

L'instabilité génétique associée aux R-loops

Une affaire d'introns

Benoit Palancade

Institut Jacques Monod CNRS, UMR 7592,
Université Paris Diderot, Sorbonne Paris Cité,
15, rue Hélène Brion, 75013 Paris, France.
benoit.palancade@ijm.fr

► Qu'elles soient isolées ou au sein d'un organisme multicellulaire, toutes les cellules eucaryotes rencontrent la même problématique : l'accomplissement de fonctions biologiques et leur adaptation aux changements physiologiques ou environnementaux nécessitent la transcription de gènes. Ce processus met nécessairement en jeu des modifications de l'organisation chromatinienne et nucléaire qui, en permettant l'accès aux machineries en charge de la synthèse des ARN, peuvent mettre en péril la stabilité du génome, notamment au cours de sa duplication. Ces considérations trouvent en fait leur source dans de nombreuses observations expérimentales réalisées depuis la fin des années 1980, et qui ont mis en évidence des taux anormalement élevés de réarrangements et de mutations sur les gènes fortement exprimés chez différentes espèces [1]. Ce n'est, cependant, que récemment que ce phénomène d'instabilité génétique associée à la transcription a trouvé des explications moléculaires, parmi lesquelles, la formation de « R-loops » sur les gènes transcrits.

Les « R-loops », une source d'instabilité génétique associée au métabolisme des ARNm

Les « R-loops » sont des structures particulières qui résultent de l'hybridation de l'ARN naissant sur son ADN matrice, laissant le brin complémentaire d'ADN sous forme non appariée (Figure 1, *panneau de gauche*). Si les R-loops peuvent contribuer favorablement à différents processus nucléaires, notamment la régulation de l'expression des gènes

ou la mise en place de réarrangements génomiques programmés, leur accumulation est néfaste pour la stabilité du génome [1]. En effet, l'ADN simple brin accessible dans ces structures peut être la cible d'attaques par des agents génotoxiques, mais également par des enzymes endogènes, et la présence des R-loops peut interférer avec la progression des fourches de réplication. Par conséquent, les gènes formant des R-loops, en particulier les gènes fortement transcrits, accumulent des mutations et des cassures double brin dans leur ADN.

Au vu de la « toxicité » potentielle de ces structures, celles-ci peuvent être éliminées par des nucléases spécialisées, les RNAses (ribonucléases)-H, ou par certaines hélicases. De plus, leur formation est elle-même limitée par les différentes étapes qui suivent la synthèse de l'ARN et qui permettent sa maturation, son assemblage en ribonucléoparticule et son export hors du noyau. Plusieurs études indépendantes avaient ainsi précédemment décrit une instabilité génétique associée aux R-loops, accrue en cas d'inactivation de composants de la machinerie d'épissage des ARNm (ARN messagers) [2].

L'épissage des introns, un paradoxe du métabolisme des ARNm

Notre vision de l'organisation des génomes a été considérablement ébranlée par la mise en évidence de l'épissage, il y a plus de 40 ans. En 1977, Richard Roberts et Philip Sharp, tous deux couronnés en 1993 par le Prix Nobel de Physiologie et Médecine pour cette

découverte, démontraient l'existence, chez les eucaryotes, des introns, des séquences non-codantes interrompant les gènes. Les introns sont reconnus spécifiquement dans l'ARN pré-messager par un complexe ribonucléoprotéique, le spliceosome, et éliminés lors d'une série de réactions biochimiques, l'épissage, aboutissant en l'assemblage précis des séquences codantes, les exons, dans l'ARNm mature [3] (→ Voir la Synthèse de G. Dujardin et al., m/s n° 12, décembre 2016, page 1103).

De multiples fonctions ont été attribuées aux introns [4], la plupart déjà anticipées dès leur découverte [5]. Leur présence autorise des étapes additionnelles de régulation de l'expression génique : la rétention d'introns dans des transcrits peut notamment induire leur dégradation de manière dépendante du type cellulaire. L'existence des introns permet aussi l'épissage alternatif, qui peut générer plusieurs protéines distinctes à partir d'un gène unique. Les introns peuvent en outre contenir eux-mêmes des unités de transcription complètes assurant la synthèse d'ARN codants ou non-codants. Enfin, l'organisation modulaire des gènes résultant de leur présence permettrait une plus grande plasticité des génomes au cours de l'évolution.

En dépit de ces nombreux rôles décrits pour les introns, leur distribution dans le génome varie considérablement selon les espèces (Tableau 1), et leur présence semble superflue dans de nombreuses situations. Il n'est pas établi que tous les cas d'épissage

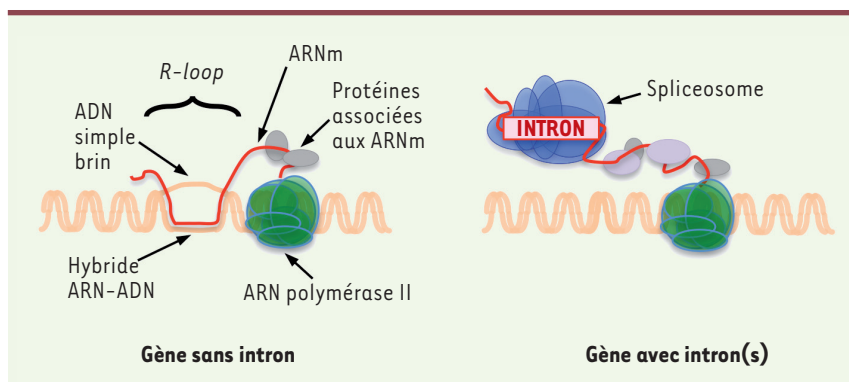


Figure 1. Les « R-loops » : structure et formation sur les gènes avec ou sans introns. À gauche, la structure d'une R-loop (cas d'un gène sans intron). L'hybride ARN:ADN formé par hybridation de l'ARN sur sa matrice ADN, ainsi que l'ADN simple brin non apparié, sont représentés. À noter que des R-loops ont été détectées sur les gènes transcrits par les trois ARN polymérases eucaryotes (ARN pol I, II, III). À droite, sur les gènes présentant des introns, le recrutement de la machinerie d'épissage limite la formation des R-loops. Les facteurs d'épissage seraient eux-mêmes susceptibles de promouvoir la liaison de protéines additionnelles sur les ARN messagers, permettant d'éviter également l'hybridation des séquences exoniques.

alternatif identifiés par les approches de séquençage à haut débit soient systématiquement associés à des mécanismes régulateurs ou à l'expression de protéines fonctionnellement distinctes. Une fraction non négligeable des jonctions intron-exon semble d'ailleurs épissée de manière constitutive chez les mammifères, soulevant la question de leur importance pour l'expression des gènes correspondants [6]. La situation semble encore plus extrême chez certaines levures telles que *Saccharomyces cerevisiae*, espèce pauvre en introns et pour laquelle la plupart d'entre eux peuvent être éliminés sans conséquence apparente sur l'expression génique ou la viabilité. Dans ce cas, pour quelles raisons les introns ont-ils colonisé nos génomes - ou ne les ont pas abandonnés - au cours de l'évolution ?

Le contenu en introns des gènes, un déterminant clé de la formation des R-loops

En raison du paradoxe précédemment énoncé, et du rôle de l'épissage dans la protection contre les R-loops, nous nous sommes demandé si l'une des fonctions des introns pourrait être de s'opposer à la formation de ces structures.

Dans un premier temps, nous nous sommes appuyés sur des études précédentes qui avaient cartographié la position des R-loops dans les génomes par l'immunoprécipitation de ces structures à l'aide d'un anticorps spécifique, suivie d'un séquençage à haut débit. En comparant les niveaux de R-loops sur des locus de niveau d'expression comparable, nous avons ainsi pu montrer que la fréquence de ces structures et leur densité le long des gènes sont toutes deux accrues sur les gènes sans introns, chez l'humain et chez la levure *S. cerevisiae*. De plus, l'accumulation de R-loops détectée sur les gènes sans introns est associée à l'apparition de lésions dans l'ADN, marquées par des modifications épigénétiques spécifiques (phosphorylation de l'histone H2A chez la levure ou du variant H2AX chez l'humain) [6]. Par la suite, la formation préférentielle des R-loops sur les locus sans intron a été également observée dans une étude indépendante dans laquelle ces structures ont été détectées sur le génome humain par modification au bisulfite de l'ADN simple-brin associé¹ [7].

¹ Le traitement au bisulfite convertit les cytosines en uraciles dans l'ADN simple brin, et cible donc le brin non apparié au niveau des R-loops. Le séquençage repère ensuite les bases modifiées et révèle l'emplacement de ces structures.

Si les gènes sans introns présentent un taux de R-loops accru, qu'en est-il des génomes des espèces pauvres en introns ? Pour répondre à cette question, nous avons systématiquement stimulé l'apparition de R-loops chez différentes espèces de levures et chez l'humain (voir Tableau 1). De manière frappante, l'accumulation des R-loops ainsi générées est plus marquée chez les espèces pauvres en introns.

Pour confirmer que l'absence d'introns est réellement le déterminant de l'instabilité génétique associée aux R-loops dans ces études génomiques, nous avons interféré avec le contenu en introns de gènes modèles en modifiant directement le génome de la levure. Nous avons ainsi pu montrer que l'élimination d'introns naturels entraîne une accumulation de R-loops sur les gènes affectés, tandis que l'insertion ectopique d'introns dans des gènes n'en comprenant pas naturellement atténue la formation de ces structures ainsi que des réarrangements associés [6]. La présence d'introns dans les gènes - et les génomes - protège donc de la formation de R-loops génotoxiques.

Par quels mécanismes les introns limitent-ils la formation des R-loops ?

Sur la base de nombreux travaux qui avaient précédemment étudié la formation des R-loops ou les mécanismes d'épissage, plusieurs hypothèses ont pu être émises pour expliquer nos observations : (1) le recrutement de la machinerie d'épissage sur les introns pourrait s'opposer stériquement à la formation des hybrides ; (2) le fait que les introns soient éliminés, et donc que la séquence de l'ARNm diverge de celle de sa matrice ADN, pourrait aussi contrecarrer la formation de ces structures.

Pour distinguer entre ces modèles, nous avons dissocié génétiquement le recrutement du spliceosome et les réactions d'épissage, et observé les conséquences en terme d'instabilité génétique associée aux R-loops, sur un gène modèle fortement exprimé chez la levure. Nous avons

	Fraction du génome contenant des introns (parmi les gènes codant des protéines, %)	Nombre moyen d'introns (par gène avec introns)	Niveaux globaux de R-loops (sur les gènes codant des protéines*)
Levures			
<i>Candida glabrata</i>	3,3	1	++
<i>Saccharomyces cerevisiae</i>	4,4	1	++
<i>Schizosaccharomyces pombe</i>	47	2,2	+/-
<i>Cryptococcus neoformans</i>	99,5	5,7	-
Homme	94,7	10,2	+/-

Tableau 1. Fréquences de distribution des introns dans les génomes eucaryotes. Le nombre de gènes contenant des introns est exprimé en pourcentage du nombre total de gènes codant des protéines, pour les différentes espèces indiquées. Le nombre moyen d'introns par gène est calculé en considérant uniquement les gènes contenant des introns [4, 6]. *Les niveaux globaux de R-loops ont été mesurés en cas d'inactivation du complexe THO/TREX, un facteur d'assemblage des particules d'ARNm dont la fonction dans la prévention des R-loops a été conservée au cours de l'évolution ([6] et données non publiées). (+) accumulation de R-loops.

ainsi pu montrer que l'insertion d'une version mutante d'intron, non compétente pour lier la machinerie d'épissage, ou d'un intron autocatalytique, capable de s'épisser sans l'aide de facteurs protéiques, ne permet pas de contrecarrer la formation des R-loops. À l'inverse, l'utilisation d'un mutant d'intron capable de recruter le spliceosome sans finaliser les réactions d'épissage, de même que l'insertion d'un autre type de séquence capable de lier des protéines sur l'ARN², sont suffisantes pour limiter l'apparition des R-loops. C'est donc le recrutement de protéines sur l'ARN qui s'oppose de manière critique à la formation des hybrides ARN:ADN (Figure 1, panneau de droite), du moins pour le gène modèle utilisé dans nos travaux.

Perspectives

La raison d'être de l'ADN non-codant qui constitue l'essentiel de notre génome est l'un des mystères actuels de la biologie moléculaire. La nouvelle fonction que nous proposons pour les séquences non-codantes que sont les introns pourrait contribuer à expliquer leur sélection par l'évolution : leur présence participerait

ainsi au maintien de la stabilité génétique dans des situations physiologiques - ou pathologiques - associées à une accumulation de R-loops. Certains facteurs cellulaires limitant la toxicité de ces structures sont effectivement inactivés dans des situations de carcinogenèse [8, 9], ou en cas d'infection par certains herpès virus [10].

À noter que les mécanismes que nous décrivons dans cette nouvelle ne sont certainement pas les seuls par lesquels l'épissage contribue au maintien de la stabilité génétique : la réponse aux dommages à l'ADN se traduit notamment par des changements globaux des profils d'épissage alternatif, et donc d'expression des gènes [11]. Il est probable que de nouveaux mécanismes connectant le métabolisme des ARN au maintien de la stabilité du génome ne manqueront pas d'émerger à l'avenir. ♦

R-loop-associated genetic instability: why introns matter?

REMERCIEMENTS

Ce travail a bénéficié de financements de la Fondation ARC pour la Recherche sur le Cancer et de la Ligue nationale contre le Cancer. Je remercie également les membres du laboratoire et mes collaborateurs pour leurs contributions aux travaux décrits, plus particulièrement Vincent Géli pour la relecture de ce manuscrit, et adresse mes excuses aux nombreux collègues dont les recherches ne peuvent être directement citées ici.

LIENS D'INTÉRÊT

L'auteur déclare n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.

RÉFÉRENCES

- Gaillard H, Herrera-Moyano E, Aguilera A. Transcription-associated genome instability. *Chem Rev* 2013 ; 113 : 8638-61.
- Chan YA, Hieter P, Stirling PC. Mechanisms of genome instability induced by RNA-processing defects. *Trends Genet* 2014 ; 30 : 245-53.
- Dujardin G, Daguinet E, Bernard DG, et al. L'épissage des ARN pré-messagers : quand le spliceosome perd pied. *Med Sci (Paris)* 2016 ; 32 : 1103-10.
- Hube F, Francastel C. Mammalian introns: when the junk generates molecular diversity. *Int J Mol Sci* 2015 ; 16 : 4429-52.
- Gilbert W. Why genes in pieces? *Nature* 1978 ; 271 : 501.
- Bonnet A, Grosso AR, Elkaoutari A, et al. Introns protect eukaryotic genomes from transcription-associated genetic instability. *Mol Cell* 2017 ; 67 : 608-21 e6.
- Dumelie JG, Jaffrey SR. Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisDRIP-seq. *Elife* 2017 ; 6. pii: e28306.
- Bhatia V, Barroso SI, Garcia-Rubio ML, et al. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* 2014 ; 511 : 362-5.
- Hatchi E, Skourti-Stathaki K, Ventz S, et al. BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Mol Cell* 2015 ; 57 : 636-47.
- Jackson BR, Noerenberg M, Whitehouse A. A novel mechanism inducing genome instability in Kaposi's sarcoma-associated herpesvirus infected cells. *PLoS Pathog* 2014 ; 10 : e1004098.
- Tresini M, Warmerdam DO, Kolovos P, et al. The core spliceosome as target and effector of non-canonical ATM signalling. *Nature* 2015 ; 523 : 53-8.

² L'intron a été remplacé par des séquences tiges-boucles issues de l'ARN du bactériophage MS2, elles-mêmes reconnues par la protéine de capsid de ce phage, exprimée de novo chez la levure.