

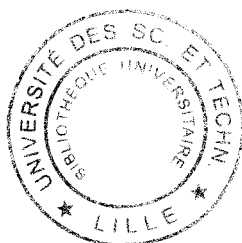
Th 2000 892 50376
2000
312

UNIVERSITE DES SCIENCES ET TECHNOLOGIES DE LILLE

N° d'ordre : 2741

THESE

Présentée pour l'obtention du grade de :
**DOCTEUR DE L'UNIVERSITE DES SCIENCES
ET TECHNOLOGIES DE LILLE
EN
SCIENCES DE LA VIE ET DE LA SANTE**



par
India LECLERCQ

**ASPECTS MOLECULAIRES DE L'INTEGRATION DU RETROVIRUS
HTLV-1 ET DE L'EXPANSION CLONALE DE SA CELLULE HOTE**

Soutenue le 26 Mai 2000 devant la Commission d'Examen :

Président : Jean-Pierre BOHIN
Rapporteurs : Thérèse ASTIER-GIN
Luc WILLEMS
Examineurs : Jean-François MOUSCADET
Antoine GESSAIN
Directeur de thèse : Eric WATTEL

Je souhaite exprimer mes sincères remerciements à Eric WATTEL qui a dirigé cette étude et m'a accordé sa confiance. Son enthousiasme pour la recherche et son aide précieuse ont beaucoup contribué à la réalisation de ces travaux.

Ce travail a été effectué à l'Institut de Recherches sur le Cancer de Lille, au sein de l'Unité INSERM 524 "Génétique Moléculaire et Approches Thérapeutiques des Hémopathies Malignes" dirigée par le Dr Jean-Pierre KERCKAERT. Je le remercie de m'avoir accueillie au sein de son unité de recherches et d'avoir encouragé la réalisation de cette étude.

Je tiens à remercier le Dr Jean-Pierre BOHIN qui m'a fait l'honneur d'accepter la présidence de ce jury de thèse.

Je remercie vivement les Dr Thérèse ASTIER-GIN et Luc WILLEMS d'avoir pris le temps de juger ce travail et d'avoir accepté d'en être les rapporteurs.

Je remercie également les Dr Antoine GESSAIN et Jean-François MOUSCADET d'avoir accepté de siéger dans ce jury.

Ce travail a été financé par le Ministère de l'Enseignement Supérieur et de la Recherche et par l'Association de Recherche contre le Cancer.

Je remercie Marielle CAVROIS, Franck MORTREUX et Anne-Sophie GABET pour leur dynamisme, leurs idées et l'aide qu'ils m'ont apportée au cours de ces années de travail.

Je tiens à remercier chaleureusement Jean-Claude D'HALLUIN pour son aide et ses précieux conseils, notamment lors de la rédaction de ce manuscrit.

Je remercie Laurence RENAUT, Ricardo GONZALEZ et Rodolphe VEREECQUE pour leur sympathie et leur bonne humeur.

Je remercie enfin tout le personnel de l'Unité INSERM 524 et de l'IRCL pour leur accueil et leur gentillesse.

Je souhaiterai maintenant adresser une pensée sincère et chaleureuse à tous mes amis, sans qui ce travail n'aurait pu aboutir. Merci à Guillaume pour son soutien et ses conseils expérimentés. Merci à tous...

Ce mémoire est enfin l'occasion pour moi de témoigner toute mon affection à mes parents, ma sœur et le reste de ma famille.

INTRODUCTION	15
GENERALITES	18
I. Le rétrovirus HTLV-1.	20
I.1. Organisation génétique du rétrovirus HTLV-1.....	20
I.1.1. Le provirus HTLV-1.	22
I.1.2. Les protéines de structure.....	22
I.1.3. Les enzymes virales.	23
I.1.4. Protéines de régulation codées par la région pX.	23
I.2. Tropisme de HTLV-1.....	31
I.3. L'ATLL.....	32
I.3.1. Données cliniques.	32
I.3.2. Effet de HTLV-1 sur la prolifération cellulaire et la transformation maligne.	33
I.4. La TSP/HAM.	36
I.5. Réplication du rétrovirus HTLV-1.....	39
I.5.1. Charge virale et stabilité génétique.	39
I.5.2. Expansion clonale des cellules infectées.....	40
I.5.3. Mutations somatiques de la séquence provirale et instabilité génétique de la cellule hôte.	40
I.5.4. Incidences pathologiques de l'expansion clonale des cellules infectées.....	41
II. L'intégration des rétrovirus.	42
II.1 La réaction d'intégration.....	42
II.1.1. Principe général de la réaction d'intégration.	44
II.1.2. Méthodes d'étude de l'intégration.	45
II.1.3. La réaction de clivage spécifique ou "3' processing".....	48
II.1.4. La réaction de transfert de brin.	49
II.1.5. La désintégration.....	50
II.2. L'intégrase.	50
II.2.1. Organisation structurale de l'IN.	50
II.2.2. La forme oligomérique de la protéine.....	55
II.3. Les complexes de préintégration ou PICs.....	56
II.3.1. Structure et composition.	56
II.3.2. Transport nucléaire des complexes de préintégration.....	57
II.4. Rôle des protéines cellulaires ou virales dans le mécanisme d'intégration.	59
II.4.1. Les complexes de préintégration.	59
II.4.2. Les IN purifiées.....	59
II.4.3. La réparation de l'ADN cible.	60
II.4.4. Les réactions "concertées".	61
II.5. Reconnaissance des substrats.....	61
II.5.1. Reconnaissance de l'ADNv.	62
II.5.2. Reconnaissance de l'ADN cible.	64
II.6. Topographie de l'intégration.....	65
II.6.1. Influence de la composition nucléotidique de l'ADN cible.....	65
II.6.2. Régions préférentielles d'intégration.	66
II.6.3. Le rôle des protéines fixées à l'ADN cible.....	67
II.6.4. Cas particulier de HTLV-1.	68

RESULTATS-DISCUSSION	69
PARTIE I : ETUDE DE LA COMPOSITION NUCLEOTIDIQUE DES SITES D'INTEGRATION DE HTLV-1 <i>IN VIVO</i>.	
I.1. Introduction.	70
I.2. Résultats.	72
I.2.1. Les sites d'intégration de HTLV-1 sont riches en nucléotides A/T.	75
I.2.2. Distribution non homogène de la composition nucléotidique des sites d'intégration de HTLV-1.	79
I.2.3. Identification des séquences ciblées par l'intégrase de HTLV-1 <i>in vivo</i>	82
I.3. Discussion.	87
PARTIE II : INFLUENCE DE LA STRUCTURE DE L'ADN HOTE SUR LE CHOIX DE L'INTEGRASE DE HTLV-1 <i>IN VITRO</i>.	
II.1. Introduction.	93
II.2. Mise au point de la stratégie expérimentale.	94
II.2.1. Expression et purification des intégrases de HTLV-1 et de HIV-1.	97
II.2.2. Test de l'activité enzymatique des IN de HTLV-1 et de HIV-1 purifiées.	98
II.2.3. Détermination de la taille réelle des fragments amplifiés.	100
II.3. La structure de l'ADN cible influence le rendement de la réaction de transfert de brin catalysée par l'IN de HTLV-1 ou l'IN de HIV-1.	102
II.3.1. Influence de la structure de l'ADN cible sur l'intégration de HTLV-1 <i>in vitro</i>	102
II.3.2. L'IN de HTLV-1 est plus spécifique que l'IN de HIV-1.	106
II.3.3. Topographie précise de l'intégration.	111
II.4. Influence des mutations survenues <i>in vivo</i> sur la réaction de transfert de brin.	111
II.4.1. Les séquences RU5 consensus ont un profil d'intégration identique.	113
II.4.2. Des mutations somatiques de l'extrémité 3' du LTR peuvent modifier sensiblement le rendement de la réaction de transfert de brin.	115
II.5. Discussion.	122
CONCLUSION-PERSPECTIVES	128
MATERIEL ET METHODES	133
I. Matériels.	134
I.1. Les échantillons étudiés.	134
I.2. Les lignées cellulaires.	134
I.3. Le phage M13mp18 et la souche bactérienne XL1.	134
I.4. Les vecteurs d'expression des intégrases de HTLV-1 et de HIV-1.	135
II. Méthodes.	136
II.1. Culture cellulaire.	136
II.2. Extraction de l'ADN.	136
II.3. Amplification enzymatique des sites d'intégration flanquant le LTR 3' du provirus HTLV-1.	137
II.3.1. PCR avec ligation d'un oligonucléotide ou LMPCR.	137
II.3.2. PCR Inverse ou IPCR.	139
II.3.3. PCR avec extension d'amorce ou <i>Run off</i>	141

II.3.4. PCR "classiques".	142
II.4. Clonage des produits de PCR et séquençage des sites d'intégration.	142
II.4.1. Production des phages recombinants.	142
II.4.2. Préparation des ampligènes.	142
II.4.3. Préparation du vecteur M13mp18.	143
II.4.4. Ligation des produits de PCR.	143
II.4.5. Electroporation des bactéries.	144
II.4.6. Sélection des phages recombinants.	145
II.4.7. Préparation des matrices M13 pour le séquençage.	145
II.5. Construction du témoin utilisé pour l'analyse de la composition nucléotidique de l'ADN non infecté.	146
II.5.1. Amplification d'un fragment du plasmide p4.39.	146
II.5.2. Préparation de l'ADN génomique non infecté.	146
II.5.3. Ligation.	147
II.6. Expression et purification des intégrases de HTLV-1 et de HIV-1.	147
II.6.1. Transformation des bactéries.	147
II.6.2. Expression de la protéine.	148
II.6.3. Purification de la protéine en conditions dénaturantes.	148
II.6.4. Electrophorèse des protéines en gel de polyacrylamide.	149
II.6.5. Renaturation.	149
II.6.6. Test de l'activité enzymatique de la protéine purifiée.	150
II.7. Réactions d'intégration <i>in vitro</i> .	150
II.7.1. Hybridation des oligonucléotides cibles et des substrats donneurs.	150
II.7.2. Réactions de transfert de brin.	151
II.7.3. Amplification des produits d'intégration.	151
II.7.4. Résolution des ampligènes sur séquenceur d'ADN ABI Prism 310.	151
II.7.5. Clonage des produits d'intégration.	151
AUTRES TRAVAUX.	153
REFERENCES BIBLIOGRAPHIQUES.	162

INDEX DES FIGURES ET DES TABLEAUX

GENERALITES

Figure 1G :	Structure de la particule virale des rétrovirus.	19
Figure 2G :	Organisation génomique de HTLV-1.	21
Figure 3G :	Rôle de la protéine Tax dans la régulation du cycle cellulaire.	35
Figure 4G :	Réactions d'intégration <i>in vivo</i> et d'autointégration.	45
Figure 5G :	Réactions de clivage, de transfert de brin et de désintégration <i>in vitro</i> .	46
Figure 6G :	Comparaison des séquences protéiques de quelques IN rétrovirales.	51
Figure 7G :	Domaines protéiques de l'IN de HIV-1.	52
Tableau 1G :	Principaux gènes activés et réprimés par Tax.	28
Tableau 2G :	Mécanismes moléculaires de l'effet activateur ou inhibiteur de Tax.	27

RESULTATS

Figure 1R :	Amplification enzymatique des sites d'intégration de HTLV-1.	72
Figure 2R :	Construction du témoin et amplification des séquences contrôles.	74
Figure 3R :	Distribution du pourcentage de nucléotides A/T des sites d'intégration de HTLV-1 et des sites témoins.	78
Figure 4R :	Analyse par "Box Plot" du pourcentage de nucléotides A/T.	79
Figure 5R :	Distribution du pourcentage de chaque nucléotide A, G, C ou T des sites d'intégration de HTLV-1.	81
Figure 6R :	Alignement du clone 224 avec le gène de l' α -énolase.	89
Figure 7R :	Facteurs influençant l'intégration de HTLV-1 <i>in vivo</i> .	92
Figure 8R :	Schéma des réactions d'intégration <i>in vitro</i> .	95
Figure 9R :	Analyse des séquences Courbe et Rigide par le programme Bend-it.	96
Figure 10R :	Purification de l'IN de HTLV-1.	98
Figure 11R :	Détermination de l'activité enzymatique optimale de l'IN de HTLV-1.	99
Figure 12R :	Vérification de l'assignation de taille calculée par le logiciel Genescan.	102
Figure 13R :	Réactions d'intégration catalysées par l'IN de HTLV-1, entre le LTR sauvage de HTLV-1 et les oligonucléotides Courbe et Rigide.	103
Figure 14R :	Intensités de fluorescence obtenues à chaque position de la séquence Courbe avec les LTR sauvages HIV-1 et HTLV-1, pour les deux IN.	105

Figure 15R :	Réactions d'intégration catalysées par l'IN de HTLV-1, entre le LTR sauvage de HTLV-1 ou de HIV-1, et l'oligonucléotide Courbe.	107
Figure 16R :	Rendements d'intégration obtenus avec les IN de HTLV-1 et de HIV-1, en présence des LTR sauvages de HTLV-1 ou de HIV-1.	108
Figure 17R :	Intensités de fluorescence obtenues à chaque position de la séquence Courbe avec les LTR sauvages HIV-1 et HTLV-1, pour les deux IN.	109
Figure 18R :	Distribution du nombre de sites d'intégration choisis par les IN de HTLV-1 et de HIV-1, sur la séquence Courbe.	110
Figure 19R :	Sites d'intégration préférentiels sur la séquence Courbe.	111
Figure 20R :	Séquences consensus et séquences mutées.	112
Figure 21R :	Rendements d'intégration obtenus avec les séquences consensus et l'IN de HTLV-1.	113
Figure 22R :	Rendements d'intégration obtenus avec les séquences consensus et l'IN de HIV-1.	114
Figure 23R :	Réactions d'intégration catalysées par l'IN de HTLV-1 entre l'oligonucléotide Courbe et les LTR SUB1 à SUB8.	116
Figure 24R :	Intensités de fluorescence obtenues à chaque position sur la séquence Courbe, avec les LTR SUB1 à SUB8.	117
Figure 25R :	Rendement d'intégration obtenu avec le LTR SUB1.	118
Figure 26R :	Rendements d'intégration obtenus avec les LTR SUB6, SUB7 et SUB8.	119
Figure 27R :	Rendements d'intégration obtenus avec les LTR SUB2, SUB3, SUB4 et SUB5.	120
Figure 28R :	Amplifications enzymatiques des LTR SUB1, SUB3 et SUB5.	121
Figure 29R :	Indices de courbure des sites d'intégrations préférentiels	124
Tableau 1R :	Répartition du nombre d'individus et du nombre de sites d'intégration en fonction du statut clinique.	73
Tableau 2R :	Compositions nucléotidiques des sites d'intégration de HTLV-1 et des sites témoins.	76
Tableau 3R :	Distribution du pourcentage de séquences HTLV-1 et témoins ayant une identité significative avec des séquences connues.	83
Tableau 4R :	Séquences HTLV-1 et séquences témoins présentant une identité significative avec des séquences non répétitives.	86
Tableau 5R :	Séquences nucléotidiques des produits d'intégration.	101
Tableau 6R :	Comparaison des rendements d'intégration obtenus avec les oligonucléotides Courbe et Rigide, pour l'IN de HIV-1.	104
Tableau 7R :	Comparaison des rendements d'intégration obtenus avec les oligonucléotides Courbe et Rigide, pour l'IN de HTLV-1.	104

MATERIEL ET METHODES

Figure 1M : Schéma de la région d'intérêt du vecteur d'expression pET22b-IN.	136
Figure 2M : Amplification des sites d'intégration de HTLV-1 par LMPCR.	138
Figure 3M : Amplification des sites d'intégration de HTLV-1 par IPCR.	140
Tableau 1M : Séquences et positions des amorces utilisées en LMPCR, IPCR, et pour la construction du témoin.	142
Tableau 2M : Séquences des oligonucléotides utilisés pour les réactions d'intégration <i>in vitro</i> .	152

LEXIQUE

Virus

ALV : Avian Leukosis Virus.

AMV : Avian Myeloblastosis Virus.

ASV : Avian Sarcoma Virus.

BLV : Virus de la Leucémie Bovine (Bovine Leukemia Virus).

FIV : Feline Immunodeficiency Virus.

HFV : Human Foamy Virus.

HIV-1 : Human Immunodeficiency Virus Type 1.

HIV-2 : Human Immunodeficiency Virus Type 2.

HTLV-1 : Virus de la Leucémie Humaine T de Type 1 (Human T-Cell Leukemia Virus Type 1).

HTLV-2 : Virus de la Leucémie Humaine T de Type 2 (Human T-Cell Leukemia Virus Type 2).

MMLV : Moloney Murine Leukemia Virus.

RSV : Rous Sarcoma Virus.

SIV : Simian Immunodeficiency Virus.

SNV : Spleen Necrosis Virus.

STLV-1/STLV-2 : Virus de la Leucémie T du Singe (Simian T-Cell Leukemia Virus Type 1/2).

Divers

ADNv : ADN viral.

AMPc : Adénosine Monophosphate Cyclique.

ATLL : Leucémie/lymphome T de l'adulte (Adult T-Cell Leukemia/Lymphoma).

ATP : Adénosine Triphosphate.

BHE : Barrière Hématoencéphalique.

bHLH : basic Helix-Loop-Helix.

CAP : cAMP Activation Protein.

CBP/p300 : p300/CREB Binding Protein.

CDK : Cyclin-Dependent Kinase.

CREB/ATF : cAMP Responsive Element Binding / Activating Transactivation Factor.

C-terminal : Carboxy-terminal.

CTL : Cellules T cytotoxiques.

HMG : High Mobility Group Proteins.

IFN : Interféron.

IHF : Integration Host Factor.

IN : Intégrase.
 INI-1 : Integrase Interactor 1.
 IPCR : PCR Inverse.
 LCR : Liquide Céphalo-Rachidien.
 LINE : Long Interspersed Nuclear Elements.
 LMPCR : PCR avec ligation d'oligonucléotides (Ligation Mediated Polymerase Chain Reaction).
 LTR : Long Terminal Repeat.
 MA : Protéine de la matrice.
 MR : Maladie résiduelle.
 NES : Nuclear Exportation Signal.
 NLS : Signal de Localisation Nucléaire (Nuclear Localisation Signal).
 N-terminal : Amino-terminal.
 ORF : Open Reading Frame.
 PBMC : Cellules mononucléées du sang périphérique.
 PCR : Amplification par Chaîne de Polymérase.
 PIC : Complexe de Préintégration (Preintégration Complex).
 Rb : Retinoblastoma tumor suppressor protein.
 RC : Rémission complète.
 RMN : Résonance Magnétique Nucléaire.
 RRE : Rex Response Element.
 RT : Transcriptase Inverse.
 SINE : Short Interspersed Nuclear Elements.
 SNC : Système Nerveux Central.
 SRF : Serum Response Factor.
 TCR : T-Cell Receptor.
 TSP/HAM : Paraparésie Spastique Tropicale/Myélopathie associée à HTLV-1 (Tropical Spastic Paraparesis/HTLV-1-Associated Myelopathy).
 UV : Ultraviolets.
 VCAM-1 : Vascular Cell Adhesion Molecule-1.
 Vpr : Protéine virale R.

Réactifs

BME : β -mercaptoéthanol.
 BSA : Serum albumine bovine.
 dNTP : Désoxynucléotides Triphosphates.
 DTT : Dithiothréitol.
 EDTA : Acide Ethylènediamine Tétra Acétique.

IPTG : Isopropyl- β -D-Thio-Galactopyranoside.
LB : Milieu Luria-Bertani.
NP40 : Nonylphénoxy polyéthoxy éthanol.
SDS : Dodécyl Sulfate de Sodium.
TCA : Acide TrichloroAcétique.
TRIS : Tris (hydroxyméthyl)-aminométhane.
X-GAL : 5-Bromo-4-Chloro-3-Indolyl- β -D-Galactopyranoside.

Unités

Ω : ohms.
 μ l : microlitre.
 μ M : micromolaire.
Ci : curie.
DO : densité optique.
F : Farad.
g : accélération de la pesanteur.
hr : heure.
kDa : Kilodalton.
M : molaire.
mg : milligramme.
mM : millimolaire.
mn : minute.
ng : nanogramme.
nm : nanomètre
nM : nanomolaire.
pb : paire de bases.
V : Volt.

COMPOSITIONS DES TAMPONS

LB :	Bacto-tryptone 10g/l ; extraits de levure 5g/l ; NaCl 10g/l ; pH 7.
2XYT :	Bacto-tryptone 16g/l ; extraits de levure 10g/l ; NaCl 5g/l ; pH 7.
TE1X :	Tris-HCl 10 mM (pH 7,7) ; EDTA 1 mM.
TBE1X :	Tris 0,089 M ; Acide Borique 0,089 M ; EDTA 0,02 M (pH 8).
Tampon A :	Gu-HCl 6M ; NaCl 0,5 M ; Tris 20 mM pH 7,9 ; EDTA 0,1 mM ; DTT 1 mM.
Tampon B :	NaCl 1M ; Hepes 40 mM (pH 7,4) ; EDTA 2 mM ; BME 10 mM ; NP40 0,4% ; Glycérol 10%.
Tampon C :	Hepes 40 mM (pH 7,4) ; NaCl 200 mM ; EDTA 0,1mM ; DTT 1 mM ; Glycérol 10% ; NP40 0,4%.
Tampon Intégration :	Hepes 25 mM (pH 7,4) ; BME 10 mM ; Glycérol 10% ; MnCl ₂ 7,5 mM.
Denhardt 100X :	FICOLL 400 10 g/l ; PVP 10 g/l ; BSA 10 g/l.
Top Agar :	LB 0,8% agarose.
Tampon de lyse :	Tris-HCl (pH 8) 10 mM ; EDTA 10 mM ; NaCl 10 mM ; SDS 0,5% ; protéinase K 100 µg/mL.
Tampon Hybridation:	50 mM Phosphate de sodium (pH 7,2) ; 750 mM NaCl ; 5 mM EDTA ; Denhardt's 5X ; Formamide 50% ; ADN de sperme de saumon dénaturé 20 µg/mL.

Les résultats exposés dans cette thèse ont été présentés dans les publications et les communications suivantes :

PUBLICATIONS.

Cavrois M, Leclercq I, Gout O, Gessain A, Wain-Hobson S, Wattel E. Persistent oligoclonal expansion of Human T-cell leukemia virus type 1-infected circulating cells in patients with Tropical spastic paraparesis/HTLV-1 associated myelopathy. *Oncogene*, 1998, 17, 77-82.

Cavrois M, Leclercq I, Mortreux F, Wattel E. Oncogenèse liée à l'infection par HTLV-1 : données récentes. *Hématoscopie*, 1997, 10, 12-18.

Leclercq I, Cavrois M, Mortreux F, Hermine O, Gessain A, Morschhauser F, Wattel E. Oligoclonal proliferation of HTLV-1 bearing T cells in ATLL without deletion of the 3' provirus integration sites. *British journal of haematology*, 1998, 101, 500-506.

Leclercq I, Mortreux F, Morschhauser F, Duthilleul P, Desgranges C, Gessain A, Cavrois M, Vernant JP, Hermine O, Wattel E. Semiquantitative analysis of residual disease in patients treated for adult T-cell leukemia/lymphoma (ATLL). *British journal of haematology* 1999, 105, 743-751.

Leclercq I, Jonsson C.B, Wattel E. Influence de la structure de l'ADN sur le choix de l'intégrase de HTLV-1. *Bulletin du Cancer*, 1999, 86(5), 433.

Leclercq I, Mortreux F, Cavrois M, Leroy A, Gessain A, Wain-Hobson S, Wattel E. Host sequences flanking the HTLV-1 provirus *in vivo*. *Journal of Virology*, 2000, 74, 2305-12.

Leclercq I, Jonsson CB, Wattel E. Naturally occurring substitutions of the Human T-cell Leukemia Virus Type 1 3' LTR influence the strand transfer reaction. *Manuscrit en préparation*.

Leclercq I, Mortreux F, Gabet AS, Jonsson CB, Wattel E. The basis of HTLV-1 target site selection. *AIDS Research and Human Retrovirology*. *Sous Presse*.

Mortreux F, Leclercq I, Gabet AS, Leroy A, Westhof E, Gessain A, Wain-Hobson, Wattel E. Somatic mutation of the HTLV-1 provirus and flanking cellular sequences during clonal expansion. *Soumis*.

Gabet AS, Mortreux F, Leclercq I, Leroy A, Talarmin A, Gessain A, Wattel E. High circulating proviral load with oligoclonal expansion of HTLV-1 bearing T cells in HTLV-1 carriers with strongyloidiasis. *Soumis*.

COMMUNICATIONS ORALES ET POSTERS.

Leclercq I., Cavrois M., & Wattel E. HTLV-1 integration *in vivo* clustered in A/T rich regions with a non homogeneous distribution of A/T nucleotides and dinucleotides. *Colloque ULB-Lille (Bruxelles). 1997.* Communication orale.

Leclercq I, Cavrois M, & Wattel E. HTLV-1 integration *in vivo* is clustered in A/T rich regions with a non homogeneous distribution of A/T nucleotides and dinucleotides. *Eight International conference on human retrovirology : HTLV, Rio de Janero, 1997.* Communication orale.

Leclercq I, Cavrois M, & Wattel E. Transcriptionally active genome regions are not preferred targets for HTLV-1 integration *in vivo*. *Eight International conference on human retrovirology : HTLV, Rio de Janero, 1997.* Communication orale.

Leclercq I, Cavrois M, Wattel E. High oligoclonal proviral load in adult t-cell leukemia without 3' deletion of HTLV-1 integration sites. *Second European Conference on Experimental AIDS research, Stockolm, 1997.* Communication orale.

Cavrois M, Leclercq I, Gout O, Gessain A, Wain-Hobson S, & Wattel E. Persistent clonal expansion of infected circulating cells in patients with Tropical spastic paraparesis/HTLV-associated myelopathy (TSP/HAM). *Eight International conference on human retrovirology : HTLV, Rio de Janero, 1997.* Communication orale.

Leclercq I, Jonsson Colleen B, & Wattel E. The influence of DNA structure on integration events directed by HTLV-1 integrase. *Nine International conference on human retrovirology: HTLV, Kagoshima, 1999.* Communication orale.

Leclercq I, Jonsson Colleen B, & Wattel E. The influence of DNA structure on integration events directed by HTLV-1 integrase. *European Tumor Virus Group. 1999.* Communication orale.

Leclercq I, Jonsson C.B, & Wattel E. Influence de la structure de l'ADN sur le choix de l'intégrase de HTLV-1. *Société Française du Cancer. 1999.* Communication orale.

Leclercq I, Jonsson Colleen B, & Wattel E. The influence of DNA structure on integration events directed by HTLV-1 integrase. *Nine International conference on human retrovirology: HTLV, Kagoshima, 1999.* Communication orale.

Leclercq I, Cavrois M, & Wattel E. HTLV-1 integration *in vivo* clustered in A/T rich regions a non homogeneous distribution of A/T nucleotides and dinucleotides. *American Society of Hematology - XIII Annual Meeting. 1996.*

Leclercq I, Cavrois M, & Wattel E. HTLV-1 integration *in vivo* clustered in A/T rich regions a non homogeneous distribution of A/T nucleotides and dinucleotides. *Cold Spring Harbor Laboratory - Retroviruses 1997*.

Cavrois M, Leclercq I, Gessain A, & Wattel E. (1997). Persistent clonal expansion of infected circulating cells in patients with Tropical spastic paraparesis/HTLV-1 associated myelopathy (TSP/HAM). *Cold Spring Harbor Laboratory - Retroviruses. 1997*.

Leclercq I, Cavrois M & Wattel E. (1997). High oligoclonal proviral load in Adult T-cell Leukemia without 3' deletion of HTLV-1 integration sites. *Cold Spring Harbor Laboratory - Retroviruses.1997*

Leclercq I, Jonsson Colleen B, & Wattel E. The influence of DNA structure on integration events directed by HTLV-1 integrase. *Cold Spring Harbor Laboratory. Retroviruses.1999*.

Leclercq I, Mortreux F, Morschauser F, Duthilleul P, Desgranges C, Gessain A, Cavrois M, Vernant JP, Hermine O, Wattel E. Semiquantitative analysis of residual disease in patients treated for adult T-cell leukemia/lymphoma (ATLL). *European Tumor Virus Group. 1999*.

Leclercq I, Jonsson C.B, & Wattel E. The influence of DNA structure on integration events directed by HTLV-1 integrase. *Nine International conference on human retrovirology: HTLV, Kagoshima. 1999*.

Leclercq I, Mortreux F, Morschhauser F, Duthilleul P, Desgranges C, Gessain A, Cavrois M, Vernant JP, Hermine O, Wattel E. Semiquantitative analysis of residual disease in patients treated for adult T-cell leukemia/lymphoma (ATLL). *Nine International conference on human retrovirology : HTLV, Kagoshima.1999*.

INTRODUCTION

Au début des années 80, une équipe américaine et une équipe japonaise découvraient le rétrovirus HTLV-1 (pour *Human T-Cell Leukemia Virus Type 1*) à partir de prélèvements tumoraux d'un patient atteint d'un lymphome cutané de type T (Poiesz *et al.*, 1980 ; Yoshida *et al.*, 1982). Plus tard, une étude épidémiologique retenait le virus HTLV-1 comme l'agent de la paraparésie spastique tropicale (TSP/HAM pour *Tropical Spastic Paraparesis/HTLV-1-Associated Myelopathy*). HTLV-1 fut également mis en cause dans d'autres maladies inflammatoires, comme l'arthropathie inflammatoire chronique ou certaines uvéites.

Ce rétrovirus humain pathogène infecte entre 15 et 25 millions d'individus dans le monde. La majorité des sujets infectés ne développe aucune affection mais ces porteurs asymptomatiques peuvent néanmoins transmettre le virus. La transmission s'effectue par voie sanguine, par voie sexuelle et essentiellement de la mère à l'enfant, via l'allaitement. Les principales zones d'endémie sont le Japon, la Mélanésie, les Caraïbes, l'Amérique du Sud et l'Afrique Intertropicale.

HTLV-1 est un rétrovirus de la famille des oncornavirus. Son cycle répliatif passe par une étape de rétrotranscription permettant la synthèse d'un ADN double brin, qui précède l'intégration de l'ADN proviral néosynthétisé dans le génome de la cellule hôte. Le provirus HTLV-1 ne s'intègre pas en un site précis de l'ADN cellulaire (Seiki *et al.*, 1984) et les mécanismes d'insertion/promotion ne peuvent être retenus dans le développement des affections, comme dans l'infection par le virus de Moloney. Le virus ne comporte pas d'oncogène cellulaire (Seiki *et al.*, 1983) et les mécanismes précis de la pathogénie liée à l'infection sont encore mal compris. La protéine virale Tax y aurait un rôle essentiel et serait directement impliquée dans le pouvoir transformant du virus.

L'intégration de l'ADN proviral dans le génome de la cellule hôte est une étape fondamentale du cycle répliatif des rétrovirus. Elle permet l'expression des gènes proviraux et la production des particules virales infectieuses. La réaction d'intégration débute par l'excision de deux nucléotides à chaque extrémité du provirus (étape de clivage) et se poursuit par l'insertion de l'ADN proviral clivé dans le génome cellulaire hôte (étape de transfert de brin).

Tout le génome hôte peut être le siège d'une intégration (Withers-Ward *et al.*, 1994). Cependant, certaines régions sont plus souvent concernées que d'autres. La structure de la

chromatine, les interactions ADN-protéines et la composition nucléotidique de l'ADN cible peuvent notamment influencer la topographie de l'intégration (Rohdewohld *et al.* 1987 ; Scherdin *et al.*, 1990 ; Pryciak *et al.*, 1992a,b,c). Ainsi, les rétrovirus s'intègrent préférentiellement dans des zones de transcription active du génome, au niveau de régions riches en nucléotides A/T. L'enroulement de l'ADN autour des nucléosomes favorise également l'intégration au niveau de certains sites (Pruss *et al.*, 1994a,b).

Les études réalisées jusqu'à présent sur le mécanisme d'intégration des rétrovirus ont pour la plupart été menées *in vitro*. L'intégrase de HTLV-1 a été caractérisée récemment (Muller *et al.*, 1999). Mais les caractéristiques nucléotidiques et structurales des séquences ciblées par l'intégrase du virus sont encore mal connues.

Ce travail a porté principalement sur l'étude de l'intégration de HTLV-1 *in vivo* et *in vitro*. L'analyse de la composition nucléotidique d'un grand nombre de sites d'intégration du virus, isolés à partir d'individus naturellement infectés, nous a permis de montrer que l'intégration de HTLV-1 ne s'effectuait pas totalement au hasard sur le plan nucléotidique. Le virus s'intègre dans des régions riches en nucléotides A/T avec une distribution nucléotidique hétérogène au voisinage de l'extrémité provirale.

Certaines caractéristiques des séquences isolées *in vivo* suggéraient que la composition nucléotidique conférait une structure particulière à la cible de l'intégrase. Pour cette raison, nous avons dans une deuxième étape, évalué l'influence du caractère rigide ou courbe de la cible sur l'intégration, au cours de réactions d'intégration effectuées *in vitro*, à l'aide d'intégrases purifiées, d'extrémités 3' provirales et de cibles oligonucléotidiques synthétiques. L'influence de la composition nucléotidique de l'extrémité du LTR sur le rendement de la réaction de transfert de brin a été relativement peu étudié *in vitro* et est inconnue *in vivo*. Ayant isolé dans une étude ultérieure des mutations au niveau du LTR 3' chez des individus naturellement infectés, nous avons évalué l'impact de ces substitutions sur la réaction de transfert de brin.

La deuxième partie de ce travail a porté sur différents aspects de la réplication de HTLV-1, basée sur l'expansion clonale des cellules infectées.

GENERALITES

Le rétrovirus HTLV-1 appartient, comme les virus HTLV-2 (*pour Human T-Cell Leukemia Virus Type 2*), BLV (*pour Bovine Leukemia Virus*) et STLV-1 et 2 (*pour Simian T-Cell Leukemia Virus Type 1 and 2*), à la famille des oncornavirus de Type C.

Le virion possède deux molécules d'ARN génomique simple brin, étroitement associées à des enzymes virales, la transcriptase inverse (ou RT pour *Reverse Transcriptase*) et l'intégrase (IN pour *Integrase*), et entourées par une capsid virale. L'ensemble est protégé par une enveloppe d'origine cellulaire et virale (Figure 1G). Dans un premier temps, la particule virale se lie à la surface de la cellule hôte par l'intermédiaire de la glycoprotéine d'enveloppe et d'un récepteur membranaire spécifique. La fusion des membranes cellulaire et virale permet la translocation de la capsid virale dans le cytoplasme. L'ARN génomique viral est rétrotranscrit à l'intérieur de la capsid, en ADN linéaire double brin, grâce à la RT. L'ADN synthétisé est transporté dans le noyau, et s'intègre dans l'ADN cellulaire. L'insertion de l'ADN proviral dans le génome de la cellule hôte, catalysée par l'IN, permet l'expression des gènes viraux et la production de nouvelles particules virales infectieuses. Toutefois, *in vivo*, l'intégration de HTLV-1 est suivie de l'expansion clonale de la cellule porteuse du provirus intégré, et la production de particules virales est extrêmement faible.

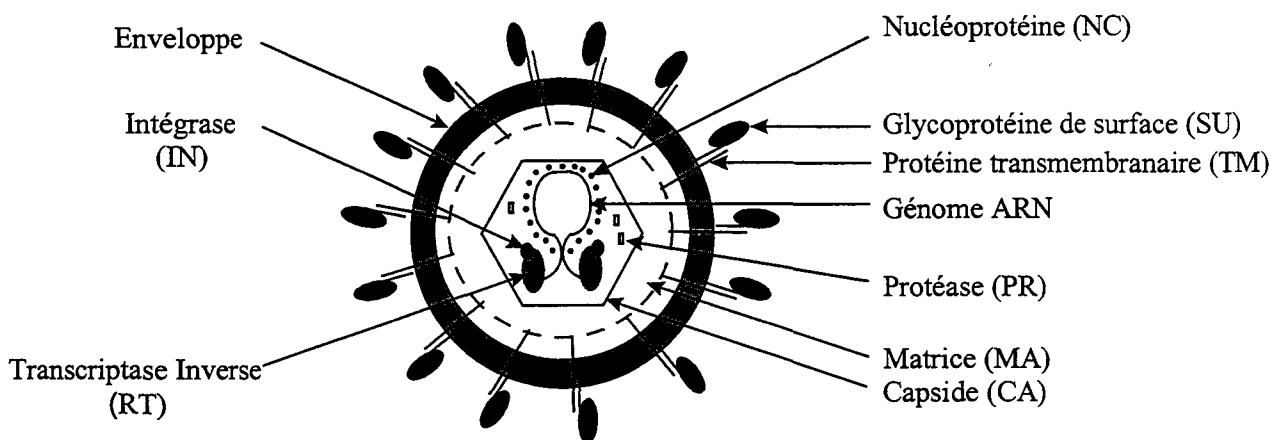


Figure 1G : Structure de la particule virale des rétrovirus (d'après Coffin, 1996).

L'intégration est donc une des étapes fondamentales du cycle réplcatif des rétrovirus et a fait l'objet de cette étude. Nous nous sommes intéressés plus particulièrement à l'étude des mécanismes moléculaires de l'intégration de HTLV-1. La première partie de ce chapitre présente le rétrovirus HTLV-1 et quelques-unes des données les plus importantes concernant sa pathogénie et son cycle réplcatif. Les données actuelles concernant l'intégration des rétrovirus seront détaillées dans la deuxième partie de ce chapitre.

I. LE RETROVIRUS HTLV-1.

Ce rétrovirus a été découvert en 1980, à partir d'une biopsie cutanée prélevée sur un patient atteint d'un lymphome cutané de type T (Poiesz *et al.*, 1980). Il infecte entre 15 et 25 millions d'individus dans le monde. Les principales zones d'endémie sont le Japon, la Mélanésie, les Caraïbes, l'Amérique du Sud et l'Afrique Intertropicale. Quelques cas ont été recensés en Italie, en Israël, en Nouvelle-Guinée et aux Etats-Unis. La transmission du virus s'effectue par voie sanguine, par voie sexuelle (préférentiellement dans le sens homme-femme) et essentiellement de la mère à l'enfant, via l'allaitement.

HTLV-1 est responsable de deux pathologies distinctes, la Paraparésie Spastique Tropicale ou TSP (Gessain *et al.*, 1985), et la Leucémie/Lymphome T de l'Adulte ou ATLL (Poiesz *et al.*, 1980 ; Yoshida *et al.*, 1982). En outre, HTLV-1 a été mis en cause dans d'autres maladies inflammatoires : polymyosites, dermatite infectieuse, syndrome de Sjögren, arthropathie, broncho-pneumopathie, uvéite (Nishioka *et al.*, 1996). La pathogénie de ces affections est incomplètement élucidée. Un élément remarquable de l'infection par HTLV-1 est que seuls 5 à 10% environ des individus porteurs du virus développent des signes cliniques. La majorité des sujets infectés reste donc asymptomatique.

I.1. Organisation génétique du rétrovirus HTLV-1.

Le génome viral d'HTLV-1 possède, comme les autres rétrovirus, les gènes *gag*, *pol* et *env*. La séquence provirale comporte en plus une région d'environ 2 kb, située dans la partie 3' de la séquence. Cette région, initialement appelée pX en raison de sa nature inconnue, code au moins 6 protéines différentes. Les protéines les mieux caractérisées sont les protéines Tax et Rex, qui jouent un rôle majeur dans la régulation du cycle réplcatif de HTLV-1. La structure génétique de HTLV-1 est représentée dans la Figure 2G.

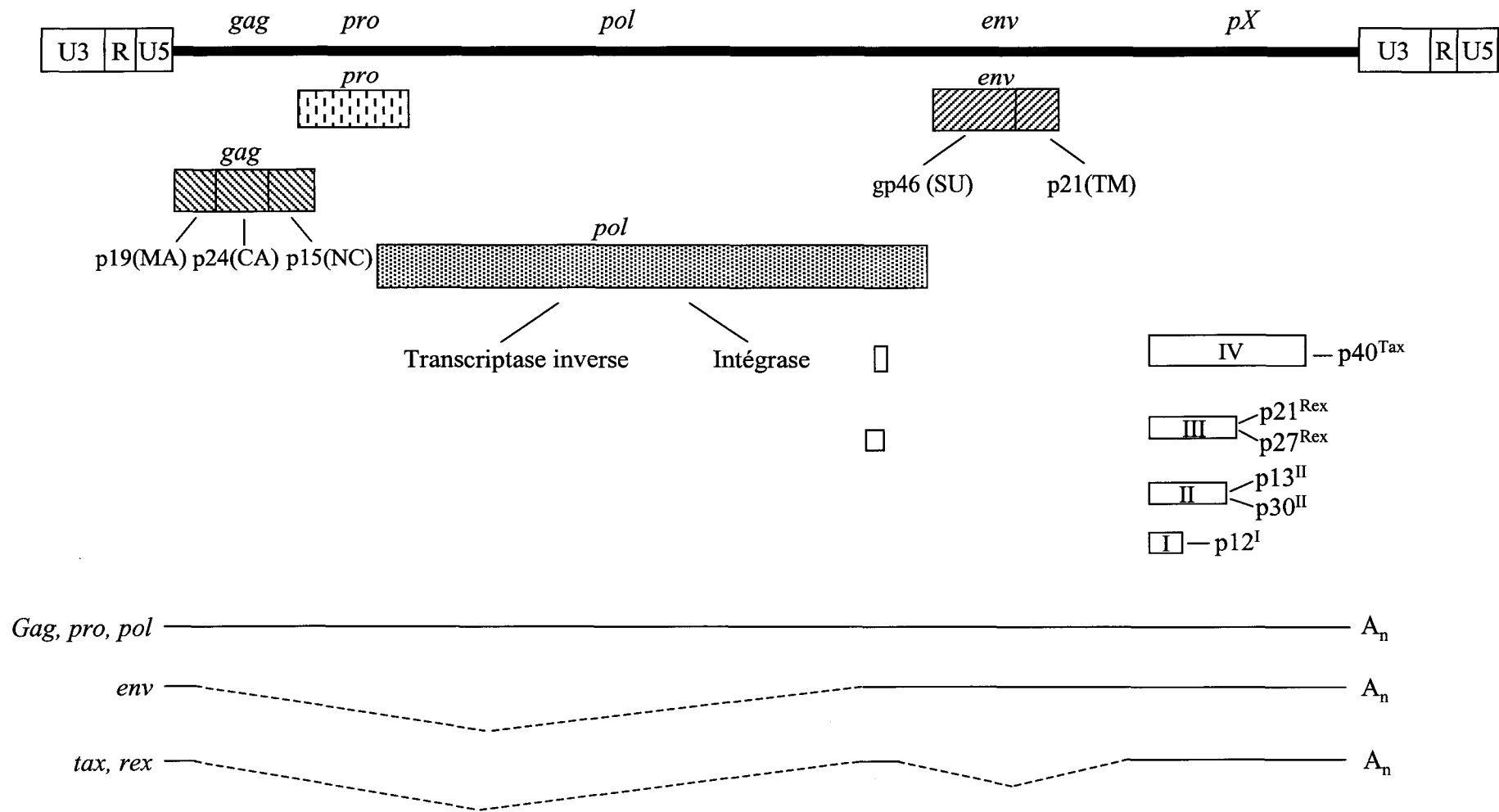


Figure 2G : Organisation génomique de HTLV-1.

I.1.1. Le provirus HTLV-1.

La structure provirale de HTLV-1 est complexe et permet de coder 11 protéines à partir d'un génome d'environ 9 kb.

Deux répétitions terminales longues (LTR pour *Long Terminal Repeat*), générées lors de la transcription inverse, encadrent la séquence provirale. Elles sont constituées des régions U3, R et U5 et comportent toutes les séquences nécessaires à l'intégration du provirus dans l'ADN génomique, et à l'expression des gènes viraux.

La région U3 du LTR 5' notamment, comporte les séquences régulatrices du promoteur viral, constituées par 3 motifs répétés de 21 pb contenant l'élément de réponse des protéines de la famille CREB/ATF (pour *cAMP Responsive Element Binding/Activating Transactivation Factor*) (Zhao *et al.*, 1992 ; Wagner *et al.*, 1993). L'activation de ce promoteur par la protéine virale Tax, entraîne la synthèse d'ARN messagers doublement, simplement ou non épissés (Sodroski *et al.*, 1984 ; Felber *et al.*, 1985). L'utilisation de ces matrices ARN et de différentes phases ouvertes de lecture permet l'expression de l'ensemble des protéines virales. Les protéines structurales et les enzymes virales, comme la RT, la protéase et l'IN sont codées par des ARN messagers non ou simplement épissés à partir des gènes *gag*, *pol* et *env* (Figure 2G).

I.1.2. Les protéines de structure.

La région *gag* est initialement traduite sous la forme d'un polypeptide précurseur avec un poids moléculaire apparent de 53 kDa (Pr53). Ce précurseur est ensuite clivé, lors du relargage des particules virales, pour former les 3 protéines de structure du virion : la protéine de la matrice (p19 MA), la protéine de la capside (p24 CA) et la protéine de la nucléocapside (p15 NC).

La protéine MA comporte un groupement myristique qui permet l'ancrage des polyprotéines Gag à la surface de la membrane plasmique, et qui facilite le bourgeonnement des nouvelles particules virales. Récemment, l'équipe de M.C. Dokhelar (Le Blanc *et al.*, 1999) a montré que la protéine p19 mature avait également un rôle déterminant dans les étapes précoces de l'infection par le rétrovirus HTLV-1, en permettant au virus de pénétrer dans la cellule hôte et de se répliquer. La transmission efficace du virus dépendrait de ces deux fonctions.

La protéine hydrophobe CA constitue la structure interne du virion.

La protéine NC est une petite protéine basique associée au génome ARN de la particule virale. Elle est également capable de se fixer à l'ADN en présence de zinc, et augmenterait l'activité de transcription inverse du rétrovirus HTLV-1 (Chechelnitsky *et al.*, 1998).

Le gène *env* code une glycoprotéine précurseur dont le poids moléculaire est de 61 kDa. Ce polypeptide est clivé pour donner la glycoprotéine de surface (gp46 SU) et la protéine transmembranaire (gp21 TM). Ces deux protéines interagissent entre elles par des liaisons non covalentes. Elles sont fixées à la surface des cellules infectées et des virions par l'intermédiaire de la gp21. Elles gouvernent l'entrée du virus dans les cellules cibles et la fusion des membranes virales et cellulaires, conduisant à la formation de syncytia (Pique *et al.*, 1992).

I.1.3. Les enzymes virales.

I.1.3.1. La protéase.

Chez HTLV-1, la protéase est codée par un cadre de lecture qui chevauche la partie 3' de la région *gag* et la partie 5' du gène *pol*. Cette protéine, maturée par autocatalyse, est responsable du clivage des précurseurs Gag en produits matures (Nam *et al.*, 1988).

I.1.3.2. Protéines codées par le gène *pol*.

La région *pol* code une protéine de 896 acides aminés (aa). Après clivage, la partie amino-terminale (N-terminale) de la protéine donne la RT, qui permet la synthèse d'ADN double brin linéaire à partir de la matrice ARN. La partie carboxy-terminale (C-terminale) de la protéine donne l'IN, qui permet l'intégration de l'ADN viral dans le génome de la cellule hôte. Ces deux enzymes sont des protéines clefs du cycle réplcatif des rétrovirus.

I.1.4. Protéines de régulation codées par la région pX.

La région pX, d'environ 2 kb, possède 4 cadres ouverts de lecture (ORF pour *Open Reading Frame*) qui codent 6 protéines différentes. Les protéines les mieux caractérisées à l'heure actuelle sont les protéines régulatrices Tax et Rex, traduites à partir des ARN messagers doublement épissés. Ces deux protéines ont un rôle déterminant dans la réplcation du virus.

La protéine Tax semble fortement impliquée dans la transformation des lymphocytes T infectés par le rétrovirus HTLV-1.

I.1.4.1. La protéine p40^{Tax}.

La protéine Tax est une phosphoprotéine nucléaire de 40 kDa qui transactive les gènes viraux et de nombreux gènes cellulaires. Elle inhibe également l'expression de certains gènes et possède une activité post-traductionnelle sur de nombreuses protéines cellulaires. Son action au sein de la cellule infectée lui confère un rôle déterminant dans la pathogénie de l'infection.

Rôle transactivateur de la protéine Tax.

La protéine Tax transactive les gènes viraux à partir du LTR 5' du provirus. La protéine ne se lie pas directement à l'ADN, mais interagit avec des facteurs de transcription de la famille CREB/ATF ; elle augmente et/ou stabilise la fixation des homodimères CREB et des hétérodimères CREB/ATF-1, sur les éléments de réponse CRE (pour *CREB Responsive Element*), présents au niveau des 3 motifs répétés de 21 pb du promoteur viral (Zhao *et al.*, 1992 ; Franklin *et al.*, 1993).

La protéine Gli2, appartenant à la famille d'oncogènes Gli et également appelée Gli2/THP (pour *Tax-Helping Protein*), interagirait également avec un autre élément de réponse à Tax, pour permettre la transcription à partir du promoteur viral (Tanimura *et al.*, 1998).

La formation du complexe nucléoprotéique, constitué de Tax et des protéines de la famille CREB/ATF, permet le recrutement du coactivateur CBP/p300 (pour *p300/CREB Binding Protein*), indispensable à la transactivation du LTR proviral (Kwok *et al.*, 1996). Ce complexe ternaire Tax-CREB-CBP pourrait recruter des facteurs régulateurs supplémentaires au niveau du LTR de HTLV-1, pour activer l'expression des gènes viraux (Giebler *et al.*, 1997). Une étude récente a montré que la protéine PCAF (pour *p300/CREB Binding Protein (CBP)-Associated Factor*) pouvait interagir avec Tax pour activer la transcription à partir du LTR de HTLV-1 (Jiang *et al.*, 1999). Les protéines CBP et PCAF interagissent *in vivo* pour activer la transcription des gènes, mais elles ne semblent pas coopérer pour la transactivation du LTR médiée par Tax. Et aucun effet synergique n'a été observé. Enfin, alors que l'activité

Histone-acétyltransférase (HAT) de CBP est nécessaire pour l'activation du LTR de HTLV-1, le domaine HAT de PCAF n'est pas indispensable (Jiang *et al.*, 1999).

Plusieurs coactivateurs semblent donc être impliqués dans la régulation de la transcription du LTR de HTLV-1.

En plus de son activité régulatrice de l'expression des gènes viraux, Tax transactive de nombreux gènes cellulaires impliqués dans la prolifération et la différenciation des lymphocytes T. Tax transactive les gènes de nombreuses cytokines, comme le GM-CSF (pour *Granulocyte Macrophage-Colony Stimulating Factor*), l'IL8, l'IL6...etc. Elle stimule notamment l'expression des gènes de l'IL2 et de la chaîne α du récepteur à l'IL2, et activerait ainsi la boucle autocrine/paracrine de la prolifération cellulaire T. Le Tableau 1G présente une liste non exhaustive des principaux gènes cellulaires transactivés par Tax. La régulation de l'expression de ces gènes par la protéine Tax fait intervenir de nombreux facteurs de transcription. Une des voies signalétiques utilisée par Tax est celle des facteurs de réponse au sérum (SRF pour *Serum Response Factor*). En effet, Tax interagit avec la protéine p67^{SRE} pour transactiver des gènes de réponse immédiate, comme *c-fos* ou *fra-1*, comportant un élément de réponse au sérum (SRE pour *Serum Responsive Element*) dans leurs régions promotrices (Fujii *et al.*, 1991 ; Tsuchiya *et al.*, 1993).

L'activation de l'expression des gènes cellulaires par Tax implique principalement la voie de transduction NF- κ B. Des lymphocytes T infectés par HTLV-1 ou transfectés par le gène codant la protéine Tax comportent des taux élevés de protéine NF- κ B dans le noyau, et Tax colocalise avec NF- κ B dans des régions subnucléaires contenant des transcrits, dont les promoteurs contiennent des sites NF- κ B (Bex *et al.*, 1997). Tax agirait donc comme un cofacteur transcriptionnel de NF- κ B dans le noyau des cellules infectées.

Tax agit également au niveau de la translocation nucléaire de NF- κ B. Elle est directement impliquée dans l'induction de l'activité kinase spécifique, nécessaire à la phosphorylation et donc à la dégradation des inhibiteurs I κ B, permettant la translocation nucléaire de NF- κ B (Brockman *et al.*, 1995 ; Kanno *et al.*, 1995). Les protéines MEKK1 (pour *Mitogen-activated Protein Kinase/Extracellular Signal-regulated Kinase Kinase 1*) et NIK (pour *NF- κ B*

Inducing Kinase), impliquées dans la phosphorylation des protéines IKKs (pour *IκB Kinases*) seraient les cibles directes de la protéine Tax (Geleziunas *et al.*, 1998 ; Uhlik *et al.*, 1998 ; Yin *et al.*, 1998). En modifiant le complexe IKK, par des mécanismes encore mal compris, et en le maintenant dans un état activé constant, la protéine Tax pourrait aboutir à l'activation constitutive de NF-κB.

Rôle trans-répresseur de la protéine Tax.

Plusieurs travaux ont montré que la protéine Tax possédait également des propriétés inhibitrices de certains gènes cellulaires, comme le gène *bax* dont le produit Bax accélère le processus d'apoptose (Brauweiler *et al.*, 1997), ou encore le gène *lck* impliqué dans la régulation de l'activation des lymphocytes T (Lemasson *et al.*, 1997). En 1990, l'équipe de S.H. Wilson a montré que le gène de la β-polymérase cellulaire, impliquée dans les mécanismes de réparation de l'ADN, était réprimé dans les cellules infectées par HTLV-1 (Jeang *et al.*, 1990).

Cette nouvelle fonction inhibitrice de Tax sur l'expression de certains gènes cellulaires fait intervenir des facteurs de transcription de type bHLH (pour *basic Helix-Loop-Helix*) (Uittenbogaard *et al.*, 1994). Cette activité repose sur l'interaction Tax-bHLH-CBP/p300 et les boîtes E, présentes au niveau du promoteur de certains gènes.

Localisation subnucléaire de Tax.

La localisation nucléaire de Tax pourrait être déterminante pour son activité. Des expériences d'immunocytochimie, de microscopies électronique et confocale ont permis de caractériser la localisation intracellulaire de la protéine Tax. Elle est détectée dans des petites zones d'ARN naissants, au niveau de structures nucléaires complexes, et colocalise avec des facteurs cellulaires divers, comme les facteurs d'épissage Sm et SC35, les protéines de la famille NF-κB, le coactivateur CBP/p300, ou encore la grande sous-unité de l'ARN polymérase II (Semmes *et al.*, 1996 ; Bex *et al.*, 1997 ; Bex *et al.*, 1998).

Les interactions protéine/protéine.

La protéine Tax interagit avec de nombreuses protéines cellulaires impliquées dans la régulation du cycle cellulaire et la prolifération des lymphocytes T. La coopération de Tax avec l'oncogène cellulaire c-Myc peut induire des transformations néoplasiques de cellules T dans des souris transgéniques (Benvenisty *et al.*, 1992). De même, la prolifération spontanée de lignées cellulaires hématopoïétiques peut être générée par l'interaction de Tax avec l'oncogène c-Myc (Miyazaki *et al.*, 1996).

Tax intervient également dans la régulation du cycle cellulaire en interagissant avec la protéine p16^{INK4A} (Suzuki *et al.*, 1996) ou encore les cyclines D3 et D2 (Neuveut *et al.*, 1998 ; Santiago *et al.*, 1999).

Le Tableau 2G résume les mécanismes moléculaires mis en jeu par la protéine Tax.

Tableau 2G : Mécanismes moléculaires de l'effet activateur ou inhibiteur de Tax sur l'expression du provirus et des gènes cellulaires.

Facteur de transcription	Coactivateur	Cible
<u>Transactivation :</u>		
CREB/ATF NF-κB SRF	CBP/p300	U3 du LTR 5' Promoteurs (IL2, RIL2...) Promoteurs (c-fos, fra-1..)
<u>Inhibition :</u>		
bHLH	CBP/p300	Promoteurs (p56 ^{lck} , β-polymérase, bax, myoD)

Rôle transactivateur

IL2
IL3
IL1- α
IL6
IL15
IL8
GM-CSF
TGF- β
TNF- α
IFN- γ

Chaîne α du récepteur à l'IL2

PTHrp (pour *Parathyroid Hormone-related Protein*)

vimentine

lymphotoxine

PCNA (pour *Proliferating Cell Nuclear Antigen*)

mdr-1

proenképhaline

VCAM-1 (pour *Vascular Cell Adhesion Molecule-1*)

FasL (pour *Fas Ligand*)

TR3/nur77

stat1 et stat5

c-fra 1

c-rel

c-myc

c-fos

egr-1

egr-2

Rôle répresseur

β -polymérase

lck

bax

Tableau 1G : Gènes transactivés ou réprimés par Tax (d'après Ressler *et al.*, 1996).

I.1.4.2. La protéine p27^{Rex}.

La protéine p27^{Rex} est une phosphoprotéine nucléolaire de 189 acides aminés, impliquée dans la régulation du cycle réplcatif de HTLV-1.

Elle agit au niveau post-transcriptionnel en permettant le transport des ARNm non ou simplement épissés dans le cytoplasme des cellules infectées. Elle favorise donc la synthèse des protéines de structure et des enzymes virales. En l'absence de Rex, ces ARNm sont séquestrés dans le noyau des cellules infectées où ils sont soit complètement épissés, soit dégradés.

La protéine Rex, ainsi que son homologue fonctionnel Rev du HIV-1 (pour *Human Immunodeficiency Virus Type 1*), est caractérisée par un domaine de fixation à l'ARN et un domaine d'activation riche en leucine, qui correspond à un signal d'exportation nucléaire (NES pour *Nuclear Exportation Signal*).

La fonction de Rex est médiée par un élément de réponse à Rex (RRE pour *Rex Response Element*) localisé dans la région R du LTR 3' viral et qui présente une structure hautement organisée. La fixation spécifique de Rex sur le RRE est régulée par phosphorylation de la protéine (Green *et al.*, 1992). Les fonctions NES de Rex et Rev seraient médiées par une protéine cellulaire, "nucleoporine-like", encore appelée Rab/hRIP (pour *Rex/Rev activation domain binding/human Rev Interaction Protein*), capable de se fixer sur les domaines d'activation de Rex et Rev (Bogerd *et al.*, 1995 ; Fritz *et al.*, 1995). Cependant, les mécanismes cellulaires et moléculaires précis, impliqués dans l'exportation des ARNm sont encore mal connus. Plus récemment, un autre cofacteur cellulaire, hCRM1 (pour *human exportin 1/Chromosomal Region Maintenance 1*) a également été impliqué dans la multimérisation et l'export nucléaire de la protéine Rex (Hakata *et al.*, 1998).

La protéine "navette" nucléocytoplasmique Rex agit également en amont de l'épissage, probablement en stabilisant les transcrits ARN (Inoue *et al.*, 1991). Une étude de Rex2 codée par HTLV-2 indique que cette classe de protéine pourrait également inhiber la formation du *spliceosome* sur le transcrit (Bakker *et al.*, 1996).

I.1.4.3. Régulation du cycle réplcatif.

Après l'infection des lymphocytes T par le rétrovirus HTLV-1 et l'intégration du provirus dans l'ADN génomique, les gènes viraux sont transcrits à partir du LTR 5', probablement par fixation de facteurs cellulaires sur le promoteur viral. La position des codons d'initiation de la traduction de Tax et de Rex conduit initialement à la traduction préférentielle de Tax. En absence de Rex, seuls les ARNm doublement épissés parviennent dans le cytoplasme. La transactivation des gènes viraux par Tax aboutit à la production massive d'ARNm Tax/Rex, et donc à la synthèse de plus en plus importante de la protéine Rex. Celle-ci va promouvoir le passage des ARNm simplement ou non épissés dans le cytoplasme et permettre la synthèse des protéines de structure et des enzymes virales. La traduction de ces ARNm va donc permettre l'assemblage des virions et leur départ de la cellule infectée.

Enfin, l'action inhibitrice de Rex sur les ARNm doublement épissés Tax/Rex (Hidaka *et al.*, 1988) aurait pour conséquence de diminuer l'effet transactivateur de Tax sur le LTR viral, entraînant la diminution de l'expression des gènes viraux. Cet effet négatif de Rex sur la régulation de la réplication du virus pourrait aboutir à la latence virale, et explique la faible virémie qui caractérise l'infection.

I.1.4.4. Autres protéines codées par la région pX.

La région pX d'HTLV-1 possède également deux ORF, pX-I et pX-II, qui codent les protéines putatives virales p12^I, p30^{II} et p13^{II}. Le rôle de ces protéines dans la réplication et la pathogenèse de HTLV-1 est encore très mal connu.

Des lymphocytes T cytotoxiques CD8⁺, dirigés contre des peptides pX-I et pX-II ont été détectés chez des individus infectés par HTLV-1. Ces protéines virales sont donc synthétisées *in vivo* et sont donc probablement impliquées dans la réplication du virus (Pique *et al.*, 2000).

La protéine p12^I.

La protéine p12 est une petite protéine hydrophobe associée aux endomembranes cellulaires. Elle peut interagir avec la sous-unité de 16 kDa de la pompe à proton vacuolaire H⁺-ATPase, comme la protéine E5 du papillomavirus bovin. Les homologues structurales de ces deux protéines suggèrent qu'elles fonctionnent de la même façon. La protéine p12 augmente le

pouvoir transformant de E5 sur les fibroblastes (Franchini *et al.*, 1993). Elle se fixe également sur les chaînes β et γ du récepteur à l'IL2 (Mulloy *et al.*, 1996).

Elle n'interviendrait pas dans la réplication virale et l'immortalisation des lymphocytes T *in vitro*. L'inhibition de l'expression des protéines p12^I, p13^{II} et p30^{II} par mutagenèse des ORF I ou II, n'a pas d'effet sur la prolifération de lymphocytes T primaires infectés et la réplication virale, par comparaison avec des lymphocytes T infectés par le virus sauvage (Robek *et al.*, 1998 ; Collins *et al.*, 1999).

Cependant, l'implication de la protéine p12 dans l'établissement d'une infection virale persistante a été démontrée *in vivo* et suggère que la protéine aurait un rôle dans les étapes précoces de l'infection par HTLV-1 (Collins *et al.*, 1998).

Les protéines p30^{II} et p13^{II}.

Leurs fonctions exactes sont encore inconnues.

La protéine p13 est localisée au niveau des membranes mitochondriales. La protéine p30 est une protéine nucléolaire. Elles ne sont pas indispensables à l'immortalisation des lymphocytes T *in vitro*. Cependant, une expérience récente réalisée *in vivo* sur des lapins infectés par un clone moléculaire HTLV-1 dont l'ORF II est muté, diminue la capacité de HTLV-1 à maintenir une charge virale élevée (Bartoe *et al.*, 2000).

I.2. Tropisme de HTLV-1.

Le rétrovirus HTLV-1 est peu infectieux et le taux de particules virales dans le sang est relativement faible. L'infection par HTLV-1 nécessite la plupart du temps des contacts cellulaires directs.

Le virus infecte principalement les lymphocytes T de phénotype CD4⁺, CD45RO⁺, CD8⁻. L'analyse de sous-populations cellulaires, provenant de patients atteints d'ATLL ou de TSP et de porteurs asymptomatiques, montre que le virus infecte principalement les lymphocytes de phénotype CD4⁺, et exceptionnellement de phénotype CD8⁺ (Hattori *et al.*, 1981). Cependant, des expériences réalisées *in vitro* par coculture de cellules cibles avec des cellules productrices de virus, montrent que le virus HTLV-1 est capable d'infecter de nombreux types cellulaires, en particulier des lymphocytes de phénotype CD8⁺, des macrophages, des

monocytes et des lymphocytes B. Les cellules endothéliales (Ho *et al.*, 1984), les cellules microgliales (Hoffman *et al.*, 1992), dendritiques (Macatonia *et al.*, 1992) et nerveuses sont également susceptibles d'être infectées par HTLV-1 *in vitro* (Akagi *et al.*, 1992). Le récepteur du virus, jusqu'ici encore inconnu, est donc présent à la surface de nombreux types cellulaires. Cependant, des études d'activité transcriptionnelle conduite par le LTR de HTLV-1 montrent que l'activation des gènes viraux est plus importante dans les lymphocytes T CD4⁺. Ceci implique que des facteurs cellulaires spécifiques de cette sous-population, activent la transcription du LTR de HTLV-1 et sont impliqués dans le potentiel transformant du virus (Newbound *et al.*, 1996).

I.3. L'ATLL.

I.3.1. Données cliniques.

L'ATLL est une affection de l'adulte qui se développe chez 2 à 5% des individus infectés, après une latence prolongée de 40 à 60 ans. Elle survient donc quasi exclusivement chez des individus infectés par l'allaitement maternel. Elle est caractérisée par une prolifération de lymphocytes T matures de phénotype CD4⁺ CD8⁻ CD45RO⁺, exprimant le récepteur à l'IL2. Les cellules malignes en prolifération sont caractérisées par un noyau multilobé pouvant prendre la forme d'un trèfle à 3 ou 4 feuilles.

Les manifestations cliniques de la maladie peuvent associer un syndrome tumoral ganglionnaire, des localisations cutanées, des lésions osseuses et une hypercalcémie. Le diagnostic de l'ATLL repose principalement sur la mise en évidence par blot de Southern, d'une intégration monoclonale du provirus dans l'ADN tumoral.

Plusieurs formes d'ATLL ont été décrites, en rapport avec le mode d'installation, la présentation clinique et l'évolution. On distingue les formes chroniques et progressives, dont les médianes de survie peuvent atteindre 2 ans, et les formes lymphomateuses et aiguës, beaucoup plus agressives, dont les médianes de survie ne dépassent pas 6 mois (Shimoyama *et al.*, 1991).

Aucune anomalie chromosomique spécifique n'est associée à la leucémie T de l'adulte. Cependant, des anomalies chromosomiques secondaires sont fréquentes : trisomies 3 et 7,

réarrangements du chromosome 6 ou perte du chromosome X. Elles surviennent au hasard au cours de l'affection mais leur fréquence augmente au fur et à mesure que la maladie progresse (Sanada *et al.*, 1987 ; Maruyama *et al.*, 1990).

Le traitement de l'ATLL repose sur la polychimiothérapie, l'association AZT-Interféron α (IFN α) ou sur l'allogreffe de cellules souches hématopoïétiques. Quelle que soit la stratégie thérapeutique utilisée, une rémission complète est obtenue chez environ 50% des patients. Ces rémissions sont souvent de courte durée et la plupart des patients traités rechute, de sorte que l'ATLL reste une affection de mauvais pronostic (Tobinai *et al.*, 1994). Des survies prolongées ont été décrites chez des patients allogreffés.

Quelque soit le sous-type d'ATLL, les prélèvements tumoraux sont toujours caractérisés par un profil clonal d'intégration du provirus (Yoshida *et al.*, 1984). L'amplification enzymatique des différentes régions de HTLV-1 et l'analyse par blot de Southern de l'ADN de patients atteints d'ATLL, met en évidence une délétion plus ou moins importante de la partie 5' du provirus, dans plus d'un tiers des cas (Korber *et al.*, 1991 ; Tamiya *et al.*, 1996). Ces délétions respectent toujours la région pX mais modifient par délétion du ou des sites donneurs en 5', l'épissage des ARN viraux. Le rôle de cette anomalie fréquente dans la réplication du virus et sa pathogénie est encore mal connu.

I.3.2. Effet de HTLV-1 sur la prolifération cellulaire et la transformation maligne.

Les PBMCs de patients prolifèrent spontanément *in vitro*, en l'absence de tout stimulus exogène (Itoyama *et al.*, 1988 ; Kramer *et al.*, 1989). La protéine Tax est nécessaire et suffisante à l'immortalisation des cellules (Grassmann *et al.*, 1992). Elle est directement impliquée dans le pouvoir transformant du rétrovirus HTLV-1. Elle est capable de transformer *in vitro*, des cellules en culture (Grassmann *et al.*, 1989 ; Pozzatti *et al.*, 1990 ; Tanaka *et al.*, 1990) et peut induire l'apparition de tumeurs chez des souris transgéniques (Nerenberg *et al.*, 1987).

Outre son activité transactivatrice de nombreux gènes cellulaires impliqués dans la prolifération, la protéine Tax interagit avec des protéines régulatrices du cycle cellulaire. En effet, HTLV-1 peut stimuler l'hyperphosphorylation de la protéine Rb (pour *Retinoblastoma tumor suppressor protein*) via deux mécanismes indépendants : L'interaction de Tax avec la

protéine p16^{INK4A} bloque sa fonction inhibitrice de l'activité kinase de CDK4 (pour *Cyclin-Dependent Kinase*) (Suzuki *et al.*, 1996). La protéine CDK4 peut alors exercer son activité kinase, et phosphoryler la protéine Rb. L'hyperphosphorylation de Rb dissocie le complexe E2F-Rb, permettant la translocation dans le noyau de la protéine E2F, qui peut alors activer la transcription de gènes impliqués dans la progression du cycle cellulaire. La transcription du gène *E2F-1* peut être directement activée par Tax, via la voie ATF/CREB (Lemasson *et al.*, 1998).

Tax peut également stimuler directement les activités kinases des protéines CDK4 et CDK6, en interagissant avec la cycline D3 (Neuveut *et al.*, 1998). L'interaction de Tax avec la cycline D3 induirait la phosphorylation de celle-ci, activant ainsi les complexes cycline-CDK et donc la progression du cycle cellulaire. De la même façon, le gène codant la cycline D2 est activé par la protéine Tax ; la surexpression de ce gène pourrait également stimuler de manière indirecte, la progression du cycle cellulaire (Santiago *et al.*, 1999).

Le cycle cellulaire est également contrôlé par la protéine suppresseur de tumeur p53, dont la synthèse permet l'arrêt des cellules en phase G1 ou leur entrée en apoptose. La fréquence des mutations du gène p53 est relativement faible dans les cellules de patients atteints d'ATLL (Sakashita *et al.*, 1992). Ces observations suggèrent que d'autres mécanismes seraient impliqués dans l'inhibition de l'expression ou de la fonction de p53 dans les lymphocytes T infectés. Les études montrent que la protéine p53 est surexprimée dans les lymphocytes T infectés par HTLV-1 (Yamato *et al.*, 1993 ; Akagi *et al.*, 1997).

Il a été proposé que la protéine Tax inhibait la transcription du gène p53 (Uittenbogaard *et al.*, 1995). Ces observations sont en contradiction avec les résultats de Yamato et Akagi. Par contre, la durée de vie de la protéine p53 augmente dans les lymphocytes T transformés par HTLV-1 (Reid *et al.*, 1993). De plus, des expériences de transfection et d'irradiation démontrent que la protéine p53 est inactive dans les lymphocytes T infectés. Seule la protéine Tax est capable de supprimer l'activité transcriptionnelle de la protéine p53 (Cereseto *et al.*, 1996 ; Mulloy *et al.*, 1998 ; Pise-Masison *et al.*, 1998a). Ces observations suggèrent l'hypothèse d'une inhibition fonctionnelle de p53 par Tax. Plusieurs mécanismes semblent intervenir dans cette inhibition :

Tax pourrait agir au niveau post-transcriptionnel en induisant la phosphorylation de p53 au niveau des résidus Ser15 et Ser392 (Pise-Masison *et al.*, 1998b). La phosphorylation de p53 bloque son interaction avec les facteurs de transcription et inactive ainsi la protéine. Tax pourrait également réguler la fonction de p53 de manière indirecte, par compétition pour des facteurs de transcription communs. Colgin et Nyborg (1998) ont ainsi montré que la protéine Tax pouvait contrer l'activité transcriptionnelle de c-Myb par compétition, en se fixant sur le coactivateur CBP. Un mécanisme similaire pourrait également être impliqué dans la dérégulation de l'activité transcriptionnelle de p53 par Tax.

Une étude récente réalisée par l'équipe de K.T. Jeang a montré que la protéine transactivatrice Tax pouvait également déréguler le cycle cellulaire en inhibant la fonction de la protéine MAD1 (pour *Mitotic Arrest Defective*), impliquée dans la transition G2-M (Jin *et al.*, 1998).

Les effets de Tax sur le cycle cellulaire sont résumés dans la Figure 3G.

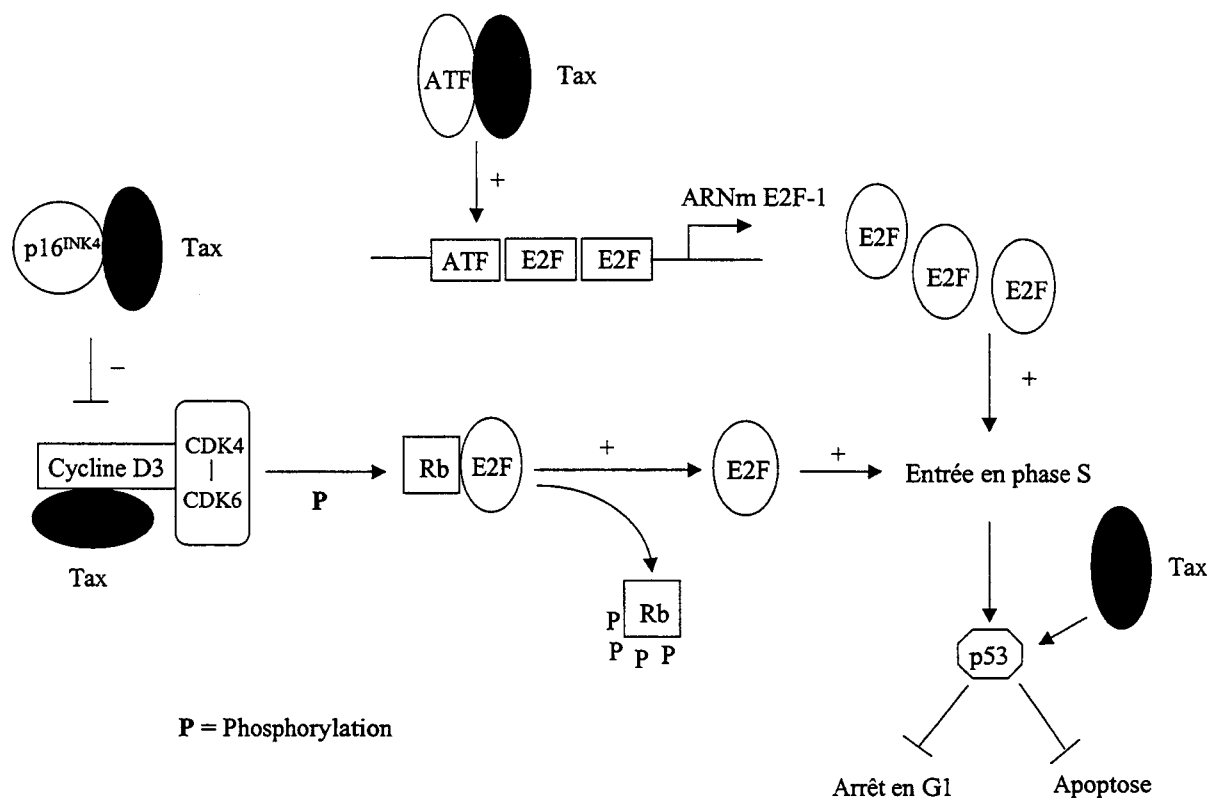


Figure 3G : Rôle de la protéine Tax dans la régulation du cycle cellulaire.

En plus de son action transactivatrice sur le cycle cellulaire, la protéine Tax inhibe certaines fonctions de réparation de l'ADN dans les cellules infectées. En effet, Tax inactive le promoteur du gène de la β -polymérase (Jeang *et al.*, 1990) et perturbe d'autres voies cellulaires impliquées dans la réparation de l'ADN (Kao *et al.*, 1999 ; Philpott *et al.*, 1999). Ces caractéristiques fonctionnelles de Tax pourrait expliquer l'effet mutagène de la protéine, récemment mis en évidence *in vitro* (Miyake *et al.*, 1999). Récemment, une étude menée au laboratoire a montré que la variabilité génétique intrapatient du virus HTLV-1 était due principalement à des mutations somatiques de la séquence provirale (Mortreux *et al.*, *soumis*).

D'autre part, les cellules T infectées par HTLV-1 en culture sont caractérisées par une phase initiale de croissance dépendante de l'IL2. Dans les phases tardives de la transformation des lymphocytes T *in vitro*, certaines lignées cellulaires infectées continuent à proliférer en absence d'IL2 (Arya *et al.*, 1984). La protéine Tax, en stimulant l'expression des gènes de l'IL2 et du récepteur à l'IL2 active dans un premier temps, la boucle autocrine/paracrine de la prolifération cellulaire T.

Elle perturbe également la transduction du signal à l'IL2 au niveau de la voie JAK-STAT, impliquée dans la transduction des signaux en réponse aux cytokines (Nakamura *et al.*, 1999). Une activation constitutive de la voie de transduction des protéines JAK-STAT est détectée dans des lymphocytes T transformés par HTLV-1 et des cellules de patients atteints d'ATLL. L'acquisition de l'indépendance à l'IL2 des cellules transformées est directement corrélée avec l'activation constitutive de cette voie de transduction (Migone *et al.*, 1995 ; Takemoto *et al.*, 1997).

I.4. La TSP/HAM.

La relation entre la TSP/HAM et le rétrovirus HTLV-1 a été établie en 1985 (Gessain *et al.*, 1985) aux Caraïbes, puis en 1986 au Japon, en Jamaïque et en Colombie (Osame *et al.*, 1986). La TSP est une affection chronique progressive du système nerveux central (SNC), rare avant 20 ans et après 70 ans. Les signes cliniques sont ceux d'une paraplégie ou d'une paraparésie spastique, avec atteinte pyramidale bilatérale. Le déficit moteur prédomine dans les territoires proximaux. L'évolution est le plus souvent invalidante, il n'existe pas à l'heure actuelle de traitement spécifique. Celui-ci repose principalement sur le traitement symptomatique de la spasticité et de ses complications.

Le diagnostic est établi à partir de critères cliniques, et sur la sérologie positive HTLV-1.

La prévalence de la TSP/HAM varie considérablement d'une région à une autre, et suggère fortement l'implication de facteurs de risque dans la survenue de la maladie. L'association entre la TSP/HAM et certains types HLA, notamment HLA-A2 (Jacobson *et al.*, 1992 ; Elovaara *et al.*, 1993 ; Koenig *et al.*, 1993 ; Kannagi *et al.*, 1994), suggère l'existence d'une prédisposition génétique dans le développement de cette affection. En effet, certains haplotypes HLA sont plus fréquents chez les patients atteints de TSP/HAM par rapport à la population générale (Sonoda *et al.*, 1996). La présence de ces haplotypes HLA-A2 diminuerait fortement le risque de développer une TSP/HAM, en permettant le contrôle de la charge virale, via la réponse T cytotoxique (Jeffery *et al.*, 1999). Toutefois, le développement de la TSP/HAM ne peut s'expliquer uniquement par des facteurs génétiques.

La recherche de mutations dans la séquence provirale, susceptibles de conduire préférentiellement à la TSP/HAM, n'a pas permis la mise en évidence de différences notables entre les différentes pathologies.

De nombreux paramètres immunologiques anormaux caractérisent la TSP et suggère une participation immunologique importante dans la genèse des lésions du SNC. En effet, le taux d'anticorps anti-HTLV-1 est significativement plus élevé dans la TSP/HAM que dans les autres conditions liées à l'infection (Dalglish *et al.*, 1988). On retrouve une distribution oligoclonale d'anticorps IgG, dirigés contre des protéines HTLV-1, dans le sang et le liquide céphalo-rachidien (LCR), ainsi qu'un taux élevé de lymphokines (IL6 et RIL2), de VCAM-1 soluble et d'IFN γ . Les marqueurs de l'inflammation comme la néoptérine et la β_2 -microglobuline sont également élevés dans le sang et le LCR.

La TSP est caractérisée par une charge provirale élevée par rapport au stade asymptomatique, et par la présence dans le sang périphérique de lymphocytes T CD4⁺ CD45RO⁺ activés, capables de proliférer spontanément *in vitro*, en l'absence de tous stimulus exogène (Itayama *et al.*, 1988 ; Kramer *et al.*, 1989). Une étude récente basée sur l'amplification en temps réel de l'ADN proviral de patients atteints de TSP ou de porteurs asymptomatiques, a ainsi montré que le pourcentage de cellules infectées était de 5% en moyenne pour les individus atteints de TSP/HAM et de 0,3% en moyenne au stade asymptomatique (Nagai *et al.*, 1998).

Enfin, la réponse T cytotoxique de phénotype CD8⁺ est très importante au cours de la TSP/HAM, par rapport au stade asymptomatique et au stade ATLL. Ces lymphocytes T cytotoxiques sont principalement dirigés contre des épitopes de la protéine Tax, en particulier le peptide 11-19, en association avec le complexe HLA-A2 (Jacobson *et al.*, 1990 ; Jacobson *et al.*, 1992 ; Parker *et al.*, 1992 ; Elovaara *et al.*, 1993 ; Koenig *et al.*, 1993).

Le nombre de cellules T cytotoxiques (CTL) présentes dans le LCR augmente au fur et à mesure que la maladie progresse, tandis que le nombre de lymphocytes T CD4⁺ et de cellules inflammatoires, comme les macrophages, se stabilise.

Le mode de transport de HTLV-1 dans le SNC a été étudié chez des patients atteints de TSP/HAM (Cavrois *et al.*, *soumis*). Les résultats montrent qu'il existe un transport cellulaire du virus dans le SNC à l'état proviral, *via* le passage de la cellule hôte au travers de la barrière hématoencéphalique (BHE). L'infection par HTLV-1 stimule l'expression des molécules d'adhésion présentes à la surface des cellules infectées (Mori *et al.*, 1994 ; Tatewaki *et al.*, 1995 ; Tanaka *et al.*, 1995, 1996). L'expression de ces protéines pourrait permettre à certains clones d'entrer en contact avec les cellules endothéliales des capillaires cérébraux pour franchir plus facilement la BHE.

L'importance des réponses immunitaires cellulaire et humorale contre HTLV-1, observée chez les patients atteints de TSP/HAM, suggère que des mécanismes immunologiques y jouent un rôle important, mais les mécanismes moléculaires et cellulaires impliqués demeurent inconnus. Trois modèles ont été proposés (Ijichi *et al.*, 1993 ; Hara *et al.*, 1994 ; Nagai *et al.*, 1996) :

Le premier modèle suggère l'existence de lymphocytes T CD4⁺ autoréactifs dont l'activation par le rétrovirus HTLV-1 leur permettrait de franchir la barrière hématoencéphalique. Les cellules et/ou composants du SNC seraient alors reconnus par les cellules autoréactives, entraînant des lésions du SNC.

Le deuxième modèle implique l'infection par HTLV-1 des cellules du SNC. L'attaque de ces cellules infectées présentant des épitopes viraux à leur surface, par le système immunitaire, entraînerait la genèse des lésions. Si le virus est capable, *in vitro*, d'infecter de nombreux

types cellulaires, l'infection de cellules nerveuses *in vivo* n'a pas encore été clairement établie et reste sujet à controverse.

Enfin, la présence de cellules infectées au sein du SNC pourrait générer des lésions par effet *bystander* sur les tissus environnants. Des interactions entre les cellules infectées et les cellules du système immunitaire entraîneraient la production de cytokines et endommageraient secondairement le tissu nerveux.

I.5. Réplication du rétrovirus HTLV-1.

I.5.1. Charge virale et stabilité génétique.

Le taux d'erreurs de la RT de HTLV-1 n'a pas été évalué. Cependant, le taux de substitutions après un cycle réplicatif de son homologue bovin, le BLV, n'est que 10 fois inférieur à celui du HIV-1 (Mansky *et al.*, 1994).

Malgré cela, l'une des caractéristiques de HTLV-1 est sa grande stabilité génétique (Gessain *et al.*, 1992). Alors que les isolats géographiques varient de 2 à 8% entre eux, la variabilité intra-individuelle de HTLV-1 est beaucoup plus faible, inférieure à 0,5% (Gessain *et al.*, 1996). Ce faible taux de mutation s'accompagne paradoxalement d'une importante charge provirale. La comparaison des séquences provirales provenant de différentes zones endémiques révèle une grande conservation de la séquence HTLV-1. Les isolats d'Afrique, du Japon, d'Amérique du Sud et des Caraïbes présentent plus de 97% d'identité. Bien que les isolats viraux soient issus de populations autochtones ayant migré depuis des milliers d'années, la divergence des variants géographiques est inférieure à 10% au niveau de la séquence nucléaire (Gessain *et al.*, 1992). Certaines régions du provirus, notamment la région pX, semblent plus variables que d'autres (Gessain *et al.*, 1993).

Une faible variabilité de la séquence provirale existe toutefois *in vivo* (Gessain *et al.*, 1992 ; Kira *et al.*, 1994 ; Niewiesk *et al.*, 1994 ; Saito *et al.*, 1995 ; Wattel *et al.*, 1995). Les mutations observées chez HTLV-1, pour la plupart issues du séquençage de segments proviraux isolés *in vivo*, bien que peu nombreuses, sont importantes dans la pathogénie de l'infection par HTLV-1, puisqu'elles peuvent modifier l'expression des gènes viraux. Ainsi,

des travaux récents ont montré que des mutations de la région pX survenant *in vivo* pouvaient d'une part altérer la reconnaissance de la protéine Tax par les CTL et d'autre part réduire son activité transactivatrice (Niewiesk *et al.*, 1995).

I.5.2. Expansion clonale des cellules infectées.

Plusieurs travaux basés sur l'analyse des sites d'intégration flanquant le LTR 3' du provirus HTLV-1, par les techniques de PCR inverse et de PCR avec ligation d'oligonucléotides (IPCR pour *Inverse Polymerase Chain Reaction* et LMPCR pour *Ligation Mediated Polymerase Chain Reaction*, détaillées dans les articles cités en références : Wattel *et al.*, 1995 ; Cavrois *et al.*, 1995 ; Cavrois *et al.*, 1996a,b ; Wattel *et al.*, 1996 ; Cavrois *et al.*, 1998 ; Leclercq *et al.*, 1998 ; Leclercq *et al.*, 1999) ont permis de caractériser le mode réplcatif particulier du rétrovirus HTLV-1. Ces études montrent qu'il existe une expansion clonale des cellules infectées chez tous les individus porteurs du virus, quelque soit le stade de l'infection (Cavrois *et al.*, 1996a,b ; Wattel *et al.*, 1996). Le virus se réplique donc via la mitose cellulaire, en même temps que les lymphocytes CD4⁺ infectés. La protéine Tax semble être l'élément clef de cette prolifération. Une prolifération continue des cellules infectées est retrouvée à tous les stades de l'infection et semble être impliquée dans la neuropathogenèse et la leucémogenèse.

Ce mode réplcatif particulier de HTLV-1 a donc permis d'expliquer l'association charge provirale élevée-stabilité génétique. C'est l'expansion clonale de la cellule hôte qui permet au virus d'atteindre une charge provirale élevée tout en restant génétiquement stable. Les techniques de LMPCR et d'IPCR permettent en outre d'apprécier semi-quantitativement le nombre de cellules au sein des différents clones cellulaires (Cavrois *et al.*, 1995). Le nombre de clones en expansion peut varier considérablement d'un individu à l'autre.

I.5.3. Mutations somatiques de la séquence provirale et instabilité génétique de la cellule hôte.

Une faible variabilité de la séquence provirale existe donc *in vivo*. Outre son action positive sur la prolifération cellulaire, la protéine Tax a un rôle inhibiteur de certains gènes impliqués dans les mécanismes de réparation de l'ADN cellulaire (Kao *et al.*, 1999 ; Philpott *et al.*, 1999), et possède un pouvoir mutagène *in vitro* (Miyake *et al.*, 1999). Le virus HTLV-1 se

réplique donc principalement à l'état proviral, via l'immortalisation de la cellule infectée, dans un contexte d'instabilité génétique généré par Tax.

Ces données suggèrent que certaines mutations de la séquence provirale acquises *in vivo* pourraient correspondre à des mutations somatiques. Récemment, une étude réalisée dans notre laboratoire, par Franck Mortreux, a montré que ces mutations étaient effectivement issues d'erreurs commises par le système réplcatif de la cellule hôte, et non par transcription inverse, comme pour les autres rétrovirus. Des mutations dans les séquences flanquantes au provirus ont également été détectées, avec une fréquence similaire (Mortreux *et al.*, *soumis*).

I.5.4. Incidences pathologiques de l'expansion clonale des cellules infectées.

La mise en évidence des clones HTLV-1 repose sur l'amplification enzymatique des sites d'intégration du virus par IPCR ou par LMPCR. Quelque soit le stade de l'infection, il existe une prolifération oligo/polyclonale des cellules T infectées (Cavrois *et al.*, 1996a,b ; Wattel *et al.*, 1996).

L'ATLL est caractérisée par la présence d'un profil monoclonal d'intégration au sein du tissu tumoral. Une précédente étude, réalisée par quadruplicate IPCR sur l'ADN de 21 patients atteints de différentes formes d'ATLL a montré qu'il existait, en plus du clone malin, une grande quantité de clones minoritaires au sein même du tissu tumoral. A titre d'exemple, 80% des clones isolés d'un ganglion correspondaient à des formes minoritaires (Cavrois *et al.*, 1996b). Le nombre et l'ampleur de ces clones sont plus importants qu'au cours de la phase asymptomatique.

L'analyse, au cours du temps, par PCR "nichée" clonotypique, de l'ADN de patients atteints de TSP/HAM a montré que la majorité des clones infectés par HTLV-1 sont stables sur une période de 11 mois à 5 ans, en absence de malignité (Cavrois *et al.*, 1998). L'expansion clonale des cellules T infectées, gouvernée par la protéine virale Tax peut persister pendant de longues périodes, malgré une forte immunité cellulaire. Une immortalisation des cellules infectées *in vivo* s'établit dès les premiers stades de l'infection.

Ces données indiquent finalement que la première et longue étape de leucémogénèse induite par HTLV-1 *in vivo* est une prolifération oligo/polyclonale continue de cellules infectées. Ainsi, l'immortalisation *in vivo* de clones cellulaires T par le virus offre un terrain propice à la

survenue de lésions génétiques secondaires. Ces mutations surviennent avec une fréquence élevée tant au niveau de la séquence provirale que du génome cellulaire hôte. Ceci explique la latence prolongée observée chez la plupart des patients entre l'infection et la déclaration de l'ATLL. Ce délai correspond vraisemblablement au temps nécessaire pour que des réarrangements surviennent au niveau de gènes impliqués dans la transformation.

II. L'INTEGRATION DES RETROVIRUS.

L'intégration est une étape fondamentale du cycle réplcatif des rétrovirus. Elle permet l'expression des gènes viraux, par l'intermédiaire de la machinerie enzymatique de la cellule, et la production de nouvelles particules virales. Elle permet aux rétrovirus de persister dans les cellules hôtes infectées, puisque les provirus intégrés sont transmis de façon stable au cours de la division cellulaire, selon un processus mendélien.

Notre connaissance des mécanismes moléculaires et biochimiques qui gouvernent cette réaction, s'est accrue considérablement au cours des dix dernières années. L'ensemble des études réalisées jusqu'à présent a porté essentiellement sur des réactions d'intégration *in vitro* ou sur l'analyse de provirus intégrés dans l'ADN de lignées cellulaires.

II.1 La réaction d'intégration.

Les premières études sur l'intégration rétrovirale ont été réalisées chez le MMLV (pour *Moloney Murine Leukemia Virus*), le SNV (pour *Spleen Necrosis Virus*) et les virus du groupe ALSV (pour *Avian Leukosis-Sarcoma Virus*). Elles ont permis d'identifier la protéine virale nécessaire à la réaction d'intégration. L'intégration du provirus dans l'ADN génomique de la cellule infectée s'effectue selon deux étapes catalysées par une protéine particulière, codée par la partie 3' du gène *pol*, l'IN. Cette protéine est nécessaire et suffisante à la catalyse de ces deux réactions (Craigie *et al.*, 1990 ; Katz *et al.*, 1990 ; Bushman *et al.*, 1991). Elle est produite au cours de l'assemblage des virions, à partir d'un précurseur polypeptidique Gag-pol clivé par la protéase virale.

Le précurseur immédiat de l'intégration est la forme linéaire du provirus, produite dans le cytoplasme par transcription inverse (Fujiwara *et al.*, 1988 ; Brown *et al.*, 1989). C'est la forme d'ADN viral (ADNv) la plus fréquemment retrouvée dans les cellules infectées.

Il existe également, dans le noyau des cellules infectées, des formes circulaires libres d'ADNv, contenant un ou deux LTR intacts. Ces formes circulaires ont été formées, soit par ligature des deux extrémités de l'ADNv linéaire (si les deux LTR sont juxtaposés), soit par recombinaison homologue entre les deux LTR (si 1 seul LTR est présent). La formation de ces molécules d'ADN circulaires ne nécessite pas l'intervention de l'IN. Une étude réalisée chez le rétrovirus HIV-1 indique que des enzymes cellulaires de recombinaison sont responsables de leur formation (Farnet *et al.*, 1991a).

Une troisième forme d'ADN circulaire peut être générée par "autointégration". En effet, les expériences d'intégration *in vitro* montrent que l'ADNv lui-même peut être utilisé comme ADN cible (Lee *et al.*, 1990). Cette réaction est suicidaire pour le rétrovirus. Au cours de celle-ci, des molécules circulaires complexes sont produites par autointégration d'une des extrémités du LTR dans l'ADNv. Cette réaction est schématisée dans la Figure 4G.

Le rôle et la fonction de ces molécules ne sont pas encore connus avec précision. Elles sont plutôt considérées à l'heure actuelle comme des produits abortifs de l'intégration.

De courtes séquences nucléotidiques répétées inversées, appelées IR (4 à 18 nucléotides selon le virus) sont présentes aux extrémités de chaque LTR. Elles définissent un locus *att* à chaque extrémité de l'ADNv néosynthétisé. Ces séquences sont indispensables à l'intégration du provirus dans l'ADN génomique. La totalité ou une partie seulement de ces séquences IR est reconnue par l'IN et participe à la réaction d'intégration.

Dans le cytoplasme, l'ADNv linéaire est associé à un complexe nucléoprotéique de grand taille, difficile à purifier à partir des extraits cytoplasmiques de cellules infectées. Ce complexe de préintégration ou PIC (pour *Preintegration Complex*), contient tous les composants nécessaires à l'intégration du provirus dans l'ADN cellulaire (Bowerman *et al.*, 1989 ; Farnet *et al.*, 1990). Il est constitué d'un certain nombre de protéines virales mais contient aussi probablement des protéines cellulaires spécifiques.

II.1.1. Principe général de la réaction d'intégration.

La réaction d'intégration se déroule en deux étapes successives. Elle est schématisée dans la Figure 4G :

- Dans un premier temps, l'activité endonucléasique de l'IN permet l'excision de deux nucléotides à l'extrémité 3' de chaque LTR. La dégradation exonucléasique des extrémités 3' s'arrête systématiquement au niveau du dinucléotide CA, conservé chez tous les rétrovirus. Cette étape de clivage, encore appelée "*3' processing*" aboutit à la formation de deux extrémités CA-OH 3'.

- Dans un second temps, les extrémités CA-OH 3' ainsi créées effectuent une "attaque nucléophile" de l'ADN hôte, en deux sites situés sur chacun des brins d'ADN cellulaire et séparés par quelques nucléotides. Les extrémités CA-OH 3' sont ensuite liées de manière covalente aux extrémités 5' de l'ADN cellulaire cible.

Le produit de cette réaction de transfert concertée présente donc des "trous" (ou *gaps*) à la jonction de l'ADN cible et de l'ADN proviral. Ces délétions sont ensuite réparées par un mécanisme encore non élucidé et qui ferait intervenir des enzymes cellulaires. La réparation de l'ADN cellulaire entraîne la formation de deux courtes séquences répétées génomiques de part et d'autre du provirus intégré. La taille de la séquence dupliquée est spécifique à chaque rétrovirus. Elle est par exemple de 6 pb pour le rétrovirus HTLV-1 (Seiki *et al.*, 1983), 5 pb pour le rétrovirus HIV-1 (Muesing *et al.*, 1985) et de 4 pb pour le MMLV (Shoemaker *et al.*, 1980).

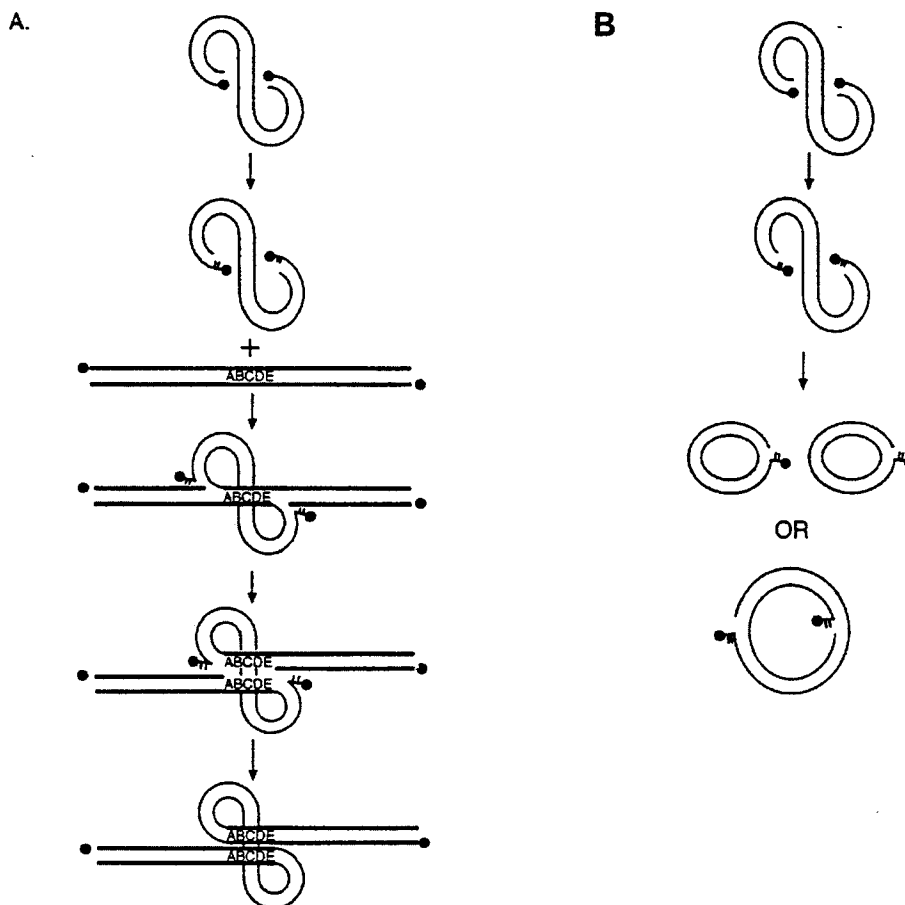


Figure 4G : Schéma des réactions d'intégration et d'autointégration (d'après Farnet et Bushman, 1996). (A) Réaction d'intégration *in vivo*. Les points noirs représentent les extrémités 5' de l'ADN. Les traits fins représentent l'ADN proviral, les traits épais représentent l'ADN cible. (B) Réaction d'autointégration. L'intégration d'une des extrémités de l'ADNv dans l'ADNv lui-même, soit dans le même brin, soit dans le brin opposé, conduit à la formation d'une seule ou de deux molécules circulaires.

II.1.2. Méthodes d'étude de l'intégration.

II.1.2.1. Les IN purifiées.

Diverses stratégies expérimentales ont été mises au point pour comprendre les aspects moléculaires et biochimiques de l'intégration des rétrovirus. Les complexes de préintégration dérivés du virion sont nécessaires à l'intégration du provirus dans l'ADN génomique *in vivo*. Cependant, les étapes biochimiques de clivage et de transfert de brin peuvent être étudiées en utilisant des tests d'intégration *in vitro* relativement simples, avec l'IN purifiée seule. Actuellement, les activités enzymatiques de 15 IN rétrovirales ont été décrites dans la

littérature. Ces enzymes ont été purifiées, le plus souvent grâce à l'utilisation d'un système d'expression procaryote, comme la bactérie *E.coli*, et ont été caractérisées en détail.

Ce sont, dans l'ordre de leur publication, les IN des virus AMV (pour *Avian Myeloblastosis Virus*), HIV-1, MMLV, ASV, HIV-2, HFV (pour *Human Foamy Virus*), FIV (pour *Feline Immunodeficiency Virus*), le rétrotransposon Ty1, virus Visna, CAEV (pour *Caprine Arthritis-Encephalitis Virus*), HERV (pour *Human Endogenous RetroVirus*), HTLV-2, EIAV (pour *Equine Infectious Anemia Virus*), HTLV-1 et SIV (pour *Simian Immunodeficiency Virus*).

II.1.2.2. Les réactions "d'autointégration" *in vitro*.

Les IN rétroviraux purifiés sont capables de catalyser *in vitro* le clivage spécifique d'oligonucléotides, dont la séquence nucléotidique correspond à celle des extrémités U3 ou U5 de l'ADNv. Le produit de cette réaction de clivage est un oligonucléotide plus court, auquel il manque deux nucléotides en 3'. L'IN catalyse ensuite l'insertion de cet oligonucléotide clivé dans une deuxième molécule : c'est la réaction de "transfert de brin", qui aboutit à la formation de produits "d'autointégration" plus longs que l'oligonucléotide de départ (Figure 5G). Un oligonucléotide différent de l'ADNv ou une molécule d'ADN plus grande, comme un plasmide, peuvent également servir de cible. Ces produits de clivage et/ou d'intégration sont facilement analysables par électrophorèse en gel de polyacrylamide dénaturant, suivie d'une autoradiographie.

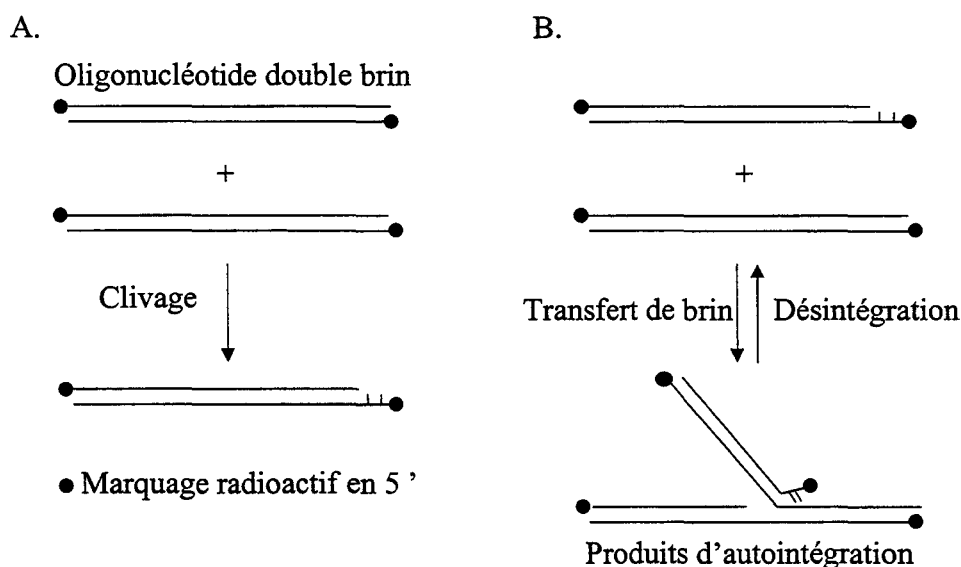


Figure 5G : Réactions d'intégration *in vitro*. (A) Réaction de clivage spécifique. (B) Réactions de transfert de brin et de désintégration.

De la même manière, le dinucléotide relargué après la réaction de clivage peut être détecté en radiomarquant l'extrémité 3' du brin clivé par l'IN (Engelman *et al.*, 1991 ; Vink *et al.*, 1991). Plusieurs produits de clivage sont obtenus, en fonction de la nature du nucléophile utilisé pour la réaction (de l'eau, du glycérol, des alcools ou encore l'extrémité 3'-OH de l'ADNv).

Les réactions de clivage spécifique et de transfert de brin ne sont pas couplées au niveau énergétique, puisque des oligonucléotides déjà clivés de manière artificielle, sont utilisés efficacement par l'IN pour la réaction de transfert de brin. Dans la plupart des cas, l'utilisation de ces oligonucléotides "pré-clivés" permet des rendements de transfert de brin supérieurs à ceux obtenus lors de l'utilisation d'oligonucléotides non clivés.

L'utilisation de cette stratégie expérimentale *in vitro* a permis de définir et de caractériser en détail les différentes activités enzymatiques de l'IN.

L'utilisation de mutants protéiques a également permis de déterminer les différents domaines fonctionnels de l'IN et leur rôle respectif dans la réaction d'intégration. Ces mutants peuvent être classés en deux catégories distinctes :

Certaines mutations affectent spécifiquement l'intégration du provirus dans l'ADN génomique alors que d'autres influencent la morphogénèse virale et la transcription inverse. En effet, certains changements de résidus au niveau de la séquence protéique de l'IN ont un impact sur la transcription inverse, qui précède l'intégration du provirus dans l'ADN génomique, mais également sur les étapes qui suivent l'intégration, comme l'assemblage et le relargage des virions (Shin *et al.*, 1994 ; Ansari-Lari *et al.*, 1995 ; Engelman *et al.*, 1995a ; Bukovsky *et al.*, 1996 ; Leavitt *et al.*, 1996).

II.1.2.3. Les réactions d'intégration *in vitro* avec des cibles complexes.

Des molécules complexes ont également été utilisées comme cibles dans des réactions d'intégration *in vitro* utilisant des IN purifiées. L'utilisation d'un plasmide comme cible de l'intégration permet d'amplifier facilement les produits d'intégration à l'aide d'oligonucléotides complémentaires d'une région connue de ce plasmide.

Des molécules d'ADN complexées à des protéines spécifiques, comme la protéine IHF (pour *Integration Host Factor*) ou la protéine CAP (pour *cAMP Activation Protein*) ont été utilisées pour étudier l'influence de la structure de l'ADN hôte, notamment la courbure de l'ADN, sur l'intégration des rétrovirus (Muller *et al.*, 1994 ; Bor *et al.*, 1995).

Enfin, des structures plus élaborées, comme les minichromosomes de levures et du SV40 (pour *Simian Virus 40*) ou les nucléosomes, ont également été utilisées comme cibles afin de se rapprocher des mécanismes moléculaires existant *in vivo* (Pryciak *et al.*, 1992a,b,c).

II.1.2.4. Réactions d'intégration "*in vivo*".

L'utilisation des IN purifiées, même si elle est simple et rapide, ne reproduit pas tous les aspects de l'intégration *in vivo*.

Le complexe de préintégration lui, est capable de réaliser, *in vitro*, une réaction d'intégration dite "concertée", dans laquelle les deux extrémités CA-OH 3' de l'ADNv sont liées aux deux extrémités 5' de l'ADN cible (Brown *et al.*, 1987). Cette réaction aboutit à la formation d'un intermédiaire non réparé, possédant les mêmes caractéristiques que le produit issu de l'intégration *in vivo*. Quelques IN purifiées réalisent cette réaction concertée mais avec un faible rendement et la plupart du temps en utilisant des molécules cibles circulaires. Mais la plupart du temps, la majorité des produits obtenus n'ont qu'une seule extrémité de l'ADNv liée à l'ADN cible. De plus, les conditions expérimentales sont différentes pour les deux systèmes. Ainsi, l'activité de l'IN purifiée est optimale en présence de Mn^{2+} , alors que les PICs ont une activité identique ou accrue en présence de Mg^{2+} . La réaction d'autointégration est très efficace *in vitro*, alors qu'elle est totalement réprimée dans le cytoplasme des cellules infectées (Lee *et al.*, 1994). Ces observations indiquent que d'autres protéines virales et cellulaires spécifiques sont importantes dans l'intégration *in vivo*.

La sélection des sites d'intégration et la caractérisation des IN ont été étudiées après infection de lignées cellulaires par des virus recombinants infectieux. L'ADN génomique est extrait puis digéré par des enzymes de restriction. Des techniques de blot de Southern, de PCR et de sélection génétique permettent ensuite la détection des provirus intégrés et l'analyse des séquences.

II.1.3. La réaction de clivage spécifique ou "*3' processing*".

Le clivage de l'ADNv s'effectue dans le cytoplasme des cellules infectées (Roth *et al.*, 1989). L'IN reconnaît les séquences IR aux deux extrémités de l'ADNv puis l'activité endonucléasique de la protéine permet l'hydrolyse du groupement phosphate situé en 3' du dinucléotide CA. L'attaque nucléophile de la liaison phosphodiester par l'eau permet

l'excision des deux nucléotides en 3' du groupement CA. L'IN peut également catalyser l'alcoolyse non spécifique de substrats non viraux (Katzman *et al.*, 1996).

Les deux résidus nucléotidiques CA sont très conservés chez toutes les IN rétrovirales. Ils sont également présents chez les rétrotransposons et les transposons de bactéries.

Leur présence est indispensable ; la délétion de ce dinucléotide aux extrémités U5 et U3 du SNV bloque l'intégration et la réplication virale (Panganiban *et al.*, 1983). Un résultat identique est obtenu lorsque l'on délète le dinucléotide CA au niveau de l'extrémité U5 du MMLV (Colicelli *et al.*, 1985).

Les IN tolèrent le changement d'un seul nucléotide au niveau de ce motif, mais les rendements de clivage sont considérablement réduits (Roth *et al.*, 1989). La position du dinucléotide CA par rapport à l'extrémité 3' de l'ADNv n'est pas un facteur limitant. En effet, l'IN du MMLV peut tolérer l'insertion de 4 bases supplémentaires sans modifier le site de clivage. Par contre, l'insertion de 10 nucléotides empêche la réaction de clivage (Colicelli *et al.*, 1988). Des résultats similaires ont été obtenus pour le rétrovirus HIV-1 (Vink *et al.*, 1991). L'IN reconnaît donc une séquence spécifique aux deux extrémités de l'ADNv et ces éléments ne doivent pas être trop éloignés de l'extrémité 3'.

II.1.4. La réaction de transfert de brin.

Les réactions de clivage et de transfert de brin sont isoénergétiques. Elles ne nécessitent la présence d'aucun facteur énergétique. Ces deux réactions sont donc probablement couplées. L'intégration ne s'effectue pas au niveau de cassures pré-existantes de l'ADN cellulaire mais au cours d'une réaction de transestérification pendant laquelle l'IN clive l'ADN, tout en liant les extrémités 3' de l'ADNv aux extrémités 5' de l'ADN cible. Cette réaction ne nécessite pas la formation d'un intermédiaire covalent entre l'ADN et la protéine. L'IN présente les extrémités CA-OH 3' de l'ADNv de tel sorte que l'atome d'oxygène accomplisse l'attaque nucléophile de la liaison phosphodiester (Engelman *et al.*, 1991).

II.1.5. La désintégration.

En plus des étapes de clivage et de transfert de brin, l'IN purifiée possède une activité supplémentaire *in vitro*, appelée "désintégration". Elle peut reconnaître un substrat oligonucléotidique mimant le produit de la réaction de transfert de brin. Le clivage de ce substrat par l'IN entraîne la formation d'un oligonucléotide clivé et du substrat cible (Figure 5G).

Des variants génétiques de l'IN, dépourvues d'activités de clivage et de transfert de brin, sont encore capables de catalyser cette réaction, facilitant ainsi l'étude fonctionnelle de ces mutants (Engelman *et al.*, 1992 ; van Gent *et al.*, 1992 ; Vincent *et al.*, 1993). Cependant, l'existence de cette activité enzymatique de l'IN reste à démontrer *in vivo*. Le provirus intégré dans l'ADN génomique n'a pas la même structure que le substrat oligonucléotidique, puisque les deux extrémités de l'ADN_v sont liées aux extrémités 5' de l'ADN cible. La réaction d'intégration *in vivo* n'est donc probablement pas réversible (Chow *et al.*, 1992).

L'étude de l'organisation structurale des IN rétrovirales et la cristallisation ou la résolution par résonance magnétique nucléaire (RMN) des différents domaines de la protéine, a également permis de préciser les mécanismes moléculaires et biochimiques de l'intégration du provirus dans l'ADN génomique.

II.2. L'intégrase.

II.2.1. Organisation structurale de l'IN.

Les IN rétrovirales comportent entre 270 et 350 acides aminés.

L'IN du rétrovirus HIV-1 (32 kDa) et celle du MMLV (46 kDa) sont les mieux caractérisées à l'heure actuelle. Chez les rétrovirus aviaires, la protéine est souvent associée à la transcriptase inverse.

La comparaison des séquences en acides aminés des différentes IN rétrovirales permet de définir plusieurs résidus acides aminés très conservés (Johnson *et al.*, 1986). La Figure 6G représente la comparaison des séquences protéiques de quelques IN rétrovirales.

HTLV-1	RTLALGTFQGRSSQAPFQALLPRLLS--RKVVYLHHVRSHTNLPDPISRLNALTDALLIT
HTLV-2	HSLAIGAFLGTSAHQTLQAALPLLQ--GKTIYLHHVRSHTNLPDPISTFNEYTDSLILA
HIV-1	QAQPDKSESELVNQIIIEQLIKKEKVV--LAWVPAHKGIGGNEQVDKLVSAGIRKVLFLDG
SIV	QSQPDHSESELVNQIIIEELIKKEKIY--LSWVPAHKGIGGNEQVDKLVSAGIRKVLFLDG
MMTV	PEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRI
RSV	LKMG---QEGVPSTAAAFILEDALSQRSAMAALVHVRSHSEVPGFTEGNDVADSQATF
	: : :
HTLV-1	PVL-QLSPAELHSFTHCGQTALTLOG--ATTTEASNILRS HA RCGNPQHQMGRGHIR
HTLV-2	PLV-PLTPQGLHGLTHCNQRALVSFG--ATPBREAKSLVQTCHTQTINSQHMPRGYIR
HIV-1	----IDKAQDEHEKYHSNWRAMASDFN--LPPVVAKEIVAS DK QLKGEAMHGQVDCSP
SIV	----IDRAQEEHERYHSNWKAMASDFN--LPPIVAKEIVAH DK QVKGEAMHGQVDCSP
MMTV	LTA-LESQAQESHALHHQNAALRFQFH--ITREQAREIVKLC PN PDWG--HAPQLGVNP
RSV	QAYPLREAKDLHTALHIGPRALSKACN--ISMQQAREVVQTC PH NSAP--ALEAGVNPR
	* * * : * *
HTLV-1	RGLLPNHIWQGDIITHFKYKNTLYRLHVWVDTFSGAISATQKRKETSSEAISSLLQAIHL
HTLV-2	RGLLPNHIWQGDVTHYKYKKYKYLHVWVDTFSGAVSVSCKKKETSCTETISAVLQAISSL
HIV-1	G-----IWQLDCTHLEGKVLIVAVHVASGYIEAEVIPAETGQET-----AYFLKLKLAGR
SIV	G-----IWQVDCTHLEGKVIIVAVHVASGYIEAEVIPAETGQET-----AYFLKLKLAGR
MMTV	RGLKPRVLWQMDVTHVSEFGKLKYVHVTVDTYSHFTFATARTGEATKDVQLQHLAQSFAYM
RSV	G-LGPLQIWQTDFTLEPRMAPRSWLAVTVDTASSAIVVTQHGRVTSVAVQHHWATAIAVL
	: ** * * : * : :
HTLV-1	GKPSYINTDNGPAYISQDFLNMCTSLAIRHTTHVPYNPTSSGLVRSNGILKTLTKYKYFT
HTLV-2	GKPLHINTDNGPAFLSQEFQEFCTSYRIKHSTHIPYNPTSSGLVERTNGVIKNNLNKYLL
HIV-1	WPVKTIHTDNGSNFTSTTVKAACWWAGIKQEFGIPYNPQSQGVVESMKNELKKIIGQVRD
SIV	WPVKTIHTDNGPNFTSAAVKAACWWADIKQEFGIPYNPQSQGVVESLNELKKIIGQVRD
MMTV	GIPQKIKTDNAPAYVSRSIQEFRLARWKISHVTGIPYNPQGQAIVERHQNIIKAQLNKLQK
RSV	GRPKAIKTDNGSCFTSKSTREWLARWGIAHTTGIPGNSQGQAMVERANRLKDRIRVLAE
	: : *** : * * : * * : ** : * :
HTLV-1	DK-----PDLPMDNALSIALWTINHLNVLTNCHKTRWQLHHSPLRQPIPETRSLSNKQT
HTLV-2	DC-----PNLPLDNAIHKALWTLNQLNVMNPSGKTRWQIIHSPPLPPIPEASTPPKPPP
HIV-1	QA-----EHLKTAVQMAVFIHNFKRKGGIG-GYSAGERIVDIIATDIQTKELQKQITKI
SIV	QA-----EHLKTAVQMAVFIHNFKRKGGIG-GYTAGERIIDIIATDIQTSLELQKQILKV
MMTV	AG-----KYTPHLLLAHALFVLNVNMDNQGHTAAERHWGPISADPKPMVMWKDLLTGT
RSV	GDGFMKRIPTSKQGELLAKAMALNHFERGE-NTKTPIQKHWRTVLTGPPVKIRIETG
	: : : :
HTLV-1	HWYY-----FKLPGLNSRQWKGPQEQALQEAAGAALIPVSASS-----AQW
HTLV-2	KWPFY-----YKLPGLTNQRWKGPLQSLQEAAGAALLSIDGS-----PRW
HIV-1	QNFR-----VYYRDSRDPLWKGPALKLLWKGEAVVIQDQNSD-----IKV
SIV	QKFR-----VYYRDSRDPWKGPATLLWKGEAVVIQDQGE-----LKV
MMTV	SWKGPDLVITAGRGYACVFPQDAETPIWVPDRFIRPFTERKEATPTPGTA-----EKT
RSV	EWEKGWNVLVWGRGYAAVKNRDTDKVIWVPDRKVKPDITQKDEVTKKDEASPLFAGISDW
	*
HTLV-1	IPWRLKRAACPRP-----VGGPADPKKDLQHHG--
HTLV-2	IPWRFLKKAACPRPD-----ASELAEHAATDHQHHG--
HIV-1	VPRRKAKIIRDYGKQ-----MAGDDCVASRQDE----
SIV	VPRRKAKIIRDYGKQ-----MAGDDCVASRQNE-----
MMTV	PPRDEKDQQESPKNESSPHQREDGLATSAGVDLRSGGGP
RSV	IPWEDEQEGLOGETASNQKQERPGEDTLAANES-----
	*

Figure 6G : Comparaison des séquences protéiques de quelques IN rétrovirales. Les domaines grisés correspondent aux domaines Nt et Ct de la protéine. Le domaine en blanc correspond au domaine catalytique. Les acides aminés en rouge correspondent aux résidus très conservés HHCC et DDE qui caractérisent les IN. Les astérisques correspondent aux autres résidus identiques et les doubles points, aux résidus conservés.

La Figure 7G représente de manière schématique l'IN de HIV-1 avec les différents domaines protéiques communs à tous les rétrovirus.

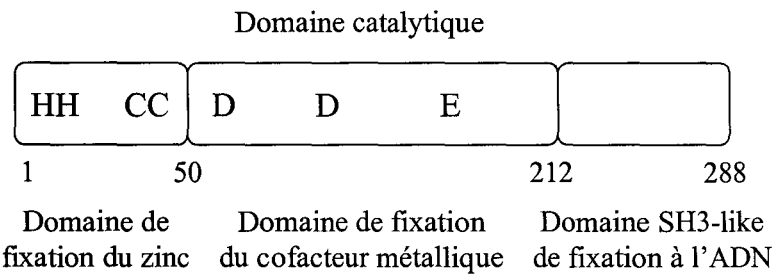


Figure 7G : Schéma des différents domaines protéiques de l'IN de HIV-1.

La présence de ces acides aminés très conservés range les IN rétrovirales dans la superfamille des nucléases et des polynucléotidyltransférases, comme la RNase H d'*E.coli*, la Résolvase RuC d'*E.coli*, les IN des rétrotransposons ou encore les transposases des transposons bactériens (Dyda *et al.*, 1994). Cette analogie de structure protéique coïncide avec les similitudes fonctionnelles observées entre les mécanismes de transposition de certains éléments génétiques mobiles et la réaction d'intégration rétrovirale.

Comme le montre la Figure 7G, les IN rétrovirales sont caractérisées par 3 domaines distincts. Les domaines les mieux conservés sont les domaines N-terminal et catalytique de la protéine. Les trois domaines de la protéine sont nécessaires pour catalyser les réactions de clivage et de transfert de brin (Drelich *et al.*, 1992 ; Bushman *et al.*, 1993). Le domaine catalytique seul est suffisant pour catalyser l'étape de désintégration (Bushman *et al.*, 1993 ; Engelman *et al.*, 1994).

II.2.1.1. Le domaine N-terminal.

Le domaine N-terminal de la protéine, d'environ 50 aa, comporte deux résidus histidine et deux résidus cystéines très conservés, formant un motif H-X₍₃₎-H-X₍₂₃₎-C-X₍₂₎-C présent chez toutes les IN rétrovirales (Figure 6G). De nombreux travaux, pour la plupart réalisés *in vitro*, prouvent que ce domaine possède un rôle fonctionnel important dans le mécanisme d'intégration des rétrovirus, mais les résultats obtenus sont variés et ne permettent pas encore de déterminer son rôle avec exactitude.

L'enchaînement des résidus histidines et cystéines crée une structure en doigt de zinc, analogue à celle de certaines protéines de fixation à l'ADN. Les domaines N-terminaux de l'IN de HIV-1 et de MMLV sont capables de fixer le zinc de manière équimolaire (Burke *et al.*, 1992 ; Zheng *et al.*, 1996) ; la fixation du zinc renforce la multimérisation de la protéine HIV-1 entière et du domaine HHCC du MMLV. Elle stimule également les réactions de clivage et de transfert de brin *in vitro* (Zheng *et al.*, 1996 ; Lee *et al.*, 1997 ; Yang *et al.*, 1999). Enfin, l'étude par RMN des domaines N-terminaux des IN de HIV-1 et de HIV-2, montrent que la fixation du zinc stabilise leur structure (Cai *et al.*, 1997 ; Eijkelenboom *et al.*, 1997). Ces domaines sont caractérisés par 4 hélices α , au lieu de la structure α -feuillet β caractéristique des motifs en doigts de zinc.

Malgré la présence de ce motif en doigt de zinc, aucune activité de fixation à l'ADN n'a été mise en évidence au niveau du domaine N-terminal (Khan *et al.*, 1991 ; Schauer *et al.*, 1992). Lorsque celui-ci est délété, la protéine purifiée conserve ses propriétés de fixation à l'ADN (Vincent *et al.*, 1993). Les expériences d'intégration *in vitro*, réalisées sur le MMLV par l'équipe de Roth (Yang *et al.*, 1999) confirment ces résultats. Le domaine N-terminal interviendrait principalement dans la multimérisation de la protéine, en association avec le domaine Ct (Donzella *et al.*, 1998) ; la structure cristalline des IN de HIV-1 et de ASV confirme le rapprochement spatial de ces deux domaines protéiques (Bizub *et al.*, 1994). La fixation du zinc au niveau des résidus HHCC entraîne un changement de conformation du domaine N-terminal, favorisant l'activité catalytique de l'IN.

II.2.1.2. Le site catalytique de l'enzyme.

Ce domaine d'environ 150 aa comporte 3 résidus très conservés (deux aspartates et 1 glutamate), déterminants pour l'activité enzymatique de la protéine (Figure 6G).

Ces résidus acides D,D(35)E sont essentiels pour toutes les activités catalytiques de l'IN et fixent les cofacteurs Mg^{2+} ou Mn^{2+} , qui interviennent dans le processus d'intégration (Drelich *et al.*, 1992 ; Engelman *et al.*, 1992 ; Kulkosky *et al.*, 1992). La modification d'un seul de ces 3 aa, par mutagenèse dirigée, inhibe ou diminue fortement les réactions de clivage, de transfert de brin et de désintégration (Engelman *et al.*, 1992 ; van Gent *et al.*, 1992 ; Leavitt *et al.*, 1993).

Le domaine catalytique de l'IN de HIV-1 est suffisant pour catalyser la réaction de désintégration mais les étapes de clivage et de transfert de brin nécessitent des fonctions

supplémentaires contenues dans la protéine complète (Bushman *et al.*, 1993). Des résultats similaires ont ensuite été obtenus avec d'autres IN rétrovirales (Bushman *et al.*, 1994a ; Katzman *et al.*, 1995).

Les structures cristallines des domaines catalytiques des IN de HIV-1 et du ASV ont été caractérisées (Dyda *et al.*, 1994 ; Bujacz *et al.*, 1995). Elles révèlent la présence d'une boucle flexible, impliquée dans les changements de conformation induits par la fixation covalente du cofacteur métallique Mg^{2+} ou Mn^{2+} . La structure cristalline du domaine central de l'IN de ASV montre que les résidus conservés DDE fixent des cations bivalents comme le Mn^{2+} , le Mg^{2+} , le Zn^{2+} , le Cd^{2+} et le Ca^{2+} (Bujacz *et al.*, 1997). Tous ces cations n'agissent pas comme des cofacteurs de l'intégration ; seuls les ions Mg^{2+} et Mn^{2+} permettent à l'IN d'exercer son activité catalytique (Engelman *et al.*, 1995b). Généralement, l'activité enzymatique de la protéine purifiée est plus efficace en présence de Mn^{2+} qu'en présence de Mg^{2+} . Le manganèse favorise la réaction de transfert de brin et l'activité endonucléasique de l'IN. Cependant, le clivage en 3' du dinucléotide CA par les IN du RSV (pour *Rous Sarcoma Virus*) ou de HIV-1 en présence de Mn^{2+} , est moins spécifique qu'avec le Mg^{2+} (Katzman *et al.*, 1989 ; Engelman *et al.*, 1995b). L'ion magnésium étant plus abondant dans la cellule, il est donc probablement utilisé *in vivo* par l'IN. Une étude récente suggère que l'ion Zn^{2+} pourrait être un cofacteur de l'IN, en stimulant exclusivement l'activité de clivage de la protéine. Cette fonction ne nécessite pas la présence du domaine N-terminal (Bujacz *et al.*, 1997).

II.2.1.3. Le domaine C-terminal.

L'alignement des séquences des domaines C-terminaux des différentes IN rétrovirales ne révèle pas d'homologie significative. Cependant, tous ces domaines C-terminaux possèdent des propriétés de fixation à l'ADN non spécifique.

Le domaine C-terminal de l'IN de HIV-1 a été caractérisé par RMN et révèle la présence inattendue d'un domaine "SH3-like" (pour *Src Homology 3 Domain*) retrouvé généralement au niveau des protéines impliquées dans la transduction des signaux cellulaires. Un sillon en forme de selle, présent au niveau du dimère, serait responsable de la fixation de la protéine à l'ADN (Eijkelenboom *et al.*, 1995 ; Lodi *et al.*, 1995). Ce domaine aurait aussi un rôle dans la multimérisation des IN de HIV-1 et de l'ASV (Andrake *et al.*, 1995 ; Jenkins *et al.*, 1996).

II.2.2. La forme oligomérique de la protéine.

Les organisations tridimensionnelles des différents domaines de la protéine sont parfaitement connues à l'heure actuelle, notamment pour la protéine HIV-1. Cependant, aucune donnée structurale valable n'a pu être établie pour les 18 derniers aa de l'enzyme. La protéine entière est très peu soluble, d'où la difficulté pour les auteurs à la purifier et à la cristalliser.

Le domaine catalytique de l'IN de ASV est soluble et peut être facilement cristallisé, sans modifications aucune (Bujacz *et al.*, 1995). Par contre, le domaine central de la protéine HIV-1 est insoluble ; il a fallu muter la protéine pour obtenir une forme protéique soluble, capable d'être cristallisée : 29 résidus hydrophobes ont été remplacés successivement, soit par des résidus alanine, soit par des résidus lysine (Jenkins *et al.*, 1995). Cette stratégie expérimentale a permis d'obtenir 3 formes cristallines différentes, dont la mieux caractérisée et la plus stable est la forme F185K (Dyda *et al.*, 1994).

Les structures des domaines N-terminal et C-terminal de l'IN de HIV-1 ont été obtenues par résonance magnétique nucléaire (RMN) (Eijkelenboom *et al.*, 1995 ; Lodi *et al.*, 1995 ; Cai *et al.*, 1997).

Les homologies de séquences entre les différentes protéines rétrovirales et la conservation des résidus HHCC et DDE, suggèrent que les structures tridimensionnelles des IN sont également conservées. Cependant, des variations de tailles entre les différents domaines des IN peuvent modifier la structure de la protéine. Ainsi, l'IN du MMLV possède 16 résidus acides aminés supplémentaires dans la partie N-terminale de la protéine (Johnson *et al.*, 1986). Par ailleurs, rien n'indique que la structure du domaine C-terminal peu conservé, soit la même pour toutes les IN rétrovirales.

La manière dont les 3 domaines de l'IN interagissent ensemble, pour former la protéine active, n'est pas encore connue. Le facteur limitant à l'heure actuelle, concerne le statut oligomérique de la protéine pendant la catalyse. Les IN recombinantes purifiées existent en solution sous forme de monomères, de dimères, de tétramères, voire plus (Sherman *et al.*, 1990 ; Jones *et al.*, 1992 ; Vincent *et al.*, 1993 ; Grandgenett *et al.*, 1994 ; Asante-Appiah *et al.*, 1998). Aucune expérience n'a encore permis de déterminer précisément le nombre de monomères nécessaire à la réaction d'intégration *in vivo*. Cependant, nous savons aujourd'hui

avec certitude, que la forme active de la protéine est un multimère, au minimum un dimère (Jones *et al.*, 1992 ; Engelman *et al.*, 1993 ; van Gent *et al.*, 1993).

Des expériences biochimiques variées ont été réalisées pour identifier les composants de l'IN impliqués dans la multimérisation de la protéine. Des études réalisées avec l'IN de ASV et l'IN de HIV-1 notamment, ont montré que les domaines C-terminal et catalytique de la protéine étaient responsables de cette activité (Andrake *et al.*, 1995 ; Jenkins *et al.*, 1996). Toutefois, chaque domaine pris indépendamment, est capable de se dimériser ; ceci suggère que les trois domaines de la protéine interagissent pour former le multimère actif.

II.3. Les complexes de préintégration ou PICs.

Dans le cytoplasme, l'ADNv linéaire est associé à un complexe nucléoprotéique de grande taille contenant tous les composants nécessaires à l'intégration du provirus dans l'ADN cellulaire.

II.3.1. Structure et composition.

Les PICs sont de grosses particules relativement denses, dans lesquelles l'ADNv est compacté (Bowerman *et al.*, 1989). Les extrémités de l'ADNv sont protégées par des protéines contre les nucléases ou les complexes de recombinaison. Elles sont maintenues proches l'une de l'autre par un complexe protéique dense, dans lequel l'IN est présente. La partie la plus interne de l'ADNv est plus exposée et moins compactée (Miller *et al.*, 1997 ; Wei *et al.*, 1997).

Les PICs sont suffisamment stables pour être purifiés à partir d'extraits cytoplasmiques de cellules infectées. On peut ainsi déterminer leur structure et leur composition. Ces PICs purifiés peuvent ensuite réaliser des réactions d'intégration *in vitro*. Les méthodes expérimentales utilisées sont basées soit sur la sélection génétique (Brown *et al.*, 1987) soit sur la PCR pour amplifier les produits d'intégration (Kitamura *et al.*, 1992 ; Pryciak *et al.*, 1992a,c). Quand les produits d'intégration sont suffisamment abondants, des méthodes plus simples et plus quantitatives, mais moins sensibles, comme l'hybridation par blot de Southern peuvent être utilisées (Fujiwara *et al.*, 1988 ; Brown *et al.*, 1989 ; Farnet *et al.*, 1990 ; Lee *et al.*, 1990).

Les différents composants du complexe nucléoprotéique de préintégration ont été déterminés principalement chez le MMLV et le HIV-1. Des expériences de purifications successives sur gradient de sédimentation, ont révélé la présence de plusieurs protéines virales, dont l'IN. La composition de ces PICs varie d'un rétrovirus à l'autre. Dans le système HIV-1, la RT, l'IN, la protéine virale R ou Vpr, et la protéine de la matrice (MA) sont présentes dans le complexe de préintégration. Des quantités relativement faibles de protéine de la capsid sont également détectées (Farnet *et al.*, 1991b ; Miller *et al.*, 1997). Mais les résultats obtenus par plusieurs équipes concernant la composition et les propriétés physiques des PICs de HIV-1 divergent. Les différences observées proviennent des méthodes expérimentales utilisées pour isoler et tester l'activité de ces PICs. La confirmation de la présence des protéines Vpr et MA dans les complexes de préintégration du rétrovirus HIV-1 est renforcée par d'autres résultats qui montrent que ces deux protéines ont un rôle déterminant dans le transport de l'ADNv dans le noyau des cellules infectées.

Dans le système MMLV par contre, seules l'IN et la protéine de la capsid sont détectées (Bowerman *et al.*, 1989).

II.3.2. Transport nucléaire des complexes de préintégration.

L'intégration du rétrovirus MMLV dans l'ADN génomique de la cellule hôte dépend de la mitose. En effet, les oncorétrovirus infectent exclusivement des cellules en division (Harel *et al.*, 1981). Le complexe nucléoprotéique tire profit de la fragilisation et de la rupture de l'enveloppe nucléaire, qui ont lieu en tout début de mitose, pour migrer dans le noyau de la cellule infectée (Roe *et al.*, 1993). L'ADNv peut alors s'intégrer dans l'ADN génomique. Cette étape ne nécessite pas forcément la synthèse de l'ADN cellulaire. Ces résultats, obtenus avec le rétrovirus MMLV, n'ont pas été confirmés pour d'autres oncorétrovirus.

Les lentivirus par contre, et en particulier HIV-1, sont capables de se répliquer dans des cellules quiescentes ou différenciées, comme les macrophages, les cellules dendritiques, les cellules microgliales ou des cellules arrêtées en phase G2 par irradiation gamma (Weinberg *et al.*, 1991 ; Lewis *et al.*, 1992 ; Schmidtmayerova *et al.*, 1997). Le complexe de préintégration de HIV-1 peut donc franchir l'enveloppe nucléaire, quelle que soit la phase du cycle cellulaire. Le signal nécessaire à l'entrée du PIC dans le noyau en interphase serait délivré soit

par la protéine de la matrice, soit par la protéine Vpr (Heinzinger *et al.*, 1994 ; Popov *et al.*, 1998).

La protéine MA du rétrovirus HIV-1, présente au niveau du PIC, possède deux propriétés distinctes:

La présence d'un groupement myristique à l'extrémité N-terminale de la protéine permet l'ancrage du polypeptide précurseur Pr55gag dans la membrane plasmique (Zhou *et al.*, 1994). Cette association est indispensable à l'assemblage des virions et à leur relargage dans le milieu extracellulaire.

D'autre part, un signal de localisation nucléaire (NLS pour *Nuclear Localisation Signal*) porté par un motif conservé de résidus basiques, serait impliqué dans le transport actif du complexe nucléoprotéique au travers des pores nucléaires, et nécessiterait une interaction moléculaire directe entre la protéine de la MA et le domaine catalytique de l'IN (Bukrinsky *et al.*, 1992, 1993 ; Gallay *et al.*, 1995). La protéine MA interagirait également avec un facteur cellulaire d'importation nucléaire, l'Importine- α qui serait associé au PIC (Gallay *et al.*, 1996, 1997).

D'autres études ont par contre montré que la protéine MA n'était pas directement impliquée dans le transport nucléaire des PICs (Freed *et al.*, 1995 ; Fouchier *et al.*, 1997).

L'IN elle-même porte un signal de localisation nucléaire (Gallay *et al.*, 1997). Toutefois, l'implication directe de ce motif dans l'importation nucléaire des PICs n'a pas encore été prouvée expérimentalement.

Ces deux types de rétrovirus, les oncorétrovirus et les lentivirus, utilisent donc des procédés différents pour franchir l'enveloppe nucléaire ; ceci pourrait expliquer pourquoi les PICs correspondants n'ont pas la même composition. Malgré tout, ces expériences ont été réalisées uniquement chez 2 rétrovirus et nécessiteraient des investigations supplémentaires.

II.4. Rôle des protéines cellulaires ou virales dans le mécanisme d'intégration.

II.4.1. Les complexes de préintégration.

Une étude réalisée en 1997 (Farnet *et al.*, 1997) a permis de mettre en évidence la présence d'une petite protéine de fixation à l'ADN, la protéine HMG I(Y) (pour *High Mobility Group Protein*), dans le PIC de HIV-1. En effet, le traitement des PICs par de fortes concentrations en sel, diminue fortement l'activité d'intégration *in vitro* du complexe, alors que l'IN y est toujours présente. L'ajout d'extraits cytoplasmiques de cellules non infectées restaure l'activité initiale du PIC. La protéine responsable est la protéine HMG I(Y). Elle a également été mise en évidence dans les PICs du MMLV (Li *et al.*, 1998).

L'étude des PICs du MMLV a également révélé la présence d'une protéine cellulaire qui empêche la réaction d'autointégration, suicidaire pour le rétrovirus. Elle a été baptisée pour cette raison BAF (pour *Barrier to Autointegration Factor*) (Lee *et al.*, 1994 ; Cai *et al.*, 1998 ; Chen *et al.*, 1998 ; Lee *et al.*, 1998).

II.4.2. Les IN purifiées.

Des protéines cellulaires stimulent l'activité des IN purifiées *in vitro*, et sont par conséquent des candidats potentiels de protéines cellulaires, importantes pour l'intégration *in vivo*. Toutefois, leur importance *in vivo* est difficile à déterminer.

L'IN du MMLV est stimulée 3 à 13 fois par la SSB d'*E.coli* (pour *Single-Stranded DNA-Binding Protein*) ou la RNase A (Lee *et al.*, 1998). De même, l'IN de HIV-1 peut être stimulée par la RNase T1, la NC et par un mélange de protéines de fixation à l'ADN (Bushman *et al.*, 1990 ; Miller *et al.*, 1995 ; Carteau *et al.*, 1997). Les protéines HMG-1 et HMG-2 stimulent légèrement l'activité catalytique de l'IN du ASV (Aiyar *et al.*, 1996).

Une autre protéine cellulaire, baptisée INI-1 (pour *Integrase Interactor 1*), et identifiée par la technique de double hybride, interagit *in vitro* avec l'IN de HIV-1, pour favoriser la réaction de transfert de brin (Kalpana *et al.*, 1994). La fonction de cette protéine INI-1 est encore mal connue, mais son homologie importante avec le facteur de transcription SNF5 de la levure *Saccharomyces cerevisiae* suggère qu'elle serait impliquée dans la réorganisation de la

chromatine, notamment au niveau des zones de transcription actives. INI-1 favoriserait ainsi, via son interaction avec l'IN, l'insertion de l'ADNv dans les zones de transcription actives du génome. Quelques travaux basés sur l'analyse des sites d'intégration du MMLV et de l'ALV ont montré que les rétrovirus s'intégraient préférentiellement dans des régions de transcription actives du génome, et donc potentiellement enrichies en protéine INI-1. Toutefois, l'implication de cette protéine dans le mécanisme d'intégration n'a pas été confirmée *in vivo*. De plus, d'autres études plus récentes n'ont pas permis d'établir une connexion étroite entre la transcription et l'intégration (Carteau *et al.*, 1998 ; Leclercq *et al.*, 2000).

II.4.3. La réparation de l'ADN cible.

L'intégration de l'ADNv dans le génome de la cellule hôte entraîne la duplication d'une courte séquence hôte, de part et d'autre du provirus intégré. Cette duplication est générée au cours de la réparation de l'intermédiaire d'intégration, après la réaction de transfert de brin. Les protéines impliquées dans la réparation de l'ADN comprennent l'IN elle-même, ainsi que des protéines cellulaires.

D'après une étude réalisée en 1998, l'IN posséderait une activité ADN polymérase au niveau du domaine C-terminal, qui serait impliquée dans les mécanismes de réparation de l'ADN génomique (Acel *et al.*, 1998).

La DNA-PK (pour *DNA-dependent Protein Kinase*) est impliquée dans le mécanisme cellulaire de réparation des cassures de l'ADN double brin. Des cellules murines *Scid*, dépourvues de fonction DNA-PK, ont été infectées par un vecteur rétroviral aviaire. Quelques heures après l'infection, les cellules meurent par apoptose. Cette réponse dépend de l'activité enzymatique de l'IN rétrovirale. En effet, les cellules ne meurent pas lorsqu'elles sont infectées par un virus déficient (Daniel *et al.*, 1999). Le modèle envisagé par les auteurs est le suivant : après la formation de l'intermédiaire proviral, les délétions générées de part et d'autre de chaque LTR sont considérées comme des dommages à l'ADN par la cellule hôte, et la voie de transduction DNA-PK est activée. La réparation de l'ADN cellulaire aboutit ainsi à la formation des séquences répétées de part et d'autre du provirus intégré.

II.4.4. Les réactions "concertées".

Des systèmes d'intégration efficaces, utilisant des IN purifiées, ont été élaborés et permettent de réaliser des réactions d'intégration dites "concertées", au cours desquelles les deux extrémités CA-OH 3' de l'ADNv sont liées aux deux extrémités 5' de l'ADN cible. La catalyse d'une telle réaction a pu être obtenue avec les IN purifiées des virus HIV-1 ou ASV mises en présence d'un plasmide cible linéaire ou circulaire de 3,4 kb, et d'un oligonucléotide portant les extrémités U3 et U5 de l'ADNv. L'addition de protéines de la famille HMG augmente l'efficacité de la réaction.

Soixante-quinze à 90% des produits obtenus sont issus d'événements d'intégration concertée et possèdent toutes les caractéristiques des provirus intégrés *in vivo* (Hindmarsh *et al.*, 1999) : les deux extrémités de l'ADNv sont liées aux extrémités 5' de l'ADN cible et sont encadrées par deux séquences répétées directes de 5 ou 6 pb.

Toutes les protéines HMG testées stimulent l'intégration. Cependant, la protéine HMG I(Y), présente dans les complexes de préintégration, est la plus efficace. Les protéines de la famille HMG sont nombreuses. Ces petites protéines non histones de fixation à l'ADN, influencent notamment la structure de la chromatine. Elles sont capables notamment de courber ou de dérouler l'ADN *in vitro*. Ces protéines permettraient le rapprochement des deux extrémités de l'ADNv, par courbure de l'ADN, favorisant ainsi la réaction concertée, caractéristique de l'intégration *in vivo*. Elles pourraient également permettre le relâchement de l'ADN aux extrémités du provirus, facilitant ainsi la fixation de l'IN.

II.5. Reconnaissance des substrats.

Les mécanismes de reconnaissance de l'ADNv et de l'ADN cible par l'IN sont encore mal compris à l'heure actuelle. Cependant, les nombreuses études *in vitro* réalisées jusqu'à présent, ont permis de dégager, malgré quelques problèmes d'interprétation liés aux expériences elles-mêmes, un certain nombre de caractéristiques communes à toutes les IN rétrovirales.

II.5.1. Reconnaissance de l'ADNv.

II.5.1.1. Les extrémités virales.

Les IN sont de manière générale plus actives sur des substrats double brin. Cependant, des mésappariements sont tolérés à l'extrémité des LTR (van den Ent *et al.*, 1994 ; Mazumder *et al.*, 1995). La taille minimale des substrats, requise pour une intégration efficace, est de 15 pb (Katzman *et al.*, 1989 ; Lafemina *et al.*, 1991 ; Sherman *et al.*, 1992).

Le dinucléotide CA est déterminant pour les étapes de clivage et de transfert de brin. L'importance de ce dinucléotide très conservé a été clairement établie grâce à des expériences de mutagenèse, réalisées avec les IN du HIV-1 (Bushman *et al.*, 1991 ; Lafemina *et al.*, 1991 ; Vink *et al.*, 1991 ; Sherman *et al.*, 1992 ; van den Ent *et al.*, 1994), du HFV (Pahl *et al.*, 1993), du virus Visna (Katzman *et al.*, 1994), de l'ASV (Kukolj *et al.*, 1995) et du MMLV (Craigie *et al.*, 1990). Ces deux résidus nucléotidiques sont déterminants pour la reconnaissance des extrémités de l'ADNv par l'IN, mais ne participent pas directement à l'activité catalytique de l'enzyme.

La nature des deux nucléotides, situés en 3' du dinucléotide CA, n'est pas primordiale. Des substitutions de ces résidus ont généralement peu d'effets sur la réplication (Colicelli and Goff 1985 ; Colicelli *et al.*, 1988 ; Roth *et al.*, 1989, 1990). Par contre, la délétion de ces deux nucléotides est létale pour le virus (Colicelli and Goff, 1985).

Les interactions fonctionnelles de l'IN avec les deux extrémités de l'ADNv sont coordonnées. En effet, des mutations au niveau du dinucléotide CA de l'extrémité U3 du MMLV inhibe la réaction de clivage spécifique au niveau de ce LTR, mais également au niveau du U5 non altéré (Murphy *et al.*, 1992).

Plusieurs travaux ont montré que les nucléotides situés en 5', à proximité de ce dinucléotide CA étaient également importants pour l'étape de clivage. Ainsi, les nucléotides -5 et -6 (si on numérote les nucléotides à partir de l'extrémité 3' de l'ADNv) sont très importants pour l'étape de clivage des IN du HIV-1, du virus Visna, et de l'AMV (Katzman *et al.*, 1989, 1996). Une restriction similaire a été observée pour les nucléotides -5 et -7 des LTR de HTLV-1, HTLV-2 et MMLV (Balakrishnan *et al.*, 1997).

Toutes les études réalisées *in vitro* avec l'IN du HIV-1 ont montré que les 6 pb terminales des régions U3 et U5 contenaient toutes les informations nécessaires pour une activité optimale de l'IN. Il en est de même pour les 9 pb terminales du LTR du MMLV. Les nucléotides situés plus en 5' dans le LTR (positions 9 à 20) ne semblent pas déterminants pour la réaction d'intégration, même si certaines substitutions peuvent affecter l'activité de l'enzyme (Vink *et al.*, 1991 ; Leavitt *et al.*, 1992 ; Sherman *et al.*, 1992 ; Balakrishnan *et al.*, 1997).

L'addition de nucléotides à l'extrémité 3' du LTR modifie très peu l'activité de l'enzyme. Ainsi, l'IN du HIV-1 clive efficacement des substrats avec 5 bases supplémentaires situées en 3' du dinucléotide CA. Cependant, une extension nucléotidique de 6 pb inhibe complètement l'étape de clivage (Vink *et al.*, 1991 ; Bushman *et al.*, 1991 ; Sherman *et al.*, 1992). Une tolérance similaire est observée pour les IN des virus Visna et MMLV (Craigie *et al.*, 1990 ; Katzman *et al.*, 1994).

La reconnaissance de l'ADNv par l'IN *in vitro* dépend donc de la séquence présente aux extrémités 3' de l'ADNv et de quelques contraintes spatiales.

In vivo, les résultats obtenus sont plus variés ; par exemple, l'importance du dinucléotide CA est moins évidente. Certains rétrovirus tolèrent des mutations au niveau de ce dinucléotide consensus (Roth *et al.*, 1989 ; Masuda *et al.*, 1995).

En général, la taille de la séquence présente à l'extrémité de l'ADNv et nécessaire à l'intégration du virus, est imparfaitement corrélée avec la longueur des IR. Par exemple, la séquence répétée inversée de 5 pb du LTR du SNV n'est pas suffisante pour l'intégration ; 12 pb au niveau du U3 et 8 pb au niveau du U5 sont nécessaires pour une intégration efficace *in vivo* (Panganiban *et al.*, 1983). Des résultats similaires sont obtenus pour d'autres rétrovirus (Cobrinik *et al.*, 1987 ; Murphy *et al.*, 1993).

La machinerie enzymatique reconnaît donc un nombre restreint de nucléotides au niveau de l'extrémité 3' des LTR. Le nombre de nucléotides situé en 3' du dinucléotide CA n'est pas critique mais il doit être relativement faible et supérieur à 0. Les deux extrémités des deux LTR U3 et U5 peuvent être identiques ou distinctes. Beaucoup de substitutions sont tolérées par les IN.

II.5.1.2. Les domaines protéiques impliqués.

Le domaine de l'IN responsable de la reconnaissance spécifique de l'ADNv a d'abord été identifié dans la région N-terminale de la protéine, puis dans le domaine Ct. Plus récemment, l'utilisation de protéines chimériques HIV-1/virus Visna a clairement établi que cette région se situait au niveau du domaine catalytique de l'enzyme. En ce qui concerne HIV-1, elle se situerait entre les résidus 50 et 212 de la protéine (Katzman *et al.*, 1998). D'autres domaines de la protéine peuvent aussi intervenir dans la reconnaissance de l'ADN.

La reconnaissance de l'ADNv nécessite la formation de complexes IN-LTR stables dans le cytoplasme des cellules infectées. En effet, c'est dans ce compartiment cellulaire qu'a lieu le clivage des extrémités virales. Ces complexes doivent être maintenus dans leur conformation puis transportés vers le noyau et l'ADN cible. La présence des deux nucléotides aux extrémités 5' des brins non clivés et la fixation des cofacteurs métalliques sont importantes pour la formation de ces complexes stables et leur durée de vie (Ellison *et al.*, 1994 ; Vink *et al.*, 1994 ; Wolfe *et al.*, 1996). La formation de ces complexes stables nécessite la présence des trois domaines de la protéine.

II.5.2. Reconnaissance de l'ADN cible.

Bien que la reconnaissance de l'ADN cible ne soit pas spécifique d'un locus particulier du génome hôte, il existe des régions préférentielles d'intégration.

Ainsi, on peut opposer la grande spécificité de l'interaction IN-ADNv à celle, moins importante, de la reconnaissance IN-ADN cible. La structure de l'ADN cible et sa composition nucléotidique influencent sa reconnaissance par l'IN. Les propriétés de fixation à l'ADN de l'IN résident principalement dans le domaine C-terminal de la protéine. On ne sait pas encore à l'heure actuelle, dans le cas où la protéine active est un dimère, s'il existe 1 seul site de fixation pour l'ADNv et l'ADN cible, ou s'il existe au contraire deux sites de fixation distincts ou encore si ces sites se chevauchent.

Un modèle se dégage toutefois des différents résultats obtenus : l'ADNv se fixe dans un premier temps, de manière non spécifique au domaine C-terminal de l'IN, puis le domaine catalytique de la protéine reconnaît spécifiquement les extrémités de l'ADNv et positionne correctement l'ADN cible pour que l'attaque nucléophile et la liaison covalente aient lieu.

Lorsque des IN rétrovirales sont en contact avec une même molécule d'ADN cible, les profils d'intégration obtenus pour chaque protéine sont différents (van Gent *et al.*, 1991 ; Pryciak *et al.*, 1992c ; Bushman *et al.*, 1994b ; Katzman *et al.*, 1995 ; Shibagaki *et al.*, 1997). L'IN a donc un rôle dans la reconnaissance de l'ADN cible. Cependant, d'autres facteurs plus importants influencent la sélection des sites d'intégration.

II.6. Topographie de l'intégration.

Concernant les rétrovirus, il apparaît que tout le génome hôte peut être le siège d'une intégration (Withers-Ward *et al.*, 1994). Cependant, l'intégration n'est pas complètement aléatoire (Shih *et al.*, 1988) et certaines régions semblent plus souvent concernées que d'autres. Plusieurs facteurs vont influencer la topographie de l'intégration.

II.6.1. Influence de la composition nucléotidique de l'ADN cible.

Les réactions d'intégration *in vitro* et l'analyse des séquences flanquantes de provirus clonés montrent que les rétrovirus s'intègrent préférentiellement dans des régions riches en nucléotides A/T (Shimotohno *et al.*, 1980 ; Grandgenett *et al.*, 1993 ; Fitzgerald *et al.*, 1994). Aucune spécificité importante des séquences cibles n'a pu être mise en évidence, au niveau nucléotidique. Cependant, il existe souvent un biais dans la composition nucléotidique des sites d'intégration, au niveau de la séquence dupliquée ou dans la séquence flanquante. Ainsi, pour le rétrovirus HIV-1, l'analyse de 112 séquences flanquantes publiées, met en évidence une faible séquence consensus au niveau de la séquence dupliquée de 5 pb : GT(A/T)AC (Vincent *et al.*, 1990 ; Stevens *et al.*, 1996 ; Carteau *et al.*, 1998).

Des études analogues réalisées chez d'autres rétrovirus, ont abouti à des résultats similaires. Ainsi, chez le MMLV, les deux positions centrales de la séquence répétée de 4 pb, sont occupés préférentiellement par des dinucléotides AA, TT ou AT (Pryciak *et al.*, 1992b). Une étude d'intégration concertée *in vitro*, réalisée chez l'AMV, montre qu'il existe également un biais dans la séquence flanquant le provirus (Fitzgerald *et al.*, 1994).

II.6.2. Régions préférentielles d'intégration.

La sélection des sites d'intégration *in vivo* n'est pas encore complètement caractérisée. Les premières études réalisées chez le MMLV, ont conduit à l'établissement d'un modèle, dans lequel les régions non compactées de la chromatine, correspondant principalement à des zones de transcription active, sont favorisées. En effet, l'analyse de provirus MMLV intégrés dans l'ADN de lignées cellulaires, indique que ce virus s'intègre préférentiellement dans des régions transcriptionnellement actives (Mooslehner *et al.*, 1990 ; Scherдин *et al.*, 1990). De plus, l'intégration semble favorisée à proximité des sites d'hypersensibilité à la DNase I de l'ADN cellulaire (Vijaya *et al.*, 1986 ; Rohdewohld *et al.*, 1987), dont la répartition dans le génome est généralement corrélée avec l'expression des gènes et les domaines de la chromatine "ouverte".

Une autre étude, réalisée *in vitro*, avec des extraits de cellules infectées par l'ALV ou avec l'IN purifiée seule, montre que l'intégration est favorisée au niveau des îlots CpG de l'ADN cellulaire (Kitamura *et al.*, 1992).

L'étude de l'intégration du rétrovirus HIV-1 indique que l'intégration survient fréquemment au niveau de séquences répétées du génome, de type LINE-1 (pour *Long Interspersed Nuclear Elements*) (Stevens *et al.*, 1994) ou Alu (Stevens *et al.*, 1996), ou encore au niveau de sites de clivage de la topoisomérase II (Howard *et al.*, 1993).

Un travail plus récent a porté sur l'analyse de 61 sites d'intégration HIV-1, générés par infection expérimentale. Par comparaison, une librairie contrôle de 104 séquences a été construite. Les fréquences respectives des unités de transcription, des éléments LINE ou des séquences Alu, n'étaient pas significativement différentes entre les sites d'intégration de HIV-1 et les séquences contrôles (Carteau *et al.*, 1998). L'intégration du rétrovirus HIV-1 par contre, est défavorisée dans les régions d'hétérochromatine centromérique, au niveau de séquences alphanucléotiques répétées. Ces zones d'hétérochromatine sont constituées d'ADN compacté, hautement organisé, et de ce fait, probablement inaccessible pour l'IN.

II.6.3. Le rôle des protéines fixées à l'ADN cible.

La fixation d'une protéine sur l'ADN peut bloquer, par encombrement stérique, l'intégration rétrovirale (Pryciak *et al.*, 1992c ; Bushman *et al.*, 1994b ; Goulaouic *et al.*, 1996 ; Katz *et al.*, 1996 ; Bushman *et al.*, 1997).

L'intégration *in vivo* ne se fait pas dans de l'ADN nu, mais préférentiellement dans des régions où l'ADN est associé à des protéines, comme les histones. Des protéines capables de courber l'ADN favorisent l'intégration. Ainsi, lorsque des minichromosomes de levure sont utilisés comme cibles, l'intégration est favorisée dans l'ADN enroulé au niveau des nucléosomes (Pryciak *et al.*, 1992b,c) par rapport à l'ADN nu. Certaines positions dans le nucléosome sont également privilégiées par l'IN.

La distribution des sites d'intégration révèle en effet une périodicité de 10 pb et un biais dans leur composition nucléotidique. La richesse en A/T et cette périodicité suggèrent que l'intégration a lieu préférentiellement au niveau du grand sillon de l'ADN, sur la face externe au corps du nucléosome. La face interne de l'ADN, en contact avec l'octamère d'histones, serait moins accessible à la machinerie d'intégration.

Des expériences plus récentes, utilisant d'autres types de préparations nucléosomiques, montrent que l'intégration a lieu préférentiellement au niveau de sites dans le grand sillon où celui-ci est le plus large et le plus accessible (Pruss *et al.*, 1994a,b).

L'utilisation d'autres complexes protéine-ADN comme cibles de l'intégration confirme ces observations. La courbure de l'ADN générée par la fixation d'autres protéines, comme la protéine IHF (Bor *et al.*, 1995) ou la protéine CAP (Muller *et al.*, 1994), entraîne également la sélection de sites d'intégration préférentiels, là où le grand sillon est le plus large. La courbure de l'ADN cible génère donc des sites d'intégration préférentiels sur la face externe de l'hélice, que l'ADN soit ou non associé à la surface d'une protéine.

La courbure de l'ADN cible est donc une des composantes majeures de la sélection des sites d'intégration par les rétrovirus.

II.6.4. Cas particulier de HTLV-1.

Plusieurs travaux ont montré que le rétrovirus HTLV-1 ne s'intégrait pas en un site précis de l'ADN hôte (Seiki *et al.*, 1984 ; Macera *et al.*, 1992 ; Cavrois *et al.*, 1995 ; Wattel *et al.*, 1995). Ces études reposaient sur des techniques d'hybridation in situ, d'hybridation d'ADN extraits d'hybridômes, de LMPCR ou d'IPCR.

L'ensemble du génome humain est constitué de nombreuses régions, d'environ 300 kb, les isochores, caractérisés par leur pourcentage en nucléotides G/C (Bernardi *et al.*, 1985). Ces isochores sont isolés par centrifugation à l'équilibre en gradient de Cs₂SO₄, en présence de ligands ADN spécifiques. Dans le génome humain, il existe 2 familles d'isochores L1 et L2, pauvres en nucléotides G/C (qui représentent 62% du génome), et 3 familles d'isochores H1, H2 et H3, riches en nucléotides G/C. L'ADN de plusieurs lignées cellulaires infectées par HTLV-1 a été fractionné par gradient de Cs₂SO₄ et la présence des séquences provirales dans les différentes fractions a été déterminée par blot de Southern. La distribution des provirus met en évidence le caractère "isopycnique" de l'intégration du rétrovirus HTLV-1 (Zoubak *et al.*, 1994) ; comme le virus de l'Hépatite B (HBV pour *Hepatitis B Virus*), le BLV et le RSV, HTLV-1 s'intègre dans des isochores de composition nucléotidique voisine de celle du provirus, avec un pourcentage en nucléotides G/C compris entre 39 et 54%. Cette étude montre que seules les régions riches en G/C du génome portent des provirus HTLV-1.

Une étude plus récente basée sur la LMPCR et sur l'analyse de 4 sites d'intégration issus de patients atteints d'ATLL, suggère une composition nucléotidique riche en nucléotides A/T (63%) dans les régions flanquantes du provirus. Il existe un biais dans la composition nucléotidique du fragment répété de 6 pb avec la présence constante d'un dinucléotide TT ou AA en une position précise de la séquence (Chou *et al.*, 1996). Le nombre de séquences étudiées est faible et l'analyse d'un nombre plus important de sites d'intégration permettrait de confirmer ces résultats.

Dans un autre travail basé sur l'analyse de 24 sites d'intégration dérivés de porteurs asymptomatiques, le pourcentage moyen en nucléotides A/T de l'hexamère dupliqué était de 59 % (Wattel *et al.*, 1995).

RESULTATS-DISCUSSION

PREMIERE PARTIE :

ETUDE DE LA COMPOSITION NUCLEOTIDIQUE DES SITES D'INTEGRATION DE HTLV-1 *IN VIVO*.

I.1. Introduction.

L'étude de la topographie de l'intégration des rétrovirus a jusqu'ici reposé sur trois grandes approches :

- L'étude de la localisation chromosomique des provirus intégrés.
- La recherche de régions préférentielles d'intégration.
- La recherche d'une composition nucléotidique particulière à proximité du site d'intégration ou dans une région plus étendue.

Si tout le génome hôte peut être le siège d'une intégration rétrovirale (Withers-Ward *et al.*, 1994), certaines régions semblent plus souvent concernées que d'autres et plusieurs facteurs influencent la topographie de l'intégration. La sélection des sites semble ainsi être favorisée au niveau de régions du génome transcriptionnellement actives, à proximité notamment des zones d'hypersensibilité à la DNase I (Vijaya *et al.*, 1986 ; Rohdewohld *et al.*, 1987). L'intégration est également conditionnée par la composition nucléotidique de l'ADN hôte, ainsi que par les interactions ADN cible-protéines (Pryciak *et al.*, 1992b,c ; Pruss *et al.*, 1994a,b ; Bor *et al.*, 1995). Il existe ainsi une préférence pour les régions riches en nucléotides A/T et l'enroulement de l'ADN autour des nucléosomes favorise l'intégration au niveau de certains sites (Pruss *et al.*, 1994a,b).

L'ensemble des études réalisées jusqu'à présent sur le mécanisme d'intégration des rétrovirus, a porté essentiellement sur des réactions d'intégration *in vitro* ou sur l'analyse de provirus intégrés dans l'ADN de lignées cellulaires.

Les données actuelles proviennent principalement de travaux dans lesquels la cible de l'intégration correspondait à une séquence connue, le plus souvent oligonucléotidique ou plasmidique (Brown *et al.*, 1987 ; Kitamura *et al.*, 1992 ; Pryciak *et al.*, 1992b,c ; Grandgenett *et al.*, 1993 ; Pruss *et al.*, 1994a,b ; Bor *et al.*, 1995 ; Balakrishnan *et al.*, 1997). *In vivo*, il existe un grand excès de sites potentiels par rapport au nombre de molécules provirales présentes dans le noyau de la cellule hôte. Or, jusqu'à présent, très peu de sites d'intégration provenant d'individus naturellement infectés ont été décrits.

L'absence de données précises sur la cible de l'intégration des rétrovirus *in vivo* nous a conduit à déterminer la composition nucléotidique des sites d'intégration de HTLV-1, à partir de l'ADN provenant d'individus infectés.

Pour cela, les sites d'intégration de 218 provirus, provenant de 29 individus infectés ont été amplifiés par IPCR ou LMPCR, puis clonés et séquencés.

Les objectifs de cette étude étaient :

- i) la recherche d'une éventuelle composition nucléotidique particulière des sites d'intégration de HTLV-1 dans les différentes conditions liées à l'infection.
- ii) la recherche d'une éventuelle corrélation entre le statut clinique et la composition nucléotidique des sites.
- iii) l'alignement des sites d'intégration sur les banques de données, afin de rechercher d'éventuelles identités avec des régions particulières du génome ou des séquences connues.

Pour éviter tout biais de composition de séquence induit par les étapes de PCR et de clonage utilisées pour isoler les sites d'intégration de HTLV-1, il était nécessaire de réaliser un témoin nous permettant d'évaluer la composition nucléotidique de l'ADN cellulaire non infecté, en utilisant la même stratégie expérimentale que celle utilisée pour les échantillons de patients ou de porteurs asymptomatiques.

I.2. Résultats.

Les 218 sites d'intégration flanquant le LTR 3' du provirus ont été isolés à partir de l'ADN de 29 échantillons infectés par HTLV-1. Les 29 individus infectés correspondaient à 12 ATLL, 10 TSP/HAM et 7 porteurs asymptomatiques. L'ADN des échantillons infectés a été extrait à partir de PBMCs, de ganglions lymphatiques, du liquide céphalo-rachidien ou de biopsies cutanées. L'ADN a ensuite été amplifié par LMPCR ou IPCR. Les produits de PCR exponentielle ont été clonés puis séquencés. La Figure 1R schématise la stratégie expérimentale utilisée pour l'obtention des sites d'intégration flanquant le LTR 3' de HTLV-1.

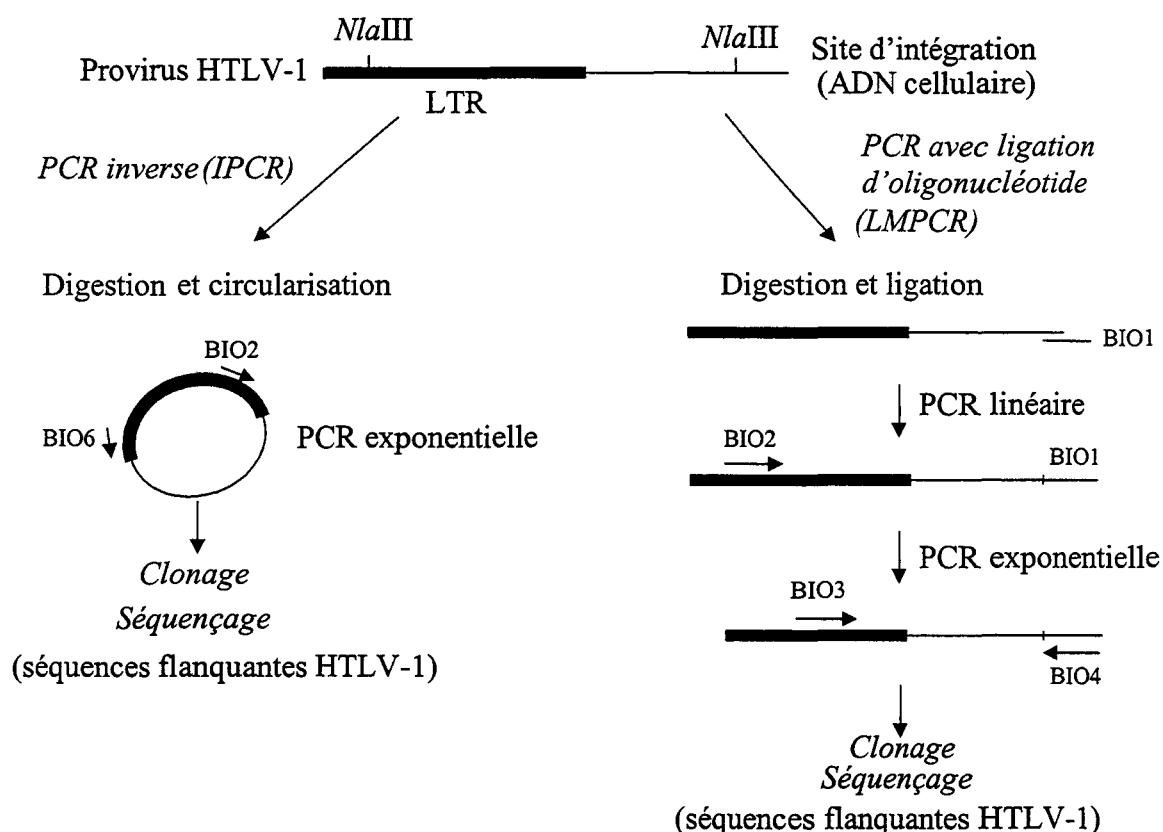


Figure 1R : Amplification de l'extrémité 3' du provirus et des sites d'intégration par LMPCR et IPCR.

La LMPCR a un seuil de détection de 20 copies sur 150 000 PBMCs. Cette technique a donc été utilisée pour les échantillons possédant une faible charge provirale, tels ceux provenant des porteurs asymptomatiques ou de certains individus atteints de TSP/HAM. L’IPCR, dont le seuil de détection est de 80 copies sur 150 000 PBMCs, a été utilisée pour les échantillons possédant une charge virale plus élevée, comme la plupart des échantillons de TSP et ceux provenant du tissu tumoral d’ATLL (Cavrois *et al.*, 1995).

Le clonage et le séquençage des produits de LMPCR et d’IPCR ont permis d’isoler 218 sites d’intégration distincts, flanquant le LTR 3’ de HTLV-1.

La distribution du nombre de sites d’intégration en fonction du statut clinique, figure dans le Tableau 1R.

Tableau 1R : Répartition du nombre d’individus et du nombre de sites d’intégration en fonction du statut clinique.

Statut clinique	Nombre d’individus étudiés	Nombre de sites d’intégration
ATLL	12	27
TSP/HAM	10	116
Asymptomatique	7	75

Comme la PCR et le clonage sont influencés par la composition nucléotidique et la taille de la séquence cible, il était nécessaire de réaliser un témoin nous permettant d’évaluer la composition nucléotidique de l’ADN cellulaire non infecté, à travers la même stratégie expérimentale que celle utilisée pour les échantillons de patients ou de porteurs asymptomatiques. La réalisation de ce témoin nous permettait ainsi d’évaluer la composition nucléotidique de l’ADN génomique non infecté, en évitant tout biais de séquence potentiellement induit par la stratégie expérimentale.

La stratégie expérimentale utilisée pour l’obtention des séquences contrôles est schématisée dans la Figure 2R.

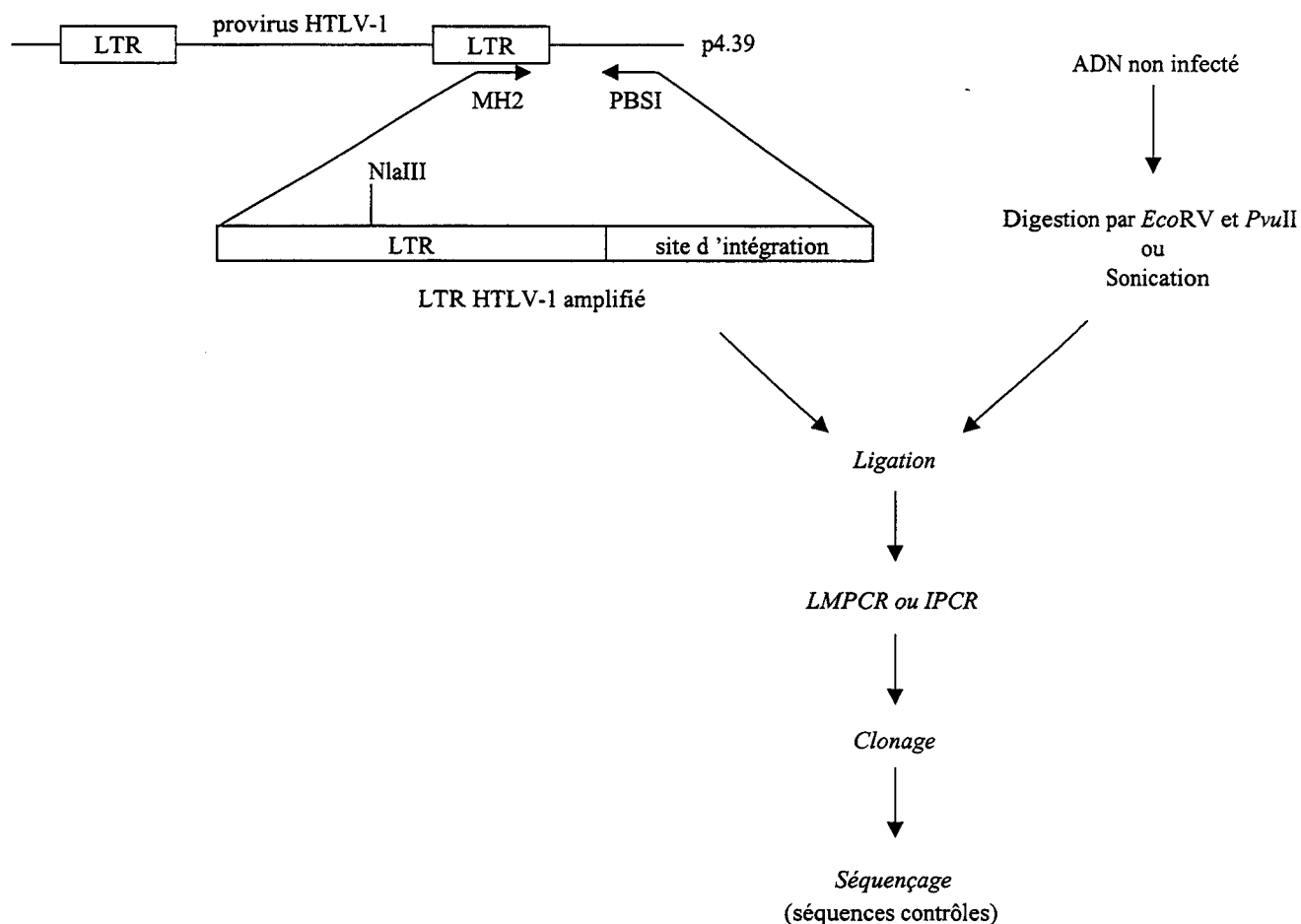


Figure 2R : Stratégie expérimentale utilisée pour la construction du témoin et l'amplification des séquences contrôles.

Un fragment de 1494 pb, généré par PCR, et comprenant l'extrémité 3' du LTR de HTLV-1, a été ligaturé à de l'ADN génomique non infecté, préalablement digéré par des enzymes de restriction ou par sonication. Ce produit de ligation a ensuite été amplifié par LMPCR ou IPCR et les produits d'amplification ont été clonés puis séquencés, de la même façon que les échantillons de patients infectés. Le clonage et le séquençage des produits d'amplification de ce témoin ont ainsi permis d'isoler 60 séquences contrôles différentes.

Toutes les séquences nucléotidiques obtenues ont été tabulées sur un logiciel de traitement de texte, puis converties et importées sur le logiciel SPSS1.6, afin de réaliser l'analyse statistique. Pour les sites d'intégration de HTLV-1 et les sites témoins, la séquence correspondant au LTR 3' a été déléetée avant l'analyse. Pour les séquences contrôles, les 31 premières bases correspondant au site d'intégration du plasmide p4.39, ainsi que les 3 bases

provenant des sites de restriction des enzymes *PvuII* et *EcoRV* ont également été déléetées avant l'analyse.

La composition nucléotidique des sites d'intégration de HTLV-1 a ensuite été déterminée puis comparée à celle des sites témoins.

Parallèlement, toutes les séquences de plus de 25 pb ont été alignées sur les banques de données GENE BANK et EMBL, ainsi que les banques de cDNA (dbEST), en utilisant les logiciels FASTA et BLAST, afin de rechercher d'éventuelles identités avec des séquences connues, répertoriées. L'utilisation de Repeat Masker (<http://ftp.genome.washington.edu/cgi-bin/RepeatMasker>) a permis l'alignement des séquences avec les différents éléments répétés du génome. Il existe en effet plusieurs catégories d'éléments répétés dans le génome. Les séquences SINE (pour *Short Interspersed Elements*) et les séquences LINE (pour *Long Interspersed Elements*) sont très nombreuses. La famille d'éléments SINE la plus représentée est la famille ALU, dont les séquences font environ 300 pb ; il en existe 300 000 à 900 000 dans le génome humain. Les séquences LINE sont majoritairement des éléments de la famille L1, et sont caractérisées par une séquence consensus d'environ 6,4 kb.

Nous avons choisi de manière arbitraire, une taille minimale de 25 pb afin d'éviter toute identité non spécifique, qui pourrait résulter de l'alignement de séquences plus courtes avec les séquences des banques de données. Les sites d'intégration de HTLV-1 ou les séquences témoins qui présentaient une identité d'au moins 80% sur 80% de leur longueur avec des séquences connues ou répertoriées, ont été retenues.

I.2.1. Les sites d'intégration de HTLV-1 sont riches en nucléotides A/T.

Deux cent dix huit sites d'intégration de HTLV-1 ont été séquencés. La taille moyenne des séquences était de 84 pb et variait de 6 à 365 pb. Nous avons obtenu entre 1 et 44 sites d'intégration par individu infecté.

Soixante sites témoins ont été séquencés. La taille moyenne des séquences était de 53 pb et variait de 12 à 150 pb.

Les compositions nucléotidiques des sites d'intégration de HTLV-1 et des sites témoins ont été déterminées puis comparées pour voir s'il existait un biais dans la composition des sites d'intégration du virus. Le Tableau 2R résume les résultats obtenus pour les deux types de séquences.

Tableau 2R : Compositions nucléotidiques des sites d'intégration de HTLV-1 et des sites témoins. Les compositions en nucléotides A/T et en nucléotides A, G, C ou T sont exprimées en pourcentage.

	Nucléotides A/T (%)	A (%)	C (%)	G (%)	T (%)
Témoin	51	25	26	23	26
HTLV-1	57	29	22	21	28

Le calcul du pourcentage de nucléotides A/T pour les séquences HTLV-1 et les séquences témoins a montré que les sites d'intégration de HTLV-1 étaient riches en nucléotides A/T par rapport aux sites témoins. La comparaison de cette composition à chaque position par un test T apparié a montré que cette différence était statistiquement significative (57% versus 51%, $p=10^{-4}$). Le pourcentage de nucléotides A/T variait de 30 à 80% pour les sites HTLV-1 et de 33 à 79% pour les sites témoins.

Le pourcentage en nucléotides A/T des séquences contrôles était de 51%, une valeur significativement inférieure à celle décrite jusqu'à ce jour pour l'ADN génomique humain, qui est de 58% (Swartz *et al.*, 1962). Les techniques de PCR et de clonage utilisées dans cette étude sous-estimaient donc le contenu en nucléotides A/T, d'où l'importance de réaliser un témoin.

La Figure 3R représente la distribution en nucléotides A/T le long des 40 premières positions, pour les sites HTLV-1 et les sites témoins. Comme le montre la figure, la richesse en A/T des sites d'intégration était supérieure à celle des séquences contrôles pour 31 positions sur 40.

La richesse en nucléotides A/T était de 57%, 59% et 55% pour les sites d'intégration provenant respectivement des patients atteints d'ATLL, de TSP et des porteurs asymptomatiques. Un test statistique montrait qu'il n'y avait pas de différence significative entre les différents statuts cliniques.

Dans une seconde étape, nous avons comparé la composition en nucléotides A, G, C ou T des sites HTLV-1 avec celle des sites témoins. Les résultats figurent dans le Tableau 2R. Seuls les pourcentages en A et C étaient significativement différents entre les sites HTLV-1 et les sites témoins (29 versus 25% pour le pourcentage en nucléotides A ; 22 versus 26% pour le pourcentage en résidus C, $p=0,001$).

Les études réalisées *in vitro* avec d'autres rétrovirus ont montré qu'il existait un biais dans la composition des sites d'intégration, principalement au niveau de la séquence dupliquée. Ainsi, il existe une faible séquence consensus au niveau de la séquence dupliquée de 5 pb du rétrovirus HIV-1 (Vincent *et al.*, 1990 ; Stevens *et al.*, 1996 ; Carteau *et al.*, 1998). Chez le MMLV, les deux positions centrales de la séquence répétée de 4 pb, sont occupées préférentiellement par des dinucléotides AA, TT ou AT (Pryciak *et al.*, 1992b). Nous nous sommes donc plus particulièrement intéressés à la composition nucléotidique des 6 premières bases des sites d'intégration de HTLV-1, correspondant à la séquence d'ADN génomique dupliquée.

Au niveau de l'hexamère dupliqué, le pourcentage de nucléotides A/T était de 56% et variait de 17 à 100%. Les compositions en nucléotides A, C, G et T étaient respectivement de 30, 24, 20 et 26%. Seul le contenu en nucléotides T était significativement plus élevé en dehors de l'hexamère dupliqué (28 versus 26%, $p=0,039$).

Il n'y avait pas de différence significative dans la distribution des autres nucléotides, entre l'hexamère dupliqué et les séquences situées en dehors de la duplication.

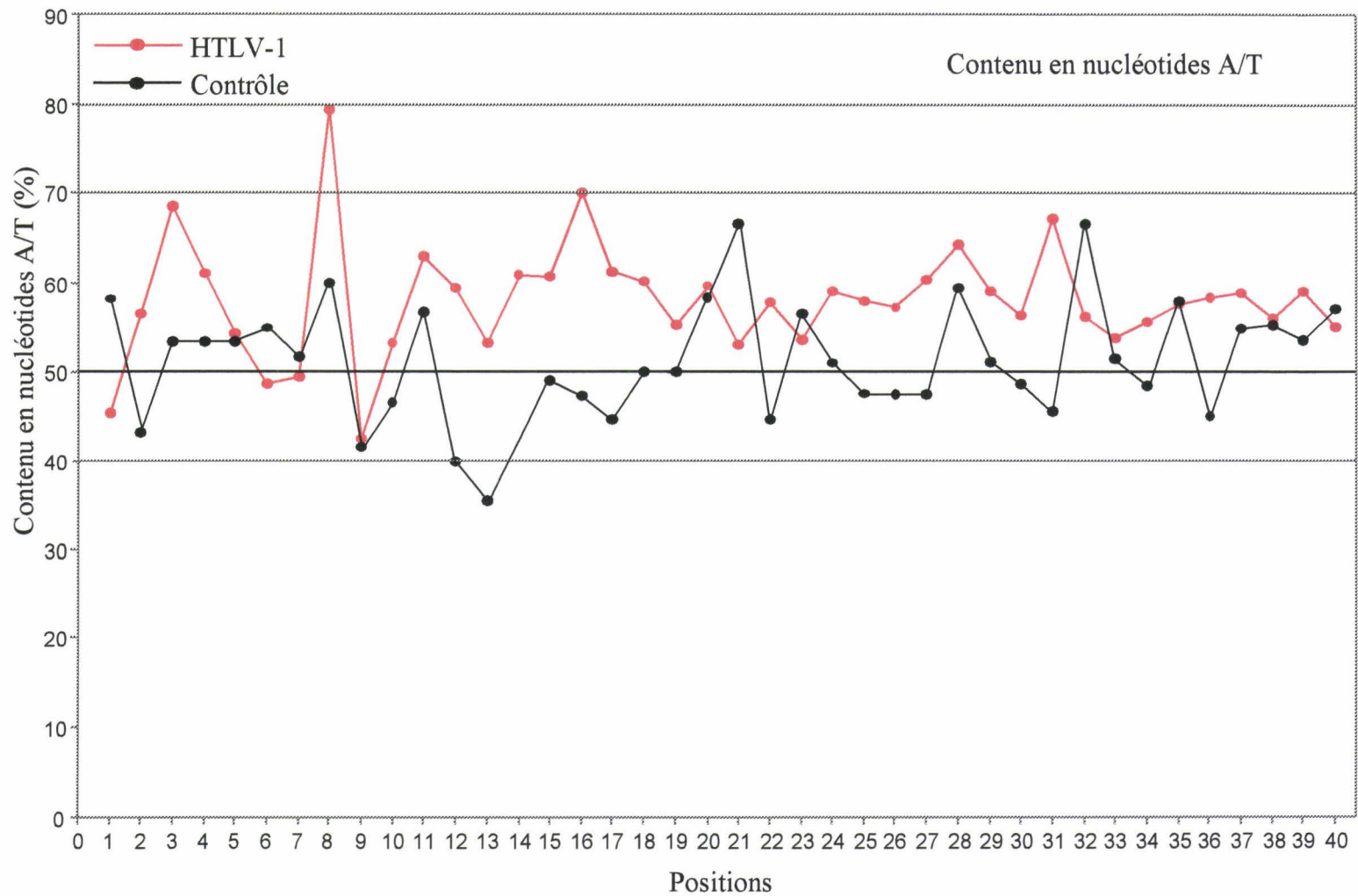


Figure 3R : Distribution du pourcentage de nucléotides A/T le long des 40 premières positions des sites d'intégration de HTLV-1 et des sites témoins. Les séquences sont alignées à partir de la première base de l'hexamère. Chaque point représente le pourcentage de A/T à chaque position de la séquence.

I.2.2. Distribution non homogène de la composition nucléotidique des sites d'intégration de HTLV-1.

La Figure 3R représente la distribution en nucléotides A/T le long des 40 premières positions, pour les sites HTLV-1 et les sites témoins.

Les positions 8 et 16 sont préférentiellement occupées par des nucléotides A ou T (79% et 70% respectivement) alors que les positions 1 et 9 sont occupées préférentiellement par des résidus G ou C (54% et 57% respectivement).

La Figure 4R représente la distribution du pourcentage de nucléotides A/T pour les sites HTLV-1 et les sites témoins. Elle permet de repérer les différentes valeurs aberrantes et facilite les comparaisons entre plusieurs distributions.

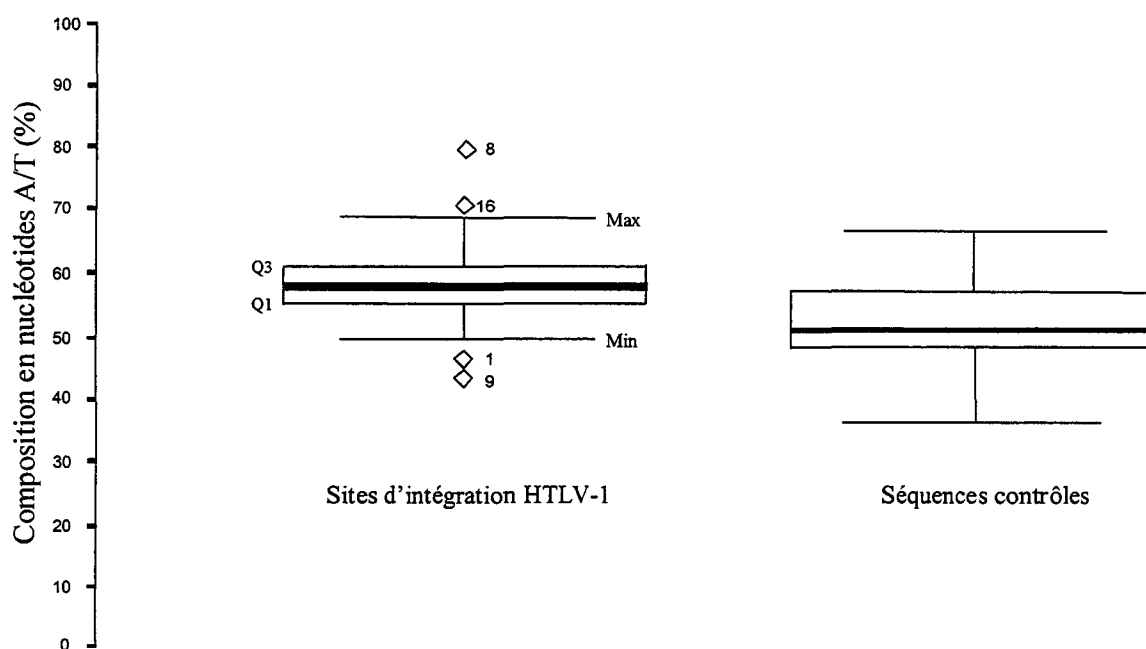


Figure 4R : Distribution du pourcentage de nucléotides A/T le long des 40 premières positions des sites HTLV-1 et des sites témoins. Le trait noir correspond à la médiane du pourcentage de nucléotides A/T. Le graphique résume la distribution à partir de ses valeurs extrêmes, de ses quartiles et de sa médiane. L'étendue (max-min) de la distribution à chacune des extrémités est indiquée par deux lignes horizontales. Le rectangle représente la distribution du pourcentage de nucléotides A/T entre le premier quartile (Q1) et le 3^{ème} quartile (Q3) et la hauteur de la boîte correspond donc à l'écart interquartile (Q3-Q1). Les valeurs supérieures à 1,5 fois l'écart interquartile sont représentées par des losanges.

Comme le montre la Figure 4R, les positions 1, 8, 9 et 16 correspondaient à des valeurs extrêmes. Pour les autres positions, la distribution en nucléotides A/T était monotone, sans grandes fluctuations. Aucune valeur extrême n'était détectée au niveau des séquences contrôles.

Cette analyse a également montré que la dispersion du pourcentage de nucléotides A/T était plus étroite que celle obtenue pour les sites témoins.

Nous avons ensuite analysé la distribution du pourcentage de chaque nucléotide A, G, C ou T le long des 40 premières positions des sites d'intégration de HTLV-1. La Figure 5R représente cette distribution du pourcentage de nucléotides A, C, G et T pour les sites d'intégration de HTLV-1. La figure montre que la richesse en nucléotides A/T en position 8 était due à un large excès de résidus A. La richesse en nucléotides A/T observée en position 16 était due à la présence de séquences riches en résidus A et T. La richesse en G/C observée en position 1 était due à un excès de nucléotides G associé à un manque de résidus T. En position 9, elle était due à un excès de résidus C associé à un manque de résidus A. La Figure 5R montre qu'il existait également un excès de résidus A au niveau des positions 4 et 28.

La distribution du pourcentage de chaque nucléotide pour les sites HTLV-1 et les séquences témoins a montré que seules les positions 1, 8, 9 et 16 correspondaient à des valeurs extrêmes. Les positions 4 et 28, bien que riches en nucléotides A, ne correspondaient pas à des valeurs extrêmes. Enfin, aucune valeur extrême n'était détectée au niveau des séquences contrôles (Résultats non montrés).

Les positions 1, 8, 9 et 16 définissent donc une séquence "consensus" caractérisée par une composition nucléotidique particulière. Plus de 34% des séquences HTLV-1 isolées observaient cette distribution pour au moins 3 de ces 4 positions (C/G, A, C, et A/T aux positions 1, 8, 9 et 16, respectivement), contre seulement 15,3% pour les séquences témoins. Cette différence était statistiquement significative ($p < 10^{-4}$). Ces résultats ont montré qu'il existait une distribution nucléotidique hétérogène dans le voisinage immédiat du provirus.

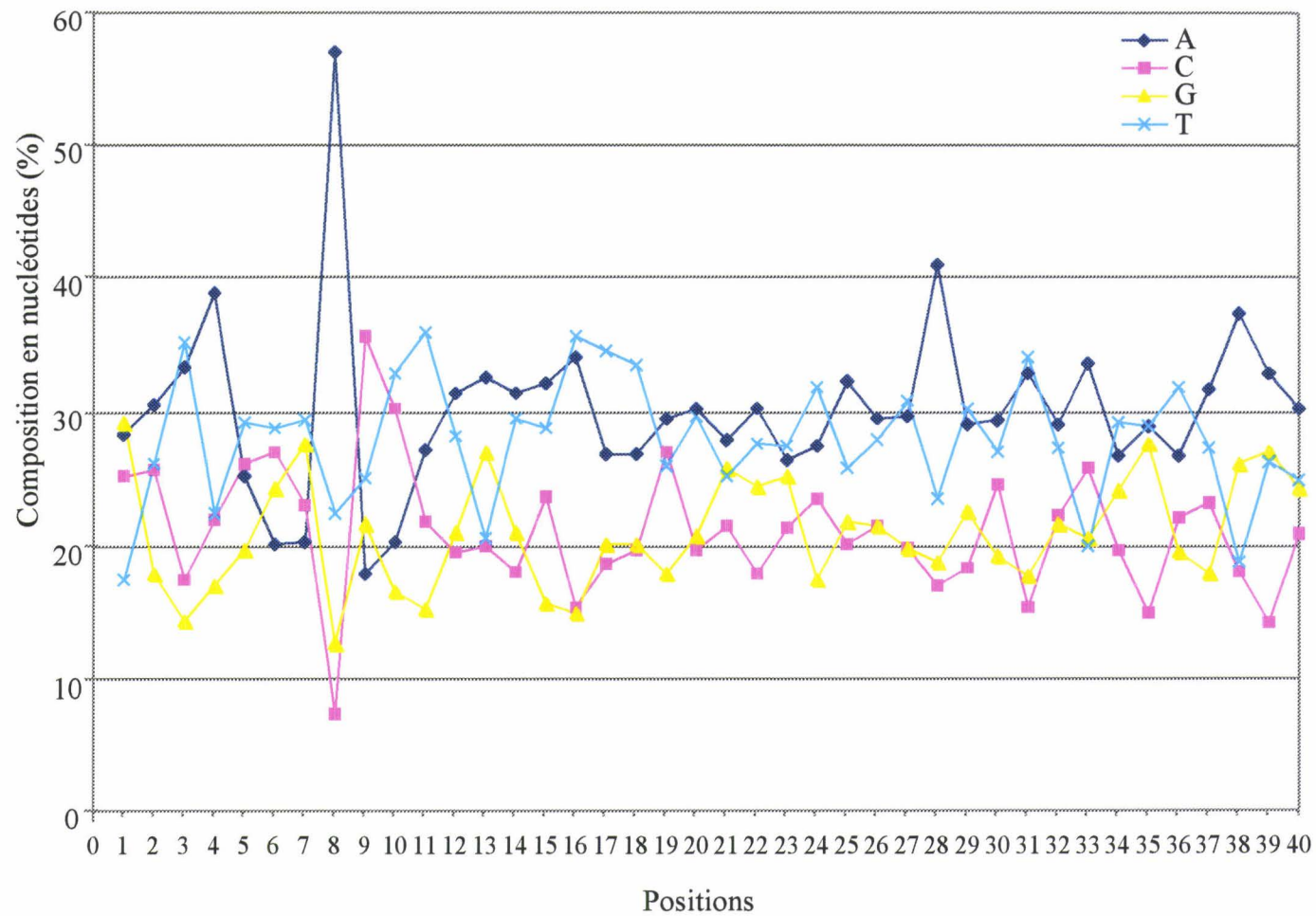


Figure 5R : Distribution du pourcentage de nucléotides A, C, G ou T le long des 40 premières positions des sites d'intégration de HTLV-1. Les séquences sont alignées à partir de la première base de l'hexamère. Chaque point représente le pourcentage de nucléotides A, C, G ou T à chaque position de la séquence.

Les sites d'intégration de HTLV-1 comportaient également de longs motifs de nucléotides A/T. En effet, sur 194 sites HTLV-1 > à 10 pb, 17% présentaient des motifs de nucléotides A/T supérieurs à 6 pb, contre 2% seulement pour les séquences contrôles. Cette différence était significative ($p=0,002$).

Une étude réalisée chez le MMLV a montré que le virus s'intégrait préférentiellement dans des îlots CpG (Kitamura *et al.*, 1992) ; la méthylation de ces îlots CpG n'inhibe pas l'intégration du provirus et crée au contraire des sites préférentiels d'intégration. La méthylation aurait, pour les auteurs de ce travail, un effet indirect sur la structure locale de l'ADN, favorisant ainsi la reconnaissance de l'ADN cible par l'IN. Nous avons donc cherché à mettre en évidence la présence d'îlots CpG au niveau des sites d'intégration de HTLV-1. L'analyse de la composition nucléotidique des sites d'intégration montrait qu'il n'y avait pas de différence significative entre les séquences témoins et les sites d'intégration, en ce qui concerne les motifs C/G ou CpG (0,5 versus 1% et 3 versus 2%, respectivement).

I.2.3. Identification des séquences ciblées par l'intégrase de HTLV-1 *in vivo*.

Les sites d'intégration de HTLV-1 et les séquences contrôles de plus de 25 pb ont été alignés sur les banques de données GENE BANK, EMBL, dbEST (banques de cDNA) et Repeat Masker. L'objectif de cette étude était de pouvoir identifier les séquences ciblées par le virus *in vivo* et de rechercher d'éventuelles régions préférentielles d'intégration. Les séquences répertoriées dans les banques de données peuvent être classées en 3 catégories distinctes : les éléments répétés, les "unités de transcription" (dbEST) et les séquences génomiques non codantes et non répétitives. Le Tableau 3R représente la distribution du nombre et du pourcentage de séquences HTLV-1 et de séquences témoins qui présentaient une identité significative avec des séquences connues ou répertoriées.

Tableau 3R : Distribution du nombre et du pourcentage de séquences présentant une identité significative avec des séquences répertoriées dans les banques de données, pour les sites HTLV-1 et les sites témoins de plus de 25 pb. Les séquences anonymes correspondent aux séquences pour lesquelles aucune identité significative n'a été détectée. Pour chaque type de séquence figure le pourcentage de nucléotides A/T correspondant.

Type de séquences	Pourcentage en nucléotides A/T	Séquences HTLV-1 (%)	Séquences contrôles (%)
Séquences répétées	55	59 (33)	13 (30)
Eléments LINE	60	20 (11)	3 (7)
Eléments SINE	52	20 (11)	6 (14)
Eléments RLE	53	11 (6)	3 (7)
ADN transposon	55	1 (1)	1 (2)
ADN satellite		3 (2)	0 (0)
Autre		4 (2)	0 (0)
Séquences non codantes non répétitives	58	15 (8)	1 (2)
Unités de transcription	55	10 (6)	3 (7)
ADNc		8 (5)	3 (7)
Gène		2 (1)	0 (0)
Total		84 (47)	17 (39)
Séquences Anonymes	59	94 (53)	27 (61)

Cent soixante dix-huit séquences correspondant aux sites d'intégration de HTLV-1 ont une taille supérieure à 25 pb. Sur ces 178 sites d'intégration, 84 séquences (soit 47%) présentaient une identité significative (au moins 80% d'homologie sur 80% de la longueur du site) avec une séquence répertoriée dans les banques de données. Ces 84 séquences correspondaient à 15 séquences non codantes et non répétitives, 10 unités de transcription (8 ADNc et 2 gènes), et 59 éléments répétés.

De même, sur les 44 séquences contrôles de plus de 25 pb, 17 (soit 39 %) présentaient une identité significative avec des séquences connues. Ces séquences correspondaient

à 1 séquence non codante et non répétitive, 3 unités de transcription (3 ADNc), et 13 éléments répétés.

Le Tableau 4R récapitule les différentes séquences connues et non répétitives, pour lesquelles les sites HTLV-1 et les sites témoins présentaient une identité significative. Les autres séquences HTLV-1 et les autres séquences contrôles n'avaient pas d'identité significative avec les séquences disponibles à l'heure actuelle dans les banques de données.

Pour les deux types de séquences, témoins et HTLV-1, la fréquence des identités avec des séquences répétées dans le génome était significativement plus élevée que celle observée avec des unités de transcription (33% versus 6% pour les sites HTLV-1 et 30% versus 7% pour les séquences contrôles) (Tableau 3R). Le Tableau 3R montre également qu'il n'existait pas de différence de répartition des identités entre les séquences HTLV-1 et les séquences témoins. Ces résultats suggèrent que toutes les régions du génome hôte sont donc accessibles au rétrovirus HTLV-1. Il n'y a pas de préférence pour les régions de transcription ou les éléments répétés du génome. Une étude récente basée sur l'analyse de 61 sites d'intégration de HIV-1, générés par infection expérimentale et isolés par PCR, a montré que le virus HIV-1 ne privilégiait pas non plus une région particulière du génome (Carteau *et al.*, 1998).

Cette étude a par ailleurs montré que le virus HIV-1 ne s'intégrait jamais dans des séquences satellites centromériques. Nous avons donc regardé s'il en était de même pour le rétrovirus HTLV-1. Après alignement des sites d'intégration de HTLV-1 sur les banques de données, 3 sites d'intégration étaient identiques à 3 séquences satellites. Le premier site d'intégration est le clone 242 et correspondait à de l'ADN satellite α localisé au niveau des chromosomes 13, 14 et 21 (Vissel *et al.*, 1992). Le site d'intégration 67 présentait une forte identité avec une séquence répétitive d'ADN, de type satellite (identification HSU10629). Enfin, le clone 43 correspondait à de l'ADN satellite β , localisé au niveau des bras courts des chromosomes acrocentriques humains (identification HUMBSATC). Aucune homologie avec des séquences satellites n'a été mise en évidence pour les séquences témoins. HTLV-1 peut donc s'intégrer dans des séquences de type satellite, à la différence du rétrovirus HIV-1.

Nous avons ensuite regardé s'il existait une corrélation entre le type de séquence et le statut clinique. Quelque soit le statut clinique, nous n'observons pas de différence significative quant aux types de séquences pour lesquels les sites HTLV-1 présentaient une identité importante (Tableau 4R).

Le contenu en nucléotides A/T a été déterminé pour les séquences anonymes, les éléments répétés, les "unités de transcription" et les séquences non codantes et non répétitives. Le Tableau 3R donne donc le pourcentage en nucléotides A/T des différents types de séquences qui présentaient une identité significative avec les séquences témoins ou les séquences HTLV-1. Les différences de richesse en nucléotides A/T étaient significatives ($p=0,0063$). Il existait une faible corrélation entre la richesse en A/T des différents types de séquences identifiées et leur fréquence avec les séquences HTLV-1 et les séquences contrôles. Ainsi, la proportion de séquences HTLV-1 qui possédaient une identité significative avec des séquences répétées de type LINE (60% de nucléotides A/T) était plus importante que pour les séquences contrôles (11 versus 7%). Il en était de même pour les séquences non codantes et non répétitives, dont la richesse en A/T était de 58% (8 versus 2%). Toutefois, ces différences n'étaient pas statistiquement significatives.

Tableau 4R : Séquences HTLV-1 et séquences témoins présentant une identité significative avec des séquences non répétitives.

a. Les clones tumoraux correspondent aux clones cellulaires malins de l'ATLL.

b. Numéro de la séquence délivré par les banques de données.

Numéro de la séquence	Echantillon	Taille (pb)	Hexamère dupliqué	Nucléotides A/T (%)	Séquences homologues Type	N° Ascession ^b
ATLL						
78	tumoral ^a	200	TTATTC	60	ADNc	HSU68704
84	non tumoral	141	AGCAAG	66	ADNc	N53238
74	tumoral	99	TCTTTC	60	Gène (Human P protein, exon 23)	HSPPROT23
85	non tumoral	28	CCTCTC	68	Non codante non répétitive	AL034422
91	tumoral	36	CAGCTG	50	Non codante non répétitive	AQ223501
Porteurs asymptomatiques						
10	PBMCs	25	TCCGCA	36	ADNc	AA451666
18	PBMCs	32	ACCCGC	41	ADNc	AA631969
27	PBMCs	25	GCAACT	60	ADNc	AC002036
39	PBMCs	25	GCAAAA	52	ADNc	AC004583
58	PBMCs	45	TTATGT	69	Non codante non répétitive	AC006054
23	PBMCs	34	GTTATA	71	Non codante non répétitive	AC006227
34	PBMCs	25	GAGAAC	52	Non codante non répétitive	AQ306143
37	PBMCs	101	CTGTGG	59	Non codante non répétitive	B17581
20	PBMCs	38	GGTGTG	42	Non codante non répétitive	HS232D4
TSP/HAM						
233	PBMCs	26	TAATAG	62	ADNc	H72803
231	PBMCs	69	CTTGGT	51	ADNc	W57727
224	PBMCs	365	GCTAGG	61	Gène (alpha enolase, exon 1)	HSENOAL1
220	PBMCs	29	CATATG	55	Non codante non répétitive	AC003693
216	LCR	25	GCTAGG	48	Non codante non répétitive	AC004505
183	PBMCs	163	ACATTT	59	Non codante non répétitive	AC005881
274	PBMCs	52	CTGAGG	44	Non codante non répétitive	AL021937
275	PBMCs	75	TCAGTC	55	Non codante non répétitive	AL022345
223	PBMCs	144	GAGAAT	70	Non codante non répétitive	AL031599
246	PBMCs	30	TCAATC	67	Non codante non répétitive	AL031683
204	LCR	41	TAAAGT	78	Non codante non répétitive	HUAC002307
Séquences contrôles						
109		64		58	ADNc	AI001768
137		25		72	ADNc	AA969105
151		30		43	ADNc	AA909212
110		45		42	Non codante non répétitive	HS431P23

I.3. Discussion.

L'analyse de 218 sites d'intégration de HTLV-1, isolés à partir de l'ADN d'individus infectés, nous a permis d'aborder les caractéristiques des séquences ciblées par le rétrovirus *in vivo*.

Pour cette étude, il était nécessaire de comparer la composition nucléotidique des sites d'intégration de HTLV-1 avec la composition nucléotidique de l'ADN génomique non infecté. Nous avons donc réalisé un témoin, capable d'évaluer la composition nucléotidique de l'ADN génomique non infecté, à travers notre stratégie expérimentale. Le pourcentage en nucléotides A/T des séquences contrôles obtenues, était de 51%, une valeur significativement inférieure à celle décrite jusqu'à ce jour pour l'ADN génomique humain, qui est de 58% (Swartz *et al.*, 1962). Les techniques de PCR et de clonage utilisées dans cette étude sous-estimaient donc le contenu en nucléotides A/T. Il était par conséquent indispensable de comparer la composition nucléotidique des sites d'intégration de HTLV-1 avec celle d'un témoin ayant subi les mêmes étapes de PCR et de clonage.

HTLV-1 est caractérisé par son mode réplcatif particulier, basé sur l'expansion clonale des cellules infectées. Le nombre de clones cellulaires T circulants reflète donc le nombre d'événements d'intégration ayant eu lieu. Le nombre d'événements d'intégration a été estimé à partir d'un porteur asymptomatique pour lequel 20 clones différents avaient été mis en évidence (Cavrois *et al.*, 1995). Le nombre de clones en expansion est plus élevé chez les patients atteints d'ATLL et les patients atteints de TSP/HAM. Toutefois, la collection de sites d'intégration analysée dans cette étude est largement représentative de la population clonale présente *in vivo*.

L'alignement des sites d'intégration de HTLV-1 sur les banques de données montrent que le virus s'intègre au hasard dans le génome de la cellule hôte. Il n'y a pas de différence de répartition des identités entre les sites HTLV-1 et les sites témoins. La plupart des régions du génome hôte sont donc accessibles au rétrovirus. Ainsi, il n'existe pas de préférence pour des régions de transcription actives du génome, contrairement à ce qui avait été suggéré pour le MMLV *in vitro* (Mooslehner *et al.*, 1990 ; Scherdin *et al.*, 1990).

De plus, contrairement au rétrovirus HIV-1, HTLV-1 peut s'intégrer au niveau de séquences d'ADN satellites centromériques constituées de chromatine compactée. Par contre, le pourcentage de séquences HTLV-1 présentant une identité significative avec des séquences répétées du génome est le même que celui précédemment obtenu pour le rétrovirus HIV-1 (Carteau *et al.*, 1998). Toutefois, ce pourcentage ne diffère pas entre les sites d'intégration et les séquences contrôles.

Dans cette étude, 6% des sites d'intégration de HTLV-1 présentent une identité significative avec un gène ou une séquence d'ADN complémentaire. Le virus HTLV-1 est donc susceptible d'altérer l'expression d'un gène, en s'intégrant dans l'ADN cellulaire. La charge provirale est élevée chez les individus infectés par HTLV-1, notamment chez les patients atteints d'ATLL. Par conséquent, le pourcentage observé n'est pas négligeable et correspond à un nombre important de clones en expansion.

Le clone 224 était virtuellement identique avec la région promotrice du gène humain de l' α -énolase (identification HSENOAL1, Tableau 4R). Ce gène est normalement exprimé dans les lymphocytes T et a été isolé à partir de la lignée T Jurkat. L'intégration du provirus a eu lieu 146 pb en 5' de la boîte TATA, l'éloignant ainsi de certains motifs de transcription comme la boîte CCAAT, les sites AP1 ou PEA2 nécessaires à la machinerie transcriptionnelle. La Figure 6R représente l'alignement du site d'intégration du clone 224 avec une partie de la séquence correspondant au gène de l' α -énolase. L'intégration du provirus pourrait inhiber la transcription du gène et l'expression de la protéine correspondante. L'altération d'un gène précieux pour le lymphocyte T, pourrait ainsi contribuer à l'expansion clonale des cellules infectées et à l'émergence d'un clone malin.

```

HSENOL  AGAGGATAAAAAATGCTAGGGGTTAAGGAAGAAAAGCAGGTTTTGCAGCTGACCA-GAAGA
          :: : ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
CL224   -----GGTTAAAGGAAGAAAAGCAGGTTTTGCAGCTGACCAAGAAGA
          10      20      30      40

          1080      1090      1100      1110      1120      1130
HSENOL  AATAATTAAATTCAAGTACATATTTGAACACCTCTTTACAAGGAAATGAGGAAAATCAGA
          ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
CL224   AATAATTAAATTCAAGTACATATTTGAACACCTCTTTACAAGGAAATGAGGAAAATCAGA
          50      60      70      80      90      100

          1140      1150      1160      1170      1180      1190
HSENOL  TGTTTTAGTTCTCTTACAGACCAGCAGTAATGTT-GAAAAATATAAA-AGTAGCCGTAAC
          :::::::::::::::::::::: :::::::::::::::::::::: ::::::::::::::::::::::
CL224   TGTTTTAGTTCTCTTACAGGCCAGCAGTAATGTTTGAAAAATATAAACAGTAGCCGTAAC
          110      120      130      140      150      160

          1200      1210      1220      1230      1240      1250
HSENOL  AGAAAACGAAGTGATCATTTTCAGCCCTCCCTTGGAG-TGGGGGGATATATTGCCGAGCTG
          :::::::::::::::::::::: :: :::::::::::::::::::::: ::::::::::::::::::::::
CL224   AGAAAACGAAGTGATCATCTCA-CCCTCCCTTGGAGGTGGGGGGATATATTGCCGAGCTG
          170      180      190      200      210      220

          1260      1270      1280      1290      1300      1310
HSENOL  AGTTTTAATGGCTATCAGTTCTTGATCAGAAAAAGTGAAATCACTCCAACCTCCTTCCGTAT
          ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
CL224   AGTTTTAATGGCTATCAGTTCTTGATC-----
          230      240      250

```

Figure 6R : Alignement du site d'intégration du clone 224 (CL224) avec une partie du gène de l' α -énolase (HSENOL).

Nos résultats suggèrent que la plupart des régions du génome hôte sont donc accessibles au rétrovirus HTLV-1. Cependant, il existe un biais dans la composition nucléotidique des sites d'intégration de HTLV-1 *in vivo*. Le calcul du pourcentage de nucléotides A/T montre en effet que le virus HTLV-1 s'intègre préférentiellement dans des régions riches en nucléotides A/T (57 % versus 51 %, $p=10^{-4}$). Une préférence pour une intégration dans des régions riches en A/T avait déjà été décrite dans la littérature pour les transposons (Sandmeyer *et al.*, 1990), les éléments Ty de levure (Oyen *et al.*, 1983), l'AMV (Fitzgerald *et al.*, 1992 ; Grandgenett *et al.*, 1993), le SNV (Shimotohno *et al.*, 1980) ou encore le MMLV (Pryciak *et al.*, 1992c). Pour le rétrovirus HTLV-1, l'analyse de 4 sites d'intégration provenant de patients atteints d'ATLL avait révélé un contenu en nucléotides A/T de 63% (Chou *et al.*, 1996). D'autre part, l'ADN de plusieurs lignées cellulaires infectées par HTLV-1 a été fractionné par gradient de Cs_2SO_4 et la présence des séquences provirales dans les différentes fractions a été déterminée par blot de Southern. La distribution des provirus montre que HTLV-1 s'intègre dans des

isochores de composition nucléotidique voisine de celle du provirus, avec un pourcentage en nucléotides G/C compris entre 39 et 54% (Zoubak *et al.*, 1994). Cette prédisposition pour les régions riches en nucléotides A/T du génome caractérisent donc l'intégration *in vitro* et *in vivo*. La richesse en nucléotides A/T pourrait faciliter la séparation des deux brins de l'ADN cible au niveau de l'hexanucléotide, favorisant ainsi l'intégration.

On observe également une distribution hétérogène du pourcentage de nucléotides tant au niveau de l'hexamère dupliqué qu'en dehors de la duplication. Des séquences consensus et un biais dans la composition nucléotidique ont déjà été décrits pour d'autres rétrovirus, comme le MMLV et le HIV-1, essentiellement au niveau des séquences dupliquées (Vincent *et al.*, 1990 ; Pryciak *et al.*, 1992b ; Fitzgerald *et al.*, 1994 ; Stevens *et al.*, 1996 ; Carteau *et al.*, 1998). En ce qui concerne HTLV-1, une faible séquence consensus a été mise en évidence dans cette étude, au niveau de l'hexamère dupliqué où seule la position 1 représente une valeur extrême. Dans notre travail, la composition nucléotidique n'est pas homogène le long des séquences flanquant le provirus HTLV-1, et on observe un biais important aux positions 1, 8, 9 et 16, occupées préférentiellement par des résidus G/C, A, C et A/T respectivement. Ces quelques résidus conservés situés à proximité du LTR 3' proviral pourraient favoriser la fixation du complexe de préintégration sur l'ADN cible ou permettre la fixation de protéines spécifiques, cellulaires ou virales, susceptibles d'interagir avec la machinerie d'intégration, et donc potentiellement impliquées dans le processus d'intégration *in vivo*.

Chez le rétrovirus AMV, des expériences d'intégration concertée *in vitro* ont permis la mise en évidence d'une distribution nucléotidique hétérogène au voisinage immédiat du provirus. En effet, le pourcentage de nucléotides G/C est faible en position 8 et significativement élevé en position 9. L'étude ne précise pas la nature exacte des nucléotides présents à ces deux positions (Fitzgerald *et al.*, 1992). La taille de la duplication chez l'AMV est de 6 pb, comme pour le rétrovirus HTLV-1. On retrouve donc un profil proche de celui obtenu pour les sites d'intégration de HTLV-1 aux positions 8 et 9. La présence des deux nucléotides A et C représente le palindrome du dinucléotide CA présent à l'extrémité 3' des LTR proviraux. Ces séquences palindromiques pourraient être reconnues spécifiquement par une protéine d'origine virale ou cellulaire, favorisant ainsi la réaction d'intégration.

La présence de longs motifs de nucléotides A/T caractérisent également les sites d'intégration de HTLV-1. De telles séquences sont fréquemment intrinsèquement courbées (Kahn *et al.*, 1992).

Toutes ces données suggèrent que la structure de l'ADN cible joue un rôle important dans la sélection des sites d'intégration du virus.

Cette hypothèse est renforcée par le fait que certains sites d'intégration isolés dans cette étude, présentent des identités importantes avec des séquences du génome qui sont naturellement courbées (Kahn *et al.*, 1992). La composition nucléotidique de l'ADN cible influence la structure intrinsèque de l'ADN. Elle confère à la molécule d'ADN une conformation particulière plus ou moins propice à l'intégration. Ainsi, des molécules présentant une structure courbée en solution, caractérisées par de courts motifs oligod(A). oligod(T), sont préférentiellement utilisées par le rétrovirus HIV-1 *in vitro* (Pruss *et al.*, 1994b). D'autres études, pour la plupart également réalisées *in vitro*, ont montré que la courbure de l'ADN favorisait l'intégration des rétrovirus au niveau de certains sites. La présence d'une distribution inhomogène de la composition nucléotidique le long des sites d'intégration de HTLV-1 (notamment aux positions 1, 8, 9 et 16) pourrait correspondre à des sites de fixation de certaines protéines cellulaires comme les protéines HMG, capables de tordre la double hélice et de conférer à la molécule d'ADN cible une conformation propice à l'intégration, accessible à la machinerie enzymatique. Ainsi, la courbure de l'ADN, induite par la composition nucléotidique de la séquence cible ou par la fixation de protéines cellulaires spécifiques, favorise l'intégration des provirus au niveau de certains sites. La Figure 7R tente de représenter les différents facteurs mis en évidence dans ce travail, qui influencent l'intégration de HTLV-1 *in vivo*.

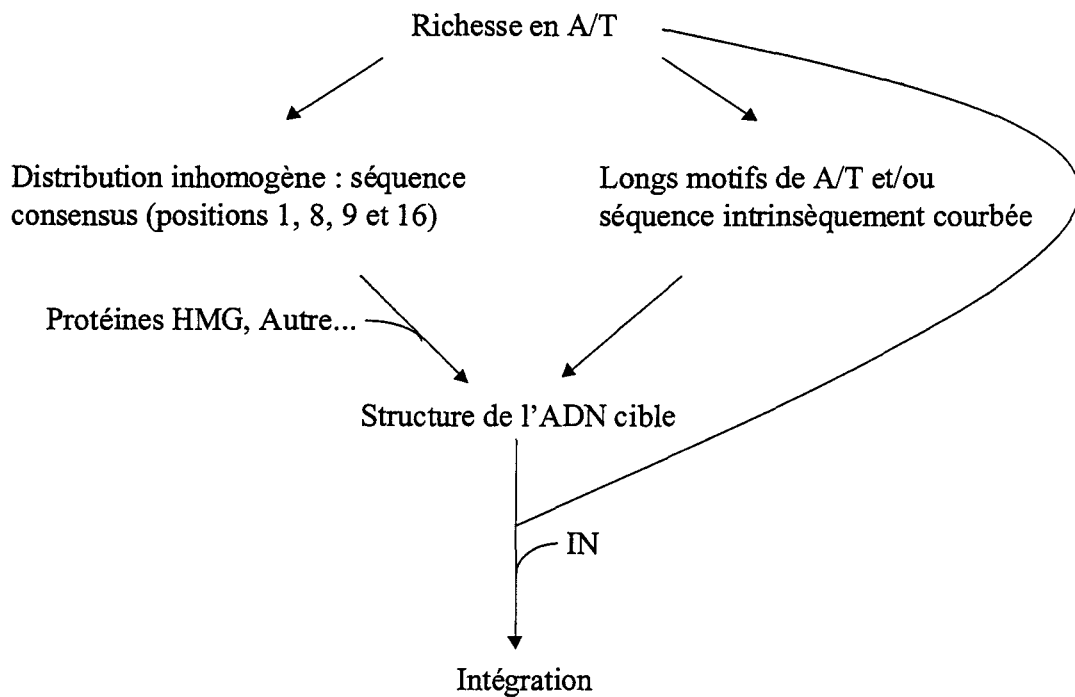


Figure 7R : Schéma récapitulatif des différents facteurs influençant l'intégration de HTLV-1 *in vivo*.

Les résultats de cette étude montrent donc que la spécificité de l'intégration *in vivo* de HTLV-1 semble déterminée principalement par des caractéristiques structurales de l'ADN cible.

DEUXIEME PARTIE :
INFLUENCE DE LA STRUCTURE DE L'ADN HOTE SUR LE CHOIX DE
L'INTEGRASE DE HTLV-1 : ETUDE *IN VITRO*.

II.1. Introduction.

Les relations entre la composition nucléotidique, la structure de l'ADN cible et l'intégration de HTLV-1, suggérées dans l'étude précédente, nous ont conduit, dans une première étape, à aborder les caractéristiques structurales influençant la réaction de transfert de brin catalysée par l'IN de HTLV-1, *in vitro*. Dans une deuxième étape, nous avons évalué l'impact de mutations présentes au niveau de l'extrémité 3' du provirus sur le rendement de l'intégration.

Des expériences d'intégration *in vitro* chez d'autres rétrovirus comme le HIV-1, l'ASV, ou le MMLV, ont montré qu'il existait une corrélation entre l'étape de transfert de brin et la structure de l'ADN cible (Milot *et al.*, 1994 ; Pruss *et al.*, 1994b ; Katz *et al.*, 1998). Concernant HTLV-1, très peu d'expériences ont été réalisées jusqu'à présent. Seules des expériences d'autointégration ont montré que l'IN de HTLV-1 pouvait catalyser les réactions de clivage et de transfert de brin à partir d'un grand nombre de substrats, correspondant aux extrémités 3' des LTR de HIV-1, de HTLV-1, de HTLV-2 et du MMLV (Balakrishnan *et al.*, 1997).

Une étude récente réalisée au laboratoire a montré que l'essentiel de la variabilité génétique de HTLV-1 résultait de mutations somatiques du provirus, acquises au cours de la prolifération des cellules infectées, et non d'erreurs commises par la RT (Mortreux *et al.*, soumis). Des expériences d'autointégration et de mutagenèse ont précédemment montré que des substitutions de l'extrémité 3' de l'oligonucléotide donneur pouvaient significativement modifier le rendement de la réaction de transfert de brin. Ayant isolé 8 séquences RU5 mutées dans leurs 54 dernières bases, nous avons évalué l'impact de ces mutations sur la réaction d'intégration *in vitro*.

Pour aborder ces questions, nous avons mis au point une technique permettant une analyse topographique et quantitative précise de l'intégration du rétrovirus HTLV-1 *in vitro*. Nous nous sommes plus particulièrement intéressés à la réaction de transfert de brin catalysée par l'IN purifiée de HTLV-1.

Cette stratégie expérimentale nous a ainsi permis, d'une part de confirmer l'influence de la structure de l'ADN cible sur le choix de l'IN de HTLV-1, et d'autre part d'étudier l'influence des mutations du LTR 3', survenues *in vivo* sur la réaction de transfert de brin.

II.2. Mise au point de la stratégie expérimentale.

La stratégie expérimentale utilisée est représentée de manière schématique dans la Figure 8R.

Des réactions d'intégration *in vitro* ont été réalisées, en utilisant 2 oligonucléotides cibles possédant des structures totalement différentes en solution. Ces deux séquences d'ADN cibles bicaténares sont riches en nucléotides A/T et existent *in vivo*. Leurs structures respectives ont été décrites en détail par l'équipe de Wolffe (Hayes *et al.*, 1991), et ces molécules cibles ont été testées pour l'intégration du rétrovirus HIV-1 (Pruss *et al.*, 1994b).

La première cible (Courbe) comprend de courts motifs répétés comprenant chacun 5 pb oligo(dA).oligo(dT) suivies de 4 à 5 pb G ou C. Ces motifs sont répétés toutes les 10-11 pb le long de la séquence Courbe. Cette séquence nucléotidique particulière confère à la molécule une structure courbée en solution.

A l'opposé, la deuxième cible (Rigide), constituée par un enchaînement de longs motifs oligo(dA) et oligo(dT), possède une structure droite et rigide en solution. Les deux oligonucléotides cibles sont encadrés par deux séquences connues et conservées de 18 et 20 pb, afin de permettre l'amplification des produits d'intégration.

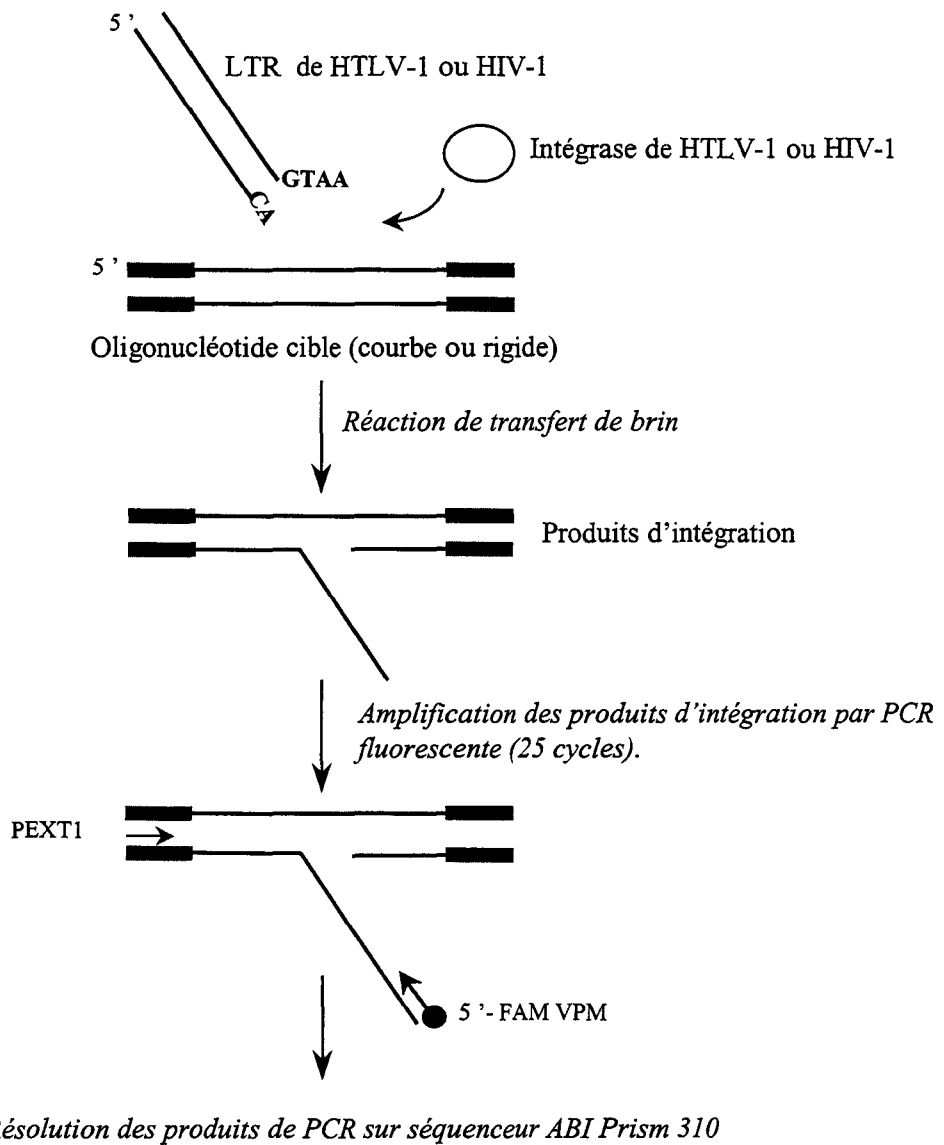


Figure 8R : Stratégie expérimentale utilisée pour les réactions d'intégration *in vitro*.

Les deux oligonucléotides Courbe et Rigide utilisés comme cibles dans cette étude ont été analysés à l'aide du programme Bend-it (http://www2.icgeb.trieste.it/~dna/bend_it.html). Ce programme permet de prédire la courbure d'une séquence nucléotidique. La Figure 9R représente les résultats obtenus.

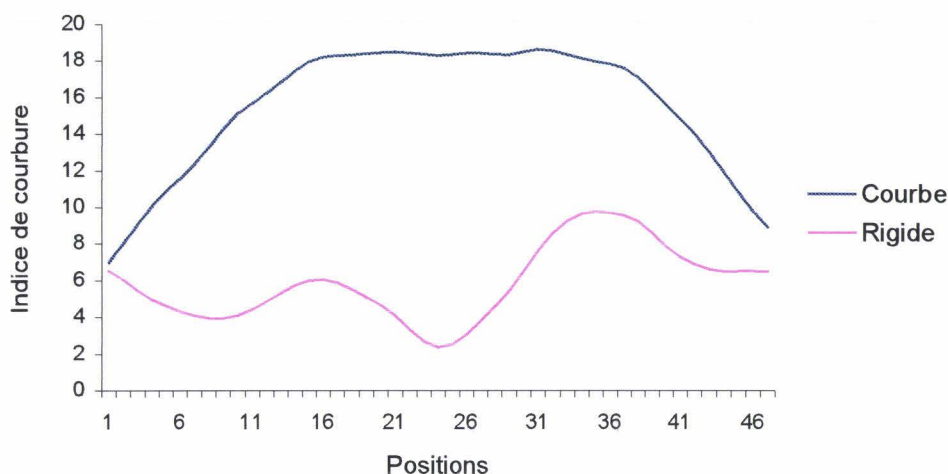


Figure 9R : Analyse des oligonucléotides Courbe et Rigide par le logiciel Bend-it. Seules les 50 positions centrales comprises entre les deux séquences conservées de 18 et 20 pb sont représentées.

La moyenne de l'indice de courbure estimé par le logiciel à chaque position est de $17,4 \pm 0,3$ pour la séquence Courbe et de $4,1 \pm 0,3$ pour la séquence Rigide. Cette différence est statistiquement significative ($p < 10^{-4}$).

Des oligonucléotides correspondant aux 54 et 74 dernières pb des LTR de HTLV-1 (séquence de référence ATK1) et de HIV-1 (séquence de référence à partir de l'isolat PV22 ; identification K02083) respectivement, ont été utilisés comme donneurs dans la réaction d'intégration. Les deux dernières bases à l'extrémité 3' de chaque LTR, au niveau du dinucléotide CA, ont été omises lors de la synthèse des oligonucléotides, pour faciliter la réaction de transfert de brin. Cette étape mime la réaction de clivage catalysée par l'IN qui précède la réaction de transfert de brin. L'utilisation de substrats LTR "pré-clivés" permet d'obtenir des rendements de transfert de brin supérieurs à ceux obtenus lors de l'utilisation d'oligonucléotides non clivés.

Les réactions d'intégration entre l'oligonucléotide donneur et la cible ont été réalisées à l'aide des IN de HTLV-1 ou de HIV-1, exprimées dans *E.coli* puis purifiées en conditions dénaturantes, comme décrit dans le Matériel et Méthodes. Ces deux enzymes ont été caractérisées précédemment par d'autres équipes (Vincent *et al.*, 1993 ; Balakrishnan *et al.*, 1997 ; Muller *et al.*, 1999).

Les produits d'intégration ont ensuite été amplifiés par PCR à l'aide d'une amorce fluorescente complémentaire de l'extrémité 5' de chaque LTR et d'une amorce complémentaire de la séquence conservée, présente à l'extrémité des oligonucléotides cibles Courbe et Rigide. Cette réaction de PCR permettait d'amplifier exclusivement les fragments d'ADN provenant de l'intégration du LTR de HTLV-1 ou de HIV-1 dans le brin négatif de l'oligonucléotide cible. Les produits amplifiés ont ensuite été séparés par électrophorèse en gel de polyacrylamide dénaturant sur séquenceur d'ADN, puis analysés grâce au logiciel Genescan (résolution de 0,15 pb) (Perkin Elmer ABI Prism 310).

Ce modèle permet de préciser rapidement la topographie de l'intégration, le nombre d'intégrants obtenus après transfert de brin et la fréquence moyenne d'intégration pour chaque cible utilisée.

II.2.1. Expression et purification des intégrases de HTLV-1 et de HIV-1.

Les IN de HTLV-1 et de HIV-1, portant à leur extrémité C-terminale une séquence de 6 résidus histidine, ont été exprimées dans la souche BL21DE3 d'*E.coli*, purifiées en conditions dénaturantes sur une colonne de nitrilotriacétate de nickel, puis renaturées par dialyse, comme décrit dans le Matériel et Méthodes.

L'expression et la pureté des protéines de fusion obtenues, ont été vérifiées sur un gel de polyacrylamide dénaturant à 12%. La coloration du gel de migration au Bleu de Coomassie a révélé la présence, dans les deux cas, d'une bande relativement pure, dont la taille correspondait à la taille attendue de la protéine d'intérêt (~32 kDa).

La Figure 10R montre les résultats obtenus lors de la purification de l'IN de HTLV-1, après migration des différentes fractions protéiques en gel de polyacrylamide. Un profil analogue a été obtenu pour la protéine HIV-1 (résultat non montré). Cependant, la quantité de protéine purifiée semblait relativement plus faible comparée à celle obtenue pour la protéine HTLV-1. Ce résultat a été confirmé après avoir effectué le dosage des différentes solutions protéiques contenant les enzymes d'intérêt et obtenues après la renaturation. Les concentrations des deux IN ont été évaluées par spectrophotométrie d'absorption à 485 nm, en utilisant comme protéine de référence, pour la réalisation de la gamme étalon, la serum albumine bovine.

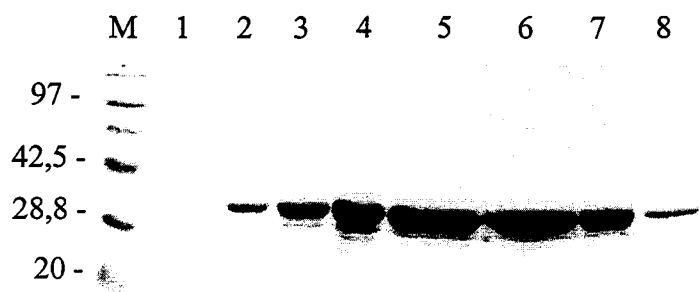


Figure 10R : Purification de l'IN de HTLV-1. Migration, en gel de polyacrylamide coloré au bleu de Coomassie, des fractions protéiques 1 à 8, obtenues après lavage de la colonne par le tampon A à pH 4,5. Les tailles des protéines du marqueur M sont exprimées en kDa.

II.2.2. Test de l'activité enzymatique des IN de HTLV-1 et de HIV-1 purifiées.

Afin de vérifier l'activité enzymatique des deux protéines purifiées, nous avons réalisé des expériences d'intégration *in vitro*, en utilisant des concentrations décroissantes de protéines, comme décrit dans le Matériel et Méthodes. Nous avons utilisé la séquence Courbe comme molécule d'ADN cible. L'IN de HTLV-1 a été testée en utilisant le LTR sauvage de HTLV-1 et l'IN de HIV-1 a été testée en utilisant le LTR sauvage de HIV-1. Les séquences des différents oligonucléotides utilisés dans cette étude figurent dans le Tableau 2M (Matériel et Méthodes, page 152). Les différents produits d'intégration ont été amplifiés par PCR fluorescente puis analysés sur séquenceur d'ADN, comme décrit dans le Matériel et Méthodes.

Les résultats obtenus pour la protéine HTLV-1 sont montrés dans la Figure 11R.

Sur l'électrophérogramme, les pics majeurs de fluorescence correspondent à des événements d'intégration. La Figure 11R montre que l'IN de HTLV-1 est active et donc compétente pour l'intégration. La protéine purifiée catalyse la réaction de transfert de brin entre le LTR sauvage de HTLV-1, clivé au niveau du dinucléotide CA, et l'oligonucléotide cible Courbe.

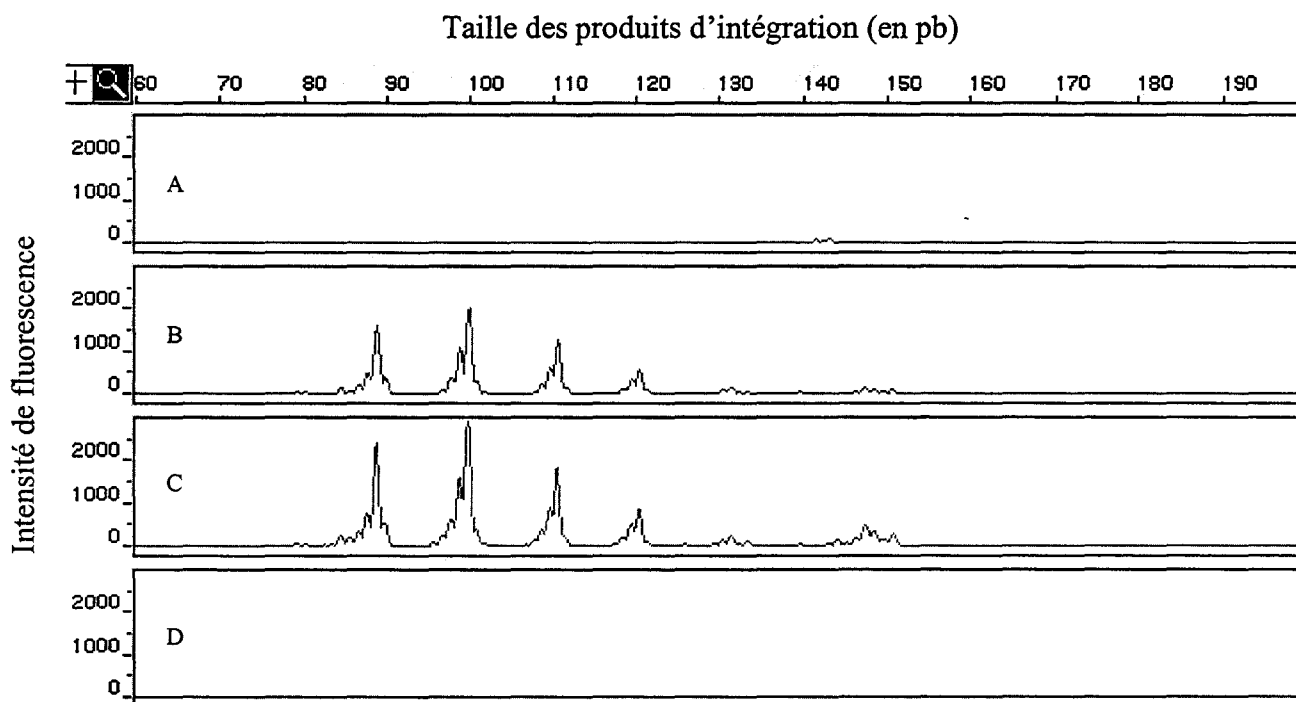


Figure 11R : Détermination de l'activité enzymatique optimale de l'IN de HTLV-1. (A) Témoin négatif, sans IN, (B) Dilution au 1/10^{ème} de la solution protéique de départ, (C) Dilution au 1/20^{ème}, (D) Dilution au 1/50^{ème}.

L'IN de HTLV-1 a une efficacité optimale lorsqu'elle est utilisée à une concentration de 23 µg/mL, soit une concentration de 0,72 µM. Cette concentration protéique sera utilisée dans toutes les expériences ultérieures. La protéine HIV-1 est moins efficace que l'IN de HTLV-1, puisqu'il faut utiliser une concentration en protéine de 400 µg/mL pour obtenir un rendement maximal, soit une concentration de 12,5 µM.

Chaque lot de protéine purifiée possède sa propre activité enzymatique. L'efficacité de la protéine purifiée va dépendre de plusieurs paramètres : le taux d'expression de la protéine dans la bactérie, la qualité de la purification et de la renaturation, la présence ou non de protéases. Par conséquent, toutes les expériences d'intégration doivent être réalisées en utilisant une protéine issue d'un même lot de purification, afin d'éviter tout biais dans l'interprétation des expériences. Si l'intensité moyenne de fluorescence des pics peut varier significativement d'un lot d'enzyme à l'autre, leur position reste toujours la même.

L'intensité de fluorescence des pics d'amplification est reproductible d'une expérience à l'autre, pour un même lot d'enzyme.

Comme le montre la Figure 11R, les pics principaux sont tous précédés d'un signal de moindre intensité. Cet aspect est caractéristique du phénomène de "bégalement" observé lors de l'amplification enzymatique de séquences répétées, comme les microsatellites. En effet, la présence de séquences répétées, à l'instar de la séquence Courbe, entraîne la formation d'artefacts de PCR au cours de l'amplification enzymatique, qui auraient pour origine le "bégalement" de la polymérase observé au cours de l'élongation. Ce phénomène a été largement décrit et discuté dans des travaux précédents (Walsh *et al.*, 1996 ; Miller M.J *et al.*, 1997). Par conséquent, seuls les pics d'intensité maximale seront pris en compte lors de l'analyse ultérieure des résultats.

L'analyse du profil d'intégration obtenu avec l'IN de HTLV-1 révèle une périodicité de 10-11 pb, reproduisant la répétition de la séquence cible. L'obtention et la signification de ce profil seront détaillés et discutés plus loin dans ce chapitre.

II.2.3. Détermination de la taille réelle des fragments amplifiés.

Tous les produits de PCR ont été analysés sur séquenceur d'ADN, à l'aide du logiciel Genescan (Perkin Elmer). Les amorces 5'Fam-VPM et PEXT1 permettent d'amplifier les fragments d'ADN qui résultent de l'intégration du LTR donneur dans le brin négatif de la séquence cible.

Les fragments d'ADN amplifiés sont soumis à une électrophorèse en gel de polyacrylamide dénaturant, en présence d'un marqueur de taille connue. La comigration de ce marqueur avec les différents produits de PCR permet d'attribuer une taille précise à chaque fragment amplifié, à 0,15 pb près.

Toutefois, la composition nucléotidique particulière des ampligènes peut influencer leur migration dans le gel dénaturant. Ainsi, la taille déduite de l'électrophorèse par le séquenceur, peut être biaisée et ne pas correspondre à la taille réelle du fragment amplifié.

Pour confirmer l'assignation de taille calculée par le logiciel, les produits de PCR obtenus lors de la détermination de l'activité enzymatique optimale de l'IN de HTLV-1 (Figure 11R), ont été clonés dans le phage M13mp18. Les différents clones moléculaires obtenus ont été

séquencés. Le clonage et le séquençage des produits de PCR ont été réalisés comme décrit dans le Matériel et Méthodes.

En ce qui concerne HTLV-1, nous avons séquencé 23 matrices M13, correspondant à 4 sites d'intégration différents. Les séquences obtenues figurent dans le Tableau 5R.

Tableau 5R : Séquences nucléotidiques des 4 clones obtenus après clonage et séquençage des produits d'intégration *in vitro*. Les séquences en bleues correspondent aux amorces 5'Fam-VPM et PEXT1 utilisées pour l'amplification. La séquence en noir correspond à l'extrémité 3' du LTR de HTLV-1. La séquence en rouge correspond à la séquence de l'oligonucléotide cible Courbe.

Nom du clone	Taille en pb	Séquence nucléotidique (5'-3')
IN 3	93	GCGAATTTCGAGCCAGCGACAGCCCATTCTATAGCACTCTCCAGGAGAGAACTTAGTACACACTATCCGAGGGGTCATCGGAACCAGAATTCGC
IN 12	104	GCGAATTTCGAGCCAGCGACAGCCCATTCTATAGCACTCTCCAGGAGAGAACTTAGTACACAATTTTAGGCCGCTATCCGAGGGGTCATCGGAAC CAGAATTCGC
IN 21	115	GCGAATTTCGAGCCAGCGACAGCCCATTCTATAGCACTCTCCAGGAGAGAACTTAGTACACAATTTTAGGGCCATTTTAGGCCGCTATCCGAGG GTCATCGGAACCAGAATTCGC
IN 1	125	GCGAATTTCGAGCCAGCGACAGCCCATTCTATAGCACTCTCCAGGAGAGAACTTAGTACACAATTTTAGGCCATTTAGCGCCATTTTAGGCCGC TATCCGAGGGGTCATCGGAACCAGAATTCGC

Les matrices M13 correspondantes ont été amplifiées par PCR, à l'aide des amorces 5' Fam-VPM et PEXT1. Les produits de PCR résultants ont été déposés sur séquenceur d'ADN puis analysés par le logiciel Genescan. La Figure 12R montre les résultats obtenus après l'électrophorèse de quelques ampligènes sur le séquenceur d'ADN.

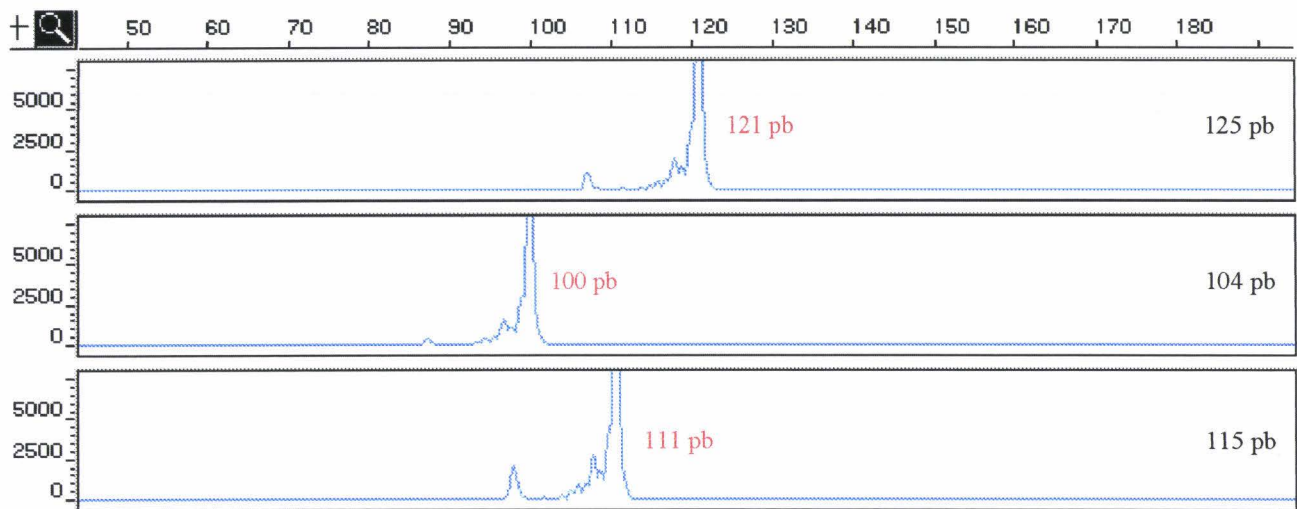


Figure 12R : Vérification de l'assignation de taille calculée par le logiciel après comigration des produits de PCR avec le marqueur de taille. La taille réelle déduite du séquençage figure en noir. La taille attribuée par le logiciel après l'électrophorèse figure en rouge.

Cette expérience montre qu'il existe, pour chaque ampligène, un décalage de 4 pb par rapport à la taille prévue par le séquençage des matrices. Ce décalage est probablement dû à la composition nucléotidique particulière de l'oligonucléotide cible qui entraîne un retard de migration dans le gel. Nous avons tenu compte de ce décalage dans l'analyse ultérieure des résultats.

II.3. La structure de l'ADN cible influence le rendement de la réaction de transfert de brin catalysée par l'IN de HTLV-1 ou l'IN de HIV-1.

II.3.1. Influence de la structure de l'ADN cible sur l'intégration de HTLV-1 *in vitro*.

L'influence de la structure de l'ADN cible sur le rendement de la réaction de transfert de brin a été évaluée en utilisant comme cibles, les oligonucléotides Courbe et Rigide, qui possèdent des structures totalement différentes en solution (Figure 9R).

La Figure 13R représente les résultats obtenus après les réactions de transfert de brin catalysées par l'IN de HTLV-1, entre le LTR sauvage de HTLV-1, et les séquences Courbe ou Rigide.

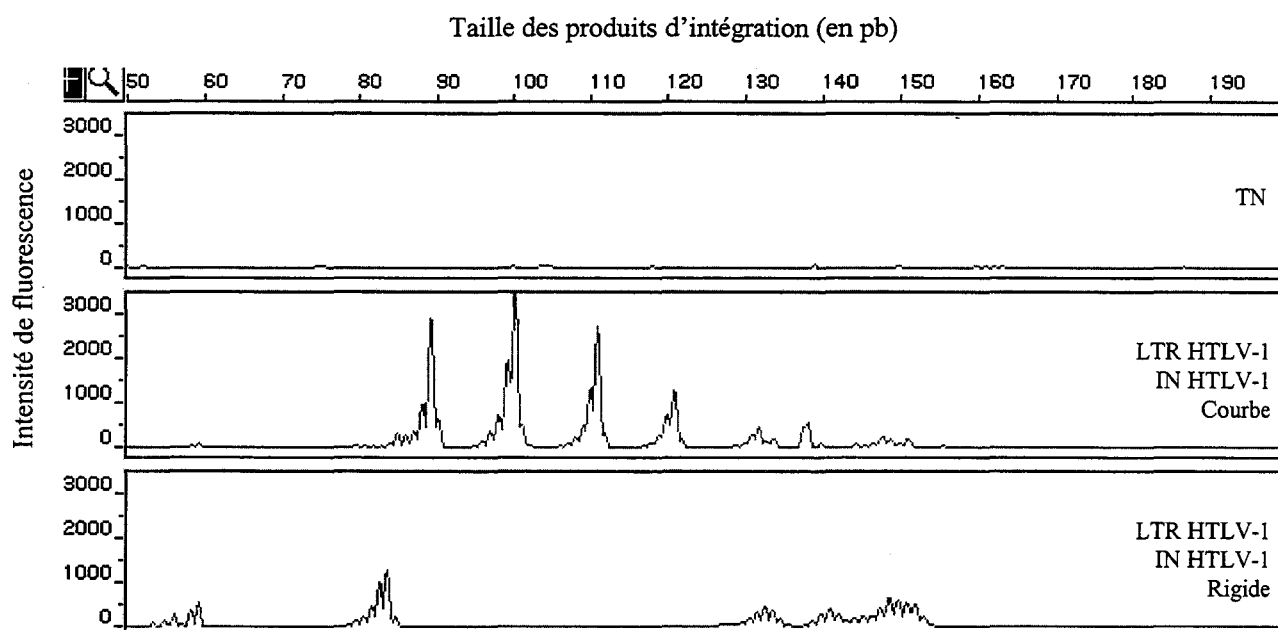


Figure 13R : Réactions d'intégration à l'aide des oligonucléotides Courbe et Rigide. L'IN de HTLV-1 catalyse la réaction de transfert de brin entre le LTR sauvage de HTLV-1 et l'oligonucléotide Courbe ou Rigide. TN : Témoin négatif, sans IN.

Un électrophérogramme similaire a été obtenu avec l'IN de HIV-1 (résultat non montré).

Chaque pic majeur de fluorescence correspond à un événement d'intégration. Pour les deux IN, les événements d'intégration sont significativement plus nombreux dans la séquence Courbe que dans la séquence Rigide, avec une périodicité de 10-11 bp entre les différents pics d'intégration, reproduisant la répétition de la séquence Courbe.

En utilisant comme oligonucléotide cible, la séquence Rigide, on observe des pics d'intégration aux extrémités de la cible. Ces pics correspondent à des événements d'intégration dans les séquences conservées de 18 et 20 pb, flanquant la séquence Rigide et utilisées pour l'amplification. Aucun événement majeur d'intégration n'a été observé dans la séquence Rigide proprement dite, au niveau des motifs oligo(dA) et oligo(dT).

Les résultats obtenus avec le LTR sauvage de HIV-1 et l'IN de HIV-1 sont les mêmes que ceux précédemment décrits en 1994 par l'équipe de Wolffe (Pruss *et al.*, 1994b). Ce résultat valide notre stratégie expérimentale.

Pour chaque réaction, nous avons déterminé le rendement moyen de la réaction de transfert de brin, correspondant à la moyenne des intensités de fluorescence de chacun des pics. La distribution de l'intensité moyenne de fluorescence et le nombre de sites créés, obtenus avec l'IN de HIV-1 pour chacune des séquences cibles et chacun des LTR sauvages, sont présentés dans le Tableau 6R. Les résultats obtenus avec l'IN de HTLV-1 sont présentés dans le Tableau 7R.

Tableau 6R: Rendements moyens des réactions de transfert de brin catalysées par l'IN de HIV-1, et nombre de sites d'intégration créés. Réactions de transfert de brin entre le LTR sauvage de HTLV-1 ou le LTR sauvage de HIV-1 et les séquences Courbe ou Rigide.

IN HIV-1				
Cible	Courbe		Rigide	
LTR	Nombre de pics	Intensité moyenne de fluorescence	Nombre de pics	Intensité moyenne de fluorescence
LTR HIV-1	5	49	3	8
LTR HTLV-1	5	11	2	1

Tableau 7R : Rendements moyens des réactions de transfert de brin catalysées par l'IN de HTLV-1, et nombre de sites d'intégration créés. Réactions de transfert de brin entre le LTR sauvage de HTLV-1 ou le LTR sauvage de HIV-1 et les séquences Courbe ou Rigide.

IN HTLV-1				
Cible	Courbe		Rigide	
LTR	Nombre de pics	Intensité moyenne de fluorescence	Nombre de pics	Intensité moyenne de fluorescence
LTR HIV-1	5	4	0	0
LTR HTLV-1	5	111	1	5

Pour HTLV-1 et HIV-1, les évènements d'intégration sont significativement plus nombreux dans la séquence Courbe que dans la séquence Rigide. Le rendement de la réaction de transfert de brin est significativement plus élevé avec la séquence Courbe qu'avec la séquence Rigide, pour les 2 IN (49 versus 8, $p < 0,05$ pour l'IN de HIV-1 ; 111 versus 5, $p < 0,05$ pour l'IN de HTLV-1).

Cinq évènements d'intégration différents ont eu lieu dans la séquence Courbe, pour les deux IN. La Figure 14R représente les intensités de fluorescence obtenues à chacune des 50 positions centrales de la séquence Courbe, déléetée des séquences conservées de 18 et 20 pb.

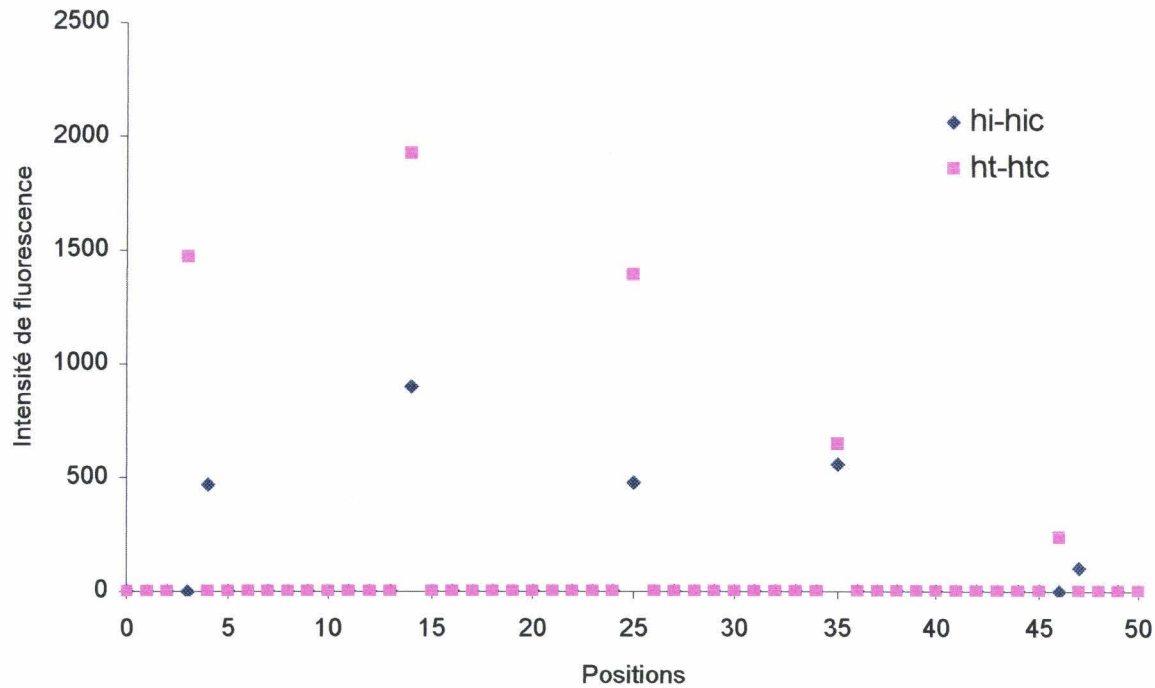


Figure 14R : Intensités de fluorescence à chacune des 50 positions de la séquence Courbe. Réaction de transfert de brin catalysée par l'IN de HTLV-1, entre la séquence Courbe et le LTR de HTLV-1, (ht-htc) ou réaction de transfert de brin catalysée par l'IN de HIV-1, entre la séquence Courbe et le LTR de HIV-1 (hi-hic).

Comme le montre la Figure 14R, les produits de la réaction catalysée par l'IN de HIV-1 correspondent à 5 sites d'intégration aux positions 4, 14, 25, 35 et 47 de la séquence Courbe. Les produits de la réaction catalysée par l'IN de HTLV-1 correspondent à 5 sites d'intégration aux positions 3, 14, 25, 35 et 46 de la séquence Courbe. L'intensité de fluorescence est supérieure aux positions 3, 14, et 25, notamment pour l'IN de HTLV-1. L'intensité de fluorescence est particulièrement élevée à la position 14 pour les deux IN. Il est vraisemblable

que les plus grandes intensités observées à ces positions puissent être dues en partie à la petite taille des fragments amplifiés, ce qui expliquerait l'absence de symétrie observée dans l'intégration. Ces variations d'intensité peuvent aussi correspondre à une meilleure accessibilité de ces positions. En effet, comme le montre la Figure 9R (page 96), dans laquelle sont présentés les résultats obtenus après l'analyse des séquences Courbe et Rigide par le logiciel Bend-it, la position 14 est caractérisée par un indice de courbure élevé comparé aux autres positions.

Les sites d'intégration choisis par l'IN de HTLV-1 sont donc les mêmes que ceux choisis par l'IN de HIV-1 aux positions 14, 25 et 35, mais différent aux positions 3 et 46.

Les deux IN n'ont pas la même activité catalytique. L'IN de HIV-1 est moins efficace que l'IN de HTLV-1. En effet, l'intensité moyenne de fluorescence est de 49 pour l'IN de HIV-1 contre 111 pour l'IN de HTLV-1, lorsqu'elles sont utilisées avec leur propre LTR. Toutefois, comme le montre la Figure 14R, la fluctuation de l'intensité de fluorescence en fonction de la position sur la séquence Courbe, est la même pour les deux protéines, excepté pour la position 35, lorsque l'on utilise l'IN de HIV-1.

Les deux IN privilégient donc de la même façon une structure particulière de l'ADN cible. L'IN de HTLV-1 privilégie ainsi une structure courbe de l'ADN, tout comme l'IN de HIV-1. Ce résultat est renforcé par l'absence d'intégration des LTR dans la séquence Rigide, dont la structure droite et non courbée, empêche l'intégration. De plus, la comparaison des pics d'intégration obtenus avec l'IN de HIV-1 et l'IN de HTLV-1 montrent que les deux rétrovirus s'intègrent le plus souvent aux mêmes positions sur la séquence Courbe. La sélection des sites d'intégration est la même quelque soit la séquence du LTR donneur, et l'IN utilisée.

II.3.2. L'IN de HTLV-1 est plus spécifique que l'IN de HIV-1.

II.3.2.1. Spécificité vis à vis du LTR.

L'IN de HTLV-1 a été utilisée pour catalyser l'intégration du LTR sauvage de HIV-1 dans la séquence Courbe. De la même façon, l'IN de HIV-1 a été utilisée pour catalyser la réaction de transfert de brin entre le LTR sauvage de HTLV-1 et la séquence Courbe. La Figure 15R

représente les électrophérogrammes obtenus après résolution des produits de PCR sur séquenceur d'ADN, pour l'IN de HTLV-1.

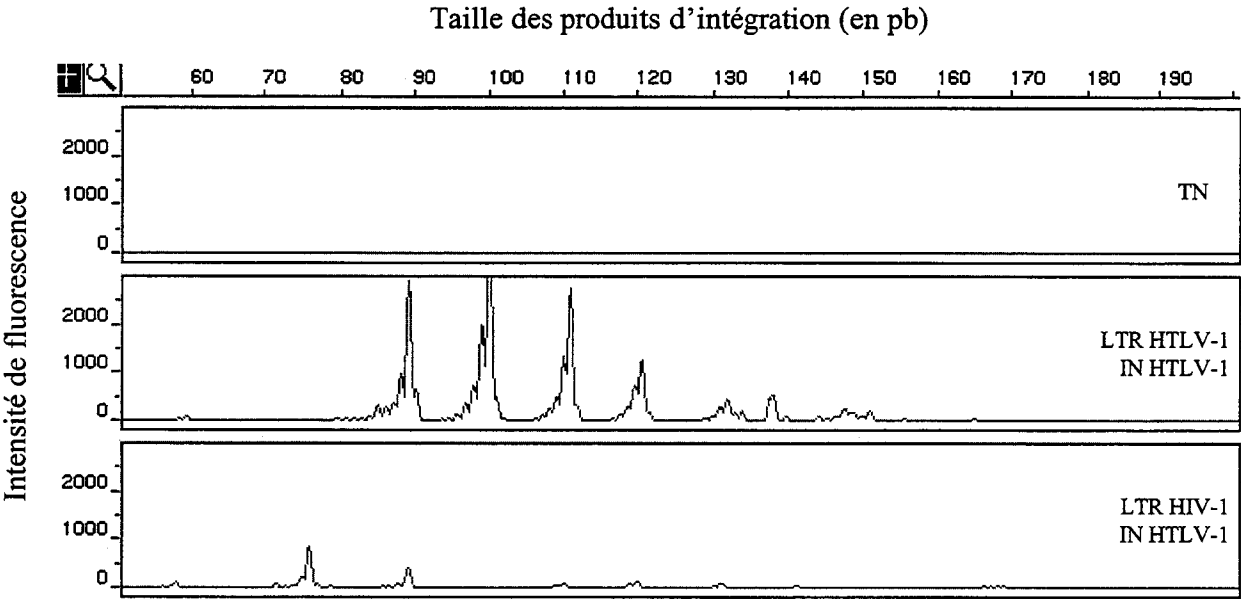


Figure 15R : Produits de PCR obtenus après les réactions de transfert de brin catalysées par l'IN de HTLV-1, entre la séquence Courbe et le LTR sauvage de HTLV-1 ou le LTR sauvage de HIV-1. TN : Témoin négatif, sans IN.

Le nombre de sites d'intégration est le même lorsque l'IN de HTLV-1 est utilisée avec le LTR de HIV-1 ou le LTR de HTLV-1. Cependant, la hauteur de chaque pic est considérablement réduite. Un électrophérogramme similaire est obtenu avec l'IN de HIV-1 (non montré).

Le rendement moyen de la réaction d'intégration, correspondant à la moyenne des intensités de fluorescence à chacune des positions, a été calculé pour chaque réaction d'intégration. Les résultats obtenus figurent dans les Tableaux 6R et 7R. Les résultats montrent qu'il y a une diminution significative de la moyenne de fluorescence lorsque l'IN de HTLV-1 est utilisée avec le LTR de HIV-1 (111 versus 4, $p<0,05$) et lorsque l'IN de HIV-1 est utilisée avec le LTR de HTLV-1 (49 versus 11, $p<0,05$).

Les rendements de transfert de brin obtenus pour chaque réaction d'intégration ont ensuite été exprimés en pourcentage (Figure 16R).

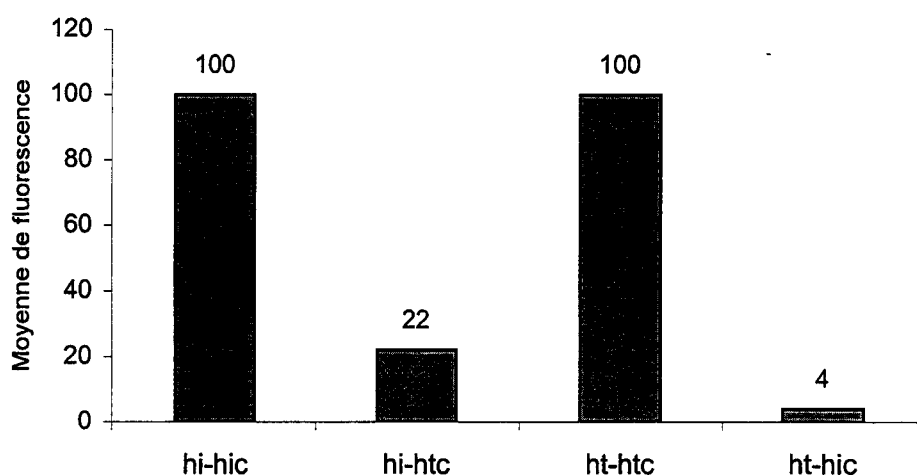


Figure 16R : Moyennes des intensités de fluorescence exprimées en pourcentage. Réactions de transfert de brin catalysées par l'IN de HIV-1 entre la séquence Courbe et le LTR de HIV-1 (hi-hic) ou le LTR de HTLV-1 (hi-htc). Réactions de transfert de brin catalysées par l'IN de HTLV-1 entre la séquence Courbe et le LTR de HTLV-1 (ht-htc) ou le LTR de HIV-1 (ht-hic).

Lorsque l'IN de HIV-1 est utilisée avec le LTR de HTLV-1, le rendement de la réaction correspond à 22% de celui obtenu avec le LTR sauvage de HIV-1. De la même manière, lorsque l'IN de HTLV-1 est utilisée avec le LTR sauvage de HIV-1, le rendement de la réaction correspond à 4% de celui obtenu avec le LTR sauvage de HTLV-1.

La diminution du rendement d'intégration est donc moins importante avec l'IN de HIV-1. Ces résultats suggèrent que l'IN de HTLV-1 est plus spécifique, et donc plus sensible à la composition nucléotidique du LTR que l'IN de HIV-1, qui semble mieux tolérer le LTR d'un autre virus.

La Figure 17R représente les intensités de fluorescence obtenues à chacune des 50 positions de la séquence Courbe, pour chaque réaction d'intégration.

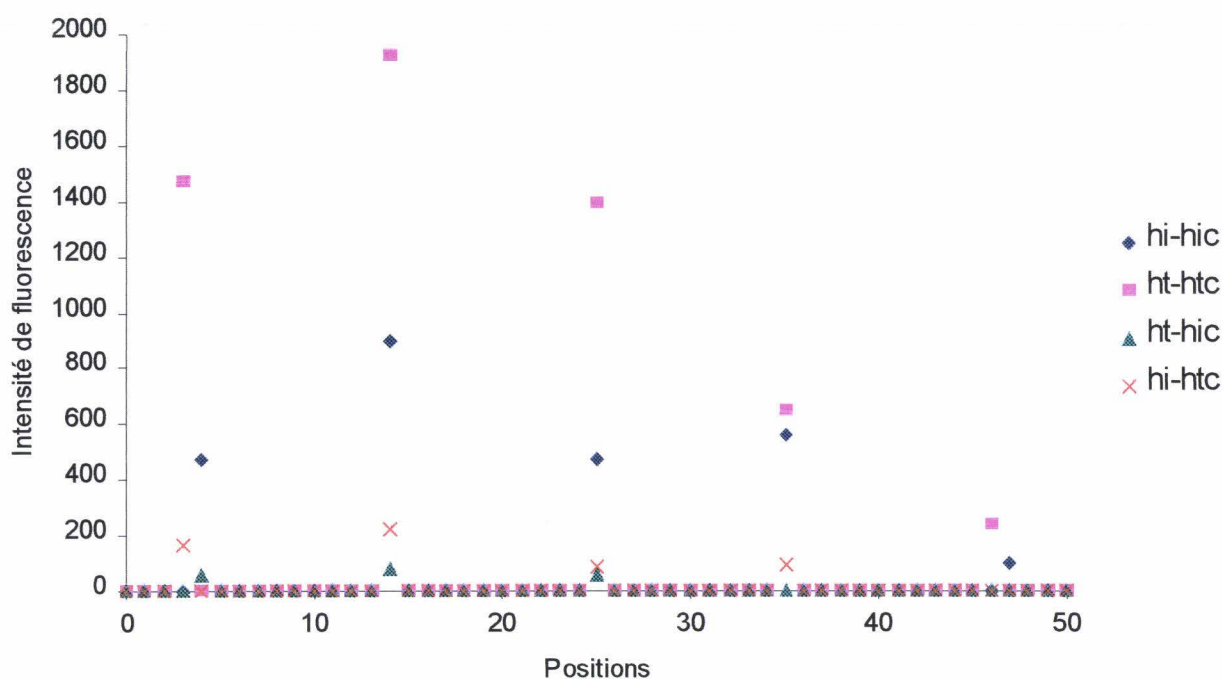


Figure 17R : Intensités de fluorescence des produits d'intégration à chacune des 50 positions centrales de la séquence Courbe. Réactions de transfert de brin catalysées par l'IN de HIV-1 entre l'oligonucléotide Courbe et le LTR sauvage de HIV-1 (hi-hic) ou le LTR sauvage de HTLV-1 (hi-htc). Réactions de transfert de brin catalysées par l'IN de HTLV-1 entre l'oligonucléotide Courbe et le LTR sauvage de HTLV-1 (ht-htc) ou le LTR sauvage de HIV-1 (ht-hic).

La fluctuation des intensités de fluorescence en fonction de la position est toujours la même. Les deux rétrovirus s'intègrent le plus souvent aux positions 3, 14, 25, 35 et 46 quel que soit le substrat utilisé. De légères variations sont observées aux positions 3 et 4 et aux positions 46 et 47.

Les deux LTR rétroviraux s'intègrent donc le plus souvent aux mêmes positions sur la séquence Courbe, quelle que soit l'IN utilisée. Et la sélection des sites d'intégration est identique quelque soit la séquence du substrat donneur.

II.3.2.2. Spécificité vis à vis de la cible.

La Figure 18R représente les différents sites d'intégration utilisés par les IN de HIV-1 et de HTLV-1, sur la séquence Courbe, après 13 réactions différentes. Les réactions d'intégration ont été réalisées en utilisant 13 LTR différents comme substrats donneurs.

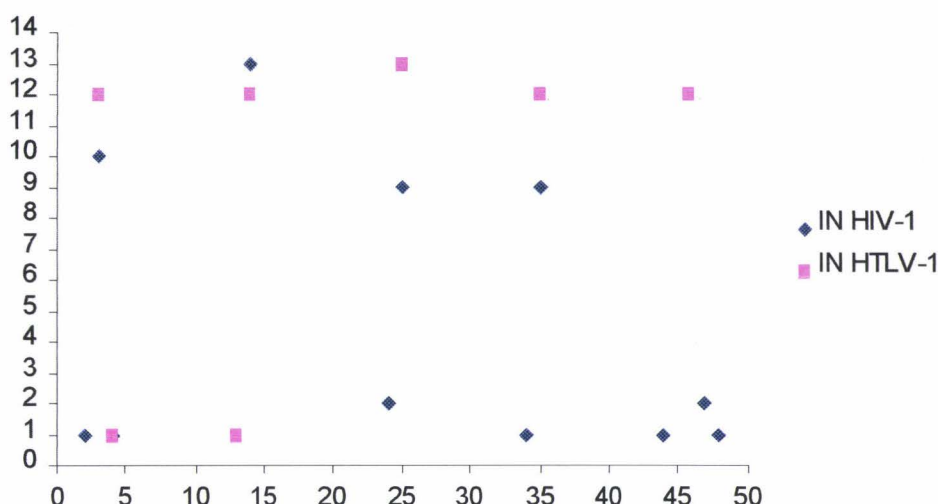


Figure 18R : Distribution du nombre de sites d'intégration choisis par l'IN de HTLV-1 ou de HIV-1, le long des 50 positions de la séquence Courbe.

Deux éléments se dégagent de ces résultats. Tout d'abord, 5 positions préférentielles (3, 14, 25, 35 et 46) correspondent à 90% de l'ensemble des événements. Ces positions correspondent donc aux sites les mieux exposés de la séquence Courbe.

Le deuxième aspect de ces résultats concerne les différences de spécificité des deux IN. 97% des sites utilisés par l'IN de HTLV-1 (soit 61 événements sur 63) correspondent aux 5 positions consensus mentionnées précédemment. Deux événements s'effectuent aux positions 4 et 13. Une dispersion plus importante des sites est observée avec l'IN de HIV-1 : 82% des sites utilisés par l'IN de HIV-1 (soit 41 événements sur 50) correspondent aux sites consensus. Neuf événements s'effectuent aux positions 2, 4, 24, 34, 44, 47 et 48. La Figure 18R montre d'autre part que la plupart des sites "aberrants" ne sont généralement distants que d'un nucléotide par rapport aux positions consensus. La différence de distribution des sites d'intégration pour les 2 IN est statistiquement significative ($p=0,005$, test de Fisher).

Ces résultats montrent que l'IN de HIV-1 présente également une spécificité moins importante que l'IN de HTLV-1 dans le choix du site d'intégration.

II.3.3. Topographie précise de l'intégration.

La comparaison des pics d'intégration dans la séquence Courbe, obtenus pour l'IN de HIV-1 et l'IN de HTLV-1 montre que les LTR des 2 rétrovirus s'intègrent pour la plupart aux mêmes sites ou à des positions voisines d'un nucléotide (Figure 18R). La taille des fragments amplifiés, et la résolution de notre système d'analyse nous ont alors permis de préciser la topographie exacte de l'intégration des deux rétrovirus dans la séquence Courbe. Ces résultats ont été confirmés par séquençage pour les positions 3, 14, 25 et 35 (Tableau 5R).

Nous avons ainsi déterminé la localisation exacte des différents sites de liaison entre le LTR donneur et la séquence Courbe, pour les deux IN. La Figure 19R représente les sites d'intégration préférentiels ciblés par les deux IN.



Figure 19R : Séquence nucléotidique de l'oligonucléotide Courbe. Les bases figurées en rouge correspondent aux séquences conservées de 18 et 20 pb utilisées pour l'amplification enzymatique des produits d'intégration. Les bases en bleu correspondent aux sites d'intégration préférentiels les plus souvent utilisés par les IN de HTLV-1 et de HIV-1.

La plupart des événements d'intégration ont donc lieu au niveau des nucléotides T, situés en aval des quadruplets de nucléotides A de la séquence Courbe.

II.4. Influence des mutations survenues *in vivo* sur la réaction de transfert de brin.

Dans une deuxième étape, nous avons évalué l'impact des mutations somatiques de l'extrémité du LTR 3', sur le rendement de la réaction de transfert de brin. Les séquences complètes des 11 séquences RU5 mutées, isolées de patients infectés par HTLV-1 (Mortreux *et al.*, soumis) et utilisées dans cette étude, figurent dans le Tableau 2M (Matériel et Méthodes).

Ces séquences ont été isolées de 4 patients infectés par HTLV-1. Les séquences P1, P2 et P5 correspondent à 3 séquences consensus différentes (2 patients ayant la même séquence consensus P1 et P6). Les séquences SUB1 à SUB8 portent des mutations somatiques et proviennent de 3 patients différents, l'un des 4 patients n'ayant pas de mutations somatiques (patient P5).

La Figure 20R représente les séquences consensus P1, P2, P5 et P6 des 4 patients infectés par HTLV-1, ainsi que les 8 LTR mutés (SUB1 à SUB8) correspondants. Ces 11 séquences de 54 pb ont été utilisées comme substrats donneurs dans des réactions d'intégration *in vitro*.

		Positions sur la séquence RU5								
Statut clinique	Nom de la séquence	9004	9006	9016	9022	9027	9036	9045	9047	9050
TSP	Consensus P1	A	G	C	T	G	A	T	T	G
P1	SUB2						G			
P1	SUB3			T						
P1	SUB4		A							
P1	SUB5									T
TSP	Consensus P2							C		
P2	SUB1					T				
ATLL	Consensus P5				C					
ATLL	Consensus P6									
P6	SUB6	-							-	
P6	SUB7								-	
P6	SUB8	-								

Figure 20R : Représentation des 11 séquences RU5 mutées dans leurs dernières 54 pb. La séquence consensus du patient P1 a servi de référence pour l'alignement. Les bases en rouge correspondent aux mutations propres à la séquence consensus, par rapport au LTR sauvage de HTLV-1. Les bases en gras correspondent aux mutations somatiques. Les séquences P1 et P6 sont identiques dans leurs 54 dernières pb. Les séquences SUB6, SUB7 et SUB8 portent des délétions (-) d'1 ou 2 bases.

Des oligonucléotides correspondant aux 3 séquences RU5 consensus et aux 8 séquences RU5 mutées (SUB1 à SUB8) ont été synthétisés puis utilisés dans des réactions d'intégration *in vitro*, de la même manière que précédemment décrit, en présence des IN de HTLV-1 ou de HIV-1.

Pour chaque séquence, la topographie et le rendement de la réaction de transfert de brin (intensité moyenne de fluorescence) ont été déterminés.

II.4.1. Les séquences RU5 consensus ont un profil d'intégration identique.

Après la réaction de transfert de brin utilisant l'IN de HTLV-1, la topographie de l'intégration pour les 4 séquences consensus provenant des 4 patients infectés, est la même sur la séquence Courbe. Les positions 3, 14, 25, 35 et 46 représentent 100% des sites d'intégration (non montré).

La Figure 21R permet de comparer les rendements de transfert de brin obtenus avec l'IN de HTLV-1, pour chaque séquence consensus et le LTR sauvage de HTLV-1.

Comme le montre la Figure 21R, les rendements de la réaction de transfert de brin ne diffèrent pas sensiblement d'une séquence consensus à une autre, et sont sensiblement identiques à celui de la séquence de référence, correspondant au LTR sauvage de HTLV-1.

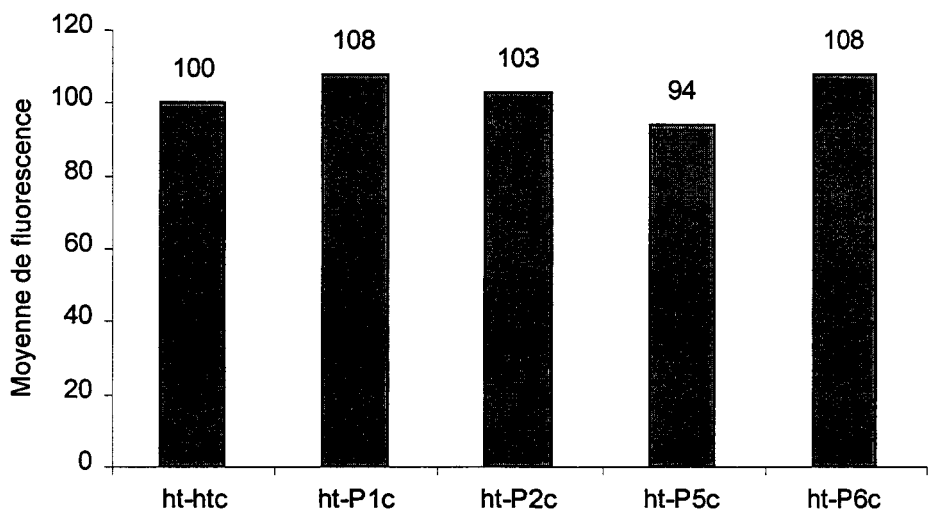


Figure 21R : Moyennes de fluorescence exprimées en pourcentage. Réactions de transfert de brin catalysées par l'IN de HTLV-1 entre l'oligonucléotide Courbe et le LTR sauvage de HTLV-1 (ht-htc), ou les LTR consensus P1, P2, P5 et P6 (ht-P1c, ht-P2c, ht-P5c et ht-P6c, respectivement).

La comparaison des intensités moyennes de fluorescence, par un test T apparié, ne montre pas de différence significative entre le rendement d'intégration obtenu avec le LTR sauvage de HTLV-1 et les rendements d'intégration obtenus avec les différents LTR consensus.

Les résultats précédents ont montré que l'IN de HIV-1 était moins spécifique que l'IN de HTLV-1. Nous avons donc comparé le rendement de la réaction de transfert de brin catalysée par l'IN de HIV-1, obtenu avec le LTR sauvage de HTLV-1 avec les rendements de transfert de brin obtenus avec les différentes séquences HTLV-1 consensus (Figure 22R).

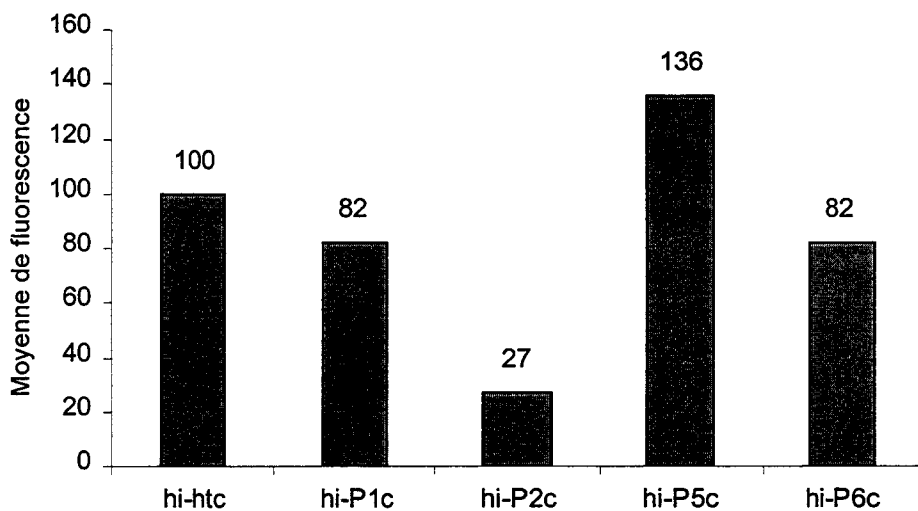


Figure 22R : Moyennes de fluorescence exprimées en pourcentage. Réactions de transfert de brin catalysées par l'IN de HIV-1 entre l'oligonucléotide Courbe et le LTR sauvage de HTLV-1 (hi-htc) ou les LTR consensus P1, P2, P5 et P6 (hi-P1c, hi-P2c, hi-P5c ou hi-P6c respectivement).

Lorsque l'IN de HIV-1 est utilisée avec les différentes séquences consensus, la topographie de l'intégration est la même aux positions 3 et 14. De légères variations sont observées aux autres positions de la séquence Courbe, dues au fait que l'IN de HIV-1 présente une spécificité moins importante que l'IN de HTLV-1 dans le choix du site d'intégration.

Concernant le rendement de la réaction de transfert de brin, la Figure 22R montre que l'intensité moyenne de fluorescence est plus faible pour 3 séquences consensus (P1, P2 et P6) par rapport à la séquence de référence, qui correspond au LTR sauvage de HTLV-1. A l'opposé, un gain de fonction est observé avec la séquence P5.

Toutefois, la comparaison des rendements moyens d'intégration par un test T apparié ne montre pas de différence significative entre le rendement d'intégration obtenu avec le LTR sauvage de HTLV-1 et les LTR consensus P1 (11 versus 9, $p>0,05$), P5 (11 versus 15, $p>0,05$) et P6 (11 versus 9, $p>0,05$). Seule une différence significative est observée entre le rendement d'intégration obtenu avec le LTR sauvage de HTLV-1 et le rendement d'intégration obtenu avec le LTR consensus P2 (11 versus 3, $p=0,05$).

Ainsi, même si l'IN de HIV-1 semble moins spécifique que l'IN de HTLV-1 vis à vis de la composition nucléotidique du LTR, certaines mutations ne sont pas tolérées par la protéine.

II.4.2. Des mutations somatiques de l'extrémité 3' du LTR peuvent modifier sensiblement le rendement de la réaction de transfert de brin.

II.4.2.1. Topographie de l'intégration.

La Figure 23R représente l'électrophérogramme obtenu avec l'IN de HTLV-1 pour les 8 séquences SUB1 à SUB8, porteuses de mutations somatiques.

La comparaison des pics d'intégration obtenus avec l'IN de HTLV-1 montre que le virus s'intègre aux mêmes positions sur la séquence Courbe.

Pour les deux oligonucléotides SUB6 et SUB7, le décalage d'une base par rapport aux autres LTR mutés s'explique par la présence des délétions au niveau de ces LTR. L'hybridation de l'amorce d'amplification 5'Fam-VP1 est décalée d'une base en 5' par rapport à l'extrémité et les ampligènes résultants sont donc plus courts d'une base (cf. Tableau 2M, page 152). Finalement, l'intégration a lieu aux mêmes sites sur la séquence Courbe que pour les autres LTR. Nous retrouvons la même périodicité de 10-11 pb que celle obtenue avec les LTR sauvages de HTLV-1 ou de HIV-1.

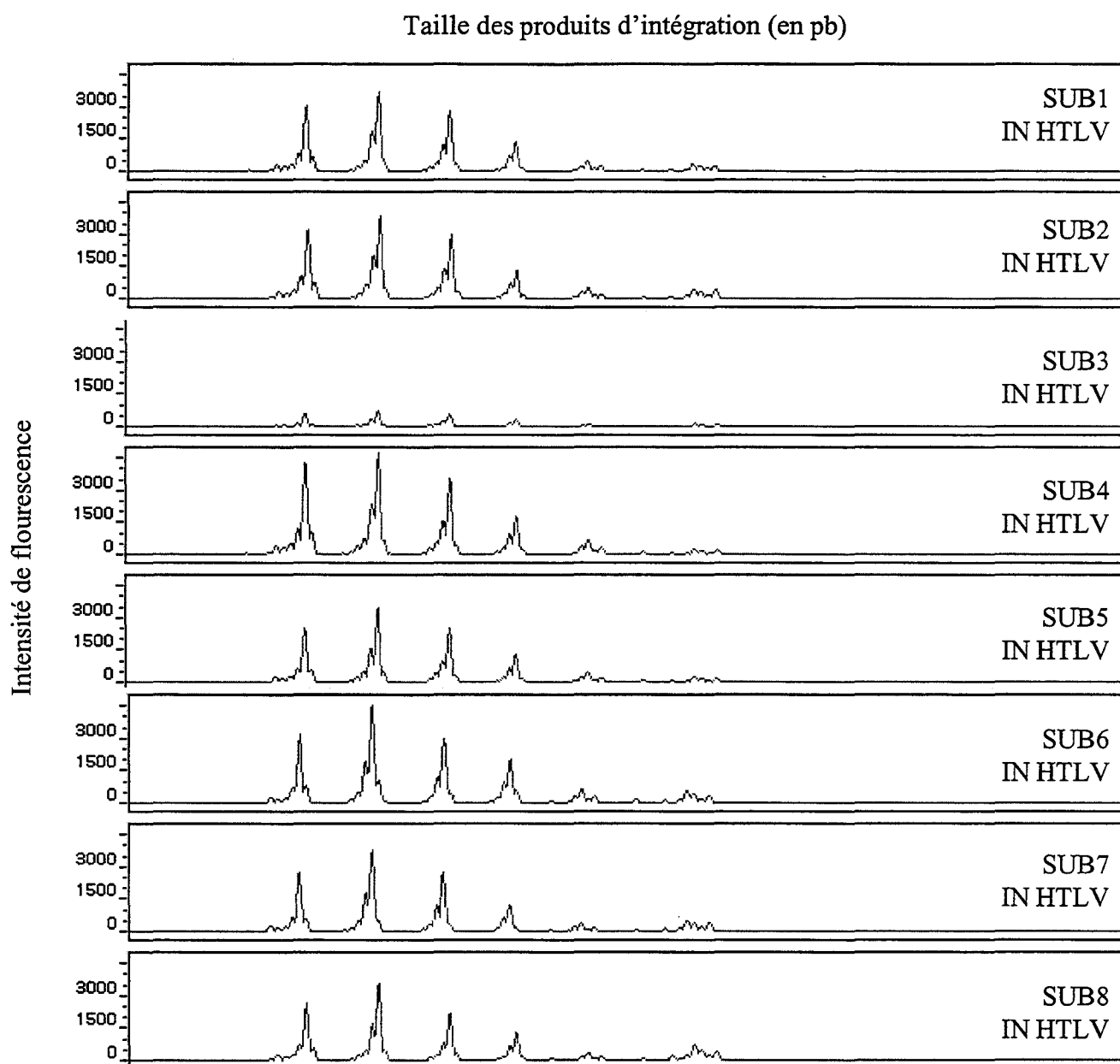


Figure 23R : Produits de PCR obtenus après les réactions de transfert de brin catalysées par l'IN de HTLV-1 entre l'oligonucléotide Courbe et les LTR SUB1 à SUB8.

La Figure 24R représente les intensités de fluorescence obtenues à chacune des 50 positions de la séquence Courbe, pour les LTR SUB1 à SUB8.

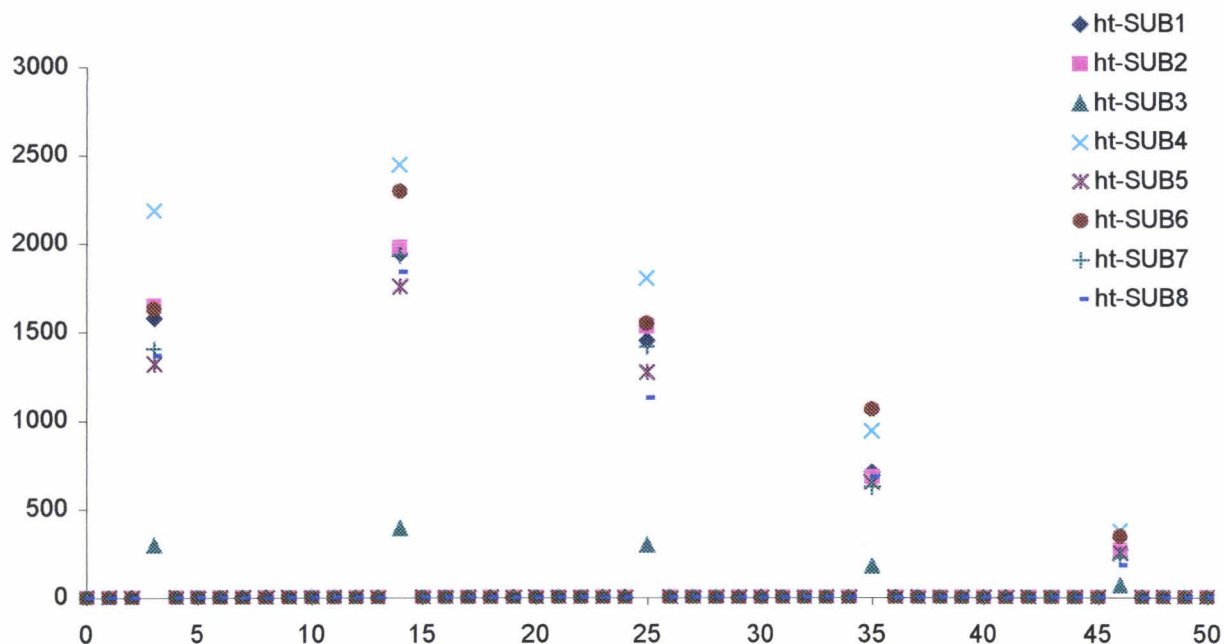


Figure 24R : Intensités de fluorescence à chacune des 50 positions de la séquence Courbe, pour les 8 LTR mutés SUB1 à SUB8. Réactions de transfert de brin catalysées par l'IN de HTLV-1 (ht-SUB1 à ht-SUB8).

Quel que soit le substrat utilisé, le virus s'intègre toujours aux positions 3, 14, 25, 35 et 46 de la séquence Courbe.

De la même façon, l'analyse topographique de l'intégration des LTR mutés SUB1 à SUB8 par l'IN de HIV-1 dans la séquence Courbe montre que le virus s'intègre la plupart du temps (29 fois sur 33) aux mêmes positions sur la séquence Courbe. Quelques variations sont observées : le LTR SUB7 ne s'intègre pas en position 3 mais en position 2. Le LTR SUB4 ne s'intègre pas en position 46 mais en position 47. Enfin, le LTR SUB8 ne s'intègre pas en position 46 mais en position 48 (résultats non montrés).

II.4.2.2. Rendement de la réaction de transfert de brin.

Afin d'évaluer l'impact des mutations somatiques sur le rendement de la réaction de transfert de brin catalysée par l'IN de HTLV-1, nous avons comparé les moyennes de fluorescence obtenues pour chaque LTR muté SUB1 à SUB8 avec la moyenne de fluorescence obtenue pour le LTR consensus correspondant.

Les Figures 25R, 26R et 27R représentent les résultats obtenus pour chaque LTR muté.

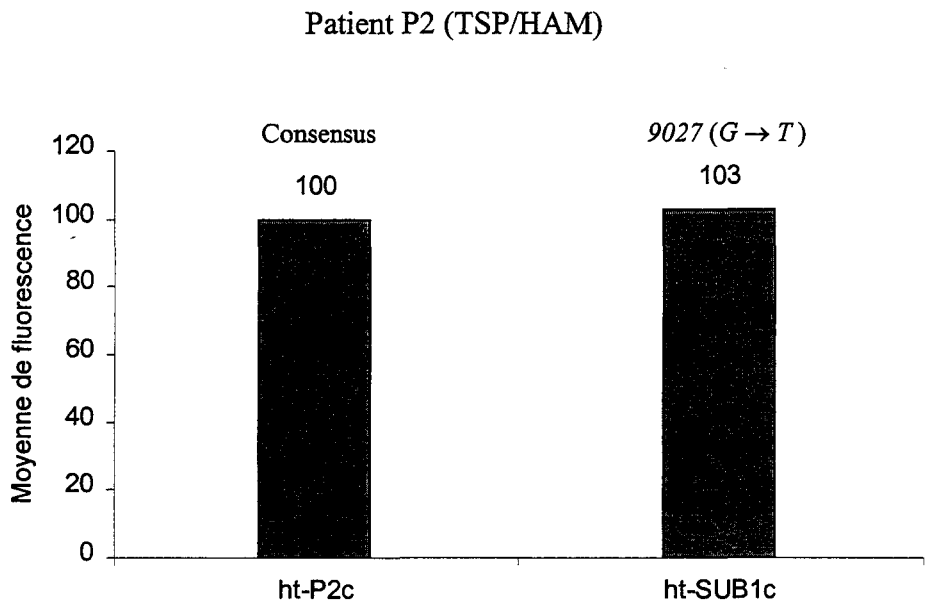


Figure 25R : Patient P2 (TSP/HAM). Moyennes des intensités de fluorescence exprimées en pourcentage. Réactions de transfert de brin catalysées par l'IN de HTLV-1 entre la séquence Courbe et le LTR consensus P2 (ht-P2c) ou le LTR SUB1 (ht-SUB1c). La position de la mutation somatique est indiquée en italique.

La comparaison des rendements moyens d'intégration par un test T apparié montre que la mutation somatique 9027 (*G → T*) de la séquence SUB1 ne modifie pas sensiblement l'efficacité de la réaction de transfert de brin (114 versus 117, $p > 0,05$).

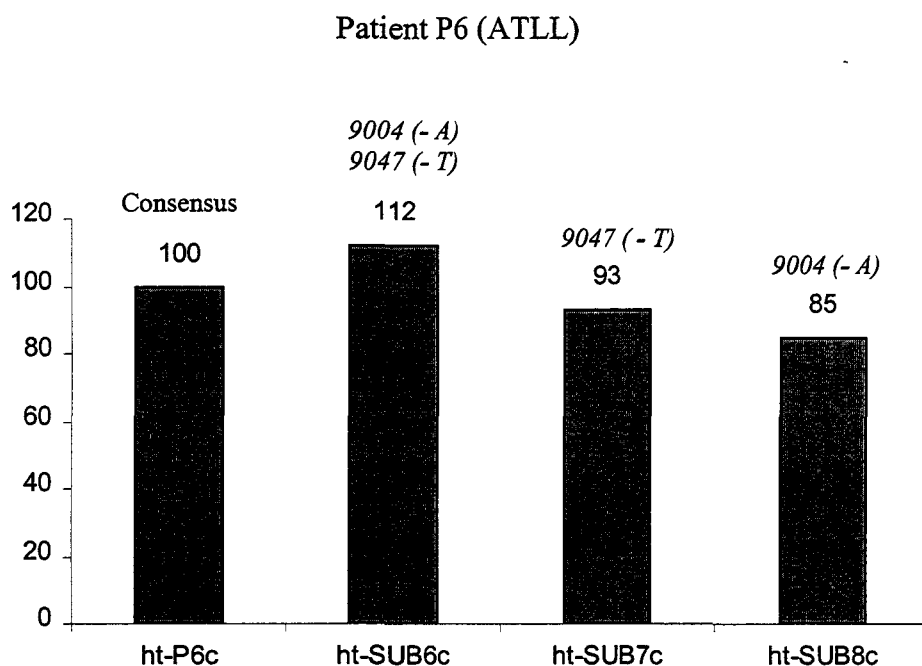


Figure 26R : Moyennes des intensités de fluorescence exprimées en pourcentage. Réactions de transfert de brin catalysées par l'IN de HTLV-1 entre la séquence Courbe et le LTR consensus P6 (ht-P6c) ou les LTR SUB6, SUB7 et SUB8 (ht-SUB6c, ht-SUB7c et ht-SUB8c respectivement).

De la même façon que précédemment, il n'y a pas de différence significative entre le rendement de la réaction de transfert de brin obtenu avec le LTR P6, et le rendement de la réaction d'intégration obtenu avec le LTR SUB6 (120 versus 135, $p > 0,05$), le LTR SUB7 (120 versus 112, $p > 0,05$) ou le LTR SUB8 (120 versus 102, $p > 0,05$).

Patient P1(TSP/HAM)

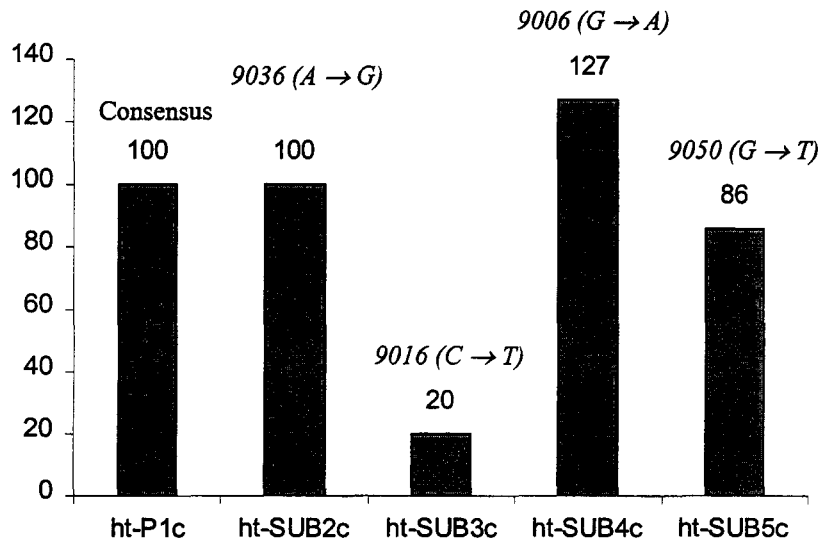


Figure 27R : Patient P1 (TSP/HAM). Moyennes des intensités de fluorescence exprimées en pourcentage. Réactions de transfert de brin catalysées par l'IN de HTLV-1 entre la séquence Courbe et le LTR consensus P1 (ht-P1c) ou les LTR SUB2, SUB3, SUB4 et SUB5 (ht-SUB2c, ht-SUB3c, ht-SUB4c et ht-SUB5c respectivement).

La comparaison des rendements moyens d'intégration par un test T apparié ne montre pas de différence significative entre le rendement d'intégration obtenu avec le LTR P1 et les LTR SUB2 et SUB5 (100% et 86% par rapport au LTR P1, respectivement).

A l'opposé, on observe une diminution significative du rendement moyen d'intégration avec le LTR SUB3 (120 versus 24, $p=0,048$) et une augmentation significative avec le LTR SUB4 (120 versus 152, $p=0,043$).

La séquence SUB3 provient d'un clone cellulaire dont l'ampleur a été estimée à 100 copies sur 150 000 PBMCs par IPCR quadruplée (Mortreux *et al.*, soumis). De même, la séquence SUB4 provient d'un clone cellulaire dont l'ampleur a été estimée à 100 copies sur 150 000 PBMCs.

La mutation 9016 (C→T) portée par la séquence SUB3 se situe au niveau de la région complémentaire de l'amorce de PCR 5'Fam-VPM. La diminution d'intensité de fluorescence obtenue après amplification enzymatique pourrait donc être issue d'un faible rendement d'amplification enzymatique et non d'une efficacité réduite de la réaction de transfert de brin.

Afin d'infirmer ou de confirmer cette hypothèse, nous avons amplifié l'oligonucléotide SUB3 en utilisant l'amorce 5'Fam-VPM et l'amorce ASVPM, complémentaire de l'extrémité 3' du LTR de HTLV-1. Parallèlement, des quantités équivalentes d'oligonucléotides SUB1 et SUB5 (non mutés au niveau de la région complémentaire de l'amorce 5'Fam-VPM) ont également été amplifiés par PCR, comme contrôles. Les produits d'amplification ont ensuite été déposés sur séquenceur d'ADN, comme décrit dans le Matériel et Méthodes. Les résultats sont présentés dans la Figure 28R.

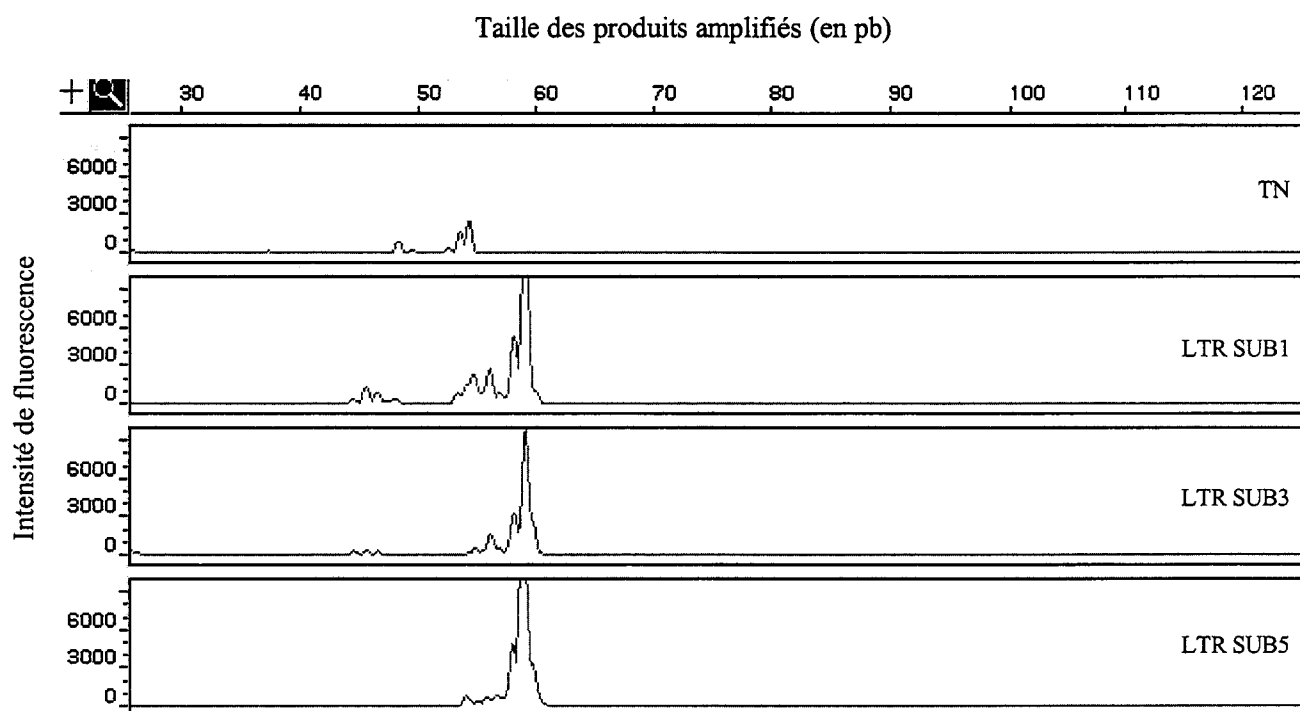


Figure 28R : Amplifications enzymatiques des LTR SUB1, SUB3 et SUB5 par PCR, à l'aide des amorces 5'Fam-VPM et ASVPM. TN : Témoin négatif.

Les rendements d'amplification enzymatique sont les mêmes, quel que soit le substrat utilisé. Ces résultats prouvent que la mutation 9016 (C→T) ne modifie pas l'efficacité de la PCR. La diminution significative de la moyenne de fluorescence observée après la réaction de transfert de brin catalysée par l'IN de HTLV-1 entre le LTR SUB3 et l'oligonucléotide Courbe est donc due à une diminution du rendement de transfert de brin.

Ces résultats montrent que des mutations, survenues *in vivo* au cours de la prolifération des cellules infectées, peuvent influencer le rendement de la réaction de transfert de brin sans modifier sensiblement la topographie de l'intégration.

II.5. Discussion.

L'étude de l'intégration du rétrovirus HTLV-1 *in vivo*, à partir de l'ADN de patients atteints d'ATLL, de TSP/HAM ou de porteurs asymptomatiques, a montré que le virus s'intégrait dans les régions du génome riches en nucléotides A/T. Les sites d'intégration sont d'autre part caractérisés par une distribution nucléotidique inhomogène dans le voisinage immédiat du provirus et par une fréquence élevée de longs motifs A/T. Ces caractéristiques suggèrent que la réaction d'intégration obéit à des contraintes structurales de l'ADN cellulaire. Ainsi, certains sites correspondent à des séquences intrinsèquement courbées alors que d'autres ont une distribution nucléotidique apparemment propice à la fixation de protéines susceptibles d'en modifier la conformation (Leclercq *et al.*, 2000).

Ces résultats nous ont conduit à aborder, dans une première étape, les caractéristiques structurales influençant la réaction de transfert de brin catalysée par l'IN de HTLV-1 *in vitro*. Pour cela, des réactions d'intégration ont été réalisées à l'aide d'IN recombinantes, de LTR synthétiques et de cibles oligonucléotidiques ayant des structures différentes en solution.

La méthode utilisée, basée sur l'amplification enzymatique des produits d'intégration à l'aide d'une amorce fluorescente et sur la résolution des ampligènes sur séquenceur d'ADN, a permis de déterminer la topographie précise de l'intégration (précision=0,15 pb) et de quantifier le rendement de la réaction de transfert de brin. Elle permet d'obtenir des produits d'intégration dans lesquels une seule extrémité de l'oligonucléotide donneur est liée à l'extrémité 5' de l'ADN cible. Cependant, une étude récente réalisée par l'équipe de Miller a montré qu'il existait une corrélation entre la réaction de transfert de brin et la réaction d'intégration *in vivo* (Hazuda *et al.*, 2000).

La topographie et le rendement de la réaction de transfert de brin ont été évalués pour deux séquences cibles ayant des structures totalement différentes en solution. Ces séquences riches en nucléotides A/T existent *in vivo* et ont déjà été testées pour l'intégration du rétrovirus HIV-1 (Pruss *et al.*, 1994b). La première cible Courbe est caractérisée par une structure courbée en solution ; à l'opposé, la deuxième cible Rigide, possède une structure droite et rigide. L'obtention de résultats identiques à ceux précédemment décrits par l'équipe de Wolffe pour l'IN de HIV-1 et le LTR de HIV-1, nous a permis de valider notre stratégie expérimentale (Pruss *et al.*, 1994b).

L'utilisation de ces deux oligonucléotides cibles Courbe et Rigide, possédant des structures totalement différentes en solution, a montré que l'IN de HTLV-1 privilégiait une structure particulière de l'ADN cible, tout comme l'IN de HIV-1. Les deux IN montraient une nette préférence pour une structure courbe de l'ADN. En effet, le rendement de la réaction de transfert de brin avec la séquence Courbe était significativement plus élevé qu'avec la séquence Rigide, pour les deux IN (111 versus 5 pour l'IN de HTLV-1 et 49 versus 8 pour l'IN de HIV-1, $p < 0,05$).

Plusieurs études, pour la plupart réalisées *in vitro*, ont montré que la structure de l'ADN cible, déterminée par sa composition nucléotidique, favorisait l'intégration au niveau de certains sites. Des protéines capables de courber l'ADN favorisent également l'intégration. Ainsi, lorsque des minichromosomes de levure sont utilisés comme cibles, l'intégration a lieu préférentiellement dans l'ADN enroulé autour des nucléosomes (Pryciak *et al.*, 1992b,c). La courbure de l'ADN, générée par la fixation de protéines spécifiques, comme la protéine IHF (Bor *et al.*, 1995) ou la protéine CAP (Muller *et al.*, 1994), entraîne également la sélection de sites d'intégration préférentiels.

En règle générale, la courbure de l'ADN cible génère donc des sites d'intégration préférentiels sur la face externe de l'hélice, que l'ADN soit ou non associé à la surface d'une protéine ; elle est donc une des composantes majeures de la sélection des sites d'intégration par les rétrovirus.

Cette étude a également montré que les deux rétrovirus s'intégraient le plus souvent, aux mêmes positions, sur la séquence Courbe. Cinq sites d'intégration préférentiels (correspondant aux positions 3, 14, 25, 35 et 46 de la séquence de 50 pb caractéristique de la séquence Courbe) correspondaient à 90% de l'ensemble des événements d'intégration réalisés lors de cette étude. Ces positions correspondaient donc aux sites les mieux exposés de la séquence Courbe. La distribution des sites d'intégration préférentiels le long de la séquence Courbe était caractérisée par une périodicité de 10 à 11 pb, reproduisant la répétition de la séquence cible. En effet, l'analyse topographique précise de l'intégration des deux rétrovirus a montré que les LTR rétroviraux s'intégraient la plupart du temps, au niveau des nucléotides T situés en aval des quadruplets de nucléotides A, qui caractérisent la molécule. Ces motifs sont séparés les uns des autres par 10-11 pb, le long de la séquence Courbe. L'analyse de la

séquence Courbe par le programme Bend-it suggère que ces résidus T possèdent un indice de courbure élevé et contribueraient par conséquent à la courbure intrinsèque de la molécule (Figure 29R). A l'opposé, aucun événement majeur d'intégration n'a été observé dans la séquence Rigide, au niveau des motifs oligo(dA) et oligo(dT), caractérisés par un indice de courbure significativement moins important.

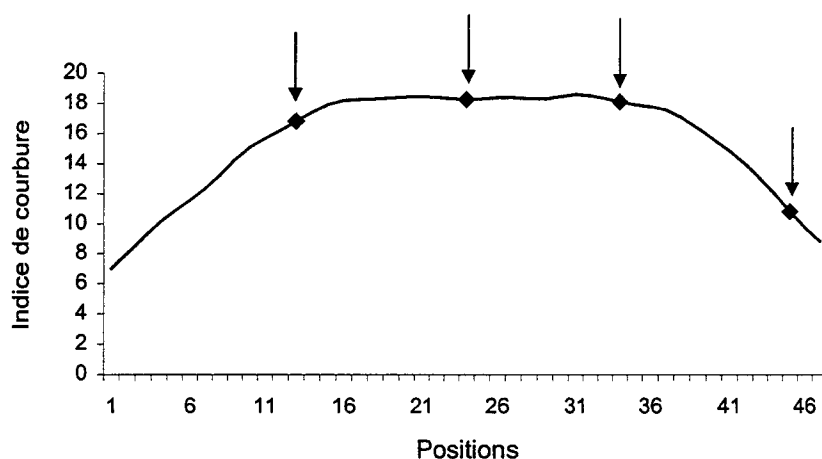


Figure 29R : Analyse de la séquence Courbe par le logiciel Bend-it. Les flèches représentent l'emplacement des nucléotides T ciblés par l'IN de HTLV-1.

L'ensemble de ces résultats confirment ainsi l'hypothèse selon laquelle l'intégration de HTLV-1 *in vivo* semble déterminée principalement par des caractéristiques structurales de l'ADN cible.

L'analyse de l'ensemble des sites d'intégration choisis par les IN de HTLV-1 et de HIV-1 au cours des différentes expériences d'intégration réalisées, a montré que les LTR rétroviraux s'intégraient le plus souvent aux positions 3, 14, 25, 35 et 46 de la séquence Courbe. Cependant, de légères variations ont été observées, notamment lors de l'utilisation de la protéine HIV-1 ; une dispersion plus importante et significative des sites était obtenue avec la protéine HIV-1 comparée à l'IN de HTLV-1. Ainsi, 97% des sites d'intégration choisis par l'IN de HTLV-1 correspondaient aux 5 positions consensus mentionnées ci-dessus, contre seulement 82% des sites utilisés par l'IN de HIV-1. Toutefois, la plupart des sites "aberrants" mis en évidence, n'étaient généralement distants que d'un nucléotide par rapport aux positions consensus.

L'analyse topographique de chaque réaction d'intégration a donc montré que la sélection des sites d'intégration était quasi identique quelle que soit la séquence du substrat donneur utilisé ; en règle générale, les LTR sauvages de HTLV-1 ou de HIV-1, et les différents LTR HTLV-1 portant des mutations "consensus" ou somatiques, sélectionnaient les mêmes sites d'intégration, quelle que soit l'IN utilisée. Des résultats similaires ont été obtenus *in vitro* après des réactions de transfert de brin réalisées avec les IN de HIV-1, de FIV ou de MMLV (Leavitt *et al.*, 1992 ; Chow *et al.*, 1994 ; Balakrishnan *et al.*, 1997). Par contre, d'autres études ont montré que lorsque plusieurs IN étaient en contact avec une même molécule d'ADN cible, les profils d'intégration obtenus pour chaque protéine étaient différents (van Gent *et al.*, 1991 ; Pryciak *et al.*, 1992c ; Bushman *et al.*, 1994b ; Katzman *et al.*, 1995 ; Shibagaki *et al.*, 1997).

Dans notre étude, la molécule d'ADN cible est relativement courte et le nombre de sites potentiels est donc relativement faible. De plus, la structure même de la séquence Courbe, favorise l'intégration préférentiellement au niveau de certains sites. Le choix de l'IN est donc relativement restreint.

L'étude de l'intégration du rétrovirus HTLV-1 *in vitro* a montré que l'IN de HTLV-1 était plus sensible à la composition nucléotidique du LTR que l'IN de HIV-1. On observe en effet une très nette diminution du rendement d'intégration lorsque l'IN de HTLV-1 était utilisée avec le LTR de HIV-1 et réciproquement ; lorsque l'IN de HIV-1 était utilisée avec le LTR de HTLV-1, le rendement de la réaction correspondait à 22% de celui obtenu avec le LTR sauvage de HIV-1. De la même manière, lorsque l'IN d'HTLV-1 était utilisée avec le LTR de HIV-1, le rendement de la réaction correspondait à 4% de celui obtenu avec le LTR sauvage de HTLV-1. L'effet était donc plus marqué pour l'IN de HTLV-1. La protéine HIV-1 semblait mieux tolérer le LTR d'un autre virus que l'IN de HTLV-1. L'IN de HTLV-1 apparaît donc plus spécifique vis à vis du LTR utilisé comme substrat donneur, et également vis-à-vis de la séquence cible.

Des expériences d'autointégration ont montré que les IN de HTLV-1 et de HTLV-2 étaient capables de catalyser les réactions de clivage et de transfert de brin à partir d'un grand nombre de substrats. Cependant, les deux protéines nécessitaient aussi une spécificité de séquence au niveau du LTR supérieure à celle requise par l'IN de HIV-1 (Balakrishnan *et al.*, 1997). Et les profils d'intégration obtenus avec les IN de HTLV-1, de HTLV-2, de HIV-1 et

du MMLV étaient identiques. Mais là encore, la molécule cible servait elle-même de substrat donneur, et le nombre de sites potentiels propices à l'intégration était donc très limité (20 pb).

Dans une deuxième étape, 11 séquences RU5 mutées, isolées de patients infectés par HTLV-1, ont été utilisées comme substrats donneurs dans la réaction de transfert de brin catalysée par l'IN de HTLV-1. L'étude de la variabilité du rétrovirus HTLV-1 a en effet permis d'isoler des séquences RU5 porteuses de mutations somatiques. Ces mutations ont été isolées *in vivo* à partir de l'ADN d'individus infectés par le virus et à partir de clones cellulaires en expansion ayant proliféré pendant plusieurs années. Les résultats ont montré que le rendement de la réaction de transfert de brin ne différait pas sensiblement d'une séquence consensus à une autre, et était sensiblement identique à celui de la séquence de référence, correspondant au LTR sauvage de HTLV-1. Les virus porteurs de ces mutations consensus ont été transmis d'un individu à l'autre ; ils constituent la forme majoritaire présente dans les lymphocytes infectés, et sont donc capables de s'intégrer et de se répliquer dans les cellules infectées.

Des mutations somatiques peuvent cependant influencer la réaction de transfert de brin. En effet, certaines mutations augmentaient le rendement de la réaction de transfert de brin ; ce fut le cas pour le LTR SUB4 qui portait une mutation G→A en position 9006. Lorsque l'IN de HTLV-1 était utilisée avec le LTR SUB4, le rendement de la réaction correspondait à 127% de celui obtenu avec le LTR correspondant à la séquence consensus du patient.

Certaines mutations au contraire, diminuaient fortement le rendement de la réaction de transfert de brin. Lorsque l'IN de HTLV-1 était utilisée avec le LTR SUB3, porteur d'une mutation C→T en position 9016, le rendement de la réaction correspondait à 20% de celui obtenu avec le LTR correspondant à la séquence consensus du patient.

Jusqu'à présent, l'ensemble des études réalisées a montré que seules les 15 dernières bases du LTR étaient importantes pour les étapes de clivage et de transfert de brin (Katzman *et al.*, 1989 ; Lafemina *et al.*, 1991 ; Sherman *et al.*, 1992). Toutefois, notre étude a montré qu'une seule substitution située en 9016, soit 39 bases en amont de l'extrémité 3' du LTR, pouvait altérer de manière significative l'efficacité de la réaction de transfert de brin. Ces résultats suggèrent que des mutations situées en aval des séquences répétées inversées (IR) ou des séquences requises pour l'intégration *in vitro* peuvent également influencer la réaction de transfert de brin.

L'acquisition de mutations *in vivo*, favorables à l'intégration du provirus dans l'ADN génomique pourrait contribuer à la propagation du virus. A l'opposé, certaines de ces mutations pourraient être délétères et empêcher ainsi la réplication du virus *in vivo*, devenu incapable de s'intégrer dans l'ADN génomique. L'incapacité du virus à s'intégrer dans l'ADN cellulaire pourrait ainsi contribuer au faible taux d'infection par rétrotranscription/intégration observé *in vivo*.

CONCLUSION-PERSPECTIVES

L'intégration est une étape clef du cycle réplcatif des rétrovirus et la présente étude montre pour la première fois qu'une composition nucléotidique particulière du génome cellulaire guide le choix de la cible pour l'intégration de HTLV-1 *in vivo*. Le virus s'intègre dans des régions riches en nucléotides A/T, avec une distribution nucléotidique particulière dans le voisinage immédiat du provirus. La réaction d'intégration semble obéir à des contraintes structurales de l'ADN cible et des relations structure-intégration ont été mises en évidence *in vitro* dans la deuxième partie de ce travail. Enfin, des mutations somatiques de l'extrémité 3' du LTR survenues *in vivo*, peuvent influencer positivement ou négativement le rendement de la réaction de transfert de brin.

Les résultats obtenus et les techniques mises au point permettent d'envisager 5 types de perspectives expérimentales.

La réaction de transfert de brin, cible thérapeutique de la réplication de HTLV-1.

Récemment, une étude a montré qu'il existait une corrélation étroite entre l'étape de transfert de brin et la réaction d'intégration *in vivo* (Hazuda *et al.*, 2000). Des inhibiteurs chimiques empêchent la réplication du rétrovirus HIV-1, en inhibant spécifiquement la réaction de transfert de brin catalysée par l'IN. La réaction de transfert de brin peut donc être utilisée pour tester des inhibiteurs potentiels de l'IN, ceci dans un objectif thérapeutique. Le rétrovirus HTLV-1 se réplique principalement par expansion clonale des cellules infectées (Cavrois *et al.*, 1996a,b ; Wattel *et al.*, 1996). Toutefois, une étude réalisée sur un modèle de singes Saïmiri infectés par HTLV-1, montre que le virus se réplique activement par rétrotranscription, pendant les 12 premiers jours qui suivent l'infection par une lignée cellulaire HTLV-1 (Kazanji *et al.*, sous presse ; Mortreux *et al.*, soumis). La recherche d'inhibiteurs de l'IN de HTLV-1 pourrait donc constituer, dans ce contexte, une approche thérapeutique intéressante. La stratégie développée ci-dessus permettrait une analyse quantitative précise de l'effet de certains inhibiteurs sur la réaction de transfert de brin.

Caractérisation moléculaire des séquences ciblées par l'IN de HTLV-1 *in vitro*.

La structure de l'ADN cible favorise l'intégration au niveau de certains sites. La courbure de l'ADN peut être induite d'une part, par la composition nucléotidique intrinsèque de la séquence cible, et d'autre part par l'interaction de la molécule cible avec des protéines spécifiques.

L'étude de l'intégration du rétrovirus HTLV-1 *in vitro* a montré que les IN de HTLV-1 et de HIV-1 privilégiaient une structure courbe de l'ADN cible. Cette étude a été réalisée en utilisant deux oligonucléotides cibles de structures totalement différentes en solution, les séquences Courbe et Rigide. Dans ce système, le choix de l'IN est donc relativement restreint comparé à ce que la cellule peut offrir *in vivo*. Le modèle que nous proposons maintenant repose sur l'utilisation d'une combinatoire de molécules cibles d'environ 150 pb dont la composition est randomisée par synthèse chimique. Ce pool d'oligonucléotides est rendu double brin par un cycle de PCR et peut être enroulé ou non autour d'histones d'érythrocytes de poulet. Plus de 10^7 molécules différentes peuvent être synthétisées et analysées simultanément, en permettant un degré élevé de variabilité. Des réactions de transfert de brin seront réalisées en utilisant comme substrat donneur un oligonucléotide biotinylé, représentant l'extrémité du LTR 3' de HTLV-1. Les produits d'intégration biotinylés ou non, seront purifiés puis amplifiés et les deux populations pourront être séquencées. Il sera alors possible de comparer, en présence et en absence d'histones, la composition nucléotidique des séquences retenues ou rejetées par l'IN. La flexibilité et la courbure des différentes cibles choisies par l'IN, pourront être évaluées par des analyses informatiques ou des tests de circularisation à la ligase. Ce test de ligation permet en effet d'évaluer rapidement la flexibilité d'une séquence et sa capacité à se tordre de façon spontanée ou non (Bailly *et al.*, 1996).

Ce travail permettra ainsi de caractériser les différents types de séquences choisies par l'IN de HTLV-1, qu'elles soient ou non associées à des histones.

Les partenaires protéiques de l'IN.

La réaction d'intégration *in vitro* décrite dans notre étude ne reproduit pas tous les aspects de l'intégration *in vivo*. Ainsi, une seule extrémité du substrat donneur est liée à l'extrémité 5' de l'ADN cible. L'utilisation des complexes de préintégration permet de réaliser des réactions d'intégration dites "concertées" dans lesquelles les deux extrémités de l'ADNv sont liées aux deux extrémités de l'ADN cible. Jusqu'à présent, seuls les complexes de préintégration des virus MMLV et HIV-1 ont été caractérisés en détail (Bowerman *et al.*, 1989 ; Farnet *et al.*, 1991b). La purification des PICs de HTLV-1 nous permettrait de déterminer les protéines cellulaires et virales présentes dans la cellule, en association avec l'IN, et donc potentiellement impliquées dans la réaction d'intégration *in vivo*. Des expériences d'intégration *in vitro* pourront être réalisées avec les PICs purifiés, en utilisant des oligonucléotides cibles comme ceux déjà utilisés précédemment, ou des cibles plus complexes comme des minichromosomes de levure.

L'étude de l'intégration du rétrovirus HTLV-1 *in vivo* a permis de mettre en évidence une distribution nucléotidique particulière au voisinage immédiat du provirus (C/G, A, C, et A/T aux positions 1, 8, 9 et 16, respectivement). Cette séquence "consensus" pourrait être propice à la fixation de protéines spécifiques, favorables à l'intégration. Une expérience préalable de « *Band Shifting* » à l'aide d'un oligonucléotide optimisé, porteur de la séquence consensus, permettra dans un premier temps de confirmer cette hypothèse. Ensuite, l'utilisation de ce même oligonucléotide, fixé sur une colonne d'affinité permettra de purifier, à partir d'un extrait nucléaire de cellules infectées ou non, les protéines capables de se fixer sur cette molécule d'ADN. Le rôle de ces protéines sur la structure de la chromatine et leur capacité à courber l'ADN pourront être évalués par un test de circularisation à la ligase (Bailly *et al.*, 1996). Enfin, l'effet de ces protéines sur le rendement de la réaction de transfert de brin sera évaluée en réalisant des expériences d'intégration *in vitro*, en présence et en absence des ces protéines.

L'intégration concertée.

Des systèmes d'intégration *in vitro* ont été élaborés récemment et permettent de réaliser des réactions d'intégration "concertées". L'addition de protéines de la famille HMG augmente l'efficacité de la réaction (Hindmarsh *et al.*, 1999). Un système équivalent utilisant l'IN de HTLV-1 et un substrat donneur porteur des régions U3 et U5 pourra être mis au point afin de permettre une intégration concertée du provirus HTLV-1. Cette étude permettra de faire la corrélation entre la réaction d'intégration *in vivo* et la réaction d'intégration *in vitro*. Le séquençage des jonctions entre l'ADN viral et l'ADN cible, permettra de déterminer si la duplication de l'ADN cible est efficace *in vitro*.

Détermination de la composition nucléotidique optimale du LTR pour la réaction de transfert de brin.

L'étude de l'intégration du rétrovirus HTLV-1 *in vitro* a montré que des mutations somatiques survenues *in vivo*, pouvaient influencer la réaction de transfert de brin. Certaines de ces mutations diminuaient le rendement de la réaction de transfert de brin alors que d'autres augmentaient au contraire l'efficacité de la réaction. Des expériences d'intégration *in vitro* comme celles que nous avons utilisé dans notre étude pourront être réalisées en utilisant une combinatoire de molécules, totalement randomisées par synthèse chimique et correspondant à l'extrémité 3' du LTR de HTLV-1 ou de HIV-1. La caractérisation d'un certain nombre de séquences, retenues ou refusées par l'IN lors de l'étape de transfert de brin, nous permettra d'identifier les résidus nucléotidiques nécessaires à cette étape. On imagine que les virus portant cette séquence consensus correspondraient à des virus possédant un avantage réplcatif certain sur d'autres souches rétrovirales. Ces virus, très efficaces pour l'intégration, pourraient être utilisés ultérieurement en thérapie génique. L'utilisation de vecteurs rétroviraux implique en effet l'intégration du provirus dans le génome des cellules cibles, et permet ainsi une expression stable du transgène.

MATERIEL ET METHODES

I. MATERIELS.

I.1. Les échantillons étudiés.

Les échantillons analysés proviennent de patients atteints de TSP/HAM ou d'ATLL, et de porteurs asymptomatiques. Les données cliniques et l'origine géographique de chaque individu ont pour la plupart déjà été publiées (Wattel *et al.*, 1992a,b ; Cavois *et al.*, 1996a,b).

Les différents prélèvements de sang et de moelle proviennent :

- du laboratoire d'épidémiologie des rétrovirus oncogènes de l'Institut Pasteur (Paris).
- du laboratoire de biologie du CHRU de Fort de France (Martinique).
- du service d'hématologie de l'hôpital Necker (Paris).
- de la fédération de neurologie de l'hôpital la Pitié Salpêtrière (Paris).

I.2. Les lignées cellulaires.

Les lignées cellulaires infectées par HTLV-1 et utilisées dans cette étude comme témoins positifs, sont les lignées BOUL et MT4. La lignée BOUL a été établie par Antoine Gessain à partir des cellules du sang périphérique d'un patient atteint de TSP/HAM. La lignée cellulaire promyéloblastique HL60 et la lignée lymphoïde Jurkat, non infectées par HTLV-1, ont été utilisées comme témoins négatifs.

I.3. Le phage M13mp18 et la souche bactérienne XL1.

Le cycle répliatif du phage M13mp18 passe par un intermédiaire simple brin, qui peut être facilement isolé à partir du surnageant de cultures bactériennes et facilite ainsi le séquençage des inserts.

La souche bactérienne XL1 (Stratagene), de phénotype : *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [*F'* *proAB lacIqZAM15 Tn10* (Tet^r) Amy Cam^r] est utilisée pour les transformations bactériennes et la culture du phage.

Le plasmide M13mp18 possède un fragment du gène de l'opéron lactose qui code le peptide α de la β -galactosidase. Sa présence a permis l'élaboration d'un test colorimétrique relativement simple, permettant de distinguer les phages recombinants des phages sauvages.

Ce gène est sous le contrôle d'un promoteur inductible par l'IPTG (Isopropyl- β -D-Thio-Galactopyranoside). Parallèlement, les bactéries XL1 possèdent un gène *lacZ* délété et ne peuvent exprimer le gène de la β -galactosidase que par α -complémentation. En présence d'IPTG et de X-gal (5-Bromo-4-Chloro-3-Indolyl- β -D-Galactopyranoside), qui est un substrat chromogénique de la β -galactosidase, les bactéries complémentent. Les plages de lyse formées par les phages sauvages sont donc bleues, le dérivé métabolisé du X-gal étant de couleur bleue. Par contre, l'insertion d'un fragment d'ADN étranger dans le site multiple de clonage du phage M13mp18, au niveau du gène de l'opéron lactose, empêche l' α -complémentation ; les phages recombinants donnent naissance à des plages de lyse non colorées.

I.4. Les vecteurs d'expression des intégrases de HTLV-1 et de HIV-1.

Les plasmides pET22b-INHT et pET22b-INHV (Novagen) portent le gène codant l'IN de HTLV-1 ou le gène codant l'IN de HIV-1, respectivement. Ces deux constructions nous ont été fournies gracieusement par Colleen Jonsson, du laboratoire de biochimie de l'Université du Nouveau Mexique (Etats-Unis). Le plasmide pET22b-INHT comporte une région de 899 pb (4312-5211) contenant le gène codant l'IN de HTLV-1. Le plasmide pET22b-INHV contient un fragment de 911 pb contenant la séquence complète du gène de l'IN de HIV-1, isolée à partir du plasmide pINT7 (Engelman *et al.*, 1992). Les deux gènes codant les IN des deux rétrovirus sont sous le contrôle du promoteur du phage T7. La souche d'*Escherichia coli* (*E.coli*) transformée par les plasmides pET22b-INHT ou pET22b-INHV est la souche BL21DE3 de phénotype : F⁻ *ompT hsdS_B (r_B⁻ m_B⁻) gal dcm* (DE3). L'ARN polymérase de cette souche bactérienne ne reconnaît pas le promoteur T7, et rend par conséquent impossible la transcription du gène d'intérêt et l'expression de la protéine.

Le génome de la souche bactérienne BL21DE3 possède le gène de l'ARN polymérase du phage T7 (λ DE3 lysogène). Ce gène est sous le contrôle du promoteur de l'opéron lactose, inductible par l'IPTG. En présence de l'inducteur, l'ARN polymérase du phage T7 est donc produite et permet la transcription du gène *pol* codant l'IN, et l'expression de la protéine. L'introduction d'une courte séquence nucléotidique codant 6 résidus histidine, en aval du gène d'intérêt, permet d'exprimer une protéine chimère, portant à son extrémité carboxy-terminale, une queue poly-histidine affine pour le nickel, qui sera utilisée ultérieurement pour

la sélection et la purification de la protéine. Une chromatographie d'affinité en conditions dénaturantes, sur colonne de nitrilotriacétate de nickel, est réalisée pour purifier l'IN produite. La Figure 1M ci-dessous schématise la construction plasmidique utilisée pour produire la protéine.

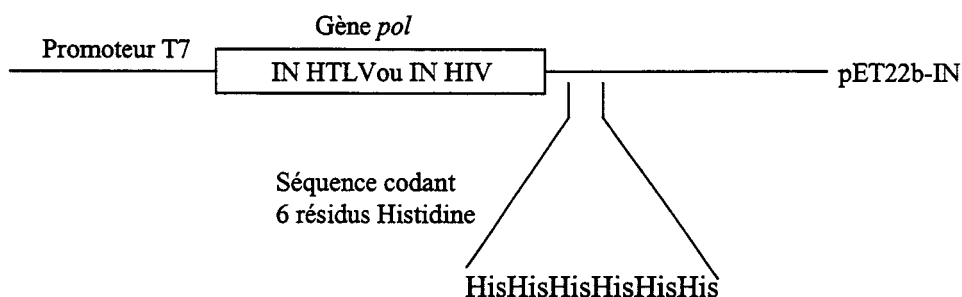


Figure 1M : Schéma de la région d'intérêt du vecteur d'expression pET22b-IN.

II. METHODES.

II.1. Culture cellulaire.

Les lignées hématopoïétiques comme les lignées HL60, Jurkat et MT4, sont cultivées dans un milieu RPMI 1640 "with Glutamax" (Gibco BRL) enrichi à 10% de sérum de veau fœtal (Boehringer), à 37°C et dans une atmosphère humide à teneur en CO₂ de 5%. La lignée BOUL dépendante de l'IL2, est cultivée dans les mêmes conditions avec une concentration finale en IL2 de 20 U/mL (Boehringer).

II.2. Extraction de l'ADN.

L'extraction de l'ADN des lignées cellulaires est effectuée après lavage des cellules dans un tampon PBS 1X. A partir des prélèvements sanguins de patients, les PBMCs ont été isolées par centrifugation sur coussin de ficoll.

Les culots cellulaires sont ensuite digérés par la protéinase K (100 µg/mL) pendant une nuit à 42°C, dans un tampon de lyse (Tris-HCl 10 mM (pH 8) ; EDTA 10 mM ; NaCl 10 mM ; SDS 0,5%).

Après purification de l'ADN par 2 extractions au phénol/chloroforme (1/1) et une extraction au chloroforme, l'ADN est précipité par deux volumes d'éthanol absolu et 1/10^{ème} de volume d'acétate de sodium 3M. Après une centrifugation à 4°C, de 30 min à 14000 g, le culot est lavé à l'éthanol 70%, séché sous vide et repris dans une solution de TE 1X. L'ADN est dosé par spectrophotométrie à 260 nm.

II.3. Amplification enzymatique des sites d'intégration flanquant le LTR 3' du provirus HTLV-1.

Deux techniques, la PCR avec ligation d'un oligonucléotide (LMPCR pour *Ligation Mediated Polymerase Chain Reaction*) et la PCR inverse (IPCR pour *Inverse Polymerase Chain Reaction*) ont été utilisées (Figure 2M et Figure 3M).

Pour éviter les contaminations de PCR, les expériences d'amplification enzymatique sont réalisées dans 3 laboratoires différents, sans qu'il puisse exister un flux de personne ou de matériel du lieu d'analyse des produits amplifiés vers les lieux où sont réalisées les étapes antérieures. Dans chaque expérience d'amplification, 2 témoins négatifs et un témoin positif sont analysés. Les séquences nucléotidiques des amorces utilisées pour l'amplification enzymatique des sites d'intégration figurent dans le Tableau 1M.

II.3.1. PCR avec ligation d'un oligonucléotide ou LMPCR.

La LMPCR consiste à digérer, dans un premier temps, l'ADN génomique de la cellule hôte infectée par HTLV-1, par une enzyme de restriction, qui va également réaliser une coupure au niveau du LTR proviral. Dans une seconde étape, la ligation d'un oligonucléotide au niveau du site de restriction, va permettre l'amplification du site d'intégration (Figure 2M). L'enzyme de restriction *NlaIII* a été choisie car elle reconnaît une séquence de 4 pb et permet donc des coupures fréquentes dans l'ADN génomique, en générant des fragments de tailles faciles à amplifier et à cloner.

De plus, la coupure enzymatique de l'ADN par *NlaIII* génère une extrémité 3' protrusive de 4 bases facilitant la ligation d'un oligonucléotide.

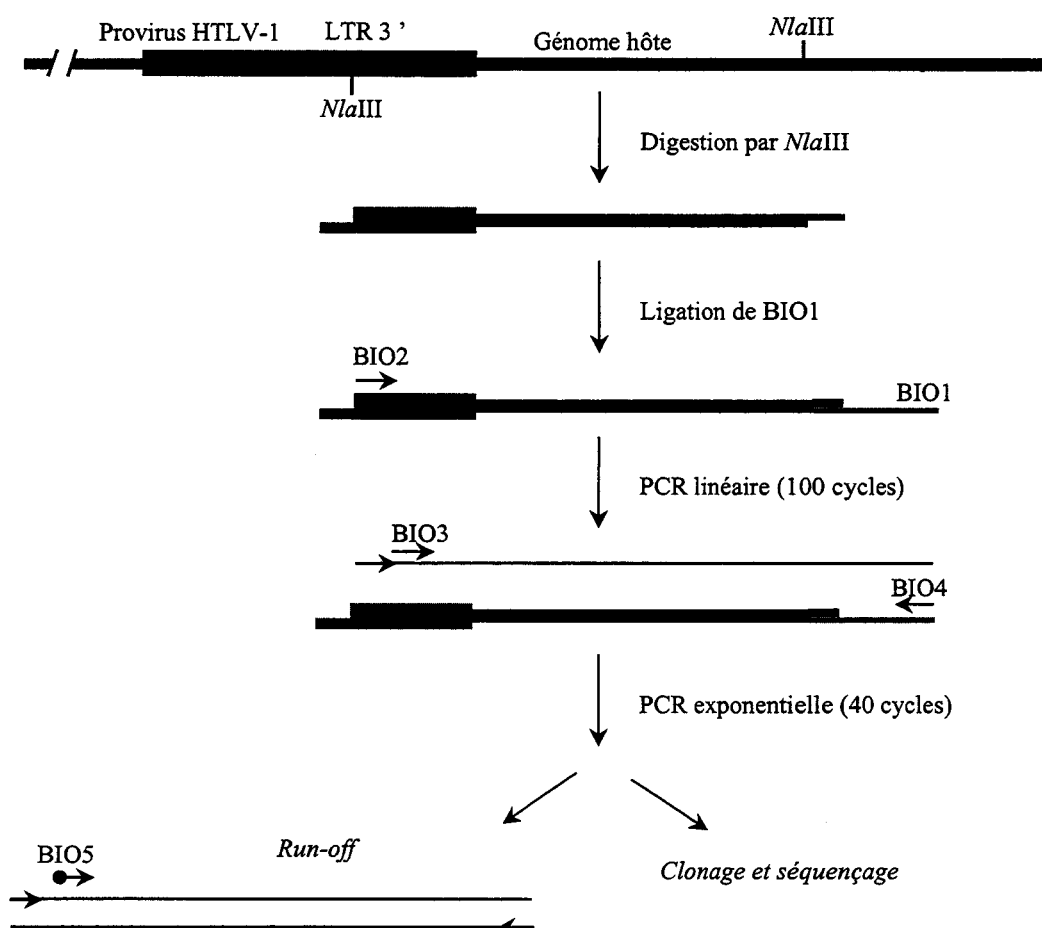


Figure 2M : Amplification des sites d'intégration flanquant le LTR 3' du provirus HTLV-1 par LMPCR.

En pratique, 2 µg d'ADN génomique sont digérés par 10 U de l'enzyme *NlaIII* (Biolabs), dans un volume réactionnel de 50 µl (Acétate de potassium 50 mM ; Acétate de Tris 20 mM ; Acétate de magnésium 10 mM ; DTT 1 mM), pendant 3 heures à 37°C. Après vérification de la digestion par électrophorèse en gel d'agarose 1%, les produits digérés sont purifiés par 2 extractions au phénol/chloroforme (1/1). Une dernière extraction au chloroforme permet d'éliminer les traces de phénol. Les produits purifiés sont ensuite précipités par 2,5 volumes d'éthanol 100% et 1/10^{ème} de volume d'acétate de sodium 3M. Après une centrifugation de 45 minutes à 14000 g, le culot d'ADN est lavé par 150 µl d'éthanol 70%, puis séché et repris dans 20 µl d'eau milliQ.

Quinze picomoles de l'amorce BIO1, dont l'extrémité 3' est complémentaire des 4 bases protrusives du site de restriction *Nla*III, sont ensuite ligaturés à 1 µg, soit 10 µl de produits digérés, pendant 1 nuit à 16°C. La réaction est catalysée par 20 U de ligase (Pharmacia) dans un volume réactionnel de 40 µl (Tris-HCl 50 mM (pH 7,8) ; MgCl₂ 10 mM ; DTT 10 mM ; ATP 1 mM ; BSA 50 µg/mL). Les produits de ligation sont ensuite purifiés par extraction au phénol/chloroforme, précipités par de l'éthanol 100%, puis repris dans 40 µl d'eau milliQ.

Une PCR linéaire "hot start" de 100 cycles est ensuite réalisée sur l'équivalent de 0,5 µg de produits de ligation, à l'aide de l'amorce BIO2, située dans la région U5 du provirus. Dans un premier temps, 25 µl du mélange réactionnel (MgCl₂ 1,5 mM ; KCl 10 mM ; Tris-HCl 10 mM (pH 8,3)) contenant 10 pmol de l'oligonucléotide BIO2 et une concentration finale en dNTP de 500 µM chacun sont placés dans un tube eppendorf de 750 µl sous une bille de paraffine (Ampliwax, Perkin Elmer). Une incubation de 8 min à 79°C suivie d'une incubation à 15°C pendant 13 min, permet l'isolement de la phase inférieure. Sur cette phase inférieure sont déposés 60 µl d'une solution (MgCl₂ 1,5 mM ; KCl 10 mM ; Tris-HCl 10 mM (pH 8,3)) contenant 0,5 µg de produits de ligation et 10 U du fragment Stoffel de la *Taq* polymérase (Perkin Elmer). Les conditions d'amplification sont : 95°C 10 min, 100 fois (95°C 1 min, 60°C 1 min, 72°C 2 min), 72°C 10 min.

Une réaction de PCR exponentielle à l'aide de l'amorce BIO3, située en aval de BIO2, et de l'amorce BIO4, colinéaire à BIO1, est réalisée sur 10 µl de produits de PCR linéaire. Cette amplification s'effectue dans un volume réactionnel de 100 µl de tampon (Tris-HCl 10 mM (pH 9) ; KCl 50 mM ; MgCl₂ 1,5 mM ; Triton X-100 0,1% ; BSA 0,2 mg/mL) avec 50 pmol de chaque amorce, une concentration finale en dNTP de 200 µM chacun, et 2 U de *Taq* polymérase (Appligene). Les paramètres d'amplification sont : 95°C 10 min, 40 fois (95°C 1 min, 58°C 1 min, 72°C 1 min), 72°C 10 min.

II.3.2. PCR Inverse ou IPCR.

L'IPCR débute comme la LMPCR, par la digestion de 2 µg d'ADN génomique par l'enzyme *Nla*III, comme précédemment décrit. Les produits de digestion sont ensuite circularisés pour permettre l'amplification (Figure 3M).

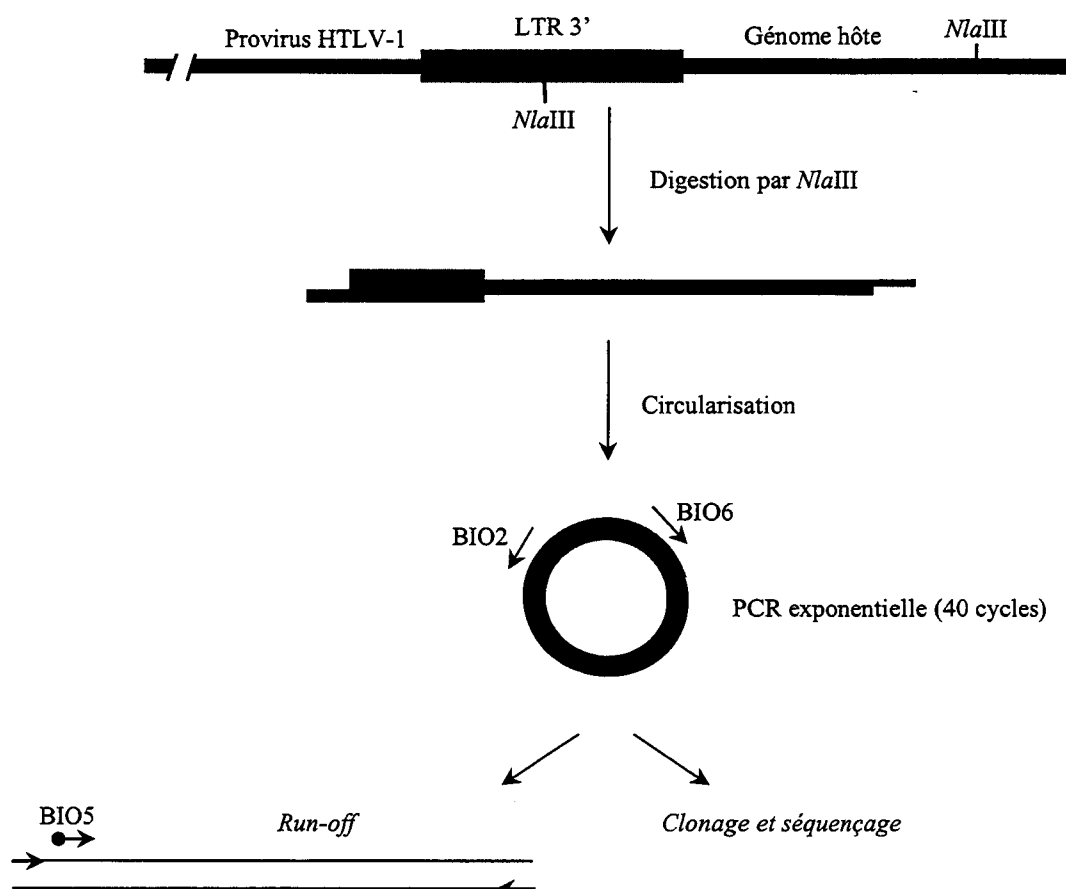


Figure 3M : Amplification des sites d'intégration flanquant le LTR 3' du provirus HTLV-1 par IPCR.

Après vérification de la digestion par électrophorèse en gel d'agarose 1%, les produits de la digestion sont purifiés par extraction au phénol/chloroforme puis précipités par de l'éthanol 100%. Après lavage et séchage, les culots d'ADN sont repris dans 20 μ l d'eau milliQ. La coupure enzymatique de l'ADN par *Nla*III génère une extrémité 3' protrusive de 4 bases facilitant la circularisation des fragments. Un microgramme de produits digérés sont ensuite circularisés par 40 U de ligase (Pharmacia) dans un volume réactionnel de 600 μ l (Tris-HCl 50 mM (pH 7,8) ; $MgCl_2$ 10 mM ; DTT 10 mM ; ATP 1 mM ; BSA 50 μ g/mL), en présence d'ATP 10 mM. La ligation s'effectue à 16°C, pendant une nuit. Les produits de ligation sont ensuite purifiés par 2 extractions au phénol/chloroforme et une extraction au chloroforme, puis précipités par 2 volumes d'éthanol 100% et 1/10^{ème} de volume d'acétate de sodium 3M. Les culots sont séchés et repris dans 40 μ l d'eau milliQ.

Cinq cent nanogrammes de produits circularisés sont ensuite amplifiés par le couple d'amorces BIO2 et BIO6, situées dans l'extrémité 3' du LTR. La réaction de PCR s'effectue avec 50 pmol de chaque oligonucléotide, une concentration finale en dNTP de 200 μ M chacun, et 2 U de *Taq* polymérase (Appligene) dans un volume réactionnel de 100 μ l (Tris-HCl 10 mM (pH 9) ; KCl 50 mM ; MgCl₂ 1,5 mM ; Triton X-100 0,1 % ; BSA 0,2 mg/mL). Les conditions d'amplification sont : 95°C 10 min, 40 fois (95°C 1 min, 58°C 1 min, 72°C 1 min), 72°C 10 min.

II.3.3. PCR avec extension d'amorce ou *Run off*.

Le polymorphisme de longueur des ampligènes est analysé par PCR radioactive. Cette PCR consiste en une PCR linéaire de 10 cycles, effectuée avec une amorce interne radiomarkée en 5'.

La réaction de marquage s'effectue sur 10 pmol de l'oligonucléotide BIO5 par incubation d'une heure à 37°C en présence de 30 μ Ci de γ ³²P ATP (Amersham) et de 10 U de Polynucléotide Kinase dans un volume réactionnel de 20 μ l (Tris-acétate 10 mM ; Acétate de magnésium 10 mM ; Acétate de potassium 50 mM). L'analyse des produits amplifiés s'effectue après électrophorèse en gel d'acrylamide dénaturant. Deux μ l de produits de PCR exponentielle sont soumis à une amplification linéaire de 10 cycles dans un volume réactionnel de 20 μ l (MgCl₂ 1,5 mM ; KCl 10 mM ; Tris-HCl 10 mM (pH 8,3)), contenant 0,1 pmol d'amorce BIO5 radiomarkée, 200 μ M de chaque dNTP et 1 U du fragment Stoffel de la *Taq* polymérase (Perkin Elmer). Les conditions d'amplification sont : 95°C 10 min, 10 fois (95°C 1 min, 58°C 1 min, 72°C 3 min), 72°C 10 min.

Les produits amplifiés, après dilution au demi dans une solution dénaturante de dépôt, sont dénaturés par chauffage pendant 5 min à 80°C. L'équivalent de 1,5 μ l de produits de *run off* est ensuite soumis à une électrophorèse en gel d'acrylamide, en conditions dénaturantes. Les ampligènes radioactifs sont ensuite révélés par autoradiographie sur phosphorimager.

Tableau 1M : Séquences et positions des amorces utilisées en LMPCR, IPCR et PCR pour l'amplification des sites d'intégration de HTLV-1, la construction du témoin et l'obtention des séquences contrôles.

Nom de l'oligonucléotide	Séquence 5'-3'	Orientation	Positions sur la séquence ATK1
BIO1	TCATGATCAATGGGACGATCACATG	anti-sens	
BIO2	CTGTTCTGCGCCGTTACAGATCGA	sens	8898-8921
BIO3	CCTTTCATTACGACTGACTGCCG	sens	8938-8961
BIO4	TCATGATCAATGGGACGATCA	anti-sens	
BIO5	TGGCTCGGAGCCAGCGACAGCCCAT	sens	8995-9019
BIO6	CTCCTGCTAGTTTATTGAGCCATA	anti-sens	8622-8598
MH2	CCCGCCAATCACTCATACAACC	sens	7592-7614
PBSI	CAGCCCCTCTTCCTTGAGTG	anti-sens	

II.3.4. PCR "classiques".

Les réactions de PCR dites classiques, sont effectuées sur 50 à 500 ng d'ADN génomique, dans un volume réactionnel de 100 µl et selon les conditions préconisées par le fournisseur de l'ADN polymérase utilisée.

II.4. Clonage des produits de PCR et séquençage des sites d'intégration.

II.4.1. Production des phages recombinants.

Les produits de PCR exponentielle, obtenus par IPCR ou LMPCR sont clonés dans le site *SmaI* du phage M13mp18, pour permettre leur séquençage. Pour favoriser l'insertion des produits de PCR de longueur variable dans le vecteur, et pour limiter la recircularisation du vecteur, les ampligènes sont phosphorylés et la forme répliquative du phage M13mp18 est déphosphorylée.

II.4.2. Préparation des ampligènes.

En général, 50 µl de produits amplifiés sont incubés 30 min à 37°C, en présence de 6 U du fragment Klenow de l'ADN polymérase I (Pharmacia) et 200 µM de chaque dNTP. Après extraction au phénol/chloroforme et précipitation par de l'éthanol 100%, l'ADN est repris dans 15 µl d'eau milliQ.

Les produits amplifiés sont ensuite purifiés par électrophorèse en gel d'agarose 1,5% à bas point de fusion (*Low melting point*, Tebu), afin d'éviter le clonage des clones "TG" et/ou des dimères d'amorces. En effet, l'intégration de HTLV-1, dont l'extrémité 3' est GTACACA, en 5' d'un dinucléotide TG, reconstitue le site de restriction de l'enzyme *NlaIII*, utilisée en LMPCR et en IPCR. Ainsi, la séquence 5'-...GTACACATG retrouvée après clonage correspond plus à une somme d'intégrations indépendantes qu'à un véritable "clone TG" en expansion. La purification est donc nécessaire pour éliminer les sites d'intégration non informatifs, ne comportant que le dinucléotide TG.

Les ampligènes ainsi purifiés sont ensuite phosphorylés par 10 U de Polynucléotide Kinase dans un volume réactionnel de 10 µl (Tris-acétate 10 mM ; Acétate de magnésium 10 mM ; Acétate de potassium 50 mM ; ATP 2 mM). La Polynucléotide Kinase est ensuite inactivée par un chauffage à 65°C, pendant 15 min. Sur ces 10 µl de produits amplifiés phosphorylés, 5 µl seront ligués dans le site *SmaI* du phage M13mp18.

II.4.3. Préparation du vecteur M13mp18.

Quatre microgrammes de la forme répllicative double brin du phage M13mp18 (Boehringer) sont digérés par 20 U de l'enzyme *SmaI* (Boehringer) pendant 2 h à 37°C, dans un volume réactionnel de 40 µl (Tris-acétate 10 mM (pH 7,5) ; Acétate de magnésium 10 mM ; Acétate de potassium 50 mM). Après vérification de la digestion par électrophorèse en gel d'agarose 1%, le vecteur est déphosphorylé par 20 U de Phosphatase Alcaline (Boehringer) pendant 30 min à 37°C. Pour permettre une meilleure accessibilité des groupements phosphates, une incubation à 55°C durant 10 min est réalisée. L'efficacité de cette réaction est améliorée par deux nouvelles incubations à 37°C et à 55°C avec 20 U supplémentaires d'enzyme. L'ADN plasmidique est purifié par extraction au phénol/chloroforme et précipité par de l'éthanol absolu. La quantité de plasmide digéré et déphosphorylé, obtenu après la purification et la précipitation, est évaluée à l'aide d'un spectrophotomètre, en mesurant la densité optique de l'échantillon à 260 nm.

II.4.4. Ligation des produits de PCR.

Cinq microlitres d'ampligènes phosphorylés sont ligués à 100 ng de vecteur M13mp18 déphosphorylé, dans un volume réactionnel de 20 µl (Tris-HCl 50 mM (pH 7,8) ; MgCl₂ 10 mM ; DTT 10 mM ; ATP 1 mM ; BSA 50 µg/mL), et en présence de 20 U de ligase

(Pharmacia). La ligation s'effectue au cours d'une première incubation d'une heure à température ambiante, suivie d'une incubation à 16°C sur la nuit. Pour augmenter l'efficacité de la réaction, 10 U de ligase supplémentaires, 2 µl de DTT 0,1 M et 1 µl d'ATP 10 mM sont ajoutés pour une nouvelle incubation de 6 h à 16°C. Après inactivation de la ligase par 10 min de chauffage à 65°C, les produits de ligation sont purifiés par extraction au phénol/chloroforme et précipités par de l'éthanol absolu. L'ADN ainsi purifié est ensuite repris dans 10 µl d'eau milliQ. La transformation des bactéries s'effectuant par électroporation, une dialyse d'une demi-heure contre de l'eau milliQ est nécessaire, pour éliminer les sels.

II.4.5. Electroporation des bactéries.

Les bactéries XL1 sont d'abord rendues compétentes pour l'électroporation, par des lavages successifs à l'eau froide.

En pratique, une préculture de bactéries XL1 estensemencée au 1/50^{ème} dans du milieu LB (Milieu Luria-Bertani) et mise en culture jusqu'à ce que la densité optique de la culture atteigne 0,8 à 600 nm. Les bactéries sont centrifugées 15 min à 4500 tours/min et reprises dans un volume identique d'eau milliQ à 4°C. Un deuxième lavage dans les mêmes conditions est effectué. Un troisième lavage est réalisé, cette fois dans 35 mL d'eau milliQ. Le culot est enfin repris volume à volume dans une solution de glycérol 10% puis dans une solution de glycérol 20%. Les bactéries sont immédiatement conservées à -80°C, aliquotées par 50 µl, après congélation instantanée dans un mélange éthanol/carboglace. Pour l'électroporation, 50 µl de bactéries compétentes et 5 µl de produits de ligation sont déposés dans une cuve à électroporation (Biorad). Après 5 minutes d'incubation dans la glace, l'électroporation est réalisée en utilisant les paramètres suivants : 2500 V, 25 µF, 700 Ω et 17,5 ms. Les bactéries électroporées sont ensuite régénérées dans 900 µl de milieu SOC-LB pendant 1 heure à 37°C. Suivant les rendements de clonage, de 0,1 à 100 µl de bactéries sont étalées en double couche dans 3 mL de Top-agar contenant 0,033% de X-Gal et 8,3 ng/mL d'IPTG. L'incubation se fait sur la nuit à 37°C.

II.4.6. Sélection des phages recombinants.

Les plages de lyse obtenues après électroporation sont criblées à l'aide de la sonde radioactive BIO5 sur des membranes de nitrocellulose. Après empreintes des boîtes de Pétri permettant le transfert des plages de lyse sur les membranes, et fixation de l'ADN aux UV, celles-ci sont préhybridées à 42°C pendant une heure dans un tampon d'hybridation (Phosphate de sodium 50 mM (pH 7,2) ; NaCl 750 mM ; EDTA 5 mM ; Denhardt's 5X ; Formamide 50% ; ADN de sperme de saumon dénaturé 20 µg/mL). Cinq picomoles d'amorce radiomarquée sont ensuite ajoutées au tampon pour une hybridation de 4 hrs minimum à 42°C. Les membranes sont lavées deux fois à température ambiante dans un tampon SSC 2X/SDS 0,1%. Une autoradiographie sur phosphorimager des membranes hybridées permet de repérer sur les boîtes de Pétri, les plages de lyse contenant les fragments d'intérêt.

II.4.7. Préparation des matrices M13 pour le séquençage.

Les matrices d'ADN à séquencer sont extraites des particules phagiques, produites lors d'une incubation à 37°C pendant 5 à 6 heures de 2,5 mL d'une préculture de XL1 diluée au 1/100^{ème} et infectée par repiquage des phages d'intérêt. Les bactéries sont sédimentées par 2 centrifugations successives de 10 min à 12000 g. 1,2 mL de surnageant contenant les particules phagiques produites sont prélevés et les phages sont précipités par 250 µl d'une solution de précipitation (PEG 6000 20% ; NaCl 2,5 M). Après une centrifugation de 15 min à 12000 g, le culot est repris par 200 µl de TE 1X. Une extraction au phénol à 50°C va permettre la purification du génome simple brin du phage. Après précipitation, les matrices sont reprises dans 20 µl de TE 1X et sont prêtes pour le séquençage. Le séquençage de 100 ng de matrice est effectué par 25 cycles d'extension à l'aide de l'amorce M13 (-20) (Fermentas) et de didéoxynucléotides fluorescents (Kit "*Bigdye Terminator Cycle Sequencing Ready Reaction*", Perkin Elmer). Les produits de chaque réaction sont finalement analysés sur séquenceur automatique ABI prism 310 ou 377 (Perkin Elmer).

Les séquences obtenues sont traitées par le logiciel PCGENE puis alignées sur les banques de données GENEbank et EMBL, grâce aux logiciels FASTA et BLAST. La recherche de séquences répétées dans le génome a été réalisée grâce au logiciel Repeat Masker (<http://ftp.genome.washington.edu/cgi-bin/RepeatMasker>).

II.5. Construction du témoin utilisé pour l'analyse de la composition nucléotidique de l'ADN non infecté.

La stratégie expérimentale utilisée pour amplifier l'ADN génomique non infecté est schématisée dans la Figure 2R (page 74). Les séquences des amorces utilisées pour la construction du témoin figurent dans le Tableau 1M.

II.5.1. Amplification d'un fragment du plasmide p4.39.

Le plasmide p4.39 a été obtenu après clonage de la séquence provirale HTLV-1 entière, encadrée des sites d'intégration d'origine cellulaire, dans le plasmide pBR327 à partir de la lignée 2060 (Nicot *et al.*, 1993). Cinq nanogrammes de plasmide p4.39 sont utilisés pour une amplification exponentielle de 40 cycles, à l'aide de l'amorce MH2, située en amont du LTR 3' et de l'amorce PBSI, complémentaire du site d'intégration du plasmide. Cette amplification est réalisée dans un volume réactionnel de 100 µl, avec 50 pmol de chaque amorce, une concentration finale de chaque dNTP de 200 µM et 2 U de *Taq* polymérase (Appligene). Les produits d'amplification exponentielle sont ensuite soumis à l'action de 4 U de Klenow, en présence de dNTP à 37°C, comme précédemment décrit. Les ampligènes sont purifiés par électrophorèse en gel d'agarose 1% à bas point de fusion (*Low melting*) puis phosphorylés par 10 U de Polynucléotide Kinase, selon le même protocole déjà détaillé ci-dessus.

II.5.2. Préparation de l'ADN génomique non infecté.

Parallèlement, 10 µg d'ADN génomique non infecté sont digérés par 100 U de l'enzyme *PvuII* (Eurogentec) et 100 U de l'enzyme *EcoRV* (Eurogentec) dans un volume réactionnel de 200 µl (Tris-HCl 50 mM ; MgCl₂ 10 mM ; NaCl 100 mM ; DTT 1 mM ; BSA 100 µg/mL), pendant 4 heures à 37°C. La digestion est vérifiée par électrophorèse en gel d'agarose à 0,8%. Cinq microgrammes d'ADN génomique non infecté sont également cassés par sonication. Dans les deux cas, l'ADN génomique digéré est purifié par extraction au phénol/chloroforme et précipité par de l'éthanol absolu. L'ADN est ensuite déphosphorylé par trois actions successives de 40 U de Phosphatase Alcaline, au cours d'incubations de 30 min à 37°C.

II.5.3. Ligation.

Deux cent nanogrammes de produits de PCR phosphorylés et 800 ng d'ADN génomique digéré et déphosphorylé, sont ligaturés sur la nuit à 16°C dans un volume réactionnel de 20 µl (Tris-HCl 50 mM ; MgCl₂ 10 mM ; DTT 10 mM ; ATP 1 mM ; BSA 50 µg/mL), par l'action de 40 U de ligase (Pharmacia). Les produits de ligation sont purifiés par extraction au phénol/chloroforme, précipité à l'éthanol absolu et le culot est repris dans 15 µl d'eau milliQ. L'ADN génomique ligué à l'extrémité 3' du LTR de HTLV-1 est ensuite amplifié par IPCR ou LMPCR. La réaction s'effectue sur 1 µg de produit de ligation, suivant la même démarche expérimentale que précédemment décrit. Cependant, les étapes de circularisation ou de ligation de l'amorce BIO1 ne sont effectuées que sur 400 ng de produit de digestion, dans les mêmes conditions de températures et de tampons. Une fois l'IPCR ou la LMPCR réalisée, les produits d'amplification sont clonés dans le vecteur M13mp18 puis séquencés, en utilisant la même approche expérimentale que pour les échantillons de patients.

II.6. Expression et purification des intégrases de HTLV-1 et de HIV-1.

II.6.1. Transformation des bactéries.

La souche bactérienne BL21DE3 (Novagen) est transformée par le plasmide pET22b-INHT lorsque l'on souhaite purifier l'IN du rétrovirus HTLV-1 et par le plasmide pET22b-INHV lorsque l'on souhaite purifier l'IN de HIV-1. En pratique, 2 ng de plasmide sont mis en présence de 20 µl de bactéries BL21DE3, compétentes pour la transformation. Les membranes bactériennes sont fragilisées par une incubation de 5 min sur la glace. Un choc thermique de 30 secondes à 42°C permet ensuite au plasmide de rentrer dans la bactérie. Les bactéries sont régénérées pendant une heure à 37°C, dans 80 µl de milieu SOC-LB (Life Technologies). Elles sont ensuite étalées sur des boîtes de Petri, contenant du LB gélosé et de l'ampicilline (50 µg/mL), pour permettre la sélection des clones recombinants.

II.6.2. Expression de la protéine.

Une préculture de bactéries BL21DE3 transformées est lancée sur la nuit, à 37°C, dans 100 mL de LB (ampicilline à 100 µg/mL). Le lendemain, les bactéries sont centrifugées pendant 15 minutes à 4500 tours/min. Le culot cellulaire est ensuite remis en suspension dans 100 mL de milieu 2XYT (ampicilline à 200 µg/mL) ; 50 mL de cette suspension bactérienne sont réensemencés dans 1 litre de milieu 2XYT (ampicilline à 200µg/mL). Après une heure de culture à 37°C sous agitation, l'expression de la protéine est induite par l'addition d'IPTG 1 mM, pendant 2 heures à 37°C.

II.6.3. Purification de la protéine en conditions dénaturantes.

Les bactéries sont récupérées par une centrifugation de 6 min à 7000 tours/min. Le culot bactérien est repris dans 80 mL de tampon A à pH 8 (Guanidine-HCl 6 M ; NaCl 0,5 M ; Tris 20 mM (pH 7,9) ; EDTA 0,1 mM ; DTT 1 mM), puis placé sous agitation pendant 1 h à température ambiante, pour permettre la lyse cellulaire et la solubilisation des protéines. L'extrait bactérien est centrifugé pendant 30 min à 20000 g, pour éliminer les débris cellulaires et les protéines insolubles.

Le surnageant contenant la protéine d'intérêt, est ensuite déposé sur une colonne d'affinité contenant 1 mL de résine NI-NTA agarose (résine nitrilotriacétate de nickel, Qiagen), préalablement équilibrée par environ 10 mL de tampon A à pH 8. Le passage du surnageant bactérien sur la colonne permet de purifier les protéines d'intérêt, portant en C-terminal une séquence poly-histidine affine pour le nickel. Après le passage de l'extrait protéique sur la colonne, les protéines non spécifiques sont éliminées par des lavages successifs avec du tampon A, dont le pH diminue progressivement ; le premier lavage, effectué avec 20 mL de tampon A à pH 8, est suivi par un lavage avec 20 mL de tampon A à pH 6,3 et enfin 20 mL de tampon A à pH 5,9. L'IN est finalement éluée après le passage de 10 mL de tampon A à pH 4,5, et récupérée en fractions de 1 ml environ.

II.6.4. Electrophorèse des protéines en gel de polyacrylamide.

L'expression et la pureté de la protéine sont vérifiées après migration des différentes fractions protéiques sur un gel de polyacrylamide dénaturant à 12% (ou "SDS Page" pour SDS *PolyAcrylamide Gel Electrophoresis*), suivie d'une coloration au bleu de Coomassie. La guanidine utilisée pour solubiliser les protéines peut gêner la migration. Il est donc nécessaire de purifier et de précipiter au préalable les protéines, avant le dépôt sur gel. En pratique, 100 µl de chaque fraction protéique sont précipités par 100 µl de TCA 10% (Acide TriChloroAcétique). Après 20 minutes d'incubation sur la glace, les protéines sont récupérées par une centrifugation de 10 min à 14000 g. Le culot protéique est ensuite lavé par 100 µl de TCA 0,1%, puis 500 µl d'éthanol à 70%, séché et repris dans 35 µl de tampon dénaturant (Urée 8M ; Glycérol 25% ; Tris-HCl 125 mM (pH 6,8) ; SDS 2% ; β-mercaptoéthanol 2% ; bleu de bromophénol 0,05%). Les protéines, dénaturées par un chauffage de 5 min à 95°C, sont déposées dans les puits d'un gel de tassement à 6% de polyacrylamide. Les protéines se concentrent à la limite du gel de tassement, avant de migrer et de se séparer en fonction de leur poids moléculaire dans un gel de séparation à 12% de polyacrylamide. L'électrophorèse est réalisée dans le tampon (Glycine 192 mM ; Tris 25 mM ; SDS 0,1 %), pendant une heure à 200 V. Un marqueur de poids moléculaire est déposé en même temps que les échantillons et permet de contrôler la migration et la taille des protéines. Le gel est ensuite coloré au bleu de Coomassie pendant une nuit à température ambiante, dans un mélange (Méthanol 50% ; Acide acétique 10 % ; R250 0,1%), puis décoloré dans le mélange méthanol/acide acétique (3/1).

Parallèlement, 5 µl de chaque fraction protéique sont dosés à l'aide du Kit Protein Assay ESL (Roche), par mesure de la densité optique à 485 nm, en spectrophotométrie d'absorption, selon la technique Biuret. Ce dosage nous permet ainsi d'évaluer la quantité totale de protéine obtenue après la purification.

II.6.5. Renaturation.

La protéine est renaturée par dialyse, organisée en plusieurs étapes. Les différentes fractions subissent des bains successifs d'urée, progressivement diluée dans du tampon B (NaCl 1 M ; Hepes 40 mM (pH 7,4) ; EDTA 2 mM ; BME 10 mM ; NP40 0,4% ; Glycérol 10%). L'agent

dénaturant est dilué toutes les 24 hrs dans du tampon B 1X, pendant 4 jours. Le dernier bain est dépourvu d'urée. Une dialyse supplémentaire des fractions est effectuée dans du tampon C 1X (Hepes 40 mM (pH 7,4) ; NaCl 200 mM ; EDTA 0,1 mM ; DTT 1 mM ; Glycérol 10% ; NP40 0,4%) pendant 36 hrs, avec renouvellement du tampon au bout de 24 hrs. Au cours de l'étape finale de dialyse, un précipité blanc se forme, que l'on récupère par centrifugation douce pendant 10 minutes. Le culot protéique est enfin repris dans un volume de tampon D (Hepes 40mM (pH 7,4) ; NaCl 500 mM ; EDTA 0,1 mM ; DTT 1 mM ; Glycérol 10%), aliquoté, puis rapidement congelé à -80°C .

II.6.6. Test de l'activité enzymatique de la protéine purifiée.

Afin de vérifier l'activité enzymatique de la protéine et de déterminer la concentration pour laquelle l'enzyme a une efficacité optimale, des réactions d'intégration sont réalisées avec différentes concentrations de l'enzyme. Des dilutions croissantes de la protéine en suspension sont effectuées dans de l'eau milliQ et une réaction de transfert de brin entre l'oligonucléotide représentant le LTR sauvage du virus correspondant, et l'oligonucléotide Courbe représentant la cible, est réalisée (Tableau 2M). Cinq microlitres et 1 μl de la suspension protéique de départ, 1 μl d'une dilution au $1/10^{\text{ème}}$ de la solution de départ, 1 μl d'une dilution au $1/20^{\text{ème}}$ et 1 μl d'une dilution au $1/50^{\text{ème}}$, sont testés. Les réactions de transfert de brin sont effectuées selon le protocole décrit ci-dessous.

II.7. Réactions d'intégration *in vitro*.

Les séquences nucléotidiques des oligonucléotides utilisés dans cette étude figurent dans le Tableau 2M et la stratégie expérimentale est schématisée dans la Figure 8R (page 95).

II.7.1. Hybridation des oligonucléotides cibles et des substrats donneurs.

Les oligonucléotides cibles, Courbe et Rigide, ainsi que les oligonucléotides représentant les extrémités 3' des LTR de HTLV-1 ou de HIV-1 sont rendus bicaténaires en présence de NaCl 0,1 M. En pratique, 20 pmol de chaque oligonucléotide monocaténaire est mis en présence de 100 pmol du deuxième oligonucléotide monocaténaire, dans un volume final de 20 μl d'eau milliQ. Une incubation de 2 min à 95°C précède une descente progressive en température, qui permet l'hybridation des deux oligonucléotides complémentaires.

II.7.2. Réactions de transfert de brin.

Une picomole de chaque oligonucléotide bicaténaire sont mises en présence de 1 µl d'IN à la concentration adéquate, dans un volume réactionnel de 15 µl (Hepes 25 mM (pH 7,4) ; BME 10 mM ; glycérol 10% ; MnCl_2 7,5 mM) pendant 1 hr à 37°C. Les produits d'intégration sont ensuite dilués dans un volume d'eau de 400 µl pour l'IN de HTLV-1 et de 75 µl pour l'IN de HIV-1.

II.7.3. Amplification des produits d'intégration.

Une PCR exponentielle à l'aide des amorces 5'-Fam VPM et PEXT1 (Figure 8R) est réalisée sur 10 µl de produits d'intégration dilués. Cette amplification s'effectue dans un volume réactionnel de 100 µl de tampon (Tris-HCl 20 mM ; MgSO_4 2 mM ; KCl 10 mM ; $(\text{NH}_4)_2\text{SO}_4$ 10 mM ; Triton X-100 0,1%) avec 25 pmol de chaque amorce, une concentration finale en dNTP de 200 µM chacun, et 3,5 U de *Pfu* polymérase (Stratagene). Les paramètres d'amplification sont : 95°C 10 min, 25 fois (95°C 1 min, 58°C 1 min, 72°C 1 min), 72°C 10 min.

II.7.4. Résolution des ampligènes sur séquenceur d'ADN ABI Prism 310.

Les produits d'amplification sont analysés sur séquenceur ABI Prism 310 grâce au logiciel Genescan (Perkin Elmer). Les ampligènes sont dilués au 1/10^{ème} et repris dans 12 µl de formamide désionisée, en présence de 1 µl de marqueur de taille TAMRA 500 (Perkin Elmer).

II.7.5. Clonage des produits d'intégration.

Le clonage des produits de PCR et le séquençage des matrices s'effectue selon le même protocole que précédemment décrit, dans le phage M13mp18. Les clones recombinants sont sélectionnés à l'aide de l'amorce VPM radioactive.

Tableau 2M : Séquences des oligonucléotides utilisés dans l'étude de l'intégration *in vitro*.

Nom de l'oligonucléotide	Taille en pb	Séquence (5'-3')
LTRHTLV	54	GAGCCAGCGACAGCCCATTCTATAGCACTCTCCAGGAGAGAACTT AGTACACA
LTRHIV	74	GAGCCAGCGACAGCCCATTCTGGTAACTAGAGATCCCTCAGACCC TTT TAGTCAGTGTGGAAAATCTCTAGCA
5'-Fam VPM	28	GCGAATTTCGAGCCAGCGACAGCCCATTC
ASVPM	20	TACTAAGTTTCTCTCCTGGA
PEXT1	28	GCGAATTCTGGTTCCGATGACCCTCGGA
SUB1	54	GAGCCAGCGACAGCCCATCTTATATCACTCTCCAGGAGAGAACTT AGTACACA
SUB2	54	GAGCCAGCGACAGCCCATCTTATAGCACTCTCCGGGAGAGAAATTT AGTACACA
SUB3	54	GAGCCAGCGACAGTCCATCTTATAGCACTCTCCAGGAGAGAAATTT AGTACACA
SUB4	54	GAACCAGCGACAGCCCATCTTATAGCACTCTCCAGGAGAGAAATTT AGTACACA
SUB5	54	GAGCCAGCGACAGCCCATCTTATAGCACTCTCCAGGAGAGAAATTT ATTACACA
SUB6	54	CGGGCCAGCGACAGCCCATCTTATAGCACTCTCCAGGAGAGAAATT AGTACACA
SUB7	54	GGAGCCAGCGACAGCCCATCTTATAGCACTCTCCAGGAGAGAAAT TAGTACACA
SUB8	54	GGGCCAGCGACAGCCCATCTTATAGCACTCTCCAGGAGAGAAATTT AGTACACA
COURBE	88	TGGTTCCGATGACCCTCGGATAGCGGCCTAAAATGGCCCTAAAATG GCCTAAAATGGCCCTAAAATGGCGCCTCGGCCTACGAATGCC
RIGIDE	88	TGGTTCCGATGACCCTCGGAGCCGAAAAAAAAAAAAACAAAAAAAA AAAAAAAAACAAAAAAAAAGCTAGCCCTCGGCCTACGAATGCC
P1CONS	54	GAGCCAGCGACAGCCCATCTTATAGCACTCTCCAGGAGAGAAATTT AGTACACA
P2CONS	54	GAGCCAGCGACAGCCCATCTTATAGCACTCTCCAGGAGAGAACT TAGTACACA
P5CONS	54	GAGCCAGCGACAGCCCATCCTATAGCACTCTCCAGGAGAGAAATT TAGTACACA
P6CONS	54	GAGCCAGCGACAGCCCATCTTATAGCACTCTCCAGGAGAGAAATTT AGTACACA

AUTRES TRAVAUX

La première partie de ce manuscrit a porté essentiellement sur l'étude des mécanismes moléculaires de l'intégration du rétrovirus HTLV-1, et la caractérisation de la cible de l'intégration du virus, *in vivo* et *in vitro*. Cette étude a constitué l'axe de recherche principal de ce projet et a donc été présentée en détail.

Ce travail de thèse a par ailleurs abordé certains aspects du mode réplcatif de HTLV-1 basé sur l'expansion clonale des cellules infectées. Ces différents travaux ont donné lieu à trois publications.

L'objectif de ce chapitre est de présenter, de manière synthétique, les résultats obtenus lors de ces différents travaux.

I. Immortalisation des clones cellulaires infectés par HTLV-1 *in vivo*.

Une première étude réalisée dans le laboratoire, a montré qu'une immortalisation des cellules infectées *in vivo* s'établissait dès les premiers stades de l'infection (Cavrois *et al.*, 1998). L'analyse, au cours du temps, par PCR "nichée" clonotypique de l'ADN de patients atteints de TSP/HAM, a montré que la majorité des clones infectés par HTLV-1 était stable sur une période de 11 mois à 5 ans, en absence de malignité. Pour ce faire, l'ADN des PBMCs de patients atteints de TSP/HAM a été amplifié à un temps donné. Les produits amplifiés ont été clonés et séquencés, et des oligonucléotides spécifiques des sites d'intégration ont été utilisés pour amplifier par PCR "nichée" l'ADN des PBMCs prélevés à d'autres temps. Par ce moyen, il a été trouvé que la majorité des clones persistait au cours du temps.

L'expansion clonale des cellules T infectées, gouvernée par la protéine virale Tax, peut donc persister pendant de longues périodes, malgré une forte immunité cellulaire. La latence prolongée de l'ATLL pourrait donc correspondre au délai nécessaire à l'obtention d'une masse critique de cellules, et donc à un état préleucémique. La croissance d'un ou plusieurs clone(s) malin(s) sur un bruit de fond de prolifération oligo/polyclonale suggère donc fortement que des événements génétiques secondaires sont nécessaires au développement de l'ATLL et à l'émergence d'un ou plusieurs clones transformants.

Persistent oligoclonal expansion of human T-cell leukemia virus type 1-infected circulating cells in patients with Tropical spastic paraparesis/HTLV-1 associated myelopathy

Marielle Cavrois^{1,7}, India Leclercq¹, Olivier Gout², Antoine Gessain³, Simon Wain-Hobson⁴ and Eric Wattel^{1,5,6}

¹Unité 124 INSERM, Institut de Recherche sur le Cancer de Lille; ²Fédération de Neurologie, Hôpital de la Pitié Salpêtrière, Paris; ³Unité d'Epidémiologie des Virus Oncogènes and ⁴Unité de Retrovirologie Moléculaire, Institut Pasteur, Paris; ⁵Unité d'Oncogenèse Virale, Centre Oscar Lambret, Lille; ⁶Service des Maladies du Sang, CHU, Lille, France

The pattern of HTLV-1 replication was assessed through PCR amplification of the 3' proviral integration sites in patients with TSP/HAM at different times. Integration sites were sequenced and oligonucleotides specific for the flanking sequences were synthesized. Together with HTLV-1 LTR specific primers, clonotypic nested PCR was performed on peripheral blood from two patients. The frequencies of five clones studied ranged from 1/300 to 1/1500 PBMCs while four clones persisted for more than 1–5 years. It would seem that Tax driven expansion of T cells may persist for considerable periods of time in TSP/HAM despite strong cellular immunity. This may provide a background for the accumulation of subsequent mutations leading to malignancy.

Keywords: HTLV-1; ATLL; leukemogenesis; TSP/HAM

Introduction

Human T-cell leukemia virus type 1 causes adult T cell leukemia/lymphoma (ATLL) (Poiesz *et al.*, 1980; Yoshida *et al.*, 1982) and TSP/HAM (Gessain *et al.*, 1985) both of which are preceded by a long asymptomatic period. Together with a remarkable genetic stability (Gessain *et al.*, 1992; Wattel *et al.*, 1995), the virus is characterized by high proviral load at each stage of the infection (Gessain *et al.*, 1990; Shinzato *et al.*, 1991; Wattel *et al.*, 1992). This results from the clonal expansion of infected cells (Wattel *et al.*, 1995). Proliferation of T-cell clones harbouring HTLV-1 has been evidenced *in vivo* in all asymptomatic carriers (Wattel *et al.*, 1996) as well as in patients with TSP/HAM (Cavrois *et al.*, 1996a) while ATLL appears on a background of HTLV-1 initiated clonal expansion (Cavrois *et al.*, 1996b).

Tax protein appears to act as a mitogen by inducing the expression of numerous genes involved in differentiation and proliferation of T-cells (Inoue *et al.*, 1986; Cross *et al.*, 1987; Fujii *et al.*, 1991; Duyao *et al.*, 1992).

In addition, through functional inactivation of p16^{INK4A}, Tax intervenes directly in the pathway controlling proliferation (Suzuki *et al.*, 1996). A corollary of Tax-driven T cell clonal expansion should be their persistence. Such clonal stability has been previously described in three patients with TSP/HAM and in two asymptomatic carriers (Furukawa *et al.*, 1992). A single clone was detected by Southern blotting in all five samples and could be found over a 4 month to 3 year period without any sign of transformation. Such cases represent less than 20% of TSP/HAM patients (Furukawa *et al.*, 1992). However, there is no doubt as to the enhanced sensitivity of PCR over Southern blotting. Indeed, it was possible through PCR based techniques to demonstrate clonal expansion of HTLV-1 harbouring T cells in all samples and disease stages (Wattel *et al.*, 1995; Cavrois *et al.*, 1996a,b). Here, the temporal stability of such clones has been examined by PCR based methods.

Results

Three HTLV-1 seropositive female patients were studied, all diagnosed TSP/HAM. Patient B was a 45-year old native of Guadeloupe at the time of diagnosis in 1977. Patient S was also a native from Guadeloupe. She was 53 year old at the time of diagnosis in 1977, while patient L was a 38-year old native from French Guyana, diagnosed in 1971. A brief clinical history is given in Figure 1 which also gives the dates when peripheral blood mononuclear cells (PBMCs) were taken (samples B1, B2, L1, L2, S1 and S2). PBMC DNA from three HTLV-1 seronegative individuals was used as a negative controls.

DNA was first analysed through quadruplicate inverse PCR (IPCR) as described in Methods. Run-off analyses of amplified products are shown in Figure 2 while Table 1 summarizes the number of clones detected in the seven samples. Although there was a wide distribution in the number of clones detected, a majority (78–100%) of clones was detected in ≤ 3 of quadruplicate amplifications of PBMC DNA meaning that clonal frequencies were $\leq 1/300$ PBMCs. Only in the case of the sample cultured from sample B2 were a majority (70%) of clones identified in quadruplicate amplifications (Table 1). All of the clones identified in the cell line had counterparts in the uncultured sample (Figure 2). Presumably *ex vivo* culture results in the selective amplification of a few clones. The total

Correspondence: E Wattel

⁷Current address: Institute of Human Virology, Medical Biotechnology Center, University of Maryland, 725 West Lombard Street, Baltimore, Maryland 21201, USA

Received 26 September 1997; revised 13 February 1998; accepted 13 February 1998

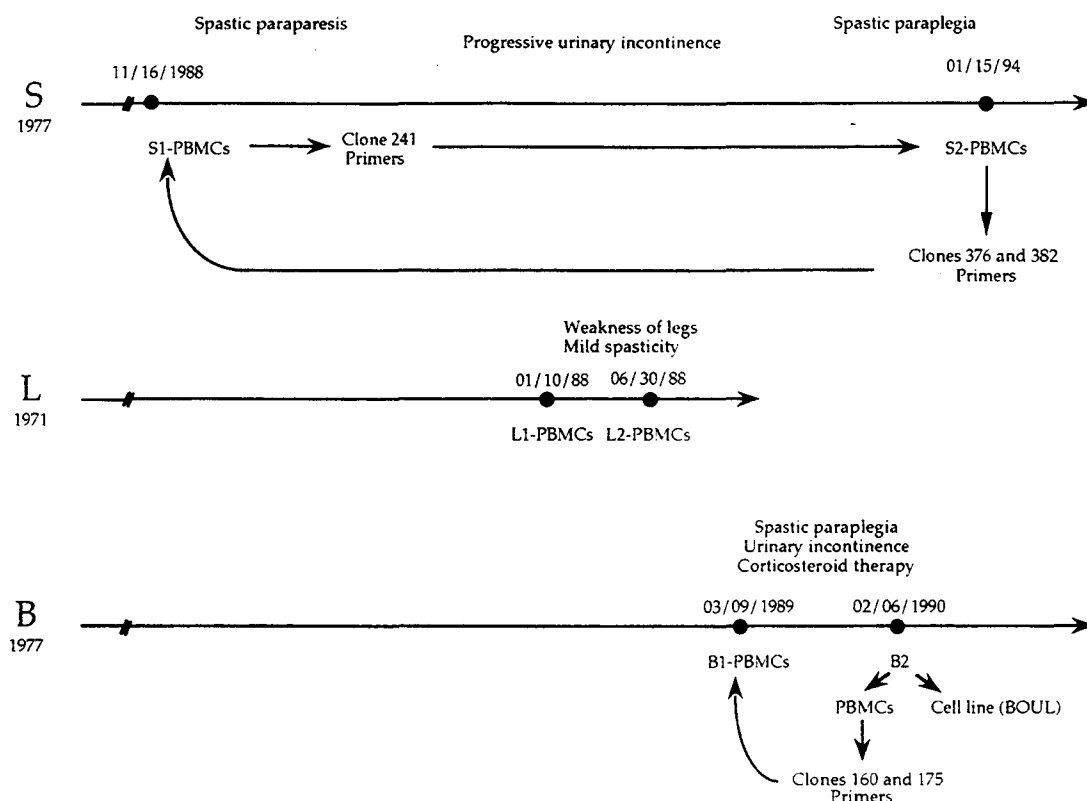


Figure 1 Profile of the three patients with TSP/HAM. The first date corresponds to that of the TSP/HAM diagnosis. Blood was taken at a number of intervals and the Ficoll-Hypaque-separated lymphocytes, referred to as S1 and S2, B1 and B2, L1 and L2, were frozen until used. The BOUL cell line was established from PBMCs of sample B2

Table 1 Distribution of the number and the abundance of detected clones after quadruplicate IPCR experiments

Sample	Date months	Total number of clones after four amplifications	Clones detected more than once after four amplifications (%)
B1, PBMCs	0	25	2 (8)
B2, PBMCs	+11	46	10 (22)
B2, cell line	+11	13	7 (54)
S1, PBMCs	0	~150	~23 (15)
S2, PBMCs	+62	6	0 (0)
L1, PBMCs	0	13	1 (8)
L2, PBMCs	+5.5	28	3 (11)

number of clones varied considerably over time. For patients B and L the number of clones detected in the periphery doubled in as little as 5.5 months (Figure 2a and Table 1) while for patient S there was a strong reduction of the number of clones over a 5 year period (~150 vs six clones). About 15% of the 150 clones present in the first sample were detected more than once after quadruplicate experiment. None of these clones were detected in the second sample S2.

The flanking cellular sequences from five clones from patients B and S are given in Figure 3. For each clone two oligonucleotides were chosen which, in conjunction with the HTLV specific oligonucleotides LTR1 and 2, allowed nested amplification of specific clones. When primers derived from clones 160 and 175 were used

with LTR1 and LTR2 to amplify DNA from samples B1 and B2, a specific band corresponding to the predicted size was obtained for both primer sets (Figure 4a). No signal was detected in DNA from the B2 derived cell line (Figure 4a). Clone 241, identified in sample S1, was undetectable in the second sample taken some 5 years later (Figure 4b). Of clones 376 and 382 identified in sample S2, only the former could be detected in the earlier sample S1 (Figure 4b and c). Assuming a detection threshold of ~ one copy sample for nested PCR (not shown), these results suggest that the frequency of the clones was <1/150 000 cells.

The frequency of the four stable clones was estimated from quadruplicate IPCR followed by radio-labelled integration site specific primer run-off of the amplification products (not shown). For clones 175 (sample B1) and 376 (sample S1) detection was 1/4. No signal was obtained for clones 160 and 382 in samples B1 and S1 respectively. Given previous results (Cavois *et al.*, 1996a,b), this indicates a frequency of $\leq 1/1500$ PBMCs.

Discussion

The results described above show considerable temporal variation in the number and frequency of IPCR detected T-cell clones harbouring HTLV-1 proviruses (Table 1). However, the sensitive clone specific nested

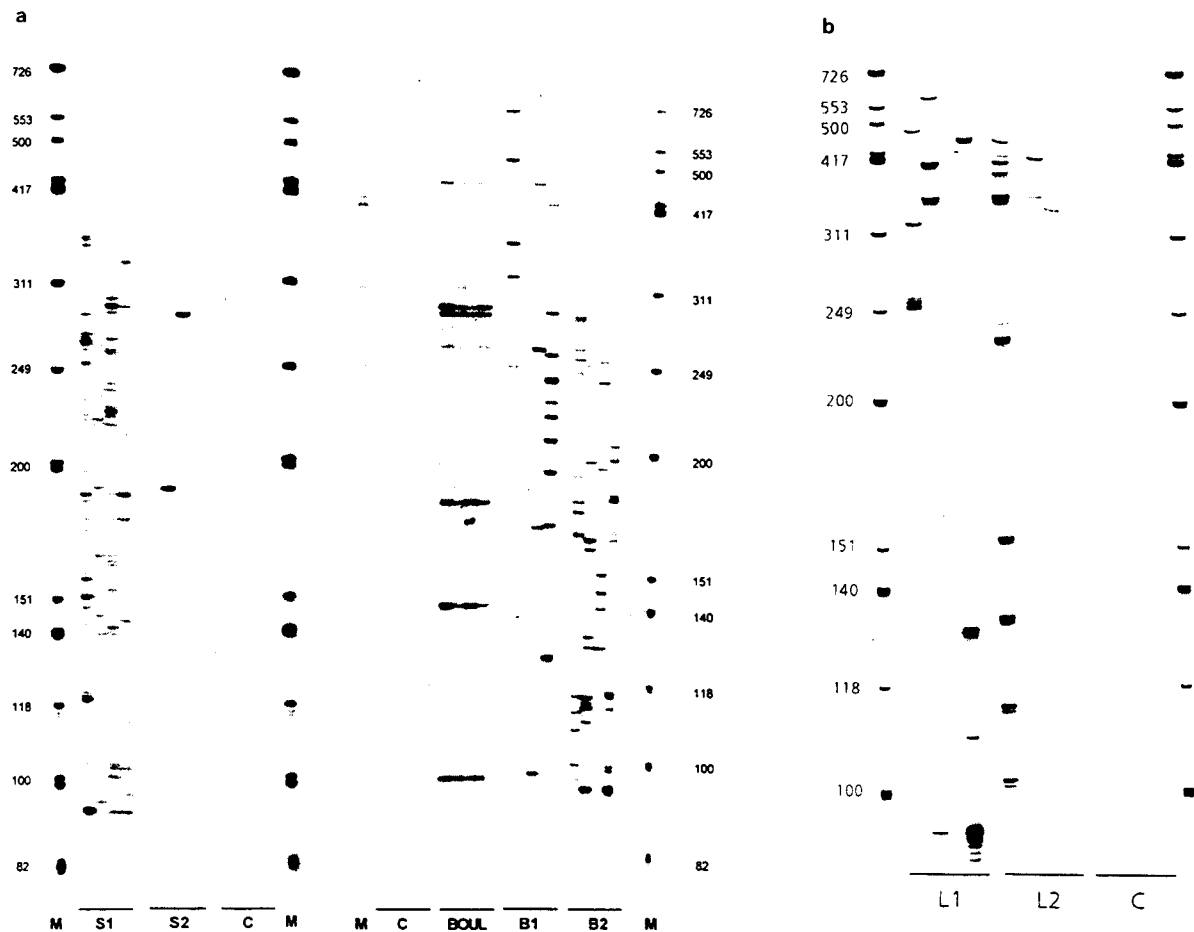


Figure 2 Quadruplicate IPCR analysis of DNA from uncultivated PBMCs of patient S, patient B (a) samples S1 and S2, B cell line ((BOUL), samples B1 and B2) and patient L (b) samples L1 and L2. The negative control, noted C, corresponds to Jurkat cell line DNA. In contrast to the BOUL T-cell line, all but one PBMCs sample harboured clones detected less than four times. Signals corresponding to clones C376 and C382 are clearly detected 1/4 in sample S2 at 194 and 267 bp, respectively. Given the high number of detected clones, clone C241 C160 and C175 corresponding to IPCR signals of 218, 196, 269 bp respectively, are difficult to identify here. However, they were detected less than four times

Sample	Clone	Flanking cellular sequence	A/T content
B2	C160 105bp	GAGAGTCAGC ATCCTTGTAG GACCTTGATT ATCAGGAGAA GGTAAGGCTG CTGTTACACG TTAAAAGAAA GGAAGAGTAT GTATAGTACC TTTGTGAAC ACATG	58%
B2	C175 178bp	ACATTTGATA AATTCAGTCC AACATTTCCA CCCCTCTCAG TCTTTGGGTT CCCGCCTCTA CATCGTGCCT TGGCATCTCA GTAGGATTAA CCATAGACTA CTGTTGAGTT AAATCTTCAA TAAATCCACC AAATAAGCTG TTAGAGTCAT TTCCAGTTGC ATG	59%
S1	C241 127bp	CTTCTCCATT TAGAAAGGAG TCATTTCAGG TCCCAGAAGC CTCCCCTGT CAGCCCCAAA ATTCCTCTTT TGCTTTCTCA CAGGATGTTT AGTGAAGAA AGCAGTTGGG AGAGGAGAAT CCTCATG	53%
S2	C376 101bp	TGTAAGGAA CAGAAATTAT AACAACTGT CTCTCAGACC ACAGTGCAAT CAAACTAGAA CTCAGGATTC AGAACTCAC TCAAACTGC TCAACTACAT G	62%
S2	C382 194bp	CCTCTGTAAT GGTATTTTCA GTTTAGGTAC ACGTGAAAAG AAAACTGCCT AATTTGAATA GTCCAGGAGA TGGTGAGAGA TTAGAAGGA GTGTTATAGA AGGGAATGAG TGAAGAAGCA GTAAATTAC TGTCAGGGTA AAATGGAATG ATGATTGACT ACTAAGAGCC AGGCCAGGAC TAAGTGCTTG CATG	60%

Figure 3 Sequences of HTLV-1 integration sites derived from patients B and S. All five sequences correspond to integration site between 3' LTR end and the first *Nla*III restriction site (bold face). Sequences underlined correspond to the oligonucleotides used to amplify flanking sequences, with LTR1 and LTR2. The size and the A/T content of the flanking sequences are given in base pairs and percentage, respectively

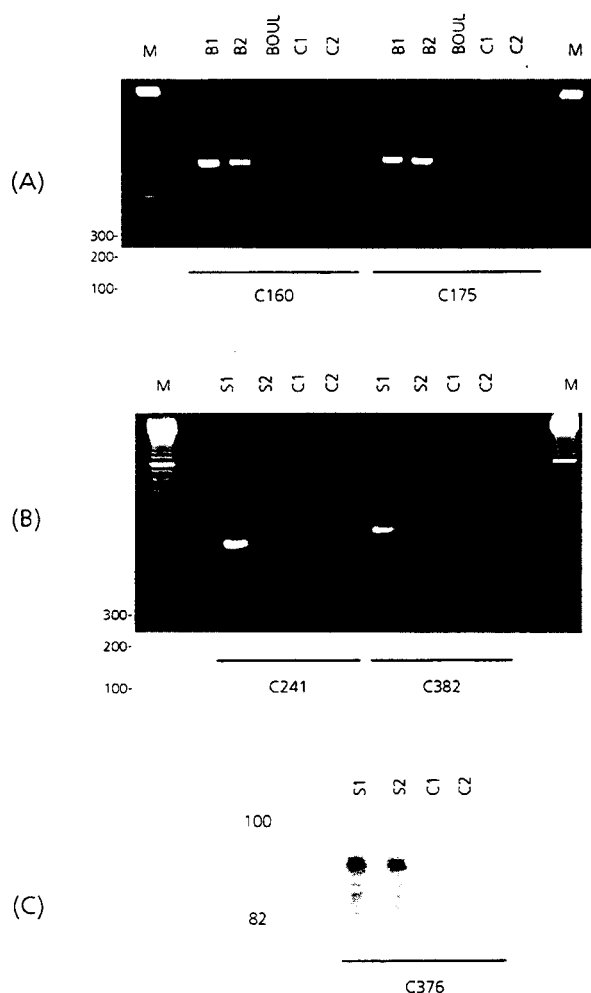


Figure 4 Nested PCR analysis of flanking sequences over time. The negative control, referred as C1 and C2, correspond to infected MT4 and non infected Jurkat cell lines, respectively. The molecular weight marker used was the 100 bp ladder (Pharmacia) except for the run off analysis. The dates the PBMCs were prepared are summarized in Figure 1. (a) Specific signals at 230 and 251 bp were obtained in samples B1 and B2 with primers derived from clones 160 and 175, respectively. These oligonucleotides failed to detect corresponding sequences in DNA from the T-cell line BOUL. (b) and (c) A specific signal corresponding to clone C241, present at 227 bp in sample S1, was not detected in sample S2. By contrast, clones C382 and C376 were present in both S1 and S2 samples at 277 and 90 bp, respectively. Run off analysis of nested PCR products was performed in order to detect specifically the clone C376 in S1 and S2 (C)

PCR using primers mapping to the 3' cellular flanking sequences showed that among the five clones analysed, four persisted over time with one being detectable for > 5 years. Temporally stable HTLV-1 clones have been previously shown in TSP/HAM patients by Southern-blotting (Furukawa *et al.*, 1992). However, rather than monoclonal expansion, TSP/HAM is characterized by oligo/polyclonal proliferation of infected cells. As four of the five clones analysed here remained stable over time, persistent clonal expansion of infected cells would seem to be the norm.

How can such clones remain stable over 1 to 6 year periods? Through interaction with several transcription factors such as CREB, NF- κ B, SRF or the transcriptional inhibitor I κ B (Zhao *et al.*, 1992; Suzuki *et al.*,

1993a,b; Fujii *et al.*, 1992; Hirai *et al.*, 1994). Tax transactivates transcription of the HTLV-1 genome and also that of many cellular genes that are involved in proliferation control of the infected cells (Inoue *et al.*, 1986; Cross *et al.*, 1987; Nagata *et al.*, 1989; Suzuki *et al.*, 1996). In addition, Tax expression *in vivo* has been evidenced by *in situ* hybridization (Gessain *et al.*, 1991; Beilke *et al.*, 1991; Cho *et al.*, 1995) technique or by reverse transcriptase-mediated polymerase chain reaction (Kinoshita *et al.*, 1989; Setoyama *et al.*, 1994; Furukawa *et al.*, 1995).

Clonal expansion is probably countered by an intense cytotoxic T cell (CTL) response which is particularly focused on Tax in cases of TSP/HAM (Jacobson *et al.*, 1990; Elovaara *et al.*, 1993; Parker *et al.*, 1994). The dynamic interplay of proliferation and destruction may partly explain the temporal variation in clonal frequencies. The stability of certain clones might simply represent an excess in this equation. However, an alternative such as clonal expansion in privileged sites, away from the wrath of the immune system, cannot be ruled out. A decrease in the CTL response may help some clone to proliferate. This was previously supposed in elderly carriers which harbour more abundant clones than do younger individuals (Cavois *et al.*, 1996a). Impairment of the CTL response may finally leads to clonal selection and emergence of an ATLL, as previously reported in several patients treated with immunosuppressive drugs (Jenks *et al.*, 1995).

HTLV-1 is the natural host of the CD45RA⁻ CD45RO⁺ CD4⁺ T-cell which is a memory cell (Richardson *et al.*, 1990). In a HTLV-1 bearing cell, viral transcription might result from any kind of host cell activation since the HTLV-1 LTR responds to some of the same physiological stimuli that mediate normal T-cell activation (Jeang *et al.*, 1988; Yin *et al.*, 1996; Newbound *et al.*, 1996). Accordingly, the specific immune activation of a HTLV-1 bearing CD45RO⁺ cell may result in viral transcription leading to Tax expression and anti-Tax cytotoxic CTL response. An intense CTL response might impaired the physiological burst of expansion of the host HTLV-1 bearing CD45RO⁺ cell as soon as viral expression occurs. Consequently, CTL control of cell proliferation may account for the impairment of immune responses in HTLV-1 infected individuals (Tachibana *et al.*, 1988; Robinson *et al.*, 1991). On the other hand, CTL response may have a positive effect on clonal proliferation and even permits expansion of some clones to the detriment of others depending upon their antigen specificity.

Any factor which triggers clonal expansion may shorten the latent phase preceding ATLL. It is well known that *Strongyloides stercoralis* (SS) infection stimulates clonal expansion in asymptomatic carriers since coinfection allows the virus to be detected via Southern-blotting (Nakada *et al.*, 1987). Interestingly, SS infected ATLL patients are significantly younger (mean age: 39 vs 70 years) than non infected patients with ATLL (Plumelle *et al.*, 1997).

Given the action of Tax on the cell cycle, temporal clonal stability is indicative of many rounds of replication which is probably countered by the cellular immunity. The former provides a propitious backdrop for the accumulation of subsequent mutations bringing

some clones closer to the transformed phenotype. The latter suggests that any decrease in cellular immunity, either naturally or resulting from the use of immunosuppressive drugs (Jenks *et al.*, 1995; Tsurumi *et al.*, 1992), may also contribute to the emergence of a malignant clone(s). This fits well with the frequent clonal change of ATLL at crisis (Tsukasaki *et al.*, 1997), suggesting that numerous premalignant clones are present when ATLL occurs.

Material and methods

Cell line

Frozen cells from sample B2 were thawed and 10 million were placed in a 50 ml culture flask containing 5 ml of RPMI 1640 medium with 20% heat-inactivated fetal calf serum, 1% L-glutamine, and 1% penicillin/streptomycin. During the 3 subsequent days, the cells were stimulated with phytohemagglutinin at 2 µg/10⁶ cells and then cultured in humidified 5% CO₂ in the same medium complemented by 10 U/ml IL-2 (Boehringer-Mannheim, Germany). After 4 months of culture, BOUL, a HTLV-1 cell line was established.

IPCR, cloning and sequencing

Inverse PCR (IPCR) and run-off analysis of amplified products were performed as previously described (Cavois *et al.*, 1996a,b). Estimation of the proportion of HTLV-1 positive clones was assessed by quadruplicate IPCR (Cavois *et al.*, 1996a,b). As previously described, there is a stochastic element to the amplification of low frequency HTLV-1 integration sites (Cavois *et al.*, 1995, 1996a,b). Quadruplicate IPCR analysis of *Nla*III digested p4.39 DNA, bearing an integrated HTLV-1 provirus, showed that above 1000 copies detection was 4/4. However at 500, 100 and ≤50 copies detection was 3/4, 1/4 and 0/4 respectively. Accordingly, an integration site detected more than once in four experiments corresponds to a clonal frequency of ~1/1500 PBMCs. Amplified products from uncultivated samples B2, S1 and S2 were purified from a low melting point agarose gel. Purified products >150 bp were blunt end ligated into M13mp18 replicative-form DNA and cloned as previously described (Vartanian

et al., 1991). Clones were screened *in situ* with a ³²P-labelled BIO5 oligonucleotide probe (Wattel *et al.*, 1995). Single-strand DNA templates were sequenced by standard dideoxy method using [α-³⁵S]dATP. Clones with large flanking cellular sequences were identified and specific oligonucleotides synthesized.

Nested PCR

For each HTLV-1 clone two oligonucleotides corresponding to the flanking cellular sequences were used in combination with HTLV-1 LTR specific primers LTR1, 5'-TCGCATCTCTCCTTCACGCG (positive strand, 8656-→8676) and LTR2, 5'-CTGACCCTGCTTGCTCAACTCTACG (positive strand, 8858-→8883). Coordinates are those of the λ ATK-1 sequence. Sequences and positions of integration sites specific oligonucleotides are shown on Figure 3. The first round of amplification was performed with 1 µg of DNA. DNA was mixed with 90 µl PCR buffer (50 mM KCl, 10 mM Tris.HCl pH 8.3), 2.5 mM MgCl₂, 200 µM of each dNTP, 0.5 µM of each external primer and 2.5 U of Taq polymerase. The samples were overlaid with 35 µl of mineral oil and amplified for 30 cycles. Thermal cycling parameters were: denaturation, 94°C for 1 min; annealing, 59°C for 1 min; extension, 72°C for 1 min. Ten microliters of amplified products were used in the second PCR with internal primers. PCR mix and thermal cycling parameters were the same as for the first round.

Cells cultures, *Nla*III digestions, ligations and first round PCR were carried out in labs which had never handled a HTLV-1 plasmid. Second round and radio-labelled amplification were made in yet another laboratory. No contamination was detected.

Acknowledgements

This work was supported by grants from the Association pour la Recherche sur le Cancer and from the Ligue Nationale contre le Cancer (Comité Pas de Calais). MC and IL were supported by bursaries from the Ministère de l'Enseignement Supérieur et de la Recherche. We would like to thank Pr Stehelin, Pr Wattre and collaborators who kindly received us in their labs for DNA extraction, digestion, ligation and PCR. We would also like to thank Michel Tavet and Mickael Vanrumbeke for assistance.

References

- Beilke MA, In DR, Gravel M, Hamilton RS, Mora CA, Leon-Monzon M, Rodgers-Johnson PE, Gajdusek DC, Gibbs Jr C and Zaninovic V. (1991). *J. Med. Virol.*, **33**, 64-711.
- Cavois M, Wain-Hobson S and Wattel E. (1995). *Res. Virol.*, **146**, 179-184.
- Cavois M, Gessain A, Wain-Hobson S and Wattel E. (1996a). *Oncogene* **12**, 2419-2423.
- Cavois M, Wain-Hobson S, Gessain A, Plumelle Y and Wattel E. (1996). *Blood*, **88**, 4646-4650.
- Cho I, Sugimoto M, Mita S, Tokunaga M, Imamura F and Ando M. (1995). *Am. J. Trop. Med. Hyg.*, **53**, 412-418.
- Cross SL, Feinberg MB, Wolf JB, Holbrook NJ, Wong-Staal F and Leonard WJ. (1987). *Cell*, **49**, 47-56.
- Duyao MP, Kessler DJ, Spicer DB and Sonenshein GE. (1992). *AIDS. Res. Hum. Retroviruses*, **8**, 752-754.
- Elovaara I, Koenig S, Brewah AY, Woods RM, Lehty T and Jacobson S. (1993). *J. Exp. Med.*, **177**, 1567-1573.
- Fujii M, Niki T, Mori T, Matsuda T, Matsui M, Nomura N and Seiki M. (1991). *Oncogene*, **6**, 1023-1029.
- Fujii M, Tsuchiya H, Chuhjo T, Akizawa T and Seiki M. (1992). *Genes Dev.*, **6**, 2066-2076.
- Furukawa Y, Fujisawa J, Osame M, Toita M, Sonoda S, Kubota R, Ijichi S and Yoshida M. (1992). *Blood*, **80**, 1012-1016.
- Furukawa Y, Osame M, Kubota R, Tara M and Yoshida M. (1995). *Blood*, **85**, 1865-1870.
- Gessain A, Barin F, Vernant JC, Gout O, Maurs L, Calender A and de-Thé G. (1985). *Lancet*, **2**, 407-410.
- Gessain A, Gallo RC and Franchini G. (1992). *J. Virol.*, **66**, 2288-2295.
- Gessain A, Louie A, Gout O, Gallo RC and Franchini G. (1991). *J. Virol.*, **65**, 1628-1633.
- Gessain A, Saal F, Gout O, Daniel MT, Flandrin G, de-Thé G, Peries J and Sigaux F. (1990). *Blood*, **75**, 428-433.
- Hirai H, Suzuki T, Fujisawa J, Inoue J and Yoshida M. (1994). *Proc. Natl. Acad. Sci. USA*, **91**, 3584-3588.
- Inoue J, Seiki M, Taniguchi T, Tsuru S and Yoshida M. (1986). *EMBO J.*, **5**, 2883-2888.

- Jacobson S, Shida H, McFarlin DE, Fauci AS and Koenig S. (1990). *Nature*, **348**, 245–248.
- Jeang KT, Boros I, Brady J, Radonovich M and Khoury G. (1988). *J. Virol.*, **62**, 4499–4509.
- Jenks PJ, Barrett WY, Raftery MJ, Kelsey SM, van-der-Walt JD, Kon SP and Breuer J. (1995). *Clin. Infect. Dis.*, **21**, 992–993.
- Kinoshita T, Shimoyama M, Tobinai K, Ito M, Ito S, Ikeda S, Tajima K, Shimotohno K and Sugimura T. (1989). *Proc. Natl. Acad. Sci. USA*, **86**, 5620–5624.
- Nagata K, Ohtani K, Nakamura M and Sugamura K. (1989). *J. Virol.*, **63**, 3220–3226.
- Nakada K, Yamaguchi K, Furugen S, Nakasone T, Nakasone K, Oshiro Y, Kohakura M, Hinuma Y, Seiki M, Yoshida M, Matutes E, Catovsky D, Ishii T and Takatsuky K. (1987). *Int. J. Cancer*, **40**, 145–148.
- Newbound GC, Andrews JM, O'Rourke JP, Brady JN and Lairmore MD. *J. Virol.*, **70**, 2101–2106.
- Parker CE, Nightingale S, Taylor GP, Weber J and Bangham CR. (1994). *J. Virol.*, **68**, 2860–2868.
- Plumelle Y, Gonin C, Edouard A, Bucher BJ, Thomas L, Brebion A and Panelatti G. (1997). *Am. J. Clin. Pathol.*, **107**, 81–87.
- Poiesz BJ, Ruscetti FW, Gasdar AF, Bunn FA, Minna JD and Gallo RC. (1980). *Proc. Natl. Acad. Sci. USA*, **77**, 7415–7421.
- Richardson JH, Edwards AJ, Cruickshank JK, Rudge P and Dalgleish AG. (1980). *J. Virol.*, **64**, 5682–5687.
- Robinson RD, Murphy EL, Wilks RJ, Neva FA, Terry SI, Hanchard B, Figueroa JP and Blattner WA. (1990). *J. Trop. Med. Hyg.*, **94**, 411–415.
- Setoyama M, Fujiyoshi T, Mizoguchi S, Katahira Y, Yashiki S, Tara M, Kanzaki T and Sonoda S. (1994). *Int. J. Cancer*, **57**, 760–764.
- Shinzato O, Ikeda S, Momita S, Nagata Y, Kamihira S, Nakayama E and Shiku H. (1991). *Blood*, **78**: 2082–2088.
- Suzuki T, Fujisawa JI, Toita M and Yoshida M. (1993a). *Proc. Natl. Acad. Sci. USA*, **90**, 610–614.
- Suzuki T, Hirai H, Fujisawa J, Fujita T and Yoshida M. (1993b). A trans-activator Tax of human T-cell leukemia virus type 1 binds to NF-kappa B p50 and serum response factor (SRF) and associates with enhancer DNAs of the NF-kappa B site and CArG box. *Oncogene*, **8**, 2391–2397.
- Suzuki T, Kitao S, Matsushime H and Yoshida M. (1996). *EMBO J.*, **15**, 1607–1614.
- Tachibana N, Okayama A, Ishizaki J, Yokota T, Shishime E, Murai K, Shioiri S, Tsuda K, Essex M and Mueller N. (1988). *Int. J. Cancer*, **42**, 829–831.
- Tsukasaki K, Tsushima H, Yamamura M, Hata T, Murata K, Maeda T, Atogami S, Sohda H, Momita S, Ideda S, Katamine S, Yamada Y, Kamihira S and Tomonaga M. (1997). *Blood*, **89**, 948–956.
- Tsurumi H, Tani K, Tsuruta T, Shirato R, Matsudaira T, Tojo A, Wada C, Uchida H, Ozawa K and Asano S. (1992). *Am. J. Hematol.*, **41**, 292–294.
- Vartanian JP, Meyerhans A, Åsjö B and Wain-Hobson S. (1991). *J. Virol.*, **65**, 1779–1788.
- Wattel E, Cavois M, Gessain A and Wain-Hobson S. (1996). *J. Acquir. Immun. Defic. Syndr. Hum. Retrovirol.*, **13** (Suppl 1), 92–99.
- Wattel E, Mariotti M, Agis F, Gordien E, Le-Coeur FF, Prin L, Rouger P, Chen IS, Wain-Hobson S and Lefrere JJ. (1992). *J. Acquir. Immune. Defic. Syndr.*, **5**, 943–946.
- Wattel E, Vartanian JP, Pannetier C and Wain-Hobson S. (1995). *J. Virol.*, **69**, 2863–2868.
- Yin MJ, Paulssen E and Gaynor RB. (1996). *J. Biol. Chem.*, **271**, 4781–4790.
- Yoshida M, Miyoshi I and Hinuma Y. (1982). *Proc. Natl. Acad. Sci. USA*, **79**, 2031–2035.
- Zanke BW, Rush DN, Jeffery JR and Israels LG. (1989). *Transplantation*, **48**, 695–697.
- Zhao LJ and Giam CZ. (1992). *Proc. Natl. Acad. Sci. USA*, **89**, 7070–7074.

II. Prolifération oligo/polyclonale des cellules T infectées par HTLV-1, en absence de réarrangements des sites d'intégration.

II.1. Introduction.

Une étude récente, basée sur l'analyse par quadruplicate IPCR, de l'ADN de 22 patients atteints d'ATLL, a montré qu'il existait une grande quantité de clones minoritaires au sein même du tissu tumoral. A titre d'exemple, 80% des clones isolés d'un ganglion lymphatique correspondaient à des formes minoritaires (Cavrois *et al.*, 1996b). Un nombre important de clones minoritaires est également présent dans les prélèvements non tumoraux, comme les PBMCs des formes lymphomateuse et progressive.

Une telle abondance de formes oligo/polyclonales pouvait s'expliquer par des réarrangements des sites d'intégration, survenant au cours de la prolifération des cellules infectées. En effet, une délétion du site d'intégration flanquant le LTR 3' du provirus pourrait modifier la longueur de la région flanquante et aboutir ainsi, après mitose et après analyse par IPCR ou LMPCR, à la détection de deux signaux, correspondant en fait à un seul clone en expansion. Récemment, deux réarrangements analogues ont été mis en évidence chez 2 porteurs asymptomatiques différents (Wattel *et al.*, 1995). Des délétions/réarrangements de la séquence flanquante peuvent donc survenir au cours de l'expansion clonale des cellules infectées, en absence de malignité. Ce type de réarrangements reste malgré tout un événement relativement rare au cours de la phase asymptomatique de l'infection. Par contre, des délétions du site d'intégration flanquant le LTR 3' du provirus pourraient être plus fréquentes au cours de l'ATLL. Il existe en effet une forte instabilité génétique au niveau du locus du provirus correspondant au clone tumoral dans l'ATLL. Ainsi, des délétions couvrant la région 5' du provirus sont retrouvées chez presque 30% des patients atteints d'ATLL, et des délétions interstitielles ont récemment été mises en évidence dans 20% des cas. De telles délétions ont par ailleurs été retrouvées au stade asymptomatique (Hiramatsu *et al.*, 1986 ; Ohshima *et al.*, 1991 ; Tamiya *et al.*, 1996 ; Tsukasaki *et al.*, 1997). Des réarrangements de la séquence génomique, flanquant le provirus, pourrait donc expliquer l'abondance des formes oligo/polyclonales observées dans l'ATLL.

Afin d'infirmer ou d'affirmer cette hypothèse, 10 clones minoritaires provenant de l'ADN tumoral de 5 patients atteints de différentes formes d'ATLL, ainsi que le ou les clone(s) majoritaires(s) correspondant(s), ont été isolés par IPCR, puis séquencés. Les différents sites d'intégration obtenus ont été alignés pour rechercher d'éventuelles identités de séquence et mettre ainsi en évidence des réarrangements des sites d'intégration.

L'ADN a été extrait à partir des PBMCs, ou à partir des ganglions lymphatiques. L'ADN a d'abord été analysé par blot de Southern puis par quadruplicate IPCR. Les produits de PCR exponentielle provenant des échantillons tumoraux ont été clonés dans le phage M13mp18, puis séquencés, comme décrit dans le matériels et méthodes. Les séquences obtenues ont été alignées entre elles grâce au logiciel PCGENE, afin de rechercher d'éventuelles identités.

II.2. Résultats et discussion.

Aucune délétion des sites d'intégration n'a été retrouvée.

Cette étude montre que, comme au stade asymptomatique, les délétions ou réarrangements des sites d'intégration flanquant le LTR 3' du provirus sont rares au cours de l'ATLL. Ils ne peuvent donc pas expliquer l'abondance des signaux détectés après IPCR. Le provirus semble donc relativement stable et ne semble pas subir de remaniements chromosomiques au cours de l'infection.

Les études précédentes ont montré que le mode de réplication par expansion clonale des cellules infectées prédominait au cours de l'infection par HTLV-1. Par conséquent, le profil réplcatif de HTLV-1 observé par IPCR au cours de l'ATLL, correspond à une prolifération monoclonale sur un bruit de fond d'expansion oligo/polyclonale de cellules non transformées. L'hypothèse d'une réintégration du provirus dans l'ADN génomique peut être envisagée pour expliquer la présence de ces clones minoritaires. Des particules virales produites au cours de l'affection pourraient réinfecter la tumeur, entraînant une expression transitoire de Tax, aboutissant ainsi au maintien du phénotype malin.

Oligoclonal proliferation of human T-cell leukaemia virus type 1 bearing T cells in adult T-cell leukaemia/lymphoma without deletion of the 3' provirus integration sites

INDIA LECLERCQ,¹ MARIELLE CAVROIS,^{1*} FRANCK MORTREUX,¹ OLIVIER HERMINE,² ANTOINE GESSAIN,³ FRANCK MORSCHHAUSER^{1,4} AND ERIC WATTEL^{1,4,5} ¹Unité 124 INSERM, Institut de Recherche sur le Cancer, Lille, ²Service d'Hématologie, Hôpital Necker, Paris, ³Unité d'Epidémiologie des Virus Oncogènes, Institut Pasteur, Paris, ⁴Service des Maladies du Sang, CHU, Lille, and ⁵Unité d'Oncogénèse Virale, Centre Oscar Lambret, Lille, France

Received 18 December 1997; accepted for publication 12 March 1998

Summary. We report a new case of an asymptomatic carrier with a deletion of a 3' HTLV-1 integration site. We further investigated whether these 3' deletions of flanking sequences may explain the oligoclonal pattern of HTLV-1 replication, evidenced by inverse PCR (IPCR) analysis of tumourous samples from patients with adult T-cell leukaemia (ATLL). 48 HTLV-1 3' integration sites, derived from tumourous DNA of five ATLL patients were sequenced. One dominant flanking sequence was obtained in the four samples harbouring a unique band after Southern-blotting. In one sample, which harboured two signals after Southern-blotting, IPCR amplification of diluted tumourous DNA revealed that these two sequences corresponded to one clone harbouring two integrated proviruses rather than to two distinct cellular

clones, a result consistent with superinfection of the tumourous sample. In addition to integration sites corresponding to malignant clones, two to six oligoclonal forms were sequenced in four samples. No flanking sequence homology was found between clones derived from each patient, indicating that integration sites deletion in the vicinity of the provirus is a rare event in ATLL. The oligoclonal pattern of HTLV-1 replication in ATLL may result from clonal expansion of non-malignant HTLV-1-bearing clones within the sample and partly from HTLV-1 superinfection of monoclonal tumour cells.

Keywords: HTLV-1, adult T-cell leukaemia, leukaemogenesis, genetic instability, deletion.

HTLV-1 causes ATLL (Poiesz *et al.* 1980; Yoshida *et al.* 1982) and tropical spastic paraparesis/HTLV-1 associated myelopathy (TSP/HAM) (Gessain *et al.* 1985). Monoclonal integration of the provirus in the DNA from tumourous cells is the hallmark of ATLL (Yoshida *et al.* 1984). In addition to Southern blotting, molecular detection of HTLV-1 bearing T-cell clones can be assessed by inverse PCR (IPCR) (Takemoto *et al.* 1994; Cavrois *et al.* 1995, 1996a, b) or ligation mediated PCR (LMPCR) (Cavrois *et al.* 1996a; Wattel *et al.* 1995). Both methods combine the sensitivity of the PCR with the analysis of integration sites restriction length polymorphism. By these means, recent works have shown that oligoclonal expansion of numerous HTLV-1 bearing cells

invariably accompanied the malignant clone(s) in tumourous samples from ATLL patients (Takemoto *et al.* 1994; Cavrois *et al.* 1996b).

As there is no preferential target site for HTLV-1 integration and assuming that there is one integrated provirus per infected cell, each signal obtained after IPCR or LMPCR corresponds to a clone of infected cells. However, mutations or deletions of the 3' integration site within a clone may modify the restriction length of the flanking region and provide, after mitosis, the detection of two distinct signals both corresponding to a single expanded clone. Such a 3' deletion has been recently found in one expanded clone from an asymptomatic carrier (Wattel *et al.* 1995). We report here a new deletion of a 3' flanking sequence in an additional carrier. This confirms that additional mutational events may occur during clonal expansion *in vivo* even in the absence of malignancy. However, these rearrangements appear to be a rare event in asymptomatic carriers.

Since there is a high level of genetic instability at the

* Present address: Institute of Human Virology, Medical Biotechnology Center, University of Maryland, 725 West Lombard Street, Baltimore, MD 21201, U.S.A.

Correspondence: Dr E. Wattel, Service des Maladies du Sang, CHU, 1 Place Verdun, 59037 Lille, France.

provirus locus harboured by the tumourous clone during ATLL (Ohshima *et al.* 1991; Shimamoto *et al.* 1994; Tamiya *et al.* 1996; Tsukasaki *et al.* 1997), one can hypothesize that microdeletions of the 3' integration sites may be more frequent in ATLL than at the asymptomatic phase of the infection. This could contribute to the IPCR pattern of oligoclonal expansion which accompanies the malignant clone in ATLL tumourous samples. Therefore we looked at the incidence of such microdeletions in ATLL by sequencing 48 integration sites derived from tumourous samples of five patients.

MATERIALS AND METHODS

Samples studied. Patients P4, P12, P20 and P25 have been described elsewhere (Cavrois *et al.* 1996b). All fulfilled ATLL diagnostic criteria and were monoclonal for HTLV-1 integration as evidenced by Southern blot. Patient P30 was a seropositive 45-year-old male from French Guyana with a 10-year history of *Strongyloides stercoralis* infection. He presented with peripheral, thoracic and abdominal lymph node enlargement. Lymph node biopsy exhibited a tumoural infiltration with pleiomorphic T-cell CD4⁺, CD25⁺ lymphoma. White blood count, blood smears, bone marrow aspirate and bone marrow biopsy were normal. LDH level was elevated (3 times normal) and calcium level was normal. Therefore he fulfilled diagnostic criteria of lymphoma ATLL. Treatment with AZT (1 g/d) and interferon (9×10^6 units/d) was started, leading to a 50% partial remission. Combined chemotherapy was then started with only a poor response, and the patient died 6 months after diagnosis. In all cases DNA was extracted from tumourous samples at the time of diagnosis. In addition to lymph node DNA, a sample from PBMCs of patient P30 was also studied. P19 was an asymptomatic carrier who has been described elsewhere (Wattel *et al.* 1995). G18 was a 45-year-old blood donor from Guadeloupe, French West Indies.

Southern blot. Approximately 10 µg of high molecular weight DNA originating from PBMCs and tumoural lymph node from patient P30 was digested with *EcoRI* or *Pst I* and then electrophoresed through a 0.6% agarose gel. After Southern blotting to a nylon membrane, the filter was hybridized with the randomly-primed ³²P-labelled PMT-2-3 probe which correspond to the 1.7 kb fragment of the gag-pol region obtained after digestion of MT2 cell line DNA by *Pst I*.

PCR amplification of HTLV-1 integration sites. IPCR analysis of HTLV-1 integration sites is a rapid and useful tool in the detection of infected clones (Takemoto *et al.* 1994; Cavrois *et al.* 1995, 1996a, b). Since the sensitivity of this method precludes the detection of a single infected cell, each signal obtained after run-off analysis of amplified products corresponds to a clone of infected cells. In addition, as previously detailed (Cavrois *et al.* 1995, 1996a, b), quadruplicate analysis of infected DNA allows a semi-quantitative detection of infected clones, since the more abundant are more frequently detected. Amplifications were performed as previously described using BIO2 and BIO6 as primers. PCR products were analysed by run-off using BIO5 as primer

(Cavrois *et al.* 1995, 1996a, b). As there is a stochastic element to the detection of HTLV-1 integration sites through IPCR (Cavrois *et al.* 1995), samples were analysed through quadruplicate experiments, as previously described (Cavrois *et al.* 1995, 1996a, b). In addition, IPCR was performed with 500, 250, 100, 150, 10, 6.6, 1 and 0.1 ng of tumourous DNA from patient P30, diluted in non-infected DNA, in order to amplify the most abundant clone(s). Sample from G18 was amplified by Linker mediated PCR (LMPCR), using BIO1-5 as primers as previously described (Wattel *et al.* 1995).

Cloning and sequencing integration sites. IPCR and LMPCR products were phosphorylated by T4 polynucleotide kinase and ligated with *SmaI*-digested and dephosphorylated M13mp18 replicative form DNA as previously described (Wattel *et al.* 1995). After transformation of *Escherichia coli* XL1 by electroporation, recombinant M13 plaques were screened by hybridization with the HTLV-1 LTR-specific ³²P-labelled oligonucleotide BIO5. For each sample, six to 22 positive plaques were picked and prepared for DNA sequencing. Single-stranded templates were carried out by the standard dideoxy method with α-³⁵S dATP and resolved on buffer gradient gel.

RESULTS

All *Nla III* digestions, ligations and first-round PCR were carried out in laboratories which had never handled an HTLV-1 plasmid. Radiolabelled amplification, cloning and sequencing were carried out in another laboratory. No contamination was detected. Results of Southern blotting and quadruplicate IPCR of HTLV-1 integration sites have been previously described for patients P4, P12, P20 and P25 (Cavrois *et al.* 1996b).

After digestion of cellular DNA from patient P30 with the endonuclease *EcoRI*, two clear bands > 9 kb were detected with the tumourous sample whereas PBMCs DNA gave a negative result. After digestion of lymph node DNA with the endonuclease *Pst I*, two internal fragments of 1.7 and 1.4 kb were detected. A faint signal at 1.7 kb was also detected by Southern blotting with *Pst I* when DNA from PBMCs was analysed. All other tumourous samples harboured one expanded clone after Southern blotting (Cavrois *et al.* 1996b).

By sequencing LMPCR products from the asymptomatic carrier G18, we detected a new deletion in the vicinity of the 3' provirus end. The cellular flanking sequences were: G18-8 AGATCAGACACATG and G18-18 AGATCAGACACTGTCTCATG. The sequences in common are underlined and the *NlaIII* site to which the BIO1 oligonucleotide was ligated is shown in bold face. Although the 5' 11 bases were identical there was no sequence homology between the remaining sequences. This was not due to a cloning artefact as both sequences extended into the 3' PCR primer and flanking M13 sequences. As previously described, in all our PCR work involving the amplification of direct repeats, deletions associated with Taq polymerase have not been noted (Wattel *et al.* 1995).

Fig 1 shows the HTLV-1 integration patterns observed after IPCR analysis of the two samples derived from patient

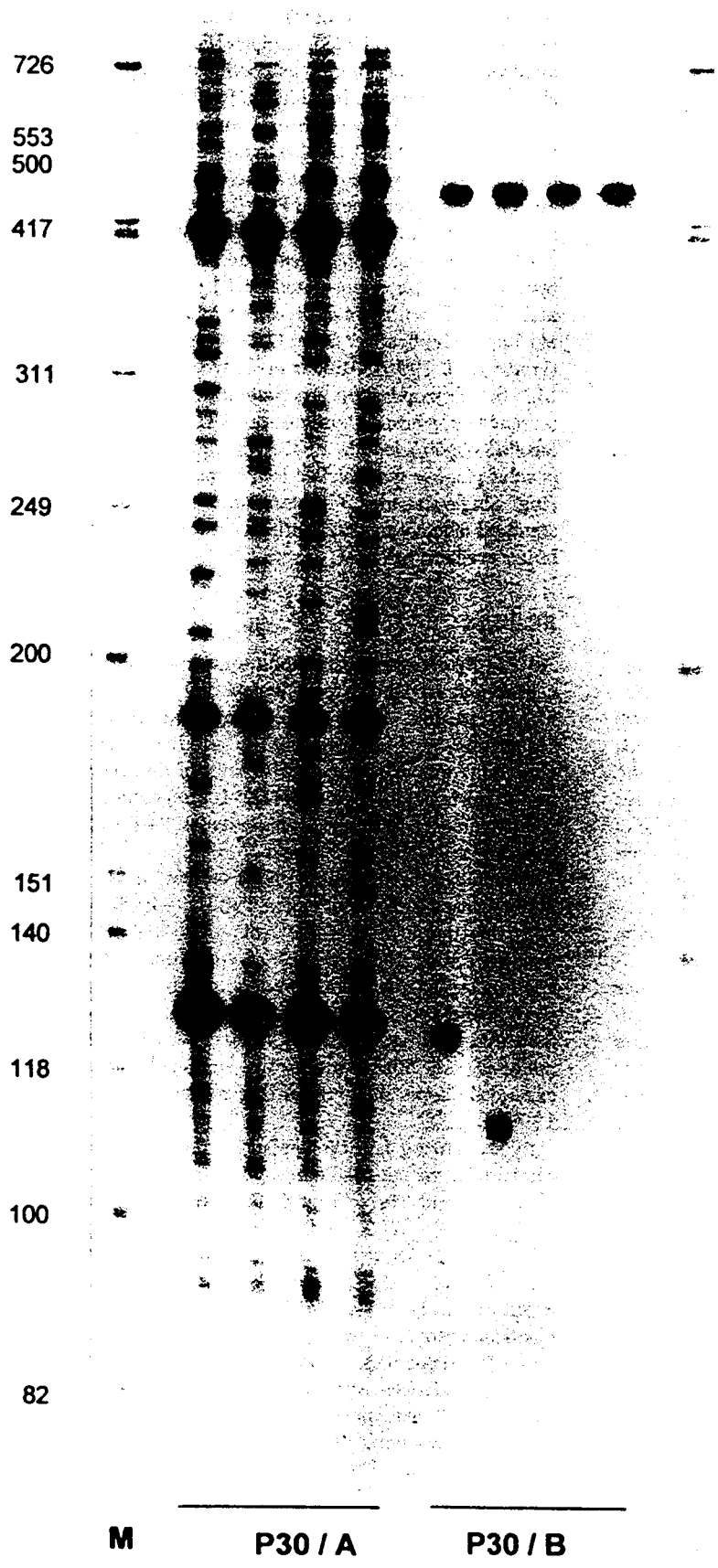


Fig 1. Quadruplicate IPCR analysis of samples from patient P30. For each sample, 2 μ g of DNA were studied in four experiments. After digestion by *Nla*III, circularization and exponential amplification, five cycles of linear PCR were performed with a 5'- 32 P-radiolabelled HTLV-1(+) strand-specific primer. Final products were loaded on a 4% sequencing gel. P30/A: after quadruplicate IPCR, ~100 HTLV-1-bearing clones were detected in lymph node DNA. Seven of these clones were detected four times. P30/B: PBMCs DNA analysis detected three clones. One of these was detected four times and did not appear to correspond to a malignant clone since it was not detected as a major form in the lymph node DNA.

P30. DNA was analysed by quadruplicate IPCR and amplified products were submitted to five linear cycles of extension using the ^{32}P radiolabelled BIO5 oligonucleotide as probe. Final products were resolved on a sequencing gel. Four intense bands at 126, 180, 420 and 450 bp dominated the tumourous sample. In addition, three clear additional bands of lower intensity were detected four times. About 100 clones of lower frequency were also detected, some of which were found only in a single sample. Three clones were detected after IPCR analysis of PBMCs DNA. A major circulating clone (detected four times) gave a signal at ~430 bp (Fig 1) which was not detected in the tumourous sample. The remaining two circulating clones were detected once.

IPCR products from all tumourous samples were cloned and sequenced as described in Material and Methods. Overall, 48 single-strand positive templates which represent 8434 bp of integration sites were sequenced. Fig 2 shows the distribution of flanking sequences corresponding to each tumourous sample. In samples from patients P4, P12, P20

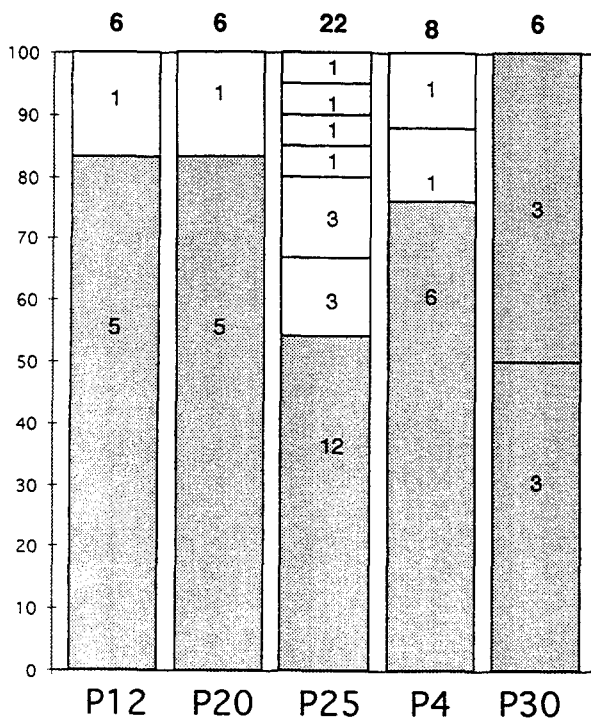


Fig 2. Distribution of the flanking sequences corresponding to the tumourous samples. Tumourous IPCR products were cloned and, in each case, six to 22 molecular clones were sequenced. The graph represents the proportion of each of the molecular clones derived from tumourous samples. The number of sequences per clone is indicated in each box. A dominant sequence (grey box) was clearly isolated in all samples from patients P4, P12, P20 and P25. In addition, one to six additional clones of lower frequency were also detected. As detailed in Results, the two distinct sequenced molecular clones of the tumourous sample derived from patient P30 correspond to a single cellular clone harbouring one integrated provirus. In each case the dominant sequence corresponded to a clone detected four times after run-off analysis (see Fig 1; Cavrois *et al.* 1996b).

and P25 the integration site derived from the malignant clone came up as a unique dominant sequence which represented at least 50% of all the purified molecular clones. Sequencing six tumourous molecular clones from patient P30 enabled the isolation of two distinct integration sites (both sequenced three times). Since Southern blotting showed two clear bands after *EcoRI* digestion, we concluded that the two isolated flanking sequences corresponded to the two signals observed after Southern blotting. To investigate whether these two sequences corresponded to two different clones or to one clone harbouring two integrated proviruses, dilutions of lymph node DNA from patient P30 were amplified by IPCR. Several signals were detected with 500 and 250 ng of tumourous P30 DNA. With 100 and 10 ng, only the two sequenced bands at 126 and 450 bp were still detected (not shown). No signal was obtained at 6.6 ng and below this concentration. Thus, we concluded that P30 tumourous DNA corresponded to one malignant clone harbouring two integrated proviruses.

Fig 3 represents the 16 unique integration sites that were sequenced, and correspond to seven major clones and nine minor clones. For 11 templates, sequencing was performed up to the terminal palindromic sequence CATG defined by the *Nla* III restriction site. The 144–164 first bp of the genomic segment were sequenced in the four remaining samples. The overall A/T content of the 16 sequences was 54% (range 43–74%). For each sample, no sequence homology was noted between integration sites derived from dominant clones and those corresponding to polyclonal forms. In addition, there was no homology between polyclonal forms in samples from patients P25 and P4.

DISCUSSION

ATLL is an aggressive malignancy characterized by a high degree of genetic instability. This results in frequent deletions and/or mutations of numerous genes such as p53 (Sakashita *et al.* 1992; Cesarman *et al.* 1992), p15 and p16 (Hatta *et al.* 1995). In addition to these mutations, the genetic instability of ATLL cells accounts for numerous genetic changes of the provirus sequence. Deletions of the 5' provirus region occur in about 30% of ATLL patients (Ohshima *et al.* 1991; Shimamoto *et al.* 1994; Tamiya *et al.* 1996; Tsukasaki *et al.* 1997) and interstitial deletions have been recently evidenced in ~20% of cases (Tamiya *et al.* 1996). Such interstitial deletions have been described at the asymptomatic phase of the infection (Saiga *et al.* 1997). In addition to proviral deletions, rearrangement of the host-flanking sequences in the vicinity of the provirus 3' end may occur during HTLV-1 infection. In a previous work we found a deletion of the 26 first bp of a 3' HTLV-1 flanking sequence, derived from PBMCs of an asymptomatic carrier (Wattel *et al.* 1996). This indicates that clonal expansion may underlie subsequent rearrangement of flanking cellular DNA in the absence of malignancy. Sequencing LMPCR products from carrier G18 detected an additional deletion of the 3' integration site.

We therefore looked for 3' microdeletion in ATLL by sequencing 48 HTLV-1 integration sites derived from five tumourous samples and found no deletion of the flanking

Patient	Clones (frequency)	Sequences (5' -> 3')
P12	80 (5)	GTATATGAAATTGGCTGTTACTAAAGTATAAGCCAATTTACAATACAATACTAAAGTATTATATTTGTTTAGGGGAGCTTTTATTA ATGCACTAAAAAATCTAAATACTTTTCTGCAATTTTGAAATTGATTAAATCTTCACGTCACAATATTAGATTTTC
	79 (1)	TTCACCGTGTTAGCCAGGATGGTCTTGATCTCTGACCTCGTGATCTGCCACCTCAGCCTCCCAAAGTGCTGGGATTACAGGTGTGA GCCACCGCGCCCGGCCAGTAATGTGGATTTTCTGAGTACAGCTTGACACAGCATT
P20	95 (5)	TGGTGAAAACCCGTTCTCTACTAAAAATACAAAAAATTAGCCGGGCGTGGTGGCGGGTGCCTGTAGTCCCAGCTACTTGGAGGGCTG AGGCAGGAGAATGGCGTGAACCCAGGAGGCGGAGTTGCAGTGAGGCGAGATGGTGCACCTGCC
	94 (1)	TGGCATTACCCAAACTCAGGCGACACAGAGAACGCCGAAGAGGAAGAGGGGTAGACCTGAGTCCCATG
P25	32 (12)	AGTATATACTATTGCATATCTAGTCAACCTCCCAAGCAGACTTGTGTAACTTTGGAGTAAATCACAAAGTTTGAAGACTACCAGA GACACTGCTTATTTCTGAGCCCCAGTTTGAAAATGTTTGAAAGCTTTTCTTTTGAATTCTAAGAAAAAAGCTGAAAAATGT CCATAGTAATTAATCGGGAATTTCTTTCCACAGTGACCTCTAAATTAACCTCTTGTATCTTTAACCTGTTCTGTTTGAATAACT TCCATACTGTCTCTGATTTTCAAAAGACCTAGAGATCATG
	33 (3)	CCTCTCAAGCATAAACAAAGCAAAACATG
	34 (3)	GTTAGGCTCCAAGAGGGCATG
	46 (1)	ACATAGTAGATCAGGAAAGACTTTCTAATTGTAGAAAATTTTGAAAAAGTAGAAACCTGAGATGGAAAGATTACTTAATAGTAT AAGTGCCAATAATACTTTAAAATCTCATTAAATAATGTCTTTTCATCCATTAGAAATAATAAGAGTCAGGACAGGCTTGCAGTCTGG TTAAAAGATAGTTATTCTTAGTCATG
	64 (1)	GAATCATG
	67 (1)	CCACCCACCTCAGCCTCCCAAAGTGTTGGGATTACAGGCGTGACGACCGTGCCTGGCTTGGACTAATTTATTTTATTTCTGTAGAG ACGATGCCTCCCTATTGTTGTCCAGGCTGGTCTCTTTCACATG
	70 (1)	TAAAGTGTCTTAAATTTATTAAATGTTAAGTGACTTTAGGCCAGGCATG
P30	84 (3)	CAGCTGCAGTTACTCACAATGGAGCTTCACACCATG
	85 (3)	AGAACTCACCTTTTACCTAACCCCTGAAAACCTAAAGCAATCAAACGTCATCTGCCTGAGCTTCCCATTCCCTTAAAGCCAGTGGAA CTTACCTCTATCTCCATCCATTCTTTCCTCCTTCCCTCCAGTTCCCTCAGAGGATAAGGTGACCATTCTCTTATTCAAGGCAATACCAC CCAATCTGCTCTGGGTACGTTGCTTAATTGATTCTCAGAAATGTTTTGTTTGTCTCACTTGTGAGTGGAAGTAGCAATAAACATTC TAAGTATAGCTTTCTCTAGTCTAATGCAACTACTGACAGTCAGAGCTCACGTGACATCATCACCTGCCGTACACAGCTGGACTCTTC ACATCCATG
P4	91 (6)	GCTTCACATCAACTTTTCCCGTGTGTCACTTTGGGTTTTGTTCTGTCCCTCATAAGGACACTTTCATTAGCTTTTACCCATCCTAATC CAGTATGATCTCATCTTTAGCCTCAGATCACATTCTGAGTTTCTGGGTGGACAATAAATTTGAGAACTATTATCAATAGACTACACCT TCATG
	88 (1)	GGAGGCGGGCAGAGAATATTTCCGAGATTTAGGGAAACAGTTCAGCAGAAACAACATG
	89 (1)	AGAGACAAGGCATTTTCAGAAGAGAATGAGAACTGTTTGGCTGGGGCTTGTGATGGTCTTAGTTGTTTAGAAGTTTGAGCCGGGTGTG GTGGCTCGCACCTTGCAATCGCAGCACAAAGGGAGGCCGAGGTGGGCGGATTACCGTAGGT

Fig 3. Flanking sequences derived from the five patients. For each patient, the first flanking sequences correspond to the malignant clone. No sequence homology was observed between integration sites derived from each sample.

sequence in the vicinity of the provirus end. This result does not preclude rearrangement of host cell DNA upstream on the flanking sequence. However, as described in asymptomatic carriers, the present study strongly suggests that deletion of the HTLV-1 flanking sequence in the vicinity of the provirus end remains a rare event in ATLL. Therefore this kind of genetic event cannot explain the oligoclonal pattern of HTLV-1 replication which accompanies the malignant clone(s) in ATLL tumourous samples.

When ATLL tumourous DNA is digested with the endonuclease *EcoRI*, which does not cut within the 9 kb of the full HTLV-1 genome, Southern blotting usually shows a single band of >9 kb. This was the case for four of our five ATLL samples. Tumourous DNA from patient P30 was found to harbour two clear bands >9 kb after Southern blot hybridization, whereas seven signals were detected four times when the same sample was analysed by quadruplicate IPCR. Only two of these seven forms were still detected after IPCR analysis of the diluted DNA, a result consistent with that of Southern blotting. This also means that five of the seven clones detected four times corresponded to oligoclonal forms within the tumourous sample.

Two or more complete proviruses can be detected by Southern blotting in ~20% of ATLL (Yoshida *et al.* 1984; Ohshima *et al.* 1991; Shimamoto *et al.* 1994; Tamiya *et al.* 1996; Tsukasaki *et al.* 1997). T-cell receptor gene rearrangement analysis shows that these ATLL are mono- or biclonal, suggesting that some of these harbour multiple copies of the HTLV-1 provirus per cell. The two clones detected here by IPCR analysis of diluted DNA from patient P30 were found to be in equal proportions after sequencing. Both the two corresponding signals were present at the same dilution points after run-off analysis but were not detected at 6.6 ng, meaning that P30 tumourous DNA corresponds to one malignant clone harbouring two integrated proviruses rather than to two distinct cellular clones, a result consistent with superinfection of the tumourous DNA derived from patient P30.

Finally, the present results suggest that the plethora of oligo/polyclonal forms which accompanied the malignant clone(s) were not associated with 3' deletions of the flanking sequences. Since clonal expansion is the dominant mode of HTLV-1 replication (Cavrois *et al.* 1996a, b; Wattel *et al.* 1996), such a pattern of replication could reflect a simple background of clonal expansion of non-malignant infected clones within the same tumourous sample. However, as evidenced here for patient P30, HTLV-1 superinfection of monoclonal tumour cells might also contribute to such a pattern.

ACKNOWLEDGMENTS

This work was supported by a grant from the Association pour la Recherche sur le Cancer. Some experiments were performed in the laboratory of Professor Wattel at Lille University Hospital. M.C., I.L. and F.M. were supported by a bursary from the Ministère de l'Enseignement Supérieur et de la Recherche. We thank Dr Yves Plumelle for some of the ATLL samples, Dr Renaud Mahieux for some

Southern blot experiments and Stephan Pouloubinsky for assistance.

REFERENCES

- Cavrois, M., Gessain, A., Wain-Hobson, S. & Wattel, E. (1996a) Proliferation of HTLV-1 infected circulating cells in vivo in all asymptomatic carriers and patients with TSP/HAM. *Oncogene*, **12**, 2419–2423.
- Cavrois, M., Wain-Hobson, S., Gessain, A., Plumelle, Y. & Wattel, E. (1996b) Adult T-cell leukemia/lymphoma on a background of clonally expanding HTLV-1 positive cells. *Blood*, **88**, 4646–4650.
- Cavrois, M., Wain-Hobson, S. & Wattel, E. (1995) Stochastic events in the amplification of HTLV-I integration sites by linker-mediated PCR. *Research in Virology*, **146**, 179–184.
- Cesarman, E., Chadburn, A., Inghirami, G., Gaidano, G. & Knowles, D.M. (1992) Structural and functional analysis of oncogenes and tumor suppressor genes in adult T-cell leukemia/lymphoma shows frequent p53 mutations. *Blood*, **80**, 3205–3216.
- Gessain, A., Barin, F., Vernant, J.C., Gout, O., Maurs, L., Calender, A. & de-The, G. (1985) Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. *Lancet*, **ii**, 407–410.
- Hatta, Y., Hiramata, T., Miller, C.W., Yamada, Y., Tomonaga, M. & Koefler, H.P. (1995) Homozygous deletions of the p15 (MTS2) and p16 (CDKN2/MTS1) genes in adult T-cell leukemia. *Blood*, **85**, 2699–2704.
- Ohshima, K., Kikuchi, M., Masuda, Y., Kobari, S., Sumiyoshi, Y., Eguchi, F., Mohtai, H., Yoshida, T., Takeshita, M. & Kimura, N. (1991) Defective provirus form of human T-cell leukemia virus type I in adult T-cell leukemia/lymphoma: clinicopathological features. *Cancer Research*, **51**, 4639–4642.
- Poiesz, B.J., Ruscetti, F.W., Gasdar, A.F., Bunn, F.A., Minna, J.D. & Gallo, R.C. (1980) Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America*, **77**, 7415–7421.
- Saiga, A., Orita, S., Minoura-Tada, N., Maeda, M., Aono, Y., Asakawa, M., Nakahara, K., Kubota, R., Osame, M. & Igarashi, H. (1997) cis-Acting inhibitory elements within the pol-env region of human T-cell leukemia virus type 1 possibly involved in viral persistence. *Journal of Virology*, **71**, 4485–4494.
- Sakashita, A., Hattori, T., Miller, C.W., Suzushima, H., Asou, N., Takatsuki, K. & Koefler, H.P. (1992) Mutations of the p53 gene in adult T-cell leukemia. *Blood*, **79**, 477–480.
- Shimamoto, Y., Suga, K., Shibata, K., Matsuzaki, M., Yano, H. & Yamaguchi, M. (1994) Clinical importance of extraordinary integration patterns of human T-cell lymphotropic virus type I proviral DNA in adult T-cell leukemia/lymphoma. *Blood*, **84**, 853–858.
- Takemoto, S., Matsuoka, M., Yamaguchi, K. & Takatsuki, K. (1994) A novel diagnostic method of adult T-cell leukemia: monoclonal integration of human T-cell lymphotropic virus type I provirus DNA detected by inverse polymerase chain reaction. *Blood*, **84**, 3080–3085.
- Tamiya, S., Matsuoka, M., Etoh, K.I., Watanabe, T., Kamihira, S., Yamaguchi, K. & Takatsuki, K. (1996) Two types of defective human T-lymphotropic virus type I provirus in adult T-cell leukemia. *Blood*, **88**, 3065–3073.
- Tsukasaki, K., Tsushima, H., Yamamura, M., Hata, T., Murata, K., Maeda, T., Atogami, S., Sohda, H., Momita, S., Ideda, S., Katamine, S., Yamada, Y., Kamihira, S. & Tomonaga, M. (1997) Integration patterns of HTLV-I provirus in relation to the clinical course of ATL: frequent clonal change at crisis from indolent disease. *Blood*, **89**, 948–956.

- Wattel, E., Cavrois, M., Gessain, A. & Wain-Hobson, S. (1996) Clonal expansion of infected cells: a way of life for HTLV-1. *Journal of Acquired Immune Deficiency Syndrome and Human Retrovirology*, **13**, 92–99.
- Wattel, E., Vartanian, J.P., Pannetier, C. & Wain-Hobson, S. (1995) Clonal expansion of human T-cell leukemia virus type I-infected cells in asymptomatic and symptomatic carriers without malignancy. *Journal of Virology*, **69**, 2863–2868.
- Yoshida, M., Miyoshi, I. & Hinuma, Y. (1982) Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. *Proceedings of the National Academy of Sciences of the United States of America*, **79**, 2031–2035.
- Yoshida, M., Seiki, M., Yamaguchi, K. & Takatsuki, K. (1984) Monoclonal integration of human T-cell leukemia provirus in all primary tumors of adult T-cell leukemia suggests causative role of human T-cell leukemia virus in the disease. *Proceedings of the National Academy of Sciences of the United States of America*, **81**, 2534–2537.

III. Analyse semi-quantitative de la maladie résiduelle au cours de la leucémie/lymphome T de l'adulte (ATLL).

III.1. Introduction.

Le traitement de l'ATLL repose sur la polychimiothérapie cytotoxique, les drogues antirétrovirales et la greffe de cellules souches hématopoïétiques.

Quelle que soit la stratégie thérapeutique utilisée, une rémission complète (RC) est obtenue chez environ 30 à 50% des patients, mais ces rémissions sont souvent de courte durée et la plupart des patients rechute, de sorte que l'ATLL reste une affection de mauvais pronostic. Moins de 10% des patients sont en vie à 1 an du diagnostic. Des travaux récents ont montré que, pour de nombreuses hémopathies malignes, des marqueurs moléculaires (translocations chromosomiques, réarrangements du récepteur des cellules T ou TCR (pour *T-Cell Receptor*)) permettaient de détecter, après traitement, des cellules malignes résiduelles chez des patients apparemment en RC. La persistance d'une telle "trace moléculaire" de la maladie est le plus souvent corrélée à une récurrence de telle sorte que l'étude de la maladie résiduelle (MR) est devenue un outil précieux dans la prise en charge de plusieurs hémopathies malignes.

Au cours de l'ATLL, aucun élément ne permet actuellement de prévoir la survenue d'une rechute chez un patient en rémission. Le traitement d'entretien ou de consolidation des patients mis en RC est donc difficile à adapter. Une méthode sensible de détection de la MR au cours de l'ATLL permettrait de mieux évaluer et de mieux comprendre la réponse au traitement, voire de détecter précocement la rechute. Ainsi, la recherche après rémission partielle ou complète, du clone malin à l'état résiduel, et la détection précoce de l'émergence d'un nouveau clone, devrait permettre de mieux comprendre l'évolution de l'ATLL et de pouvoir ajuster le traitement en conséquence.

Les techniques classiques, comme le blot de Southern, l'Hybridation par Fluorescence in Situ (ou FISH pour *Fluorescence in Situ Hybridization*) ou encore les méthodes de cytogénétique classique ne sont pas appropriées pour la détection de la MR au cours de l'ATLL, principalement en raison de leur faible sensibilité et de l'absence d'anomalies chromosomiques spécifiques, comme les translocations.

Les méthodes plus sensibles, comme la PCR et l'amplification du segment VDJ des gènes de TCR ou d'immunoglobuline, sont incapables de détecter un changement de clone intervenant au cours de l'infection. Or, de tels événements semblent fréquents au cours de l'ATLL. De plus, la faible masse tumorale collectée après traitement étant très faible, cette dernière technique peut entraîner l'amplification de segments VDJ présentés par des cellules non infectées ou des cellules CD8⁺ en expansion.

Nous avons donc utilisé la technique sensible de LMPCR, afin d'étudier le profil de clonalité et le devenir des cellules tumorales infectées par HTLV-1, après un traitement médicamenteux ou une greffe de moelle. L'étude a porté sur 8 patients atteints de différentes formes d'ATLL. Les patients ont été traités soit par chimiothérapie combinatoire, soit par administration de zidovudine et d'interféron (AZT/IFN) ou encore par allogreffe de moelle. Un patient a reçu de manière tout à fait exceptionnelle, une allogreffe de moelle de sa sœur séropositive pour HTLV-1.

III.2. Résultats et discussion.

L'analyse par quadruplicate IPCR de 31 échantillons de patients, prélevés au cours du temps avant et après traitement, a permis de mettre en évidence différents profils évolutifs de clonalité sous traitement :

- Le plus fréquent correspond à la persistance d'au moins un clone tumoral à une fréquence élevée, même chez les patients en RC. Ce profil de clonalité est toujours associé à une mauvaise évolution de la maladie.
- Un changement de clone, survenant au cours du traitement, a été observé et décrit chez un patient atteint d'une forme aiguë d'ATLL. D'autre part, la fréquence des clones circulants fluctue considérablement chez la plupart des patients étudiés. Ces observations suggèrent que les clones cellulaires présents dans le sang présentent des sensibilités différentes aux traitements. Tax transactive le promoteur du gène *mdr-1* (pour *Multi-Drug Resistance*) qui confère la résistance à la chimiothérapie (Lau *et al.*, 1998). Des variants naturels de Tax peuvent modifier sa fonction transactivatrice. Les cellules qui présentent des variants de Tax et capables d'activer le gène *mdr-1*, peuvent être sélectionnées par le traitement, conduisant à l'expansion de clones résistants.

- Finalement, une rémission moléculaire a été observée après transplantation de moelle allogénique d'un donneur séropositif. En effet, le clone tumoral, présent au diagnostic chez le receveur, disparaît totalement après la transplantation de moelle osseuse. Le clone cellulaire détecté après la greffe provient du donneur séropositif HTLV-1. La présence des cellules T cytotoxiques anti-HTLV-1 provenant du donneur asymptomatique pourrait être impliquée dans le contrôle de la MR et la disparition du clone malin.

Les réponses très variables, obtenues lors du traitement de l'ATLL, sont donc probablement dues à une hétérogénéité clonale. En effet, certains clones cellulaires semblent mieux résister aux traitements que d'autres. Par sa sensibilité, cette approche offre donc la possibilité de détecter et de quantifier la MR au cours de l'ATLL.

Nos résultats suggèrent que l'analyse de la MR par quadruplicate LMPCR permet de prédire l'évolution de l'ATLL puisque même chez les patients en RC, la persistance d'au moins 1 clone tumoral à une fréquence $> 1/300$ PBMCs est corrélée à une récurrence de l'hémopathie. Des traitements additionnels, comme la transplantation de cellules souches hématopoïétiques ou l'immunothérapie pourraient être d'un intérêt certain chez ces répondeurs après une première ligne de traitement. Ces résultats permettent donc de contribuer à la décision thérapeutique chez les patients traités pour l'ATLL.

Une rémission moléculaire prolongée et une possible guérison ont été obtenues après une allogreffe de cellules souches hématopoïétiques, avec un donneur séropositif pour HTLV-1. Il est tentant de proposer une participation immunologique au maintien de la RC après l'allogreffe de moelle. En effet, pour de nombreuses hémopathies, il a été montré que l'allogreffe de cellules souches hématopoïétiques s'accompagnait souvent d'un effet de "*Graft versus Leukemia*" (ou GVL), dont la présence et l'intensité étaient corrélées à la durée de la RC. Un tel mécanisme pourrait donc participer au maintien de la RC dans l'ATLL.

Dans le cas du patient P22 de notre étude, le donneur était séropositif pour HTLV-1. Cette donnée suggère l'intervention d'un mécanisme supplémentaire dans le contrôle de la MR post-allogreffe. En effet, la réponse CTL anti-HTLV-1 est importante chez les individus asymptomatiques (Parker *et al.*, 1992 ; Daenke *et al.*, 1996) alors qu'elle est profondément réduite au cours de l'ATLL (Kannagi *et al.*, 1994 ; Katahira *et al.*, 1995). Le donneur et le receveur de notre étude partagent le même génotype HLA de classe I et II. Par conséquent, sachant que les antigènes viraux sont fréquemment exprimés à la surface des cellules infectées (Jacobson *et al.*, 1990 ; Kannagi *et al.*, 1991 ; Parker *et al.*, 1992, 1994 ; Katahira *et al.*, 1995 ; Hanon *et al.*, 2000), l'allogreffe pourrait empêcher l'émergence d'un nouveau clone malin, en restaurant une réponse T cytolytique chez le receveur. Des arguments récents confortent cette hypothèse : l'apparition d'ATLL chez des individus immunodéprimés par la cyclosporine (Zanke *et al.*, 1989 ; Tsurumi *et al.*, 1992 ; Incan *et al.*, 1994 ; Jenks *et al.*, 1995), ou encore l'émergence de l'hémopathie chez le rat déplété en lymphocytes T (Hanabuchi *et al.*, 2000).

Semiquantitative analysis of residual disease in patients treated for adult T-cell leukaemia/lymphoma (ATLL)

INDIA LECLERCQ,¹ FRANCK MORTREUX,¹ FRANCK MORSCHHAUSER,^{1,2} PATRICK DUTHILLEUL,³ CLAUDE DESGRANGES,⁴ ANTOINE GESSAIN,⁵ MARIELLE CAVROIS,^{1*} JEAN-PAUL VERNANT,⁶ OLIVIER HERMINE⁷ AND ERIC WATTEL^{1,2,8}

¹Unité 124 INSERM, Institut de Recherche sur le Cancer de Lille, ²Service des Maladies du Sang, CHU, Lille, ³Département d'Hématologie, Immunologie et Cytogénétique, Centre Hospitalier, Valenciennes, ⁴U271 INSERM, Lyon, ⁵Unité d'Epidémiologie des Virus Oncogènes, Institut Pasteur, Paris, ⁶Service d'Hématologie, Hôpital de la Pitié Salpêtrière, Paris, ⁷Service d'Hématologie, Hôpital Necker, Paris, and ⁸Unité d'Oncogénèse Virale, Centre Oscar Lambret, Lille, France

Received 9 October 1998; accepted for publication 12 February 1999

Summary. Many adult T-cell leukaemia/lymphoma (ATLL) patients who respond to induction treatment, then relapse. Knowing the clonality pattern of residual tumourous clones during treatment could help understand disease evolution and aid therapeutic decisions. We developed a sensitive and semi-quantitative molecular analysis of these clones in ATLL patients. DNA samples from PBMCs derived from eight ATLL patients were studied over time by quadruplicate linker mediated PCR (LMPCR) amplification of HTLV-1 integration sites. Patients were treated with combination chemotherapy, zidovudine–interferon-alpha and/or by peripheral stem cell transplantation or allogeneic bone marrow transplantation. Persistence of tumourous clones at a high frequency (>1/300 PBMCs) was frequently observed, even in complete responders, and was invariably correlated with relapse and/or poor outcome. Fluctuation in the frequency of some tumourous clones was observed with evidence for clonal change under treatment in one patient, indicating that

treatment of ATLL can result in the selection of resistant clones. Finally, allogeneic bone marrow transplantation (BMT) using an HTLV-1 infected sibling as donor was found to be associated with long-lasting disappearance of tumourous clones and a possible cure of the disease. Long-term persistent clonal expansion of circulating HTLV-1 bearing T cells which derived from the donor bone marrow was evidenced in this patient. In conclusion, variable success in treatment of ATLL is probably due to the clonal heterogeneity which results in the selection of resistant clones. Semi-quantitative assessment of residual disease (RD) through LMPCR may predict treatment failure. Accordingly, additional therapy may be tailored to the clonality pattern observed after first-line therapy.

Keywords: adult T-cell leukaemia/lymphoma, HTLV-1, clonality, residual disease, bone marrow transplantation.

Adult T-cell leukaemia/lymphoma (ATLL) is a malignant T-cell disorder caused by HTLV-1 (Poiesz *et al.* 1980; Yoshida *et al.* 1982). About 30–50% of patients treated for ATLL achieve complete remission after combination chemotherapy or treatment by zidovudine and interferon-alpha (Shimamoto *et al.* 1990; Gill *et al.* 1995; Hermine *et al.* 1995). Many of these patients relapse and the overall survival remains short (Shimamoto *et al.* 1990). However,

long-term disease-free survival can be obtained after consolidation by allogeneic bone marrow transplantation (BMT) (Borg *et al.* 1996; Sobue *et al.* 1987).

Routine molecular techniques, such as Southern blotting or PCR amplification of the VDJ segment of the TCR gene, are inadequate for extensive monitoring of residual disease (RD) in ATLL, due to their low sensitivity or their inability to detect clonal change (Shimamoto *et al.* 1993; Shimoyama *et al.* 1991). Accordingly, therapeutic decision making, especially regarding intensive consolidation (Borg *et al.* 1996; Sobue *et al.* 1987), remains difficult for the patients who respond to induction therapy, calling for a test which creates a better understanding of disease evolution at the molecular level.

We used a sensitive and semi-quantitative linker mediated

* Present address: Institute of Human Virology, Medical Biotechnology Center, University of Maryland at Baltimore, 725 West Lombard Street, Baltimore, MD 21201, U.S.A.

Correspondence: Dr E. Wattel, Service des Maladies du Sang, CHU, Hôpital Huriez, 1 place de Verdun, 59037 Lille, France. e-mail: wattel@lille.inserm.fr.

polymerase chain reaction (LMPCR) assay to assess the clonality pattern of tumourous cells during treatment in eight ATLL patients. In some cases relapse or resistance to treatment were found to result from clonal selection. Persistence of the tumourous clones at a high frequency correlated with poor outcome and clearance of these clones, obtained after BMT, was associated with prolonged disease-free survival.

PATIENTS AND METHODS

Patients and samples studied. DNA samples from PBMCs derived from eight ATLL patients were studied over time. The diagnosis of ATLL was based on clinical features, cell surface marker analysis, serum HTLV-1 antibodies and clonal integration of HTLV-1 proviruses in the tumour sample as evidenced by Southern blotting. Patients were classified according to lymphoma study group criteria (Shimoyama *et al*, 1991). The clinical history of patients 3 and 4 has been previously detailed (Hermine *et al*, 1995). The initial characteristic and clinical history of all the eight patients studied is summarized in Table I and Fig 1, respectively. As shown in Fig 1, patients were treated with either combination chemotherapy, zidovudine and interferon-alpha, autologous peripheral stem cell transplantation or BMT.

Table I. Initial characteristics of the eight patients studied.

UPN	Age (yr)	Sex	Geographic origin	ATLL subtype
3	39	M	French West Indies	Acute leukaemia
21	37	F	French West Indies	Acute leukaemia
19	42	F	Senegal	Acute leukaemia
6	40	M	French West Indies	Acute leukaemia
11	39	F	French Guiana	Acute leukaemia
4	22	F	French Guiana	Smouldering
17	51	M	Portugal	Acute leukaemia
22	37	M	French West Indies	Acute leukaemia
AS49	41	F	French West Indies	(Asymptomatic)

Complete remission (CR) was defined as the resolution of all malignant disease (including disappearance of circulating leukaemic cells) for at least 4 weeks. Partial remission (PR) was defined as a reduction of at least 50% or more in measurable indexes of disease, for at least 4 weeks without the development of new lesions. Progressive disease was defined as an increase of 25% or more in measurable disease or in the number of circulating leukaemic cells.

PCR, cloning and sequencing. LMPCR and run-off were performed as described (Fig 2, Wattel *et al*, 1995; Cavois *et al*, 1996). Briefly, DNA was digested with *Nla*III in 1 × *Nla*III buffer for 3 h at 37°C.

Digested DNA was ligated with 10 pmol of BIO1 primer using 20 U of T4 DNA ligase. This was followed by a phenol/chloroform extraction and precipitation. Ligated DNA was amplified for 100 cycles using the BIO2 primer alone. 10 µl

of this linear PCR reaction was used in a classic PCR amplification using the BIO3 and BIO4 primer pair. Sequences of the oligonucleotides have been published previously (Wattel *et al*, 1995). Amplification conditions were as before with 50 pmol of each primer, again in a final volume of 100 µl.

The proportion of HTLV-1-positive clones was estimated by quadruplicate experiments (Cavois *et al*, 1995, 1996). There is a stochastic element to the amplification of low frequency HTLV-1 integration sites (Cavois *et al*, 1995). Quadruplicate LMPCR analysis of *Nla*III digested p4.39 DNA, bearing an integrated HTLV-1 provirus (Cavois *et al*, 1995), showed that at >500 copies (diluted in 1 µg, ~150 000 cell equivalents), detection was 4/4. However at 100, 50 and ≤10 copies detection was 2/4, 1/4 and 0/4 respectively. Accordingly, an integration site detected four times in four experiments corresponds to a clonal frequency of at least 1/300 PBMCs. Such clones were defined as major clones.

All the amplified products from samples derived from patient 22 and from his donor (carrier AS49) were cloned and sequenced as described (Wattel *et al*, 1995). Specific amplification of a tumourous clone derived from sample 22-1 was performed over time by nested PCR as previously described (Cavois *et al*, 1998) using ARBIS, AR3, BIO5 and BIO3 as primers (see Results for details).

RESULTS

Clonal evolution in patients who respond to treatment

Patient 3 had an acute ATLL with two major clones (Figs 1 and 3, sample 3-1). After 8 weeks treatment with zidovudine and interferon-alpha the patient was in PR with 25 10⁷ ATLL cells/l. At that time, four major clones were detected: the two major clones present at diagnosis and two new ones (Fig 3, sample 3-2, signals at ~80 and ~90 bp). An identical pattern was observed 1 month later while the patient was still in PR (Fig 3, sample 3-3). Patient 3 discontinued his treatment and relapsed at 5 months from diagnosis. At that time the blood cell count (WBC) was normal and the number of atypical circulating cells was stable (6%). However, the appearance of these cells was that of large-cell lymphoma rather than that of typical ATLL (Hermine *et al*, 1995). Interestingly, as shown in Fig 3 (sample 3-4), only one of the two clones present at diagnosis could be detected. Its clonal frequency was down to 1/3000 from 1/300 cells. In contrast, the two clones corresponding to signals at ~80 and ~90 bp, previously detected in samples 3-2 and 3-3 but not at diagnosis, were still present as the sole major clones. CR was obtained after two courses of an anthracyclin containing regimen. At that time (sample 3-5), the two major clones were still detected at the same frequency. The patient relapsed 6 weeks later and died at 13 months from diagnosis. At the time of second relapse (sample 3-6) the two signals at ~80 and ~90 bp still remained detectable at the same frequency. In contrast, Fig 3A (sample 3-6) shows that only one of the two major clones present at diagnosis was still detected but at a lower frequency.

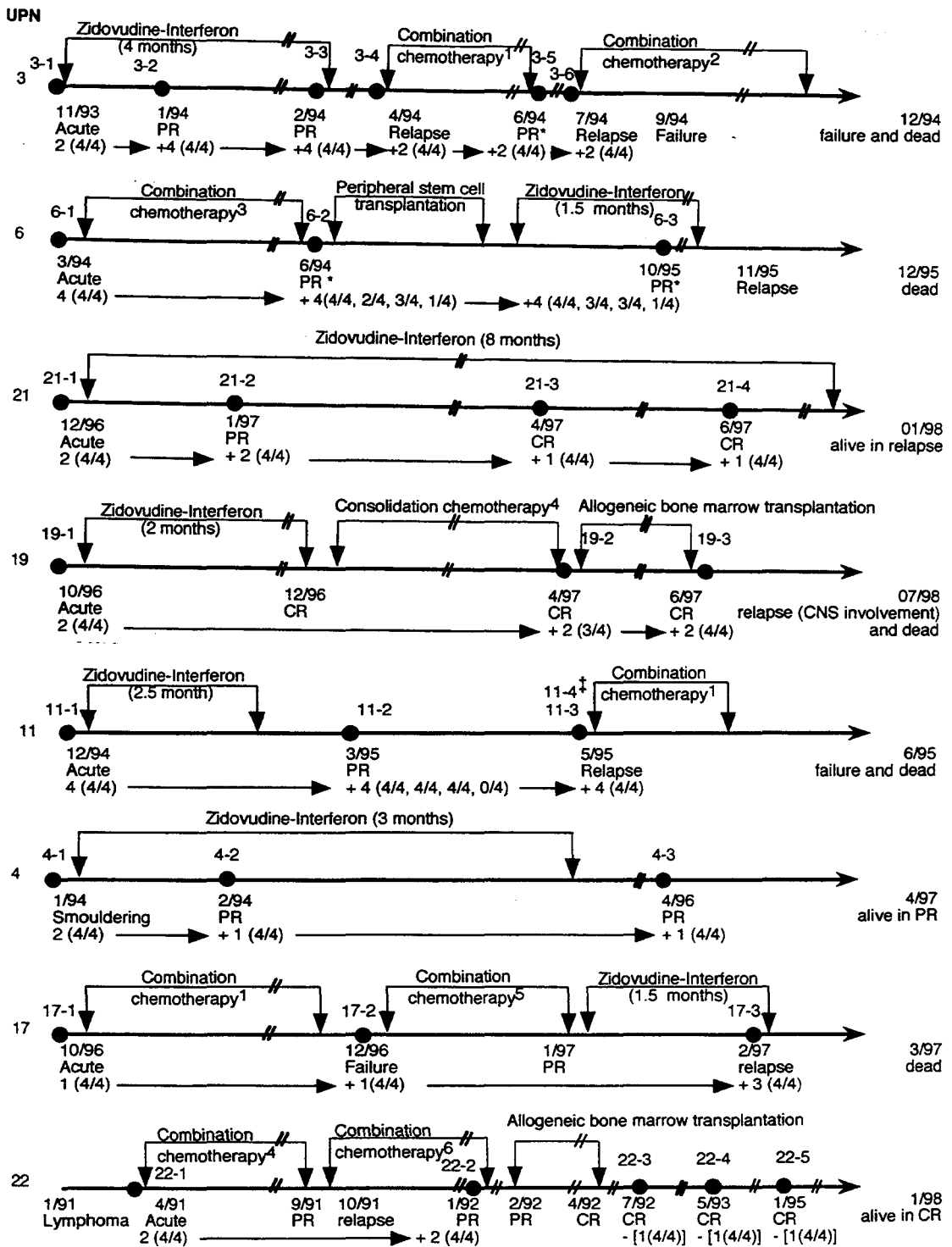


Fig 1. Profile of the eight patients with ATLL. Patients are identified by their unique patient number (UPN). Treatment is summarized for each case. Black circles correspond to the date the samples were analysed. Each sample is represented by a number. The first line below the thick arrow corresponds to the date the samples were prepared and/or the clinical status was evaluated. The second line represents the corresponding clinical status. The third line represents the number of clones detected four times after quadruplicate LMPCR analysis. Persistent tumourous clones are represented by thin horizontal arrows. 1: CHOP regimen (cyclophosphamide, vincristine, adriamycin, prednisone); 2: DHAP regimen (dexamethasone, cisplatin, cytarabine); 3: one course of CEOP regimen (cyclophosphamide, epirubicin, vincristine, and prednisone) followed by two courses of ECVBP (cyclophosphamide, epirubicin, vindesine, prednisone); 4: adriamycin, VP16; 5: CEEP regimen (cyclophosphamide, VP16, eldisine and prednisone); 6: ACVBP regimen (adriamycin, cyclophosphamide, vindesine, bleomycin and prednisone). * PR achievement with normalization of the WBC without detectable circulating leukaemic cells. ‡ Sample 11-3 corresponds to a skin biopsy.

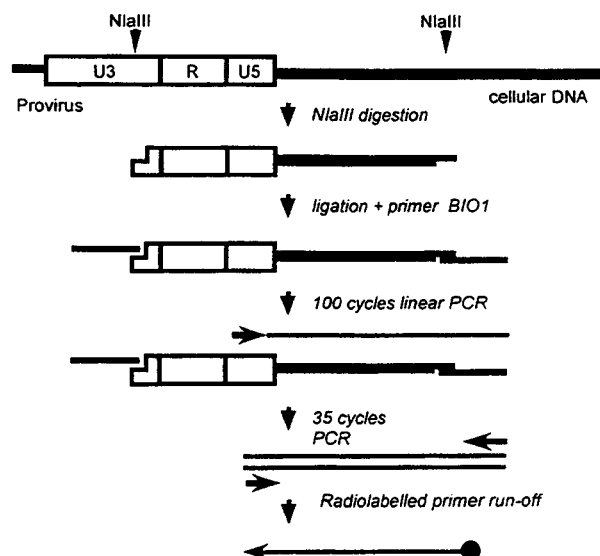


Fig 2. LMP-PCR amplification of HTLV-1 flanking sequences.

Patient 21 had an acute ATLL with two major clones (Figs 1 and 3, sample 21-1, WBC $20 \times 10^9/l$, 30% leukaemic cells). After 4 weeks treatment with zidovudine and interferon-alpha, the WBC was $6 \times 10^9/l$ with 18% leukaemic cells and the two clones remained detectable at the same frequency (Fig 3, sample 21-2). 3 months later the patient was in CR. Of the two major clones present at diagnosis, one was still present at a frequency of $>1/300$ cells (Fig 3, sample 21-3, signal at ~ 80 bp) but the remaining tumourous clone was not detected. 6 months after diagnosis the patient was in persistent CR and the same clone was still detected at the same frequency (sample 21-4). 5 months later the patient had a leukaemic relapse (WBC $2.9 \times 10^9/l$; 6% leukaemic cells). 13 months after diagnosis she remained alive in relapse.

Patient 19 had an acute ATLL with two major clones (Figs 1 and 3, WBC $84 \times 10^9/l$; 84% leukaemic cells). CR was achieved after 2 months of treatment with zidovudine and interferon-alpha. This treatment was discontinued because of a psychosis attributed to interferon and the patient was given two courses of combination chemotherapy. 6 months after diagnosis she remained in CR (Figs 1 and 3, sample 19-2). At that time the two clones present at diagnosis were detected at a lower frequency and three additional major clones were present (Fig 3C, sample 19-2). The patient received an allogeneic T-depleted BMT from an HLA-identical HTLV-1 seronegative sister with cyclophosphamide and total body irradiation as conditioning and methotrexate (MTX) as graft-versus-host disease (GVHD) prophylaxis. The patient engrafted without acute GVHD. Sample 19-3 was collected 1 month after transplantation, when the patient was in persistent CR. The two clones present at diagnosis were detected at the frequency of $>1/300$ (Fig 3C, sample 19-3). Only one of the three major clones present in sample 19-2 was still detected but at a lower frequency. Finally, one additional major clone was

detected. One month later the patient relapsed with CNS involvement and died.

Patient 6 had an acute ATLL with four major clones. Two courses of combination chemotherapy followed by G-CSF infusion and peripheral blood stem cell (PBSC) collection resulted in PR. The WBC was normal without detectable leukaemic cells. All the biological parameters were normal except for the bone marrow trephine which revealed a weak tumoural infiltration. Sample 6-2 was collected at that time and all the four clones present at diagnosis were found to persist. Only one was detected as a major clone. After peripheral stem-cell transplantation the patient remained in PR and was treated with zidovudine and interferon alpha. 19 months after diagnosis there were no circulating leukaemic cells. However, three of the four major clones present at diagnosis were detected in sample 6-3. The same clone as in sample 6-2 was detected as a major form while the remaining two corresponded to clones of lower frequency. One month later the patient relapsed with diffuse localizations. He died 21 months following diagnosis.

Clonal evolution in patients who did not respond to treatment

Patient 11 had an acute ATLL with four major clones (Figs 1 and 3, sample 11-1, WBC $29 \times 10^9/l$, 83% ATLL cells). He was first treated with zidovudine-interferon-alpha and achieved PR. At that time, three of the four clones present at diagnosis were still detected at the same frequency (sample 11-2). Patient 11 relapsed at 5 months from diagnosis with multiple cutaneous localizations and an increase in circulating leukaemic cells. Two of the three major clones previously detected in blood were also evidenced in the skin lesion (sample 11-3, signals at ~ 200 and ~ 230 bp): one as a major clone and the other at a clonal frequency between $1/1500$ to $1/300$ cells. The clonal pattern in blood sample 11-4, collected the day the skin biopsy was performed, was the same as at diagnosis (not shown).

Patient 4 had a smouldering ATLL with two major clones (not shown). PR was achieved after treatment (Fig 1) with persistent circulating leukaemic cells. Only one of the two major clones was detected at a clonal frequency of $>1/300$ cells in samples 4-2 and 4-3 which were collected 1 and 3 months after diagnosis, respectively.

Patient 17 had an acute ATLL with CNS involvement and a unique major circulating clone (Fig 1, sample 17-1, WBC $37 \times 10^9/l$; 30% leukaemic cells). No clinical response was obtained after two courses of an anthracycline-containing regimen combined with intrathecal chemotherapy. At that time, three major clones were detected, including the initial one (Fig 3D, sample 17-2). The two additional clones were also present in sample 17-1 but at a lower frequency. Patient 17 failed a second-line combination chemotherapy and a 6 weeks treatment with zidovudine and interferon-alpha. At that time the clonality pattern was the same as at diagnosis.

At the time of the study, patient 22 had a leukaemic relapse of a lymphoma-type ATLL with two major clones (Figs 1 and 3, sample 22-1). A PR was obtained after four courses of combination chemotherapy. The patient relapsed 1 month later and achieved a second PR. Physical and

laboratory examination were normal, except for the persistence of circulating flower cells (<1%, normal WBC). At that time, of the two major clones present at diagnosis (Fig 3E, sample 22-2, signal at 246 bp), one was observed at the same frequency, and the remaining was detected at a frequency of 1/1500 (signal at ~290 bp).

Clonal evolution after allogeneic BMT from a HTLV-1 seropositive donor

Patient 22 was allografted in PR. He received an allogeneic non-T-depleted BMT from AS49, his unique HLA identical sibling who was also HTLV-1 seropositive. Conditioning regimen consisted of total body irradiation, VP16 and cyclophosphamide. GVHD prophylaxis consisted of cyclosporine A (CSA) and MTX. Patient 22 experienced grade II acute GVHD and later moderate chronic GVHD which was controlled by steroid and CSA. After haematological reconstitution, examination of blood smears did not detect leukaemic cells. In contrast, a major clone corresponding to the signal at 117 bp was observed in DNA derived from samples 22-3 (not shown), 22-4 and 22-5 (Figs 1 and 3E). This persistent clone was not detected before BMT. This was confirmed by sequencing LMPCR products derived from all

the five samples. However, a signal at the same size was detected four times in PBMCs derived from AS49 (Fig 3). Sequencing LMPCR products from the AS49 PBMCs sample enabled the detection of the same dominant HTLV-1 proviral flanking sequence as that detected in samples 22-3 to 22-5: 5'-GAGAGGAATCAGTGTGGATTGGGGCCCCAATCCATG-3'. Therefore the persistent major clone detected in patient 22 PBMCs samples collected after BMT derived from the donor bone marrow and was also detected in the donor PBMC DNA at the same frequency.

A sensitive nested PCR assay was used in order to assess whether major tumourous clones could have disappeared after BMT. Sequencing tumourous molecular clones derived from samples 22-1 and 22-2 isolated a unique dominant flanking sequence: 5'-GCTCCCTGGGTACTACTGGTTAGCAAGCTATTTTATCACAGTTGTATACTTCTTGGTTAATAATGAATGCACAAAATTAACCTTTAAATGAATTATAAAATTCCTCAAAGTTCAATTTTACATAAAATGCCATCGGGAAGTATCAAAATACGGTGGCAACCCATG-3'. After run-off analysis, the predicted size of this 167 bp integration site corresponded to that of the signal at 246 bp (clone IS167) detected after LMPCR analysis of samples 22-1 and 22-2. Sequencing six molecular clones failed to isolate another

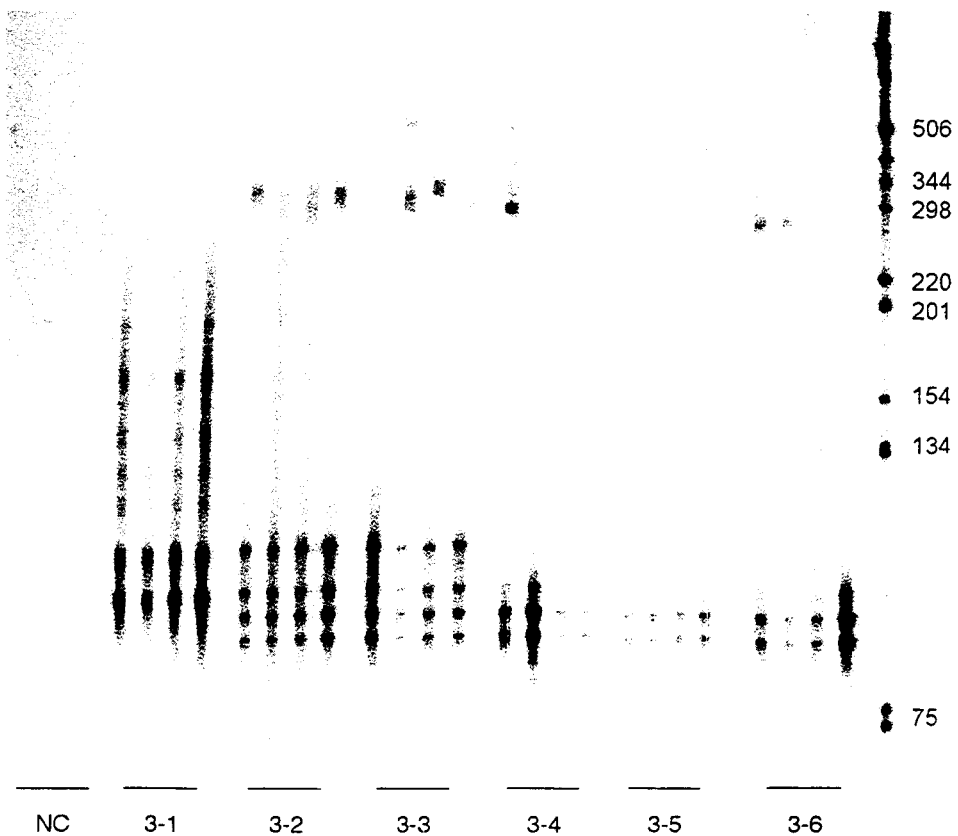
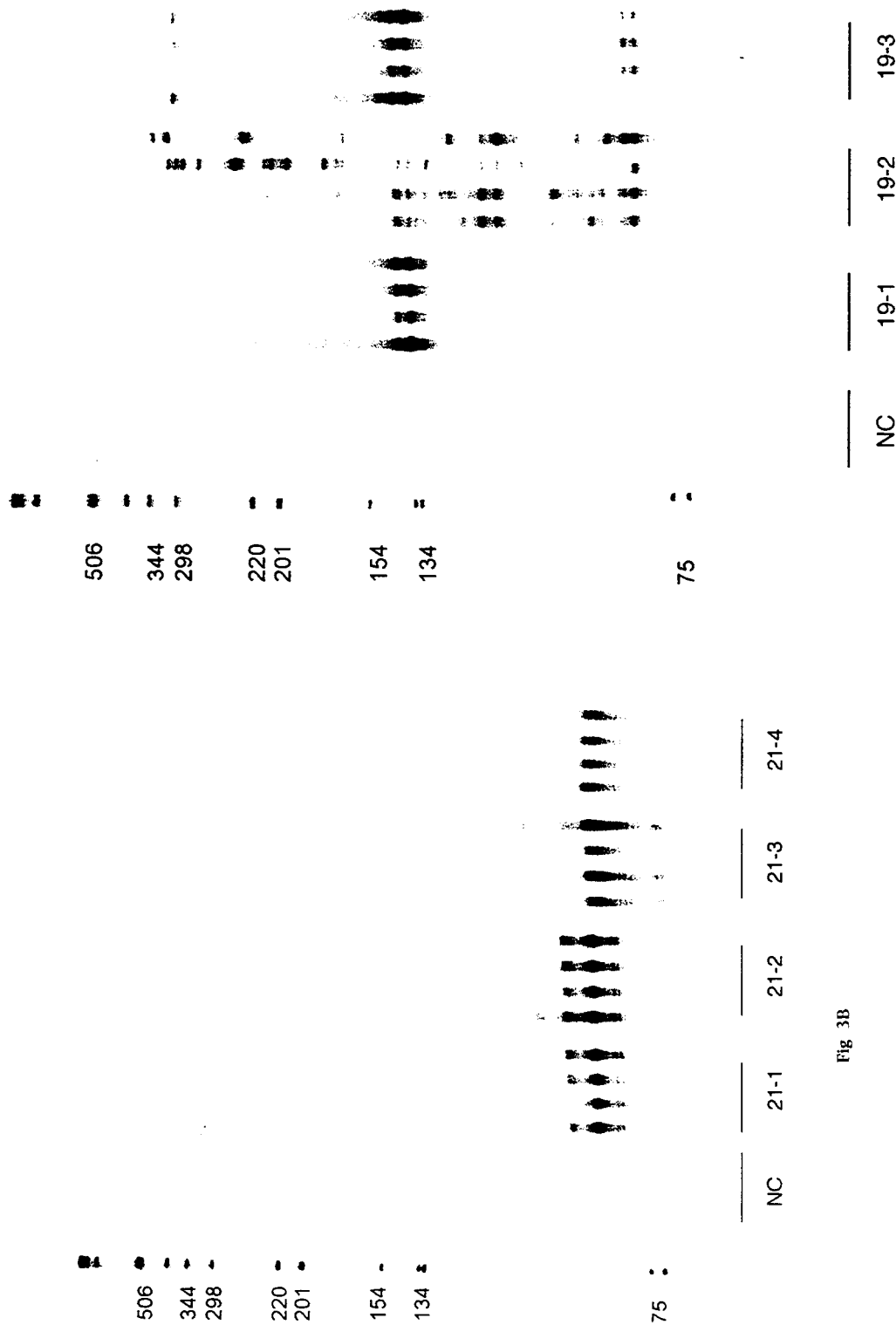
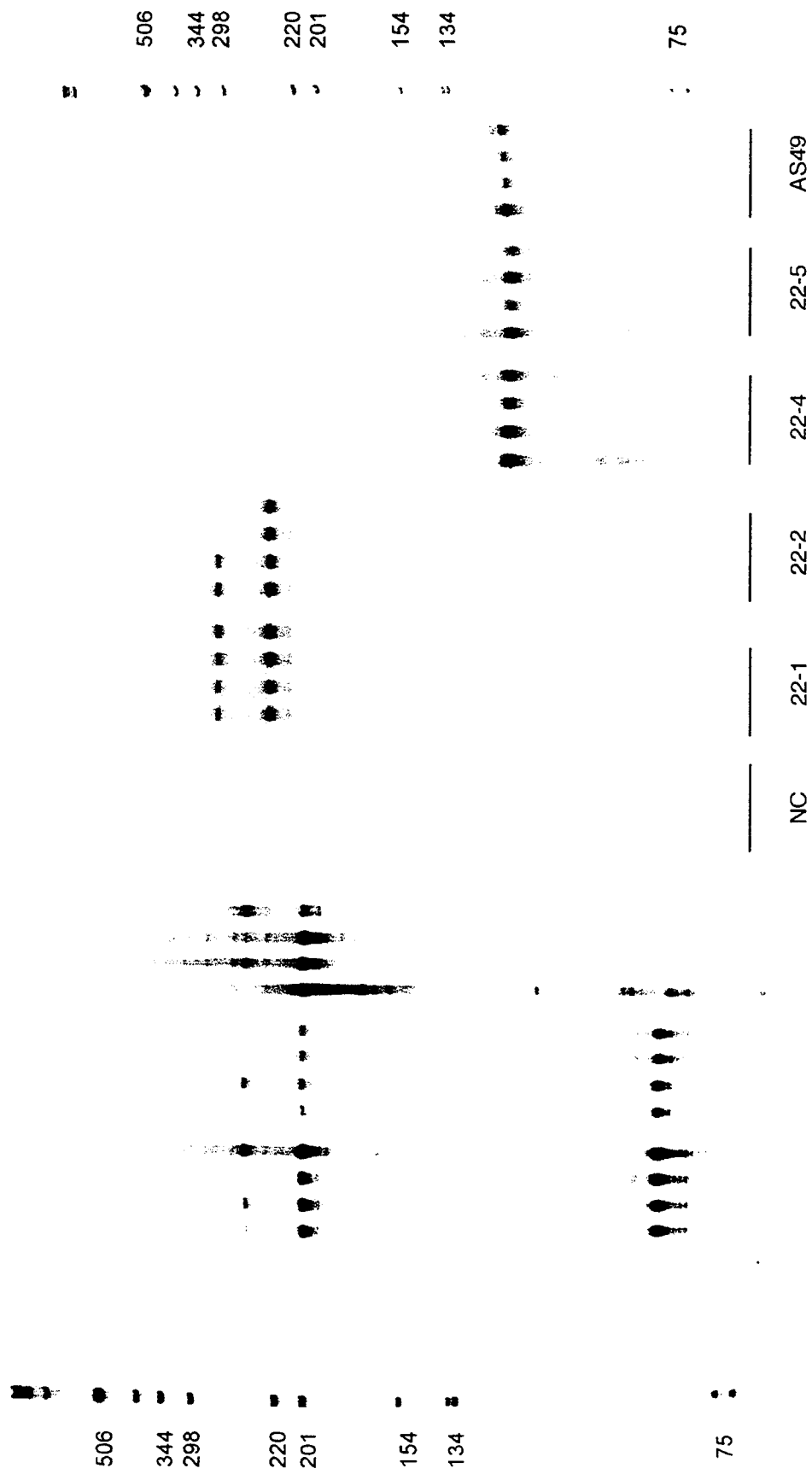


Fig 3A

Fig 3. Quadruplicate LMPCR analysis of DNA samples collected over time in patients treated for ATLL. Figs 3A–E correspond to the clonality pattern of patients 3, 21, 19, 11 and 22, respectively. Except for sample 11-3 which corresponded to a skin biopsy, DNA was extracted from PBMCs in all remaining cases.





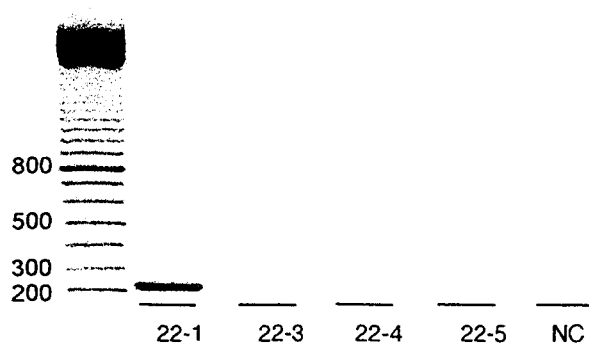


Fig 4. Clonotypic nested PCR analysis of PBMCs samples derived from patient 22. Oligonucleotides corresponding to the flanking sequence harboured by a tumorous clone was used in conjunction with LTR-specific primers.

sequence, indicating that IS167 corresponded to the major clone within the tumorous DNA. Two oligonucleotides ARBIS (5'-GGGTTGCCACCGTATTTTGATAC-3') and AR3 (5'-GTATCAAAATACGGTGGCAACCC-3') were chosen which, in conjunction with the HTLV-specific oligonucleotides BIO3 and 5 (Wattel *et al.* 1995), allowed specific nested amplification of the clone IS167. As shown in Fig 4, nested PCR amplification of sample 22-1 gave a signal at 208 bp corresponding to the predicted size. Negative results were obtained with samples 22-3 and 22-4.

DISCUSSION

In contrast to many other haematological malignancies, the lack of recurrent chromosomal translocation (Shimoyama *et al.* 1987) in ATLL precludes the use of junctional sequence created by such a translocation as a target to detect RD. PCR amplification of the V-D-J segment of the TCR gene is an alternative useful tool in the detection of RD in T-cell malignancies (Chan *et al.* 1996). However, in ATLL samples with low tumoral burden, such as those collected after treatment, this method may result in the amplification of V-D-J segments harboured by non-infected or CD8 expanded cells (Furukawa *et al.* 1994). This drawback may be circumvented by the use of primers specific to the tumorous rearrangement(s) (Chan *et al.* 1996). However, this more sensitive and specific approach cannot detect clonal change during the course of the disease (Tsukasaki *et al.* 1997; Shimamoto *et al.* 1993). Specific detection and semiquantitation of circulating infected clones with an absolute detection threshold of 2/15 000 (Cavrois *et al.* 1995) can be obtained with low amount of DNA by the quadruplicate LMPCR amplification of HTLV-1 integration sites. By this means, it was possible to follow RD in all the eight patients from the present study.

There are few reports of ATLL harbouring distinct clones (Tsukasaki *et al.* 1997; Shimamoto *et al.* 1993; Kondo *et al.* 1995; Shibata *et al.* 1995), but patient 3 is the first documented case of clonal change occurring during treatment. Fluctuation of the frequency of circulating clones was also observed in all other patients. Together, these results

reinforce the idea that there are multiple premalignant clones in the tumoural tissue from patients with ATLL (Tsukasaki *et al.* 1997; Cavrois *et al.* 1996) and suggest that some clones harbour distinct sensitivities to treatment. Tax transactivates the promoter of the multidrug resistance gene (MDR-1, Chuang *et al.* 1997) which confers resistance to chemotherapy. It has been shown that naturally occurring variants of Tax protein can modify its transactivation function (Niewiesk *et al.* 1995). Cells harbouring Tax variants able to activate MDR-1 may be selected under treatment, leading to the selection of resistant clones.

Normalization of the WBC together with disappearance of circulating leukaemic cells, was observed in five patients. However, except in patient 22 after BMT, LMPCR analysis of RD invariably revealed that at least one of the initial tumorous clones was found to persist in a major form after treatment. Even in complete responders this pattern of clonality was associated with poor outcome. In patient 22, BMT allowed the clearance of tumorous clones from the peripheral blood. Acute and chronic GVHD which were observed after transplantation and were possibly associated with a graft-versus-leukaemia effect (Michallet, 1996). In addition, transplantation of normal donor anti-HTLV-1 cytotoxic T-cell precursors (Jacobson *et al.* 1990) may have accounted for the control of RD after BMT.

Finally, the study suggests that variable success in treatment of ATLL is probably due to clonal heterogeneity. Assessment of RD by LMPCR appeared to constitute a useful tool in the management of ATLL. Even in responders, persistence of tumorous clones at a high frequency, i.e. >1/300, was correlated with poor outcome. Accordingly, additional treatment such as BMT or immunotherapy (Waldmann *et al.* 1995) may be of interest in those responders after first line therapy. The role of the donor HTLV-1 status in tumour control following BMT remains to be elucidated.

ACKNOWLEDGMENTS

We thank Simon Wain-Hobson for helpful discussions and review of the manuscript. This work was supported by grants from the Association pour la Recherche sur le Cancer, from the Fédération Nationale des Groupements des Entreprises Françaises dans la Lutte contre le Cancer (FEGEFLUC) and from the Ligue Nationale contre le Cancer (Comité Pas de Calais). I.L. and E.M. were supported by bursaries from the Ministère de l'Enseignement Supérieur et de la Recherche. We thank Isabelle Allard, Aspasia Stamatoullas, Anne-Chantal Braud, Frédéric Viret and Dominique Jaubert for a number of samples from ATLL patients. We also thank Professor Wattre and collaborators who kindly allowed us to use their laboratories for DNA extraction, digestion, ligation and PCR.

REFERENCES

- Borg, A., Yin, J.A., Johnson, P.R., Tosswill, J., Saunders, M. & Morris, D. (1996) Successful treatment of HTLV-1-associated acute adult T-cell leukaemia lymphoma by allogeneic bone marrow transplantation. *British Journal of Haematology*, **94**, 713-715.

- Cavrois, M., Leclercq, I., Gout, O., Gessain, A., Wain-Hobson, S., Wattel, E. (1998) Persistent oligoclonal expansion of human T-cell leukemia virus type 1 infected circulating cells in patients with tropical spastic paraparesis/HTLV-1 associated myelopathy. *Oncogene*, 17, 77–82.
- Cavrois, M., Wain-Hobson, S., Gessain, A., Plumelle, Y. & Wattel, E. (1996) Adult T-cell leukemia/lymphoma on a background of clonally expanding HTLV-1 positive cells. *Blood*, 88, 4646–4650.
- Cavrois, M., Wain-Hobson, S. & Wattel, E. (1995) Stochastic events in the amplification of HTLV-I integration sites by linker-mediated PCR. *Research in Virology*, 146, 179–184.
- Chan, D.W., Liang, R., Kwong, Y.L. & Chan, V. (1996) Detection of T-cell receptor delta gene rearrangement in T-cell malignancies by clonal specific polymerase chain reaction and its application to detect minimal residual disease. *American Journal of Hematology*, 52, 171–177.
- Chuang, S.E., Doong, S.L., Lin, M.T. & Cheng, A.L. (1997) Tax of the human T-lymphotropic virus type transactivates promoter of the MDR-1 gene. *Biochemical and Biophysical Research Communications*, 238, 482–486.
- Furukawa, K., Mori, M., Ohta, N., Ikeda, H., Shida, H., Furukawa, K. & Shiku, H. (1994) Clonal expansion of CD8⁺ cytotoxic T lymphocytes against human T cell lymphotropic virus type I (HTLV-I) genome products in HTLV-I-associated myelopathy/tropical spastic paraparesis patients. *Journal of Clinical Investigation*, 94, 1830–1839.
- Gill, P.S., Harrington, W., Jr, Kaplan, M.H., Ribeiro, R.C., Bennett, J.M., Liebman, H.A., Bernstein-Singer, M., Espina, B.M., Cabral, L., Allen, S., Kornblau, S., Pike, M.C. & Levine, A.M. (1995) Treatment of adult T-cell leukemia-lymphoma with a combination of interferon alpha and zidovudine. *New England Journal of Medicine*, 332, 1744–1748.
- Hermine, O., Bouscary, D., Gessain, A., Turlure, P., Leblond, V., Franck, N., Buzyn-Veil, A., Rio, B., Macintyre, E., Dreyfus, F. & Bazarbachi, A. (1995) Treatment of adult T-cell leukemia-lymphoma with zidovudine and interferon alpha. *New England Journal of Medicine*, 332, 1749–1751.
- Jacobson, S., Shida, H., McFarlin, D.E., Fauci, A.S. & Koenig, S. (1990) Circulating CD8⁺ cytotoxic T lymphocytes specific for HTLV-I pX in patients with HTLV-I associated neurological disease. *Nature*, 348, 245–248.
- Kondo, S., Kotani, T., Tsumori, S., Narahara, S., Aratake, Y., Kobayashi, M., Takahashi, M., Inoue, S. & Ohtaki, S. (1995) Identification of biclonal (duplex) leukaemic cells expressing either CD4⁺/CD8⁺ or CD4⁺/CD8⁺ from a patient with adult T-cell leukaemia/lymphoma. *British Journal of Haematology*, 89, 669–671.
- Leclercq, I., Cavrois, M., Mortreux, F., Hermine, O., Gessain, A., Morschhauser, F. & Wattel, E. (1998) Oligoclonal proliferation of human T-cell leukaemia virus type 1 bearing T cells in adult T-cell leukaemia/lymphoma without deletion of the 3' provirus integration sites. *British Journal of Haematology*, 101, 500–506.
- Michallet, M. (1996) Graft-versus-host disease and graft-versus-leukemia. *Hematological Cell Therapy*, 38, 459–460.
- Niewiesk, S., Daenke, S., Parker, C.E., Taylor, G., Weber, J., Nightingale, S. & Bangham, C.R. (1995) Naturally occurring variants of human T-cell leukemia virus type I Tax protein impair its recognition by cytotoxic T lymphocytes and the transactivation function of Tax. *Journal of Virology*, 69, 2649–2653.
- Poiesz, B.J., Ruscetti, F.W., Gasdar, A.E., Bunn, F.A., Minna, J.D. & Gallo, R.C. (1980) Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America*, 77, 7415–7421.
- Shibata, K., Shimamoto, Y., Suga, K., Watanabe, M., Kikuchi, M. & Yamaguchi, M. (1995) Adult T-cell leukemia/lymphoma with two distinct clones in the peripheral blood and lymph node. *American Journal of Hematology*, 48, 116–119.
- Shimamoto, Y., Kikuchi, M., Funai, N., Suga, K., Matsuzaki, M. & Yamaguchi, M. (1993) Spontaneous regression in adult T-cell leukemia/lymphoma. *Cancer*, 72, 735–740.
- Shimamoto, Y., Suga, K., Nishimura, J., Nawata, H. & Yamaguchi, M. (1990) Major prognostic factors of Japanese patients with lymphoma-type adult T-cell leukemia. *American Journal of Hematology*, 35, 232–237.
- Shimoyama, M. (1991) Diagnostic criteria and classification of clinical subtypes of adult T-cell leukaemia-lymphoma: a report from the Lymphoma Study Group (1984–87). *British Journal of Haematology*, 79, 428–437.
- Shimoyama, M., Abe, T., Miyamoto, K., Minato, K., Tobinai, K., Nagoshi, H., Matsunaga, M., Nomura, T., Tsubota, T. & Ohnishi, T. (1987) Chromosome aberrations and clinical features of adult T cell leukemia-lymphoma not associated with human T cell leukemia virus type I. *Blood*, 69, 984–989.
- Sobue, R., Yamauchi, T., Miyamura, K., Sao, H., Tahara, T., Yoshikawa, H., Yoshikawa, S., Morishima, Y. & Kodera, Y. (1987) Treatment of adult T cell leukemia with mega-dose cyclophosphamide and total body irradiation followed by allogeneic bone marrow transplantation. *Bone Marrow Transplantation*, 2, 441–444.
- Tsukasaki, K., Tsushima, H., Yamamura, M., Hata, T., Murata, K., Maeda, T., Atogami, S., Sohda, H., Momita, S., Ieda, S., Katamine, S., Yamada, Y., Kamihira, S. & Tomonaga, M. (1997) Integration patterns of HTLV-I provirus in relation to the clinical course of ATL: frequent clonal change at crisis from indolent disease. *Blood*, 89, 948–956.
- Waldmann, T.A., White, J.D., Carrasquillo, J.A., Reynolds, J.C., Paik, C.H., Gansow, O.A., Brechbiel, M.W., Jaffe, E.S., Fleisher, T.A. & Goldman, C.K. (1995) Radioimmunotherapy of interleukin-2R alpha-expressing adult T-cell leukemia with Yttrium-90-labeled anti-Tac. *Blood*, 86, 4063–4075.
- Wattel, E., Vartanian, J.P., Pannetier, C. & Wain-Hobson, S. (1995) Clonal expansion of human T-cell leukemia virus type I-infected cells in asymptomatic and symptomatic carriers without malignancy. *Journal of Virology*, 69, 2863–2868.
- Yoshida, M., Miyoshi, I. & Hinuma, Y. (1982) Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. *Proceedings of the National Academy of Sciences of the United States of America*, 79, 2031–2035.

REFERENCES BIBLIOGRAPHIQUES

- Acel, A., Udashkin, B. E., Wainberg, M. A. & Faust, E. A. (1998). Efficient gap repair catalyzed *in vitro* by an intrinsic DNA polymerase activity of human immunodeficiency virus type 1 integrase. *J Virol* 72, 2062-71.
- Aiyar, A., Hindmarsh, P., Skalka, A. M. & Leis, J. (1996). Concerted integration of linear retroviral DNA by the avian sarcoma virus integrase *in vitro*: dependence on both long terminal repeat termini. *J Virol* 70, 3571-80.
- Akagi, T., Hoshida, Y., Yoshino, T., Teramoto, N., Kondo, E., Hayashi, K. & Takahashi, K. (1992). Infectivity of human T-lymphotropic virus type I to human nervous tissue cells *in vitro*. *Acta Neuropathol* 84, 147-52.
- Akagi, T., Ono, H., Tsuchida, N. & Shimotohno, K. (1997). Aberrant expression and function of p53 in T-cells immortalized by HTLV- I Tax1. *FEBS Lett* 406, 263-6.
- Andrake, M. D. & Skalka, A. M. (1995). Multimerization determinants reside in both the catalytic core and C terminus of avian sarcoma virus integrase. *J Biol Chem* 270, 29299-306.
- Ansari-Lari, M. A., Donehower, L. A. & Gibbs, R. A. (1995). Analysis of human immunodeficiency virus type 1 integrase mutants [published erratum appears in *Virology* 1995 Nov 10;213(2):680]. *Virology* 211, 332-5.
- Arya, S. K., Gallo, R. C., Hahn, B. H., Shaw, G. M., Popovic, M., Salahuddin, S. Z. & Wong-Staal, F. (1984). Homology of genome of AIDS-associated virus with genomes of human T- cell leukemia viruses. *Science* 225, 927-30.
- Asante-Appiah, E., Merkel, G. & Skalka, A. M. (1998). Purification of untagged retroviral integrases by immobilized metal ion affinity chromatography. *Protein Expr Purif* 12, 105-10.
- Bailly, C., Minnock, A. & Waring, M. J. (1996). A simple ligation assay to detect effects of drugs on the curvature/flexibility of DNA. *FEBS Lett* 396, 253-6.
- Bakker, A., Li, X., Ruland, C. T., Stephens, D. W., Black, A. C. & Rosenblatt, J. D. (1996). Human T-cell leukemia virus type 2 Rex inhibits pre-mRNA splicing *in vitro* at an early stage of spliceosome formation. *J Virol* 70, 5511-8.
- Balakrishnan, M. & Jonsson, C. B. (1997). Functional identification of nucleotides conferring substrate specificity to retroviral integrase reactions. *J Virol* 71, 1025-35.
- Bartoe, J. T., Albrecht, B., Collins, N. D., Robek, M. D., Ratner, L., Green, P. L. & Lairmore, M. D. (2000). Functional role of pX open reading frame II of human T-lymphotropic virus type 1 in maintenance of viral loads *in vivo*. *J Virol* 74, 1094-100.
- Benvenisty, N., Ornitz, D. M., Bennett, G. L., Sahagan, B. G., Kuo, A., Cardiff, R. D. & Leder, P. (1992). Brain tumours and lymphomas in transgenic mice that carry HTLV-I LTR/c- myc and Ig/tax genes. *Oncogene* 7, 2399-405.

- Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M. & Rodier, F. (1985). The mosaic genome of warm-blooded vertebrates. *Science* 228, 953-8.
- Bex, F., McDowall, A., Burny, A. & Gaynor, R. (1997). The human T-cell leukemia virus type 1 transactivator protein Tax colocalizes in unique nuclear structures with NF-kappaB proteins. *J Virol* 71, 3484-97.
- Bex, F., Yin, M. J., Burny, A. & Gaynor, R. B. (1998). Differential transcriptional activation by human T-cell leukemia virus type 1 Tax mutants is mediated by distinct interactions with CREB binding protein and p300. *Mol Cell Biol* 18, 2392-405.
- Bizub-Bender, D., Kulkosky, J. & Skalka, A. M. (1994). Monoclonal antibodies against HIV type 1 integrase: clues to molecular structure. *AIDS Res Hum Retroviruses* 10, 1105-15.
- Bogerd, H. P., Fridell, R. A., Madore, S. & Cullen, B. R. (1995). Identification of a novel cellular cofactor for the Rev/Rex class of retroviral regulatory proteins. *Cell* 82, 485-94.
- Bor, Y. C., Bushman, F. D. & Orgel, L. E. (1995). *In vitro* integration of human immunodeficiency virus type 1 cDNA into targets containing protein-induced bends. *Proc Natl Acad Sci U S A* 92, 10334-8.
- Bowerman, B., Brown, P. O., Bishop, J. M. & Varmus, H. E. (1989). A nucleoprotein complex mediates the integration of retroviral DNA. *Genes Dev* 3, 469-78.
- Brauweiler, A., Garrus, J. E., Reed, J. C. & Nyborg, J. K. (1997). Repression of bax gene expression by the HTLV-1 Tax protein: implications for suppression of apoptosis in virally infected cells. *Virology* 231, 135-40.
- Brockman, J. A., Scherer, D. C., McKinsey, T. A., Hall, S. M., Qi, X., Lee, W. Y. & Ballard, D. W. (1995). Coupling of a signal response domain in I kappa B alpha to multiple pathways for NF-kappa B activation. *Mol Cell Biol* 15, 2809-18.
- Brown, P. O., Bowerman, B., Varmus, H. E. & Bishop, J. M. (1987). Correct integration of retroviral DNA *in vitro*. *Cell* 49, 347-56.
- Brown, P. O., Bowerman, B., Varmus, H. E. & Bishop, J. M. (1989). Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. *Proc Natl Acad Sci U S A* 86, 2525-9.
- Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A. & Skalka, A. M. (1995). High-resolution structure of the catalytic domain of avian sarcoma virus integrase. *J Mol Biol* 253, 333-46.
- Bujacz, G., Alexandratos, J., Wlodawer, A., Merkel, G., Andrade, M., Katz, R. A. & Skalka, A. M. (1997). Binding of different divalent cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity. *J Biol Chem* 272, 18161-8.

- Bukovsky, A. & Gottlinger, H. (1996). Lack of integrase can markedly affect human immunodeficiency virus type 1 particle production in the presence of an active viral protease. *J Virol* 70, 6820-5.
- Bukrinsky, M. I., Sharova, N., Dempsey, M. P., Stanwick, T. L., Bukrinskaya, A. G., Haggerty, S. & Stevenson, M. (1992). Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. *Proc Natl Acad Sci U S A* 89, 6580-4.
- Bukrinsky, M. I., Haggerty, S., Dempsey, M. P., Sharova, N., Adzhubel, A., Spitz, L., Lewis, P., Goldfarb, D., Emerman, M. & Stevenson, M. (1993). A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells [see comments]. *Nature* 365, 666-9.
- Burke, C. J., Sanyal, G., Bruner, M. W., Ryan, J. A., LaFemina, R. L., Robbins, H. L., Zeff, A. S., Middaugh, C. R. & Cordingley, M. G. (1992). Structural implications of spectroscopic characterization of a putative zinc finger peptide from HIV-1 integrase. *J Biol Chem* 267, 9639-44.
- Bushman, F. D. & Craigie, R. (1991). Activities of human immunodeficiency virus (HIV) integration protein *in vitro*: specific cleavage and integration of HIV DNA. *Proc Natl Acad Sci U S A* 88, 1339-43.
- Bushman, F. D., Fujiwara, T. & Craigie, R. (1990). Retroviral DNA integration directed by HIV integration protein *in vitro*. *Science* 249, 1555-8.
- Bushman, F. D., Engelman, A., Palmer, I., Wingfield, P. & Craigie, R. (1993). Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. *Proc Natl Acad Sci U S A* 90, 3428-32.
- Bushman, F. D. & Wang, B. (1994a). Rous sarcoma virus integrase protein: mapping functions for catalysis and substrate binding. *J Virol* 68, 2215-23.
- Bushman, F. D. (1994b). Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. *Proc Natl Acad Sci U S A* 91, 9233-7.
- Bushman, F. D. & Miller, M. D. (1997). Tethering human immunodeficiency virus type 1 preintegration complexes to target DNA promotes integration at nearby sites. *J Virol* 71, 458-64.
- Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G. M. & Gronenborn, A. M. (1997). Solution structure of the N-terminal zinc binding domain of HIV-1 integrase [published erratum appears in Nat Struct Biol 1997 Oct;4(10):839-40]. *Nat Struct Biol* 4, 567-77.
- Cai, M., Huang, Y., Zheng, R., Wei, S. Q., Ghirlando, R., Lee, M. S., Craigie, R., Gronenborn, A. M. & Clore, G. M. (1998). Solution structure of the cellular factor BAF responsible for protecting retroviral DNA from autointegration. *Nat Struct Biol* 5, 903-9.

- Carteau, S., Batson, S. C., Poljak, L., Mouscadet, J. F., de Rocquigny, H., Darlix, J. L., Roques, B. P., Kas, E. & Auclair, C. (1997). Human immunodeficiency virus type 1 nucleocapsid protein specifically stimulates Mg²⁺-dependent DNA integration *in vitro*. *J Virol* 71, 6225-9.
- Carteau, S., Hoffmann, C. & Bushman, F. (1998). Chromosome structure and human immunodeficiency virus type 1 cDNA integration: centromeric alphoid repeats are a disfavored target. *J Virol* 72, 4005-14.
- Carteau, S., Gorelick, R. J. & Bushman, F. D. (1999). Coupled integration of human immunodeficiency virus type 1 cDNA ends by purified integrase *in vitro*: stimulation by the viral nucleocapsid protein. *J Virol* 73, 6670-9.
- Cavrois, M., Wain-Hobson, S. & Wattel, E. (1995). Stochastic events in the amplification of HTLV-I integration sites by linker-mediated PCR. *Res Virol* 146, 179-84.
- Cavrois, M., Gessain, A., Wain-Hobson, S. & Wattel, E. (1996a). Proliferation of HTLV-1 infected circulating cells *in vivo* in all asymptomatic carriers and patients with TSP/HAM. *Oncogene* 12, 2419-23.
- Cavrois, M., Wain-Hobson, S., Gessain, A., Plumelle, Y. & Wattel, E. (1996b). Adult T-cell leukemia/lymphoma on a background of clonally expanding human T-cell leukemia virus type-1-positive cells. *Blood* 88, 4646-50.
- Cavrois, M., Gessain, A., Gout, O., Wain-Hobson, S., & Wattel, E. (1997). HTLV-1 crosses the blood brain barrier via infected cells at the proviral state. *Submitted*.
- Cavrois, M., Leclercq, I., Gout, O., Gessain, A., Wain-Hobson, S. & Wattel, E. (1998). Persistent oligoclonal expansion of human T-cell leukemia virus type 1- infected circulating cells in patients with Tropical spastic paraparesis/HTLV-1 associated myelopathy. *Oncogene* 17, 77-82.
- Cereseto, A., Diella, F., Mulloy, J. C., Cara, A., Michieli, P., Grassmann, R., Franchini, G. & Klotman, M. E. (1996). p53 functional impairment and high p21waf1/cip1 expression in human T- cell lymphotropic/leukemia virus I-transformed T cells. *Blood* 88, 1551-60.
- Chechelnitsky, G. & Priel, E. (1998). DNA binding properties of the nucleocapsid protein from human T-cell leukemia virus type-I. *Int J Oncol* 13, 749-55.
- Chen, H. & Engelman, A. (1998). The barrier-to-autointegration protein is a host factor for HIV type 1 integration. *Proc Natl Acad Sci U S A* 95, 15270-4.
- Chou, K. S., Okayama, A., Su, I. J., Lee, T. H. & Essex, M. (1996). Preferred nucleotide sequence at the integration target site of human T- cell leukemia virus type I from patients with adult T-cell leukemia. *Int J Cancer* 65, 20-4.

- Chow, S. A., Vincent, K. A., Ellison, V. & Brown, P. O. (1992). Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. *Science* 255, 723-6.
- Chow, S. A. & Brown, P. O. (1994). Substrate features important for recognition and catalysis by human immunodeficiency virus type 1 integrase identified by using novel DNA substrates. *J Virol* 68, 3896-907.
- Cobrinik, D., Katz, R., Terry, R., Skalka, A. M. & Leis, J. (1987). Avian sarcoma and leukosis virus pol-endonuclease recognition of the tandem long terminal repeat junction: minimum site required for cleavage is also required for viral growth. *J Virol* 61, 1999-2008.
- Coffin, J. (1996). Retroviridae : The viruses and their Replication. *Fields Virology*, 58, 1767.
- Colgin, M. A. & Nyborg, J. K. (1998). The human T-cell leukemia virus type 1 oncoprotein Tax inhibits the transcriptional activity of c-Myb through competition for the CREB binding protein. *J Virol* 72, 9396-9.
- Colicelli, J. & Goff, S. P. (1985). Mutants and pseudorevertants of Moloney murine leukemia virus with alterations at the integration site. *Cell* 42, 573-80.
- Colicelli, J. & Goff, S. P. (1988). Sequence and spacing requirements of a retrovirus integration site. *J Mol Biol* 199, 47-59.
- Collins, N. D., Newbound, G. C., Albrecht, B., Beard, J. L., Ratner, L. & Lairmore, M. D. (1998). Selective ablation of human T-cell lymphotropic virus type 1 p12I reduces viral infectivity *in vivo*. *Blood* 91, 4701-7.
- Collins, N. D., D'Souza, C., Albrecht, B., Robek, M. D., Ratner, L., Ding, W., Green, P. L. & Lairmore, M. D. (1999). Proliferation response to interleukin-2 and Jak/Stat activation of T cells immortalized by human T-cell lymphotropic virus type 1 is independent of open reading frame I expression. *J Virol* 73, 9642-9.
- Craigie, R., Fujiwara, T. & Bushman, F. (1990). The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration *in vitro*. *Cell* 62, 829-37.
- Daenke, S., Kermode, A. G., Hall, S. E., Taylor, G., Weber, J., Nightingale, S. & Bangham, C. R. (1996). High activated and memory cytotoxic T-cell responses to HTLV-1 in healthy carriers and patients with tropical spastic paraparesis. *Virology* 217, 139-46.
- Dagleish, A., Richardson, J., Matutes, E., Cruickshank, K., Newell, A., Sinclair, A., Thorpe, R., Brasher, M., Weber, J., Catovsky, D. & et al. (1988). HTLV-1 infection in tropical spastic paraparesis: lymphocyte culture and serologic response. *AIDS Res Hum Retroviruses* 4, 475-85.
- Daniel, R., Katz, R. A. & Skalka, A. M. (1999). A role for DNA-PK in retroviral DNA integration. *Science* 284, 644-7.

- D'Incan, M., Combemale, P., Verrier, B., Garin, D., Audoly, G., Brunot, J., Desgranges, C. & Flechaire, A. (1994). Transient adult T-cell leukemia/lymphoma picture during varicella infection in an HTLV-1 carrier. *Leukemia* 8, 682-7.
- Donzella, G. A., Leon, O. & Roth, M. J. (1998). Implication of a central cysteine residue and the HHCC domain of Moloney murine leukemia virus integrase protein in functional multimerization. *J Virol* 72, 1691-8.
- Drelich, M., Wilhelm, R. & Mous, J. (1992). Identification of amino acid residues critical for endonuclease and integration activities of HIV-1 IN protein *in vitro*. *Virology* 188, 459-68.
- Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R. & Davies, D. R. (1994). Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases [see comments]. *Science* 266, 1981-6.
- Eguchi, K., Matsuoka, N., Ida, H., Nakashima, M., Sakai, M., Sakito, S., Kawakami, A., Terada, K., Shimada, H., Kawabe, Y. & et al. (1992). Primary Sjogren's syndrome with antibodies to HTLV-I: clinical and laboratory features. *Ann Rheum Dis* 51, 769-76.
- Eijkelenboom, A. P., Lutzke, R. A., Boelens, R., Plasterk, R. H., Kaptein, R. & Hard, K. (1995). The DNA-binding domain of HIV-1 integrase has an SH3-like fold. *Nat Struct Biol* 2, 807-10.
- Eijkelenboom, A. P., van den Ent, F. M., Vos, A., Doreleijers, J. F., Hard, K., Tullius, T. D., Plasterk, R. H., Kaptein, R. & Boelens, R. (1997). The solution structure of the amino-terminal HHCC domain of HIV-2 integrase: a three-helix bundle stabilized by zinc. *Curr Biol* 7, 739-46.
- Ellison, V. & Brown, P. O. (1994). A stable complex between integrase and viral DNA ends mediates human immunodeficiency virus integration *in vitro*. *Proc Natl Acad Sci U S A* 91, 7316-20.
- Elovaara, I., Koenig, S., Brewah, A. Y., Woods, R. M., Lehky, T. & Jacobson, S. (1993). High human T cell lymphotropic virus type 1 (HTLV-1)-specific precursor cytotoxic T lymphocyte frequencies in patients with HTLV-1-associated neurological disease. *J Exp Med* 177, 1567-73.
- Engelman, A., Mizuuchi, K. & Craigie, R. (1991). HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell* 67, 1211-21.
- Engelman, A. & Craigie, R. (1992). Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function *in vitro*. *J Virol* 66, 6361-9.
- Engelman, A., Bushman, F. D. & Craigie, R. (1993). Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. *Embo J* 12, 3269-75.

- Engelman, A., Hickman, A. B. & Craigie, R. (1994). The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. *J Virol* 68, 5911-7.
- Engelman, A., Englund, G., Orenstein, J. M., Martin, M. A. & Craigie, R. (1995a). Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. *J Virol* 69, 2729-36.
- Engelman, A. & Craigie, R. (1995b). Efficient magnesium-dependent human immunodeficiency virus type 1 integrase activity. *J Virol* 69, 5908-11.
- Farnet, C. M. & Haseltine, W. A. (1990). Integration of human immunodeficiency virus type 1 DNA *in vitro*. *Proc Natl Acad Sci U S A* 87, 4164-8.
- Farnet, C. M. & Haseltine, W. A. (1991a). Circularization of human immunodeficiency virus type 1 DNA *in vitro*. *J Virol* 65, 6942-52.
- Farnet, C. M. & Haseltine, W. A. (1991b). Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. *J Virol* 65, 1910-5.
- Farnet, C. M. & Bushman, F. D. (1996). HIV cDNA integration: molecular biology and inhibitor development. *Aids* 10, S3-11.
- Farnet, C. M. & Bushman, F. D. (1997). HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes *in vitro*. *Cell* 88, 483-92.
- Felber, B. K., Paskalis, H., Kleinman-Ewing, C., Wong-Staal, F. & Pavlakis, G. N. (1985). The pX protein of HTLV-I is a transcriptional activator of its long terminal repeats. *Science* 229, 675-9.
- Fitzgerald, M. L., Vora, A. C., Zeh, W. G. & Grandgenett, D. P. (1992). Concerted integration of viral DNA termini by purified avian myeloblastosis virus integrase. *J Virol* 66, 6257-63.
- Fitzgerald, M. L. & Grandgenett, D. P. (1994). Retroviral integration: *in vitro* host site selection by avian integrase. *J Virol* 68, 4314-21.
- Fouchier, R. A., Meyer, B. E., Simon, J. H., Fischer, U. & Malim, M. H. (1997). HIV-1 infection of non-dividing cells: evidence that the amino-terminal basic region of the viral matrix protein is important for Gag processing but not for post-entry nuclear import. *Embo J* 16, 4531-9.
- Franchini, G., Mulloy, J. C., Koralnik, I. J., Lo Monaco, A., Sparkowski, J. J., Andresson, T., Goldstein, D. J. & Schlegel, R. (1993). The human T-cell leukemia/lymphotropic virus type I p12I protein cooperates with the E5 oncoprotein of bovine papillomavirus in cell transformation and binds the 16-kilodalton subunit of the vacuolar H⁺ ATPase. *J Virol* 67, 7701-4.

- Franklin, A. A., Kubik, M. F., Uittenbogaard, M. N., Brauweiler, A., Utaisincharoen, P., Matthews, M. A., Dynan, W. S., Hoeffler, J. P. & Nyborg, J. K. (1993). Transactivation by the human T-cell leukemia virus Tax protein is mediated through enhanced binding of activating transcription factor-2 (ATF-2) ATF-2 response and cAMP element-binding protein (CREB). *J Biol Chem* 268, 21225-31.
- Freed, E. O., Englund, G. & Martin, M. A. (1995). Role of the basic domain of human immunodeficiency virus type 1 matrix in macrophage infection. *J Virol* 69, 3949-54.
- Fritz, C. C., Zapp, M. L. & Green, M. R. (1995). A human nucleoporin-like protein that specifically interacts with HIV Rev. *Nature* 376, 530-3.
- Fujii, M., Niki, T., Mori, T., Matsuda, T., Matsui, M., Nomura, N. & Seiki, M. (1991). HTLV-1 Tax induces expression of various immediate early serum responsive genes. *Oncogene* 6, 1023-9.
- Fujiwara, T. & Mizuuchi, K. (1988). Retroviral DNA integration: structure of an integration intermediate. *Cell* 54, 497-504.
- Fujiwara, T. & Craigie, R. (1989). Integration of mini-retroviral DNA: a cell-free reaction for biochemical analysis of retroviral integration. *Proc Natl Acad Sci U S A* 86, 3065-9.
- Gallay, P., Swingler, S., Song, J., Bushman, F. & Trono, D. (1995). HIV nuclear import is governed by the phosphotyrosine-mediated binding of matrix to the core domain of integrase. *Cell* 83, 569-76.
- Gallay, P., Stitt, V., Mundy, C., Oettinger, M. & Trono, D. (1996). Role of the karyopherin pathway in human immunodeficiency virus type 1 nuclear import. *J Virol* 70, 1027-32.
- Gallay, P., Hope, T., Chin, D. & Trono, D. (1997). HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway. *Proc Natl Acad Sci U S A* 94, 9825-30.
- Geleziunas, R., Ferrell, S., Lin, X., Mu, Y., Cunningham, E. T., Jr., Grant, M., Connelly, M. A., Hambor, J. E., Marcu, K. B. & Greene, W. C. (1998). Human T-cell leukemia virus type 1 Tax induction of NF-kappaB involves activation of the IkappaB kinase alpha (IKKalpha) and IKKbeta cellular kinases. *Mol Cell Biol* 18, 5157-65.
- Gessain, A., Barin, F., Vernant, J. C., Gout, O., Maurs, L., Calender, A. & de The, G. (1985). Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. *Lancet* 2, 407-10.
- Gessain, A., Gallo, R. C. & Franchini, G. (1992). Low degree of human T-cell leukemia/lymphoma virus type I genetic drift *in vivo* as a means of monitoring viral transmission and movement of ancient human populations. *J Virol* 66, 2288-95.

- Gessain, A., Boeri, E., Yanagihara, R., Gallo, R. C. & Franchini, G. (1993). Complete nucleotide sequence of a highly divergent human T-cell leukemia (lymphotropic) virus type I (HTLV-I) variant from melanesia: genetic and phylogenetic relationship to HTLV-I strains from other geographical regions. *J Virol* 67, 1015-23.
- Gessain, A., Mahieux, R. & de The, G. (1996). Genetic variability and molecular epidemiology of human and simian T cell leukemia/lymphoma virus type I. *J Acquir Immune Defic Syndr Hum Retrovirol* 13, S132-45.
- Giebler, H. A., Loring, J. E., van Orden, K., Colgin, M. A., Garrus, J. E., Escudero, K. W., Brauweiler, A. & Nyborg, J. K. (1997). Anchoring of CREB binding protein to the human T-cell leukemia virus type 1 promoter: a molecular mechanism of Tax transactivation. *Mol Cell Biol* 17, 5156-64.
- Goulaouic, H. & Chow, S. A. (1996). Directed integration of viral DNA mediated by fusion proteins consisting of human immunodeficiency virus type 1 integrase and Escherichia coli LexA protein. *J Virol* 70, 37-46.
- Grandgenett, D. P. & Goodarzi, G. (1994). Folding of the multidomain human immunodeficiency virus type-I integrase. *Protein Sci* 3, 888-97.
- Grandgenett, D. P., Inman, R. B., Vora, A. C. & Fitzgerald, M. L. (1993). Comparison of DNA binding and integration half-site selection by avian myeloblastosis virus integrase. *J Virol* 67, 2628-36.
- Grassmann, R., Dengler, C., Muller-Fleckenstein, I., Fleckenstein, B., McGuire, K., Dokhelar, M. C., Sodroski, J. G. & Haseltine, W. A. (1989). Transformation to continuous growth of primary human T lymphocytes by human T-cell leukemia virus type I X-region genes transduced by a Herpesvirus saimiri vector. *Proc Natl Acad Sci U S A* 86, 3351-5.
- Grassmann, R., Berchtold, S., Radant, I., Alt, M., Fleckenstein, B., Sodroski, J. G., Haseltine, W. A. & Ramstedt, U. (1992). Role of human T-cell leukemia virus type 1 X region proteins in immortalization of primary human lymphocytes in culture. *J Virol* 66, 4570-5.
- Green, P. L., Yip, M. T., Xie, Y. & Chen, I. S. (1992). Phosphorylation regulates RNA binding by the human T-cell leukemia virus Rex protein. *J Virol* 66, 4325-30.
- Hakata, Y., Umemoto, T., Matsushita, S. & Shida, H. (1998). Involvement of human CRM1 (exportin 1) in the export and multimerization of the Rex protein of human T-cell leukemia virus type 1. *J Virol* 72, 6602-7.
- Hanabuchi, S., Ohashi, T., Koya, Y., Kato, H., Takemura, F., Hirokawa, K., Yoshiki, T., Yagita, H., Okumura, K. & Kannagi, M. (2000). Development of human T-cell leukemia virus type 1-transformed tumors in rats following suppression of T-cell immunity by CD80 and CD86 blockade. *J Virol* 74, 428-35.

- Hanon, E., Hall, S., Taylor, G. P., Saito, M., Davis, R., Tanaka, Y., Usuku, K., Osame, M., Weber, J. N. & Bangham, C. R. (2000). Abundant tax protein expression in CD4+ T cells infected with human T-cell lymphotropic virus type I (HTLV-I) is prevented by cytotoxic T lymphocytes. *Blood* 95, 1386-92.
- Hara, H., Morita, M., Iwaki, T., Hatae, T., Itoyama, Y., Kitamoto, T., Akizuki, S., Goto, I. & Watanabe, T. (1994). Detection of human T lymphotropic virus type I (HTLV-I) proviral DNA and analysis of T cell receptor V beta CDR3 sequences in spinal cord lesions of HTLV-I-associated myelopathy/tropical spastic paraparesis. *J Exp Med* 180, 831-9.
- Harel, J., Rassart, E. & Jolicoeur, P. (1981). Cell cycle dependence of synthesis of unintegrated viral DNA in mouse cells newly infected with murine leukemia virus. *Virology* 110, 202-7.
- Hattori, T., Uchiyama, T., Toibana, T., Takatsuki, K. & Uchino, H. (1981). Surface phenotype of Japanese adult T-cell leukemia cells characterized by monoclonal antibodies. *Blood* 58, 645-7.
- Hayes, J. J., Bashkin, J., Tullius, T. D. & Wolffe, A. P. (1991). The histone core exerts a dominant constraint on the structure of DNA in a nucleosome. *Biochemistry* 30, 8434-40.
- Hazuda, D. J., Felock, P., Witmer, M., Wolfe, A., Stillmock, K., Grobler, J. A., Espeseth, A., Gabryelski, L., Schleif, W., Blau, C. & Miller, M. D. (2000). Inhibitors of strand transfer that prevent integration and inhibit HIV- 1 replication in cells. *Science* 287, 646-50.
- Heinzinger, N. K., Bukinsky, M. I., Haggerty, S. A., Ragland, A. M., Kewalramani, V., Lee, M. A., Gendelman, H. E., Ratner, L., Stevenson, M. & Emerman, M. (1994). The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. *Proc Natl Acad Sci U S A* 91, 7311-5.
- Heuer, T. S. & Brown, P. O. (1997). Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photo-cross- linking. *Biochemistry* 36, 10655-65.
- Hidaka, M., Inoue, J., Yoshida, M. & Seiki, M. (1988). Post-transcriptional regulator (rex) of HTLV-1 initiates expression of viral structural proteins but suppresses expression of regulatory proteins. *Embo J* 7, 519-23.
- Hindmarsh, P., Ridky, T., Reeves, R., Andrade, M., Skalka, A. M. & Leis, J. (1999). HMG protein family members stimulate human immunodeficiency virus type 1 and avian sarcoma virus concerted DNA integration *in vitro*. *J Virol* 73, 2994-3003.
- Hiramatsu, K. & Yoshikura, H. (1986). Frequent partial deletion of human adult T-cell leukemia virus type I proviruses in experimental transmission: pattern and possible implication. *J Virol* 58, 508-12.

- Ho, D. D., Rota, T. R. & Hirsch, M. S. (1984). Infection of human endothelial cells by human T-lymphotropic virus type I. *Proc Natl Acad Sci U S A* 81, 7588-90.
- Hoffman, P. M., Dhib-Jalbut, S., Mikovits, J. A., Robbins, D. S., Wolf, A. L., Bergey, G. K., Lohrey, N. C., Weislow, O. S. & Ruscetti, F. W. (1992). Human T-cell leukemia virus type I infection of monocytes and microglial cells in primary human cultures. *Proc Natl Acad Sci U S A* 89, 11784-8.
- Howard, M. T. & Griffith, J. D. (1993). A cluster of strong topoisomerase II cleavage sites is located near an integrated human immunodeficiency virus. *J Mol Biol* 232, 1060-8.
- Ijichi, S., Izumo, S., Eiraku, N., Machigashira, K., Kubota, R., Nagai, M., Ikegami, N., Kashio, N., Umehara, F., Maruyama, I. & et al. (1993). An autoaggressive process against bystander tissues in HTLV-I-infected individuals: a possible pathomechanism of HAM/TSP. *Med Hypotheses* 41, 542-7.
- Inoue, J., Itoh, M., Akizawa, T., Toyoshima, H. & Yoshida, M. (1991). HTLV-1 Rex protein accumulates unspliced RNA in the nucleus as well as in cytoplasm. *Oncogene* 6, 1753-7.
- Itoyama, Y., Minato, S., Kira, J., Goto, I., Sato, H., Okochi, K. & Yamamoto, N. (1988). Spontaneous proliferation of peripheral blood lymphocytes increased in patients with HTLV-I-associated myelopathy. *Neurology* 38, 1302-7.
- Jacobson, S., Shida, H., McFarlin, D. E., Fauci, A. S. & Koenig, S. (1990). Circulating CD8+ cytotoxic T lymphocytes specific for HTLV-I pX in patients with HTLV-I associated neurological disease. *Nature* 348, 245-8.
- Jacobson, S., McFarlin, D. E., Robinson, S., Voskuhl, R., Martin, R., Brewah, A., Newell, A. J. & Koenig, S. (1992). HTLV-I-specific cytotoxic T lymphocytes in the cerebrospinal fluid of patients with HTLV-I-associated neurological disease. *Ann Neurol* 32, 651-7.
- Jeang, K. T., Widen, S. G., Semmes, O. J. t. & Wilson, S. H. (1990). HTLV-I trans-activator protein, tax, is a trans-repressor of the human beta-polymerase gene. *Science* 247, 1082-4.
- Jeffery, K. J., Usuku, K., Hall, S. E., Matsumoto, W., Taylor, G. P., Procter, J., Bunce, M., Ogg, G. S., Welsh, K. I., Weber, J. N., Lloyd, A. L., Nowak, M. A., Nagai, M., Kodama, D., Izumo, S., Osame, M. & Bangham, C. R. (1999). HLA alleles determine human T-lymphotropic virus-I (HTLV-I) proviral load and the risk of HTLV-I-associated myelopathy. *Proc Natl Acad Sci U S A* 96, 3848-53.
- Jenkins, T. M., Hickman, A. B., Dyda, F., Ghirlando, R., Davies, D. R. & Craigie, R. (1995). Catalytic domain of human immunodeficiency virus type 1 integrase: identification of a soluble mutant by systematic replacement of hydrophobic residues. *Proc Natl Acad Sci U S A* 92, 6057-61.

- Jenkins, T. M., Engelman, A., Ghirlando, R. & Craigie, R. (1996). A soluble active mutant of HIV-1 integrase: involvement of both the core and carboxyl-terminal domains in multimerization. *J Biol Chem* 271, 7712-8.
- Jenks, P. J., Barrett, W. Y., Raftery, M. J., Kelsey, S. M., van der Walt, J. D., Kon, S. P. & Breuer, J. (1995). Development of human T-cell lymphotropic virus type I-associated adult T-cell leukemia/lymphoma during immunosuppressive treatment following renal transplantation. *Clin Infect Dis* 21, 992-3.
- Jiang, H., Lu, H., Schiltz, R. L., Pise-Masison, C. A., Ogryzko, V. V., Nakatani, Y. & Brady, J. N. (1999). PCAF interacts with tax and stimulates tax transactivation in a histone acetyltransferase-independent manner. *Mol Cell Biol* 19, 8136-45.
- Jin, D. Y., Spencer, F. & Jeang, K. T. (1998). Human T cell leukemia virus type 1 oncoprotein Tax targets the human mitotic checkpoint protein MAD1. *Cell* 93, 81-91.
- Johnson, M. S., McClure, M. A., Feng, D. F., Gray, J. & Doolittle, R. F. (1986). Computer analysis of retroviral pol genes: assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes. *Proc Natl Acad Sci U S A* 83, 7648-52.
- Jones, K. S., Coleman, J., Merkel, G. W., Laue, T. M. & Skalka, A. M. (1992). Retroviral integrase functions as a multimer and can turn over catalytically. *J Biol Chem* 267, 16037-40.
- Kahn, J. D. & Crothers, D. M. (1992). Protein-induced bending and DNA cyclization. *Proc Natl Acad Sci U S A* 89, 6343-7.
- Kalpana, G. V., Marmon, S., Wang, W., Crabtree, G. R. & Goff, S. P. (1994). Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5 [see comments]. *Science* 266, 2002-6.
- Kannagi, M., Harada, S., Maruyama, I., Inoko, H., Igarashi, H., Kuwashima, G., Sato, S., Morita, M., Kidokoro, M., Sugimoto, M. & et al. (1991). Predominant recognition of human T cell leukemia virus type I (HTLV-I) pX gene products by human CD8+ cytotoxic T cells directed against HTLV-I-infected cells. *Int Immunol* 3, 761-7.
- Kannagi, M., Matsushita, S., Shida, H. & Harada, S. (1994). Cytotoxic T cell response and expression of the target antigen in HTLV- I infection. *Leukemia* 8 Suppl 1, S54-9.
- Kanno, T., Brown, K. & Siebenlist, U. (1995). Evidence in support of a role for human T-cell leukemia virus type I Tax in activating NF-kappa B via stimulation of signaling pathways. *J Biol Chem* 270, 11745-8.
- Kao, S. Y. & Marriott, S. J. (1999). Disruption of nucleotide excision repair by the human T-cell leukemia virus type 1 Tax protein. *J Virol* 73, 4299-304.

- Katahira, Y., Yashiki, S., Fujiyoshi, T., Nomura, K., Tara, M., Mori, M., Setoyama, M., Kanzaki, T., Shida, H. & Sonoda, S. (1995). *In vitro* induction of cytotoxic T lymphocytes against HTLV-I-infected T-cells from adult T-cell leukemia patients, asymptomatic HTLV-I carriers and seronegative healthy donors. *Jpn J Cancer Res* 86, 21-7.
- Katz, R. A., Merkel, G., Kulkosky, J., Leis, J. & Skalka, A. M. (1990). The avian retroviral IN protein is both necessary and sufficient for integrative recombination *in vitro*. *Cell* 63, 87-95.
- Katz, R. A., Merkel, G. & Skalka, A. M. (1996). Targeting of retroviral integrase by fusion to a heterologous DNA binding domain: *in vitro* activities and incorporation of a fusion protein into viral particles. *Virology* 217, 178-90.
- Katz, R. A., Gravuer, K. & Skalka, A. M. (1998). A preferred target DNA structure for retroviral integrase *in vitro*. *J Biol Chem* 273, 24190-5.
- Katzman, M., Katz, R. A., Skalka, A. M. & Leis, J. (1989). The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the *in vivo* sites of integration. *J Virol* 63, 5319-27.
- Katzman, M. & Sudol, M. (1994). *In vitro* activities of purified visna virus integrase. *J Virol* 68, 3558-69.
- Katzman, M. & Sudol, M. (1995). Mapping domains of retroviral integrase responsible for viral DNA specificity and target site selection by analysis of chimeras between human immunodeficiency virus type 1 and visna virus integrases. *J Virol* 69, 5687-96.
- Katzman, M. & Sudol, M. (1996). Influence of subterminal viral DNA nucleotides on differential susceptibility to cleavage by human immunodeficiency virus type 1 and visna virus integrases. *J Virol* 70, 9069-73.
- Katzman, M. & Sudol, M. (1998). Mapping viral DNA specificity to the central region of integrase by using functional human immunodeficiency virus type 1/visna virus chimeric proteins. *J Virol* 72, 1744-53.
- Kazanji, M., Ureta-Vidal, A., Ozden, S., Tangy, F., de Thoisy, B., Fiette, L., Talarmin, A., Gessain, A., & de The, G. (2000). Lymphoid organs as a major reservoir for Human T-Cell Leukemia Virus Type 1 in experimentally infected Squirrel monkeys (*Saimiri sciureus*): Provirus expression, persistence, and humoral and cellular immune response. *J Virol. in Press*.
- Khan, E., Mack, J. P., Katz, R. A., Kulkosky, J. & Skalka, A. M. (1991). Retroviral integrase domains: DNA binding and the recognition of LTR sequences [published erratum appears in *Nucleic Acids Res* 1991 Mar 25;19(6):1358]. *Nucleic Acids Res* 19, 851-60.

- Kira, J., Koyanagi, Y., Yamada, T., Itoyama, Y., Tateishi, J., Akizuki, S., Kishikawa, M., Baba, E., Nakamura, M., Suzuki, J. & et al. (1994). Sequence heterogeneity of HTLV-I proviral DNA in the central nervous system of patients with HTLV-I-associated myelopathy [see comments]. *Ann Neurol* 36, 149-56.
- Kitamura, Y., Lee, Y. M. & Coffin, J. M. (1992). Nonrandom integration of retroviral DNA *in vitro*: effect of CpG methylation. *Proc Natl Acad Sci U S A* 89, 5532-6.
- Koenig, S., Woods, R. M., Brewah, Y. A., Newell, A. J., Jones, G. M., Boone, E., Adelsberger, J. W., Baseler, M. W., Robinson, S. M. & Jacobson, S. (1993). Characterization of MHC class I restricted cytotoxic T cell responses to tax in HTLV-1 infected patients with neurologic disease. *J Immunol* 151, 3874-83.
- Korber, B., Okayama, A., Donnelly, R., Tachibana, N. & Essex, M. (1991). Polymerase chain reaction analysis of defective human T-cell leukemia virus type I proviral genomes in leukemic cells of patients with adult T-cell leukemia. *J Virol* 65, 5471-6.
- Kramer, A., Jacobson, S., Reuben, J. F., Murphy, E. L., Wiktor, S. Z., Cranston, B., Figueroa, J. P., Hanchard, B., McFarlin, D. & Blattner, W. A. (1989). Spontaneous lymphocyte proliferation in symptom-free HTLV-I positive Jamaicans [letter]. *Lancet* 2, 923-4.
- Kubota, S., Furuta, R., Maki, M., Siomi, H. & Hatanaka, M. (1993). Long cellular repeats flanking a defective HTLV-I provirus: implication for site-targeted integration. *Oncogene* 8, 2873-7.
- Kukolj, G. & Skalka, A. M. (1995). Enhanced and coordinated processing of synapsed viral DNA ends by retroviral integrases *in vitro*. *Genes Dev* 9, 2556-67.
- Kulkosky, J., Jones, K. S., Katz, R. A., Mack, J. P. & Skalka, A. M. (1992). Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol Cell Biol* 12, 2331-8.
- Kwok, R. P., Laurance, M. E., Lundblad, J. R., Goldman, P. S., Shih, H., Connor, L. M., Marriott, S. J. & Goodman, R. H. (1996). Control of cAMP-regulated enhancers by the viral transactivator Tax through CREB and the co-activator CBP. *Nature* 380, 642-6.
- LaFemina, R. L., Callahan, P. L. & Cordingley, M. G. (1991). Substrate specificity of recombinant human immunodeficiency virus integrase protein. *J Virol* 65, 5624-30.
- Lau, A., Gant, T. W. & Cann, A. J. (1998). The mechanism of trans-activation of the MDR1 gene by human T-cell leukemia virus. *Biochem Biophys Res Commun* 249, 397-404.
- Le Blanc, I., Rosenberg, A. R. & Dokhelar, M. C. (1999). Multiple functions for the basic amino acids of the human T-cell leukemia virus type 1 matrix protein in viral transmission. *J Virol* 73, 1860-7.

- Leavitt, A. D., Rose, R. B. & Varmus, H. E. (1992). Both substrate and target oligonucleotide sequences affect *in vitro* integration mediated by human immunodeficiency virus type 1 integrase protein produced in *Saccharomyces cerevisiae*. *J Virol* 66, 2359-68.
- Leavitt, A. D., Shiue, L. & Varmus, H. E. (1993). Site-directed mutagenesis of HIV-1 integrase demonstrates differential effects on integrase functions *in vitro*. *J Biol Chem* 268, 2113-9.
- Leavitt, A. D., Robles, G., Alesandro, N. & Varmus, H. E. (1996). Human immunodeficiency virus type 1 integrase mutants retain *in vitro* integrase activity yet fail to integrate viral DNA efficiently during infection. *J Virol* 70, 721-8.
- Leclercq, I., Cavrois, M., Mortreux, F., Hermine, O., Gessain, A., Morschhauser, F. & Wattel, E. (1998). Oligoclonal proliferation of human T-cell leukaemia virus type 1 bearing T cells in adult T-cell leukaemia/lymphoma without deletion of the 3' provirus integration sites. *Br J Haematol* 101, 500-6.
- Leclercq, I., Mortreux, F., Morschhauser, F., Duthilleul, P., Desgranges, C., Gessain, A., Cavrois, M., Vernant, J. P., Hermine, O. & Wattel, E. (1999). Semiquantitative analysis of residual disease in patients treated for adult T-cell leukaemia/lymphoma (ATLL). *Br J Haematol* 105, 743-51.
- Leclercq, I., Mortreux, F., Cavrois, M., Leroy, A., Gessain, A., Wain-Hobson, S. & Wattel, E. (2000). Host sequences flanking the human T-cell leukemia virus type 1 provirus *In vivo* [In Process Citation]. *J Virol* 74, 2305-12.
- Lee, Y. M. & Coffin, J. M. (1990). Efficient autointegration of avian retrovirus DNA *in vitro*. *J Virol* 64, 5958-65.
- Lee, M. S. & Craigie, R. (1994). Protection of retroviral DNA from autointegration: involvement of a cellular factor. *Proc Natl Acad Sci U S A* 91, 9823-7.
- Lee, S. P., Xiao, J., Knutson, J. R., Lewis, M. S. & Han, M. K. (1997). Zn²⁺ promotes the self-association of human immunodeficiency virus type- 1 integrase *in vitro*. *Biochemistry* 36, 173-80.
- Lee, M. S. & Craigie, R. (1998). A previously unidentified host protein protects retroviral DNA from autointegration. *Proc Natl Acad Sci U S A* 95, 1528-33.
- Lemasson, I., Robert-Hebmann, V., Hamaia, S., Duc Dodon, M., Gazzolo, L. & Devaux, C. (1997). Transrepression of lck gene expression by human T-cell leukemia virus type 1-encoded p40tax. *J Virol* 71, 1975-83.
- Lemasson, I., Thebault, S., Sardet, C., Devaux, C. & Mesnard, J. M. (1998). Activation of E2F-mediated transcription by human T-cell leukemia virus type I Tax protein in a p16(INK4A)-negative T-cell line. *J Biol Chem* 273, 23598-604.

- Lewis, P., Hensel, M. & Emerman, M. (1992). Human immunodeficiency virus infection of cells arrested in the cell cycle [published erratum appears in EMBO J 1992 Nov;11(11):4249]. *Embo J* 11, 3053-8.
- Lewis, P. F. & Emerman, M. (1994). Passage through mitosis is required for oncoretroviruses but not for the human immunodeficiency virus. *J Virol* 68, 510-6.
- Li, L., Farnet, C. M., Anderson, W. F. & Bushman, F. D. (1998). Modulation of activity of Moloney murine leukemia virus preintegration complexes by host factors *in vitro*. *J Virol* 72, 2125-31.
- Lodi, P. J., Ernst, J. A., Kuszewski, J., Hickman, A. B., Engelman, A., Craigie, R., Clore, G. M. & Gronenborn, A. M. (1995). Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry* 34, 9826-33.
- Low, K. G., Dorner, L. F., Fernando, D. B., Grossman, J., Jeang, K. T. & Comb, M. J. (1997). Human T-cell leukemia virus type 1 Tax releases cell cycle arrest induced by p16INK4a. *J Virol* 71, 1956-62.
- Macatonia, S. E., Cruickshank, J. K., Rudge, P. & Knight, S. C. (1992). Dendritic cells from patients with tropical spastic paraparesis are infected with HTLV-1 and stimulate autologous lymphocyte proliferation. *AIDS Res Hum Retroviruses* 8, 1699-706.
- Macera, M. J., Szabo, P. & Verma, R. S. (1992). Chromosomal localization of HTLV-1 viral integration sites using in situ hybridization: detection of a novel IL2R fragment. *Mol Gen Genet* 234, 466-74.
- Mansky, L. M. & Temin, H. M. (1994). Lower mutation rate of bovine leukemia virus relative to that of spleen necrosis virus. *J Virol* 68, 494-9.
- Marini, J. C., Effron, P. N., Goodman, T. C., Singleton, C. K., Wells, R. D., Wartell, R. M. & Englund, P. T. (1984). Physical characterization of a kinetoplast DNA fragment with unusual properties. *J Biol Chem* 259, 8974-9.
- Maruyama, K., Fukushima, T., Kawamura, K. & Mochizuki, S. (1990). Chromosome and gene rearrangements in immortalized human lymphocytes infected with human T-lymphotropic virus type I. *Cancer Res* 50, 5697S-5702S.
- Masuda, T., Planelles, V., Krogstad, P. & Chen, I. S. (1995). Genetic analysis of human immunodeficiency virus type 1 integrase and the U3 att site: unusual phenotype of mutants in the zinc finger-like domain. *J Virol* 69, 6687-96.
- Mazumder, A. & Pommier, Y. (1995). Processing of deoxyuridine mismatches and abasic sites by human immunodeficiency virus type-1 integrase. *Nucleic Acids Res* 23, 2865-71.
- Migone, T. S., Lin, J. X., Cereseto, A., Mulloy, J. C., O'Shea, J. J., Franchini, G. & Leonard, W. J. (1995). Constitutively activated Jak-STAT pathway in T cells transformed with HTLV-I. *Science* 269, 79-81.

- Miller, M. D., Bor, Y. C. & Bushman, F. (1995). Target DNA capture by HIV-1 integration complexes. *Curr Biol* 5, 1047-56.
- Miller, M. D., Farnet, C. M. & Bushman, F. D. (1997). Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J Virol* 71, 5382-90.
- Miller, M. J. & Yuan, B. Z. (1997). Semiautomated resolution of overlapping stutter patterns in genomic microsatellite analysis. *Anal Biochem* 251, 50-6.
- Milot, E., Belmaaza, A., Rassart, E. & Chartrand, P. (1994). Association of a host DNA structure with retroviral integration sites in chromosomal DNA. *Virology* 201, 408-12.
- Miyake, H., Suzuki, T., Hirai, H. & Yoshida, M. (1999). Trans-activator Tax of human T-cell leukemia virus type 1 enhances mutation frequency of the cellular genome. *Virology* 253, 155-61.
- Miyazaki, T., Liu, Z. J. & Taniguchi, T. (1996). Selective cooperation of HTLV-1-encoded p40tax-1 with cellular oncoproteins in the induction of hematopoietic cell proliferation. *Oncogene* 12, 2403-8.
- Mochizuki, M., Tajima, K., Watanabe, T. & Yamaguchi, K. (1994). Human T lymphotropic virus type 1 uveitis. *Br J Ophthalmol* 78, 149-54.
- Mooslehner, K., Karls, U. & Harbers, K. (1990). Retroviral integration sites in transgenic Mov mice frequently map in the vicinity of transcribed DNA regions. *J Virol* 64, 3056-8.
- Mori, N., Murakami, S., Oda, S. & Eto, S. (1994). Human T-cell leukemia virus type I tax induces intracellular adhesion molecule-1 expression in T cells [letter]. *Blood* 84, 350-1.
- Mortreux, F., Leclercq, I., Gabet, A.S., Leroy, A., Westhof, E., Gessain, A., Wain-Hobson, S., & Wattel, E. Somatic mutations of the HTLV-1 provirus and flanking cellular sequences during clonal expansion. *Submitted*.
- Mortreux, F., Kazanji, M., Gabet, A.S., & Wattel, E. HTLV-1 replicates in squirrel monkey (*Saimiri sciureus*) as in human. *Submitted*.
- Muesing, M. A., Smith, D. H., Cabradilla, C. D., Benton, C. V., Lasky, L. A. & Capon, D. J. (1985). Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus. *Nature* 313, 450-8.
- Muller, H. P. & Varmus, H. E. (1994). DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. *Embo J* 13, 4704-14.
- Muller, B. & Krausslich, H. G. (1999). Characterization of human T-cell leukemia virus type I integrase expressed in *Escherichia coli*. *Eur J Biochem* 259, 79-87.

- Mulloy, J. C., Crownley, R. W., Fullen, J., Leonard, W. J. & Franchini, G. (1996). The human T-cell leukemia/lymphotropic virus type 1 p12I proteins bind the interleukin-2 receptor beta and gamma chains and affects their expression on the cell surface. *J Virol* 70, 3599-605.
- Mulloy, J. C., Kislyakova, T., Cereseto, A., Casareto, L., LoMonico, A., Fullen, J., Lorenzi, M. V., Cara, A., Nicot, C., Giam, C. & Franchini, G. (1998). Human T-cell lymphotropic/leukemia virus type 1 Tax abrogates p53- induced cell cycle arrest and apoptosis through its CREB/ATF functional domain. *J Virol* 72, 8852-60.
- Murphy, J. E. & Goff, S. P. (1992). A mutation at one end of Moloney murine leukemia virus DNA blocks cleavage of both ends by the viral integrase *in vivo*. *J Virol* 66, 5092-5.
- Murphy, J. E., De Los Santos, T. & Goff, S. P. (1993). Mutational analysis of the sequences at the termini of the Moloney murine leukemia virus DNA required for integration. *Virology* 195, 432-40.
- Nagai, M., Yashiki, S., Fujiyoshi, T., Fujiyama, C., Kitze, B., Izumo, S., Osame, M. & Sonoda, S. (1996). Characterization of a unique T-cell clone established from a patient with HAM/TSP which recognized HTLV-I-infected T-cell antigens as well as spinal cord tissue antigens. *J Neuroimmunol* 65, 97-105.
- Nagai, M., Usuku, K., Matsumoto, W., Kodama, D., Takenouchi, N., Moritoyo, T., Hashiguchi, S., Ichinose, M., Bangham, C. R., Izumo, S. & Osame, M. (1998). Analysis of HTLV-I proviral load in 202 HAM/TSP patients and 243 asymptomatic HTLV-I carriers: high proviral load strongly predisposes to HAM/TSP. *J Neurovirol* 4, 586-93.
- Nakamura, N., Fujii, M., Tsukahara, T., Arai, M., Ohashi, T., Wakao, H., Kannagi, M. & Yamamoto, N. (1999). Human T-cell leukemia virus type 1 Tax protein induces the expression of STAT1 and STAT5 genes in T-cells. *Oncogene* 18, 2667-75.
- Nam, S. H., Kidokoro, M., Shida, H. & Hatanaka, M. (1988). Processing of gag precursor polyprotein of human T-cell leukemia virus type I by virus-encoded protease. *J Virol* 62, 3718-28.
- Nerenberg, M., Hinrichs, S. H., Reynolds, R. K., Khoury, G. & Jay, G. (1987). The tat gene of human T-lymphotropic virus type 1 induces mesenchymal tumors in transgenic mice. *Science* 237, 1324-9.
- Neuveut, C., Low, K. G., Maldarelli, F., Schmitt, I., Majone, F., Grassmann, R. & Jeang, K. T. (1998). Human T-cell leukemia virus type 1 Tax and cell cycle progression: role of cyclin D-cdk and p110Rb. *Mol Cell Biol* 18, 3620-32.
- Newbound, G. C., Andrews, J. M., O'Rourke, J. P., Brady, J. N. & Lairmore, M. D. (1996). Human T-cell lymphotropic virus type 1 Tax mediates enhanced transcription in CD4+ T lymphocytes. *J Virol* 70, 2101-6.

- Nicot, C., Astier-Gin, T., Edouard, E., Legrand, E., Moynet, D., Vital, A., Londos-Gagliardi, D., Moreau, J. P. & Guillemain, B. (1993). Establishment of HTLV-I-infected cell lines from French, Guianese and West Indian patients and isolation of a proviral clone producing viral particles. *Virus Res* 30, 317-34.
- Niewiesk, S., Daenke, S., Parker, C. E., Taylor, G., Weber, J., Nightingale, S. & Bangham, C. R. (1994). The transactivator gene of human T-cell leukemia virus type I is more variable within and between healthy carriers than patients with tropical spastic paraparesis. *J Virol* 68, 6778-81.
- Niewiesk, S., Daenke, S., Parker, C. E., Taylor, G., Weber, J., Nightingale, S. & Bangham, C. R. (1995). Naturally occurring variants of human T-cell leukemia virus type I Tax protein impair its recognition by cytotoxic T lymphocytes and the transactivation function of Tax. *J Virol* 69, 2649-53.
- Nishioka, K., Sumida, T. & Hasunuma, T. (1996). Human T lymphotropic virus type I in arthropathy and autoimmune disorders. *Arthritis Rheum* 39, 1410-8.
- Ohshima, K., Kikuchi, M., Masuda, Y., Kobari, S., Sumiyoshi, Y., Eguchi, F., Mohtai, H., Yoshida, T., Takeshita, M. & Kimura, N. (1991). Defective provirus form of human T-cell leukemia virus type I in adult T-cell leukemia/lymphoma: clinicopathological features. *Cancer Res* 51, 4639-42.
- Osame, M., Usuku, K., Izumo, S., Ijichi, N., Amitani, H., Igata, A., Matsumoto, M. & Tara, M. (1986). HTLV-I associated myelopathy, a new clinical entity [letter]. *Lancet* 1, 1031-2.
- Oyen, T. B. & Gabrielsen, O. S. (1983). Non-random distribution of the Ty1 elements within nuclear DNA of *Saccharomyces cerevisiae*. *FEBS Lett* 161, 201-6.
- Pahl, A. & Flugel, R. M. (1993). Endonucleolytic cleavages and DNA-joining activities of the integration protein of human foamy virus. *J Virol* 67, 5426-34.
- Panganiban, A. T. & Temin, H. M. (1983). The terminal nucleotides of retrovirus DNA are required for integration but not virus production. *Nature* 306, 155-60.
- Parker, C. E., Daenke, S., Nightingale, S. & Bangham, C. R. (1992). Activated, HTLV-1-specific cytotoxic T-lymphocytes are found in healthy seropositives as well as in patients with tropical spastic paraparesis. *Virology* 188, 628-36.
- Parker, C. E., Nightingale, S., Taylor, G. P., Weber, J. & Bangham, C. R. (1994). Circulating anti-Tax cytotoxic T lymphocytes from human T-cell leukemia virus type I-infected people, with and without tropical spastic paraparesis, recognize multiple epitopes simultaneously. *J Virol* 68, 2860-8.
- Philpott, S. M. & Buehring, G. C. (1999). Defective DNA repair in cells with human T-cell leukemia/bovine leukemia viruses: role of tax gene [see comments]. *J Natl Cancer Inst* 91, 933-42.

- Pique, C., Pham, D., Tursz, T. & Dokhelar, M. C. (1992). Human T-cell leukemia virus type I envelope protein maturation process: requirements for syncytium formation. *J Virol* 66, 906-13.
- Pique, C., Ureta-Vidal, A., Gessain, A., Chancerel, B., Gout, O., Tamouza, R., Agis, F. & Dokhelar, M. C. (2000). Evidence for the Chronic *In vivo* Production of Human T Cell Leukemia Virus Type I Rof and Tof Proteins from Cytotoxic T Lymphocytes Directed against Viral Peptides. *J Exp Med* 191, 567-572.
- Pise-Masison, C. A., Choi, K. S., Radonovich, M., Dittmer, J., Kim, S. J. & Brady, J. N. (1998a). Inhibition of p53 transactivation function by the human T-cell lymphotropic virus type 1 Tax protein. *J Virol* 72, 1165-70.
- Pise-Masison, C. A., Radonovich, M., Sakaguchi, K., Appella, E. & Brady, J. N. (1998b). Phosphorylation of p53: a novel pathway for p53 inactivation in human T- cell lymphotropic virus type 1-transformed cells. *J Virol* 72, 6348-55.
- Poiesz, B. J., Ruscetti, F. W., Gazdar, A. F., Bunn, P. A., Minna, J. D. & Gallo, R. C. (1980). Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc Natl Acad Sci U S A* 77, 7415-9.
- Popov, S., Rexach, M., Zybarth, G., Reiling, N., Lee, M. A., Ratner, L., Lane, C. M., Moore, M. S., Blobel, G. & Bukrinsky, M. (1998). Viral protein R regulates nuclear import of the HIV-1 pre-integration complex. *Embo J* 17, 909-17.
- Pozzatti, R., Vogel, J. & Jay, G. (1990). The human T-lymphotropic virus type I tax gene can cooperate with the ras oncogene to induce neoplastic transformation of cells. *Mol Cell Biol* 10, 413-7.
- Pruss, D., Bushman, F. D. & Wolffe, A. P. (1994a). Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. *Proc Natl Acad Sci U S A* 91, 5913-7.
- Pruss, D., Reeves, R., Bushman, F. D. & Wolffe, A. P. (1994b). The influence of DNA and nucleosome structure on integration events directed by HIV integrase. *J Biol Chem* 269, 25031-41.
- Pryciak, P. M., Muller, H. P. & Varmus, H. E. (1992a). Simian virus 40 minichromosomes as targets for retroviral integration *in vivo*. *Proc Natl Acad Sci U S A* 89, 9237-41.
- Pryciak, P. M., Sil, A. & Varmus, H. E. (1992b). Retroviral integration into minichromosomes *in vitro*. *Embo J* 11, 291-303.
- Pryciak, P. M. & Varmus, H. E. (1992c). Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell* 69, 769-80.

- Reid, R. L., Lindholm, P. F., Mireskandari, A., Dittmer, J. & Brady, J. N. (1993). Stabilization of wild-type p53 in human T-lymphocytes transformed by HTLV-I. *Oncogene* 8, 3029-36.
- Ressler, S., Connor, L. M. & Marriott, S. J. (1996). Cellular transformation by human T-cell leukemia virus type I. *FEMS Microbiol Lett* 140, 99-109.
- Robek, M. D., Wong, F. H. & Ratner, L. (1998). Human T-cell leukemia virus type 1 pX-I and pX-II open reading frames are dispensable for the immortalization of primary lymphocytes. *J Virol* 72, 4458-62.
- Roe, T., Reynolds, T. C., Yu, G. & Brown, P. O. (1993). Integration of murine leukemia virus DNA depends on mitosis. *Embo J* 12, 2099-108.
- Rohdewohld, H., Weiher, H., Reik, W., Jaenisch, R. & Breindl, M. (1987). Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. *J Virol* 61, 336-43.
- Roth, M. J., Schwartzberg, P. L. & Goff, S. P. (1989). Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence on IN function and terminal DNA sequence. *Cell* 58, 47-54.
- Roth, M. J., Schwartzberg, P., Tanese, N. & Goff, S. P. (1990). Analysis of mutations in the integration function of Moloney murine leukemia virus: effects on DNA binding and cutting. *J Virol* 64, 4709-17.
- Saito, M., Furukawa, Y., Kubota, R., Usuku, K., Sonoda, S., Izumo, S., Osame, M. & Yoshida, M. (1995). Frequent mutation in pX region of HTLV-1 is observed in HAM/TSP patients, but is not specifically associated with the central nervous system lesions. *J Neurovirol* 1, 286-94.
- Sakashita, A., Hattori, T., Miller, C. W., Suzushima, H., Asou, N., Takatsuki, K. & Koeffler, H. P. (1992). Mutations of the p53 gene in adult T-cell leukemia. *Blood* 79, 477-80.
- Sanada, I., Ishii, T., Matsuoka, M., Kumagai, E., Nishimura, H., Yamaguchi, K. & Takatsuki, K. (1987). Chromosomal abnormalities in non-Hodgkin lymphoma with peripheral T-cell type: effect of HTLV-I infection. *Hematol Oncol* 5, 157-66.
- Sandmeyer, S. B., Hansen, L. J. & Chalker, D. L. (1990). Integration specificity of retrotransposons and retroviruses. *Annu Rev Genet* 24, 491-518.
- Santiago, F., Clark, E., Chong, S., Molina, C., Mozafari, F., Mahieux, R., Fujii, M., Azimi, N. & Kashanchi, F. (1999). Transcriptional up-regulation of the cyclin D2 gene and acquisition of new cyclin-dependent kinase partners in human T-cell leukemia virus type 1-infected cells. *J Virol* 73, 9917-27.

- Sato, K., Maruyama, I., Maruyama, Y., Kitajima, I., Nakajima, Y., Higaki, M., Yamamoto, K., Miyasaka, N., Osame, M. & Nishioka, K. (1991). Arthritis in patients infected with human T lymphotropic virus type I. Clinical and immunopathologic features. *Arthritis Rheum* 34, 714-21.
- Schauer, M. & Billich, A. (1992). The N-terminal region of HIV-1 integrase is required for integration activity, but not for DNA-binding. *Biochem Biophys Res Commun* 185, 874-80.
- Scherdin, U., Rhodes, K. & Breindl, M. (1990). Transcriptionally active genome regions are preferred targets for retrovirus integration. *J Virol* 64, 907-12.
- Schmidtayerova, H., Nuovo, G. J. & Bukrinsky, M. (1997). Cell proliferation is not required for productive HIV-1 infection of macrophages. *Virology* 232, 379-84.
- Schmitt, I., Rosin, O., Rohwer, P., Gossen, M. & Grassmann, R. (1998). Stimulation of cyclin-dependent kinase activity and G1- to S-phase transition in human lymphocytes by the human T-cell leukemia/lymphotropic virus type 1 Tax protein. *J Virol* 72, 633-40.
- Scottoline, B. P., Chow, S., Ellison, V. & Brown, P. O. (1997). Disruption of the terminal base pairs of retroviral DNA during integration. *Genes Dev* 11, 371-82.
- Seiki, M., Hattori, S., Hirayama, Y. & Yoshida, M. (1983). Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. *Proc Natl Acad Sci U S A* 80, 3618-22.
- Seiki, M., Eddy, R., Shows, T. B. & Yoshida, M. (1984). Nonspecific integration of the HTLV provirus genome into adult T-cell leukaemia cells. *Nature* 309, 640-2.
- Semmes, O. J. & Jeang, K. T. (1996). Localization of human T-cell leukemia virus type 1 tax to subnuclear compartments that overlap with interchromatin speckles. *J Virol* 70, 6347-57.
- Sherman, P. A., Dickson, M. L. & Fyfe, J. A. (1992). Human immunodeficiency virus type 1 integration protein: DNA sequence requirements for cleaving and joining reactions. *J Virol* 66, 3593-601.
- Sherman, P. A. & Fyfe, J. A. (1990). Human immunodeficiency virus integration protein expressed in *Escherichia coli* possesses selective DNA cleaving activity. *Proc Natl Acad Sci U S A* 87, 5119-23.
- Shibagaki, Y. & Chow, S. A. (1997). Central core domain of retroviral integrase is responsible for target site selection. *J Biol Chem* 272, 8361-9.
- Shih, C. C., Stoye, J. P. & Coffin, J. M. (1988). Highly preferred targets for retrovirus integration. *Cell* 53, 531-7.

- Shimotohno, K. & Temin, H. M. (1980). No apparent nucleotide sequence specificity in cellular DNA juxtaposed to retrovirus proviruses. *Proc Natl Acad Sci U S A* 77, 7357-61.
- Shimoyama, M. (1991). Diagnostic criteria and classification of clinical subtypes of adult T-cell leukaemia-lymphoma. A report from the Lymphoma Study Group (1984- 87). *Br J Haematol* 79, 428-37.
- Shin, C. G., Taddeo, B., Haseltine, W. A. & Farnet, C. M. (1994). Genetic analysis of the human immunodeficiency virus type 1 integrase protein. *J Virol* 68, 1633-42.
- Shoemaker, C., Goff, S., Gilboa, E., Paskind, M., Mitra, S. W. & Baltimore, D. (1980). Structure of a cloned circular Moloney murine leukemia virus DNA molecule containing an inverted segment: implications for retrovirus integration. *Proc Natl Acad Sci U S A* 77, 3932-6.
- Sodroski, J. G., Rosen, C. A. & Haseltine, W. A. (1984). Trans-acting transcriptional activation of the long terminal repeat of human T lymphotropic viruses in infected cells. *Science* 225, 381-5.
- Sonoda, S., Fujiyoshi, T. & Yashiki, S. (1996). Immunogenetics of HTLV-I/II and associated diseases. *J Acquir Immune Defic Syndr Hum Retrovirol* 13, S119-23.
- Stevens, S. W. & Griffith, J. D. (1994). Human immunodeficiency virus type 1 may preferentially integrate into chromatin occupied by L1Hs repetitive elements. *Proc Natl Acad Sci U S A* 91, 5557-61.
- Stevens, S. W. & Griffith, J. D. (1996). Sequence analysis of the human DNA flanking sites of human immunodeficiency virus type 1 integration. *J Virol* 70, 6459-62.
- Suzuki, T., Kitao, S., Matsushime, H. & Yoshida, M. (1996). HTLV-1 Tax protein interacts with cyclin-dependent kinase inhibitor p16INK4A and counteracts its inhibitory activity towards CDK4. *Embo J* 15, 1607-14.
- Swartz, M.N., Trautner, T.A., & Kornberg, A. (1962). Enzymatic synthesis of deoxyribonucleic acid. XI. Further studies on nearest neighbor base sequences in deoxyribonucleic acid. *J Biol Chem* 237, 1961-1967.
- Takemoto, S., Mulloy, J. C., Cereseto, A., Migone, T. S., Patel, B. K., Matsuoka, M., Yamaguchi, K., Takatsuki, K., Kamihira, S., White, J. D., Leonard, W. J., Waldmann, T. & Franchini, G. (1997). Proliferation of adult T cell leukemia/lymphoma cells is associated with the constitutive activation of JAK/STAT proteins. *Proc Natl Acad Sci U S A* 94, 13897-902.
- Tamiya, S., Matsuoka, M., Etoh, K., Watanabe, T., Kamihira, S., Yamaguchi, K. & Takatsuki, K. (1996). Two types of defective human T-lymphotropic virus type I provirus in adult T-cell leukemia. *Blood* 88, 3065-73.

- Tanaka, A., Takahashi, C., Yamaoka, S., Nosaka, T., Maki, M. & Hatanaka, M. (1990). Oncogenic transformation by the tax gene of human T-cell leukemia virus type I *in vitro*. *Proc Natl Acad Sci U S A* 87, 1071-5.
- Tanaka, Y., Fukudome, K., Hayashi, M., Takagi, S. & Yoshie, O. (1995). Induction of ICAM-1 and LFA-3 by Tax1 of human T-cell leukemia virus type 1 and mechanism of down-regulation of ICAM-1 or LFA-1 in adult-T- cell-leukemia cell lines. *Int J Cancer* 60, 554-61.
- Tanaka, Y., Hayashi, M., Takagi, S. & Yoshie, O. (1996). Differential transactivation of the intercellular adhesion molecule 1 gene promoter by Tax1 and Tax2 of human T-cell leukemia viruses. *J Virol* 70, 8508-17.
- Tanimura, A., Dan, S. & Yoshida, M. (1998). Cloning of novel isoforms of the human Gli2 oncogene and their activities to enhance tax-dependent transcription of the human T-cell leukemia virus type 1 genome. *J Virol* 72, 3958-64.
- Tatewaki, M., Yamaguchi, K., Matsuoka, M., Ishii, T., Miyasaka, M., Mori, S., Takatsuki, K. & Watanabe, T. (1995). Constitutive overexpression of the L-selectin gene in fresh leukemic cells of adult T-cell leukemia that can be transactivated by human T- cell lymphotropic virus type 1 Tax. *Blood* 86, 3109-17.
- Tobinai, K. & Shimoyama, M. (1994). [Recent progress in the treatment of malignant lymphoma]. *Gan To Kagaku Ryoho* 21, 1151-6.
- Tsuchiya, H., Fujii, M., Niki, T., Tokuhara, M., Matsui, M. & Seiki, M. (1993). Human T-cell leukemia virus type 1 Tax activates transcription of the human fra-1 gene through multiple cis elements responsive to transmembrane signals. *J Virol* 67, 7001-7.
- Tsukasaki, K., Tsushima, H., Yamamura, M., Hata, T., Murata, K., Maeda, T., Atogami, S., Sohda, H., Momita, S., Ideda, S., Katamine, S., Yamada, Y., Kamihira, S. & Tomonaga, M. (1997). Integration patterns of HTLV-I provirus in relation to the clinical course of ATL: frequent clonal change at crisis from indolent disease. *Blood* 89, 948-56.
- Tsurumi, H., Tani, K., Tsuruta, T., Shirato, R., Matsudaira, T., Tojo, A., Wada, C., Uchida, H., Ozawa, K. & Asano, S. (1992). Adult T-cell leukemia developing during immunosuppressive treatment in a renal transplant recipient. *Am J Hematol* 41, 292-4.
- Uhlik, M., Good, L., Xiao, G., Harhaj, E. W., Zandi, E., Karin, M. & Sun, S. C. (1998). NF-kappaB-inducing kinase and IkappaB kinase participate in human T- cell leukemia virus I Tax-mediated NF-kappaB activation. *J Biol Chem* 273, 21132-6.
- Uittenbogaard, M. N., Armstrong, A. P., Chiaramello, A. & Nyborg, J. K. (1994). Human T-cell leukemia virus type I Tax protein represses gene expression through the basic helix-loop-helix family of transcription factors. *J Biol Chem* 269, 22466-9.

- Uittenbogaard, M. N., Giebler, H. A., Reisman, D. & Nyborg, J. K. (1995). Transcriptional repression of p53 by human T-cell leukemia virus type I Tax protein. *J Biol Chem* 270, 28503-6.
- Valenzuela, P., Bell, G. I., Masiarz, F. R., DeGennaro, L. J. & Rutter, W. J. (1977). Nucleotide sequence of the yeast 5S ribosomal RNA gene and adjacent putative control regions. *Nature* 267, 641-3.
- van den Ent, F. M., Vink, C. & Plasterk, R. H. (1994). DNA substrate requirements for different activities of the human immunodeficiency virus type 1 integrase protein. *J Virol* 68, 7825-32.
- van Gent, D. C., Elgersma, Y., Bolk, M. W., Vink, C. & Plasterk, R. H. (1991). DNA binding properties of the integrase proteins of human immunodeficiency viruses types 1 and 2. *Nucleic Acids Res* 19, 3821-7.
- van Gent, D. C., Groeneger, A. A. & Plasterk, R. H. (1992). Mutational analysis of the integrase protein of human immunodeficiency virus type 2. *Proc Natl Acad Sci U S A* 89, 9598-602.
- van Gent, D. C., Vink, C., Groeneger, A. A. & Plasterk, R. H. (1993). Complementation between HIV integrase proteins mutated in different domains. *Embo J* 12, 3261-7.
- Vijaya, S., Steffen, D. L. & Robinson, H. L. (1986). Acceptor sites for retroviral integrations map near DNase I- hypersensitive sites in chromatin. *J Virol* 60, 683-92.
- Vincent, K. A., York-Higgins, D., Quiroga, M. & Brown, P. O. (1990). Host sequences flanking the HIV provirus. *Nucleic Acids Res* 18, 6045-7.
- Vincent, K. A., Ellison, V., Chow, S. A. & Brown, P. O. (1993). Characterization of human immunodeficiency virus type 1 integrase expressed in *Escherichia coli* and analysis of variants with amino- terminal mutations. *J Virol* 67, 425-37.
- Vink, C., van Gent, D. C., Elgersma, Y. & Plasterk, R. H. (1991). Human immunodeficiency virus integrase protein requires a subterminal position of its viral DNA recognition sequence for efficient cleavage. *J Virol* 65, 4636-44.
- Vink, C., Oude Groeneger, A. M. & Plasterk, R. H. (1993). Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type I integrase protein. *Nucleic Acids Res* 21, 1419-25.
- Vink, C., Lutzke, R. A. & Plasterk, R. H. (1994). Formation of a stable complex between the human immunodeficiency virus integrase protein and viral DNA. *Nucleic Acids Res* 22, 4103-10.
- Vissel, B., Nagy, A. & Choo, K. H. (1992). A satellite III sequence shared by human chromosomes 13, 14, and 21 that is contiguous with alpha satellite DNA. *Cytogenet Cell Genet* 61, 81-6.

- Wagner, S. & Green, M. R. (1993). HTLV-I Tax protein stimulation of DNA binding of bZIP proteins by enhancing dimerization. *Science* 262, 395-9.
- Walsh, P. S., Fildes, N. J. & Reynolds, R. (1996). Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. *Nucleic Acids Res* 24, 2807-12.
- Wattel, E., Mariotti, M., Agis, F., Gordien, E., Le Coeur, F. F., Prin, L., Rouger, P., Chen, I. S., Wain-Hobson, S. & Lefrere, J. J. (1992a). Quantification of HTLV-1 proviral copy number in peripheral blood of symptomless carriers from the French West Indies. *J Acquir Immune Defic Syndr* 5, 943-6.
- Wattel, E., Mariotti, M., Agis, F., Gordien, E., Prou, O., Courouce, A. M., Rouger, P., Wain-Hobson, S., Chen, I. S. & Lefrere, J. J. (1992b). Human T lymphotropic virus (HTLV) type I and II DNA amplification in HTLV-I/II-seropositive blood donors of the French West Indies. *J Infect Dis* 165, 369-72.
- Wattel, E., Vartanian, J. P., Pannetier, C. & Wain-Hobson, S. (1995). Clonal expansion of human T-cell leukemia virus type I-infected cells in asymptomatic and symptomatic carriers without malignancy. *J Virol* 69, 2863-8.
- Wattel, E., Cavrois, M., Gessain, A. & Wain-Hobson, S. (1996). Clonal expansion of infected cells: a way of life for HTLV-I. *J Acquir Immune Defic Syndr Hum Retrovirol* 13, S92-9.
- Wei, S. Q., Mizuuchi, K. & Craigie, R. (1997). A large nucleoprotein assembly at the ends of the viral DNA mediates retroviral DNA integration. *Embo J* 16, 7511-20.
- Weinberg, J. B., Matthews, T. J., Cullen, B. R. & Malim, M. H. (1991). Productive human immunodeficiency virus type 1 (HIV-1) infection of nonproliferating human monocytes. *J Exp Med* 174, 1477-82.
- Withers-Ward, E. S., Kitamura, Y., Barnes, J. P. & Coffin, J. M. (1994). Distribution of targets for avian retrovirus DNA integration *in vivo*. *Genes Dev* 8, 1473-87.
- Wolfe, A. L., Felock, P. J., Hastings, J. C., Blau, C. U. & Hazuda, D. J. (1996). The role of manganese in promoting multimerization and assembly of human immunodeficiency virus type 1 integrase as a catalytically active complex on immobilized long terminal repeat substrates. *J Virol* 70, 1424-32.
- Yamato, K., Oka, T., Hiroi, M., Iwahara, Y., Sugito, S., Tsuchida, N. & Miyoshi, I. (1993). Aberrant expression of the p53 tumor suppressor gene in adult T-cell leukemia and HTLV-I-infected cells. *Jpn J Cancer Res* 84, 4-8.
- Yang, F., Leon, O., Greenfield, N. J. & Roth, M. J. (1999). Functional interactions of the HHCC domain of moloney murine leukemia virus integrase revealed by nonoverlapping complementation and zinc- dependent dimerization. *J Virol* 73, 1809-17.

- Yin, M. J., Christerson, L. B., Yamamoto, Y., Kwak, Y. T., Xu, S., Mercurio, F., Barbosa, M., Cobb, M. H. & Gaynor, R. B. (1998). HTLV-I Tax protein binds to MEKK1 to stimulate IkappaB kinase activity and NF-kappaB activation. *Cell* 93, 875-84.
- Yoshida, M., Miyoshi, I. & Hinuma, Y. (1982). Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. *Proc Natl Acad Sci U S A* 79, 2031-5.
- Yoshida, M., Seiki, M., Yamaguchi, K. & Takatsuki, K. (1984). Monoclonal integration of human T-cell leukemia provirus in all primary tumors of adult T-cell leukemia suggests causative role of human T-cell leukemia virus in the disease. *Proc Natl Acad Sci U S A* 81, 2534-7.
- Zanke, B. W., Rush, D. N., Jeffery, J. R. & Israels, L. G. (1989). HTLV-1 T cell lymphoma in a cyclosporine-treated renal transplant patient. *Transplantation* 48, 695-7.
- Zhao, L. J. & Giam, C. Z. (1992). Human T-cell lymphotropic virus type I (HTLV-I) transcriptional activator, Tax, enhances CREB binding to HTLV-I 21-base-pair repeats by protein-protein interaction. *Proc Natl Acad Sci U S A* 89, 7070-4.
- Zheng, R., Jenkins, T. M. & Craigie, R. (1996). Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc Natl Acad Sci U S A* 93, 13659-64.
- Zhou, W., Parent, L. J., Wills, J. W. & Resh, M. D. (1994). Identification of a membrane-binding domain within the amino-terminal region of human immunodeficiency virus type 1 Gag protein which interacts with acidic phospholipids. *J Virol* 68, 2556-69.
- Zoubak, S., Richardson, J. H., Rynditch, A., Hollsberg, P., Hafler, D. A., Boeri, E., Lever, A. M. & Bernardi, G. (1994). Regional specificity of HTLV-I proviral integration in the human genome. *Gene* 143, 155-63.

Host Sequences Flanking the Human T-Cell Leukemia Virus Type 1 Provirus In Vivo

INDIA LECLERCQ,¹ FRANCK MORTREUX,¹ MARIELLE CAVROIS,^{1*} ARNAUD LEROY,²
ANTOINE GESSAIN,³ SIMON WAIN-HOBSON,⁴ AND ERIC WATTEL^{5,6*}

Unité 524 INSERM, Institut de Recherche sur le Cancer de Lille,¹ and Unité d'Oncogénèse Virale, Centre Oscar Lambret,² Lille, Unité d'Epidémiologie des Virus Oncogènes,³ and Unité de Rétrovirologie Moléculaire,⁴ Institut Pasteur, Paris, and Unité d'Oncogénèse Virale, UMR5537 CNRS-Université Claude Bernard, Centre Léon Bérard, Lyon,⁵ and Service des Maladies du Sang, CHU 59037 Lille,⁶ France

Received 26 March 1999/Accepted 23 November 1999

Human pathogenic retroviruses do not have common loci of integration. However, many factors, such as chromatin structure, transcriptional activity, DNA-protein interaction, CpG methylation, and nucleotide composition of the target sequence, may influence integration site selection. These features have been investigated by *in vitro* integration reactions or by infection of cell lines with recombinant retroviruses. Less is known about target choice for integration *in vivo*. The present study was conducted in order to assess the characteristics of cellular sequences targeted for human T-cell leukemia virus type 1 (HTLV-1) integration *in vivo*. Sequencing integration sites from ≥ 200 proviruses (19 kb of sequence) isolated from 29 infected individuals revealed that HTLV-1 integration is not random at the level of the nucleotide sequence. The virus was found to integrate in A/T-rich regions with a weak consensus sequence at positions within and without of the hexameric repeat generated during integration. These features were not associated with a preference for integration near active regions or repeat elements of the host chromosomes. Most or all of the regions of the genome appear to be accessible to HTLV-1 integration. As with integration *in vitro*, integration specificity *in vivo* seems to be determined by local features rather than by the accessibility of specific regions.

In the course of retrovirus infection, a DNA copy of the viral RNA genome is synthesized, and that DNA is then permanently inserted into the genome of the host cell. The integration of the viral DNA into the host genome is a crucial step in the life cycle: it is important for the efficient expression of progeny virus, and it is responsible for the ability of these viruses to persist and cause disease (36). If most or all of the regions of the host genome are accessible to retroviral integration (14, 36, 46), the frequency of use of potential sites varies considerably. This integration specificity correlates with local DNA structural features, which govern the accessibility of specific regions (37, 50).

Indeed, integration may be favored near DNase I-hypersensitive sites (32, 33) or CpG islands (23), suggesting a preference for transcriptionally active regions (26, 37). Human immunodeficiency virus type 1 (HIV-1) integration preferentially occurs near Alu elements (41) or topoisomerase cleavage sites (19). Recently, centromeric alphoid repeats were found to be selectively absent at HIV-1 integration sites (6). A nonrandom, compartmentalized integration in GC-rich isochores has been previously described for human T-cell leukemia virus type 1 (HTLV-1) (52), bovine leukemia virus (22), hepatitis B virus (51), and Rous sarcoma virus (34), while mouse mammary tumor virus has been found to integrate into GC-poor regions (35) of the host genome. At the nucleotide level, *in vitro* integration reactions and analysis of the flanking sequence of cloned viruses have shown that there is a preference for integration in A/T-rich regions (15, 17, 27, 39). A consensus sequence at the direct repeat flanking HIV-1 provirus genome

has been shown by sequencing integration sites derived from cells infected *in vitro* (6) (41, 47).

The data summarized above result from *in vitro* studies, and only a few flanking sequences derived from naturally infected samples have been described to date. HTLV-1 is the causative agent of adult T-cell leukemia/lymphoma (ATLL), an aggressive T-cell malignancy and of tropical spastic paraparesis/HTLV-1 associated myelopathy (TSP/HAM), a chronic progressive neuromyelopathy. A study based on the analysis of four HTLV-1 integration sites derived from uncultured ATLL cells has suggested that the regions flanking the provirus were A/T-rich with a nucleotide composition bias in the 6-bp direct repeat generated by integration (13). In a previous work based on the analysis of 24 distinct HTLV-1 integration sites derived from asymptomatic carriers, the mean A/T content of the hexameric repeat was found to be 59% (49).

An inventory of 218 distinct HTLV-1 integration sites was made from infected individuals. The virus was found to integrate in A/T-rich regions and a weakly conserved sequence was identified at *in vivo* integration sites. Database analysis revealed that these features were not associated with any preference for integration in transcriptionally active regions or in repeat elements of the host genome.

MATERIALS AND METHODS

Samples studied. Samples from 29 HTLV-1-infected individuals were analyzed. These samples corresponded to 12 ATLL, 10 TSP/HAM, and 7 asymptomatic carriers. DNA was phenol-chloroform extracted and ethanol precipitated from peripheral blood mononuclear cells (PBMCs), cerebrospinal fluid (CSF), lymph nodes, or skin biopsies.

PCR. Our strategy is summarized in Fig. 1. DNA was amplified as described by ligation-mediated PCR (LMPCR) or inverse PCR (IPCR) (7–10, 49). Both of these two methods allow the amplification of the HTLV-1 3' extremities, together with their cellular flanking sequences. For LMPCR, DNA was digested with *Nla*III in 1× *Nla*III buffer for 3 h at 37°C. Digestion was controlled by gel electrophoresis. DNA was phenol-chloroform extracted and ethanol precipitated. Digested DNA was ligated with BIO1 primer (49) by using T4 DNA ligase.

* Corresponding author. Mailing address: Unité d'Oncogénèse Virale, UMR5537-CNRS-Université Claude Bernard, Centre Léon-Bérard, 28, rue Laënnec, 69373 Lyon Cedex 08, France. Phone: 334-78-78-26-69. Fax: 334-78-78-27-17. E-mail: ewattel@easynet.fr.

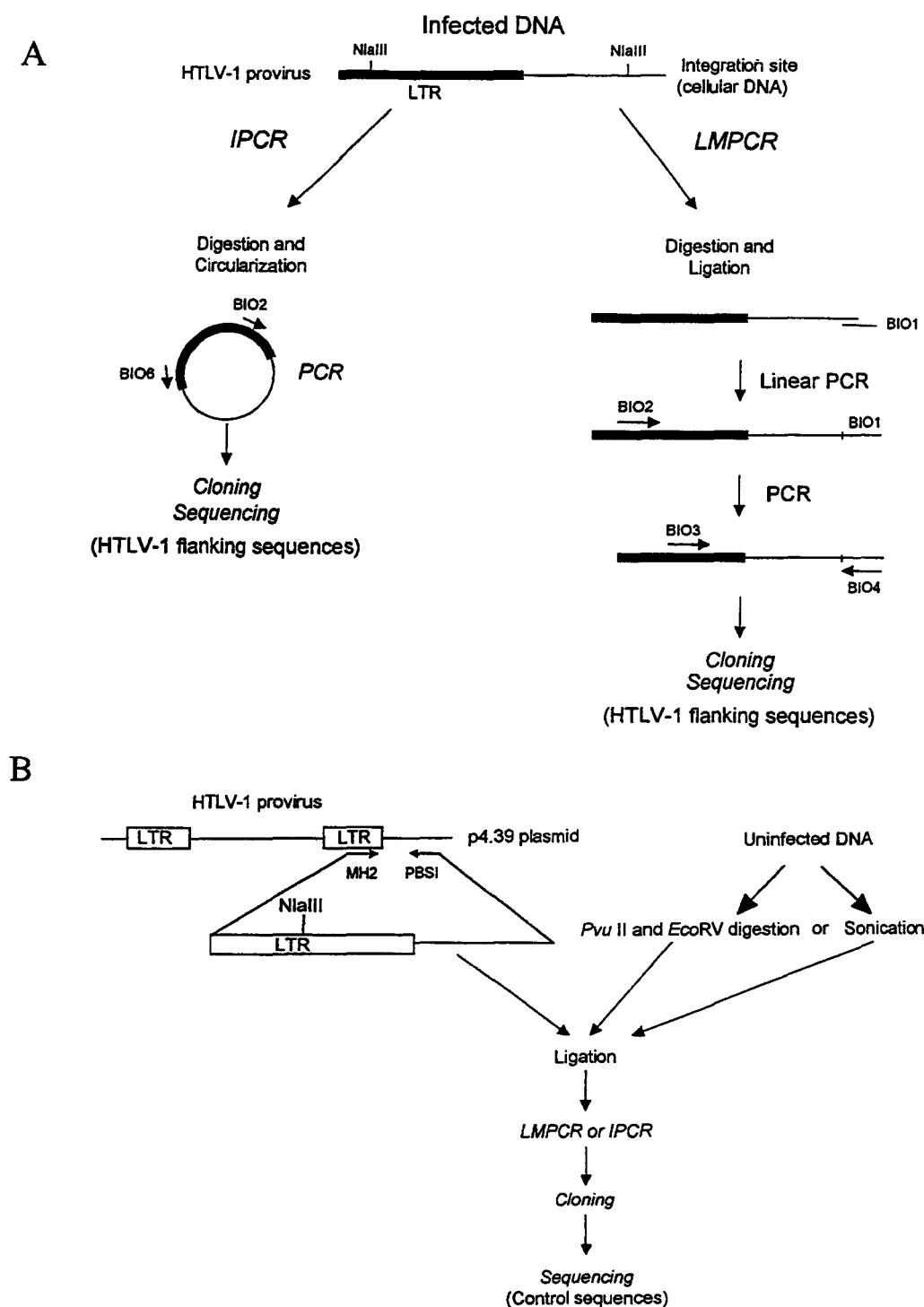


FIG. 1. Strategy to construct HTLV-1 flanking sequences and control sequences libraries. (A) Construction of HTLV-1 integration sites library. HTLV-1-infected DNA was amplified by LMPCR or IPCR. Amplified products were cloned and sequenced as detailed in Materials and Methods. (B) A ~1,500-bp segment spanning the 3' extremity of the provirus was generated by PCR amplification of a HTLV-1 provirus cloned in p4.39 plasmid (10). High-molecular-weight uninfected human DNA was first sonicated or digested with *Pvu*II and *Eco*RV which cut blunt ends. Sonicated or digested DNA was then ligated with the 3' HTLV-1 purified construct as detailed in Materials and Methods. Then, LMPCR, IPCR, cloning, and sequencing were performed in the same conditions as with the infected samples.

This was followed by two phenol-chloroform extractions and precipitation. Ligated DNA was amplified for 100 cycles with the BIO2 primer alone (49). Conditions were 1× Stoffel DNA polymerase buffer, 1.5 mM MgCl₂, 50 pmol BIO2, a 150 μM concentration of each deoxynucleoside triphosphate (dNTP), and 10 U of Stoffel fragment of *Taq* DNA polymerase (Perkin-Elmer Cetus) in a final volume of 85 μl. First, 25 μl of a 1× PCR buffer containing dNTPs and

primers were first loaded into a 750-μl tube, and an AmpliwaX PCR Gem 100 (Cetus) was added to each tube. After wax layer formation by incubation at 75°C 10 min and cooling at room temperature for 15 min, 60 μl of the remaining reagent and ligated products were loaded. Thermal cycling parameters were as follows: 1 cycle of 94°C for 10 min and 100 cycles of 95°C for 45 s, 60°C for 45 s, and 72°C for 2 min, followed by a final elongation step of 10 min at 72°C. Ten

microliters of this linear PCR reaction was used in a classical PCR amplification with the BIO3 and BIO4 primer pair (49). Amplification conditions were as before, with 40 pmol of each primer and 2.5 U of *Taq* polymerase all in a final volume of 100 μ l. Thermal cycling parameters were as follows: 1 cycle of 94°C for 10 min and 35 cycles of 95°C for 45 s, 58°C for 45 s, and 72°C for 1 min, followed by a final elongation step of 10 min at 72°C.

For inverse PCR, DNA was first digested by *Nla*III as in LMPCR. Digested DNA was circularized with T4 DNA ligase in 600 μ l for 16 h at 14°C. This was followed by two phenol-chloroform extractions and precipitation. Circularized DNA was amplified for 40 cycles by using the BIO2 and BIO6 primer pair (10). Amplification conditions and thermal cycling parameter were as follows: 1 cycle of 95°C for 10 min and 40 cycles of 95°C for 45 s, 58°C for 45 s, and 72°C for 1 min, followed by a final elongation step of 10 min at 72°C.

The IPCR and LMPCR thresholds are 100 and 20 copies/sample, respectively (10). Since high proviral loads are invariably associated with ATLL and TSP/HAM, corresponding PBMC samples were analyzed by IPCR, while PBMC samples from asymptomatic carriers and CSF samples from TSP/HAM patients were analyzed by LMPCR. Two to four samples of noninfected DNA served as negative controls in all experiments.

Cloning and sequencing. PCR products were phosphorylated by T4 polynucleotide kinase and ligated with *Sma*I-digested and dephosphorylated M13mp18 replicative-form DNA as described earlier (49). After transformation of *Escherichia coli* XL1 by electroporation, recombinant M13 plaques were transferred in situ to nitrocellulose filters and screened by hybridization with the HTLV-1 long terminal repeat (LTR)-specific ³²P-labeled oligonucleotide BIO5 (49). Filters were first prehybridized at 42°C for 2 h in 5 $\times$ SSC (1 $\times$ SSC is 0.15 M NaCl plus 0.015 M sodium citrate)-1 $\times$ Denhardt's solution-10 μ g of denatured salmon sperm DNA per ml. Hybridization was carried out in a fresh solution at 42°C, and filters were washed in 2 $\times$ SSC-0.1% sodium dodecyl sulfate at 42°C. Positive plaques were picked and prepared for DNA sequencing. Single-stranded templates were sequenced by using fluorescent dideoxynucleotides. The products were resolved on an Applied Biosystems 377A DNA sequencer with 377 software.

Control experiments. A control was used in order to compare the nucleotide composition of HTLV-1 flanking sequences with that of the uninfected DNA. As shown in Fig. 1B, a 1,494-bp segment spanning the 3' extremity of the provirus was first generated by PCR. This was performed by using an integration site-specific primer, PBSI (10), and an LTR-specific primer, MH2 (5'-CCCGCCAA TCATCATACAACC-3') that encompassed the 3' extremity of the HTLV-1 provirus cloned in the plasmid p4.39 (10). Amplification conditions and thermal cycling parameters were as follows: 1 cycle of 95°C for 10 min and 35 cycles of 95°C for 1 min, 58°C for 45 s, and 72°C for 2 min, followed by a final elongation step of 10 min at 72°C. Amplified products were purified and treated with the Klenow fragment of the DNA polymerase. As shown in Fig. 1, high-molecular-weight uninfected human DNA that derived from an uninfected blood donor was blunt-end digested with *Pvu*II and *Eco*RV or sonicated. These three different methods were used in order to avoid any bias in the selection of DNA sequence during fragmentation. Digested or sonicated DNA was then ligated with the purified 3' HTLV-1 end. Then, LMPCR, IPCR, cloning, and sequencing were performed under the same conditions as with samples that were derived from infected individuals.

Comparison of integration sites with database sequences. Sequences with integration sites longer than 25 bp were analyzed by comparison to the nonredundant human sequence (nr) database (2), the human cDNA (dbEST) database (3), and the MONTH (January 1999) database (42) by using BLASTN with Search Launcher (1), FASTA (28), and Repeat Masker (See A. F. A. Smit and P. Green, RepeatMasker, at <http://ftp.genome.washington.edu/cgi-bin/RepeatMasker>) (6). Default parameters were used. The 25-bp cutoff was arbitrary chosen in order to avoid nonspecific homologies that could result from the alignment of shorter sequences with those of databases. A total of 18,118 bp of human DNA corresponding to 178 integration sites that were $\geq$ 25 bp were analyzed. For the 44 control sequences of $\geq$ 25 bp, 2,991 bp were analyzed. These 21,109 bp corresponded to 96% of the 22,018 bp we have sequenced. The lengths of flanking human DNA sequences analyzed ranged from 25 to 350 bp. The lengths of control sequences analyzed ranged from 25 to 150 bp. Similarities to repeated sequences were ranked in accordance with the Smith-Waterman parameter generated by RepeatMasker or by the probability of matching by chance generated by BLASTN (1) and by FASTA (29).

RESULTS

Construction of control sequences and HTLV-1 integration sites libraries. The goal of the study was to analyze HTLV-1 integration sites isolated by LMPCR or IPCR. Since PCR and cloning are influenced by target sequence composition and sequence size, a control was performed in order to assess the nucleotide composition of the uninfected human genome by using the same experimental procedure as that used in the analysis of HTLV-1 integration sites (see Materials and Meth-

ods and Fig. 1). This was done by sequencing flanking sequences of an HTLV-1 construct ligated in uninfected DNA. The three different methods-used in the control DNA fragmentation process (*Pvu*II digestion, *Eco*RV digestion, and sonication) ruled out any bias in the nucleotide composition of the control sequences. HTLV-1 flanking sequences isolated from infected individuals were subsequently compared to that of the control library.

Sixty distinct control molecular clones were sequenced. The mean sequence size was 53 bp, ranging from 12 to 150 bp (median, 37 bp). A total of 218 distinct HTLV-1 integration sites were sequenced. The mean sequence size was 84 bp (median, 69 bp; range, 6 to 365 bp). Totals of 75, 116, and 27 sequences were derived from asymptomatic carriers, TSP/HAM, and ATLL, respectively. We obtained 1 to 44 distinct integration sites (mean, 8) from each infected individual.

LMPCR products were found to be shorter than IPCR products. In addition, a weak correlation was observed between the nucleotide content and the length of flanking sequences, the longest being the more A/T-rich. Based on these results, all sequence comparisons were adjusted for size.

HTLV-1 flanking sequences are A/T-rich. The A/T content was 51% (median, 50%; standard deviation [SD], 11%; range, 30 to 80%) after control experiments, a value significantly lower than that initially described for the human DNA (58%) (43). DNA digestion, amplification, and cloning may have accounted for the selection of such sequences. Based on these results, we performed a comparative analysis of HTLV-1 integration sites and control sequences.

For the control sequences, the percentages of A, C, G, and T nucleotides were 25, 26, 23, and 26%, respectively. The overall A/T content of HTLV-1 integration sites was 57% (median, 57%; SD, 9%; range, 33 to 79%), a value significantly higher than that of control sequences ($P = 10^{-4}$, independent samples *t* test). The percentages of A, C, G, and T nucleotides were 29, 22, 21, and 28%, respectively. Only the A ($P = 0.001$) and C ($P = 0.001$) contents were found to be significantly different between HTLV-1 and control sequences. All of these differences remained significant when HTLV-1 and control sequences were adjusted for size. The A/T content of the hexameric repeat was 56% (median, 50%; SD, 20%; range, 17 to 100%). The percentages of A, C, G, and T nucleotides of the HTLV-1 flanking hexameric repeats were 30, 24, 20, and 26%, respectively. The T content without the hexameric repeats were 30, 24, 20, and 26%, respectively. The T content without the hexameric repeat was significantly higher than that within, i.e., 28 versus 25% ($P = 0.039$, paired samples *t* test). There was no significant difference in the distribution of the remaining nucleotides within and without the hexameric repeat. The A/T contents of flanking sequences isolated from ATLL, TSP/HAM, and carriers were 57, 59, and 55%, which were not significantly different values.

Nonhomogenous distribution of the nucleotide composition within and without the hexameric repeat. Figure 2 represents the distribution of A/T nucleotides along the 40 first positions of both the HTLV-1 integration sites and control sequences. Deletion of the 3' LTR sequence was performed prior to sequence analysis. In addition, deletion of the first 31 bases of the p4.39 integration site, together with the 3 bases that correspond to *Pvu*II and *Eco*RV restriction sites, was also performed when corresponding control sequences were studied. For the box plot analysis, points more than 1.5 times the interquartile range from the ends of the box were labeled as outliers. Points more than 3 times the interquartile range from the ends of the box were labeled as extreme values (see legend of Fig. 2).

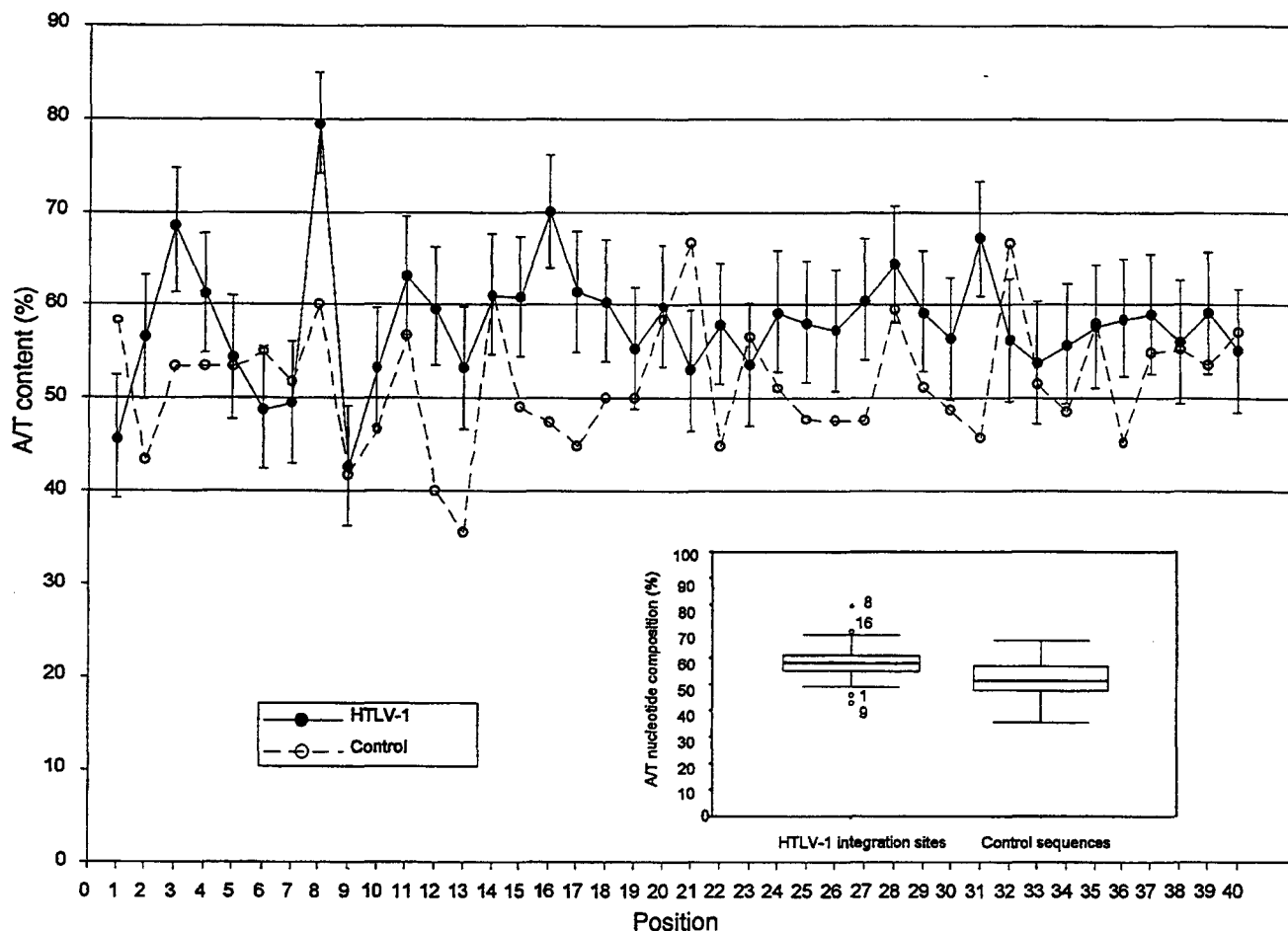


FIG. 2. Distribution of A/T nucleotides at each of the first 40 positions of the flanking sequences. The 218 HTLV-1 integration sites (●) and the 60 control sequences (○) were aligned, as described in the text, relative to the first base of the flanking sequences 3' to the provirus. The standard deviation of the A/T content is shown at each position of the HTLV-1 flanking sequences. The insert shows the distribution of the AT content along the flanking sequences. The boxes represent the difference between the 75th percentile and the 25th percentile of the AT distribution (i.e., the interquartile range). Within the inset box, the median is represented by a thick horizontal line. Lines from the ends of the box extend as far as the most extreme values not considered outliers. Points more than 1.5 times the interquartile range from the ends of the box are labeled as outliers (○) or as extreme values (*).

Figure 2 shows that the A/T content of HTLV-1 integration sites was higher than that of control sequences in 31 of 40 positions ($P < 10^{-4}$). In addition, positions 8 and 16 were found to be preferentially occupied by A or T residues (79 and 70%, respectively), while positions 1 and 9 were preferentially occupied by G or C residues (54 and 57%, respectively). These four positions corresponded to outliers (positions 1, 9, and 16) or extreme values (position 8) after box plot analysis. The distribution of A/T residues was monotonous at the remaining positions. As shown in Fig. 2, no extreme value was observed along the control sequences. Finally, the box plot analysis revealed that the A/T content spread of the integration sites library was narrower than that of the control library. However, such difference in the spread of the A/T content might be explained in part by the small sample size of the control library.

Figure 3 shows the distribution of each residue along the first 40 positions of HTLV-1 integration sites. The figure shows that the significant C/G richness observed in Fig. 2 at position 1 of HTLV-1 integration sites resulted from an excess of G residues associated with a lack of T residues. The A/T richness at position 8 resulted from a large excess of A residues, while the C/G richness at position 9 corresponded to a significant excess of C residues combined with a lack of A residues. The

A/T richness at position 16 resulted from a combination of A- and T-rich sequences. In addition, Fig. 3 shows that there was an excess of A residues at positions 4 and 28. However, only the A and C contents at positions 8 and 9 and the T content at position 1 were detected as outliers after box plot analysis (not shown). No additional extreme position was noted along HTLV-1 integration sites. The percentage of HTLV-1 and control sequences harboring at least three of the four identified significant hot spots (C/G, A, C, and A/T at positions 1, 8, 9, and 16, respectively) were 34.5 and 15.3%, respectively ($P < 10^{-4}$).

An additional characteristic of the HTLV-1 library was the presence of integration sites harboring long A/T stretches. Indeed, 17% of the 194 HTLV-1 flanking sequences that were longer than 10 bp harbored A/T stretches of ≥ 6 bp compared to only 2% of the control sequences ($P = 0.002$). By contrast, the frequencies of C/G and CpG long stretches were not significantly different between the two libraries (0.5 versus 1% and 3 versus 2%, respectively).

Identification of the host sequences targeted for HTLV-1 integration in vivo. The matches to known sequences are summarized in Table 1, while Table 2 shows the distribution of HTLV-1 3' integration sites and control sequences with signif-

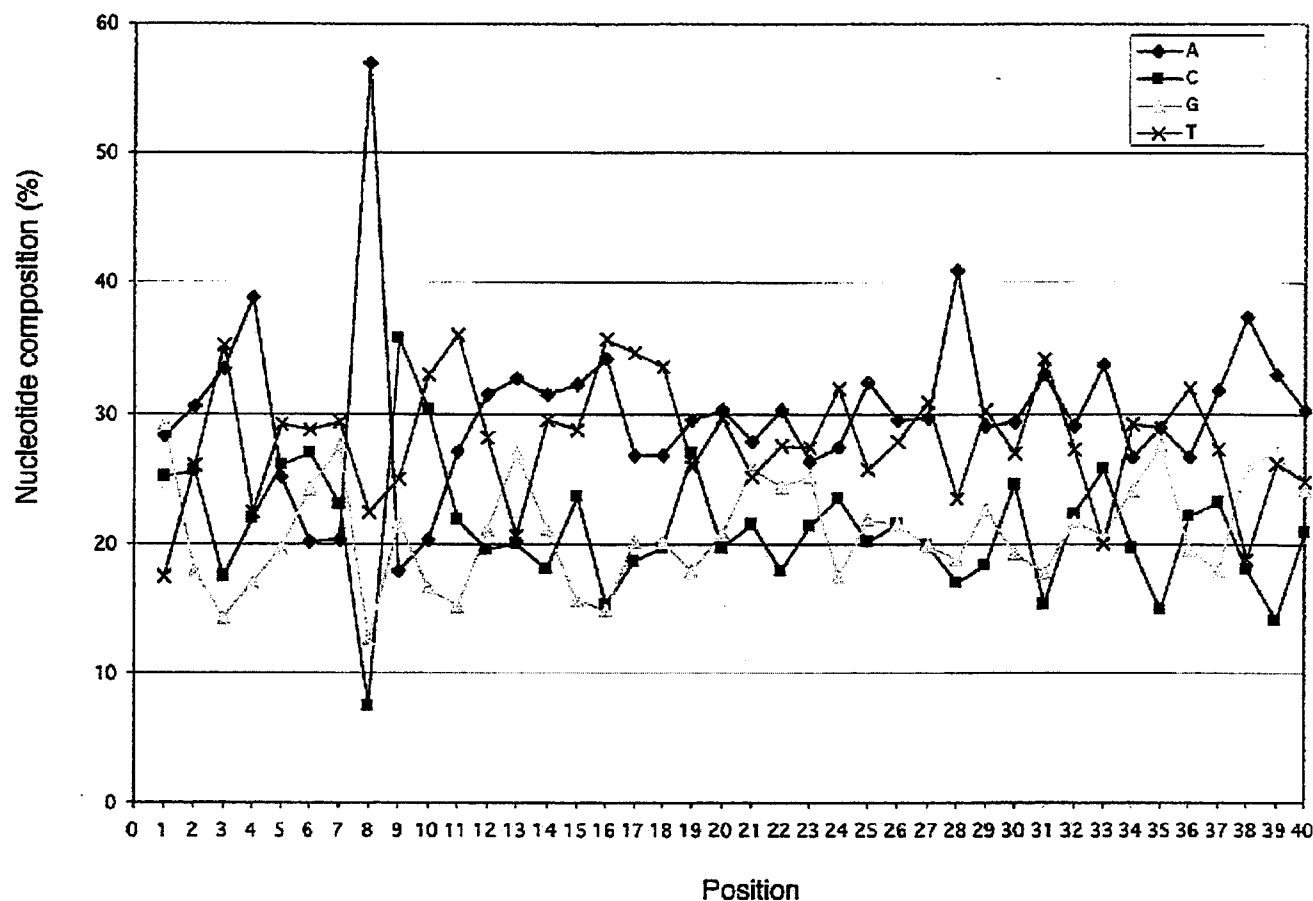


FIG. 3. Nucleotide composition of the HTLV-1 flanking sequences. Sequences isolated after cloning LMPCR and IPCR products were aligned relative to the first base of the hexameric repeat sequences flanking the provirus.

icant homologies to known nonrepetitive elements from databases. The sequences were classified as genomic noncoding nonrepetitive elements; most of these corresponded to GenBank high-throughput genomic sequences, transcription units, and repeat elements. Overall, the mean size of anonymous sequences was not significantly different than that of other sequences (100 versus 99 bp). As shown in Table 1, 84 (47%) of the 178 HTLV-1 sequences of ≥ 25 bp were found to match to known sequences. Of the 44 control sequences of ≥ 25 bp, 17 (39%) matched known sequences. For both libraries, the frequency of matches to repeat sequences was significantly higher than that for transcription units. Table 2 also shows that there was no significant difference in the distribution of matches between the HTLV-1 integration sites and the control sequences. In addition, there was no correlation between the clinical status accompanying to the DNA samples and the type of matches (not shown).

Centromeric alphoid repeats have been previously found to be unfavorable targets for HIV-1 integration in a cell line (6). In the present study, three HTLV-1 integration sites showed significant homologies with such satellite DNA sequences. The first clone corresponded to alpha-satellite DNA located on chromosomes 13, 14, and 21 (48). The second clone was homologous to a human middle repetitive DNA sequence (P. P. Ratnasinghe and P. R. Musich, unpublished data), while the third clone corresponded to beta-satellite DNA located on the distal and proximal short arms of the human acrocentric chromosomes (18). No satellite DNA was observed in the control library.

The A/T contents of anonymous sequences, repeat elements, genomic noncoding nonrepetitive elements, and transcription units were 59, 55, 58, and 55%, respectively. These differences were statistically significant ($P = 0.0063$, Kruskal-Wallis one-way analysis-of-variance). Among repeat elements,

TABLE 1. Distribution of the number and the percentage of matches to known sequences in HTLV-1 integration sites and control sequences longer than 25 bp

Sequence type	No. of HTLV-1 integration sites (%)	No. of control sequences (%)
Genomic sequences	15 (8)	1 (2)
Transcription units	10 (6)	3 (7)
cDNA	8 (5)	3 (7)
Genes	2 (1)	0 (0)
Repeat elements	59 (33)	13 (30)
SINE elements	20 (11)	6 (14)
LINE elements	20 (11)	3 (7)
RLE elements	11 (6)	3 (7)
DNA transposon	1 (1)	1 (2)
DNA satellite	3 (2)	0 (0)
Simple repeat	4 (2)	0 (0)
Anonymous	94 (53)	27 (61)
Total of matched sequences	84 (47)	17 (39)

TABLE 2. Integration sites and control sequences having significant homologies to known nonrepetitive elements

Group and sequence no.	Sample origin	Length (bp)	Hexameric repeat	AT (%)	Identified similarities	
					Type	Accession no. ^b
ATLL						
78	Tumorous ^a	200	TTATTC	60	cDNA	HSU68704
84	Nontumorous	141	AGCAAG	66	cDNA	N53238
74	Tumorous	99	TCTTTC	60	Gene (human P protein, exon 23)	HSPPROT23
85	Nontumorous	28	CCTCTC	68	Noncoding nonrepetitive	AL034422
91	Tumorous	36	CAGCTG	50	Noncoding nonrepetitive	AQ223501
Asymptomatic carriers						
10	PBMCs	25	TCCGCA	36	cDNA	AA451666
18	PBMCs	32	ACCCGC	41	cDNA	AA631969
27	PBMCs	25	GCAACT	60	cDNA	AC002036
39	PBMCs	25	GCAAAA	52	cDNA	AC004583
58	PBMCs	45	TTATGT	69	Noncoding nonrepetitive	AC006054
23	PBMCs	34	GTTATA	71	Noncoding nonrepetitive	AC006227
34	PBMCs	25	GAGAAC	52	Noncoding nonrepetitive	AQ306143
37	PBMCs	101	CTGTGG	59	Noncoding nonrepetitive	B17581
20	PBMCs	38	GGTGTG	42	Noncoding nonrepetitive	HS232D4
TSP/HAM						
233	PBMCs	26	TAATAG	62	cDNA	H72803
231	PBMCs	69	CTTGGT	51	cDNA	W57727
224	PBMCs	365	GCTAGG	61	Gene (alpha enolase, exon 1)	HSENOAL1
220	PBMCs	29	CATATG	55	Noncoding nonrepetitive	AC003693
216	CSF	25	GCTAGG	48	Noncoding nonrepetitive	AC004505
183	PBMCs	163	ACATTT	59	Noncoding nonrepetitive	AC005881
274	PBMCs	52	CTGAGG	44	Noncoding nonrepetitive	AL021937
275	PBMCs	75	TCAGTC	55	Noncoding nonrepetitive	AL022345
223	PBMCs	144	GAGAAT	70	Noncoding nonrepetitive	AL031599
246	PBMCs	30	TCAATC	67	Noncoding nonrepetitive	AL031683
204	CSF	41	TAAAGT	78	Noncoding nonrepetitive	HUAC002307
Control sequences						
109		64		58	cDNA	AI001768
137		25		72	cDNA	AA969105
151		30		43	cDNA	AA909212
110		45		42	Noncoding nonrepetitive	HS431P23

^a Tumorous clones correspond to malignant ATLL cellular clones.^b GenBank accession number of the corresponding database sequences (2).

the A/T content of LINE elements was significantly higher than that of the SINE element, i.e., 60 versus 52% ($P < 10^{-4}$). The A/T contents of retrovirus-like elements (RLE) and other repeat elements were 53 and 55%, respectively. There was a weak correlation between the A/T content of identified database sequences and their frequencies of match with HTLV-1 integration sites or control sequences. Indeed, the proportion of integration sites matching with the A/T-rich LINE elements was higher than that for control sequences: 11 versus 7% (Table 1). Similarly, the proportion of integration site matching with the C/G-rich LINE elements was lower than that for control sequences: 11 versus 14% (Table 1). However, these differences were not statistically significant.

DISCUSSION

In the infected cell, integrase functions as a component of the preintegration complex derived from the virus core. The main factors which may influence the choice for a chromosomal target site during infection include chromatin structure (26), target DNA sequence (15), host cell proteins, and virus-encoded proteins other than integrase (4). The present study was conducted in order to assess the selection of target sites during integration of the HTLV-1 provirus *in vivo*. To this end,

HTLV-1 integration sites, isolated by IPCR or LMPCR amplification from naturally infected individuals, were sequenced. The nucleotide composition of the flanking sequences was analyzed, and their homologies to known sequences were identified. The nature of the HTLV-1 flanking sequences was compared to that of uninfected DNA. Since PCR and cloning are influenced by the structure and the nucleotide content of the target sequence, we designed a control experiment in order to isolate uninfected DNA sequences by using a strategy identical to that used for the analysis of HTLV-1 flanking sequences. The A/T content was found to be 51% after this control experiment, a value significantly lower than that initially described for human DNA (58%) (43). This indicates that the A/T content is underestimated by IPCR and LMPCR and emphasizes the need to perform a comparative analysis of HTLV-1 integration sites with sequences isolated by the same manner. Results from the comparative analysis show that there is a nucleotide composition bias at HTLV-1 integration sites. The virus integrates into A/T-rich regions with a nonhomogeneous distribution of residues 3' to the provirus. These characteristics do not appear to be associated with a preference for integration in transcription units or repeat elements.

Preference for A/T-rich regions as a common feature for *in vivo* and *in vitro* integration. HTLV-1 flanking sequences

(~19 kb) isolated from naturally infected individuals were found to be significantly more A/T-rich than control sequences. A preference for A/T-rich regions has been previously described for insertion by transposons (36) or Ty elements (28) and for integration of adenovirus (24), spleen necrosis virus (40), avian myeloblastosis virus (16, 17) and Moloney murine leukemia virus (31). By hybridization of a viral probe with compositional fractions of HTLV-1 cultured cells DNA, HTLV-1 sequences have been found to be distributed in the 46 to 61% A/T range of the host genome (52). Recently, four HTLV-1 flanking sequences (183 bp) isolated from four ATLL patients were found to have an average A/T content of 63% (13). Altogether, these data indicate that a preference for A/T-rich regions characterizes both *in vitro* and *in vivo* integration.

The target DNA structure affects HTLV-1 integration site selection. Integration of proviral DNA into the host genome generates short direct repeats of 4 to 6 bp as a result of DNA repair to the cellular sequence flanking the integrated provirus (45). Analysis of the nucleotide sequence surrounding the integrated provirus indicates that the size of the direct repeat is characteristic of each virus. For MMLV, the middle two positions of the 4-bp direct repeat are preferentially occupied by AA, TT, or AT dinucleotides (31). Similarly, a preferential AT pairing toward the central portion of the direct repeat has been previously described in direct repeat generated by the insertion of yeast Ty elements (28) and in the target of site-specific recombination by FLP recombinase (44). For HIV, analysis of the 112 published direct repeats shows that there is a preference for a G residue at the first position, for T at the second position, for A at the third and fourth positions, and for C at the fifth position (6, 41, 47). Here, alignment of the 218 HTLV-1 clones showed a weak consensus sequence within the hexameric repeat flanking the HTLV-1 provirus. However, only the C/G content at position 1 of the hexameric repeat corresponded to an outlier as shown in Fig. 2. In fact, we have found a significant nucleotide composition bias at positions within and without the hexameric repeat. Indeed, there was a preference for C or G residues at position 1, for A residues at position 8, for C residues at position 9, and for A or T residues at position 16. These weakly conserved motifs observed in the vicinity of the provirus 3' end may constitute a propitious binding site for the preintegration complex. Alternatively, they might correspond to binding sites for other cellular or viral factors that interact favorably with the integration machinery.

DNA bending creates favored sites for retroviral integration (4, 27). Such distortion may result from DNA-protein interaction (4). Alternatively, the target sequence itself may result in intrinsic DNA bent (27). In the present study, the presence of long runs of A/T residues was almost restricted to HTLV-1 integration sites compared to control sequences. Indeed, some of the HTLV-1 flanking sequences isolated here were found to have strong homologies with intrinsically bent A-tract DNA (21). Therefore, in addition to a preference for A/T-rich regions of the host cell chromosomes, the target choice for HTLV-1 integration appears to be influenced by the target DNA structure. In addition to an intrinsically bent target, it is possible that the consensus motifs observed along HTLV-1 integration sites may contribute to the binding of proteins that induce DNA bends.

Lack of evidence for favored HTLV-1 integration near known sequences. Nucleotide composition of the database sequences with significant homologies with HTLV-1 integration sites reflected the nature of the corresponding matches. Indeed, transcription unit sequences harbored the lowest A/T range, which correspond to DNA regions of high gene concen-

tration. However, there was no significant difference in the frequency and distribution of matches to known sequences between HTLV-1 and control sequences. In contrast to results from a previous analysis of HIV-1 integration in SupT1 cells (6), we found that HTLV-1 can integrate into aliphoid repeats *in vivo*. The frequency of HTLV-1 integration in repeat element was 33%, a value similar to that previously described for HIV integration in cultured cells (6). The frequency of matches to known transcription units was not significantly different between HTLV-1 and control sequences. This suggests that transcriptionally active regions are not preferred targets for HTLV-1 integration *in vivo*.

HTLV-1 integration may alter gene expression *in vivo*. In contrast to Ty3 elements (5, 11, 12), Ty1 elements (20), and retroviruses that induce chronic leukemia in animals (25), HTLV-1 integration does not preferentially occur near cellular sequences such as proto-oncogenes (38). However, about 6% of the HTLV-1 proviruses analyzed in the present work were found to be integrated into transcription units. This suggests that in some cells, HTLV-1 integration may alter gene expression *in vivo*. Given the elevated proviral load in HTLV-1-infected individuals, a substantial number of clonally expanded cells must have disrupted transcription units (6% of a large number is substantial). It is possible that some of these events could contribute further to clonal expansion.

Basis of HTLV-1 target site selection *in vivo*. HTLV-1 replicates mainly via the mitosis of its host cells (49) and the number of distinct circulating clones of infected CD4 T cells reflects the number of integration events. In a previous work, the overall number of such events has been estimated in an asymptomatic carrier in whom 20 distinct clones of infected cells were evidenced (10). Although this number is higher in TSP/HAM and in ATLL (7, 9), the present collection of HTLV-1 flanking sequences appears to correspond to a representative sample of *in vivo* HTLV-1 integration events.

At the level of nucleotide sequence, the present study shows that HTLV-1 integration is not completely random *in vivo*. The differences observed between HTLV-1 integration sites and control sequences indicate that the integration machinery shows subtle preferences *in vivo*. Given the huge excess of potential sites, the observed subset may reflect those with a kinetic advantage in recognition, which is influenced by the sequence within and without the hexanucleotide. Present results suggest that the structure of the target DNA plays an important role for integration *in vivo*. However, the structural constraints evidenced here among the 218 HTLV-1 flanking sequences do not appear to be linked to a preference for integration in specific regions of host cell chromosomes. Therefore, most or all regions of the host cell genome seem to be accessible to HTLV-1 integration *in vivo*.

ACKNOWLEDGMENTS

This work was supported by grants from the Association pour la Recherche sur le Cancer, from the Ligue Nationale contre le Cancer (Comité Pas de Calais), and from the Fondation Contre la Leucémie. I.L. and F.M. were supported by bursaries from the Ministère de l'Enseignement Supérieur et de la Recherche.

We thank P. Wattre and collaborators, who kindly received us in their laboratories for DNA extraction, digestion, ligation, and PCR. We also thank Marie-Dominique Reynaud for assistance.

REFERENCES

1. Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389-3402.
2. Bilofsky, H. S., and C. Burks. 1988. The GenBank genetic sequence data bank. *Nucleic Acids Res.* 16:1861-1863.

3. Boguski, M. S., T. M. Lowe, and C. M. Tolstoshev. 1993. dbEST database for "expressed sequence tags." *Nat. Genet.* 4:332-333.
4. Bor, Y. C., F. D. Bushman, and L. E. Orgel. 1995. In vitro integration of human immunodeficiency virus type 1 cDNA into targets containing protein-induced bends. *Proc. Natl. Acad. Sci. USA* 92:10334-10338.
5. Brodeur, G. M., S. B. Sandmeyer, and M. V. Olson. 1983. Consistent association between sigma elements and tRNA genes in yeast. *Proc. Natl. Acad. Sci. USA* 80:3292-3296.
6. Carreau, S., C. Hoffmann, and F. Bushman. 1998. Chromosome structure and human immunodeficiency virus type 1 cDNA integration: centromeric alphoid repeats are a disfavored target. *J. Virol.* 72:4005-4014.
7. Cavois, M., A. Gessain, S. Wain-Hobson, and E. Wattel. 1996. Proliferation of HTLV-1 infected circulating cells in vivo in all asymptomatic carriers and patients with TSP/HAM. *Oncogene* 12:2419-2423.
8. Cavois, M., I. Leclercq, O. Gout, A. Gessain, S. Wain-Hobson, and E. Wattel. 1998. Persistent oligoclonal expansion of human T-cell leukemia virus type 1 infected circulating cells in patients with Tropical spastic paraparesis/HTLV-1 associated myelopathy. *Oncogene* 17:77-82.
9. Cavois, M., S. Wain-Hobson, A. Gessain, Y. Plumelle, and E. Wattel. 1996. Adult T-cell leukemia/lymphoma on a background of clonally expanding HTLV-1 positive cells. *Blood* 88:4646-4650.
10. Cavois, M., S. Wain-Hobson, and E. Wattel. 1995. Stochastic events in the amplification of HTLV-1 integration sites by linker-mediated PCR. *Res. Virol.* 146:179-184.
11. Chalker, D. L., and S. B. Sandmeyer. 1990. Transfer RNA genes are genomic targets for de novo transposition of the yeast retrotransposon Ty3. *Genetics* 126:837-850.
12. Chalker, D. L., and S. B. Sandmeyer. 1992. Ty3 integrates within the region of RNA polymerase III transcription initiation. *Genes Dev.* 6:117-128.
13. Chou, K. S., A. Okayama, L. J. Su, T. H. Lee, and M. Essex. 1996. Preferred nucleotide sequence at the integration target site of human T-cell leukemia virus type I from patients with adult T-cell leukemia. *Int. J. Cancer* 65:20-24.
14. Craigie, R. 1992. Hotspots and warm spots: integration specificity of retroelements. *Trends Genet.* 8:187-190.
15. Fitzgerald, M. L., and D. P. Grandgenett. 1994. Retroviral integration: in vitro host site selection by avian integrase. *J. Virol.* 68:4314-4321.
16. Fitzgerald, M. L., A. C. Vora, W. G. Zeh, and D. P. Grandgenett. 1992. Concerted integration of viral DNA termini by purified avian myeloblastosis virus integrase. *J. Virol.* 66:6257-6263.
17. Grandgenett, D. P., R. B. Inman, A. C. Vora, and M. L. Fitzgerald. 1993. Comparison of DNA binding and integration half-site selection by avian myeloblastosis virus integrase. *J. Virol.* 67:2628-2636.
18. Greig, G. M., and H. F. Willard. 1992. Beta satellite DNA: characterization and localization of two subfamilies from the distal and proximal short arms of the human acrocentric chromosomes. *Genomics* 12:573-580.
19. Howard, M. T., and J. D. Griffith. 1993. A cluster of strong topoisomerase II cleavage sites is located near an integrated human immunodeficiency virus. *J. Mol. Biol.* 232:1060-1068.
20. Ji, H., D. P. Moore, M. A. Blomberg, L. T. Braiterman, D. F. Voytas, G. Natsoulis, and J. D. Boeke. 1993. Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences. *Cell* 73:1007-1018.
21. Kahn, J. D., and D. M. Crothers. 1992. Protein-induced bending and DNA cyclization. *Proc. Natl. Acad. Sci. USA* 89:6343-6347.
22. Kettmann, R., M. Meunier-Rotival, J. Cortadas, G. Cuny, J. Ghysdael, M. Mammerickx, A. Burny, and G. Bernardi. 1979. Integration site of bovine leukemia virus DNA in the bovine genome. *Arch. Int. Physiol. Biochim.* 87: 818-819.
23. Kitamura, Y., Y. M. Lee, and J. M. Coffin. 1992. Nonrandom integration of retroviral DNA in vitro: effect of CpG methylation. *Proc. Natl. Acad. Sci. USA* 89:5532-5536.
24. Knoblauch, M., J. Schrer, B. Schmitz, and W. Doerfler. 1996. The structure of adenovirus type 12 DNA integration sites in the hamster cell genome. *J. Virol.* 70:3788-3796.
25. Kung, H. J., C. Boerkoel, and T. H. Carter. 1991. Retroviral mutagenesis of cellular oncogenes: a review with insights into the mechanisms of insertional activation. *Curr. Top. Microbiol. Immunol.* 171:1-25.
26. Mooslehner, K., U. Karls, and K. Harbers. 1990. Retroviral integration sites in transgenic Mice frequently map in the vicinity of transcribed DNA regions. *J. Virol.* 64:3056-3058.
27. Müller, H. P., and H. E. Varmus. 1994. DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. *EMBO J.* 13:4704-4714.
28. Oyen, T. B., and O. S. Gabrielsen. 1983. Non-random distribution of the Ty1 elements within nuclear DNA of *Saccharomyces cerevisiae*. *FEBS Lett.* 161: 201-206.
29. Pearson, W. R., and D. J. Lipman. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* 85:2444-2448.
30. Pruss, D., F. D. Bushman, and A. P. Wolffe. 1994. Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. *Proc. Natl. Acad. Sci. USA* 91:5913-5917.
31. Pryciak, P. M., and H. E. Varmus. 1992. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell* 69:769-780.
32. Robinson, H. L., and G. C. Gagnon. 1986. Patterns of proviral insertion and deletion in avian leukosis virus-induced lymphomas. *J. Virol.* 57:28-36.
33. Rohdewohld, H., H. Weiher, W. Reik, R. Jaenisch, and M. Breindl. 1987. Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. *J. Virol.* 61: 336-343.
34. Rynditch, A., F. Kadi, J. Geryk, S. Zoubak, J. Svoboda, and G. Bernardi. 1991. The isopycnic, compartmentalized integration of Rous sarcoma virus sequences. *Gene* 106:165-172.
35. Salinas, J., M. Zerial, J. Filipinski, M. Crepin, and G. Bernardi. 1987. Non-random distribution of MMTV proviral sequences in the mouse genome. *Nucleic Acids Res.* 15:3009-3022.
36. Sandmeyer, S. B., L. J. Hansen, and D. L. Chalker. 1990. Integration specificity of retrotransposons and retroviruses. *Annu. Rev. Genet.* 24:491-518.
37. Scheridin, U., C. Rhodes, and M. Breindl. 1990. Transcriptionally active genome regions are preferred targets for retrovirus integration. *J. Virol.* 64: 907-912.
38. Seiki, M., R. Eddy, T. Shows, and M. Yoshida. 1984. Nonspecific integration of the HTLV provirus genome into adult T-cell leukemia cells. *Nature* 309: 640-642.
39. Shih, C. C., J. P. Stoye, and J. M. Coffin. 1988. Highly preferred targets for retrovirus integration. *Cell* 53:531-537.
40. Shimotohno, K., and H. M. Temin. 1980. No apparent nucleotide sequence specificity in cellular DNA juxtaposed to retrovirus proviruses. *Proc. Natl. Acad. Sci. USA* 77:7357-7361.
41. Stevens, S. W., and J. D. Griffith. 1996. Sequence analysis of the human DNA flanking sites of human immunodeficiency virus type 1 integration. *J. Virol.* 70:6459-6562.
42. Stoesser, G., M. A. Tuli, R. Lopez, and P. Sterk. 1999. The EMBL nucleotide sequence database. *Nucleic Acids Res.* 27:18-24.
43. Swartz, M. N., T. A. Trautner, and A. Kornberg. 1962. Enzymatic synthesis of deoxyribonucleic acid. XI. Further studies on nearest neighbor base sequences in deoxyribonucleic acid. *J. Biol. Chem.* 237:1961-1967.
44. Umlauf, S. W., and M. M. Cox. 1988. The functional significance of DNA sequence structure in a site-specific genetic recombination reaction. *EMBO J.* 7:1845-1852.
45. Varmus, H. E. 1983. Using retroviruses as insertional mutagens to identify cellular oncogenes. *Prog. Clin. Biol. Res.* 119:23-35.
46. Varmus, H. E., and P. O. Brown. 1989. Mobile DNA, p. 53-108. In M. M. Howe and D. E. Berg (ed.), *Retroviruses*. American Society of Microbiology, Washington, D.C.
47. Vincent, K. A., D. York Higgins, M. Quiroga, and P. O. Brown. 1990. Host sequences flanking the HIV provirus. *Nucleic Acids Res.* 18:6045-6047.
48. Vissel, B., A. Nagy, and K. H. Choo. 1992. A satellite III sequence shared by human chromosomes 13, 14, and 21 that is contiguous with alpha satellite DNA. *Cytogenet. Cell Genet.* 61:81-86.
49. Wattel, E., J. P. Vartanian, C. Pannetier, and S. Wain-Hobson. 1995. Clonal expansion of human T-cell leukemia virus type I-infected cells in asymptomatic and symptomatic carriers without malignancy. *J. Virol.* 69:2863-2868.
50. Withers-Ward, E. S., Y. Kitamura, J. P. Barnes, and J. M. Coffin. 1994. Distribution of targets for avian retrovirus DNA integration in vivo. *Genes Dev.* 8:1473-1478.
51. Zerial, M., J. Salinas, J. Filipinski, and G. Bernardi. 1986. Genomic localization of hepatitis B virus in a human hepatoma cell line. *Nucleic Acids Res.* 14:8373-8386.
52. Zoubak, S., J. H. Richardson, A. Rynditch, P. Hollsberg, D. A. Hafler, E. Boeri, A. M. Lever, and G. Bernardi. 1994. Regional specificity of HTLV-I proviral integration in the human genome. *Gene* 143:155-163.

