HAL
open science

# Double approche à la thérapie anti-tumorale à l'aide de vecteurs lentiviraux

Tata Safiatou Coulibaly

## To cite this version:

Tata Safiatou Coulibaly. Double approche à la thérapie anti-tumorale à l'aide de vecteurs lentiviraux. Cancer. Université de Strasbourg, 2017. Français. NNT : 2017STRAJ087 . tel-01763161

**HAL Id: tel-01763161**

**https://theses.hal.science/tel-01763161v1**

Submitted on 10 Apr 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DE STRASBOURG

École doctorale des sciences de la vie et de la santé

CNRS UPR 9002

# THÈSE

Présentée par

**Tata Safiatou COULIBALY**

soutenue le 13 Octobre 2017

pour obtenir le grade de : **Docteur de l'Université de Strasbourg**

Discipline : Virologie : Aspects moléculaires et médicaux

# Double approche à la thérapie anti-tumorale à l'aide de vecteurs lentiviraux

THÈSE dirigée par :

Monsieur Matteo NEGRONI — Docteur, Université de Strasbourg

Rapporteurs :

Monsieur Eric ROBINET — Docteur, EFS Besançon

Monsieur Jean-Christophe PAGES — Professeur, Université de Tours

---

Autre membre du jury :

Madame Christiane MOOG — Docteur, Université de Strasbourg

# Remerciements

*Je remercie tout d'abord les membres de mon jury, le Pr Jean-Christophe PAGES, le Dr Christiane MOOG et le Dr Éric ROBINET d'avoir accepté d'évaluer mon travail de thèse.*

*Merci à Matteo de m'avoir permis d'intégrer ton équipe, de m'avoir guidée tout au long de cette aventure qu'est la thèse. Tes conseils, tes remarques et tes critiques m'ont permis de progresser.*

*Merci à tous les membres du labo 337. Paola, tu m'as supervisée lors de mes premiers pas au sien du laboratoire et je t'en remercie. Daniela, merci pour ta disponibilité. Malgré un emploi du temps chargé, tu as toujours pris du temps lorsque j'ai eu besoin de tes conseils. Flore, ta bonne humeur constante me manquera énormément, j'ai beaucoup appris à ton contact, tu es une personne formidable. Pierre toutes ces poses café à admirer la « vue » sur le campus me manqueront. Romain, merci (et surtout bravo) d'avoir supporté mes blagues  et se n'est pas rien. Marine, merci pour les séances jogging, sans toutes les endorphines libérées au cours de ces séances je pense que la rédaction aurait été beaucoup plus laborieuse (au taquet pour un nouveau semi-marathon !). Alexis merci pour les fous rires, les séances de danses dans les couloirs, et les soirées complaintes. Je te souhaite le meilleur pour la suite.*

*Merci à tous les membres de l'UPR 9002 ainsi qu'à tous les membres de l'IBMC.*

*Merci à mes amis, ma blonde, Asso, Malick, Sarah, Yanina, Martina, Anne-Sophie, Djibril. On dit que l'on a deux familles, celles qu'on a à la naissance et celle que l'on se choisi. Vous êtes ma deuxième famille et je ne serais jamais arrivé jusqu'ici sans vous.*

*Merci à Ibrahima, ta présence et ton soutien en particulier pendant la période de rédaction. Ce n'était pas une période facile pour toi mais tu m'as toujours soutenu et tu as su me remotiver dans les moments de fatigue. Merci d'être là.*

*Je voudrais enfin remercier ma famille. Merci à mes parents, mes sœurs, mes frères et mon neveu. Pendant cette période mon nouveau surnom a été « fantomas »  car je n'ai pas beaucoup appelé. En revanche chaque conversation avec vous m'a redonné confiance en moi et m'a permis d'aller au bout de cette aventure. Merci pour votre confiance, votre soutien et vos conseils. Chacun de vous est un modèle pour moi.*

Tata Safiatou COULIBALY

# Double approche à la thérapie anti-tumorale à l'aide de vecteurs lentiviraux


## Résumé

Le traitement du cancer par thérapie génique nécessite d'une part des gènes suicides efficaces et, d'autre part, l'adressage spécifique de ces gènes aux cellules cancéreuses. J'ai d'abord caractérisé un nouveau gène suicide dérivé de la désoxycytidine kinase humaine (dCK) : le M36. Comparé à la dCK, le M36 permet une meilleure sensibilisation des certaines cellules cancéreuses aux traitements avec différents chimiothérapeutiques comme la gemcitabine et la cytarabine. Ces résultats sont particulièrement encourageants pour l'élimination des cellules cancéreuses résistantes à ces traitements du fait d'un défaut de la dCK. Dans une deuxième partie, je me suis intéressée à l'adressage spécifique des transgènes aux cellules cancéreuses par les vecteurs lentiviraux. J'ai travaillé à la preuve de concept qu'une enveloppe (Env) VIH modifiée peut permettre un tel ciblage. J'ai généré une Env qui a fortement diminué son tropisme naturel et qui comporte un motif liant le marqueur tumoral modèle HER2.


Mots-clés : Thérapie génique des cancers, gène suicide, enveloppe VIH-1, ciblage spécifique des cellules cancéreuses


## Abstract

Cancer gene therapy requires the use of an effective suicide gene and the specific targeting of cancer cells. In my PhD work, I have first characterized a new potential suicide gene derived from human deoxycytidine kinase (dCK): M36. Compared to dCK, M36 improves sensitization of certain cancer cells to treatment with chemotherapeutic compounds as gemcitabine and AraC. These results are particularly encouraging for the elimination of cancer cells resistant to the treatment because of a defect with dCK. In a second part, I have worked at the proof of concept that a modified HIV envelope can allow specific targeting of cancer cells by lentiviral vectors. During this work, I have generated a CD4i envelope with a strongly diminished natural tropism and that carries a motif known to bind the model cell surface cancer marker HER2. This envelope constitutes a good starting material to be improved by evolution in cell culture to obtain specific targeting of HER2+ cells.


Keywords : Cancer gene therapy, suicide gene, HIV-1 envelope, specific targeting of cancer cells

## Sommaire

**PARTIE I : LA THERAPIE GENIQUE DANS LE DOMAINE DU CANCER** ..... 3

**I GENERALITES SUR LE CANCER** ..... 3
I.1 PROCESSUS DE CANCERISATION ..... 3
I.2 MICROENVIRONNEMENT TUMORAL ..... 5

**II TRAITEMENTS CLASSIQUES DES CANCERS** ..... 7
II.1 CHIRURGIE ..... 7
II.2 RADIOTHERAPIE ..... 8
II.3 CHIMIOTHERAPIE ..... 8

**III LA THERAPIE GENIQUE DANS LE DOMAINE DU CANCER** ..... 9
III.1 GENERALITES SUR LA THERAPIE GENIQUE ..... 9
III.2 IMMUNOTHERAPIE DES CANCERS ..... 11
*III.2.1 Le transfert de lymphocytes T infiltrant* ..... *11*
*III.2.2 La TCR thérapie* ..... *12*
*III.2.3 La thérapie CAR-T* ..... *13*
III.3 LA THERAPIE PAR GENES SUICIDES ..... 15
*III.3.1 La thymidine kinase du virus herpes simplex 1* ..... *16*
*III.3.2 La désoxycytidine kinase humaine* ..... *17*
*III.3.3 Nouvelle stratégie d'évolution moléculaire : la rétrovolution* ..... *20*

**PARTIE II : LES OUTILS DU TRANSFERT DE GENE** ..... 24

**I LES VECTEURS NON VIRAUX** ..... 24
I.1 LE TRANSFERT PHYSIQUE ..... 24
I.2 LE TRANSFERT PAR TRANSPORTEURS CHIMIQUES ..... 25

**II LES VECTEURS VIRAUX** ..... 27
II.1 LES ADENOVIRUS ..... 28
*II.1.1 Cycle de réplication* ..... *28*
*II.1.2 Utilisation en thérapie génique* ..... *30*
II.2 LES VIRUS ADENO-ASSOCIES ..... 30
*II.2.1 Cycle de réplication* ..... *30*
*II.2.2 Utilisation en thérapie génique* ..... *31*

**III LES VECTEURS LENTIVIRAUX** ..... 33
III.1 ORGANISATION DU GENOME ..... 33
*III.1.1 Les régions non codantes* ..... *34*
*III.1.2 Les protéines de structures* ..... *35*
*III.1.3 Les protéines régulatrices et les protéines auxiliaires* ..... *36*
III.2 CYCLE DE REPLICATION ..... 38
*III.2.1 La phase précoce* ..... *39*
*III.2.2 La phase tardive* ..... *43*
III.3 REDIRECTION DE L'ENTREE VIRALE ..... 46
*III.3.1 Les mécanismes moléculaires de l'entrée* ..... *46*
*III.3.2 Les enveloppes CD4 indépendantes* ..... *50*
*III.3.3 La redirection de l'entrée virale* ..... *51*
*III.3.4 L'introduction d'un peptide dans l'enveloppe virale* ..... *52*

**PARTIE III : OBJECTIFS DU TRAVAIL DE THESE** ..... 54

**PARTIE IV : ETUDE DES MUTANTS DE LE DCK** ……56

**I OBJECTIFS** ..... **56**

**II MATERIELS ET METHODES** ..... **58**
II.1 LIGNEES CELLULAIRES ..... 58
II.2 CONSTRUCTIONS DES MUTANTS DE LA DCK ..... 58
II.3 PRODUCTION DES PARTICULES VIRALES ..... 59
II.4 ETABLISSEMENT DES LIGNEES CELLULAIRES ..... 60
II.5 TESTS DE SENSIBILITE AUX MEDICAMENTS ..... 60
II.6 SIARN ET SHARN ..... 60
II.7 ANALYSE DES WESTERN BLOT ..... 61

**III RESULTATS** ..... **61**
III.1 ARTICLE ..... 61
III.2 ETUDE DU MUTANT M36 ..... 63
*III.2.1 Capacité à sensibiliser différentes lignées cellulaires cancéreuses* ..... *63*
*III.2.2 Sensibilisation des cellules par une expression du M36 et une réduction de l'expression de la RRM1* ..... *66*
*III.2.3 Caractérisation des différentes mutations du M36* ..... *69*
*III.2.4 Modification du M36 sur une base rationnelle en vue d'identifier un meilleur variant de la protéine* ..... *70*
*III.2.5 Etude structurelle de la protéine M36* ..... *72*

**IV DISCUSSION** ..... **74**

**PARTIE V : MODIFICATION DE L'ENVELOPPE POUR UN CIBLAGE SPECIFIQUE DES CELLULES** ……80

**I OBJECTIFS** ..... **80**

**II MATERIELS ET METHODES** ..... **80**
II.1 MODIFICATION DE L'ENVELOPPE ..... 81
II.2 CELLULES ..... 81
II.3 PRODUCTIONS DES PARTICULES VIRALES ..... 81
II.4 WESTERN BLOT ..... 82
II.5 TEST D'ENTREE ..... 82

**III RESULTATS** ..... **83**
III.1 CONSTRUCTION D'UNE ENVELOPPE CD4I ..... 83
*III.1.1 Augmenter la capacité à réaliser l'entrée dans les cellules CD4-* ..... *83*
*III.1.2 Réduire la capacité à réaliser l'entrée dans les cellules exprimant le récepteur CD4-* ..... *85*
III.2 INTRODUCTION D'UN LIGAND SPECIFIQUE POUR LE MARQUEUR TUMORAL ..... 88
*III.2.1 Choix du marqueur tumoral à utiliser comme modèle* ..... *88*
*III.2.2 Choix du site d'introduction du ligand* ..... *88*
*III.2.3 Choix de la manière dont introduire le ligand* ..... *89*

**IV DISCUSSION** ..... **92**

**PARTIE VI : CONCLUSIONS ET PERSPECTIVES** ……97

**PARTIE VII : SUIVI DES CELLULES SOUCHES EMBRYONNAIRE DANS LE MODELE DU POISSON ZEBRE** ……103

# Liste des abréviations

*Å* Ångstrom
*Aa* Acide aminé
*AAV* Virus adéno-associés
*Ad* : Adénovirus
*ADN* Acide désoxyribonuléique
*AHNP* Anti-HER2/neu peptide
*Alix* ALG-2 interacting protein X
*APOBEC3G* Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G
*AraC* Cytarabine
*Arg* Arginine
*ARN* Acide ribonucléique
*ARNm* Acide ribonucléique messager
*ARNt* Acide ribonucléique de transfert
*Asn* Asparagine
*Asp* Acide aspartique
*ATP* Adénosine triphosphate
*BST2* Antigène 2 issu des cellules stromales de la moelle osseuse
*CA* Capside
*CAR-T* Récepteur antigénique chimérique des lymphocytes T
*CCR5* Récepteur C-C chimiokine de type 5
*CD4* Cluster de différentiation 4
*CD4i* CD4 indépendant
*cm* Centimètre
*CMH* Complexe majeur d'histocompatibilité
*CPSF* Facteur spécifique au clivage et à la polyadénylation
*CsTF* Facteur de stimulation du clivage
*CXCR4* C-X-C chimiokine de type 4
*dA* Désoxyadénosine
*dC* Désoxycytidine
*dCK* Désoxycytidine kinase
*dG* Désoxyguanosine
*ELISA* Dosage d'immuno-absorbance par enzyme liée
*FBS* Sérum de veau faetal
*Gem* Gemcitabine
*Glu* Acide glutamique
*gp* Glycoprotréine
*HCT116* Cellules de tumeur du colon humain
*HEK* Cellules rénales embryonnaires humaines
*HER2* Récepteur du facteur de croissance épidermique humain
*hsTK* Thymidine kinase du virus herpes simplex 1
*iCasp9* Caspase 9 inductible
*IN* Intégrase
*ITR* Région répétée inversée
*Kb* kilobase
*LTR* Longue région terminale répétée
*MA* Matrice
*MW* Poids moléculaire
*NC* Nucléocapside
*Nef* Facteur négatif
*nt* Nucléotide
*PCR* Réaction de polymérisation en chaîne
*PEI* Polyéthylènimine
*Phe* Phénylalanine
*PIC* Complexe de préintégration
*PLL* Polylysine
*PVDF* Polyfluorure de vynilidène
*R* Séquence répétée
*RE* Réticulum endoplasmique
*Rev* Régulateur de l'expression des protéines
*RRE* Eléments régulateur de Rev
*RRM1* Sous-unité 1 de la ribonucléotide réductase
*RT* Transcriptase inverse
*RTC* Complexe de transcription inverse
*SA* Site accepteur d'épissage
*SD* Site donneur d'épissage
*Ser* Sérine
*SIN* Auto inactivant
*SP* Espaceur
*Tat* Trans-activateur de la transcription
*TCR* Récepteur des cellules T
*Trp* Tryptophane
*U3* Séquence unique en 3'
*U5* Séquence unique en 5'
*Vif* Facteur d'infection des virions
*VIH-1* Virus de l'immunodéficience humaine -1
*Vpr* Protéine virale R
*VSV* Virus de la stomatite vésiculeuse
*WT* Type sauvage

# Partie I : La thérapie génique dans le domaine du cancer

## I Généralités sur le cancer

### I.1 Processus de cancérisation

Le cancer est aujourd'hui la deuxième cause de mortalité dans les pays occidentaux après les maladies cardiovasculaires et constitue un enjeu majeur de santé publique[1]. Le cancer résulte de l'acquisition de mutations conduisant à une prolifération anormale de la cellule au sein du tissu sain. Cette accumulation de mutation entraine une dérégulation du cycle cellulaire, ce qui conduit notamment à une croissance incontrôlée des cellules, à la perte d'inhibition de contact et à la perte de la capacité de différenciation. Une seule cellule transformée est à l'origine de la masse tumorale (figure 1).

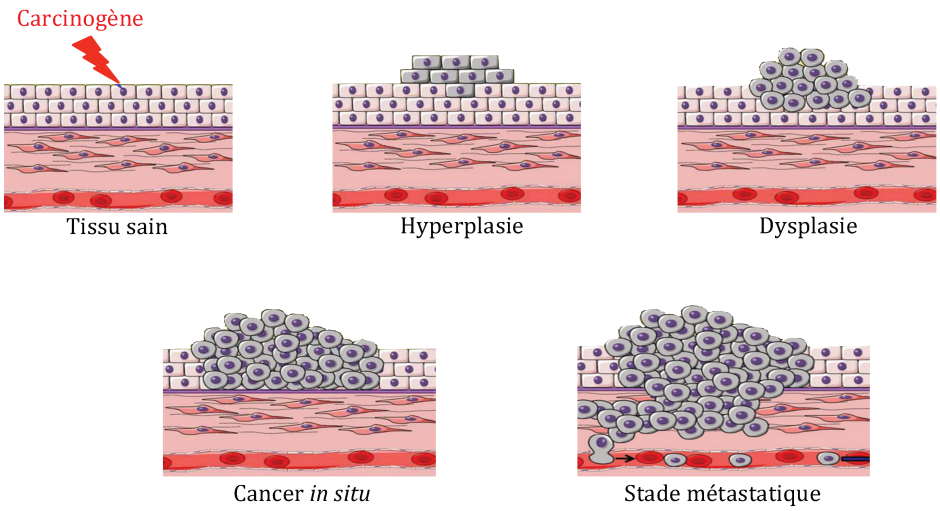



**Figure 1 : Les différents stades de développement du cancer [2].**

Si la masse tumorale est initialement clonale, les cellules filles commencent rapidement à diverger de la cellule initialement mutée. Au début du processus, la divergence entre les cellules de la masse tumorale est relativement faible et on parle d'hyperplasie. Lorsque la divergence devient plus grande on parle alors de dysplasie. S'en suit la phase de cancer in situ, au cours de laquelle la masse continue de croître tout en restant localisée au niveau du tissu où elle a pris naissance. Puis vient la phase de métastase, lorsque certaines cellules de la masse tumorale se détachent et entrent dans la circulation sanguine, conduisant à des sites secondaires de cancer dans des tissus plus ou moins éloignés du site primaire. Le

processus de développement du cancer est ainsi progressif et complexe. En fonction du tissu où se localise le tumeur, les caractéristiques cellulaires peuvent être différentes. Toutefois, la plus part des types de cancers possèdent des caractéristiques communes (figure 2).

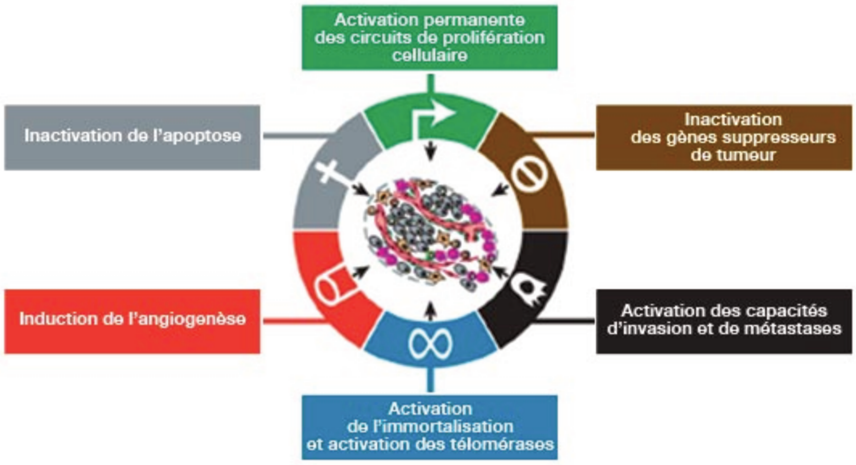



**Figure 2 : Les caractéristiques communes aux différents types de cancer [3].**

Ces caractéristiques ont été définies afin d'avoir une vision plus simplifiée de ce processus complexe et sont au nombre de 6[2] :

i. La prolifération de manière incontrôlée. Un point commun à toutes les cellules cancéreuses est qu'elles possèdent des mutations dans des gènes jouant un rôle clé dans la stimulation de la division cellulaire et ces gènes font le plus souvent partie de la famille des proto-oncogènes[3]. Lorsqu'un proto-oncogène est muté il est alors appelé oncogène et les oncogènes présentent une activité accrue comparée aux gènes sauvages. L'expression des oncogènes induit une forte stimulation de la croissance cellulaire via la production de facteurs de croissance et des récepteurs qui leur sont associés.

ii. L'échappement aux signaux inhibiteurs de croissance. Dans la très grande majorité des cancers, il a également été mis en évidence des mutations dans les gènes suppresseurs de tumeur[4]. Comme l'indique leur nom, ces gènes codent pour des protéines qui régulent négativement la croissance cellulaire, et les mutations qui leur sont associées visent principalement à inhiber leur expression.

iii. L'échappement aux signaux de mort cellulaire. L'apoptose est un processus naturel qui contribue à l'élimination des cellules et qui intervient tout au long du cycle de développement et de la vie de l'individu. Ce processus intervient également au cours du développement de la masse tumorale et vise à éliminer les cellules modifiées afin de prévenir le développement du cancer. Le dérèglement du cycle cellulaire en jeu dans les cellules cancéreuses devrait entraîner une activation de la voie apoptotique dans celles-ci, mais des mutations affectant les gènes impliqués dans le mécanisme d'apoptose permettent aux cellules cancéreuses d'échapper à ce système de contrôle[5,6].

iv. La prolifération de manière illimitée. Les cellules saines réalisent normalement un nombre de division limitée, en partie, du fait de la taille des télomères qui se raccourcissent après chaque division. Les télomères protègent les extrémités chromosomiques et sont dégradés à chaque division cellulaire jusqu'à atteindre une longueur minimale, ce qui conduit à l'activation de mécanismes de mort cellulaire programmée. Dans le cas des cellules cancéreuses cependant, l'activité télomérase est réactivée, ce qui permet de maintenir une longueur suffisante des télomères grâce à une activité accrue de la télomérase, évitant ainsi l'activation des signaux de mort cellulaire[7].

v. La capacité angiogénique. Comme pour les cellules saines, le développement des cellules cancéreuses nécessite un apport en nutriments et en oxygène qui est réalisé à travers les vaisseaux sanguins. Il a été mis en évidence que dès le début du processus de cancérisation, les mécanismes permettant l'angiogenèse, à savoir la croissance de nouveaux vaisseaux sanguins, sont déjà mis en place[8].

vi. La formation des métastases. Il semble que la capacité des cellules cancéreuses à former des métastases est un processus lent et progressif qui se met en place de manière précoce dans le développement de la tumeur[9]. L'analyse de nombreuses cellules métastatiques indique en effet que dès l'apparition des premières cellules cancéreuses, la masse cancéreuse tend à faire des métastases, étape finale du développement cancéreux.

## I.2 Microenvironnement tumoral

L'acquisition des 6 caractéristiques précédemment citées se fait via l'introduction de différentes mutations au début du processus de cancérisation. Leur maintien et le

développement de la masse tumorale impliquent cependant de nombreux autres facteurs et types cellulaires à proximité du site développement de la tumeur, ce qui crée un microenvironnement particulier autour de la tumeur (figure 3).

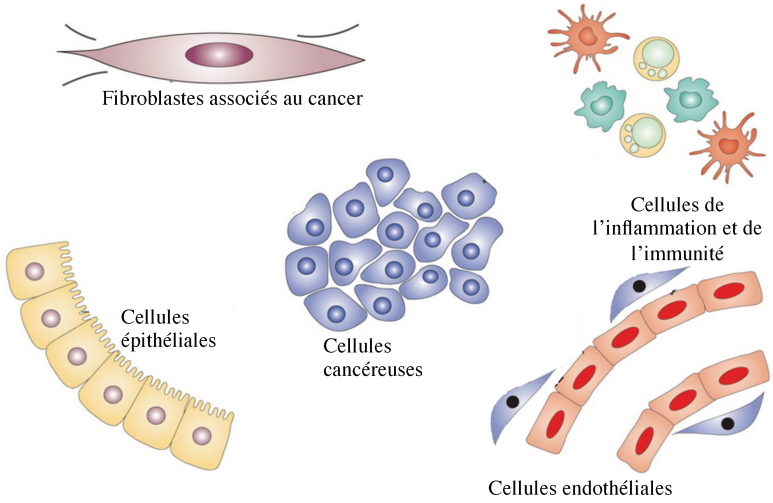



**Figure 3 : Composition du microenvironnement tumoral [10].**

Ce microenvironnement est spécifique au tissu où se situe la tumeur et prend en compte des éléments tel que les différents types cellulaires présents, les facteurs sécrétés au niveau de la matrice extracellulaire et l'interaction entre tous ces éléments. Cette notion de microenvironnement tumoral est particulièrement importante pour la mise en place de stratégies thérapeutiques. En effet, la composition de cet environnement à un impact direct sur la sensibilité des cellules aux stratégies d'immunothérapie et de thérapie génique. De manière générale, les masses tumorales sont composées, en plus des cellules cancéreuses, de :

i. Cellules souches cancéreuses. De telles cellules sont retrouvées dans une grande variété de cancers[10–13]. La notion de cellules souches cancéreuses est liée au fait que certaines cellules de la masse tumorale sont capables de se renouveler et de croître de façon à maintenir la masse tumorale comme feraient les cellules souches pour le maintien d'un organe. Ces cellules ont été caractérisées au début des années 2000 et sont considérées comme étant à l'origine de la pathologie et du développement de la tumeur.

ii. Cellules endothéliales associées à la tumeur. Ces cellules font partie de la composition des vaisseaux sanguins et participent au processus d'angiogenèse. Il semble que les facteurs sécrétés par les cellules endothéliales participent également au développement de la masse tumorale[14,15].

iii. Péricytes. Ces cellules situées au niveau de la lame basale des endothéliums participent normalement au maintien de ceux-ci via la sécrétion de divers facteurs. Dans le cadre de tumeurs, il semblerait que l'interaction entre les péricytes et les cellules endothéliales soit altérée, déstabilisant l'intégrité de la lame basale et facilitant ainsi l'évasion des cellules cancéreuses. Les péricytes participent donc au développement du cancer vers le stade métastatique[16].

iv. Cellules de l'inflammation et de l'immunité. Leur présence au niveau du microenvironnement tumoral a été mise en évidence au début des années 90. Il apparaît que dans cet environnement, ces cellules sécrètent des facteurs favorisant la croissance des cellules et participant également au processus de métastase[17,18]. L'importance de ces cellules d'un point de vue clinique est notamment étudiée dans le cas du cancer du colon par exemple.

v. Fibroblastes. Le rôle des fibroblastes au niveau du microenvironnement tumoral n'est pas totalement élucidé mais il semblerait qu'ils sécrètent des facteurs participant à la composition de la matrice extracellulaire.

## II Traitements classiques des cancers

Les traitements classiques pour le cancer sont la chirurgie, la radiothérapie et la chimiothérapie. La chirurgie et la radiothérapie sont des traitements locaux, alors que la chimiothérapie est une méthode systémique du fait du mode d'administration des composés utilisés. La chirurgie et la radiothérapie présentent certaines limites qui en revanche ne sont pas présentent dans le cas de la chimiothérapie.

### II.1 Chirurgie

La chirurgie est la première méthode décrite pour le traitement du cancer. L'utilisation de la chirurgie pour le traitement des tumeurs est déjà reportée dans des papyrus égyptiens datant de 1700 ans avant Jésus Christ. Les méthodes chirurgicales ont cependant bien évolué avec notamment l'approfondissement des connaissances sur l'anatomie du corps

humain ainsi que sur la biologie même des cancers. Le traitement du cancer par chirurgie consiste principalement à éliminer tout ou une partie de la masse tumorale avec des protocoles cliniques bien définis. Aujourd'hui la chirurgie est aussi bien utilisée dans le cadre d'un traitement curatif que dans le cadre d'un traitement palliatif. Suite à l'ablation de la masse tumorale, il se peut cependant que certaines cellules cancéreuses ne soient pas éliminées. Ainsi, pour limiter les risques de récidives ou de métastases, la radiothérapie ou la chimiothérapie sont classiquement utilisées en complément de la chirurgie.

## II.2 Radiothérapie

Le but de la radiothérapie est d'induire la mort cellulaire en causant des dommages au niveau de l'ADN via l'utilisation de rayons ionisants de forte énergie tel que les rayons X ou les protons. L'application des radiations peut conduire à des dommages directs sur les brins d'ADN ou à la formation de radicaux libres qui eux, génèreront les dommages. Il existe deux principales méthodes d'application de la radiothérapie : une application externe, et une stratégie de radiothérapie interne. Cette dernière stratégie consiste à utiliser des composés radioactifs qui sont mis en contact direct de la masse tumorale sous forme de billes ou de fil introduit dans le corps du patient. La radiothérapie impacte directement la réplication cellulaire et vise principalement les cellules à division rapide. Dans une grande partie des cas, les cellules cancéreuses ont un métabolisme cellulaire plus rapide que la plus part des cellules saines, elles seront donc prioritairement touchées par le traitement. Tout comme la chirurgie, les stratégies de radiothérapie ont considérablement évolués afin d'éliminer au plus les effets secondaires et d'améliorer l'efficacité du traitement. Ces évolutions concernent principalement la nature des rayons utilisés ainsi que le développement d'appareil de délivrance de ces rayons.

## II.3 Chimiothérapie

Le but de la radiothérapie est d'induire la mort cellulaire en causant des dommages au niveau de l'ADN via l'utilisation de rayons ionisants de forte énergie tel que les rayons X ou les protons. L'application des radiations peut conduire à des dommages directs sur les brins d'ADN ou à la formation de radicaux libres qui eux, génèreront les dommages. Il existe deux principales méthodes d'application de la radiothérapie : une application externe, et une stratégie de radiothérapie interne. Cette dernière stratégie consiste à utiliser des composés radioactifs qui sont mis en contact direct de la masse tumorale sous forme de billes ou de fil

introduit dans le corps du patient. La radiothérapie impacte directement la réplication cellulaire et vise principalement les cellules à division rapide. Dans une grande partie des cas, les cellules cancéreuses ont un métabolisme cellulaire plus rapide que la plus part des cellules saines, elles seront donc prioritairement touchées par le traitement. Tout comme la chirurgie, les stratégies de radiothérapie ont considérablement évolués afin d'éliminer au plus les effets secondaires et d'améliorer l'efficacité du traitement. Ces évolutions concernent principalement la nature des rayons utilisés ainsi que le développement d'appareil de délivrance de ces rayons.

## III La thérapie génique dans le domaine du cancer

En complément des trois stratégies classiques de traitement du cancer décrites précédemment, une nouvelle stratégie de traitement des cancers a émergé au cours des dernières années : la thérapie génique. Le développement de cancer étant lié à une accumulation de mutations génétiques, c'est tout naturellement que les chercheurs se sont intéressés à la thérapie génique dans le cadre du traitement du cancer. La principale limite de cette stratégie est que les mutations à l'origine du cancer touchent de multiples gènes ce qui rend La correction du défaut difficile dans ces conditions. Dans le cadre du traitement du cancer, les stratégies de thérapie génique vont donc principalement viser à insérer un gène permettant de sensibiliser et d'éliminer les cellules cancéreuses de manière spécifique.

### III.1 Généralités sur la thérapie génique

La thérapie génique est une méthode thérapeutique qui utilise des acides nucléiques recombinant (ADN ou ARN) afin de remplacer ou de réguler des gènes pour prévenir ou traiter les maladies[19]. La notion de thérapie génique commence à émerger au cours des années 1960. A cette époque, les recherches menées ont permis d'avoir de meilleures connaissances dans le domaine de la génétique et notamment sur les mécanismes d'expression des gènes ainsi que sur les défauts génétiques liés au développement de certaines pathologies. La stratégie de thérapie génique devient également envisageable grâce aux études menées sur les mécanismes de transfert de gènes entre les organismes.

En 1928 Frederick Griffith, un biologiste britannique, met pour la première fois en évidence le fait qu'un agent biologique est responsable de la transformation d'une souche sensible de la bactérie pneumocoque en une souche résistante de la bactérie[20]. Cette transformation

fait intervenir un transfert de matériel génétique, avec une expression stable de celui-ci permettant l'acquisition de nouveau phénotype pour la bactérie transformée. En 1932-1933 sans avoir de plus amples informations sur le matériel génétique à l'origine de ce transfert de fonction entre les organismes, le médecin J. Lionel Alloway caractérise la capacité de transfert par le principe de la transformation[21,22]. Ce n'est qu'en 1944 que Avery et *al.* Démontrent par des études physico-chimiques que l'élément permettant la transformation bactérienne est un acide désoxyribonucléique : l'ADN[23]. Commence alors une longue série d'études sur la caractérisation de l'ADN, support de l'information génétique, ainsi que sur les mécanismes de transfert de cette information génétique d'un organisme à l'autre. Différents mécanismes de transfert sont identifiés dans les années qui suivent, et en 1952, est introduit le terme de « transduction » [24]. Ce terme désigne un transfert d'ADN entre deux bactéries par l'intermédiaire d'un virus. Il est alors admis que les virus pourraient être des vecteurs pour le transfert de gènes et la notion de thérapie génique commence à émerger. L'idée que le transfert d'ADN puisse être une approche thérapeutique pour le traitement de nombreuses maladies d'origine génétique commence alors à être acceptée par la communauté scientifique de l'époque et les virus sont alors étudiés pour leur capacité naturelle à transférer leur matériel génétique dans les cellules hôtes. La première étude de thérapie génique chez l'homme a été réalisée en 1973 et avait pour but de palier au déficit de l'activité arginase des patients en leur inoculant le virus du papillome de Shope. Ce virus ayant été caractérisé comme induisant une augmentation de l'activité arginase chez le lapin[25]. Même si cette première tentative n'a pas été concluante, l'ensemble des études réalisées par la suite a permis d'aboutir en 1990 au premier essai clinique de thérapie génique à être officiellement approuvé pour l'homme par la FDA[19].

Aujourd'hui, plus de 2400 études cliniques sont en cours dans le domaine de la thérapie génique. 64,6 % de ces études cliniques sont consacrées aux traitements du cancer[26] (figure 4) et différentes stratégies sont explorées. Ici, je ne me focaliserai que sur les deux stratégies en rapport avec les études que j'ai menées au cours de ce travail de thèse : l'immunothérapie et la thérapie génique par transfert de gènes.

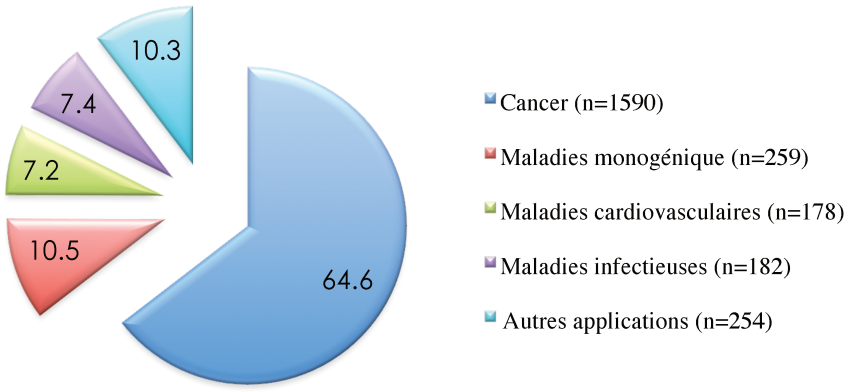



**Figure 4 : Utilisation de la thérapie génique dans les essaies cliniques**[28].

## III.2 Immunothérapie des cancers

L'immunothérapie consiste à rétablir ou à stimuler le système immunitaire en vue d'éliminer les cellules cancéreuses. Les nombreuses mutations à l'origine de la transformation des cellules engendrent, en plus du dérèglement du cycle cellulaire, la génération de nouveaux antigènes tumoraux qui peuvent être reconnus par le système immunitaire. Ces nouveaux antigènes peuvent par la suite être présentés au niveau des molécules du CMH I et induire une réponse immunitaire contre les cellules modifiées via leur reconnaissance par les lymphocytes T. Cependant, la présentation des antigènes tumoraux est perturbée au niveau du microenvironnement tumoral, ce qui conduit à une évasion immunitaire des cellules cancéreuses via l'installation d'une immunotolérance ou via une immunosuppression au niveau du site développement de la tumeur[27]. L'immunothérapie dans le domaine du cancer vise donc à restimuler les cellules immunitaires et cela peut se faire de différentes manières. Historiquement, la vaccination est l'une des premières stratégies d'immunothérapie contre le cancer à avoir été tentée[28]. Dans le domaine du cancer, la stratégie de vaccination consiste principalement à stimuler la réponse immunitaire par l'utilisation de virus oncolytique. Cette stimulation du système immunitaire passe notamment par la libération d'antigènes tumoraux suite à la lyse des cellules par les virus. Plus récemment, les recherches dans le domaine de l'immunothérapie se sont orientées vers l'utilisation et la modification des lymphocytes T.

### III.2.1 Le transfert de lymphocytes T infiltrant

L'une des premières stratégies impliquant les lymphocytes T à être appliquée est celle utilisant les lymphocytes infiltrant les tumeurs. En 1986, il a été mis en évidence qu'il est possible d'extraire les lymphocytes T infiltrés au niveau de la masse tumorale, de les mettre en expansion *in vitro*, et que le transfert de la population cellulaire obtenue permet ensuite une diminution significative de la masse tumorale. Une observation qui a notamment été faite pour différents types de cancers[34–37].

Cette stratégie se base sur le fait que certaines des cellules T ayant été en contact avec les cellules cancéreuses sont capables de reconnaître des antigènes tumoraux mais sont inhibées du fait de l'environnement immunosuppresseur du microenvironnement tumoral. La stimulation *ex vivo* permet ainsi de réactiver les cellules afin qu'elles puissent initier la réponse immunitaire. L'inconvénient de la méthode tient au fait qu'il est difficile d'isoler les lymphocytes T spécifiques à la tumeur et que l'expansion des cellules est un processus qui demande du temps. Ainsi, dans le souci d'améliorer ces stratégies de thérapie cellulaire adoptive, deux axes principaux de recherches ont émergés au cours des dernières années : la modification génétique des récepteurs des cellules T (TCR thérapie) et la génération de récepteur antigéniques chimériques (CAR-T thérapie).

### III.2.2 La TCR thérapie

De façon générale, les lymphocytes T se distinguent des autres cellules du système immunitaire par leurs récepteurs TCR (figure 5). Ces récepteurs sont un complexe de protéines permettant la reconnaissance de l'antigène présenté par les molécules du CMH, ainsi que la transduction du signal pour l'activation et la prolifération des cellules. La partie du TCR responsable de la reconnaissance de l'antigène est un hétérodimère constitué d'une sous unité α et d'une sous unité β.

La TCR thérapie consiste dans un premier temps à identifier les récepteurs de lymphocytes T spécifique pour l'antigène tumoral et à les isoler. Après le clonage des séquences codantes pour ces récepteurs spécifiques dans un vecteur d'expression, des lymphocytes T naïfs sont transduits et mis en expansion de manière à générer une population capable de reconnaître spécifiquement l'antigène tumoral[33]. Les principales difficultés de cette stratégie sont d'une part le choix de l'antigène à cibler, et d'autre part l'identification de la séquence des TCR spécifique à ces antigènes. Concernant le choix de l'antigène, les recherches s'orientent depuis de nombreuses années vers l'identification de bio-marqueurs tumoraux, qui permettraient d'orienter spécifiquement la réponse vers les cellules tumorales. Les

différents marqueurs identifiés au cours des dernières années sont pour la plus part des protéines également exprimées à la surface des cellules saines, ce qui ne permet pas toujours une discrimination suffisante des cellules saines et des cellules cancéreuses. Concernant l'identification des séquences de TCR, différents outils ont été développés pour permettre une identification plus rapide des TCR[34,35] mais le processus demande tout de même du temps.

Les premiers essais cliniques réalisés avec cette technique semblent encourageants même s'il reste encore à améliorer certains aspects du système. Cela est notamment le cas en ce qui concerne l'efficacité de l'approche du fait de la présence de nombreux facteurs suppresseurs au niveau du site de la tumeur et de la persistance des cellules modifiées chez le patient[36]. La biosécurité de l'approche doit également être améliorée car des cas de réactions croisées ont été mis en évidence lors de certains essais cliniques, allant même jusqu'à la perte du patient[37,38].

### III.2.3 La thérapie CAR-T

Une autre stratégie d'immunothérapie impliquant une modification des lymphocytes T est celle utilisant les cellules CAR-T. Les cellules CAR-T sont des lymphocytes T prélevés chez le patient et modifiés afin d'exprimer des récepteurs chimériques. En 1987, des chercheurs japonais ont démontré qu'il était possible de générer des cellules T exprimant des récepteurs chimériques, dont l'une des sous unités est substituée par une chaine issue d'un anticorps[39]. Ces études sont à la base de la stratégie de thérapie CAR-T. Ainsi, les récepteurs exprimés à la surface des cellules CAR-T sont composés d'un domaine extra cellulaire constitué d'un fragment de chaîne variable lui même dérivé d'un anticorps, fusionné à un domaine transmembranaire et à un ou plusieurs domaines intracellulaires[33,40] (figure 5). Le domaine extracellulaire permet la reconnaissance d'antigènes sans besoin d'une présentation par les molécules du CMH I, comme cela est le cas pour les TCR classiques. Le fait de s'affranchir de la présentation de l'antigène permet d'une part de ne pas être limité par les différences de fond génétique entre le donneur et le receveur, et d'autre part de reconnaître des molécules de natures diverses (protéines, carbohydrates glycolipides)[40], permettant ainsi de cibler une plus grande diversité de marqueurs tumoraux. Les domaines intracellulaires des CAR-T permettent quant à eux la transmission du signal suite à la

reconnaissance de l'antigène, conduisant ainsi à l'activation, au maintien, et à la prolifération des cellules CAR-T.

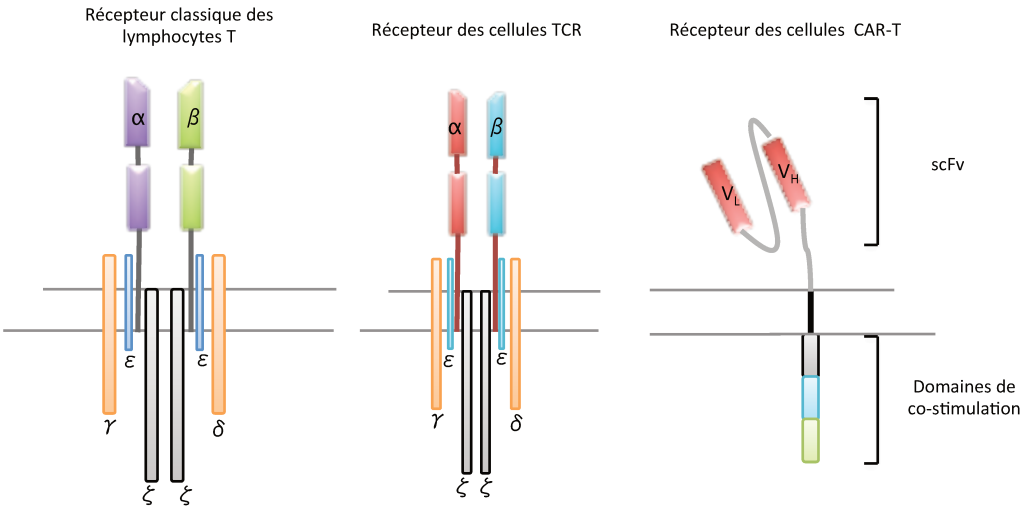



**Figure 5 : Représentation des récepteurs utilisés en immunothérapie.**

L'utilisation des cellules CAR-T a permis une meilleure prise en charge du traitement des lymphomes et des leucémies[41–43] et est actuellement en essai pour le traitement de tumeurs solides. Dans le cadre des tumeurs solides, la grande majorité des marqueurs utilisés pour le ciblage des cellules sont des protéines naturellement exprimées au niveau de différents tissus sains et qui dans le cas des cancers sont très fortement surexprimées.

Les stratégies de thérapie basées sur la modification et le transfert de lymphocytes T semblent prometteuses mais des progrès doivent encore être fait notamment en ce qui concerne la biosécurité de ces approches. Comme indiqué précédemment, les marqueurs utilisés pour le traitement des tumeurs solides par les CAR-T sont principalement des protéines également exprimées à la surface de cellules saines. Même si l'expression des marqueurs tumoraux est faible dans les cellules saines, elle est tout de même suffisante pour que les tissus sains soient également ciblé par les CAR-T, induisant des effets secondaires sévères allant jusqu'à la perte du patient[44–46]. Il faudrait donc augmenter la spécificité de reconnaissances des cellules cancéreuses. Pour ce faire, une stratégie

envisagée est de modifier les récepteurs afin que les cellules T soient activées non pas par une seule molécule tumorale mais par deux[47,48].

En attendant la mise en place et l'évaluation clinique de telles stratégies, la méthode actuellement utilisée consiste à éliminer les cellules modifiées en cas de problème. Afin de permettre cette élimination des cellules modifiées, un gène suicide est introduit en même temps que le gène permettant l'expression des récepteurs. Ce gène code pour une enzyme qui, associée à un traitement, permet l'élimination spécifique des cellules qui expriment l'enzyme. Les gènes suicides sont intéressants comme gènes de sécurité dans le cadre des stratégies de transfert de cellules modifiées et sont également d'un grand intérêt dans les stratégies de thérapies anticancéreuses par gène suicide comme agent direct de l'élimination des cellules cancéreuses.

### III.3 La thérapie par gènes suicides

Parce que les stratégies de traitement par chimiothérapie ou radiothérapie peuvent présenter certaines limites, différentes stratégies de traitement alternatives sont en cours d'études. L'une de ces pistes de recherche est l'étude et le développement de thérapie ciblée par gène suicide. Cette stratégie consiste à insérer un gène codant une enzyme capable de transformer un composé inactif non toxique (prodrogue) en composé toxique pour la cellule. La stratégie de thérapie par gène suicide repose sur deux éléments essentiels. Le premier élément est le fait que le médicament administré ne peut être activé et donc toxique que dans les cellules exprimant le gène suicide. De ce fait, si le gène est adressé et exprimé uniquement dans les cellules cancéreuses, le traitement sera spécifique pour ces cellules. Le second élément est que le choix du couple gène suicide/prodrogue repose en partie sur la forte affinité de l'enzyme pour le médicament, afin que de faibles doses de médicaments soient suffisantes pour induire la mort des cellules. Ainsi, cette stratégie vise à répondre aux problèmes liés à l'utilisation des traitements classiques.

Dans le cadre de la stratégie de thérapie génique par gène suicide, différents types de couple enzyme/prodrogue ont été étudiés. Ici je me focaliserai sur deux enzymes : la thymidine kinase du virus Herpes simplex 1 (hsTK), l'enzyme qui est actuellement la plus largement utilisée et la désoxycytidine kinase humaine (dCK), l'enzyme clé pour l'activation de certains composés utilisés en chimiothérapie sur laquelle j'ai travaillé au cours de ma thèse.

### III.3.1 La thymidine kinase du virus herpes simplex 1

Le couple hsTK/ GCV a été largement étudié en recherche clinique pour une utilisation comme gène suicide dans le domaine du cancer mais également comme gène de sécurité pour des applications de thérapie par modification des lymphocytes T. Le GCV est un analogue nucléosidique utilisé pour le traitement des infections au virus appartenant à la famille des Herpesviridae. Administré sous forme inactive, le GCV doit être tri-phosphorylé afin d'être actif (figure 6A). Sous sa forme active, il est intégré dans les brins d'ADN néo-synthétisés, bloquant ainsi la synthèse de l'ADN et conduisant à la mort cellulaire par apoptose[49]. L'étape clé de l'activation, et donc de l'efficacité du GCV, est la première phosphorylation qui est réalisée par la TK. La TK virale présente une affinité 1000 fois supérieure pour le GCV comparée aux TK humaines[50], ce qui fait de la TK virale l'enzyme clé pour l'activation de ce composé.

Au cours des dernières années, les études menées sur le système hsTK/GCV ont permis d'augmenter son efficacité via l'identification de variantes de l'enzyme possédant une meilleure affinité pour le GCV. La dernière variante de l'enzyme développée est appelée TK.007. Ces variantes dont la séquence protéique a été optimisée avec notamment des modifications visant à éliminer les sites d'épissages, permet non seulement une plus grande sensibilité des cellules transduites au GCV mais permet également d'avoir un meilleur effet « bystander »[51]. L'effet bystander est la capacité des molécules activées à induire la mort des cellules avoisinantes et non transduites (figure 6B). Cet effet est important pour les stratégies utilisant un gène suicide pour les cancers solides, car dans les protocoles actuels seulement 10% des cellules de la masse tumorale sont transduites[52]. De plus les protocoles actuels empêchent la délivrance du médicament dans toutes les cellules de la masse tumorale du fait de sa pauvre vascularisation. Ainsi, pour pouvoir éliminer la tumeur, il faut que le composé actif soit capable d'induire la mort des cellules adjacentes non transduites. La mort des cellules adjacences se fait non seulement par l'activation des cellules du système immunitaire suite à la mort des cellules transduites mais également grâce à la capacité du composé activé à passer les jonctions communicantes entres les cellules[53,54].

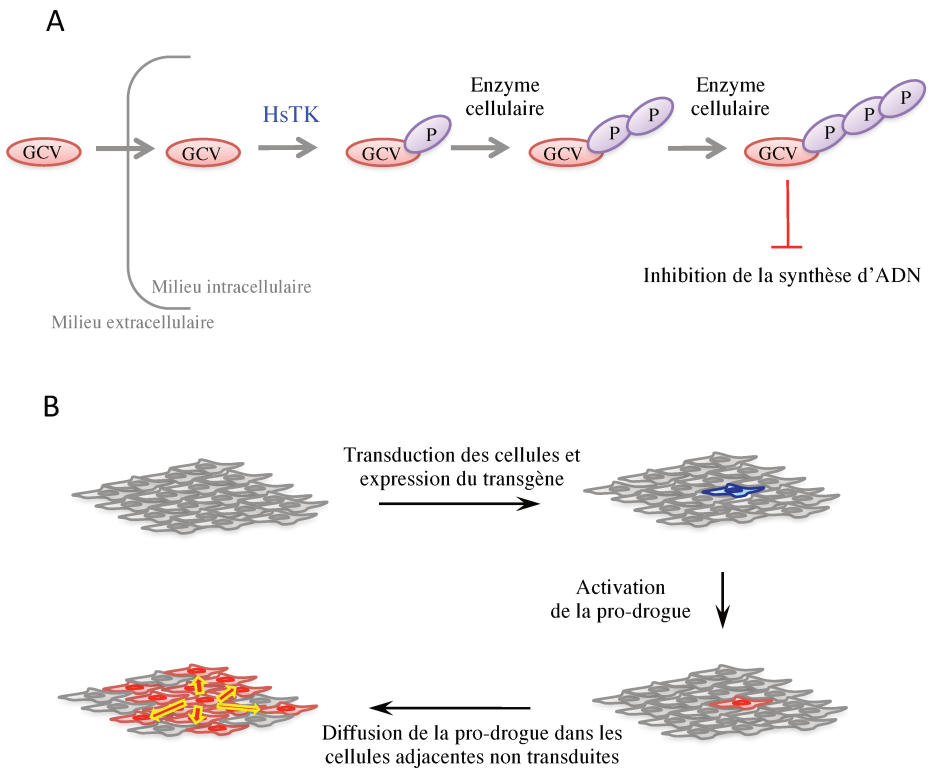



**Figure 6 : Mécanisme d'action du GCV.**

Le système hsTK/GCV a été utilisé pour le traitement du cancer de la prostate et du foie[55,56], mais il est surtout très largement utilisé pour les stratégies de thérapie génique par modifications des lymphocytes T. Dans le cadre de cette dernière application, l'origine virale de la protéine TK constitue cependant une limite de l'application. Il a en effet été montré que l'injection des cellules modifiées exprimant le hsTK conduit à une réponse immunitaire dirigée contre celles-ci, conduisant à l'élimination d'une partie des cellules[57,58] et à une diminution de l'efficacité de l'approche. Un moyen d'éviter la mise en place d'une réponse immunitaire contre les cellules modifiées serait l'utilisation de protéines d'origine humaine. A l'inverse des protéines bactériennes et virales, les protéines humaines ne devraient pas générer de réponses immunitaires contre les cellules modifiées et ainsi augmenter l'efficacité de la stratégie. Les principales protéines humaines étudiées à cet effet sont la désoxycytidine kinase humaine (dCK) et plus récemment la caspase 9. Au cours de mon travail de thèse je me suis intéressée à la caractérisation d'une variante de la dCK humaine identifiée au laboratoire grâce à la Rétrovolution[59].

### III.3.2 La désoxycytidine kinase humaine

La dCK est une protéine composée de 260 acides aminées et est active sous forme d'homodimère. Chaque monomère de dCK est composé de 5 feuillets béta et de 10 hélices alpha (figure 7).

Les dimères de dCK se forment via l'interaction des hélices α4 et α7 de chaque monomère. L'hélice α10 quant à elle contient une région appelé « *base sensing loop* » qui influe sur le repliement de la protéine et donc sur la capacité de la protéine à se fixer au donneur de phosphate. C'est une enzyme clé impliquée dans la voie de biosynthèse et de recyclage des désoxyribonucléotides nécessaires à la biosynthèse de l'ADN. Les substrats naturels de la dCK sont la désoxycytidine (dC), la désoxyadénosine (dA) ainsi que la désoxyguanosine (dG), même si l'enzyme présente avec une meilleure affinité pour la dC. La dCK a été largement étudiée au cours des dernières années car elle joue un rôle clé pour l'activation de nombreux analogues nucléosidiques utilisés en chimiothérapie et pour les thérapies antirétrovirales. Parmi les composés de chimiothérapie activés par la dCK, l'on peut citer la Gemcitabine (Gem) un analogue de la dC (figure 8) utilisé pour le traitement de divers cancers tel que le cancer du sein, du pancréas ou encore des ovaires[60–62].

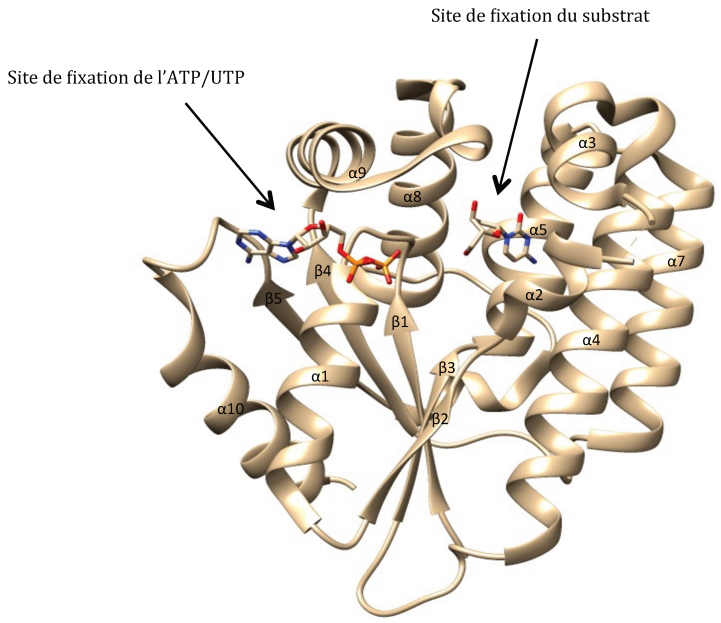



**Figure 7 : Structure d'un monomère de dCK[67].**

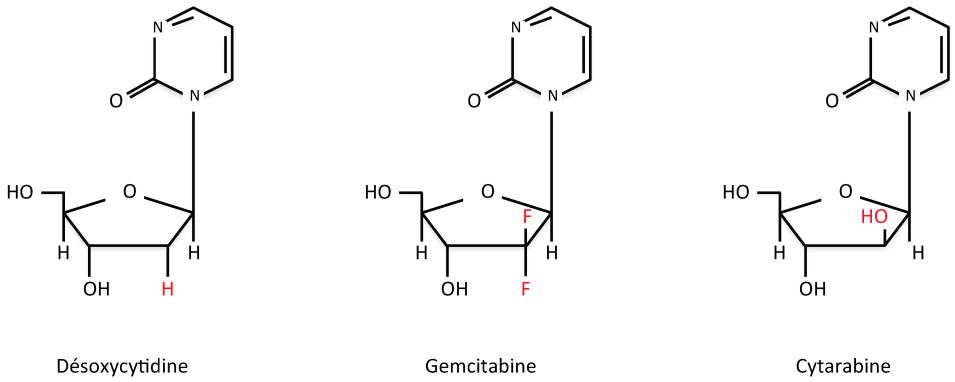



**Figure 8 : La désoxycytidine et deux de ses analogues utilisés en chimiothérapie.**

Le mécanisme d'action de la dCK est semblable à celui de le TK présenté précédemment. Les composés utilisés sont principalement des terminateurs de chaînes qui après leur insertion dans le brin d'ADN néo-synthétisé bloquent la réaction et induisent la mort cellulaire par apoptose. Comme le GCV, les composés chimiothérapeutiques sont administrés sous forme inactive et doivent être tri-phosphorylés pour être actifs (Figure 9). La première étape de phosphorylation réalisée par la dCK constitue l'étape clé de l'activation et détermine en grande partie l'efficacité des traitements. La relation entre le niveau d'expression de la protéine et la résistance ou la sensibilité des cellules au traitement a été démontrée dans différents types de cancers[63–67], ce qui a naturellement conduit à considérer la dCK comme un bon candidat gène suicide. Hormis la dCK, d'autres mécanismes de résistance aux médicaments peuvent intervenir au niveau des cellules. C'est par exemple le cas de l'augmentation de l'efflux qui permet d'éliminer les composés avant que ceux-ci ne soient activés. Il est également décrit que la ribonucléotide réductase et plus particulièrement sa sous unités 1 (RRM1) est également identifiée dans certains cas de cellules résistantes. La RRM est l'enzyme qui assure la production des désoxyribonucléotides (figure 10), ce qui fait de cette enzyme un pilier de la synthèse des nouveaux brins d'ADN. Il est montré que la Gem-diphosphate est capable de se fixer à la RRM1 inhibant ainsi l'activité de l'enzyme. Une surexpression de cette sous unité RRM1 permet ainsi aux cellules de résister à l'action de la Gem aussi bien *in vitro* qu'*in vivo*[68–71].

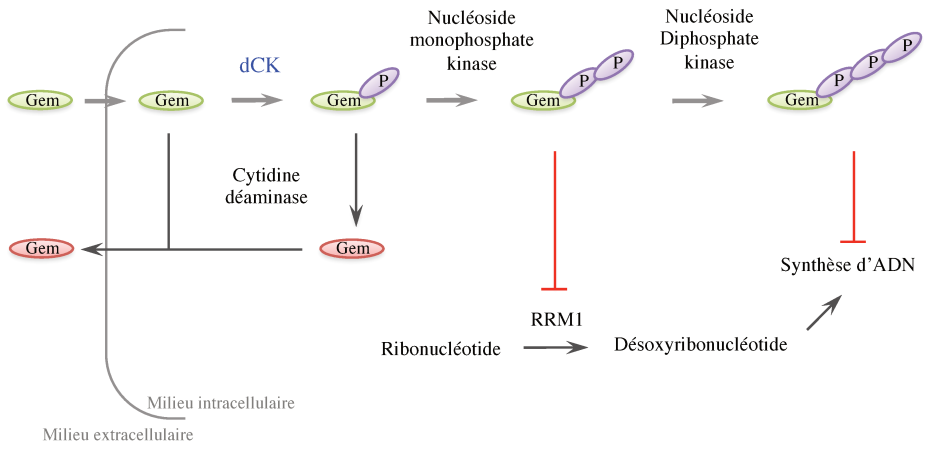



**Figure 9 : Mécanisme d'action de la Gem.**

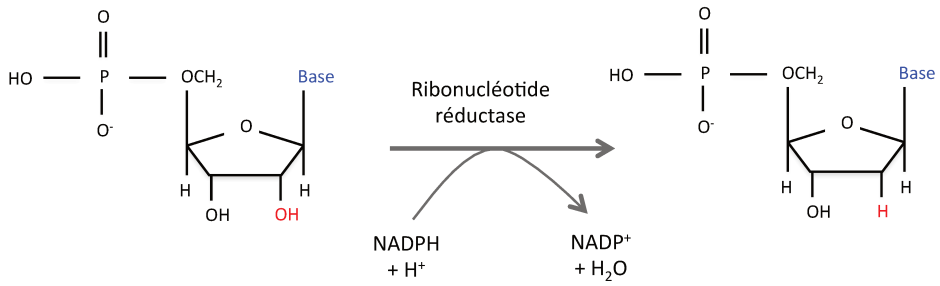



**Figure 10 : Mécanisme d'action de la ribonucléotide réductase.**

Les études ont également démontré que, comme dans le cas du GCV, le traitement avec la Gem permet un effet « bystander » qui tel qu'indiqué dans le chapitre précédent est un critère important pour l'utilisation d'un médicament en thérapie par gène suicide[72,73].

Les premières études menées sur une utilisation possible de la dCK comme gène suicide ont mis en évidence que l'une des limites du système est la plus grande affinité de l'enzyme pour son substrat naturel comparé au médicament. Cela a donc conduit différents laboratoires à rechercher des variantes de l'enzyme présentant une meilleure capacité à activer les médicaments[74–78].

### III.3.3 Nouvelle stratégie d'évolution moléculaire : la rétrovolution

Afin d'identifier un variant de la dCK présentant une meilleure affinité pour les médicaments que pour la dC, le laboratoire a utilisé une méthode d'évolution mise en place au sein même

du laboratoire et appelé « rétrovolution »[59]. Appliquée à la dCK, la rétrovoluion a permis l'identification de deux variants de la dCK capable de sensibiliser les cellules résistantes à de très faibles concentrations de Gem.

La Rétrovolution est une méthode d'évolution moléculaire basée sur la capacité de la polymérase du VIH-1 à introduire des mutations au cours du cycle réplicatif. Cette stratégie consiste à introduire la séquence du gène à faire évoluer dans un vecteur lentiviral à réplication conditionnel, et à mimer plusieurs cycles infectieux[59] (figure11).

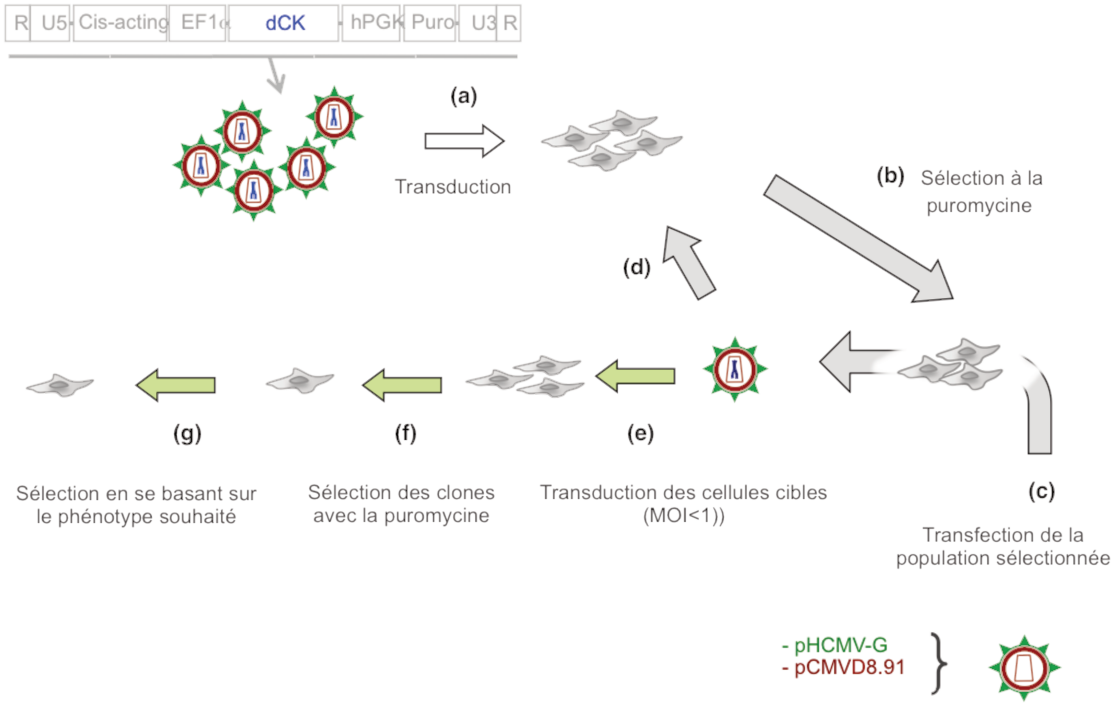
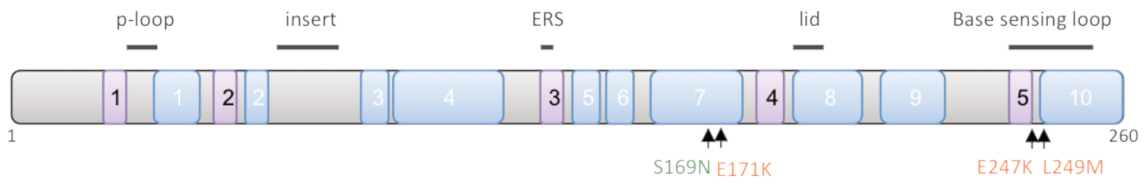



**Figure 11 : Représentation schématique du système de rétrovolution[63].** Le gène de la dCK est inséré dans un vecteur lentiviral défectif pour la réplication et produit par la transfection des cellules HEK 293T. (a) Les vecteurs produits sont utilisés pour transduire de nouvelles cellules HEK 293T. (b) Les cellules transduitent et donc ayant intégré le génome viral sont par la suite sélectionnées pour leur résistance à la puromycine dont le gène de résistance est présent dans le génome viral. (c) Les cellules précédemment sélectionnées sont à leur tour transfectées avec d'une part le plasmide permettant l'expression des enzymes et protéines structurales du VIH-1 (en rouge) et d'autre part le plasmide permettant l'expression de la protéine d'enveloppe du VSV (en vert). Cette étape de transfection conduit à la production de nouveaux vecteurs contenant la séquence précédemment intégrée. (d) Ce cycle est répété afin de permettre l'accumulation de mutation dans la séquence de la dCK et conduit à la génération d'une banque de mutant. (e) La banque est ensuite analysée en transduisant les cellules Messa 10k résistantes à la Gem. (f)Les cellules transduitent sont sélectionnées par la puromycine et (g) les clones d'intérêt sont ensuite sélectionnés sur leur capacité à sensibiliser les cellules transduites à de faibles concentrations de Gem.

A chaque cycle, le génome viral contenant la séquence du gène à faire évoluer est soumis au processus de transcription inverse afin de pouvoir être intégré dans le génome cellulaire et permettre la production de nouvelles particules virales. L'étape de transcription inverse du génome viral est réalisée par la polymérase virale qui au cours de chaque cycle de réplication introduit des mutations dans l'ADN proviral générant ainsi une banque de mutants pouvant être directement testée dans les cellules pour identifier le phénotype souhaité[59]. Appliqué à la dCK, la rétrovolution a ainsi permis de générer de nombreux mutants de la dCK qui ont été testés pour leur capacité à sensibiliser des cellules résistantes à de faibles concentrations de Gem. Parmi les mutants obtenus, le mutant G12 est celui qui présentait la meilleure capacité à sensibiliser les cellules à la Gem. Le G12 se caractérise par 3 mutations comparé à la dCK sauvage (figure 12). Les analyses biochimiques réalisées sur le mutant G12 ont démontré que celui-ci a perdu la capacité à phosphoryler son substrat naturel la dC conduisant ainsi à une meilleure activation de la Gem dans les cellules[59].

La localisation des trois mutations caractéristiques du G12 au niveau de la structure de la protéine montre qu'elles se situent dans des régions éloignées du site actif. De ce fait, elles sont difficilement prédictibles sur une base rationnelle, ce qui constitue un atout majeur du processus de rétrovolution. Le mutant G12 a été identifié suite à 17 cycles de rétrovolution. Afin de pouvoir isoler un autre mutant qui serait plus efficace que le G12, celui-ci a également été soumis à la rétrovolution, et après 11 cycles de rétrovolution supplémentaires, les tests de sensibilités réalisés ont permis l'identification d'un autre mutant appelé M36. Comparé au G12 ce mutant se caractérise par une mutation supplémentaire en position 169 (figure 12).




**<u>Figure 12 : Représentation schématique de la dCK et localisation des mutations du G12 et du M36.</u>** Le schéma représente la séquence nucléotidique de la dCK avec indiqué en bleu les régions organisées en feuillet β et en violet les régions organisées en hélice α. Sont également indiqués les régions clés pour l'activité de la protéine *: p-loop* : intéragit avec les groupes phosphoryles α et β du donneur de phosphate ; *ERS* : interagit avec l'ion magnésium ; *lid* : participe à la fixation du donneur de phosphate et à sa catalyse. Les trois mutations caractéristiques du mutant G12 sont indiquées en rouge et la mutation supplémentaire du M36 est indiquée en vert.

## Partie II : Les outils du transfert de gène

En thérapie génique, les vecteurs sont les véhicules du gène thérapeutique dans les cellules cibles. Ils doivent permettre de délivrer le transgène au niveau des cellules et conduire à l'expression de celui-ci en vue d'atteindre le but thérapeutique. Le vecteur idéal se définit en fonction du type cellulaire que l'on souhaite atteindre, du type d'expression souhaité (stable ou transitoire) ainsi que de la taille de la séquence à insérer. Classiquement les vecteurs sont classés en deux groupes : les vecteurs viraux et les vecteurs non viraux. Les vecteurs viraux sont ceux largement utilisés dans les essais cliniques du fait de l'efficacité naturelle des virus à transférer leur génome dans les cellules. Les virus sont cependant des pathogènes de l'homme qui détournent la machinerie cellulaire afin de permettre leur réplication. Leur utilisation peut donc conduire à des problèmes de toxicité et d'immunogénicité. Le premier incident ayant conduit à la perte d'un patient au cours des stratégies de thérapie génique était notamment lié à une réaction inflammatoire contre le vecteur adénoviral utilisé pour le traitement[79]. Les essais cliniques réalisés au cours des années ont également mis en évidence le risque de mutations insertionnelles liées à l'utilisation de vecteurs intégratif pour une expression stable du transgène. Ce problème a été mis en évidence au cours d'études réalisées avec les rétrovirus dans le cadre du traitement de l'immunodéficience combinées sévères[80,81]. Ces rapports d'effets secondaires liés à l'utilisation des vecteurs viraux poussent à rechercher des alternatives à l'utilisation des virus comme vecteurs. De nombreux laboratoires s'intéressent ainsi au développement de vecteurs non viraux. Le faible coût de production des vecteurs non viraux ainsi que leur biosécurité relative en fait des outils attractifs. Ces vecteurs ont cependant une très faible capacité à transfecter les cellules *in vivo*, conduisant ainsi à une expression insuffisante et transitoire du transgène[82].

### I Les vecteurs non viraux

Les vecteurs non viraux sont classiquement de l'ADN nu complexé avec des liposomes, des polymères ou des protéines afin de faciliter l'entrée de l'ADN dans la cellule[83]. Il existe deux principaux systèmes pour la délivrance des vecteurs non viraux : la méthode physique et la méthode utilisant des transporteurs chimiques.

#### I.1 Le transfert physique

La méthode physique de transfert consiste à utiliser une force physique pour faciliter la délivrance de l'ADN à travers les membranes cellulaire. Les forces utilisées pour l'entrée peuvent être de différentes natures, magnétiques, électriques, ou encore hydrodynamiques. Ici je ne présenterai que brièvement les méthodes électriques et hydrodynamique.

-Electroporation : la méthode consiste à appliquer un champ électrique de sorte à générer un pore à l'endroit de la cellule où le champ électrique est appliqué. L'ADN peut alors passer la membrane cellulaire à travers ce pore qui se forme pour un court laps de temps. La perméabilité de la membrane dépend de la puissance du champ électrique et de la durée d'exposition, et la génération du pore doit être réversible afin de garder les cellules en vie. L'électroporation permet de délivrer le transgène par une application intradermique, intramusculaire ou intra tumorale[84].

-Hydroporation : cette méthode consiste à appliquer une forte pression hydrodynamique via l'injection de grand volume de solution d'ADN en un laps de temps court. Cela permet d'augmenter la perméabilité des cellules endothéliales situées au niveau des capillaires conduisant à l'entrée de l'ADN exogène dans les cellules. Cette technique est notamment utilisée dans la recherche sur les cellules hépatiques [85].

### I.2 Le transfert par transporteurs chimiques

Le transfert de gène par l'intermédiaire de transporteurs chimiques est plus courant que la méthode de transfert physique[86]. Les transporteurs chimiques peuvent être classés en deux grands groupes : les lipoplexes et les polyplexes. Les lipoplexes sont issus de la compaction entre les acides nucléiques et les lipides ou micelles (figure13 A), alors que les polyplexes, sont issus de la compaction entre les acides nucléiques et des polymères (naturelles ou synthétiques) (figure13 B).

De nombreux systèmes de lipoplexes ont été décrits, mais ils possèdent tous la même structure de base à savoir:

- une tête hydrophile qui permet l'interaction avec les groupements phosphate des acides nucléiques,
- une queue hydrophobe qui permet la formation de la particule.

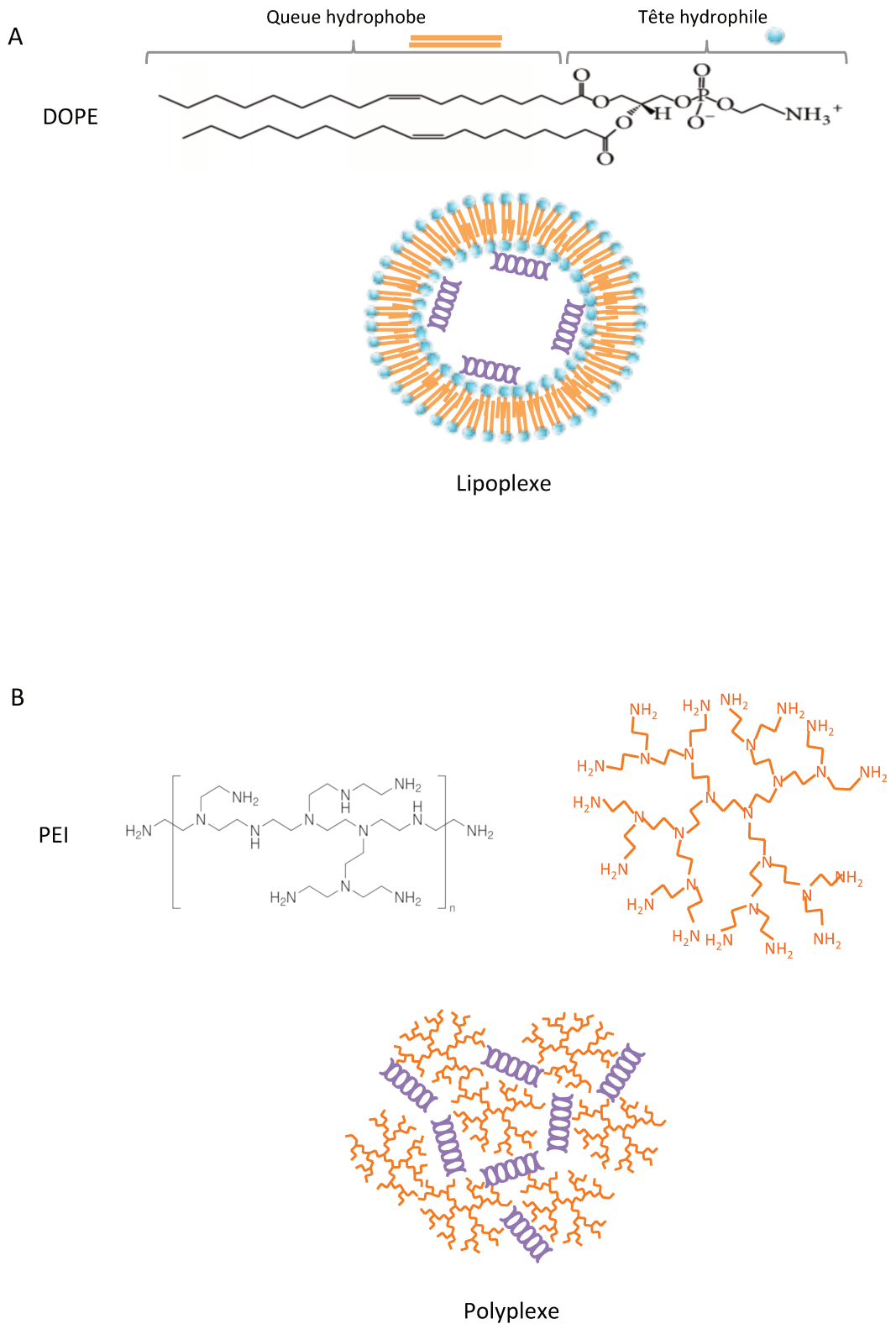



**Figure 13 : Représentation des lipoplexes et polyplexes utilisés comme vecteurs non viraux.**

Les polyplexes sont quant à eux plus stables que les lipoplexes et sont classés en deux catégories dépendamment de la nature du polymère : naturel (protéine, polysaccharides ou peptides) ou synthétique (PEI, PLL)[87].

Le faible risque de déclencher une réponse immunitaire avec les vecteurs non viraux en fait une alternative plus sure à l'utilisation des vecteurs viraux. Cependant, de nombreux progrès doivent encore être faits pour obtenir une bonne efficacité de transfection et d'expression du transgène. En effet, les vecteurs non viraux présentent une faible capacité à transférer le transgène entre autre dû à leur dégradation suite à leur administration ou encore à leur faible capacité à franchir les barrières cellulaires[88]. De ce fait, peu de vecteurs parviennent à atteindre les cellules cibles. Afin de permettre un ciblage spécifique, certains laboratoires s'intéressent à la possibilité de greffer à la surface des particules différents ligands qui permettraient une interaction spécifique de ces vecteurs avec les cellules cibles[89–94]. De même, les recherches continuent sur la possibilité d'avoir une intégration des plasmides permettant ainsi une expression stable du transgène [95,96].

## II Les vecteurs viraux

La grande majorité des vecteurs utilisés en thérapie génique sont des vecteurs dérivés des virus (FIGURE 14).

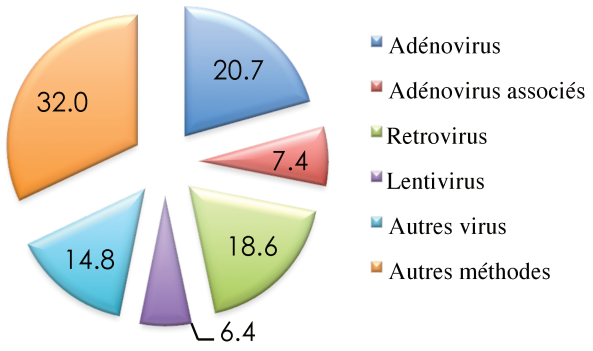



**Figure 14 : Les vecteurs utilisés en thérapie génique[19].**

L'évolution et la pression de sélection ont poussé les virus à optimiser leur capacité d'entrée dans les cellules cibles afin de répliquer leur matériel génétique, ce qui en fait un outil de choix pour le transfert de gènes dans les cellules. Etant des pathogènes de l'homme, il existe cependant pour la plupart des virus une immunité préexistante conduisant à une élimination prématurée des vecteurs viraux et donc à une diminution de l'efficacité de la stratégie.

L'intégration de certains vecteurs dans le génome cellulaire peut également conduire à l'introduction de mutations délétère pour les cellules ou conduire à une perturbation du cycle cellulaire. Parmi les virus majoritairement utilisés comme vecteurs, on compte les adénovirus, les adénovirus associés ainsi que les rétrovirus. C'est à cette dernière famille de virus que je m'intéresse et plus particulièrement aux lentivirus.

Dans cette partie, je décrirai de façon succincte les vecteurs adénoviraux et adéno-associés, et de façon plus détaillée les vecteurs lentiviraux.

## II.1 Les adénovirus

### II.1.1 Cycle de réplication

Les adénovirus (Ad) ont été découverts en 1953 par Wallace P Rowe[97]. Les virus de la famille des *Adénoviridae* sont des virus non enveloppés à capside icosaédrique et à ADN double brin linéaire (figure15). Ils sont associés à des maladies respiratoires aigües, à des conjonctivites ou encore à des gastro-entérites, mais chez la plupart des personnes immunocompétentes, l'infection à Ad reste asymptomatique.

Il existe actuellement 84 sérotypes humain référencés pour les adénovirus[98]. Pour les applications de thérapie génique, le sérotype le plus utilisé et le plus étudié est le sérotype 2 que je présenterai ici.

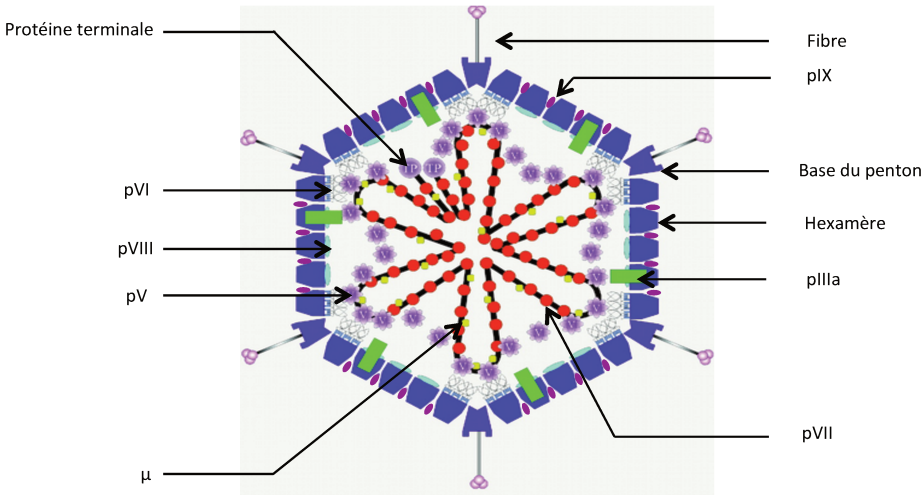



**Figure 15 : Structure de la particule virale des adénovirus[21].**

Le génome viral est flanqué de chaque côté par des séquences répétées inversées d'environ 100nt appelées ITR pour *inverted terminal repeat* à partir des quels se fait la réplication virale, et qui contiennent les séquences nécessaire à l'encapsidation. Ces régions répétées inversées sont protégées à leurs extrémités par les protéines terminales qui via leur interaction, font apparaître le génome viral comme circulaire. Le génome est très compacté avec la présence de beaucoup de chevauchements, peu de régions non-codantes, et peut être divisé en deux unités de transcription principales (figure 16). L'annotation des gènes est basée sur la phase au cours de laquelle ceux-ci sont exprimés. Les gènes exprimés au cours de la phase précoce de l'infection sont annotés de E1 à E4 et codent pour des protéines de même nom impliquées dans le contrôle du cycle cellulaire et dans la réplication de l'ADN viral. La protéine E1A induit la mitose des cellules et stimule l'expression des autres gènes. La protéine E2 quant à elle permet la réplication virale alors que les protéines E3 et E4 participent à l'immuno-évasion virale. Les gènes exprimés au cours de la phase tardive de l'infection sont annotés de L1 à L5 et codent quant à eux pour les protéines structurelles. En plus des unités de transcription E et L, les Ad codent également pour 4 protéines exprimées au cours de la phase intermédiaire. Parmi ces protéines exprimées au cours de la phase intermédiaire, la protéine pIX est impliquée dans la formation de la capside. Les autres protéines participent à l'encapsidation du génome et à la maturation des nouvelles particules virales.

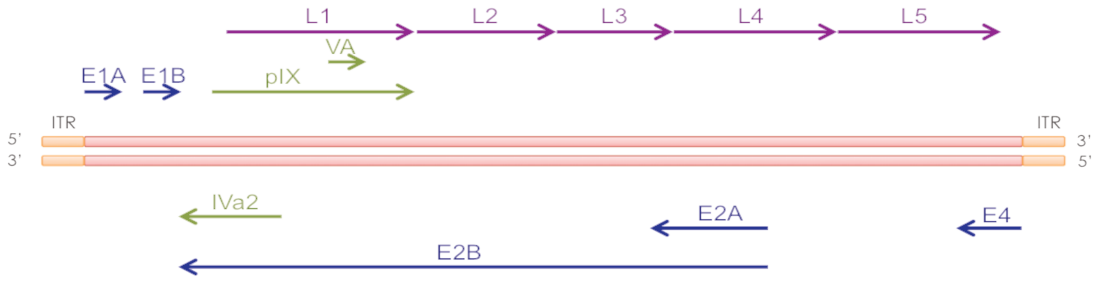



**Figure 16 : Représentation schématique du génome des adénovirus.**

L'entrée du virus dans la cellule se fait par une première interaction entre les protéines de la capside et les récepteurs cellulaires. Un exemple de tel récepteur est la protéine CAR (*Coxackie Adénovirus Receptor*) qui est une protéine d'adhésion cellulaire. Cette interaction

conduit à une invagination de la membrane cellulaire, permettant l'entrée des virus par endocytose. Suite à l'acidification de l'endosome, le virus est libéré dans le cytoplasme et l'ADN transféré au noyau pour l'expression des gènes et la réplication. Le tropisme naturel du virus est relativement large mais des modifications des protéines de capside et notamment de la protéine pIX pour des finalités biotechnologiques peuvent permettre une entrée spécifique des vecteurs dans les cellules cibles[99,100].

### II.1.2 Utilisation en thérapie génique

Un des avantages majeurs à l'utilisation des Adénovirus comme vecteur est la capacité d'empaqueter un ADN de grande taille. Les vecteurs adénoviraux utilisés peuvent être classées en deux groupes : les vecteurs réplicatifs et les vecteurs non réplicatifs. Au cours des dernières années, de nombreuses améliorations ont été apportées aux vecteurs adénoviraux pour des raisons de biosécuritté [101]. On distingue ainsi les vecteurs de première génération qui sont délétés de la région E1 et de la région E3 ; les vecteurs de deuxième génération qui comprennent des mutations supplémentaires dans la région E2 ; ainsi que les vecteurs de troisième génération qui sont ceux actuellement utilisés, et qui sont délétés de la totalité du génome viral à l'exception des régions ITR.

L'inconvénient majeur de ces vecteurs reste l'immunité préexistante contre les virus qui conduit à leur élimination rapide. De plus, même si le patient ne possède pas d'anticorps neutralisants contre le vecteur, la première injection de celui-ci conduit à une forte réponse immunitaire, ce qui fait qu'il ne peut pas être injecté une seconde fois au risque d'être directement éliminé ou d'entrainer une réponse immunitaire et inflammatoire[79].

## II.2 Les virus adéno-associés

### II.2.1 Cycle de réplication

Les virus adéno-associés (AAV) ont été découverts en 1965 dans des préparations adénovirales[102,103]. L'AAV fait partie de la famille des *Parvovirus*, et du genre *Dependovirus*. Les AAV sont des virus non enveloppés à capside icosaédrique qui renferme un génome d'ADN simple brin d'environ 5kb. Le génome des AAV contient deux séquences codantes et est flanqué à chacune de ces extrémités par des régions répétées inversées ITR (figure 17). La première région codante appelé Rep code pour 4 protéines impliquées dans la régulation

de la réplication du génome viral. La deuxième séquence codante quant à elle Cap, code pour trois protéines impliquées dans la formation de la capside virale. Ces virus sont appelés virus adéno-associés car ils ne peuvent se répliquer sans la présence dans la cellule d'un virus *helper* à savoir un adénovirus ou encore un herpes virus[104]. Ces virus *helper* interviennent de façon différentes dans le cycle réplicatif des AAV mais permettent en tous cas la régulation du cycle cellulaire, rendant la cellule permissive à l'infection par les AAV[105]. En absence des virus *helper*, la réplication virale est réprimée de même que l'expression des gènes viraux conduisant à l'entrée du virus en phase de latence et/ou à l'intégration dans le génome de la cellule[106]. Cette intégration se fait préférentiellement dans le chromosome 19 au niveau d'un locus nommé AAVS1 et fait intervenir des facteurs cellulaires ainsi que les séquences ITR[107]. Cette intégration site spécifique ayant été observée *in vitro*, aucunes preuves ne permet pour l'instant de croire que ce mécanisme a également lieu *in vivo*. Certaines études tendent en effet à dire que l'intégration n'a pas toujours lieu au niveau du site AAVS1 et qu'elle peut également se faire dans des régions transcriptionnellement actives ce qui pourrait influer sur l'expression de certains gènes cellulaires[108].

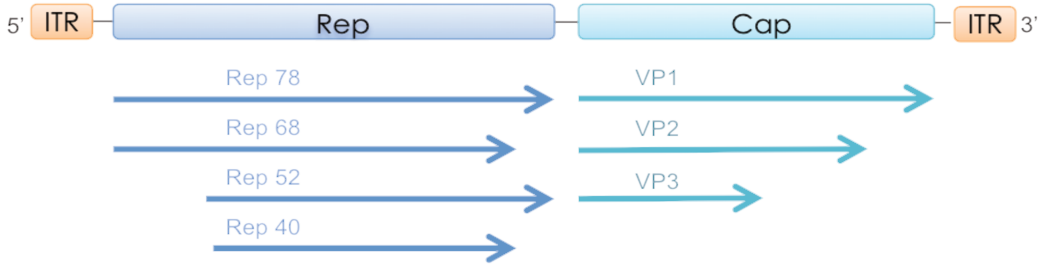



**Figure 17 : Représentation schématique du génome des AAV.**

### II.2.2 Utilisation en thérapie génique

Il existe actuellement 20 sérotypes d'AAV répertoriés et le sérotype 2 est celui qui est le plus utilisé pour les applications de thérapie génique. L'existence de ces différents sérotypes du virus possédant chacun un tropisme différent et relativement restreint à certains tissus, présente un avantage à l'utilisation de ces virus en thérapie génique. En effet, même si la structure de base des vecteurs utilisés est issue des AAV-2, le changement des protéines de capsides par celles d'un autre sérotype permet de réorienter l'entrée virale vers des tissus différents dépendamment du sérotype d'origine des protéines de capsides[109–111] (figure 18).

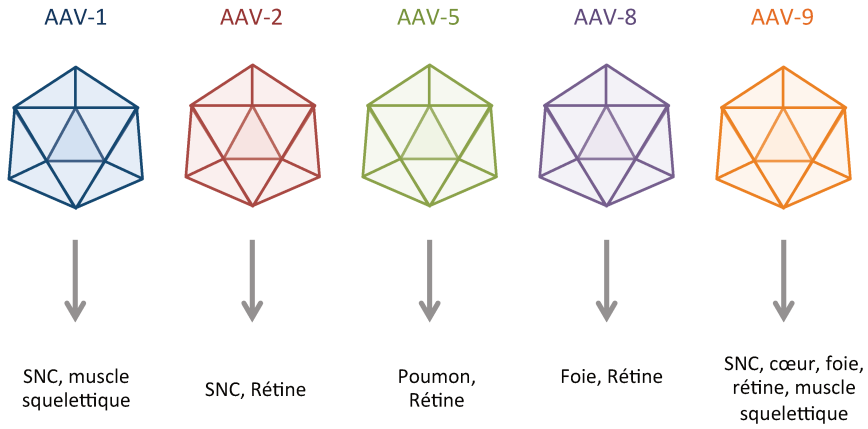



**Figure 18 : tropisme de différents sérotypes de virus adéno-associés.**

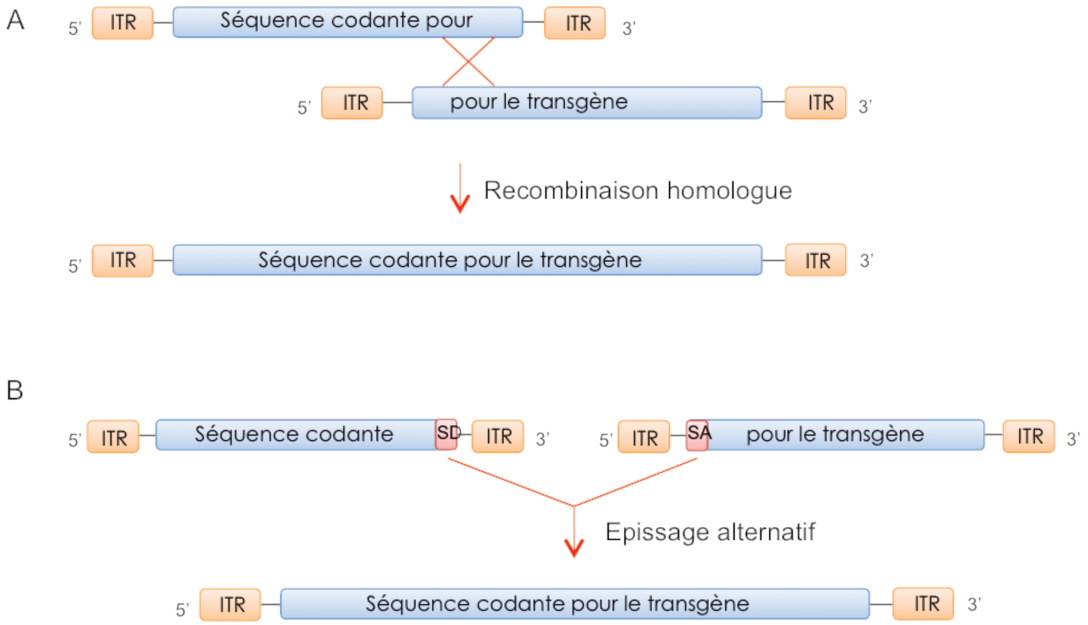



**Figure 19 : Stratégies d'expression de gène de plus de 4,5 kb par les vecteurs AAV.**

SD : site donneur de l'épissage. SA : site accepteur de l'épissage.

Enfin, la principale limite à l'utilisation des virus adéno-associés comme vecteurs est la faible taille des gènes pouvant être encapsidés. Le génome viral ne faisant que 4,5kb, la taille de la séquence pouvant être introduite entre les séquences ITR ne peut excéder 4,5kb. Afin de remédier à cette limite, certaines stratégies d'expression des gènes ont été développées afin de permettre l'expression de gènes jusqu'à 9kb[112]. Il s'agit notamment de transduire les cellules avec deux vecteurs différents exprimant chacun une région du transgène. Ces stratégies reposent notamment sur le mécanisme de recombinaison par homologie de séquence ou encore sur le mécanisme d'épissage[113] (figure 19). Les vecteurs ainsi construits ont cependant une plus faible efficacité pour l'expression du transgène que les vecteurs classiques.

## III Les vecteurs lentiviraux

Les vecteurs lentiviraux actuellement utilisés dérivent principalement du virus de l'immunodéficience humain de type 1 (VIH-1), et dans ce rapport de thèse je ne parlerai donc que de ce virus.

Les Lentivirus font partis de la famille des *Retroviridae* et de la sous famille des *Orthoretroviridae*. Cette famille regroupe des virus enveloppés à ARN positif et est caractérisée par :

- la présence d'une ADN polymérase ARN/ADN dépendante appelée la transcriptase inverse qui permet la synthèse de l'ADN proviral.
- Une organisation génomique de base, identique pour tous les Retroviridae, avec la présence de 3 gènes principaux (Gag, Pol, Env) qui codent les protéines de structure et les enzymes virales.
- La présence à chaque extrémité de l'ADN proviral de séquences répétées (LTR : *Long Terminal Repeat*) essentielles à la transcription inverse, à l'intégration de l'ADN proviral dans le génome cellulaire, à la transcription des gènes viraux et à la synthèse de l'ARN génomique.

### III.1 Organisation du génome

Le génome viral se compose d'un ARN de polarité positive qui possède une coiffe à son extrémité 5’ et est polyadénylé à son extrémité 3’. A chaque extrémité du génome est présente une séquence non codante répétée R et une séquence non codante unique U5 pour l’extrémité 5’ et U3 pour l’extrémité 3’ (figure 20).

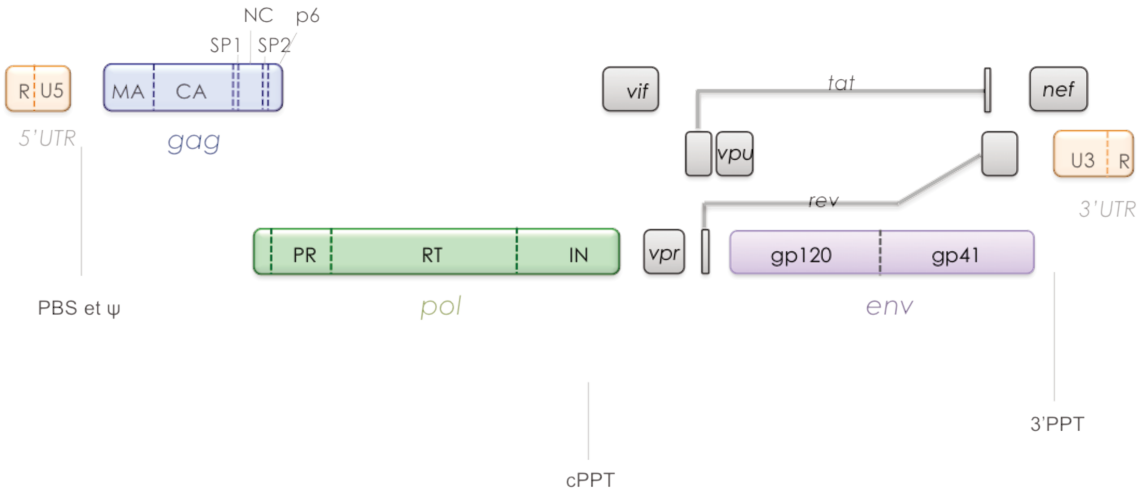



**Figure 20 : Représentation schématique du génome du VIH-1[38].**

### III.1.1 Les régions non codantes

Région R : La région répétée R composée de 100 nt est présente à chaque extrémité de l’ARN génomique et joue un rôle important dans le mécanisme de transcription inverse. Elle permet en effet le saut nécessaire pour la synthèse du premier brin d’ADN de l'extrémité 5' à l’extrêmité 3' de l'ARN génomique. Elle contient également le site d’initiation de la transcription ainsi que le signal de polyadénylation.

Région U3 : La région U3 est une région unique composée de 450 nt qui contient le promoteur pour la transcription virale. Elle joue également un rôle important pour l’intégration de l’ADN proviral car elle contient à son extrémité 5’ une séquence *att* reconnue par l’intégrase virale et nécessaire pour l’intégration du génome viral.

Région U5 : La région U5 est une région unique composée de 80 nt, importante pour le démarrage de la transcription inverse et contient comme la région U3 une séquence *att*

nécessaire à l'intégration de l'ADN viral. Cette région contient également la tige boucle TAR qui joue un rôle important pour la transcription des gènes viraux.

### III.1.2 Les protéines de structures

Comme indiqué au début de cette partie, les rétrovirus possèdent tous 3 gènes essentiels (Gag, Pol, Env) qui codent pour les protéines de structure ainsi que pour les enzymes virales (figure 20). Le gène Pol est toujours exprimé sous forme de protéine de fusion avec le gène Gag par un mécanisme de décalage du cadre de lecture et les trois gènes essentiels sont exprimés sous la forme de polyprotéines.

Le gène *gag* : gag est exprimé sous forme de précurseurs $Pr55^{Gag}$ ou $Pr160^{Gag\text{-}Pol}$ qui sont clivés par la protéase virale au cours de la maturation des virions suite au bourgeonnement de ceux-ci. Ce gène code pour la protéine de matrice (MA), la protéine de capside (CA), la protéine de nucléocapside (NC) et la protéine p6. Dans les virions matures, la protéine NC est associée au brin d'ARN formant ainsi un complexe ribonucléoprotéique. Les hexamères de protéine CA forment un complexe de forme conique renfermant les complexes ribonucléoprotéiques appelé la capside virale. Les protéines MA quant à elles restent associées aux protéines de l'enveloppe virale et notamment à la partie intra-virale de la sous unité gp41. La protéine p6 quant à elle possède les sites de fixation pour de nombreuses protéines cellulaires et virales et elle est impliquée dans différentes étapes du cycle infectieux. Le clivage du précurseur $Pr55^{Gag}$ génère également des petits peptides : SP1, SP2. La protéine SP1 joue un rôle important pour l'assemblage des virions et leur maturation [114,115] alors que le rôle de la protéine SP2 n'est pour l'instant pas encore élucidé même si sa séquence est bien conservée parmi les différentes souches de VIH-1.

Le gène *pol* : Le gène *pol* est également exprimé sous forme de polyprotéine et code pour la protéase virale (PR), pour la transcriptase inverse (RT) et pour l'intégrase (IN). La protéase virale est indispensable pour la maturation des virions et donc pour le cycle infectieux du virus[116]. Du fait de son rôle clé dans l'infection viral, elle est la protéine la plus largement ciblée dans les stratégies de thérapie antivirale. La transcriptase inverse est caractéristique de tous les rétrovirus et est une ADN polymérase ARN/ADN dépendant agissant sous forme d'hétérodimère. Cette enzyme possède une activité RNaseH qui permet la dégradation de

l'ARN génomique une fois recopié au cours de la synthèse du premier brin d'ADN[117]. La RT est une polymérase présentant un fort taux d'introduction de mutations[118]. Elle est notamment responsable des grandes variations de séquences qui sont observées parmi les virions, et responsable de l'apparition des quasi-espèces au cours de l'infection. La notion de quasi-espèce désigne dans un organisme l'apparition d'une population de virons hétérogènes issus de l'infection par un seul virus fondateur.

Le gène *env* : Le gène *env* code les protéines de l'enveloppe virale gp120 et gp41. Ce gène est également exprimé sous forme d'un précurseur gp160 qui est glycosylé et clivé dans l'appareil de Golgi avant d'être adressé au niveau de la membrane cellulaire[119]. Suite au clivage du précurseur, les protéines gp120 et gp41 produites restent en interaction grâce à des liaisons non covalentes[120]. La protéine gp120 est responsable de l'interaction entre la particule virale et la cellule, alors que la protéine gp41 est responsable de la fusion entre les membranes virale et cellulaire et donc de l'entrée du virus dans la cellule[121]. Le mécanisme d'entrée sera décrit plus en détail dans la suite de ce manuscrit.

### III.1.3 Les protéines régulatrices et les protéines auxiliaires

Les lentivirus sont considérés comme étant des rétrovirus complexes car en plus des protéines communes à tous les virus de la famille, ils possèdent des gènes codant des protéines régulatrices du cycle infectieux ou participant à la virulence.

Tat (trans-activateur de la transcription). La protéine Tat permet d'augmenter le niveau de transcription de l'ADN proviral via l'augmentation de la processivité de la polymérase cellulaire qui permet l'expresison des gènes viraux. La protéine Tat se fixe au niveau de la boucle TAR présente dans la région R[122] de l'ARN néo-transcrit et interagit également avec divers facteurs de transcription cellulaire tel que NF-kB et Sp-1[123,124].

Rev (régulateur de l'expression des protéines). La protéine Rev lie une structure d'ARN présente au niveau de l'ARN génomique. Elle contient à la fois des signaux d'import et d'export nucléaire. La présence du NLS conduit à l'adressage de la protéine au noyau où elle intervient dans le mécanisme d'export nucléaire des ARN non épissés ou partiellement épissés possédant la séquence RRE[125].

Vif (facteur d'infection des virions). La protéine Vif joue principalement un rôle dans la neutralisation de la protéine APOBEC3G, un facteur de restriction cellulaire induisant des mutations pouvant conduire à l'inactivation des particules virales[126]. En se fixant à la protéine APOBEC3G, Vif induit la dégradation de la protéine empêchant ainsi son incorporation dans les virions[127,128]. Elle est capable d'interagir avec l'ARN génomique ainsi que les précurseurs protéiques Gag-Pro et Gag-Pro-Pol, ce qui permet son encapsidation dans les virions[129].

Vpr (protéine virale R). La protéine Vpr joue un rôle important au niveau de différentes étapes du cycle infectieux. Il a été montré qu'elle est impliquée dans l'inhibition de certains mécanismes visant à éliminer les cellules infectées comme par exemple en limitant la sécrétion de cytokines tel que le TNF[130]. Elle agit également comme co-activateur de la transcription via son interaction avec le facteur de transcription cellulaire SP1[131] et bloque le cycle cellulaire en phase G2/M afin de favoriser une transcription optimale pour le virus[132,133]. Il a également été démontré qu'au cours du processus de transcription inverse, la protéine Vpr permet de limiter les mutations dans le génome viral et qu'elle participe également au transport nucléaire du complexe de pré-intégration[133,134].

Nef (Facteur négatif) : La protéine Nef participe à l'échappement au système immunitaire des cellules infectées en inhibant l'expression du CMH-1 à la surface cellulaire, évitant ainsi la présentation des antigènes viraux aux cellules du système immunitaire[135]. Elle provoque également l'endocytose des récepteurs CD4 et leur destruction dans les lysosomes ce qui contribue à éviter les cas de surinfection qui pourraient perturber l'étape de libération des particules virales[136–138]. Elle contrecarre aussi l'effet des facteurs de restriction SERINC3 et SERINC5 qui modifient le membrane virale en rendant le virus moins infectieux[139,140].

Vpu : la protéine Vpu permet de diminuer l'expression du CD4 à la surface des cellules infectées via leur dégradation au niveau du réticulum endoplasmique (RE)[141]. Vpu est exprimée à partir du même ARN messager que celui des protéines Env et il a été démontré que l'interaction de Vpu avec le CD4 dans le RE permet de limiter les complexes entre le précurseur d'enveloppe gp160 et le CD4[142]. Au niveau du RE la formation de complexes entre la gp160 et le CD4 conduit à une diminution du clivage du précurseur de l'enveloppe et donc à une réduction de l'expression des protéines d'enveloppes virales à la surface de la

cellule[142]. La protéine Vpu joue donc un rôle clé dans le cycle viral en permettant la maturation et l'adressage des protéines d'enveloppe au niveau de la membrane cytoplasmique. Vpu participe également au bourgeonnement des virions en contrecarrant l'action de la tétherine/BST2/CD317, un facteur de restriction cellulaire qui empêche le relargage des particules virales des cellules infctées[143].

### III.2 Cycle de réplication

Le cycle réplicatif du virus conduit à la production de nouveaux virions. Cette phase commence au premier contact entre la particule virale et la membrane cellulaire et se termine avec la libération des nouveaux virions (Figure 21). Au cours de ce cycle, les lentivirus vont libérer la capside virale dans le cytoplasme après la fusion membranaire. Se poursuit alors la transcription inverse et commence l'acheminement du génome viral vers le noyau dans lequel a lieu l'étape d'intégration. Suite à cette intégration, les protéines virales sont exprimées, le génome répliqué ce qui permet la formation de nouveau virions libérés au niveau de la membrane cytoplasmique par un mécanisme de bourgeonnement. Le cycle réplicatif du virus peut être décomposé en deux phases : la phase précoce et la phase tardive. La phase précoce comprend les étapes d'entrée du virus de décapsidation et de reverse transcription ainsi que les étapes d'adressage au noyau et d'intégration du génome virale. La phase tardive quant à elle comprend toutes les étapes qui font suite à l'intégration à savoir l'expression des gènes viraux, la formation des nouveaux virions à la membrane cytoplasmique ainsi que leur libération.

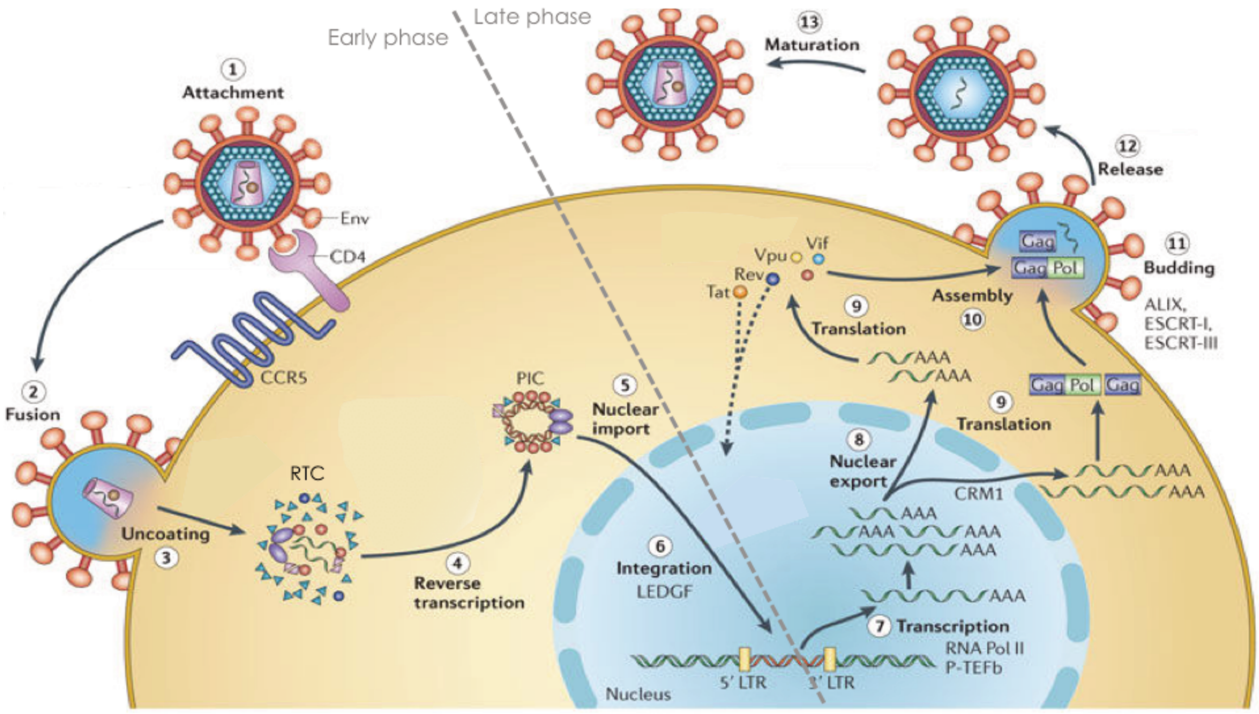



**Figure 21 : Cycle réplicatif du VIH-1[40].**

### III.2.1 La phase précoce

#### III.2.1.1 L'entrée

Le VIH-1 infecte les cellules CD4+ du système immunitaire ce qui inclus les lymphocytes T ainsi que les monocytes et il est également capable d'infecter les cellules du système nerveux[144]. L'entrée du virus est réalisée par les protéines de l'enveloppe organisées sous forme de trimères d'hétérodimères gp120/gp41. Le processus d'entrée nécessite la présence du récepteur CD4, exprimé en grande quantité à la surface des cellules T CD4+ et en plus petite quantité à la surface des macrophages et des cellules dendritiques. En plus de ce récepteur, la présence d'un corécepteur est également nécessaire pour conduire les changements conformationnels indispensables à la fusion membranaire. Les principaux corécepteurs pour l'entrée virale sont les protéines CCR5 (ou R5) et CXCR4 (ou X4). L'enveloppe virale est issue de la membrane cellulaire de la cellule productrice, et possède également certaines protéines membranaires cellulaires[145]. Au cours de mon travail de thèse, je me suis particulièrement intéressée à cette étape d'entrée, et elle sera décrite de manière plus détaillée dans la partie portant sur la redirection de l'entrée virale.

#### III.2.1.2 La transcription inverse

Suite à la fusion des membranes, la particule de capside est libérée dans cytoplasme et commencent alors les étapes de décapsidation et de transcription inverse (Figure 21). L'ordre de réalisation de ces étapes n'est pour l'instant pas définitivement établi mais les études montrent qu'une première étape de décapsidation est nécessaire pour l'initiation de la transcription inverse[146,147]. Les deux étapes semblent ensuite se poursuivent parallèlement au sein d'un complexe appelé le complexe de transcription inverse (RTC). Ce complexe est composé du génome viral associé aux protéines NC, ainsi que de toutes les protéines virales et cellulaires encapsidées au moment de la formation de la particule virale[148–150]. L'étape de transcription inverse est une étape clé du cycle réplicatif du virus car elle permet la synthèse de l'ADN proviral qui sera intégré et servira de base pour l'expression des gènes viraux. La synthèse de l'ADN proviral débute par l'utilisation de l'ARNt$^{Lys}$ qui sert d'amorce et qui est hybridé à l'ARN génomique au niveau du site PBS situé en 3' de la région U5 (figure 22). Suite à cette fixation démarre la synthèse du premier brin d'ADN par la polymérase virale. Au cours de cette synthèse, la matrice ARN déjà recopiée est dégradée par la RT qui porte une activité RNaseH. Seule deux régions de l'ARN génomique riches en purine son résistantes à l'activité RNaseH de la RT et ne sont pas dégradées. Ces régions appelées PPT pour *polypurine tract* servent toutes deux d'amorces pour la synthèse du deuxième brin d'ADN. Cette seconde synthèse étant réalisée à partir de deux amorces, elle génère à la fin de l'étape de réplication du génome viral, un ADN triple brin sur 90 nt. Cette structure triple brin de l'ADN est appelée « flap » et semble participer au transport nucléaire de l'ADN viral[151,152].

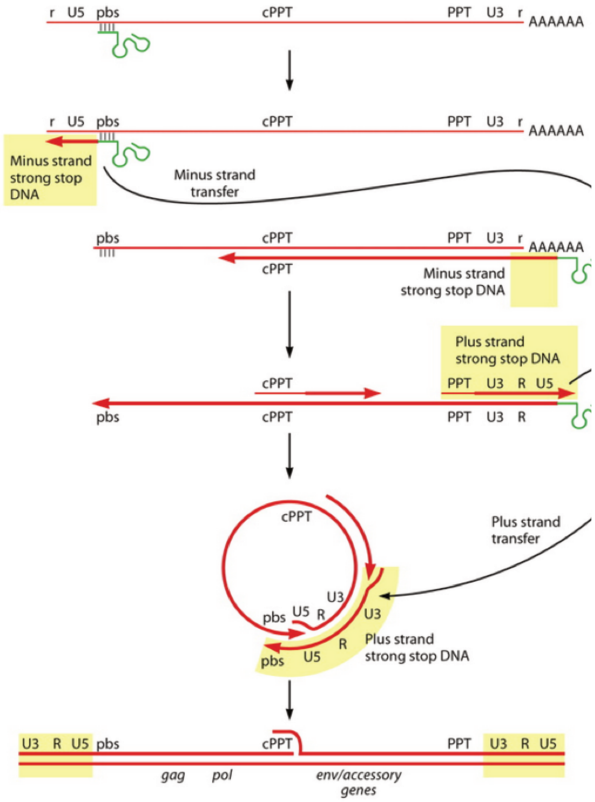



**Figure 22 : Mécanisme de transcription inverse[78]**

### III.2.1.3 L'import nucléaire

Une fois la synthèse de l'ADN proviral achevée, le RTC mature est appelé PIC pour « complexe de pré-intégration » et est transporté vers le noyau cellulaire pour l'étape d'intégration. Le transfert du PIC est réalisé par un mécanisme de transport actif faisant intervenir les protéines cellulaires constitutive des pores nucléaire tel que les importines α et β ainsi que les nucléoporines[153–156]. La capacité du PIC à utiliser les systèmes de transferts actifs cellulaires permet notamment aux lentivirus d'infecter, outre les cellules en division, les cellules quiescentes. En plus des protéines cellulaires mobilisées, les protéines virales présentent dans le complexe PIC semblent également participer à ce transfert[155,157–161]. Une fois dans le noyau cellulaire, l'ADN viral peut être intégré de façon stable dans le génome cellulaire.

### III.2.1.4 L'intégration

L'intégration de l'ADN viral est catalysée par l'intégrase virale (IN) et se fait en deux étapes (figure 23). Le processus de transcription inverse génère aux extrémités de l'ADN proviral

des séquences répétées appelées LTR pour *Long Terminal Reapeat* constituée des régions U3, R et U5 et qui sont caractéristiques des rétrovirus. L'intégrase grâce à son activité endonucléase effectue un clivage à l'extrémité de chaque LTR, ce qui conduit à la synthèse d'un groupement hydroxyle réactif 3'CA à chaque extrémité de l'ADN proviral. L'IN réalise ensuite un clivage non spécifique de l'ADN cellulaire générant ainsi des groupements phosphates. L'étape de ligation entre l'ADN cellulaire et l'ADN proviral, connu sous le nom de transfert de brin, se fait ensuite par une réaction de trans-estérification qui permet de générer des liaisons phosphodiesters entre les deux séquences. Ce mécanisme de transfert de brin est réalisé par les enzymes cellulaires impliquées dans le processus de réparation de l'ADN.

Le clivage de l'ADN cellulaire ne se fait pas au niveau d'un site particulier, de ce fait, l'intégration des lentivirus a longtemps été considérée comme aléatoire. Il a toutefois été démontré par la suite que cette intégration se fait préférentiellement dans les régions transcriptionnellement actives du génome[162,163]. La grande majorité de l'ADN proviral ne s'intègre cependant pas dans le génome cellulaire. Cet ADN proviral est capable de se maintenir dans le noyau sous différentes formes non intégrées[164–166]. Ces différentes formes permettent également l'expression des protéines virales mais avec un taux d'expression plus faibles que l'expression à partir de la forme proviral[167]. Ces formes épisomales présentent un grand intérêt pour l'utilisation des vecteurs lentiviraux, afin d'éviter les problèmes liés à l'insertion du vecteur dans le génome cellulaire.

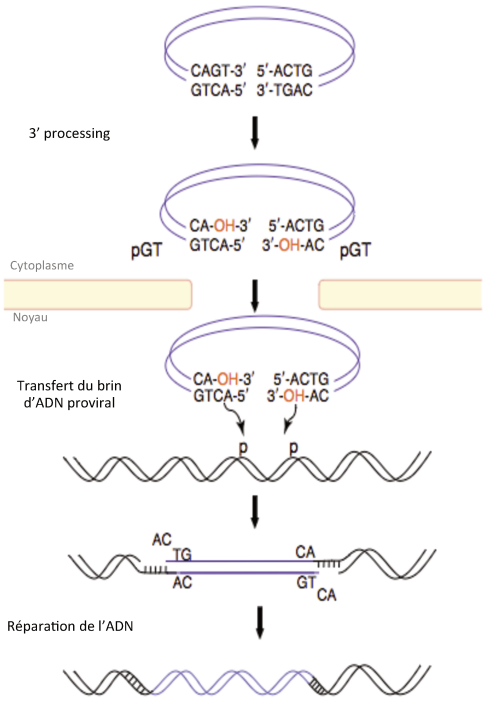



**Figure 23 : Mécanisme d'intégration[93].**

### III.2.2 La phase tardive

La phase tardive du cycle réplicatif du VIH commence suite à l'intégration de l'ADN proviral. Cette phase se compose des étapes d'expression des gènes (figure 24) et de formation des nouveaux virions.

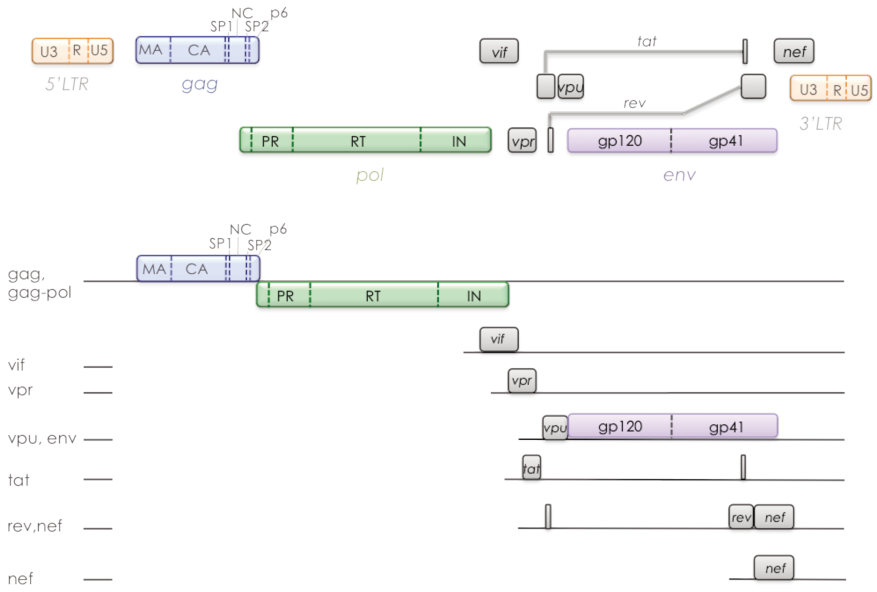



**Figure 24 : Epissage alternatif et expression des protéines virales.**

#### III.2.2.1 La transcription

La transcription du provirus est réalisée par la machinerie cellulaire et est catalysée par l'ARN polymérase II cellulaire. Les séquences promotrices et terminatrices de la transcription virale sont localisées au niveau des régions LTR. La région U3 contient la séquence promotrice de la transcription ainsi que les sites de fixation de nombreux facteurs cellulaires tels que les protéines Sp1, NF-kB et NFAT[168] et le signal de polyadénylation est quant à lui situé au niveau de la région R. Les LTR étant strictement identiques en 5' et en 3' du génome viral, le signal de polyadénylation se trouve également situé au début du génome mais est réprimé à ce niveau. Cette répression est liée à différents facteurs présents dans cette région notamment la présence de la « cap » en amont[169] du signal, la présence du site donneur de l'épissage en aval du signal[170,171] et la stabilité de la structure en tige boucle contenant le signal dans cette région[172,173]. La transcription du génome viral démarre à partir du premier nucléotide de la région R du LTR 5', et se termine au dernier nucléotide de la région R présente dans le LTR 3' où le signal de polyadénylation est accessible pour la machinerie cellulaire. La séquence de polyadénylation AAUAAA est reconnue par le complexe CPSF ce qui conduit au recrutement du complexe CsTF qui va stimulé le clivage du l'ARNm 10-30 nt en aval du signal. Suite au clivage, l'addition de la queue poly(A) est alors catalysée par la protéine PAP[174].

#### III.2.2.2 La traduction l'assemblage et le bourgeonnement

Au début du cycle de transcription, tous les ARNm produits sont coiffés, polyadénylés et totalement épissés avant d'être exportés dans le cytoplasme par le mécanisme classique d'export nucléaire en vue de leur traduction. Ces ARN multi-épissés permettent l'expression des protéines régulatrices Tat, Rev et Nef. Les protéines Tat et Rev possèdent des NLS, ce qui leur permet d'être tout de suite adressées au noyau une fois produites.

La protéine Tat se fixe au niveau de la tige-boucle TAR formée sur l'ARN néo-transcrit dans la région R de la LTR 5'. Elle permet ensuite de recruter le complexe cellulaire P-TEFb qui via la phosphorylation de l'extrémité C terminale de l'ARN polymérase va augmenter la processivité de l'enzyme et stimuler la synthèse des ARN viraux[175–178]. Cette activation de la transcription permet notamment d'atteindre un taux de transcription suffisant de l'ARN génomique pour la poursuite du cycle réplicatif.

En se fixant à la séquence structurée RRE présente sur l'ARN génomique, la protéine Rev permet l'export nucléaire des différents ARN non ou partiellement épissés[179] via la voie d'export CRM1/RANGTP[180,181]. Classiquement, seule les ARN complètement épissés, coiffés et polyadénylés peuvent être exportés au cytoplasme pour être traduit. Les protéines de structures du virus ainsi que les protéines auxiliaires sont cependant exprimées à partir d'ARN mono ou non épissés. La protéine Rev est donc indispensable pour l'expression de ces protéines virale.

Produite à partir d'un ARNm mono-épissé, la protéine gp160 est adressée à l'appareil de Golgi où elle sera clivée et glycosylée avant d'être acheminée au niveau de la membrane cytoplasmique par le système de transport vésiculaire de la cellule. Il a été démontré que la glycosylation des protéines d'enveloppes est une étape clé pour la fonctionnalité de l'enveloppe car elle affecte aussi bien le repliement de la protéine que sa capacité à lier le récepteur CD4 et les corécepteurs[182–187].

Les précurseurs Gag et Gag-Pol permettent quant à eux le recrutement et l'acheminement des ARN génomiques au niveau de la membrane plasmide. Ce recrutement s'effectue grâce à l'interaction des différentes protéines du précurseur avec l'ARN génomique ou des facteurs cellulaires. L'interaction avec les ARN génomiques se fait grâce au domaine contenant la protéine NC qui possède deux domaines zinc finger. Le complexe est ensuite adressé au niveau de la membrane cellulaire à laquelle il s'ancre grâce à son domaine N-terminal myristylé. La partie C-terminale du précurseur $Pr55^{Gag}$ quant à elle contient la protéine P6 qui permet notamment l'interaction avec les protéines TSG101 et ALIX. Ces protéines font parties du complexe cellulaire ESCRT qui participe au bourgeonnement des nouvelles particules virales[188] qui est un processus ATP-dépendant (figure 25). De manière succincte, ce processus démarre avec la multimérisation des précurseurs gag qui permet une invagination de la membrane cytoplasmique. Les protéines ESCRT-III s'assemblent ensuite pour former une structure en spirale à la base des bourgeons, et le mouvement de ces filaments permet de rapprochement des bords par un mécanisme non élucidé pour l'instant. Suite à ce rapprochement, la protéine VPS4 qui possède une activité ATPase catalyse l'hydrolyse de l'ATP couplé à une déstabilisation des membranes, permettant ainsi la libération des virions[189].

Les virions libérés sont dits « immatures » car ils ne sont pas infectieux. Pour qu'ils soient infectieux, il faut que les virions soient maturés, et cette maturation est réalisée par la protéase virale. Par une action en cis sur les précurseurs $Pr55^{Gag\text{-}Pol}$, la protéase (active même lorsqu'elle fait partie de ce précurseur) se libère d'abord de la polyprotéine et catalyse ensuite le clivage des autres protéines. La protéase virale est active sous forme de dimère et réalise les clivages en des sites bien définis au niveau des précurseurs et dans un ordre également bien défini. Cette étape de maturation permet aux virions d'acquérir des caractéristiques permettant de les rendre infectieux.

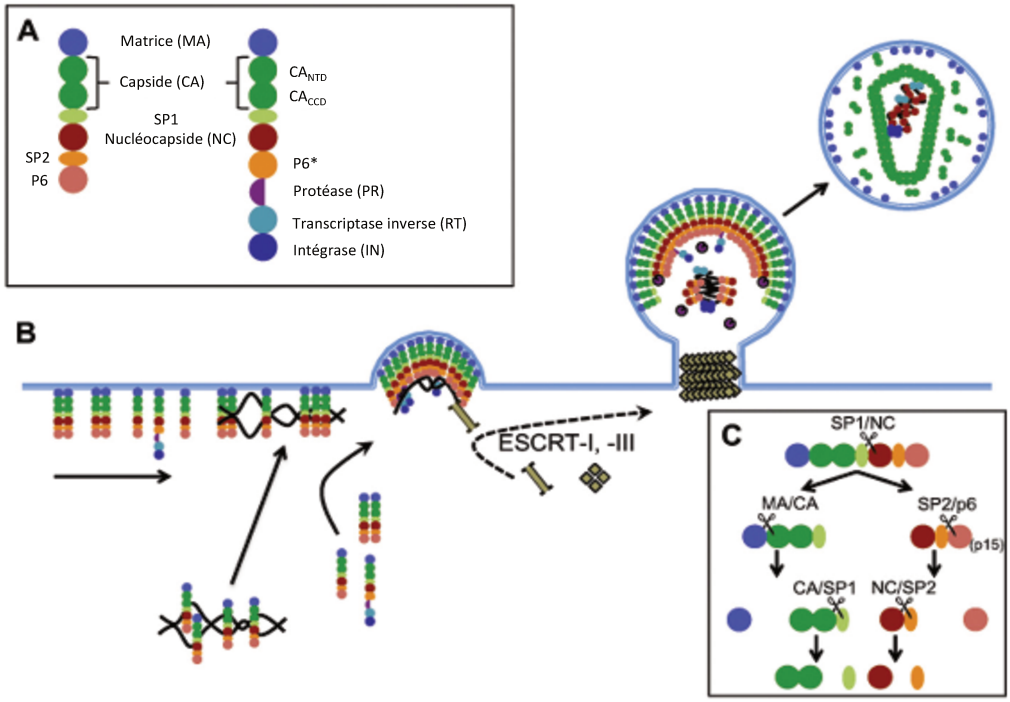



**Figure 25 : Assemblage, bourgeonnement et maturation des particules virales[115].**

## III.3 Redirection de l'entrée virale

### III.3.1 Les mécanismes moléculaires de l'entrée

Les particules virales portent entre 7 et 14 trimères des protéines d'enveloppe à leur surface[190]. Les protéines d'enveloppe sont, comme décrit précédemment, des hétérodimères composés par les protéines gp120 et gp41. Ces deux protéines sont reliées entre elles par des interactions non covalentes majoritairement entre les extrémités N-terminale et C-terminale de la gp120 et le domaine extra membranaire de la gp41[191,192]. La protéine gp120 est celle qui interagit avec le récepteur CD4 et successivement avec les

corécepteurs CCR5/CXCR4. Chacune de ces interactions induisent des changements conformationnels qui aboutissent à la fusion des membranes grâce à la protéine gp41.

La protéine gp120 est composée de 5 régions variables et de 5 régions constantes. Ces régions sont nommées sur la base de l'analyse de leurs séquences et sur la conservation de celles-ci entre les différentes souches virales. Les régions constantes constituent le cœur des trimères alors que les régions variables sont plus exposées à l'extérieur de la protéine (figure 26 B). La capacité des régions variables à être plus permissives aux mutations et aux modifications post-traductionnelles explique notamment la capacité du virus à échapper au système immunitaire de l'hôte via l'altération des épitopes qui pourraient être reconnus par celui-ci[193–195]. La résolution de la structure de la protéine gp120 a mis en évidence le fait que la protéine se replie en formant une cavité dans laquelle se fait l'interaction avec le récepteur CD4. Cette interaction fait intervenir 22 aa du récepteur CD4 situé entre les résidus 25 et 64 au niveau de l'extrémité N-terminale de la protéine avec 26 résidus de la protéine gp120, repartis sur 6 domaines différents de la protéine[196]. Parmi ces interactions, certaines sont considérées comme centrales car elles sont conservées chez tous les virus du genre. Ces interactions essentielles font notamment intervenir les résidus Phe 43 et Arg 59 du CD4 et les résidus Asp 368, Glu 370 et Trp 427 de la gp120. Une fois le CD4 entré dans la cavité, le résidu 59 du CD4 établit des liaisons hydrogènes avec les résidu Asp 368 et Val 430 de la gp120. De même, le résidu Phe 43 interagit avec les résidus Trp 427 et Asp 368. Aucune évidence directe n'est pour l'instant établie mais la comparaison des conformations entre les trimères libres ou en interaction avec des anticorps semble indiquer que les interactions entre le CD4 et la gp120 permettent la création à la base de la cavité d'une structure secondaire en feuillet β appelée le « bridging sheet »[196–198]. L'interaction avec le récepteur entraine également la délocalisation de la région V3 vers la membrane cellulaire.

Ce sont ces deux éléments, le « bridging sheet » ainsi que la délocalisation de la région V3 (figure 27) qui sont les éléments centraux de la liaison au corécepteur. La région V3 est celle qui va définir le tropisme viral à savoir l'utilisation du CCR5 ou du CXCR4 comme corécepteur[199]. De nombreuses méthodes visant à prédire le tropisme des virus sont basées sur la séquence de la région V3. Une de ces méthodes est notamment basée sur la charge globale de la région en se focalisant principalement sur les acides aminés en position 11 et 25 de la région 25[200]. Cette méthode de prédiction considère notamment que si en position 11 ou 25 de la région V3 se trouve un acide aminé de charge positive, alors l'enveloppe sera de tropisme CXCR4.

Quoi qu'il en soit la liaison du corécepteur permet d'exposer le peptide de fusion de la gp41 qui va s'ancrer à la membrane cellulaire. Ce peptide de fusion est situé à l'extrémité N-terminale de la protéine et est suivi de deux régions organisées sous forme d'hélice alpha qui sont respectivement nommée HR1 et HR2 (figure 26 C).

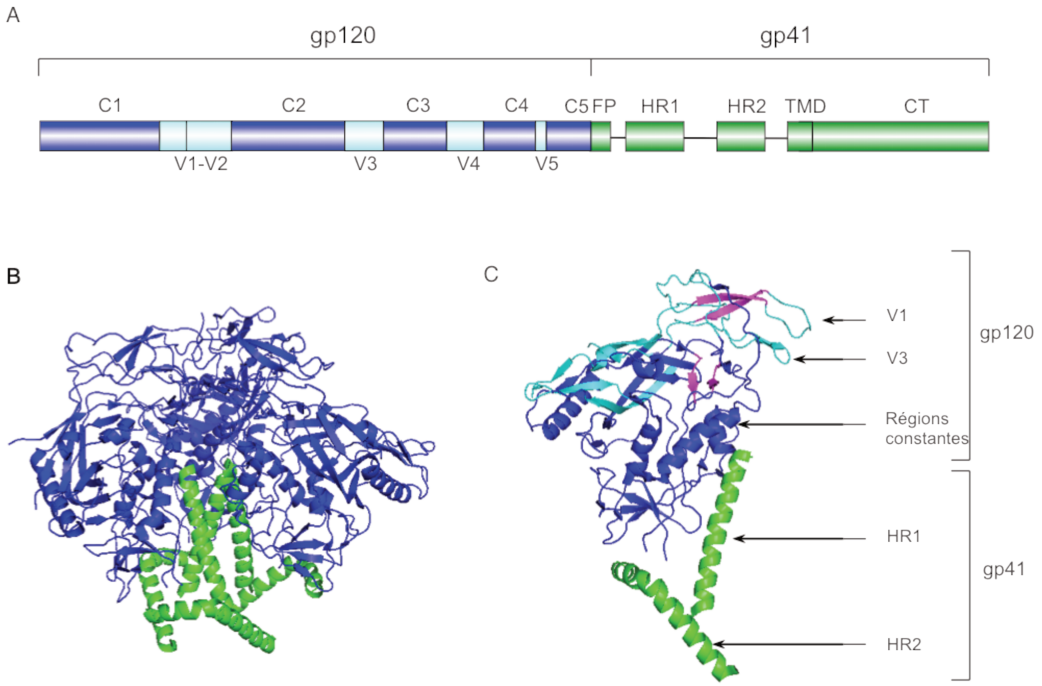



**Figure 26 : Structure de l'enveloppe virale**. A) La protéine gp160 est représentée de manière linéaire avec en bleu la protéine gp120 (les régions variables en bleu clair et les régions constantes en bleu foncé) et en vert la protéine gp41. B) La structure du trimère de l'enveloppe virale (PDB ID : 3J5M.) est représentée avec en bleu la protéine gp120 et en vert la gp41. C) La structure d'un hétérodimère gp120-gp41 avec en bleu clair les régions variables de la gp120 et en bleu foncé les régions constantes de la gp120. la protéine gp41est représenté en vert, et sont indiqués en violet les feuillets β impliqués dans la formation du « *bridging sheet* ».

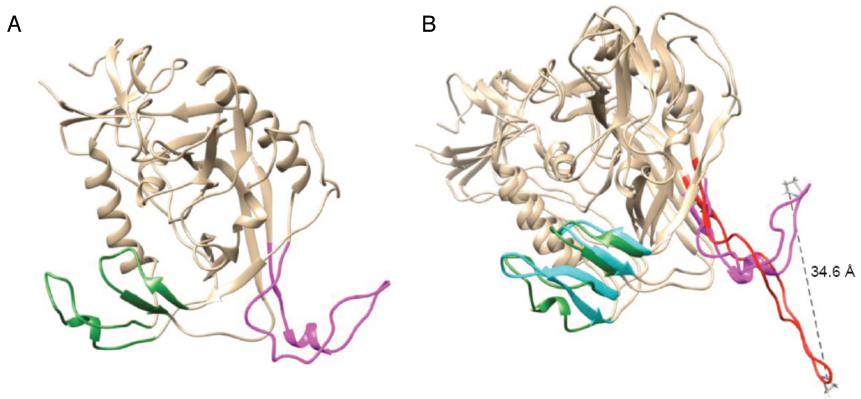



**Figure 27 : Structure comparée de la région variable V3[126]**. A) Structure d'un monomère de protéine gp120 native. En violet est indiqué la région V3 et en vert les régions impliquées dans la formation du « *bridging sheet* ». B) superposition entre la structure d'une protéine gp120 native et d'une protéine gp120 liée au récepteur CD4. Sont indiqué en vert, le « *bridging sheet* » dans la gp120 native et en bleu, le « *bridging sheet* » dans la protéine liée. En violet la région V3 dans la protéine native et en rouge la région V3 dans la protéine liée.

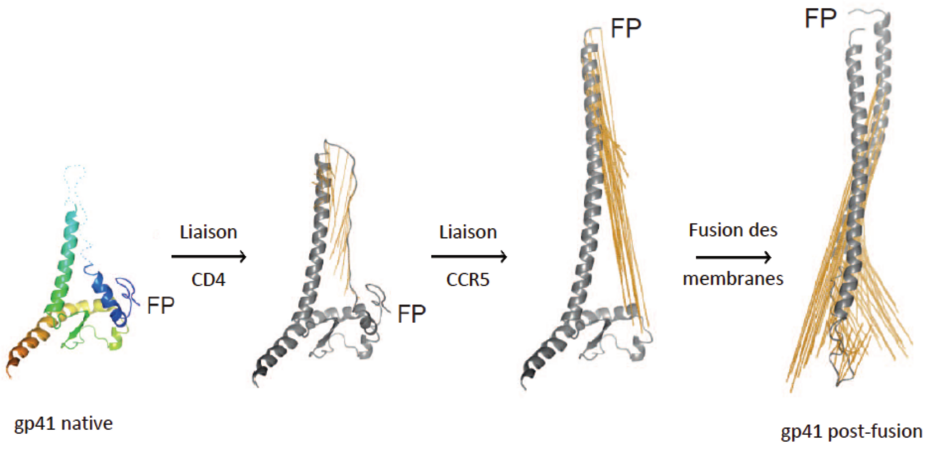



**Figure 28 : Modification de la structure de la protéine gp41 au cours de l'entrée virale[127].**

Dans son état natif, lorsque l'enveloppe virale n'est pas en interaction avec les récepteurs et corécepteurs viraux, la protéine gp41 est dans un état condensé et non stable qui est maintenu via les interactions qu'elle réalise avec la sous unité gp120. Les changements conformationnels qui affectent la gp120 au cours du processus d'entrée permettent de libérer la protéine des contraintes subies, ce qui induit un relâchement qui permet dans un premier temps l'ancrage du peptide de fusion dans la membrane cellulaire (Figure 28). La perte des interactions stables entre la gp120 et la gp41 induit également un détachement de la gp120 qui se retrouve alors dans le milieu extracellulaire. Ce processus de relargage de la

gp120 est appelé le « *shedding* »[201]. Dans un deuxième temps, les trois domaines HR1 du trimère de gp41 s'associent aux trois domaines HR2 afin de former une « super hélice » antiparallèle, ce qui entraine le rapprochement de l'enveloppe virale et de la membrane cellulaire permet la formation d'un pore à travers lequel est libéré la capside virale dans le cytoplasme.

### III.3.2 Les enveloppes CD4 indépendantes

Comme décrit dans la partie précédente, l'entrée virale nécessite une première interaction entre la gp120 et le récepteur CD4 afin d'engager les changements conformationnels nécessaires à la fusion des membranes et au démarrage du cycle infectieux. L'efficacité d'infection est bien meilleure en présence du récepteur CD4 mais il est décrit dans la littérature des virus VIH-2 capable d'infecter les cellules en présence uniquement du corécepteur[202]. Ces enveloppes dites CD4 indépendantes (CD4i) semblent plus abondantes pour le VIH-2 ainsi que pour les virus simiens comparés au VIH-1. Il a tout de même été isolé chez des patients, des virus VIH-1 CD4i[203]. Des souches de laboratoires du VIH-1 ont également été adaptées afin de pouvoir infecter les cellules en absence du récepteur CD4[204–206], et les études réalisées sur l'enveloppe virale ont permis d'identifier certaines mutations capable d'abolir la nécessité de liaison au récepteur CD4 pour l'infection[182,207–210]. Une partie des études décrit notamment des mutations au niveau des résidus Asp368, Glu 370 et Trp 427 de la gp120[208,209] caractérisés comme étant essentielles pour l'interaction avec le CD4 comme indiqué précédemment. Les acides aminés Ser190 et Asn197 sont également décrits comme participant au phénotype CD4i de même que la région V1/V2[182,207]. Les enveloppes CD4i isolées ainsi que celles adaptées en laboratoire sont cependant toutes décrites comme étant peut efficaces pour l'infection comparées aux enveloppes nécessitant le CD4. Ces enveloppes CD4i gardent également une capacité résiduelle à infecter les cellules CD4+. Dans la partie précédente, concernant les mécanismes moléculaires de l'entrée virale, il est mis en évidence qu'en plus de permettre l'initiation des changements conformationnels nécessaires à l'entrée du virus dans la cellule, la liaison au récepteur CD4 permet également un ancrage solide des virus à la surface cellulaire facilitant toutes les interactions et changements conformationnels nécessaires à l'entrée virale. Dans le cas des virus CD4i, la baisse de la capacité à réaliser l'entrée peut donc être expliquée par la perte de cet ancrage solide. D'autre part, dans les enveloppes CD4i, l'interaction entre les protéines gp120 et gp41 est fragilisée afin de permettre à l'enveloppe d'avoir une conformation capable de lier

le corécepteur. La fragilité de l'interaction entre les deux protéines pourrait conduire à un détachement de la gp120 et donc à un *shedding* précoce. Cela aurait pour conséquence d'affecter la capacité de l'enveloppe à lier le corécepteur et à réaliser l'entrée.

Dans le cadre de mon travail de projet de thèse, nous avons souhaité généré une enveloppe VIH-1 modifiée capable de reconnaître préférentiellement les cellules cancéreuses et pour ce faire, nous nous sommes appuyés sur les recherches concernant les enveloppes CD4i.

### III.3.3 La redirection de l'entrée virale

En ce qui concerne les vecteurs lentiviraux, le pseudotypage est la méthode communément utilisée afin de réorienter l'entrée des vecteurs vers des types cellulaires différents. Dans le cas du traitement contre le cancer, une des stratégies étudiées consiste à utiliser des enveloppes capables de reconnaître spécifiquement les marqueurs tumoraux exprimés à la surface des cellules cancéreuses. Il est très difficile de trouver des enveloppes possédant un tropisme restreint pour certains types cellulaires et presqu'impossible d'en trouver présentant un tropisme spécifique pour les marqueurs tumoraux exprimés à la surface des cellules cancéreuses. Ainsi, l'option la plus simple pour atteindre notre objectif, est d'utiliser des enveloppes dont le tropisme peut être facilement modifiable via d'une part, l'introduction de mutations permettant d'abolir le tropisme naturelle de l'enveloppe, et d'autre part, l'introduction de peptides capables de reconnaître spécifiquement les marqueurs tumoraux. Les précédents travaux réalisés dans ce domaine se sont principalement focalisés sur l'utilisation des glycoprotéines d'enveloppe du virus de la rougeole et celle du virus Nipah. L'intérêt pour ces deux virus est lié au fait que leurs protéines d'enveloppe sont des hétérodimères pour lesquels les fonctions d'attachement à la membrane cellulaire et de fusion des membranes sont assurées séparément. Il est donc plus facile pour ce type d'enveloppe de modifier le tropisme sans pour autant affecter la capacité d'entrée. Cependant, en ce qui concerne les protéines d'enveloppe du virus de la rougeole, le principal inconvénient est qu'une partie de la population possède une immunité préexistante contre ces virus, ce qui conduit à l'élimination rapide des vecteurs après leur injection[211]. Concernant, les glycoprotéines d'enveloppe du virus Nipah, les études menées ont montré que l'efficacité d'entrée des vecteurs lentiviraux pseudotypés avec ces protéines est dépendante de la distance entre l'épitope ciblé et la membrane cellulaire[212]. Cette

contrainte limite donc le choix des ligands pouvant être utilisés pour cibler le marqueur tumoral à la surface des cellules cibles.

Pour notre part, nous nous sommes intéressés aux enveloppes des virus VIH-1 qui présentent également la caractéristique d'être des hétérodimères avec un partage des fonctions de reconnaissance et de fusion entre les protéines gp120 et gp41 (cf Partie II-III.3.1).

### III.3.4 L'introduction d'un peptide dans l'enveloppe virale

Afin que cette réorientation de l'entrée virale soit effective, il faut être capable d'insérer dans l'enveloppe virale un motif peptidique capable de reconnaître les récepteurs exprimés spécifiquement à la surface des cellules cibles.

Les études menées sur l'enveloppe virale du VIH s'intéressent principalement à l'immunogénicité de la protéine et à la neutralisation du virus via l'utilisation d'anticorps dirigés vers différentes parties de la gp120. Afin de mieux comprendre l'efficacité des anticorps à neutraliser le virus quelques laboratoires ont tenté d'introduire des épitopes connus et faciles à cibler sur la protéine. D'autres laboratoires se sont également intéressés à l'insertion de tag dans l'enveloppe virale afin de suivre le traffic des protéines dans les cellules infectées. Mon travail c'est basé sur les résultats préalablement obtenus dans ces recherches.

Les régions variables de la protéine de surface présentent une grande diversité aussi bien parmi les différentes souches virales que parmi les quasis espèces présentes chez un même patient.

Les régions V1/V2 se situent au sommet des trimères et les études menées sur l'analyse de leurs séquences montrent qu'ils présentent une grande diversité de séquences et de tailles entre les différentes souches virales mais également entre les différentes quasis espèces que l'on peut trouver chez un même patient[213,214]. La région V3 est située en dessous des régions V1/V2 et est impliquée dans la reconnaissance du corécepteur[196,215]. La fonction des régions V4 et V5 n'est pour l'instant pas encore bien définie, mais certaines études ont montrés que la délétion d'une région variable de la protéine est suffisante pour abolir sa capacité à promouvoir l'entrée virale[216,217]. Dans les structures récemment publiée, on constate que les régions V2 et V4 ne sont pas résolues, ce qui suggère leurs grande flexibilité[218–220]. Les

régions V1/V2 et V3 sont associées au sommet du trimer et son décrites comme permettant de stabiliser la structure, alors que les régions V4 et V5 sont éloignées des sites reconnus comme importants pour la reconnaissance des récepteurs et corécepteurs. Avec toutes ces observations, il semble donc que la région V4 soit une région optimale pour l'insertion d'un motif peptidique sans pour autant déstabiliser la structure et la fonction de l'enveloppe virale.

Les recherches actuelles utilisent différentes stratégies d'insertion de peptide[221–224]. Certaines stratégies substituent la séquence virale par celle du peptide en vue de ne pas trop modifier la taille de la région. Cette stratégie donne lieu à de grande variation dans la fonctionnalité de l'enveloppe en fonction du site où est inséré le motif [223]. D'autres stratégies insèrent le peptide en lui ajoutant une séquence « linker » de part et d'autre. Avec cette stratégie, la fonctionnalité n'est affectée qu'en fonction de la région dans laquelle le peptide est inséré et non pas de la position exacte de cette insertion, ce qui est beaucoup moins contraignant[224]. Dans le cas de deux protéines produites en fusion, les séquences *linker* servent à créer un lien entre les protéines tout en permettant de les éloigner suffisamment pour favoriser le bon repliement de chacune d'elles. Il existe différents types de séquences *linker* qui sont classées en trois catégories : les flexibles, les rigides et celles pouvant être clivées *in vivo*[225]. Les linker flexibles sont ceux qui nous intéressent particulièrement et ils sont principalement composés de glycine et de sérine. Du fait de leur petite taille et de leur polarité, ces acides aminés permettent d'assurer la flexibilité et la solubilité des protéines de fusion produites sans apporter de charge supplémentaire[225,226].

## Partie III : Objectifs du travail de thèse

Les gènes suicides présentent un grand intérêt pour les stratégies de thérapie génique. Du fait de son importance dans l'activation de nombreux analogues nucléosidiques utilisés dans les traitements chimiothérapeutiques, la dCK humaine a longtemps été étudiée dans le but d'identifier une variante de la protéine qui permettrait de sensibiliser les cellules à de faibles concentrations de médicament et pourrait être utilisée comme gène suicide dans les stratégies anti-tumorales. Le laboratoire grâce à une méthode innovante d'évolution moléculaire mise en place au sein même de l'équipe et appelé la Rétrovolution, à identifié deux variants de la dCK capable de sensibiliser les cellules résistantes à des concentrations de Gem jusqu'à 900 fois inférieures de celles requises pour les cellules résistantes.

Le premier mutant appelé le G12 a été caractérisé et a fait l'objet d'une publication avant mon arrivé dans l'équipe[59]. Au cours de la première partie de mon travail de thèse, j'ai complété les études déjà effectuées sur le G12 en évaluant l'importance de chacune de ces trois mutations caractéristiques pour le phénotype observé. J'ai principalement travaillé sur la caractérisation de la seconde protéine identifiée appelée M36 qui se positionnait comme un bon candidat gène suicide. Le M36 a été identifié après 11 cycles de rétrovolution supplémentaires à partir du G12 et a démontré une meilleure capacité à sensibiliser les cellules à la Gem. Comparé à la dCK sauvage, le M36 est caractérisé par 4 mutations ponctuelles dont trois lui sont communes avec le G12.

Dans la première partie de mon travail de thèse je me suis donc intéressée à la caractérisation d'un potentiel nouveau gène suicide qui, du fait de son origine humaine devrait être une bonne alternative au système hsTK/GCV qui présentent des problème d'immunogénicité du fait e l'origine viral de la protéine.

Indépendamment du gène suicide utilisé, pour qu'une stratégie de thérapie par gène suicide puisse être optimale, il faut que le gène puisse être exprimé spécifiquement dans les cellules cibles. Les vecteurs utilisés pour le transfert de gène doivent donc être capables de cibler spécifiquement les cellules, et c'est à cet aspect des stratégies de thérapie que je me suis intéressée au cours de la seconde partie de mon travail de thèse. L'adressage des gènes aux cellules peut être réalisé grâce à l'utilisation de vecteurs viraux ou non viraux. Pour ma part, je me suis intéressée aux vecteurs lentiviraux et plus particulièrement aux enveloppes des virus VIH-1.

Dans le cadre de mon projet, nous avons émis l'hypothèse que l'utilisation d'enveloppes VIH-1 CD4i auxquelles serait greffé un peptide capable de reconnaître une protéine membranaire spécifique à un type cellulaire, permettrait de modifier le tropisme du virus et de réorienter son entrée vers un autre type cellulaire. En effet, les enveloppes CD4i n'ayant pas besoin de l'interaction avec le récepteur CD4 pour être capable d'interagir avec le co-récepteur et réaliser l'entrée, elles ne seront pas restreintes aux cellules exprimant le récepteur. En revanche, la greffe d'un peptide capable de reconnaître une protéine membranaire spécifique aux cellules cibles permettra d'une part de réorienter l'entrée préférentiellement vers ces cellules permettra également l'entrée dans les cellules conditions que celles-ci expriment un des co-récepteur du VIH-1. Si le ligand possède une bonne affinité pour le récepteur, l'ancrage solide assuré par la liaison au CD4 pour les enveloppes sauvages sera rétablie, permettant ainsi d'améliorer la capacité des enveloppes CD4i qui sont décrites comme moins efficaces pour l'entrée que les enveloppes sauvages.

# Partie IV : Etude des mutants de la dCK humaine

## I Objectifs

Au cours de cette première partie de ma thèse j'avais plusieurs objectifs en relation avec les mutants de la dCK humaine isolés au laboratoire. Mon travail a eu pour but d'avoir une meilleure compréhension de leurs mécanismes de fonctionnement et visait également à une amélioration de ces mutants. A mon arrivé au laboratoire, les protéines G12 et M36 avaient déjà été identifiées par le processus de rétrovolution. La capacité du mutant G12 à sensibiliser les cellules Messa 10K résistantes à la Gem, de même que sa capacité à sensibiliser à la Gem d'autres lignées cellulaires cancéreuses avait également été démontrée.

Un premier objectif était lié à une meilleure caractérisation du mutant G12. Celui-ci a été identifié après 17 cycles de rétrovolution (et donc de mutagénèse) et est caractérisé par trois mutations comparé à la dCK sauvage. Le processus de rétrovolution est une méthode d'évolution aléatoire qui consiste dans un premier temps à générer des mutations dans le gène à faire évoluer, avant d'évaluer l'effet des mutants sur les cellules cibles. Au cours du processus, il est possible qu'une ou deux des mutations du G12 aient été introduites avant de cribler la banque, mais qu'elles ne soient pas nécessaires pour observer le phénotype de la protéine. Le criblage de la banque de mutant générée ne se faisant qu'à partir d'un certains nombres de cycles, l'importance relative des trois mutations individuelles du G12 pour son efficacité restait encore indéterminée. Au début de ce travail, j'ai donc étudié la contribution de chacune des mutations du G12 pour son efficacité. Pour ce faire, j'ai construit des variantes de la dCK possédant une ou deux des mutations du G12 en couvrant toutes les combinaisons possibles, et ai testé la sensibilité à la Gem des lignées cellulaires stables exprimant chacune une variante de la protéine.

Un deuxième objectif était de modifier la nature des vecteurs lentiviraux utilisés dans nos essais en culture de cellules. Pour le processus de rétrovolution, les vecteurs utilisés devaient posséder des séquences LTR fonctionnelles. Afin de se rapprocher des conditions d'utilisation des vecteurs lentiviraux en thérapie génique, j'ai construit des vecteurs contenant des séquences LTR tronquées et donc non fonctionnelles appelés vecteur SIN

pour « *self inactivating* » et comparé le comportement des ces vecteurs à celui des vecteurs possédant des LTR fonctionnels.

Le troisième objectif a consisté en une caractérisation extensive du mutant M36. Dans le but d'identifier un mutant plus performant que le G12, celui-ci avait également servi de point de départ pour la rétrovolution, ce qui a permis après 11 cycles supplémentaires l'identification de la protéine M36. A mon arrivée, la protéine M36 n'avait été décrite dans aucunes des publications du laboratoire et seul sa capacité à sensibiliser les cellules Messa 10K à la Gem avait été démontrée. En ce qui concerne ce sujet, mon travail a donc principalement été l'étude du M36 et cela a porté sur les points suivants :

i. La capacité de la protéine M36 à sensibiliser les cellules à des médicaments autres que la Gem. Pour ce faire, j'ai testé la sensibilité des cellules à la Fludarabine et à l'AraC, deux composés chimiothérapeutiques analogues de la dC activés par la dCK. J'ai également mis en place un test permettant de tester la capacité de la protéine à renforcer l'action de deux analogues de la dC utilisés comme antiviraux : la ddC et le 3TC.
ii. La capacité du M36 à sensibiliser des cellules cancéreuses autres que les cellules Messa 10K aux médicaments.
iii. La contribution de la mutation supplémentaire S169N du M36 pour son efficacité. La protéine M36, étant issue du mutant G12, elle possède les trois mutations de celui-ci ainsi qu'une mutation supplémentaire en position 169. J'ai ainsi introduit cette mutation dans toutes les constructions précédemment réalisées afin d'évaluer la contribution des mutations individuelles du G12 pour effectuer la même analyse en présence de la mutation supplémentaire S169N.
iv. Une amélioration de l'efficacité du M36 via une approche combinée de l'expression du M36 et de réduction de l'expression de la RRM1 qui, comme présenté dans la première partie de ce manuscrit, intervient dans les mécanismes de résistances des cellules aux médicaments.
v. Une modification de la séquence du M36 sur une base rationnelle dans le but d'identifier une variante de la protéine présentant une meilleure efficacité.

# II Matériels et méthodes

## II.1 Lignées cellulaires

Les cellules HEK-293T ont été cultivées dans le milieu DMEM (Gibco) additionné de 10% de FBS et de 100 U / ml de pénicilline-100 mg / ml de streptomycine (Invitrogen, CA, USA). Les cellules Messa10K, Jurkat, MT4, THP-1, K562, HCT116 et DAMI ont été cultivées dans du milieu RPMI (Gibco) supplémenté avec 10% de FBS et de pénicilline-streptomycine. Les cellules HT 29 ont été cultivées dans un milieu Mc Coy's 5A (Gibco) additionné de 10% de FBS et de 100 U / ml de pénicilline-100 mg / ml de streptomycine. Les cellules ont été maintenues en culture dans un incubateur à 37°C avec 5% de $CO_2$.

## II.2 Constructions des mutants de la dCK

Les différents mutants ont été construits par PCR reconstituve dont le schéma est présenté dans la figure 29 et comme décrit précédemment[227]. Le gène de la dCK a été amplifié par deux couples d'amorces afin de générer deux fragments du gène. Les amorces internes utilisées contiennent la/les mutations à introduire et sont complémentaires afin de générer une région homologue sur 30 nt. Les deux fragments générés ont été purifiés sur colonne avant d'utiliser 0,25pmol de chaque fragment pour la seconde PCR. Les 8 premiers cycles de la seconde PCR ont été réalisés en absence d'amorces afin de favoriser la reconstitution d'un ADN de taille complète par extension des deux fragments qui servent réciproquement d'amorce l'un pour l'autre. Les amorces externes précédemment utilisées ont ensuite été ajoutées et 22 cycles supplémentaires ont été réalisés. Les produits de PCR ont ensuite été introduits après une digestion enzymatique par les enzymes BamHI et XhoI, dans un plasmide pSDYΔLTR[228] qui permet l'expression de l'ARN génomique des vecteurs lentiviraux (figure 30).

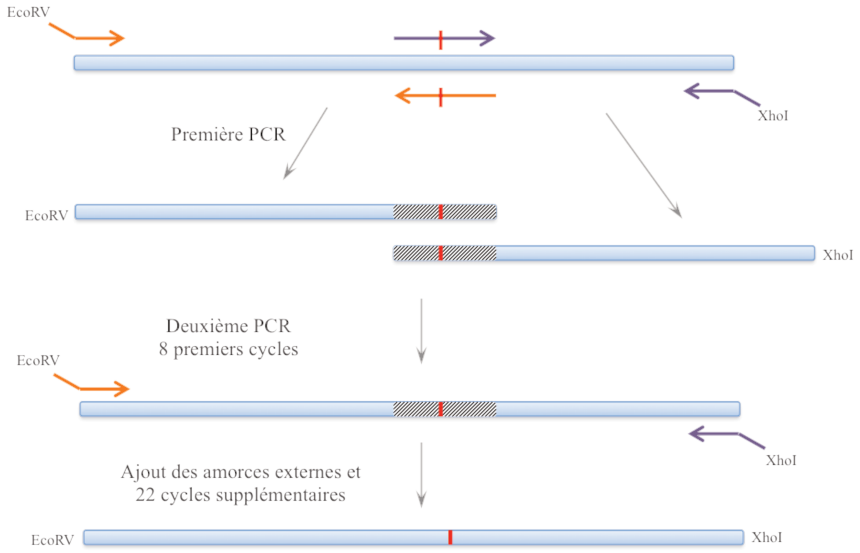



**Figure 29 : Représentation schématique de la réaction de PCR reconstituée.**

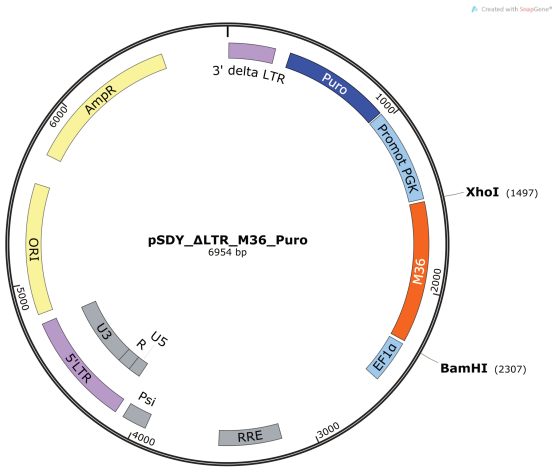



**Figure 30 : Carte du plasmide pSDY permettant l'expression de l'ARN génomique des vecteurs.**

## II.3 Production des particules virales

Les particules lentivirales ont été produites par la co-transfection des cellules HEK 293T. Les cellules ont été ensemencées la veille dans des boites de culture de 10 cm de diamètre. Seize heures après l'ensemencement, lorsque les cellules sont à 80% de confluence, le milieu de transfection a été rajouté au milieu de culture des cellules. Ce milieu de transfection est composé de 2 µg de plasmide p8.91 (GenBank ID : JQ585717.1) qui code les protéines de capside, la polymérase ainsi que l'intégrase virale, de 2 µg de plasmide pSDY qui permet la

synthèse de l'ARN génomique et d'1 µg de plasmide pVSV-G (GenBank ID : AJ318514.1) qui code pour la glycoprotéine d'enveloppe G du virus de la stomatite vésiculeuse, ainsi que du polyéthylènimine (PEI) linéaire MW~25 000 (Polysciences, Inc). Le surnageant viral a été récolté 48 h après la transfection, filtré à l'aide d'un filtre de 0,45 µm et concentré à l'aide de Vivaspin®20 (Sartorius Stedim Biotech). Le titre viral a ensuite été estimé par la quantification de la protéine de capside p24, par test ELISA (InnotestTM HIV Antigen mAB, International Genetic Technologies, Chilly-Mazarin, France).

### II.4 Etablissement des lignées cellulaires

Les cellules ont été mises en contact avec des quantités croissantes de vecteurs dans une plaque de culture de 6 puits en présence de 8 µg/mL de polybrène (Sigma-Aldrich). Vingt-quatre heures post-transduction le milieu de culture des cellules a été remplacé par du milieu contenant 0,6 µg/mL de puromycine afin de sélectionner les cellules ayant intégré le provirus. Celles-ci sont ensuite maintenues en culture et mise en expansion.

### II.5 Tests de sensibilité aux médicaments

Les cellules ont été ensemencées à 5000 cellules par puits dans une plaque à 96 puits et cultivées toute la nuit à 37 ° C. Deux rangées ont été utilisées pour chaque population. Douze heures après l'ensemencement, des concentrations croissantes de gemcitabine (0-400 nM ou 0-100 µM), ou AraC (0-10 mM) ont été ajoutés à chaque puits et laissés en culture pendant encore 72 h. La viabilité cellulaire a été mesurée par un test MTT (*Cell Titer 96 Non-Radioactive Cell Proliferation Assay*, Promega) et le nombre de cellules vivantes dans chaque puits a été évalué en mesurant la DO à 570 nm. Pour chaque population, la fraction de cellules vivantes a été calculée en faisant le rapport entre la DO mesurée pour la concentration X de Gem et la DO mesurée en absence de Gem dans le puit. Ce qui permet d'estimer la sensibilité de la population à la prodrogue.

### II.6 siARN et shARN

Les séquences des siARN ont été choisies afin de cibler la région 5' du gène codant la RRM1. Les siARN ont été transfectées avec le kit DharmaFECT (Dharmacon). Le siARN contrôle

utilisé est le « *Universale negative control #1* » (SIGMA). Les séquences des siARN ainsi que celle des shARN sont présentés dans *le tableau 1*.

**Tableau 1 : Séquences utilisées pour la répression de RRM1**

| | |
|---|---|
| siARN 1 | 5' AAUUACAUCUCGAAUCCAGUU 3' |
| siARN 2 | 5' AAUAUGGAUUUUGUUGAUCCUGUU3' |
| siARN 3 | 5' AACUAGAUACUUUGGCUGCUU3' |
| siARN 4 | 5' AAGUGUUCAGUGAUGUGAUGGUU3' |
| shARN 1 | 5' CCGGTAATTACATCTCGAATCCAGTTCAAGAGACTGGATTCGAGATGTAATTTTTTTGGAAG 3' |
| shARN 4 | 5' CCGGTAAGTGTTCAGTGATGTGATGGTTCAAGAGACCATCACATCACTGAACACTTTTTTTGGAG 3' |

### II.7 Analyse des western blot

Les cellules ont été lysées dans un tampon RIPA 1X, et 30 ou 60 µg de protéines totales (évaluées par Bradford) sont chargés sur un gel de polyacrylamide à 10% pour chaque extrait cellulaire. Après le transfert sur une membrane PVDF, les protéines RRM1 ont été révélées avec une dilution 1: 4000 d'un anticorps polyclonal anti-RRM1 (lapin, Sigma-Aldrich) et un anticorps secondaire anti-lapin conjugué à la HRP (BioRad) 1: 5000 additionné d'un anticorps anti-actine 1: 20 000 (BioRad). La détection a été réalisée par autoradiographie. L'analyse du niveau d'expression des protéines a été réalisée avec le logiciel Image Lab™ Software de BioRad. Le niveau d'expression de la RRM1 a été normalisé pour chaque échantillon en fonction du niveau d'expression de l'actine dans le même échantillon.

## III Résultats

Les résultats obtenus sur l'étude des mutations de la protéine G12 ainsi que l'étude de la capacité du M36 à sensibiliser les cellules à d'autres analogues de la dC font l'objet d'un article paru en 2015 et joint ci-dessous.

### III.1 Article

Résumé de l'article

Identifier une enzyme qui, une fois exprimée dans les cellules cancéreuses permettrait de conduire à leur élimination spécifique avec de faibles doses de médicaments, est l'un des objectifs principaux des stratégies de lutte anti-tumorale. Le laboratoire a identifié un mutant de la dCK appelé G12 qui est capable de sensibiliser différentes lignées cellulaires à la Gem, un analogue nucléosidique de la dC utilisé en chimiothérapie. L'identification de ce mutant a été possible grâce à une approche originale, appelée la rétrovolution et mise en place par le laboratoire. La rétrovolution se base sur l'utilisation de vecteurs lentiviraux à réplication conditionnelle pour introduire des mutations dans le gène à faire évoluer. Le mutant G12 ainsi que l'approche ayant permis sa découverte ont été décrit dans un précédent papier[59]. Dans cette article, le mutant G12 a également été utilisé comme point de départ pour une nouvelle procédure de rétrovolution, et cela à conduit à l'identification d'une autre variante de la dCK, appelé M36, qui permet une meilleure sensibilisation des cellules, comparé au G12. Par rapport au G12, le M36 présente une mutation supplémentaire située au niveau de l'interface de dimérisation de la protéine. La présence de cette mutation supplémentaire permet de diminuer l'$IC_{50}$ de la Gem et donc d'augmenter la sensibilité des cellules au médicament. De plus, l'utilisation de vecteurs SIN permet une augmentation de l'efficacité de la protéine, ce qui est cohérent avec le phénomène d'interférence transcriptionnelle sur les promoteurs internes décrit lorsque les séquences LTR sont transcriptionnellement actives. Enfin, l'effet le plus marqué observé avec le mutant M36 est qu'il permet de diminuer de 10 000 fois l'$IC_{50}$ de la cytarabine (AraC), un composé utilisé pour le traitement des leucémies myéloïdes aigües. En permettant de sensibiliser les cellules à la Gem et à l'AraC, deux composés couramment utilisés pour le traitement de cancer à faible pronostic, le M36 se positionne comme un candidat gène suicide prometteur.

PLOS | ONE

RESEARCH ARTICLE

# Potent Sensitisation of Cancer Cells to Anticancer Drugs by a Quadruple Mutant of the Human Deoxycytidine Kinase

**Safiatou T. Coulibaly[1☯], Paola Rossolillo[1☯], Flore Winter[1], Franziska K. Kretzschmar[1], Mélanie Brayé[1], Darren P. Martin[2,3], Daniela Lener[1], Matteo Negroni[1]***

**1** Architecture et Réactivité de l'ARN, CNRS, IBMC, Université de Strasbourg, 15 rue René Descartes, 67084 Strasbourg, Cedex, France, **2** Centre for High-Performance Computing, Rosebank, Cape Town, South Africa, **3** Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa

☯ These authors contributed equally to this work.
* m.negroni@ibmc-cnrs.unistra.fr

CrossMark
click for updates

OPEN ACCESS

**Citation:** Coulibaly ST, Rossolillo P, Winter F, Kretzschmar FK, Brayé M, Martin DP, et al. (2015) Potent Sensitisation of Cancer Cells to Anticancer Drugs by a Quadruple Mutant of the Human Deoxycytidine Kinase. PLoS ONE 10(10): e0140741. doi:10.1371/journal.pone.0140741

**Editor:** Jean-Luc EPH Darlix, Institut National de la Santé et de la Recherche Médicale, FRANCE

**Received:** June 11, 2015

**Accepted:** September 30, 2015

**Published:** October 20, 2015

**Copyright:** © 2015 Coulibaly et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This work was supported by funding from the Ligue Contre le Cancer (http://www.ligue-cancer.net), Appel d'offre Conférence de Coordination InterRégionale du Grand Est 2012 to MN. S. Coulibaly is recipient of a fellowship from the French "Ministère de l'Enseignement Supérieur et de la Recherche".

**Competing Interests:** The retrovolution method and the mutants described in the present paper were

## Abstract

Identifying enzymes that, once introduced in cancer cells, lead to an increased efficiency of treatment constitutes an important goal for biomedical applications. Using an original procedure whereby mutant genes are generated based on the use of conditional lentivector genome mobilisation, we recently described, for the first time, the identification of a human deoxycytidine kinase (dCK) mutant (G12) that sensitises a panel of cancer cell lines to treatment with the dCK analogue gemcitabine. Here, starting from the G12 variant itself, we generated a new library and identified a mutant (M36) that triggers even greater sensitisation to gemcitabine than G12. With respect to G12, M36 presents an additional mutation located in the region that constitutes the interface of the dCK dimer. The simple presence of this mutation halves both the IC50 and the proportion of residual cells resistant to the treatment. Furthermore, the use of vectors with self-inactivating LTRs leads to an increased sensitivity to treatment, a result compatible with a relief of the transcriptional interference exerted by the U3 promoter on the internal promoter that drives the expression of M36. Importantly, a remarkable effect is also observed in treatments with the anticancer compound cytarabine (AraC), for which a 10,000 fold decrease in IC50 occurred. By triggering the sensitisation of various cancer cell types with poor prognosis to two commonly used anticancer compounds M36 is a promising candidate for suicide gene approaches.

## Introduction

Inducing death of cancer cells through gene therapy is a major goal for biomedical applications. The main obstacle to this achievement is the possibility of delivering specifically into cancer cells transgenes that lead to cell death: a process achieved with poor efficiency *in vivo* [1]. An additional difficulty is constituted by the efficiency at which the transgene induces the death of the modified cell. Cell death can be induced either constitutively or in response to exposure to chemical compounds, such as an anticancer drug [1]. The approach of using a gene with an

covered by patents, for more information regarding licenses please contact the corresponding author.

inducible toxicity has the advantage of being of interest also in the field of safety genes that allow negative selection of transplanted cells in gene therapy approaches. In these cases, the presence of a safety gene is essential in the event of a neoplastic transformation after engraftment of the engineered cells or, for adoptive immunotherapy, in the case of the development of a graft versus host response [2]. The suicide gene most frequently employed so far for these purposes has been the thymidine kinase gene of herpes simplex virus (hsTK) coupled with gancyclovir treatment [3–5]. However, due to its viral origin, the hsTK presents the drawback of inducing an immune response to TK-derived epitopes [6], suggesting that its replacement by less immunogenic proteins could potentially increase the efficiency of this approach. For these reasons, a human enzyme would constitute an ideal candidate.

In this regard, human deoxycytidine kinase (dCK) has attracted a lot of interest. Besides being involved in the salvage pathway that converts recycled deoxyribonucleosides into dNTP [7], this enzyme catalyzes the first rate-limiting, phosphorylation step for the activation of different deoxycytidine analogs (dCa) used in clinical treatments of various cancers. Gemcitabine is one of these drugs. It is generally used for the treatment of several cancers such as pancreatic cancer, metastatic non-small cell lung and breast cancers, as well as ovarian cancers; all of which are associated with poor prognosis [8–13]. The dCK is also involved in the activation of other compounds, that are structurally related to deoxycytidine and are used as anti-cancer or anti-viral drugs. One such drugs is AraC (cytosine arabinoside), which is mostly used for the treatment of leukaemias, such as acute myeloid leukaemia (AML) [14].

In many cases the dCK determines the degree of sensitivity of cells to treatment with gemcitabine or AraC. A correlation between the level of dCK expression and sensitivity to treatment with dCa has in fact been observed in patients treated for hairy-cell leukaemia and chronic lymphocytic leukaemia [15] and in various breast cancer cell lines [16]. A link between the expression of dCK, the accumulation and retention of intracellular gemcitabine pool levels and gemcitabine incorporation into DNA has also been reported [17–19], as well as an association between pre-treatment levels of intracellular dCK and sensitivity to gemcitabine [20]. A similar correlation has been reported for AraC [21]. Moreover, the exogenous over expression of dCK in glioma cells by retroviral or adenoviral vectors leads to increased sensitivity to the cytotoxic action of cytosine arabinoside *in vitro* and *in vivo* [22]. Conversely, deficiencies in dCK activity lead to resistance to treatment either with gemcitabine [18, 23, 24] or with AraC [25–27].

Thereby, the possibility of foreseeing gene therapy approaches based on the insertion of extra copies of the dCK gene in cancer cells appears promising and the possibility of generating variants of dCK with improved dCa phosphorylation activity is a major goal in the field. Rational attempts to design dCK mutants with such properties have been made in the past, either based on structural information [28] or on the effect of post-translational modifications of the protein that have been observed to enhance its enzymatic activity *in vivo* [29, 30]. Although these studies succeeded in obtaining dCK variants with an improved capacity to phosphorylate and activate gemcitabine, this activity was paralleled by a much more marked enhancement of phosphorylation of the natural substrate dC [28, 30]. Since competition for the dCK enzyme between natural substrate and incoming drug molecules is crucial for the efficiency of cell death induction [31, 32], this feature suggested that such mutants once introduced into cancer cells would not have induced a sensitisation to the treatment. For one of these (the triple mutant A100V, R104M, D133A) [28], it was indeed reported that its introduction into Messa10K cells, uterine sarcoma cells resistant to gemcitabine [33], resulted in no sensitisation to this drug [34].

We recently developed an experimental approach (retrovolution) where a gene of interest is inserted into conditional replication-defective VSV-pseudotyped HIV-1 derived vectors that are used to mimic multiple infectious cycles, during which the sequence of interest accumulates

mutations due to the error-prone nature of the HIV-1 replication machinery [34]. This generates a library of variants of the gene of interest that are already inserted into a biological vector appropriate for an efficient delivery of the transgene to human cells, which can then be directly screened for the desired phenotype. Using, as a sequence of interest, the cDNA coding for the human dCK, we isolated a mutant, named G12, which induces a 300-fold sensitisation to gemcitabine in cells originally resistant to this drug (Messa 10K), an effect 60 times stronger than that observed in controls cells where an extra copy of the wt dCK had been inserted [34]. This mutant is characterised by the presence of three mutations, located in positions of the protein distant from the active site and which are therefore unlikely to have been predicted on a rational basis. One of these mutations, E171K, is located in a region involved in the generation of the interface of the dCK dimer. The other two mutations (E247K and L249M) instead are located in the "base-sensing loop" that modulates the folding of the protein upon binding of the phosphate donor (ATP or UTP), (S1 Fig) [28]. No alteration of the degree of dimerization was however observed *in vitro* for the mutant, suggesting that the reason for the increased gemcitabine sensitivity that it induces is not attributable to a change in its expression expression level [34], but is possibly due to a modification of the overall structure of the enzyme.

G12 was isolated after screening of the library resulting from 17 successive cycles of retrovolution [34]. This way, mutations were introduced randomly in the gene with no external selection pressure applied during the procedure that would favour the isolation of the desired phenotype. The question of the possibility of isolating a mutant with an even greater degree of gemcitabine sensitisation activity than G12 is the starting point of the present study.

## Materials and Methods

### Cell lines

HEK-293T cells were obtained from the American Type Culture Collection and grown in Dulbecco's Modified Eagle's Medium (Gibco, Thermo Fischer Scientific, Waltham, MA, USA) supplemented with 10% FBS and 100U/ml pennicillin-100 mg/ml streptomycin. Messa10K cells were kindly provided by L.P.Jordheim (Lyon, France) and grown in RPMI medium supplemented with 10% FBS and penicillin-streptomycin.

### Retrovolution cycles by lentivector particles

For starting the experiments leading to the generation of the F27-RL (where RL stands for random library), eight 10-cm plates containing $5x10^6$ cells from the population of HEK 293T cells that was used to identify the G12 clone (F16, reference [34]) were transfected, using the Calcium Phosphate protocol, with 10 μg of pCMVΔR8.91 plasmid [35] and 5 μg of pHCMV-G plasmid [36], in order to generate the viral-like population F17-RL. The same procedure was used for starting the experiments leading to the generation of the F11-DL, but in this case the initial transfection with pCMVΔR8.91 and pHCMV-G plasmids was performed on a HEK 293T cell line previously generated to stably express the lentivector genomic RNA coding G12 [34]. In this case the viral-like population generated was called F1-DL (where DL stands for directed library). These viral-like populations (F17-RL and F1-DL) were then used to independently transduce fresh HEK 293T cells to continue with the retrovolution procedure in parallel, as described in reference [34]. We performed 11 further retrovolution cycles, leading to the generation of the F27-RL and of the F11-DL cell populations, respectively. Transfections during the cycles of retrovolution were achieved with 10 μg of pCMVΔR8.91 and 5 μg of pHCMV-G plasmid on 8 10-cm plates containing $5x10^6$ cells from the population of HEK 293T cells from the previous retrovolution cycle. For the transduction procedure during the cycles of retrovolution, the lentivector-containing supernatant was collected 48h after

transfection, filtered through a 0.45 μm filter and concentrated 40 times in Vivaspin 20 columns (MWCO 50 KDa, Sartorius Stedim Biotech, Aubagne, France). Transduction conditions differed when performed to generate genetic diversity (retrovolution cycles) and when performed to isolate single clones for screening of the library. For the retrovolution cycles, $5x10^6$ HEK-293T cells were transduced with 1 ml of concentrated lentiviral vectors (see above) and puromycin selection was applied 24h after transduction. Under these conditions 90–100% of the transduced cells were resistant to puromycin, indicating that at least for the majority of the cells, the multiplicity of transduction was higher than one. Transductions for the isolation of single clones were, instead, performed at varying dilutions of the lentiviral vectors based on the titers estimated by ELISA directed against the HIV-1 p24 capsid protein (Innotest HIV Antigen mAb, Innogenetics, Gent, Belgium). A hundred cells were then seeded per well in 96-well plates in DMEM. Puromycin selection was applied 24h after transduction by adding puromycin at concentrations of 0.6 μg/ml for HEK-293T cells, 0.6 μg/ml for HEK-293T-$CD4^+$ cells, 0.5 μg/ml for Messa10K cells. We selected conditions where, on average, 5 wells per plate resulted positive to cell growth in the presence of puromycin. Under these conditions, the multiplicity of transduction was of $5x10^{-4}$ per cell ensuring that in each well the cell population was generated by transduction with a single lentiviral vector.

## Amplification and sequencing of the dCK-encoding portion of the proviral DNA

Puromycin resistant clones from the F27-RL and F11-DL generations were isolated by limiting dilution and expanded in 6 wells plates. Cells from the individual clones were then recovered and lysed with 250 μl of Cell Direct PCR (Viagen Biotech, Los Angeles, CA, USA). The dCK transgene was amplified from 1 μl of the lysate with oligonucleotides on the vector, EF1 (5'-gatgtcgtgtactggctccg-3') and PGK (5-gatgtggaatgtgtgcgagg-3'), flanking the transgene. Sequencing of the PCR fragment was performed by the GATC sequencing service (GATC Biotech, Konstanz, Germany).

## Calculation of the frequency of mutations in the libraries

For the F27-RL and F11-DL libraries the number of mutations found in the dCK coding sequence was calculated starting from the data given in Table 1 as: (m/(783 x c))/y, where m is the number of mutated positions, 783 is the size of the dCK coding-sequence in base pairs, c is the number of clones sequenced and y is the number of cycles of retrovolution carried out. For the F11-DL library the starting sequence for calculating the mutation rates was the sequence of G12.

## Permutation tests for the occurrence of synonymous and non synonymous mutations along the dCK protein

A permutation test was performed by simulating the occurrence of mutations in the two halves of the protein coding sequence (i.e. on either side of codon 130) based on the observed frequency of occurrence of random mutations and of preferential mutations in APOBEC3F consensus sequences (29.1 and 70.9%, respectively, see Table 1). The relative incidence of synonymous and non-synonymous mutations obtained experimentally was compared to that of 10,000 simulated datasets. The proportion of simulated datasets with non synonymous/synonymous (dN/dS) mutation ratios lower than or equal to those of the real dataset were then taken as the probability that there was a significant tendency in the region considered for the occurrence of more synonymous mutations than was expected on a stochastic basis.

**Table 1. Mutation panel found in the dCK coding-sequence in F27-RL and in F10-DL.**

| | Clones sequenced | Positions mutated | Frequency of mutation (per nt) | % of G>A mutations | Mutations/ clone | Cycles of retrovolution |
|---|---|---|---|---|---|---|
| **F27-RL** | 41 | 57 | $6.34 \times 10^{-5}$ | 70.9 | 1–11 | 28 |
| **F11-DL** | 43 | 29 | $7.83 \times 10^{-5}$ | 75.9 | 1–5 | 11 |

Frequencies of mutation are calculated as described in Materials and Methods; %G>A mutations: % of transitions from G to A among the number of mutated positions found; mutations/clone: range of mutations found in single clones (for the F10-DL library the three mutations present in G12 are not considered).

doi:10.1371/journal.pone.0140741.t001

Conversely, the proportion of simulated datasets with dN/dS ratios higher than or equal to those of the real dataset was then taken as the probability that there was a significant tendency in the region considered for the occurrence of more non-synonymous mutations than was expected on a stochastic basis.

## Tests of sensitivity to anticancer drugs

Cells harbouring integrated lentiviral vector genomes (G12, M36, single or double mutants) were seeded at 5 000 cells per well in a 96-well plate and grown overnight at 37°C. For each population experiments were carried out in duplicate. Twelve hours after seeding, increasing concentrations of gemcitabine (0–400 nM), or AraC (0–10 mM) were added to each well and left in culture for a further 72h. Cell viability was measured using the MTT test (CellTiter 96 Non-Radioactive Cell Proliferation Assay, Promega, Fitchburg, WI, USA) and the number of living cells in each well was evaluated by measuring the OD at 570 nm. For each population, the fraction of living cells was calculated as OD570 concentration X Gem/OD570 concentration 0 Gem, to estimate the sensitivity of the population to the prodrug.

## Tests of sensitivity to antiviral drugs

Lentiviral vectors were generated by transfection, using the calcium phosphate protocol, of HEK-293T cells with 10 μg of pCMV ΔR8.91 plasmid [35], 5 μg of pHCMV-G plasmid [36], and 10 μg of genomic plasmid SDY-dCK (described in Rossolillo et al. [34]) in its SIN version (see results) encoding either for the wild type sequence of the human dCK or the M36 variant. Several transductions, using different amounts of lentiviral vector, were performed on $1 \times 10^6$ HEK-293T-CD4+ cells [37] and transduced cells were selected in the presence of 0.6 μg/ml of puromycin, added 24h after transduction. Only transduction conditions giving 50 to 200 individual puromycin-resistant clones (each clone derived from a cell transduced by a single lentiviral vector) were retained and all the clones were pooled and expanded to obtain a polyclonal HEK-293T-CD4+ cell population either expressing one extra copy of the wt dCK enzyme or one copy of M36 (HEK 293T/$CD4^+$-SINvec-wtdCK or HEK 293T/$CD4^+$-SINvec-M36, respectively). Vectors for monitoring the efficiency of transduction in the presence of antiviral drugs were generated by transfection of twelve 10-cm plates containing $5 \times 10^6$ HEK 293T cells with 10 μg of pTopo plasmid into which the sequence coding for the HIV-1 envelope ADA [38] had been inserted [39], and 10 μg of pNL4.3-$Env^-$Luc+, that encodes the firefly luciferase, as previously described [40]. The resulting lentivectors were concentrated as previously described [34]. Transduction of $2.5 \times 10^5$ HEK 293T/$CD4^+$-SINvec-wtdCK cells and HEK 293T/$CD4^+$-SINvec-M36 cells were performed in parallel using 1 ml each of concentrated lentiviral vector. Five hours after transduction cells were seeded at $5 \times 10^4$ cells/well in a 6 well plate in the presence of increasing concentrations of the anti-viral compounds ddC or 3TC. The efficiency of

transduction was then monitored 48 hours after transduction through expression of the luciferase gene. For this, the medium was removed, the cells were washed twice in PBS, lysed, centrifuged and the supernatant was then used to measure luciferase activity according to the manufacturer's protocol (Promega, Fitchburg, WI, USA) with a Glomax luminometer (Promega, Fitchburg, WI, USA), as previously described [40].

### Western blot analyses

Messa10K cell lines expressing either M36, both in the context of a lentiviral vector DNA with a functional LTR or with an inactivated LTR were lysed in 1X RIPA buffer (25mM Tris-HCl, 1% IGEPAL, 0,1% SDS, 0,5% Sodium deoxycholate, 150mM NaCl, protease inhibitor) and 30, 60, and 120 μg of total protein (evaluated by Bradford assay) for each cell type were loaded on a 12% bis-tricine gel (Invitrogen, Carlsbad, NM, USA). After transfer on PVDF membrane, the dCK proteins were analysed by western Blot with 1:4000 dilution of a polyclonal anti-dCK antibody (rabbit, Sigma-Aldrich, St. Louis, MO, USA) and 1:3000 dilution of an anti-rabbit HRP conjugated secondary antibody (BioRad, Hercules, CA, USA) with binding being detected by chemiluminescence (Pierce ECL Western Blotting substrate, Thermo Scientific).

### Production and purification of recombinant dCK proteins

The wt-dCK and M36 sequences were cloned in the pET14b plasmid and expressed in *E. coli* cells BL21 DE3 pLysE. Protein expression was induced by adding 0.1mM IPTG and cells were collected after 4h of growth at 37°C. His-tagged proteins were eluted with 250 mM imidazole from His-Trap TM FF Columns (GE Healthcare Life Science, Little Chalfont, UK), the Histidine tag was removed using the S-Tag Thrombin Purification Kit (Novagen, Madison, WI, USA), and dCK and M36 were further purified by gel filtration on S-200 Sephacryl columns (GE Healthcare Life Science, Little Chalfont, UK).

### Phosphorylation tests *in vitro*

The efficiency of phosphorylation of the natural substrate deoxycytidine and of the prodrugs gemcitabine and AraC were measured for purified wt-dCK and M36 in a NADH-based assay as previously described [41]. All reagents were purchased from Sigma-Aldrich (St. Louis, MO, USA) except gemcitabine (Lilly, Indianapolis, IN, USA). Enzymes were assayed at room temperature at a concentration of 0.9 μM for dC, and 0.3 μM for both gemcitabine and AraC. dC was used at concentrations between 5 and 50 μM, gemcitabine at concentrations between 20 μM and 1mM and AraC at concentrations between 10 μM and 0.5 mM. ATP was used at a concentration of 4 mM.

### Statistical analyses

To evaluate whether the average amounts of cell death occurring in the Messa10K populations containing different dCK variants were significantly different from the value shown by Messa10K wt-dCK, a two sample t-test was applied for the different concentrations of Gemcitabine.

## Results

### Contribution of the individual mutations of G12 to sensitisation to gemcitabine

Using a new method of directed human gene evolution that we have developed in our laboratory, we previously described the isolation and characterisation of a dCK mutant, named G12,

that induces sensitisation of cancer cells to treatment with the anticancer compound gemcitabine [34]. G12 is characterized by three mutations with respect to the wild type dCK enzyme. However, it has remained undetermined whether the sensitisation activity observed in G12 depends on the presence of all three mutations and, if so, what the relative contribution of each of these mutations to the G12 phenotype is. To address these issues, we constructed six mutants, three containing either the E171K, the E247K, or the L249M mutation alone, and three containing two mutations each, covering all the possible combinations of double mutants (E171K/E247K, E171K/L249M, and E247K/L249M). The mutants were inserted in the previously described pSDY plasmid [34], and used for transfection with transcomplementation plasmids as described in the Materials and Methods, in order to generate lentiviral vectors. With respect to the pSDY plasmids we previously described [34], the plasmids we used (pSDY-SIN) contained a partially deleted and therefore non-functional version of the 3' U3 sequence (Fig 1A). The resulting lentiviral vectors will generate, after reverse transcription, non-functional LTR sequences and are therefore called self-inactivating (SIN) vectors (SDY-SIN in our case). These lentivectors were chosen for this test for their pertinence for biomedical applications. As controls, the G12 and the wt dCK sequences were also inserted in pSDY-SIN plasmid. For each variant of dCK, the lentivectors were used to establish a polyclonal population of Messa 10K cells. These polyclonal populations (SINvec-G12 population for G12, SINvec-wt dCK for wt dCK, SINvec-E171K for the E171K mutants, and so on) were generated by transducing $10^6$ HEK 293T cells with different amounts of lentiviral vector in order to obtain, after puromycin selection, 50 to 200 individual clones. This way, each clone was almost certainly derived from a cell containing a single integrated copy of proviral DNA. The clones were then pooled ensuring that the resulting population was constituted by cells carrying the same transgene, but with the gene integrated at different locations in the genome, thereby averaging over possible effects that might be attributable to the integration site. The fact that the cells contained a single copy of the transgene allowed the exclusion, when comparing populations one with the other, of possible dose effects due to the number of transgene integrated. Sensitisation to gemcitabine treatment was then evaluated by MTT tests. To this end, the different Messa 10K polyclonal populations were μded into 96-well plates, exposed to varying concentrations of gemcitabine, and cell survival was measured for each condition as previously described [34].

The three single mutants (grey symbols Fig 1B) yielded response curves in the same range as the reference wt dCK population (white squares). The same results were obtained for the double mutants E171K/E247K, and E247K/L249M (grey triangles and circles in Fig 1C, respectively). All three of the single mutants and two of the three double mutants that were tested were unable to induce gemcitabine sensitisation more efficiently than was achievable with an extra copy of the wt dCK gene. The third double mutant, E171K/L249M (grey squares in Fig 1C), yielded a result similar to that observed for wt Messa 10K cells (white triangles in the Fig), suggesting that an inactivation of dCK activity occurred for this mutant, leaving the cells deprived of dCK activity (as is the case with the original Messa 10K line). In conclusion, since none of the single and double mutants displayed a response curve similar to that of G12 (white circles in the Fig), the simultaneous presence of the three mutations is apparently required to achieve the sensitisation phenotype observed with G12.

## Generation of libraries using "random" versus "directed" evolution

The G12 mutant had been isolated, starting from the wt dCK sequence, from a library of dCK variants obtained after 17 generations of retrovolution (called F16, the first generation being constituted by the parental vectors P [34]). Here we investigated whether performing retrovolution starting from the G12 sequence instead of the wt dCK sequence, would yield mutants

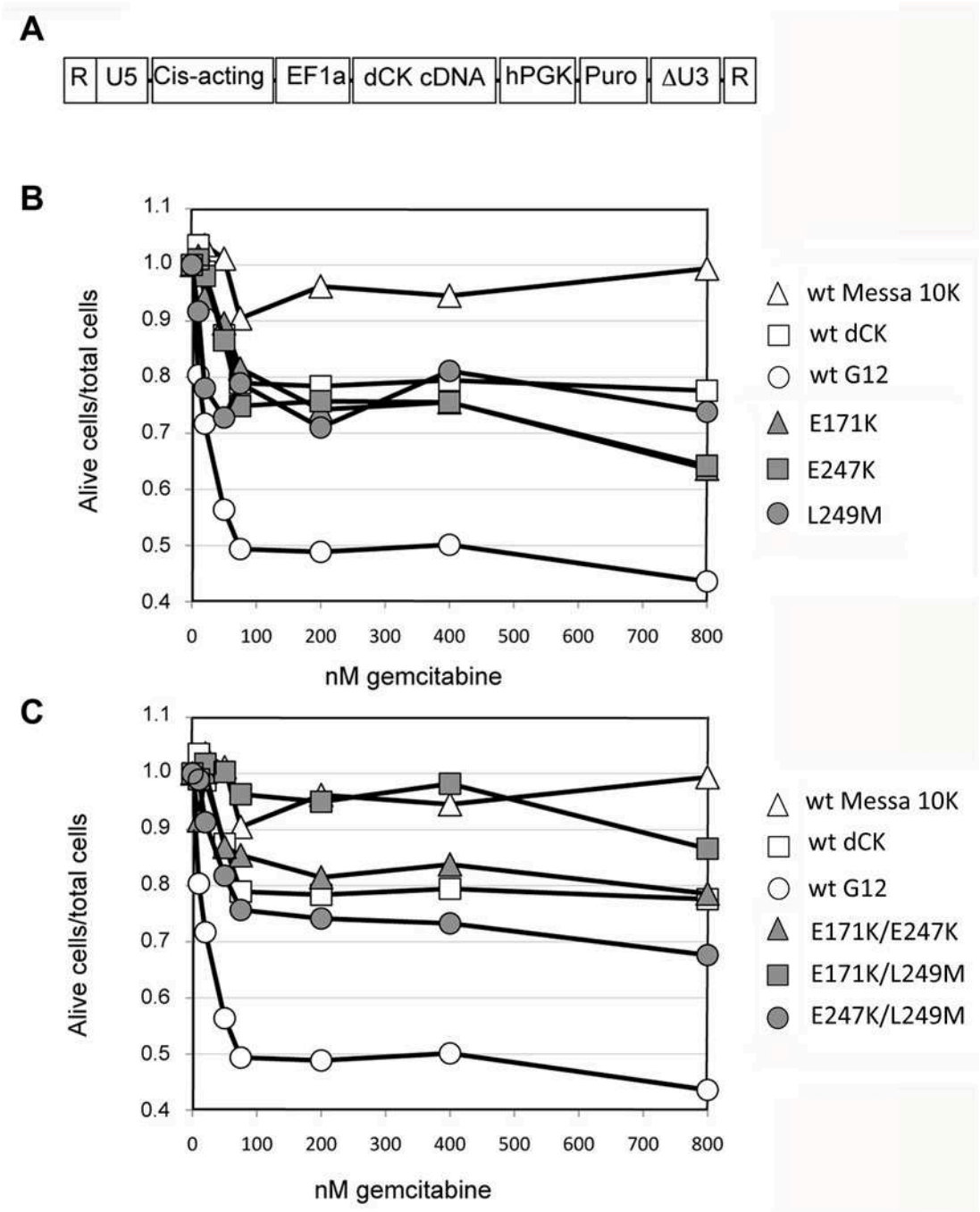



**Fig 1. Contribution of the individual mutations of G12 to the observed phenotype.** Panel A. Structure of the genomic RNA generated by transcription after transfection of cells with pSDY-SIN plasmids. R, repeated sequence from HIV-1 genome; U5, 5' unique sequence of HIV-1 LTR; Cis-acting, sequences required for packaging and reverse transcription of the genomic RNA; EF1-alpha, human elongation factor 1-alpha promoter; hPGK, human phosphoglycerate kinase promoter; Puro, puromycin N-acetyl-transferase gene; ΔU3, partially deleted version of the 3' unique sequence of HIV-1 LTR. Panels B and C. The proportion of alive cells over the total number of cells as function of the concentration in gemcitabine is given as a ratio with respect to the alive cells observed in the absence of drug. White symbols represent reference populations in both panels and. grey symbols represent single and double mutants (as described in the main text). Since single and double mutants were tested in parallel in the same experiment. The data are the average of 3 independent experiments. Error bars are not shown for the sake of clarity.

doi:10.1371/journal.pone.0140741.g001

with a further improved phenotype. For this purpose, the HEK 293T cell line stably expressing G12 that we previously generated [34], was used to start 11 cycles of retrovolution, by transfection with the transcomplementation plasmids pCMVΔR 8.91 and pHCMV-G [35, 36] as described in Materials and Methods and outlined in Fig 2. This led to the generation of the population of HEK 293T cell library we call F11-DL (for F11 directed library). In parallel, in order to evaluate the possibility of isolating mutants with an improved phenotype with respect to G12 simply by continuing the retrovolution procedure that led to the isolation of G12, we also performed a further 11 cycles of retrovolution on the entire F16 population from which G12 was isolated [34], so as to reach the F27 generation (we call F27-RL, for F27 random library).

To isolate individual clones from each library, for the last cycle of retrovolution the F10-DL and F26-RL virus-like populations were used to independently transduce fresh HEK 293T cells at a multiplicity of vector particles per cell that was substantially lower than 1. Transduced cells were seeded into 96-well plates, and selection with puromycin was applied as described in Materials and Methods, in such a way that each clone contained a single copy of the transgene. Individual clones were expanded and used for sequencing analysis. For the functional screening of the libraries, the HEK 293T clones isolated by this method were transfected with the transcomplementation plasmids, and the resulting virus-like populations were used to

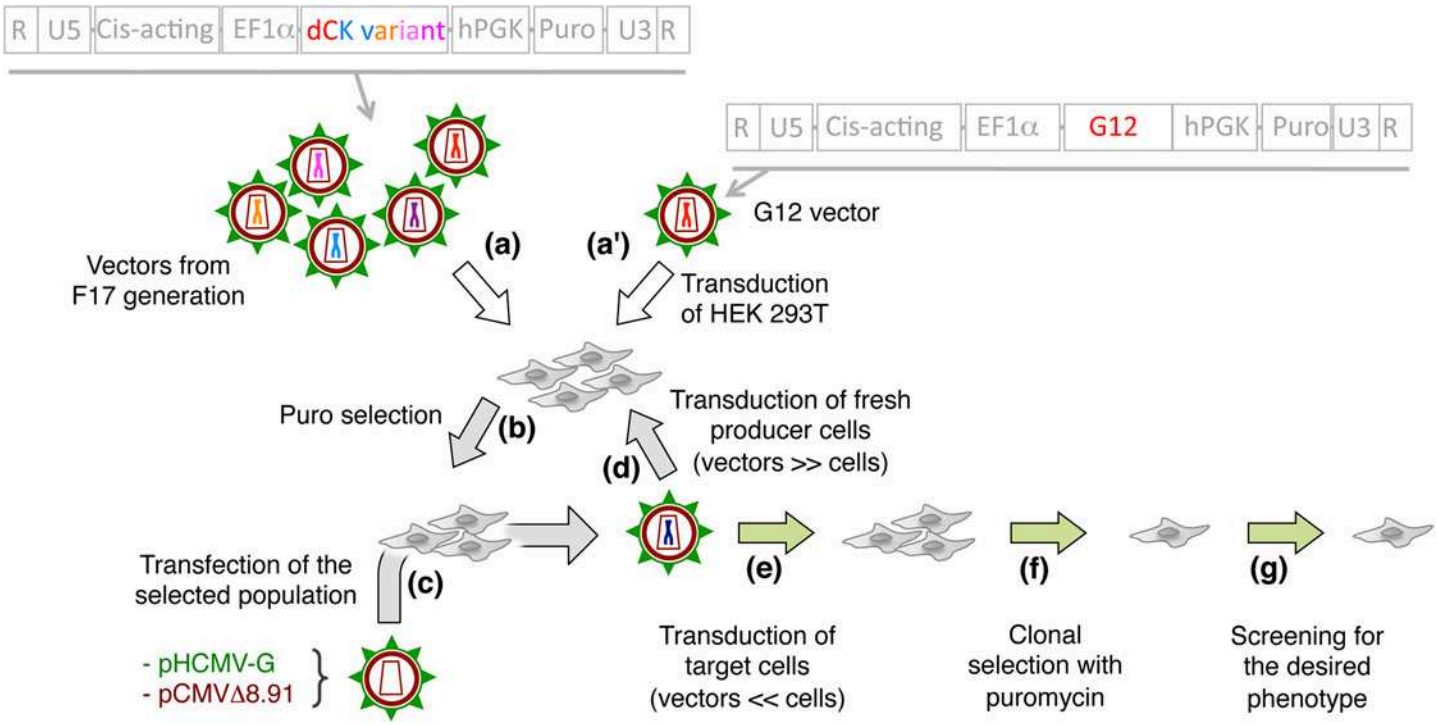



**Fig 2. Scheme of the procedure followed to generate F27-RL and F11-DL.** The general structure of the genomic RNAs used for retrovolution is given at the top, on the left for the F16 population and on the right for the population containing only the G12 variant. The procedure has been previously described [34] and consists in the repeated transduction of HEK 293T cells with the lentiviral vectors followed by the selection, with puromycin, of cell populations that have stably integrated the proviral DNA that has been produced by reverse transcription. Transfection of this population with the transcomplementation plasmids pHCMV-G and pCMVΔ8.91 (see main text) allows the production of the next vector generation that is then used to transduce fresh HEK 293T cells during repeated cycles of selection, transfection and transduction. For screening purposes, target cells are transduced by less than one lentivector particle per cell, and individual clones were isolated in the presence of puromycin. Screening for sensitisation to various compounds is carried out as described in the main text.

doi:10.1371/journal.pone.0140741.g002

transduce Messa 10K cells to generate polyclonal populations, as described above. The resulting populations were used for testing for the sensitivity to anticancer and antiviral compounds.

## Sequencing of the libraries

To evaluate the complexity of the library and characterise the nature of the mutations generated during the 11 cycles of retrovolution that were carried out in the DL and in the RL libraries, 41 HEK 293T clones from the F27-RL and 43 from the F11-DL were sequenced in the dCK-coding region, after PCR amplification from proviral DNA. The clones of the F27-RL library presented point mutations distributed all along the dCK gene, with 57 positions that were mutated (Table 1 and S1 Fig), resulting in a mutation frequency of 6.34 x $10^{-5}$ nt/cycle per site (a total of 28 cycles having been performed to create this library). G>A transitions represented 70.9% of the mutations, in line with the mutation bias of HIV-1 reverse transcriptase. Of note, the majority of these transitions fell in the typical recognition motif of APOBEC3F (GA>AA).

These mutations generated aminoacid changes all along the protein. However, while the distribution of the mutations was not different (p = 0.3657) between the two halves of the protein (aa 1–130 and 131–260) with 26 mutations in the first half, and 29 in the second, the majority of the mutations falling in the first half of the gene surprisingly corresponded to silent mutations (in black in Fig 3A). Namely, while in the first part 13 of the 25 positions mutated (52.0%) generated synonymous mutations (one mutation generating a stop codon was not considered), only 2 synonymous mutations were present among the 27 positions (7.4%) that were mutated in the second half of the protein (two mutations generating stop codons were not considered). In the first half of the protein, the ratio of non-synonymous mutations over synonymous (dN/dS) was significantly lower than expected under random mutation (p = 0.0019, as defined by permutation tests, see Material and Methods) indicating the occurrence of purifying selection on the N-ter half of the protein during the retrovolution cycles in HEK 293T cells. In sharp contrast, the C-ter half of the protein yielded dN/dS ratios indicative of the occurrence of significant (p = 0.00143) positive selection.

The analysis of the clones sequenced from the directed evolution library F11-DL (S2 Fig and Fig 3B) revealed that 14 clones out of 43 did not present additional mutations with respect to the three mutations observed in G12. The other 29 clones presented 1 to 5 additional nucleotide mutations (1 to 3 aminoacidic substitutions) for a total of 29 mutated positions, giving a mutation frequency of 7.83 x$10^{-5}$ (Table 1). As for the RL library, most mutations (75.9%) were G>A transitions. Overall, the pattern and frequency of mutations in the two libraries were similar to one another other. Also in this case an unbalanced distribution of non-synonymous versus synonymous mutations in the two halves of the protein was observed, with 40.0% of mutations in the first half being synonymous as opposed to only 14.3% being synonymous in the second half (note that the three mutations found in G12 were not considered here since they were present in the starting sequence).

## Screening of the libraries

Forty-one and twenty-three Messa 10K polyclonal populations from F27-RL and F11-DL, respectively, each derived from an individual clone of HEK 293T isolated as described above, were then screened for the induction of an increased sensitivity to the antitumor compound gemcitabine. In the F27-RL, no clones inducing sensitisation to treatment with gemcitabine were isolated. Only one clone presented the 3 G12 mutations together, but the association with other mutations generated an aberrant protein with no sensitisation phenotype. The analysis of the F11-DL library revealed that the association of the G12 mutations with the majority of the

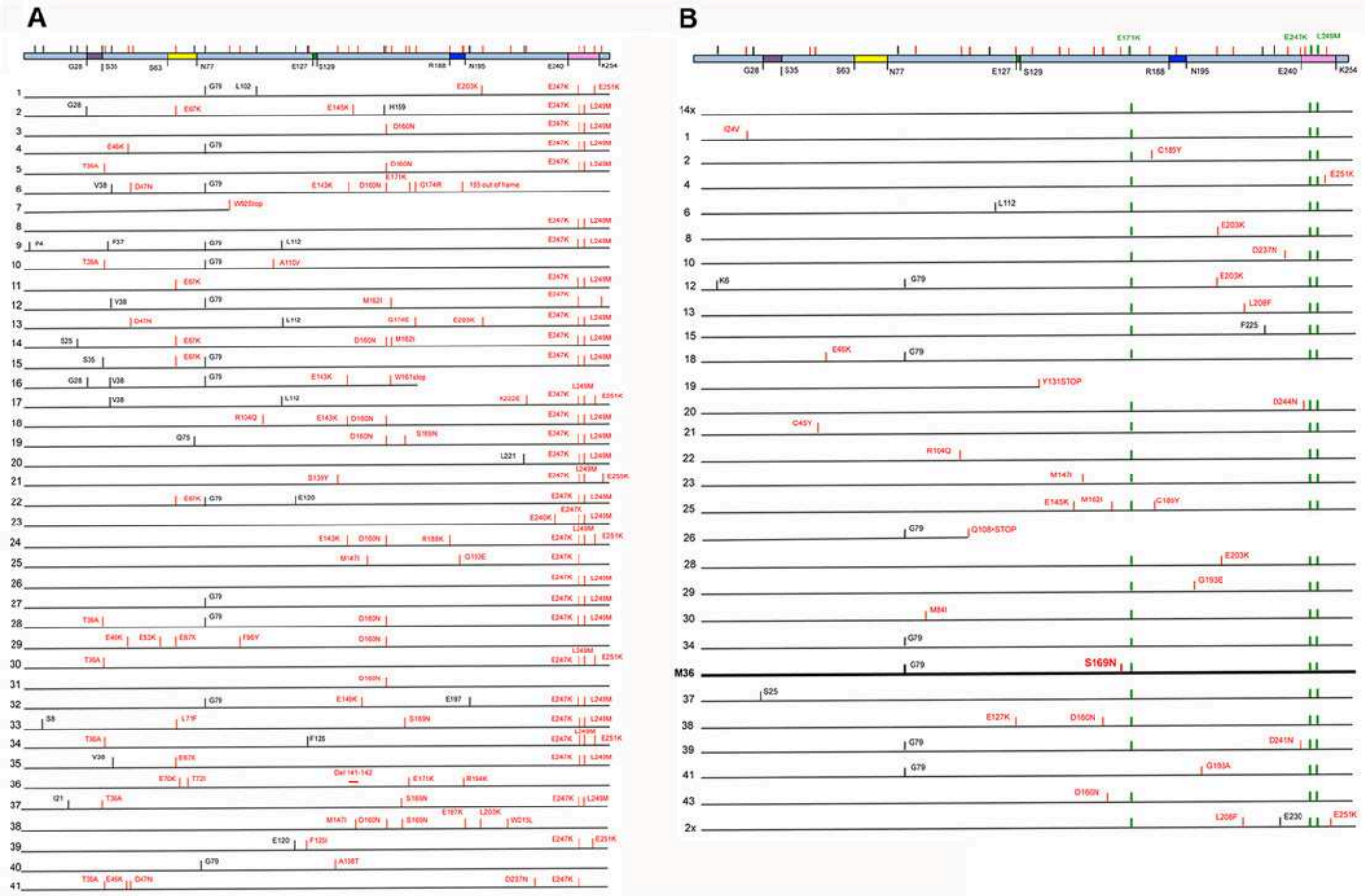



**Fig 3. Position of the mutations found in the individual clones.** In each panel, the dCK protein is represented (top of the drawing) as a pale blue box with the main structural motifs given in color (the first and last aminoacid of each motif is given below the box): purple, P-loop; yellow, insert; green, ERS; blue, lid; pink, base sensing loop; as defined in [28]. For each panel, a cumulative map of the positions that were mutated is provided at the top of the box with aminoacid substitutions given in red, while aminoacids that did not change but with associated codons that carried synonymous mutations given in black. Panels A and B respectively show the patterns observed for the random and directed libraries, respectively. Two deletion were also found: one single nucleotide deletion in clone 6 indicated as "193 out of frame", and a two-nucleotide deletion (positions 141–142) in clone 36 indicated by a red horizontal line. In Panel B, clone 36 (corresponding to M36 in the main text, for Mutant 36) is given in bold.

doi:10.1371/journal.pone.0140741.g003

other mutations along the protein either did not change the phenotype with respect to G12 or it completely abolished sensitisation to gemcitabine. One variant, though, named M36 (represented in Fig 3B), increased Messa 10K cell sensitivity to gemcitabine to a degree greater than that of G12 ($p<0.005$), which was tested in parallel (Fig 4A). Besides the three mutations typical of G12, this mutant presented the additional substitution S169N (Fig 3B). Therefore, four mutations clustered in pairs in the 169–171 (S169N, E171K) and in the 247–249 (E247K, L249M) regions (Fig 4B and S3 Fig) characterise the M36 mutant isolated from F11-DL.

## Characterisation of M36 in cell culture with a ΔLTR lentiviral vector

The need for functional LTRs is intrinsic to the retrovolution procedure. However, as mentioned above, the use of vectors carrying a self-inactivating LTR (called SIN vectors) is required in view of prospective biomedical applications. The expression of the transgene from a proviral DNA deprived of a functional promoter for the synthesis of the viral genomic RNA could, however, be altered, leading to a change in the profile of sensitisation of the transgenic cells. To

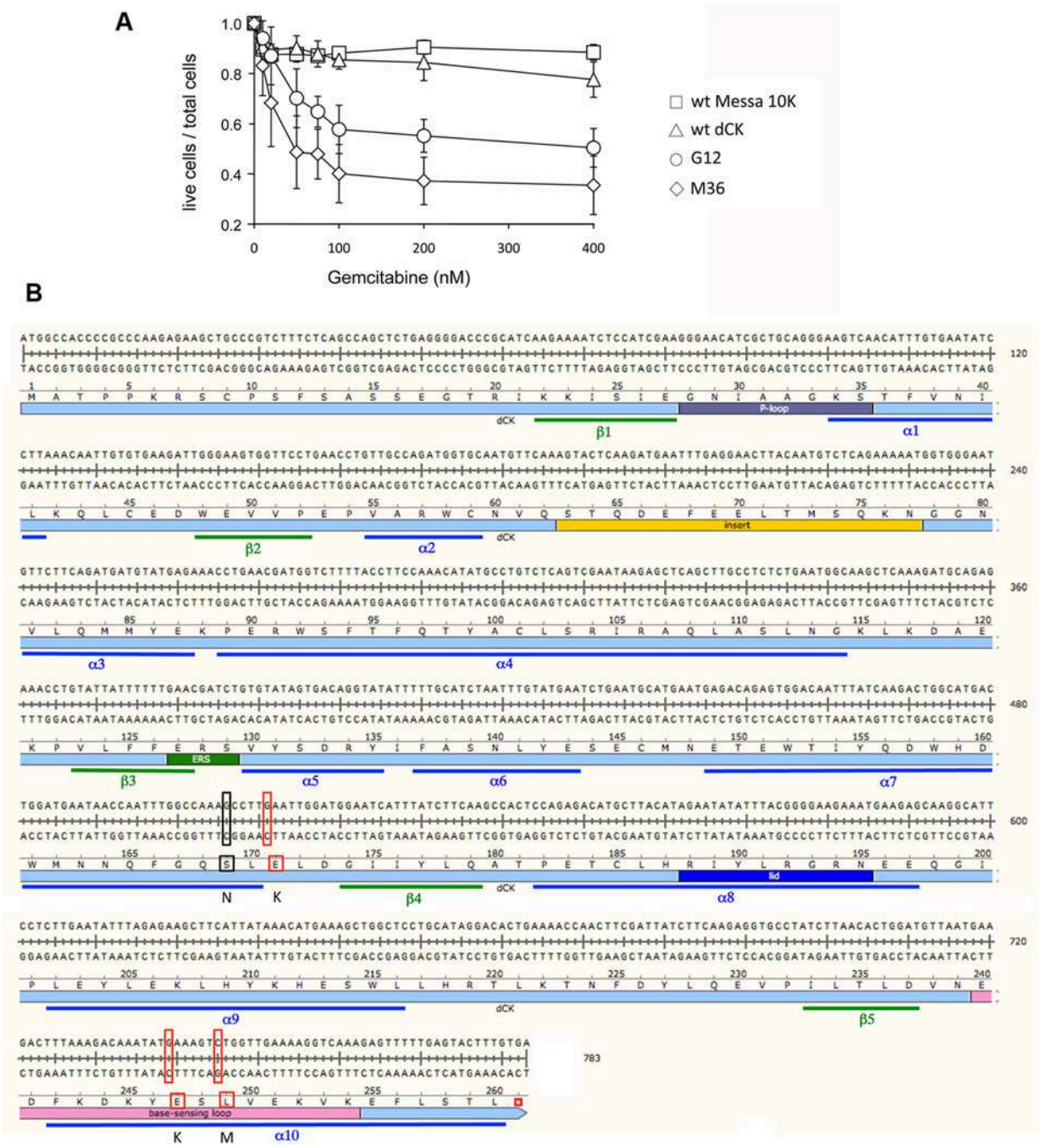



**Fig 4. The M36 mutant.** Panel A. Sensitivity to gemcitabine of Messa10K cells induced by M36. The ratio of living cells over the total number of cells is plotted as a function of the concentration of gemcitabine. Data are the average of five independent experiments. Panel B. Localisation of the mutations in the human dCK that characterise M36. The nucleotide sequence (double stranded) is given above, with the aminoacid sequence below. The full-length protein is given by a pale blue bar with the major functional domains, as defined in [28], indicated in colour.

doi:10.1371/journal.pone.0140741.g004

evaluate this possibility, the M36-coding sequence has been inserted in a SIN version of pSDY (Fig 1A) that has been used to generate lentiviral vectors by transfection of HEK 293T cells with the transcomplementation plasmids, as described above. These vectors have been used to establish a polyclonal population of Messa 10K cells (SINvec-M36 population), on which cell

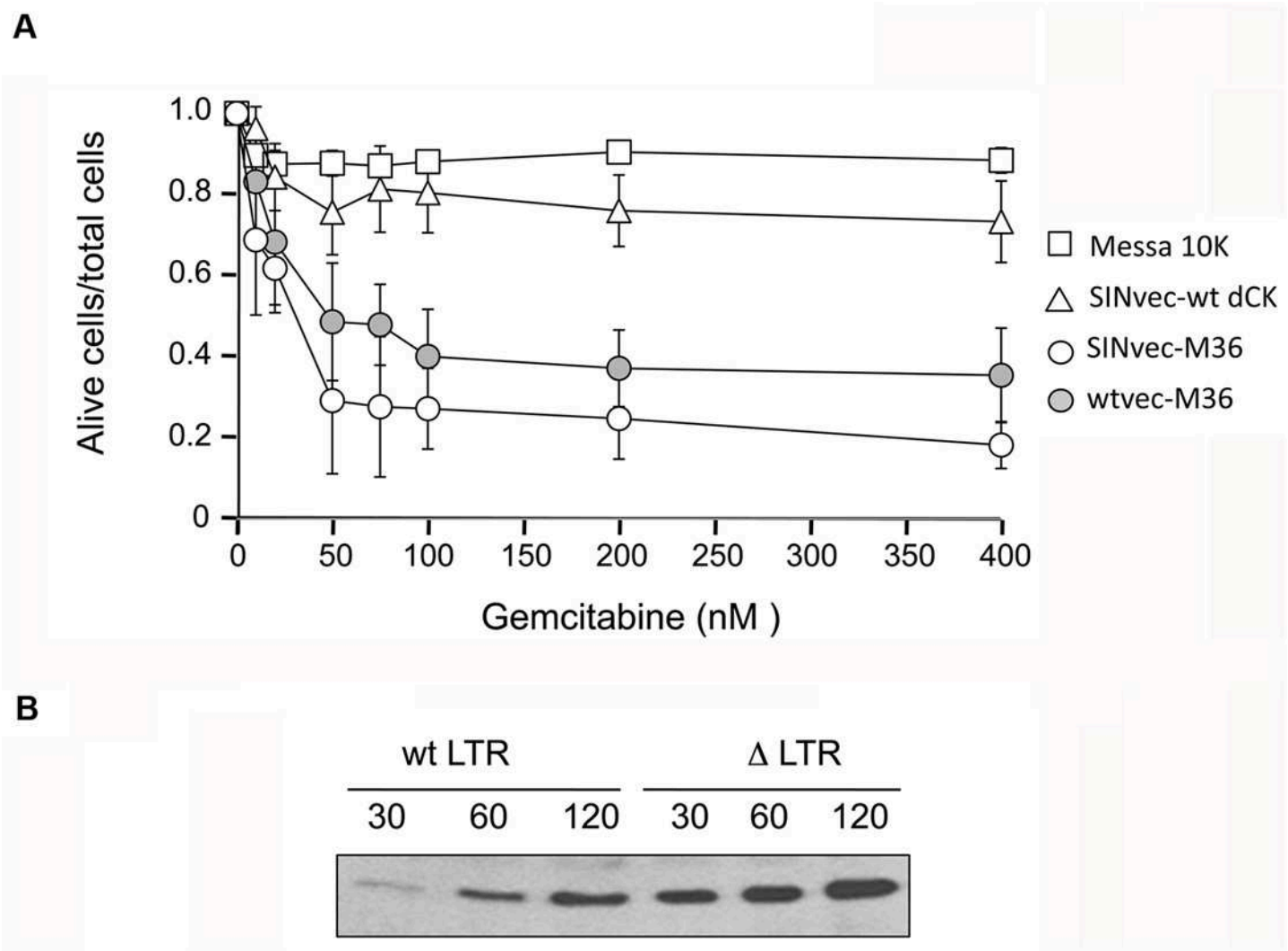



**Fig 5. Use of a self-inactivating (SIN) LTR sequence.** Panel A, sensitivity of cells harbouring a proviral DNA carrying SIN LTR and expressing M36 (white circles); of cells harbouring a proviral DNA carrying wt LTR and expressing M36 (grey circles, the data for this curve are reported as reference from Fig 4A, where they appear as white diamonds); of cells harbouring a proviral DNA carrying SIN LTR and expressing the wt dCK (white triangles); of wt Messa 10K cells (white squares). The proportion of living cells over the total number of cells is plotted as a function of the concentration of gemcitabine. The averages of three independent experiments are presented. Panel B, Western blot analysis of the expression of wt dCK and of M36 in wt-LTR and in SIN-LTR vectors. The band corresponding to the human dCK is shown for protein extracts from a population of Messa 10K cells expressing the M36 variant from a proviral DNA containing either a wt LTR or a SIN LTR. Three different amounts of total protein extract were loaded for each sample, as indicated (in μg) above each lane. The picture shows a representative result of 3 independent Western blot experiments.

doi:10.1371/journal.pone.0140741.g005

death was then measured in response to varying concentrations of gemcitabine. As a control, the survival of the SINvec-wt dCK Messa 10K polyclonal population (described above) was evaluated in parallel. As shown in Fig 5A, a strong sensitisation was observed relative to both wt Messa cells and to the SINvec-wt dCK population (white squares and white triangles in Fig 5A, respectively), in cells expressing the M36 (white circles) transgene from proviral DNA produced following reverse transcription of SIN vectors. Noteworthy, sensitisation was stronger ($p<0.05$) in the SINvec-M36 population than it was in a polyclonal population expressing M36 from a vector carrying a wt LTR (wtvec-M36 population reported as reference in the figure with grey circles). With respect to the wtvec-M36 population, the IC50 for the SINvec-M36 population was reduced from 50 to 24 nM. In addition, the plateau of residual survival

observed was also reduced in the SINvec-M36 population, from around 40% to around 20% (Fig 5A). These efficiencies of cell death are in the range of those observed with the use of SIN vectors for combination hsTK + gangcyclovir treatment [42].

Since an increased degree of sensitisation in the SINvec-M36 population could be attributable to increased M36 production in the presence of a ΔLTR variant, we investigated by Western blot whether a higher intracellular concentration of dCK/M36 was present in the SINvec-M36 polyclonal population of Messa10K cells than in the wtvec-M36 cell population. As shown in Fig 5B, more protein was detected in SINvec-M36 cells than was detected in wtvec-M36 cells. This difference in the amount of protein could account for the increased sensitisation observed for the cells transduced with SIN vectors.

## M36 and sensitisation to antiviral compounds

We then investigated whether M36 could induce sensitisation to drugs used for antiviral treatments. Specifically, two deoxycytidine analogs used as antiviral drugs that are activated through phosphorylation by dCK were tested, 3TC and ddC. To this end, we developed the antiviral assay depicted in Fig 6A. As described in the Materials and Methods, we generated HEK 293T/CD4$^+$ cell lines stably expressing either M36-SIN or wt dCK-SIN as a control; these cell lines were respectively named HEK 293T/CD4$^+$-SINvec-M36 and the HEK 293T/CD4$^+$-SINvec-wtdCK. For testing resistance to infection of each of these cell lines, cells were cultivated in the presence of varying concentrations of antiviral drug for 24 hours and then transduced with lentiviral vectors carrying the HIV-1 ADA envelope and encoding the luciferase gene, as detailed in the Materials and Methods and outlined in Fig 6A. Cells were maintained in the presence of the antiviral compound, and luciferase levels were measured 48 hours after transduction (Fig 6B and 6C). In cells expressing M36, no significant increase in transduction resistance was observed in the presence of ddC (Fig 6B). A significant ($p<0.05$) decrease in the efficiency of transduction was instead observed for 3TC (Fig 6C), for which 164 nM of 3TC was required to avoid 50% of transduction in the presence of an extra copy of dCK while only 15 nM of the drug was sufficient to reduce transduction by 50% in cells expressing M36.

## M36 and sensitisation of cells to other anticancer compounds

The possibility that M36 might sensitise cancer cells to anticancer compounds other than gemcitabine was then tested by evaluating sensitisation of Messa 10K cells to two compounds used in the clinical treatment of cancer that are structurally related to gemcitabine: fludarabine and AraC. The tests were performed as previously described for gemcitabine on SINvec-M36 and, as control, on SINvec-wt dCK Messa 10K polyclonal populations. Markedly different results were obtained with the two compounds. Although no sensitisation by M36 was observed (Fig 7A) in the case of fludarabine, strong sensitisation to AraC was observed in SINvec-M36 cells (Fig 7B, black squares) relative to both SINvec-wt dCK cells (grey diamonds) and to wt Messa 10K cells (white circles). A 10,000 fold decrease in IC50 ($p<0.001$) was observed for SINvec-M36 cells relative to SINvec-wt dCK cells.

## Biochemical characterisation of phosphorylation of gemcitabine and AraC by M36

To characterise the biochemical properties of M36 and, as a control, of wt dCK, the two enzymes were produced as recombinant, N-terminal His tagged proteins in *E. coli*. The purified enzymes were tested for their efficiency at phosphorylating dC (the natural substrate), gemcitabine and AraC *in vitro*, as previously described [34]. For the three substrates, a decrease in the initial rate of phosphorylation ($V_{ini}$) as a function of the concentration of substrate was

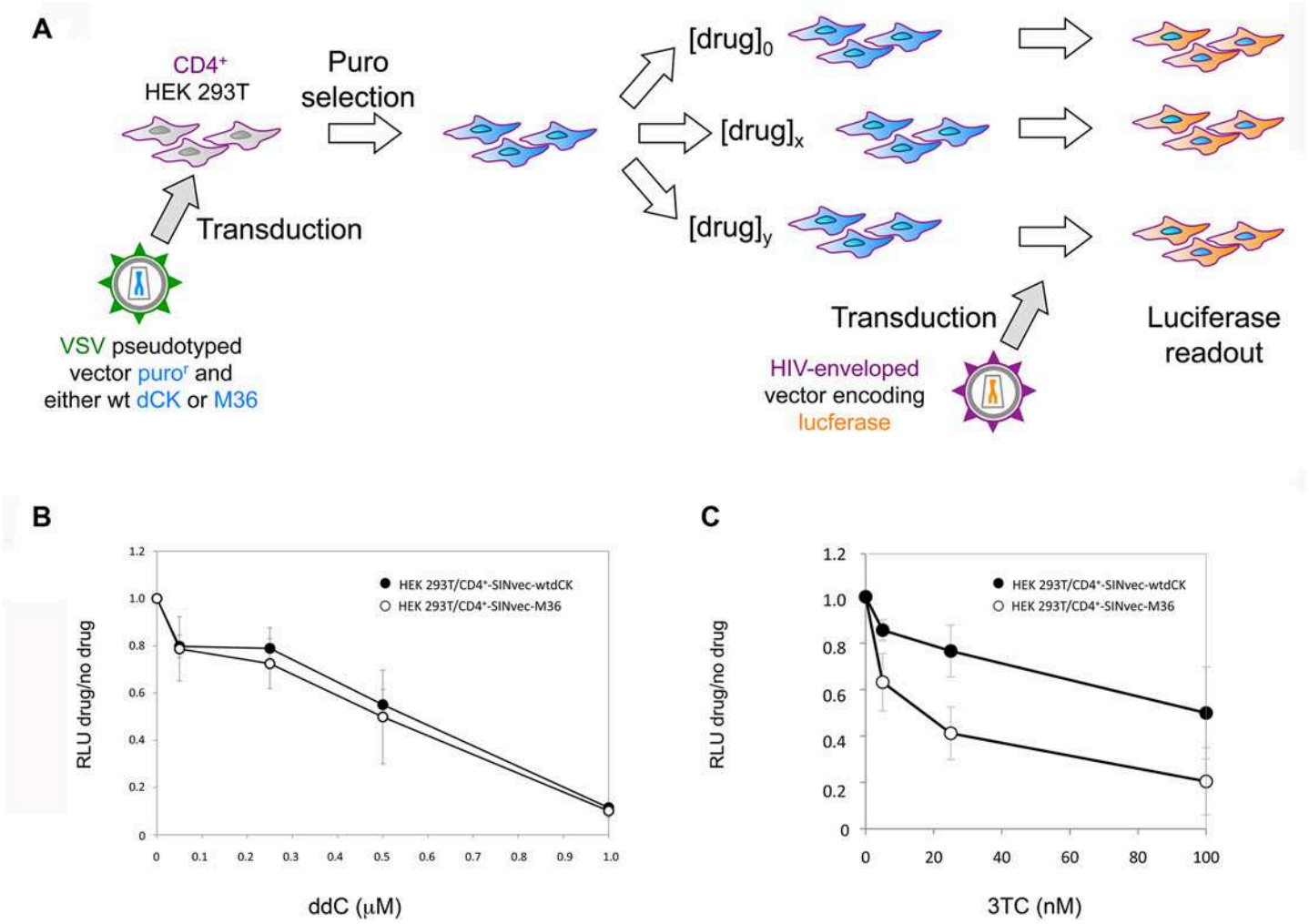



**Fig 6. M36 and sensitisation to anti HIV drugs.** Panel A. Schematic representation of the experimental procedure used to test the level of resistance to infection by HIV-derived vectors of cells expressing constitutively M36 and treated in the presence of the anti-HIV compounds ddC and 3TC. Panels B and C. Sensitivity to transduction with HIV-derived vectors in the presence of varying concentrations of ddC (panel B), or 3TC (panel C), of HEK 293T-CD4+ cells encoding M36 (white circles), or an extra copy of the wt dCK gene (black circles). Data are the average of 3 independent experiments. RLU, relative luciferase units.

doi:10.1371/journal.pone.0140741.g006

observed for M36 with respect to the wt enzyme (red and black, respectively, in Fig 8 panels A to C) with a decrease in Vmax that was more important for dC (3-fold) than for gemcitabine and AraC (1.6 and 1.5 -fold, respectively).

A comparative analysis of the kinetic parameters of dCK and M36 with respect to phosphorylation of dC, gemcitabine and AraC is given in Table 2. While the Kcat/Km, indicative of the overall efficiency of the reaction, was 2.6 times lower for M36 than for the wt dCK when considering the natural substrate, it was only 1.3 and 1.1 lower for gemcitabine and for AraC, respectively. The decrease in efficiency of phosphorylation of M36 specifically for dC is expected to bias phosphorylation in favour of the drugs providing a likely explanation for the sensitisation of the cells that express M36 to these drugs.

To understand the molecular bases of the M36 mediated sensitisation of cells to gemcitabine relative to that of G12, Km and Kcat values for dC and gemcitabine were compared between these two mutants by dividing the values found with M36 by those we previously reported for

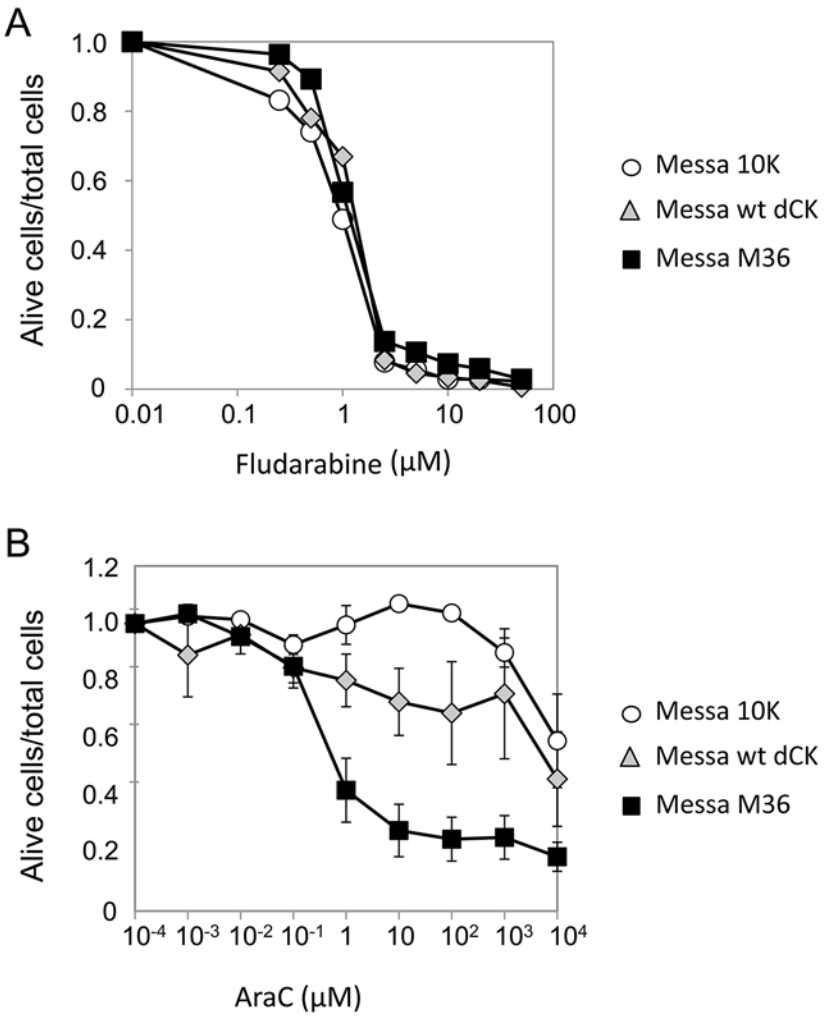



**Fig 7. M36 and sensitisation to anticancer compounds other than gemcitabine.** Sensitivity to anticancer compounds of Messa10K cells induced by M36. The percentage of living cells over the total number of cells is plotted as a function of the concentration of the anticancer compound (the x axis is given in log scale). Panel A, sensitivity to fludarabine (average of three independent experiments, error bars are not shown for the sake of clarity). Panel B, sensitivity to AraC. Data in panel B give the average of five independent experiments.

doi:10.1371/journal.pone.0140741.g007

G12 [34]; (AraC was not tested with G12 in that work). Relative to G12, M36 has an increased Km (by a factor of 3.5, black in Fig 8D) and Kcat (by a factor of 9.5, grey in the figure) for the natural substrate dC. For gemcitabine, however, the two enzymes displayed almost identical Km and Kcat values (ratios of 1 between M36 and G12). The fact that the kinetic parameters for gemcitabine are unaltered between the two mutants suggests that the different levels of sensitisation to gemcitabine observed in cell culture must be due to differences in their relative abilities to phosphorylate the natural substrate. In this respect, the values of Kcat and those of Km of the two dCK mutants point in two opposite directions. The differences observed with dC imply that M36 has a lower affinity for the natural substrate than G12, while it has improved catalytic activity that will facilitate phosphorylation of the natural substrate compared to G12. If we compare the time needed for the transformation of one molecule of substrate into product (1/Kcat) G12 will bind 3–4 times more efficiently the natural substrate than M36, but it will require a 9–10 longer time to complete phosphorylation (G12 1/Kcat = 526 sec,

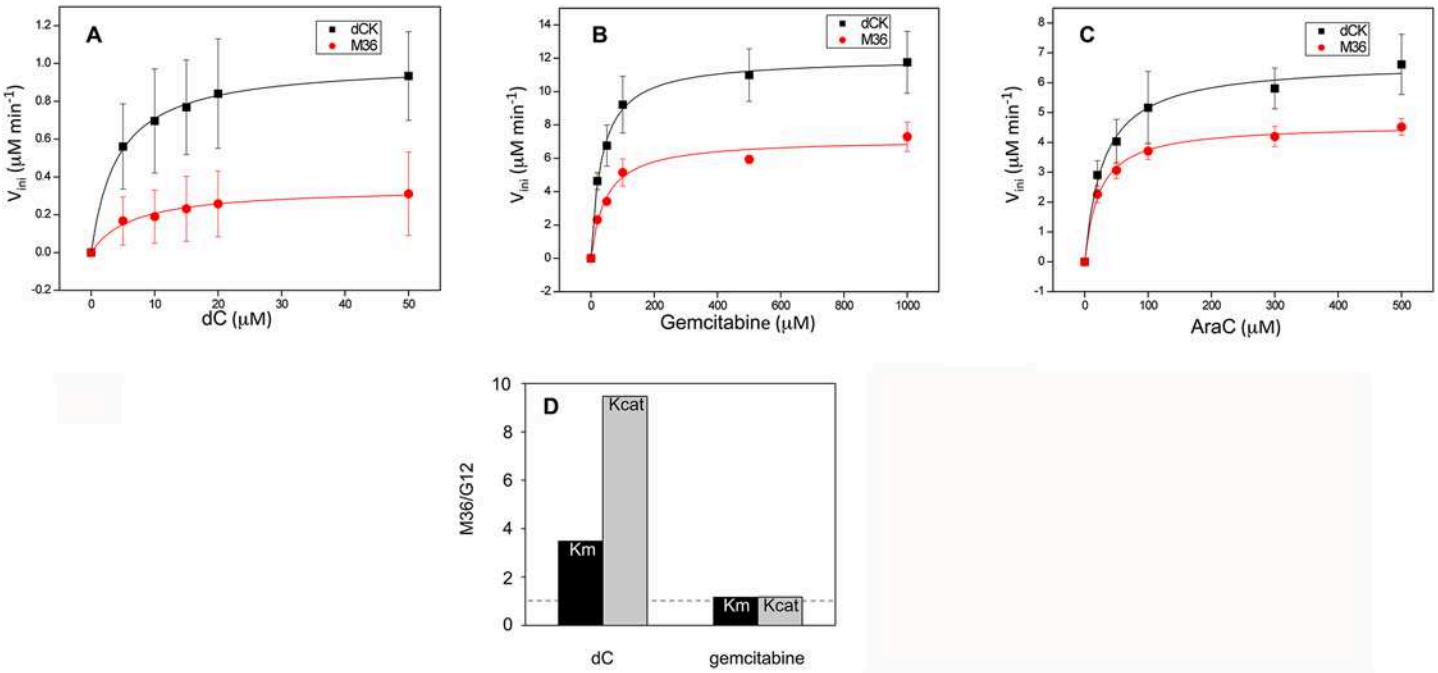



**Fig 8. Biochemical comparative characterisation of phosphorylation of AraC by wt dCK and M36.** Panels A and C. Phosphorylation kinetics of purified wt dCK (black squares) and M36 (red circles). Steady state kinetic data were fitted according to the Michaelis-Menten equation. Panel A, phosphorylation of dC (average of four independent experiments). Panel B, phosphorylation of gemcitabine (average of four independent experiments). Panel C, phosphorylation of AraC (average of three independent experiments). Panel D, ratio of Km (black) and of Kcat (grey) for M36 vs G12, with respect to dC and to gemcitabine. The dotted line gives the reference of a ratio of 1.

doi:10.1371/journal.pone.0140741.g008

M36 1/Kcat = 55 sec). These results indicate that G12 will be blocked in a non productive state that will impede the phosphorylation of the drug, while M36 will proceed to a new round of catalysis.

## Discussion

Suicide gene therapy is an approach of interest for many biomedical applications such as in the clinical treatment of cancers and as a safety gene in cell therapy. We previously described the identification of G12, a triple mutant of the human dCK that sensitises a panel of cancer cells

**Table 2. Kinetic parameters of the M36 mutant with respect to dC, gemcitabine and AraC.**

| | | dC | |
|---|---|---|---|
| | Km *μM* | $K_{cat}$ $s^{-1}$ | $K_{cat}$/Km $(M^{-1}\ s^{-1})\ x\ 10^3$ |
| **dCK** | 5.42 ± 1.89 | 0.035 ± 0.010 | 6.46 |
| **M36** | 7.31 ± 2.11 | 0.018 ± 0.005 | 2.46 |
| | | **Gemcitabine** | |
| | Km *μM* | $K_{cat}$ $s^{-1}$ | $K_{cat}$/Km $(M^{-1}\ s^{-1})\ x\ 10^3$ |
| **dCK** | 46.7 ± 18.9 | 0.57 ± 0.15 | 12.21 |
| **M36** | 38.7 ± 7.69 | 0.37 ± 0.05 | 9.56 |
| | | **AraC** | |
| | Km *μM* | $K_{cat}$ $s^{-1}$ | $K_{cat}$/Km $(M^{-1}\ s^{-1})\ x\ 10^3$ |
| **dCK** | 28.2 ± 3.80 | 0.38 ± 0.08 | 13.48 |
| **M36** | 20.5 ± 3.76 | 0.25 ± 0.01 | 12.20 |

doi:10.1371/journal.pone.0140741.t002

to treatment with gemcitabine at doses significantly lower than those required for induction of wild type cell death [34]. Here, by identifying a new mutant (M36), we extend these findings in two main directions. We observed an improved sensitisation to gemcitabine with respect to G12 and we report that a dramatic sensitisation is also observed with AraC, an anticancer compound that, as with gemcitabine, is currently used in clinics for the treatment of poor prognosis cancers. Furthermore, we also report that cells expressing the transgene are resistant to transduction with HIV-1 derived vectors at concentrations of 3TC significantly lower than those required for wt cells, which is consistent with the activation of analogues of dC by M36.

The G12 mutant was identified by screening a library of dCK variants that was created by using retrovolution, a new procedure developed in our laboratory, where controlled successive cycles of transduction are carried out using conditional replication-defective HIV-1 derived viral vectors that contain the dCK coding sequence in their genomic RNA. As the number of cycles progresses, mutations accumulate in the target gene due to the error-prone nature of the HIV-1 reverse transcription process. We observed the presence of synonymous mutations at higher than expected frequencies in the N-terminus half of the protein, indicative of purifying selection acting during the retrovolution cycles. Since the cycles are performed in HEK 293T cells that contain functional dCK, it is possible that this effect is due to the removal of dominant negative mutants that might be generated in this part of the protein during the procedure. In this regard, the N-ter half of the protein encompasses a large part of the dimer interface (S4 Fig), and the mutants that are counter-selected could interfere with the correct formation of dimers of the endogenous dCK dimers that are presumably present in HEK 293T cells. In contrast, positive selection was observed in the second half of the protein, indicative of a potential advantage for the expression of dCK variants with mutations in this region. This part of the protein contains the substrate binding site as well as the phosphate donor site. It is possible that modifications in this region favoured the growth of HEK 293T cells containing the transgene, during the retrovolution cycles. However, irrespective of the underlying mechanism, documenting the occurrence of purifying selection underscores an interesting asset of this system: mutants that are toxic to the cells are removed during the procedure. Such mutants might be toxic even in the absence of the drug, and therefore engineered cells containing them would be of little use for inducing controlled cell death in response to drug treatment.

We performed 11 further cycles of retrovolution either on the entire library from which G12 was isolated (random library) or starting from the G12 sequence alone (directed library) aiming at isolating a dCK mutant with greater sensitisation activity than G12. Whereas screening of the random library failed to identify new mutants of interest, either with a genotype related to G12, or with a completely different set of mutations a mutant with a related genotype to G12 but with an improved sensitisation phenotype was isolated from the directed library. This result indicates that productively extending the mutational pathway of G12 was likely more easily achievable than was developing a mutant along a completely different mutational pathway. Although gemcitabien sensitisation conferring dCK mutants that have a totally different genotype to that of G12 and M36 are likeky possible, this might require a set of mutations that are too complex to be easily identified through retrovolution. The difficulty in obtaining a mutant conferring improved sensitisation to anticancer drug treatment is further underlined by the following considerations. One is that the analysis of the directed library revealed that in the majority of cases the association of the G12 mutations with single mutations in other regions of the protein completely abolished the gemcitabine sensitisation phenotype (data not shown). Another is the absolute need for the three mutations that characterize G12 to be present simultaneously to obtain sensitisation of cells to the drug: mutants containing only one or two of the G12 mutations failed to induce the sensitisation phenotype (Fig 1, panels B and C).

The extra mutation (S169N) characterizing M36 with respect to G12 is located two aminoacids upstream one of the three G12 mutations (E171K). This results in a protein with four mutations clustered by pairs in a region in close proximity of the surface of the protein involved in the formation of the dimer and in the "base sensing loop" region (S3 Fig). Each cluster contains one mutation that leads to a drastic change in charge (E replaced by K, for positions 171 and 247) and another pair that is conservative of the physicochemical properties of the residues involved (replacement between non-polar residues at position 249 and between polar residues at position 169). These two "mild" mutations, albeit evidently important, probably reflect fine-tuning adjustments in the local folding of the regions that harbour the more drastic changes due to the two E to K substitutions. It is noteworthy that the mild mutation that differentiates M36 and G12, S169N, removes a potential phosphorylation site, possibly impacting the activity of the enzyme altering its post-translational modifications *in vivo*. It should however be stressed that phosphorylation of S169, if it indeed occurs, is probably not relevant for the activity of the enzyme since the characterisation of the dCK phosphorylation sites has identified T3, S11, S15, and, more importantly, S74 but not S169 as phosphorylation sites *in vivo* [43]. Therefore, loss of the potential phosphorylation site at S169 is unlikely to be responsible of the differences in activity observed between M36 and G12.

A feature emerging from the biochemical characterisation of M36 is that the mutant displays, both with respect to the wild type enzyme and to G12, markedly modified kinetic parameters relative to the natural substrate dC but only mild changes with respect to either gemcitabine or AraC. Indeed, the efficiency of phosphorylation (expressed as Kcat/Km values) for gemcitabine and for AraC is not significantly different between M36 and wt dCK (Table 2) while the mutant phosphorylates the natural substrate four times less efficiently than wt dCK. As a result, M36 phosphorylates the drugs with increased efficiency relative to the natural substrate. Decreased ability to phosphorylate the natural substrate seems to be the preferred path to inducing increased sensitivity to anticancer drugs. In support of this view, an alteration in its capacity to phosphorylate the natural substrate also seems to explain the improved phenotype of M36 with respect to G12: relative to M36, G12 appears to both bind the natural substrate more avidly, and takes longer periods once bound before it releases the phosphorylated molecule, and therefore has lower availability for binding the drug molecules.

Besides the fact that increases in sensitisation to drug treatment that are induced by G12 and M36 are strongly impacted by the efficiencies with which these mutants hosphorylate the natural substrate, the characteristics of specific drug molecules are also likely highly relevant. Indeed, while M36 increases cellular sensitivity to both gemcitabine and AraC (for which the effect is markedly stronger than for gemcitabine), it failed to detectably sensitise cells to the other anticancer drug tested, fludarabine (an analogue of dA), an observation that might be accounted for by the fact that dA is not the preferential substrate of dCK [44]. In addition to activating anticancer drugs, M36 also activated one of the two antiviral compounds tested, 3TC, effectively protecting M36-containing cells from transduction with HIV-1 derived vectors at 3TC doses eleven times lower than those required to protect wt cells. The lack of effect observed with ddC is probably due to this molecule being too structurally dissimilar to its natural substrate.

We also report here that the use of the SIN vectors for inserting M36 into the target cells, instead of vectors carrying a wt LTR, further sensitises cells to gemcitabine exposure. This is probably due to the relief of transcriptional interference that is expected to occur with the functional U3 promoter [45]. In support to such a view, we detected higher intracellular quantities of synthesised M36, when the protein was expressed from a proviral DNA derived from a SIN vector than when it was synthetised from a wt vector. In line with the reported direct correlation between the amounts of dCK protein that are present in a cell and the degree to which a

cell is sensitive to anticancer compounds [21], such vector-mediated increases in cellular concentrations of sensitisation mutant protein such as M36, should lead to an increased intracellular concentration of activated drug molecules and therefore increased degrees of cellular.

In conclusion, these results describe the identification and molecular characterisation of a new variant of the human dCK enzyme that could be of great interest as either a safety gene with utility in cell therapy approaches, or as suicide gene for cancer treatment. The ability of M36 to strongly sensitise cells to two of the drugs that are routinely employed in the treatment of different types of cancer with poor prognoses makes this dCK mutant a particularly promising candidate for suicide gene approaches.

## Supporting Information

**S1 Fig. Schematic representation of the nucleotide substitutions generated in the 41 clones of the F27-RL.** Colour code: red G>A; green C>T; violet T>C; light blue C>A; pink T>A; dark blue A>C; black A>T. A cumulative map of the positions mutated along the protein is given at the top of the figure. The light blue segments indicate regions of 100 nt.
(TIF)

**S2 Fig. Schematic representation of the nucleotide substitutions generated in the 43 clones of the F11-DL.** The representation and colour code is the same as in S1 Fig.
(TIF)

**S3 Fig. Schematic diagram of a dCK monomer.** The G12 mutations are outlined in red and the additional M36 mutation is outlined in yellow.
(TIF)

**S4 Fig. Structure of a human dCK dimer.** The part in blue constitutes the N-ter half of the protein; in grey the C-ter half and in red highlighted the positions mutated in M36.
(TIF)

## Acknowledgments

This work was supported by funding from the Ligue Contre le Cancer (http://www.ligue-cancer.net), Appel d'offre CCIR_GE 2012 to MN. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors wish to thank S. Bernacchi for fitting the curves of the enzymatic reactions and E. Pencreach for the kind gift of Gemcitabine. S. Coulibaly is recipient of a fellowship from the French "Ministère de l'Enseignement Supérieur et de la Recherche".

## Author Contributions

Conceived and designed the experiments: STC PR DL MN. Performed the experiments: STC PR FW FKK MB. Analyzed the data: STC PR FW FKK MB DPM DL MN. Contributed reagents/materials/analysis tools: DPM. Wrote the paper: STC PR DPM DL MN.

## References

1. Duarte S, Carle G, Faneca H, de Lima MC, Pierrefite-Carle V. Suicide gene therapy in cancer: where do we stand now? Cancer Lett. 2012; 324(2):160–70. doi: 10.1016/j.canlet.2012.05.023 PMID: 22634584
2. Bonini C, Bondanza A, Perna SK, Kaneko S, Traversari C, Ciceri F, et al. The suicide gene therapy challenge: how to improve a successful gene therapy approach. Mol Ther. 2007; 15(7):1248–52. PMID: 17505474

3. Bonini C, Ferrari G, Verzeletti S, Servida P, Zappone E, Ruggieri L, et al. HSV-TK gene transfer into donor lymphocytes for control of allogeneic graft-versus-leukaemia. Science. 1997 Jun 13; 276 (5319):1719–24. PMID: 9180086
4. Immonen A, Vapalahti M, Tyynela K, Hurskainen H, Sandmair A, Vanninen R, et al. AdvHSV-tk gene therapy with intravenous ganciclovir improves survival in human malignant glioma: a randomised, controlled study. Mol Ther. 2004; 10(5):967–72. PMID: 15509514
5. Shalev M, Kadmon D, Teh BS, Butler EB, Aguilar-Cordova E, Thompson TC, et al. Suicide gene therapy toxicity after multiple and repeat injections in patients with localized prostate cancer. J Urol. 2000; 163(6):1747–50. PMID: 10799174
6. Berger C, Flowers ME, Warren EH, Riddell SR. Analysis of transgene-specific immune responses that limit the in vivo persistence of adoptively transferred HSV-TK-modified donor T cells after allogeneic hematopoietic cell transplantation. Blood. 2006; 107(6):2294–302. PMID: 16282341
7. Reichard P. Interactions between deoxyribonucleotide and DNA synthesis. Annu Rev Biochem. 1988; 57:349–74. PMID: 3052277
8. Abratt RP, Bezwoda WR, Falkson G, Goedhals L, Hacking D, Rugg TA. Efficacy and safety profile of gemcitabine in non-small-cell lung cancer: a phase II study. J Clin Oncol. 1994; 12(8):1535–40. PMID: 8040664
9. Anderson H, Lund B, Bach F, Thatcher N, Walling J, Hansen HH. Single-agent activity of weekly gemcitabine in advanced non-small-cell lung cancer: a phase II study. J Clin Oncol. 1994; 12(9):1821–6. PMID: 8083706
10. Carmichael J, Possinger K, Phillip P, Beykirch M, Kerr H, Walling J, et al. Advanced breast cancer: a phase II trial with gemcitabine. J Clin Oncol. 1995; 13(11):2731–6. PMID: 7595731
11. Casper ES, Green MR, Kelsen DP, Heelan RT, Brown TD, Flombaum CD, et al. Phase II trial of gemcitabine (2,2'-difluorodeoxycytidine) in patients with adenocarcinoma of the pancreas. Invest New Drugs. 1994; 12(1):29–34. PMID: 7960602
12. Lund B, Hansen OP, Theilade K, Hansen M, Neijt JP. Phase II study of gemcitabine (2',2'-difluorodeoxycytidine) in previously treated ovarian cancer patients. J Natl Cancer Inst. 1994; 86(20):1530–3. PMID: 7932808
13. Moore M. Activity of gemcitabine in patients with advanced pancreatic carcinoma. A review. Cancer. 1996; 78(3 Suppl):633–8. PMID: 8681302
14. Lee MW, Parker WB, Xu B. New insights into the synergism of nucleoside analogs with radiotherapy. Radiat Oncol. 2013; 8:223. doi: 10.1186/1748-717X-8-223 PMID: 24066967
15. Kawasaki H, Carrera CJ, Piro LD, Saven A, Kipps TJ, Carson DA. Relationship of deoxycytidine kinase and cytoplasmic 5'-nucleotidase to the chemotherapeutic efficacy of 2-chlorodeoxyadenosine. Blood. 1993; 81(3):597–601. PMID: 8094016
16. Geutjes EJ, Tian S, Roepman P, Bernards R. Deoxycytidine kinase is overexpressed in poor outcome breast cancer and determines responsiveness to nucleoside analogs. Breast Cancer Res Treat. 2012; 131(3):809–18. doi: 10.1007/s10549-011-1477-3 PMID: 21465168
17. Beausejour CM, Gagnon J, Primeau M, Momparler RL. Cytotoxic activity of 2',2'-difluorodeoxycytidine, 5-aza-2'-deoxycytidine and cytosine arabinoside in cells transduced with deoxycytidine kinase gene. Biochem Biophys Res Commun. 2002; 293(5):1478–84. PMID: 12054682
18. Bergman AM, Pinedo HM, Peters GJ. Determinants of resistance to 2',2'-difluorodeoxycytidine (gemcitabine). Drug Resist Updat. 2002; 5(1):19–33. PMID: 12127861
19. Blackstock AW, Lightfoot H, Case LD, Tepper JE, Mukherji SK, Mitchell BS, et al. Tumor uptake and elimination of 2',2'-difluoro-2'-deoxycytidine (gemcitabine) after deoxycytidine kinase gene transfer: correlation with in vivo tumor response. Clin Cancer Res. 2001; 7(10):3263–8. PMID: 11595723
20. Kroep JR, Loves WJ, van der Wilt CL, Alvarez E, Talianidis I, Boven E, et al. Pretreatment deoxycytidine kinase levels predict in vivo gemcitabine sensitivity. Mol Cancer Ther. 2002; 1(6):371–6. PMID: 12477049
21. Hapke DM, Stegmann AP, Mitchell BS. Retroviral transfer of deoxycytidine kinase into tumor cell lines enhances nucleoside toxicity. Cancer Res. 1996; 56(10):2343–7. PMID: 8625309
22. Manome Y, Wen PY, Dong Y, Tanaka T, Mitchell BS, Kufe DW, et al. Viral vector transduction of the human deoxycytidine kinase cDNA sensitizes glioma cells to the cytotoxic effects of cytosine arabinoside in vitro and in vivo. Nat Med. 1996; 2(5):567–73. PMID: 8616717
23. Bergman AM, Giaccone G, van Moorsel CJ, Mauritz R, Noordhuis P, Pinedo HM, et al. Cross-resistance in the 2',2'-difluorodeoxycytidine (gemcitabine)-resistant human ovarian cancer cell line AG6000 to standard and investigational drugs. Eur J Cancer. 2000; 36(15):1974–83. PMID: 11000580

24. Galmarini CM, Clarke ML, Jordheim L, Santos CL, Cros E, Mackey JR, et al. Resistance to gemcitabine in a human follicular lymphoma cell line is due to partial deletion of the deoxycytidine kinase gene. BMC Pharmacol. 2004; 4:8. PMID: 15157282

25. Bhalla K, Nayak R, Grant S. Isolation and characterization of a deoxycytidine kinase-deficient human promyelocytic leukemic cell line highly resistant to 1-beta-D- arabinofuranosylcytosine. Cancer Res. 1984; 44(11):5029–37. PMID: 6091869

26. Cohen A, Lee JW, Gelfand EW. Selective toxicity of deoxyguanosine and arabinosyl guanine for T-leukemic cells. Blood. 1983; 61(4):660–6. PMID: 6600944

27. Flasshove M, Strumberg D, Ayscue L, Mitchell BS, Tirier C, Heit W, et al. Structural analysis of the deoxycytidine kinase gene in patients with acute myeloid leukaemia and resistance to cytosine arabinoside. Leukaemia. 1994; 8(5):780–5.

28. Sabini E, Ort S, Monnerjahn C, Konrad M, Lavie A. Structure of human dCK suggests strategies to improve anticancer and antiviral therapy. Nat Struct Biol. 2003; 10(7):513–9. PMID: 12808445

29. Hazra S, Szewczak A, Ort S, Konrad M, Lavie A. Post-translational phosphorylation of serine 74 of human deoxycytidine kinase favors the enzyme adopting the open conformation making it competent for nucleoside binding and release. Biochemistry. 2011; 50(14):2870–80. doi: 10.1021/bi2001032 PMID: 21351740

30. McSorley T, Ort S, Hazra S, Lavie A, Konrad M. Mimicking phosphorylation of Ser-74 on human deoxycytidine kinase selectively increases catalytic activity for dC and dC analogues. FEBS Lett. 2008; 582 (5):720–4. doi: 10.1016/j.febslet.2008.01.048 PMID: 18258203

31. Momparler RL, Chu MY, Fischer GA. Studies on a new mechanism of resistance of L5178Y murine leukaemia cells to cytosine arabinoside. Biochim Biophys Acta. 1968; 161(2):481–93. PMID: 4174433

32. Momparler RL, Momparler LF. Chemotherapy of L1210 and L1210/ARA-C leukaemia with 5-aza-2'-deoxycytidine and 3-deazauridine. Cancer Chemother Pharmacol. 1989; 25(1):51–4. PMID: 2480188

33. Jordheim LP, Galmarini CM, Dumontet C. Gemcitabine resistance due to deoxycytidine kinase deficiency can be reverted by fruitfly deoxynucleoside kinase, DmdNK, in human uterine sarcoma cells. Cancer Chemother Pharmacol. 2006; 58(4):547–54. PMID: 16463058

34. Rossolillo P, Winter F, Simon-Loriere E, Gallois-Montbrun S, Negroni M. Retrovolution: HIV-driven evolution of cellular genes and improvement of anticancer drug activation. PLoS Genet. 2012; 8(8): e1002904. doi: 10.1371/journal.pgen.1002904 PMID: 22927829

35. Naldini L, Blomer U, Gallay P, Ory D, Mulligan R, Gage FH, et al. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science. 1996; 272(5259):263–7. PMID: 8602510

36. Yee JK, Miyanohara A, LaPorte P, Bouic K, Burns JC, Friedmann T. A general method for the generation of high-titer, pantropic retroviral vectors: highly efficient infection of primary hepatocytes. Proc Natl Acad Sci U S A. 1994; 91(20):9564–8. PMID: 7937806

37. Baker AM, Sauliere A, Gaibelet G, Lagane B, Mazeres S, Fourage M, et al. CD4 interacts constitutively with multiple CCR5 at the plasma membrane of living cells. A fluorescence recovery after photobleaching at variable radii approach. J Biol Chem. 2007; 282(48):35163–8. PMID: 17855336

38. Pleskoff O, Treboute C, Brelot A, Heveker N, Seman M, Alizon M. Identification of a chemokine receptor encoded by human cytomegalovirus as a cofactor for HIV-1 entry. Science. 1997; 276(5320):1874–8. PMID: 9188536

39. Simon-Loriere E, Galetto R, Hamoudi M, Archer J, Lefeuvre P, Martin DP, et al. Molecular mechanisms of recombination restriction in the envelope gene of the human immunodeficiency virus. PLoS Pathog. 2009; 5(5):e1000418. doi: 10.1371/journal.ppat.1000418 PMID: 19424420

40. Hamoudi M, Simon-Loriere E, Gasser R, Negroni M. Genetic diversity of the highly variable V1 region interferes with Human Immunodeficiency Virus type 1 envelope functionality. Retrovirology. 2013; 10:114. doi: 10.1186/1742-4690-10-114 PMID: 24156625

41. Agarwal KC, Miech RP, Parks RE, Jr. Guanylate kinases from human erythrocytes, hog brain, and rat liver. Methods Enzymol. 1978; 51:483–90. PMID: 211390

42. Preuss E, Muik A, Weber K, Otte J, von Laer D, Fehse B. Cancer suicide gene therapy with TK.007: superior killing efficiency and bystander effect. J Mol Med (Berl). 2011; 89(11):1113–24.

43. Smal C, Vertommen D, Bertrand L, Ntamashimikiro S, Rider MH, Van Den Neste E, et al. Identification of in vivo phosphorylation sites on human deoxycytidine kinase. Role of Ser-74 in the control of enzyme activity. J Biol Chem. 2006; 281(8):4887–93. PMID: 16361699

44. Van Rompay AR, Johansson M, Karlsson A. Substrate specificity and phosphorylation of antiviral and anticancer nucleoside analogues by human deoxyribonucleoside kinases and ribonucleoside kinases. Pharmacol Ther. 2003; 100(2):119–39. PMID: 14609716

45. Karn J, Stoltzfus CM. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. Cold Spring Harb Perspect Med. 2012; 2(2):a006916. doi: 10.1101/cshperspect.a006916 PMID: 22355797

## III.2 Etude du mutant M36

### III.2.1 Capacité à sensibiliser différentes lignées cellulaires cancéreuses

Un des points forts de la rétrovolution est de pouvoir tester directement dans les cellules cibles la banque de mutants obtenue. Les cellules utilisées pour cette étude sont les cellules Messa 10K qui ont été adaptées en laboratoire afin d'être résistantes à la Gem[229]. Ces cellules représentent un bon modèle d'étude car elles expriment une dCK inactive, et fournissent donc un fond génétique « vierge » sur lequel il est plus facile d'apprécier les propriétés des mutants isolés. Le mutant M36 originairement identifié dans les cellules HEK 293T a démontré sa capacité à sensibiliser les cellules Messa10K, et nous avons souhaité savoir s'il est également capable de sensibiliser d'autres lignées cellulaires cancéreuses exprimant une dCK endogène fonctionnelle. J'ai notamment utilisé des lignées cellulaires issues de cancer solides (HT29, HCT116) ou de leucémies (Jurkat, MT4, THP-1, DAMI, K562). A partir de ces cellules, j'ai établi des lignées stables exprimant soit une copie supplémentaire de la dCK sauvage, soit le M36, comme décrit dans le Matériels et Méthodes, et testé leur sensibilité à la Gem et/ou à l'AraC (figure 31). Les résultats obtenus montrent qu'aucune des lignées testées ne semble présenter une sensibilisation accrue aux médicaments lorsqu'elles expriment le M36.

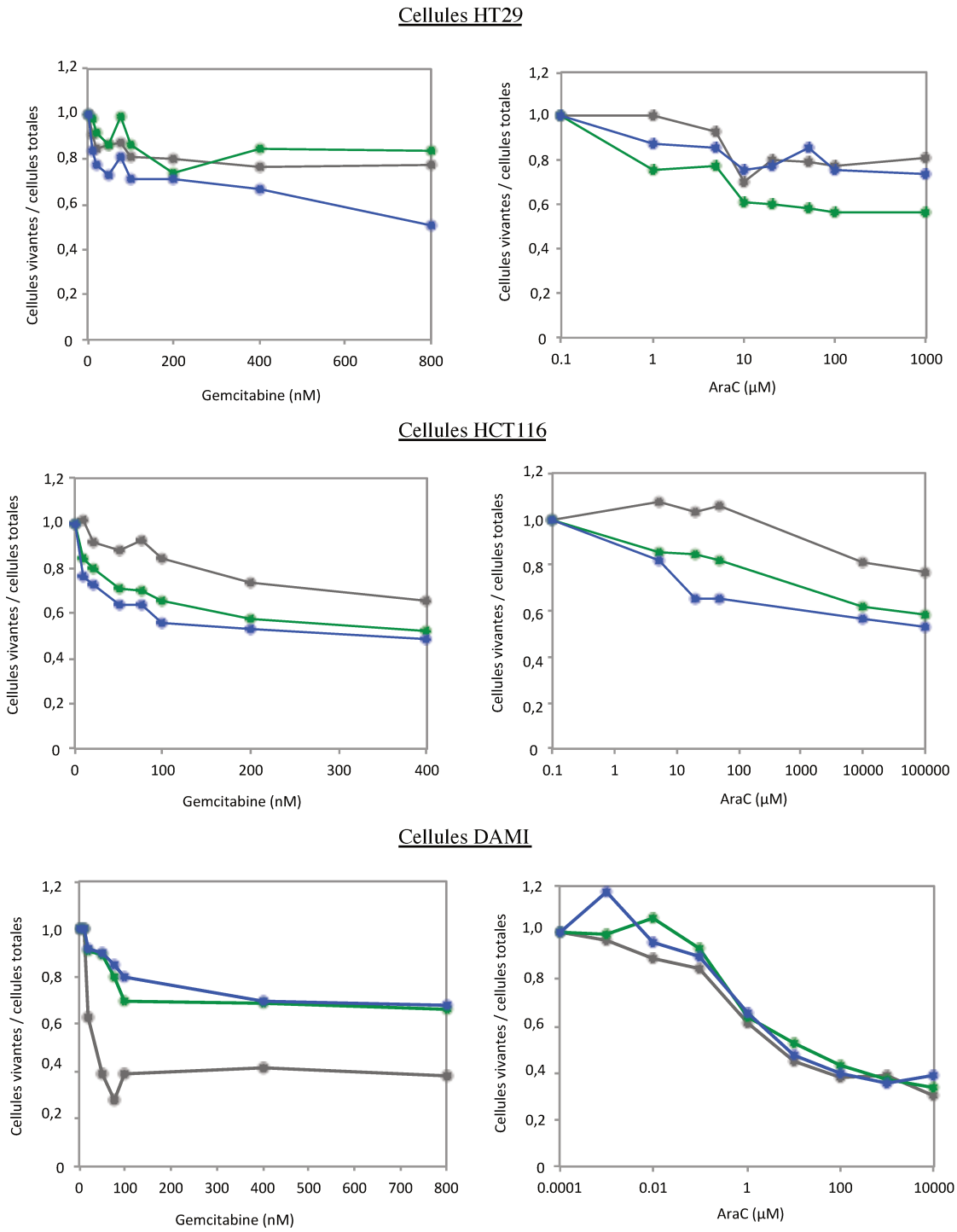



**Figure 31 : Capacité du M36 à sensibiliser d'autres cellules cancéreuses.** Pour chacune des lignées cellulaires testées, des lignées stables ont été établi de manière à exprimer soit une copie supplémentaire de la dCK soit le mutant M36. Les cellules non transduites sont représentées en gris, les cellules exprimant une copie supplémentaire de la dCK en vert et les cellules exprimant le M36 en bleu. n=2. Les barres n'erreurs n'apparaissent pas pour plus de clarté.

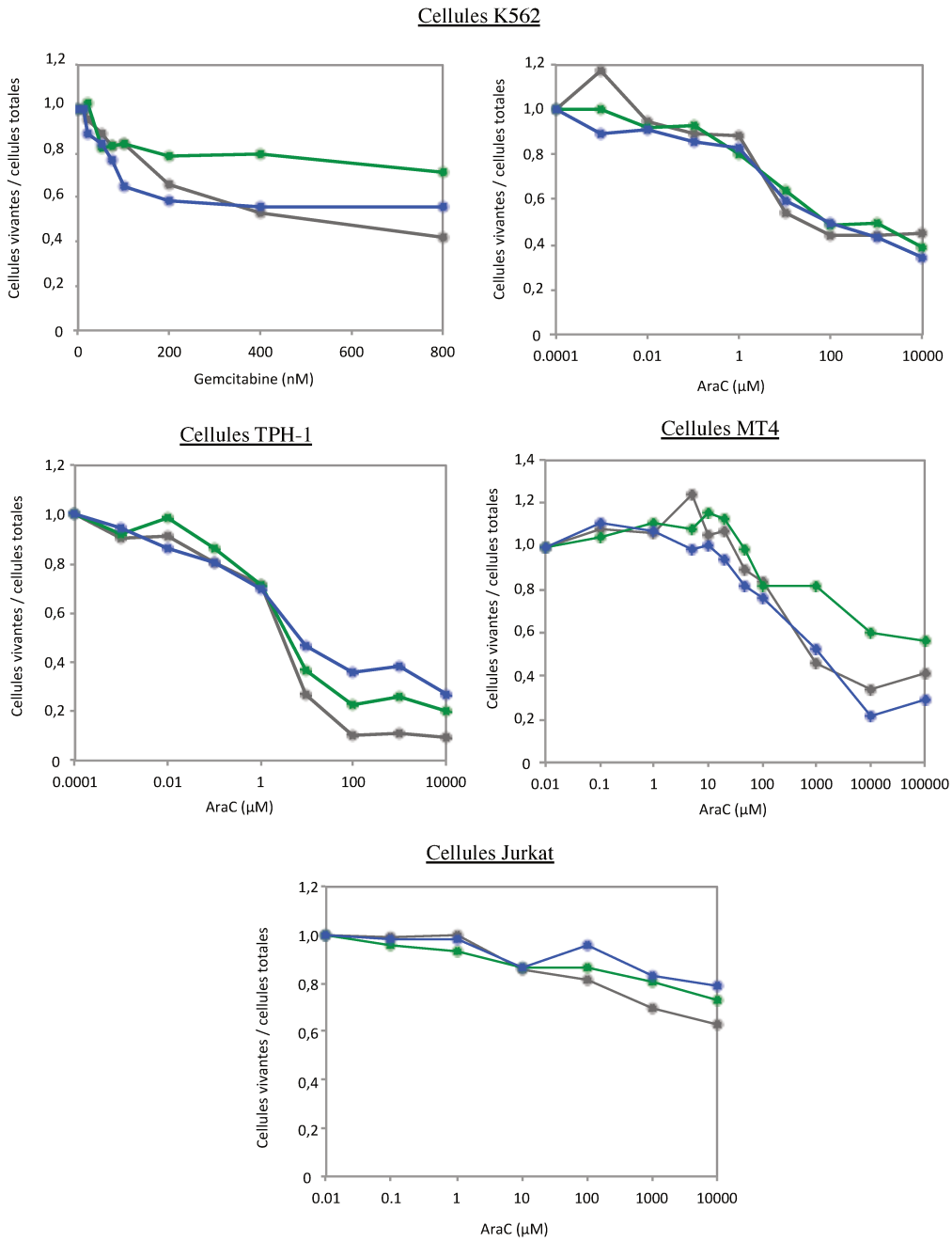



**Figure 31 : suite**

### III.2.2 Sensibilisation des cellules par une expression du M36 et une réduction de l'expression de la RRM1

La ribonucléotide réductase est une enzyme impliquée dans la production des désoxyribonucléotides à partir des ribonucléotides (figure 10). Cette enzyme est composée de deux sous unités et il a été démontré que la répression de l'expression de la sous-unité 1 (RRM1) permet de sensibiliser les cellules aux composés chimiothérapeutiques (cf Partie I-III.3.2). Nous avons donc voulu tester la sensibilité des cellules exprimant le M36 en réduisant en même temps, l'expression de la RRM1. Pour ce faire, j'ai dans un premier temps généré 4 siARN dirigés contre l'extrémité 5' du gène (Matériels et Méthodes), et déterminé par western blot leur capacité à réprimer l'expression de la RRM1 dans les cellules transfectées (figure 32). L'analyse du résultat de western blot montre que les siARN 1 et 4 sont ceux qui permettent une meilleure répression de l'expression de la protéine. L'objectif de cette étude était de pouvoir utiliser des vecteurs permettant à la fois l'expression du M36 et la répression de l'expression de la RRM1. Pour qu'une séquence siARN puisse réprimer l'expression de la RRM1 via l'utilisation d'un vecteur lentiviral, il faut qu'elle soit produite à partir d'un précurseur shARN, qui serait introduit dans l'ARN génomique du vecteur. Pour atteindre notre objectif, il fallait tout d'abord tester la capacité des shARN correspondant aux siARN 1 et 4, à réprimer l'expression de la RRM1. A partir des séquences des siARN, j'ai donc construit un plasmide permettant l'expression de chaque shARN sous le contrôle d'un promoteur U6 (figure 33A) et évalué la capacité de ces shARN à induire une réduction de l'expression de la RRM1. Deux jours après la transduction, les cellules ont été récoltées, lysées et le niveau d'expression de la RRM1 a été évalué par western blot, effectué à partir de 30 µg de protéines totales (figure 33B). L'analyse du niveau d'expression de la RRM1 dans chacun des extraits cellulaires (figure 33C) montre cependant que l'expression de la protéine n'est pas réprimée par les shARN et cela, même en augmentant la quantité de vecteurs utilisés pour la transduction. Ce résultat ne permet pas de déterminer si le manque de répression de la RRM1 est dû à un défaut de production des siARN ou alors à une faible expression de ceux-ci.

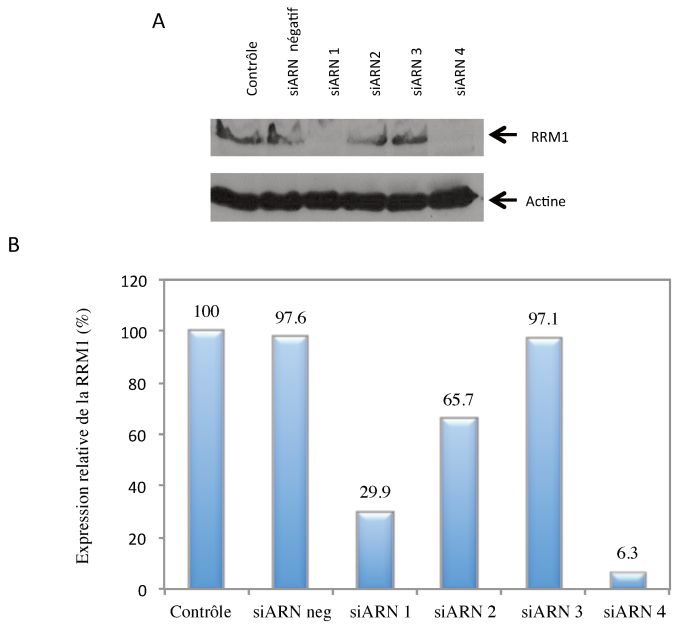
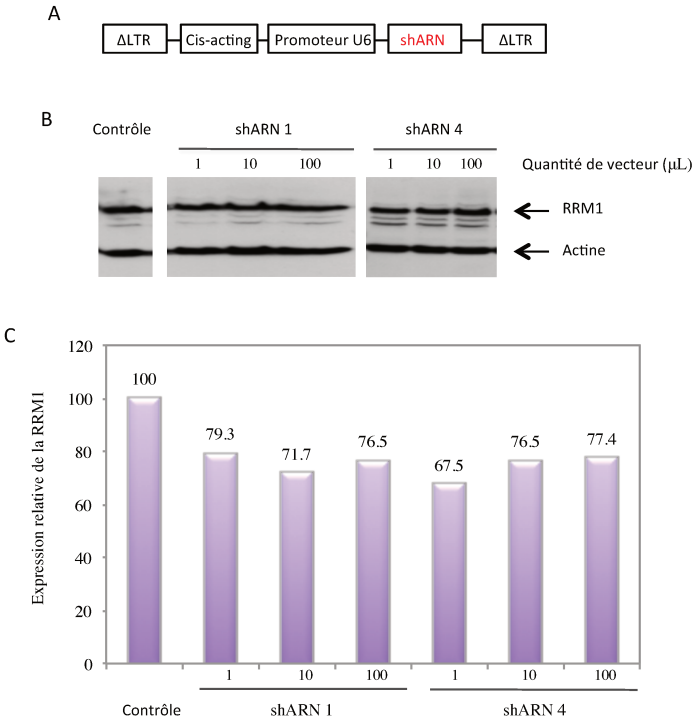



**Figure 32 : Répression de l'expression de la RRM1 par les siRNA.** Les cellules HEK 293T ont été transfectées avec l'un des 4 siARN générés afin de cibler l'extrémité 5' du gène codant la RRM1. Le contrôle correspond aux cellules non transfectées. Le siARN neg est un siARN ne ciblant aucune partie du gène (Matériels et Méthodes). A) Western blot réalisé avec 60 µg de protéines totales de chaque extrait cellulaire. B) Pourcentage d'expression de la RRM1 dans chaque extrait en fonction de l'expression de la protéine dans les cellules non transfectées (Matériels et Méthodes). Ce pourcentage est indiqué au dessus des histogrammes pour chaque extrait.

**Figure 33 : Répression de l'expression de la RRM1 par les shARN.** A) Schéma représentatif de l'ADN proviral exprimé à partir des vecteurs. Sont indiquées les régions ΔLTR permettant l'intégration dans le génome cellulaire et la région Cis acting permettant l'encapsidation de l'ARN génomique dans les particules virale. Le contrôle correspond aux cellules non transduites. B) Western blot réalisé avec 30 µg de protéines totales de chaque extrait cellulaire. C) Pourcentage d'expression de la RRM1 dans chaque extrait en fonction de l'expression de la protéine dans les cellules non transfectées (Matériels et Méthodes). Ce pourcentage est indiqué au dessus des histogrammes pour chaque extrait.



Toutefois, afin de déterminer si une réduction de la RRM1 couplée à l'expression du M36 permettrait une meilleure sensibilisation des cellules aux médicaments, nous avons utilisé les siRNA 1 et 4 pour transfecter les cellules Messa exprimant de manière stable le M36. Le niveau d'expression de la RRM1 dans les cellules transfectées a été évalué deux jours après la transfection et un test de sensibilité à la Gem et à l'AraC a été réalisé. Comme vu précédemment, l'utilisation des siARN permet de réduire considérablement l'expression de la RRM1 dans les cellules (figure 34A), mais elle en permet pas d'accroitre la sensibilité des cellules à la Gem (figure 34B).

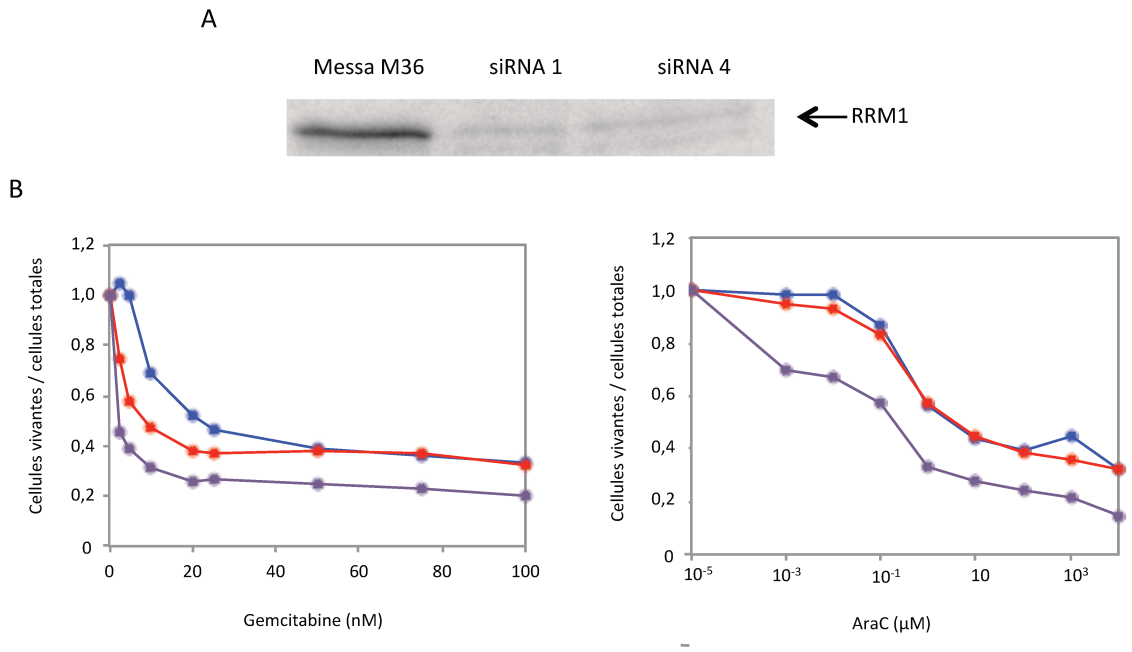



**Figure 34 : Sensibilisation des cellules exprimant le M36 couplé à une réduction de l'expression de la RRM1**. Les cellules Messa exprimant la protéine M36 ont été transfectées avec les siARN 1 et 4. A) La réduction de l'expression de la RRM1 a été évaluée par western blot deux jours post-transfection à partir de 30 µg de protéines totales de chaque extrait cellulaire. B) Les cellules transfectées ou non ont été utilisées pour réaliser un test MTT avec la Gem (à gauche) ou l'AraC (à droite). En bleu sont représentées les cellules Messa M36 non transfectées, en rouge les cellules transfectées avec le siARN 1 et en violet, les cellules transfectées avec le siARN 4. n=1, chaque lignée est testée en duplicat.

### III.2.3 Caractérisation des différentes mutations du M36

La protéine M36 étant issue du processus de rétrovolution à partir du G12, elle possède les trois mutations du G12 précédemment décrites et étudiées[59,228], en plus d'une mutation supplémentaire en position 169. Dans le cas du G12, j'avais déjà établi que toutes les mutations sont nécessaires pour l'efficacité de la protéine[228]. Concernant le M36, il nous fallait déterminer si la mutation S169N permet éventuellement de compenser l'absence d'une ou de plusieurs des mutations précédentes. Pour ce faire, la mutation S169N a été introduites dans les précédents mutants simples et doubles du G12, générant ainsi les mutants doubles et triples couvrant toutes les combinaisons possibles à partir des quatre mutations du M36 (figure 35A). A partir des constructions réalisées, les lignées stables exprimant chacune une des variantes de la protéine ont été établies et leur sensibilité à la Gem a été testée (figure 35B). Comme dans le cas du mutant G12, cette étude des différentes combinaisons de mutation du M36 a permis de déterminer que toutes les mutations du M36 sont requises pour l'effet observé.

A

| | Mutations du G12 | | | |
|---|---|---|---|---|
| Position | 169 | 171 | 247 | 249 |
| M36 | N | K | K | M |
| Mutant 1 (Δ) | N | K | - | - |
| Mutant 2 (□) | N | - | K | - |
| Mutant 3 (○) | N | - | - | M |

| | Mutations du G12 | | | |
|---|---|---|---|---|
| Position | 169 | 171 | 247 | 249 |
| M36 | N | K | K | M |
| Mutant 4 (Δ) | N | K | K | - |
| Mutant 5 (□) | N | - | K | M |
| Mutant 6 (○) | N | K | - | M |

B

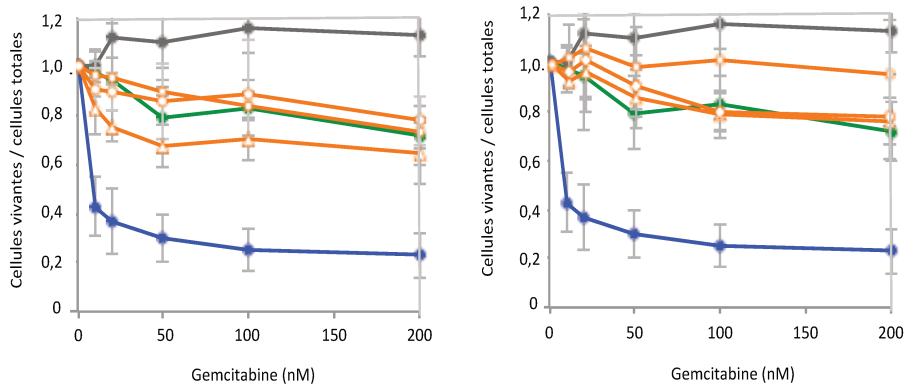



**Figure 35 : Contribution de la mutation S169N du M36.** A) Les tableaux indiquent les différentes combinaisons de mutations testées. Les cellules Messa10K ont été transduites avec des vecteurs permettant l'expression de ces mutants afin d'établir les lignées stables. B) Chacune de ces lignées a été utilisée pour réaliser des tests MTT avec la Gem. Sur les graphiques sont représentés en gris les cellules Messa10k non transduites, en vert les cellules exprimant une copie exogène de la dCK sauvage et en bleu les cellules exprimant le M36. en orange sont représentés les cellules exprimant les mutants testés. A gauche (tableau et graphique) sont reportés les mutants doubles et à droite (tableau et graphique) les mutants triples. n=3

### III.2.4 Modification du M36 sur une base rationnelle en vue d'identifier un meilleur variant de la protéine

Après avoir démontré que toutes les 4 mutations du M36 sont nécessaires pour son efficacité, nous avons voulu savoir s'il était possible d'améliorer la protéine sur une base rationnelle ou si la variante trouvée était la meilleure que nous puissions obtenir (par mutagénèse de ces positions).

Au cours de la rétrovolution, la génération des mutants n'est pas complètement aléatoire et est majoritairement influencée par deux facteurs. Le premier facteur est la nature des mutations introduites par la transcriptase inverse. En effet, certaines mutations tel que la

substitution des G par des A, sont favorisées par la transcriptase inverse du VIH-1, de même que les transitions sont plus fréquentes que les transversions[230]. Le deuxième facteur est le nombre de nucléotides qui doivent être mutés dans le codon d'origine pour obtenir le codon du nouvel acide aminé. En effet, si le changement d'un seul nucléotide est requis pour obtenir un acide aminé différent, le mutant généré par cette mutation unique sera plus fréquent qu'un autre mutant pour lequel le changement de deux ou trois nucléotides est requis.

En ce qui concerne la dCK, L'analyse de la nature des mutations montre que les mutations en position 171 et 247 conduisent à un changement de charge de l'acide aminé (K à la place de E), celle en position 169 conduit à un changement de polarité (N à la place de S) et celle en position 247 conserve la nature de l'acide aminé présent dans la dCK (M à la place de L). En analysant la séquence nucléotidique au niveau des mutations, on constate que les changements en acides aminés résultent tous de la substitution d'une seule base dans les codons. Le processus de rétrovolution semble donc indiquer que pour obtenir une meilleure capacité de la dCK à activer les médicaments, il faudrait en position 171 et 247 un acide aminé basique, en position 169 un acide aminé polaire et laisser un acide aminé hydrophobe en position 247.

Nous avons donc avancé l'hypothèse que d'autres acides aminés avec les mêmes caractéristiques (basiques, polaires ou hydrophobes) aux mêmes positions pourraient avoir un meilleur effet sur l'activité du mutant. Dans notre hypothèse, ces mutants n'ont cependant pas été générés au cours de la rétrovolution à cause de la difficulté biochimique et génétique pour leur génération.

Ainsi, nous avons modifié la séquence du gène codant le M36 de façon à conserver la nature des modifications présentent dans la protéine M36, tout en introduisant au niveau des positions mutées, des acides aminés avec une distance génétique plus éloignée de ceux présents dans la dCK sauvage. Le tableau récapitulatif des modifications effectuées est présenté dans la figure 36A. Chacune des positions a été modifiée dans la protéine M36, comme indiqué dans le tableau, générant ainsi 4 mutants ponctuels du M36 et un quadruple mutant. Après avoir établi les lignées stables permettant leur expression, un test de sensibilité à la Gem a été réalisé (figure 36B). On observe que les 4 mutants ponctuels ne permettent pas de sensibiliser les cellules aux médicaments et, qu'au contraire, l'introduction d'une seule de ces modifications dans le M36 est suffisante pour abolir toute

sensibilisation des cellules. Le quadruple mutant quant à lui présente un phénotype intermédiaire entre la dCK sauvage et le M36, avec une meilleure capacité à sensibiliser les cellules que la dCK sauvage, mais sans pour autant atteindre le niveau de sensibilisation du M36.

A

| Position | dCK | M36 | variant du M36 |
|---|---|---|---|
| 169 | Ser AGC | Asn AAC | Gln CAG |
| 171 | Glu GAA | Lys AAA | Arg CGC |
| 247 | | | |
| 249 | Leu CTG | Met ATG | Ile ATC |

| Position | 169 | 171 | 247 | 249 |
|---|---|---|---|---|
| M36 | N | K | K | M |
| Mutant 1 (○) | Q | K | K | M |
| Mutant 2 (□) | N | R | K | M |
| Mutant 3 (△) | N | K | R | M |
| Mutant 4 (◇) | N | K | K | I |
| Mutant 5 (●) | Q | R | R | I |

B

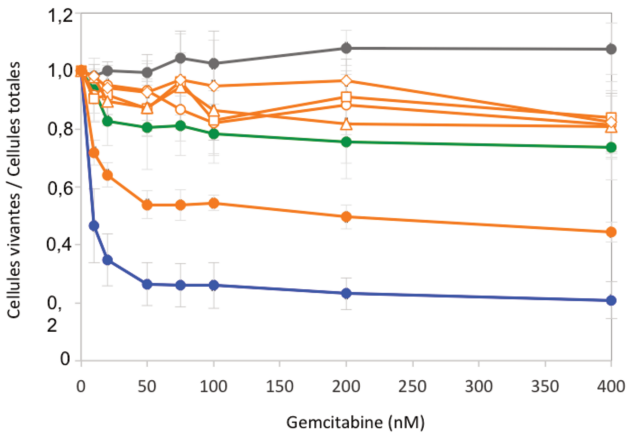



**Figure 36 : Modification du M36 sur une base rationnelle**. A) Tableau récapitulatif des variants du M36 testés. B) Tests de sensibilité des lignées stables exprimant les différents variant de la dCK. En gris sont représentés les cellules Messa 10K non modifiées, en vert les cellules exprimant la dCK exogène et en bleu les cellules exprimant le M36. En orange sont représentées les cellules exprimant les différents mutants ponctuels du M36 et en violet le quadruple mutant. n=3

### III.2.5 Etude structurelle de la protéine M36

Les tests biochimiques réalisés avec le mutant M36 ont montré que le M36 garde une capacité à phosphoryler la Gem et l'AraC comparable à celle de la dCK, alors qu'il perd la capacité à phosphoryler son substrat naturel la dC. Ainsi, la compétition entre la dC et les médicaments devient favorable à une activation des médicaments en présence du M36. Ce comportement avait notamment été observé avec le mutant G12 et la Gem[59]. Dans le but

d'avoir une meilleure compréhension des mécanismes moléculaires à l'origine de cette perte de capacité à phosphoryler la dC, nous avons entamé une collaboration avec l'équipe de Marc Ruff à l'IGBMC de Strasbourg, afin de réaliser la structure de la dCK et du M36 en présence de la dC ou de la Gem. En superposant les structures obtenues avec les deux substrats pour la même enzyme, on constate que dans le cas de la dCK (figure 37B), les deux substrats se positionnent de façon pratiquement identique au sein de la protéine. En ce qui concerne le M36 (figure 37 C), on constate pour la dC, spécifiquement, une délocalisation du groupement accepteur de phosphate, ce qui entraîne un éloignement de celui-ci de 0,7Å. Cette délocalisation pourrait expliquer la plus faible capacité du M36 à phosphoryler la dC comparé à la dCK.

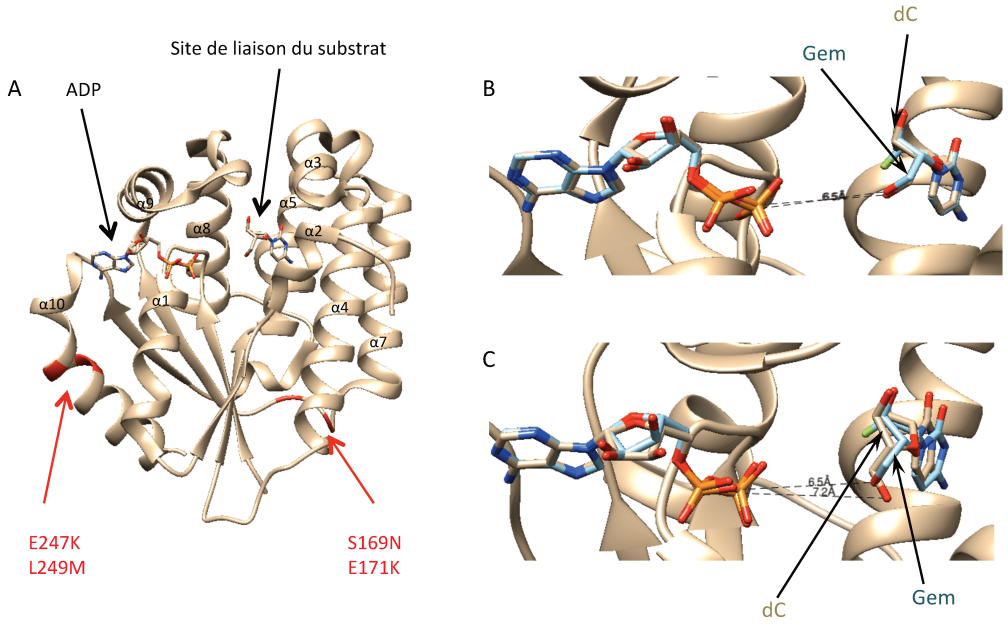



**Figure 37 : Structure comparative de la dCK et du M36 avec la dC et la Gem.** A) Représentation de la structure d'un monomère de dCK avec indiqué en rouge les quatre mutations du M36. Sont également indiqués les sites de fixation pour le donneur de phosphate et le site de fixation du substrat. B) Structure du site actif de la dCK avec une superposition de la dC (en beige) et de la Gem (en bleu) au niveau de ce site. La structure de la dCK avec la dC dans le site actif a été obtenue avec une résolution de 2,2 Å et celle de la dCK avec la Gem dans le site actif avec une résolution de 1,9 Å. C) Structure du site actif du M36 avec une superposition de la dC (en beige) et de la Gem (en bleu) au niveau du site. La structure du M36 avec la dC dans le site actif a été obtenue avec une résolution de 2,9 Å, et celle avec la Gem dans le site actif, avec une résolution de 2,3 Å. Les pointillés correspondent à la distance entre les sites donneur et accepteur de phosphate. Les figures ont été obtenues avec le logiciel PyMol et les distances entre les différents groupements ont été déterminées en utilisant ce même logiciel.

## IV Discussion

La thérapie génique est une approche intéressante pour différents domaines de biomédecine tel que la médecine régénérative ou le traitement du cancer. Grâce à la rétrovolution, une nouvelle méthode d'évolution moléculaire mise en place au laboratoire, deux mutants de la dCK potentiellement intéressant comme gène suicide ont été identifiés. Le premier, appelé G12 a démontré sa capacité à sensibiliser les cellules à des faibles concentrations de Gem[59]. Avec le G12 comme point de départ pour la rétrovolution, un second mutant de la dCK a également été identifié et est appelé M36. Le G12 est caractérisé par 3 mutations (E171K, E247K, L249M) comparé à la dCK sauvage, et le M36 contient une mutation supplémentaire (S169N) comparé au G12. Les premières études réalisées sur les mutants G12 et M36 ont été réalisées avec des vecteurs lentiviraux portant des séquences LTR fonctionnelles du fait du mécanisme de fonctionnement de la rétrovolution qui a permis leur identification (figure 11). Nous avons montré ici que l'utilisation de vecteurs SIN qui est requise pour une utilisation des vecteurs lentiviraux pour des applications biomédicales, permet une meilleure efficacité de la protéine (article-figure 5). Les raisons pouvant expliquer cette meilleure efficacité avec les vecteurs SIN, comme évoqué dans la discussion de l'article, sont vraisemblablement liées à un effet d'interférence transcriptionnelle entre le promoteur interne qui porte à l'expression du transgène et le promoteur viral contenu dans les LTR. Au cours de ce travail, nous avons également mis en évidence le fait qu'en plus de sensibiliser les cellules à la Gem, le M36 permet également de les sensibiliser fortement à l'AraC (Article-figure 7). Cet effet n'avait pas été mis en évidence pour le G12 mais les conditions dans lesquelles les tests avec le G12 avaient été effectués étant un peu différentes on ne peut pas exclure qu'une certaine sensibilisation ne soit pas présente avec le G12 également. Quoi qu'il en soit, la possibilité de sensibiliser les cellules à deux médicaments différents (comme pour le M36) est très intéressant car elle élargit l'intérêt de ce mutant à plusieurs types de cancers comme les cancers solides du pancréas, du sein et du colon (par le biais de la Gem) et des leucémies (par le biais de l'AraC).

Le M36 permet également une meilleure activation comparé à la dCK sauvage du 3TC, un analogue de la dC activé par la dCK et utilisé pour le traitement des infections aux VIH (Article-figure 6). La stratégie que nous envisageons dans le contexte de l'utilisation du M36 pour des finalités antivirales est de modifier des cellules souches hématopoïétiques pour

qu'elles expriment le transgène. Une fois réimplantées chez le patient, ces cellules, et celles auxquelles elles auraient donné origine, seraient plus résistantes à l'infection lors d'un traitement par l'antiviral. Cette stratégie à déjà fait ses preuves dans un contexte différent : pour la mutation CCR5 $\Delta$ 32, qui empêche l'expression du corécepteur majeur pour le VIH. Il a en effet été mis en évidence, en 2009, que la mutation CCR5 $\Delta$ 32 permet aux cellules implantées d'être résistantes à l'infection des souches VIH R5 tropiques[231]. La réduction de l'expression de CCR5, atteinte par le biais de vecteurs lentiviraux permettant l'expression de shARN[232,233], ou encore des nucléases en doigt de zinc[234,235] dirigés contre le gène codant le CCR5, a permis d'atteindre une réduction de l'expression de CCR5 dans les cellules transduites allant jusqu'à 90%, et cette inhibition de l'expression de la protéine a été corrélée avec la résistance des cellules à l'infection[233]. En complément de ces observations, les résultats obtenus avec le 3TC et les cellules exprimant le M36 permettent d'envisager une stratégie combinée dans laquelle les vecteurs utilisés permettraient à la fois une expression stable du M36 et une inhibition de l'expression du CCR5. L'utilisation de tels vecteurs pourraient permettre de renforcer l'effet du traitement par la délétion du récepteur CCR5.

La lignée cellulaire utilisée comme modèle pour la sélection des mutants générés par la rétrovolution est la lignée Messa 10K. Cette lignée cellulaire issue d'un carcinome utérin a été adaptée en laboratoire afin d'être résistante à la Gem et exprime une dCK inactive[229]. De ce fait, la lignée Messa 10K permet d'avoir une bonne appréciation de l'efficacité des mutants identifiés car elle présente un fond génétique « vierge » en ce qui concerne l'activité de la dCK. Physiologiquement, ces cellules constituent un modèle pour les cellules résistantes au traitement qui émergent fréquemment lors des thérapies anticancéreuses. Exprimée dans les cellules Messa 10K, le M36 permet de sensibiliser les cellules à des concentrations de Gem 900 fois inférieures à celles requises pour les cellules non modifiées, et à des concentrations d'AraC jusqu'à 10 000 fois inférieures (article). Toutefois, la capacité du M36 à sensibiliser à la Gem ou à l'AraC, d'autres lignées cellulaires exprimant une dCK endogène fonctionnelle, n'a pour l'instant pas encore été démontré (figure 28). Cette observation est en adéquation avec les résultats obtenus par d'autres équipes. Il est en effet décrit que, en fonction de la lignée cellulaire utilisée, l'apport d'une copie supplémentaire de la dCK ou d'une de ses variantes ne permet pas d'atteindre des niveaux de sensibilisation équivalents des cellules aux médicaments[59,77,236]. Afin de pallier à ces problèmes, il serait

nécessaire soit d'introduire plusieurs copies du transgène (objectif qui peut être atteint en culture de cellules en utilisant une MOI plus élevée, mais qui est difficilement atteignable *in vivo*), soit d'utiliser un promoteur plus fort que le promoteur EF1-$\alpha$ utilisé dans nos expériences, pour conduire à l'expression du transgène. Le promoteur EF1-$\alpha$ est un promoteur constitutif, et une expression du transgène à partir d'un promoteur plus fort et spécifique aux types de cellules cancéreuses ciblées (comme par exemple le promoteur PSA pour le cancer de la prostate ou le promoteur MIA pour les mélanomes)[237], pourrait être envisagé pour une meilleure expression du transgène.

Une autre possibilité pour augmenter l'efficacité du M36 dans les cellules serait d'utiliser une stratégie combinée entre l'expression du M36 et la répression de la RRM1. Nous avons montré qu'une telle stratégie utilisant des siARN dirigés contre le gène codant la RRM1, permet une meilleure sensibilisation des cellules à la Gem (figure 34). Toutefois, lorsque les siARN doivent être générés à partir de leurs précurseurs shARN eux même introduit via l'utilisation de vecteurs lentiviraux, aucun effet n'est observé. Il se peut que dans ce contexte, l'expression du siARN à partir du précurseur shARN ne soit pas suffisante pour observer une sensibilisation des cellules. Il serait donc intéressant d'optimiser nos précurseurs pour augmenter l'efficacité de génération des siARN, si cela est l'origine du problème.

Le mutant M36 est caractérisé par 4 mutations organisées par paire. Les mutations S169N et E171K sont situées au niveau de l'interface de dimérisation de la protéine alors que les mutations E247K et L249M sont situées au niveau de l'hélice $\alpha$10 qui influence sur le repliement de la protéine ainsi que sur sa capacité à fixer les substrats. Ici, nous avons mis en évidence pour le M36, comme pour le G12 auparavant (article-figure 1), la nécessité de l'ensemble des mutations car. En effet, lorsqu'une mutation du M36 est éliminée, le mutant généré perd la capacité, à sensibiliser les cellules au traitement (figure 34). Le fait que la réponse des cellules soit comparable à celle des cellules transduites avec une dCK sauvage et non à celle des cellules Messa 10K non transduites, suggère que les mutants ont quand-même gardé une activité semblable à celle de la dCK sauvage (figure 34).

Au cours de cette étude nous avons aussi tenté d'améliorer l'efficacité du M36 sur une base rationnelle. Notre hypothèse a été qu'en gardant la nature des acides aminés introduits par

la rétrovolution (basique, hydrophobe, polaire), aux mêmes positions (169, 171, 247 et 249), mais en allant plus loin au niveau de la divergence génétique au niveau de la séquence des codons, on pourrait obtenir une variante de la dCK encore plus efficace que le M36. Lorsque, sur cette base, nous avons réalisé la substitution d'une seule des 4 positions mutées dans le M36, cela a également conduit à une perte de l'efficacité des 4 mutants générés et à une activité de ces mutants qui est comparable à celle de la dCK sauvage (figure 35 lignées représentées avec des ronds vides). En revanche, lorsque les 4 positions sont substituées simultanément, on obtient un phénotype de sensibilisation intermédiaire entre celui du M36 et celui de la dCK sauvage (figure 35 lignées représentées en orange avec des ronds pleins). Il serait intéressant de poursuivre l'étude du mutant afin notamment d'obtenir une meilleure compréhension de l'impact de chacune des mutations sur la structure de la protéine. Il semble en effet, suite à l'étude de l'importance des mutations que j'ai effectuée, que les codons les plus adaptés pour l'effet que nous souhaitions observer aient été sélectionnés au cours du processus de rétrovolution.

Les tests biochimiques réalisés (Article-Tableau 2) indiquent que l'efficacité du M36 serait liée à une baisse de sa capacité à phosphoryler la dC et non à une hausse de sa capacité à phosphoryler la Gem comparé à la dCK. Cette observation semble confirmée par la comparaison des structures de la dCK et du M36 en présence de la dC ou de la Gem. En effet, il apparait que dans la protéine M36, le site accepteur de phosphate de la dC est plus éloigné du donneur de phosphate de 0,7 Å comparé à sa position dans la dCK. Cela rendrait la phosphorylation de la dC par le M36 potentiellement plus difficile, alors que la Gem garde sensiblement la même position dans la dCK et dans le M36. Il faut toutefois considérer le fait que la structure du M36 en interaction avec la dC a été obtenue avec une résolution de 2,9Å qui reste une résolution de faible qualité[238,239], nous obligeant donc à rester prudents avec les conclusions tirées. La structure de la protéine semble être difficile à obtenir ce qui explique que la première structure obtenue avec une bonne résolution n'ait été publiée qu'en 2003 et ait nécessité de réaliser différentes substitutions dans la protéine[74]. Une des structures publiées de la dCK a notamment nécessité une substitution de la position 247 (E247R), afin de pouvoir être résolue[240]. Si l'introduction d'une mutation en cette position est nécessaire pour l'obtention de la structure la protéine, cela suggère que cette position est impliquée dans la stabilité et donc probablement dans l'activité de la protéine. Sachant que l'une des mutations caractéristiques du G12 et du M36 est la mutation E247K, cela

suggère la possibilité que certaines des mutations introduites au cours de la rétrovolution pourraient avoir un effet sur la stabilité de la protéine. Il est également reporté que les modifications post-traductionnelles influent sur les propriétés structurelles et sur l'activité de la protéine. Afin de pouvoir lier le substrat, la protéine adopte une conformation « ouverte ». Une fois la liaison effectuée, elle passe à une conformation « fermée » qui la rend compétente pour le transfert de phosphate, puis adopte à nouveau la conformation « ouverte » pour la libération du substrat phosphorylé. Il est décrit que la phosphorylation de la Ser en position 74 influe sur la cinétique de ces changements de conformation au cours de la réaction affectant ainsi l'activité de la protéine[75]. Si les modifications introduites au cours de la rétrovolution influent sur les modifications post-traductionnelles du M36, on pourrait envisagé que l'efficacité de la protéine soit liée à des modifications au niveau de la cinétique de phosphorylation.

La protéine M36 pourrait présenter un grand intérêt comme gène suicide dans les stratégies de lutte contre le cancer. Elle pourrait également être utilisé comme gène suicide de sécurité dans les stratégies de transfert de cellules modifiées en cas de complications liés à l'injection des cellules modifiées tel que les cas de réaction contre le greffon (GvHD) reportés dans différents essaies cliniques[241]. Actuellement le gène suicide le plus fréquemment utilisé est la TK du virus herpes simplex 1 (hsTK), mais les problèmes d'immunité contre les cellules modifiées exprimant cette protéine virale poussent à la recherche de nouvelles protéines. Les protéines d'origine humaines intéressent particulièrement la recherche car les variantes de ces protéines ne devraient avoir qu'un effet immunologique minime. Tout comme la hsTK, le mécanisme d'action du M36 est basé sur le métabolisme des cellules, et les cellules à multiplication rapide sont donc les plus sensibles aux traitements. Récemment, un autre gène suicide a été caractérisé : la iCasp9 (caspase 9 inductible)[15] qui dérive de la caspase 9 humaine, une protéine impliquée dans la voie apoptotique. Le mécanisme d'action de la protéine est donc différent de celui de la hsTK et du M36 car il n'est pas en relation avec la réplication de l'ADN. Une fois activée, la iCasp9 permet d'éliminer rapidement jusqu'à 99% des cellules modifiées. Le mécanisme d'action de la iCasp9 constitue cependant une limite pour son utilisation dans certaines stratégies tel que la médecine régénérative. En effet, dans le cadre d'une greffe de cellules modifiées pour la reconstruction d'un organe, si une partie des cellules modifiées connait une croissance incontrôlée alors que la majorité des cellules modifiées gardent un

comportement normal, il sera souhaitable d'éliminer préférentiellement les cellules modifiées à croissance incontrôlée. Dans ce contexte, l'utilisation de la iCasp9, du fait de son mécanisme d'action, ne permettra pas de discriminer les deux types de cellules modifiées et va conduire à l'élimination de la totalité d'entre elles. En revanche, l'utilisation d'un gène suicide tel que le M36 dont l'action est basée sur la réplication des cellules permettra de cibler préférentiellement les cellules ayant un comportement anormal tout en ayant un effet limité (dépendamment des doses de médicament utilisées) sur celles ayant un comportement normal.

# Partie V : Modification de l'enveloppe pour un ciblage spécifique des cellules

## I Objectifs

Pour les stratégies de lutte anti-tumorale par insertion de gènes suicide, l'une des difficultés majeures est le transfert spécifique du gène dans les cellules cancéreuses. Pour ce faire, différentes stratégies peuvent être appliquées en fonction du vecteur utilisé. En ce qui concerne les vecteurs lentiviraux, qui m'intéressent plus particulièrement, la principale stratégie actuellement utilisée consiste au pseudotypage, à savoir l'utilisation des glycoprotéines d'enveloppe d'autres virus pour habiller les vecteurs lentiviraux. Au cours de mon travail de thèse, j'ai travaillé à la preuve de concept qu'une enveloppe VIH peut être modifiée afin de permettre aux vecteurs lentiviraux un ciblage préférentiel des cellules cancéreuses. Ce ciblage se ferait grâce à l'introduction dans cette enveloppe d'un ligand spécifique au marqueur tumoral des cellules cibles. Ce travail se base d'une part sur les connaissances du laboratoire sur l'enveloppe VIH-1 ainsi que sur la littérature, et d'autre part, sur l'identification au cours des dernières années de nombreux marqueurs tumoraux exprimés à la surface des cellules cancéreuses.

Afin de faire la preuve de concept d'une telle utilisation de l'enveloppe VIH il a fallu réaliser des tests préliminaires, ce que j'ai fait au cours de cette deuxième partie de mon travail de thèse. Il s'agissait notamment de :

1. construire une enveloppe VIH-1 ne nécessitant pas le récepteur CD4 pour réaliser l'entrée (enveloppe CD4 indépendante, ou CD4i)
2. réduire le tropisme résiduel des enveloppes CD4i pour le récepteur CD4 car, il est montré qu'elle garde une capacité résiduelle à entrer dans les cellules CD4+[182,208]
3. déterminer la région de l'enveloppe virale dans laquelle introduire le ligand du marqueur tumoral
4. déterminer la manière dont ce ligand doit être introduit afin de ne pas affecter la fonctionnalité de l'enveloppe

## II Matériels et méthodes

## II.1 Modification de l'enveloppe

Les mutations ponctuelles ont été introduites par PCR reconstitutive dont le principe est présenté dans la figure 29 et comme décrit précédemment[227]. De manière succincte, le gène codant l'enveloppe est amplifié afin de générer deux fragments possédant une partie homologue d'environ 30 nt incluant la position de la mutation à introduire. Pour ce faire, les amorces internes au gène utilisées pour chacune des deux PCR sont complémentaires et contiennent la mutation à introduire. La taille et la pureté des fragments générés sont analysées par migration sur gel d'agarose de 1% avant d'être purifiés sur colonne UltraClean® GelSpin® DNA extraction kit (Qiagen, CA, USA). La seconde PCR permettant de générer un gène complet contenant les mutations a été réalisée avec 0,25 pmol de chaque fragment. Pour ce faire, les 8 premiers cycles de la réaction sont réalisés en absence d'amorces afin de favoriser l'appariement entre les deux fragments sur la région homologue. Les amorces externes utilisées pour la première réaction de PCR sont ensuite rajoutées et la réaction est menée à terme avec 22 cycles supplémentaires. Toutes les PCR sont réalisées avec le kit « *Phusion High Fidelity PCR Kit* » (Thermo Fischer). Le produit final de la PCR est analysé sur gel d'agarose 1% avant d'être purifié et cloné dans un plasmide pcDNA3.1 TOPO (Invitrogen, CA, USA). Le clonage est effectué par restriction enzymatique avec les enzymes EcoRV et XhoI.

La modification de la région gp41 à quant à elle été réalisée par restriction enzymatique avec les enzymes XhoI et AleI à partir des plasmides pcDNA-ADA et pcDNA-HxB2.

## II.2 Cellules

Les cellules HEK 293T, les cellules HEK 293T $CD4^+CCR5^+$ ainsi que les cellules HEK 293T $CCR5^+$ ont été mises en culture dans le milieu DMEM (Gibco) supplémenté de 10% de FBS et de 100 U / ml de pénicilline-100 mg / ml de streptomycine (Invitrogen, CA, USA). Les cellules sont maintenues en culture dans un incubateur à 37°C avec 5% de $CO_2$.

## II.3 Productions des particules virales

Les particules virales sont produites par transfection des cellules HEK 293T avec 2 µg de plasmide p8.91 (GenBank ID : JQ585717.1) qui code pour les protéines de capside, la polymérase ainsi que pour l'intégrase virale, de 2 µg de plasmide pSDY-Luc qui permet de générer l'ARN génomique contenant le gène codant la luciférase, et 1 µg de plasmide

pcDNA-Env. Cinq millions de cellules sont ensemencées la veille dans des boîtes de culture de 10 cm de diamètre. À 80 % de confluence, les cellules sont transfectées avec le mix précédemment décrit et contenant du polyéthylènimine (PEI) linéaire MW~25 000 (Polysciences, Inc). Le milieu de culture des cellules est changé 6 heures après la transfection. Deux jours post-transfection, le milieu de culture est récolté, filtré avec des filtres de 0,45 µm et concentré à l'aide de Vivaspin®20 (Sartorius Stedim Biotech). Le titre viral est estimé par la quantification de la protéine de capside p24, par test ELISA (InnotestTM HIV Antigen mAB, International Genetic Technologies, Chilly-Mazarin, France).

### II.4 Western blot

L'expression des enveloppes chimériques ainsi que leur présence au niveau des particules virales sont vérifiés par western blot avant de réaliser les tests d'entrée. Pour ce faire, les cellules HEK 293T productrices sont récoltées et lysées avec une solution de RIPA 1X. Après avoir récupéré le surnageant viral, les cellules sont lavées par du PBS et récupérées dans ce même tampon. Les cellules sont centrifugées à 1200 rpm pendant 3 min, le milieu est éliminé et les cellules re-suspendues dans 100 µL de RIPA et gardées sur glace pendant 15 min. Les tubes sont ensuite centrifugés 20 min à 12 000 g et à 4°C après quoi le surnageant est collecté et conservé à -20°C. Un dosage des protéines totales de chaque extrait cellulaire est ensuite réalisé par la méthode de Bradford et 30 µg de protéines totales sont utilisées pour vérifier l'expression des enveloppes modifiées à partir du plasmide pcDNA-Env utilisé pour la transfection. Après la quantification de la protéine de capside p24 (Partie IV- II.1.3), le volume de suspension correspondant à 8 µg de p24 est également utilisé pour vérifier la présence des enveloppes au niveau des particules virales et ceux pour chaque suspension.
La migration des échantillons se fait sur un gel de polyacrylamide de 10% et les protéines virales sont révélées en utilisant du sérum de patient inactivé à 95°C pendant 5 min et contenant des anticorps dirigés contres les virus VIH-1 du sous types B à une dilution de 1/5000. L'anticorps secondaire utilisé est l'anti-humain produit chez la chèvre (Sigma-Aldrich, St-Louis, USA) à une dilution de 1/20 000. La détection des protéines est ensuite réalisée par autoradiographie et la quantification des bandes est faite par le logiciel Image Lab™ Software de BioRad.

### II.5 Test d'entrée

Afin de tester la capacité des enveloppes modifiées à réaliser l'entrée, les cellules HEK 293T CD4+ R5+ ou les cellules HEK 293T CCR5+ sont transduites avec les différents vecteurs produits. La transduction est réalisée en utilisant 125 000 cellules et 12,5 ng de p24 de chaque suspension virale en présence de 2 µg/mL de polybrène (Sigma-Aldrich, St-Louis, USA) dans des plaques de culture de 24 puits. Deux jours post-transduction, les cellules sont récoltées et lysées afin d'évaluer le niveau d'expression de la luciférase. Les cellules sont récoltées dans du PBS et centrifugées à 1200 rpm pendant 3 min. Le surnageant est éliminé et les cellules remises en suspension dans 150 µL de tampon de lyse *Reporter Lysis 5X buffer* (Promega, Madison WI, USA) dilué au cinquième. Les tubes sont alors placés à -80°C jusqu'à la congélation totale et décongelé à température ambiante. Après une centrifugation de 5 min à 13 000 g, 50 µL du surnageant sont utilisés pour le test. Celui-ci est réalisé avec un luminomètre Glomax® 96 puits (Promega, Madison WI, USA) et en utilisant, par échantillon, 100 µL de luciférine provenant du kit « *Luciferase assay system* » (Promega, Madison WI, USA).

# III Résultats

## III.1 Construction d'une enveloppe CD4i

Les variantes d'enveloppe aux quelles nous nous sommes intéressés au cours de ce travail sont dérivées de la souche ADA et de la souche HxB2 du VIH-1. La souche ADA a été isolée à partir de PBMC de patient et utilise la protéine CCR5 (R5) comme corécepteur pour l'entrée (Partie II-III.2.1.1). La souche HxB2, quant à elle, a été isolée d'une lignée cellulaire infectée en laboratoire et utilise la protéine CXCR4 (X4) pour réaliser l'entrée. La numérotation des acides aminée de ces deux souches étant légèrement différente, j'utiliserais uniquement la numérotation de la souche HxB2 pour plus de simplicité et de clarté. Le but final de ce travail est de généré une enveloppe CD4i capable de réaliser l'entrée en utilisant le corécepteur X4 qui est surexprimé dans plus d'une vingtaine de cancer[242–245]. Cependant, les informations les plus prometteuses concernant la génération d'une enveloppe CD4i disponible dans la littérature concernent les enveloppes de tropisme R5. Ainsi, dans ce travail préliminaire, j'ai utilisé une enveloppe de tropisme R5 qui pourra être convertie par la suite en enveloppe X4 tropique (Partie II-III.3.1).

### III.1.1 Augmenter la capacité à réaliser l'entrée dans les cellules CD4-

A partir des enveloppes disponibles, il fallait générer des enveloppes CD4i (cf Partie II-III.3.2), et pour cela, je me suis basée sur la littérature qui est riche en informations concernant l'enveloppe du VIH-1 et son interaction avec le CD4. Je me suis particulièrement intéressée à deux enveloppes décrites comme ayant une bonne capacité à lier le corécepteur R5, sans besoin de liaison préalable avec le récepteur CD4. La première enveloppe que j'ai considérée contient une substitution de l'arginine en position 197 par une sérine (N197S, figure 38A), ce qui conduit à la perte de la glycosylation de la gp120 en cette position. La seconde est une enveloppe chimérique qui possède la région gp120 de la souche ADA, la région gp41 de la souche HxB2 ainsi que la mutation ponctuelle en position 197 (N197S-HxB2, figure 38A).

J'ai dans un premier temps évalué la capacité de ces deux enveloppes à réaliser l'entrée afin de confirmer les résultats précédemment décrits. Les tests que j'ai effectués se basent sur l'expression relative de la luciférase dans les cellules transduites avec les vecteurs portant les différentes enveloppes. Dans ces tests, l'expression de la luciférase dans les cellules transduites est proportionnelle à la capacité de l'enveloppe à réaliser l'entrée puisque le gène codant la luciférase est porté par l'ARN génomique des vecteurs. Les résultats obtenus et présentés dans la figure 33B montrent que les cellules CD4+R5+ transduites avec le vecteur portant l'enveloppe WT expriment fortement la luciférase, alors que les cellules R5+ transduites avec ce même vecteur expriment la luciférase à un niveau significativement plus faible (p=0,00078) et non significativement différentes (p=0,26) des témoins négatifs (pcDNA3.1 dans la figure). Le mutant contenant la mutation ponctuelle N197S, tout en gardant sa capacité à entrer dans les cellules CD4+R5+, montre une incapacité à entrer dans les cellules CD4-R5+, contrairement à ce qui a été décrit dans la littérature. En revanche, l'enveloppe chimérique N197S-HxB2 permet d'augmenter l'expression de la luciférase de 30 fois dans les cellules n'exprimant que le corécepteur R5 comparé à l'enveloppe WT, tout en gardant une capacité à entrer dans les cellules CD4+ comparable à celle du WT, en accord avec ce qui a déjà été décrit. Compte tenu de ces résultats, nous avons poursuivi les expériences avec l'enveloppe N197S-HxB2.

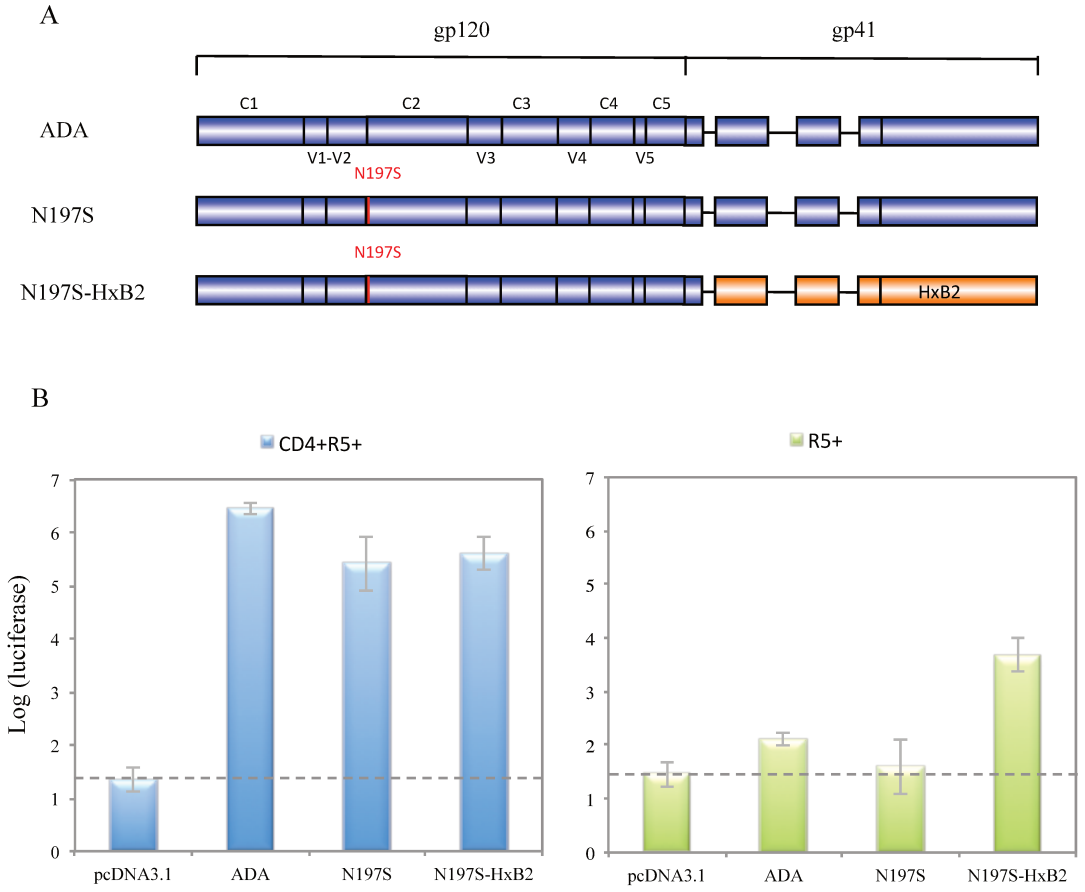


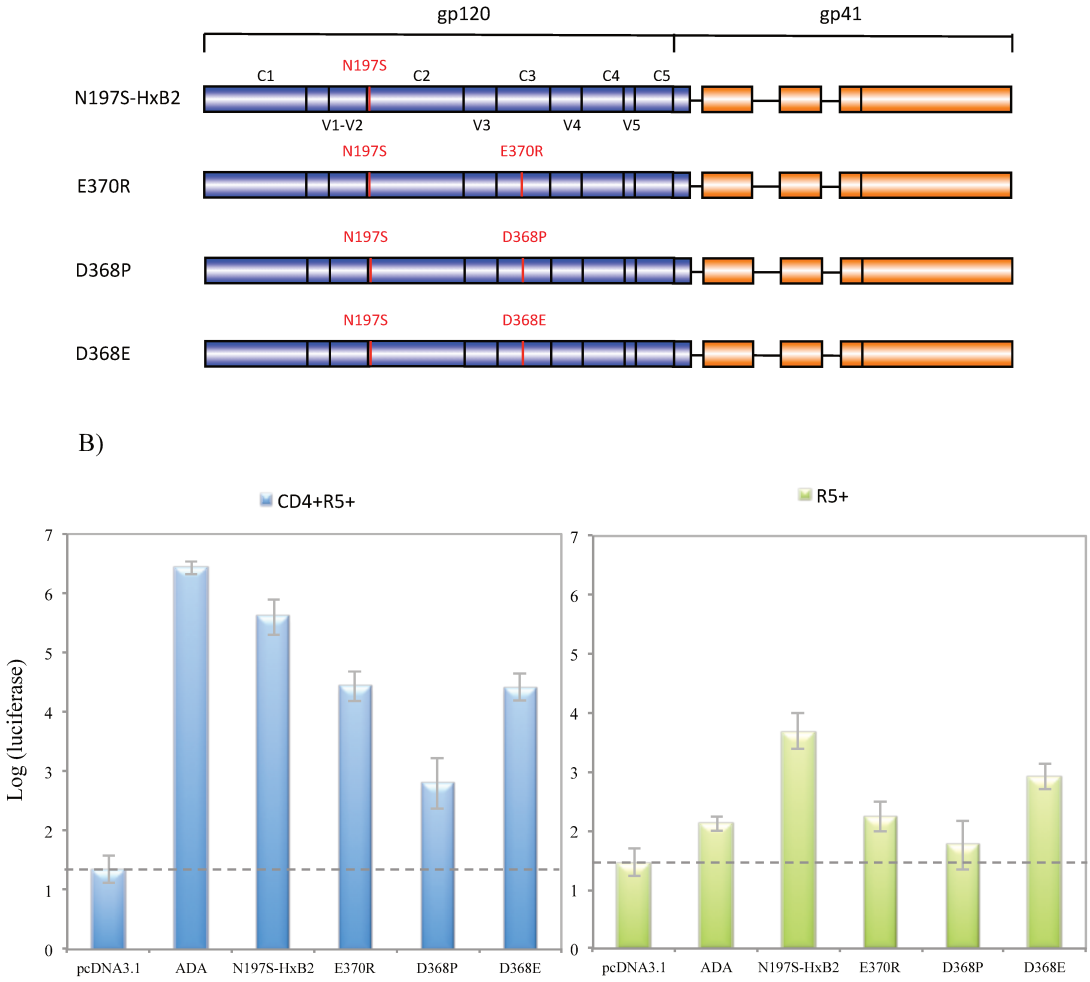

**Figure 38 : Augmentation de la capacité de l'enveloppe à entrer dans les cellules CD4-.** A) schéma représentatif de l'enveloppe ADA et de ces différentes régions constantes et variables. Sont également représentés les mutants de l'enveloppe qui ont été testées. B) Résultats des tests d'entrée réalisés (n=3) avec les trois enveloppes représentées, sur les cellules exprimant soit le récepteur CD4 et le corécepteur CCR5 (CD4+R5+) soit uniquement le corécepteur (R5+). Le pcDNA3.1 correspond aux vecteurs ne portant pas d'enveloppe virale et constitue le contrôle négatif de nos expériences. Le trait pointillé représente la limite d'expression de la luciférase en dessous de laquelle les résultats sont considérés comme étant négatifs. Il est défini par la valeur maximale de luciférase dans les échantillons pcDNA3.1.

### III.1.2 Réduire la capacité à réaliser l'entrée dans les cellules exprimant le récepteur CD4

Pour une utilisation de l'enveloppe VIH pour le ciblage spécifique des cellules cancéreuses, il faut que celle-ci perde le plus possible sa capacité naturelle à accomplir

l'entrée dans les cellules exprimant le CD4. Afin de réduire cette capacité de l'enveloppe N197S-HxB2 précédemment générée, j'ai évalué l'effet de l'introduction de mutations supplémentaires, décrites dans la littérature dans cette enveloppe.

En effet, en plus de la position 197 de la gp120, les positions 368 et 370 de la protéine ont également été identifiées comme affectant la capacité de l'enveloppe à lier le récepteur CD4 et différentes substitutions baissant cette capacité sont décrites dans la littérature au niveau de ces deux positions[207]. Après l'analyse des résultats de ces différentes substitutions, j'ai choisi d'en tester trois qui semblaient plus prometteuses pour les finalités de mon travail (figure 39A). En ce qui concerne la position 368, les deux mutations choisies (D368P et D368E) sont celles qui présentaient la plus faible capacité à interagir avec le CD4 tout en gardant une très bonne capacité de maturation de la gp160, le précurseur de l'enveloppe, et un faible taux de *shedding* (Partie II-III.3.1). En ce qui concerne la position 370, le mutant testé (E370R) est celui qui présentait la plus faible capacité à lier le CD4 même s'il présentait une plus faible maturation du précurseur de l'enveloppe[207]. Les vecteurs possédant les différentes enveloppes modifiées ont été produits et utilisés comme décrit précédemment pour réaliser les tests d'entrée (figure 39 B). Les trois enveloppes montrent toutes une baisse de la capacité à entrer dans des cellules CD4+ par rapport à l'enveloppe N197S-HxB2. Alors que les mutants D368P et E370R montrent une baisse importante de la capacité à réaliser l'entrée dans les cellules n'exprimant que le corécepteur R5, l'enveloppe D368E garde une capacité d'entrée dans ces cellules qui est comparable ($p=0,05$) à celle de l'enveloppe N197S-HxB2. Cette enveloppe apparaît donc comme la plus prometteuse et, en ce qui concerne l'entrée dans les cellules CD4+R5+, elle est 10 fois moins efficace pour l'entrée comparée à l'enveloppe N197S-HxB2. Comparée à l'enveloppe ADA WT, l'enveloppe D368E permet d'une part, de réduire d'environ 100 fois l'entrée dans les cellules CD4+ et d'autre part d'augmenter la capacité à réaliser l'entrée dans les cellules CD4- de 10 fois.




**<u>Figure 39 : Réduction de la capacité de l'enveloppe à entrer dans les cellules CD4+.</u>** A) Schéma représentatif des enveloppes testées. B) le graphique représente les résultats obtenus pour les tests d'entrée réalisés (n=3) avec les différentes enveloppes représentées. Les enveloppes ont été testées pour leur capacité à entrer dans les cellules exprimant le récepteur et le corécepteur (CD4-R5) ou dans les cellules ne possédant que le corécepteur (R5). Le pcDNA3.1 correspond aux vecteurs ne portant pas d'enveloppe virale et constitue le contrôle négatif de nos expériences. Le trait pointillé représente la limite d'expression de la luciférase en dessous de laquelle les résultats sont considérés comme étant négatifs. Il est défini par la valeur maximale de luciférase dans les échantillons pcDNA3.1.

## III.2 Introduction d'un ligand spécifique pour le marqueur tumoral

### III.2.1 Choix du marqueur tumoral à utiliser comme modèle

Afin de tester notre hypothèse de pouvoir réorienter l'entrée des enveloppes de virus VIH-1 en se basant sur l'expression d'un marqueur tumoral exprimé à la surface des cellules cibles, il nous fallait un marqueur modèle. Nous avons choisi la protéine HER2 qui est une protéine membranaire surexprimée dans différents types de cancers (Figure 40A). Elle joue notamment un rôle important dans le diagnostic et le traitement du cancer du sein et a, de ce fait, été largement étudiée. Différents ligands présentant une bonne affinité pour la protéine ont été décrit. Nous avons choisi d'utiliser pour nos tests le ligand AHNP de 12 acides aminés (YCDGFYACYMDV)[246] (Figure 40B). Ce peptide construit suite à la comparaison de la structure de différents anticorps monoclonaux capables de lier la protéine HER2 présente une bonne affinité pour la protéine avec un $K_D$ de 300nM. Notre choix c'est notamment porté sur le peptide AHNP du fait de sa petite taille, en vue de limiter la perturbation de la fonctionnalité de l'enveloppe suite à son introduction.

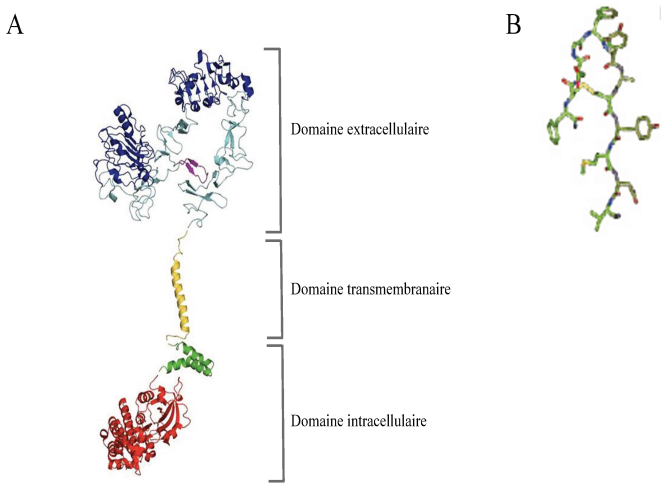



**Figure 40 : Représentation de la protéine HER2 (A) et de son ligand AHNP (B)[9,10].**

### III.2.2 Choix du site d'introduction du ligand

Une fois le choix de la protéine modèle ainsi que de son ligand effectué, il nous fallait déterminer la région de l'enveloppe dans laquelle insérer le peptide ligand. La protéine gp120 de l'enveloppe virale est composée de 5 régions constantes (régions C1 à C5) et de 5

régions variables (V1 à V5, figure 23). Les régions variables sont les plus permissives aux modifications. L'introduction de peptides dans ces différentes régions variables a déjà été réalisée par d'autres équipes. L'analyse de ces différentes études montre que les régions V1 et V4 sont celles qui sont les plus intéressantes pour l'introduction d'un ligand. De ce fait, à partir de l'enveloppe D368E précédemment générée, j'ai évalué l'effet de l'insertion du peptide AHNP dans la région V1 ou V4 (figure 41A) sur la capacité de notre enveloppe à réaliser l'entrée. L'entrée par le D368E étant meilleure dans les cellules CD4+R5+, ces cellules ont été choisies pour les tests d'entrée dont les résultats sont présentés dans la figure 41B. Lorsque le peptide AHNP est inséré dans la région V1 de la gp120 (D368E-AHNP-V1), cela conduit à une perte quasi totale de l'expression de la luciférase dans les cellules transduites comparée au niveau d'expression dans les cellules transduites qui s'avoisine au niveau d'expression observé avec les vecteurs portant d'enveloppe (pcDNA3.1). Par contre, lorsque le peptide AHNP est introduit dans la région V4 (D368E-AHNP-V4), cela ne réduit pas l'expression de la luciférase dans les cellules transduites comparée à l'expression dans les cellules transduites avec les vecteurs portant l'enveloppe D368E sans AHNP. L'introduction du peptide dans la région V4 permet donc de mieux conserver la capacité de l'enveloppe à réaliser l'entrée.

### III.2.3 Choix de la manière dont introduire le ligand

Différentes stratégies d'insertion des peptides ont été décrites dans la littérature (Partie II-III.3.3). L'insertion de tout peptide dans une protéine peut porter à une perturbation de la structure de la protéine. Afin de minimiser cette perturbation, une stratégie courante est d'effectuer l'insertion en rajoutant, aux bordures de la séquence exogène une région flexible qui agit comme « *linker* ». Ici, j'ai choisi de comparer l'effet d'une insertion avec ou sans *linker* sur la capacité de l'enveloppe à réaliser l'entrée. Le *linker* SGGG choisi est décrit comme permettant de garder une certaine flexibilité en ce qui concerne le repliement des deux protéines de fusion[225]. Dans le plasmide codant l'enveloppe D368R, le peptide AHNP a donc été introduit entre les acides aminés 404 et 405 situés dans la région V4 (figure 42A). La modélisation de L'enveloppe D368E contenant le motif AHNP avec ou sans *linker* a été réalisée et semble confirmer que l'ajout du *linker* permet une meilleure exposition du motif comparé à l'exposition en absence de séquence *linker* (figure 42B). Les vecteurs portant l'une ou l'autre des enveloppes ont été utilisés comme précédemment décrit pour réaliser

des tests d'entrée dont les résultats sont présentés dans la figure 43. L'insertion de la séquence *linker* permet de restaurer, par rapport à la protéine contenant le AHNP sans *linker*, un niveau d'entrée comparable à celui de la protéine ne contenant pas de motif AHNP (p=0,74).

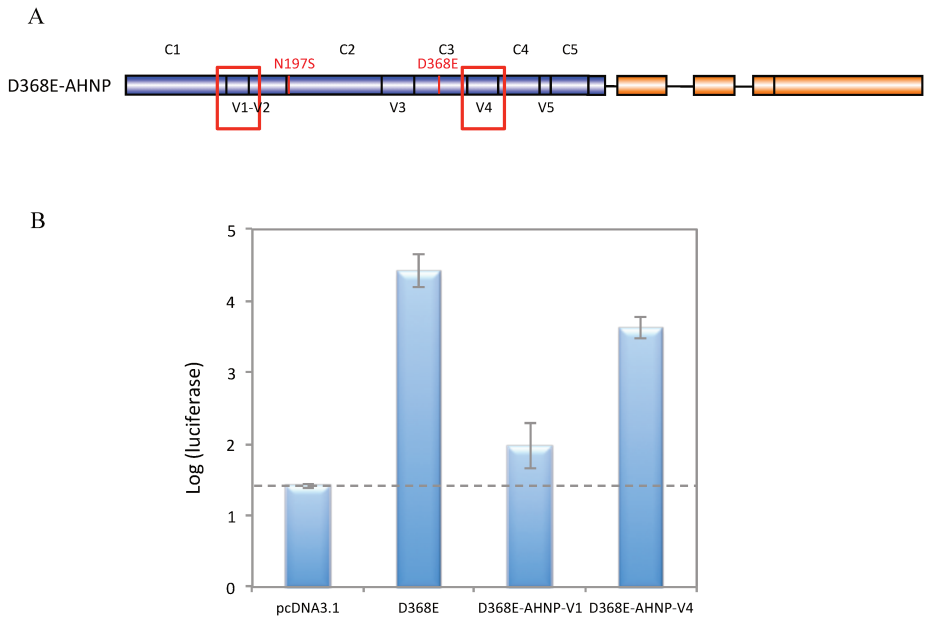
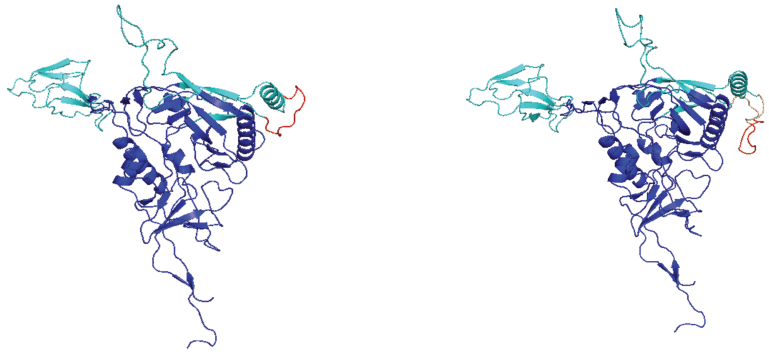



**Figure 41 : Détermination de la régions de l'enveloppe dans laquelle introduire le ligand.** A) Schéma représentatif des régions testées (encadré rouge) pour l'introduction du peptide ligand. B) Le graphique représente les résultats obtenus pour les tests d'entrée réalisés (n=3) avec les différentes enveloppes générées. Les enveloppes ont été testées pour leur capacité à entrer dans les cellules exprimant le récepteur et le corécepteur (CD4+R5+). Le pcDNA3.1 correspond aux vecteurs ne portant pas d'enveloppe virale et constitue le contrôle négatif de nos expériences. Le trait pointillé représente la limite d'expression de la luciférase en dessous de laquelle les résultats sont considérés comme étant négatifs. Il est défini par la valeur de luciférase dans les échantillons pcDNA3.1.

A)

| | |
|---|---|
| D368E | LTQSNG$^{404}$T$^{405}$EGND |
| D368E-AHNP | LTQSNG$^{404}$YCDGFYACYMDVT$^{405}$EGND |
| D368E-AHNP-Linker | LTQSNG$^{404}$SGGGGYCDGFYACYMDVGGGGST$^{405}$EGND |

B)



D368E-AHNP D368E-AHNP-Linker

**Figure 42 : Insertion du peptide AHNP dans la région V4 de la gp120.** A) Séquence protéique de la partie de la région V4 de la gp120 où a été effectué l'insertion du motif AHNP. La séquence de l'enveloppe est représentée en bleu, le peptide AHNP en rouge et le *linker* en beige. B) Les modèles de la protéine gp120 ont été obtenus à partir des séquences codantes en utilisant l'outil Swiss-Model. Sont représentées en bleu foncés les régions constantes, en bleu clair les régions variables et en vert la région gp41. En rouge est indiqué le peptide AHNP et en beige la séquence *linker*.

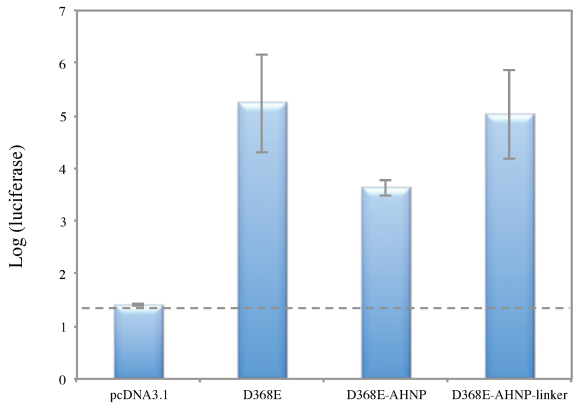



**Figure 43 : Effet du mode d'introduction du ligand sur la fonctionnalité de l'enveloppe.** Le graphique représente les résultats obtenus pour les tests d'entrée réalisés sur les cellules exprimant à la fois le récepteur CD4 et le corécepteur CCR5 (n=3). Le trait pointillé représente la limite d'expression de la luciférase en dessous de laquelle les résultats sont considérés comme négatifs, il est défini par la valeur maximale de luciférase dans les échantillons pcDNA3.1.

## IV Discussion

Pour les stratégies de thérapie génique nécessitant une intervention *in vivo*, la capacité des vecteurs à cibler spécifiquement les cellules cancéreuses est l'un des points clés. Au cours de ce travail de thèse, j'ai démarré les études préliminaires nécessaire pour faire la preuve de concept qu'une enveloppe VIH-1 modifiée peut permettre un ciblage spécifique des cellules cancéreuses exprimant à leur surface un marqueur tumoral. Ces études préliminaires nous ont permis dans un premier temps de générer une enveloppe CD4i, capable de réaliser l'entrée dans des cellules n'exprimant que le corécepteur R5. Nous avons ensuite déterminé que la région variable V4 de la protéine d'enveloppe gp120 est la plus adéquate pour insérer le ligand de la protéine HER2 utilisée comme modèle d'étude. Enfin, nous avons déterminé que l'insertion du peptide ligand avec une séquence *linker* SGGG, permet de ne pas affecter la capacité de l'enveloppe à réaliser l'entrée.

En ce qui concerne l'enveloppe CD4i générée, nous sommes partis des enveloppes WT des virus VIH-1 disponibles au laboratoire. Celles-ci nécessitent une première interaction avec le récepteur CD4, ce qui induit les changements de conformation nécessaires pour la liaison au corécepteur, à savoir, la délocalisation de la région V3 (figure 27) et le démasquage du site de liaison au corécepteur. Dans le cas des enveloppes CD4i cependant, elles sont présentes dans une conformation qui leur permet de lier directement le corécepteur et de poursuivre le processus d'entrée. En suivant les indications de la littérature, nous avons d'abord généré, à partir de l'enveloppe issue de la souche ADA du VIH-1, une enveloppe appelée N197S qui contenait la mutation ponctuelle en position N197S. Dans les tests que nous avons réalisés, cette enveloppe ne permettait cependant pas de s'affranchir de la liaison au CD4 pour réaliser l'entrée. Nous avons donc construit une deuxième enveloppe, appelée N197S-HxB2, qui, en plus de la mutation ponctuelle N197S, comportait la substitution de la gp41 par celle de l'enveloppe de la souche HxB2. Cette deuxième enveloppe (N197S-HxB2) a quant à elle montré une capacité à entrer dans les cellules en absence du récepteur CD4. Il semblerait donc que ce soit la substitution de la région gp41 de la souche ADA par celle issue de la souche HxB2 qui permet d'aboutir à la conformation de l'enveloppe requise pour une liaison directe au corécepteur R5. Cette hypothèse semble supportée par la modélisation des enveloppes chimériques réalisées à partir de leurs séquences codantes (figure 44). Sur la base de cette modélisation, on constate que l'enveloppe N197S présente une conformation

semblable à l'enveloppe ADA WT en ce qui concerne la localisation de la région V3 de la gp120. Par contre, dans l'enveloppe N197S-HxB2, on observe une délocalisation de la région V3 qui est semblable à celle observée dans les enveloppes WT une fois que celles-ci se sont liées au récepteur CD4 (figure 27). Cela laisse supposer que la divergence entre les protéines gp120 et gp41 constituant l'enveloppe chimérique conduit dans notre cas à une conformation permettant une interaction directe avec le corécepteur. Il est en effet décrit dans la littérature qu'en plus d'induire des changements de conformation dans la gp120 (avec notamment la délocalisation de la région V3), la liaison de la gp120 au CD4 fragilise également les interactions entre les protéines gp120 et gp41[220]. Ces interactions entre les deux protéines font intervenir les régions C1 et C5 de la gp120 et 17 résidus de la protéine gp41, et lorsque l'on compare les séquences protéiques des gp41 issues de la souche ADA ($gp41_{ADA}$) et de la souche HxB2 ($gp41_{HxB2}$), on constate que 20 résidus sont modifiés entre ces deux souches. Parmi ces 20 résidus, 9 correspondent aux résidus décrits comme impliqués dans l'interaction de la gp41 avec la gp120 ou sont adjacents à ceux-ci. Cette observation laisse envisager que dans l'enveloppe chimérique, du fait de la modification des 9 résidus impliqués dans l'interaction gp120-gp41, l'interaction entre la $gp120_{ADA}$ et la $gp41_{HxB2}$ pourrait être déstabilisée. Cette déstabilisation de l'interaction induirait alors des changements de conformation dans les deux protéines, permettant ainsi de générer une enveloppe présente dans une conformation intermédiaire, similaire à celle des enveloppes ayant lié le CD4. Cela expliquerait la capacité de notre enveloppe à réaliser l'entrée dans les cellules CD4-R5+.

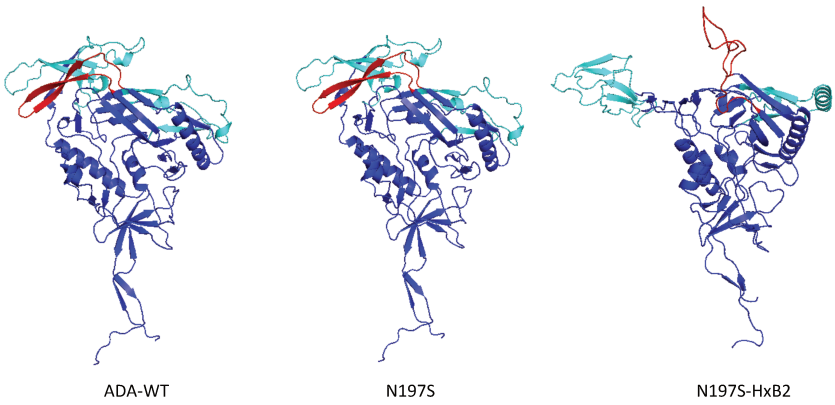



**Figure 44 : Position de la région V3 dans la protéine gp120.** La modélisation de la structure des différentes enveloppes a été obtenue avec l'outil Swiss-Model à partir des séquences codantes pour chacune des enveloppes. Sont représentées les régions variables (en bleu clair) les régions constantes (en bleu foncé) ainsi que la région V3 (en rouge).

Les précédentes études réalisées sur les enveloppes CD4i ont montré que ces enveloppes gardent tout de même une meilleure capacité à entrer dans les cellules exprimant le récepteur CD4 comparé aux cellules n'exprimant que le corécepteur[182,208]. Il n'est donc pas surprenant que malgré sa capacité à réaliser l'entrée dans les cellules R5, l'enveloppe N197S-HxB2 garde une meilleure capacité à entrer dans les cellules CD4+ (figure 38). Afin de réduire cette entrée résiduelle, nous avons introduit dans l'enveloppe N197S-HxB2, une deuxième mutation (D368E, elle aussi décrite dans la littérature), générant ainsi l'enveloppe qu'on appelle D368E, qui est celle, parmi les mutants que nous avons testés, qui montre les meilleurs résultats d'entrée dans les cellules CD4-. Cette enveloppe D368E garde tout de même une capacité d'entrée résiduelle dans les cellules CD4+ (figure 39). Cette capacité de l'enveloppe à entrer dans les cellules CD4+ est un problème pour son utilisation, car elle pourrait potentiellement conduire à des effets secondaires au niveau des cellules CD4+. Il est donc primordial que l'enveloppe modifiée perde au maximum sa capacité à entrer dans les cellules CD4+. Deux stratégies sont envisagées pour atteindre ce but : la mutagénèse dirigée et l'évolution en culture de cellules, et seront présentées dans la partie « Perspectives ».

En plus d'abolir le tropisme naturel de l'enveloppe, la réorientation de l'entrée de l'enveloppe nécessite également l'insertion dans celle-ci d'un peptide ligand capable de reconnaître spécifiquement le marqueur tumoral exprimé à la surface des cellules cibles. Afin de déterminer la région de l'enveloppe dans laquelle l'insertion du peptide peut se faire sans pour autant affecter la fonctionnalité de l'enveloppe, nous avons testé l'effet de cette insertion dans les régions variables V1 et V4 de la gp120. Nous avons ainsi pu déterminer que la région V4 est la plus adaptée pour l'insertion du ligand. La région V1/V2 est impliquée dans la reconnaissance du récepteur CD4[196,215] et certaines enveloppes de phénotype CD4i décrites ont été générées par la délétion de cette région. Cela suggérait que la délétion ou la modification de cette région n'aurait pas du affecter la capacité de l'enveloppe à réaliser l'entrée. Nos résultats montrent toutefois que l'insertion du motif AHNP dans cette région conduit à une baisse de la fonctionnalité de l'enveloppe (figure 41). La région V4 est quant à elle non résolue dans les structures publiées jusqu'aujourd'hui, ce qui suggère une grande flexibilité et donc probablement une capacité à être modifiée sans perturber l'entrée virale. Cela est notamment confirmé par les résultats que nous avons obtenus qui montrent que l'insertion du motif AHNP dans cette région n'impacte pas significativement la capacité de l'enveloppe à réaliser l'entrée (figure 41).

Lorsque l'on souhaite produire des protéines en fusion, il est courant d'utiliser des séquences permettant d'espacer les deux protéines afin que celles-ci ne soient pas mutuellement perturbées dans leur repliement. Ces séquences sont appelées des séquences *linker*. Ici, nous avons testé l'effet sur la fonctionnalité de l'enveloppe, de l'introduction du motif AHNP avec ou sans séquence *linker* de part et d'autre de la séquence du motif. Les résultats obtenus montrent que l'insertion de la séquence *linker* permet de ne pas affecter la fonctionnalité de l'enveloppe comparée à une insertion du peptide sans *linker* (figure 43).

Concernant la nature du ligand à insérer, nous avons sélectionné parmi les différents ligands décrits pour la protéine HER2, le ligand AHNP du fait de sa petite taille, de façon à limiter les perturbations éventuelles que son insertion aurait pu causer sur la fonctionnalité de l'enveloppe. Une étude récemment publiée montre cependant que l'insertion d'une séquence minimale de la GFP fonctionnelle (plus de 200 aa) dans la région V4 de la gp120 avec le même *linker* que nous avons utilisé, est possible[224]. Dans cette étude, les tests de fusion de membranes réalisés ont permis aux auteurs de déterminer que l'enveloppe gardait sa fonctionnalité au sein de la protéine de fusion, et que la GFP gardait également ses propriétés de fluorescence. Ajouté à nos résultats préliminaires, ces données nous permettent d'envisager un remplacement du motif AHNP par un autre ligand (de plus grande taille) qui pourrait présenter une meilleure capacité à interagir avec la protéine HER2. Il faut néanmoins garder à l'esprit que l'introduction de modifications dans l'enveloppe peut avoir un impact sur la stabilité des interactions entre la gp120 et la gp41, et conduire à un fort taux de *shedding* de l'enveloppe modifiée. En ce qui concerne l'étude précédemment citée par contre, les tests ont été effectués dans le contexte de fusion cellules-cellules, mais les enveloppes n'ont pas été produites sur des particules virales afin d'évaluer leur capacité à réaliser l'entrée. Nous n'avons donc aucune indication sur l'effet de l'introduction d'un peptide de si grande taille sur la stabilité des enveloppes produites. Quoi qu'il en soit, les résultats obtenus avec l'insertion du *linker* nous permettent d'envisager de tester des ligands de la protéine HER2+ plus grands ou nécessitant une plus grande flexibilité que le ligand AHNP. Un ligand qui serait notamment intéressant d'utiliser dans notre étude est le domaine variable (scFv) du Trastuzumab, un anticorps monoclonal utilisé dans le cadre du traitement du cancer du sein HER2+[247]. Cet anticorps présente une très forte affinité pour la protéine HER2 et serait idéal pour tester la spécificité de nos enveloppes modifiées pour les cellules HER2+. L'expression des scFv en fusion avec une autre protéine demande une très grande flexibilité pour le repliement du motif ainsi que pour la fixation à la protéine

cible, et cette flexibilité peut être apportée par l'insertion de la séquence *linker*. De plus, la taille de cette séquence peut être augmentée afin d'atteindre la flexibilité nécessaire au bon fonctionnement des protéines[225].

## Partie VI : Conclusions et perspectives

La thérapie génique est une stratégie visant à utiliser les acides nucléiques (ADN ou ARN) afin de corriger dans les cellules le défaut génétique à l'origine de la pathologie, ou afin d'éliminer les cellules comme par exemple dans le cadre des stratégies contre le cancer.

Au cours de la première partie de mon travail de thèse j'ai principalement effectué la caractérisation du M36 (un mutant de la dCK humaine) qui est capable de sensibiliser certaines lignées cellulaires cancéreuses à de faibles concentrations de Gem et d'AraC. L'efficacité du M36 à sensibiliser des cellules résistantes au traitement a notamment été montrée sur les cellules Messa 10K. La lignée Messa 10K a été adaptée en laboratoire afin d'être résistantes à la Gem et possède une dCK inactive. Elle constitue un bon modèle d'étude, car elle permet de reproduire l'apparition de cellules résistantes qui est observée au cours des traitements de chimiothérapie. Il arrive en effet qu'au cours du traitement, certaines cellules deviennent résistantes au médicament, et que cette résistance soit liée à une inactivation de la dCK. Les résultats de sensibilisation obtenus sur les cellules Messa 10K permettent ainsi d'envisager une utilisation du M36 pour le traitement et l'élimination des cellules résistantes pouvant apparaître au cours du traitement. Il faut toutefois garder à l'esprit que les cellules Messa 10k ont été produites en laboratoire, il serait donc intéressant d'évaluer l'efficacité du M36 sur des cellules tumorales issues d'explants de patients. La résistance des cellules aux traitement pouvant être due à différents facteurs (inactivation du médicament par les désaminases, faible expression de la dCK ou expression d'une dCK inactive, surexpression de la RRM1), il faudra que les tests soient réalisés sur des cellules dont la résistance est liée à un défaut au niveau de la dCK.

La deuxième application envisageable pour le M36 est son utilisation comme gène suicide de sécurité pour la thérapie génique. Dans le cadre de cette application, les cellules souches présentant le défaut génétique à soigner seront corrigées pour ce défaut et, en parallèle, il leur sera introduit le transgène de sécurité (M36). Une fois la correction effectuée, les cellules seront réimplantées chez le patient. Suite à l'implantation des cellules, un traitement avec la Gem ou l'AraC pourrait permettre la déplétion des cellules modifiées dans le cas où celles-ci présenteraient une prolifération non contrôlée. Une telle utilisation du M36 nécessite par contre de confirmer au préalable son efficacité sur des cellules plus

pertinente d'un point de vue biologique, comme par exemple les cellules mononuclées sanguine périphérique.

Parallèlement à ces tests, il sera intéressant d'approfondir la caractérisation moléculaire du M36. Les résultats préliminaires que nous avons obtenus avec la comparaison entre les structures de la dCK en présence de dC et de Gem et celles du M36 en présence des mêmes substrats, ne nous a pas permis de comprendre d'un point de vue moléculaire l'implication des mutations introduites, sur l'activité de la protéine. Certaines pistes sont toutefois apparues, avec notamment la délocalisation éventuelle du site accepteur de phosphate de la dC dans le M36. L'obtention d'une structure à plus haute résolution est une perspective essentielle pour pouvoir effectuer une analyse plus concluante à ce sujet. La structure de la dCK est relativement difficile à obtenir du fait de la faible solubilité de la protéine, mais il est décrit dans la littérature, que l'introduction de 4 mutations (C9S, C45S, C59S et C146S) permet de la stabiliser en solution et de faciliter sa cristallisation. Il est ainsi envisageable d'introduire ces 4 mutations ponctuelles dans la séquence de la dCK et du M36. Une fois les mutations introduites, il faudra vérifier avec des tests de sensibilité au médicament en culture cellulaire et par des tests biochimiques, qu'elles n'impactent pas les caractéristiques de sensibilisation des cellules conférées par le M36 et par la dCK sauvage, avant de procéder aux tentatives de cristallisation.

Un autre point intéressant à développer dans le contexte de la caractérisation moléculaire du M36 concerne le fait que, dans la structure que nous avions précédemment déterminée, la protéine contenait au niveau du site de fixation du donneur de phosphate une molécule d'ADP. La résolution de la structure de la protéine avec un analogue non hydrolysable de l'ATP tel que l'AMP-PNP, nous permettrait d'avoir une vision plus proche du contexte physiologique en ce qui concerne le positionnement des différentes molécules (donneur de phosphate et substrat) au sein de la protéine. Une telle structure serait particulièrement intéressante pour confirmer la délocalisation précédemment observée, et permettrait également de mieux évaluer si l'éloignement qui résulte de cette délocalisation est significatif ou non. Ce qui, le cas échéant, expliquerait la baisse de la capacité du M36 à phosphoryler la dC.

Au cours de la seconde partie de mon travail de thèse, j'ai réalisé les études préliminaires pour faire la preuve de concept qu'une enveloppe VIH modifiée peut être utilisée pour un ciblage spécifique des cellules d'intérêt (notamment certaines cellules tumorales). L'enveloppe CD4i que j'ai réussi à développer pendant ma thèse, malgré sa capacité à entrer dans les cellules en absence de CD4, garde tout de même une meilleure capacité à entrer dans les cellules qui expriment le récepteur CD4. Ainsi, une perspective prioritaire est d'améliorer le phénotype CD4i de notre enveloppe. Il est envisageable, dans un premier temps d'introduire une nouvelle mutation ponctuelle en position 427 et de déterminer par des tests d'entrée l'effet de cette mutation supplémentaire sur la capacité d'entrée dans les cellules CD4+ par rapport aux CD4-. En effet, en plus de la mutation en position 368 introduite dans notre enveloppe, deux autres mutations présentant une bonne efficacité à altérer l'interaction de l'enveloppe avec le CD4 sont décrites dans la littérature. Il s'agit des mutations au niveau des résidus 370 et 427 de la gp120[208,209]. Nous avons déjà déterminé que l'introduction d'une mutation en position 370 de l'enveloppe abolit complètement l'infection des cellules (figure 37), ce qui va à l'encontre du phénotype que nous souhaitons obtenir. Il serait maintenant intéressant de tester l'effet de l'introduction d'une mutation supplémentaire en position 427 sur l'efficacité de l'enveloppe D368E à entrer dans les cellules CD4+ et dans les cellules CD4-. La présence simultanée des mutations en position 368 et 427 pourrait permettre de réduire la capacité d'entrée de notre enveloppe dans les cellules CD4+. En complément de cette approche, une procédure d'évolution de l'enveloppe en culture de cellules, comme indiqué plus loin, pourra être utilisée pour optimiser cette caractéristique de la protéine.

Pour cette preuve de concept, nous avons choisi comme marqueur tumoral modèle la protéine HER2 qui est surexprimé dans le cas de certains cancers et notamment dans certains types de cancer du sein. Cette protéine a été largement étudiée et différents ligands qui lui sont spécifiques sont décrit dans la littérature, ce qui fait d'elle un bon modèle pour notre étude. Comme cellules à utiliser pour ces expériences, deux types de cellules sont prévus. Pour les expériences de routine nécessitant que les cellules soient transfectées, l'on utilisera les cellules HEK 293T (qui sont facilement transfectables) exprimant la protéine HER2 de façon stable (après transduction avec un vecteur lentiviral exprimant HER2) ou transitoire (après transfection, ce qui porterait à un niveau d'expression plus élevé que par transduction). En revanche, pour les expériences portant sur l'évaluation de l'entrée

spécifique des vecteurs générés, on pourra utiliser des lignées cellulaires issues de cancer du sein HER2+, comme par exemple la lignée SKBR-3, qui surexpriment donc la protéine à des niveaux physiologiquement pertinents.

Une étape préliminaire aux tests d'entrée qu'il faudra optimiser est le choix du ligand de HER2 à insérer dans l'enveloppe. Dans les premiers tests que j'ai réalisés, j'ai utilisé le ligand AHNP du fait de sa petite taille, afin de ne pas trop perturber la fonctionnalité de l'enveloppe. Ces tests, de mêmes qu'un article récemment publié, semblent indiquer que l'addition de séquence *linker* de part et d'autre de la séquence du ligand permet d'envisager l'utilisation de peptides plus grand et présentant une meilleure affinité pour le HER2. Les différents peptides pouvant être testés sont présenté dans la figure 45.

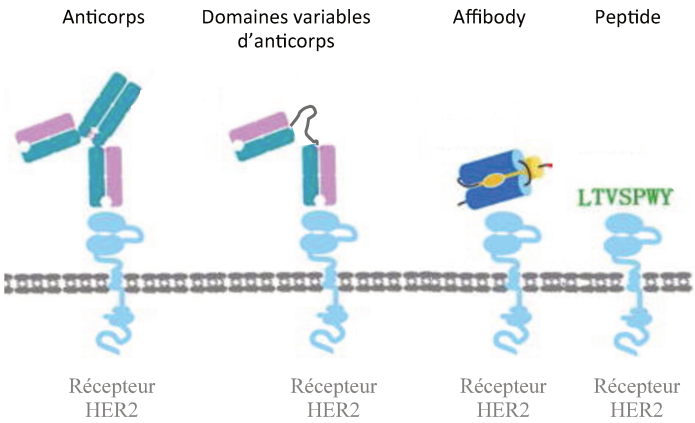



**Figure 45 : Nature de différents ligand capables de lier efficacement la protéine HER2[3].**

La surexpression de la protéine HER2 dans les cellules cancéreuses est associée à une surexpression de la protéine CXCR4 (X4), un des corécepteurs du VIH-1. Dans notre étude cependant, nous avons utilisé une enveloppe de tropisme CCR5 (R5), car les modifications dans ces souches sont celles qui sont le mieux décrites dans la littérature. Une étape préliminaire avant le test de spécificité sera donc de convertir l'enveloppe afin qu'elle utilise le corécepteur X4 au lieu du R5 pour l'entrée. La région V3 de la g120 est celle qui majoritairement détermine le corécepteur utilisé pour l'entrée et il est décrit que la substitution de la région V3 d'une souche de tropisme R5 par celle d'une souche de tropisme X4 permet de convertir la souche R5 en souche X4 tropique. Une deuxième méthode, moins drastique consisterait à introduire deux mutations ponctuelles dans la région V3. Il existe en

fait différentes méthodes de prédiction de l'utilisation des corécepteurs et l'une d'elle est notamment basée sur la nature des acides aminés présents en position 11 et 25 de la région V3, et sur la charge globale de cette région[200]. Cette méthode de prédiction considère que si en position 11 ou 25 de la région V3 il se trouve un acide aminé de charge positive alors l'enveloppe sera de tropisme X4. Lorsque l'on compare la séquence protéique de la région V3 des souches ADA (enveloppe de départ de notre étude, R5 tropique) et HxB2 (X4 tropique), on constate que la méthode de prédiction est validée. Ainsi pour changer le tropisme de l'enveloppe CD4i que nous avons générée nous avons aussi la possibilité d'introduire simplement des modifications ponctuelles dans la région V3 de notre enveloppe (comme montré dans la figure 46).

| | |
|---|---|
| HxB2-X4 | $C^{1}$TRPNNNTRKR$^{11}$IRIQRGPGRAFVTI$^{25}$GKIGNMRQAH$^{35}$ |
| ADA-R5 | $C^{1}$TRPNNNTRKS$^{11}$IHIGPGRAFYTTGD$^{25}$IIGDIRQAH$^{34}$ |
| ADA-X4 | $C^{1}$TRPNNNTRKR$^{11}$IHIGPGRAFYTTGK$^{25}$IIGDIRQAH$^{34}$ |

**Figure 46 : Séquence protéique de la région V3 des souches ADA et HxB2 du VIH-1.**

Une fois que, sur une base rationnelle, on sera parvenu à un stade de l'étude permettant d'atteindre le maximum de critères souhaités, on pourra envisager de soumettre l'enveloppe à la rétrovolution afin d'améliorer ces critères sur une base aléatoire et, enfin, identifier un mutant d'intérêt pour lequel l'entrée dans les cellules CD4+ serait très fortement diminué, avec éventuellement une augmentation de sa capacité d'entrée dans les cellules HER2+X4+. Nous espérons en particulier:

1. diminuer encore un peu la capacité d'entrée de l'enveloppe dans les cellules CD4+ via l'introduction de la mutation supplémentaire en position 427
2. convertir le tropisme de l'enveloppe de sorte qu'elle utilise le corécepteur X4 pour l'entrée
3. greffer sur l'enveloppe un meilleur ligand pour la protéine HER2 sans perturber sa fonctionnalité

Pour exploiter la rétrovolution, il s'agira de réaliser le passage des vecteurs portant l'enveloppe générée sur des cellules HEK 293T HER2+ X4+ afin de sélectionner les mutants

dont la capacité d'entrée dans ces cellules sera améliorée. La rétrovolution étant une procédure qui repose sur la mutagénèse virale, elle demande un certain taux de réplication virale pour pouvoir avoir lieu. Dans un premier temps donc, il faudra que les passages soient réalisés sur un mélange de cellules CD4+X4+ et HER2+X4+ afin de garantir un certain titre viral de départ. Au cours du processus, nous diminuerons la quantité de cellules CD4+X4+ afin de ne sélectionner que les enveloppes capables de rentrer dans les cellules sans une liaison préalable au CD4. Et enfin, suite à la sélection des mutants générés, la comparaison entre leur efficacité à entrer dans les cellules CD4+X4+ et celle à entrer dans les cellules HER2+X4+, nous permettra d'identifier que les enveloppes entrant préférentiellement dans les cellules cibles. Cette méthode d'adaptation utilisant des proportions variables de cellules a déjà fait ces preuves auparavant pour convertir des enveloppes WT en enveloppes CD4i et pour adapter des isolats du VIH-1 à la réplication dans les cellules exprimant la protéine TRIM5α du macaque rhésus[206,248].

Enfin, il est décrit que l'introduction de modifications dans l'enveloppe virale peut conduire à un *shedding* de l'enveloppe, à savoir la perte de la protéine gp120 au niveau des trimères, conduisant ainsi à une diminution de la capacité d'entrée. Sachant qu'il est possible de produire des particules virales possédant deux types d'enveloppes différentes à leur surface, par co-transfection de celles-ci, et qu'il semble que la présence d'une seule protéine fonctionnelle dans le trimère suffit pour permettre l'entrée, on peut envisager de contourner le problème de *shedding* s'il se présente. Pour ce faire, on pourra utiliser simultanément deux types d'enveloppe : une enveloppe modifiée pour accrocher le marqueur tumoral efficacement mais non fonctionnelle pour l'entrée (car elle contiendra des mutations afin que le précurseur gp160 ne soit pas clivé pour éviter tout problème de *shedding*) et une enveloppe CD4i. La co-transfection de ces deux enveloppes devrait générer une majorité d'enveloppes mixtes à la surface des vecteurs. Au sein de ces enveloppes mixtes, l'accrochage efficace des particules à la surface des cellules sera renforcé principalement grâce à l'enveloppe gp160, ce qui facilitera la fusion des membranes (à savoir l'entrée) assuré par l'enveloppe CD4i générée.

## Partie VII : Suivi des cellules souches embryonnaires dans le modèle du poisson zèbre

En marge des sujets principaux de ma thèse, j'ai participé à une collaboration avec l'équipe de Manuela Tavian à l'Inserm UMR-S949 de Strasbourg et celle de Karima Kissa à l'UMR 5235 (Université de Montpellier) en vue d'étudier le devenir des cellules souches hématopoïétiques humaines (CSH) CD34+ dans le modèle de l'embryon du poisson zèbre. Ces cellules sont d'un grand intérêt pour la recherche clinique et notamment dans le domaine de la thérapie cellulaire et génique. En effet, du fait de leur multipotence, elles sont à l'origine de toutes les cellules sanguines, et de leur capacité d'auto-renouvèlement, constituant ainsi un outil essentiel pour la mise en place de thérapies dans le traitement des hémopathies (leucémies et maladies génétiques du système sanguin).

Le poisson zèbre est un modèle très prometteur pour l'étude de différentes maladies humaines et notamment pour l'étude du processus d'hématopoïèse et des maladies associées.
Ce modèle suscite l'intérêt en partie du fait de son génome, qui contient des homologues d'environ 70% des gènes humains et 80% de ceux qui sont impliqués dans des maladies, Ce grand intérêt est également dû au fait que, comparé à la souris qui est le modèle fréquemment utilisé, les embryons de poisson-zèbre (et potentiellement les adultes) sont optiquement transparents, permettant ainsi un suivi aisé par imagerie directe des cellules transplantées lorsqu'elles sont marquée avec des protéines fluorescentes. De nombreuses lignées de cellules de leucémie humaine ainsi que des cellules leucémiques directement issues des patients ont été xénogreffées avec succès dans le poisson zèbre.
Toutefois la xénogreffe chez le poisson zèbre de CSH humaine saines est très peu documentée et semble présenter des obstacles majeurs, à l'heure actuelle non identifiés.

Afin d'effectuer cette étude, il a fallu dans un premier temps optimiser les différentes étapes : le tri des cellules, leur marquage et leur transplantation dans le poisson. Au sein de cette collaboration, mon rôle a été d'optimiser la méthode de transduction des cellules afin de réduire au maximum les temps nécessaires pour une transduction efficace par des vecteurs lentiviraux. Ces vecteurs portent la séquence codante la GFP (green fluorescent

protein) et permettent de rendre les CSH vertes en vue de leur suivi après transplantation chez le poisson.

Dans les protocoles classiques, les temps de transduction des cellules humaines peuvent aller jusqu'à deux jours de contact entre les cellules et les vecteurs, alors que dans le cadre de l'étude des cellules souches, le temps est un facteur important. En effet, il est crucial de limiter le temps pendant lequel les CSH sont maintenues en culture, car cette étape peut conduire les cellules à se différencier et ainsi perdre leurs capacités de multipotence et d'auto-renouvellement.

Dans ce but, j'ai testé différentes conditions de transduction en modifiant les quantités de vecteurs utilisées, la durée de la transduction ainsi que la méthode de transduction. J'ai ainsi pu développer un nouveau protocole de transduction de 2 h, tout en conservant des taux de transduction comparables à ceux qui sont généralement obtenus en suivant les protocoles classiques. La méthode que nous avons mise en place et qui englobe toute les étapes de l'études à savoir, le tri cellulaire, le protocole de transduction ainsi que la méthode de transplantation, fait actuellement l'objet d'un article en cours de rédaction qui est présenté en annexe 2 de ce manuscrit. Dans cet article, nous montrons dans une première partie que les progéniteurs hématopoïétiques humains disparaissent très rapidement après la transplantation dans le poisson. Grâce à l'utilisation des poissons transgéniques nous montrons dans une deuxième partie que la cause de leur élimination est liée à la présence des macrophages du poisson. En effet, lorsque les macrophages sont éliminés suite à des modifications génétiques (utilisation d'oligonucléotides antisens = morpholino) ou à un traitement chimique (utilisation d'un composé phagocyté par les macrophages et conduisant à la mort cellulaire), cela permet le maintien des cellules humaines transplantées.

# MACROPHAGE DEPLETION ALLOWS HUMAN HEMATOPOIETIC CELL ENGRAFTMENT IN ZEBRAFISH EMBRYOS

Reine El Omar[1], Naoill Abdellaoui[1], Safiatou Coulibaly[2], Laura Fontenille [4], Matteo Negroni[2], Karima Kissa[3], Manuela Tavian[1]

[1]Université de Strasbourg, INSERM BPPS UMR-S949, EFS Grand Est, F-67065 France
[2]Université de Strasbourg, CNRS UPR 9002, IBMC, F-67000 France
[3]Université de Montpellier, CNRS UMR 5235, DIMNP, F-34095 France
[4] AZELEAD SA,' Université de Montpellier, DIMNP, F-34095 France

## INTRODUCTION

Hematopoietic stem cells (HSCs) are self-renewing multipotent blood cells capable of producing all blood cell types during the entire life of an individual. Clinically, these cells are the relevant component of bone marrow transplants, which are used to treat patients with blood disorders and malignant diseases (Kumar and Geiger, 2017). Successful transplantation requires the ability of HSC to migrate toward suitable niches as well as to identify signals that promote their self-renewal and differentiation (Crane et al., 2017). A number of important experimental breakthroughs underlie the success of HSC biology. Data on HSC activity and potentialities have been collected mainly through in vitro assays or by tracking in vivo the fate of transplanted HSC in various wild-type and humanized animal models (Babovic and Eaves, 2014; Benveniste et al., 2010; Biasco et al., 2016; Cheung et al., 2013; North et al., 2002; Notta et al., 2011; Wright et al., 2001). Historically, mouse has always been viewed as the "gold standard" for xenotransplantation experiments. Nonetheless, this model has major limitations. For example, it is complicated to follow-up the initial steps of engraftment, self-renewal and differentiation of transplanted cells in mouse, even by using highly sensitive bioluminescent techniques (Gansner et al., 2017). The monitoring of mice after transplantation is lengthy, requires a relatively big number of cells and high costs. Moreover, mouse xenograft experiments need the use of immunocompromised animals, since the human cells xenografts would be rejected by the mouse immune system (Wertman et al., 2016). This may mask pathways of tolerance or rejection that exist in humans (Chong et al., 2013).

The importance of zebrafish (Danio rerio) as a research organism has grown rapidly over the last two decades (Collin and Martin, 2017). Its genome, fully sequenced in 2013, contains homologues of around 70% of human genes, and 80% of those are involved in diseases (Howe et al., 2013). Moreover, this model has many other characteristics like the short generation time (approximately three months), the high fecundity (several hundred eggs per male per week), an external embryogenesis and a fast development stages. Zebrafish embryos (and potentially adults for the "casper" zebrafish strain) are optically transparent allowing direct noninvasive live imaging with much greater sensitivity reaching the single-cell level, something generally not achievable in mouse (Gansner et al., 2017). Furthermore, zebrafish is known for its ease of handling and genetic manipulation, and the availability of a highly diversified transgenic lines (Collin and Martin, 2017; Iwanami et al., 2017). These attributes and many others have established zebrafish as a powerful *in vivo* tool for investigating hematopoiesis- where there is strong conservation of key lineages, genes and developmental stages with humans- and modeling human hematopoietic disorders (Rasighaemi et al., 2015). Many human leukemia cell lines (K562, Jurkat, NB4) (Bentley et al., 2015; Corkery et al., 2011; Pruvot et al., 2011; Zhang et al., 2016) and primary leukemia cells form patients (Bentley et al., 2015; Pruvot et al., 2011) have been successfully xenografted into zebrafish. Whereas, engraftment of healthy human hematopoietic stem cells (HSC), in order to investigate their potentialities, is very poorly documented and appears to be more complicated. Recently, Pruvot et al. showed that CD34+ hematopoietic cells derived from human cord blood rapidly disappear after transplantation in zebrafish embryo (Pruvot et al., 2011).

In this study, we wanted to "humanize" the zebrafish embryo by engrafting human CD34+ hematopoietic stem/progenitor cells derived from umbilical cord blood, and especially to understand the reason for the lack of data reporting human healthy and primary hematopoietic xenografts into zebrafish. First, to visualize transplanted CD34+ cells in zebrafish embryos, we have developed a rapid method of lentiviral transduction. We then attempted to engraft transduced cells into zebrafish embryos and tried to decipher the reason of engraftment failure. Our results showed that the hematopoietic engraftment success depends on the depletion of the zebrafish embryo macrophages.

## MATERIAL AND METHODS

### Umbilical cord blood CD34+ cells sorting

Umbilical cord blood (UCB) units were collected from full-term deliveries after receiving a written consent, according t the guidelines of the French National Ethics Committee. UCB was processed within 24 hours after reception. Mononuclear cells (MNC) were isolated by Ficoll-Histopaque (1,077g/ml) (10771-500ML, Sigma-Aldrich) density centrifugation. Low-density cells were washed twice in Dulbecco's Phosphate-Buffered Saline (PBS, 14190-094, Gibco) supplemented with 2% fetal bovine serum (FBS) (P30-3302, Pan Biotech) and then processed for CD34+ cell sorting. Briefly, MNC were resuspended in PBS containing 2% FBS and incubated with FITC mouse anti-human CD45 (Clone 5B1, 130-080-202, Miltenyi) and APC mouse anti-human CD34 (clone AC136, 130-090-954, Miltenyi) for 30 min at 4°C. Cells were then washed with 2 ml of PBS 2% FBS and resuspended in a solution of 7-Aminoactinomycin (7-AAD, 2µg/mL) (A9400, Sigma-Aldrich). Fluorescence-activated cell sorting (FACS) was performed on a BD Aria II flow cytometer. Sorted CD34+ cells were reanalyzed to establish purity, and it was not less than 95%. Isolated CD34+ cells were incubated overnight in serum-free expansion medium (SFEM) consisting of StemSpan (09650, StemCell Technologies) containing recombinant human Stem Cell Factor (SCF, 50 ng/ml, 255-SC-050, R&D Systems), recombinant human Thrombopoietin (TPO, 50 ng/ml, 288-TP-025, R&D Systems) and recombinant human Flt-3 Ligand (Flt-3l, 308-FK-025, R&D Systems), and then transduced by lentiviral vector.

### Cell Lines

Cell lines were obtained from the American Type Culture Collection (ATCC). HEK 293T were cultured in Dulbecco's Modified Eagle's Medium (Gibco, thermo Fischer scientific, Waltham, MA, USA) supplemented with 10% FCS and 1% penicillin/streptomycin. Jurkat cells were cultured in Roswell Park Memorial Institute medium (Gibco, thermo Fischer scientific, Waltham, MA, USA) supplemented with 10% FCS and 1% penicillin/streptomycin. Cells were maintained at 37°C with 5% $CO_2$.

### Construction of pSDY-nGFP

pSDY-nGFP was obtained by replacing, in the pSDY-dCK-Puro plasmid (Rossolillo et al PLoS Genetics 2012) the sequence encoding the human deoxycytidine kinase by a cassette containing the eGFP preceded by the nuclear localization signal of the histone H2B. In detail, the eGFP coding sequence contains (in the 5' to 3' orientation) the sequence from nt 75 to 453 of the homo sapiens cDNA, FLJ9211 (gene bank AK311849.1) followed by the sequence encoding the linker (in aminoacid sequence) GSGGGGSGGGGSGGGGSGGGGLACRNSISSLSDP and by the eGFP coding sequence.

### Viral vector production and titer

For lentiviral vector production, 5x106 HEK 293T cells were plated in 10 cm diameter plates and, when at 80% of confluence, were transfected with a mix containing 2ug of pCMVΔR8.91 (Naldini L, et al. Science. 1996; 272:263–7), 1ug pHCMV-G (Yee JK et al, PNAS, 1994; 91:9564–8), 2ug of pSDY-nGFP, Opti-MEM (Gibco, thermo Fischer scientific, Waltham, MA, USA) and polyethylenimine (Polysciences, INC, Warmington, PA, USA). Vectors were collected 48h after transfection and filtered on 0.45um filters before concentration with Vivaspin, 1,000,000 MWXCO PES (Sartorius Stedium Biotech). Viral vector titer was evaluated by transduction of HEK 293T cells with different volumes of vector preparation, estimating the proportion of GFP positive cells estimated by visual inspection of four independent pictures of the cell culture 48h after transduction.

**Stem cells transduction**

Transductions were carried out on variable amounts cells (depending on their availability) and the corresponding appropriate number of viral vectors to reach the MOI indicated in each case, in a final volume of 1 mL of cell culture medium in the presence of hexadimethrine bromide (polybrene) at a final concentration of 8 ug/ml, by spinning at 800g for 2h at 32°C (spin inoculation). After transduction, the medium was removed the cells resuspended in fresh medium and transferred in 24-well plates at 37°C with 5% CO2.

**Zebrafish transgenic lines**

Wild-type AB and *Tg(Mpeg1:mCherry-f)* zebrafish (Ellett et al., 2011) were maintained in compliance with the French Institutional Animal Care (CEEA-LR-13007). Embryos were kept in the presence of 1-phenyl-2-thiourea (PTU) to prevent melanin pigmentation (Westerfield 2000) and staged as described by Kimmel and collaborators (Kimmel et al., 1995)

**Cell transplantation**

Transduced Jurkat cells were transplanted in the caudal vein of a Wild-type AB and Tg(mpeg1:mCherry-f) zebrafish embryo at 24 hours post-fertilization (h.p.f) stage using a CellTrame micromanipulator (Eppendorf).

**Macrophage chemical and genetic depletion**

Macrophages were depleted by injecting 5 nl of L-clodronate into the caudal vein as described previously (Travnickova et al., 2015). Embryos with L-clodronate aggregates causing vessel occlusions and shortcuts were discarded.
Specific PU1 antisense targeting morpholino (MO, Gene Tools) was resuspended in morpholino buffer – 12 mM KCl and 20 mM HEPES – in a 2 mM stock concentration. The MO was injected at 1-4 cell stage in zebrafish embryos using the micro-injector system (Tritech Research Inc.).

**Microscopy**

Embryos were anesthetized with tricaine (0.016%) and mounted on a glass cover dish with 0.7 % low melting agarose (lateral views, rostral to the left). Fluorescence and Time-lapse images were acquired using Zeiss LSM510 at 20X or 40X magnification (Fig. 2i, 3a, 3h, and Supplementary Movie 2, 3 and 4) or Zeiss AxioImager (Figure 2 (a-h), 3b-g, 4 5a-b, 5d-e and Supplementary Movies. 1). Temperature was maintained at 28°C by placing the dish in a temperature-control chamber during time-lapse acquisitions

## RESULTS

**Lentivirus transduction and spin inoculation: Transduction efficiency in human hematopoietic progenitor cells and Jurkat cells**

Human CD34+ cells were cultured with cytokines for 24h before transduction and then transduced using an HIV1-based lentiviral vector that expresses a nuclear version of the enhanced green fluorescent protein (nGFP), as described in the Methods section. Transduction was carried out using the spin inoculation method described above (mild centrifugation for 2 hours at 800g) (Figure 1a). The Jurkat cell line was used as a positive control for transduction. Two different multiplicity of infection (MOI) were tested: 30 and 120. A MOI of 30 yielded slightly higher transduction rates than the MOI of 120 with efficiencies of 33.8 ± 3.5 % in cord blood CD34+ cells and of 77.2% ± 4.8 % in Jurkat cells. More interestingly, an MOI of 30 maintained significantly higher CD34 expression rates (determined at 5 days after transduction of CD34+ cells) (92.2 ± 1.8, at MOI 30 vs 83.1 ± 3.0 a MOI 120, $p<0.05$) (Figure 1c). Therefore, overall, our data show that a low MOI preserves a stable stem/progenitor phenotype of human cord blood CD34+ cells. Furthermore, Figure 1c shows that CD34+ and Jurkat cells have normal morphology after lentiviral transduction by spin inoculation under our conditions, and that transduced cells express an important level of GFP as also confirmed by flow cytometry (Figure1e)

**Fates of transplanted cells into the zebrafish embryo**

1. *Functional humanization with a proliferation of human cells and organs colonization.*

Transduced Jurkat cells were transplanted in the caudal vein of a wildtype (WT) zebrafish embryo line at 36 hours post-fertilization (h.p.f) stage. The distribution of Jurkat cells in zebrafish embryos was checked immediately after injection and then followed over time. Within zebrafish embryos, transplanted human cells displayed two different behaviors.

The first case was rare. In the transplanted embryos, right after injection, cells are mostly localized at the injection site, but some cells were also found distributed in the caudal hematopoietic tissue (CHT) (Figure 2a). At one-day post-transplantation (d.p.t), Jurkat cells were found mainly in the CHT (Figure 2b). From 2 d.p.t to 7 d.p.t, live imaging microscopy revealed that fluorescent cells circulate actively in the vascular system of the zebrafish embryo, extravagate the blood vessels (Figure 2c and 2d, supplementary movie 1) and colonize the different hematopoietic organs: CHT (Figure 2e), thymus (Figure 2f), fetal liver (Figure 2g), kidney (Figure 2h). These results suggest that the transplanted human Jurkat cells can survive in the zebrafish embryo.

Additionally, in many cases, we observed an increase in the size of clusters of transplanted cells within the zebrafish embryos. To determine whether this size increase was due to a cell proliferation or cells spreading, we tried to examine the behavior of a single cell by live cell imaging. As shown in the Figure 2i (Supplementary Movie 2), directly after injection, we can detect a cell undergoing mitosis and giving two daughter cells in almost 20 minutes after transplantation. Together, these observations reveal a possible humanization of zebrafish embryo and that the zebrafish hematopoietic environment may provide efficient signals for the survival, proliferation and homing of hematopoietic cells.

2. *Engraftment failure and transplanted cell disappearance*

The major situation observed was the decrease in the number of transplanted human cells (CD34+ and Jurakt) in the few hours following the injection and the complete disappearance of CD34+ and Jurkat cells few hours and one-day post-transplantation respectively. In a remarkable way, cells seem to be lysed from the half hour following transplantation (Figure 3a). Closer examination of human cells

in the embryos using live confocal microscopy showed that cells are fragmented into small vesicles before they disappear (Supplementary Movie 3). These vesicles are suspected to be phagolysosomes and we speculated that primitive macrophages are responsible for the clearance of the transplanted cells.

**Phagocytosis by zebrafish embryonic macrophages at the origin of transplantation failure**

Zebrafish embryos, like other vertebrate embryos, have early macrophages that appear as early as 20 h.p.f, differentiate in the yolk sac before other leukocytes arise and then migrate into the mesenchyme and developing organs and vasculature (Herbomel et al., 1999, 2001). Zebrafish primitive embryonic macrophages display an important phagocytic activity towards pathogens and can mount efficiently an innate immune response that is similar to the responses in mammalian systems (Davis et al., 2002; Herbomel et al., 1999; O'Toole et al., 2004; Prajsnar et al., 2008; van der Sar et al., 2006). In order to verify if the primitive embryonic macrophages were responsible for the cell lysis and thus the transplantation failure, human Jurkat cells were injected in a Tg(mpeg1:mCherry) zebrafish embryos line (Ellett et al., 2011) in which the Mpeg1 promoter drives the specific expression of membrane targeted red fluorescent protein in macrophages allowing visualization of macrophages behavior *in vivo*. First, we analyzed the distribution pattern of macrophages in a 48 h.p.f *Tg(Mpeg1:mCherry)* zebrafish embryo not injected (Figure 3b, left panel) and another one injected (Figure 3b,right panel) with Jurlat cells. In vivo imaging showed an important inflammation with a recruitment of macrophages at the injection site and a considerable proliferation in the injected embryos comparing to the non-injected control embryo (Figure 3b). A simultaneous observation of the red macrophages and the green Jurkat cells showed a remarkable co-localization of macrophages and injected fluorescent cells in zebrafish embryos at 2 d.p.t (Figure 3c). As shown in Figure 3d and Supplementary Video 4, a time-lapse confocal imaging between 0 hours post-transplantation (h.p.t) and 15 h.p.t confirmed our hypothesis and showed that embryonic primitive zebrafish macrophages (shown in red) are active and able to phagocyte efficiently Jurkat cells (shown in green). The same behavior was observed when transplanting fluorescent human umbilical cord blood CD34+ under the same conditions, except that these cells disappeared completely in the few hours following the injection (data not shown). We reasoned that primitive CD34+ cells are smaller than the Jurkat cells and they do not form cell aggregates like the latter, therefore they are more easily encapsulated and phagocytized by the macrophages (We could not image the CD34+ cells phagocytosis by live-confocal microscopy given the speed of this phenomenon).

**Longer cell survival and proliferation with macrophages escape**

Afterwards, we searched for an injection site allowing a physical escape from macrophages. First, we attempted to inject fluorescent cells (100-200 Jurkat or CD34+cells) in the swim bladder of 30 h.p.f Tg (mpeg1: mCherry) (Figure 4a-c). Few hours after injection, an important inflammation state in the zebrafish embryos was induced with an extensive recruitment of macrophages to the injection site (Figure 4b-c, arrow), followed by the disappearance of cells (data not shown). Then, we injected fluorescent cells in the hindbrain ventricle of 30 h.p.f *Tg(Mpeg1:mCherry)* (Figure 4d-f), a site that we sought it would be inaccessible to macrophages. Same observations as injecting in the swim bladder were made (Figure 4e-f). The yolk sac is the most common location used for injection of human cells in zebrafish embryos. The benefits of this injection site are its large size allowing an easy transplantation of a sufficient number of cells (Wertman et al., 2016), its acellular, partly avascular and nutrient-rich environment which can be supportive of cell proliferation (Haldi et al., 2006). Taking all these factors into consideration, we supposed that the yolk sac would be a site not accessible by

macrophages. Transduced cells were injected into the yolk sac of 30 h.p.f *Tg(Mpeg1: mCherry)* zebrafish embryo. We observed that macrophages had a homogeneous localization at the epidermal layer of the yolk sac (Figure 4g). No macrophages were detected inside the yolk sac in a direct contact with human cells (Figure 4h-i). Furthermore, the proliferation of both human cell types was detected three days post-transplantation (data not shown). Taken together, these observations prove that the absence of macrophages is suitable for the survival, the maintenance, and proliferation of human cells in the zebrafish embryo. Since injected cells remain sequestered in the yolk sac and cannot reach the blood circulation, the use of this site does not allow the study of the target cells behavior.

**Macrophage depletion allows human hematopoietic cell xenograft success**

In order to characterize precisely the role of macrophages in xenograft success, we genetically and chemically depleted primitive macrophages before cell engraftment.

First, we used a genetic approach for macrophages depletion by injecting antisense morpholinos against the myeloid transcription factor gene PU.1, thus excluding the role of both zebrafish primitive macrophages and neutrophils. When using PU.1 morpholino, morphant embryos (Figure 5a) showed a complete and efficient depletion of macrophages compared to the PU.1 non-treated control embryos (Figure 5b). Human Jurkat cell were injected in the duct of Cuvier (a temporary vessel at the rostral side of the embryo which returns blood to the heart) of PU.1 treated and non-treated *Tg(Mpeg1: mCherry)* zebrafish embryos at 24 h.p.f. As shown in **Figure 5c**, 100% of morphant zebrafish embryos (24/24) (red dash line) contains transplanted cells 1 d.p.t, unlike the control group (blue dash line), where Jurkat cells were detected in only 48% of the embryos (14/29). At 2 d.p.t, the percentage of PU.1-treated zebrafish with detectable Jurkat cells decreased to 86% Vs 41% for the control group. At 5 d.p.t the % of embryos still containing Jurkat cells is the same (≈33% on average) in both the groups of embryos. Although macrophages depletion by the morpholino PU.1 has significantly improved the transplanted cells uptake and survival for the first two days of transplanted human cells, its effect was temporary. In fact, by 2 d.p.t, some macrophages reappear in PU.1-treated embryos (data not shown) indicating the rescue of the myeloid development, which correlates with an apparent reduction in the number of transplanted GFP-cells (Figure 5c). These results confirm that macrophages are responsible for cell transplantation failure by phagocyting transplanted cells.

Then, a chemical approach was applied for the selective depletion of macrophages into embryos. We recently demonstrated that Liposome encapsulated clodronate (Lipo-C) induces specifically the death of phagocytic macrophages in zebrafish embryos in *vivo* (Travnickova et al., 2015). Injection of Lipo-C in the caudal vein of 24 h.p.f Tg (mpeg1: mCherry) zebrafish embryos induced an efficient macrophages depletion (Figure 5d) compared to the non-treated embryos (Figure 5e). Transplantation of Jurkat cells in the Duct of Cuvier of Lipo-C-treated 30 h.p.f *Tg(Mpeg1:mCherry)* embryos reproduced the results obtained with an injection into the yolk sac, a condition where transplanted cells are isolated from the zebrafish embryo macrophages (Figure 5f). Indeed, transplanted Jurkat cells are detected in the vascular system of all the Lipo-C treated embryos (9/9) at 2 d.p.t., unlike the non-treated group where the percentage of embryos containing $GFP^+$ cells is reduced to32.5% on average (in 3 different experiences). Similar results were obtained when human $CD34^+$ cells are transplanted in the Duct of Cuvier of Lipo-C-treated 30 h.p.f *Tg(Mpeg1:mCherry)* embryos. At 2 d.p.t., non-treated group showed a drastic decrease in the percentage of embryos containing $CD34^+$ GFP-cells (14.5%), whereas Lipo-C treated group maintained transplanted cells at a level similar to an injection into the yolk sac (87% on average) (Figure 5g).

## DISCUSSION

To the best of our knowledge, we are the first to decipher the reasons behind the lack of data reporting the xenotransplantation of primary human stem/progenitor cells in zebrafish embryo model. Most of studies have described the use of leukemic cell lines, primary leukemic cells isolated from patients or other cancer cells for xenotransplantation [reviewed in (Konantz et al., 2012; Wertman et al., 2016) ].

In our study, we chose to "humanize" zebrafish embryo because this organism model has emerged as an extraordinary vertebrate model for transplantation assays, thanks to many characteristics like the external fertilization, the transparency of embryos, the rapid generation of results by using a large-scale transplantation-based screening, and the existence of a variety of transgenic lines.

The first key factor to consider in xenotransplantation assays is distinguishing the transplanted target cells from the host cells. Cell membrane stains have been used extensively due to their ease of use, the rapid staining procedures and the large array of colors they provide. However, they may present high toxicity depending on the cellular type and the concentration to use in order to have a stable labelling. Moreover, these stains are diluted with each cell division, can be released into the tissue or be transferred to other cells which may not be suitable for *in vivo* live imaging (Progatzky et al., 2013). Thus, we wanted to label our cells by endogenously expressing green fluorescent protein. Lentiviruses have been shown to transduce both proliferating and non-proliferating cells and therefore are capable of infecting certain types of quiescent cells (Vigna and Naldini, 2000) as well as producing a stable expression of transgene with relatively low cytotoxicity in a variety of cell types including hematopoietic stem cells (HSCs) (Barrette et al., 2000; Blomer et al., 1997; Naldini et al., 1996; Varma et al., 2011; Vigna and Naldini, 2000). In our study, we compared the efficiency of lentiviral transduction using two different MOIs (30 and 120). The protocol described in this study was confirmed with a human Jurkat cell line, and then was applied to transduce human cord blood CD34+ hematopoietic progenitors. Our results showed that an efficient transduction may be achieved with a 24 hours pre-stimulation with a moderate (50 ng/ml) concentration of cytokines, and that a low MOI (30) is preferable to maintain a stable expression of CD34 over time, which is markedly lower than those reported by some other groups (Aiuti et al., 2013; Case et al., 1999; Charrier et al., 2007; Di Nunzio et al., 2007; Guenechea et al., 2000). Another important factor to consider is the duration of contact between target cells and the viral suspension. According to the technique of spin-inoculation described in our protocol, two hours were enough to generate good transduction efficiencies compared to long duration exposure described in other studies (Case et al., 1999; Millington et al., 2009). A previous work had addressed the issue of the efficiency of spin inoculation on CD34+ cells, but reaching less clear conclusions since the efficiency of transduction was not directly estimated on the transduced cells, but only after progenitor cell assays (Park et al., 2014). Shortening times of transduction by spin inoculation provides the potential to reduce ex vivo manipulation of the target cells with the aim of preserving their cellular functions and their undifferentiated state before transplantation.

Another important factor to take into consideration is the location of injection. The yolk sac is considered the ideal injection site and is the most common location described in the literature. It is known to be a nutrient-rich environment which is supportive of cell survival and proliferation (Haldi et al., 2006). Our findings came to strengthen these data by adding a role of protection for the yolk sac. Indeed, when human cells were injected in the Duct of Cuvier to reach the blood circulation, they disappear rapidly after injection. It is worth mentioning that human umbilical cord CD34+ disappear from the blood circulation more rapidly than Jurkat cells which are cleared within two days following transplantation. Our observations are consistent with the results of Pruvot et al showing that human umbilical cord CD34+ cells are detected only at 1 dpt and disappear rapidly, whereas Jurkat cells can be detected in the circulation up to 4 dpt. We here prove by using a Tg (mpeg1:mCherry) zebrafish

embryo line that presence of human cells in the blood circulation of zebrafish embryo induce an important inflammation state with an extensive recruitment and proliferation of primitive macrophages that phagocytose thereafter target cells (Pruvot et al., 2011). The invasive nature of cancer cells may have contributed to the longer time needed for the clearance of Jurkat cells as they can resist more than CD34+ cells to the phagocytosis by macrophages.

By keeping human cells not in a direct contact with macrophages in the yolk sac, they survive and proliferate demonstrating the role of protection provided by the yolk sac. To further confirm the role of macrophages in the human cells engraftment failure, we tried to deplete zebrafish embryos macrophages by means of two different strategies (chemical and genetic). We found that, as long as macrophages are absent, transplanted human cells can survive in the zebrafish embryos showing that the zebrafish environment provides efficient factors and signals for the survival of human hematopoietic progenitors. Furthermore, the zebrafish hematopoietic environment may provide signals for proliferation and homing of human hematopoietic progenitors given that, in many rare cases, we showed that these cells circulate actively in the vascular system of the zebrafish embryo, proliferate and colonize hematopoietic niches until 7 dpt. We were not able to verify the state of macrophages in these few embryos but they may suffer from a genetic defect affecting the primitive hematopoiesis and the generation of primitive macrophages. The latter observations are in contradiction with the study of Pruvot et al, who speculated that zebrafish environment is not suitable for human primary hematopoietic progenitors (Pruvot et al., 2011); while they are in accordance with the reports of Pozzoli et al. and Staal et al, the only two studies representing examples of successful human hematopoietic progenitors xenotransplantation in zebrafish (Pozzoli et al., 2011; Staal et al., 2016). In their study, Staal and colleagues, used transparent casper mutant zebrafish embryos. It will be interesting to characterize primitive macrophages of this strain and to compare them with those of wild-type AB strain embryos in order to understand the differences observed in their results and ours.

To our knowledge, the present study is the first to reveal a mechanism explaining the failure of human primitive cell engraftment. Although the adaptive immune system is absent in the early zebrafish embryo, our results showed that, for a successful xenotransplantation in this model, it is crucial to control the primitive wave of hematopoiesis producing the myelo-eryhtroid cells, especially the early macrophages that appear as early as 20 hours post fertilization (Herbomel et al., 1999). Xenografts of cancer cells in the yolk sac is an excellent strategy to explore the pharmacological activity of anticancer drugs; while xenografts in other sites may help to explore the interactions between transplanted cells and host cells, to determine the zebrafish-derived signals that support human primary stem survival, proliferation, homing and differentiation. Given the importance of macrophages in embryonic development and tissue homeostasis (Wynn et al., 2013), it will be useful to establish a transgenic line in which the PU1 transgene expression would be temporally tamoxifen-controlled (Akerberg et al., 2014). In this context, our results may have important fundamental and translational implications for defining the potentialities oh human primitive somatic and stem cells.

## Acknowledgements

We thank E. Stainnack and Emmanuelle Julien for technical support. This work was supported by grants from the Fondation ARC pour la recherche sur le Cancer and from ANR to M.T. and from the EFS-Grand ESTand from the Chercheur d'Avenir - Région Occitanie, FRM and ATIP-Avenir to K.K. N.A. was supported by a fellowship from the FRM (FDT20150532507) and ATIP-Avenir. R.E.O. was supported by a fellowship from ANR.

## FIGURE LEGENDS

**Figure 1. Transgene transfer and expression after transduction of human CD34 $^{+}$ cells and Jurkat cell line with an HIV-derived vector.**

(**a**) Human CD34+ cells from umbilical cord blood and Jurkat cells were transduced by spin inoculation with lentiviral vectors encoding a nuclear variant of the eGFP at multiplicities of infection (MOI) of 30 and 120 and cultured in vitro. After 5 days, cells were analyzed for GFP expression by both flow cytometry and fluorescence microscopy. (**b**) Percentage of GFP+ cells are analyzed. For both cell types, higher percentage of transduction were at MOI of 30. (**c**) Higher percentage of cells expressing CD34 detected after 5 days of culture are observed when cells are treated with a MOI at 30 (92.23 ± 1.759 N=6) compared to 120 (83.08 ± 2.980 N=9). (**d-e**) A representative experiment showing expression of the GFP transgene after transduction of human CD34 + cells or Jurkat cells at a MOI of 30. In (**e**) results are represented as histograms of GFP fluorescence intensity (x-axis, log scale) versus cell number (y-axis, linear).

**Figure 2. Transplantation of human Jurkat cells in the zebrafish embryo, colonization of hematopoietic organs and proliferation.**

**a.** Transplantation of transduced Jurkat cells in the caudal vein of a wildtype zebrafish embryo at 36 h.p.f. **b-d**. From 1 d.p.t to 7 d.p.t, proliferation and colonization of zebrafish hematopoietic tissues. Colonization of the CHT (**b, e**), thymus (**f**), fetal liver (**g**), kidney (**h**). See **Supplemental Movie 1**. **i**. Live imaging of human Jurkat cells migration (asterisk) and proliferation (Time code 00:15 and 00:24, arrow) in the CHT. See **Supplementary Movie 2**.
Scale bar: 50 µm. a-d: 250 µm. Time code in minute.

**Figure 3. Human Jurkat cells phagocytosis by zebrafish embryonic macrophages at the origin of transplantation failure.**

**a**. Live imaging of human Jurkat cells fate injected in a zebrafish embryo at 30 h.p.f. Starting from 33 min, some green followed cells are fragmented (arrow). Inset reveals a high magnification of fragments cells (arrow) and round vesicles (asterisk). See also **Supplementary Movie 3.**
**b-g**. Macrophages proliferation in a 48 h.p.f *Tg (Mpeg1:mCherry)* zebrafish embryo after Jurkat cell injection (**e-g**) compared with an embryo injected with PBS (**b-d**). **e-g**. Recruitment of macrophages at the injection site (arrow) compared to the non-injected control embryo (**b-d**, arrow).
**g**. Merged pictures show the accumulation of macrophages around transplanted Jurkat cells.
**h**. Live imaging of Mcherry+ macrophage behavior in a zebrafish embryos at 2 d.p.t injected with GFP+ Jurkat cells reveals the accumulation of macrophages around green cells and their phagocytosis at 117 min. See also **Supplementary Movie 4**.
Scale bar: a: 25 µm, b-g: 250 µm, h: 20 µm. Time code in minute.

**Figure 4. Comparison of injection sites allowing a physical escape from macrophages and consequently the xenograft success of human cells in the zebrafish embryo.**

Injection of human Jurkat cells in swim bladder (**a-c**), hindbrain ventricle (**d-f**) and yolk sac (**g-i**), of 30 h.p.f *Tg(Mpeg1:mCherry)* zebrafish embryos. (**c, f, i**) Merged pictures show macrophages accumulation in the swim bladder (**c**) and the hindbrain ventricle (**f**) but not in the yolk sac (**i**).
Scale bar: 250 µm.

**Figure 5. Depletion of macrophages allows human hematopoietic cells xenograft**

Primitive macrophages were genetically and chemically depleted before cell engraftment in Tg (mpeg1: mCherry) zebrafish embryos.

(**a-b** and **d-e**) Complete and efficient depletion of macrophages is observed in morphante PU.1 (a) and in Lipo-C-treated (**e**) embryos at 35 h.p.f and 48 h.p.f, respectively, compared to controls (**a** and **d**) which contain fluorescent macrophages (head arrows). (**c**) Fluorescent human Jurkat cell are injected in PU.1 treated (red dash line), and non-treated (blue dash line), zebrafish embryos are followed for 6 days. Higher percentage of engraftment is observed in treated embryos and is correlated with the depletion of macrophages. (**f-g**) Transplantation of Jurkat (**f**) and $CD34^+$ (**g**) cells in the Duct of Cuvier of Lipo-C-treated embryos shows a percentage of engraftment comparable to that obtained with an injection into the yolk sac. Transplanted GFP-cells are detected in the vascular system of almost all the Lipo-C treated embryos until two days post-transplantation, unlike the control non-treated.

Scale bar: **a-b**, **d-e**: 250 µm

- ■ Injection in the yolk sac of non-treated embryos.
- □ Injection into the Duct of Cuvier of non-treated embryos.
- □ Injection into the Duct of Cuvier of Lipo-C-treated embryos.

**Supplemental Movie 1.**

Time-lapse imaging of human Jurkat cells fate injected in a WT zebrafish embryos at 36 h.p.f (1 frame/3 min) during 7 days. Survival, maintenance and proliferation of human cells in the zebrafish embryo.

Scale bar: 50 µm. Time is indicated in minutes.

**Supplementary Movie 2.**

Time-lapse imaging of human Jurkat cells migration (asterisk) and proliferation (Time code 00:15 and 00:24, arrow) in the CHT. Confocal microscopy (1 frame/3 min). Maximum projection of 3 plans - CHT 3 focus.

Scale bar: 50 µm. Time is indicated in minutes.

**Supplementary Movie 3.**

Time-lapse imaging of human Jurkat cells fate injected in a zebrafish embryo at 30 h.p.f. Starting from 33 min some green followed cells are fragmented (arrow). Inset reveals a high magnification of fragments cells (arrow) and round vesicles (asterisk).

**Supplementary Movie 4.**

Time-lapse imaging of Mcherry+ macrophage behavior in a zebrafish embryos at 2 d.p.t injected with GFP+ Jurkat cells reveals the accumulation of macrophages around green cells and their phagocytosis (1 frame/3 min).

Scale bar: 50 µm. Time is indicated in minutes.

## References


Aiuti, A., Biasco, L., Scaramuzza, S., Ferrua, F., Cicalese, M.P., Baricordi, C., Dionisio, F., Calabria, A., Giannelli, S., Castiello, M.C*., et al.* (2013). Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome. Science *341*, 1233151.

Akerberg, A.A., Stewart, S., and Stankunas, K. (2014). Spatial and temporal control of transgene expression in zebrafish. PloS one *9*, e92217.

Babovic, S., and Eaves, C.J. (2014). Hierarchical organization of fetal and adult hematopoietic stem cells. Exp Cell Res *329*, 185-191.

Barrette, S., Douglas, J.L., Seidel, N.E., and Bodine, D.M. (2000). Lentivirus-based vectors transduce mouse hematopoietic stem cells with similar efficiency to moloney murine leukemia virus-based vectors. Blood *96*, 3385-3391.

Bentley, V.L., Veinotte, C.J., Corkery, D.P., Pinder, J.B., LeBlanc, M.A., Bedard, K., Weng, A.P., Berman, J.N., and Dellaire, G. (2015). Focused chemical genomics using zebrafish xenotransplantation as a pre-clinical therapeutic platform for T-cell acute lymphoblastic leukemia. Haematologica *100*, 70-76.

Benveniste, P., Frelin, C., Janmohamed, S., Barbara, M., Herrington, R., Hyam, D., and Iscove, N.N. (2010). Intermediate-term hematopoietic stem cells with extended but time-limited reconstitution potential. Cell Stem Cell *6*, 48-58.

Biasco, L., Pellin, D., Scala, S., Dionisio, F., Basso-Ricci, L., Leonardelli, L., Scaramuzza, S., Baricordi, C., Ferrua, F., Cicalese, M.P*., et al.* (2016). In Vivo Tracking of Human Hematopoiesis Reveals Patterns of Clonal Dynamics during Early and Steady-State Reconstitution Phases. Cell Stem Cell *19*, 107-119.

Blomer, U., Naldini, L., Kafri, T., Trono, D., Verma, I.M., and Gage, F.H. (1997). Highly efficient and sustained gene transfer in adult neurons with a lentivirus vector. J Virol *71*, 6641-6649.

Case, S.S., Price, M.A., Jordan, C.T., Yu, X.J., Wang, L., Bauer, G., Haas, D.L., Xu, D., Stripecke, R., Naldini, L*., et al.* (1999). Stable transduction of quiescent CD34(+)CD38(-) human hematopoietic cells by HIV-1-based lentiviral vectors. Proc Natl Acad Sci U S A *96*, 2988-2993.

Charrier, S., Dupre, L., Scaramuzza, S., Jeanson-Leh, L., Blundell, M.P., Danos, O., Cattaneo, F., Aiuti, A., Eckenberg, R., Thrasher, A.J*., et al.* (2007). Lentiviral vectors targeting WASp expression to hematopoietic cells, efficiently transduce and correct cells from WAS patients. Gene Ther *14*, 415-428.

Cheung, A.M., Nguyen, L.V., Carles, A., Beer, P., Miller, P.H., Knapp, D.J., Dhillon, K., Hirst, M., and Eaves, C.J. (2013). Analysis of the clonal growth and differentiation dynamics of primitive barcoded human cord blood cells in NSG mice. Blood *122*, 3129-3137.

Collin, J., and Martin, P. (2017). Zebrafish as a Research Organism. In Basic Science Methods for Clinical Researchers, (Elsevier), pp. 235–261.

Corkery, D.P., Dellaire, G., and Berman, J.N. (2011). Leukaemia xenotransplantation in zebrafish--chemotherapy response assay in vivo. Br J Haematol *153*, 786-789.

Crane, G.M., Jeffery, E., and Morrison, S.J. (2017). Adult haematopoietic stem cell niches. Nature reviews Immunology.

Davis, J.M., Clay, H., Lewis, J.L., Ghori, N., Herbomel, P., and Ramakrishnan, L. (2002). Real-time visualization of mycobacterium-macrophage interactions leading to initiation of granuloma formation in zebrafish embryos. Immunity *17*, 693-702.

Di Nunzio, F., Piovani, B., Cosset, F.L., Mavilio, F., and Stornaiuolo, A. (2007). Transduction of human hematopoietic stem cells by lentiviral vectors pseudotyped with the RD114-TR chimeric envelope glycoprotein. Hum Gene Ther *18*, 811-820.

Ellett, F., Pase, L., Hayman, J.W., Andrianopoulos, A., and Lieschke, G.J. (2011). mpeg1 promoter transgenes direct macrophage-lineage expression in zebrafish. Blood *117*, e49-56.

Guenechea, G., Gan, O.I., Inamitsu, T., Dorrell, C., Pereira, D.S., Kelly, M., Naldini, L., and Dick, J.E. (2000). Transduction of human CD34+ CD38- bone marrow and cord blood-derived SCID-repopulating cells with third-generation lentiviral vectors. Mol Ther *1*, 566-573.

Haldi, M., Ton, C., Seng, W.L., and McGrath, P. (2006). Human melanoma cells transplanted into zebrafish proliferate, migrate, produce melanin, form masses and stimulate angiogenesis in zebrafish. Angiogenesis *9*, 139-151.

Herbomel, P., Thisse, B., and Thisse, C. (1999). Ontogeny and behaviour of early macrophages in the zebrafish embryo. Development *126*, 3735-3745.

Herbomel, P., Thisse, B., and Thisse, C. (2001). Zebrafish early macrophages colonize cephalic mesenchyme and developing brain, retina, and epidermis through a M-CSF receptor-dependent invasive process. Developmental biology *238*, 274-288.

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F. (1995). Stages of embryonic development of the zebrafish. Dev Dyn *203*, 253-310.

Konantz, M., Balci, T.B., Hartwig, U.F., Dellaire, G., Andre, M.C., Berman, J.N., and Lengerke, C. (2012). Zebrafish xenografts as a tool for in vivo studies on human cancer. Ann N Y Acad Sci *1266*, 124-137.

Kumar, S., and Geiger, H. (2017). HSC Niche Biology and HSC Expansion Ex Vivo. Trends in molecular medicine.

Millington, M., Arndt, A., Boyd, M., Applegate, T., and Shen, S. (2009). Towards a clinically relevant lentiviral transduction protocol for primary human CD34 hematopoietic stem/progenitor cells. PloS one *4*, e6461.

Naldini, L., Blomer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F.H., Verma, I.M., and Trono, D. (1996). In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science *272*, 263-267.

North, T.E., de Bruijn, M.F., Stacy, T., Talebian, L., Lind, E., Robin, C., Binder, M., Dzierzak, E., and Speck, N.A. (2002). Runx1 expression marks long-term repopulating hematopoietic stem cells in the midgestation mouse embryo. Immunity *16*, 661-672.

Notta, F., Doulatov, S., Laurenti, E., Poeppl, A., Jurisica, I., and Dick, J.E. (2011). Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. Science *333*, 218-221.

O'Toole, R., Von Hofsten, J., Rosqvist, R., Olsson, P.E., and Wolf-Watz, H. (2004). Visualisation of zebrafish infection by GFP-labelled Vibrio anguillarum. Microb Pathog *37*, 41-46.

Park, SW., Pyo, CW., Choi, SY. (2014). High efficiency lentiviral transduction of primary human CD34+ hematopoetic cells with low-dose viral inocula. Biotechnol Lett 37, 281-288.

Pozzoli, O., Vella, P., Iaffaldano, G., Parente, V., Devanna, P., Lacovich, M., Lamia, C.L., Fascio, U., Longoni, D., Cotelli, F.*, et al.* (2011). Endothelial fate and angiogenic properties of human CD34+ progenitor cells in zebrafish. Arterioscler Thromb Vasc Biol *31*, 1589-1597.

Prajsnar, T.K., Cunliffe, V.T., Foster, S.J., and Renshaw, S.A. (2008). A novel vertebrate model of Staphylococcus aureus infection reveals phagocyte-dependent resistance of zebrafish to non-host specialized pathogens. Cell Microbiol *10*, 2312-2325.

Progatzky, F., Dallman, M.J., and Lo Celso, C. (2013). From seeing to believing: labelling strategies for in vivo cell-tracking experiments. Interface Focus *3*, 20130001.

Pruvot, B., Jacquel, A., Droin, N., Auberger, P., Bouscary, D., Tamburini, J., Muller, M., Fontenay, M., Chluba, J., and Solary, E. (2011). Leukemic cell xenograft in zebrafish embryo for investigating drug efficacy. Haematologica *96*, 612-616.

Rasighaemi, P., Basheer, F., Liongue, C., and Ward, A.C. (2015). Zebrafish as a model for leukemia and other hematopoietic disorders. J Hematol Oncol *8*, 29.

Staal, F.J., Spaink, H.P., and Fibbe, W.E. (2016). Visualizing Human Hematopoietic Stem Cell Trafficking In Vivo Using a Zebrafish Xenograft Model. Stem cells and development *25*, 360-365.

Travnickova, J., Tran Chau, V., Julien, E., Mateos-Langerak, J., Gonzalez, C., Lelievre, E., Lutfalla, G., Tavian, M., and Kissa, K. (2015). Primitive macrophages control HSPC mobilization and definitive haematopoiesis. Nature communications *6*, 6227.

van der Sar, A.M., Stockhammer, O.W., van der Laan, C., Spaink, H.P., Bitter, W., and Meijer, A.H. (2006). MyD88 innate immune function in a zebrafish embryo infection model. Infection and immunity *74*, 2436-2441.

Varma, N., Janic, B., Ali, M., Iskander, A., and Arbab, A. (2011). Lentiviral Based Gene Transduction and Promoter Studies in Human Hematopoietic Stem Cells (hHSCs). J Stem Cells Regen Med *7*, 41-53.

Vigna, E., and Naldini, L. (2000). Lentiviral vectors: excellent tools for experimental gene transfer and promising candidates for gene therapy. J Gene Med *2*, 308-316.

Westerfield M. 2000. The zebrafish book: A guide for the laboratory use of zebrafish (Danio rerio), 4th ed. University of Oregon Press, Eugene, OR.

Wertman, J., Veinotte, C.J., Dellaire, G., and Berman, J.N. (2016). The Zebrafish Xenograft Platform: Evolution of a Novel Cancer Model and Preclinical Screening Tool. Adv Exp Med Biol *916*, 289-314.

Wright, D.E., Wagers, A.J., Gulati, A.P., Johnson, F.L., and Weissman, I.L. (2001). Physiological migration of hematopoietic stem and progenitor cells. Science *294*, 1933-1936.

Wynn, T.A., Chawla, A., and Pollard, J.W. (2013). Macrophage biology in development, homeostasis and disease. Nature *496*, 445-455.

Zhang, B., Shimada, Y., Hirota, T., Ariyoshi, M., Kuroyanagi, J., Nishimura, Y., and Tanaka, T. (2016). Novel immunologic tolerance of human cancer cell xenotransplants in zebrafish. Transl Res *170*, 89-98 e83.

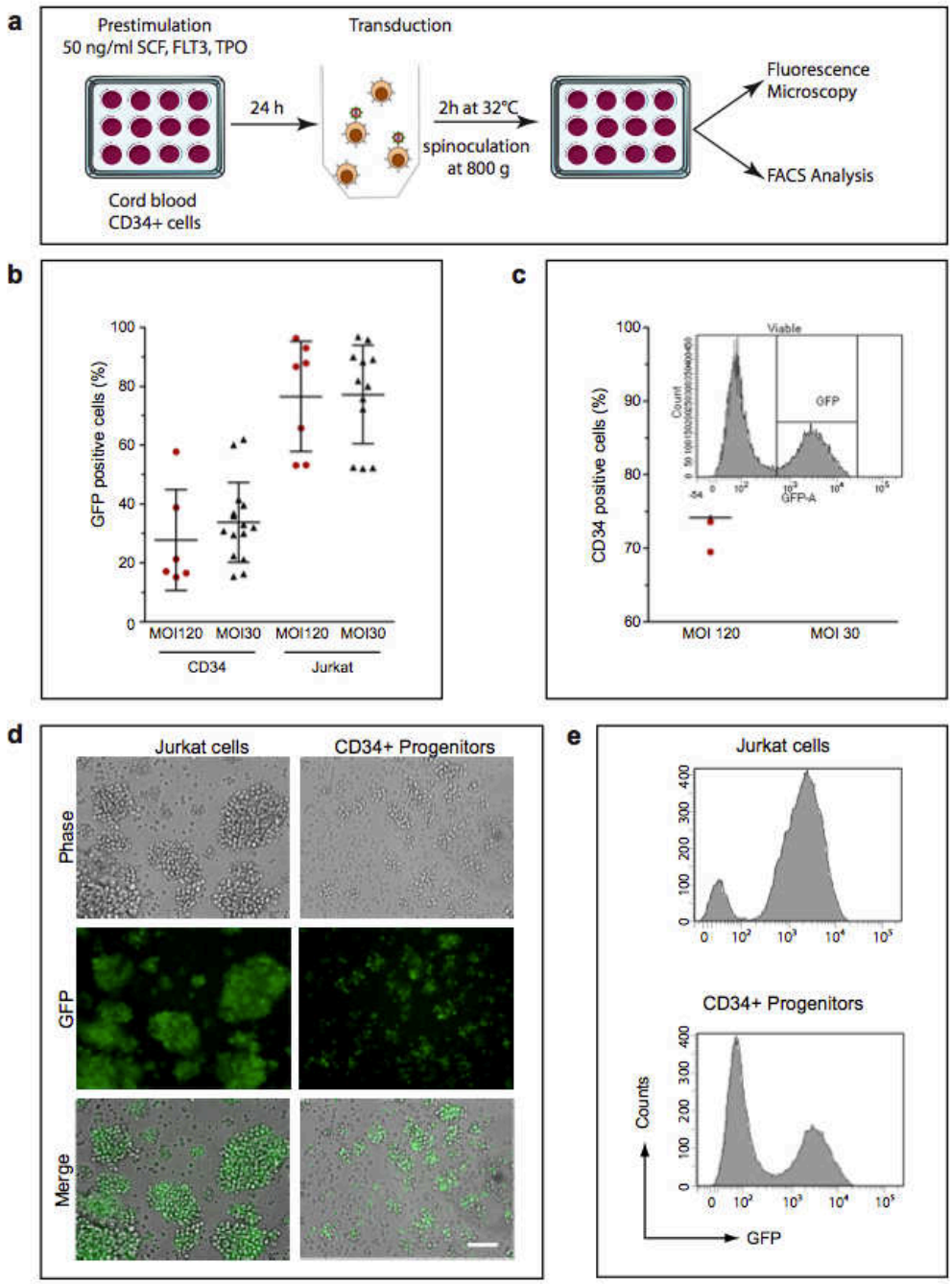



**Figure 1. Transgene transfer and expression after transduction of human CD34$^+$ cells and Jurkat cell line with an HIV-derived vector.**

(**a**) Human CD34+ cells from umbilical cord blood and Jurkat cells were transduced by spin inoculation with lentiviral vectors encoding a nuclear variant of the eGFP at multiplicities of infection (MOI) of 30 and 120 and cultured in vitro. After 5 days, cells were analyzed for GFP expression by both flow cytometry and fluorescence microscopy. (**b**) Percentage of GFP+ cells are analyzed. For both cell types, higher percentage of transduction were at MOI of 30. (**c**) Higher percentage of cells expressing CD34 detected after 5 days of culture are observed when cells are treated with a MOI at 30 (92.23 ± 1.759 N=6) compared to 120 (83.08 ± 2.980 N=9). (**d-e**) A representative experiment showing expression of the GFP transgene after transduction of human CD34 + cells or Jurkat cells at a MOI of 30. In (**e**) results are represented as histograms of GFP fluorescence intensity (x-axis, log scale) versus cell number (y-axis, linear).

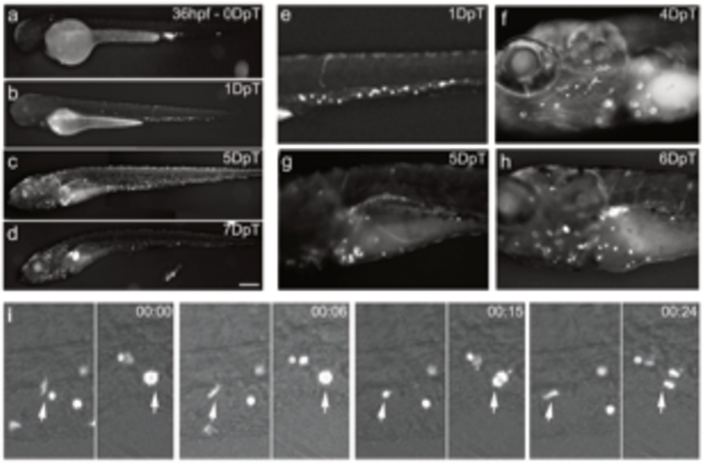



**Figure 2. Transplantation of human Jurkat cells in the zebrafish embryo, colonization of hematopoietic organs and proliferation.**

**a.** Transplantation of transduced Jurkat cells in the caudal vein of a wildtype zebrafish embryo at 36 h.p.f. **b-d**. From 1 d.p.t to 7 d.p.t, proliferation and colonization of zebrafish hematopoietic tissues. Colonization of the CHT (**b, e**), thymus (**f**), fetal liver (**g**), kidney (**h**). See **Supplemental Movie 1**. **i**. Live imaging of human Jurkat cells migration (asterisk) and proliferation (Time code 00:15 and 00:24, arrow) in the CHT. See **Supplementary Movie 2**.
Scale bar: 50 µm. a-d: 250 µm. Time code in minute.

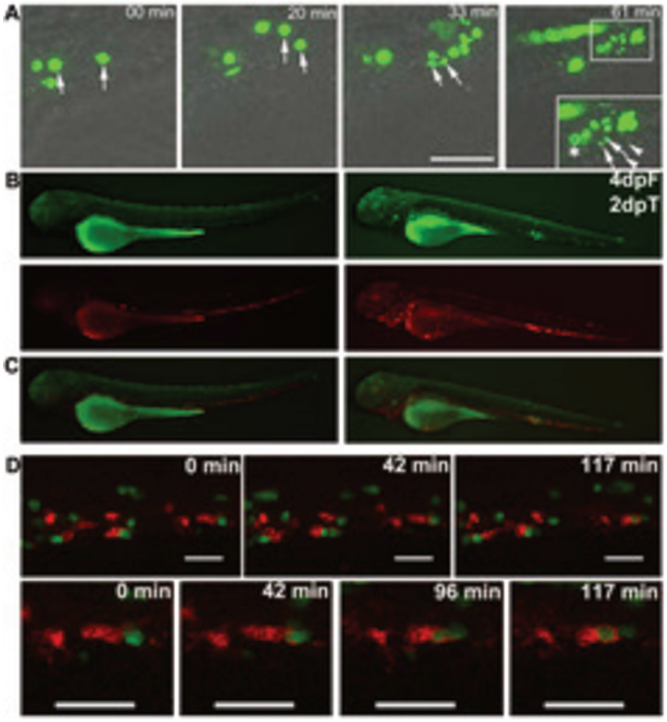



**Figure 3. Human Jurkat cells phagocytosis by zebrafish embryonic macrophages at the origin of transplantation failure.**

**a**. Live imaging of human Jurkat cells fate injected in a zebrafish embryo at 30 h.p.f. Starting from 33 min, some green followed cells are fragmented (arrow). Inset reveals a high magnification of fragments cells (arrow) and round vesicles (asterisk). See also **Supplementary Movie 3.**
**b-g**. Macrophages proliferation in a 48 h.p.f *Tg (Mpeg1:mCherry)* zebrafish embryo after Jurkat cell injection (**e-g**) compared with an embryo injected with PBS (**b-d**). **e-g**. Recruitment of macrophages at the injection site (arrow) compared to the non-injected control embryo (**b-d**, arrow).
**g**. Merged pictures show the accumulation of macrophages around transplanted Jurkat cells.

**h**. Live imaging of Mcherry+ macrophage behavior in a zebrafish embryos at 2 d.p.t injected with GFP+ Jurkat cells reveals the accumulation of macrophages around green cells and their phagocytosis at 117 min. See also **Supplementary Movie 4**. Scale bar: a: 25 µm, b-g: 250 µm, h: 20 µm. Time code in minute.

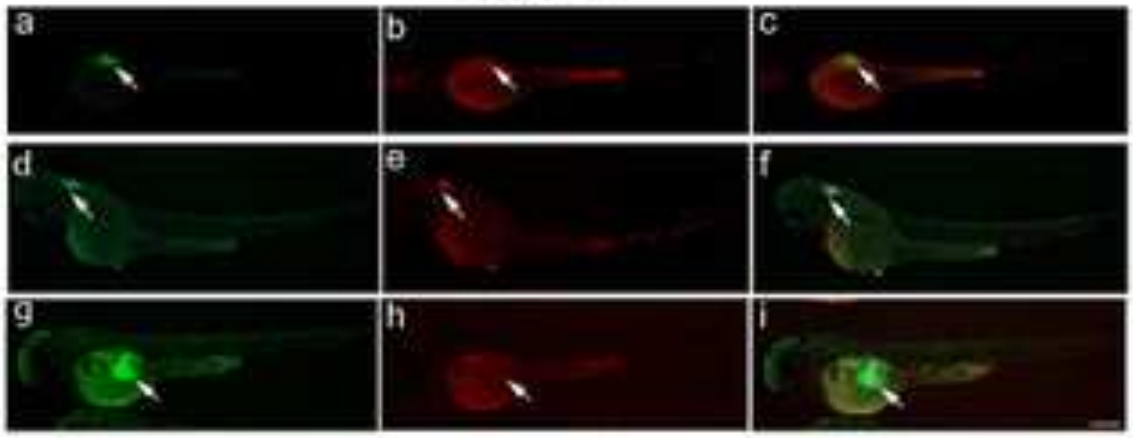



**Figure 4. Comparison of injection sites allowing a physical escape from macrophages and consequently the xenograft success of human cells in the zebrafish embryo.**

Injection of human Jurkat cells in swim bladder (**a-c**), hindbrain ventricle (**d-f**) and yolk sac (**g-i**), of 30 h.p.f *Tg(Mpeg1:mCherry)* zebrafish embryos. (**c, f, i**) Merged pictures show macrophages accumulation in the swim bladder (**c**) and the hindbrain ventricle (**f**) but not in the yolk sac (**i**).
Scale bar: 250 µm.

**Figure 5. Depletion of macrophages allows human hematopoietic cells xenograft**

Primitive macrophages were genetically and chemically depleted before cell engraftment in Tg (mpeg1: mCherry) zebrafish embryos.
(**a-b** and **d-e**) Complete and efficient depletion of macrophages is observed in morphante PU.1 (a) and in Lipo-C-treated (**e**) embryos at 35 h.p.f and 48 h.p.f, respectively, compared to controls (**a** and **d**) which contain fluorescent macrophages (head arrows). (**c**) Fluorescent human Jurkat cell are injected in PU.1 treated (red dash line), and non-treated (blue dash line), zebrafish embryos are followed for 6 days. Higher percentage of engraftment is observed in treated embryos and is correlated with the depletion of macrophages. (**f-g**) Transplantation of Jurkat (**f**) and $CD34^+$ (**g**) cells in the Duct of Cuvier of Lipo-C-treated embryos shows a percentage of engraftment comparable to that obtained with an injection into the yolk sac. Transplanted GFP-cells are detected in the vascular system of almost all the Lipo-C treated embryos until two days post-transplantation, unlike the control non-treated.
Scale bar: **a-b**, **d-e**: 250 µm

- ■ Injection in the yolk sac of non-treated embryos.
- □ Injection into the Duct of Cuvier of non-treated embryos.
- □ Injection into the Duct of Cuvier of Lipo-C-treated embryos.

**Supplemental Movie 1.**

Time-lapse imaging of human Jurkat cells fate injected in a WT zebrafish embryos at 36 h.p.f (1 frame/3 min) during 7 days. Survival, maintenance and proliferation of human cells in the zebrafish embryo.
Scale bar: 50 µm. Time is indicated in minutes.

**Supplementary Movie 2.**

Time-lapse imaging of human Jurkat cells migration (asterisk) and proliferation (Time code 00:15 and 00:24, arrow) in the CHT. Confocal microscopy (1 frame/3 min). Maximum projection of 3 plans - CHT 3 focus.
Scale bar: 50 µm. Time is indicated in minutes.

**Supplementary Movie 3.**

Time-lapse imaging of human Jurkat cells fate injected in a zebrafish embryo at 30 h.p.f. Starting from 33 min some green followed cells are fragmented (arrow). Inset reveals a high magnification of fragments cells (arrow) and round vesicles (asterisk).

**Supplementary Movie 4.**

Time-lapse imaging of Mcherry+ macrophage behavior in a zebrafish embryos at 2 d.p.t injected with GFP+ Jurkat cells reveals the accumulation of macrophages around green cells and their phagocytosis (1 frame/3 min).
Scale bar: 50 µm. Time is indicated in minutes.

# Références bibliographiques


1. OMS | Cancer. *WHO* Available at: http://www.who.int/mediacentre/factsheets/fs297/fr/. (Accessed: 17th July 2017)
2. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144,** 646–674 (2011).
3. Prior, I. A., Lewis, P. D. & Mattos, C. A comprehensive survey of Ras mutations in cancer. *Cancer Res.* **72,** 2457–2467 (2012).
4. Du, W. & Searle, J. S. The Rb Pathway and Cancer Therapeutics. *Curr. Drug Targets* **10,** 581–589 (2009).
5. Adams, J. M. & Cory, S. The Bcl-2 apoptotic switch in cancer development and therapy. *Oncogene* **26,** 1324–1337 (2007).
6. Lowe, S. W., Cepero, E. & Evan, G. Intrinsic tumour suppression. *Nature* **432,** 307–315 (2004).
7. Blasco, M. A. Telomeres and human disease: ageing, cancer and beyond. *Nat. Rev. Genet.* **6,** 611–622 (2005).
8. Raica, M., Cimpean, A. M. & Ribatti, D. Angiogenesis in pre-malignant conditions. *Eur. J. Cancer* **45,** 1924–1934 (2009).
9. Fidler, I. J. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nat. Rev. Cancer* **3,** 453–458 (2003).
10. Prince, M. E. *et al.* Identification of a subpopulation of cells with cancer stem cell properties in head and neck squamous cell carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 973–978 (2007).
11. Singh, S. K. *et al.* Identification of human brain tumour initiating cells. *Nature* **432,** 396–401 (2004).
12. Li, C. *et al.* Identification of Pancreatic Cancer Stem Cells. *Cancer Res.* **67,** 1030–1037 (2007).
13. Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 3983–3988 (2003).
14. Pasquale, E. B. Eph receptors and ephrins in cancer: bidirectional signaling and beyond. *Nat. Rev. Cancer* **10,** 165–180 (2010).
15. Carmeliet, P. & Jain, R. K. Angiogenesis in cancer and other diseases. *Nature* **407,** 249–257 (2000).
16. Gerhardt, H. & Semb, H. Pericytes: gatekeepers in tumour cell metastasis? *J. Mol. Med.* **86,** 135–144 (2008).
17. Qian, B. & Pollard, J. W. Macrophage Diversity Enhances Tumor Progression and Metastasis. *Cell* **141,** 39–51 (2010).
18. Murdoch, C., Muthana, M., Coffelt, S. B. & Lewis, C. E. The role of myeloid cells in the promotion of tumour angiogenesis. *Nat. Rev. Cancer* **8,** 618–631 (2008).
19. Wirth, T., Parker, N. & Ylä-Herttuala, S. History of gene therapy. *Gene* **525,** 162–169 (2013).
20. Griffith, F. The Significance of Pneumococcal Types. *J. Hyg. (Lond.)* **27,** 113–159 (1928).
21. Alloway, J. L. THE TRANSFORMATION IN VITRO OF R PNEUMOCOCCI INTO S FORMS OF DIFFERENT SPECIFIC TYPES BY THE USE OF FILTERED PNEUMOCOCCUS EXTRACTS. *J. Exp. Med.* **55,** 91–99 (1932).
22. Alloway, J. L. FURTHER OBSERVATIONS ON THE USE OF PNEUMOCOCCUS EXTRACTS IN EFFECTING TRANSFORMATION OF TYPE IN VITRO. *J. Exp. Med.* **57,** 265–278 (1933).
23. Avery, O. T., MacLeod, C. M. & McCarty, M. STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES. *J. Exp. Med.* **79,** 137–158 (1944).
24. Tatum, E. L. & Lederberg, J. Gene Recombination in the Bacterium Escherichia coli. *J. Bacteriol.* **53,** 673–684 (1947).

25. Rogers, S., Lowenthal, A., Terheggen, H. G. & Columbo, J. P. Induction of Arginase Activity with the Shope Papilloma Virus in Tissue Culture Cells from an Argininemic Patient. *J. Exp. Med.* **137,** 1091–1096 (1973).
26. Gene Therapy Clinical Trials Worldwide. Available at: http://www.abedia.com/wiley/indications.php. (Accessed: 22nd June 2017)
27. Vinay, D. S. *et al.* Immune evasion in cancer: Mechanistic basis and therapeutic strategies. *Semin. Cancer Biol.* **35, Supplement,** S185–S198 (2015).
28. Farkona, S., Diamandis, E. P. & Blasutig, I. M. Cancer immunotherapy: the beginning of the end of cancer? *BMC Med.* **14,** 73 (2016).
29. Rosenberg, S. A., Spiess, P. & Lafreniere, R. A new approach to the adoptive immunotherapy of cancer with tumor-infiltrating lymphocytes. *Science* **233,** 1318–1321 (1986).
30. Rosenberg, S. A. *et al.* Use of tumor-infiltrating lymphocytes and interleukin-2 in the immunotherapy of patients with metastatic melanoma. A preliminary report. *N. Engl. J. Med.* **319,** 1676–1680 (1988).
31. Loi, S. Tumor-infiltrating lymphocytes, breast cancer subtypes and therapeutic efficacy. *Oncoimmunology* **2,** (2013).
32. Sato, E. *et al.* Intraepithelial CD8+ tumor-infiltrating lymphocytes and a high CD8+/regulatory T cell ratio are associated with favorable prognosis in ovarian cancer. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 18538–18543 (2005).
33. Sharpe, M. & Mount, N. Genetically modified T cells in cancer therapy: opportunities and challenges. *Dis. Model. Mech.* **8,** 337–350 (2015).
34. Wu, A. R. *et al.* Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods* **11,** 41–46 (2014).
35. Redmond, D., Poran, A. & Elemento, O. Single-cell TCRseq: paired recovery of entire T-cell alpha and beta chain transcripts in T-cell receptors from single-cell RNAseq. *Genome Med.* **8,** (2016).
36. Ping, Y., Liu, C. & Zhang, Y. T-cell receptor-engineered T cells for cancer treatment: current status and future directions. *Protein Cell* 1–13 (2017). doi:10.1007/s13238-016-0367-1
37. Morgan, R. A. *et al.* Cancer regression and neurologic toxicity following anti-MAGE-A3 TCR gene therapy. *J. Immunother. Hagerstown Md 1997* **36,** 133–151 (2013).
38. Linette, G. P. *et al.* Cardiovascular toxicity and titin cross-reactivity of affinity-enhanced T cells in myeloma and melanoma. *Blood* **122,** 863–871 (2013).
39. Kuwana, Y. *et al.* Expression of chimeric receptor composed of immunoglobulin-derived V resions and T-cell receptor-derived C regions. *Biochem. Biophys. Res. Commun.* **149,** 960–968 (1987).
40. Maher, J. Immunotherapy of Malignant Disease Using Chimeric Antigen Receptor Engrafted T Cells. *ISRN Oncol.* **2012,** (2012).
41. Brentjens, R. *et al.* CD19-targeted T cells rapidly induce molecular remissions in adults with chemotherapy-refractory acute lymphoblastic leukemia. *Sci. Transl. Med.* **5,** 177ra38 (2013).
42. Davila, M. L. *et al.* Efficacy and Toxicity Management of 19-28z CAR T Cell Therapy in B Cell Acute Lymphoblastic Leukemia. *Sci. Transl. Med.* **6,** 224ra25 (2014).
43. Tasian, S. K. & Gardner, R. A. CD19-redirected chimeric antigen receptor-modified T cells: a promising immunotherapy for children and adults with B-cell acute lymphoblastic leukemia (ALL). *Ther. Adv. Hematol.* **6,** 228–241 (2015).
44. Grupp, S. A. *et al.* Chimeric Antigen Receptor–Modified T Cells for Acute Lymphoid Leukemia. *N. Engl. J. Med.* **368,** 1509–1518 (2013).
45. Parkhurst, M. R. *et al.* T Cells Targeting Carcinoembryonic Antigen Can Mediate Regression of Metastatic Colorectal Cancer but Induce Severe Transient Colitis. *Mol. Ther.* **19,** 620–626 (2011).
46. Morgan, R. A. *et al.* Case Report of a Serious Adverse Event Following the Administration of T Cells Transduced With a Chimeric Antigen Receptor Recognizing ERBB2. *Mol. Ther.* **18,** 843–851 (2010).

47. Hegde, M. *et al.* A bispecific chimeric antigen receptor molecule enhances T cell activation through dual immunological synapse formation and offsets antigen escape in glioblastoma. *J. Immunother. Cancer* **3,** O3 (2015).
48. Grada, Z. *et al.* TanCAR: A Novel Bispecific Chimeric Antigen Receptor for Cancer Immunotherapy. *Mol. Ther. Nucleic Acids* **2,** e105 (2013).
49. Beltinger, C. *et al.* Herpes simplex virus thymidine kinase/ganciclovir-induced apoptosis involves ligand-independent death receptor aggregation and activation of caspases. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 8699–8704 (1999).
50. Karjoo, Z., Chen, X. & Hatefi, A. Progress and Problems with the Use of Suicide Genes for Targeted Cancer Therapy. *Adv. Drug Deliv. Rev.* **99,** 113–128 (2016).
51. Preuß, E. *et al.* Cancer suicide gene therapy with TK.007: superior killing efficiency and bystander effect. *J. Mol. Med.* **89,** 1113 (2011).
52. Zhang, J., Kale, V. & Chen, M. Gene-Directed Enzyme Prodrug Therapy. *AAPS J.* **17,** 102–110 (2014).
53. The bystander effect in the HSVtk/ganciclovir system and its relationship to gap junctional communication. *Publ. Online 03 Dec. 1998 Doi101038sjgt3300784* **5,** (1998).
54. Freeman, S. M., Ramesh, R. & Marrogi, A. J. Immune system in suicide-gene therapy. *The Lancet* **349,** 2–3 (1997).
55. Nasu, Y. *et al.* Suicide Gene Therapy With Adenoviral Delivery of HSV-tK Gene for Patients With Local Recurrence of Prostate Cancer After Hormonal Therapy. *Mol. Ther.* **15,** 834–840 (2007).
56. Li, N. *et al.* Adjuvant Adenovirus-Mediated Delivery of Herpes Simplex Virus Thymidine Kinase Administration Improves Outcome of Liver Transplantation in Patients with Advanced Hepatocellular Carcinoma. *Clin. Cancer Res.* **13,** 5847–5854 (2007).
57. Berger, C., Flowers, M. E., Warren, E. H. & Riddell, S. R. Analysis of transgene-specific immune responses that limit the in vivo persistence of adoptively transferred HSV-TK–modified donor T cells after allogeneic hematopoietic cell transplantation. *Blood* **107,** 2294–2302 (2006).
58. Burt, R. K. *et al.* Herpes simplex thymidine kinase gene–transduced donor lymphocyte infusions. *Exp. Hematol.* **31,** 903–910 (2003).
59. Rossolillo, P., Winter, F., Simon-Loriere, E., Gallois-Montbrun, S. & Negroni, M. Retrovolution: HIV–Driven Evolution of Cellular Genes and Improvement of Anticancer Drug Activation. *PLoS Genet* **8,** e1002904 (2012).
60. Carmichael, J. *et al.* Advanced breast cancer: a phase II trial with gemcitabine. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **13,** 2731–2736 (1995).
61. Carmichael, J. *et al.* Phase II study of gemcitabine in patients with advanced pancreatic cancer. *Br. J. Cancer* **73,** 101–105 (1996).
62. Lund, B., Hansen, O. P., Theilade, K., Hansen, M. & Neijt, J. P. Phase II Study of Gemcitabine (2′,2′-Difluorodeoxycytidine) in Previously Treated Ovarian Cancer Patients. *JNCI J. Natl. Cancer Inst.* **86,** 1530–1533 (1994).
63. Klanova, M. *et al.* Downregulation of deoxycytidine kinase in cytarabine-resistant mantle cell lymphoma cells confers cross-resistance to nucleoside analogs gemcitabine, fludarabine and cladribine, but not to other classes of anti-lymphoma agents. *Mol. Cancer* **13,** 159 (2014).
64. Bhalla, K., Nayak, R. & Grant, S. Isolation and Characterization of a Deoxycytidine Kinase-deficient Human Promyelocytic Leukemic Cell Line Highly Resistant to 1-β-d-Arabinofuranosylcytosine. *Cancer Res.* **44,** 5029–5037 (1984).
65. Galmarini, C. M. *et al.* Resistance to gemcitabine in a human follicular lymphoma cell line is due to partial deletion of the deoxycytidine kinase gene. *BMC Pharmacol.* **4,** 8 (2004).
66. Bergman, A. M., Pinedo, H. M. & Peters, G. J. Determinants of resistance to 2′,2′-difluorodeoxycytidine (gemcitabine). *Drug Resist. Updat.* **5,** 19–33 (2002).
67. Geutjes, E.-J., Tian, S., Roepman, P. & Bernards, R. Deoxycytidine kinase is overexpressed in poor outcome breast cancer and determines responsiveness to nucleoside analogs. *Breast Cancer Res. Treat.* **131,** 809–818 (2012).

68. Bergman, A. M. *et al. In vivo* Induction of Resistance to Gemcitabine Results in Increased Expression of Ribonucleotide Reductase Subunit M1 as the Major Determinant. *Cancer Res.* **65,** 9510–9516 (2005).
69. Jordheim, L. P., Guittet, O., Lepoivre, M., Galmarini, C. M. & Dumontet, C. Increased expression of the large subunit of ribonucleotide reductase is involved in resistance to gemcitabine in human mammary adenocarcinoma cells. *Mol. Cancer Ther.* **4,** 1268–1276 (2005).
70. Davidson, J. D. *et al.* An Increase in the Expression of Ribonucleotide Reductase Large Subunit 1 Is Associated with Gemcitabine Resistance in Non-Small Cell Lung Cancer Cell Lines. *Cancer Res.* **64,** 3761–3766 (2004).
71. Bepler, G. *et al.* RRM1 Modulated In Vitro and In Vivo Efficacy of Gemcitabine and Platinum in Non–Small-Cell Lung Cancer. *J. Clin. Oncol.* **24,** 4731–4737 (2006).
72. Cottin, S., Ghani, K., de Campos-Lima, P. O. & Caruso, M. Gemcitabine intercellular diffusion mediated by gap junctions: new implications for cancer therapy. *Mol. Cancer* **9,** 141 (2010).
73. Garcia-Rodríguez, L. *et al.* Connexin-26 Is a Key Factor Mediating Gemcitabine Bystander Effect. *Mol. Cancer Ther.* **10,** 505–517 (2011).
74. Sabini, E., Ort, S., Monnerjahn, C., Konrad, M. & Lavie, A. Structure of human dCK suggests strategies to improve anticancer and antiviral therapy. *Nat. Struct. Mol. Biol.* **10,** 513–519 (2003).
75. Hazra, S., Szewczak, A., Ort, S., Konrad, M. & Lavie, A. Post-translational phosphorylation of serine 74 of human deoxycytidine kinase favors the enzyme adopting the open conformation making it competent for nucleoside binding and release. *Biochemistry (Mosc.)* **50,** 2870–2880 (2011).
76. McSorley, T., Ort, S., Hazra, S., Lavie, A. & Konrad, M. Mimicking phosphorylation of Ser-74 on human deoxycytidine kinase selectively increases catalytic activity for dC and dC analogues. *FEBS Lett.* **582,** 720–724 (2008).
77. Neschadim, A. *et al.* Cell Fate Control Gene Therapy Based on Engineered Variants of Human Deoxycytidine Kinase. *Mol. Ther.* **20,** 1002–1013 (2012).
78. Vernejoul, F. *et al.* Gene Therapy Based on Gemcitabine Chemosensitization Suppresses Pancreatic Tumor Growth. *Mol. Ther.* **14,** 758–767 (2006).
79. Hollon, T. Researchers and regulators reflect on first gene therapy death. *Nat. Med.* **6,** 6–6 (2000).
80. Hacein-Bey-Abina, S. *et al.* LMO2-Associated Clonal T Cell Proliferation in Two Patients after Gene Therapy for SCID-X1. *Science* **302,** 415–419 (2003).
81. Hacein-Bey-Abina, S. *et al.* A Serious Adverse Event after Successful Gene Therapy for X-Linked Severe Combined Immunodeficiency. *N. Engl. J. Med.* **348,** 255–256 (2003).
82. Glover, D. J., Lipps, H. J. & Jans, D. A. Towards safe, non-viral therapeutic gene expression in humans. *Nat. Rev. Genet.* **6,** 299–310 (2005).
83. Hardee, C. L., Arévalo-Soliz, L. M., Hornstein, B. D. & Zechiedrich, L. Advances in Non-Viral DNA Vectors for Gene Therapy. *Genes* **8,** (2017).
84. André, F. & Mir, L. M. DNA electrotransfer: its principles and an updated review of its therapeutic applications. *Gene Ther.* **11,** S33–S42 (2004).
85. Herweijer, H. & Wolff, J. A. Gene therapy progress and prospects: Hydrodynamic gene delivery. *Gene Ther.* **14,** 99–107 (2006).
86. Nayerossadat, N., Maedeh, T. & Ali, P. A. Viral and nonviral delivery systems for gene delivery. *Adv. Biomed. Res.* **1,** (2012).
87. Yin, H. *et al.* Non-viral vectors for gene-based therapy. *Nat. Rev. Genet.* **15,** 541–555 (2014).
88. Polyethylenimine/DNA complexes shielded by transferrin target gene expression to tumors after systemic application. *Publ. Online 29 January 2001 Doi101038sjgt3301351* **8,** (2001).
89. Hood, J. D. *et al.* Tumor Regression by Targeted Gene Delivery to the Neovasculature. *Science* **296,** 2404–2407 (2002).
90. Hofland, H. E. J. *et al.* Folate-Targeted Gene Transfer in Vivo. *Mol. Ther.* **5,** 739–744 (2002).

91. Plank, C., Zatloukal, K., Cotten, M., Mechtler, K. & Wagner, E. Gene transfer into hepatocytes using asialoglycoprotein receptor mediated endocytosis of DNA complexed with an artificial tetra-antennary galactose ligand. *Bioconjug. Chem.* **3,** 533–539 (1992).
92. Chain Length of the Polylysine in Receptor-targeted Gene Transfer Complexes Affects Duration of Reporter Gene Expression Both in Vitro and in Vivo. Available at: http://www.jbc.org/content/274/8/4908.long. (Accessed: 5th June 2017)
93. Buschle, M. *et al.* Receptor-mediated gene transfer into human T lymphocytes via binding of DNA/CD3 antibody particles to the CD3 T cell receptor complex. *Hum. Gene Ther.* **6,** 753–761 (1995).
94. Chiu, S.-J., Ueno, N. T. & Lee, R. J. Tumor-targeted gene delivery via anti-HER2 antibody (trastuzumab, Herceptin®) conjugated polyethylenimine. *J. Controlled Release* **97,** 357–369 (2004).
95. Ortiz-Urda, S. *et al.* Stable nonviral genetic correction of inherited human skin disease. *Nat. Med.* **8,** 1166–1170 (2002).
96. Olivares, E. C. *et al.* Site-specific genomic integration produces therapeutic Factor IX levels in mice. *Nat. Biotechnol.* **20,** 1124–1128 (2002).
97. Rowe, W. P., Huebner, R. J., Gilmore, L. K., Parrott, R. H. & Ward, T. G. Isolation of a Cytopathogenic Agent from Human Adenoids Undergoing Spontaneous Degeneration in Tissue Culture. *Proc. Soc. Exp. Biol. Med.* **84,** 570–573 (1953).
98. HAdV Working Group.
99. Vellinga, J. *et al.* Efficient incorporation of a functional hyper-stable single-chain antibody fragment protein-IX fusion in the adenovirus capsid. *Gene Ther.* **14,** 664–670 (2007).
100. Dmitriev, I. P., Kashentseva, E. A. & Curiel, D. T. Engineering of Adenovirus Vectors Containing Heterologous Peptide Sequences in the C Terminus of Capsid Protein IX. *J. Virol.* **76,** 6893–6899 (2002).
101. Dormond, E., Perrier, M. & Kamen, A. From the first to the third generation adenoviral vector: What parameters are governing the production yield? *Biotechnol. Adv.* **27,** 133–144 (2009).
102. Melnick, J. L., Mayor, H. D., Smith, K. O. & Rapp, F. Association of 20-Millimicron Particles with Adenoviruses. *J. Bacteriol.* **90,** 271–274 (1965).
103. Mayor, H. D., Jamison, R. M., Jordan, L. E. & Melnick, J. L. Structure and Composition of a Small Particle Prepared from a Simian Adenovirus. *J. Bacteriol.* **90,** 235–242 (1965).
104. Hoggan, M. D., Blacklow, N. R. & Rowe, W. P. Studies of small DNA viruses found in various adenovirus preparations: physical, biological, and immunological characteristics. *Proc. Natl. Acad. Sci. U. S. A.* **55,** 1467–1474 (1966).
105. Daya, S. & Berns, K. I. Gene Therapy Using Adeno-Associated Virus Vectors. *Clin. Microbiol. Rev.* **21,** 583–593 (2008).
106. Kotin, R. M. *et al.* Site-specific integration by adeno-associated virus. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 2211–2215 (1990).
107. Kotin, R. M., Linden, R. M. & Berns, K. I. Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. *EMBO J.* **11,** 5071–5078 (1992).
108. Nakai, H. *et al.* AAV serotype 2 vectors preferentially integrate into active genes in mice. *Nat. Genet.* **34,** 297–302 (2003).
109. Chao, H. *et al.* Several Log Increase in Therapeutic Transgene Delivery by Distinct Adeno-Associated Viral Serotype Vectors. *Mol. Ther.* **2,** 619–623 (2000).
110. Bish, L. T. *et al.* Adeno-Associated Virus (AAV) Serotype 9 Provides Global Cardiac Gene Transfer Superior to AAV1, AAV6, AAV7, and AAV8 in the Mouse and Rat. *Hum. Gene Ther.* **19,** 1359–1368 (2008).
111. Rabinowitz, J. E. *et al.* Cross-Packaging of a Single Adeno-Associated Virus (AAV) Type 2 Vector Genome into Multiple AAV Serotypes Enables Transduction with Broad Specificity. *J. Virol.* **76,** 791–801 (2002).
112. Yan, Z., Zhang, Y., Duan, D. & Engelhardt, J. F. Trans-splicing vectors expand the utility of adeno-associated virus for gene therapy. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 6716–6721 (2000).

113. Zacchigna, S., Zentilin, L. & Giacca, M. Adeno-Associated Virus Vectors as Therapeutic and Investigational Tools in the Cardiovascular System. *Circ. Res.* **114,** 1827–1846 (2014).
114. de Marco, A. *et al.* Role of the SP2 Domain and Its Proteolytic Cleavage in HIV-1 Structural Maturation and Infectivity. *J. Virol.* **86,** 13708–13716 (2012).
115. Pettit, S. C. *et al.* The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *J. Virol.* **68,** 8017–8027 (1994).
116. Engelman, A. & Cherepanov, P. The structural biology of HIV-1: mechanistic and therapeutic insights. *Nat. Rev. Microbiol.* **10,** 279–290 (2012).
117. Mölling, K. *et al.* Association of Viral Reverse Transcriptase with an Enzyme degrading the RNA Moiety of RNA-DNA Hybrids. *Nature* **234,** 240–243 (1971).
118. Abram, M. E., Ferris, A. L., Shao, W., Alvord, W. G. & Hughes, S. H. Nature, Position, and Frequency of Mutations Made in a Single Cycle of HIV-1 Replication. *J. Virol.* **84,** 9864–9878 (2010).
119. McCune, J. M. *et al.* Endoproteolytic cleavage of gp160 is required for the activation of human immunodeficiency virus. *Cell* **53,** 55–67 (1988).
120. Pancera, M. *et al.* Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 1166–1171 (2010).
121. Chan, D. C., Fass, D., Berger, J. M. & Kim, P. S. Core Structure of gp41 from the HIV Envelope Glycoprotein. *Cell* **89,** 263–273 (1997).
122. Dingwall, C. *et al.* Human immunodeficiency virus 1 tat protein binds trans-activation-responsive region (TAR) RNA in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 6925–6929 (1989).
123. Demarchi, F., d'Adda di Fagagna, F., Falaschi, A. & Giacca, M. Activation of transcription factor NF-kappaB by the Tat protein of human immunodeficiency virus type 1. *J. Virol.* **70,** 4427–4437 (1996).
124. Suñé, C. & García-Blanco, M. A. Sp1 transcription factor is required for in vitro basal and Tat-activated transcription from the human immunodeficiency virus type 1 long terminal repeat. *J. Virol.* **69,** 6572–6576 (1995).
125. Askjaer, P., Jensen, T. H., Nilsson, J., Englmeier, L. & Kjems, J. The Specificity of the CRM1-Rev Nuclear Export Signal Interaction Is Mediated by RanGTP. *J. Biol. Chem.* **273,** 33414–33422 (1998).
126. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418,** 646–650 (2002).
127. Sheehy, A. M., Gaddis, N. C. & Malim, M. H. The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. *Nat. Med.* **9,** 1404–1407 (2003).
128. Yu, X. *et al.* Induction of APOBEC3G Ubiquitination and Degradation by an HIV-1 Vif-Cul5-SCF Complex. *Science* **302,** 1056–1060 (2003).
129. Bouyac, M. *et al.* Human immunodeficiency virus type 1 Vif protein binds to the Pr55Gag precursor. *J. Virol.* **71,** 9358–9365 (1997).
130. Ayyavoo, V. *et al.* HIV-1 Vpr suppresses immune activation and apoptosis through regulation of nuclear factor κB. *Nat. Med.* **3,** 1117–1123 (1997).
131. Amini, S., Saunders, M., Kelley, K., Khalili, K. & Sawaya, B. E. Interplay between HIV-1 Vpr and Sp1 Modulates p21WAF1 Gene Expression in Human Astrocytes. *J. Biol. Chem.* **279,** 46046–46056 (2004).
132. Le Rouzic, E. & Benichou, S. The Vpr protein from HIV-1: distinct roles along the viral life cycle. *Retrovirology* **2,** 11 (2005).
133. Zhao, R. Y. & Bukrinsky, M. I. HIV-1 Accessory Proteins: VpR. *Methods Mol. Biol. Clifton NJ* **1087,** 125–134 (2014).
134. Popov, S. *et al.* Viral protein R regulates nuclear import of the HIV-1 pre-integration complex. *EMBO J.* **17,** 909–917 (1998).

135. Schwartz, O., Maréchal, V., Gall, S. L., Lemonnier, F. & Heard, J.-M. Endocytosis of major histocompatibility complex class I molecules is induced by the HIV–1 Nef protein. *Nat. Med.* **2,** 338–342 (1996).
136. Aiken, C., Konner, J., Landau, N. R., Lenburg, M. E. & Trono, D. Nef induces CD4 endocytosis: Requirement for a critical dileucine motif in the membrane-proximal CD4 cytoplasmic domain. *Cell* **76,** 853–864 (1994).
137. Rhee, S. S. & Marsh, J. W. Human immunodeficiency virus type 1 Nef-induced down-modulation of CD4 is due to rapid internalization and degradation of surface CD4. *J. Virol.* **68,** 5156–5163 (1994).
138. Downregulation of cell-surface CD4 expression by simian immunodeficiency virus Nef prevents viral super infection. *J. Exp. Med.* **177,** 1561–1566 (1993).
139. Rosa, A. *et al.* HIV-1 Nef promotes infection by excluding SERINC5 from virion incorporation. *Nature* **526,** 212–217 (2015).
140. Usami, Y., Wu, Y. & Göttlinger, H. G. SERINC3 and SERINC5 restrict HIV-1 infectivity and are counteracted by Nef. *Nature* **526,** 218–223 (2015).
141. Bour, S., Schubert, U. & Strebel, K. The human immunodeficiency virus type 1 Vpu protein specifically binds to the cytoplasmic domain of CD4: implications for the mechanism of degradation. *J. Virol.* **69,** 1510–1520 (1995).
142. Kimura, T., Nishikawa, M. & Ohyama, A. Intracellular membrane traffic of human immunodeficiency virus type 1 envelope glycoproteins: vpu liberates Golgi-targeted gp160 from CD4-dependent retention in the endoplasmic reticulum. *J. Biochem. (Tokyo)* **115,** 1010–1020 (1994).
143. Neil, S. J. D., Zang, T. & Bieniasz, P. D. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* **451,** 425–430 (2008).
144. Joseph, S. B., Arrildt, K. T., Sturdevant, C. B. & Swanstrom, R. HIV-1 target cells in the CNS. *J. Neurovirol.* **21,** 276–289 (2015).
145. Geijtenbeek, T. B. H. *et al.* DC-SIGN, a Dendritic Cell–Specific HIV-1-Binding Protein that Enhances trans-Infection of T Cells. *Cell* **100,** 587–597 (2000).
146. Forshey, B. M., Schwedler, U. von, Sundquist, W. I. & Aiken, C. Formation of a Human Immunodeficiency Virus Type 1 Core of Optimal Stability Is Crucial for Viral Replication. *J. Virol.* **76,** 5667–5677 (2002).
147. Tang, S. *et al.* Human Immunodeficiency Virus Type 1 N-Terminal Capsid Mutants That Exhibit Aberrant Core Morphology and Are Blocked in Initiation of Reverse Transcription in Infected Cells. *J. Virol.* **75,** 9357–9366 (2001).
148. Iordanskiy, S. & Bukrinsky, M. Reverse transcription complex: the key player of the early phase of HIV replication. *Future Virol.* **2,** 49–64 (2007).
149. Fassati, A. & Goff, S. P. Characterization of Intracellular Reverse Transcription Complexes of Human Immunodeficiency Virus Type 1. *J. Virol.* **75,** 3626–3635 (2001).
150. Bukrinsky, M. I. *et al.* Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection. *Proc. Natl. Acad. Sci.* **90,** 6125–6129 (1993).
151. Rivière, L., Darlix, J.-L. & Cimarelli, A. Analysis of the Viral Elements Required in the Nuclear Import of HIV-1 DNA. *J. Virol.* **84,** 729–739 (2010).
152. Zennou, V. *et al.* HIV-1 Genome Nuclear Import Is Mediated by a Central DNA Flap. *Cell* **101,** 173–185 (2000).
153. Ao, Z. *et al.* Contribution of Host Nucleoporin 62 in HIV-1 Integrase Chromatin Association and Viral DNA Integration. *J. Biol. Chem.* **287,** 10544–10555 (2012).
154. Fouchier, R. A. M. *et al.* Interaction of the Human Immunodeficiency Virus Type 1 Vpr Protein with the Nuclear Pore Complex. *J. Virol.* **72,** 6004–6013 (1998).
155. Nitahara-Kasahara, Y. *et al.* Novel Nuclear Import of Vpr Promoted by Importin α Is Crucial for Human Immunodeficiency Virus Type 1 Replication in Macrophages. *J. Virol.* **81,** 5284–5293 (2007).

156. Gallay, P., Stitt, V., Mundy, C., Oettinger, M. & Trono, D. Role of the karyopherin pathway in human immunodeficiency virus type 1 nuclear import. *J. Virol.* **70,** 1027–1032 (1996).
157. Depienne, C. *et al.* Cellular distribution and karyophilic properties of matrix, integrase, and Vpr proteins from the human and simian immunodeficiency viruses. *Exp. Cell Res.* **260,** 387–395 (2000).
158. Gallay, P., Swingler, S., Song, J., Bushman, F. & Trono, D. HIV nuclear import is governed by the phosphotyrosine-mediated binding of matrix to the core domain of integrase. *Cell* **83,** 569–576 (1995).
159. Yamashita, M. & Emerman, M. Capsid Is a Dominant Determinant of Retrovirus Infectivity in Nondividing Cells. *J. Virol.* **78,** 5670–5678 (2004).
160. Depienne, C. *et al.* Characterization of the Nuclear Import Pathway for HIV-1 Integrase. *J. Biol. Chem.* **276,** 18102–18107 (2001).
161. Gallay, P., Hope, T., Chin, D. & Trono, D. HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 9825–9830 (1997).
162. Schröder, A. R. W. *et al.* HIV-1 Integration in the Human Genome Favors Active Genes and Local Hotspots. *Cell* **110,** 521–529 (2002).
163. Scherdin, U., Rhodes, K. & Breindl, M. Transcriptionally active genome regions are preferred targets for retrovirus integration. *J. Virol.* **64,** 907–912 (1990).
164. Sharkey, M. E. *et al.* Persistence of episomal HIV-1 infection intermediates in patients on highly active anti-retroviral therapy. *Nat. Med.* **6,** 76–81 (2000).
165. Teo, I. *et al.* Circular forms of unintegrated human immunodeficiency virus type 1 DNA and high levels of viral protein expression: association with dementia and multinucleated giant cells in the brains of patients with AIDS. *J. Virol.* **71,** 2928–2933 (1997).
166. Pang, S. *et al.* High levels of unintegrated HIV-1 DNA in brain tissue of AIDS dementia patients. *Nature* **343,** 85–89 (1990).
167. Schröder, A. R. W. *et al.* HIV-1 Integration in the Human Genome Favors Active Genes and Local Hotspots. *Cell* **110,** 521–529 (2002).
168. Karn, J. & Stoltzfus, C. M. Transcriptional and Posttranscriptional Regulation of HIV-1 Gene Expression. *Cold Spring Harb. Perspect. Med.* **2,** (2012).
169. Cherrington, J. & Ganem, D. Regulation of polyadenylation in human immunodeficiency virus (HIV): contributions of promoter proximity and upstream sequences. *EMBO J.* **11,** 1513–1524 (1992).
170. Ashe, M. P., Griffin, P., James, W. & Proudfoot, N. J. Poly(A) site selection in the HIV-1 provirus: inhibition of promoter-proximal polyadenylation by the downstream major splice donor site. *Genes Dev.* **9,** 3008–3025 (1995).
171. Ashe, M. P., Pearson, L. H. & Proudfoot, N. J. The HIV-1 5' LTR poly(A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site. *EMBO J.* **16,** 5752–5763 (1997).
172. Das, A. T., Klaver, B. & Berkhout, B. A Hairpin Structure in the R Region of the Human Immunodeficiency Virus Type 1 RNA Genome Is Instrumental in Polyadenylation Site Selection. *J. Virol.* **73,** 81–91 (1999).
173. Gee, A. H., Kasprzak, W. & Shapiro, B. A. Structural differentiation of the HIV-1 polyA signals. *J. Biomol. Struct. Dyn.* **23,** 417–428 (2006).
174. Colgan, D. F. & Manley, J. L. Mechanism and regulation of mRNA polyadenylation. *Genes Dev.* **11,** 2755–2766 (1997).
175. Zhou, Q., Chen, D., Pierstorff, E. & Luo, K. Transcription elongation factor P-TEFb mediates Tat activation of HIV-1 transcription at multiple stages. *EMBO J.* **17,** 3681–3691 (1998).
176. Southgate, C. D. & Green, M. R. The HIV-1 Tat protein activates transcription from an upstream DNA-binding site: implications for Tat function. *Genes Dev.* **5,** 2496–2507 (1991).
177. Isel, C. & Karn, J. Direct evidence that HIV-1 tat stimulates RNA polymerase II carboxyl-terminal domain hyperphosphorylation during transcriptional elongation11Edited by A. R. Fersht. *J. Mol. Biol.* **290,** 929–941 (1999).

178. Kim, Y. K., Bourgeois, C. F., Isel, C., Churcher, M. J. & Karn, J. Phosphorylation of the RNA Polymerase II Carboxyl-Terminal Domain by CDK9 Is Directly Responsible for Human Immunodeficiency Virus Type 1 Tat-Activated Transcriptional Elongation. *Mol. Cell. Biol.* **22,** 4622–4637 (2002).
179. Fischer, U., Huber, J., Boelens, W. C., Mattajt, L. W. & Lührmann, R. The HIV-1 Rev Activation Domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell* **82,** 475–483 (1995).
180. Neville, M., Stutz, F., Lee, L., Davis, L. I. & Rosbash, M. The importin-beta family member Crm1p bridges the interaction between Rev and the nuclear pore complex during nuclear export. *Curr. Biol.* **7,** 767–775 (1997).
181. Fornerod, M., Ohno, M., Yoshida, M. & Mattaj, I. W. CRM1 Is an Export Receptor for Leucine-Rich Nuclear Export Signals. *Cell* **90,** 1051–1060 (1997).
182. Kolchinsky, P., Kiprilov, E., Bartley, P., Rubinstein, R. & Sodroski, J. Loss of a Single N-Linked Glycan Allows CD4-Independent Human Immunodeficiency Virus Type 1 Infection by Altering the Position of the gp120 V1/V2 Variable Loops. *J. Virol.* **75,** 3435–3443 (2001).
183. Dumonceaux, J., Goujon, C., Joliot, V., Briand, P. & Hazan, U. Determination of essential amino acids involved in the CD4-independent tropism of the X4 human immunodeficiency virus type 1 m7NDK isolate: role of potential N glycosylations in the C2 and V3 regions of gp120. *J. Virol.* **75,** 5425–5428 (2001).
184. Ly, A. & Stamatatos, L. V2 Loop Glycosylation of the Human Immunodeficiency Virus Type 1 SF162 Envelope Facilitates Interaction of This Protein with CD4 and CCR5 Receptors and Protects the Virus from Neutralization by Anti-V3 Loop and Anti-CD4 Binding Site Antibodies. *J. Virol.* **74,** 6769–6776 (2000).
185. Malenbaum, S. E. *et al.* The N-terminal V3 loop glycan modulates the interaction of clade A and B human immunodeficiency virus type 1 envelopes with CD4 and chemokine receptors. *J. Virol.* **74,** 11008–11016 (2000).
186. Li, Y., Luo, L., Rasool, N. & Kang, C. Y. Glycosylation is necessary for the correct folding of human immunodeficiency virus gp120 in CD4 binding. *J. Virol.* **67,** 584–588 (1993).
187. Huang, X. *et al.* Highly conserved HIV-1 gp120 glycans proximal to CD4-binding region affect viral infectivity and neutralizing antibody induction. *Virology* **423,** 97–106 (2012).
188. Carlton, J. G. & Martin-Serrano, J. Parallels Between Cytokinesis and Retroviral Budding: A Role for the ESCRT Machinery. *Science* **316,** 1908–1912 (2007).
189. Schmidt, O. & Teis, D. The ESCRT machinery. *Curr. Biol.* **22,** R116–R120 (2012).
190. Zhu, P. *et al.* Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature* **441,** 847–852 (2006).
191. Wyatt, R. *et al.* Analysis of the interaction of the human immunodeficiency virus type 1 gp120 envelope glycoprotein with the gp41 transmembrane glycoprotein. *J. Virol.* **71,** 9722–9731 (1997).
192. Helseth, E., Olshevsky, U., Furman, C. & Sodroski, J. Human immunodeficiency virus type 1 gp120 envelope glycoprotein regions important for association with the gp41 transmembrane glycoprotein. *J. Virol.* **65,** 2119–2123 (1991).
193. Back, N. K. T. *et al.* An N-Glycan within the Human Immunodeficiency Virus Type 1 gp120 V3 Loop Affects Virus Neutralization. *Virology* **199,** 431–438 (1994).
194. Burton, D. R. *et al.* HIV vaccine design and the neutralizing antibody problem. *Nat. Immunol.* **5,** 233–236 (2004).
195. Reitter, J. N., Means, R. E. & Desrosiers, R. C. A role for carbohydrates in immune evasion in AIDS. *Nat. Med.* **4,** 679–684 (1998).
196. Kwong, P. D. *et al.* Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393,** 648–659 (1998).
197. Wilen, C. B., Tilton, J. C. & Doms, R. W. HIV: Cell Binding and Entry. *Cold Spring Harb. Perspect. Med.* **2,** (2012).
198. Chen, B. *et al.* Structure of an unliganded simian immunodeficiency virus gp120 core. *Nature* **433,** 834–841 (2005).

199. Hwang, S. S., Boyle, T. J., Lyerly, H. K. & Cullen, B. R. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. *Science* **253,** 71–74 (1991).
200. Fouchier, R. A. *et al.* Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. *J. Virol.* **66,** 3183–3187 (1992).
201. Schneider, J., Kaaden, O., Copeland, T. D., Oroszlan, S. & Hunsmann, G. Shedding and Interspecies Type Sero-reactivity of the Envelope Glycopolypeptide gp120 of the Human Immunodeficiency Virus. *J. Gen. Virol.* **67,** 2533–2538 (1986).
202. Reeves, J. D. *et al.* Primary Human Immunodeficiency Virus Type 2 (HIV-2) Isolates Infect CD4-Negative Cells via CCR5 and CXCR4: Comparison with HIV-1 and Simian Immunodeficiency Virus and Relevance to Cell Tropism In Vivo. *J. Virol.* **73,** 7795–7804 (1999).
203. Xiao, P. *et al.* Characterization of a CD4-independent clinical HIV-1 that can efficiently infect human hepatocytes through chemokine (C-X-C motif) receptor 4: *AIDS* **22,** 1749–1757 (2008).
204. Dumonceaux, J. *et al.* Spontaneous Mutations in the env Gene of the Human Immunodeficiency Virus Type 1 NDK Isolate Are Associated with a CD4-Independent Entry Phenotype. *J. Virol.* **72,** 512–519 (1998).
205. Hoffman, T. L. *et al.* Stable exposure of the coreceptor-binding site in a CD4-independent HIV-1 envelope protein. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 6359–6364 (1999).
206. Kolchinsky, P. *et al.* Adaptation of a CCR5-Using, Primary Human Immunodeficiency Virus Type 1 Isolate for CD4-Independent Replication. *J. Virol.* **73,** 8120–8126 (1999).
207. Olshevsky, U. *et al.* Identification of individual human immunodeficiency virus type 1 gp120 amino acids important for CD4 receptor binding. *J. Virol.* **64,** 5701–5707 (1990).
208. LaBranche, C. C. *et al.* Determinants of CD4 Independence for a Human Immunodeficiency Virus Type 1 Variant Map outside Regions Required for Coreceptor Specificity. *J. Virol.* **73,** 10310–10319 (1999).
209. Douagi, I. *et al.* Influence of Novel CD4 Binding-Defective HIV-1 Envelope Glycoprotein Immunogens on Neutralizing Antibody and T-Cell Responses in Nonhuman Primates. *J. Virol.* **84,** 1683–1695 (2010).
210. Cordonnier, A., Montagnier, L. & Emerman, M. Single amino-acid changes in HIV envelope affect viral tropism and receptor binding. *Nature* **340,** 571–574 (1989).
211. Kneissl, S. *et al.* Measles Virus Glycoprotein-Based Lentiviral Targeting Vectors That Avoid Neutralizing Antibodies. *PLOS ONE* **7,** e46667 (2012).
212. Bender, R. R. *et al.* Receptor-Targeted Nipah Virus Glycoproteins Improve Cell-Type Selective Gene Delivery and Reveal a Preference for Membrane-Proximal Cell Attachment. *PLOS Pathog.* **12,** e1005641 (2016).
213. Zolla-Pazner, S. & Cardozo, T. STRUCTURE–FUNCTION RELATIONSHIPS OF HIV-1 ENVELOPE SEQUENCE-VARIABLE REGIONS PROVIDE A PARADIGM FOR VACCINE DESIGN. *Nat. Rev. Immunol.* **10,** 527–535 (2010).
214. Curlin, M. E. *et al.* HIV-1 Envelope Subregion Length Variation during Disease Progression. *PLOS Pathog.* **6,** e1001228 (2010).
215. Huang, C. *et al.* Structure of a V3-Containing HIV-1 gp120 Core. *Science* **310,** 1025–1028 (2005).
216. Yuan, T., Li, J. & Zhang, M.-Y. HIV-1 Envelope Glycoprotein Variable Loops Are Indispensable for Envelope Structural Integrity and Virus Entry. *PLOS ONE* **8,** e69789 (2013).
217. Wyatt, R. *et al.* Functional and immunologic characterization of human immunodeficiency virus type 1 envelope glycoproteins containing deletions of the major variable regions. *J. Virol.* **67,** 4557–4565 (1993).
218. Julien, J.-P. *et al.* Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science* **342,** (2013).
219. Lyumkis, D. *et al.* Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 Env trimer. *Science* **342,** 1484–1490 (2013).
220. Pancera, M. *et al.* Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature* **514,** 455–461 (2014).

221. Ren, X., Sodroski, J. & Yang, X. An Unrelated Monoclonal Antibody Neutralizes Human Immunodeficiency Virus Type 1 by Binding to an Artificial Epitope Engineered in a Functionally Neutral Region of the Viral Envelope Glycoproteins. *J. Virol.* **79,** 5616–5624 (2005).
222. Yang, X., Lipchina, I., Cocklin, S., Chaiken, I. & Sodroski, J. Antibody Binding Is a Dominant Determinant of the Efficiency of Human Immunodeficiency Virus Type 1 Neutralization. *J. Virol.* **80,** 11404–11408 (2006).
223. Wallace, A. & Stamatatos, L. Introduction of Exogenous Epitopes in the Variable Regions of the Human Immunodeficiency Virus Type 1 Envelope Glycoprotein: Effect on Viral Infectivity and the Neutralization Phenotype. *J. Virol.* **83,** 7883–7893 (2009).
224. Nakane, S., Iwamoto, A. & Matsuda, Z. The V4 and V5 Variable Loops of HIV-1 Envelope Glycoprotein Are Tolerant to Insertion of Green Fluorescent Protein and Are Useful Targets for Labeling. *J. Biol. Chem.* **290,** 15279–15291 (2015).
225. Chen, X., Zaro, J. & Shen, W.-C. Fusion Protein Linkers: Property, Design and Functionality. *Adv. Drug Deliv. Rev.* **65,** 1357–1369 (2013).
226. Reddy Chichili, V. P., Kumar, V. & Sivaraman, J. Linkers in the structural biology of protein–protein interactions. *Protein Sci. Publ. Protein Soc.* **22,** 153–167 (2013).
227. Simon-Loriere, E. *et al.* Molecular Mechanisms of Recombination Restriction in the Envelope Gene of the Human Immunodeficiency Virus. *PLoS Pathog* **5,** e1000418 (2009).
228. Coulibaly, S. T. *et al.* Potent Sensitisation of Cancer Cells to Anticancer Drugs by a Quadruple Mutant of the Human Deoxycytidine Kinase. *PLoS ONE* **10,** (2015).
229. Jordheim, L. P., Galmarini, C. M. & Dumontet, C. Gemcitabine resistance due to deoxycytidine kinase deficiency can be reverted by fruitfly deoxynucleoside kinase, DmdNK, in human uterine sarcoma cells. *Cancer Chemother. Pharmacol.* **58,** 547–554 (2006).
230. Abram, M. E., Ferris, A. L., Shao, W., Alvord, W. G. & Hughes, S. H. Nature, Position, and Frequency of Mutations Made in a Single Cycle of HIV-1 Replication. *J. Virol.* **84,** 9864–9878 (2010).
231. Hütter, G. *et al.* Long-Term Control of HIV by CCR5 Delta32/Delta32 Stem-Cell Transplantation. *N. Engl. J. Med.* **360,** 692–698 (2009).
232. Anderson, J. S., Javien, J., Nolta, J. A. & Bauer, G. Preintegration HIV-1 inhibition by a combination lentiviral vector containing a chimeric TRIM5 alpha protein, a CCR5 shRNA, and a TAR decoy. *Mol. Ther. J. Am. Soc. Gene Ther.* **17,** 2103–2114 (2009).
233. Liang, M. *et al.* Inhibition of HIV-1 infection by a unique short hairpin RNA to chemokine receptor 5 delivered into macrophages through hematopoietic progenitor cell transduction. *J. Gene Med.* **12,** 255–265 (2010).
234. Perez, E. E. *et al.* Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat. Biotechnol.* **26,** 808–816 (2008).
235. Holt, N. *et al.* Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. *Nat. Biotechnol.* **28,** 839–847 (2010).
236. Hapke, D. M., Stegmann, A. P. A. & Mitchell, B. S. Retroviral Transfer of Deoxycytidine Kinase into Tumor Cell Lines Enhances Nucleoside Toxicity. *Cancer Res.* **56,** 2343–2347 (1996).
237. Pleshkan, V. V., Alekseenko, I. V., Zinovyeva, M. V., Vinogradova, T. V. & Sverdlov, E. D. Promoters with Cancer Cell-Specific Activity for Melanoma Gene Therapy. *Acta Naturae* **3,** 13–21 (2011).
238. Wlodawer, A., Minor, W., Dauter, Z. & Jaskolski, M. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J.* **275,** 1–21 (2008).
239. Palamini, M., Canciani, A. & Forneris, F. Identifying and Visualizing Macromolecular Flexibility in Structural Biology. *Front. Mol. Biosci.* **3,** (2016).
240. Sabini, E., Hazra, S., Ort, S., Konrad, M. & Lavie, A. Structural basis for substrate promiscuity of dCK. *J. Mol. Biol.* **378,** 607–621 (2008).
241. Garnett, C., Apperley, J. F. & Pavlů, J. Treatment and management of graft-versus-host disease: improving response and survival. *Ther. Adv. Hematol.* **4,** 366–378 (2013).

242. Chatterjee, S., Azad, B. B. & Nimmagadda, S. The Intricate Role of CXCR4 in Cancer. *Adv. Cancer Res.* **124,** 31–82 (2014).
243. Müller, A. *et al.* Involvement of chemokine receptors in breast cancer metastasis. *Nature* **410,** 50–56 (2001).
244. Balkwill, F. The significance of cancer cell expression of the chemokine receptor CXCR4. *Semin. Cancer Biol.* **14,** 171–179 (2004).
245. Darash-Yahana, M. *et al.* Role of high expression levels of CXCR4 in tumor growth, vascularization, and metastasis. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **18,** 1240–1242 (2004).
246. Park, B.-W. *et al.* Rationally designed anti-HER2/neu peptide mimetic disables P185HER2/neu tyrosine kinases in vitro and in vivo. *Nat. Biotechnol.* **18,** 194–198 (2000).
247. Hortobagyi, G. N. Trastuzumab in the Treatment of Breast Cancer. *N. Engl. J. Med.* **353,** 1734–1736 (2005).
248. Pacheco, B., Finzi, A., Stremlau, M. & Sodroski, J. Adaptation of HIV-1 to Cells Expressing Rhesus Monkey TRIM5α. *Virology* **408,** 204–212 (2010).