

L'intégration des rétrovirus : faits et croyances

Depuis plusieurs années, le développement de la thérapie génique [1] et l'implication du VIH dans le SIDA [2] ont placé les rétrovirus sous les feux de la rampe. Deux étapes de leur cycle infectieux caractérisent cette classe de virus: l'ARN génomique est converti en ADN double brin grâce à la transcriptase inverse virale, et la molécule produite, l'intermédiaire d'intégration, s'insère ensuite de manière permanente dans le génome des cellules hôtes. C'est ce dernier point qui est abordé dans cette mini-synthèse.

Le cycle infectieux de tous les rétrovirus peut être décomposé de la manière suivante (pour revues, consulter les références [3-5]): (1) ils se lient aux cellules susceptibles par l'interaction entre la glycoprotéine d'enveloppe virale et un récepteur membranaire spécifique; (2) la fusion des membranes cellulaire et virale permet la translocation de la capsid virale dans le cytoplasme; (3) à l'intérieur de la capsid, toujours localisée dans le cytoplasme, s'effectue la transcription inverse; (4) l'ADN ainsi synthétisé est ensuite transporté, au sein d'un complexe nucléoprotéique de grande taille, dans le noyau où il s'intègre au génome de la cellule hôte; (5) c'est à partir du provirus que sont produites les nouvelles particules virales, en utilisant les machineries cellulaires de transcription, de maturation des ARN pré-messagers et de traduction; (6) les particules virales sont assemblées dans la cellule puis expulsées sans provoquer la lyse des cellules (*figure 1*). Si les différentes étapes du cycle rétroviral sont bien identifiées, il n'en est pas toujours de même des mécanismes biochimiques impliqués. Certains, cependant, comme la transcription inverse ou

l'intégration, sont maintenant bien documentés... ce qui n'empêche pas certaines idées fausses d'être encore communément véhiculées, y compris par des manuels d'enseignement... Nous nous attacherons ici à montrer que, d'une part, la synthèse d'ADN cellulaire, par elle-même, n'est pas nécessaire à l'intégration provirale et que, d'autre part, l'intermédiaire dit « proximal » est une molécule linéaire et non pas circulaire. Pour une information exhaustive et des références, le lecteur pourra se référer aux excellentes revues de Varmus et Brown [6], Goff [7] et Whitcomb et Hughes [8].

La transcription inverse engendre une molécule d'ADN double brin linéaire.

L'ARN génomique viral est synthétisé par l'ADN polymérase II des cellules infectées, à partir du provirus intégré. Il ressemble de ce fait à un ARN messager classique présentant, respectivement, une coiffe et une queue de polyA à ses extrémités 5' et 3'. Il est, en outre, limité par de courtes répétitions directes (16 à 70 nucléotides suivant le rétrovirus): la région R.

Dans chaque virion sont empaquetées deux molécules d'ARN génomiques étroitement associées ainsi que 50 à 100 molécules de chacune des enzymes virales: la transcriptase inverse (RT) et l'intégrase (In) virales; cette dernière est l'enzyme clé de l'intégration. Moins de quelques heures après l'infection, les deux ARN génomiques sont transformés en une seule copie d'ADN proviral par un mécanisme complexe (non détaillé ici) appelé transcription inverse [8]. Fait remarquable, la copie d'ADN double brin linéaire

est plus longue que l'ARN génomique. Cet accroissement de taille résulte de l'association et de la duplication, aux deux extrémités du provirus, des séquences terminales de l'ARN rétroviral (pour les détails, voir la *figure 2* (2.1 et 2.2)). La structure répétée qui borde le provirus est appelée LTR pour *long terminal repeat*. Les LTR comprennent 300 à 1 300 nucléotides suivant le virus. Ils se composent de la succession des motifs U3-R-U5; U3 correspond à l'extrémité 3' dupliquée de l'ARN viral et U5 à l'extrémité 5' dupliquée. Une autre caractéristique importante des LTR est que chacun est limité par de courtes séquences répétées inversées appelées IR (6 à 13 nucléotides suivant le virus) qui sont absolument nécessaires à l'intégration du provirus (*figure 2.3* et 2.4). Outre cette fonction, les LTR sont importants pour l'expression du génome viral. Ils portent, en effet, les signaux d'initiation de la transcription (*enhancer*, boîte CAAT, boîte TATA), fonctionnels dans le LTR gauche, et un signal de polyadénylation, fonctionnel dans le LTR droit (*figure 2.3*). La transcription du provirus commence à l'extrémité 5' de R dans le LTR gauche et la polyadénylation intervient à l'extrémité 3' de R dans le LTR droit, ce qui explique la répétition de cette séquence en bordure des ARN rétroviraux.

Dans le cytoplasme, l'ADN néosynthétisé est associé à un complexe nucléoprotéique de grande taille (environ 160S). Ce complexe ne peut être purifié qu'au prix de gros efforts. C'est ce qui explique que sa composition précise soit encore inconnue. Cependant, des expériences *in vitro* ont montré que, dans le cytoplasme, il possède déjà tous les

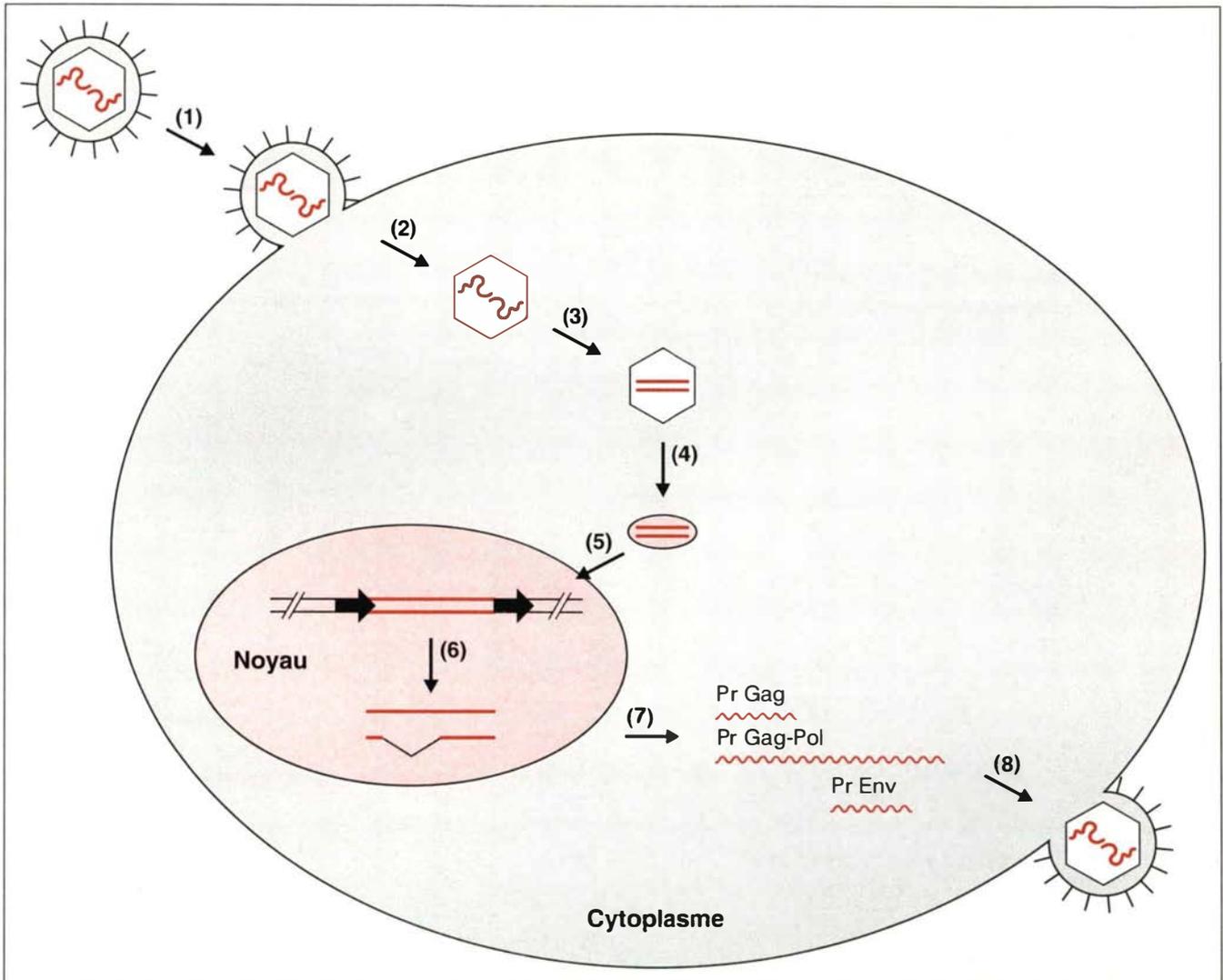


Figure 1. **Cycle infectieux des rétrovirus.** Le cycle infectieux des rétrovirus se déroule en plusieurs étapes. (1) Le rétrovirus reconnaît un récepteur spécifique à la surface des cellules cibles grâce à sa glycoprotéine d'enveloppe. (2) Après fusion des membranes virale et cellulaire, la nucléocapside virale est libérée dans le cytoplasme. (3) Le dimère d'ARN génomique est transcrit en un seul provirus, par transcription inverse, dans la capsidie toujours localisée dans le cytoplasme. (4) Le provirus migre vers le noyau au sein d'un complexe nucléoprotéique de haut poids moléculaire contenant des protéines de capsidie et l'intégrase rétrovirale (160 à 170S pour les virus de mammifères). (5) Le provirus s'intègre dans l'ADN de la cellule hôte. (6) Les ARN génomiques et sous-génomiques viraux sont synthétisés à partir du provirus intégré. (7) Dans le cytoplasme, ils sont traduits en précurseurs Gag, Gag-Pol et Env qui sont subséquentement protéolysés pour donner les protéines virales. (8) Ces dernières sont utilisées pour constituer les particules virales qui sont relarguées dans le milieu extracellulaire par bourgeonnement (c'est-à-dire sans lyse cellulaire).

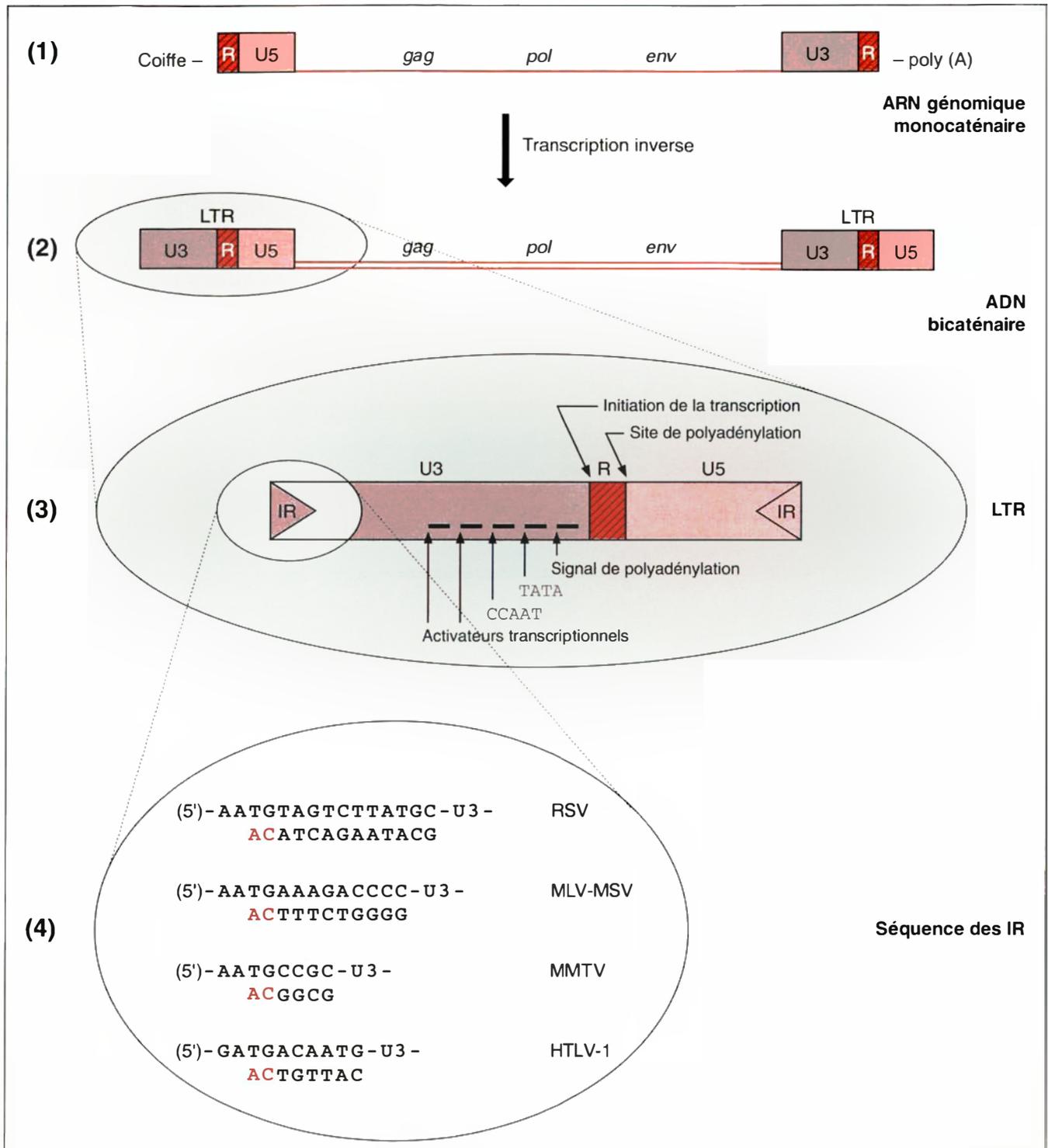


Figure 2. **Genèse et structure des LTR rétroviraux.** Dès le début de l'infection, (1) les deux molécules d'ARN génomique monocaténaire codant pour les protéines enzymatiques (Pol), les protéines structurales de capsid (Gag) et d'enveloppe (Env) donnent naissance à une seule molécule d'ADN bicaténaire, grâce à la transcriptase inverse. (2) Le provirus est limité par deux LTR (long terminal repeats) (3) subdivisés en trois régions U3, R et U5 et qui sont eux-mêmes bordés par de courtes séquences inversées répétées (IR). Toutes les informations génétiques nécessaires à l'initiation de la transcription (activateurs, boîte CAAT, boîte TATA) et à la polyadénylation des ARN viraux sont contenues dans U3, à l'exception des virus leucémogènes et sarcomatogènes murins chez lesquels le signal de polyadénylation est situé dans R. Les séquences des IR de quelques virus sont données en (4). On peut observer que la dégradation exonucléasique des extrémités 3' précédant l'intégration s'arrête systématiquement au niveau du dinucléotide CA. RSV: Rous sarcoma virus; MLV-MSV: murine leukemia-sarcoma virus; MMTV: mouse mammary tumor virus; HTLV-1: human T cell leukemia virus.

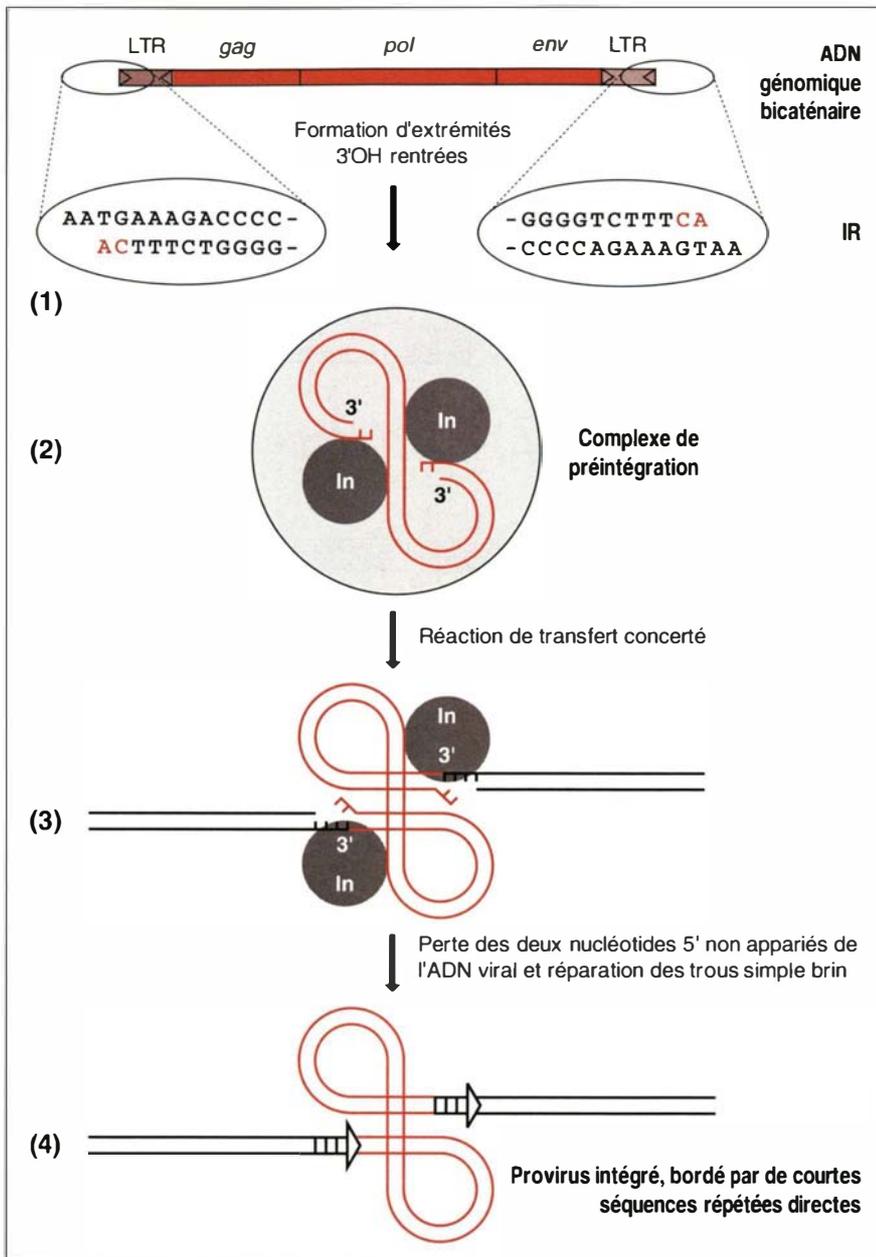


Figure 3. **Réaction d'intégration.** (1) Après transcription inverse de l'ADN génomique, l'intégrase engendre des extrémités 3' OH rentrées au niveau de chacune des séquences inversées répétées (IR) terminales du provirus en provoquant la perte de deux nucléotides. Leur excision s'arrête à un motif CA conservé dans tous les rétrovirus. (2) Cette molécule migre ensuite vers le noyau au sein d'un complexe protéique de haut poids moléculaire dans lequel seules l'intégrase virale (In) et les protéines de capsid ont pu être identifiées. (3) Par une réaction de transfert concerté, les extrémités 3' OH de l'ADN viral sont liées de manière covalente à l'ADN de l'hôte. « L'attaque » de ce dernier par le provirus s'effectue en deux sites situés sur chacun des brins d'ADN et séparés par quelques nucléotides. (4) Une dernière étape, probablement réalisée par les enzymes cellulaires, permet la réparation des trous (gaps) à la jonction de l'ADN cible et de l'ADN viral après la perte des deux nucléotides 5' terminaux du provirus. Résultant de l'attaque décalée sur les deux brins d'ADN et de la réparation, deux courtes séquences répétées, dont la longueur est caractéristique de chaque virus, sont retrouvées de part et d'autre du provirus intégré (4 à 6 paires de bases).

composants nécessaires à l'intégration. Les seules protéines qui ont pu y être reconnues à ce jour sont des protéines de capsid (Ca) et l'intégrase rétrovirale. A l'exception du cas du rétrovirus murin MoMuLV, le mécanisme par lequel la particule de 160S pénètre dans le noyau est inconnu (voir ci-dessous).

Le mécanisme de l'intégration

Il est maintenant établi que le véritable substrat d'intégration est la forme linéaire du provirus produite dans le cytoplasme. Cependant, des formes circulaires libres d'ADN viral contenant un ou deux LTR intacts sont aussi retrouvées dans les noyaux des cellules infectées. Ces molécules sont formées, soit par ligature des extrémités des LTR (quand les deux LTR sont juxtaposés), soit par recombinaison homologue entre les deux LTR (quand un seul LTR est présent). Il existe, en outre, des molécules circulaires plus complexes, produites par autointégration d'une des extrémités dans l'ADN viral. Aucune fonction n'est connue ni même soupçonnée pour toutes ces molécules. La tendance actuelle est plutôt de les considérer comme des produits abortifs de l'intégration.

Des approches à la fois génétiques et biochimiques ont permis de montrer que les IR et l'intégrase sont indispensables à l'intégration du provirus. Le processus complet peut être décomposé en trois étapes (figure 3): (1) dans un premier temps, deux bases sont éliminées de l'extrémité 3' de chaque chaîne d'ADN proviral. Il semble que l'intégrase soit la seule enzyme impliquée et que la réaction puisse déjà s'effectuer dans le cytoplasme (figure 3.1 et 3.2); (2) dans un second temps, les nouvelles extrémités 3' de l'ADN viral sont liées à l'ADN cible à quelques bases (4 à 6) de distance sur les brins opposés. Pour autant qu'on puisse l'affirmer à partir d'expériences conduites *in vitro*, l'intégration ne s'effectue pas au niveau de cassures pré-existantes de l'ADN. Il s'agit d'une réaction concertée de transfert entre les deux substrats qui

ne nécessite pas d'énergie exogène (en d'autres termes, il ne s'agit pas de la succession d'une coupure et d'une ligature effectuées indépendamment l'une de l'autre, mais de « l'attaque » du génome de l'hôte par les extrémités 3' de l'ADN viral) (pour les détails, voir la figure 3.3) ; (3) enfin, le produit initial de la réaction d'intégration est réparé. Il présente en effet deux trous, liés, d'une part, à la création d'une coupure décalée de l'ADN hôte qui engendre de courtes séquences simple brin complémentaires, et, d'autre part, deux courtes queues (2 bases) non appariées, liées à la perte des deux nucléotides 5' des extrémités de l'ADN proviral. On ne sait pas encore si la réparation est assurée par des enzymes cellulaires ou bien par un complexe impliquant l'intégrase et la transcriptase inverse. Il n'en reste pas moins qu'en fin de processus le provirus intégré est flanqué de courtes répétitions directes de séquences hôtes : 4 nucléotides pour les virus de la leucémie murine et féline, 5 pour les virus aviaires et VIH et 6 pour le virus de la tumeur mammaire de la souris (figure 3.4).

Pendant longtemps, on a pensé que l'intermédiaire proximal de l'intégration était l'ADN circulaire à deux LTR. Si, à l'origine, on possédait bien des arguments indirects d'ordre expérimental pour le supposer, il faut admettre que certains modèles de recombinaison bactérienne ont fortement aidé à la popularisation de cette idée. En outre, en 1984, Panganiban et Temin [9] ont rapporté que la jonction palindromique des deux LTR des provirus circulaires du virus aviaire SNV pouvait être utilisée pour l'intégration. En fait, cette observation n'a jamais pu être reproduite avec d'autres rétrovirus et l'on sait maintenant que, même si cette structure est utilisable dans le modèle SNV, elle est infiniment moins efficace que la forme linéaire.

L'intégrase

Dans les rétrovirus, la transcriptase inverse et l'intégrase sont codées par le gène *pol*. La génétique a permis

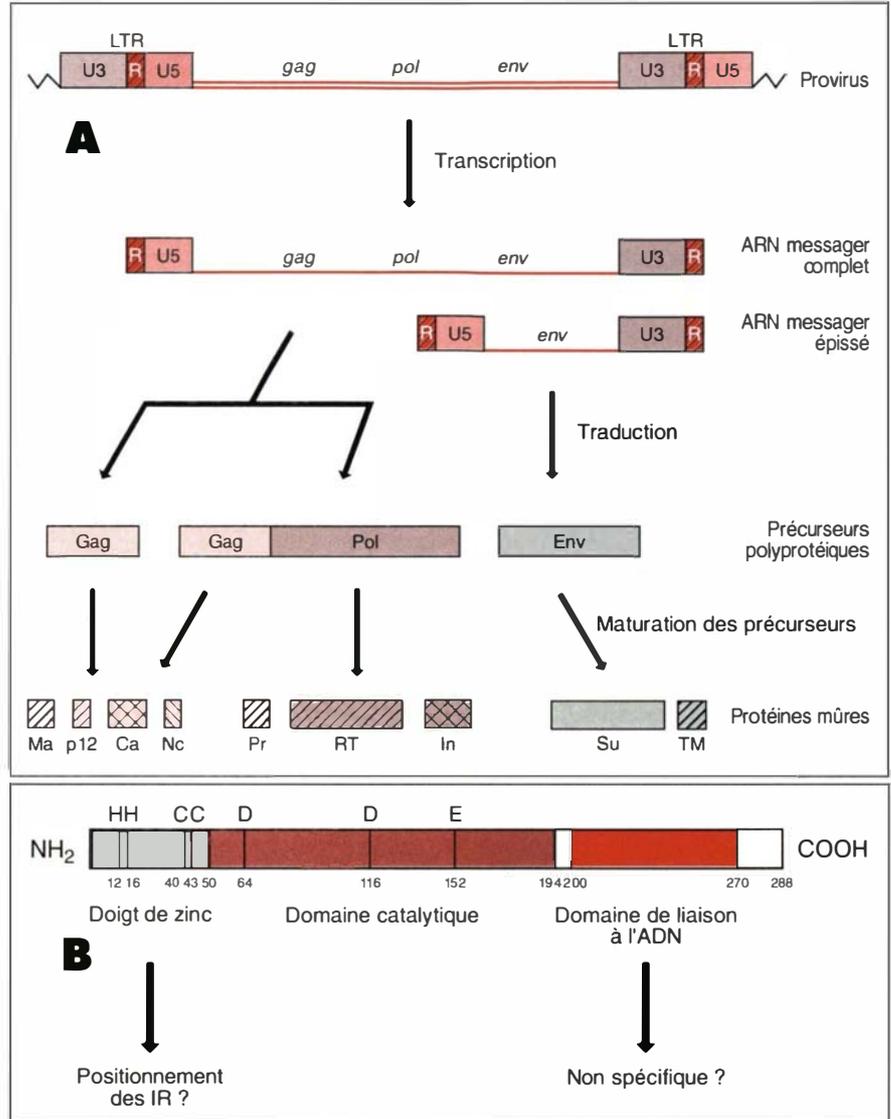


Figure 4. **Genèse et structure de l'intégrase.** (A) **Genèse de l'intégrase** (cas des virus murins et aviaires non défectueux). La transcription de l'ADN proviral engendre deux transcrits : un ARN complet qui sera utilisé, soit comme ARN génomique, soit comme ARN messager, et un ARN messager épissé. Puis, par traduction de ces deux ARN, trois précurseurs polyprotéiques (Gag, Gag-Pol et Env) donneront naissance aux différentes protéines virales par protéolyse spécifique. Les protéines de capsid (Ma : matrice, p12, Ca : capsid et Nc : nucléocapsid) sont issues des précurseurs Gag et Gag-Pol. Trois protéines sont codées par le gène *pol* : Pr (protéase), RT (transcriptase inverse) et In (intégrase). Deux protéines sont produites à partir du précurseur *env* pour constituer les protéines d'enveloppe : Su (glycoprotéine de surface) et TM (protéine transmembranaire). (B) **Structure de l'intégrase.** La structure de l'intégrase du VIH, présentée ci-dessus, est représentative des intégrases rétrovirales et est constituée de 288 acides aminés. Elle peut être divisée en trois régions selon leur fonction. En position N-terminale, un doigt de zinc présomptif, comprenant des acides aminés conservés histidine (H) et cystéine (C), joue un rôle dans la reconnaissance des séquences inversées répétées (IR) de l'ADN viral et dans le positionnement correct des LTR pour la réaction finale. Une région centrale (acides aminés 50 à 194) constitue le domaine catalytique impliqué dans le transfert du brin d'ADN viral. Enfin, un domaine de liaison à l'ADN, non spécifique, est situé en position C-terminale.

de localiser la fonction « intégrase » à l'extrémité 3' de ce dernier. La protéine In est produite au cours de l'assemblage des virions à partir d'un précurseur polyprotéique Gag-Pol de haut poids moléculaire grâce à la protéase virale (Pr), elle-même codée par Pol (figure 4A). Chez les rétrovirus infectant les mammifères, une protéine In indépendante (p46 pour MoMuLV, p32 pour VIH) est produite. En revanche, chez les rétrovirus aviaires, si la fonction In existe dans une protéine indépendante de 32 kDa (pp32), elle est aussi associée à la transcriptase inverse (sous-unité β). Il semble que la dimérisation de In, et peut-être même sa tétramérisation pour certains virus, soit nécessaire à la réaction d'intégration. Les intégrases rétrovirales possèdent deux régions conservées : un doigt de zinc présumé à l'extrémité N-terminale, et un motif caractéristique de la région centrale, retrouvé aussi dans les transposases bactériennes (qui permettent, elles aussi, l'intégration d'ADN dans de l'ADN bactérien ou phagique) (figure 4B). La mutagenèse des résidus histidine et cystéine du doigt de zinc a montré que celui-ci assure un rôle crucial dans la fonction d'intégration en permettant probablement la reconnaissance des extrémités des LTR et leur positionnement correct au cours de la réaction. L'autre motif consiste en trois acides aminés arrangés de la façon suivante Asp-Asp-(X)₃₅-Glu. Leur mutation a montré qu'ils sont essentiels à la catalyse du transfert polynucléotidique. Enfin, les protéines In possèdent une activité de liaison à l'ADN, non spécifique, portée par l'extrémité C-terminale (figure 4B).

Les séquences IR

Aucune séquence interne à l'ADN viral n'est nécessaire à l'intégration. Seuls sont requis les motifs IR (figure 2). En fait, les 5 à 10 nucléotides terminaux, selon le rétrovirus, sont véritablement importants. Chez certains virus, les exigences à chacune des deux extrémités sont même légèrement différentes. La machinerie d'intégration semble

assez tolérante au changement. Ainsi l'ont montré l'isolement de mutants naturels, des expériences de mutagenèse dirigée, et le fait que, *in vitro*, l'In de MoMuLV soit capable d'utiliser les IR de VIH, le contraire n'est toutefois pas vrai. En fait, le seul motif nécessaire et invariant des IR est le dinucléotide CpA au niveau duquel s'arrête l'élimination des deux nucléotides 3' terminaux (figures 2.4 et 3.1). C'est le nucléotide A qui est universellement lié à l'ADN de l'hôte.

L'intégration n'est pas complètement aléatoire

Les rétrovirus s'intègrent en d'innombrables sites. Cependant, l'intégration n'est pas complètement aléatoire (pour revue et références, voir [10]). La préférence semble essentiellement gouvernée par l'accessibilité du domaine d'insertion puisque aucune spécificité de séquences cibles dans l'ADN hôte n'a pu être mise en évidence, que ce soit au niveau primaire ou secondaire. La majorité des intégrations rétrovirales intervient dans les régions transcrites de la chromatine, c'est-à-dire dans les zones de faible condensation des chromosomes. Les régions d'hypersensibilité aux DNAses (régions associées aux sites de régulation de l'expression des gènes), dépourvues de nucléosomes, sont préférentiellement visées. Cependant, même dans la chromatine active, il existe de grandes disparités. Ainsi, les insertions rétrovirales sont, d'après certains auteurs, privilégiées au niveau de certains proto-oncogènes dont elles peuvent parfois activer le potentiel tumorigène. C'est, par exemple, le cas du proto-oncogène *c-myc* chez le poulet. *A contrario*, elles sont plutôt rares dans des régions actives comme le gène *hprt** de la souris.

La synthèse d'ADN n'est pas nécessaire à l'intégration

La notion couramment admise que l'intégration provirale intervient au cours de la phase S du cycle cellulaire est fautive. Une première

preuve convaincante est que VIH est capable d'infecter des macrophages différenciés [11], des cellules arrêtées en phase G2 par irradiation gamma [12] et même des lymphocytes T quiescents [13]. Ces observations suggèrent que le complexe nucléoprotéique proviral de VIH est capable de franchir la barrière que constitue l'enveloppe nucléaire, quelle que soit la phase du cycle cellulaire. Il n'en est clairement pas de même pour tous les rétrovirus. En utilisant des cultures de cellules synchrones, Roe *et al.* [14] ont fortement suggéré que l'exigence que manifeste le virus murin MoMuLV vis-à-vis des cellules pour son intégration ne semble pas être la synthèse d'ADN mais la mitose. Il semble que ce soit la simple destruction de l'enveloppe nucléaire pendant cette période qui permette l'accès à l'ADN de la cellule hôte. Cette observation suggère, en second lieu, que les complexes proviraux cytoplasmiques des différents rétrovirus ne sont pas tous équivalents du point de vue biochimique et fonctionnel pour la pénétration nucléaire. Ainsi, on peut supposer que dans le cas de VIH, soit certaines protéines virales, soit certaines protéines cellulaires associées à la particule de 160S, possèdent des séquences de localisation nucléaire (NLS) permettant d'emprunter les pores nucléaires.

En conclusion, le concept erroné de la nécessité de synthèse d'ADN pour l'intégration provirale ne provient pas de faits expérimentaux clairs et établis mais, peut-être plus, de la surinterprétation par d'autres auteurs d'observations originales effectuées par Varmus *et al.* [15]. Ces derniers avaient rapporté que des fonctions cellulaires étaient nécessaires à la transcription inverse et à l'intégration du virus aviaire ASV. En particulier, ils notaient que l'intégration n'était intervenue que

* gène *hprt*: gène de l'hypoxanthine guanine phosphoribosyl transférase. L'étude de la fréquence de l'intégration du provirus MoMuLV dans ce gène a montré une intégration 100 fois plus faible que ne le voudrait une intégration au hasard.

dans les cellules qui avaient dupliqué leur ADN. En fait, les cellules s'étaient aussi divisées, mais les modèles alors en cours sur la recombinaison génétique ont certainement favorisé l'hypothèse de la nécessité de réplication. Il est cependant maintenant raisonnable de penser, à la lumière des résultats de Roe *et al.* [14], ainsi que de la réinterprétation d'autres données de la littérature, que la véritable exigence pour ASV est la même que celle de MoMuLV, c'est-à-dire la destruction de l'enveloppe nucléaire au cours de la mitose. Attendons-en, toutefois, la démonstration formelle pour l'affirmer... ■

RÉFÉRENCES

1. Kahn A. *Thérapie génique : l'ADN médicament*. Paris: John Libbey Eurotext, 1993, 172 p.
2. Levy JP. Traitement du SIDA: recherche de nouveaux médicaments et élaboration de thérapies géniques. *médecine/sciences* 1991; 7: 830-41.
3. Darlix JL. Eléments de structure et variabilité du rétrovirus humain HIV-1. *Bull Institut Pasteur* 1991; 89: 211-42.
4. Lazo PA, Tschlis PN. Biology and pathogenesis of retroviruses. *Semin Oncol* 1990; 17: 269-94.
5. Cann AJ, Karn J. Molecular biology of HIV: new insights into the virus life-cycle. *AIDS* 1989; 3 (suppl 1): 519-34.
6. Varmus H, Brown P. Retroviruses. In: Mobile DNA, Berg DE, Howe MM, eds. Washington DC, *Am Soc Microbiol* 1989: 53-108.
7. Goff S. Genetics of retroviral integration. *Annu Rev Genet* 1992; 26: 527-44.

8. Whitcomb JM, Hughes SH. Retroviral reverse transcription and integration: progress and problems. *Annu Rev Cell Biol* 1992; 8: 275-306.
9. Panganiban AT, Temin HM. Circles with two tandem LTR are precursors to integrated retrovirus DNA. *Cell* 1984; 36: 673-9.
10. Sandmeyer SB, Hansen IJ, Chalker DL. Integration specificity of retrotransposons and retroviruses. *Annu Rev Genet* 1990; 24: 491-518.
11. Weinberg JB, Matthews TJ, Cullen BR, Malim MH. Productive human deficiency virus type 1 (HIV-1) infection in non-proliferative monocytes. *J Exp Med* 1991; 174: 1477-82.
12. Lewis P, Hensel M, Emerman M. Human immunodeficiency virus infection of cells arrested in the cell cycle. *EMBO J* 1992; 11: 3053-8.
13. Li G, Simm M, Potash MJ, Volsky DJ. Human immunodeficiency virus type 1 DNA synthesis, integration and efficient viral replication in growth-arrested T cells. *J Virol* 1993; 67: 3969-77.
14. Roe T, Reynolds TC, Yu G, Brown PO. Integration of murine leukemia virus DNA depends on mitosis. *EMBO J* 1993; 12: 2099-108.
15. Varmus H, Padgett T, Heasley S, Simon G, Bishop M. Cellular functions are required for the synthesis and integration of avian sarcoma virus-specific DNA. *Cell* 1977; 11: 307-19.

Mariana Marin
Maryse Étienne-Julan
Marc Piechaczyk
Danièle Noël

*Institut de génétique moléculaire, Cnrs
UMR 9942, BP 5051, route de Mende,
34033 Montpellier Cedex 01, France.*

TIRÉS A PART

D. Noël.