Mrc1 dans notre système n'a pas pu être étudié (Osborn and Elledge 2003; Tourriere et al. 2005; Crabbe et al. 2010).



**Figure 37: Rad53 est phosphorylé dans les mutants *rrm3Δ* et *pif1Δrrm3Δ*, mais pas dans le mutant *pif1Δ*.** (A) Western blot réalisé à gauche: avec un anticorps anti-Rad53 dans une souche WT, *rrm3Δ*, et *pif1Δrrm3Δ* (O'Rourke *et al.* Gene 2005). A droite: avec un anticorps anti-Myc, à différents temps (min) après l'entrée en phase S dans une souche WT ou *rrm3Δ* (Hashash *et al.* J. Cell Sci. 2011). (B) Phosphorylation de Rad53 au cours de la phase S dans une souche *pif1Δ* contenant CEB1-1.8 dans l'orientation I (ORT6146-2), traitées ou non par Phen-DC3. La flèche indique la position où la forme phosphorylée est attendue, grâce à un contrôle positif pour la phosphorylation de Rad53 (traité 3h au MMS).

**Article 4 :**

**Stimulation of Gross Chromosomal Rearrangements  
by the Human CEB1 and CEB25 Minisatellites in  
*S. cerevisiae* Depends on G-quadruplexes or Cdc13**



## **II.D. Article 4: The induction of Gross Chromosomal Rearrangements by the Human GC-Rich Minisatellites CEB1 and CEB25 Depends On G-quadruplexes and Cdc13**

### **II.D.1. Contexte**

Les extrémités chromosomiques humaines (5-10% terminaux) subissent plus souvent des réarrangements que les parties internes des chromosomes, chez les individus comme à l'échelle de l'évolution chez les primates ((Linardopoulou et al. 2005; Shaffer et al. 2006; Baldwin et al. 2008; Rudd et al. 2009; Kaminsky et al. 2011). Les facteurs impliqués dans l'instabilité de ces séquences ne sont pas connus. Il est intéressant de noter que les minisatellites riches en GC sont enrichis au niveau de ces extrémités chromosomiques. Sont-ils à même d'induire les réarrangements fréquents à leur niveau? Présentent-ils une fragilité particulière? Si oui, quelles caractéristiques de séquences modulent cette fragilité? Pour apporter des éléments de réponse à ces questions, et tenter de généraliser à un autre phénotype d'instabilité les observations faites sur les expansion/contractions, j'ai introduit les minisatellites subtélomériques humains riches en GC CEB1, CEB25, et HRAS dans une nouvelle construction chromosomique chez *S. cerevisiae*.





**Stimulation of gross chromosomal rearrangements by the human HRAS1, CEB1 and CEB25 minisatellites in *S. cerevisiae* depends on G-quadruplexes or Cdc13**

Aurèle Piazza<sup>1</sup>, Alexandre Serero<sup>1</sup>, Jean-Baptiste Boulé<sup>1</sup>, Patricia Legoix-Né<sup>2</sup>, Judith Lopes<sup>1</sup>, Alain Nicolas<sup>1\*</sup>

<sup>1</sup> Recombinaison et Instabilité Génétique, Institut Curie Centre de Recherche, CNRS UMR3244, Université Pierre et Marie Curie, 26 rue d'Ulm, 75248 Paris Cedex 05, France

<sup>2</sup> NGS Platform, Institut Curie, 26 rue d'Ulm, 75248 Paris Cedex 05, France

Keywords: G-quadruplex, telomere healing, genomic instability, minisatellite, yeast, Pif1, Cdc13

\* Corresponding author: [anicolas@curie.fr](mailto:anicolas@curie.fr)

## Abstract

Genomes contain tandem repeats that are at risk of internal rearrangements and a threat to genome integrity. Here, we investigated the behavior of the human subtelomeric minisatellites HRAS1, CEB1, and CEB25 in *S. cerevisiae*. In mitotically growing wild type cells, these GC rich tandem arrays stimulate the rate of gross chromosomal rearrangements (GCR) by 20, 1620, and 276,000-fold, respectively. In the absence of the Pif1 helicase, known to inhibit GCR by telomere addition and to unwind G-quadruplexes, the GCR rate is further increased in the presence of CEB1, by 385-fold compared to the *pif1* $\Delta$  control strain. The behavior of CEB1 is strongly dependent on its capacity to form G-quadruplexes since the treatment of WT cells with the Phen-DC<sub>3</sub> G-quadruplex ligand has a 52-fold stimulating effect while the mutation of the G-quadruplex-forming motif reduced the GCR rate 30-fold in WT, and 100-fold in *pif1* $\Delta$  cells. The GCR events are telomere additions within CEB1. Differently, the extreme stimulation of CEB25 GCR depends on its affinity for Cdc13, which binds the TG-rich ssDNA telomere overhang. This property confers a biased orientation-dependent behavior to CEB25, while CEB1 and HRAS1 increase GCR similarly in either orientation. Furthermore, we analyzed the minisatellites distribution in the human genome and discuss their potential role to trigger subtelomeric rearrangements.

## Author Summary

All genomes contain particular DNA sequences that are prone to break and rearrange. They can be lost or rescued at the expense of sequence variations and complex rearrangements. Using a sensitive yeast model system, we examined the fragility of the HRAS1, CEB1 and CEB25 GC-rich human minisatellites (tandem repetition of motifs from 10 to 100 bp long). We observed that they all stimulate Gross Chromosomal Rearrangements but to very different extents, both in wild type and cells deficient for the Pif1 helicase. Several intrinsic sequence features can account for these discrepancies: the total number of repeats, the ability to form G-quadruplex secondary structures or to bind with high affinity the telomerase cofactor Cdc13. The orientation on the chromosome dictates the type of GCR (telomere addition versus other structural rearrangements) while not affecting the GCR rate in most cases. Being enriched in the subtelomeric regions of the human chromosomes, this class of GC-rich minisatellite has the potential to trigger a large variety of human genome rearrangements.

## Introduction

Some chromosomal regions are more prone to rearrangement than others and thus are the source of genetic diseases and cancer. Among “at risks” sequences, tandem repeats like microsatellites and minisatellites that differ by the length of their repeat unit (1-10 nt and 10-100 nt, respectively) are prone to changes in repeat number (expansion and contraction of the array) [1]. Mechanistically, this instability can be explained by the propensity of the motifs to misalign during template-directed repair of endogenous lesions, occurring stochastically or promoted by the nucleotide sequence themselves, which, for example, can perturb replication. Consistently, their instability is exacerbated by defects of replication proteins (like Rad27 or Pol $\delta$ ) that ubiquitously affect genome integrity [2-7].

Intrinsic features of repeated sequences also play a role in the formation of rearrangements [1]. Microsatellite instability caused by hairpin formation during replication has been well documented [8] but less is known about minisatellite instability. Sequence composition and its ability to interact with endogenous factors and/or to adopt secondary structures can be invoked. Among these are G-quadruplexes. They are four-stranded structures that some G-rich nucleic acids form spontaneously in physiological salt and pH conditions *in vitro* [9]. A growing body of evidence implicates these structures in several biological processes, like directed genome rearrangements [10,11], telomere capping [12,13], and control of gene expression at the transcriptional and post-transcriptional levels [14,15]. Recently, we showed that the GC-rich human minisatellite CEB1 forms G-quadruplexes *in vitro* and demonstrated that Pif1, a conserved 5'-3' helicase, unwinds these G-quadruplexes [16]. In *S. cerevisiae*, Pif1 prevents the formation of G-quadruplex-dependent CEB1 internal rearrangements during leading-strand replication and, consistently, the treatment of WT cells with the potent G-quadruplex binder Phen-DC<sub>3</sub> mimicked the absence of Pif1 [16,17,18].

A different but perhaps related feature of the human GC-rich minisatellites with respect to genome stability is their clustering in the chromosomal subtelomeric regions ([19,20] that are subjected to pathological terminal truncations [21-24]. The genomic factors involved in the highly dynamic behavior of terminal regions being poorly identified, here we examined the fragility of the subtelomeric human minisatellites HRAS1 [25], CEB1 [26] and CEB25 [27] and the role of their specific sequence features in the induction of Gross Chromosomal Rearrangements (GCR) in *S. cerevisiae*. To this end, we employed the GCR assay developed by R. Kolodner and colleagues [28] that measures the rate of the yeast chromosome V terminal deletion. We showed that the three minisatellites and sequence variants stimulated the formation of GCR in WT cells to different extents depending on several factors: the number of motifs in the tandem array, the ability to form G-quadruplexes, the presence of Cdc13 binding sites, their orientation which yields different type of

rearrangements, and/or the activity of Pif1 and of the homologous recombination pathway. Altogether, these results point to GC-rich minisatellites as major at-risk regions of the genome not only for changes in repeat number but also for their propensity to generate structural variants.

## Results

### Experimental system

To study the behavior of human GC-rich minisatellites in the formation of GCR, we employed the genetic assay developed by Chen and Kolodner [28]. In this sensitive assay, the left arm of chromosome V was engineered to measure the rate of the simultaneous loss of the *CAN1* and *URA3* markers located in the terminal non-essential part of the chromosome V. Cells that undergo a GCR event that results in the simultaneous loss of *URA3* and *CAN1* are recovered on media containing canavanine and 5-fluoro-orotic acid (5-FOA). Fluctuation analysis of the number of growing colonies provide a very sensitive GCR assay (see Materials and Methods), ranging over several order of magnitude since in WT cells, the GCR rate is approximately  $10^{-10}$  events per generation [28]. We inserted the minisatellites centromere-proximal to *CAN1* within the non-essential *NPR2* locus, together with the Hygromycin resistance gene (*hphMX*) (Figure 1A). Importantly, the  $\text{HYG}^R$  cassette (1797 bp) has a GC-content of 58%, does not share homology with the yeast genome, and is devoid of potential G-quadruplex-forming sequences or Cdc13 binding sites. Hereafter, to compare strains with similar size inserts, the *hphMX* construct constitutes our “no minisatellite” control strain.

Altogether, we examined three subtelomeric GC-rich human minisatellites: CEB1 [26], CEB25 [19], and the minisatellite located in the promoter of the *HRAS1* gene [25]. They are tandem arrays with motif lengths of 39, 52, and 28 nt, respectively. The sequence of the consensus motif and additional features of these minisatellites are indicated in Table 1. Furthermore, it is known that the CEB1 and CEB25, but not the *HRAS1* motifs can form stable G-quadruplex structures *in vitro* [16,27]. All three minisatellites were inserted in both chromosomal orientations at the same locus. In the orientation “G”, the G-rich strands of CEB1 is on the same strand as the G-rich 3' ssDNA overhang of the chromosome V left-arm telomere (distance is approximately 45 Kb), while in the orientation “C”, the C-rich strand is on the same strand as the G-rich 3' overhang (Figure 1B). All the rates measured throughout this study are reported in Table S3. Hereafter, unless otherwise stated, the inserts we refer to are in the “G” orientation.

### The minisatellites increase the GCR rate in WT cells, to very different extents

First, we examined GCR rates in the absence of minisatellite inserts. Previous studies provided an estimated rate of  $3.5 \cdot 10^{-10}$  events/generation for WT cells [28]. Consistently, the control strain (*npr2::hphMX*) exhibits the same GCR rate as the parental (*NPR2<sup>+</sup>*) RDKY3615 haploid strain ( $4.3 \cdot 10^{-10}$  vs.  $4.2 \cdot 10^{-10}$  events/generation, respectively). Thus, adding approximately 1.8 Kb of a non-repeated GC-rich DNA to the 13 kb region permissive for rearrangements (located between *CAN1* and the first centromere-proximal essential gen, *PCM1*) has no detectable effect. Then, we measured the consequence of the insertion of the CEB1-WT-1.7 allele containing 43 motifs [16], CEB25-WT-0.7 (13 motifs) and HRAS1-0.7 (26 motifs). Compared to the control strain (*hphMX*), these minisatellites strongly but differentially increased the GCR rate in WT cells: 20-fold for HRAS1 ( $8.48 \cdot 10^{-9}$  events/generation), 1,620-fold for CEB1 ( $6.97 \cdot 10^{-7}$  events/generation) and 276,000-fold for CEB25 ( $1.16 \cdot 10^{-4}$  events/generation) (Figure 1C).

### **Pif1 differentially suppress minisatellite-induced GCR formation**

Pif1 is a conserved 5'-3' helicase that suppresses GCR events by telomere healing [29,30] through direct removal of the telomerase from DNA ends [31]. Pif1 is also involved in G-quadruplex unwinding [16]. We constructed *pif1Δ* cells carrying the minisatellites. Consistent with previous findings [29,32], in the “no-insert” and in our control insert strain, the GCR rates are increased approximately 1500-2250-fold ( $6.63 \cdot 10^{-7}$  and  $1.01 \cdot 10^{-6}$  events/generation, respectively) in the *pif1Δ* strain compared to WT. The presence of the minisatellites had various quantitative effects. Compared to the control *pif1Δ* strains, HRAS1, CEB1 and CEB25 stimulated the GCR rate 3.6-fold ( $3.68 \cdot 10^{-6}$  events/generation), 385-fold ( $3.89 \cdot 10^{-4}$  events/generation) and 120-fold ( $1.21 \cdot 10^{-4}$  events/generation), respectively (Figure 1D). If we now compare the WT and the *pif1Δ* cells carrying the same minisatellite, the absence of Pif1 increases the GCR rate of HRAS1 and CEB1 approximately 500- and 558-fold, but has no effect on CEB25. This insensitivity to Pif1 reflects the already high rate of GCR induced by CEB25 in WT cells. The heterogeneous behavior of this set of minisatellites suggests that specific sequence features modulate their propensity to trigger GCR, in both WT and *pif1Δ* cells.

### **The G-quadruplex-forming sequences of CEB1 stimulate the formation of GCR**

The CEB1 motif forms G-quadruplexes that are efficiently unwound by Pif1 *in vitro* [16,18]. To determine the role of the G-quadruplex forming sequences of CEB1 on GCR rate, we first examined the behavior of the CEB1-Gmut-1.7 array which does not form G-quadruplex (Figure 2A)[16]. In the WT strain background, the insertion of CEB1-Gmut-1.7 yields a GCR rate of  $2.06 \cdot 10^{-8}$  events/generation. This is 65-fold higher than in the control strain but 30-fold lower than in the CEB1-WT-1.7 cells carrying the same number of G quadruplex forming motifs (Figure 2B). These

results indicate that the effect of CEB1 on GCR rate is both G-quadruplex-independent and – dependent. Similarly, we examined the behavior of the CEB1-Gmut-1.7 allele in *pif1Δ* cells. The GCR rate was stimulated 6-fold ( $6.32 \times 10^{-6}$  events/generation) compared to the control *pif1Δ* strain, but was 62-fold lower than in the CEB1-WT-1.7 cells (Figure 2C). This level is similar to the GCR rate induction observed with the HRAS1-0.7 minisatellite also devoid of G-quadruplex-forming sequence. We conclude that, in both WT and *pif1Δ* cells, the induction of GCR by CEB1 strongly depends on its potential to form G-quadruplexes.

To confirm the stimulating role of G-quadruplex, we compared the rate of GCR in cells treated or not with the G-quadruplex-stabilizing ligand Phen-DC<sub>3</sub> [33]. The treatment of WT cells bearing CEB1-WT-1.7 with 10 μM Phen-DC<sub>3</sub> yielded a GCR rate of  $3.65 \times 10^{-5}$  events/generation, 52-fold higher than in the untreated cells (Figure 2D). We verified that this induction was not due to a better growth rate of cells having performed a GCR in the presence of the ligand (Figure S1). In contrast, Phen-DC<sub>3</sub> failed to increase the GCR rate in CEB1-Gmut-1.7 cells ( $3.56 \times 10^{-8}$  events/generation) (Figure 2D). We also assayed concentration effects and treatment with Phen-DC<sub>6</sub>, a compound related to Phen-DC<sub>3</sub> [33]. Clearly, the extent of GCR rate induction in WT cells carrying the CEB1-WT-1.7 minisatellite was stimulated by both ligands and is dependent on their concentration (Figure 2E).

Finally, since our previous studies examined G-quadruplex-dependent expansion/contraction of CEB1 in different chromosomal locations [16-18], we determined the frequencies of CEB1 expansion/contraction in this chromosome V location. As previously observed on Chr. III and VIII, the CEB1-WT-1.7 array was rather stable in WT cells (3/192 rearrangements) and become frequently rearranged upon treatment with Phen-DC<sub>3</sub> (39/192, p-value vs. untreated=  $8.8 \times 10^{-10}$ ) or *PIF1* deletion (16/192, p-value vs. WT=  $3.55 \times 10^{-3}$ ) (Table 2). This depends on the presence of the G-quadruplex-forming sequences, since the CEB1-Gmut-1.7 allele remained stable in the above conditions (Table 2).

We conclude that the impairment of the G-quadruplex unwinding capability of the cell, either by adding G-quadruplex-stabilizing ligands in WT cells or by deleting *PIF1*, stimulates the propensity of the G-quadruplex-prone CEB1 minisatellite to undergo a high level of expansion/contraction and to a lesser extent GCRs.

### **CEB1 induces GCR in a size-dependent manner**

Next, we examined the relationship linking the total number of motifs in the CEB1 array and the GCR rate, both in WT and *pif1Δ* cells. We observe that the rate of GCR in WT cells is positively correlated to the number of repeats (p-value=  $2.8 \times 10^{-3}$ , Spearman's correlation test)(Figure 2F), with rates ranging from  $1.1 \times 10^{-8}$  events/generation for the allele of 0.66 kb (17 motifs) to  $1.59 \times 10^{-5}$  events/generation (37,000-fold higher) for the longest allele of 2.7 kb (≈70 motifs). The straight slope

in logarithmic scale suggests that the relationship linking the motif number and the GCR rate is roughly exponential. Similarly, CEB1-Gmut also induces the formation of GCR in a size-dependent manner ( $p\text{-value}=2.8\times 10^{-3}$ ) (Figure 2F), but with a lower slope: an allele of 1.9 kb ( $\approx 49$  motifs) induced a GCR rate only 4-fold higher than a 0.9 kb allele (23 motifs) ( $3.52\times 10^{-8}$  and  $8.64\times 10^{-9}$  events/generation, respectively). In the absence of Pif1, the GCR rates also increased exponentially with the number of CEB1-WT repeats ( $p\text{-values}=3.97\times 10^{-4}$ ) (Figure 2G). Hence, we conclude that the number of repetition of the minisatellite motif is an aggravating factor of the fragility of these sequences, being steeper with the G-quadruplex-forming ones.

### **The GCR induced by CEB1 are mainly CEB1 truncation healed by telomere addition**

To determine the nature of the GCR events induced by CEB1-WT-1.7, we isolated a set of Can/5FOA resistant colonies from independent cultures to avoid sibling events and analyzed their genomic DNA by Southern blot. The DNA was digested with a restriction enzyme cutting in the centromere proximal part of CEB1 and successively visualized with a CEB1 and a telomeric probe on the same blot. In the majority of colonies isolated in the WT strain background (29/31, 94%) it revealed a smeared CEB1 hybridizing band, which co-hybridized with the telomeric probe (Figure 3A). Similar events and proportion were found for the WT strain treated with Phen-DC<sub>3</sub> (18/18), *pif1* $\Delta$  cells (18/19) (Figure 4C), and WT cells carrying the CEB1-Gmut-1.7 array (8/10 events). Thus, these GCR are likely telomere addition (telomere have variable length in the cell population) associated with a variable number of residual CEB1 motifs.

Analysis of the median length of the smeared band allowed us to roughly determining the number of remaining CEB1 motifs. In untreated WT cells, the events were evenly distributed along the 43 CEB1-WT motifs with the median telomere addition at the 25<sup>th</sup> motif (Figure 3B). In contrast, upon treatment of WT cells with Phen-DC<sub>3</sub>, or deletion of *PIF1*, telomere addition sites shifted significantly toward small fragments, with a median of 11 ( $p\text{-value} = 6\times 10^{-4}$ ) and 15 ( $p\text{-value} = 9.1\times 10^{-3}$ ) motifs, respectively (Figure 3B). These results indicate that (i) irrespectively of the nature of the CEB1 array, most GCR events are telomere addition within CEB1, (ii) telomere addition can occur at numerous places within the CEB1 array thus leaving a variable number of CEB1 motifs, and (iii) impairing the ability of cells to unwind G-quadruplexes (Phen-DC<sub>3</sub> and *pif1* $\Delta$ ) is associated with an increased loss of CEB1 motifs.

To gain higher resolution mapping of the telomere healing events within the CEB1 motifs, we sequenced a set of CEB1-telomere junctions using Ion Torrent Next-Generation Sequencing technology after purification of appropriate DNA bands on agarose gel (see Materials and Methods). We identified the CEB1-Tel junctions from 15 untreated and 12 Phen-DC<sub>3</sub>-treated WT cells (Figure 3C and Figure S2, respectively).



Telomere additions occur mainly at regions of the CEB1 motif that exhibit limited homology to the yeast telomeric sequence. Precisely, 10/27 CEB1-telomere junctions lie in the longest sequence of homology between CEB1 and the telomeric sequence (GGGTGG) and 24/27 junctions have at least two nucleotides in common between CEB1 and the telomeric sequence (shown in blue Figure 3C). This result is consistent with previous observations showing that for *de novo* telomere addition to occur, homology to telomeric sequence of 2-bp (TG, GG, and GT dinucleotide) is sufficient and that a longer homology facilitates telomere healing [30,34,35]. The fact that 62% of the telomere additions occur in, or at, the junction with the G-quadruplex-forming sequence of CEB1 (red lines Figure 3C) is consistent with the fact that 60% of the TG, GG, and GT dinucleotide overlaps this sequence. The distribution of the telomere addition within the CEB1 motif is not significantly different in untreated and Phen-DC<sub>3</sub>-treated WT cells (Figure S2). Hence, although the Phen-DC<sub>3</sub> treatment strongly increases the rate of GCR (Figure 2D) and affects the position of the telomere addition in the array (Figure 3B), the position of the CEB1-telomere junction remains unaffected and mainly lies in the G-quadruplex-forming sequence (Figure 3C and Figure S2).

Altogether, these results suggest that the G-quadruplexes present within the CEB1 array in conditions where the capacity of the cell to unwind G-quadruplexes is impaired (upon Phen-DC<sub>3</sub> treatment or *Pif1* deletion) stimulate the formation of GCR associated with a decreased number of CEB1 motifs remaining in the final repair product.

### **Rad51- and Rad52-dependent and –independent telomere additions in CEB1**

Telomere healing may occur by *de novo* telomere addition to a 3' ssDNA extremity, especially in the absence of Pif1 [29,34-36], leaving a specific pattern of telomeric sequences [34]. However, among the 27 junctions we sequenced, we do not notice any obvious addition of a particular pattern of telomeric sequence in CEB1. On the other hand, telomere addition could occur by capture of endogenous telomeric sequences by break-induced replication (BIR) [28,37,38]. We examined the effect of the deletion of the *RAD51* or *RAD52* genes that are required for BIR [37,38] but not for direct telomere addition by telomerase. It causes a 2-fold decrease of the GCR formation in strains bearing CEB1-WT-1.7, with rates of  $2.92 \times 10^{-7}$  and  $3.54 \times 10^{-7}$  events/generation, respectively (Figure 3D). The extent of the decrease is similar (3- to 5-fold) upon Phen-DC<sub>3</sub> treatment, with GCR rates of  $1.13 \times 10^{-5}$  and  $7.12 \times 10^{-6}$  events/generation in the *rad51Δ* and *rad52Δ* mutants, respectively (Figure 3D). Interestingly, the molecular analyses of the nature of the events provided additional information. We found that the drop of the GCR rate in the absence of Rad52 is associated with a specific decrease of GCRs by telomere addition within CEB1 (Figure S3) while the analysis of the CEB1-telomere junction sequences recovered from untreated or Phen-DC<sub>3</sub>-treated WT cells revealed the presence of SNPs around the junction in 4/6 strains (Figure 3E). These SNPs are found either in

the telomeric sequence only, or both in the CEB1 and the telomeric sequence around the junction (Figure 3E). These intriguing observations suggest that in WT cells roughly half of the telomere healing events in CEB1 occur by BIR on an ectopic telomere sharing a region of limited homology with the CEB1 motif [28]. SNPs found at the junction may result from the correction of the heteroduplex formed between CEB1 and the telomeric sequence, and/or by misincorporation of nucleotides in the early BIR steps [39].

### **The structure of the GCR, but not the rate, depends on the orientation of CEB1**

CEB1 strands strongly differ with respect to their GC composition (GC-bias = 76.6%) and the density of TG/GG/GT dinucleotide (bias is 87%) that seeds GCR by telomere healing (Table 1). We examined the behavior of CEB1 placed in the opposite orientation (orientation C) relatively to the distal telomere (Figure 1B). Strikingly, in WT cells, the GCR rates of CEB1-WT-1.7 are similar in either orientation ( $6.97$  and  $7.47 \times 10^{-7}$  events/generation)(Figure 4A) and alike the G-strand, the GCR rates increase according to the total size of the array (Figure S4). Similarly, although occurring at various absolute rates, there is no significant orientation-dependent difference in all the other strains and conditions that we assayed (Figures 4A and 4B, Table S3). Namely, in WT cells carrying the CEB1-WT-1.7 array treated with Phen-DC<sub>3</sub> ( $3.65$  and  $1.66 \times 10^{-5}$  events/generation), CEB1-WT-1.7 in *pif1*Δ cells ( $3.89$  and  $4.6 \times 10^{-4}$  events/generation), CEB1-Gmut-1.7 in WT ( $2.77$  and  $2.07 \times 10^{-8}$  events/generation) and *pif1*Δ cells ( $6.32$  and  $3.05 \times 10^{-6}$  events/generation) nor in cells carrying HRAS1-0.7 in WT ( $8.48 \times 10^{-9}$  and  $1.1 \times 10^{-8}$  events/generation,) and *pif1*Δ cells ( $3.68$  and  $3.2 \times 10^{-6}$  events/generation) (Figures 4A and 4B, Table S3). Hence, both in the WT and *pif1*Δ cells, the GCR rates induced by CEB1-WT-1.7, CEB1-Gmut 1-7, and HRAS1-0.7 are not affected by the minisatellite orientation on the chromosome. However, the pattern of rearrangements in the G and C orientations is very different (Figures 3A and 4C). In WT cells bearing CEB1-WT-1.7 in the orientation C, only 2/18 rearrangements are smears indicative of telomere healing. The DNA of two other colonies migrate at the size expected for an unaltered region (lane 8), thus likely carrying SNPs or small insertion/deletion within the *CAN1* and *URA3* genes. The majority of events (14/18) appear as discrete bands of various sizes, and among them 10 hybridize with both the *hphMX* and the CEB1 probes (Figures 4C). However, whereas the intensity of the signal is not very different from one lane to the other with the *hphMX* probe, intensities vary greatly with the CEB1 probe. It indicates that the amount of remaining CEB1 sequence in the rearranged chromosomes is different from one strain to the other (for example, compare lanes 2 and 6 in Figure 4C). It is interesting to note that in some cases (lane 3) two bands hybridizing both the CEB1 and *hphMX* probes are visible, indicating a more complex rearrangement that copied the CEB1-*hphMX* region at another locus. The Phen-DC<sub>3</sub> treatment increased the rate of GCR in the orientation C but does not change the pattern of rearrangements (Figure 4C). In sum,

contrary to the prominent telomere additions observed in the G orientation, GCR induced by CEB1 in the C orientation are diverse and sometimes complex, as observed among spontaneous GCR events [28].

Furthermore, to address the genetic requirements of these GCR events, we examined the role of the non-homologous end joining (NHEJ) and homologous recombination (HR) pathways. The GCR rate remains unchanged in the *dnf4Δ* mutant (Figure S5A) while we observed a small but significant 4-fold decrease of the GCR rate in the *rad51Δ* and *rad52Δ* mutants (Figure S5A). In the absence of Rad52, the remaining events are telomere additions (8/9 events) (Figure S5B) suggesting that the HR pathway plays a major role in the formation of the structural but not telomere addition events generated by CEB1 in the C-orientation.

### **The GCR rate induced by CEB25 depends on its ability to bind Cdc13**

We next asked what could be the molecular reasons for the high GCR rate induced by CEB25 in orientation G, and the inability of Pif1 to suppress GCR induced by this construct in WT cells (Figures 1C and 1D). Interestingly, we found that contrary to CEB1, the GCR rate induced by CEB25 strongly depends on its orientation: the inversion of CEB25 caused a 516-fold decrease of the GCR rate in WT cells ( $2.24 \times 10^{-7}$  events/generation). In *pif1Δ* cells, the GCR rate of CEB25 in the orientation C was close to the “no insert” control strain ( $2.41 \times 10^{-6}$  vs.  $1.01 \times 10^{-6}$  events/generation). Additionally, GCR are all (11/11) telomere additions within CEB25 in WT cells, only in the orientation G (data not shown). This strong orientation-dependency prompted us to investigate the sequence composition of the CEB25 motif.

CEB25 has a GC content of 58% and exhibits an absolute GC-bias and GT/GG/TG dinucleotide bias (Table 1). Interestingly, it contains several consensus-binding sites for the 3' telomeric overhang binding protein Cdc13 (underlined in Figure 5A) [40,41]. Cdc13, together with Stn1 and Ten1, is a part of the CST complex involved in telomere capping and mutagenic DSB repair by addition of telomeric repeats at a 3' ssDNA end [32,42-44]. This unique feature, compared to CEB1 and HRAS1, led us to suspect that the recruitment of Cdc13 on CEB25 could be responsible for its GCR effect. To test this hypothesis, we conducted both *in vitro* and *in vivo* experiments. *In vitro*, we determined the affinity of the purified Cdc13 for the CEB25 motif upon gel shift assay (Figure 5A). Cdc13 binds with high affinity to the CEB25 motif (CEB25-WT), with a  $K_d = 6.4 \times 10^{-11} \pm 10^{-11}$  M. Mutations of the Cdc13 binding sites present in the CEB25 motif (CEB25-Cdc13mut) resulted in a 44-fold lower affinity for Cdc13 ( $K_d = 2.8 \times 10^{-9} \pm 3 \times 10^{-10}$  M)(Figure 5A).

Then, to address the possibility that the high affinity of Cdc13 for CEB25 is responsible for the high GCR rate induced by this minisatellite only when the G-rich strand is in the same molecule than the

distal telomere (and thus can be directly extended by telomerase), we constructed and introduced in yeast a 1.4 kb CEB25 allele mutated for its Cdc13-binding sites (CEB25-Cdc13mut-1.4, same motif as in Figure 5A) that kept the same GC content and did not change the G-triplets potentially involved in the G-quadruplex formation (see below). Remarkably, in the orientation G, this construct induced a GCR rate of  $3.07 \times 10^{-7}$  events/generation. This is 713-fold higher than in the “no minisatellite” control strain, and 377-fold lower than with CEB25-WT-0.7 in the same orientation (Figure 5B). Contrary to CEB25-WT, the GCR rate was not affected by the inversion of CEB25-Cdc13mut-1.4 ( $2.95 \times 10^{-7}$  events/generation), indicating that the strong orientation dependency observed with CEB25-WT relies on the presence of the Cdc13-binding sites (Figure 5B). Additionally, in the absence of Pif1, CEB25-Cdc13mut-1.4 also shows a decreased GCR rate compared to CEB25-WT-0.7 in the orientation G (60-fold)(Figure 5C). Again, the GCR rate induced by CEB25-Cdc13mut-1.4 was similar in both the orientations G and C ( $3.86$  and  $2.89 \times 10^{-6}$  events/generation, respectively), and close to the control *pif1* $\Delta$  strain ( $1.01 \times 10^{-6}$  events/generation)(Figure 5C). Hence, the orientation-dependent and Pif1-independent behavior of CEB25-WT is associated with the ability of its motifs to bind the accessory telomerase subunit Cdc13 with high affinity.

#### **CEB25 does not induce G-quadruplex-dependent GCR**

CEB25 contains a consensus G-quadruplex-forming motif (Table 1) that forms a monomorphic G-quadruplex whose structure has been recently solved by NMR [27]. To investigate the potential involvement of G-quadruplexes in the fragility of CEB25, we first examined the GCR rate of CEB25-Cdc13mut-1.4 in the WT and *pif1* $\Delta$  strains (mutations of the Cdc13 binding sites does not change the G-triplets involved in G-quadruplex formation). We found that GCR rates were (i) similar in these strains (Figure 5C), (ii) occurred at a low level comparable to CEB1-Gmut-1.7 (Figure 4B) and HRAS1 (Figure 1D) and (iii) lower than for CEB1-WT-1.7 (Figure 4B). To investigate the potential role of the CEB25 G-quadruplex forming sequences, we synthesized a CEB25 allele mutated for both the G-tracts and the Cdc13 binding sites (CEB25-Cdc13mut-Gmut-1.4). Clearly, the Phn-DC<sub>3</sub> treatment of WT cells bearing CEB25-Cdc13mut-1.4 and CEB25-Cdc13mut-Gmut-1.4 alleles in both orientations yielded no increase of the GCR rates (Figure 5D). This did not depend on the absence of intact Cdc13 binding sites since the CEB25-WT-0.7 allele in the orientation C also remained insensitive to Phn-DC<sub>3</sub> (Figure 5D). Rather, the G-quadruplex-forming and the G-mutated versions of CEB25-Cdc13mut exhibited exactly the same rates of GCR in WT cells. This absence of effect of Phn-DC<sub>3</sub> contrasts with the 22- to 52-fold inductions observed with CEB1-WT upon WT cells treatment (Figure 4A). We then combined the deletion of *PIF1* to the Phn-DC<sub>3</sub> treatment, conditions that yielded synergistic destabilization of CEB1 [18]. We observed a weak 5.5-, 2.3- and 4.6-fold induction of the GCR rates upon treatment of cells bearing CEB25-WT-0.7 in the orientation C, and CEB25-Cdc13mut-1.4 in the

orientations G or C, respectively (Figure 5E). No induction was seen upon treatment of cells bearing the G-mutated version of CEB25-Cdc13mut (Figure 5E). These extreme conditions revealed a slight G-quadruplex-dependent GCR induction by CEB25.

### **GC-rich and G-quadruplex-forming minisatellites cluster at chromosome ends in the human genome**

Since the minisatellites studied here induced the formation of GCR, we wished to gain more insights into the GC-rich minisatellite representation and localization in the human genome. Using Tandem Repeat Finder [45], we determined a list of 353,460 minisatellites (Table S4). They are not evenly distributed along chromosome arms (Figure 6A)[20], being enriched in the 10 and 5% terminal arm regions (Figure 6B). Interestingly, it seems to relate to their GC-content since the 85,222 minisatellites (24%) that have a GC-content higher than 50% preferentially localize at the most terminal parts of the chromosome, whereas the other minisatellites with a lower GC-content are evenly distributed along the arms (Figures 6A and 6B). A similar bias has been previously reported for chromosome 22 [19]. Then, we examined the minisatellites having potential G-quadruplex-forming sequences. Five percent (18,906) of the minisatellites bear at least one G-quadruplex-forming sequence (see Materials and Methods), and 96% (18,191) of these G-quadruplex-forming minisatellites are GC-rich (Table S5). Among the 504 minisatellites that contain at least 30 G-quadruplex-forming sequences due to their tandem repeated structure, 60% (313/504) lie within the terminal 10% of chromosome arms, among which 80% (253/313) lie within the terminal 5%, while keeping a constant GC-content (Figure S6). Hence, GC-rich and G-quadruplex-forming minisatellites appear to preferentially cluster towards the chromosomal extremities (Figure 6C). The mutagenic behavior of HRAS1, CEB1 and CEB25 arrays in yeast described here raises the possibility that the human GC-rich minisatellites play a role in GCRs of the terminal part of human chromosomes.

### **Discussion**

In this study, we assayed the fragility of three GC-rich human minisatellites and mutant derivatives in *S. cerevisiae*. All these minisatellites stimulated the formation of GCRs but at rates varying by several orders of magnitude. We found that the rate depends on several intrinsic sequence features: the total number of repeats, the ability or not to form G-quadruplex secondary structures (case of CEB1) or to bind with high affinity the telomere ssDNA binding protein Cdc13 (case of CEB25). These features also explain their different levels of responsiveness to the Pif1 helicase controlling telomere elongation and G quadruplex unwinding. CEB1 and CEB25 are also differentially responsive to their orientation on the chromosome; it drastically affects the GCR rate of CEB25 but not HRAS1 or CEB1,

and in all cases dictates the type of GCR (telomere addition versus other structural rearrangements). Thus, the behavior of these minisatellites is largely specific. We uncovered here their sequence features.

### **Roles of Pif1 and Cdc13**

Spontaneous GCR in WT cells occurs at a very low rate ( $10^{-10}$ ). It yields a variety of rearrangements that delete the non-essential distal chromosomal region and rescue the chromosome by telomere addition at breaks that contain limited homology to telomere-like seed sequences as well as through more complex genome rearrangements [28]. Two factors may increase the rate of GCR: an increased number of initiating lesions or defects in the repair pathways [28,29]. Regarding the later possibility, as previously reported, we observed that Pif1 plays an important role in suppressing the formation of GCR by telomere healing [29,30,32,46]. In all but one of our minisatellite insertions, GCR rates were increased by several orders of magnitude upon *PIF1* deletion. However, in sharp contrast, the extreme GCR rate stimulated by CEB25 in WT cells remained roughly the same in *pif1Δ* cells. This insensitivity to Pif1 depends on the orientation of CEB25 relative to the distal telomere (G-strand in the same orientation as the single-stranded telomere G-overhang is the most active) in agreement with the ability of the motif to bind the endogenous Cdc13 yeast protein with high affinity (Figures 5A and 5B). Clearly, the mutation of the three Cdc13-binding sites yields a  $\approx 380$ -fold reduction in GCR, consequently abolishing the CEB25 orientation-dependent behavior. The simplest interpretation of these results is that the recruitment of Cdc13 to CEB25 is sufficient to overcome the suppressive effect exerted by Pif1 to prevent the recruitment of the telomerase [43]. This is consistent with the Pif1-independent *de novo* telomere addition at a long internal telomeric tract  $(TG)_{81}$  introduced near an unrepairable HO break [32]. In our assay, due to its motif sequence and its organization in tandem repeats, the human CEB25 minisatellite fortuitously resembles a pseudo-telomere.

### **Role of G-quadruplexes**

On the other hand, the HRAS1, CEB1, CEB1-Gmut, CEB25-Cdc13mut and CEB25-Cdc13mut-Gmut tandem arrays devoid of Cdc13 binding sites also induce GCR but at various rates and in an orientation-independent manner. Among the parameters potentially involved in the fragility of CEB1, its ability to form G-quadruplexes appeared as an important destabilizing feature. Compared to the CEB1-Gmut-1.7 construct, the G-quadruplex-prone CEB1-WT-1.7 allele stimulates the GCR rate in WT cells 30-fold (Figure 2B) and accordingly the conditions that shift the equilibrium toward the folded G-quadruplex state increase the GCR rate : 52-fold upon treatment with the G-quadruplex stabilizing ligand Phen-DC<sub>3</sub> and 558-fold in the absence of the G-quadruplex unwinding helicase Pif1 (Figures 2B,

2C, 3D and 3E). However, it should be emphasized that a predictive G-quadruplex-dependent phenotype cannot be safely ascertained from the presence of a consensus G-quadruplex motif in a given sequence, nor its ability to form stable G-quadruplexes *in vitro*. Indeed, contrary to CEB1, the CEB25 array did not respond *in vivo* to the three conditions that affect G-quadruplex-dependent events (G quadruplex motif mutation, treatment with PhenDC<sub>3</sub> or Pif1 deletion) except slightly, when combining the Phen-DC<sub>3</sub> treatment to the *PIF1* deletion (Figure 5E). This synergistic combination previously observed for CEB1 [18] appears as an extreme hypersensitive condition that may lead to the rare accumulation of unprocessed CEB25 G-quadruplexes. The distinct behavior of CEB1 and CEB25 may rely on different conformations of their respective G-quadruplexes affecting their folding and/or their processing *in vivo*.

### **Role of the repeats number**

Besides sequence affinity to Cdc13 and potential to form G-quadruplexes, a third aggravating factor stimulating the GCR rate is the total number of motifs. Thanks to the sensitivity of this GCR assay, we found that the GCR rate of CEB1 arrays increased exponentially with the number of motifs without an apparent threshold in both WT and *pif1Δ* cells (Figures 2F and 2G). Interestingly, a similar exponential relationship between the number of motifs and the propensity of the triplex-forming (GAA)<sub>n</sub> repeats [47] to form GCRs [48] and expansions [49] has also been reported in yeast. It suggests the intriguing possibility that the capacity of tandem arrays to form secondary structures is a relevant feature. Along this line, we know that a tandem of two and three CEB25 motifs is able to form a pearl-necklace monomorphic G-quadruplexes structure [27]. If CEB1 is also able to form a pearl-necklace G-quadruplexes structure, the size-dependent exponential increase of the GCR rate may reflect the cooperative behavior between the CEB1 motifs to fold into G-quadruplexes. Mechanistically, we recently reported that the CEB1 G-quadruplex prone array perturbs replication and lead to expansion and contraction events [17]. As we proposed, the blockage of the DNA polymerase(s) at the first G-quadruplex may be sufficient to trigger the accumulation of ssDNA between the replication forks and the polymerase and thus enhance the formation of G-quadruplexes per cell and per molecule in a manner related to the total number of repeats. This situation may be similar to the Pol2 slowdown observed at single G-quadruplex-forming motifs under treatment of Pif1-deficient cells with the replication inhibitor hydroxyurea [50].

### **Other factors**

In addition to the effect of G-quadruplexes, other non B-DNA secondary structures can be the source of sequence fragility [8]. However, we found that the HRAS1-0.7 and CEB25-Cdc13mut-Gmut-1.4 minisatellites, devoid of potential G-quadruplex or other secondary structures, also stimulated the

GCR rate by 20- and 700-fold in WT cells, respectively. In addition, once the G-quadruplex-forming capacity of CEB1 was removed by site-directed mutations, we noted that the CEB1-Gmut-1.7 construct was still able to stimulate GCRs at a substantial level ( $\approx 2 \times 10^{-8}$  events/generation), approximately 60-fold higher than in the control WT strain. Similarly, the structure-free (ATTCT)<sub>n</sub> microsatellite has been reported recently to induce chromosomal fragility in WT yeast cells, which increase with the number of motifs [51]. However, the slope of this length-dependent effect could not be derived from these experiments since only two different allele sizes have been assessed [51]. The analysis of CEB1-Gmut allele of various lengths (23-70 motifs) revealed a length-dependent fragility in WT cells in an almost linear manner (multiplying the number of motifs by two increased the GCR rate by 4), in sharp contrast with the exponential slope observed with CEB1-WT (Figure 2F). This difference suggests that the G-quadruplex-independent fraction of the CEB1 fragility does not involve a cooperative behavior between the motifs. What remaining sequence properties could account for this structure-independent fragility? The GC-richness *per se* can be invoked, since it has been shown to slow down DNA polymerases *in vitro* [52]. In the case of our minisatellites, however, three reasons argue against its essential role. First, with similar size arrays, the GCR induction is not clearly correlated to the GC-richness: HRAS1 (GC=67%) and CEB1-Gmut (72%) both stimulated the GCR rate  $\approx 20$ -fold compared to the no insert strain, but  $\approx 35$ -fold less than CEB25-Cdc13mut (GC=56%). Second, the *hphMX* insert, whose size and GC content is similar to CEB25-Cdc13mut-1.4 ( $\approx 58\%$ ), did not stimulate GCR above the no-insert control strain. And third, the density of TG/GG/GT dinucleotides that can seed telomere addition is similar in the CEB25-Cdc13mut, HRAS and *hphMX* insertions (Table 1). These observations suggest that the GC-richness is not *per se* the determinant triggering GCR. The remaining shared feature of these sequences is their organization in tandem. By itself, it may perturb the normal progression of replication due to the high local concentration of homologous templates or create long range specific chromatin structures that might be processed at the expense of maintaining genome stability.

### **Comparison between minisatellite-induced GCR and internal rearrangement frequencies**

In addition to inducing truncated arrays and motifs by GCR, CEB1 also varies in size by increasing or decreasing the total number of motifs via SDSA and/or template switch without involving the flanking regions [16-18]. These events are extremely frequent, being detected in 8.3 and 20.3% of the cells upon deletion of *PIF1* or Phen-DC<sub>3</sub> treatment, respectively (Table 2). This is 100 - 1000 fold higher than the GCR rates ( $3 \times 10^{-4}$  and  $3.6 \times 10^{-5}$  events/generation, respectively) of the same construct. Thus quantitatively, expansion/contraction is the major outcome of CEB1 instability with the advantage to avoid the formation of potentially detrimental structural rearrangements. This is in agreement with numerous reports that compared internal rearrangements and GCR induced by



different microsatellites [2,53-55]. Mechanistically, since the presence of CEB1 perturbs replication [17], GCR events might result from the rare situations in which the template directed intra-motif interactions failed, allowing break-induced replication on an ectopic telomere sequence [48] or the recruitment of the telomerase to act. Consistent with a role of the homologous recombination pathway, we observed that the deletion of the *RAD51* or *RAD52* genes yield a  $\approx 4$ -fold decrease of the GCR rate (Figure 3D and S4A). This is true in both orientations although the nature of the GCR events is different. The insufficient absolute frequency of GCR events ( $<10^{-4}$ ) prevented us to determine whether or not the variation of the GCR rates were compensated by an increase of the expansion/contraction events that can be detected by Southern blot analyses of individual or small pool of colonies.

### **Role and consequences of the preferential location of the human GC-rich minisatellites in chromosomal subtelomeric regions**

Chromosomal rearrangements are potentially detrimental for cell functions and are the source of genetic diseases and cancer. Remarkably, subtelomeric regions are highly dynamic in primate and altered in approximately a third of the human pathologies involving chromosomal rearrangements [21,23,24,56,57]. However, the factors involved in the high propensity of these regions to break and rearrange have not been identified. The intergenic CEB1 and HRAS1, as well as the intronic CEB25 minisatellites assayed here are located 400 kb – 1.4 Mb away from the telomeres (Table 1), representative of the enrichment for GC-rich and the G-quadruplex-forming minisatellites at chromosome terminal regions in the human genome (Figures 6A and 6B). In yeast, the orientation does not affect the fragility *per se* but the nature of the GCR. Hence, given the high number of GC-rich minisatellites clustering at chromosome ends in the human genome irrespectively of their orientation, these sequences are likely implicated in the generation of the various subtelomeric rearrangements [58,59]. But why these harmful sequences are massively present in the human genome? And what could be the reasons of their terminal clustering? A positively selected function could be to signal defects in replicating G-quadruplex-forming sequences [17,50]. In this regard, the arrangement in tandem of G-quadruplex-forming motifs presents at least two advantages. First, they would act as severe “tandem of problems” for replication machinery as revealed by their exponential size-dependent fragility. Hence, cells with a decreased ability to remove G-quadruplexes will experience replication difficulties preferentially at these G-quadruplex-forming minisatellites rather than at unique sequences present throughout the genome and enriched in proto-oncogenes [17,50,60]. Second, owing to the higher local concentration of homologous template compare to unique sequences, they will preferentially undergo internal rearrangements rather than inducing structural variations. Thus, we envision that GC-rich and G-quadruplex-forming minisatellites help

signaling deficient replication machineries, and their clustering at chromosome ends and repetitive nature overall limit the potential formation of detrimental structural rearrangements.

## Materials and Methods

### Strains

The genotypes of the *Saccharomyces cerevisiae* strains (S288C background) used in this study are reported in Table S1. All strains have been derived from RDKY3615 (WT strains) [28] or RDKY4399 (*pif1Δ* strains) [29] by regular lithium-acetate transformation. Correct insertion of the hphMX cassette with or without minisatellite at *NPR2* (position 804, BamHI site), as well as the minisatellite size, have been verified by Southern blot. The CEB1-WT-1.7 and CEB1-Gmut-1.7 minisatellites have been synthesized previously [16]. Contractions and expansions of these minisatellites have been generated during the insertion procedure at the *NPR2* locus and are thus independent clones. The CEB25-WT-0.7, CEB25-Cdc13mut-1.4, and CEB25-Cdc13mut-Gmut-1.4 minisatellites have been synthesized *in vitro* using PCR-based method as previously described [16]. The HRAS1 minisatellite of 0.7 kb (HRAS1-0.7) has been amplified from human genomic DNA. The motifs of the minisatellites used in this study are presented in Table S2. Deletion of *RAD51*, *RAD52*, and *DNL4* has been performed by transformation of the corresponding KMX cassettes amplified from the EUROSCARF deletants collection [61]. Primer sequences are available upon request.

### Media

Liquid synthetic complete (SC) and solid Yeast-Peptone-Dextrose (YPD) media have been prepared according to standard protocols [62]. Plates containing Canavanine (Sigma-Aldrich) and 5FOA (Euromedex) have been prepared according to standard protocols [63] with minor differences: because *npr2Δ* cells exhibit a decreased resistance to acidic pH (<4.0) [64] we adjusted the pH to 4.5-4.8 (instead of 2.8-3.0) and compensated the decreased penetration of 5FOA at this pH by using it at a slightly higher concentration ( $\approx 1.5X$  instead of  $1X$ ). SC liquid media containing Phen-DC<sub>3</sub> (1, 5, or 10  $\mu M$ ) and Phen-DC<sub>6</sub> (1 or 5  $\mu M$ ) have been prepared as previously described [18].

### Fluctuation analysis

The GCR rate has been determined by fluctuation analysis of 5FOA and canavanine-resistant cells. A *ura+* colony is used to inoculate at least 10 independent cultures at a concentration of  $\approx 10^{2-3}$  cells/mL in 2-50 mL of SC media and grown with shaking at 30°C. When they have reached saturation (2

days), cells are spread on 5FOA/canavanine-containing plates and on YPD plates. A maximum of  $10^8$  cells was spread on 85 mm plates, and  $10^9$  cells on 145 mm plates. The number of cells spread was adjusted in order not to exceed 100 colonies per plate. For G-quadruplex ligands-containing SC media, cells undergo an overnight preculture in SC prior to inoculation with the ligand, and are grown at 30°C up to saturation. For *pif1Δ* cells bearing CEB1-WT, which exhibit an inherently high level of CEB1 internal rearrangements, which can influence the GCR rate (strains ORT6543-1, ORT7153-9, and ORT6592-22), the size of the parental minisatellite is determined by Southern blot from individual colonies plated on YPD. The colonies bearing the parental size of CEB1-WT are directly spread on YPD and 5FOA/Can-containing plates without additional liquid culture. After 4 days at 30°C, the number of 5FOA/Can-resistant colonies ( $r$ ) is counted, as well as the total number of viable cells spread ( $N_t$ ) derived from the number of colonies formed on YPD. The GCR rate ( $M$ ) as well as the upper and lower 95% confidence intervals (95% CI) have been calculated from  $r$  and  $N_t$  with Falcor [65] using the Lea and Coulson method of the median. For each strain and condition, 10 to 45 independent cultures have been performed, in at least two independent experiments. The rates, 95% confidence intervals, and the number of independent cultures performed are listed in the Table S3.

#### **CEB1 instability measurement**

Colonies grown on YPD plates after the 2 days culture in SC media are inoculated in 96-well megaplaque for 24-48 hours. Pools of 4-16 colonies were made right before DNA extraction. DNA was digested with XbaI/EcoNI (leaving 414 bp of flanking sequence) and migrated O/N in a 0.8% agarose-TBE 1X gel at 50V. Digestion products were analyzed by Southern blot using a CEB1 radiolabeled probe. Blots were scanned using a Storm Phosphorimager (Molecular Dynamics) or a Typhoon Phosphorimager (GE Healthcare), and quantified using ImageQuant 5.2 as described in [18].

#### **Analysis of the nature of the GCR**

In order to avoid sibling events, DNA of 5FOA/Canavanine-resistant colonies from separate cultures is extracted, digested using either SacI or XbaI, and migrated in a 0.8% agarose-TBE 1X gel overnight at 50 V. Digestion products were analyzed by Southern blot as described previously using a radio-labeled CEB1, hphMX (from pAG34), or telomeric (from pCT300) probe. The position of the telomere addition is estimated by measuring the size of the center of the smear, and subtracting both 50 bp of flanking sequence plus the mean telomere size (300 bp in WT cells and 400 bp in *pif1Δ* cells [30]).

#### **Sequencing of CEB1-telomere junctions**

DNA of colonies bearing a CEB1-telomere smear identified by Southern blot was digested by XbaI and migrated in a 0.8% agarose-TBE 1X gel overnight at 50 V. After staining of the DNA with BET, the DNA

fragments containing the CEB1-telomere junction were cut and extracted from the gel using the Nucleospin Extract II (Macherey-Nagel) kit. Fragments were quantified, pooled, and precipitated. Samples were prepared for Ion Torrent Personal Genome Machine (PGM™, Applied Biosystems). Sequencing has been performed according to manufacturer instructions on a 314R chip. Reads have been validated and aligned on the S288c genome (R64-1-1, 2011-02-03) and custom CEB1-telomere templates using the in-built Torrent Suite 1.5.1. Reads matching both the CEB1 and the telomeric sequences have been isolated and analyzed manually using Tablet 1.11.11.01 [66] and Microsoft Excel 2007.

### ***In silico* minisatellite analysis**

The list of minisatellites (motif comprised between 10 and 100 bp) and their associated characteristics has been obtained from the Tandem Repeat Database [67](list generated on the 2010-10-31 by the TRF algorithm [45] on the Homo Sapiens hg19 release). Overlapping duplicates of the same repeat due to uncertainties in the algorithm have been eliminated. The human minisatellites are listed in **Table S4**. The number of non-overlapping G-quadruplex-forming sequences per minisatellite have been determined using R software. The custom algorithm searches for 4 runs of 3 Gs in a window of 30 nt, with a minimal loop size of 1 nt, and consequently a maximal loop size of 16 nt [68]. They are listed in the **Table S5**.

### **Cdc13 gel shift assay**

A full length version of CDC13 WT was cloned into a pYES2 vector and expressed as a fusion with a C-terminal tag consisting of a 8 glycine linker, 5 strepII-tags (IBA, Germany) and a HAT-tag (Clontech). Cdc13 overexpression was induced in 2% galactose for 16 hours at 30°C according to the method described by P.M. Burgers [69]. Briefly, after grinding cell pellets in liquid nitrogen, the lysate was clarified from DNA by precipitation in 0.1% polyethyleneimine, and the proteins were precipitated with ammonium sulfate at 60% saturation. After resuspension in 50 mM Tris pH8.0, 300 mM NaCl, 10 % glycerol, the soluble fraction was loaded successively on a streptactin column (IBA, germany) followed by a Talon column (Clontech). Purified protein was dialysed against storage buffer 2X without glycerol, and concentrated and stored at -80°C in 1x storage buffer (25 mM tris-HCl pH 8.0, 250 mM NaCl, 0.5 mM DTT, 50% Glycerol). This procedure yielded homogeneous CDC13 estimated more than 90% pure by coomassie blue staining after protein separation by SDS-PAGE.

Gel shift was carried out by incubating 20 pM of the 52-mer CEB25 WT oligonucleotide or the 52\_mer-Cdc13mut version, end-labeled at the 5' end using  $\gamma$ -ATP and T4 polynucleotide kinase, with indicated amount of CDC13, in the following buffer: 5 mM Tris pH 8.0, 2.5 mM MgCl<sub>2</sub>, 0.1 mM EDTA, 2 mM DTT, 0.1  $\mu$ g/ $\mu$ l BSA (NEB), 50 mM NaCl, 0.2 M LiCl. After incubation at room temperature for

30 minutes, binding reactions were supplemented with 3% Ficoll and run on a 6% native polyacrylamide gel (37.5:1 acrylamide/polyacrylamide ratio), at 4°C and 8V/cm. Gels were dried on DE81 paper and quantified using a Typhoon phosphorimager. Data were fitted to a one-site-specific binding equation ( $Y = B_{max} * X / (K_d + X)$ ) using Prism software (Graphpad), yielding  $R^2$  values for goodness of fit of 0.91 and 0.95 for CEB25-WT and CEB25-Cdc13mut, respectively.

### Statistical analysis

Statistical tests have been performed with R software 2.13.1 [70] or Graphpad Prism 5.0b. The  $\alpha$ -cutoff for statistical significance was set to 0.05. Rearrangement frequencies of CEB1 have been compared using a two-tailed Fisher's exact test. Correlation between the number of CEB1 motifs and the rate of GCR has been determined using the Spearman correlation test. GCR rates, as well as the distributions of the position of telomere addition in the CEB1 array have been compared using a non-parametric test (Mann-Whitney-Wilcoxon, two-tailed). A one-tailed  $\chi^2$  test has been used to determine the enrichment of minisatellites in the 10 and 5 terminal percent of chromosome arms.

### Acknowledgements

We thank Dr. Rodney Rothstein for critical reading of the manuscript and Dr. Gael Millot for advices on statistical analysis.

### References

1. Gemayel R, Vences MD, Legendre M, Verstrepen KJ (2010) Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annu Rev Genet* 44: 445-477.
2. Freudenreich CH, Kantrow SM, Zakian VA (1998) Expansion and length-dependent fragility of CTG repeats in yeast. *Science* 279: 853-856.
3. Kokoska RJ, Stefanovic L, Tran HT, Resnick MA, Gordenin DA, et al. (1998) Destabilization of yeast micro- and minisatellite DNA sequences by mutations affecting a nuclease involved in Okazaki fragment processing (rad27) and DNA polymerase delta (pol3-t). *Mol Cell Biol* 18: 2779-2788.
4. Lopes J, Ribeyre C, Nicolas A (2006) Complex minisatellite rearrangements generated in the total or partial absence of Rad27/hFEN1 activity occur in a single generation and are Rad51 and Rad52 dependent. *Mol Cell Biol* 26: 6675-6689.
5. Tishkoff DX, Filosi N, Gaida GM, Kolodner RD (1997) A novel mutation avoidance mechanism dependent on *S. cerevisiae* RAD27 is distinct from DNA mismatch repair. *Cell* 88: 253-263.
6. Maleki S, Cederberg H, Rannug U (2002) The human minisatellites MS1, MS32, MS205 and CEB1 integrated into the yeast genome exhibit different degrees of mitotic instability but are all stabilised by RAD27. *Curr Genet* 41: 333-341.
7. Schweitzer JK, Livingston DM (1998) Expansions of CAG repeat tracts are frequent in a yeast mutant defective in Okazaki fragment maturation. *Hum Mol Genet* 7: 69-74.
8. Mirkin EV, Mirkin SM (2007) Replication fork stalling at natural impediments. *Microbiol Mol Biol Rev* 71: 13-35.

9. Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res* 34: 5402-5415.
10. Cahoon LA, Seifert HS (2009) An alternative DNA structure is necessary for pilin antigenic variation in *Neisseria gonorrhoeae*. *Science* 325: 764-767.
11. Duquette ML, Pham P, Goodman MF, Maizels N (2005) AID binds to transcription-induced structures in c-MYC that map to regions associated with translocation and hypermutation. *Oncogene* 24: 5791-5798.
12. Paeschke K, Juranek S, Simonsson T, Hempel A, Rhodes D, et al. (2008) Telomerase recruitment by the telomere end binding protein-beta facilitates G-quadruplex DNA unfolding in ciliates. *Nat Struct Mol Biol* 15: 598-604.
13. Paeschke K, Simonsson T, Postberg J, Rhodes D, Lipps HJ (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures *in vivo*. *Nat Struct Mol Biol* 12: 847-854.
14. Bugaut A, Balasubramanian S (2012) 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res*.
15. Maizels N (2006) Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat Struct Mol Biol* 13: 1055-1059.
16. Ribeyre C, Lopes J, Boule JB, Piazza A, Guedin A, et al. (2009) The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences *in vivo*. *PLoS Genet* 5: e1000475.
17. Lopes J, Piazza A, Bermejo R, Kriegsman B, Colosio A, et al. (2011) G-quadruplex-induced instability during leading-strand replication. *EMBO J* 30: 4033-4046.
18. Piazza A, Boule JB, Lopes J, Mingo K, Largy E, et al. (2010) Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in *Saccharomyces cerevisiae*. *Nucl Acids Res* 38: 4337-4348.
19. Vergnaud G, Denoeud F (2000) Minisatellites: mutability and genome architecture. *Genome Res* 10: 899-907.
20. Amarger V, Gauguier D, Yerle M, Apiou F, Pinton P, et al. (1998) Analysis of distribution in the human, pig, and rat genomes points toward a general subtelomeric origin of minisatellite structures. *Genomics* 52: 62-71.
21. Baldwin EL, Lee JY, Blake DM, Bunke BP, Alexander CR, et al. (2008) Enhanced detection of clinically relevant genomic imbalances using a targeted plus whole genome oligonucleotide microarray. *Genet Med* 10: 415-429.
22. Kaminsky EB, Kaul V, Paschall J, Church DM, Bunke B, et al. (2011) An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. *Genet Med* 13: 777-784.
23. Shaffer LG, Kashork CD, Saleki R, Rorem E, Sundin K, et al. (2006) Targeted genomic microarray analysis for identification of chromosome abnormalities in 1500 consecutive clinical cases. *J Pediatr* 149: 98-102.
24. Shao L, Shaw CA, Lu XY, Sahoo T, Bacino CA, et al. (2008) Identification of chromosome abnormalities in subtelomeric regions by microarray analysis: a study of 5,380 cases. *Am J Med Genet A* 146A: 2242-2251.
25. Krontiris TG, Devlin B, Karp DD, Robert NJ, Risch N (1993) An association between the risk of cancer and mutations in the HRAS1 minisatellite locus. *N Engl J Med* 329: 517-523.
26. Buard J, Vergnaud G (1994) Complex recombination events at the hypermutable minisatellite CEB1 (D2S90). *Embo J* 13: 3203-3210.
27. Amrane S, Adrian M, Heddi B, Serero A, Nicolas A, et al. (2012) Formation of pearl-necklace monomorphic G-quadruplexes in the human CEB25 minisatellite. *J Am Chem Soc*.
28. Chen C, Kolodner RD (1999) Gross chromosomal rearrangements in *Saccharomyces cerevisiae* replication and recombination defective mutants. *Nat Genet* 23: 81-85.
29. Myung K, Chen C, Kolodner RD (2001) Multiple pathways cooperate in the suppression of genome instability in *Saccharomyces cerevisiae*. *Nature* 411: 1073-1076.

30. Schulz VP, Zakian VA (1994) The *saccharomyces* PIF1 DNA helicase inhibits telomere elongation and *de novo* telomere formation. *Cell* 76: 145-155.
31. Boulé JB, Vega LR, Zakian VA (2005) The yeast Pif1p helicase removes telomerase from telomeric DNA. *Nature* 438: 57-61.
32. Zhang W, Durocher D (2010) *De novo* telomere formation is suppressed by the Mec1-dependent inhibition of Cdc13 accumulation at DNA breaks. *Genes Dev* 24: 502-515.
33. De Cian A, Delemos E, Mergny JL, Teulade-Fichou MP, Monchaud D (2007) Highly efficient G-quadruplex recognition by bisquinolinium compounds. *J Am Chem Soc* 129: 1856-1857.
34. Kramer KM, Haber JE (1993) New telomeres in yeast are initiated with a highly selected subset of TG1-3 repeats. *Genes Dev* 7: 2345-2356.
35. Putnam CD, Pennaneach V, Kolodner RD (2004) Chromosome healing through terminal deletions generated by *de novo* telomere additions in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 101: 13262-13267.
36. Pennaneach V, Putnam CD, Kolodner RD (2006) Chromosome healing by *de novo* telomere addition in *Saccharomyces cerevisiae*. *Mol Microbiol* 59: 1357-1368.
37. Bosco G, Haber JE (1998) Chromosome break-induced DNA replication leads to nonreciprocal translocations and telomere capture. *Genetics* 150: 1037-1047.
38. Llorente B, Smith CE, Symington LS (2008) Break-induced replication: what is it and what is it for? *Cell Cycle* 7: 859-864.
39. Deem A, Keszthelyi A, Blackgrove T, Vayl A, Coffey B, et al. (2011) Break-induced replication is highly inaccurate. *PLoS Biol* 9: e1000594.
40. Bourns BD, Alexander MK, Smith AM, Zakian VA (1998) Sir proteins, Rif proteins, and Cdc13p bind *Saccharomyces telomeres in vivo*. *Mol Cell Biol* 18: 5600-5608.
41. Hughes TR, Weilbaecher RG, Walterscheid M, Lundblad V (2000) Identification of the single-strand telomeric DNA binding domain of the *Saccharomyces cerevisiae* Cdc13 protein. *Proc Natl Acad Sci U S A* 97: 6457-6462.
42. Puglisi A, Bianchi A, Lemmens L, Damay P, Shore D (2008) Distinct roles for yeast Stn1 in telomere capping and telomerase inhibition. *EMBO J* 27: 2328-2339.
43. Bianchi A, Negrini S, Shore D (2004) Delivery of yeast telomerase to a DNA break depends on the recruitment functions of Cdc13 and Est1. *Mol Cell* 16: 139-146.
44. Dewar JM, Lydall D (2012) Similarities and differences between "uncapped" telomeres and DNA double-strand breaks. *Chromosoma* 121: 117-130.
45. Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27: 573-580.
46. Makovets S, Blackburn EH (2009) DNA damage signalling prevents deleterious telomere addition at DNA breaks. *Nat Cell Biol* 11: 1383-1386.
47. Mariappan SV, Catasti P, Silks LA, 3rd, Bradbury EM, Gupta G (1999) The high-resolution structure of the triplex formed by the GAA/TTC triplet repeat associated with Friedreich's ataxia. *J Mol Biol* 285: 2035-2052.
48. Kim HM, Narayanan V, Mieczkowski PA, Petes TD, Krasilnikova MM, et al. (2008) Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair. *EMBO J* 27: 2896-2906.
49. Shishkin AA, Voineagu I, Matera R, Cherg N, Chernet BT, et al. (2009) Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. *Mol Cell* 35: 82-92.
50. Paeschke K, Capra JK, Zakian VA (2011) DNA replication through G-quadruplex motifs is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase. *Cell* 145: 678-691.
51. Cherg N, Shishkin AA, Schlager LI, Tuck RH, Sloan L, et al. (2011) Expansions, contractions, and fragility of the spinocerebellar ataxia type 10 pentanucleotide repeat in yeast. *Proc Natl Acad Sci U S A* 108: 2843-2848.
52. Manosas M, Spiering MM, Ding F, Croquette V, Benkovic SJ (2012) Collaborative coupling between polymerase and helicase for leading-strand synthesis. *Nucleic Acids Res.*

53. Balakumaran BS, Freudenreich CH, Zakian VA (2000) CGG/CCG repeats exhibit orientation-dependent instability and orientation-independent fragility in *Saccharomyces cerevisiae*. *Hum Mol Genet* 9: 93-100.
54. Callahan JL, Andrews KJ, Zakian VA, Freudenreich CH (2003) Mutations in yeast replication proteins that increase CAG/CTG expansions also increase repeat fragility. *Mol Cell Biol* 23: 7849-7860.
55. Kerrest A, Anand RP, Sundararajan R, Bermejo R, Liberi G, et al. (2009) SRS2 and SGS1 prevent chromosomal breaks and stabilize triplet repeats by restraining recombination. *Nat Struct Mol Biol* 16: 159-167.
56. Linardopoulou EV, Williams EM, Fan Y, Friedman C, Young JM, et al. (2005) Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature* 437: 94-100.
57. Rudd MK, Endicott RM, Friedman C, Walker M, Young JM, et al. (2009) Comparative sequence analysis of primate subtelomeres originating from a chromosome fission event. *Genome Res* 19: 33-41.
58. Bonaglia MC, Giorda R, Beri S, De Agostini C, Novara F, et al. (2011) Molecular mechanisms generating and stabilizing terminal 22q13 deletions in 44 subjects with Phelan/McDermid syndrome. *PLoS Genet* 7: e1002173.
59. Luo Y, Hermetz KE, Jackson JM, Mulle JG, Dodd A, et al. (2011) Diverse mutational mechanisms cause pathogenic subtelomeric rearrangements. *Hum Mol Genet* 20: 3769-3778.
60. Eddy J, Maizels N (2006) Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res* 34: 3887-3896.
61. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, et al. (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 285: 901-906.
62. Treco DA, Lundblad V (2001) Preparation of yeast media. *Curr Protoc Mol Biol* Chapter 13: Unit13 11.
63. Putnam CD, Kolodner RD (2010) Determination of gross chromosomal rearrangement rates. *Cold Spring Harb Protoc* 2010: pdb prot5492.
64. Mira NP, Lourenco AB, Fernandes AR, Becker JD, Sa-Correia I (2009) The *RIM101* pathway has a role in *Saccharomyces cerevisiae* adaptive response and resistance to propionic acid and other weak acids. *FEMS Yeast Res* 9: 202-216.
65. Hall BM, Ma CX, Liang P, Singh KK (2009) Fluctuation analysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics* 25: 1564-1565.
66. Milne I, Bayer M, Cardle L, Shaw P, Stephen G, et al. (2010) Tablet--next generation sequence assembly visualization. *Bioinformatics* 26: 401-402.
67. Gelfand Y, Rodriguez A, Benson G (2007) TRDB--the Tandem Repeats Database. *Nucleic Acids Res* 35: D80-87.
68. Guedin A, Gros J, Alberti P, Mergny JL (2010) How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res* 38: 7858-7868.
69. Burgers PM (1999) Overexpression of multisubunit replication factors in yeast. *Methods* 18: 349-355.
70. Team RDC (2008) R: A language and environment for statistical computing.

## Figure Legends



**Figure 1: GC-rich minisatellites induce the formation of GCR in WT and *pif1Δ* cells.** (A) Schematic representation of the GCR assay on the *S. cerevisiae* chromosome V. *URA3* has been inserted at the *HXT13* locus (*hxt13::URA3*) [28]. The *CAN1* and *URA3* genes confer sensitivity to canavanine (Can) and 5-fluoroorotic acid (5FOA), respectively. Plating cells on media containing both Can and 5FOA allows detecting the simultaneous loss of both markers, occurring upon genomic rearrangements (GCR) in the non-essential region between the first essential gene *PCM1* and *CAN1* ( $\approx 13$  kb, dotted line) [28]. The rate of GCR is determined by fluctuation analysis of 5FOA/Can-resistant colonies occurrence (see Materials and Methods). A cassette containing the *hphMX* gene and the minisatellite of interest is inserted in the closest centromere-proximal position to the *CAN1* and *URA3* genes, at the *NPR2* locus. The example of a telomere addition in the tandem repeat associated with the loss of the distal part of the chromosome V is shown. (B) Schematic representation of the two orientations in which CEB1 has been inserted relatively to the distal telomere. (C) GCR rates in WT strains bearing no insert (RDKY3615), no minisatellite (control *npr2::hphMX* strain, ORT6531), HRAS1-0.7 (ORT7182), CEB1-WT-1.7 (ORT6542-6), or CEB25-WT-0.7 (ORT6558) in the orientation G. The fold increase over the “no minisatellite” control strain is indicated. (D) GCR rates in *pif1Δ* cells bearing either no insert (RDKY4399), no minisatellite (control *npr2::hphMX* strain, ORT6568), HRAS1-0.7 (ORT7122), CEB1-WT-1.7 (ORT6543-1), or CEB25-WT-0.7 (ORT6559-5) in the orientation G. Other legends as in (C).

**Figure 2: Size-dependent minisatellite fragility is aggravated by its ability to form G-quadruplexes, especially in G-quadruplex-stabilizing conditions.** (A) Sequence of the CEB1-WT and CEB1-Gmut motifs used in this study [16,18]. G-tracts potentially involved in the G-quadruplex formation are in bold and underlined. Single nucleotide mutations are depicted in bold. (B) GCR rates in WT cells bearing either CEB1-WT-1.7 (ORT6542-6) or CEB1-Gmut-1.7 (ORT6550-2). The dotted line indicates the GCR rate in the “no minisatellite” WT control strain, in which *hphMX* alone has been inserted at *NPR2* (ORT6531)(Figure 1C). (C) GCR rates in *pif1Δ* cells bearing either CEB1-WT-1.7 (ORT6543-1) or CEB1-Gmut-1.7 (ORT6551-1). The dotted line indicates the GCR rate in the “no minisatellite” *pif1Δ* control strain, in which *hphMX* alone has been inserted at *NPR2* (ORT6568)(Figure 1D). (D) GCR rates in untreated (grey) and Phen-DC<sub>3</sub>-treated (white) WT cells bearing either CEB1-WT-1.7 (ORT6542-6) or CEB1-Gmut-1.7 (ORT6550-2). The increase of the GCR rate upon treatment with Phen-DC<sub>3</sub> is indicated. (E) GCR rates in the WT strain bearing CEB1-WT-1.7 in the orientation G (ORT6542-6) treated with 1, 5, and 10  $\mu$ M Phen-DC<sub>3</sub> (white circles) and 1 and 5  $\mu$ M Phen-DC<sub>6</sub> (with squares). GCR rates in WT (F) and *pif1Δ* (G) cells as a function of the size of CEB1-WT (open) and CEB1-Gmut (black). For each genotype, the point at 0 is the “no minisatellite” control strain (ORT6531 and ORT6568, respectively).

**Figure 3: GCR are mainly telomere additions in CEB1.** (A) The top panel schematically represents the genomic region surrounding CEB1 with the XbaI restriction site and the CEB1 (green) and TG<sub>1-3</sub> (red) probes used to study rearrangements of the region. The size of the unaltered region upon digestion is indicated. Upon telomere addition, digestion with XbaI is expected to produce a fragment that can be hybridized by both the CEB1 and the telomeric (TG<sub>1-3</sub>) probe. The bottom panels show the rearrangements present in independent 5FOA/Can-resistant colonies obtained from WT cells bearing CEB1-WT-1.7 in the orientation G (ORT6542-6). Stars indicate smears hybridizing both the CEB1 and the telomeric probes (Orientation G lanes 1,2,4,5,6,7,8,10). Arrows indicate endogenous telomeres. Ambiguities remained for CEB1 smears co-migrating with endogenous telomeres (lane 9). One lane shows no CEB1 signal (lane 3). (B) Distribution of the telomere addition positions within CEB1-WT-1.7 estimated from the mean molecular weight determined by Southern blot (see Materials and Methods) in untreated and Phen-DC<sub>3</sub>-treated WT cells (ORT6542-6), and *pif1Δ* cells (ORT6543-1). Grey bars show the median of the distributions. Distributions were compared using a non-parametric test (Mann-Whitney-Wilcoxon). (C) Analysis of the CEB1-telomere junctions in WT cells (ORT6542-6). Each CEB1-telomere junctions have been obtained from 15 independent 5FOA/Can-resistant colonies. The sequences are oriented 5'-3'. The template CEB1 motif is shown in green, and the flanking motifs (n-1 and n+1) are shown in grey. The G-quadruplex forming motifs are underlined in red. Nucleotides shared by both the CEB1 and the telomeric sequence at the junction are in blue. The telomeric sequence is in red. The CEB1 sequence above the junctions is identical to the CEB1 reference, and is not shown for each molecule. A schematic representation of the length of shared nucleotides between the CEB1 and the telomeric sequences is shown in blue below the reference CEB1 motif. An arrow indicates the junction in which no nucleotide is shared. (D) GCR rates in WT (ORT6542-6), *rad51Δ* (ORT7189), and *rad52Δ* (ORT7310-2) cells bearing CEB1-WT-1.7 upon treatment (right) or not (left) with Phen-DC<sub>3</sub> at 10 μM. The scale is linear. (E) Analysis of the single nucleotide polymorphisms (SNPs) around the CEB1-telomere junctions covered by at least 3 different reads. SNPs present in 100% of the reads are shown in black and underlined. Other legends as in (C).

**Figure 4: The nature of the GCR, but not the GCR rate, depends on the orientation of CEB1 in WT cells.** (A) GCR rates in untreated (grey) and Phen-DC<sub>3</sub>-treated (white) WT cells bearing CEB1-WT-1.7 in the orientation G (ORT6542-6) and C (ORT6591-1), and CEB1-Gmut-1.7 in the orientation G (ORT6550-2) and C (ORT6548). The fold increases of the GCR rate upon treatment with Phen-DC<sub>3</sub> is indicated. (B) GCR rates in *pif1Δ* cells bearing CEB1-WT-1.7 in the orientation G (ORT6543-1) and C (ORT7153-9), and CEB1-Gmut-1.7 in the orientation G (ORT6551-1) and C (ORT6549). The dotted line indicates the GCR rate in the “no minisatellite” control *pif1Δ* strain (ORT6568) (see Figure 1D). Other legends as in Figure 1D. (C) The top panel schematically represent the genomic region surrounding

CEB1 with the *SacI* restriction site and CEB1 (green) and *hphMX* (blue) probes used to study rearrangements of the region. The size of the unaltered region upon digestion is indicated. The example of a translocation within CEB1 is shown, and is expected to produce a fragment longer than 1.7 kb. The bottom panels show the rearrangements present in 5FOA/Can-resistant colonies obtained from independent cultures of WT cells bearing CEB1-WT-1.7 in the orientation C (ORT6591-1). In most lanes the bands hybridized both the CEB1 and the *hphMX* probes, except in lane 5 in which CEB1 has been lost. (D) Nature of the GCR determined by Southern blot analysis of independent 5FOA/Can-resistant colonies derived of cells bearing CEB1-WT-1.7 in the orientations G or C, in WT cells treated or not with Phen-DC<sub>3</sub> 10 μM, or in *pif1Δ* cells. The results are presented as a percentage of the total number (n) of 5FOA/Can-resistant colonies analyzed. Telomere additions are shown in red, and their locations relative to CEB1 are indicated by different motifs: within CEB1 (no motif), in proximal (cross), or in distal (lines) position to CEB1. Discrete rearrangements are shown in dark grey. We observed no structural alteration of the region (no rearr.) in two colonies (light grey). In some instances, colonies lost both CEB1 and *hphMX* (unknown, white). Distributions were compared using the Fisher's exact test.

**Figure 5: The high rate of GCR induced by CEB25 in an orientation-dependent manner relies on its high affinity Cdc13-binding sites.** (A) Analyses of the Cdc13 affinity for CEB25 motifs. The sequences of the CEB25-WT and CEB25-Cdc13mut oligonucleotides used for the gel shift assay experiment are shown. The Cdc13-binding sites are underlined. Mutations of the Cdc13-binding sites in CEB25-Cdc13mut are shown in red (16 mutations). The quantification of the proportion of the Cdc13-bound oligonucleotides of CEB25-WT (black) and CEB25-Cdc13mut (red) as a function of the Cdc13 concentration is shown. A second order binomial fit has been applied. Absolute  $K_d$  values  $\pm$  SD and fold difference compared to the CEB25-WT motif are indicated. (B) GCR rates in the control “no minisatellite” WT strain (ORT6531), and WT cells bearing CEB25-WT-0.7 in the orientation G (ORT6558) and C (ORT6556), and CEB25-Cdc13mut-1.4 in the orientation G (ANT1181-1) and C (ANT1180-5). Other legends as in Figure 1C. (C) GCR rates in the control “no minisatellite” *pif1Δ* strain (ORT6568), and *pif1Δ* cells bearing CEB25-WT-0.7 in the orientation G (ORT6559-5) and C (ORT6557-1), and CEB25-Cdc13mut-1.4 in the orientation G (ANT1185-4) and C (ANT1184-1). Other legends as in Figure 1D. (D) GCR rates in WT cells bearing CEB25-WT-0.7 in the orientation C, CEB25-Cdc13mut-1.4 in the orientation G and C, or CEB25-Cdc13mut-Gmut-1.4 in the orientation G and C treated with 10 μM Phen-DC<sub>3</sub> (white) or not (grey). (E) GCR rates in *pif1Δ* cells bearing CEB25-WT-0.7 in the orientation C (ORT6557-1), CEB25-Cdc13mut-1.4 in the orientation G (ANT1184-1) and C (ANT1185-4), or CEB25-Cdc13mut-Gmut-1.4 in the orientation G (ANT1187-1) and C (ANT1186-1) treated with 10 μM Phen-DC<sub>3</sub> (white) or not (grey).

**Figure 6: GC-rich and G-quadruplex forming minisatellites are enriched at chromosome ends in the human genome.** (A) Distribution of the human minisatellites along a normalized chromosome arm in the human genome. The centromere (CEN) is at 0, and the telomere (TEL) at 1. (B) Proportion of the GC-rich (GC content > 50%), GC-poor (GC content ≤ 50%), and all the minisatellites in the terminal 10 and 5% on each chromosome arms. n indicates the number of minisatellites. A star (\*) indicates a significant enrichment (all p-values < 2e<sup>-16</sup>). No enrichment is observed for GC-poor minisatellites. (C) Proportion of minisatellites in the terminal 10% (continuous line) and 5% (dotted line) of chromosome arms as a function of the number of non-overlapping G-quadruplex-forming sequences per minisatellite.

**Table 1: Human minisatellites used in this study and their sequence composition properties.**

Minisatellite	Location	Distance to Telomere	Motif size	% GC	GC-bias	TG/GG/ GT-bias	G-rich strand motif sequence (5'-3')
CEB1	2q37.3	0.4 Mb	39 bp	77%	77%	87%	<b>GGGGGGAGGGAGGG</b> TGGCCTGCGGAGGT CCCT <b>GGG</b> CTGA
CEB25	10q26.3	1.4 Mb	52 bp	56%	100%	100%	A <b>AGGGTGGGT</b> GTAAAGTGT <b>GGTGGGT</b> GTGAGTGT <b>GGG</b> TGTGGAGGTAGATGT
HRAS1	11p15.5	0.5 Mb	28 bp	68%	74%	80%	CCCCTGGAGAGA <b>AGGGGG</b> AGTGTGGCGT

Guanines potentially involved in the formation of G-quadruplexes are shown in bold.

**Table 2: Internal rearrangement frequencies of CEB1 in WT cells treated or not with 10 μM Phen-DC<sub>3</sub>, and in *pif1Δ* cells.**

Minisatellite	Genotype	Untreated	Phen-DC <sub>3</sub> 10 μM
CEB1-WT-1.7	WT	3/192 (1.6%)	39/192 (20.3%)*
	<i>pif1Δ</i>	16/192 (8.3%)°	ND
CEB1-Gmut-1.7	WT	1/192 (0.5%)	0/96
	<i>pif1Δ</i>	0/192	ND

\*p-value vs. Control < 0.05

° p-value vs. WT < 0.05

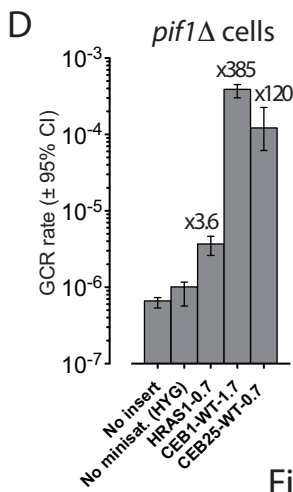
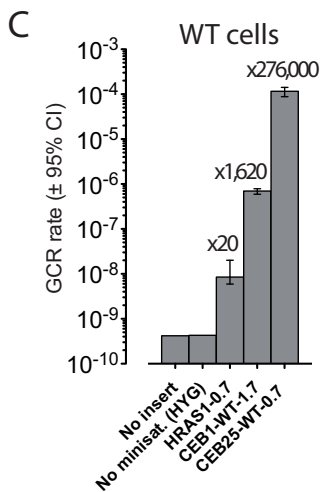
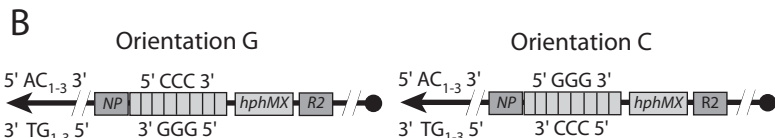
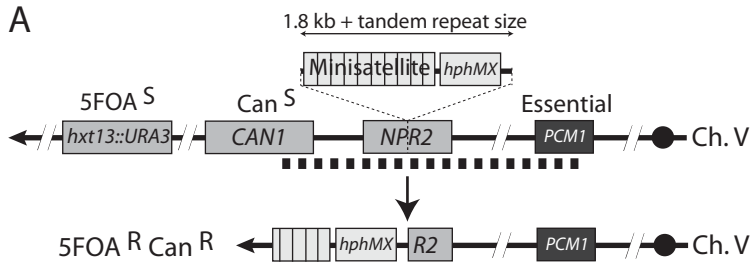


Figure 1

**A** CEB1-WT : GGGGGGAGGGAGGGGGGTGGCCCTGCGGAGGTCCTCTGGGCTGA  
 CEB1-Gmut : GCGCGGAGTGAGAAGTGGCCCTGCGGAGGTCCTCTGCCTGA

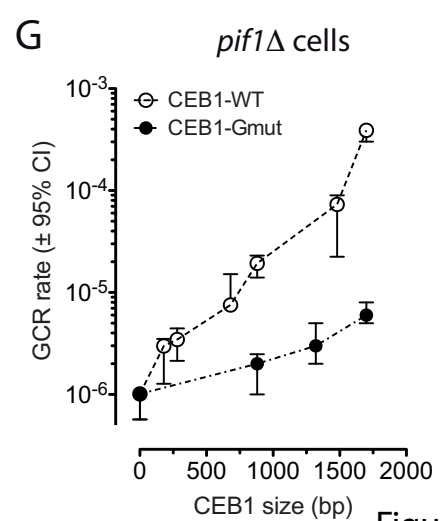
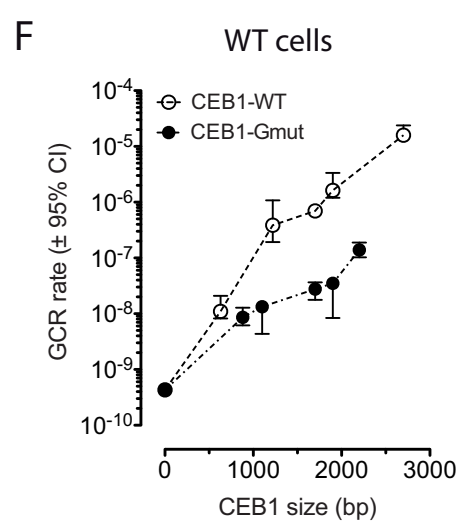
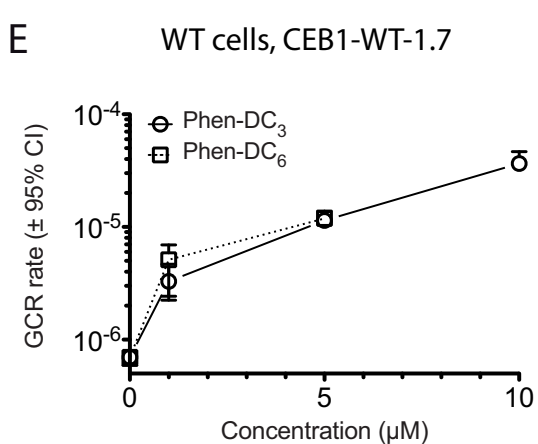
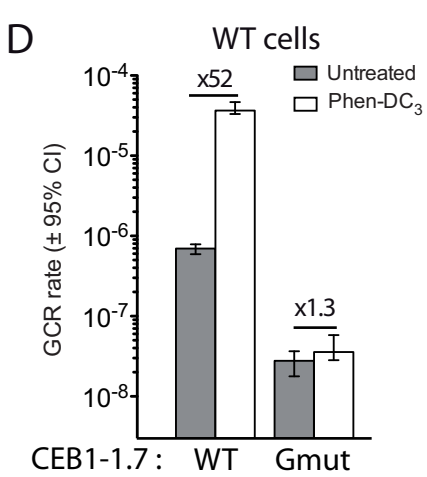
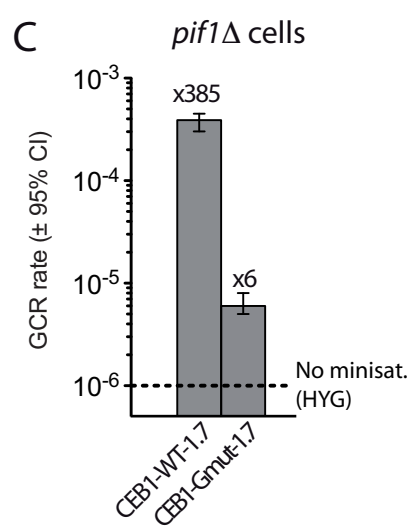
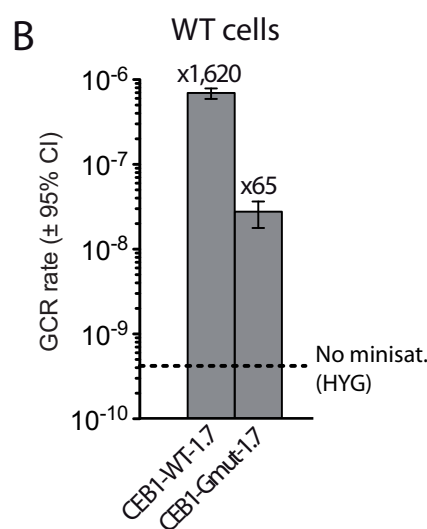


Figure 2



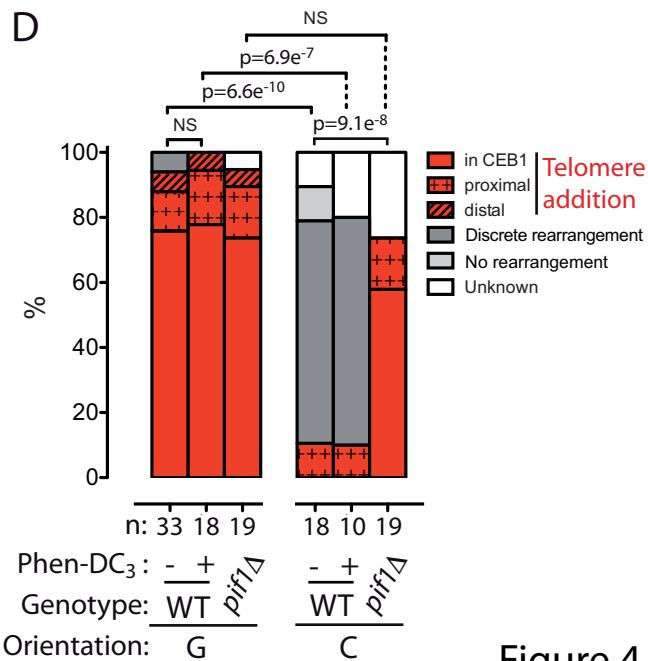
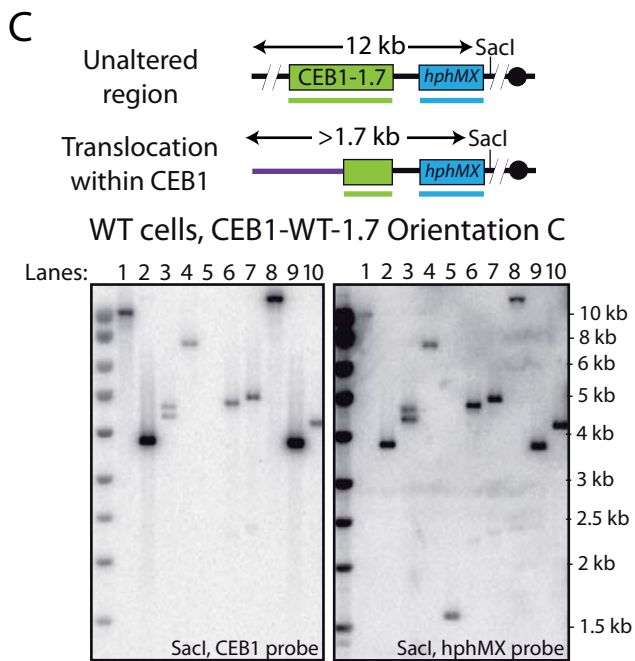
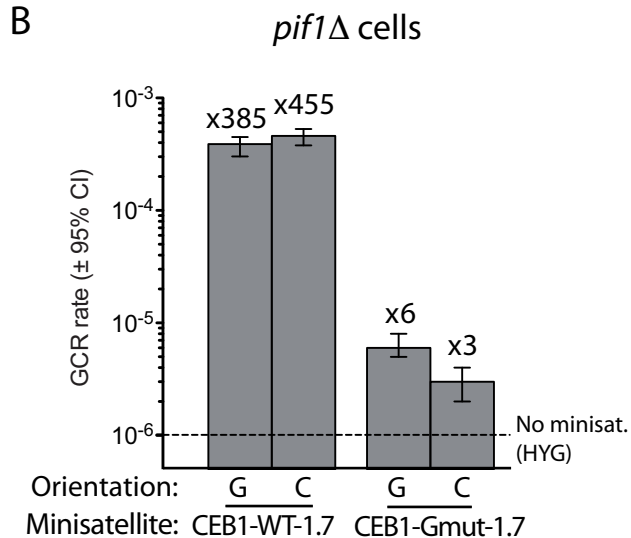
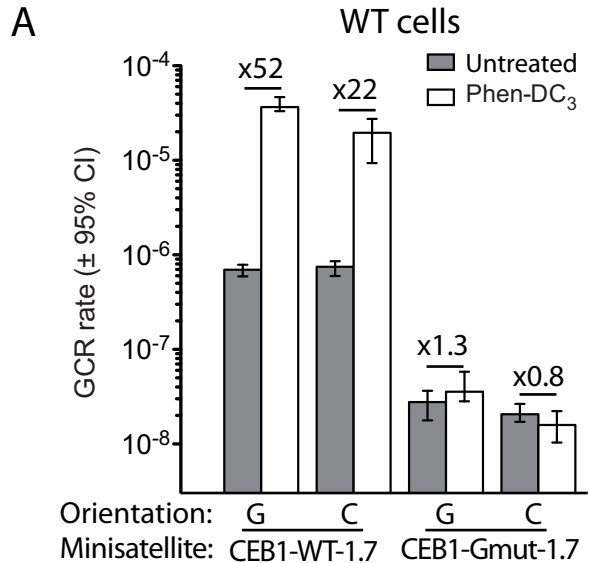


Figure 4



**A** CEB25 WT TAAGGGTGGGTGTAAGTGTGGGTGGGTGTGAGTGTGGGTGTGGAGGTAGATG  
 CEB25-Cdc13mut TCAGGGCGGGTCTAAGTTAGGGCGGGATTTAGTCAAGGGACTCGAGGATGATG

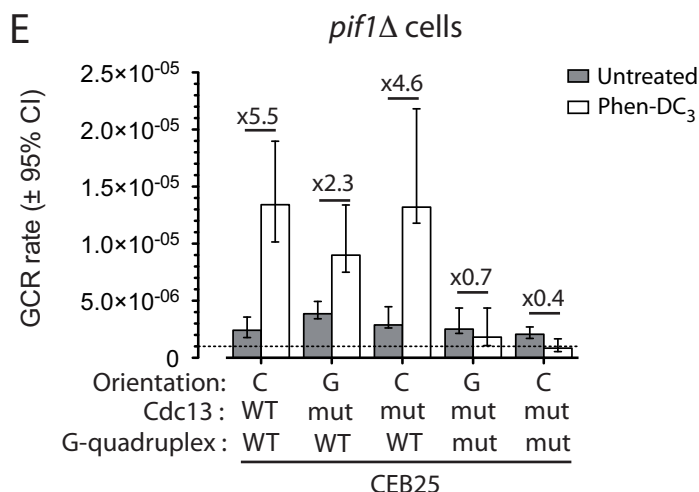
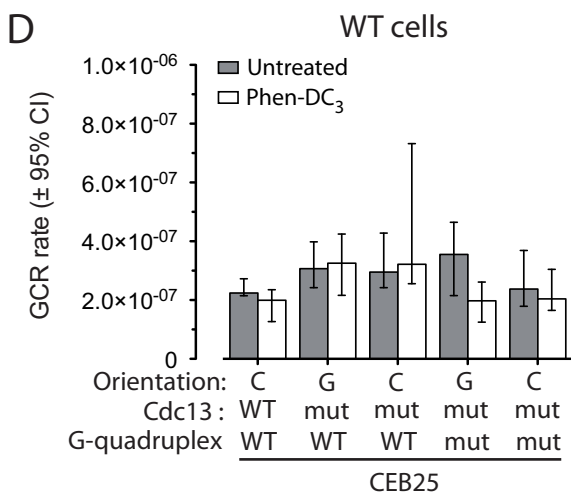
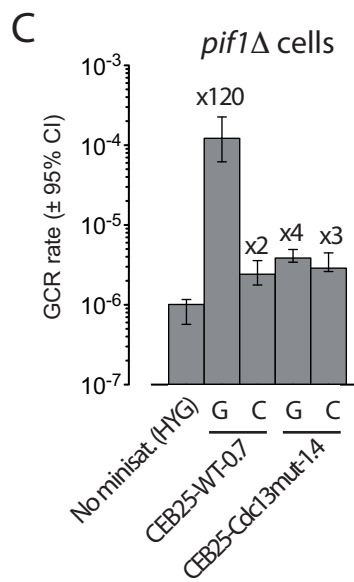
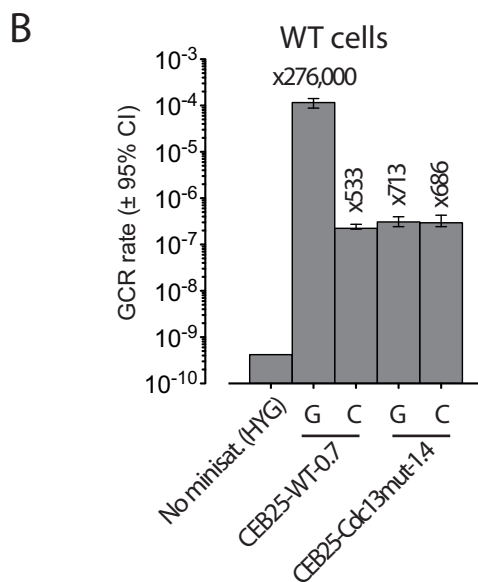
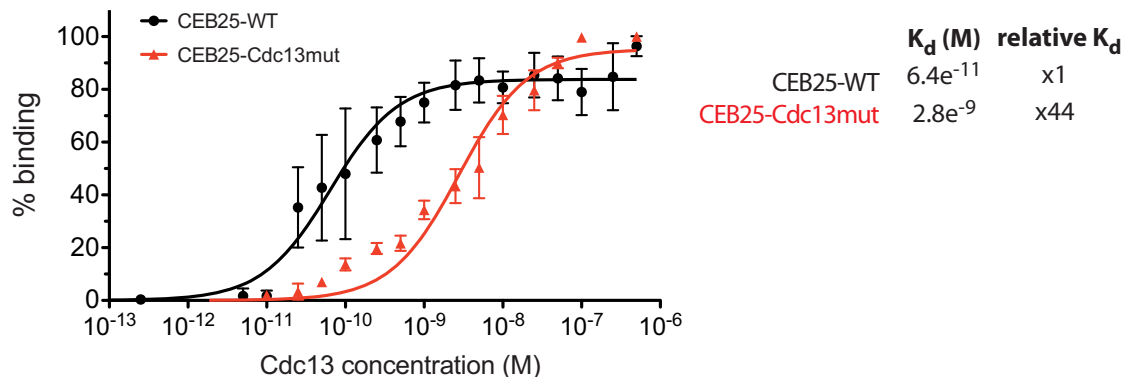
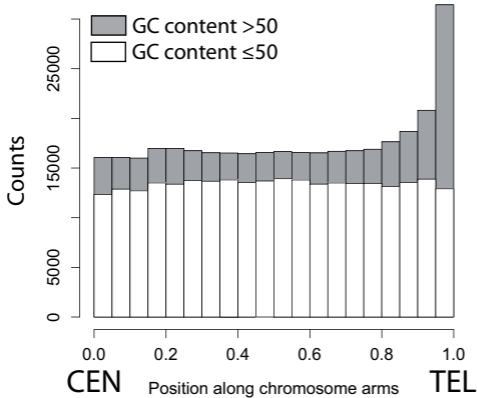
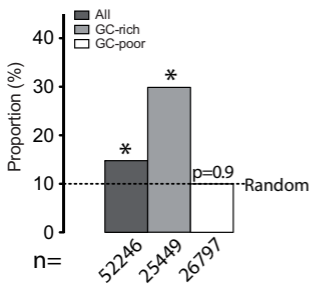


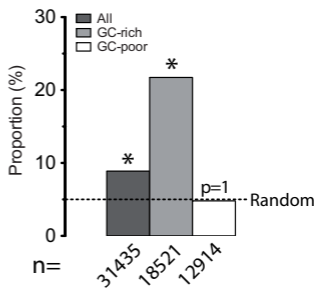
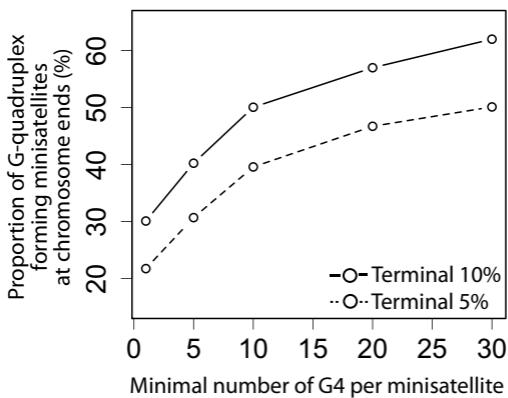
Figure 5

**A****B**

Terminal 10%



Terminal 5%

**C****Figure 6**

## Supplemental data

### Supplementary Figures and Tables Legends

**Figure S1:** The increased GCR rate in WT cells bearing CEB1-WT upon treatment with Phen-DC<sub>3</sub> can not be explained by a better growth rate of cells having performed a GCR in presence of Phen-DC<sub>3</sub> 10 μM. Growth of WT cells bearing CEB1-WT-1.7 in the orientation G (WT, ORT6542-6) and a derived strain bearing a CEB1 minisatellite fused to a telomere (GCR) have been monitored by measurement of the optical density at 600 nm (OD600) as described in [1]. The generation time has been measured during the exponential growth phase, between the 4 and 8h time points. Error bars show SD (n=3).

**Figure S2:** Analysis of the telomere addition site in the CEB1-WT motif in Phen-DC<sub>3</sub>-treated WT cells (ORT6542-6). Each CEB1-telomere junctions have been obtained from 12 5FOA/Can-resistant colonies from independent cultures. Other legends as in Figure 3C.

**Figure S3:** A subset of the telomere additions in CEB1 depends on Rad52. Proportions of the different types of GCR characterized by Southern blot analysis in WT (ORT6542-6) and *rad52Δ* (ORT7310-2) cells bearing CEB1-WT-1.7 in the orientation G. Other legend as in Figure 4D.

**Figure S4:** Size effect of CEB1-WT in wild-type and *pif1Δ* cells, in orientation G or C.

**Figure S5:** GCR are mainly Rad51- and Rad52-dependent translocations in WT cells bearing CEB1-WT-1.7 in the orientation C. (A) GCR rates in WT (ORT6591-1), *dnI4Δ* (ORT7309-3), *rad51Δ* (ORT7191-3), and *rad52Δ* (ORT7312-5) cells bearing CEB1-WT-1.7 in the orientation C upon treatment (right) or not (left) with 10 μM of Phen-DC<sub>3</sub>. Other legends as in Figure 1C. (B) Proportions of the different types of GCR characterized by Southern blot analysis in WT (ORT6591-1) and *rad52Δ* (ORT7312-3) cells bearing CEB1-WT-1.7 in the orientation C. Other legends as in Figure 4D.

**Figure S6:** The number of G-quadruplex forming sequence per minisatellite does not affect The GC-content of G-quadruplex-forming minisatellites. Error bars show SD.

**Table S1:** Strains used in this study.

**Table S2:** Minisatellites used in this study.

**Table S3:** GCR rates measured in the untreated cells (untreated sheet) and in cells treated with G-quadruplex ligands (Phen-DC sheet). The upper and lower 95% confidence intervals, as well as the number of independent cultures performed (n) are indicated. The number after the minisatellite name (CEB1-WT-) indicates its size in kb. The number of motifs has been determined by sequencing

of the array for CEB1-Gmut-1.7 [2], HRAS1-0.7, CEB25-WT-0.7, CEB25-Cdc13mut-1.4 and CEB25-Cdc13mut-Gmut-1.4, and estimated based on their size for the various CEB1-WT and CEB1-Gmut alleles.

**Table S4:** Human minisatellites (motif comprised between 10 and 100 nt) determined using Tandem Repeat Finder (see **Material and Methods**) [3,4]. Information concerning the position, motif size, copy number of the motif in the reference hg19 release, GC bias between the strands, total array length, and GC content of each minisatellite is provided.

**Table S5:** G-quadruplex forming human minisatellites. In addition to the basic information provided in Table S4, this table give the number of non-overlapping potential G-quadruplex-forming sequences for each minisatellite when searching for 4 runs of 3 Gs in a maximum of 30, 40, 50, or 100 bp sliding window ( $G_4 \leq 30$ ,  $G_4 \leq 40$ ,  $G_4 \leq 50$ ,  $G_4 \leq 100$ , respectively)(see Materials and Methods). The number of G-quadruplex-forming sequence per motif is also provided for the search performed with each window (last 4 lanes). The column "G4strand" indicates if the G-quadruplex-forming sequence was present in the Watson strand (G) or the Crick strand (C). Analyses presented in the Results section and in Figure 6 have been performed using data obtained with the most stringent search parameter (window  $\leq 30$  bp).

## Supplementary References

1. Piazza A, Boule JB, Lopes J, Mingo K, Largy E, et al. (2010) Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in *Saccharomyces cerevisiae*. *Nucl Acids Res* 38: 4337-4348.
2. Ribeyre C, Lopes J, Boule JB, Piazza A, Guedin A, et al. (2009) The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences *in vivo*. *PLoS Genet* 5: e1000475.
3. Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27: 573-580.
4. Gelfand Y, Rodriguez A, Benson G (2007) TRDB--the Tandem Repeats Database. *Nucleic Acids Res* 35: D80-87.

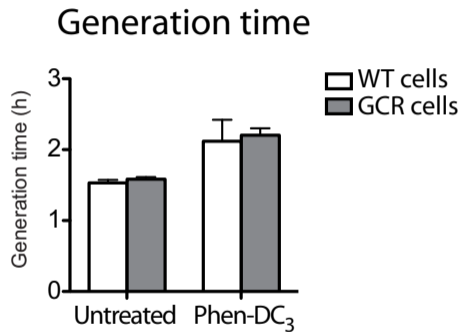
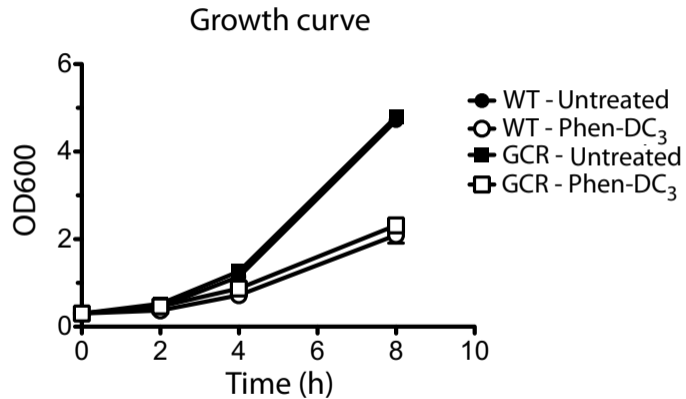


Figure S1

# WT cells + Phen-DC<sub>3</sub>

- #1 GGGTGTGGGTGTCGGGT
- #2 GGGTGAGTGGGTGGTGGTTCGTGGT
- #3 GGGTGGTGGGTGCGTG
- #4 TCGGGATGGGTGTG
- #5 GGTGGGGTGGTGCCTGGTGGGTG
- #6 GGGTGTGGGTGTGTGGGTGTGG
- #7 GGGTGTGGGTGTGTGGGTGTGG
- #8 GGGTGTGGGTGTGGGTGTGTGGGT
- #9 GGGTGGGTGTGGGTGTGGGTGTGGGTGT
- #10 GGTGTGTGGGTGTGGGTGTGG
- #11 GGTGTGGGTGTGGGTGTGGGT
- #12 TGGGTGTGGGTGTGTGGGT



Figure S2

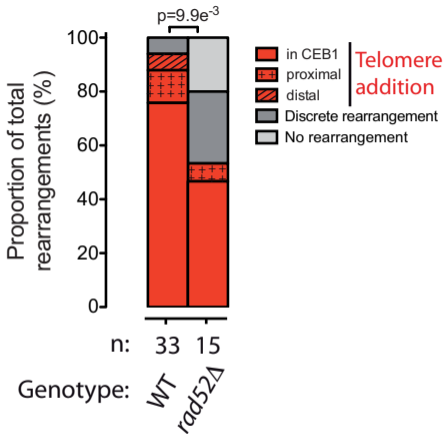


Figure S3

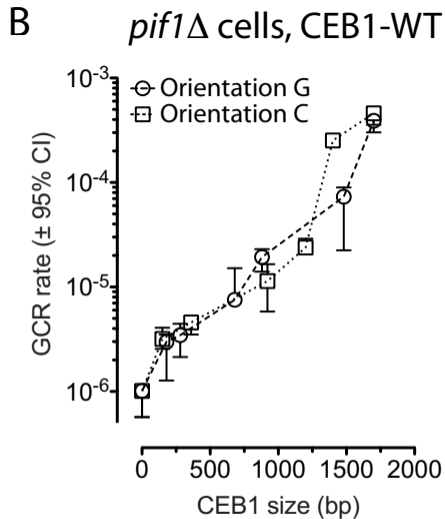
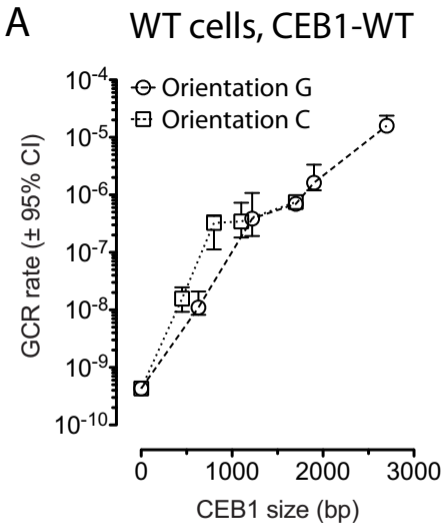


Figure S4



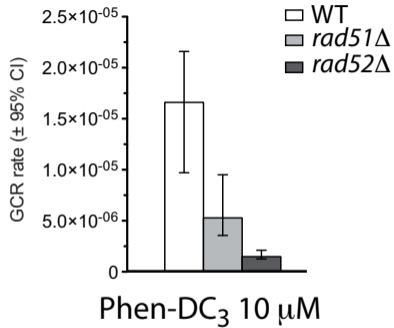
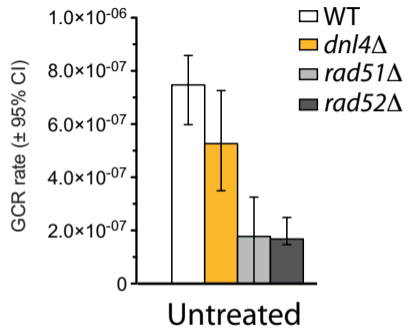
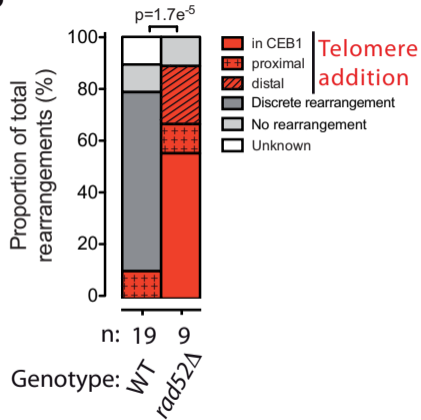
**A****B**

Figure S5

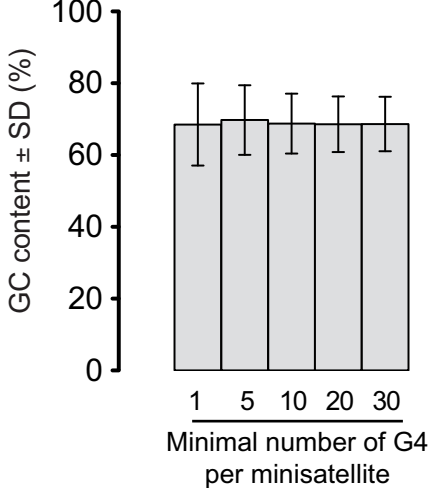


Figure S6

**Table S1**

<b>Strain</b>	<b>Relevant genotype</b>	<b>Source</b>
RDKY3615	MAT a, <i>ura3-52, his3Δ200, leu2Δ1, trp1Δ63, lys2ΔBgl, hom3-10, ade2Δ1, ade8, hxt13::URA3</i>	ref. 28
RDKY4399	<i>pif1::KMX</i>	ref. 29
ORT6531	<i>npr2::hphMX</i>	This study
ORT6568	<i>pif1::KMX, npr2::hphMX</i>	This study
ORT6542-8	<i>npr2::hphMX-CEB1-WT-0.8 orientation G</i>	This study
ORT6542-5	<i>npr2::hphMX-CEB1-WT-1.4 orientation G</i>	This study
ORT6542-6	<i>npr2::hphMX-CEB1-WT-1.7 orientation G</i>	This study
ORT6542-4	<i>npr2::hphMX-CEB1-WT-1.9 orientation G</i>	This study
ORT6542-1	<i>npr2::hphMX-CEB1-WT-2.7 orientation G</i>	This study
ORT6591-9	<i>npr2::hphMX-CEB1-WT-0.2 orientation C</i>	This study
ORT6591-4	<i>npr2::hphMX-CEB1-WT-0.45 orientation C</i>	This study
ORT6591-3	<i>npr2::hphMX-CEB1-WT-0.8 orientation C</i>	This study
ORT6591-8	<i>npr2::hphMX-CEB1-WT-1.1 orientation C</i>	This study
ORT6591-1	<i>npr2::hphMX-CEB1-WT-1.7 orientation C</i>	This study
ORT6543-10	<i>pif1::KMX, npr2::hphMX-CEB1-WT-0.16 orientation G</i>	This study
ORT7154-2	<i>pif1::KMX, npr2::hphMX-CEB1-WT-0.28 orientation G</i>	This study
ORT7154-7	<i>pif1::KMX, npr2::hphMX-CEB1-WT-0.76 orientation G</i>	This study
ORT7154-3	<i>pif1::KMX, npr2::hphMX-CEB1-WT-0.88 orientation G</i>	This study
ORT6543-4	<i>pif1::KMX, npr2::hphMX-CEB1-WT-1.4 orientation G</i>	This study
ORT6543-1	<i>pif1::KMX, npr2::hphMX-CEB1-WT-1.7 orientation G</i>	This study
ORT6541-1	<i>pif1::KMX, npr2::hphMX-CEB1-WT-0.16 orientation C</i>	This study
ORT6592-1	<i>pif1::KMX, npr2::hphMX-CEB1-WT-0.4 orientation C</i>	This study
ORT6592-4	<i>pif1::KMX, npr2::hphMX-CEB1-WT-0.96 orientation C</i>	This study
ORT6592-18	<i>pif1::KMX, npr2::hphMX-CEB1-WT-1.2 orientation C</i>	This study
ORT6592-22	<i>pif1::KMX, npr2::hphMX-CEB1-WT-1.4 orientation C</i>	This study
ORT7153-9	<i>pif1::KMX, npr2::hphMX-CEB1-WT-1.7 orientation C</i>	This study
ORT7189-1	<i>rad51::KMX, npr2::hphMX-CEB1-WT-1.7 orientation G</i>	This study
ORT7191-3	<i>rad51::KMX, npr2::hphMX-CEB1-WT-1.7 orientation C</i>	This study
ORT7309-3	<i>dnl4::KMX, npr2::hphMX-CEB1-WT-1.7 orientation C</i>	This study
ORT7310-2	<i>rad52::KMX, npr2::hphMX-CEB1-WT-1.7 orientation G</i>	This study
ORT7312-5	<i>rad52::KMX, npr2::hphMX-CEB1-WT-1.7 orientation C</i>	This study
ORT6550-2	<i>npr2::hphMX-CEB1-Gmut-1.7 orientation G</i>	This study
ORT6548	<i>npr2::hphMX-CEB1-Gmut-1.7 orientation C</i>	This study
ORT6551-1	<i>pif1::KMX, npr2::hphMX-CEB1-Gmut-1.7 orientation G</i>	This study
ORT6549	<i>pif1::KMX, npr2::hphMX-CEB1-Gmut-1.7 orientation C</i>	This study
ORT7182	<i>npr2::hphMX-HRAS1-0.7 orientation G</i>	This study
ORT7183	<i>npr2::hphMX-HRAS1-0.7 orientation C</i>	This study
ORT7322	<i>pif1::HIS3, npr2::hphMX-HRAS1-0.7 orientation G</i>	This study
ORT7323	<i>pif1::HIS3, npr2::hphMX-HRAS1-0.7 orientation C</i>	This study
ORT6558	<i>npr2::hphMX-CEB25-WT-0,7 orientation G</i>	This study
ORT6556	<i>npr2::hphMX-CEB25-WT-0,7 orientation C</i>	This study
ANT1181-1	<i>npr2::hphMX-CEB25-Cdc13mutG4wt-1.4 orientation G</i>	This study
ANT1180-5	<i>npr2::hphMX-CEB25-Cdc13mutG4wt-1.4 orientation C</i>	This study
ANT1183-1	<i>npr2::hphMX-CEB25-Cdc13mutG4mut-1.4 orientation G</i>	This study
ANT1182-1	<i>npr2::hphMX-CEB25-Cdc13mutG4mut-1.4 orientation C</i>	This study
ORT6559-5	<i>pif1::KMX, npr2::hphMX-CEB25-WT-0,7 orientation G</i>	This study
ORT6557-1	<i>pif1::KMX, npr2::hphMX-CEB25-WT-0,7 orientation C</i>	This study
ANT1185-4	<i>pif1::KMX, npr2::hphMX-CEB25-Cdc13mutG4wt-1.4 orientation G</i>	This study
ANT1184-1	<i>pif1::KMX, npr2::hphMX-CEB25-Cdc13mutG4wt-1.4 orientation C</i>	This study
ANT1187-1	<i>pif1::KMX, npr2::hphMX-CEB25-Cdc13mutG4mut-1.4 orientation G</i>	This study
ANT1186-1	<i>pif1::KMX, npr2::hphMX-CEB25-Cdc13mutG4mut-1.4 orientation C</i>	This study

Table S2

Name	Sequence (5'-3')	Nb. of mutations	TG/GG/GT density		GC content (%)
			G strand	C strand	
CEB1-WT	GGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTGA	NA	0,52	0,08	77
CEB1-Gmut	GCGCGAGTGAGAGTGGCCTGCGGAGGTCCCTGCGCTGA	5	0,31	0,08	72
CEB25-WT	AAGGGTGGGTGTAAGTGTGGGTGGGTGTGAGTGTGGGTGTGGAGGTAGATGT	NA	0,75	0,00	56
CEB25-Cdc13mut	CAGGGCGGGTCTAAGTTAGGGCGGGATTTAGTCAGGACTCGAGGATGATGT	16	0,33	0,06	56
CEB25-Cdc13mut-Gmut	CAGCGCGCTCTAAGTTAGCGCGGATTTAGTCAGGACTCGAGGATGATGT	20	0,18	0,06	56
hRAS1	CCCTGGAGAGAAGGGGAGTGTGGCGT	NA	0,44	0,11	68

Table S3 (Untreated)

Strain	Minisatellite	Number of motif	Orientation	Genotype	GCR rate	95% IC up	95% IC low	n
ORT6531	None			WT	4.30E-10	4.63E-09	4.30E-10	11
ORT6568	None			<i>pif1Δ</i>	1.01E-06	1.56E-07	4.42E-07	20
ORT6542-8	CEB1-WT-0.66	17	G	WT	1.10E-08	9.95E-09	2.75E-09	25
ORT6542-5	CEB1-WT-1.2	31	G	WT	3.86E-07	6.93E-07	1.93E-07	15
ORT6542-6	CEB1-WT-1.7	43	G	WT	6.97E-07	8.87E-08	1.05E-07	45
ORT6542-4	CEB1-WT-1.9	49	G	WT	1.63E-06	1.72E-06	4.25E-07	10
ORT6542-1	CEB1-WT-2.7	70	G	WT	1.59E-05	7.92E-06	7.33E-07	15
ORT6591-4	CEB1-WT-0.45	11	C	WT	1.57E-08	9.03E-09	6.44E-09	10
ORT6591-3	CEB1-WT-0.8	20	C	WT	3.25E-07	1.15E-07	2.12E-07	10
ORT6591-8	CEB1-WT-1.1	28	C	WT	3.45E-07	3.83E-07	1.63E-07	10
ORT6591-1	CEB1-WT-1.7	43	C	WT	7.48E-07	1.10E-07	1.50E-07	20
ORT6543-10	CEB1-WT-0.2	5	G	<i>pif1Δ</i>	2.99E-06	5.18E-07	1.72E-06	15
ORT7154-2	CEB1-WT-0.3	8	G	<i>pif1Δ</i>	3.45E-06	9.93E-07	1.31E-06	10
ORT7154-7	CEB1-WT-0.7	18	G	<i>pif1Δ</i>	7.54E-06	7.62E-06	7.54E-06	15
ORT7154-3	CEB1-WT-0.9	23	G	<i>pif1Δ</i>	1.93E-05	3.70E-06	5.28E-06	15
ORT6543-4	CEB1-WT-1.5	38	G	<i>pif1Δ</i>	7.31E-05	1.67E-05	5.06E-05	10
ORT6543-1	CEB1-WT-1.7	43	G	<i>pif1Δ</i>	3.89E-04	6.05E-05	8.67E-05	16
ORT6541-1	CEB1-WT-0.15	4	C	<i>pif1Δ</i>	3.18E-06	8.89E-07	6.28E-07	10
ORT6592-1	CEB1-WT-0.36	9	C	<i>pif1Δ</i>	4.57E-06	7.46E-07	1.06E-06	10
ORT6592-4	CEB1-WT-0.9	23	C	<i>pif1Δ</i>	1.14E-05	5.12E-06	5.57E-06	10
ORT6592-18	CEB1-WT-1.2	31	C	<i>pif1Δ</i>	2.39E-05	5.10E-06	2.11E-06	10
ORT6592-22	CEB1-WT-1.4	36	C	<i>pif1Δ</i>	2.53E-04	6.22E-05	1.03E-04	12
ORT7153-9	CEB1-WT-1.7	43	C	<i>pif1Δ</i>	4.60E-04	7.05E-05	8.10E-05	14
ORT7189-1	CEB1-WT-1.7	43	G	<i>rad51Δ</i>	2.92E-07	1.04E-07	1.08E-07	11
ORT7191-3	CEB1-WT-1.7	43	C	<i>rad51Δ</i>	1.77E-07	1.48E-07	1.77E-07	11
ORT7310-2	CEB1-WT-1.7	43	G	<i>rad52Δ</i>	3.55E-07	5.10E-08	1.01E-07	11
ORT7312-5	CEB1-WT-1.7	43	C	<i>rad52Δ</i>	1.68E-07	8.09E-08	2.13E-08	10
ORT7309-3	CEB1-WT-1.7	43	C	<i>dnl4Δ</i>	5.26E-07	2.00E-07	1.77E-07	11
ORT7188-3	CEB1-Gmut-0.9	23	G	WT	8.64E-09	4.16E-09	2.46E-09	11
ORT6550-2	CEB1-Gmut-1.7	42	G	WT	2.77E-08	8.76E-09	9.95E-09	40
ORT6550-7	CEB1-Gmut-1.9	49	G	WT	3.52E-08	6.42E-09	2.68E-08	10
ORT7188-2	CEB1-Gmut-2.2	56	G	WT	1.39E-07	4.99E-08	3.67E-08	11
ORT6548	CEB1-Gmut-1.7	42	C	WT	2.07E-08	5.90E-09	3.53E-09	20
ORT6551-5	CEB1-Gmut-0.9	23	G	<i>pif1Δ</i>	2.30E-06	4.78E-07	8.90E-07	10
ORT6551-4	CEB1-Gmut-1.3	33	G	<i>pif1Δ</i>	3.25E-06	1.67E-06	1.19E-06	10
ORT6551-1	CEB1-Gmut-1.7	42	G	<i>pif1Δ</i>	6.32E-06	1.98E-06	8.41E-07	15
ORT6549	CEB1-Gmut-1.7	42	C	<i>pif1Δ</i>	3.05E-06	7.34E-07	1.27E-06	20
ORT7182	hRAS1-0.7	26	G	WT	8.48E-09	1.16E-08	2.55E-09	11
ORT7183	hRAS1-0.7	26	C	WT	1.11E-08	5.91E-09	1.11E-08	10
ORT7322	hRAS1-0.7	26	G	<i>pif1Δ</i>	3.68E-06	9.56E-07	1.07E-06	15
ORT7323	hRAS1-0.7	26	C	<i>pif1Δ</i>	3.21E-06	4.27E-07	1.62E-06	9
ORT6558-1	CEB25-WT-0.7	13	G	WT	1.16E-04	2.55E-05	2.78E-05	15
ORT6556-1	CEB25-WT-0.7	13	C	WT	2.24E-07	4.84E-08	9.53E-09	10
ORT6559-5	CEB25-WT-0.7	13	G	<i>pif1Δ</i>	1.22E-04	1.04E-04	6.01E-05	15
ORT6557-1	CEB25-WT-0.7	13	C	<i>pif1Δ</i>	2.42E-06	1.16E-06	6.44E-07	15
ANT1181-1	CEB25-Cdc13mut-1.4	27	G	WT	3.07E-07	9.13E-08	6.43E-08	20
ANT1180-5	CEB25-Cdc13mut-1.4	27	C	WT	2.95E-07	1.33E-07	5.26E-08	18
ANT1185-4	CEB25-Cdc13mut-1.4	27	G	<i>pif1Δ</i>	3.86E-06	1.08E-06	4.32E-07	20
ANT1184-1	CEB25-Cdc13mut-1.4	27	C	<i>pif1Δ</i>	2.89E-06	1.59E-06	2.74E-07	19
ANT1183-1	CEB25-Cdc13mut-Gmut-1.4	27	G	WT	3.55E-07	1.09E-07	1.40E-07	20
ANT1182-1	CEB25-Cdc13mut-Gmut-1.4	27	C	WT	2.37E-07	1.32E-07	5.84E-08	18
ANT1187-1	CEB25-Cdc13mut-Gmut-1.4	27	G	<i>pif1Δ</i>	2.52E-06	1.84E-06	3.77E-07	18
ANT1186-1	CEB25-Cdc13mut-Gmut-1.4	27	C	<i>pif1Δ</i>	2.07E-06	6.45E-07	3.70E-07	25

Table S3 (Phen-DC-treated)

Strain	Minisatellite	Number of motif	Orientation	Genotype	Treatment	GCR rate	95% IC up	95% IC low	n	Fold vs.	Fold vs.
										untreated	WT
ORT6531	None			WT	Phen-DC <sub>3</sub> 10 μM	7.18E-09	3.33E-09	2.50E-09	11	17	NA
ORT6542-6	CEB1-WT-1.7	43	G	WT	Phen-DC <sub>3</sub> 1 μM	3.29E-06	1.39E-06	1.04E-06	10	5	NA
ORT6542-6	CEB1-WT-1.7	43	G	WT	Phen-DC <sub>3</sub> 5 μM	1.14E-05	5.95E-07	1.06E-06	10	16	NA
ORT6542-6	CEB1-WT-1.7	43	G	WT	Phen-DC <sub>3</sub> 10 μM	3.66E-05	1.00E-05	3.44E-06	25	52	NA
ORT6542-6	CEB1-WT-1.7	43	G	WT	Phen-DC <sub>6</sub> 1 μM	5.11E-06	1.82E-06	2.68E-06	10	7	NA
ORT6542-6	CEB1-WT-1.7	43	G	WT	Phen-DC <sub>6</sub> 5 μM	1.19E-05	1.93E-06	2.63E-06	10	17	NA
ORT6591-1	CEB1-WT-1.7	43	C	WT	Phen-DC <sub>3</sub> 10 μM	1.66E-05	4.99E-06	6.89E-06	20	22	NA
ORT6550-2	CEB1-Gmut-1.7	42	G	WT	Phen-DC <sub>3</sub> 10 μM	3.57E-08	2.24E-08	7.39E-09	25	1.3	NA
ORT6550-2	CEB1-Gmut-1.7	42	G	WT	Phen-DC <sub>6</sub> 5 μM	1.38E-08	1.71E-08	1.38E-08	10	0.5	NA
ORT6548	CEB1-Gmut-1.7	42	C	WT	Phen-DC <sub>3</sub> 10 μM	1.59E-08	6.34E-09	5.51E-09	22	0.8	NA
ORT6556-1	CEB25-WT-0.7	13	C	WT	Phen-DC <sub>3</sub> 10 μM	2.00E-07	3.56E-08	7.28E-08	10	0.9	NA
ORT6557-1	CEB25-WT-0.7	13	C	<i>pif1Δ</i>	Phen-DC <sub>3</sub> 10 μM	1.34E-05	5.56E-06	3.26E-06	10	5.5	NA
ANT1181-1	CEB25-Cdc13mut-1.4	27	G	WT	Phen-DC <sub>3</sub> 10 μM	3.25E-07	9.95E-08	1.09E-07	10	1.1	NA
ANT1180-5	CEB25-Cdc13mut-1.4	27	C	WT	Phen-DC <sub>3</sub> 10 μM	3.22E-07	4.11E-07	6.60E-08	10	1.1	NA
ANT1185-4	CEB25-Cdc13mut-1.4	27	G	<i>pif1Δ</i>	Phen-DC <sub>3</sub> 10 μM	9.00E-06	4.40E-06	1.49E-06	10	2.3	NA
ANT1184-1	CEB25-Cdc13mut-1.4	27	C	<i>pif1Δ</i>	Phen-DC <sub>3</sub> 10 μM	1.32E-05	8.60E-06	1.40E-06	10	4.6	NA
ANT1183-1	CEB25-Cdc13mut-Gmut-1.4	27	G	WT	Phen-DC <sub>3</sub> 10 μM	1.97E-07	6.39E-08	7.26E-08	10	0.6	NA
ANT1182-1	CEB25-Cdc13mut-Gmut-1.4	27	C	WT	Phen-DC <sub>3</sub> 10 μM	2.04E-07	1.00E-07	3.93E-08	10	0.9	NA
ANT1187-1	CEB25-Cdc13mut-Gmut-1.4	27	G	<i>pif1Δ</i>	Phen-DC <sub>3</sub> 10 μM	1.82E-06	2.55E-06	7.43E-07	10	0.7	NA
ANT1186-1	CEB25-Cdc13mut-Gmut-1.4	27	C	<i>pif1Δ</i>	Phen-DC <sub>3</sub> 10 μM	8.45E-07	8.16E-07	2.96E-07	13	0.4	NA
ORT7189-1	CEB1-WT-1.7	43	G	<i>rad51Δ</i>	Phen-DC <sub>3</sub> 10 μM	1.13E-05	5.92E-06	7.61E-07	16	39	0.31
ORT7191-3	CEB1-WT-1.7	43	C	<i>rad51Δ</i>	Phen-DC <sub>3</sub> 10 μM	5.27E-06	4.23E-06	1.71E-06	21	30	0.14
ORT7310-2	CEB1-WT-1.7	43	G	<i>rad52Δ</i>	Phen-DC <sub>3</sub> 10 μM	7.12E-06	6.45E-06	7.12E-06	10	20	0.19
ORT7312-5	CEB1-WT-1.7	43	C	<i>rad52Δ</i>	Phen-DC <sub>3</sub> 10 μM	1.48E-06	6.31E-07	2.25E-07	6	9	0.04



### II.D.3. Résultats supplémentaires

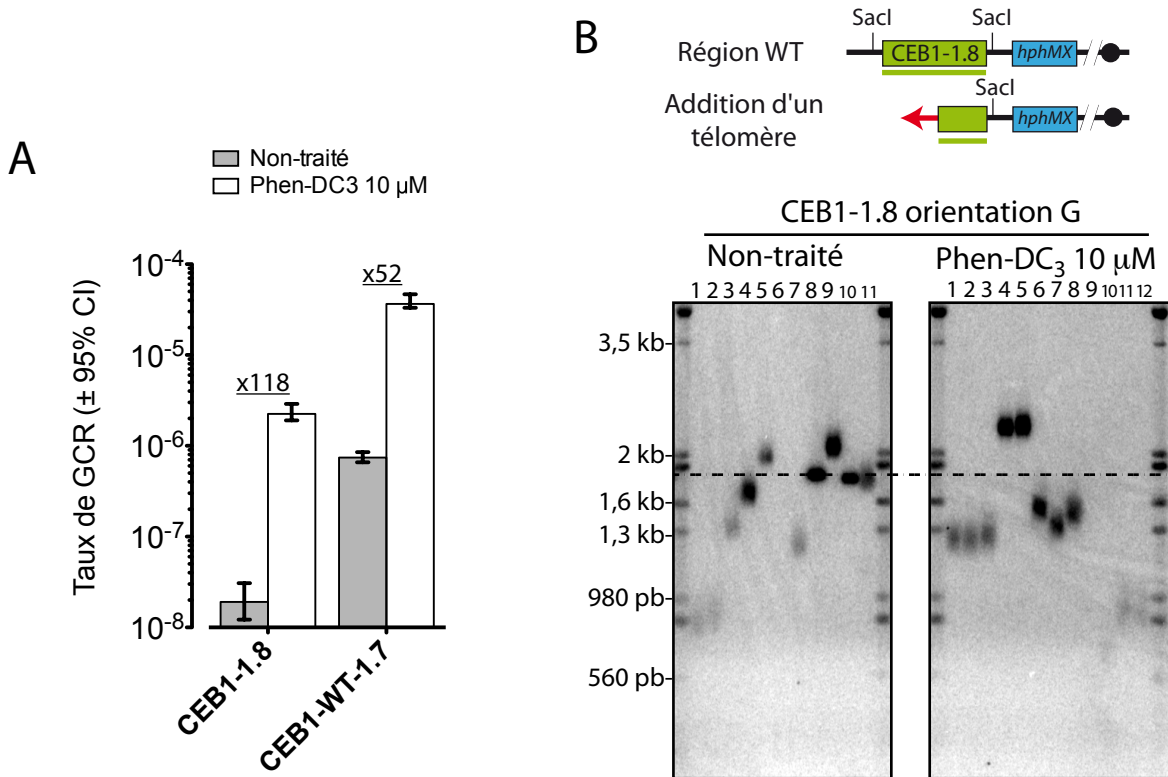
#### II.D.3.a. L'allèle naturel CEB1-1.8 induit des GCR à des taux plus faible que l'allèle artificiel

L'allèle naturel CEB1-1.8, dont l'instabilité a été étudiée aux locus *ARG4* (**articles 1 et 2**) et *ARS305* (**article 3**), a également été inséré au locus *NPR2* dans l'orientation G. Dans cette souche, le taux de GCR est  $1,9 \times 10^{-8}$  événements/génération ( $n=15$ ). Le traitement avec Phen-DC<sub>3</sub> 10  $\mu$ M résulte en une augmentation du taux de GCR d'un facteur 118 ( $2,2 \times 10^{-6}$  événements/génération,  $n=15$ ) (**Figure 38A**). Respectivement 12 et 11 GCRs obtenus à partir des souches traitées et non-traitées ont été analysés par Souther blot (**Figure 38B**). En absence de traitement, 10 réarrangements sur 11 contiennent la séquence CEB1 : 8/10 sont des smears dont le poids moléculaire indique une addition de télomères dans CEB1. Les deux réarrangements restants sont des bandes discrètes : l'une migre à la taille attendue pour la région non-altérée (clone 8), et l'autre à une taille différente (clone 10). La nature du réarrangement ayant perdu CEB1 n'a pas été déterminée. Des réarrangements induits par Phen-DC<sub>3</sub>, 11/12 sont des additions de télomères dans CEB1. Un réarrangement a perdu CEB1 et sa nature n'a pas été déterminée (ligne 9). Ainsi, de manière similaire à l'allèle CEB1 artificiel de même taille (CEB1-WT-1,7), le taux de GCR en présence d'un allèle contenant des séquences G4P est fortement stimulé par le traitement avec Phen-DC<sub>3</sub>, et les réarrangements induits dans l'orientation G sont majoritairement des ajouts de télomère dans CEB1. Cependant, le taux de GCR observé avec CEB1-1.8 est  $\approx 35$  inférieur à ce qui est mesuré avec CEB1-WT-1,7 en absence de traitement, et 16 fois inférieur en présence de Phen-DC<sub>3</sub>. Ainsi, que le type de réarrangements étudié soit des GCRs ou des expansions/contractions, l'allèle artificiel est plus instable que l'allèle naturel. Cette différence est discutée au **Chapitre III.D.2**.

#### II.D.3.b. Phen-DC<sub>3</sub> stimule la capacité de la répétition télomérique humaine à induire des GCR

La répétition télomérique humaine (TTAGGG)<sub>n</sub> se comporte comme un site fragile dans des cellules humaines (Sfeir et al. 2009) et induit des réarrangements chromosomiques quand elle est insérée dans un gène dans des cellules CHO (Kilburn et





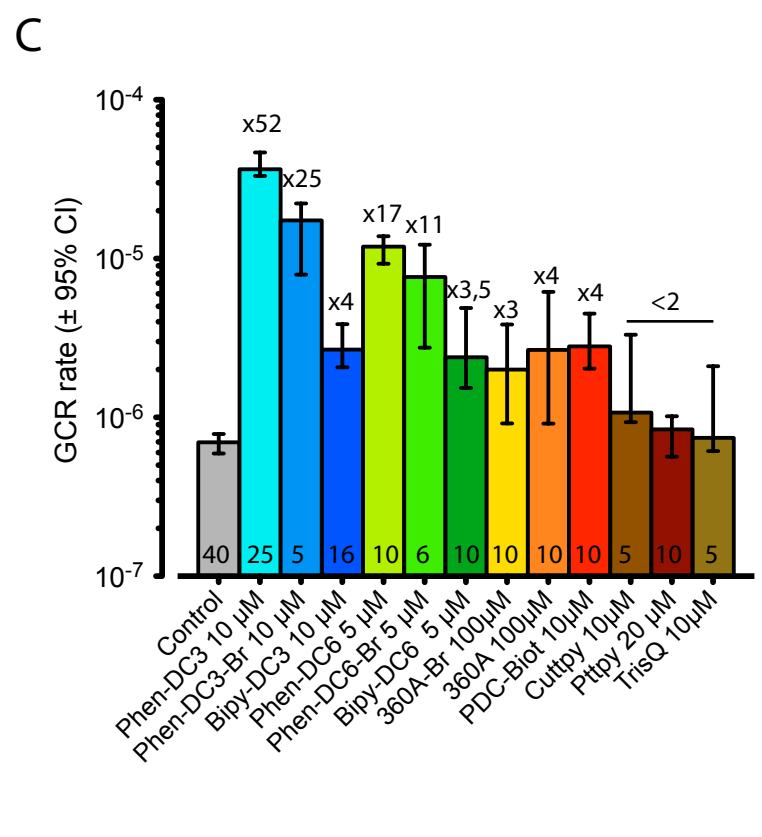
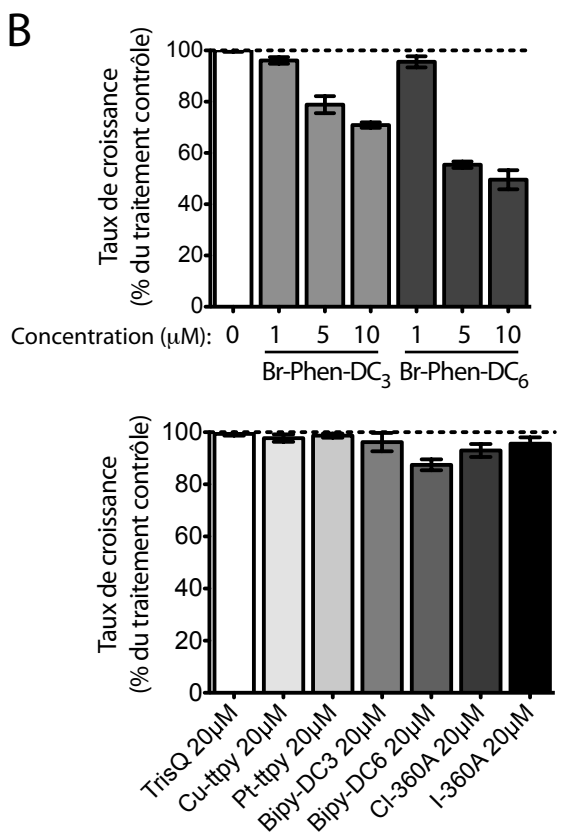
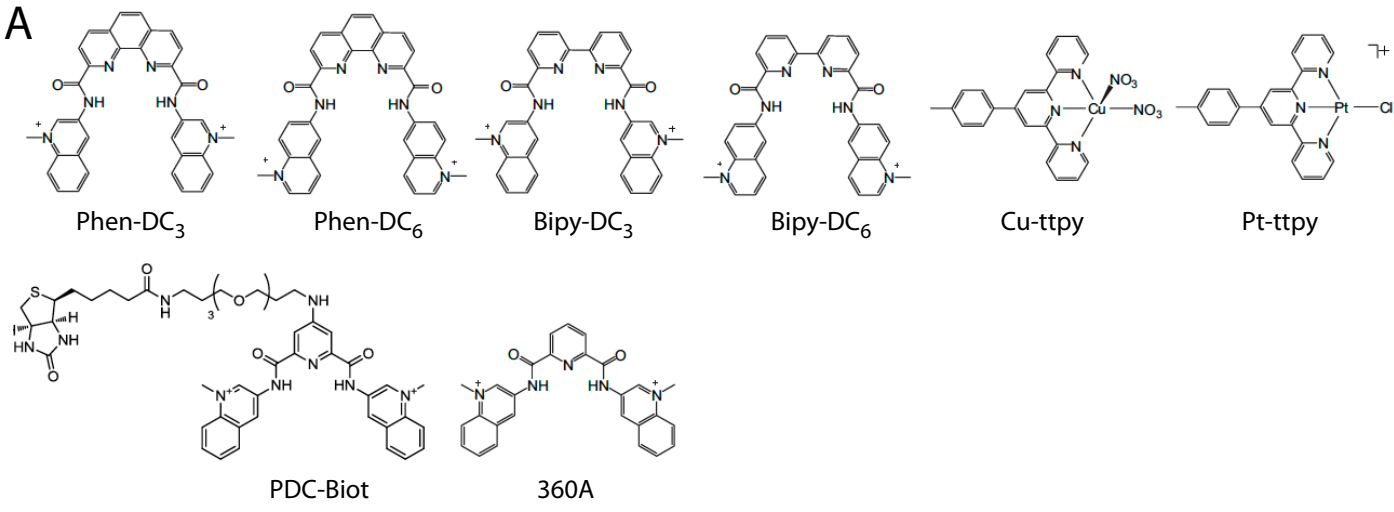
**Figure 38: L'allèle naturel CEB1-1.8 stimule moins le taux de GCR que l'allèle artificiel de même de taille dans une souche WT.** (A) Taux de GCR d'une souche WT avec l'allèle naturel CEB1-1.8 inséré dans l'orientation G au locus *NPR2* (ORT6518-2) en présence ou en absence de Phen-DC<sub>3</sub>. L'augmentation du taux de GCR induite par le traitement est indiquée. Le taux de GCR en présence de l'allèle artificiel CEB1-WT-1.7 (ORT6542-6) dans les mêmes conditions est présenté pour comparaison (**Article 4, Figure 2D**). (B) Le schéma présente les sites de restrictions à proximité de CEB1-1.8. Le fragment non-réarrangée mesure 1840 pb (ligne en pointillés). Des fragments de poids moléculaires au maximum supérieur à 1840 + 300 pb et qui forment un smear sont attendues en cas d'ajout de télomère dans CEB1 (voir **Article 4, Figure 3A**). Les Southern blots montrent les GCRs obtenus dans une souche WT contenant CEB1-1.8 dans l'orientation G, après traitement ou non par Phen-DC<sub>3</sub>.

al. 2001). Cette instabilité pourrait être due à sa capacité à former des G-quadruplexes (Parkinson et al. 2002; Tran et al. 2010). Puisque Phen-DC<sub>3</sub> lie avidement le G-quadruplexe télomérique humain *in vitro* (De Cian et al. 2007a; Monchaud et al. 2008; Largy et al. 2011), j'ai testé si le traitement d'une souche WT contenant un fragment (TTAGGG)<sub>n</sub> pourrait stimuler le taux de GCR, comme observé avec CEB1. Pour cela, j'ai introduit une répétition télomérique de 0,7 kb (TTAGGG-0,7) dans l'orientation G au locus NPR2. Cette séquence induit un taux de GCR de  $3.01.10^{-6}$  évènements/génération,  $\approx 7,200$  fois supérieur à ce qui est observé dans une souche sans minisatellite. Le traitement par Phen-DC<sub>3</sub> stimule le taux de GCR d'un facteur 6 ( $1.92.10^{-5}$  évènement/génération). Ainsi, bien que la part du taux de GCR due à la capacité à former des G-quadruplexes par la répétition télomérique ne puisse être déterminée en absence de traitement (on ne peut exclure que la capacité à lier la protéine subtélomérique essentielle Tbf1 puisse être impliquée dans ce taux élevé de GCR (Brigati et al. 1993; Koering et al. 2000)), la fragilité de cette séquence est stimulée lorsque les cellules sont traitées avec un ligand spécifique et affin des G-quadruplexes. Ce résultat supporte l'idée que la séquence télomérique humaine forme des G-quadruplexes *in vivo*, et que leur persistance est source de fragilité.

### II.D.3.c. Effet d'autres ligands de G-quadruplexes

Le test GCR est un système expérimental simple et sensible qui permet de tester la sélectivité de ligands de G-quadruplexes *in vivo*, en comparant l'induction de GCR en présence d'une séquence G4P et de sa contrepartie mutée. La spécificité de ces molécules pour certaines conformations de G-quadruplexes pourrait également être estimée *in vivo* en comparant leur effet déstabilisateur sur des souches contenant différentes séquences G4P. Ce projet est mené en collaboration avec l'équipe de chimistes de MP. Teulade-Fichou, afin de tester l'efficacité *in vivo* de nombreux ligands de G-quadruplexes qu'ils synthétisent (**Figure 39A**).

A l'exception des dérivés bromés de Phen-DC<sub>3</sub> et Phen-DC<sub>6</sub>, les autres ligands testés n'affectent pas ou peu (<10%) la croissance des cellules en milieu SC (**Figure 39B**). La capacité de ces ligands à induire des GCR a été testée sur une souche WT contenant CEB1-WT-1.7 dans l'orientation G (ORT6542-6). Les traitements par des concentrations sub-toxiques de Br-Phen-DC<sub>3</sub> (à 10  $\mu$ M) et Br-Phen-DC<sub>6</sub> (à 5  $\mu$ M)



**Figure 39: Effet de divers ligands de G-quadruplexe sur la croissance et le taux de GCR d'une souche WT de *S. cerevisiae*.** (A) Structure des ligands de G-quadruplexes utilisés. Tiré de (Monchaud et al. Biochimie, 2008) et communication personnelle E. Largy (MP Teulade-Fichou). (B) Effet de différents ligands de G-quadruplexes sur la croissance de *S. cerevisiae*. Les cellules ont subi une pré-culture en milieu SC sur la nuit, pour atteindre la saturation au matin. Elles sont re-diluées (DO=0,3) dans du milieu contenant ou non le ligand à la concentration indiquée, dilué et filtré extemporanément. Le taux de croissance est déterminé en phase exponentielle (2 - 6h après dilution) et exprimé en pourcent de la croissance du traitement contrôle (DMSO 0,04%). La moyenne et l'écart-type de 3 expériences indépendantes sont présentés. (C) Taux de GCR dans une souche WT contenant CEB1-WT-1.7 dans l'orientation G (ORT6542-6) après traitement par différents ligands de G-quadruplexe. Le protocole est le même que celui utilisé dans l'article 4 pour Phen-DC<sub>3</sub> et Phen-DC<sub>6</sub>. L'augmentation du taux de GCR et le nombre de cultures indépendantes réalisées sont indiqués au sommet et au bas de chaque barre, respective-

induisent des taux de GCR respectivement 25 et 11 fois supérieurs à ce qui est observé en absence de traitement ( $1,7 \times 10^{-5}$  et  $7,7 \times 10^{-6}$  évènements/génération). Ces inductions sont similaires à celles observées avec les versions non-bromées (**Figure 39C**). Les Bipy-DC<sub>3</sub> 10 μM, Bipy-DC<sub>6</sub> 10 μM, 360A et Br-360A 100 μM et Biot-360A (=PDC-biot) à 10 μM stimulent le taux de GCR d'un facteur  $\approx 4$ . Cu-ttpty, Pt-ttpty et TrisQ à 10 μM n'ont pas d'effet sur le taux de GCR (**Figure 39C**). Le niveau d'induction de GCR est positivement corrélé à l'inhibition de croissance (p-value =  $2,8 \times 10^{-3}$ , test de corrélation de Spearman), ce qui suggère soit que l'inhibition de croissance est directement liée à la capacité stabilisatrice des ligands *in vivo*, soit que les ligands ne démontrant d'effet ni sur la croissance, ni sur le taux de GCR, ne pénètrent pas dans les cellules. En faveur de la seconde hypothèse, les ligands les plus efficaces pour la déstabilisation de CEB1 (les Phen-DC) sont les molécules les plus planes, ce qui favoriserait leur passage à travers la membrane plasmique. L'utilisation de mutants de détoxification (*pdr5Δ*, *erg6Δ*, ...) couplé à l'utilisation d'agents perméabilisants (poly-godial) pourrait augmenter la sensibilité du système expérimental.



**Structure-Function Relationship of  
G-quadruplexes to Genomic Instability**



## **II.E. Structure-Function Relationship of G-quadruplexes to Genomic Instability** (collaboration with M. Adrian and AT. Phan)

### **II.E.1. Loop Length Affects the Instability of G-Quadruplex-Forming Sequences in *S. cerevisiae***

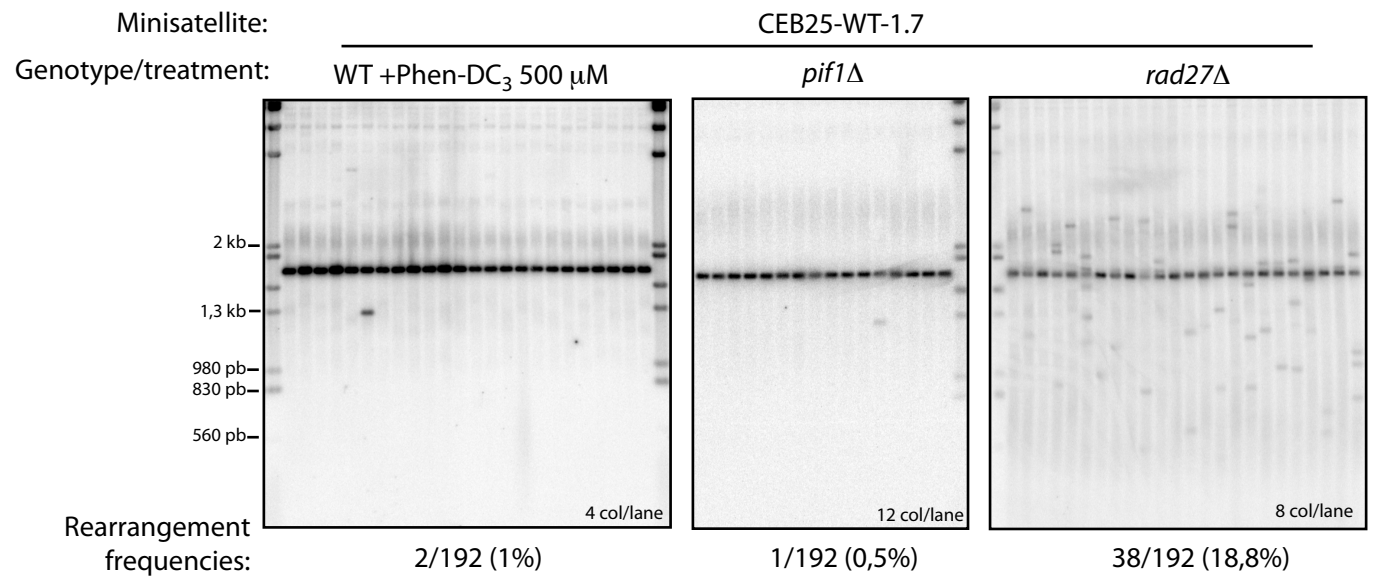
#### **II.E.1.a. CEB25 is stable in the absence of Pif1 or upon treatment with Phen-DC<sub>3</sub>**

In order to examine the instability of other G-quadruplex-forming sequences, we introduced the CEB25 minisatellite (**Chapitre I.C.1.a.ii**) into the genome of *S. cerevisiae*. This analysis, initiated by A. Serero at the *ARG4* locus on Chr. VIII (**Figure 40**) revealed that CEB25 remained stable in conditions that destabilize CEB1: the rearrangement frequency of the synthetic allele CEB25-WT-1,7 kb ( $\approx 33$  motifs) is null in a WT strain treated with Phen-DC<sub>3</sub> and in the *pif1* $\Delta$  mutant (**Figure 40**). The *PIF1* deletion combined to the Phen-DC<sub>3</sub> treatment does not induce rearrangement neither (0/192). The ability of CEB25 to bind Cdc13 with high affinity is not responsible for this stable behavior, since the CEB25-Cdc13mut-1,6 allele, mutated for its Cdc13-binding sites ( $\approx 31$  motifs, sequence see **Article 4 Table S2**) is stable in the absence of Pif1 (**Table 5**). This is not due to an intrinsic inability of CEB25 to produce rearranged alleles, since both the natural and artificial CEB25 minisatellites are destabilized in a *rad27* $\Delta$  mutant (**Figure 40, Table 5**).

In order to study a potential effect of the insertion site and to increase the sensitivity of our experimental system, I inserted CEB25 in close proximity to the *ARS305* (Chr. III) where the rearrangement frequency of CEB1 in the orientation I is  $\approx 10$ -times higher than at the *ARG4* locus (**Article 3**). I also inserted CEB25 at the *NPR2* locus (Chr. V) that allows to measure the formation of GCR (**Article 4**).

Near *ARS305*, like at *ARG4*, the synthetic CEB25 allele (CEB25-WT-0,7, 13 motifs) inserted in the orientation I is stable upon treatment with Phen-DC<sub>3</sub> (0/192) or in the absence of Pif1 (1/192) (**Figure 41B, Table 6**). However, a synthetic allele of CEB1 of equivalent size (CEB1-WT-1,1, 26 motifs) is strongly destabilized in these conditions (12 and 59%, respectively) (**Figure 42B**).





**Figure 40: CEB25 instability at the *ARG4* locus in WT strains treated by Phen-DC3, or in *pif1*Δ and *rad27*Δ mutants.** Southern blot analysis of the instability of the synthetic CEB25-WT-1,7 allele (33 motifs) inserted in the promoter of the *ARG4* gene in a WT strain (AND1420-E8) treated with Phen-DC3 at 500 μM (YPD medium), *pif1*Δ strain (AND1420-E9), and a *rad27*Δ strain (ANT1417-3F). Digestion: *Apa*I/*Spe*I. Probe: CEB25.

**Table 5: Rearrangement frequencies of CEB25 inserted near *ARG4* in WT, *pif1Δ*, and *rad27Δ* strains.**

Minisatellite	CEB25-2,5	CEB25-WT-1,7	CEB25-Cdc13mut-1,6
WT	2/264 (0,8%)	ND	0/192
<i>pif1Δ</i>	0/192	1/192 (0,5%)	1/180 (0,5%)
<i>rad27Δ</i>	25/170 (14,7%)*	38/192 (18,8%)	25/372 (6,7%)*

Results obtained by A. Serero

\* p-value vs. WT < 0,05

**Table 6: Instability of different CEB25 alleles inserted near the *ARS305* in the orientation I in WT strains ± Phen-DC3, and in *pif1Δ* strains.**

Minisatellite	CEB25-WT-0,7	CEB25-L1T-0,4	CEB25-Cdc13mut-	CEB25-Cdc13mut-
			L1T-0,6	Gmut-L1T-0,8
Central loop size (nt)	9	1	1	1
G4P sequence	wt	wt	wt	mut
Cdc13-binding sites	wt	wt	mut	mut
WT	0/192	0/96	5/90 (5.5%)	ND
WT + Phen-DC <sub>3</sub>	0/192	10/94 (10.6%)*	32/176 (18%)*	ND
<i>pif1Δ</i>	1/192 (0.5%)	17/38 (45%)*	6/15 (40%)*	0/192

\* p-value vs. CEB25-WT < 0,05

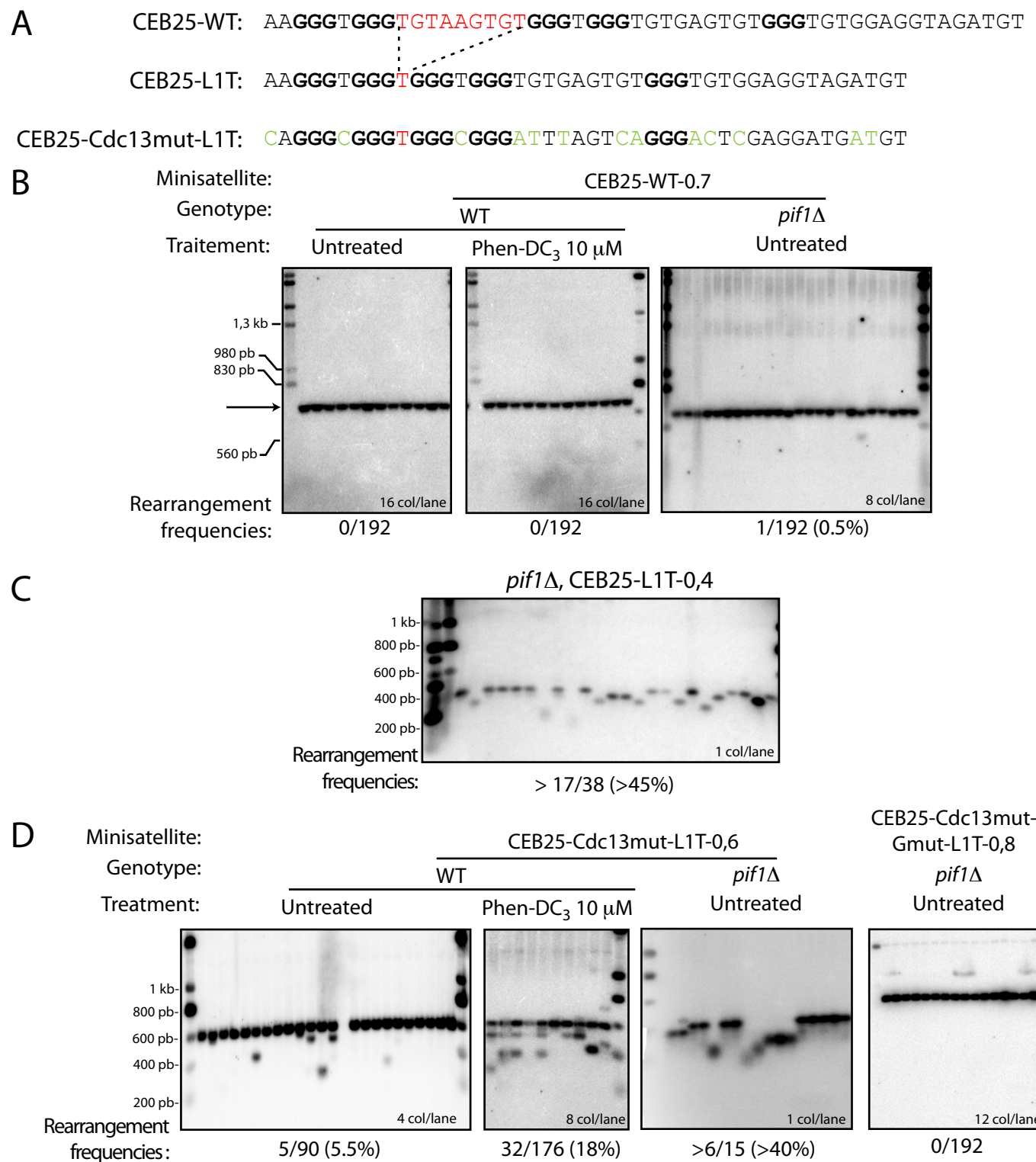
At the *NPR2* locus, the GCR rate of a *pif1Δ* strain bearing CEB25 is not different from the "no insert" *pif1Δ* control strain, and the Phen-DC<sub>3</sub> treatment does not stimulate the GCR formation (**Article 4, Figure 5**). However, a weak GCR rate induction (x2 - 5) is observed in *pif1D* strains treated by Phen-DC<sub>3</sub> (**Article 4, Figure 5E**). These conditions do not stimulate the GCR formation of a G-mutated CEB25 allele (**Article 4, Figure 5E**). Thus, a non-negligible G-quadruplex-dependent GCR induction is detected with CEB25, but only when combining Phen-DC<sub>3</sub> treatment with *PIF1* deletion, conditions known to synergistically destabilize CEB1 (**Article 2, Figure 4**).

If CEB1 and CEB25 both form stable G-quadruplexes *in vitro*, which properties specific to these minisatellites may explain their different behavior *in vivo*? Does the CEB25 G-quadruplex form less often, or is more easily unwound, than the CEB1 G-quadruplex?

#### II.E.1.b. The central loop of the CEB25 G-quadruplex stabilizes the minisatellite *in vivo*

The CEB25 motif forms a single G-quadruplex, parallel (**Figure 4A**) (Amrane 2012) whereas the CEB1 motif is a mixture of at least 5 G-quadruplexes, for which two structures have been determined (see **Chapitre II.E.2.a**) (**Figure 43**). The central loop of 9 nt is a particularity of the CEB25 G-quadruplex, the maximum loop length of the G4P sequence of CEB1 being of 4 nt. To test the role of this 9 nt loop in the stability of CEB25 *in vivo*, I realized complementary mutagenesis experiments. First, I shortened the CEB25 loop from 9 to 1 nt. Second, I introduced the 9 nt loop of CEB25 into the G4P sequence of CEB1.

Two alleles of CEB25 containing a central loop of a single nucleotide have been synthesized, either in a "WT" allele (CEB25-L1T-0.4, 8 motifs) or in an allele mutated for its Cdc13-binding sites (CEB25-Cdc13mut-L1T-0.6, 12 motifs)(**Figure 41A**). They have been inserted in the orientation I near *ARS305*. Spectacularly, the CEB25-L1T-0.4 is strongly destabilized in the absence of the Pif1 helicase (45%, 17/38)(**Figure 41C**). The CEB25-Cdc13mut-L1T-0.6 exhibits a slight instability in a WT strain (5%, 5/90), which is exacerbated by the Phen-DC<sub>3</sub> treatment (18%, 32/176, p-value = 4.6e<sup>-3</sup>) or by the absence of Pif1 (40%, 6/15, p-value vs. WT = 9.5e<sup>-5</sup>) (**Figure 41D**). The CEB25-Cdc13mut-L1T allele mutated for its 4 G-triplets (CEB25-Cdc13mut-Gmut-L1T-0.8, 15



**Figure 41: The shortening of the 9-nt loop of the CEB25 G-quadruplex induces its instability.** (A) Sequences of the motifs of the synthetic minisatellites CEB25-WT, CEB25-L1T (9-nt loop of CEB25-WT replaced by a T), and CEB25-Cdc13mut-L1T (mutation of the high-affinity Cdc13-binding sites of CEB25-L1T). The G-triplets are in bold. (B) Southern blot analysis of the CEB25-WT-0.7 instability in a WT strain (ORT7167-1) treated or not with Phen-DC3, and in a *pif1*Δ strain (ORT7175-1). (C) Analysis of the CEB25-L1T-0.4 instability in a *pif1*Δ strain (ORT7301-1). (D) Analysis of the CEB25-Cdc13mut-L1T-0.6 instability in a WT strain (ORT7302-6) treated or not with Phen-DC3, and in a *pif1*Δ strain (ORT7303-1). On the right, CEB25-Cdc13mut-Gmut-L1T-0.8 in a *pif1*Δ strain (ORT7318-2). (B-D) All the minisatellites have been integrated near the *ARS305* in the orientation I. Digestion: EcoRI.

motifs) is stable in the absence of Pif1 (0/192)(**Figure 41D**). Thus, the instability observed with CEB25-Cdc13mut-L1T-0.6 depends on its G4P sequences. In conclusion, a CEB25 allele whose central loop has been shortened from 9 to 1 nucleotide behave like CEB1 *in vivo*.

I realized the reciprocal experiment that consists in inserting a 9 nt loop of CEB25 within the G4P sequence of CEB1 (**Figure 42A**). A CEB1 allele in which the "Cdc13mut" version of the 9 nt loop has also been constructed. The resulting alleles are CEB1-loopCEB25-1.0 (21 motifs) and CEB1-loopCEB25cdc13mut-1.0 (21 motifs), respectively. During transformation into yeast, the CEB1-loopCEB25-1.0 underwent an expansion at 1.2 kb ( $\approx$ 24 motifs). These two minisatellites are stable in a WT strain, treated or not with Phen-DC3, as well as in a *pif1* $\Delta$  strain (**Figure 42B**). Thus, inserting the CEB25 loop into the G4P sequence of CEB1 is sufficient to abolish its instability. The loop sequence, and especially the presence of a Cdc13 binding site, does not affect the stabilization of the minisatellite. These results demonstrate that a G4P sequence bearing a long loop is not able to induce the instability of the underlying sequence. Hence, a simple structural characteristic of G-quadruplexes that poorly affects their thermal stability *in vitro* (Guedin et al. 2010a; Amrane et al. 2012) drastically modifies their behavior *in vivo*.

How a 9 nt loop within the G-quadruplex stabilizes the minisatellite? Does the G-quadruplex become substrate for other unwinding activities than the one of Pif1? To test this hypothesis, and better describe the effect of the loop length on the CEB25 G-quadruplex on its instability, I am building alleles containing central loops of 3, 5, and 7 nt. Their instabilities will be measured in WT cells treated with Phen-DC<sub>3</sub> or in the absence of Pif1.



## II.E.2. Multiple G-quadruplexes Formed by CEB1

### Contributes to its Instability

#### II.E.2.a. The CEB1 motif adopt several G-quadruplex conformations

(M. Adrian & AT Phan)

The initial study of the CEB1 motif ability to form G-quadruplex suggested that several conformations were possible, for two reasons. First, the number and the position of the G provided to the structure by the run of 6 Gs ( $G_{2-7}$ ) may vary. Second, the repetitive nature of CEB1 may allow additional “inter-motif” G-quadruplexes (two contiguous motifs are shown separated by dotted lines, and nucleotides are numbered **Figure 43A**) (**Article 1**). Collaboration with the team of AT Phan has been initiated to unravel the diversity of these conformations. The NMR spectrometry analysis of the full motif revealed the presence of a mixture of different conformations in solution ( $K^+$ ). In order to characterize each of them, the CEB1 motif has been cut into several smaller pieces, excluding or mutating some Gs. This narrowing procedure allowed to determine 5 different G-quadruplex conformations and the G-runs involved in each form. A summary of the thermal stability and strand orientations is shown **Figure 43A**. The core G4P sequence (between  $G_{34}$  and  $G_{18}$  of motif n-1 and n, respectively) form the majority of the possible G-quadruplexes (forms 1, 3, 3', and 4), but inter-motif G-quadruplexes also exist (forms 2). Except form 3', none of the oligonucleotides match a consensus G4P sequence (**Chapitre I.A.3.a.i**): oligonucleotides 2 contains a 18 nt-long loop, and oligonucleotides 1, 3 and 4 do not bear 4 G-triplets separated by 3 loops of at least 1 nt (**Figure 43A**).

Three-dimensional structures of form 1 and 4 have been determined by NMR (**Figure 43B** et **43C**). The stability of both forms depends on the oligonucleotide concentration, suggesting formation of an inter-molecular G-quadruplex. Regarding form 1, it adopts an all parallel strand conformation with 3 lateral loops of 1 nt and all but one bases in the *anti* conformation ( $G_3$  is *syn*). Two particular features must be highlighted:

- One G-strand is interrupted, formed by the  $G_{17}G_{18}$  doublet and by an unused G from the 6-Gs run ( $G_3$ ), which snaps back into the G-quadruplex core.

A

CEB1 motif

33 34 35 36 37 38 39 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39

TGGGCTGAGGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTG

AGGGGGGAGGGAGGGTGG

AGGGGGGAGGGAGGGTTTCCCTGCGGAGGTCCCTGGG

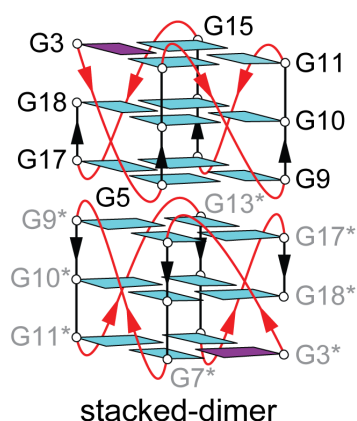
TGGGCTGAGGGGGGA~~TTT~~AGGGTGG

TGGGCTGAGGGGGGAGGGAGGGT

TGGGCTGAGGGGGGAGGGA

G4	Tm (°C)	Strands	Notes
Form 1	65	Parallèle	Dimer
Form 2	45-56	Parallèle	Long loop (18 nt)
Form 3	39	Anti/3+1	-
Form 3'+4	ND	Anti/3+1	Mix of 2 forms
Form 4	43-47	Parallèle	Bimolecular dimer

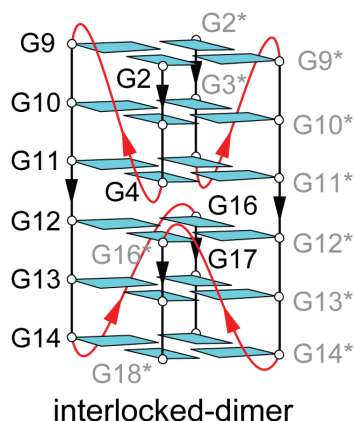
B



Form 1

AGGGGGGAGGGAGGGTGG

C



Form 4

TGGGCTGAGGGGGGAGGGA

**Figure 43: The CEB1 motif adopt several G-quadruplex conformations (M Adrian & AT Phan, personal communication).** (A) Oligonucleotides able to form one or several G-quadruplexes *in vitro*. The CEB1 motif can adopt five different conformations : 1, 2, 3, 3', 4 . Tm (°C), as well as strand arrangements (parallel, antiparallel, 3+1), are indicated. The Tm of the form 1 is sensitive to the oligonucleotide concentration: the value shown has been determined with 5  $\mu$ M of oligonucleotide. (B) Tridimensional structure of the G-quadruplexe form 1, determined by NMR. The structure is a stacked dimer. Monomers are arranged in a head-to-head manner. Each monomer is a parallel G-quadruplex, with external loops of 1 nt. One of the four strand, composed of G3, G17, and G18 is interrupted. It seems that G4 can substitute to G3 in a fraction of the G-quadruplex population. (C) Tridimensional structure of the form 4, determined by NMR. The structure is a bi-molecular G-quadruplex, arranged as an interlocked dimer. Each monomer is a parallel G-quadruplex. The top monomer has 2 external loops of 4 nt, and the bottom monomer has 2 external loops of 1 nt. For (B) and (C), the oligonucleotides used are shown with the Gs involved highlighted in blue (*anti* conformation) or purple (*syn* conformation).



- Two G-quadruplexes subunits are stacked head-to-head. This dimer formation is responsible for the unique strand concentration-dependent stability of the structure.

This type of stacked dimeric structure is unusual and have been reported once for the HIV inhibitor T30177 (**Chapter I.A.1.b.vi, Figure 5**)(Mukundan NAR 2011). The G-strand interruption, although not matching the consensus G4P sequence, have already been reported for few structures (**Chapter I.A.1.b.vi, Figure 5**).

Regarding form 4, strands are all parallel, and bases in the *anti* conformation. Consistent with its formation being dependent on the oligonucleotide concentration, the structure is inter-molecular. It consists in an interlocked bi-molecular G-quadruplex (**Figure 43C**). The structure is not symmetric, and bears only two loops per monomer: the top monomer bears the loops of 4 nt (CTGA) and the bottom monomer bears the loops of a single nucleotide (A). However, this G-quadruplex may not be the only one formed by this oligonucleotide. Indeed, at low concentration a G-quadruplex is still formed, suggesting the existence of an intramolecular form 4. This imply a 0 nt loop for this monomeric form. Experiments are ongoing to determine its structure.

These results underline the diversity of structures adopted by very G-rich strands, which can involve Gs not present in G-triplets. Does these unusual G-quadruplexes form *in vivo*? Are they implicated in the instability of CEB1? In addition, this analysis focused on the G-quadruplexes formed by the more common motif of CEB1 minisatellite, but the end-motif polymorphisms of the natural CEB1 allele (which add or several guanines and interruption of the run of 6Gs likely) are likely to increase the structural diversity of G-quadruplexes this sequence can adopt (**Figure 30B**).

#### II.E.2.b. Point mutagenesis of CEB1 reveals G-runs involved in its instability

Based on this analysis, we examined if these various G-quadruplexes may induce CEB1 instability *in vivo*. To this aim we synthesized CEB1 alleles mutated for particular Gs and unable to adopt certain conformations. This approach is theoretical able to address whether one or several of these forms are involved in the CEB1 instability *in vivo*, and to which extent. The results are summarized **Figure 47**.

### *II.E.2.b.i. G-triplets are all required for the instability of CEB1*

First, we tested if the isolated G-triplets (G<sub>34-36</sub>, G<sub>9-11</sub>, and G<sub>13-15</sub>) are required for the CEB1 instability. We independently mutated the central G of each G-triplets (G->T). The resulting minisatellites of  $\approx$  1 kb (CEB1-G35T, CEB1-G10T, and CEB1-G14T, respectively, sequences see **Figure 44A**), together with a same-sized CEB1-WT allele, were inserted near the *ARS305* in the orientation I, and their instability assayed in WT cells treated or not with Phen-DC<sub>3</sub> and in *pif1* $\Delta$  cells. Contrary to the CEB1-WT minisatellite, which is strongly destabilized upon Phen-DC<sub>3</sub> treatment or *PIF1* deletion (12 and 59%, respectively)(**Figure 42B**), the mutated minisatellites remained fairly stable in any conditions (**Figure 44B-D**): except a barely significant instability induction for CEB1-G35T in the *pif1* $\Delta$  mutant (3,2%, 3/84, p-value vs. CEB1-WT = 0,033), the rearrangement frequencies remained lower than 1% (**Figure 44B-D**). These results demonstrate that the 3 independent G-triplets, previously mutated at once in the initial CEB1-Gmut construct (**Articles 1 - 4**), are critical to destabilize CEB1. The G-quadruplex(es) forms 2 and 3' are the only conformations that require all these G-triplets (**Figure 43A**) suggesting that at least one of these forms is required to destabilize CEB1.

### *II.E.2.b.ii. The G-doublet involved in two G-quadruplex forms is partially required for the CEB1 instability*

To test the role of the G-doublet G<sub>17</sub>G<sub>18</sub> involved in form 1 and 3, I constructed two alleles: one mutated for the G-doublet (CEB1-G17,18T) and the other mutated only for G<sub>18</sub> (CEB1-G18T) (sequences see **Figure 45A**). Both alleles are stable in WT cells (**Figure 45B**). However, they are significantly and similarly destabilized upon Phen-DC<sub>3</sub> treatment (4,2% both, p-value vs. untreated =  $7,3e^{-3}$ ) or *PIF1* deletion (18,4 and 12,5%, p-value vs. WT =  $2,8e^{-9}$  and  $4,3e^{-7}$ )(**Figure 45B**). Still, these rearrangement frequencies are significantly lower than what is measured with CEB1-WT in all pairwise comparison (Phen-DC<sub>3</sub>-treated WT cells, p-values =  $7,8e^{-3}$  both; *pif1* $\Delta$  cells, p-values =  $2,4e^{-7}$  and  $4,4e^{-10}$ , respectively). Thus, the G-doublet G<sub>17-18</sub> involved in forms 1 and 3 is not neutral, and likely participates in a subset of the CEB1 instability.

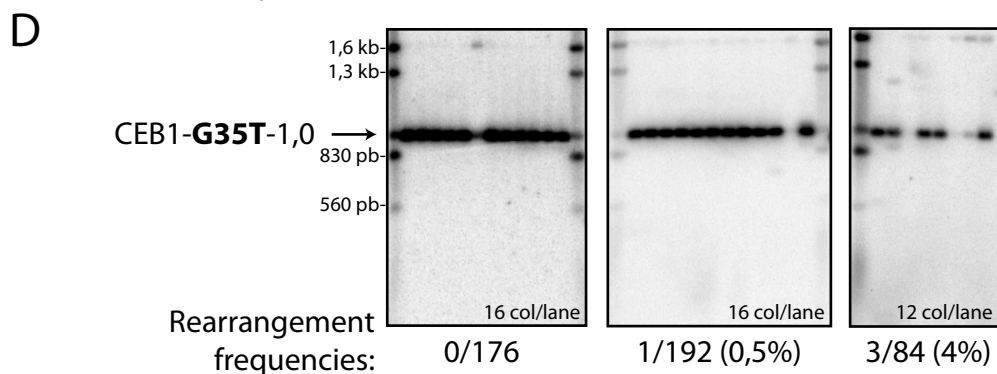
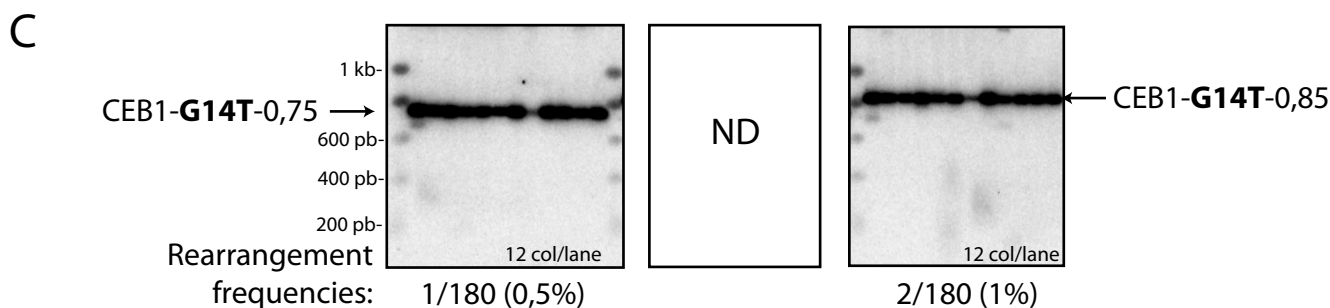
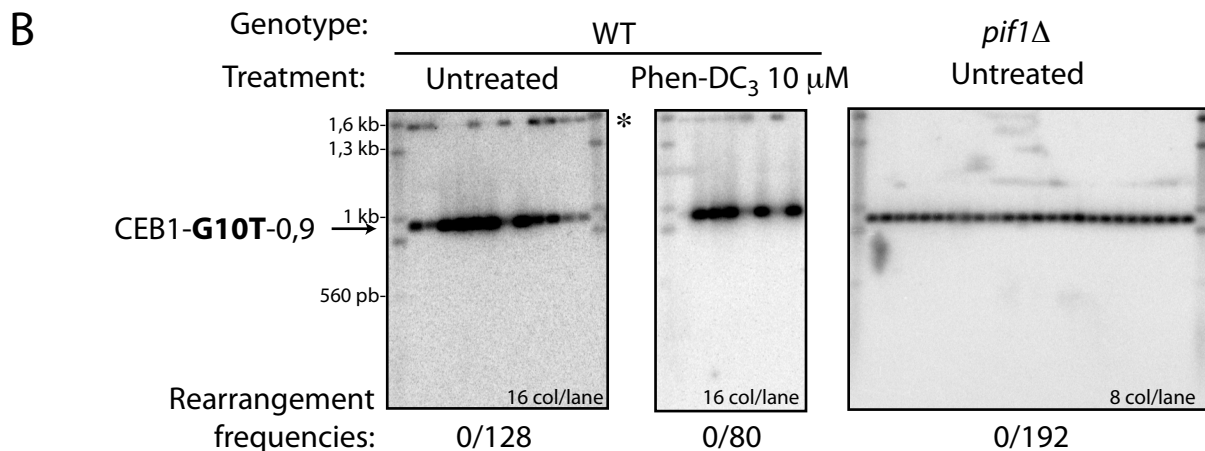
### *II.E.2.b.iii. The 6 Gs puzzle*

**A**

n-1 | CEB1 motif

33 34 35 36 37 38 39 | 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39

CEB1-WT: TGGGCTGAGGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTG  
 CEB1-Gmut: TGGCCTGAGCGCGGAGTGAGAGTGGCCTGCGGAGGTCCCTGGCCTG  
 CEB1-G10T: TGGGCTGAGGGGGGAGTGGAGGGTGGCCTGCGGAGGTCCCTGGGCTG  
 CEB1-G14T: TGGGCTGAGGGGGGAGGGAGTGTGGCCTGCGGAGGTCCCTGGGCTG  
 CEB1-G35T: TGTGCTGAGGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGTGCTG



**Figure 44: Instability of the CEB1 alleles mutated for their G-triplets.** (A) Sequences of the WT motif of CEB1, of the 5Gmut allele (used in the **Articles 1-4**), and of the alleles mutated for a single G in each of the G-triplet of CEB1 (red). (B) Southern blot analysis of the instability of CEB1-G10T-0.9 in a WT strain (ORT7130-3) treated or not with Phen-DC3, or in a *pif1*Δ strain (ORT7136-2). \* partially digested DNA. Lanes containing too few DNA due to incomplete DNA digestion (not shown on the gel) have been removed from the analysis. (C) Southern blot analysis of the instability of CEB1-G14T-0.75 in a WT strain (ORT7199-2) and of CEB1-G14T-0.85 in a *pif1*Δ strain (ORT7300-12). (D) Southern blot analysis of the instability of CEB1-G35T-1.0 in a WT strain (ORT7122-3) treated or not with Phen-DC3, or in a *pif1*Δ strain (ORT7126-1). (B-D) All minisatellites have been inserted in the orientation I near *ARS305*. Digestion: EcoRI.



The run of 6 Gs is predicted to be fully required for the formation of all G-quadruplexes identified: replacement by 6 T abolishes the G-quadruplex formation *in vitro* (data not shown). An allele in which the 4 central Gs of the 6G-run have been mutated (CEB1-G<sub>3-6</sub>T, sequence see **Figure 46A**) confirm this prediction: it remained fully stable in all conditions (Phen-DC<sub>3</sub> and *pif1Δ*)(**Figure 46B**). We next started to examine the requirement for individual Gs or combination of Gs into the 6G-run. To test if the involvement of this run of 6Gs is just to provide a 4<sup>th</sup> G-triplet in addition to the 3 previously described, we constructed three CEB1 alleles: two that leave a G-triplet (CEB1-G4T and CEB2,3,7T), and one that leave 4 Gs but no G-triplets (G4,6T)(**Figure 46A**).

The mutation of the single G<sub>4</sub> (CEB1-G4T), does not prevent the instability of CEB1: it is significantly destabilized upon Phen-DC<sub>3</sub> treatment (7,8%, p-value vs. WT = 4e<sup>-4</sup>) and in the absence of Pif1 (44%, p-value vs. WT = 1,7e<sup>-14</sup>)(**Figure 46B**), not significantly different from what has been measured with CEB1-WT (p-value vs. WT = 0,23 and 0,51, respectively). Thus, G<sub>4</sub> is not involved in any form responsible for the CEB1 instability, and indicate that the run of 6G can be interrupted without consequence for the destabilization of CEB1.

To test if the remaining G-triplet (G<sub>5-7</sub>) in the allele CEB1-G4T is involved in the structure destabilizing CEB1, we designed the CEB1-G4,6T allele (**Figure 46A**). Treatment with Phen-DC<sub>3</sub> did not destabilize it (0/192)(**Figure 46B**). Thus, the G-triplet G<sub>5-7</sub> is essential for the CEB1 instability.

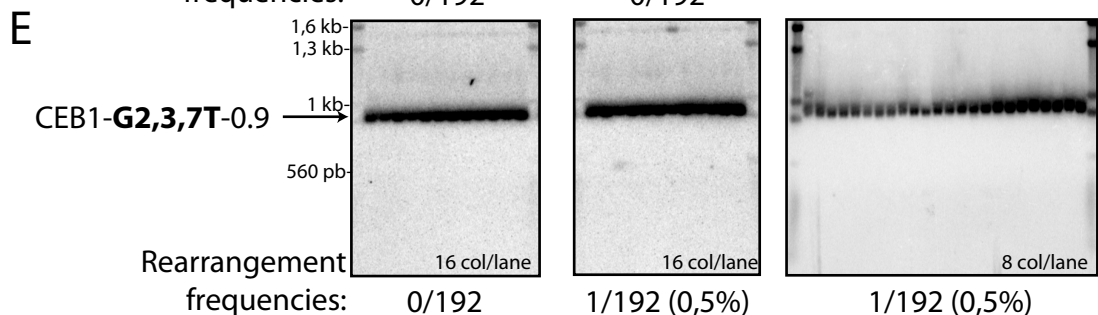
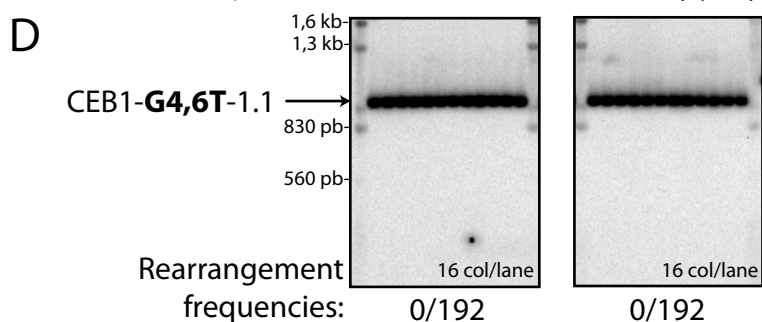
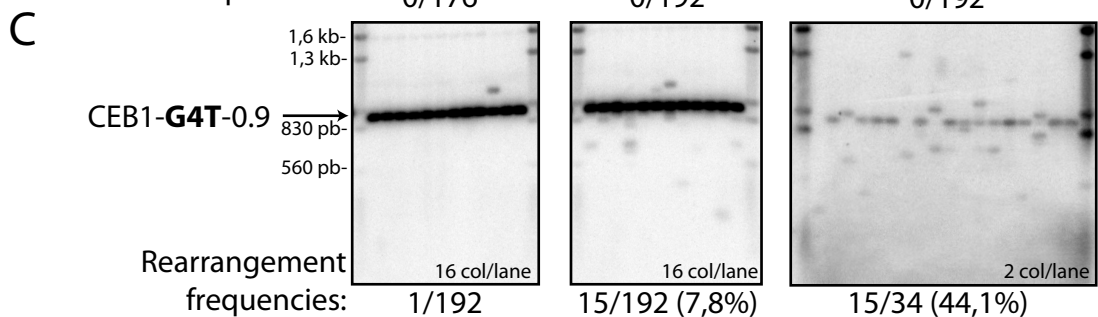
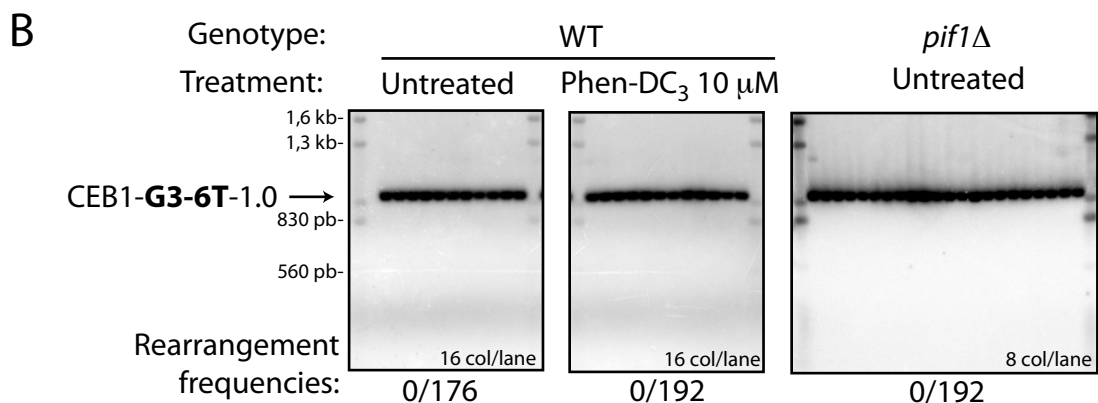
If a G-triplet in the run of 6Gs is required, is it sufficient? Indeed, the G4T mutation leaves a G-doublet in addition to a G-triplet, and a G-doublet have been implicated in a subset of the CEB1 instability (**Figure 45B**). To answer this question, we constructed the CEB1-G2,3,7T allele, which leaves only a centered G-triplet in the run of 6Gs (**Figure 46A**). Strikingly, this minisatellite is stable in both WT-treated cells (0/192) and in *pif1Δ* cells (0,5%, 1/192)(**Figure 46B**). It suggests that a single G-triplet in the run of 6Gs is not sufficient the destabilization of CEB1 or that the position of the G-triplet in the run of 6 Gs is critical for the formation of a destabilizing G-quadruplex (shifted of 1 nt in the CEB1-G2,3,7T allele compared to CEB1-G4T). To distinguish between these two possibilities we will synthesize a CEB1-G2-4T allele bearing a single G-triplet at the same position than in the CEB1-G4T allele.

**A**

n-1 | CEB1 motif

33 34 35 36 37 38 39 | 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39

CEB1-WT: TGGGCTGAGGGGGGAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTG  
 CEB1-G3-6T: TGGGCTGAG**TTTT**GAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTG  
 CEB1-G4T: TGGGCTGAG**GGTGGG**AGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTG  
 CEB1-G4,6T: TGGGCTGAG**GGTGT**GAGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTG  
 CEB1-G2,3,7T: TGGGCTGAG**TTGGGT**AGGGAGGGTGGCCTGCGGAGGTCCCTGGGCTG

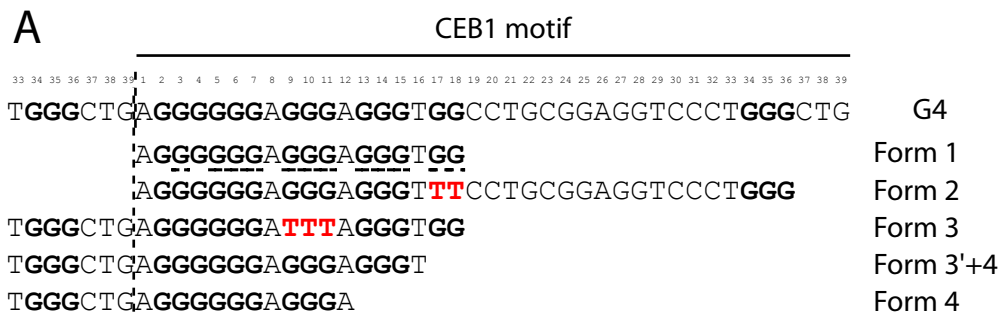


**Figure 46: Instability of the CEB1 alleles mutated for the 6 contiguous G.** (A) Sequences of the CEB1 motifs mutated for the run of 6 Gs (red). (B) Southern blot analysis of the instability of CEB1-G3-6T-1.0 in a WT strain (ORT7158-1) treated or not with Phen-DC3, and in a *pif1*Δ strain (ORT7162-2). (C) Southern blot analysis of the instability of the CEB1-G4T-0.9 allele in a WT strain (ORT7122-3) treated or not with Phen-DC3, and in a *pif1*Δ strain (ORT7126-1). (D) Southern blot analysis of the instability of CEB1-G4,6T-1,1 in a WT strain (ORT7166-1) treated or not with Phen-DC3. (E) Southern blot analysis of the instability of CEB1-G2,3,7T-0.9 in a WT strain (ORT7159-1) treated or not with Phen-DC3, and in a *pif1*Δ strain (ORT7163-3). (B-E) All the minisatellites have been inserted in the orientation I near *ARS305*. Digestion: EcoRI.

### II.E.2.c. Does several G-quadruplexes cooperate to induce the instability of CEB1?

The mutagenesis approach allowed narrowing the G-runs involved in the formation of the G-quadruplex(es) responsible for the CEB1 instability *in vivo*. These results are summarized **Figure 47B**. They show that all the G-triplets are required for CEB1 instability to occur. Unexpectedly, they also show that a G-doublet participates in the CEB1 instability, although to a lower extent than G-triplets. Finally, they show that all the Gs in the run of 6 Gs are not required to destabilize CEB1. The number and the position of the Gs required in this run of 6 Gs remains to be deciphered.

Can we rationalized this mutagenesis approach with the G-quadruplex conformations determined *in vitro* (**Figure 47A**)? First, the G-triplet mutation reveals that the G-quadruplexes 1, 3, or 4 alone are not able to induce CEB1 instability (alleles CEB1-G35T, CEB1-G10T, and CEB1-G14T, respectively)(**Figure 47B**). Furthermore, the CEB1-G4T allele that behaves like CEB1-WT revealed that form 4 is dispensable to CEB1 instability. Rather, forms 2 and/or 3' that are eliminated by each G-triplet mutations are the forms suspected to induce CEB1 instability. However, the CEB1-G2,3,7T allele, theoretically able to fold into form 2 and/or 3', is stable. In addition, none of the forms 2 and 3' requires the G-doublet G<sub>17,18</sub>. However, this G-doublet (involved only in forms 1 and 3) participates in a subset of the CEB1 instability. Hence, trying to connect the mutagenesis results with the G-quadruplex structures determined *in vitro* reveals incompatibilities regarding the forms involved in the CEB1 instability. Consequently, none of the G-quadruplex conformations alone account for the CEB1-WT instability. Rather, it may results from the combination of several forms. Indeed, whereas the forms 2+3' (that cannot be distinguish by mutagenesis) or 2+3+3' are not sufficient by themselves to trigger the instability of CEB1 (see CEB1-G2,3,7T and CEB1-G4,6T, respectively), addition of form 1 to form 2+3+3' (CEB1-G4T) rescue a level of instability, not significantly different from what is observed with CEB1-WT. Likewise, addition of form 4 to forms 2+3' (CEB1-G18T and CEB1-G17,18T) results in a significant induction of the instability. This instability induction, although not as important as for the CEB1-WT allele, is synergistic. Thus, although the forms are independtly unable to induce CEB1 instability, the combination of forms can partially or fully rescue the level



**B**

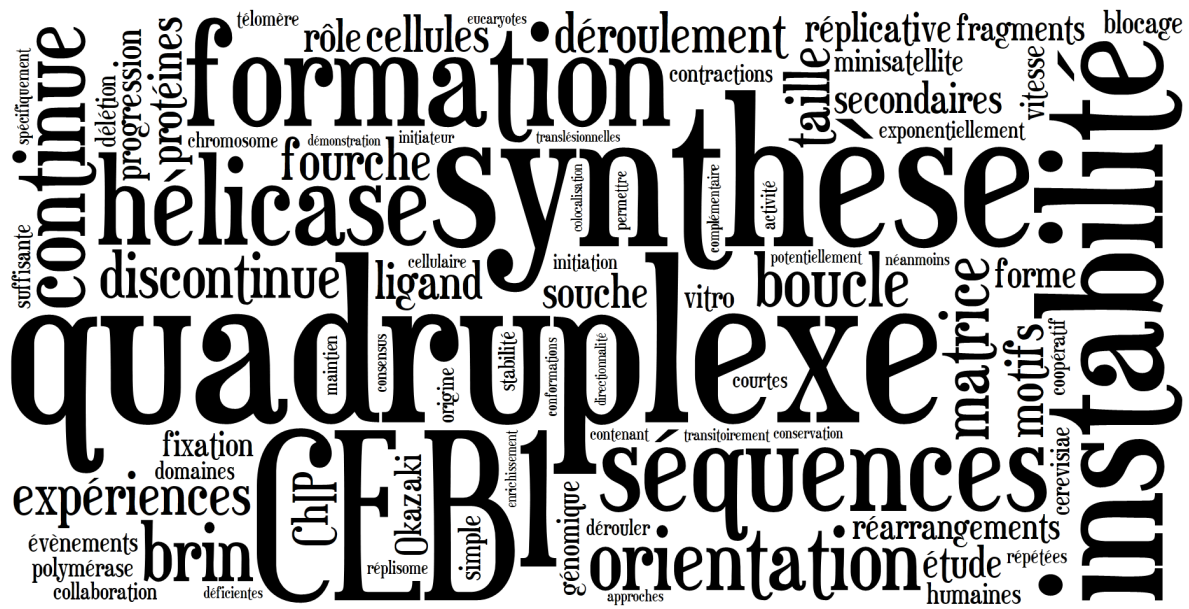
Mutation	Max. loop size (kb)	Allele size (kb)	Instability			Forms remaining	Forms affected
			WT	WT + Phen-DC3	<i>pif1</i> Δ		
WT	4	1	4/159 (2,5%)	23/192 (12%)	39/66 (59%)*	All	None
5G-mut	-	1,7	ND	ND	0/384	None	All
G10T	5	0,9	0/128	0/80	0/192	3	1,2,3',4
G14T	5	0,75	1/180 (0,5%)	ND	2/180 (1%)	4	1,2,3,3'
G35T	1	1	0/176	1/192 (0,5%)	3/84 (3,2%)	1	2,3,3',4
G18T	4	1,1	0/192	8/192 (4,2%)	16/87 (18,4%)	2,3,4	1,3
G17,18T	4	1,1	0/192	6-8/192	9-12/96	2,3,4	1,3
G3-6T	11	1	0/176	0/192	0/192	None	All
G4T	7	1	1/192 (0,5%)	15/192 (7,8%)	15/34 (44%)	1,2,3,3'	4
G4,6T	11	1,1	0/192	0/192	ND	(2,3,3')	1,4
G2,3,7T	6	0,9	0/192	0/192	1/192 (0,5%)	2,3'	1,3,4

**Figure 47: Summary of the behavior of the CEB1 mutant alleles *in vivo*.** (A) Oligonucleotides able to form one or several G-quadruplexes *in vitro*. Five conformations can be distinguished: 1, 2, 3, 3', 4 (M. Adrian and AT Phan, personal communication). (B) Instability of several mutated CEB1 alleles in WT (+/- Phen-DC3 10 μM) and *pif1*Δ strains upon insertion of the arrays near the *ARS305* in the orientation I (G-strand template for the leading strand replication machinery). Mutation that **abolish**, **significantly decrease**, and **does not significantly decrease** the instability compared to the CEB1-WT array.



observed with the CEB1-WT allele. These results suggest that the CEB1 instability is the result of a cooperation of the various G-quadruplexes that it can form.

### III. DISCUSSION





### III. Discussion et perspectives

#### III.A. Formation des G-quadruplexes à la fourche de réplication

L'étude exposée dans l'**Article 3** révèle que les G-quadruplexes formés et non-déroulés au sein de CEB1 conduisent à son instabilité lors de la réplication. Un résultat important de cette étude réside dans la déstabilisation spécifique à l'orientation de CEB1 vis-à-vis de la direction de la réplication : le minisatellite n'est instable que lorsque le brin riche en G est la matrice de la machinerie de synthèse continue (modèle : **Article 3, Figure 5**). Ce résultat pose plusieurs questions fondamentales concernant les mécanismes d'instabilité génomique causés par des structures secondaires de l'ADN : Où et quand se forment les G-quadruplexes ? Se forment-ils sur la matrice de la machinerie de synthèse discontinue ? Quels mécanismes permettent le redémarrage de la synthèse sur ces brins ?

##### III.A.1. **Formation du G-quadruplexe entre l'hélicase répllicative et Pole**

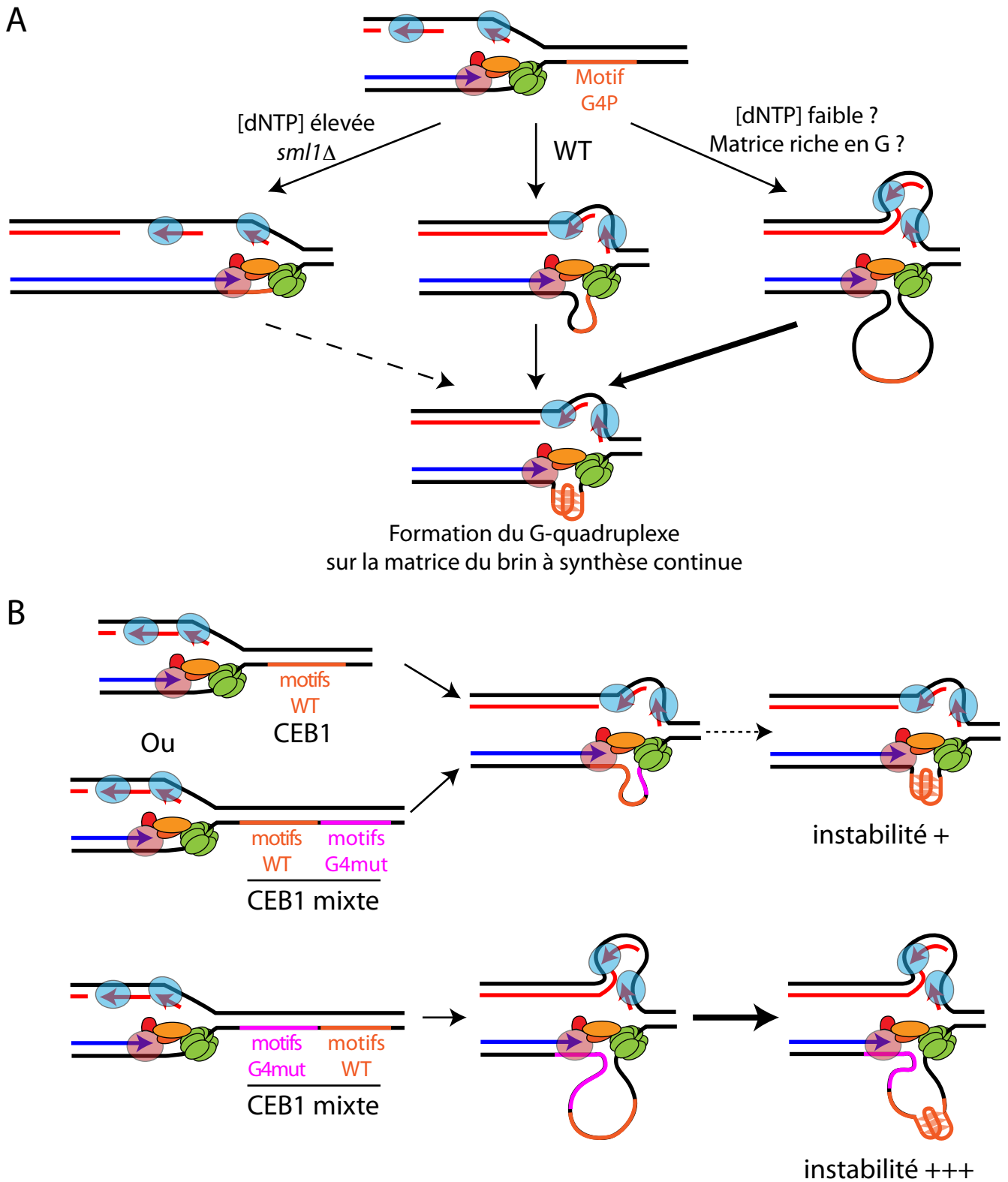
Dans un contexte cellulaire, la situation de compétition avec la forme double-brin de l'ADN défavorise la formation du G-quadruplexe, plus encore dans des conditions de superhélicité positive générées en aval de la fourche de réplication (**Chapitre I.A.1.c.iii**). La préformation d'un G-quadruplexe en amont de la fourche et sa persistance jusqu'à la machinerie de synthèse continue étant peu probable, le G-quadruplexe doit obligatoirement se former entre l'hélicase et la Pole. Ceci implique une taille d'ADNsb suffisante entre l'hélicase répllicative et le site actif de l'ADN Pole. Les données structurales de la Pole et du complexe MCM2-7 suggèrent une taille non-nulle entre le site d'ouverture du duplex et le site de synthèse, au moins au sein des protéines (**Figure 20, Chapitre I.B.1.b**). Cette taille d'ADNsb est minimale dans des conditions où la progression de la polymérase est limitée par la vitesse de progression de l'hélicase, et peut théoriquement augmenter si l'hélicase progresse plus rapidement que la polymérase, par exemple en présence d'HU (Sogo et al. 2002; Feng et al. 2006). Ainsi, la taille de matrice simple-brin pour la machinerie de synthèse continue, et donc l'opportunité pour un G-quadruplexe de se former, dépend des vitesses relatives de

déroulement et de synthèse. Dans une cellule WT, la vitesse de synthèse n'est pas optimale: elle peut être augmentée d'un facteur 2 lorsque la concentration intracellulaire de dNTP augmente, en dé-réprimant la RNR (Poli et al. 2012). Aussi est-il probable que cette carence (relative) en précurseurs nucléotidiques conduise à la formation de régions d'ADNsb non-négligeables entre la Pol $\epsilon$  et l'hélicase dans une souche sauvage (**Figure 48A**). Le fait que la délétion de *RRM3* cause une augmentation de la vitesse de réplication (Poli et al. 2012) et supprime l'instabilité observée dans le mutant *pif1 $\Delta$*  est en faveur de cette hypothèse (**Tables 3, Chapitre II.C.3.c**). L'effet d'une augmentation de la concentration de dNTP, et donc de la vitesse de synthèse, sur l'instabilité de CEB1 sera testé plus directement dans un mutant *sm11 $\Delta$*  ou en condition de sur-expression de la RNR. A l'inverse, l'effet d'une diminution de dNTP sera mesuré en présence d'HU dans une souche *pif1 $\Delta$* .

La richesse en GC des motifs de CEB1, largement supérieure à celle du génome (respectivement, 77% et 38%), pourrait également favoriser la formation d'ADNsb entre Pol $\epsilon$  et hélicase répliquative. En effet, des expériences de molécule unique *in vitro* montrent que la vitesse de progression des ADN polymérase diminue quand la richesse GC du substrat augmente (Manosas et al. 2012). De plus, l'instabilité des répétitions trinuécléotidiques (CAG/CTG) dans les cellules humaines est positivement corrélée à la richesse GC des régions flanquantes (Brock et al. 1999). Cette hypothèse peut être simplement testée en construisant des allèles hybrides de CEB1: une moitié constituée de motifs CEB1-Gmut (riche en G mais incapable de former des G-quadruplexes) et la seconde moitié de motifs CEB1-WT. La position respective des parties WT et Gmut vis-à-vis de la réplication est cruciale: une induction de l'instabilité de la partie CEB1-WT n'est attendue que lorsqu'elle est précédée de la partie Gmut (**Figure 48B**).

### **III.A.2. Découplage entre machinerie de synthèse continue et hélicase répliquative**

Que se passe-t-il une fois le G-quadruplexe formé sur la matrice de Pol $\epsilon$ ? La formation de G-quadruplexe ne semble pas associée à la pause de la fourche de réplication au niveau de CEB1 (**Article 3 Figure S4**), ni au niveau de séquences G4P uniques (Paeschke et al. 2011). Ceci suggère que la formation d'un G-quadruplexe à la fourche n'est pas suffisante pour bloquer la progression du réplisome. Des expériences

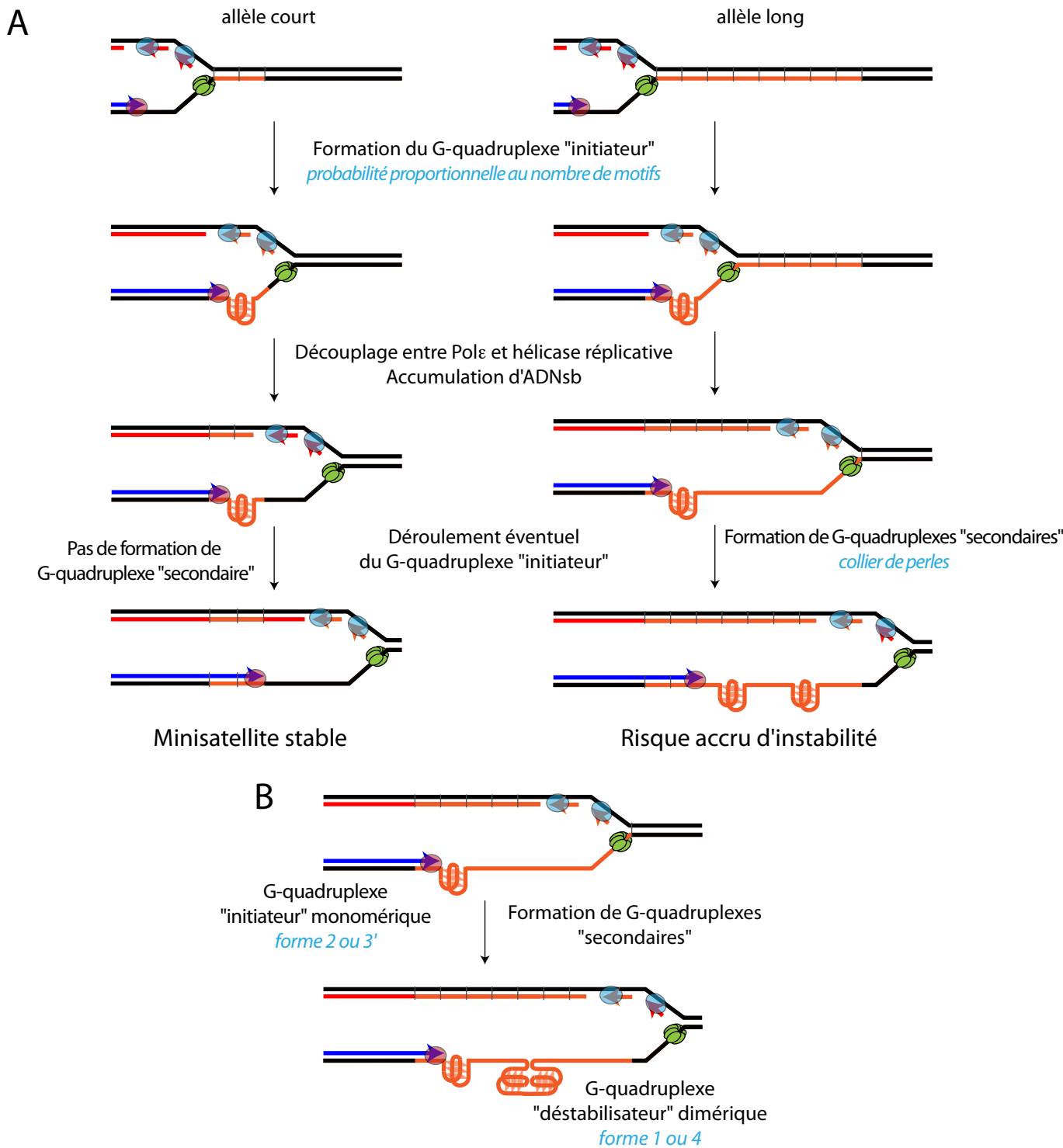


**Figure 48: L'accumulation de la matrice du brin à synthèse continue sous forme simple-brin favorise la formation de G-quadruplexe.** (A) Effet des facteurs agissant en *cis* (richesse GC) et en *trans* (concentration en dNTP) sur la vitesse de synthèse par Pole, et effet sur la taille de l'ADNsb présent sur la matrice du brin à synthèse continue. Plus l'ADNsb généré est long, plus la probabilité de former un G-quadruplexe est élevée (épaisseur des flèches). (B) Expérience envisagée pour tester si la richesse GC de CEB1 favorise la formation de G-quadruplexes. Un allèle mixte CEB1-WT/G4mut est comparé à un allèle CEB1-WT seul. La partie G4mut, qui conserve la richesse GC (77%), est ajoutée soit en amont (bas), soit en aval (milieu) de la partie CEB1-WT. Seul l'ajout en amont de la partie WT est censé causer une augmentation de l'instabilité en comparaison de l'allèle CEB1 normal (haut). Légendes de la fourche, voir Figure 19.

de ChIP-chip et ChIP-qPCR suggèrent un ralentissement de la progression de Pol2 au niveau de certaines séquences G4P chez la levure en absence de Pif1 (**Figure 50B**) (Paeschke et al. 2011). Cependant, l'absence de plusieurs expériences contrôles importantes dans cette étude (vérification de la formation *in vitro* de G-quadruplexe par les motifs déterminés *in silico* avec un consensus très libéral (Capra et al. 2010), mutagénèse, et traitement avec un ligand) empêchent de conclure définitivement quant au blocage de Pol $\epsilon$  sur un G-quadruplexe *in vivo*. Des expériences réalisées au laboratoire sont néanmoins en faveur de ce mécanisme. En effet, une forte accumulation d'ADNsb est détectée par ChIP-qPCR anti-RPA au cours de la phase S, uniquement dans des souches *pif1* $\Delta$  ou WT traitées par Phen-DC<sub>3</sub> et en présence de CEB1-1.8 dans l'orientation I, (J. Lopès, données non-montrées). Des expériences supplémentaires sont requises pour déterminer si l'accumulation de simple-brin concerne les deux molécules parentales, ou seulement le brin matrice de la synthèse continue. Néanmoins, ces résultats suggèrent fortement qu'une fois la Pol $\epsilon$  bloquée sur un G-quadruplexe, l'ADNsb s'accumule entre elle et l'hélicase réplivative : la machinerie de synthèse continue et l'hélicase réplivative sont découplées (**Figure 49A**). De tels découplages entre la machinerie de synthèse continue et hélicase en cas d'obstacle sur l'ADN matrice ont déjà été caractérisés chez la bactérie et la levure (**Chapitre I.B.1.c.ii**). De manière complémentaire, l'analyse par microscopie électronique de plasmides contenant CEB1 au cours de la réplication devrait permettre de mettre en évidence la formation asymétrique d'ADNsb à la fourche.

### **III.A.3.      Coopération entre motifs et aggravation des problèmes de réplication**

Une observation récurrente dans nos études concerne la relation plus que linéaire entre le nombre de motifs de CEB1 et son instabilité, que ce soit pour la formation d'expansions/contractions (**Article 1 Figure 3C ; Article 3 Figure 1E**) ou de GCR (**Article 4, Figure 2**). Cet effet, sous-estimé en raison de la méthode de comptage dans le cas des expansions/contractions et clairement exponentiel dans le cas des GCR, implique une collaboration entre motifs pour la formation des lésions à l'origine de ces réarrangements divers. Pourtant, l'étude de la formation de G-quadruplexes par des motifs de CEB25 en tandem *in vitro* suggère qu'un G-quadruplexe formé sur un motif n'affecte pas la formation d'un G-quadruplexe sur le motif adjacent : isolés par une



**Figure 49: Mécanismes putatifs de collaboration entre motifs en mesure d'expliquer l'effet de taille sur l'instabilité de CEB1.** (A) Le découplage entre Polε et l'hélicase répliquative consécutif au blocage de Pole sur un G-quadruplexe initiateur favorise la formation de G-quadruplexes "secondaires". Dans ce modèle, la probabilité de former le G-quadruplexe initiateur est proportionnelle au nombre de motifs. A cette probabilité s'ajoute celle de former les G-quadruplexes secondaires, fonction du nombre de motifs sous forme simple-brin. La probabilité d'induire une lésion (considérée proportionnelle au nombre de G-quadruplexes formés) augmente donc plus que linéairement avec la taille du minisatellite. (B) Mécanisme putatif de collaboration entre les différentes conformations de G-quadruplexe formés par CEB1. Le G-quadruplexe "initiateur" monomérique (formes 2 ou 3') nécessite peu d'ADNs pour se former, mais est facilement déroulé par Polε ou d'autres protéines redondantes à Pif1: il n'induit pas l'instabilité de CEB1, mais donne l'opportunité aux G-quadruplexes "déstabilisateurs" de se former. Ces G-quadruplexes déstabilisateurs sont les formes dimériques 1 et/ou 4. Elles nécessitent une quantité substantielle d'ADNs pour se former (2 motifs, soit environ 80 nt).



séquence « linker » neutre, ils se forment indépendamment les uns des autres dans l'ADNsb en un « collier de perles » observé par AFM (**Chapitre I.A.1.b.v, Figure 4C**) (Amrane et al. 2012). S'il n'existe pas de collaboration entre structures au sein de l'ADNsb, quels autres mécanismes pourraient être impliqués dans ce comportement coopératif ?

Le découplage entre la machinerie de synthèse continue et l'hélicase répliquative suite au blocage de Pol $\epsilon$  sur le G-quadruplexe « initiateur » pourrait constituer l'événement moléculaire à la base de cette coopération : l'accumulation d'ADNsb causée par le blocage de Pol $\epsilon$  sur le G-quadruplexe « initiateur », même s'il est à terme déroulé, augmente la probabilité pour les motifs suivants de former à leur tour un G-quadruplexe (**Figure 49A**). L'ADNsb généré pourrait voir s'y former un « collier de perle » de G-quadruplexes, comme observé *in vitro* pour CEB25 (**Figure 49A**). Plus l'incapacité de la cellule à dérouler les G-quadruplexes serait patente et plus le « collier de perles » fourni pourrait induire des lésions nombreuses à l'origine de réarrangements fréquents, complexes et courts, comme observé dans une souche *pif1 $\Delta$*  traitée par Phen-DC<sub>3</sub> (**Article 2, Figure 5**).

Une collaboration entre les différentes conformations de G-quadruplexe formées par CEB1 pourrait également participer à cette augmentation plus que linéaire du taux réarrangements ou de GCR. En effet, indépendamment, aucune des conformations caractérisées *in vitro* n'est suffisante pour induire l'instabilité de CEB1. En revanche, la combinaison de formes (2+3' et 4, ou 2+3' et 1) conduit à une augmentation synergique de l'instabilité de CEB1 (**Figure 47**) suggérant que ces différentes formes collaborent pour stimuler l'instabilité du minisatellite. La façon dont procède cette collaboration est inconnue, mais deux hypothèses me semblent envisageables.

Soit l'absence d'une forme n'est pas compensée par la formation d'une autre forme, ce qui résulte en une diminution nette du nombre de G-quadruplexes que peut former le minisatellite, et partant de l'effet coopératif.

Soit deux types de G-quadruplexes sont formés au sein de CEB1. Certains G-quadruplexes « initiateurs » se forment, inhibent transitoirement Pol $\epsilon$ , mais sont déroulés même en absence de Pif1. Ils ne nécessitent pas d'employer une voie de template-switch pour être contournés. Cependant, l'accumulation transitoire de

simples-brins autorise la formation d'une conformation « déstabilisatrice » qui, persistante, requiert un redémarrage de la synthèse par template-switch. Ces G-quadruplexes « déstabilisateurs » pourraient être les G-quadruplexes dimériques (formes 1 et 4), dont la formation nécessite au moins deux motifs sous forme simple brin (**Figure 49B**).

#### **III.A.4. Les G-quadruplexes se forment-ils sur la matrice du brin à synthèse discontinue ?**

La faible mais significative instabilité mesurée avec CEB1 orientation II dans des souches WT traitées avec Phen-DC<sub>3</sub> ou des souches *pif1Δ* est-elle bien due à la formation de G-quadruplexes sur le brin matrice de la machinerie de synthèse discontinue ? Ou est-elle due à une instabilité qui survient en réalité au niveau d'une machinerie de synthèse continue, émanant à faible fréquence d'*ARS306* ? Nous avons encore peu d'informations pour trancher entre ces deux possibilités, mais un résultat récent apporte une première indication : la délétion de *SML1* réduit fortement l'instabilité de CEB1 orientation I en présence de Phen-DC<sub>3</sub>, alors qu'elle n'a pas d'effet sur l'instabilité de CEB1 dans l'orientation II (J. Lopes, résultat non-montré). Ce résultat suggère que l'instabilité observée dans les orientations I et II dans une souche WT traitée n'est pas causée par la formation d'un G-quadruplexe ayant en définitive lieu uniquement sur le brin matrice de la synthèse continue. Une expérience réciproque consisterai à sous-exprimer Pol $\alpha$ , censée diminuer le nombre d'initiation de fragments d'Okazaki, et donc augmenter leur taille et la taille d'ADNsb générée entre eux (Gordenin et al. 1993; Callahan et al. 2003; Lemoine et al. 2005). Le résultat attendu est une augmentation de l'instabilité de CEB1 spécifiquement dans l'orientation II.

Dans ce cas, comment expliquer qu'en présence d'une quantité substantielle d'ADNsb formée entre deux fragments d'Okazaki et propice à la formation de G-quadruplexes, l'instabilité de CEB1 soit si faible dans l'orientation II ? Je pense que cette stabilité inattendue provient d'une protection intrinsèque du mode de synthèse discontinue au processus d'aggravation et de découplage, responsables de la forte instabilité observée dans l'orientation I. En effet, même si un G-quadruplexe se forme entre deux fragments d'Okazaki, l'inhibition de la polymérisation qui en découle ne favorisera pas la formation d'un G-quadruplexe en aval, où se trouve déjà un ADNdb

répliqué (le fragment d'Okazaki précédent), ni en amont, la synthèse d'un primer ARN par Pol $\alpha$  et de son élongation par Pol $\delta$  étant indépendante d'un blocage de synthèse d'un autre fragment d'Okazaki (Pages and Fuchs 2003; Lopes et al. 2006b). Ainsi, en absence de ce mécanisme coopératif propre à la machinerie de synthèse continue, CEB1 dans l'orientation II se comporte comme une succession indépendante de CEB1 de 165 nt (quatre motifs) dans l'orientation I. L'instabilité de CEB1 en fonction de sa taille (en pb) dans l'orientation II est donc proportionnelle à l'instabilité d'un allèle d'environ 165 pb.

Cette différence intrinsèque aux modes de synthèse continue et discontinue pourrait constituer la base mécanistique de la plus grande tolérance aux séquences G4P de la machinerie de synthèse discontinue. Elle a des implications évidentes pour expliquer la conservation du positionnement du brin riche en G des répétitions télomériques chez pratiquement tous les organismes eucaryotes connus à l'heure actuelle (Gilson and Geli 2007; Tran et al. 2010). Au-delà des G-quadruplexes, cette plus grande tolérance de la machinerie de synthèse discontinue a des implications pour les séquences capables de former d'autres types de structures secondaires et invite à revisiter les modèles classiques d'instabilité génomique.

### **III.A.5. Implications pour l'instabilité génomique causée par d'autres structures secondaires**

Les modèles traditionnels d'instabilité génomique par des structures secondaires sont principalement issus de l'étude de séquences capables de former des structures triplex (GAA<sub>n</sub>) (Mariappan et al. 1999) et « hairpins »/cruciformes (CNG<sub>n</sub> et palindromes) (Mirkin 2007). La plupart de ces modèles postulent que la structuration de l'ADNsb de la matrice du brin discontinu, entre deux fragments d'Okazaki, est à l'origine des défauts de réplication et de l'instabilité de la séquence sous-jacente. Cet *a priori* raisonnable n'a cependant jamais été démontré. Contrairement à ces structures, la formation d'un G-quadruplexe est sans contrepartie possible sur le brin complémentaire. Notre démonstration de l'implication d'une structure secondaire de l'ADN dans l'instabilité de séquences répétées spécifiquement sur le brin matrice de la machinerie de synthèse continue lors de la réplication, par des approches de mutagénèse et d'interférence chimique, est à ma connaissance sans précédent dans le domaine. Ces résultats, ainsi que certaines données de la littérature dont les modèles

actuels ne rendent pas compte, invitent à un changement paradigmatique concernant l'instabilité génomique induite par des structures secondaires lors de la réplication.

Premièrement, l'instabilité et la fragilité de ces séquences augmentent exponentiellement avec leur taille, de manière similaire à ce qui est observé avec CEB1 (Freudenreich et al. 1998; Shishkin et al. 2009). Aucun mécanisme n'a été proposé pour expliquer cet effet de taille à partir de structures formées sur la matrice du brin à synthèse discontinue. Le phénomène d'aggravation par accumulation d'ADNsb au niveau de la machinerie de synthèse continue pourrait également être à l'origine de cet effet exponentiel de l'instabilité pour ces autres séquences répétées.

Deuxièmement, les pauses de réplication observées au niveau des séquences palindromiques (**Figure 23**) (Samadashwily et al. 1997; Lobachev et al. 1998; Balakumaran et al. 2000; Pelletier et al. 2003; Zhang and Freudenreich 2007; Voineagu et al. 2008; Kerrest et al. 2009; Voineagu et al. 2009; Anand et al. 2012) ou triplex (Krasilnikova and Mirkin 2004; Kim et al. 2008; Shishkin et al. 2009; Tang et al. 2011; Chandok et al. 2012), généralement associées à l'instabilité de ces séquences, suggère que ces phénotypes sont les deux faces d'un même mécanisme, initié par la formation d'une structure secondaire. Cette association entre pause de réplication et instabilité constitue une différence notable avec ce qui est observé au niveau des séquences G4P (**Article 3**) (Paeschke et al. 2011). Des couplages entre machinerie de synthèse discontinue et continue, puis entre machinerie continue et hélicase réplivative ont été invoqués pour expliquer l'arrêt de la fourche après formation d'une structure secondaire sur le simple-brin entre deux fragments d'Okazaki (Voineagu et al. 2008; Voineagu et al. 2009). C'est ignorer la capacité de ré-initiation intrinsèque à la machinerie de synthèse discontinue et l'absence de couplage entre machineries de synthèses continue et discontinue (**Chapitre I.B.1.c.ii**). Contrairement aux G-quadruplexes, ces structures peuvent se former à partir d'ADNdb. La pré-formation de la structure dans l'ADNdb parental et sa persistance à l'approche de la fourche une fois liée par une protéine piégée, comme le complexe MutS $\beta$  (Msh2-Msh3) sur les « hairpins » imparfaits CNG/GNC (Owen et al. 2005; Lang et al. 2011) est probablement responsable des pauses du réplisome. Ce blocage s'apparente à ceux observés au niveau de barrières protéine-ADN, comme en présence de Fob1 au site RFB du rDNA (Brewer et al. 1992;

Kobayashi et al. 1998; Ivessa et al. 2002), ou Lacl sur des répétitions de LacO (Dubarry et al. 2011).

En résumé, alors que le G-quadruplexe ne serait qu'inhibiteur des machineries de synthèse (principalement la machinerie de synthèse continue), les structures de type « hairpin »/cruciforme et triplex inhiberaient deux composants du réplisome : la machinerie de déroulement pour les structures formées dans l'ADNdb parental, et une machinerie de synthèse lors de la formation dans l'ADNsb à la fourche. Cette inhibition de la machinerie de synthèse se ferait suivant des modalités similaires à ce qui est observé pour les G-quadruplexes : leur formation sur la matrice de la machinerie de synthèse continue conduirait à une instabilité qui augmente exponentiellement avec la taille du minisatellite.

### **III.B. Reprise de la synthèse et formation des réarrangements génomiques au niveau de CEB1**

Plusieurs mécanismes peuvent ré-établir une synthèse continue processive (**Chapitre I.B.3.b**), mais seul un mécanisme qui utilise l'information de séquence d'une autre molécule est à même de générer des expansions ou des contractions d'une séquence répétée en tandem (**Chapitre I.B.3.b.iii**). En absence de cassure visible par gel d'électrophorèse en champ pulsé, les expansions/contractions de CEB1 sont probablement produites par un mécanisme de template-switch Rad51- et Rad52-dépendant (c'est-à-dire sans régression de fourche) initiés à partir d'une région simple-brin directement à la fourche, ou laissée en amont de la fourche après un évènement de ré-initiation. Ainsi, le blocage de la synthèse continue imposée par le G-quadruplexe peut être contourné par un mécanisme de template-switch. Est-ce à dire que le template-switch est la seule voie apte à lever ce blocage ? L'effet inhibiteur de Rev1 sur la formation des réarrangements de CEB1, et la mutagénèse d'une séquence G4P sur plasmide observée dans des cellules déficientes pour Pif1 (Paeschke et al. 2011), suggère que des polymérases translésionnelles pourraient également être impliquées dans la répllication des séquences formant des G-quadruplexes (**Figure 34, Chapitre I.C.3.b**). Cette voie du PRR, dite « error-prone », pourrait contribuer à limiter la formation de réarrangements chromosomiques médiés par des séquences répétées (**Chapitre I.B.4**).

Qu'en est-il des GCR, qui sont soit des additions de télomères (orientation G), soit des évènements de translocation non-réciproques (orientation C) au niveau de CEB1 ? Quelle que soit l'orientation, une majorité de ces évènements dépend des protéines Rad51 et Rad52 (**Article 4**) suggérant, malgré leurs différences de nature, que ces évènements sont produits par un mécanisme similaire. En effet, la seule séquence susceptible de générer une invasion ectopique est CEB1, sur un télomère. Dans l'orientation G, un évènement de BIR conduira à l'ajout d'un télomère qui stabilisera immédiatement le chromosome. Dans l'orientation C, le BIR sur un télomère conduira à la copie du chromosome, à la formation d'un chromosome dicentrique, et à une instabilité génomique chronique avant stabilisation du chromosome (Paek et al. 2009). Les réarrangements supplémentaires consécutifs aux cassures du chromosome dicentrique font principalement intervenir les abondantes séquences répétées dispersées de type Ty1 (**Article 4, Figure S5**)(Putnam et al. 2009; Chan and Kolodner 2011).

### **III.C. Rôle de Pif1 dans la réplication des séquences G4P et le maintien de la stabilité du génome**

Nos études démontrent un rôle majeur de Pif1 pour la suppression des grands réarrangements chromosomiques au niveau du minisatellite CEB1. Pif1 agit sur ces séquences à la fois pour limiter la formation de lésions en déroulant les structures G-quadruplexes, et pour inhiber la formation de GCR par ajout d'un télomère au niveau de ces lésions (**Article 4**) (Schulz and Zakian 1994; Myung et al. 2001). Ainsi, un mutant *pif1Δ* contenant CEB1-WT exhibe un taux de GCR un million de fois supérieur à celui observé pour une souche WT sans minisatellite (**Article 4**).

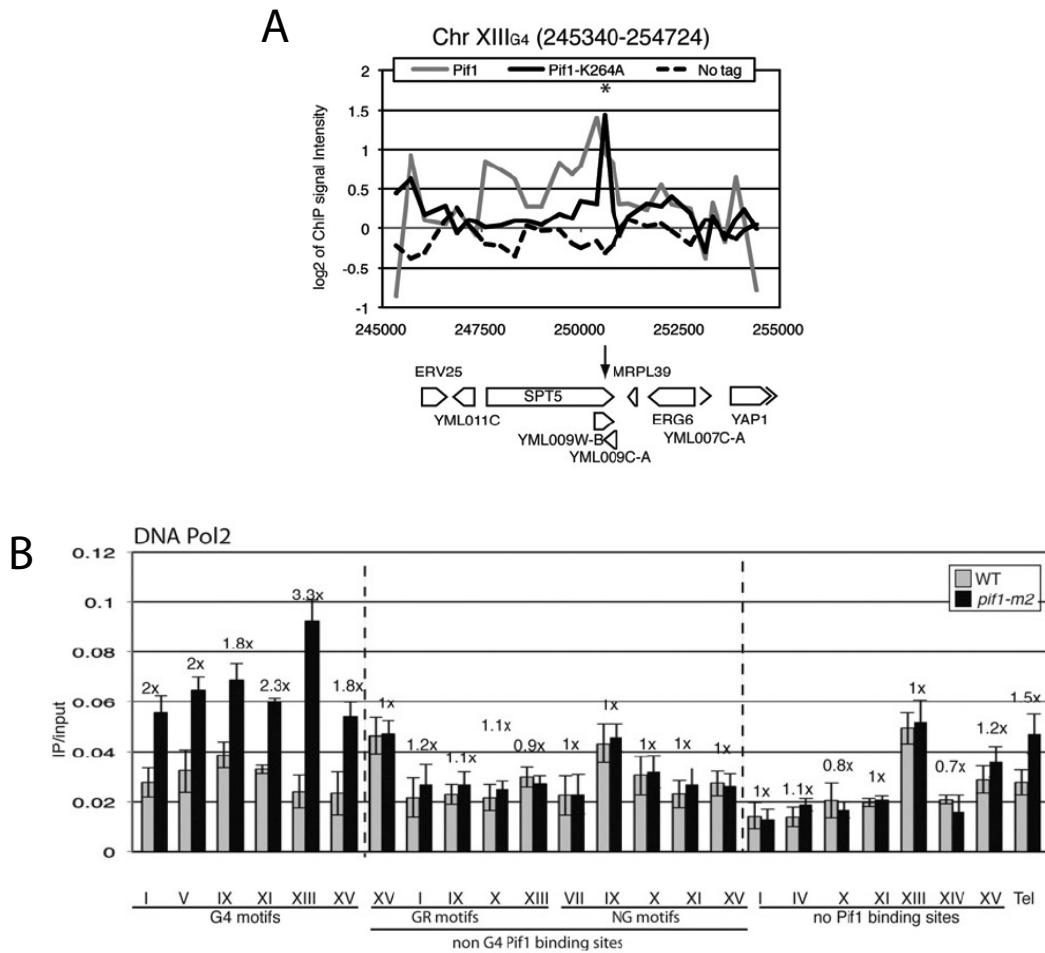
#### **III.C.1. Pif1 promeut la réplication des séquences G4P endogènes de *S. cerevisiae***

Une étude récente du laboratoire de VA Zakian (Paeschke et al. 2011) a identifié de nombreux sites de fixation préférentiels pour l'hélicase Pif1 au sein du génome de *S. cerevisiae* par des expériences de ChIP-chip: Pif1 et sa contrepartie dévolue d'activité translocase Pif1-K264A lient respectivement 1584 et 1123 régions du génome (**Figure 50A**). Les pics obtenus avec Pif1-K264A sont discrets, suggérant que le signal diffus obtenu avec la forme sauvage de Pif1 est dû à sa translocation autour de son site de

fixation initial. Le génome de levure compte peu de séquences G4P si l'on applique le consensus traditionnel  $G_{3-5}N_{1-7}G_{3-5}N_{1-7}G_{3-5}N_{1-7}G_{3-5}$  (**Chapitre I.A.3.a.i**). Aussi un consensus plus libéral (boucles  $N_{1-25}$ ) a été utilisé pour déterminer des séquences G4P potentielles (Capra et al. 2010). Sur l'ensemble des pics Pif1-K264A ( $\pm 500$  pb), 11% (128/1123) recourent un motif G4P. Cet enrichissement est significatif et équivaut à 25% (138/558) des séquences G4P liées par Pif1 (certains pics de Pif1 recouvrent plusieurs séquences G4P). Quel rôle Pif1 joue-t-il au niveau des séquences G4P? Nos études, et particulièrement l'**article 3**, révèle un rôle de Pif1 pour permettre la réplication fidèle de CEB1, sans formation d'intermédiaire de recombinaison à son niveau au cours de la phase S. En accord avec ces résultats, l'étude par Paeschke *et al.* (Paeschke et al. 2011) montre que dans une souche déficiente pour Pif1 (*pif1-m2*), Pol2 s'accumule au niveau des séquences G4P préférentiellement liées par Pif1 (**Figure 50B**). Cependant, les auteurs ne testent pas si le recrutement de Pif1 sur ces séquences dépend vraiment de leur capacité à former un G-quadruplexe (capacité à former un G-quadruplexe *in vitro*, mutagénèse du motif, traitement avec un ligand de G-quadruplexe). Des expériences de ChIP-qPCR préliminaires réalisées au laboratoire sont néanmoins en faveur de ce modèle : elles révèlent un enrichissement de Pif1 dans les régions flanquants directement CEB1 (la qPCR ne peut être réalisée dans CEB1 en raison de son caractère répété), uniquement quand CEB1 est dans l'orientation I et en présence du ligand Phen-DC<sub>3</sub>, c'est-à-dire des conditions où CEB1 perturbe la réplication (J. Lopès, non-montré). Ainsi, même si une démonstration formelle reste à fournir, les résultats de Paeschke *et al.* conjointement aux nôtres suggèrent qu'un des rôles de Pif1 est d'assurer la réplication et le maintien d'au moins une partie des séquences capables de former des G-quadruplexes chez *S. cerevisiae*.

### **III.C.2. Le rôle de déroulement des G-quadruplexes par Pif1 est-il conservé chez les eucaryotes ?**

Pif1 est une hélicase conservée chez les eucaryotes (Bochman et al. 2010). La conservation de l'activité de déroulement de G-quadruplexe chez l'homologue humain hPif1 suggère une conservation du rôle biologique de déroulement des G-quadruplexes *in vivo* (Sanders 2010). Cependant, des expériences de fractionnement de lysat cellulaire (HeLa) montrent que la fraction contenant l'hélicase 3'-5' G4R1, mais pas hPif1, porte l'essentiel de l'activité de déroulement des G-quadruplexes intra- et inter-moléculaires



**Figure 50: Pif1 est enrichi au niveau de certaines séquences G4P, et limite l'accumulation de Pol2 spécifiquement à ces sites. (A) Profil de ChIP-chip anti-Myc sur une souche Pif1-Myc (gris), Pif1-K264A-Myc (noir) ou Pif1 non taggée (pointillés) au niveau d'une région du chromosome XIII. La flèche indique la position d'une séquence G4P. (B) Augmentation de la présence de Pol2 dans le mutant *pif1-m2* spécifiquement au niveau de séquences G4P normalement enrichies en Pif1.**

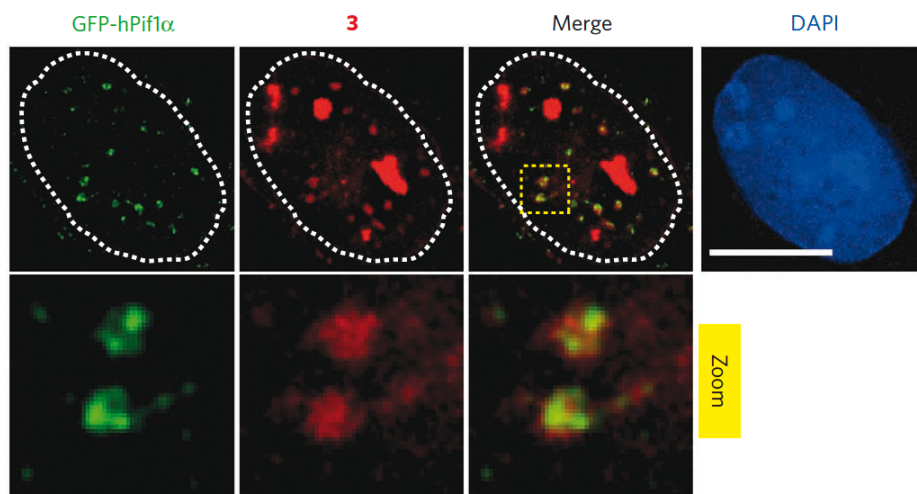
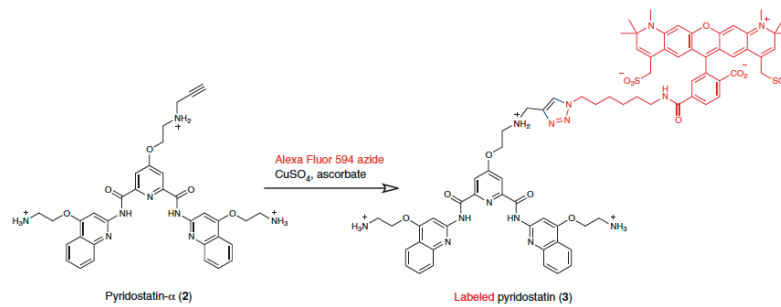


(**Chapitre I.A.3.c.ii**) (Vaughn et al. 2005; Giri et al. 2011). Ceci n'exclut pas pour autant un tel rôle pour hPif1, dont la directionnalité opposée à G4R1 pourrait conférer une spécificité d'action au cours de la réplication (voir **Chapitre III.D**).

Récemment, le traitement de cellules humaines immortalisées (U2OS) avec de la pyridostatine, un ligand de G-quadruplexe dont la structure est proche du bisquinolinium 360A (Muller et al. 2010), induit la formation de foyers  $\gamma$ H2AX dans les cellules humaines. Des expériences de ChIP-seq anti- $\gamma$ H2AX montrent que ces foyers sont formés principalement dans le corps de gènes enrichis en séquences G4P (Rodriguez et al. 2012). Ce ligand est fonctionnalisé : un groupement fluorescent peut y être attaché par « click-chemistry » (**Figure 51**) (Muller et al. 2010). Ce marquage sur cellules fixées révèle une colocalisation (qualifiée de « très forte » par les auteurs mais non-quantifiée), de hPif1-GFP avec le ligand quand les cellules vivantes sont traitées avec le ligand, puis fixées, ou quand le ligand est ajouté après fixation des cellules (**Figure 51**). Ce résultat suggère que le ligand et hPif1 ciblent les mêmes régions dans les cellules humaines. La quantification de la colocalisation et une expérience de ChIP-chip anti-GFP similaire à celle réalisée avec  $\gamma$ H2AX auraient néanmoins permis de solidifier ce résultat d'intérêt. Ces observations encouragent à tester le rôle de déroulement des G-quadruplexes par Pif1 chez l'homme, en essayant de reproduire les résultats d'instabilité de CEB1 (ou d'autres minisatellites) dans des souches déficientes pour Pif1 ou traitée par un ligand pendant plusieurs générations.

### **III.C.3. Pif1 est-elle la seule hélicase impliquée dans le maintien des séquences G4 chez *S. cerevisiae* ?**

La comparaison des instabilités par génération suggère que le traitement par Phen-DC3 (8 générations de traitement) déstabilise plus CEB1 que l'absence de Pif1 (30-40 générations). Ce résultat suggère qu'une plus grande quantité de G-quadruplexes sont formés et persistent en présence du ligand qu'en absence de Phen-DC<sub>3</sub>. Cette instabilité supérieure pourrait-elle être due à l'inhibition du déroulement des G-quadruplexes de CEB1 par d'autres protéines que Pif1 ? Nous avons exploré cette hypothèse en mutant d'autres hélicases qui déroulent ou dont un homologue déroule les G-quadruplexes *in vitro* (Sgs1, Dna2, et Chl1)(**Chapitre I.A.5.b**), et Rrm3, l'autre hélicase de la famille Pif1 dont les activités biochimiques ne sont pas connues. Aucun de ces



**Figure 51: Le ligand de G-quadruplexe pyridostatin et hPif1 colocalisent dans des cellules humaines.** En haut, le ligand pyridostatin et sa forme taggée fluorescente (composé 3). En bas, images de microscopie de fluorescence à haute résolution de noyaux en présence du ligand taggé (rouge) et de hPif1-GFP (vert).

mutants ne conduit à l'instabilité de CEB1 et, à l'exception de Sgs1 dont l'interprétation est encore peu claire, ils n'induisent pas une instabilité supplémentaire en absence de Pif1 ou en présence de Phen-DC<sub>3</sub>. Alternativement, cette instabilité plus importante en présence de Phen-DC<sub>3</sub> pourrait être due à la capacité du ligand, similaire en structure et en propriétés stabilisatrices à 360A, à catalyser la formation de G-quadruplexes qui, en son absence, ne se seraient pas formés *in vivo* (De Cian and Mergny 2007). Aussi, il n'existe à l'heure actuelle aucune preuve de l'existence d'une autre hélicase que Pif1 en mesure de dérouler les G-quadruplexes impliqués dans l'instabilité de CEB1 chez *S. cerevisiae*.

### **III.D. Structure des G-quadruplexes et spécificité des hélicases pour le maintien des séquences sous-jacentes**

#### **III.D.1. Différences d'instabilité entre les allèles naturels et artificiels de CEB1**

Dans nos deux tests (expansion/contraction et GCR) l'allèle naturel de CEB1 induit moins de réarrangements que l'allèle artificiel. Ces résultats suggèrent que l'allèle naturel induit moins de lésions que l'allèle artificiel. L'étude de la relation structure-fonction des G-quadruplexes à l'instabilité génomique suggère que des polymorphismes de séquence entre allèle naturel et allèle artificiel seraient à l'origine de ces différences quantitatives d'instabilité *in vivo*:

- Premièrement, la mutagenèse ponctuelle de CEB1 montre que le triplet G<sub>13-15</sub> est essentiel à l'instabilité du minisatellite (**Chapitre II.E.2.b.i**). Ce triplet de G est muté dans 9 motifs sur 42 dans l'allèle naturel (**Figure 30B**).

- Deuxièmement, nous avons montré que la taille de la boucle au sein du G-quadruplexe est un déterminant important de l'instabilité du minisatellite *in vivo* (**Chapitre II.E.1**). L'allèle artificiel de CEB1 est la mise en tandem, sans polymorphisme, du motif le plus représenté dans CEB1-1.8 (exemple: motifs 2-4, **Figure 30B**) et qui compte parmi les motifs les plus courts (39 nt). Un polymorphisme de l'allèle naturel qui concerne 14 motifs sur 42 est l'ajout d'un triplet de G dans la séquence G<sub>4P</sub> du motif (exemple: motif 12 **Figure 30B**). *In vitro*, un oligonucléotide contenant cinq répétitions de la séquence télomérique humaine voit le triplet de G surnuméraire inclus dans une boucle (Yue et al.

2011). Ainsi, les motifs de CEB1 naturels contenant un triplet de G surnuméraire ont la capacité de former des G-quadruplexes contenant une longue boucle, une situation stabilisatrice *in vivo*. Cette hypothèse pourrait être testée simplement en synthétisant un minisatellite artificiel à partir de ce motif contenant un triplet de G supplémentaire de l'allèle naturel.

Ainsi, il est probable que ces polymorphismes de séquence propres à l'allèle naturel de CEB1, qui concernent 21 motifs sur 42, lui confèrent une plus grande stabilité que l'allèle artificiel.

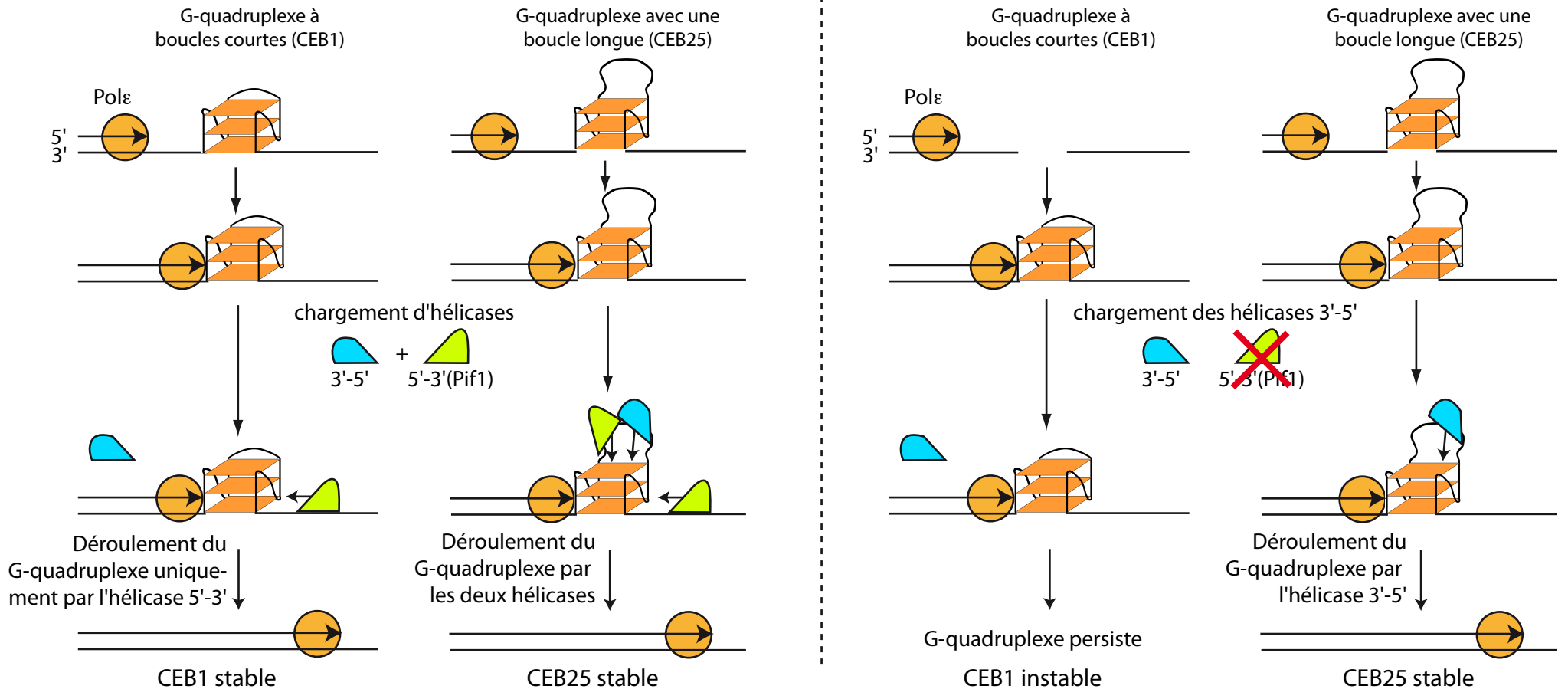
### **III.D.2. Un G-quadruplexe contenant une longue boucle n'induit pas l'instabilité du minisatellite**

L'inhibition de la progression de Pol $\epsilon$ , proposée au **Chapitre III.A.1**, suppose l'inaccessibilité du G-quadruplexe par son extrémité 3'. En effet, le déroulement du G-quadruplexe par une hélicase 3'-5' (comme Sgs1), nécessite le déplacement de Pol $\epsilon$ , et probablement la résection de l'ADN néo-synthétisé pour générer une quantité d'ADNsb suffisante pour charger l'hélicase. Ainsi, cette accessibilité réduite fait du G-quadruplexe un substrat privilégié des hélicases de directionnalité 5'-3', qui peut se charger directement sur l'ADNsb généré en aval du G-quadruplexe (**Figure 52**).

L'analyse des propriétés structurales de CEB1 et de CEB25 montre qu'une longue boucle (9 nt) au sein de la séquence G4P est suffisante pour abolir l'instabilité du minisatellite (**Chapitre II.E.1**). Cette boucle, dont le changement de taille affecte peu la stabilité thermique de la structure mais grandement l'instabilité de la séquence sous-jacente *in vivo*, suggère que des mécanismes qui la ciblent et facilitent le déroulement de la structure sont à l'œuvre dans la cellule. Cette boucle simple-brin pourrait constituer une plate-forme de chargement pour des hélicases de polarité 3'-5' (dont le site de chargement est généralement d'environ 5 nt) qui n'avaient jusque là pas accès à la structure: le G-quadruplexe serait alors substrat d'autres protéines avec qui Pif1 deviendrait redondante (**Figure 52**). Chez *C. elegans*, la délétion de *him-6* (homologue de BLM) ou de *chl-1* n'induit pas de délétion au niveau de l'homopolymère G<sub>20</sub> de l'intron 4 du gène *vab-1*. En revanche, ces mutants combinés à l'inactivation de *dog-1* causent une augmentation synergique de ces délétions (Youds et al. 2006; Chung et al. 2011). En effet, le minisatellite G<sub>20</sub> peut former des G-quadruplexes dont la taille

## Souche WT

## Souche *pif1*Δ



**Figure 52: Pif1 est redondant avec d'autres hélicases sur les G-quadruplexes à boucle longue.** Dans une souche WT (gauche), un G-quadruplexe à boucles courtes (CEB1) n'est accessible que par le simple-brin en 3'. Il est uniquement substrat de l'hélicase Pif1. Un G-quadruplexe à boucle longue est substrat de Pif1, et d'une autre hélicase 3'-5' inconnue: les deux hélicases sont redondantes. Les deux hélicases peuvent se charger et dérouler le G-quadruplexe à partir de la grande boucle. Dans une souche *pif1*Δ (droite), le G-quadruplexe à boucles courtes n'est plus substrat d'aucune hélicase, persiste, et induit une lésion à l'origine du réarrangement de CEB1. En revanche, un G-quadruplexe à boucle longue peut toujours être déroulé par une hélicase 3'-5' qui se charge sur la grande boucle: CEB25 reste stable.

cumulée des boucles est compris entre 3 à 8 nt. De plus, le mutant *dog-1* n'induit des délétions qu'au niveau des séquences G4P contenant des boucles très courtes (Kruisselbrink et al. 2008). Ainsi, *dog-1* pourrait être essentielle au déroulement de G-quadruplexes à boucles courtes, et redondante avec *him-6* et *chl-1* sur les G-quadruplexes à boucles plus longues. Il serait intéressant d'étudier la stabilité des séquences G4P à boucles plus longues dans ces double-mutants.

Les hélicases ne sont peut-être pas les seules à même de dérouler cette structure. En effet, la fixation de RPA à l'ADNsb est initiée par une première fixation des domaines A et B de la sous-unité RPA1 (homologue humain de Rfa1) à une région de 8 nt, qui promeut ensuite la fixation des domaines C et D à une région d'ADNsb finale de 30 nt (Jiang et al. 2006). De manière intéressante, les domaines D et E de hRPA semblent fixer préférentiellement des structures G-quadruplexes, et l'association des domaines CDE le déroule efficacement (Prakash et al. 2011). Cette fixation séquentielle, initiée par un premier évènement de fixation sur la boucle de 9 nt, pourrait forcer le déroulement du G-quadruplexe, une hypothèse testable *in vitro*.

Étant donné la redondance fonctionnelle potentiellement très importante de protéines capables de se charger sur, et de dérouler le G-quadruplexe à partir de cette boucle (il existe une centaine d'hélicases putatives chez *S. cerevisiae*), les approches génétiques visant à les identifier semblent inappropriées. Une expérience de spectrométrie de masse sur les protéines capables de lier le G-quadruplexe de CEB25 à boucle longue, mais pas le G-quadruplexe à boucle courte, devrait permettre d'identifier des protéines candidates capables de se fixer au G-quadruplexe via cette grande boucle.

### **III.D.3. Versatilité évolutive et physiologique des G-quadruplexes**

La diversité conformationnelle des G-quadruplexes en fait une entité apte à former des interactions spécifiques entre acide nucléique et protéine. A la différence des interactions protéine-ADN "classiques" par reconnaissance d'une séquence nucléotidique dans la double-hélice, la reconnaissance d'une structure secondaire par une protéine présente l'avantage d'être dépendante des conditions de formation de cette structure. En effet, la formation, la stabilité, et éventuellement la conformation du G-quadruplexe dépendent des propriétés physico-chimiques du milieu (**Chapitre I.A.1**). *In vivo*, des variations de faibles amplitudes de la température ou de la concentration en

Sodium et Potassium modulent des phénotypes G-quadruplexes-dépendants tels que la traduction chez la bactérie ou le transport axonal d'ARNm chez l'homme, respectivement (Wieland and Hartig 2007; Subramanian et al. 2011). Ces deux exemples suggèrent un rôle de senseur environnemental pour le G-quadruplexe, qui agit comme un interrupteur de l'interaction entre une protéine et un acide nucléique.

La gamme de sensibilité de ce senseur dépend de la séquence primaire: plus la taille des boucles augmente, plus la gamme thermique de transition entre ADNsb et G-quadruplexe (c'est-à-dire le  $T_m$ ) s'abaisse (**Chapitre I.A.1**). De simples mutations de la séquence primaire permettent de moduler la sensibilité d'un processus faisant intervenir un G-quadruplexe. C'est le cas de CEB25, où la taille de la boucle au sein du G-quadruplexe modifie drastiquement le phénotype d'instabilité de la séquence *in vivo* (**Chapitre II.E.1**). De manière plus drastique encore, la mutation d'une seule guanine requise pour la formation du G-quadruplexe abolit le phénotype associé, comme observé par mutagenèse ponctuelle du motif de CEB1 (**Chapitre II.E.2**). L'existence de mécanismes "error-prone" de réplication de ces structures, qui reste à démontrer, serait un puissant moteur évolutif d'extinction d'une interaction ou d'un phénotype dépendant d'un G-quadruplexe.

Ces propriétés font du G-quadruplexe une structure versatile d'un point de vue évolutif et d'un point de vue physiologique. Cette versatilité d'interaction entre acides nucléiques et protéine est probablement à l'origine de la diversité des rôles, toujours plus nombreux, identifiés pour ces structures à de nombreux niveaux du métabolisme de l'ADN et des ARN.

## Conclusions

À l'issue de ma thèse, je pense avoir contribué à établir dans la communauté une approche standard permettant de déterminer si un phénotype dépend de la formation d'un G-quadruplexe ou non *in vivo*. Cette approche établie, j'ai pu m'intéresser au « côté obscur » de la biologie des G-quadruplexes : leur capacité à induire l'instabilité du génome. Cette étude fondamentale a permis (i) de démontrer que les G-quadruplexes se forment chez *S. cerevisiae*, (ii) que l'hélicase Pif1 est impliquée dans leur déroulement, (iii) que leur persistance pathologique perturbe la réplication, (iv) que cette perturbation concerne principalement le brin à synthèse continue qui (v) crée un substrat recombinogène conduisant à l'instabilité de la séquence sous-jacente, et (vi) que finalement ces perturbations et l'induction de l'instabilité génomique dépendent de certaines propriétés structurales du G-quadruplexe. Ainsi, même si des rôles « positifs » toujours plus nombreux sont découverts pour ces structures dans les cellules, leur formation doit être régulée pour les empêcher d'interférer avec la réplication et d'induire la formation de réarrangements chromosomiques.





## Références

- Aguilera A, Gomez-Gonzalez B. 2008. Genome instability: a mechanistic view of its causes and consequences. *Nature Rev Genet* 9: 204-217.
- Amarger V, Gauguier D, Yerle M, Apiou F, Pinton P, Giraudeau F, Monfouilloux S, Lathrop M, Dutrillaux B, Buard J et al. 1998. Analysis of distribution in the human, pig, and rat genomes points toward a general subtelomeric origin of minisatellite structures. *Genomics* 52: 62-71.
- Amrane S, Adrian M, Heddi B, Serero A, Nicolas A, Mergny JL, Phan AT. 2012. Formation of pearl-necklace monomorphic G-quadruplexes in the human CEB25 minisatellite. *J Am Chem Soc*.
- Anand RP, Shah KA, Niu H, Sung P, Mirkin SM, Freudenreich CH. 2012. Overcoming natural replication barriers: differential helicase requirements. *Nucleic Acids Res* 40: 1091-1105.
- Arnott S, Chandrasekaran R, Marttila CM. 1974. Structures for polyinosinic acid and polyguanylic acid. *Biochem J* 141: 537-543.
- Arnoult N, Shin-Ya K, Londono-Vallejo JA. 2008. Studying telomere replication by Q-CO-FISH: the effect of telomestatin, a potent G-quadruplex ligand. *Cytogenet Genome Res* 122: 229-236.
- Ashton TM, Mankouri HW, Heidenblut A, McHugh PJ, Hickson ID. 2011. Pathways for Holliday junction processing during homologous recombination in *Saccharomyces cerevisiae*. *Mol Cell Biol* 31: 1921-1933.
- Asturias FJ, Cheung IK, Sabouri N, Chilkova O, Wepplo D, Johansson E. 2006. Structure of *Saccharomyces cerevisiae* DNA polymerase epsilon by cryo-electron microscopy. *Nat Struct Mol Biol* 13: 35-43.
- Atkinson J, McGlynn P. 2009. Replication fork reversal and the maintenance of genome stability. *Nucleic Acids Res* 37: 3475-3492.
- Balakumaran BS, Freudenreich CH, Zakian VA. 2000. CGG/CCG repeats exhibit orientation-dependent instability and orientation-independent fragility in *Saccharomyces cerevisiae*. *Hum Mol Genet* 9: 93-100.
- Baldwin EL, Lee JY, Blake DM, Bunke BP, Alexander CR, Kogan AL, Ledbetter DH, Martin CL. 2008. Enhanced detection of clinically relevant genomic imbalances using a targeted plus whole genome oligonucleotide microarray. *Genet Med* 10: 415-429.
- Bando M, Katou Y, Komata M, Tanaka H, Itoh T, Sutani T, Shirahige K. 2009. Csm3, Tof1, and Mrc1 form a heterotrimeric mediator complex that associates with DNA replication forks. *J Biol Chem* 284: 34355-34365.
- Baran N, Lapidot A, Manor H. 1991. Formation of DNA triplexes accounts for arrests of DNA synthesis at d(TC)<sub>n</sub> and d(GA)<sub>n</sub> tracts. *Proc Natl Acad Sci U S A* 88: 507-511.
- Barlow JH, Lisby M, Rothstein R. 2008. Differential regulation of the cellular response to DNA double-strand breaks in G1. *Mol Cell* 30: 73-85.
- Basundra R, Kumar A, Amrane S, Verma A, Phan AT, Chowdhury S. 2010. A novel G-quadruplex motif modulates promoter activity of human thymidine kinase 1. *FEBS J* 277: 4254-4264.
- Beaudoin JD, Perreault JP. 2010. 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res* 38: 7022-7036.
- Bejugam M, Sewitz S, Shirude PS, Rodriguez R, Shahid R, Balasubramanian S. 2007. Trisubstituted isoalloxazines as a new class of G-quadruplex binding ligands: small molecule regulation of c-kit oncogene expression. *J Am Chem Soc* 129: 12926-12927.
- Blume SW, Guarcello V, Zacharias W, Miller DM. 1997. Divalent transition metal cations counteract potassium-induced quadruplex assembly of oligo(dG) sequences. *Nucleic Acids Res* 25: 617-625.
- Boan F, Blanco MG, Barros P, Gonzalez AI, Gomez-Marquez J. 2004. Inhibition of DNA synthesis by K<sup>+</sup>-stabilised G-quadruplex promotes allelic preferential amplification. *FEBS Lett* 571: 112-118.

- Bochman ML, Sabouri N, Zakian VA. 2010. Unwinding the functions of the Pif1 family helicases. *DNA Repair (Amst)* 9: 237-249.
- Bochman ML, Schwacha A. 2008. The Mcm2-7 complex has in vitro helicase activity. *Mol Cell* 31: 287-293.
- . 2009. The Mcm complex: unwinding the mechanism of a replicative helicase. *Microbiol Mol Biol Rev* 73: 652-683.
- Borde V, Lin W, Novikov E, Petrini JH, Lichten M, Nicolas A. 2004. Association of Mre11p with double-strand break sites during yeast meiosis. *Mol Cell* 13: 389-401.
- Bosco G, Haber JE. 1998. Chromosome break-induced DNA replication leads to nonreciprocal translocations and telomere capture. *Genetics* 150: 1037-1047.
- Boule JB, Vega LR, Zakian VA. 2005. The yeast Pif1p helicase removes telomerase from telomeric DNA. *Nature* 438: 57-61.
- Boule JB, Zakian VA. 2007. The yeast Pif1p DNA helicase preferentially unwinds RNA DNA substrates. *Nucleic Acids Res* 35: 5809-5818.
- Bourdoncle A, Estevez Torres A, Gosse C, Lacroix L, Vekhoff P, Le Saux T, Jullien L, Mergny JL. 2006. Quadruplex-based molecular beacons as tunable DNA probes. *J Am Chem Soc* 128: 11094-11105.
- Branzei D. 2011. Ubiquitin family modifications and template switching. *FEBS Lett* 585: 2810-2817.
- Brassart B, Gomez D, De Cian A, Paterski R, Montagnac A, Qui KH, Temime-Smaali N, Trentesaux C, Mergny JL, Gueritte F et al. 2007. A new steroid derivative stabilizes g-quadruplexes and induces telomere uncapping in human tumor cells. *Mol Pharmacol* 72: 631-640.
- Brewer BJ, Lockshon D, Fangman WL. 1992. The arrest of replication forks in the rDNA of yeast occurs independently of transcription. *Cell* 71: 267-276.
- Brigati C, Kurtz S, Balderes D, Vidali G, Shore D. 1993. An essential yeast gene encoding a TTAGGG repeat-binding protein. *Mol Cell Biol* 13: 1306-1314.
- Brock GJ, Anderson NH, Monckton DG. 1999. Cis-acting modifiers of expanded CAG/CTG triplet repeat expandability: associations with flanking GC content and proximity to CpG islands. *Hum Mol Genet* 8: 1061-1067.
- Brooks TA, Kendrick S, Hurley L. 2010. Making sense of G-quadruplex and i-motif functions in oncogene promoters. *Febs J* 277: 3459-3469.
- Brown RV, Danford FL, Gokhale V, Hurley LH, Brooks TA. 2011. Demonstration that drug-targeted down-regulation of MYC in non-Hodgkins lymphoma is directly mediated through the promoter G-quadruplex. *J Biol Chem* 286: 41018-41027.
- Broxson C, Beckett J, Tornaletti S. 2011. Transcription arrest by a G quadruplex forming-trinucleotide repeat sequence from the human c-myc gene. *Biochemistry* 50: 4162-4172.
- Buard J, Bourdet A, Yardley J, Dubrova Y, Jeffreys AJ. 1998. Influences of array size and homogeneity on minisatellite mutation. *EMBO J* 17: 3495-3502.
- Buard J, Vergnaud G. 1994. Complex recombination events at the hypermutable minisatellite CEB1 (D2S90). *Embo J* 13: 3203-3210.
- Budd ME, Reis CC, Smith S, Myung K, Campbell JL. 2006. Evidence Suggesting that Pif1 Helicase Functions in DNA Replication with the Dna2 Helicase/Nuclease and DNA Polymerase  $\delta$ . *Mol Cell Biol* 26: 2490-2500.
- Bugaut A, Balasubramanian S. 2012. 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.*
- Burger AM, Dai F, Schultes CM, Reszka AP, Moore MJ, Double JA, Neidle S. 2005. The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. *Cancer Res* 65: 1489-1496.
- Bussen W, Raynard S, Busygina V, Singh AK, Sung P. 2007. Holliday junction processing activity of the BLM-Topo IIIalpha-BLAP75 complex. *J Biol Chem* 282: 31484-31492.
- Byun TS, Pacek M, Yee MC, Walter JC, Cimprich KA. 2005. Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. *Genes Dev* 19: 1040-1052.

- Bzymek M, Thayer NH, Oh SD, Kleckner N, Hunter N. 2011. Double Holliday junctions are intermediates of DNA break repair. *Nature* 464: 937-941.
- Cahoon LA, Seifert HS. 2009. An alternative DNA structure is necessary for pilin antigenic variation in *Neisseria gonorrhoeae*. *Science* 325: 764-767.
- Callahan JL, Andrews KJ, Zakian VA, Freudenreich CH. 2003. Mutations in yeast replication proteins that increase CAG/CTG expansions also increase repeat fragility. *Mol Cell Biol* 23: 7849-7860.
- Capra JA, Paeschke K, Singh M, Zakian VA. 2010. G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in *Saccharomyces cerevisiae*. *PLoS Comput Biol* 6: e1000861.
- Carlile CM, Pickart CM, Matunis MJ, Cohen RE. 2009. Synthesis of free and proliferating cell nuclear antigen-bound polyubiquitin chains by the RING E3 ubiquitin ligase Rad5. *J Biol Chem* 284: 29326-29334.
- Cejka P, Cannavo E, Polaczek P, Masuda-Sasa T, Pokharel S, Campbell JL, Kowalczykowski SC. 2010a. DNA end resection by Dna2-Sgs1-RPA and its stimulation by Top3-Rmi1 and Mre11-Rad50-Xrs2. *Nature* 467: 112-116.
- Cejka P, Plank JL, Bachrati CZ, Hickson ID, Kowalczykowski SC. 2010b. Rmi1 stimulates decatenation of double Holliday junctions during dissolution by Sgs1-Top3. *Nat Struct Mol Biol* 17: 1377-1382.
- Chan JE, Kolodner RD. 2011. A genetic and structural study of genome rearrangements mediated by high copy repeat Ty1 elements. *PLoS Genet* 7: e1002089.
- Chandok GS, Patel MP, Mirkin SM, Krasilnikova MM. 2012. Effects of Friedreich's ataxia GAA repeats on DNA replication in mammalian cells. *Nucleic Acids Res* 40: 3964-3974.
- Chang M, Bellaoui M, Zhang C, Desai R, Morozov P, Delgado-Cruzata L, Rothstein R, Freyer GA, Boone C, Brown GW. 2005. RMI1/NCE4, a suppressor of genome instability, encodes a member of the RecQ helicase/Topo III complex. *EMBO J* 24: 2024-2033.
- Chang M, Luke B, Kraft C, Li Z, Peter M, Lingner J, Rothstein R. 2009. Telomerase is essential to alleviate pif1-induced replication stress at telomeres. *Genetics* 183: 779-791.
- Chen C, Kolodner RD. 1999. Gross chromosomal rearrangements in *Saccharomyces cerevisiae* replication and recombination defective mutants. *Nat Genet* 23: 81-85.
- Chen JM, Cooper DN, Ferec C, Kehrer-Sawatzki H, Patrinos GP. 2010. Genomic rearrangements in inherited disease and cancer. *Semin Cancer Biol* 20: 222-233.
- Cheng X, Dunaway S, Ivessa AS. 2007. The role of Pif1p, a DNA helicase in *Saccharomyces cerevisiae*, in maintaining mitochondrial DNA. *Mitochondrion* 7: 211-222.
- Cheung I, Schertzer M, Rose A, Lansdorp PM. 2002. Disruption of dog-1 in *Caenorhabditis elegans* triggers deletions upstream of guanine-rich DNA. *Nat Genet* 31: 405-409.
- Chung G, O'Neil NJ, Rose AM. 2011. CHL-1 provides an essential function affecting cell proliferation and chromosome stability in *Caenorhabditis elegans*. *DNA Repair (Amst)* 10: 1174-1182.
- Clerici M, Mantiero D, Lucchini G, Longhese MP. 2005. The *Saccharomyces cerevisiae* Sae2 protein promotes resection and bridging of double strand break ends. *J Biol Chem* 280: 38631-38638.
- Cobb JA, Shimada K, Gasser SM. 2004. Redundancy, insult-specific sensors and thresholds: unlocking the S-phase checkpoint response. *Curr Opin Genet Dev* 14: 292-300.
- Cogoi S, Xodo LE. 2006. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res* 34: 2536-2549.
- Cordeiro-Stone M, Makhov AM, Zaritskaya LS, Griffith JD. 1999. Analysis of DNA replication forks encountering a pyrimidine dimer in the template to the leading strand. *J Mol Biol* 289: 1207-1218.
- Crabbe L, Thomas A, Pantesco V, De Vos J, Pasero P, Lengronne A. 2010. Analysis of replication profiles reveals key role of RFC-Ctf18 in yeast replication stress response. *Nat Struct Mol Biol* 17: 1391-1397.
- Creacy SD, Routh ED, Iwamoto F, Nagamine Y, Akman SA, Vaughn JP. 2008. G4 resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source

- of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates. *J Biol Chem* 283: 34626-34634.
- Dai J, Carver M, Yang D. 2008. Polymorphism of human telomeric quadruplex structures. *Biochimie* 90: 1172-1183.
- Dai J, Chen D, Jones RA, Hurley LH, Yang D. 2006. NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res* 34: 5133-5144.
- Daigaku Y, Davies AA, Ulrich HD. 2010. Ubiquitin-dependent DNA damage bypass is separable from genome replication. *Nature* 465: 951-955.
- De Armond R, Wood S, Sun D, Hurley LH, Ebbinghaus SW. 2005. Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1alpha promoter. *Biochemistry* 44: 16341-16350.
- De Cian A, Delemos E, Mergny JL, Teulade-Fichou MP, Monchaud D. 2007a. Highly efficient G-quadruplex recognition by bisquinolinium compounds. *J Am Chem Soc* 129: 1856-1857.
- De Cian A, Gros J, Guedin A, Haddi M, Lyonnais S, Guittat L, Riou JF, Trentesaux C, Sacca B, Lacroix L et al. 2008a. DNA and RNA quadruplex ligands. *Nucleic Acids Symp Ser (Oxf)*: 7-8.
- De Cian A, Guittat L, Kaiser M, Sacca B, Amrane S, Bourdoncle A, Alberti P, Teulade-Fichou MP, Lacroix L, Mergny JL. 2007b. Fluorescence-based melting assays for studying quadruplex ligands. *Methods* 42: 183-195.
- De Cian A, Lacroix L, Douarre C, Temime-Smaali N, Trentesaux C, Riou JF, Mergny JL. 2008b. Targeting telomeres and telomerase. *Biochimie* 90: 131-155.
- De Cian A, Mergny JL. 2007. Quadruplex ligands may act as molecular chaperones for tetramolecular quadruplex formation. *Nucleic Acids Res* 35: 2483-2493.
- Debrauwere H, Buard J, Tessier J, Aubert D, Vergnaud G, Nicolas A. 1999. Meiotic instability of human minisatellite CEB1 in yeast requires DNA double-strand breaks. *Nat Genet* 23: 367-371.
- Decorsiere A, Cayrel A, Vagner S, Millevoi S. 2011. Essential role for the interaction between hnRNP H/F and a G quadruplex in maintaining p53 pre-mRNA 3'-end processing and function during DNA damage. *Genes Dev* 25: 220-225.
- Deem A, Keszthelyi A, Blackgrove T, Vayl A, Coffey B, Mathur R, Chabes A, Malkova A. 2011. Break-induced replication is highly inaccurate. *PLoS Biol* 9: e1000594.
- Delagoutte E, von Hippel PH. 2001. Molecular mechanisms of the functional coupling of the helicase (gp41) and polymerase (gp43) of bacteriophage T4 within the DNA replication fork. *Biochemistry* 40: 4459-4477.
- Dhar A, Lahue RS. 2008. Rapid unwinding of triplet repeat hairpins by Srs2 helicase of *Saccharomyces cerevisiae*. *Nucleic Acids Res* 36: 3366-3373.
- Diamant N, Hendel A, Vered I, Carell T, Reissner T, de Wind N, Geacinov N, Livneh Z. 2012. DNA damage bypass operates in the S and G2 phases of the cell cycle and exhibits differential mutagenicity. *Nucleic Acids Res* 40: 170-180.
- Dixon IM, Lopez F, Esteve JP, Tejera AM, Blasco MA, Pratviel G, Meunier B. 2005. Porphyrin derivatives for telomere binding and telomerase inhibition. *Chembiochem* 6: 123-132.
- Do NQ, Lim KW, Teo MH, Heddi B, Phan AT. 2011. Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity. *Nucleic Acids Res* 39: 9448-9457.
- Du Z, Zhao Y, Li N. 2008. Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. *Genome Res* 18: 233-241.
- Dua R, Levy DL, Campbell JL. 1999. Analysis of the essential functions of the C-terminal protein/protein interaction domain of *Saccharomyces cerevisiae* pol epsilon and its unexpected ability to support growth in the absence of the DNA polymerase domain. *J Biol Chem* 274: 22283-22288.
- Dubarry M, Loiodice I, Chen CL, Thermes C, Taddei A. 2011. Tight protein-DNA interactions favor gene silencing. *Genes Dev* 25: 1365-1370.

- Duquette ML, Handa P, Vincent JA, Taylor AF, Maizels N. 2004. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev* 18: 1618-1629.
- Duquette ML, Huber MD, Maizels N. 2007. G-rich proto-oncogenes are targeted for genomic instability in B-cell lymphomas. *Cancer Res* 67: 2586-2594.
- Duquette ML, Pham P, Goodman MF, Maizels N. 2005. AID binds to transcription-induced structures in c-MYC that map to regions associated with translocation and hypermutation. *Oncogene* 24: 5791-5798.
- Eddy J, Maizels N. 2006. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res* 34: 3887-3896.
- . 2008. Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic Acids Res* 36: 1321-1333.
- . 2009. Selection for the G4 DNA motif at the 5' end of human genes. *Mol Carcinog* 48: 319-325.
- Eddy J, Vallur AC, Varma S, Liu H, Reinhold WC, Pommier Y, Maizels N. 2011. G4 motifs correlate with promoter-proximal transcriptional pausing in human genes. *Nucleic Acids Res* 39: 4975-4983.
- Elvers I, Johansson F, Groth P, Erixon K, Helleday T. 2011. UV stalled replication forks restart by re-priming in human fibroblasts. *Nucleic Acids Res* 39: 7049-7057.
- Enemark EJ, Joshua-Tor L. 2006. Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* 442: 270-275.
- Fachinetti D, Bermejo R, Cocito A, Minardi S, Katou Y, Kanoh Y, Shirahige K, Azvolinsky A, Zakian VA, Foiani M. 2010. Replication termination at eukaryotic chromosomes is mediated by Top2 and occurs at genomic loci containing pausing elements. *Mol Cell* 39: 595-605.
- Feng W, Collingwood D, Boeck ME, Fox LA, Alvino GM, Fangman WL, Raghuraman MK, Brewer BJ. 2006. Genomic mapping of single-stranded DNA in hydroxyurea-challenged yeasts identifies origins of replication. *Nat Cell Biol* 8: 148-155.
- Forget AL, Kowalczykowski SC. 2012. Single-molecule imaging of DNA pairing by RecA reveals a three-dimensional homology search. *Nature* 482: 423-427.
- Foury F, Kolodynski J. 1983. pif mutation blocks recombination between mitochondrial rho+ and rho- genomes having tandemly arrayed repeat units in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 80: 5345-5349.
- Foury F, Lahaye A. 1987. Cloning and sequencing of the PIF gene involved in repair and recombination of yeast mitochondrial DNA. *EMBO J* 6: 1441-1449.
- Freudenreich CH, Kantrow SM, Zakian VA. 1998. Expansion and length-dependent fragility of CTG repeats in yeast. *Science* 279: 853-856.
- Fry M, Loeb LA. 1994. The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahelical structure. *Proc Natl Acad Sci U S A* 91: 4950-4954.
- Fu YV, Yardimci H, Long DT, Ho TV, Guainazzi A, Bermudez VP, Hurwitz J, van Oijen A, Scharer OD, Walter JC. 2011. Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase. *Cell* 146: 931-941.
- Futami K, Shimamoto A, Furuichi Y. 2007. Mitochondrial and nuclear localization of human Pif1 helicase. *Biol Pharm Bull* 30: 1685-1692.
- Gacy AM, Goellner GM, Spiro C, Chen X, Gupta G, Bradbury EM, Dyer RB, Mikesell MJ, Yao JZ, Johnson AJ et al. 1998. GAA instability in Friedreich's Ataxia shares a common, DNA-directed and intraallelic mechanism with other trinucleotide diseases. *Mol Cell* 1: 583-593.
- Gambus A, van Deursen F, Polychronopoulos D, Foltman M, Jones RC, Edmondson RD, Calzada A, Labib K. 2009. A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase alpha within the eukaryotic replisome. *EMBO J* 28: 2992-3004.
- Gangloff S, McDonald JP, Bendixen C, Arthur L, Rothstein R. 1994. The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase. *Mol Cell Biol* 14: 8391-8398.
- Gellert M, Lipsett MN, Davies DR. 1962. Helix formation by guanylic acid. *Proc Natl Acad Sci U S A* 48: 2013-2018.

- Georgescu RE, Kurth I, O'Donnell ME. 2012. Single-molecule studies reveal the function of a third polymerase in the replisome. *Nat Struct Mol Biol* 19: 113-116.
- Gilmour DS. 2009. Promoter proximal pausing on genes in metazoans. *Chromosoma* 118: 1-10.
- Gilson E, Geli V. 2007. How telomeres are replicated. *Nat Rev Mol Cell Biol* 8: 825-838.
- Giraldo R, Rhodes D. 1994. The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. *EMBO J* 13: 2411-2420.
- Giraldo R, Suzuki M, Chapman L, Rhodes D. 1994. Promotion of parallel DNA quadruplexes by a yeast telomere binding protein: a circular dichroism study. *Proc Natl Acad Sci U S A* 91: 7658-7662.
- Giri B, Smaldino PJ, Thys RG, Creacy SD, Routh ED, Hantgan RR, Lattmann S, Nagamine Y, Akman SA, Vaughn JP. 2011. G4 resolvase 1 tightly binds and unwinds unimolecular G4-DNA. *Nucleic Acids Res* 39: 7161-7178.
- Gomez D, O'Donohue MF, Wenner T, Douarre C, Macadre J, Koebel P, Giraud-Panis MJ, Kaplan H, Kolkes A, Shin-ya K et al. 2006a. The G-quadruplex ligand telomestatin inhibits POT1 binding to telomeric sequences in vitro and induces GFP-POT1 dissociation from telomeres in human cells. *Cancer Res* 66: 6908-6912.
- Gomez D, Wenner T, Brassart B, Douarre C, O'Donohue MF, El Khoury V, Shin-Ya K, Morjani H, Trentesaux C, Riou JF. 2006b. Telomestatin-induced telomere uncapping is modulated by POT1 through G-overhang extension in HT1080 human tumor cells. *J Biol Chem* 281: 38721-38729.
- Gordenin DA, Lobachev KS, Degtyareva NP, Malkova AL, Perkins E, Resnick MA. 1993. Inverted DNA repeats: a source of eukaryotic genomic instability. *Mol Cell Biol* 13: 5315-5322.
- Gowan SM, Harrison JR, Patterson L, Valenti M, Read MA, Neidle S, Kelland LR. 2002. A G-quadruplex-interactive potent small-molecule inhibitor of telomerase exhibiting in vitro and in vivo antitumor activity. *Mol Pharmacol* 61: 1154-1162.
- Granotier C, Pennarun G, Riou L, Hoffschir F, Gauthier LR, De Cian A, Gomez D, Mandine E, Riou JF, Mergny JL et al. 2005. Preferential binding of a G-quadruplex ligand to human chromosome ends. *Nucleic Acids Res* 33: 4182-4190.
- Guedin A, De Cian A, Gros J, Lacroix L, Mergny JL. 2008. Sequence effects in single-base loops for quadruplexes. *Biochimie* 90: 686-696.
- Guedin A, Gros J, Alberti P, Mergny JL. 2010a. How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res* 38: 7858-7868.
- Guedin A, Lacroix L, Mergny JL. 2010b. Thermal melting studies of ligand DNA interactions. *Methods Mol Biol* 613: 25-35.
- Guirola M, Barreto L, Pagani A, Romagosa M, Casamayor A, Atrian S, Arino J. 2010. Lack of DNA helicase Pif1 disrupts zinc and iron homeostasis in yeast. *Biochem J* 432: 595-605.
- Gunaratnam M, Green C, Moreira JB, Moorhouse AD, Kelland LR, Moses JE, Neidle S. 2009. G-quadruplex compounds and cis-platin act synergistically to inhibit cancer cell growth in vitro and in vivo. *Biochem Pharmacol* 78: 115-122.
- Guo K, Pourpak A, Beetz-Rogers K, Gokhale V, Sun D, Hurley LH. 2007. Formation of pseudosymmetrical G-quadruplex and i-motif structures in the proximal promoter region of the RET oncogene. *J Am Chem Soc* 129: 10220-10228.
- Hagihara M, Yoneda K, Yabuuchi H, Okuno Y, Nakatani K. 2010. A reverse transcriptase stop assay revealed diverse quadruplex formations in UTRs in mRNA. *Bioorg Med Chem Lett* 20: 2350-2353.
- Haider S, Parkinson GN, Neidle S. 2002. Crystal structure of the potassium form of an Oxytricha nova G-quadruplex. *J Mol Biol* 320: 189-200.
- Halder K, Benzler M, Hartig JS. 2012a. Reporter assays for studying quadruplex nucleic acids. *Methods*.
- Halder R, Riou JF, Teulade-Fichou MP, Frickey T, Hartig JS. 2012b. Bisquinolinium compounds induce quadruplex-specific transcriptome changes in HeLa S3 cell lines. *BMC Res Notes* 5: 138.
- Han H, Cliff CL, Hurley LH. 1999a. Accelerated assembly of G-quadruplex structures by a small molecule. *Biochemistry* 38: 6981-6986.

- Han H, Hurley LH, Salazar M. 1999b. A DNA polymerase stop assay for G-quadruplex-interactive compounds. *Nucleic Acids Res* 27: 537-542.
- Hanakahi LA, Sun H, Maizels N. 1999. High affinity interactions of nucleolin with G-G-paired rDNA. *J Biol Chem* 274: 15908-15912.
- Haracska L, Prakash S, Prakash L. 2002. Yeast Rev1 protein is a G template-specific DNA polymerase. *J Biol Chem* 277: 15546-15551.
- Haracska L, Unk I, Johnson RE, Johansson E, Burgers PM, Prakash S, Prakash L. 2001. Roles of yeast DNA polymerases delta and zeta and of Rev1 in the bypass of abasic sites. *Genes Dev* 15: 945-954.
- Harrington C, Lan Y, Akman SA. 1997. The identification and characterization of a G4-DNA resolvase activity. *J Biol Chem* 272: 24631-24636.
- Hashash N, Johnson AL, Cha RS. 2011. Regulation of fragile sites expression in budding yeast by MEC1, RRM3 and hydroxyurea. *J Cell Sci* 124: 181-185.
- Heddi B, Phan AT. 2011. Structure of human telomeric DNA in crowded solution. *J Am Chem Soc* 133: 9824-9833.
- Heller RC, Marians KJ. 2006. Replication fork reactivation downstream of a blocked nascent leading strand. *Nature* 439: 557-562.
- Henderson E, Hardin CC, Walk SK, Tinoco I, Jr., Blackburn EH. 1987. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell* 51: 899-908.
- Hershman SG, Chen Q, Lee JY, Kozak ML, Yue P, Wang LS, Johnson FB. 2008. Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. *Nucleic Acids Res* 36: 144-156.
- Heyer WD, Li X, Rolfmeier M, Zhang XP. 2006. Rad54: the Swiss Army knife of homologous recombination? *Nucl Acids Res* 34: 4115-4125.
- Higgins NP, Kato K, Strauss B. 1976. A model for replication repair in mammalian cells. *J Mol Biol* 101: 417-425.
- Ho CK, Mazon G, Lam AF, Symington LS. 2010. Mus81 and Yen1 promote reciprocal exchange during mitotic recombination to maintain genome integrity in budding yeast. *Mol Cell* 40: 988-1000.
- Hopfner KP, Craig L, Moncalian G, Zinkel RA, Usui T, Owen BA, Karcher A, Henderson B, Bodmer JL, McMurray CT et al. 2002. The Rad50 zinc-hook is a structure joining Mre11 complexes in DNA recombination and repair. *Nature* 418: 562-566.
- Huber MD, Duquette ML, Shiels JC, Maizels N. 2006. A conserved G4 DNA binding domain in RecQ family helicases. *J Mol Biol* 358: 1071-1080.
- Huber MD, Lee DC, Maizels N. 2002. G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. *Nucleic Acids Res* 30: 3954-3961.
- Huppert JL. 2010. Structure, location and interactions of G-quadruplexes. *Febs J* 277: 3452-3458.
- Huppert JL, Balasubramanian S. 2005. Prevalence of quadruplexes in the human genome. *Nucl Acids Res* 33: 2908-2916.
- . 2007. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res* 35: 406-413.
- Huppert JL, Bugaut A, Kumari S, Balasubramanian S. 2008. G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res* 36: 6260-6268.
- Incles CM, Schultes CM, Kempinski H, Koehler H, Kelland LR, Neidle S. 2004. A G-quadruplex telomere targeting agent produces p16-associated senescence and chromosomal fusions in human prostate cancer cells. *Mol Cancer Ther* 3: 1201-1206.
- Ip SC, Rass U, Blanco MG, Flynn HR, Skehel JM, West SC. 2008. Identification of Holliday junction resolvases from humans and yeast. *Nature* 456: 357-361.
- Ishmael FT, Trakselis MA, Benkovic SJ. 2003. Protein-protein interactions in the bacteriophage T4 replisome. The leading strand holoenzyme is physically linked to the lagging strand holoenzyme and the primosome. *J Biol Chem* 278: 3145-3152.



- Ivessa AS, Lenzmeier BA, Bessler JB, Goudsouzian LK, Schnakenberg SL, Zakian VA. 2003. The *Saccharomyces cerevisiae* helicase Rrm3p facilitates replication past nonhistone protein-DNA complexes. *Mol Cell* 12: 1525-1536.
- Ivessa AS, Zhou JQ, Schulz VP, Monson EK, Zakian VA. 2002. *Saccharomyces* Rrm3p, a 5' to 3' DNA helicase that promotes replication fork progression through telomeric and subtelomeric DNA. *Genes Dev* 16: 1383-1396.
- Ivessa AS, Zhou JQ, Zakian VA. 2000. The *Saccharomyces* Pif1p DNA helicase and the highly related Rrm3p have opposite effects on replication fork progression in ribosomal DNA. *Cell* 100: 479-489.
- Jiang X, Klimovich V, Arunkumar AI, Hysinger EB, Wang Y, Ott RD, Guler GD, Weiner B, Chazin WJ, Fanning E. 2006. Structural mechanism of RPA loading on DNA during activation of a simple pre-replication complex. *EMBO J* 25: 5516-5526.
- Kamath-Loeb AS, Loeb LA, Johansson E, Burgers PM, Fry M. 2001. Interactions between the Werner syndrome helicase and DNA polymerase delta specifically facilitate copying of tetraplex and hairpin structures of the d(CGG)<sub>n</sub> trinucleotide repeat sequence. *J Biol Chem* 276: 16439-16446.
- Kaminsky EB, Kaul V, Paschall J, Church DM, Bunke B, Kunig D, Moreno-De-Luca D, Moreno-De-Luca A, Mulle JG, Warren ST et al. 2011. An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. *Genet Med* 13: 777-784.
- Kaplan DL, Davey MJ, O'Donnell M. 2003. Mcm4,6,7 uses a "pump in ring" mechanism to unwind DNA by steric exclusion and actively translocate along a duplex. *J Biol Chem* 278: 49171-49182.
- Karras GI, Jentsch S. 2010. The RAD6 DNA damage tolerance pathway operates uncoupled from the replication fork and is functional beyond S phase. *Cell* 141: 255-267.
- Katou Y, Kanoh Y, Bando M, Noguchi H, Tanaka H, Ashikari T, Sugimoto K, Shirahige K. 2003. S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex. *Nature* 424: 1078-1083.
- Kerrest A, Anand RP, Sundararajan R, Bermejo R, Liberi G, Dujon B, Freudenreich CH, Richard GF. 2009. SRS2 and SGS1 prevent chromosomal breaks and stabilize triplet repeats by restraining recombination. *Nat Struct Mol Biol* 16: 159-167.
- Kesti T, Flick K, Keranen S, Syvaoja JE, Wittenberg C. 1999. DNA polymerase epsilon catalytic domains are dispensable for DNA replication, DNA repair, and cell viability. *Mol Cell* 3: 679-685.
- Kilburn AE, Shea MJ, Sargent RG, Wilson JH. 2001. Insertion of a telomere repeat sequence into a mammalian gene causes chromosome instability. *Mol Cell Biol* 21: 126-135.
- Kim HM, Narayanan V, Mieczkowski PA, Petes TD, Krasilnikova MM, Mirkin SM, Lobachev KS. 2008. Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair. *EMBO J* 27: 2896-2906.
- Kim NW, Piatyszek MA, Prowse KR, Harley CB, West MD, Ho PL, Coviello GM, Wright WE, Weinrich SL, Shay JW. 1994. Specific association of human telomerase activity with immortal cells and cancer. *Science* 266: 2011-2015.
- Kobayashi T, Heck DJ, Nomura M, Horiuchi T. 1998. Expansion and contraction of ribosomal DNA repeats in *Saccharomyces cerevisiae*: requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I. *Genes Dev* 12: 3821-3830.
- Koering CE, Fourel G, Binet-Brasselet E, Laroche T, Klein F, Gilson E. 2000. Identification of high affinity Tbf1p-binding sites within the budding yeast genome. *Nucleic Acids Res* 28: 2519-2526.
- Koster DA, Crut A, Shuman S, Bjornsti MA, Dekker NH. 2010. Cellular strategies for regulating DNA supercoiling: a single-molecule perspective. *Cell* 142: 519-530.
- Kouzine F, Sanford S, Elisha-Feil Z, Levens D. 2008. The functional response of upstream DNA to dynamic supercoiling in vivo. *Nat Struct Mol Biol* 15: 146-154.
- Kramer KM, Haber JE. 1993. New telomeres in yeast are initiated with a highly selected subset of TG1-3 repeats. *Genes Dev* 7: 2345-2356.

- Krasilnikova MM, Mirkin SM. 2004. Replication stalling at Friedreich's ataxia (GAA)<sub>n</sub> repeats in vivo. *Mol Cell Biol* 24: 2286-2295.
- Kruisselbrink E, Guryev V, Brouwer K, Pontier DB, Cuppen E, Tijsterman M. 2008. Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCD1-defective *C. elegans*. *Curr Biol* 18: 900-905.
- Kumar N, Maiti S. 2008. A thermodynamic overview of naturally occurring intramolecular DNA quadruplexes. *Nucleic Acids Res* 36: 5610-5622.
- Kumar N, Sahoo B, Varun KA, Maiti S, Maiti S. 2008. Effect of loop length variation on quadruplex-Watson Crick duplex competition. *Nucleic Acids Res* 36: 4433-4442.
- Kuryavyi V, Patel DJ. 2010. Solution structure of a unique G-quadruplex scaffold adopted by a guanosine-rich human intronic sequence. *Structure* 18: 73-82.
- Lacroix L, Seosse A, Mergny JL. 2010. Fluorescence-based duplex-quadruplex competition test to screen for telomerase RNA quadruplex ligands. *Nucl Acids Res* 39: e21.
- Lahaye A, Leterme S, Foury F. 1993. PIF1 DNA helicase from *Saccharomyces cerevisiae*. Biochemical characterization of the enzyme. *J Biol Chem* 268: 26155-26161.
- Lahaye A, Stahl H, Thines-Sempoux D, Foury F. 1991. PIF1: a DNA helicase in yeast mitochondria. *EMBO J* 10: 997-1007.
- Lambert S, Mizuno K, Blaisonneau J, Martineau S, Chanet R, Freon K, Murray JM, Carr AM, Baldacci G. 2010. Homologous recombination restarts blocked replication forks at the expense of genome rearrangements by template exchange. *Mol Cell* 39: 346-359.
- Lane AN. 2012. The stability of intramolecular DNA G-quadruplexes compared with other macromolecules. *Biochimie* 94: 277-286.
- Lang WH, Coats JE, Majka J, Hura GL, Lin Y, Rasnik I, McMurray CT. 2011. Conformational trapping of mismatch recognition complex MSH2/MSH3 on repair-resistant DNA loops. *Proc Natl Acad Sci U S A* 108: E837-844.
- Largy E, Hamon F, Teulade-Fichou MP. 2011. Development of a high-throughput G4-FID assay for screening and evaluation of small molecules binding quadruplex nucleic acid structures. *Anal Bioanal Chem* 400: 3419-3427.
- Laughlan G, Murchie AI, Norman DG, Moore MH, Moody PC, Lilley DM, Luisi B. 1994. The high-resolution crystal structure of a parallel-stranded guanine tetraplex. *Science* 265: 520-524.
- Leach DR. 1994. Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair. *Bioessays* 16: 893-900.
- Lee C, Liachko I, Bouten R, Kelman Z, Tye BK. 2010. Alternative mechanisms for coordinating polymerase alpha and MCM helicase. *Mol Cell Biol* 30: 423-435.
- Lehmann AR. 1972. Postreplication repair of DNA in ultraviolet-irradiated mammalian cells. *J Mol Biol* 66: 319-337.
- Lemoine FJ, Degtyareva NP, Lobachev K, Petes TD. 2005. Chromosomal translocations in yeast induced by low levels of DNA polymerase a model for chromosome fragile sites. *Cell* 120: 587-598.
- Levin DS, Bai W, Yao N, O'Donnell M, Tomkinson AE. 1997. An interaction between DNA ligase I and proliferating cell nuclear antigen: implications for Okazaki fragment synthesis and joining. *Proc Natl Acad Sci U S A* 94: 12863-12868.
- Li JL, Harrison RJ, Reszka AP, Brosh RM, Jr., Bohr VA, Neidle S, Hickson ID. 2001. Inhibition of the Bloom's and Werner's syndrome helicases by G-quadruplex interacting ligands. *Biochemistry* 40: 15194-15202.
- Lia G, Michel B, Allemand JF. 2012. Polymerase exchange during Okazaki fragment synthesis observed in living cells. *Science* 335: 328-331.
- Linardopoulou EV, Williams EM, Fan Y, Friedman C, Young JM, Trask BJ. 2005. Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature* 437: 94-100.
- Lipps HJ. 1980. In vitro aggregation of the gene-sized DNA molecules of the ciliate *Stylonychia mytilus*. *Proc Natl Acad Sci U S A* 77: 4104-4107.

- Liu G, Chen X, Bissler JJ, Sinden RR, Leffak M. 2010a. Replication-dependent instability at (CTG)<sub>x</sub> (CAG) repeat hairpins in human cells. *Nat Chem Biol* 6: 652-659.
- Liu JQ, Chen CY, Xue Y, Hao YH, Tan Z. 2010b. G-quadruplex hinders translocation of BLM helicase on DNA: a real-time fluorescence spectroscopic unwinding study and comparison with duplex substrates. *J Am Chem Soc* 132: 10521-10527.
- Liu Y, Kao HI, Bambara RA. 2004. Flap endonuclease 1: a central component of DNA metabolism. *Annu Rev Biochem* 73: 589-615.
- Lobachev KS, Shor BM, Tran HT, Taylor W, Keen JD, Resnick MA, Gordenin DA. 1998. Factors affecting inverted repeat stimulation of recombination and deletion in *Saccharomyces cerevisiae*. *Genetics* 148: 1507-1524.
- Lopes J, Debrauwere H, Buard J, Nicolas A. 2002. Instability of the human minisatellite CEB1 in *rad27Δ* and *dna2-1* replication-deficient yeast cells. *EMBO J* 21: 3201-3211.
- Lopes J, Ribeyre C, Nicolas A. 2006a. Complex minisatellite rearrangements generated in the total or partial absence of Rad27/hFEN1 activity occur in a single generation and are Rad51 and Rad52 dependent. *Mol Cell Biol* 26: 6675-6689.
- Lopes M, Cotta-Ramusino C, Pellicoli A, Liberi G, Plevani P, Muzi-Falconi M, Newlon CS, Foiani M. 2001. The DNA replication checkpoint response stabilizes stalled replication forks. *Nature* 412: 557-561.
- Lopes M, Foiani M, Sogo JM. 2006b. Multiple mechanisms control chromosome integrity after replication fork uncoupling and restart at irreparable UV lesions. *Mol Cell* 21: 15-27.
- Lou H, Komata M, Katou Y, Guan Z, Reis CC, Budd M, Shirahige K, Campbell JL. 2008. Mrc1 and DNA polymerase epsilon function together in linking DNA replication and the S phase checkpoint. *Mol Cell* 32: 106-117.
- Lydeard JR, Lipkin-Moore Z, Jain S, Eapen VV, Haber JE. 2010. Sgs1 and exo1 redundantly inhibit break-induced replication and de novo telomere addition at broken chromosome ends. *PLoS Genet* 6: e1000973.
- Makovets S, Blackburn EH. 2009. DNA damage signalling prevents deleterious telomere addition at DNA breaks. *Nat Cell Biol* 11: 1383-1386.
- Makovets S, Herskowitz I, Blackburn EH. 2004. Anatomy and dynamics of DNA replication fork movement in yeast telomeric regions. *Mol Cell Biol* 24: 4019-4031.
- Maleki S, Cederberg H, Rannug U. 2002. The human minisatellites MS1, MS32, MS205 and CEB1 integrated into the yeast genome exhibit different degrees of mitotic instability but are all stabilised by RAD27. *Curr Genet* 41: 333-341.
- Malkova A, Naylor ML, Yamaguchi M, Ira G, Haber JE. 2005. RAD51-dependent break-induced replication differs in kinetics and checkpoint responses from RAD51-mediated gene conversion. *Mol Cell Biol* 25: 933-944.
- Manosas M, Spiering MM, Ding F, Croquette V, Benkovic SJ. 2012. Collaborative coupling between polymerase and helicase for leading-strand synthesis. *Nucleic Acids Res*.
- Manosas M, Xi XG, Bensimon D, Croquette V. 2010. Active and passive mechanisms of helicases. *Nucleic Acids Res* 38: 5518-5526.
- Marians KJ. 2008. Understanding how the replisome works. *Nat Struct Mol Biol* 15: 125-127.
- Mariappan SV, Catasti P, Silks LA, 3rd, Bradbury EM, Gupta G. 1999. The high-resolution structure of the triplex formed by the GAA/TTC triplet repeat associated with Friedreich's ataxia. *J Mol Biol* 285: 2035-2052.
- Marrero VA, Symington LS. 2010. Extensive DNA end processing by exo1 and sgs1 inhibits break-induced replication. *PLoS Genet* 6: e1001007.
- McCune HJ, Danielson LS, Alvino GM, Collingwood D, Delrow JJ, Fangman WL, Brewer BJ, Raghuraman MK. 2008. The temporal program of chromosome replication: genomewide replication in *clb5Δ* *Saccharomyces cerevisiae*. *Genetics* 180: 1833-1847.
- McInerney P, O'Donnell M. 2004. Functional uncoupling of twin polymerases: mechanism of polymerase dissociation from a lagging-strand block. *J Biol Chem* 279: 21543-21551.
- McLuckie KI, Waller ZA, Sanders DA, Alves D, Rodriguez R, Dash J, McKenzie GJ, Venkitaraman AR, Balasubramanian S. 2011. G-quadruplex-binding benzo[a]phenoxazines down-

- regulate c-KIT expression in human gastric carcinoma cells. *J Am Chem Soc* 133: 2658-2663.
- Membrino A, Cogoi S, Pedersen EB, Xodo LE. 2011. G4-DNA formation in the HRAS promoter and rational design of decoy oligonucleotides for cancer therapy. *PLoS One* 6: e24421.
- Mergny JL, De Cian A, Ghelab A, Sacca B, Lacroix L. 2005. Kinetics of tetramolecular quadruplexes. *Nucleic Acids Res* 33: 81-94.
- Mergny JL, Phan AT, Lacroix L. 1998. Following G-quartet formation by UV-spectroscopy. *FEBS Lett* 435: 74-78.
- Meyne J, Ratliff RL, Moyzis RK. 1989. Conservation of the human telomere sequence (TTAGGG)<sub>n</sub> among vertebrates. *Proc Natl Acad Sci U S A* 86: 7049-7053.
- Mikami-Terao Y, Akiyama M, Yuza Y, Yanagisawa T, Yamada O, Yamada H. 2008. Antitumor activity of G-quadruplex-interactive agent TMPyP4 in K562 leukemic cells. *Cancer Lett* 261: 226-234.
- Millevoi S, Moine H, Vagner S. 2012. G-quadruplexes in RNA biology. *Wiley Interdiscip Rev RNA* 3: 495-507.
- Minca EC, Kowalski D. 2010. Multiple Rad5 activities mediate sister chromatid recombination to bypass DNA damage at stalled replication forks. *Mol Cell* 38: 649-661.
- Mine-Hattab J, Rothstein R. 2012. Increased chromosome mobility facilitates homology search during recombination. *Nat Cell Biol* 14: 510-517.
- Mirkin EV, Mirkin SM. 2007. Replication fork stalling at natural impediments. *Microbiol Mol Biol Rev* 71: 13-35.
- Mirkin SM. 2007. Expandable DNA repeats and human disease. *Nature* 447: 932-940.
- Miyoshi D, Matsumura S, Nakano S, Sugimoto N. 2004. Duplex dissociation of telomere DNAs induced by molecular crowding. *J Am Chem Soc* 126: 165-169.
- Miyoshi D, Sugimoto N. 2008. Molecular crowding effects on structure and stability of DNA. *Biochimie* 90: 1040-1051.
- Mohaghegh P, Karow JK, Brosh RM, Jr., Bohr VA, Hickson ID. 2001. The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic Acids Res* 29: 2843-2849.
- Monchaud D, Allain C, Bertrand H, Smargiasso N, Rosu F, Gabelica V, De Cian A, Mergny JL, Teulade-Fichou MP. 2008. Ligands playing musical chairs with G-quadruplex DNA: a rapid and simple displacement assay for identifying selective G-quadruplex binders. *Biochimie* 90: 1207-1223.
- Moyzis RK, Buckingham JM, Cram LS, Dani M, Deaven LL, Jones MD, Meyne J, Ratliff RL, Wu JR. 1988. A highly conserved repetitive DNA sequence, (TTAGGG)<sub>n</sub>, present at the telomeres of human chromosomes. *Proc Natl Acad Sci U S A* 85: 6622-6626.
- Mukundan VT, Do NQ, Phan AT. 2011. HIV-1 integrase inhibitor T30177 forms a stacked dimeric G-quadruplex structure containing bulges. *Nucleic Acids Res* 39: 8984-8991.
- Mullen JR, Nallaseth FS, Lan YQ, Slagle CE, Brill SJ. 2005. Yeast Rmi1/Nce4 controls genome stability as a subunit of the Sgs1-Top3 complex. *Mol Cell Biol* 25: 4476-4487.
- Muller S, Kumari S, Rodriguez R, Balasubramanian S. 2010. Small-molecule-mediated G-quadruplex isolation from human cells. *Nat Chem* 2: 1095-1098.
- Myung K, Chen C, Kolodner RD. 2001. Multiple pathways cooperate in the suppression of genome instability in *Saccharomyces cerevisiae*. *Nature* 411: 1073-1076.
- Nassif N, Penney J, Pal S, Engels WR, Gloor GB. 1994. Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. *Mol Cell Biol* 14: 1613-1625.
- Nedelcheva MN, Roguev A, Dolapchiev LB, Shevchenko A, Taskov HB, Shevchenko A, Stewart AF, Stoyanov SS. 2005. Uncoupling of unwinding from DNA synthesis implies regulation of MCM helicase by Tof1/Mrc1/Csm3 checkpoint complex. *J Mol Biol* 347: 509-521.
- Nedelcheva-Velva MN, Krastev DB, Stoyanov SS. 2006. Coordination of DNA synthesis and replicative unwinding by the S-phase checkpoint pathways. *Nucleic Acids Res* 34: 4138-4146.
- Neidle S, Balasubramanian SU. 2006. *Quadruplex nucleic acids*. RSC Pub.

- Nelson JR, Lawrence CW, Hinkle DC. 1996. Deoxycytidyl transferase activity of yeast REV1 protein. *Nature* 382: 729-731.
- Nick McElhinny SA, Gordenin DA, Stith CM, Burgers PM, Kunkel TA. 2008. Division of labor at the eukaryotic replication fork. *Mol Cell* 30: 137-144.
- O'Rourke TW, Doudican NA, Mackereth MD, Doetsch PW, Shadel GS. 2002. Mitochondrial dysfunction due to oxidative mitochondrial DNA damage is reduced through cooperative actions of diverse proteins. *Mol Cell Biol* 22: 4086-4093.
- O'Rourke TW, Doudican NA, Zhang H, Eaton JS, Doetsch PW, Shadel GS. 2005. Differential involvement of the related DNA helicases Pif1p and Rrm3p in mtDNA point mutagenesis and stability. *Gene* 354: 86-92.
- Ohshima K, Kang S, Wells RD. 1996. CTG triplet repeats from human hereditary diseases are dominant genetic expansion products in *Escherichia coli*. *J Biol Chem* 271: 1853-1856.
- Osborn AJ, Elledge SJ. 2003. Mrc1 is a replication fork component whose phosphorylation in response to DNA replication stress activates Rad53. *Genes Dev* 17: 1755-1767.
- Owen BA, Yang Z, Lai M, Gajec M, Badger JD, 2nd, Hayes JJ, Edelman W, Kucherlapati R, Wilson TM, McMurray CT. 2005. (CAG)<sub>n</sub>-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. *Nat Struct Mol Biol* 12: 663-670.
- Paek AL, Kaochar S, Jones H, Elezaby A, Shanks L, Weinert T. 2009. Fusion of nearby inverted repeats by a replication-based mechanism leads to formation of dicentric and acentric chromosomes that cause genome instability in budding yeast. *Genes Dev* 23: 2861-2875.
- Paeschke K, Capra JK, Zakian VA. 2011. DNA replication through G-quadruplex motifs is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase. *Cell* 145: 678-691.
- Paeschke K, Juranek S, Simonsson T, Hempel A, Rhodes D, Lipps HJ. 2008. Telomerase recruitment by the telomere end binding protein-beta facilitates G-quadruplex DNA unfolding in ciliates. *Nat Struct Mol Biol* 15: 598-604.
- Paeschke K, Simonsson T, Postberg J, Rhodes D, Lipps HJ. 2005. Telomere end-binding proteins control the formation of G-quadruplex DNA structures *in vivo*. *Nat Struct Mol Biol* 12: 847-854.
- Pages V, Fuchs RP. 2003. Uncoupling of leading- and lagging-strand DNA replication during lesion bypass *in vivo*. *Science* 300: 1300-1303.
- Palm W, de Lange T. 2008. How shelterin protects mammalian telomeres. *Annu Rev Genet* 42: 301-334.
- Palumbo SL, Memmott RM, Uribe DJ, Krotova-Khan Y, Hurley LH, Ebbinghaus SW. 2008. A novel G-quadruplex-forming GGA repeat region in the c-myc promoter is a critical regulator of promoter activity. *Nucleic Acids Res* 36: 1755-1769.
- Pâques F, Haber JE. 1999. Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* 63: 349-404.
- Parkinson GN, Ghosh R, Neidle S. 2007. Structural basis for binding of porphyrin to human telomeres. *Biochemistry* 46: 2390-2397.
- Parkinson GN, Lee MP, Neidle S. 2002. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* 417: 876-880.
- Pavlov YI, Newlon CS, Kunkel TA. 2002. Yeast origins establish a strand bias for replicational mutagenesis. *Mol Cell* 10: 207-213.
- Payet L, Huppert JL. 2012. Stability and structure of long intramolecular G-quadruplexes. *Biochemistry* 51: 3154-3161.
- Pearson CE, Wang YH, Griffith JD, Sinden RR. 1998. Structural analysis of slipped-strand DNA (S-DNA) formed in (CTG)<sub>n</sub>. (CAG)<sub>n</sub> repeats from the myotonic dystrophy locus. *Nucleic Acids Res* 26: 816-823.
- Pelletier R, Krasilnikova MM, Samadashwily GM, Lahue R, Mirkin SM. 2003. Replication and expansion of trinucleotide repeats in yeast. *Mol Cell Biol* 23: 1349-1357.
- Pennaneach V, Putnam CD, Kolodner RD. 2006. Chromosome healing by de novo telomere addition in *Saccharomyces cerevisiae*. *Mol Microbiol* 59: 1357-1368.
- Pennarun G, Granotier C, Gauthier LR, Gomez D, Hoffschir F, Mandine E, Riou JF, Mergny JL, Mailliet P, Boussin FD. 2005. Apoptosis related to telomere instability and cell cycle

- alterations in human glioma cells treated by new highly selective G-quadruplex ligands. *Oncogene* 24: 2917-2928.
- Petraccone L, Spink C, Trent JO, Garbett NC, Mekmaysy CS, Giancola C, Chaires JB. 2011. Structure and stability of higher-order human telomeric quadruplexes. *J Am Chem Soc* 133: 20951-20961.
- Phan AT, Kuryavyi V, Burge S, Neidle S, Patel DJ. 2007. Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. *J Am Chem Soc* 129: 4386-4392.
- Phan AT, Kuryavyi V, Darnell JC, Serganov A, Majumdar A, Ilin S, Raslin T, Polonskaia A, Chen C, Clain D et al. 2011. Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nat Struct Mol Biol* 18: 796-804.
- Phan AT, Mergny JL. 2002. Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix. *Nucleic Acids Res* 30: 4618-4625.
- Phatak P, Cookson JC, Dai F, Smith V, Gartenhaus RB, Stevens MF, Burger AM. 2007. Telomere uncapping by the G-quadruplex ligand RHPS4 inhibits clonogenic tumour cell growth in vitro and in vivo consistent with a cancer stem cell targeting mechanism. *Br J Cancer* 96: 1223-1233.
- Pike JE, Burgers PM, Campbell JL, Bambara RA. 2009. Pif1 helicase lengthens some Okazaki fragment flaps necessitating Dna2 nuclease/helicase action in the two-nuclease processing pathway. *J Biol Chem* 284: 25170-25180.
- Poli J, Tsaponina O, Crabbe L, Keszthelyi A, Pantesco V, Chabes A, Lengronne A, Pasero P. 2012. dNTP pools determine fork progression and origin usage under replication stress. *EMBO J* 31: 883-894.
- Pollard LM, Sharma R, Gomez M, Shah S, Delatycki MB, Pianese L, Monticelli A, Keats BJ, Bidichandani SI. 2004. Replication-mediated instability of the GAA triplet repeat mutation in Friedreich ataxia. *Nucleic Acids Res* 32: 5962-5971.
- Popuri V, Bachrati CZ, Muzzolini L, Mosedale G, Costantini S, Giacomini E, Hickson ID, Vindigni A. 2008. The Human RecQ helicases, BLM and RECQ1, display distinct DNA substrate specificities. *J Biol Chem* 283: 17766-17776.
- Prakash A, Natarajan A, Marky LA, Ouellette MM, Borgstahl GE. 2011. Identification of the DNA-Binding Domains of Human Replication Protein A That Recognize G-Quadruplex DNA. *J Nucleic Acids* 2011: 896947.
- Prakash S, Johnson RE, Prakash L. 2005. Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. *Annu Rev Biochem* 74: 317-353.
- Putnam CD, Hayes TK, Kolodner RD. 2009. Specific pathways prevent duplication-mediated genome rearrangements. *Nature* 460: 984-989.
- Putnam CD, Pennaneach V, Kolodner RD. 2004. Chromosome healing through terminal deletions generated by de novo telomere additions in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 101: 13262-13267.
- Qin Y, Fortin JS, Tye D, Gleason-Guzman M, Brooks TA, Hurley LH. 2010. Molecular cloning of the human platelet-derived growth factor receptor beta (PDGFR-beta) promoter and drug targeting of the G-quadruplex-forming region to repress PDGFR-beta expression. *Biochemistry* 49: 4208-4219.
- Qin Y, Hurley LH. 2008. Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie* 90: 1149-1171.
- Qin Y, Rezler EM, Gokhale V, Sun D, Hurley LH. 2007. Characterization of the G-quadruplexes in the duplex nuclease hypersensitive element of the PDGF-A promoter and modulation of PDGF-A promoter activity by TMPyP4. *Nucleic Acids Res* 35: 7698-7713.
- Rachwal PA, Brown T, Fox KR. 2007a. Effect of G-tract length on the topology and stability of intramolecular DNA quadruplexes. *Biochemistry* 46: 3036-3044.
- . 2007b. Sequence effects of single base loops in intramolecular quadruplex DNA. *FEBS Lett* 581: 1657-1660.
- Rachwal PA, Findlow IS, Werner JM, Brown T, Fox KR. 2007c. Intramolecular DNA quadruplexes with different arrangements of short and long loops. *Nucleic Acids Res* 35: 4214-4222.
- Rachwal PA, Fox KR. 2007. Quadruplex melting. *Methods* 43: 291-301.

- Raghuraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, Conway A, Lockhart DJ, Davis RW, Brewer BJ, Fangman WL. 2001. Replication dynamics of the yeast genome. *Science* 294: 115-121.
- Randall GL, Zechiedrich L, Pettitt BM. 2009. In the absence of writhe, DNA relieves torsional stress with localized, sequence-dependent structural failure to preserve B-form. *Nucleic Acids Res* 37: 5568-5577.
- Rankin S, Reszka AP, Huppert J, Zloh M, Parkinson GN, Todd AK, Ladame S, Balasubramanian S, Neidle S. 2005. Putative DNA quadruplex formation within the human c-kit oncogene. *J Am Chem Soc* 127: 10584-10589.
- Reinhold WC, Mergny JL, Liu H, Ryan M, Pfister TD, Kinders R, Parchment R, Doroshow J, Weinstein JN, Pommier Y. 2010. Exon array analyses across the NCI-60 reveal potential regulation of TOP1 by transcription pausing at guanosine quartets in the first intron. *Cancer Res* 70: 2191-2203.
- Ren J, Chaires JB. 1999. Sequence and structural selectivity of nucleic acid binding ligands. *Biochemistry* 38: 16067-16075.
- Richard GF, Kerrest A, Dujon B. 2008. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol Mol Biol Rev* 72: 686-727.
- Riou JF, Guittat L, Mailliet P, Laoui A, Renou E, Petitgenet O, Megnin-Chanet F, Helene C, Mergny JL. 2002. Cell senescence and telomere shortening induced by a new series of specific G-quadruplex DNA ligands. *Proc Natl Acad Sci U S A* 99: 2672-2677.
- Risitano A, Fox KR. 2004. Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Res* 32: 2598-2606.
- Rizzo A, Salvati E, Porru M, D'Angelo C, Stevens MF, D'Incalci M, Leonetti C, Gilson E, Zupi G, Biroccio A. 2009. Stabilization of quadruplex DNA perturbs telomere replication leading to the activation of an ATR-dependent ATM signaling pathway. *Nucleic Acids Res* 37: 5353-5364.
- Roca J. 2011. The torsional state of DNA within the chromosome. *Chromosoma* 120: 323-334.
- Rodriguez R, Miller KM, Forment JV, Bradshaw CR, Nikan M, Britton S, Oelschlaegel T, Xhemalce B, Balasubramanian S, Jackson SP. 2012. Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat Chem Biol* 8: 301-310.
- Rose AM, Joyce PB, Hopper AK, Martin NC. 1992. Separate information required for nuclear and subnuclear localization: additional complexity in localizing an enzyme shared by mitochondria and nuclei. *Mol Cell Biol* 12: 5652-5658.
- Rudd MK, Endicott RM, Friedman C, Walker M, Young JM, Osoegawa K, de Jong PJ, Green ED, Trask BJ. 2009. Comparative sequence analysis of primate subtelomeres originating from a chromosome fission event. *Genome Res* 19: 33-41.
- Ruskin B, Fink GR. 1993. Mutations in POL1 increase the mitotic instability of tandem inverted repeats in *Saccharomyces cerevisiae*. *Genetics* 134: 43-56.
- Sacca B, Lacroix L, Mergny JL. 2005. The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. *Nucleic Acids Res* 33: 1182-1192.
- Salas TR, Petrusseva I, Lavrik O, Bourdoncle A, Mergny JL, Favre A, Saintome C. 2006. Human replication protein A unfolds telomeric G-quadruplexes. *Nucleic Acids Res* 34: 4857-4865.
- Salvati E, Leonetti C, Rizzo A, Scarsella M, Mottolese M, Galati R, Sperduti I, Stevens MF, D'Incalci M, Blasco M et al. 2007. Telomere damage induced by the G-quadruplex ligand RHPS4 has an antitumor effect. *J Clin Invest* 117: 3236-3247.
- Samadashwily GM, Raca G, Mirkin SM. 1997. Trinucleotide repeats affect DNA replication in vivo. *Nat Genet* 17: 298-304.
- Sanchez-Berrondo J, Mesa P, Ibarra A, Martinez-Jimenez MI, Blanco L, Mendez J, Boskovic J, Montoya G. 2012. Molecular architecture of a multifunctional MCM complex. *Nucleic Acids Res* 40: 1366-1380.
- Sanders CM. 2010. Human Pif1 helicase is a G-quadruplex DNA-binding protein with G-quadruplex DNA-unwinding activity. *Biochem J* 430: 119-128.
- Sarkies P, Reams C, Simpson LJ, Sale JE. 2010. Epigenetic instability due to defective replication of structured DNA. *Mol Cell* 40: 703-713.

- Schaffitzel C, Berger I, Postberg J, Hanes J, Lipps HJ, Pluckthun A. 2001. In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylonychia lemnae* macronuclei. *Proc Natl Acad Sci U S A* 98: 8572-8577.
- Schulz VP, Zakian VA. 1994. The *saccharomyces* PIF1 DNA helicase inhibits telomere elongation and *de novo* telomere formation. *Cell* 76: 145-155.
- Schwartz EK, Heyer WD. 2011. Processing of joint molecule intermediates by structure-selective endonucleases during homologous recombination in eukaryotes. *Chromosoma* 120: 109-127.
- Sen D, Gilbert W. 1988. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* 334: 364-366.
- Sfeir A, Kosiyatrakul ST, Hockemeyer D, MacRae SL, Karlseder J, Schildkraut CL, de Lange T. 2009. Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell* 138: 90-103.
- Shachar S, Ziv O, Avkin S, Adar S, Wittschieben J, Reissner T, Chaney S, Friedberg EC, Wang Z, Carell T et al. 2009. Two-polymerase mechanisms dictate error-free and error-prone translesion DNA synthesis in mammals. *EMBO J* 28: 383-393.
- Shaffer LG, Kashork CD, Saleki R, Rorem E, Sundin K, Ballif BC, Bejjani BA. 2006. Targeted genomic microarray analysis for identification of chromosome abnormalities in 1500 consecutive clinical cases. *J Pediatr* 149: 98-102.
- Shida T, Yokoyama K, Tamai S, Sekiguchi J. 1991. Self-association of telomeric short oligodeoxyribonucleotides containing a dG cluster. *Chem Pharm Bull (Tokyo)* 39: 2207-2211.
- Shimada K, Pasero P, Gasser SM. 2002. ORC and the intra-S-phase checkpoint: a threshold regulates Rad53p activation in S phase. *Genes Dev* 16: 3236-3252.
- Shirude PS, Okumus B, Ying L, Ha T, Balasubramanian S. 2007. Single-molecule conformational analysis of G-quadruplex formation in the promoter DNA duplex of the proto-oncogene *c-kit*. *J Am Chem Soc* 129: 7484-7485.
- Shishkin AA, Voineagu I, Matera R, Cherng N, Chernet BT, Krasilnikova MM, Narayanan V, Lobachev KS, Mirkin SM. 2009. Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. *Mol Cell* 35: 82-92.
- Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. 2002. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress *c-MYC* transcription. *Proc Natl Acad Sci U S A* 99: 11593-11598.
- Simonsson T. 2001. G-quadruplex DNA structures--variations on a theme. *Biol Chem* 382: 621-628.
- Sissi C, Gatto B, Palumbo M. 2011. The evolving world of protein-G-quadruplex recognition: a medicinal chemist's perspective. *Biochimie* 93: 1219-1230.
- Smith DJ, Whitehouse I. 2012. Intrinsic coupling of lagging-strand synthesis to chromatin assembly. *Nature* 483: 434-438.
- Smith JS, Chen Q, Yatsunyk LA, Nicoludis JM, Garcia MS, Kranaster R, Balasubramanian S, Monchaud D, Teulade-Fichou MP, Abramowitz L et al. 2011. Rudimentary G-quadruplex-based telomere capping in *Saccharomyces cerevisiae*. *Nat Struct Mol Biol* 18: 478-485.
- Sogo JM, Lopes M, Foiani M. 2002. Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects. *Science* 297: 599-602.
- Stegle O, Payet L, Mergny JL, MacKay DJ, Leon JH. 2009. Predicting and understanding the stability of G-quadruplexes. *Bioinformatics* 25: i374-382.
- Subramanian M, Rage F, Tabet R, Flatter E, Mandel JL, Moine H. 2011. G-quadruplex RNA structure as a signal for neurite mRNA targeting. *EMBO Rep* 12: 697-704.
- Sun D, Guo K, Rusche JJ, Hurley LH. 2005. Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res* 33: 6070-6080.



- Sun D, Hurley LH. 2009. The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. *J Med Chem* 52: 2863-2874.
- . 2010. Biochemical techniques for the characterization of G-quadruplex structures: EMSA, DMS footprinting, and DNA polymerase stop assay. *Methods Mol Biol* 608: 65-79.
- Sun D, Thompson B, Cathers BE, Salazar M, Kerwin SM, Trent JO, Jenkins TC, Neidle S, Hurley LH. 1997. Inhibition of human telomerase by a G-quadruplex-interactive compound. *J Med Chem* 40: 2113-2116.
- Sun H, Bennett RJ, Maizels N. 1999. The *Saccharomyces cerevisiae* Sgs1 helicase efficiently unwinds G-G paired DNAs. *Nucl Acids Res* 27: 1978-1984.
- Sun H, Karow JK, Hickson ID, Maizels N. 1998. The Bloom's syndrome helicase unwinds G4 DNA. *J Biol Chem* 273: 27587-27592.
- Sun H, Yabuki A, Maizels N. 2001. A human nuclease specific for G4 DNA. *Proc Natl Acad Sci U S A* 98: 12444-12449.
- Tang W, Dominska M, Greenwell PW, Harvanek Z, Lobachev KS, Kim HM, Narayanan V, Mirkin SM, Petes TD. 2011. Friedreich's ataxia (GAA)<sub>n</sub>\*(TTC)<sub>n</sub> repeats strongly stimulate mitotic crossovers in *Saccharomyces cerevisiae*. *PLoS Genet* 7: e1001270.
- Tauchi T, Shin-Ya K, Sashida G, Sumi M, Nakajima A, Shimamoto T, Ohyashiki JH, Ohyashiki K. 2003. Activity of a novel G-quadruplex-interactive telomerase inhibitor, telomestatin (SOT-095), against human leukemia cells: involvement of ATM-dependent DNA damage response pathways. *Oncogene* 22: 5338-5347.
- Tercero JA, Longhese MP, Diffley JF. 2003. A central role for DNA replication forks in checkpoint activation and response. *Mol Cell* 11: 1323-1336.
- Thomas M, White RL, Davis RW. 1976. Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proc Natl Acad Sci U S A* 73: 2294-2298.
- Todd AK, Johnston M, Neidle S. 2005. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res* 33: 2901-2907.
- Todd AK, Neidle S. 2008. The relationship of potential G-quadruplex sequences in cis-upstream regions of the human genome to SP1-binding elements. *Nucleic Acids Res* 36: 2700-2704.
- Tornaletti S, Park-Snyder S, Hanawalt PC. 2008. G4-forming sequences in the non-transcribed DNA strand pose blocks to T7 RNA polymerase and mammalian RNA polymerase II. *J Biol Chem* 283: 12756-12762.
- Tourriere H, Versini G, Cordon-Preciado V, Alabert C, Pasero P. 2005. Mrc1 and Tof1 promote replication fork progression and recovery independently of Rad53. *Mol Cell* 19: 699-706.
- Tran PL, Mergny JL, Alberti P. 2010. Stability of telomeric G-quadruplexes. *Nucl Acids Res*.
- Ulrich HD. 2011. Timing and spacing of ubiquitin-dependent DNA damage bypass. *FEBS Lett* 585: 2861-2867.
- Ulrich HD, Walden H. 2010. Ubiquitin signalling in DNA replication and repair. *Nat Rev Mol Cell Biol* 11: 479-489.
- Usdin K, Woodford KJ. 1995. CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis in vitro. *Nucleic Acids Res* 23: 4202-4209.
- Vannier JB, Pavicic-Kaltenbrunner V, Petalcorin MI, Ding H, Boulton SJ. 2012. RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. *Cell* 149: 795-806.
- Vaughn JP, Creacy SD, Routh ED, Joyner-Butt C, Jenkins GS, Pauli S, Nagamine Y, Akman SA. 2005. The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. *J Biol Chem* 280: 38117-38120.
- Venczel EA, Sen D. 1993. Parallel and antiparallel G-DNA structures from a complex telomeric sequence. *Biochemistry* 32: 6220-6228.
- Voineagu I, Narayanan V, Lobachev KS, Mirkin SM. 2008. Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins. *Proc Natl Acad Sci U S A* 105: 9936-9941.

- Voineagu I, Surka CF, Shishkin AA, Krasilnikova MM, Mirkin SM. 2009. Replisome stalling and stabilization at CGG repeats, which are responsible for chromosomal fragility. *Nat Struct Mol Biol* 16: 226-228.
- Walter J, Sun L, Newport J. 1998. Regulated chromosomal DNA replication in the absence of a nucleus. *Mol Cell* 1: 519-529.
- Wang H, Nora GJ, Ghodke H, Opresko PL. 2011a. Single molecule studies of physiologically relevant telomeric tails reveal POT1 mechanism for promoting G-quadruplex unfolding. *J Biol Chem* 286: 7479-7489.
- Wang Q, Liu JQ, Chen Z, Zheng KW, Chen CY, Hao YH, Tan Z. 2011b. G-quadruplex formation at the 3' end of telomere DNA inhibits its extension by telomerase, polymerase and unwinding by helicase. *Nucleic Acids Res* 39: 6229-6237.
- Wang Y, Patel DJ. 1993. Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure* 1: 263-282.
- Wang Y, Zhang H, Ligon LA, McGown LB. 2009. Association of insulin-like growth factor 2 with the insulin-linked polymorphic region in cultured fetal thymus cells. *Biochemistry* 48: 8189-8194.
- Washington MT, Minko IG, Johnson RE, Haracska L, Harris TM, Lloyd RS, Prakash S, Prakash L. 2004. Efficient and error-free replication past a minor-groove N2-guanine adduct by the sequential action of yeast Rev1 and DNA polymerase zeta. *Mol Cell Biol* 24: 6900-6906.
- Watson JD, Crick FH. 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171: 737-738.
- Weitzmann MN, Woodford KJ, Usdin K. 1997. DNA secondary structures and the evolution of hypervariable tandem arrays. *J Biol Chem* 272: 9517-9523.
- Wheelhouse RT, Sun D, Han H, Han FX, Hurley LH. 1998. Cationic Porphyrins as Telomerase Inhibitors: the Interaction of Tetra-(N-methyl-4-pyridyl)porphine with Quadruplex DNA. *J Am Chem Soc* 120: 3261-3262.
- Wieland M, Hartig JS. 2007. RNA quadruplex-based modulation of gene expression. *Chem Biol* 14: 757-763.
- Williamson JR, Raghuraman MK, Cech TR. 1989. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell* 59: 871-880.
- Wiltrout ME, Walker GC. 2011. The DNA polymerase activity of *Saccharomyces cerevisiae* Rev1 is biologically significant. *Genetics* 187: 21-35.
- Wlodarczyk A, Grzybowski P, Patkowski A, Dobek A. 2005. Effect of ions on the polymorphism, effective charge, and stability of human telomeric DNA. Photon correlation spectroscopy and circular dichroism studies. *J Phys Chem B* 109: 3594-3605.
- Wong HM, Stegle O, Rodgers S, Huppert JL. 2010. A toolbox for predicting g-quadruplex formation and stability. *J Nucleic Acids* 2010.
- Wood A, Garg P, Burgers PM. 2007. A ubiquitin-binding motif in the translesion DNA polymerase Rev1 mediates its essential functional interaction with ubiquitinated proliferating cell nuclear antigen in response to DNA damage. *J Biol Chem* 282: 20256-20263.
- Woodford KJ, Howell RM, Usdin K. 1994. A novel K(+)-dependent DNA synthesis arrest site in a commonly occurring sequence motif in eukaryotes. *J Biol Chem* 269: 27029-27035.
- Wu X, Maizels N. 2001. Substrate-specific inhibition of RecQ helicase. *Nucleic Acids Res* 29: 1765-1771.
- Wu Y, Sommers JA, Khan I, de Winter JP, Brosh RM, Jr. 2012. Biochemical characterization of Warsaw breakage syndrome helicase. *J Biol Chem* 287: 1007-1021.
- Wyatt JR, Davis PW, Freier SM. 1996. Kinetics of G-quartet-mediated tetramer formation. *Biochemistry* 35: 8002-8008.
- Xiao J, Carter JA, Frederick KA, McGown LB. 2009. A genome-inspired DNA ligand for the affinity capture of insulin and insulin-like growth factor-2. *J Sep Sci* 32: 1654-1664.
- Xiao J, McGown LB. 2009. Mass spectrometric determination of ILPR G-quadruplex binding sites in insulin and IGF-2. *J Am Soc Mass Spectrom* 20: 1974-1982.
- Yabuki N, Terashima H, Kitada K. 2002. Mapping of early firing origins on a replication profile of budding yeast. *Genes Cells* 7: 781-789.

- Yang Z, Lau R, Marcadier JL, Chitayat D, Pearson CE. 2003. Replication inhibitors modulate instability of an expanded trinucleotide repeat at the myotonic dystrophy type 1 disease locus in human cells. *Am J Hum Genet* 73: 1092-1105.
- Ying L, Green JJ, Li H, Klenerman D, Balasubramanian S. 2003. Studies on the structure and dynamics of the human telomeric G quadruplex by single-molecule fluorescence resonance energy transfer. *Proc Natl Acad Sci U S A* 100: 14629-14634.
- Youds JL, O'Neil NJ, Rose AM. 2006. Homologous recombination is required for genome stability in the absence of DOG-1 in *Caenorhabditis elegans*. *Genetics* 173: 697-708.
- Yu HQ, Miyoshi D, Sugimoto N. 2006a. Characterization of structure and stability of long telomeric DNA G-quadruplexes. *J Am Chem Soc* 128: 15461-15468.
- . 2006b. Properties of long human telomeric DNAs under cell-mimicking conditions. *Nucleic Acids Symp Ser (Oxf)*: 207-208.
- Yue DJ, Lim KW, Phan AT. 2011. Formation of (3+1) G-quadruplexes with a long loop by human telomeric DNA spanning five or more repeats. *J Am Chem Soc* 133: 11462-11465.
- Zahler AM, Williamson JR, Cech TR, Prescott DM. 1991. Inhibition of telomerase by G-quartet DNA structures. *Nature* 350: 718-720.
- Zaug AJ, Podell ER, Cech TR. 2005. Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proc Natl Acad Sci U S A* 102: 10864-10869.
- Zechiedrich EL, Khodursky AB, Bachellier S, Schneider R, Chen D, Lilley DM, Cozzarelli NR. 2000. Roles of topoisomerases in maintaining steady-state DNA supercoiling in *Escherichia coli*. *J Biol Chem* 275: 8103-8113.
- Zhang F, Carvalho CM, Lupski JR. 2009. Complex human chromosomal and genomic rearrangements. *Trends Genet* 25: 298-307.
- Zhang H, Freudenreich CH. 2007. An AT-rich sequence in human common fragile site FRA16D causes fork stalling and chromosome breakage in *S. cerevisiae*. *Mol Cell* 27: 367-379.
- Zhang ML, Tong XJ, Fu XH, Zhou BO, Wang J, Liao XH, Li QJ, Shen N, Ding J, Zhou JQ. 2010. Yeast telomerase subunit Est1p has guanine quadruplex-promoting activity that is required for telomere elongation. *Nat Struct Mol Biol* 17: 202-209.
- Zhang R, Lin Y, Zhang CT. 2008. Greglist: a database listing potential G-quadruplex regulated genes. *Nucleic Acids Res* 36: D372-376.
- Zhang W, Durocher D. 2010. De novo telomere formation is suppressed by the Mec1-dependent inhibition of Cdc13 accumulation at DNA breaks. *Genes Dev* 24: 502-515.
- Zhao X, Chabes A, Domkin V, Thelander L, Rothstein R. 2001. The ribonucleotide reductase inhibitor Sml1 is a new target of the Mec1/Rad53 kinase cascade during growth and in response to DNA damage. *EMBO J* 20: 3544-3553.
- Zheng KW, Chen Z, Hao YH, Tan Z. 2010. Molecular crowding creates an essential environment for the formation of stable G-quadruplexes in long double-stranded DNA. *Nucleic Acids Res* 38: 327-338.
- Zhou J, Monson EK, Teng S, Schulz VP, Zakian VA. 2000. Pif1p helicase, a catalytic inhibitor of telomerase in yeast. *Science* 289: 771-774.
- Zhu Z, Chung WH, Shim EY, Lee SE, Ira G. 2008. Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. *Cell* 134: 981-994.
- Zimmerman SB, Cohen GH, Davies DR. 1975. X-ray fiber diffraction and model-building study of polyguanylic acid and polyinosinic acid. *J Mol Biol* 92: 181-192.
- Zou H, Rothstein R. 1997. Holliday junctions accumulate in replication mutants via a RecA homolog-independent mechanism. *Cell* 90: 87-96.
- Zou L, Elledge SJ. 2003. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science* 300: 1542-1548.
- Zou L, Liu D, Elledge SJ. 2003. Replication protein A-mediated recruitment and activation of Rad17 complexes. *Proc Natl Acad Sci U S A* 100: 13827-13832.