

Modelage du génome des mammifères par la transcriptase inverse

La transcriptase inverse catalyse la conversion d'ARN en ADN. Cette opération pourrait avoir eu un rôle essentiel dans la mise en place des systèmes vivants. Plus du dixième du génome d'un mammifère est composé de séquences engendrées ainsi, dont la plus grande partie est constituée de séquences répétées (SINE, LINE). Ces séquences jouent un rôle évolutif important en favorisant des recombinaisons entre régions non homologues. Des séquences codantes dépourvues d'intron, elles aussi issues de transcription inverse, peuvent être fonctionnelles. Enfin, des rétrovirus endogènes peuvent être impliqués dans la régulation de gènes avoisinants. Ces vestiges d'une activité de transcription inverse montrent que l'enzyme qui la porte a contribué à conférer aux génomes actuels une plasticité propice à leur évolution, qui se manifeste encore actuellement à travers des maladies génétiques.

Pascal Nouvel

Le monde vivant présente un paradoxe lorsqu'on l'envisage sous l'angle de sa création. D'une part, l'ADN est nécessaire à la formation des protéines, puisque c'est lui qui engrange la forme canonique de leur mémoire; d'autre part, les protéines sont nécessaires à la formation de l'ADN, polymère stable, dont la synthèse paraît nécessiter l'intervention d'une catalyse enzymatique. Relevant le paradoxe, Crick et Orgel, ont proposé en 1968 que l'ARN pourrait avoir été l'un des constituants majeurs du milieu prébiotique [1, 2]. L'ARN, intermédiaire instable entre l'ADN et la protéine, aurait pu constituer la base moléculaire d'un système répliatif élémentaire.

C'est l'hypothèse dite du *RNA world*. En 1982, Cech *et al.* [3] puis en 1983 Altman *et al.* [4] montrèrent que certains ARN peuvent avoir une activité catalytique. La propriété ainsi mise en évidence est certes assez éloignée du réseau réticulé des interactions qui caractérise les systèmes cellulaires, mais elle permet de concevoir une solution au paradoxe: dans la molécule d'ARN, se trouvent réunies les deux propriétés nécessaires à l'établissement d'un système évolutif. Composé de nucléotides susceptibles de s'apparier entre eux par le jeu des complémentarités moléculaires, l'ARN pourrait avoir constitué un substrat autorépliatif; capable de réaliser des opérations de catalyse spécifique, le même ARN aurait pu don-

ADRESSE

P. Nouvel : docteur ès sciences. Unité de génétique des mammifères. Institut Pasteur, 25, rue du Docteur-Roux, 75724 Paris Cedex 15, France.

ner prise à un processus de sélection. Mais, si des systèmes prébiotiques composés d'ARN ont prélué à la vie terrestre, c'est alors le passage au stade actuel qui devient problématique. Comment s'est élaboré le monde vivant, tel que nous le connaissons avec sa hiérarchie insistante ADN → ARN → protéine ?

En 1971, Temin et Baltimore ont montré, indépendamment, que les rétrovirus produisaient une enzyme capable de synthétiser de l'ADN à partir d'ARN : la transcriptase inverse (*reverse transcriptase* RT) [5, 6]. A partir des propriétés de cette enzyme, on peut concevoir un scénario de conversion d'un *RNA world* en une forme plus proche du monde biologique actuel. L'information génétique, initialement portée et sélectionnée sous forme d'ARN, molécule instable dans les conditions standard, aurait ainsi été figée sous une forme plus stable, l'ADN. Le scénario est imprécis et présente essentiellement une valeur heuristique ; en revanche, l'examen des génomes actuels et de leur composition en séquences dérivant de transcription inverse permet de retracer, parfois assez précisément, les événements de rétrotransposition aux conséquences évolutives éventuellement importantes.

Le génome des organismes eucaryotes et notamment des vertébrés supérieurs est parsemé de séquences engendrées par transcription inverse. On estime que ces séquences représentent près de 10 % de la totalité du génome. La part la plus abondante de ces séquences peut être divisée en deux groupes principaux : les SINE (*small interspersed nuclear elements*) et les LINE (*long interspersed nuclear elements*) fortement répétés dans le génome et dispersés [7, 8]. On trouve, par exemple, 50 000 à 100 000 copies de chacun de ces deux types de séquences dans un génome de souris. Chez l'homme, les séquences Alu sont les SINE les plus abondantes. On trouve, par ailleurs, dans les génomes de mammifères, des copies de gènes cellulaires qui présentent les marques caractéristiques d'une rétrotransposition (absence d'intron, présence d'une séquence polyadénylée en 3') et qui sont, pour cette raison,

nommés rétroposons ou rétro-pseudogènes. Enfin, on trouve des séquences faiblement répétées qui correspondent à des insertions rétrovirales endogènes ou à des insertions de rétrotransposons.

Séquences engendrées par transcription inverse et organisation des génomes complexes

La diversité des rétroéléments présents dans les génomes eucaryotes resterait assez anecdotique si on ne pouvait entrevoir un dialogue entre ces éléments et les génomes des organismes qui les hébergent. On peut distinguer deux modes de « dialogue ». L'un, indirect, concerne les effets de la présence de rétroéléments dans le génome en tant qu'ils favorisent des événements de recombinaison entre régions homologues. Le second, direct, concerne les effets de l'insertion de ces rétroéléments en tant qu'ils produisent ou peuvent affecter les propriétés activatrices de la région où ils sont insérés. Chacun de ces cas peut être illustré par quelques exemples.

Les effets indirects

Seuls des effets indirects dans les recombinaisons entre SINE ont pu être mis en évidence jusqu'à présent. Ces séquences répétées offrent des points d'homologie partielle qui parsement le génome. Ils permettent d'élargir le jeu des mécanismes de brassage de l'information génétique, essentiellement basé sur l'échange entre régions homologues. Deux exemples d'anomalies génétiques chez l'homme ont confirmé le rôle des séquences répétées dans certains remaniements génétiques. Il s'agit, d'une part, d'une duplication de 7 exons du gène récepteur de la LDL (*low density lipoprotein*) chez un sujet atteint d'hypercholestérolémie et, d'autre part, d'un réarrangement entre séquences Alu chez un sujet dit « mâle XX ».

Le récepteur des LDL

Les LDL sont des protéines transporteurs du cholestérol. Des récep-

teurs spécifiques, situés à la surface cellulaire, permettent l'internalisation des LDL. Le récepteur est une protéine membranaire codée par un gène comportant 18 exons (*figure 1A*). Le cas d'un sujet atteint d'hypercholestérolémie familiale (FH) a pu être analysé au niveau moléculaire [9]. Chez ce patient (FH 295), la mutation du récepteur du LDL est la conséquence de la duplication de 7 exons (exons 2 à 8), consécutive à une recombinaison entre deux séquences Alu. L'allèle muté du gène produit une protéine de grande taille dont les propriétés fonctionnelles sont fortement altérées. Le mécanisme proposé pour l'apparition de ce mutant est un *crossing-over* inégal (*figure 1B*) entre deux séquences Alu présentes dans les introns du gène. La recombinaison s'est produite dans une courte région d'une trentaine de nucléotides parfaitement homologues.

Stérilité XX

Un autre cas de maladie associée à une recombinaison entre séquences Alu, cette fois entre deux chromosomes non homologues, a été récemment mis en évidence. Il s'agit d'une anomalie génétique humaine, nommée stérilité XX. On trouve chez certains adultes de phénotype mâle un caryotype XX, généralement associé au sexe femelle. L'analyse plus précise des chromosomes X chez ces individus montre que l'un des deux chromosomes a, en fait, subi une mutation, conservant l'extrémité centromérique du chromosome Y et présentant une extrémité télomérique de chromosome X. La partie de chromosome Y retenue dans le chromosome remanié contient le TDF (*testis determining factor*) qui confère un phénotype mâle normal aux porteurs de cette mutation. En revanche, plusieurs gènes impliqués dans la fertilité mâle sont absents, et il en résulte une stérilité. L'analyse moléculaire a permis de montrer que la recombinaison entre les chromosomes X et Y a eu lieu au niveau d'une séquence Alu qui a probablement servi de substrat à un processus de recombinaison [10]. Les séquences fortement répétées sont

RÉFÉRENCES

1. Crick FHC. The origin of the genetic code. *J Mol Biol* 1968; 38: 367-79.
2. Orgel LE. Evolution of the genetic apparatus. *J Mol Biol* 1968; 38: 381-93.
3. Kruger K, Grabowski PJ, Sands J, Gottschling DE, Cech TR. Self-splicing RNA: Autoexcision and autocyclisation of the ribosomal RNA intervening sequence of tetrahymena. *Cell* 1982; 31: 147-57.
4. Guerrier-Takada C, Gardnier K, Marsh T, Pace N, Altman S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 1983; 35: 849-57.
5. Baltimore D. RNA-dependant DNA polymerase in virions of RNA tumour viruses. *Nature* 1970; 226: 1209-11.
6. Temin HM, Mizutani S. RNA-directed DNA polymerase in virions of Rous Sarcoma virus. *Nature* 1970; 226: 1211-3.
7. Hutchison CA3, Hardies SC, Loeb DD, Ronald-Shehee W, Edgell MH. LINEs and related retroposons: long interspersed repeated sequences in the eucaryotic genome. In: Berg and Howe, eds. *Mobile DNA*. Washington, DC: American Society for Microbiology, 1989: 593-617.
8. Deininger PL. SINES: short interspersed repeated DNA elements in higher eucaryotes. In: Berg and Howe, eds. *Mobile DNA*. Washington, DC: American Society for Microbiology, 1989: 619-36.
9. Lehrman MA, Goldstein JL, Russell DW, Brown MS. Duplication of seven exons in LDL receptor gene caused by Alu-Alu recombination in a subject with familial hypercholesterolemia. *Cell* 1987; 48: 827-35.
10. Rouyer F, Simmler MC, Page DC, Weissenbach J. A sex chromosome rearrangement in a human XX male caused by Alu-Alu recombination. *Cell* 1987; 51: 417-25.
11. Fitch DH, Bailey WJ, Tagle DA, Goodman M, Sieu L, Slightom JL. Duplication of the gamma-globin gene mediated by L1 long interspersed repetitive elements in an early ancestor of simian primates. *Proc Natl Acad Sci USA* 1991; 88: 7396-400.

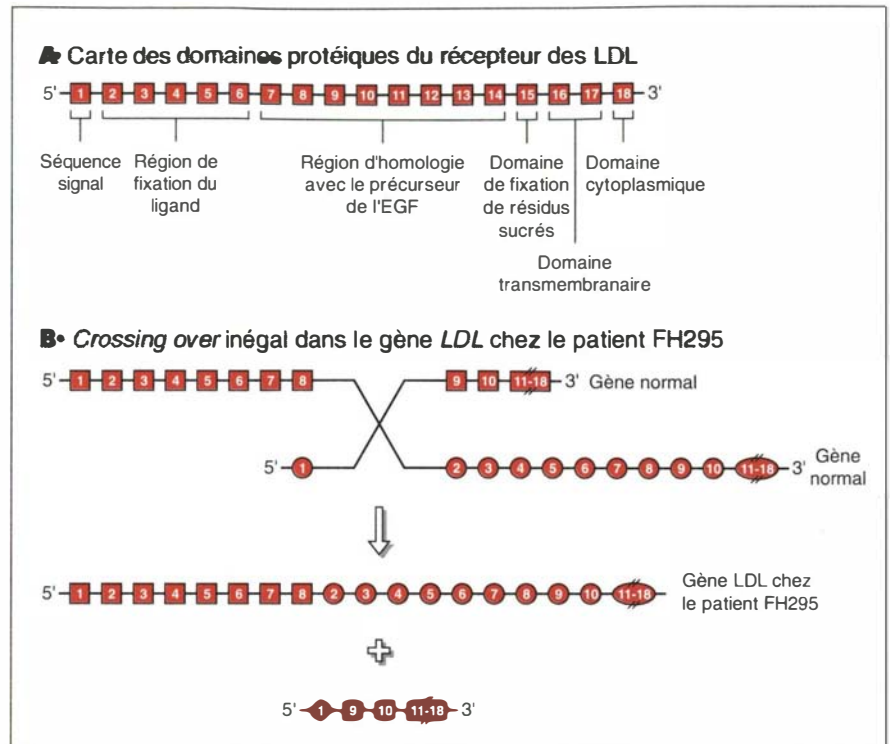


Figure 1. **Remaniements du gène du récepteur des LDL conduisant à une hypercholestérolémie [9]. A.** Structure du récepteur au LDL avec les 18 exons caractérisés et les fonctions ou homologues qui leur ont été trouvées. **B.** Crossing-over inégal dans le gène du récepteur LDL, dont on postule qu'il est à l'origine de la structure observée. Le gène présente, chez le patient FH295, une duplication de huit exons (exons 2 à 10).

donc susceptibles de jouer un rôle important dans des remaniements génétiques de plus ou moins grande ampleur.

Le gène γ -globine des primates

Le gène γ -globine est dupliqué dans le génome des primates, alors qu'il est unique dans le génome de la souris. Les comparaisons de séquences ont permis de montrer que la duplication du gène humain s'était produite à la faveur d'une recombinaison entre deux séquences répétées de type LINE [11], l'une située en amont, et l'autre en aval, de l'unique gène γ -globine de la souris. Le mécanisme de duplication est donc formellement identique à celui qui a été rapporté pour la mutation pathologique du gène humain du récepteur des LDL. Ainsi, des remaniements, qui peuvent avoir des conséquences pathologiques dans cer-

tains contextes, peuvent aussi conduire à des configurations génétiques conservées par l'évolution dans d'autres contextes.

Les effets directs

L'insertion de rétroélément dans le génome peut également jouer un rôle direct, en interrompant des phases ouvertes de lecture, ou en engendrant de nouvelles associations de séquence. McClintock a montré, en 1956, que des éléments génétiques transposables peuvent faire varier l'expression de gènes [12]. A l'époque, les preuves moléculaires à l'appui d'une telle assertion étaient assez minces. En 1974, Fincham et Sastry proposaient que de tels éléments aient pu jouer un rôle majeur au cours de l'évolution, en tant qu'agents aptes à remanier la structure ou à affecter l'activité des systèmes génétiques stables [13]. En

s'insérant en des sites variés du génome, les rétroéléments peuvent altérer des gènes. Mais, compte tenu de la forte dilution des informations structurellement significatives dans le génome, il est rare que ces altérations conduisent à un phénotype très remarquable. Parmi un grand nombre de mutations décrites avec des degrés de précision variables, on peut citer deux exemples bien documentés de mutation induite par un rétrovirus ayant entraîné des modifications phénotypiques. Il s'agit, d'une part, de la mutation affectant la couleur du poil chez la souris DBA et, d'autre part, d'une mutation touchant le gène du collagène.

Souris DBA

Le pelage clair des souris DBA est déterminé par un gène unique baptisé *dilute*. Jenkins *et al* [14] ont montré en 1981 que cette mutation ségrège avec une insertion provirale endogène. Deux ans plus tard, Copeland *et al* [15] publièrent l'analyse moléculaire de révertants. Ils montraient que les rares réversions de la mutation étaient associées à l'excision du provirus, par suite, vraisemblablement, d'une recombinaison homologue entre les deux séquences LTR (séquences répétées situées aux deux extrémités du provirus). En effet, ils trouvaient chez les révertants un résidu de séquence virale correspondant à une séquence LTR. C'est précisément la structure qu'on s'attendrait à trouver au site d'insertion si une recombinaison homologue était à l'origine de l'excision. Ces données achevèrent de démontrer le lien causal entre le phénotype d'altération de la couleur du poil et la présence d'un rétrovirus endogène dans un gène.

La mutation *Mov-13*

Jaenisch obtint expérimentalement en 1983 une mutation insertionnelle à la suite de l'infection d'embryons de souris par un rétrovirus leucémogène. La mutation affectait le gène du collagène et entraînait la létalité au jour 13 de la gestation [16]. L'analyse moléculaire du *locus* muté permit de montrer que l'insertion s'était produite dans le premier

intron du gène du collagène I. L'analyse de la mutation réservait plusieurs surprises. Il fut d'abord montré que le provirus était intégré dans le premier intron du gène, puisqu'il n'altérait pas l'expression du gène dans certaines cellules (odontoblastes) alors qu'il l'abolissait dans d'autres (fibroblastes). Chez les souris mutées, les odontoblastes, pré-curseurs des dents, produisaient un collagène intact et en quantité normale [17]. La mutation *Mov-13*, qui paraissait relativement simple, débouchait sur la mise en évidence d'une régulation complexe. Le provirus, bien que localisé dans un intron, altérait la régulation de l'expression du gène, mais seulement dans certains tissus.

Mutations induites par des insertions de séquences engendrées par transcription inverse

Plusieurs cas de maladies génétiques se sont révélés être associés à des insertions de séquences mobiles dans le génome humain. Ainsi, l'acholinestérasémie, maladie bénigne qui se manifeste par une apnée prolongée après injection de succinylcholine, est due à l'insertion d'une séquence Alu qui inactive le gène de la cholinestérase [18]. De même, la neurofibromatose de type 1, maladie caractérisée par la présence de neurofibromes et de quelques autres manifestations mineures, est le résultat de l'insertion d'un élément Alu dans un intron du gène *NF1* qui conduit à la délétion d'un exon au cours de l'épissage [19]. D'autres mutations sont induites par des insertions de séquence de type LINE: insertion dans le gène du facteur VIII qui cause une hémophilie A [20, 21], insertion dans le gène de la dystrophine, qui est l'une des causes de la myopathie de Duchenne [22].

Déplacement d'une région codante

Parmi les remaniements directs du message génétique induit par les rétrovirus, on peut distinguer, à nouveau, deux cas, selon qu'on observe le déplacement d'une phase codante par l'intermédiaire d'un phénomène de rétroposition, ou le

déplacement d'une région activatrice. Cette distinction reflète la dichotomie du génome en régions *cis* et *trans* actives. Les rétrocopies d'ARN messagers, une fois intégrées dans le génome, peuvent, occasionnellement, être exprimées et conduire à la formation de protéine si les séquences situées au voisinage du site d'insertion se trouvent capables d'assurer l'initiation de la transcription. En effet, la rétrocopie d'un gène est, en principe, dépourvue de séquences activatrices (qui sont généralement situées en amont de la partie transcrite), de sorte que son activation dépend étroitement des propriétés de la région dans laquelle elle s'insère et de la configuration de l'insertion. Les rétrocopies de deux gènes ubiquitaires impliqués dans le métabolisme des sucres, le gène de la phosphoglycérate kinase et le gène codant pour la sous-unité $E1\alpha$ de la pyruvate déshydrogénase ($E1\alpha$) chez l'homme, fournissent un exemple de ce type d'activation.

Le gène *pgk*

Le gène *pgk1* code pour l'enzyme phosphoglycérate kinase impliquée dans la glycolyse. Il comporte 10 introns. Localisé sur le chromosome X chez l'homme et chez la souris, il est, soit absent dans les spermatoocytes mâles, soit inactivé dans les spermatoocytes femelles. En revanche, il est exprimé de manière ubiquitaire dans toutes les cellules somatiques. Le gène *pgk2* est une rétrocopie du gène *pgk1* dans laquelle la phase codante est entièrement conservée tandis que les introns sont délétés. Ces particularités structurales, jointes à l'analyse des divergences entre les deux séquences, montrent que l'ARN issu du gène *pgk1* [23, 24] a donné naissance au gène *pgk2* par suite d'une rétrotransposition il y a environ 100 millions d'années.

Ce dernier, localisé sur un autosome, est exprimé dans les cellules de sperme à partir du stade pachytène, ce qui laisse supposer que la protéine pour laquelle il code est nécessaire à la survie des gamètes. L'expression de la phosphoglycérate kinase dans les cellules de sperme

RÉFÉRENCES

12. McClintock B. Controlling elements and the gene. *Cold Spring Harbor Symp Quant Biol* 1956; 21: 197-216.
 13. Fincham JR, Sastry GR. Controlling elements in maize. *Ann Rev Genet* 1974; 8: 15-50.
 14. Jenkins NA, Copeland NG, Taylor BA, Lee BK. Dilute (d) coat colour mutation of DBA/2J mice is associated with the site of integration of an ecotropic MuLV genome. *Nature* 1981; 293: 370-4.
 15. Copeland NG, Hutchison KW, Jenkins NA. Excision of the DBA ecotropic provirus in dilute coat-color revertants of mice occurs by homologous recombination involving the viral I.TRs. *Cell* 1983; 33: 379-87.
 16. Jaenisch R, Harbers K, Schnieke A, Lohler J, Chumakov I, Jahner D, Grotkopp D, Hoffmann E. Germline integration of moloney murine leukemia virus at the Mov13 locus leads to recessive lethal mutation and early embryonic death. *Cell* 1983; 32: 209-16.
 17. Kratochwil K, von der Mark K, Kollar EJ, Jaenisch R, Mooslehner K, Schwarz M, Haase K, Gmachl I, Harbers K. Retrovirus-induced insertional mutation in Mov13 mice affects collagen I expression in a tissue-specific manner. *Cell* 1989; 57: 807-16.
 18. Muratani K, Hada T, Yamamoto Y, Kaneko T, Shigeto Y, Ohue T, Furuyama J, Higashino K. Inactivation of the cholinesterase gene by Alu insertion: possible mechanism for human gene transposition. *Proc Natl Acad Sci USA* 1991; 88: 11315-9.
 19. Wallace MR, Andersen LB, Saulino AM, Gregory PE, Glover TW, Collins FS. A de novo Alu insertion results in neurofibromatosis type 1. *Nature* 1991; 353: 864-6.
 20. Kazazian HHJ, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* 1988; 332: 164-6.
 21. Dombroski BA, Scott AF, Kazazian HHJ. Two additional potential retrotransposons isolated from a human L1 subfamily that contains an active retrotransposable element. *Proc Natl Acad Sci USA* 1993; 90: 6513-7.
- est assurée par le gène *pgk2*, rétrocopie du gène *pgk1* qui prend le relais du premier au moment où celui-ci est inactivé ou absent.
- ### La sous-unité E1 α de la pyruvate déshydrogénase
- Une situation remarquablement similaire est trouvée avec le gène de la sous-unité E1 α de la pyruvate déshydrogénase humaine. Ce gène, localisé sur le chromosome X, comporte 10 introns et s'étend sur 17 kb. Une rétrocopie de ce gène, identique à 85 % au gène ancestral, est insérée sur le chromosome 4 [25]. La pyruvate déshydrogénase est une enzyme impliquée dans le métabolisme général et, là encore, la rétrocopie du gène initial fait figure de « copie de sauvegarde ». Cette seconde copie est exprimée exclusivement dans les testicules alors que la copie initiale est exprimée dans tous les autres tissus.
- La similitude des deux situations qui viennent d'être décrites a conduit à avancer l'hypothèse que, par suite d'un remaniement massif touchant la régulation de l'expression des gènes du chromosome X, une copie de « sauvegarde » des informations essentielles situées sur ce chromosome a été sélectionnée au cours de l'évolution. Cette opération semble avoir mis à contribution, au moins pour partie, des mécanismes de rétrotransposition par lesquels une copie d'un gène peut être déplacée dans un environnement génétique différent. Exportée dans le génome, la nouvelle copie peut s'affranchir des contraintes imposées à la région du génome d'où elle est originaire.
- ### Déplacement d'une région activatrice
- La régulation de gènes cellulaires, comme le gène *slp* ou les gènes du complexe amylase, font intervenir l'insertion d'éléments rétroviraux.
- ### Gène *slp*
- Le gène *slp* de la souris présente de fortes homologies avec le gène *C4* du complément, dont il dérive probablement par duplication. Sa fonction est inconnue mais la protéine pour laquelle il code est trouvée spécifiquement chez les souris mâles.
- La cartographie de la sensibilité à la DNase I a révélé une région sensible en amont du gène *slp* qui paraissait pouvoir correspondre à une séquence activatrice. L'analyse plus précise de la région a révélé l'existence d'une insertion rétrovirale en amont de la partie codante du gène *slp* [26]. L'étude détaillée du locus montre que le gène est sous contrôle de l'activateur viral. La région activatrice présente une homologie importante avec certains rétrovirus murins, or on sait que pour plusieurs d'entre eux (MMTV (*murine mammary tumor virus*) [27] et MoMuSV (*Moloney murine sarcoma virus*) [28]) la transcription peut être activée par les hormones androgènes. L'insertion rétrovirale constitue un locus rétroviral endogène, vraisemblablement sélectionné pour sa capacité à mettre en place une régulation spécifique. Toutefois, la fonction du gène *slp* n'étant pas connue, il est difficile de préciser en quoi cette régulation présente un avantage sélectif.
- ### Le complexe amylase
- Un cas intéressant, et très complet, d'analyse de séquences régulatrices apportées par un rétrovirus endogène est fourni par l'étude du locus de l'amylase chez les primates. Chez l'homme, les gènes *amylase* sont regroupés dans un complexe (*cluster*) comportant cinq gènes (*figure 2A*). Ces gènes sont fortement exprimés dans le pancréas et dans la glande salivaire. En fait, deux des gènes du complexe sont exprimés dans le pancréas (*AMY2A* et *AMY2B*) tandis que les trois autres (*AMY1A*, *AMY1B*, et *AMY1C*) sont exprimés dans la glande salivaire (*figure 2B*). Les gènes pancréatiques et salivaires présentent une forte similitude (93 %) [29].
- Deux types de rétroéléments ont été repérés en amont des gènes de l'amylase humaine. Il s'agit, d'une part, d'une rétrocopie du gène de la γ actine et, d'autre part, d'une séquence rétrovirale endogène. Chez la souris, les gènes de l'amylase sont également disposés sous forme d'un complexe de plusieurs copies homologues, mais la région ne comporte pas de séquence homologue du gène de la γ actine ni

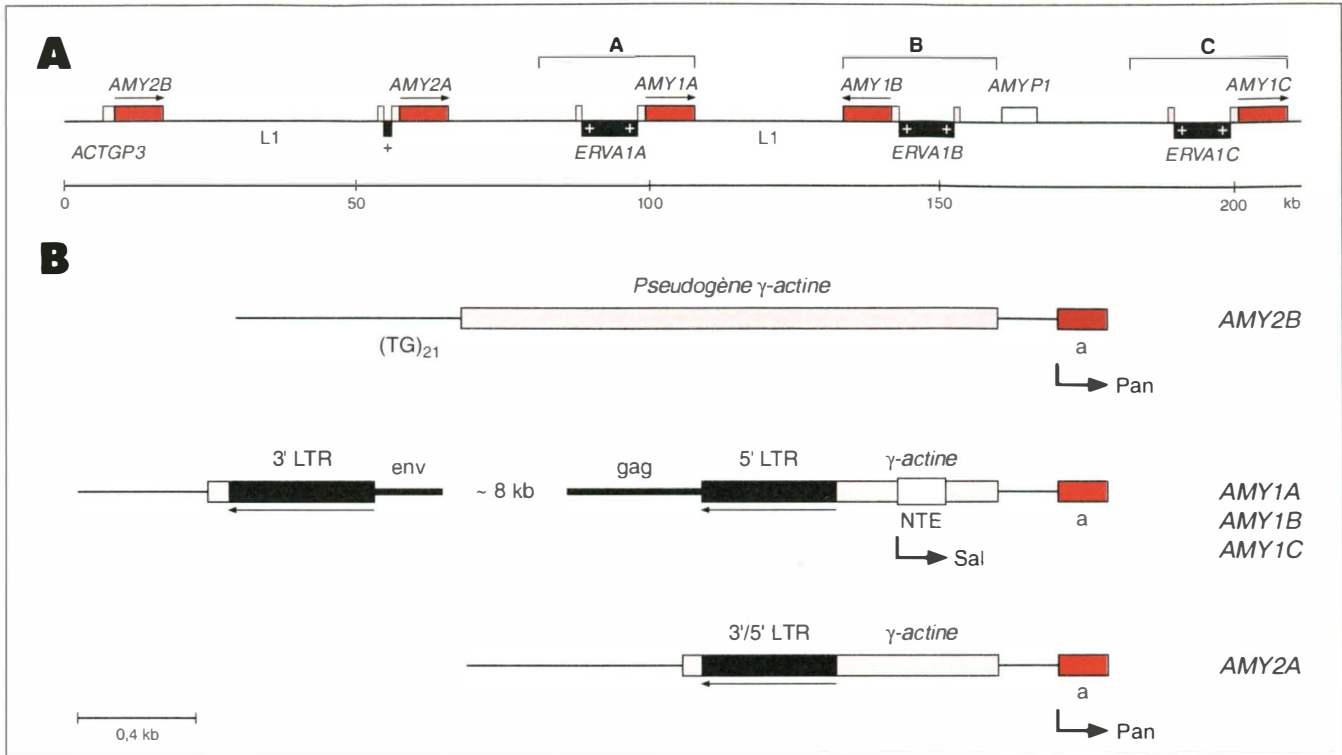


Figure 2. **Organisation du locus amylase humain** [29]. **A.** Le locus amylase humain s'étend sur plus de 200 kb et contient six copies du gène amylase, dont une non fonctionnelle (AMY P1). Deux types de rétroséquences sont observés en amont des gènes, une rétrocopie du gène de la γ -actine et une séquence rétrovirale endogène (ERVA1). Parmi les cinq copies fonctionnelles, deux sont exprimées dans le pancréas (AMY 2 A et B) et trois dans les glandes salivaires (AMY 1 A, B et C). Ces trois gènes comportent un rétrovirus endogène inséré dans la région régulatrice. **B.** Détail des structures des cinq gènes amylase fonctionnels. La figure indique les sites de démarrage de la transcription qui sont différents pour les gènes pancréatiques (pan) et salivaires (sal). Un provirus complet est inséré dans les trois gènes s'exprimant dans les glandes salivaires AMY 1 A, B et C. En revanche, on ne trouve aucune insertion rétrovirale dans le gène AMY 2 B et seulement un reliquat d'insertion rétrovirale (correspondant à un LTR) dans le gène AMY 2 A. Cette dernière insertion est vraisemblablement la conséquence d'une délétion de la séquence virale après insertion. (NTE : exon non traduit).

de rétrovirus endogène. L'insertion de ces deux rétroéléments a donc dû avoir lieu après la séparation des phylums homme et souris. L'analyse complète du locus fait apparaître une étroite corrélation entre la présence d'une insertion rétrovirale et l'expression dans les glandes salivaires chez l'homme : les cinq gènes du complexe possèdent un résidu de séquence codant pour la γ -actine en amont, les trois gènes exprimés dans les glandes salivaires possèdent, en outre, insérée dans cette séquence, une insertion rétrovirale endogène qui appartient à une famille de rétrovirus endogènes humains, et ont été nommés ERVA1.

L'un des deux gènes exprimés dans le pancréas (AMY2A) contient éga-

lement un reliquat d'insertion rétrovirale dans la région codant pour la γ -actine. L'insertion ne correspond pas à un rétrovirus entier, mais à une séquence LTR de rétrovirus ERVA1. La délétion de la région centrale du génome viral est vraisemblablement consécutive à une recombinaison entre les deux LTR. La transcription des gènes pancréatiques AMY2B et AMY2A est initiée à partir du premier exon du gène tandis que, pour les gènes salivaires AMY1 A/B/C, la transcription est initiée un peu en amont, à partir d'un exon non traduit situé dans la séquence codant pour la γ -actine (figure 2B), ce qui suggère l'existence de séquences régulatrices distinctes pour les deux types de gènes.

Afin de démontrer l'implication des séquences rétrovirales dans la régulation de l'expression des gènes AMY2, Samuelson *et al.* ont entrepris une série de constructions génétiques qu'ils ont testées chez des souris transgéniques [30]. La séquence clonée, comportant l'insertion virale suivie du gène AMY1C, a été déléetée de toute la région située en amont de l'insertion virale, ou de la plus grande partie du virus lui-même. Avec les constructions qui conservent la partie virale, l'expression est retrouvée dans les glandes salivaires, alors que la délétion des séquences virales abolit l'expression. Une autre série de constructions, dans lesquelles la séquence rétrovirale est fusionnée à un gène hété-

RÉFÉRENCES

22. Narita N, Nishio H, Kitoh Y, Ishikawa Y, Ishikawa Y, Minami R, Nakamura H, Matsuo M. Insertion of a 5' truncated L1 element into the 3' end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. *J Clin Invest* 1993; 91: 1862-7.

23. Boer PH, Adra CN, Lau YF, McBurney MW. The testis-specific phosphoglycerate kinase gene pgk-2 is a recruited retroposon. *Mol Cell Biol* 1987; 7: 3107-12.

24. McCarrey JR, Thomas K. Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. *Nature* 1987; 326: 501-5.

25. Dahl HH, Brown RM, Hutchison WM, Maragos C, Brown GK. A testis-specific form of the human pyruvate dehydrogenase E1 alpha subunit is coded for by an intronless gene on chromosome 4. *Genomics* 1990; 8: 225-32.

26. Stavenhagen JB, Robins DM. An ancient provirus has imposed androgen regulation on the adjacent mouse sex-limited protein gene. *Cell* 1988; 55: 247-54.

27. Yamamoto KR. Steroid receptor regulated transcription of specific genes and gene networks. *Annu Rev Genet* 1985; 19: 209-52.

28. Miksicsek R, Hebert A, Schmid W, Danesch U, Posseckert G, Beato M, Schütz G. Glucocorticoid responsiveness of the transcriptional enhancer of Moloney murine sarcoma virus. *Cell* 1986; 46: 283-90.

29. Samuelson LC, Wiebauer K, Snow CM, Meisler MH. Retroviral and pseudogene insertion sites reveal the lineage of human salivary and pancreatic amylase genes from a single gene during primate evolution. *Mol Cell Biol* 1990; 10: 2513-20.

30. Ting CN, Rosenberg MP, Snow CM, Samuelson LC, Meisler MH. Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes Dev* 1992; 6: 1457-65.

31. Britten RJ. Rates of DNA sequence evolution differ between taxonomic groups. *Science* 1986; 231: 1393-8.

32. Nouvel P. The mammalian genome shaping activity of reverse-transcriptase. *Genetica* 1994 (sous presse).

33. Xiong Y, Eickbush TH. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *Embo J* 1990; 9: 3353-62.

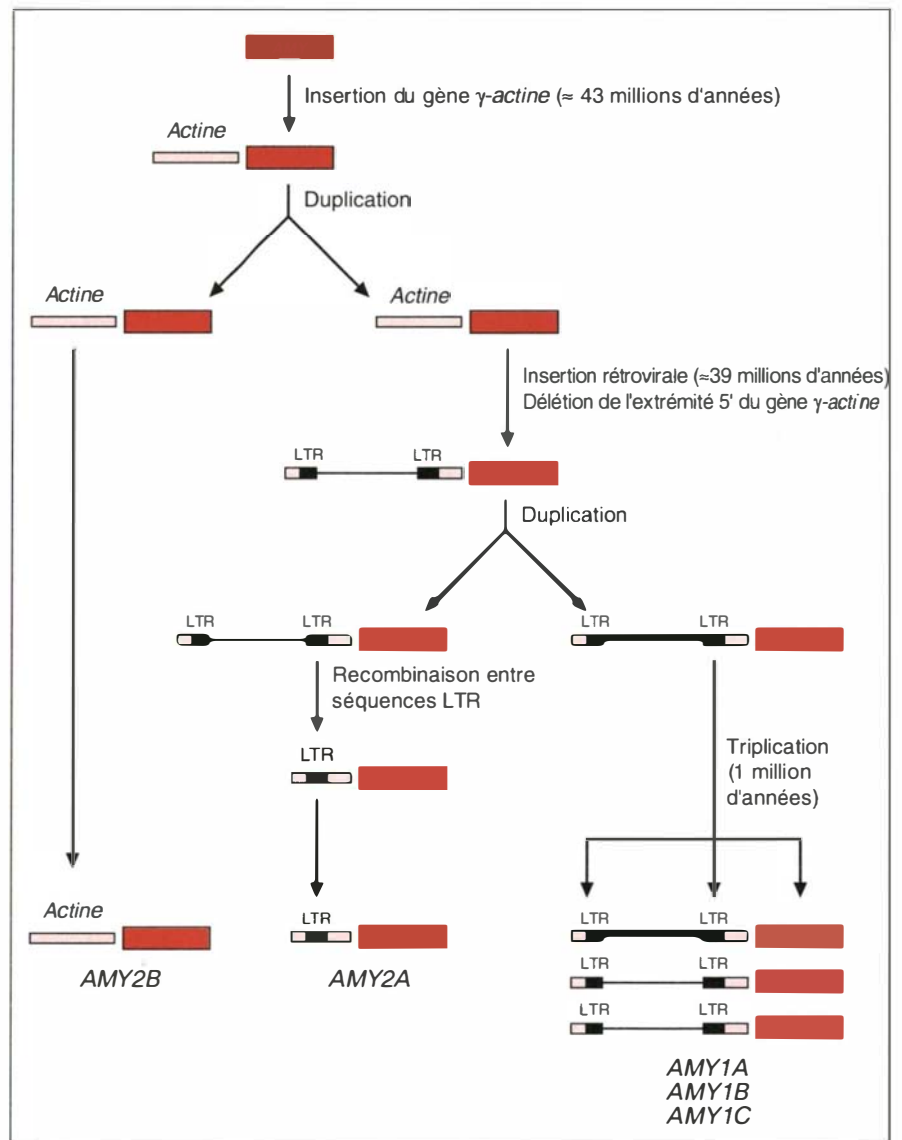


Figure 3. **Évolution moléculaire du complexe amylase** [29]. Séquence d'événements postulée pour rendre compte de la structure actuelle du locus amylase humain. Il y a environ 43 millions d'années, s'est produite l'insertion d'une rétrocopie du gène de la γ -actine, en amont du gène de l'amylase. Par la suite, l'ensemble ainsi formé s'est trouvé dupliqué. L'insertion d'une séquence rétrovirale en amont de l'un des deux gènes a vraisemblablement engendré une configuration génétique qui permet son expression dans les glandes salivaires. Une duplication de cette séquence, puis la perte de la séquence virale à l'exception d'une séquence LTR, ont conduit à la structure actuelle. Les gènes AMY 2 A et AMY 2 B sont exprimés dans le pancréas. Les gènes AMY 1 A, B et C sont exprimés dans les glandes salivaires.

rologue (hGH, *human growth hormone*), a permis de montrer que cette séquence possède bien les éléments nécessaires et suffisants pour assurer une expression dans les glandes salivaires.

On peut, finalement, dresser une carte assez précise des événements moléculaires majeurs ayant façonné le *locus* de l'amylase au cours de l'évolution des primates (*figure 3*). L'histoire moléculaire du *locus* amylase paraît pouvoir se résumer ainsi : il y a environ 43 millions d'années, s'est produite l'insertion d'une rétrocopie du gène de la γ -actine en amont du gène de l'amylase dans le phylum qui allait donner naissance aux primates. Par la suite, l'ensemble ainsi formé s'est trouvé dupliqué. L'insertion d'une séquence rétrovirale en amont de l'un des deux gènes permettant de faire basculer son expression dans les glandes salivaires a ensuite été vraisemblablement sélectionnée. Une nouvelle duplication de cette séquence, puis finalement la perte de la séquence virale à l'exception d'une séquence LTR, a conduit aux trois structures actuellement trouvées dans le génome des primates. Deux d'entre elles sont exprimées dans le pancréas et une dans les glandes salivaires. Enfin, chez l'homme, une triplification de la séquence ayant conservé l'insertion virale complète a conduit à la structure actuelle du *locus* humain. On parvient ainsi, à partir d'événements moléculaires simples, à rendre compte de la structure du *locus* amylase actuel.

En fait, chez la souris, dont le complexe amylase ne contient aucune insertion rétrovirale, le gène est également exprimé dans les glandes salivaires. Ce résultat indique que les régions régulatrices des gènes amylase ont suivi une évolution fonctionnelle convergente entre la souris et l'homme tout en recourant à des « bricolages moléculaires » différents. Le détail du bricolage mis en jeu dans l'élaboration de la fonctionnalité des gènes amylase de souris est inconnu. En revanche, la situation est assez bien démêlée dans le cas de l'homme, chez lequel il apparaît que les moyens mis en œuvre pour agencer régions régulatrices et régions codantes ont fait appel à des

déplacements d'informations dans le génome relayés par des rétro-transpositions. Ce cas est, actuellement, unique en son genre, dans la mesure où il permet de suivre avec une relative précision, le devenir et les transformations d'une région du génome.

Conclusion

Les exemples qui viennent d'être rapportés montrent que la rétrotransposition a joué un rôle dans la mise en forme des génomes de mammifères au cours de l'évolution. De nombreux effets directs ou indirects de son activité passée et présente sont repérables. Il peut s'agir, soit de recombinaisons entre séquences engendrées par transcription inverse, soit de déplacements de séquences codantes dans le génome, soit encore de déplacements de séquences activatrices. En fait, la rétrotransposition est le seul mécanisme connu de transposition de gènes chez les mammifères (pas de transposition *via* un intermédiaire ADN) et paraît être relativement importante d'un point de vue quantitatif, au moins sur des durées de l'ordre du million d'années (25 000 événements de rétrotransposition de séquence LINE par million d'années dans le génome de la souris). Il est toutefois difficile d'en évaluer l'importance évolutive relativement à d'autres types de mutations. Comme n'importe quel système en évolution, un génome doit maintenir l'équilibre entre ordre et désordre. On peut donc s'attendre à ce que l'apparition d'organismes composés d'un grand nombre de cellules se soit accompagnée de l'augmentation de la fidélité répliquative. Une diminution du taux de mutation ponctuelle est effectivement observée au cours de l'évolution [31]. Or cette dernière impose une fixité accrue du génome, déséquilibrant de ce fait le rapport ordre/désordre. Dans ces conditions, il est concevable que d'autres modes de variation aient été sélectionnés, en raison de leur capacité à rétablir l'équilibre en faveur d'un accroissement de désordre. Dans cette hypothèse, il est envisageable que la transcriptase inverse ait été

initialement sélectionnée en raison même de ses propriétés de modelage, adaptées à des génomes complexes [32]. Les rétrovirus et rétrotransposons, entités biologiques plus sophistiquées, apparus plus tardivement [33], représenteraient un sous-produit de cette activité ■

Summary

Shaping of the mammalian genome by reverse-transcriptase

Reverse-transcriptase catalyses the conversion of RNA into DNA. This operation might have played an essential role in primitive living systems and seems to have largely contributed to the evolution of complex genomes. More than 10 % of a mammalian genome is composed of sequences with reverse-transcribed origin, most of which consist in repeated sequences (SINEs, LINEs). In spite of their simplicity, these sequences may play a key role in evolution by favoring illegitimate recombination. Genes depleted from intervening sequences, known as pseudogenes, can be functional. Finally, endogenous retroviruses can be involved in the regulation of neighboring genes. These hallmarks of a reverse transcription activity indicate that it has largely contributed to the fluidity of modern genomes which remains apparent in genetic pathologies.

TIRÉS A PART

P. Nouvel.