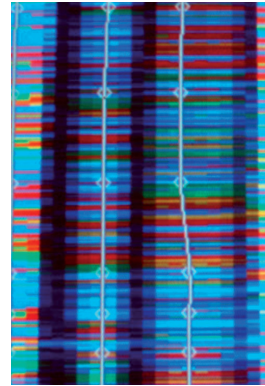


► La duplication correcte du génome humain est sous la dépendance d'un programme spatio-temporel qui détermine où et quand commencent les fourches de réplication. Cette régulation s'opère donc principalement au niveau de leurs points de départ, appelés origines de réplication. Environ 50 000 origines sont déclenchées dans une cellule humaine au cours de la phase S. La progression normale, ou perturbée, des fourches de réplication a un impact sur le déroulement de la réplication. Récemment, de nombreux travaux ont révélé le rôle d'une structure non canonique de l'ADN, le G-quadruplex, dans le processus de duplication des génomes chez les vertébrés. Nous décrivons dans cette revue, l'impact majeur de cette structure sur le départ et sur la progression des fourches de réplication. ◀

Le génome des mammifères est une mosaïque de segments d'ADN, de l'ordre de la mégabase, qui se distinguent par leur richesse en bases guanines et cytosines (G et C). Cette organisation a été observée bien avant le séquençage massif de nombreux génomes, les segments identifiés étant alors appelés isochores [1]. Les isochores qui sont riches ou pauvres en GC, correspondent à des régions chromosomiques aux propriétés bien distinctes. Les isochores riches en GC sont en effet riches en gènes et en séquences régulatrices. Les isochores pauvres en GC, mais riches en AT (adénine et thymine), présentent, eux, peu de gènes. Cette organisation est étroitement liée au programme temporel de la duplication des génomes [2]. Chez l'ensemble des eucaryotes, les génomes sont en effet répliqués selon un programme temporel bien précis. Chez l'homme, les domaines correspondant approximativement aux régions riches en GC et à celles riches en AT, sont répliqués respectivement en début et en fin de phase S (Figure 1). Les régions de composition intermédiaire sont plus flexibles. Elles seront répliquées selon le type cellulaire [3]. Ces régions

G-quadruplex : acteurs majeurs de la duplication du génome humain

Jérémy Poulet-Benedetti², Anne-Laure Valton¹, Marie-Noëlle Prioleau²



¹Howard Hughes Medical Institute, Program in systems biology, Department of biochemistry and molecular pharmacology, University of Massachusetts Medical School, 368 Plantation Street, Worcester, MA 01605-0103, États-Unis.

²Institut Jacques Monod, CNRS UMR7592, université Paris Diderot, Équipe labellisée ARC (Association pour la recherche sur le cancer), 75013 Paris, France.

marie-noelle.prioleau@ijm.fr

se trouvent également dans les zones

de transition, entre domaines précoces et tardifs. Bien qu'universel, la fonction biologique du programme temporel de réplication reste énigmatique. À ce programme, s'ajoute un programme spatial qui détermine d'où partent les fourches de réplication (origines de réplication, ORI). Au cours de la phase S, la coordination du déclenchement des origines de réplication assure la duplication correcte du génome, selon un programme temporel relativement précis. Des études pionnières, fondées sur le marquage radiographique de bulles de réplication¹, ont permis d'estimer la distance moyenne entre deux origines de réplication à environ 70 kilobases (kb), et ainsi d'évaluer le nombre d'origines actives lors de la phase S, à 50 000 origines par génome dans des cellules humaines [4]. Plus récemment, de nombreuses approches génomiques ont rendu possible la cartographie des sites préférentiels d'initiation de la réplication [5, 6] (→). Les origines de réplication ne sont pas réparties de façon homogène le long du génome. Les régions chromosomiques riches en GC et en gènes sont en effet également riches en origines fortes de réplication (Figure 1). Inversement, les régions répliquées en fin de phase S, comprennent peu d'origines fortes de réplication. Les origines fortes se déclenchent dans une grande proportion de cellules au sein d'une population et, par conséquent, sont détectables dans les analyses génomiques. Les origines faibles, qui sont utilisées au contraire dans une petite fraction de cellules, ne pourront être détectées en utilisant certaines approches. Les régions tardives pourraient être répliquées soit passivement, par des fourches initiées dans les régions voisines, soit

(→) Voir la Synthèse de B. Miotto, m/s n° 2, février 2017, page 143.

¹ La réplication commence aux origines de réplication (ORI) puis s'étend sous la forme de bulles de réplication. Chaque bulle comporte deux fourches.

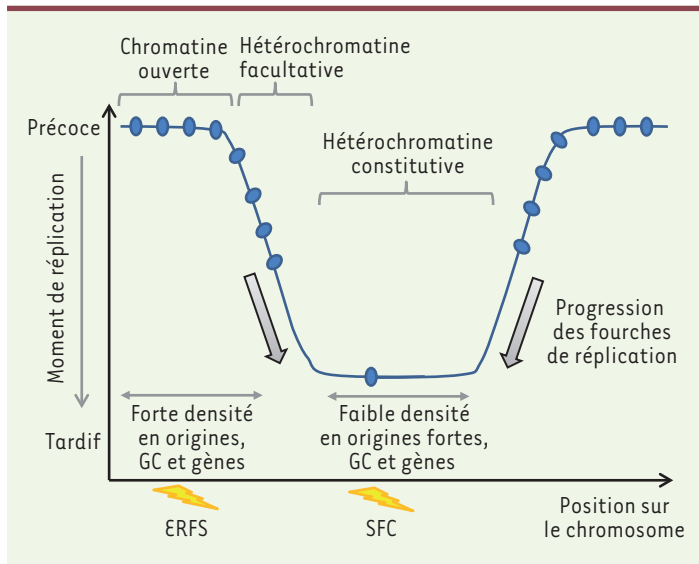


Figure 1. Représentation schématique des domaines de moment de répliation. Les domaines précoces et tardifs ont une taille de 200 kilobases à 2 mégabases et sont séparés par des régions intermédiaires où les fourches de répliation transitent majoritairement des domaines précoces vers les domaines tardifs. Les sites fragiles communs (SFC) sont dans les domaines tardifs de répliation alors que les « *early replicating fragile sites* » (ERFS) se trouvent dans les régions précoces. Les caractéristiques des domaines précoces et tardifs sont bien distinctes. Les domaines précoces sont riches en GC, en gènes et en origines fortes ; de plus, ils contiennent de nombreuses marques actives de la chromatine. Inversement, les domaines tardifs sont pauvres en GC, en gènes et en origines fortes et présentent les caractéristiques de l'hétérochromatine : peu de marques ouvertes de la chromatine et un enrichissement dans la marque H3K9me3 (histone H3 tri-méthylée sur la lysine [K] 9).

par des origines peu efficaces, qui se déclenchent très aléatoirement dans une population de cellules. Comprendre comment ces régions précoces et tardives sont répliquées est important car des sites de cassures récurrents, identifiés dans des cellules cancéreuses, sont situés dans ces deux types de régions (SFC, sites fragiles communs, et ERFS, *early replicating fragile sites* ; Figure 1) [7] (→).

(→) Voir la Nouvelle de A. Letessier et al., m/s n° 8-9, août-septembre 2011, page 707

Ces cassures résulteraient de défauts de répliation. Le bon déroulement de la duplication du génome dépend, de plus, non seulement du bon contrôle des points de départ de la répliation, mais également de la progression normale des fourches de répliation dans les régions dépourvues d'origines très actives. Dans cette revue, nous nous focaliserons sur une structure particulière de l'ADN, le G-quadruplex (G4), apparue récemment comme un élément régulateur important de la répliation de l'ADN. Ces structures participent à de nombreuses autres fonctions, telles que la transcription, la traduction, la stabilité des ARN messagers (ARNm) ainsi qu'au fonctionnement des télomères. Des revues récentes décrivent le rôle potentiel des G4 dans ces multiples fonctions [8, 9].

Les G-quadruplex sont des structures de l'ADN non conventionnelles et sont répartis de façon non homogène le long du génome

L'ADN, constitué d'une succession de bases nucléiques (A, C, G et T), est normalement structuré en double hélice. Les G-quadruplex d'ADN, ou G4, sont des agencements de quatre brins qui peuvent adopter les acides nucléiques riches en guanine (G) [41] (→).

(→) Voir la nouvelle de D. Monchaud, page 1042 ce numéro

Le G4 repose sur l'appariement de quatre guanines par des liaisons de type Hoogsteen² formant un plateau, appelé quartet (Figure 2A).

² L'appariement de Hoogsteen est un type d'appariement canonique impliquant les purines trouvées dans les acides nucléiques. Les liaisons hydrogènes dans les paires Hoogsteen impliquent une face de la purine différente de la face Watson-Crick et en particulier les positions N7 et le groupement oxo ou amino en position 6.

Le « G4 classique » est constitué de l'empilement de trois quartets (Figure 2C).

À la suite du séquençage du génome humain, des logiciels ont été développés afin de localiser les séquences ayant le potentiel de se structurer en G4 (pG4). Initialement, cette recherche reposait sur l'identification d'un motif du type G₃N₁₋₇G₃N₁₋₇G₃N₁₋₇G₃, N pouvant être n'importe quelle base, dont une guanine ; on verra que la taille des boucles entre les séries de guanines joue un rôle important dans la stabilité des G4. Environ 370 000 pG4 ont ainsi été identifiés [10]. Plus récemment, une approche fondée sur l'utilisation du séquençage massif, couplé à la propriété des G4 de bloquer les ADN polymérases, a permis d'identifier 700 000 pG4 dans le génome humain [11]. Les pG4 ne sont pas répartis de façon homogène. En effet, leur richesse en guanines implique qu'ils sont fortement représentés dans les domaines riches en GC. Ils sont également particulièrement enrichis au niveau des promoteurs de gènes, notamment ceux des oncogènes. Ils s'accumulent aussi dans les régions 5' non traduites des ARNm, ainsi qu'au niveau des télomères où des répétitions du motif TTAGGG sont retrouvées. La surreprésentation des pG4 dans ces régions spécifiques du génome suggère qu'au cours de l'évolution, une pression de sélection les a maintenus et donc, que les pG4 joueraient un rôle important dans la dynamique des génomes. Bien que la démonstration de la formation des G4 *in vitro*, dans des conditions physiologiques, soit relativement ancienne, les preuves *in vivo* sont assez récentes. La première visualisation de la présence de G4 *in vivo* a été réalisée grâce à des anticorps spécifiques des G4 télomériques, chez un cilié qui a la particularité d'avoir un macro-noyau contenant une quantité massive de télomères portant

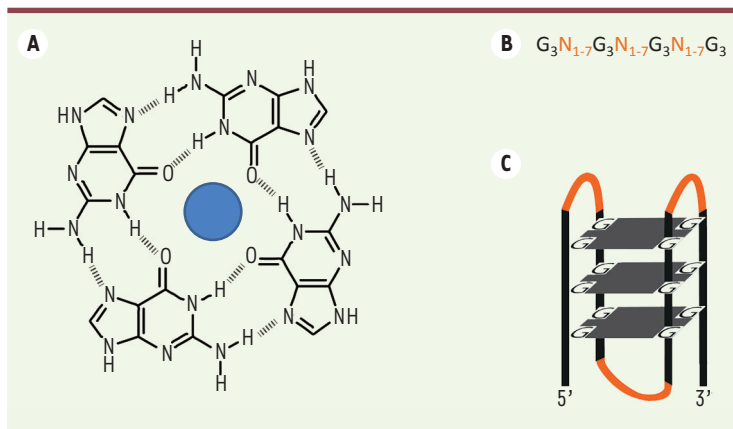


Figure 2. Séquence et structure d'un G-quadruplex. **A.** Structure d'un G-quartet, le cation monovalent qui le stabilise est représenté par un cercle bleu. **B.** Séquence consensus d'un G-quadruplex. **C.** Représentation d'un G4 intramoléculaire de type anti-parallèle. Il existe de nombreuses autres formes de type intermoléculaire et/ou parallèle où les brins ont la même orientation au sein de la structure.

des pG4 [12]. L'utilisation de drogues spécifiques, qui reconnaissent les G4, a ensuite permis d'étendre cette observation aux régions non télomériques, dans des cellules humaines [13]. Enfin, des expériences d'immuno-précipitation de la chromatine, utilisant des anticorps dirigés contre les G4, ont mis en évidence environ 10 000 sites de formation de G-quadruplex *in vivo* [14]. Ces G4 sont fortement associés à des promoteurs activement transcrits, ce qui suggère un lien entre formation de G4 et activation de la transcription *in vivo*.

Les pG4, une source potentielle d'instabilité génomique

La formation de G-quadruplex est favorisée par des processus qui supposent l'ouverture transitoire de la double hélice de l'ADN. La transcription et la réplication peuvent participer à cette ouverture. On ne peut cependant pas exclure que des contraintes topologiques fortes, induites par la fixation de facteurs liant l'ADN, puissent également être impliquées et suffire à l'ouverture et à la formation des G4. Une fois formés, les G4 sont extrêmement stables. Ils peuvent bloquer, *in vitro*, la progression des ADN polymérases. L'inactivation de gènes impliqués dans le déroulement des G4 (gènes codant des hélicases spécifiques des G-quadruplex) a permis de mettre en évidence le rôle potentiel de ces structures dans le blocage de la progression des fourches de réplication *in vivo*. La première démonstration découle de l'observation que la délétion du gène *DOG1* (qui code l'hélicase FANCF [*Fanconi anemia complementation group J*]) chez *Caenorhabditis elegans*, induit l'accumulation de petites délétions d'environ 50 à 300 nucléotides en amont de pG4 particuliers constitués principalement d'une succession de guanines [15]. La majorité des jonctions 3' de ces délétions sont localisées dans une petite zone de 3 paires de bases (pb), en aval de la première guanine du pG4. Cette observation est en accord avec l'hypothèse selon laquelle la jonction est définie par l'incapacité de l'ADN polymérase à progresser correctement à travers un G4 structuré (Figure 3). L'ensemble de ces données suggère ainsi

qu'en l'absence de FANCF, les pG4 ayant une petite boucle peuvent se structurer en G4 *in vivo*, au moins chez cet organisme. En accord avec cette observation réalisée chez *C. elegans*, des cellules humaines invalidées pour le gène *FANCF* accumulent également de larges délétions à proximité de certains pG4 [16]. La démonstration directe de l'implication de FANCF dans la réplication des G4 a été obtenue récemment, dans un système *in vitro* reproduisant la progression des brins naissants [17]. En parallèle, des études réalisées chez la levure *Saccharomyces cerevisiae*, ont montré le rôle crucial de Pif1 (*petite integration frequency protein 1*), une hélicase connue pour dérouler des structures G4 *in vitro*, dans le maintien de l'intégrité génomique [18]. La délétion du gène *PIF1* induit des réarrangements dans une séquence mini-satellite humaine riche en guanines (CEB1) répétée et insérée dans le génome de *S. cerevisiae* [19]. Cette instabilité peut également être induite

en ajoutant, dans le milieu de culture, un ligand fort des G4 (le Phen-DC₃) qui inhibe efficacement *in vitro*, le déroulement des G4 par l'hélicase Pif1. Cette instabilité est observée spécifiquement lorsque le brin riche en guanines se trouve sur le brin d'ADN direct, et non sur le brin rétrograde (Figure 3). Différentes mutations de nucléotides, réalisées dans les pG4 localisés dans la séquence CEB1, ont montré que des boucles très courtes (moins de 4 nucléotides) se révèlent être les structures les plus efficaces pour induire de l'instabilité génomique [20]. Le raccourcissement des boucles est également lié à une augmentation de la stabilité thermique du G4 *in vitro*. De ces résultats, environ 18 000 pG4 ont été décrits comme étant potentiellement délétères dans un contexte de cellules défectueuses pour une activité hélicase en charge de dérouler les G4.

pG4 et instabilité épigénétique

Des expériences réalisées dans des cellules de poulet (*Gallus gallus*) ont montré que la réplication retardée d'un pG4 placé sur le brin direct était suffisante pour induire la formation d'une brèche d'ADN simple brin en amont du pG4 (Figure 3) [21]. Ce blocage induit un découplage entre réplication de l'ADN et recyclage des histones parentales. Il est ainsi à l'origine de la perte progressive de l'information épigénétique sur une région de 4,5 kb en amont du pG4. Ce processus peut être suivi par la répression progressive d'un locus activement transcrit, *BU-1*⁴. L'extinction de Bu-1A, l'un

³ C₃₄H₂₆N₆O₂ · 2 CF₃O₃S.

⁴ BU-1 est exprimé par les précurseurs précoces des lymphocytes.

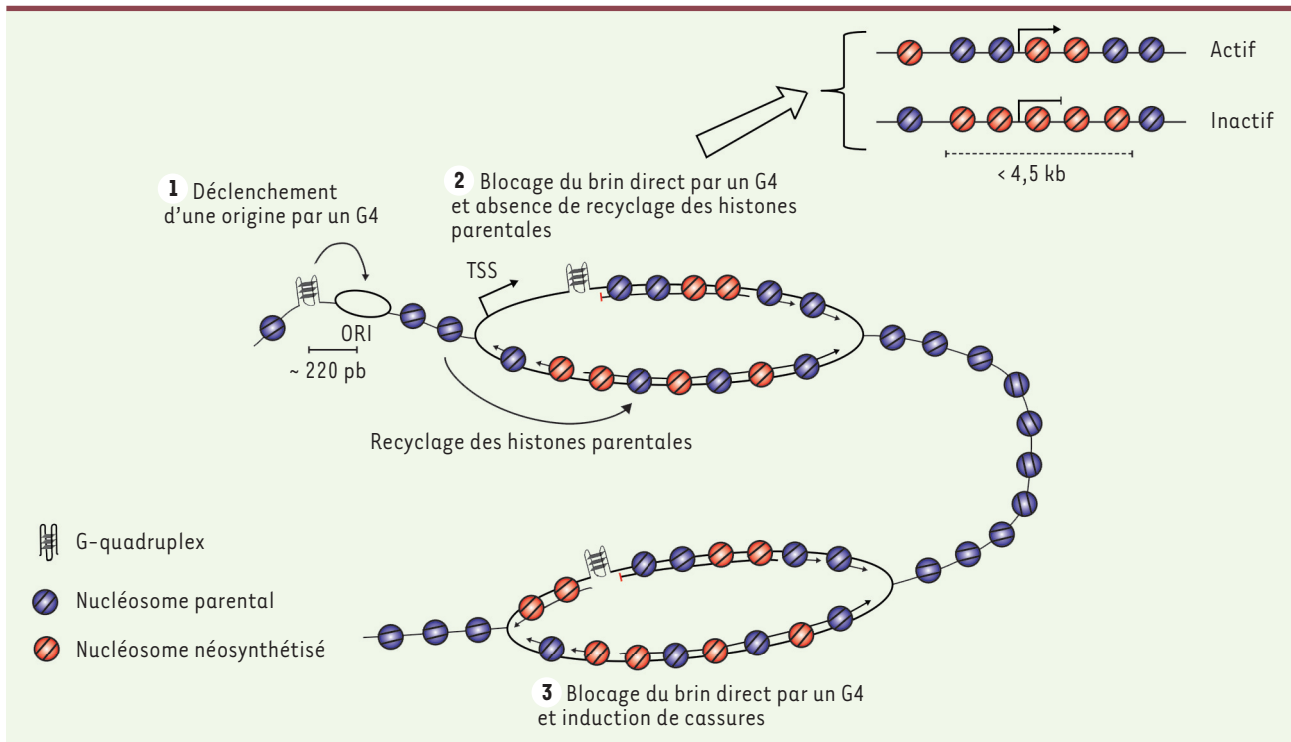


Figure 3. Rôles des pG4 dans la régulation de la réplication. Les pG4 jouent un rôle direct dans l'initiation de la réplication. Environ 80 % des origines fortes contiennent un pG4 localisé environ 220 paires de bases en amont du site d'initiation (ORI) (1). Dans des contextes génétiques déficients pour une activité hélicase spécifique des G4, la progression des fourches de réplication sur le brin direct peut être bloquée par la formation d'un G4. Cela peut induire alors une instabilité épigénétique liée à un délai entre la réplication et le recyclage des histones parentales (2). Un blocage de fourche peut également induire des cassures chromosomiques qui seront réparées mais pourront induire de l'instabilité génomique (3). pG4 : potentiel G-quadruplex ; TSS : transcription start site.

des alloantigènes⁵ de Bu-1, est associée à la perte de la marque de chromatine active H3K4me3 (tri-méthylation de la lysine 4 de l'histone H3) dans la région promotrice. Cette répression, dans un contexte mutant, est sous le contrôle d'un unique pG4 localisé 3,5 kb en aval du site d'initiation de la transcription (TSS, transcription start site), et le pG4 peut agir sur une distance maximale de 4,5 kb (Figure 3). Le suivi de la perte d'expression de Bu-1A a été réalisé dans plusieurs contextes mutants. L'absence de la polymérase REV1, ou des hélicases BLM (*Bloom syndrome protein*) et WRN (*Werner syndrome protein*) ou FANCD1, est suffisante pour induire une inactivation stochastique de Bu-1A au cours des divisions cellulaires [21-23]. Une étude plus récente, utilisant la même approche, a montré que l'ADN polymérase primase-polymérase (PrimPol) est également impliquée dans le passage à travers les G4 [24]. PrimPol a initialement été identifiée pour être impliquée dans le passage à travers les lésions photo-oxydatives grâce à ses activités de polymérase translésionnelle et de réamorçage. Dans cette nouvelle étude, les auteurs ont montré que PrimPol ne pouvait pas répliquer efficacement un G4 structuré, mais qu'elle avait, néanmoins, la capacité de s'y lier et d'amorcer la synthèse d'ADN dans la région voisine. PrimPol semble avoir une affinité supérieure pour

les pG4 avec une boucle courte et une forte stabilité thermique. En accord avec cette observation, ces pG4 induisent une plus forte instabilité de l'expression de Bu-1A dans un contexte mutant *PrimPol*. L'ensemble de ces travaux suggère que PrimPol assure le réamorçage de la réplication en amont d'un G4 structuré, permettant ainsi à la réplication bloquée de redémarrer. Ces études ont permis d'identifier de nouvelles voies impliquées dans la réplication correcte des G4 et ont confirmé que les G4 avec une boucle courte et une forte stabilité thermique sont plus fréquemment associés à des sites de pause de la réplication.

Les pG4 régulent la spécification des origines de réplication

Le développement récent de méthodes de cartographie des origines de réplication à haut débit a révélé un rôle inattendu des pG4 dans la réplication de l'ADN [25]. La méthode la plus résolutive est fondée sur la purification et la quantification des petits brins naissants (PBN), molécules spécifiques des origines de réplication [6]. Les PBN de 1,5 à 2 kb sont purifiés après extraction de

⁵ Molécule codée par différents allèles présents chez les individus d'une même espèce et dont l'antigénicité peut varier.

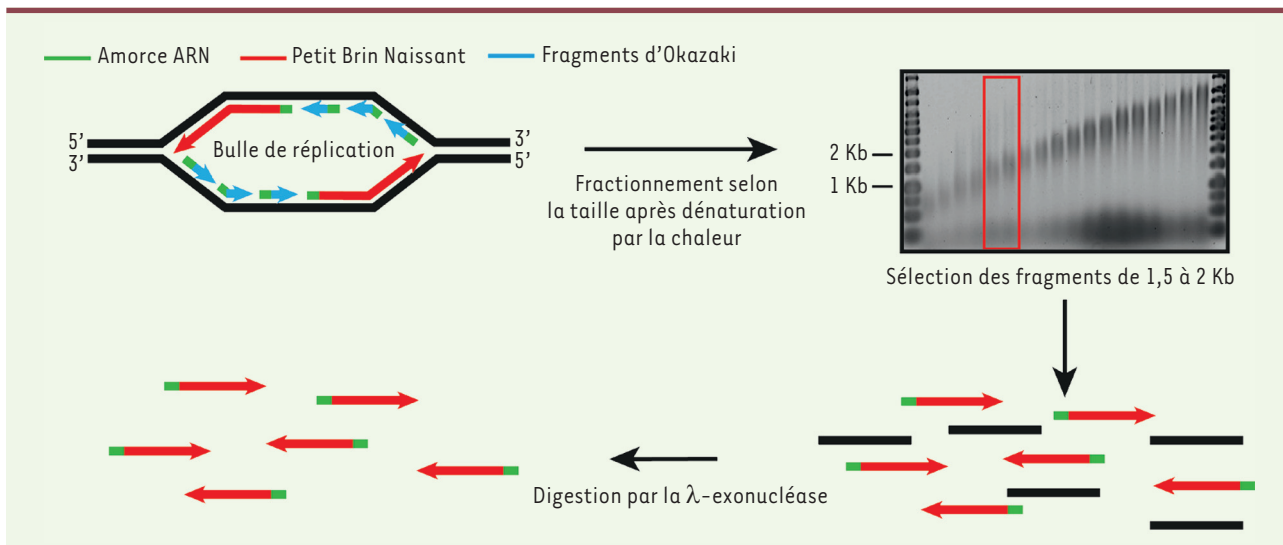


Figure 4. Préparation des petits brins naissants. Les petits brins naissants (PBN) sont des molécules caractéristiques des origines de réplication. Ils possèdent en 5' une amorce ARN synthétisée par la primase lors de l'initiation de la synthèse du brin complémentaire. Cette propriété permet de les enrichir après digestion par la λ -exonucléase. Après traitement par une kinase, l'ADN cassé contaminant est, lui, dégradé car il possède une extrémité 5' phosphate reconnue par la λ -exonucléase alors que les PBN sont protégés par leur amorce ARN. Pour préparer les PBN, de l'ADN génomique est extrait de cellules en phase exponentielle. Après dénaturation à la chaleur, l'ADN est fractionné sur gradient de sucrose. L'ADN compris entre 1,5 et 2 kilobases est sélectionné et soumis à une phosphorylation et à un traitement à la λ -exonucléase. Les origines de réplication sont caractérisées par un fort enrichissement en PBN qui peut être détecté à l'échelle du génome par séquençage massif ou localement par PCR (*polymerase chain reaction*) quantitative.

l'ADN génomique, dénaturation et séparation sur gradient de sucrose, puis digestion par la λ -exonucléase⁶ (Figure 4). Cette étape est nécessaire pour assurer la digestion du large excès d'ADN génomique cassé, les PBN étant protégés par une petite amorce ARN incorporée à leur extrémité 5' par la primase. La première cartographie à grande échelle de PBN, réalisée sur 1 % du génome humain, a apporté des données importantes sur les mécanismes moléculaires qui participent potentiellement à l'initiation de la réplication [26]. Les origines fortes de réplication ne sont pas distribuées au hasard. Elles sont particulièrement enrichies au niveau d'éléments régulateurs de la transcription, comme les TSS et les activateurs (ou « *enhancers* »). Les îlots CpG (CGI), qui sont associés à environ 1/3 des promoteurs, constituent les origines les plus efficaces, c'est-à-dire celles qui vont se déclencher dans une grande proportion de cycles cellulaires [27]. Dans une lignée cellulaire donnée, environ 35 % des origines sont associées à des CGI. Ces analyses ont été confirmées par des approches impliquant tout le génome. Elles ont également révélé l'enrichissement des origines en pG4 chez l'homme et la souris [28, 29]. Environ 80 % des origines sont co-localisées avec un pG4, mais tous les pG4 ne sont pas associés à des origines, suggérant qu'ils pourraient être des déterminants importants des origines de réplication mais qu'ils ne sont pas suffisants [30]. L'observation que les pG4 pouvaient potentiellement contrôler le départ des fourches de réplication a été une surprise, compte tenu du rôle connu des pG4 dans le blocage de ces mêmes fourches. Il apparaissait

alors concevable que de nombreux sites d'initiation de la réplication soient choisis de manière stochastique dans des zones d'initiation, et que ceux proches des pG4 soient anormalement enrichis à cause d'une pause transitoire de la fourche au niveau d'un G4. Cette hypothèse est en accord avec les pics d'enrichissement en PBN retrouvés dans les analyses à haut débit en 3' des pG4. La chute brutale d'enrichissement en PBN au niveau d'un pG4 suggère qu'il y a, en effet, une pause très transitoire au niveau des pG4 des origines. Des zones d'initiation de la réplication ont également été caractérisées par d'autres méthodes de cartographie d'origines, qui ont une plus faible résolution mais qui peuvent capturer les événements d'initiation à faible fréquence. Au vu de l'ensemble des résultats, il semblerait que l'initiation de la réplication chez les vertébrés soit plus efficace au niveau de sites spécifiques enrichis en pG4 dans des zones où se trouvent distribuées des origines moins efficaces. Il reste à estimer plus précisément la contribution de ces différents types d'origines dans le contrôle de la duplication des génomes de vertébrés.

Un test fonctionnel d'activité origine a été développé récemment afin de comprendre cette contribution [31]. Il a permis de déterminer si l'abondance forte en PBN, à proximité du pG4 d'une origine modèle, était effectivement caractéristique d'un site fort d'initiation de

⁶ Exonucléase isolée de *Escherichia coli*, infectée par le bactériophage lambda.

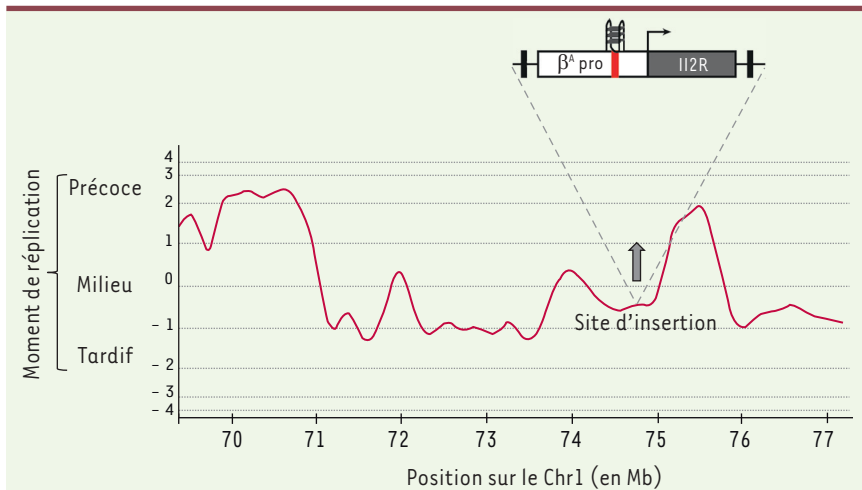


Figure 5. Nouveau test fonctionnel d'une activité origine. Une origine modèle (promoteur du gène β^A globine du poulet) associée au gène rapporteur IL-2R (récepteur de l'interleukine-2) a été insérée par recombinaison homologue dans la lignée aviaire DT40. Le graphe représente le moment de répliation d'une portion du chromosome 1 (Chr1) dans laquelle se trouve le site d'insertion du construit à tester. Ce site d'insertion est dans une région répliquée naturellement en milieu-fin de phase S. L'origine β^A contient un pG4 (ligne rouge). L'ajout de part et d'autre de sites de fixation du facteur USF (*upstream stimulating factor* –

rectangles noirs) permet à ce construit d'avancer localement le moment de répliation d'environ 45 min. Cet avancement dépend à la fois d'une origine fonctionnelle et des sites de fixation du facteur USF.

la répliation. Ce nouveau test est fondé sur la capacité d'une origine active à avancer le moment de répliation d'une région naturellement répliquée tardivement (Figure 5). En effet, l'avancement du moment de répliation d'une région chromosomique sera détectable dans une population de cellules, uniquement si le signal origine introduit assure un déclenchement dans la majorité des cellules. L'origine modèle utilisée se situe dans le promoteur du gène β^A globine de poulet et contient un pG4. Elle a été insérée par recombinaison homologue dans une région répliquée en milieu-fin de phase S grâce au fort taux de recombinaison homologue spécifique de la lignée aviaire DT40 (Figure 5). L'origine modèle β^A , seule, est capable d'induire un pic d'enrichissement en PBN, ce qui montre que ce fragment d'ADN de 1 kb est capable de fonctionner en position ectopique. Cependant elle est incapable d'avancer le moment de répliation, suggérant que son moment de déclenchement reste similaire à celui de la région modifiée. Lorsque des sites de fixation pour le facteur USF (*upstream stimulating factor*), qui a la capacité de recruter des facteurs modificateurs de la chromatine, sont placés de part et d'autre de l'origine β^A , un avancement du moment de répliation peut être observé [32]. Ce nouveau test fondé sur l'avancement du moment de répliation confirme bien que l'origine β^A est non seulement fonctionnelle mais très efficace. Le rôle du pG4 portant des petites boucles a été analysé en détail. Des mutations ponctuelles, diminuant la stabilité du G4 *in vitro*, ont également un impact fort sur l'activité origine, telle que testée par l'approche classique d'enrichissement en PBN, mais aussi avec le nouveau test fonctionnel du moment de répliation. Ces résultats ont montré sans équivoque qu'un G4 joue un rôle positif majeur dans le contrôle de l'origine modèle, et ceci indépendamment d'un rôle potentiel de blocage de fourche. Cette étude a également montré que l'orientation du pG4 au sein de l'origine β^A déterminait la position du site d'initiation de la répliation, en accord avec les données génomiques, qui montrent un pic d'enrichissement en PBN environ 220 pb en 3' des pG4 (Figure 3). Cette analyse a finalement montré qu'une coopération entre le pG4 et un module de 250 pb était nécessaire pour qu'une origine soit fonctionnelle. La nature de l'infor-

mation fournie par ce deuxième module reste à déterminer. La délétion de pG4 sur une autre origine modèle a montré l'importance de ces modules dans son contrôle. De plus, il a été observé que deux pG4 pouvaient être actifs de manière synergique dans cette origine. Dans leur ensemble, ces travaux apportent la démonstration formelle, au moins pour deux origines modèles, du rôle important des pG4 dans le contrôle des origines de répliation. C'est une avancée conceptuelle importante dans la compréhension des mécanismes moléculaires impliqués dans la duplication des génomes de vertébrés. Se pose désormais la question de comprendre comment les pG4 peuvent influencer positivement le déclenchement des origines de répliation. Cette régulation peut s'effectuer à deux niveaux correspondant à deux phases distinctes du cycle cellulaire : le premier se déroule en phase G1, durant laquelle les pré-complexes de répliation (pré-RC) se forment ; le deuxième durant la phase S, lorsque les pré-RC sont déclenchés par des kinases spécifiques de la phase S. À ce moment, les pré-RC ne peuvent en revanche plus se former, ce qui évite la répliation de régions qui ont déjà été dupliquées. Dans chaque cellule, seule une sous-fraction des pré-RC formés en excès en phase G1 est déclenchée au cours de la phase S suivante. Le choix des pré-RC déclenchés, ainsi que leur moment d'activation, sont également régulés. Les G4 pourraient donc favoriser le recrutement des pré-RC en phase G1. Cette facilitation pourrait reposer sur la capacité des pG4 à exclure les nucléosomes (Figure 3) [33]. Les régions dépourvues de nucléosomes pourraient, en effet, être plus favorablement reconnues et fixées par le pré-RC. Cette hypothèse est soutenue par l'observation, chez *S. cerevisiae*, que les régions dépourvues de nucléosomes sont des sites plus favorables à

l'assemblage des pré-RC [34, 35]. La formation de G4 pourrait néanmoins également favoriser l'ouverture de la double hélice et donc l'initiation de la réplication en phase S. Les pG4 pourraient également être reconnus par des facteurs spécifiques, importants pour le déclenchement d'une origine de réplication. En accord avec cette dernière hypothèse, le complexe ORC (*origin recognition complex*) humain, qui est le premier facteur à s'assembler pour former un pré-RC, n'a aucune spécificité de séquence *in vitro*, mais il reconnaît toutefois plus fortement les ARN structurés en G4 ainsi que les G4 d'ADN [36]. À noter que le facteur de déclenchement des origines RecQL4 (*RecQ-like helicase 4*) présente une affinité qui est importante pour les G4 [37].

Conclusions et perspectives

Au cours de cette dernière décennie, les G-quadruplex sont apparus comme des éléments *cis*-régulateurs majeurs de processus qui se déroulent sur le génome humain. Dans cette revue, nous n'avons discuté que de leur rôle dans la réplication de l'ADN. Il est maintenant formellement établi que des hélicases et des polymérases spécialisées sont impliquées dans la réplication correcte des pG4. En cas de défaillance, l'induction des cassures qui, par la suite, pourront être réparées, est une source d'instabilité génomique. Plusieurs maladies génétiques associées à une perte de fonction de ces facteurs causent un vieillissement prématuré et/ou un accroissement du risque de développer un cancer [8]. Une étude récente, utilisant le système modèle du gène *BU-1*, a montré qu'un stress réplicatif récurrent, induit par une limitation de la réserve de déoxynucléotides, était source d'instabilité épigénétique [38]. Il est donc vraisemblable que les G4 deviennent problématiques dans des conditions de stress réplicatifs induits par l'activation d'oncogènes dans les tumeurs primaires. Cette hypothèse est en accord avec l'observation que de nombreuses jonctions de translocations sont proches de pG4 dans les cellules cancéreuses [39, 40]. Tous les pG4 ne sont pas équivalents quant à leur potentielle dangerosité, ceux ayant une forte stabilité thermique associée à de petites boucles semblent en effet les plus problématiques à répliquer. La découverte récente que les pG4 participent au déclenchement des origines de réplication a permis de réévaluer leur rôle dans la duplication des génomes eucaryotes. Considérés exclusivement comme des sites de blocage des fourches de réplication, leur nouvelle fonction dans le déclenchement des origines de réplication leur confère un rôle régulateur majeur dans la duplication des génomes chez les vertébrés. Des approches génétiques sur des origines modèles permettront de déterminer si une sous-classe de pG4 est particulièrement impliquée dans ce processus, et de révéler quels éléments *cis* supplémentaires participent à la formation d'une origine efficace. Un enjeu majeur sera de comprendre quels sont les mécanismes moléculaires impliqués dans le contrôle des origines par les pG4. ♦

SUMMARY

G-quadruplex: key controllers of human genome duplication

The correct duplication of the human genome is under the control of a spatiotemporal program that determines where and when replication forks start. This regulation thus mainly operates on replication start

sites named replication origins. During the S-phase, about 50 000 origins fire in one human cell. However, the normal or perturbed progression of replication forks also strongly impacts on replication. Recently, several studies have put forward the role of a noncanonical DNA structure, the G-quadruplex, in the control of genome duplication. In this review, we describe the major impact of this structure on starting points and on the progression of replication forks. ♦

LIENS D'INTÉRÊT

Les auteurs déclarent n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.

RÉFÉRENCES

1. Costantini M, Musto H. The isochores as a fundamental level of genome structure and organization: a general overview. *J Mol Evol* 2017 ; 84 : 93-103.
2. Rhind N, Gilbert DM. DNA replication timing. *Cold Spring Harb Perspect Med* 2013 ; 3 : 1-26.
3. Hiratani I, Ryba T, Itoh M, et al. Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol* 2008 ; 6 : e245.
4. Berezney R, Dubey DD, Huberman JA. Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci. *Chromosoma* 2000 ; 108 : 471-84.
5. Miotto B. Comment l'approche génomique aide à comprendre le processus d'initiation de la réplication. *Med Sci (Paris)* 2017 ; 33 : 143-50.
6. Prioleau MN, MacAlpine DM. DNA replication origins-where do we begin? *Genes Dev* 2016 ; 30 : 1683-97.
7. Letessier A, Birnbaum D, Debatisse M, et al. La pauvreté en sites d'initiation de la réplication rend-elle fragile certaines régions du génome ? *Med Sci (Paris)* 2011 ; 27 : 707-9.
8. Rhodes D, Lipps HJ. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res* 2015 ; 43 : 8627-37.
9. Hänsel-Hertsch R, Di Antonio M, Balasubramanian S. DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat Rev Mol Cell Biol* 2017 ; 18 : 279-84.
10. Huppert JL, Balasubramanian S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res* 2005 ; 33 : 2908-16.
11. Chambers VS, Marsico G, Boutell JM, et al. High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat Biotechnol* 2015 ; 33 : 877-81.
12. Schaffitzel C, Berger I, Postberg J, et al. In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylomychia lemnae* macronuclei. *Proc Natl Acad Sci USA* 2001 ; 98 : 8572-7.
13. Rodriguez R, Miller KM, Forment JV, et al. Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat Chem Biol* 2012 ; 8 : 301-10.
14. Hänsel-Hertsch R, Beraldi D, Lensing SV, et al. G-quadruplex structures mark human regulatory chromatin. *Nat Genet* 2016 ; 48 : 1267-72.
15. Krusselbrink E, Guryev V, Brouwer K, et al. Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCD1-defective *C. elegans*. *Curr Biol* 2008 ; 18 : 900-5.
16. London TBC, Barber LJ, Mosedale G, et al. FANCD1 is a structure-specific DNA helicase associated with the maintenance of genomic G/C Tracts. *J Biol Chem* 2008 ; 283 : 36132-9.
17. Castillo Bosch P, Segura-Bayona S, Koole W, et al. FANCD1 promotes DNA synthesis through G-quadruplex structures. *EMBO J* 2014 ; 33 : 2521-33.
18. Paeschke K, Capra JA, Zakian VA. DNA replication through G-Quadruplex motifs is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase. *Cell* 2011 ; 145 : 678-91.
19. Lopes J, Piazza A, Bermejo R, et al. G-quadruplex-induced instability during leading-strand replication: G-quadruplex-induced instability. *EMBO J* 2011 ; 30 : 4033-46.
20. Piazza A, Adrian M, Samazan F, et al. Short loop length and high thermal stability determine genomic instability induced by G-quadruplex-forming minisatellites. *EMBO J* 2015 ; 34 : 1718-34.

RÉFÉRENCES

21. Schiavone D, Guilbaud G, Murat P, *et al.* Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells. *EMBO J* 2014 ; 33 : 2507-20.
22. Sarkies P, Murat P, Phillips LG, *et al.* FANCD1 coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. *Nucleic Acids Res* 2012 ; 40 : 1485-98.
23. Sarkies P, Reams C, Simpson LJ, *et al.* Epigenetic instability due to defective replication of structured DNA. *Mol Cell* 2010 ; 40 : 703-13.
24. Schiavone D, Jozwiakowski SK, Romanello M, *et al.* PrimPol is required for replicative tolerance of G quadruplexes in vertebrate cells. *Mol Cell* 2016 ; 61 : 161-9.
25. Méchali M, Yoshida K, Coulombe P, *et al.* Genetic and epigenetic determinants of DNA replication origins, position and activation. *Curr Opin Genet Dev* 2013 ; 23 : 124-31.
26. Cadoret JC, Meisch F, Hassan-Zadeh V, *et al.* Genome-wide studies highlight indirect links between human replication origins and gene regulation. *Proc Natl Acad Sci USA* 2008 ; 105 : 15837-42.
27. Sequeira-Mendes J, Díaz-Uriarte R, Apedaile A, *et al.* Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLoS Genet* 2009 ; 5 : e1000446.
28. Cayrou C, Coulombe P, Vigneron A, *et al.* Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res* 2011 ; 21 : 1438-49.
29. Besnard E, Babled A, Lapasset L, *et al.* Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat Struct Mol Biol* 2012 ; 19 : 837-44.
30. Picard F, Cadoret JC, Audit B, *et al.* The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. *PLoS Genet* 2014 ; 10 : e1004282.
31. Valton AL, Hassan-Zadeh V, Lema I, *et al.* G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *EMBO J* 2014 ; 33 : 732-46.
32. Hassan-Zadeh V, Chilaka S, Cadoret JC, *et al.* USF binding sequences from the HS4 insulator element impose early replication timing on a vertebrate replicator. *PLoS Biol* 2012 ; 10 : e1001277.
33. Fenouil R, Cauchy P, Koch F, *et al.* CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Res* 2012 ; 22 : 2399-408.
34. Eaton ML, Galani K, Kang S, *et al.* Conserved nucleosome positioning defines replication origins. *Genes Dev* 2010 ; 24 : 748-53.
35. Berbenetz NM, Nislow C, Brown GW. Diversity of eukaryotic DNA replication origins revealed by genome-wide analysis of chromatin structure. *PLoS Genet* 2010 ; 6 : e1001092.
36. Hoshina S, Yura K, Teranishi H, *et al.* Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA. *J Biol Chem* 2013 ; 288 : 30161-71.
37. Keller H, Kiosze K, Sachsenweger J, *et al.* The intrinsically disordered amino-terminal region of human RecQL4: multiple DNA-binding domains confer annealing, strand exchange and G4 DNA binding. *Nucleic Acids Res* 2014 ; 42 : 12614-27.
38. Papadopoulou C, Guilbaud G, Schiavone D, *et al.* Nucleotide pool depletion induces G-quadruplex-dependent perturbation of gene expression. *Cell Rep* 2015 ; 13 : 2491-503.
39. De S, Michor F. DNA secondary structures and epigenetic determinants of cancer genome evolution. *Nat Struct Mol Biol* 2011 ; 18 : 950-5.
40. Bose P, Hermetz KE, Conneely KN, *et al.* Tandem repeats and G-rich sequences are enriched at human CNV breakpoints. *PLoS One* 2014 ; 9 : e101607.
41. Monchard D. Quadruplexes d'ADN : structures, fonctions et détection. *Med Sci (Paris)* 2017 ; 33 : 1042-5.

TIRÉS À PART
M.N. Prioleau



SAVE THE DATE

July 6 - 10, 2018 | Vienna, Austria



ICNMD
2018



**15TH INTERNATIONAL CONGRESS
ON NEUROMUSCULAR DISEASES**

www.icnmd2018.org



Ludwig Boltzmann Institute
Experimental and Clinical Traumatology