

## La relation SINE/LINE : un exemple de parasitisme moléculaire ?

Les éléments transposables sont généralement considérés comme des parasites des génomes eucaryotes. Cette vision des choses, qui présente les éléments transposables comme des envahisseurs perturbant le fonctionnement des gènes, est fortement réductrice et passe sous silence les nombreuses interconnexions entre ces derniers et plusieurs processus cellulaires fondamentaux (revue dans [1]). En réalité, il est fort possible que, d'un point de vue évolutif, les envahisseurs soient les gènes ! En effet, il est probable que les premiers organismes vivants présentaient un génome (ARN ?) de taille très faible et qu'une des étapes essentielles de leur évolution ait consisté à augmenter la taille de leur matériel génétique. Dans ce contexte, les fonctions de transposition (répliquative) ont pu être les premières à s'imposer. L'apparition de fonctions de transposition a nécessité, en réaction, la mise en place de systèmes de régulation de façon à limiter l'expansion des éléments transposables et, ainsi, à contrer une augmentation excessive de la taille des génomes. Ces fonctions de régulation ont pu évoluer des fonctions même de la transposition. Dans cette hypothèse, l'émergence des premiers « gènes » réglant la transposition serait une réponse à la pression exercée par les éléments transposables. Les fonctions géniques modernes pourraient donc avoir comme origine les fonctions primitives de répliquative/transposition des premiers éléments transposables. Les gènes auraient progressivement « envahi » un matériel génétique primitif composé majoritairement de ces fonctions de répliquative/transposition.

Ce scénario, impossible à vérifier, implique que les génomes eucaryotes modernes sont essentiellement composés, dans leurs régions intergéniques, d'éléments transposables (plus ou moins récents), ce qui semble se confirmer chez les mammifères et chez les plantes [2, 3]. Les éléments transposables ne seraient donc pas des « envahisseurs » mais des composants essentiels des génomes participant encore aujourd'hui à la dynamique évolutive, à la transmission et à l'expression du programme génétique. Bien sûr, ceci n'exclut pas que les éléments transposables puissent jouer le rôle d'agents mutagènes. Ainsi plusieurs maladies chez l'homme résultent de l'intégration de deux types d'éléments transposables, les rétroposons SINE (*short interspersed nuclear elements*) et LINE (*long interspersed nuclear elements*) ([1] et *m/s*, 1996, n°6, p. 833).

### Les rétroposons SINE et LINE

Les éléments transposables peuvent être séparés en deux grands groupes [4]. Les éléments de classe I portent le nom générique de rétrotransposons. Comme leur nom l'indique, ces éléments transposables ont un mode de transposition répliquatif fondé sur la transcription inverse d'un intermédiaire ARN. Les éléments de classe II portent le nom générique de transposons et transposent par un mécanisme de type « copier/coller », qui peut être ou non répliquatif. Parmi les éléments de classe I, les rétroposons de type SINE et LINE sont considérés comme les plus primitifs [4]. Ces éléments ont généralement eu beaucoup de succès au cours de l'évolu-

tion puisqu'ils occupent une part importante des génomes eucaryotes modernes. Par exemple, 15 % du génome humain sont représentés par des éléments LINE L1 ( $10^5$  copies) et 10 % par des éléments SINE Alu ( $10^6$  copies) [2]. Les éléments LINE complets sont autonomes car ils codent pour les fonctions nécessaires à leur rétroposition (*figure 1*). Leur transcription par l'ARN polymérase II produit un transcrite qui donnera naissance, avec les protéines codées par ce transcrite, à un complexe ribonucléoprotéique similaire à une particule rétrovirale. Ce complexe, appelé VLP (*virus-like particle*) est responsable de la transcription inverse de l'ARN messager LINE en ADNc et de son intégration à une nouvelle position chromosomique. D'un point de vue évolutif, les LINE peuvent être séparés en 12 lignées distinctes, chaque lignée ayant comme origine le précambrien (~ 800 millions d'années) [5]. Le LINE L1 de mammifère appartient à une de ces 12 lignées de LINE (appelée la lignée L1). Par opposition aux LINE, les éléments SINE sont visiblement des éléments non-autonomes puisqu'ils ne possèdent pas de capacité de codage. Les SINE ont un certain nombre de points communs avec les gènes codant soit pour les ARN de transfert, soit pour l'ARN 7SL, et sont transcrits par l'ARN polymérase III (*figure 1*). Une de leurs particularités est qu'ils ne possèdent pas de terminateur de transcription pour la polymérase III (qui est habituellement une séquence d'au moins quatre thymines sur le brin matriciel). La transcription d'un élément SINE se termine donc dans sa région génomique d'aval et l'ARN ainsi pro-

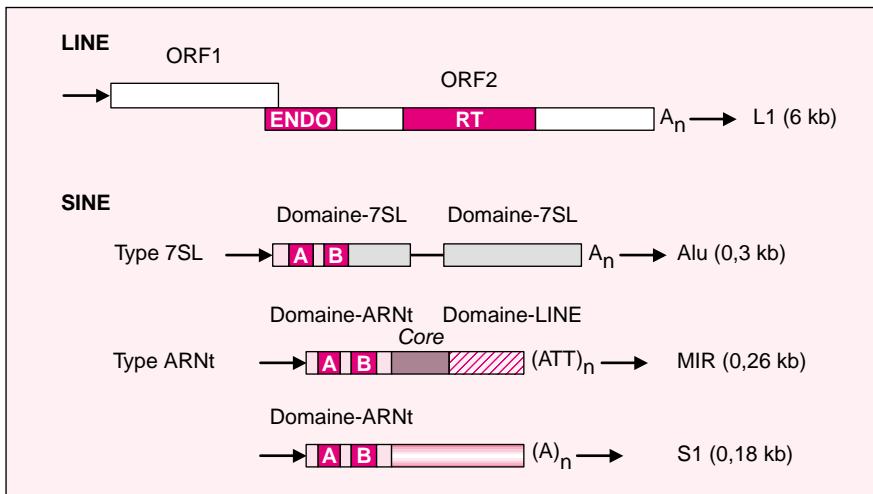


Figure 1. **Structure des éléments SINE et LINE.** Les éléments LINE possèdent deux cadres de lecture (ORF pour open reading frame) permettant la synthèse des protéines nécessaires à leur transposition. Les éléments SINE possèdent des domaines apparentés à l'ARN 7SL ou aux ARN de transfert. Les SINE possédant un domaine Core présentent également une région 3' homologue à la région 3' d'éléments LINE. Les noms et tailles de différents LINE et SINE sont donnés à titre d'exemples. A et B, promoteurs pour l'ARN polymérase III; Core, région conservée de fonction inconnue; ENDO, domaine endonucléase; RT, transcriptase inverse.

duit comprendre une portion non-apparentée au SINE qui varie d'un site génomique à l'autre. Ainsi, la transposition des éléments SINE nécessite une sélection de leur transcrit par un complexe de transcription inverse codé par un autre élément. Récemment, l'hypothèse selon laquelle la transposition des SINE s'effectue grâce à l'utilisation de la machinerie de rétroposition LINE a été émise [6-8, 10, 14]. Selon cette hypothèse, les SINE seraient donc considérés comme des « parasites » de la VLP des LINE.

### Relation SINE/LINE : arguments et contre-arguments

Quels sont les arguments qui permettent de proposer une relation entre SINE et LINE chez les mammifères ? Tout d'abord, une analyse statistique des sites d'intégration du SINE Alu et du LINE L1 a montré qu'ils possédaient les mêmes caractéristiques moléculaires [7]. De plus, l'analyse biochimique de l'endonucléase codée par le LINE L1 a révélé que cette enzyme possède une spécificité de coupure qui pourrait expliquer à

la fois les événements d'intégration des éléments L1 et ceux des éléments Alu [9]. Ce serait donc cette enzyme qui serait responsable de l'intégration de ces éléments. Un autre argument peut être tiré du travail du groupe de Thierry Heidmann qui a montré que la VLP du LINE L1 pouvait s'associer à des ARNm cellulaires, les transcrire en ADNc de manière inverse et produire, à la suite de leur intégration, des pseudogènes épissés [10]. Ce résultat implique que le complexe de rétroposition L1 peut non seulement rétroposer l'ARNm L1 utilisé pour sa propre traduction (rétroposition en *cis*), mais également rétroposer d'autres ARNm cellulaires (rétroposition en *trans*). Finalement, les rétrovirus endogènes ne peuvent être les partenaires des SINE car les transcriptases inverses rétrovirales sont incapables de produire les empreintes caractéristiques de la rétroposition [10, 11]. Ainsi, malgré l'absence de preuve directe, le LINE L1 est généralement considéré comme le partenaire de transposition du SINE Alu. Cependant, un obstacle majeur à cette hypothèse est

l'incroyable succès évolutif de la sélection en *trans* de l'ARN Alu (permettant de produire  $10^6$  copies de cet élément dans le génome humain) malgré la forte préférence du complexe de rétroposition de l'élément L1 pour la rétroposition en *cis* [10]. Les protéines codées par le LINE L1 lient en effet l'ARN messager utilisé pour leur propre traduction (en *cis*) 20 fois plus efficacement qu'un autre ARN messager (en *trans*) [10]. On peut envisager que l'insertion d'un élément de petite taille comme le SINE Alu est moins destructrice que celle d'un élément L1. Ainsi, malgré une fréquence de transposition plus faible, l'élément Alu aurait pu s'accumuler dans le génome avec un nombre de copies plus élevé. Dans cette hypothèse, un problème reste entier : pourquoi la sélection de l'ARN Alu, qui est généralement peu abondant, est-elle plus efficace que celle d'autres ARN bien plus abondants comme les ARN de transfert ou l'ARN 7SL ? En d'autres termes, pourquoi l'ARN Alu serait-il un partenaire privilégié de la VLP du LINE L1 ? Une façon d'expliquer le succès de la rétroposition des éléments Alu est de proposer que leurs ARN interagissent avec les protéines L1 dans une configuration moléculaire équivalente à la relation en *cis* de l'ARN LINE. Dans ce cas, il faudrait que l'ARN Alu soit ciblé spécifiquement au niveau des ribosomes pendant la traduction de l'ARN L1 (figure 2A). Or, il a été montré que l'ARN Alu peut adopter une structure secondaire similaire à celle de l'ARN 7SL [12] et interagir avec certaines protéines du complexe SRP (*signal recognition particle*) [13]. Le complexe SRP, auquel appartient l'ARN 7SL, est nécessaire à la liaison des ribosomes au réticulum endoplasmique pendant la synthèse des protéines. Comme le complexe SRP est lui-même lié aux ribosomes, près de la sortie du peptide naissant, l'ARN Alu serait donc en position idéale pour interagir avec les protéines L1. Cette interaction se ferait au niveau de la région poly (A) des deux transcrits. Cette hypothèse, appelée la *poly (A) connection* [6], peut également s'appliquer à d'autres éléments SINE qui s'apparentent aux ARN de trans-

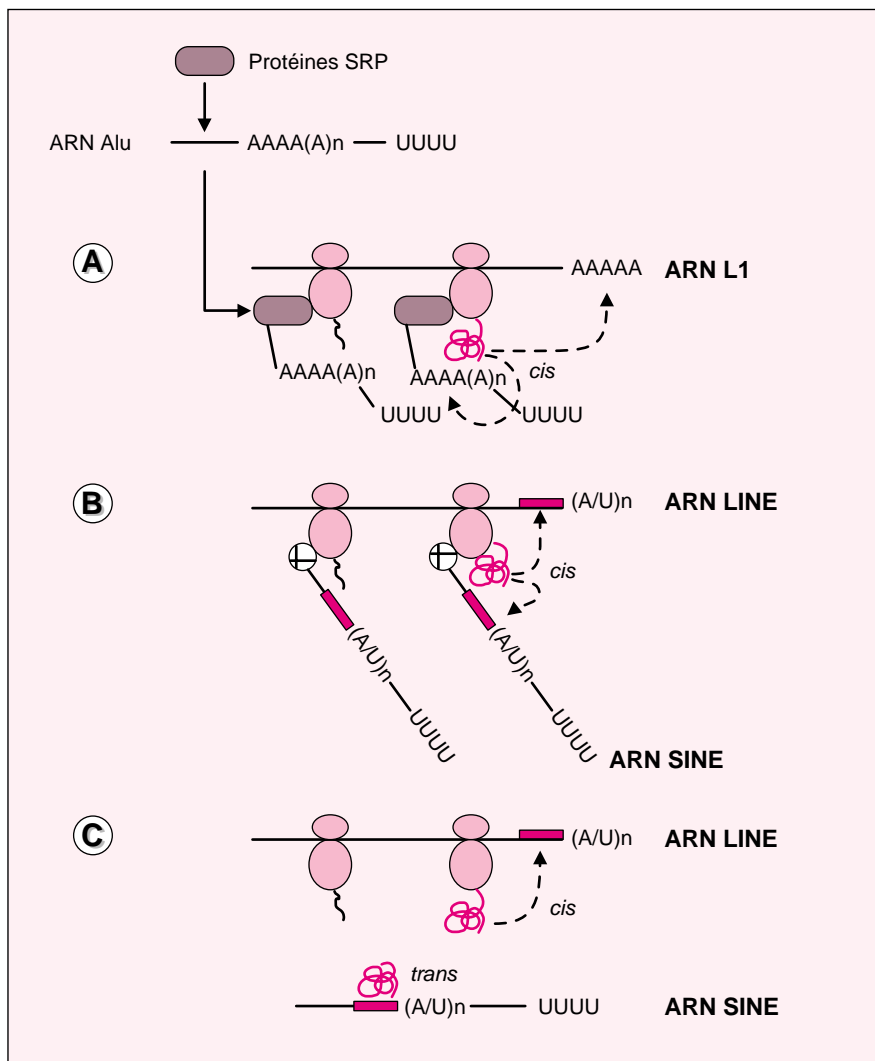


Figure 2. **Modèles d'interaction entre SINE et LINE.** A. Illustration de la poly (A) connexion. Les ARN Alu seraient ciblés au niveau des ribosomes grâce à leur liaison aux protéines du complexe SRP (signal recognition particule) [13]. Les protéines L1 (en rouge), dès leur formation, pourraient interagir soit avec la région poly (A) de l'ARN L1, soit avec la région poly (A) de l'ARN Alu. En principe la poly (A) connexion peut également s'appliquer aux SINE qui s'apparentent aux ARN de transfert et qui possèdent une queue poly (A) puisque leur structure de type ARNt peut également leur permettre de se positionner efficacement au niveau des ribosomes. B. et C. Illustration d'un deuxième mode d'interaction possible entre SINE et LINE fondé sur une identité de séquence autre que la queue poly (A). Le partage d'une même région 3' permettrait aux protéines LINE de reconnaître en cis l'ARN LINE et d'avoir une action dans une configuration moléculaire de type cis (B) ou trans (C) sur l'ARN SINE. L'interaction de type « cis » serait possible grâce à l'association du domaine ARN de transfert du SINE avec des protéines impliquées dans la traduction. On peut noter que les LINE et les SINE qui partagent une identité de séquences au niveau de leur région 3' ne possèdent pas de longue répétition poly (A) mais de très courtes répétitions d'un motif variable, habituellement riche en A/T.

fert et qui possèdent une queue poly (A) puisque leur structure de type ARNt pourrait également leur permettre de se positionner efficacement au niveau des ribosomes. Il reste cependant beaucoup à faire pour valider ce modèle, notamment expliquer les mécanismes permettant aux ARN SINE de quitter le noyau, et démontrer qu'ils sont effectivement associés aux ribosomes. Il reste également à prouver ce rôle essentiel de la queue poly (A) et à définir le mode de transcription inverse des ARN LINE et SINE, qui doit débiter à partir de la région poly (A) située à l'extrémité de l'ARN LINE mais en position interne pour l'ARN SINE [14].

Un second type de relation SINE/LINE reposant sur une homologie de séquence primaire autre que la queue poly (A) a été proposé [8]. En effet, certains SINE apparentés aux ARN de transfert partagent dans leur région 3' une forte identité de séquence (entre 85 % et 95 %, habituellement sur 30 à 50 paires de bases) avec la région 3' de LINE appartenant uniquement aux lignées CR1 et RTE (deux lignées LINE comprenant des éléments de vertébrés et d'invertébrés [5]) mais pas à la lignée L1 [8]. Dans ces exemples, les couples SINE/LINE ne possèdent pas de longue répétition poly (A), comme c'est le cas pour Alu et L1, mais de très courtes répétitions d'un motif variable, habituellement riche en A/T (figure 1). L'origine de ces éléments SINE serait la fusion entre une région de type ARNt et la région 3' d'un LINE, fusion rendue possible par la présence d'une région centrale d'origine inconnue appelée séquence « Core » [15] (figure 1). Dans ce type de relation SINE/LINE, il a été proposé que le partage d'une séquence commune entre les deux éléments pourrait faciliter la reconnaissance, par le complexe de rétroposition LINE, de l'ARN SINE et donc leur interaction en trans [8] (figure 2C). Cependant, il a été montré récemment dans le cas du LINE L1 que les homologies de séquence primaires n'influaient pas sur la fréquence de rétroposition en trans [10]. On peut toutefois envisager que les LINE des lignées CR1 et RTE

se comportent différemment des LINE de la lignée L1 et utilisent une sélection en *trans* fondée sur une homologie de séquence primaire. Ceci reste à déterminer expérimentalement. Une autre possibilité est que, tout comme dans le cas de la *poly (A) connection*, l'association préférentielle des ARN SINE aux ribosomes leur permettent d'interagir dans une configuration moléculaire équivalente à l'interaction en *cis* de l'ARN L1 (figure 2B). Dans ce cas, la seule différence entre les deux modèles de relation SINE/LINE serait la séquence impliquée dans l'interaction qui est, soit la région poly (A), soit la portion 3' homologue (figure 2, A et B). Dans tous les cas, les mécanismes permettant aux ARN SINE de quitter le noyau et de s'associer aux ribosomes restent à déterminer.

A l'heure actuelle, l'hypothèse selon laquelle les éléments SINE sont des parasites du complexe de rétroposition des éléments LINE, bien que plausible, souffre d'un manque de preuves directes. Les deux modèles proposés d'interaction entre ces éléments doivent être testés expérimentalement. Il est donc nécessaire de développer des systèmes d'étude *in vivo* permettant de tester la réalité et les différentes modalités de cette relation ■

## RÉFÉRENCES

1. Deragon JM, Cappy P. Impact of transposable elements on the human genome. *Ann Med* 2000; 32: 264-73.
2. Smit AFA. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr Opin Genet Dev* 1999; 9: 657-63.
3. SanMiguel P, Tikhonov A, Jin YK, et al. Nested retrotransposons in the intergenic regions of the maize genome. *Science* 1996; 274: 765-8.
4. Cappy P, Bazin C, Higuier D, et al. *Dynamics and evolution of transposable elements*. Springer, Austin Texas, USA: R.G. Landes Company, 1998: 37-47.
5. Malik HM, Burke WD, Eickbush TH. The age and evolution of non-LTR retrotransposable elements. *Mol Biol Evol* 1999; 16: 793-805.
6. Boeke JD. LINEs and Alus-the polyA connection. *Nat Genet* 1997; 16: 6-7.
7. Jurka J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retrotransposons. *Proc Natl Acad Sci USA* 1997; 94: 1872-7.
8. Okada N, Hamada M, Ogiwara I, et al. SINEs and LINEs share common 3' sequences: a review. *Gene* 1997; 205: 229-43.
9. Cost GJ, Boeke JD. Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. *Biochemistry* 1998; 37: 18081-93.
10. Esnault C, Maestre J, Heidmann, T. Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* 2000; 24: 363-7.
11. Dornburg R, Temin HM. cDNA genes formed after infection with retroviral vector particles lack the hallmarks of natural processed. *Mol Cell Biol* 1991; 10: 68-74
12. Sinnott D, Richer C, Deragon JM, et al. Alu RNA secondary structure consists of two independent 7SL RNA-like folding units. *J Biol Chem* 1991; 266: 8675-8.
13. Hsu K, Chang DY, Marais RJ. Human signal recognition particle (SRP) Alu-associated protein also binds Alu interspersed repeat sequence RNAs. Characterization of human SRP9. *J Biol Chem* 1995; 270: 10179-86.
14. Weiner AM. Do all SINEs lead to LINEs? *Nat Genet* 2000; 24: 332-3.
15. Gilbert N, Labuda D. CORE-SINEs: eukaryote short interspersed retroposing elements with common sequence motifs. *Proc Natl Acad Sci USA* 1999; 96: 2869-74

### Jean-Marc Deragon

Cnrs UMR6547 et GDR2157, Biomove, Université Blaise-Pascal Clermont-Ferrand II, 63177 Aubière Cedex, France.

### TIRÉS À PART

J.M. Deragon.