



**HAL**  
open science

# Nouvelle génération d'AAV : un mélange de chimie organique et de vectorologie

Pierre-Alban Lalys

► **To cite this version:**

Pierre-Alban Lalys. Nouvelle génération d'AAV : un mélange de chimie organique et de vectorologie. Chimie organique. Nantes Université, 2023. Français. NNT : 2023NANU4048 . tel-04612356

**HAL Id: tel-04612356**

**<https://theses.hal.science/tel-04612356>**

Submitted on 14 Jun 2024

**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.

# THESE DE DOCTORAT

NANTES UNIVERSITE

ECOLE DOCTORALE N° 596

*Matière, Molécules, Matériaux et Géosciences*

Spécialité : Chimie moléculaire et Macromoléculaire

Par

**Pierre-Alban Lalys**

**Nouvelle génération d'AAV : un mélange de chimie organique et de vectorologie**

Thèse présentée et soutenue à Nantes, le 21 juin 2023

Unité de recherche : Laboratoire CEISAM, CNRS UMR 6230, 2 rue de la Houssinière, Nantes

Laboratoire TaRGeT, INSERM UMR 1089, 22 boulevard Benoni-Goullin, Nantes

## Rapporteurs avant soutenance :

Dr. Sébastien VIDAL  
Dr. Nicolas FERRY

Directeur de recherche CNRS, Université Paris-Saclay  
Directeur de recherche INSERM

## Composition du Jury :

Examineurs : Dr. Valérie DESVERGNES  
Pr. Emilie ALLARD-VANNIER

Chargée de recherche CNRS, Université de Bordeaux  
Professeure des Universités, Université de Tours

Dir. de thèse : Pr. David DENIAUD  
Co-dir. de thèse : Dr. Mathieu MEVEL

Professeur des Universités, Nantes Université  
Chercheur, Nantes Université

*Pour ma grand-mère Maminou*



# Remerciements

---

Débutons par des remerciements formels, Tout d'abord, je souhaite remercier l'Université de Nantes d'avoir financé les travaux décrits dans ce manuscrit. Je tiens aussi à remercier Jean-Michel Bouler et Oumeya Adjali pour m'avoir accueilli dans leur laboratoire respectif, le laboratoire CEISAM UMR CNRS 6230 et le laboratoire TaRGeT, UMR INSERM 1089. Je remercie aussi mes encadrants, David et Mathieu de m'avoir choisi et fait confiance, et ce depuis le master, pour réaliser ce projet à l'interface chimie/biologie. Ainsi que pour leurs conseils et leurs aides tout au long de cette thèse non seulement dans les diverses présentations que j'ai pu réaliser mais surtout dans la rédaction qui s'est avérée être une étape difficile... Merci au Jury, Sébastien VIDAL, Nicolas FERRY, Valérie DESVERGNES et Emilie ALLARD-VANNIER pour avoir accepté de rapporter ce travail, et pour votre enthousiasme lors de la soutenance. Je remercie Muriel Pipelier, et Nicolas boisgerault pour leur participation aux différents comités de suivi de thèse.

Merci à tous les collaborateurs :

Audrey et Caroline pour leur expertise concernant l'étude *in vivo* ainsi que pour la relecture de cette partie durant l'écriture de mon manuscrit. Thibaut Larcher pour sa disponibilité et ces analyses histologiques qui nous ont permis d'avancer grandement sur le projet, en plus d'avoir de superbes images à présenter. L'entreprise Glycodiag<sup>®</sup> pour leur savoir-faire dans l'obtention de mesure d'affinité entre une lectine et un ligand. La plateforme PAC pour l'extraction et le dosage d'ADN et d'ARN ainsi que pour la quantification de la GFP (merci Karine et Tiphaine). Merci au CPV pour la préparation des vecteurs. Et pour finir, Merci au laboratoire MICHELET pour la préparation des différentes solutions tampons et autres.

Dimitri Alvarez Dorta et Mohammed qui m'ont appris des techniques mais également des connaissances que j'ai eu plaisir à réaliser et développer au laboratoire. Je tiens énormément à les remercier ici, Dimitri pour toute nos discussions de chimie organique et Mohammed pour ton calme, ta patience et pour ton excellente habileté technique. Merci à eux.

Enfin merci à Chloé et Héloïse vous avez été d'excellentes stagiaires et je suis très fier d'avoir pu vous apprendre le peu de choses que je connais.

Je tiens également à remercier les personnes suivantes qui ont contribué au bon déroulement de cette thèse :

Julie, Laurence et Isabelle, pour leurs analyses et le sourire qu'elles affichent en permanence. Virginie, Benoît Charrier et Aurélie BERNARD pour avoir géré la RMN malgré les nombreuses pannes de

compresseur durant ces trois dernières années ainsi que nos bêtises courantes. Julien et Stéphane pour tous les problèmes informatiques. Les secrétaires, en particulier Nathalie qui m'aura sauvé la mise plus d'une fois, Karine pour sa maîtrise dans le domaine de la gestion des dépenses doctorales en congrès et Anita pour sa capacité à retrouver une feuille de papier d'importance crucial au milieu de 1000 de ses congénère. Aurélien qui s'occupe des déchets et toutes les commandes du CEISAM, le tout dans la bonne humeur ainsi que Yann et Errol pour les solvants.

Enfin, un grand merci également à toutes les personnes du laboratoire TaRGeT et CEISAM.

Continuons, je vais adresser des remerciements un peu moins formel, mais en version « thèse publique ». Pour tous ceux et celles qui lisent ces lignes, il existe des remerciements non-formels et plus personnels, je vous invite à me les demander.

Merci aux copains de biologie Juliette, Malo, Johanna, Guigui, Elise, Clémence, Rebecca, Thomas Sénéchal pour les soirées, les fous rires et les très peu « solution tampon » que l'on a pu boire ensemble.

Merci aux copains de Chimie : Paul, Louis, Guillaume, Johan, Jimmy, Krystal, Ehu ... la liste est longue et je suis désoler d'en oublier.

Merci aussi à tous les autres membres de l'équipe CORAIL : Mireia, Erwan, Vincent, Françoise, Isabelle, Pierrick, Elvina et François-Xavier pour m'avoir accueillie dans l'équipe et ou partagé les pauses café du vendredi ou dans le couloir.

Merci également aux copains des « foot du jeudi » : Baptiste, Louis, Thibault, David, Hichem, Hamissou, Alexis, Charlotte, Clément, Judicalel, Thamer, Federico ... La liste est longue merci à tous et toutes !

Merci à la superbe équipe DREAM TEAM DDe, Denis, Sébastien (gouin), Roxanne, Mélyne la petite nouvelle, Karine, Valérie, Yoan, Aurélien, Christophe, Elodie et tous les stagiaires que nous avons pu avoir. Une très belle équipe très bien organisé que je recommande !

Un plus grand merci Marianne, Maia, Coralie (sapin), Mathieu pour votre amitié fidèle, pour tous ces moments passé ensemble et tout ce que l'on a pu se dire, je ne doute pas que l'on se reverra et je suis très fier de vous compter parmi mes amis.

Un grand merci à ma famille, papa, maman, ma sœur, pierre, tonton, tata, Valérie et Fabrice et à tout le reste de ma grande et belle famille pour leur soutien indéfectible. Je vous aime. J'écris ces lignes après l'oral et vous voir réunie dans cet amphithéâtre a été le jour le plus beau de toute ma vie.

Enfin, merci à Sébastien Depienne, parce que c'est un chic type.

Et pour finir, je vais paraphraser Mathieu Scalabrini et finalement me remercier un peu, aussi, pour se travaille déjà, mais également pour avoir continué naïvement à faire mon petit bonhomme de chemin dans les études et avoir pu rencontrer des gens incroyables alors que dans les premières années de ma vie personne n'aurait parié sur moi. Ne doutez jamais de vous et si vous voulez faire une thèse faite là car c'est une expérience qui restera à jamais dans votre mémoire.

Je clôture ici les derniers mots de ma thèse et me souhaite également bon courage pour cette nouvelle aventure en Nouvelle-Zélande !

# Abréviation

---

AA : Acides Aminés	kDa : kilo-dalton
AAV : Adeno Associated Virus	MOI : Multiplicity Of Infection
rAAV : Adeno Associated Virus recombinant	NAbs : Anticorps Neutralisants
ADN : Acide Désoxyribonucléique	NHEJ : Non-Homologous End-Joining
ARN : Acide Ribonucléique	OMS : Organisation Mondiale de la Santé
ASGP : Asialoglycoprotéine	PAGES : PolyAcrylamide Gel Electrophoresis
ASGPr : Asialoglycoprotéine recepteur	PCR : Polymerase Chain Reaction
Cap : Capside	PTAD : PhénylTriAzolineDione
CAR-T cells : Chimeric Antigen Receptor	qPCR : quantitative Polymerase Chain Reaction
CRD : Carbohydrate Recognition Domain	PEG : PolyEthylene Glycol
CMV : Promoteur du Cytomégalovirus	PDB : Protein Data Bank
CRIPSR: Clustered Regularly Interspaced Short Palindromic Repeats	Py : Pyridine
DCM : Dichlorométhane	Rep : Réplicase
DICS : Déficit Immunitaire Combiné Sévère	RMN : Résonance Magnétique Nucléaire
DMAP : 4-diméthylaminopyridine	RT : Room Temperature
DMF : N,N-Diméthylformamide	SBA : SoyBean Agglutinin
DMSO : Diméthylsulfoxyde	SDS : Sodium Dodecyl Sulphate
Eq : Equivalent	SARS-CoV-2 : Severe Acute Respiratory Syndrome Coronavirus 2
FACS : Fluorescence Activated Cell Sorting	SV40 : Virus Simien 40
FITC : Fluorescéine d'IsoThioCyanate	TCO : Trans-CycloOctène
GalNAc : <i>N</i> -acétylgalactosamine	TEA : Triéthylamine
GFP : Green Fluorescent Protein	TIL : Lymphocytes d'Infiltration Tumoral
HDR : Homology-Directed Repair	Tris : Tris(hydroxyméthyl)aminomethane
HEK293 : Human Embryonic Kidney cell	VIH : Virus de l'Immunodéficience Humaine
HRMS : High-resolution mass spectra	VP : Viral Protein
HSPG : Heparan Sulfate ProteoGlycan	VR : Variable Regions
IEDDA : Diels-Alder à Demande Inverse	WB : Western Blot
ITR : Inverted Terminal Repeat	WFA : Wisteria Floribunda
kb : kilo-bases	





# Table des matières

---

<b>Remerciements</b> .....	3
<b>Abréviation</b> .....	6
Table des matières .....	8
<b>Introduction</b> .....	12
Contexte .....	14
1. La Thérapie génique.....	16
A. Maladies génétiques.....	16
B. Histoire et principe .....	17
C. Les différents vecteurs pour la thérapie génique.....	24
2. Les virus adéno-associés .....	32
A. L'utilisation des AVV comme vecteurs en thérapie génique .....	32
B. Les limites des AAVr.....	42
C. Optimisation des vecteurs .....	46
3. Optimisation chimique des AVV .....	51
A. Principe .....	51
B. Bioconjugaison de capsid d'AAVr .....	54
C. Travaux antérieurs à la thèse.....	57
4. Ciblage du foie .....	61
A. Le foie.....	61
B. Les hépatocytes .....	64
C. Récepteurs ASGP .....	65
D. Mais quel ligand utiliser ? .....	67
5. Etude préliminaire et objectifs de la thèse.....	70
Synthèse des ligands .....	74
1. Introduction .....	76
2. Synthèse du ligand PMS-02.....	78
A. Synthèse du motif bicyclique.....	78
B. Introduction de l'espaceur et synthèse du ligand final. ....	88
3. Modélisation .....	91
A. Introduction .....	91
B. Chimiothèque et docking.....	91
C. Résultats du docking.....	93
D. Mode d'interaction et optimisation .....	96
E. Conclusion.....	98
4. Synthèse du composé PMS-02bis.....	99
A. Voie A.....	104

B.	Voie B .....	108
A.	Voie C .....	111
B.	Voie D .....	116
5.	Synthèse des ligands PMS-03 et PMS-04 .....	122
6.	Conclusion .....	129
Bioconjugaison, caractérisations des vecteurs chimiquement modifiés et étude <i>in vivo</i> .....		132
1.	Introduction .....	134
2.	Couplage chimique et caractérisations des AAV .....	135
A.	Bioconjugaison des différents ligands .....	135
B.	Caractérisations des nouveaux vecteurs modifiés chimiquement .....	137
C.	Reconnaissance du vecteur .....	145
3.	Infectiosité des vecteurs modifiés <i>in vitro</i> .....	151
A.	Lignée cellulaire non hépatocytaire .....	151
B.	Lignées cellulaires hépatocytaires .....	152
4.	Etude <i>in vivo</i> des nouveaux vecteurs chimiquement modifiés. ....	156
A.	Quantification des génomes viraux dans le foie des souris injectées avec les différents vecteurs AAV2 .....	157
B.	Quantification de l'ARN messager GFP dans le foie des souris injectées avec les différents vecteurs AAV2 .....	159
C.	Evaluation de l'expression de la protéine GFP dans le foie des souris injectées avec les différents vecteurs AAV2 .....	162
D.	Quantification de l'ARN messager GFP dans les organes non ciblés des souris injectées avec les différents vecteurs AAV2 .....	169
5.	Conclusion .....	172
Conclusion et Perspective .....		176
Conclusion .....		178
Perspectives : .....		180
A.	Axe 1 : synthèse organique .....	180
B.	Axe 2 : caractérisations .....	184
C.	Axe 3 : étude biologique .....	185
Materials and methods .....		188
1.	Chemistry .....	190
A.	Peg linker .....	191
B.	Ligand PMS-02 .....	192
C.	Ligand PMS-02bis .....	205
D.	Ligand PMS-03 .....	217
E.	Ligand PMS-04 .....	222
F.	Molecular modeling .....	226
2.	Biology .....	227
A.	AAV2 production and purification .....	227
B.	Coupling and purification .....	227

C.	Titration of AAV vector genomes.....	228
D.	Western blot and silver staining .....	228
E.	Dot blot .....	229
F.	Transduction of HEK cells HEK293, HUH7 and HEPG2.....	229
G.	Transduction of primary mouse hepatocytes.....	229
H.	Glycodiag© experiment.....	230
I.	Animal care and welfare .....	230
J.	Intravenous injection .....	230
K.	Absolute quantification of vector genomes by qPCR in mice tissue samples .....	231
L.	Relative quantification of GFP expression by RT-qPCR in mice tissue samples .....	231
M.	Western blot of GFP expression .....	232
N.	Immunohistochemistry GFP staining in mice .....	233
	Bibliographie :.....	234
	Annexes .....	249



# Introduction

---

Hurair Vasken Aposhian déclarait en 1969 que « Si le rôle d'un médicament est de restaurer la fonction d'un processus physiologique particulier, alors l'ADN devrait être considéré comme le médicament ultime ». Depuis cette déclaration, la thérapie génique est devenue en quelques années un outil thérapeutique incontournable avec plus de 3000 essais cliniques réalisés ou en cours. L'optimisation des vecteurs pour le transfert d'ADN ou ARN ainsi que l'essor des techniques d'édition de gène, récompensé en 2020 par le prix Nobel de chimie, ont permis à la thérapie génique de passer de la théorie à une réalité clinique.

Parmi les vecteurs couramment utilisés, un petit virus de la famille des *Parvoviridae*, le virus adéno-associé ou AAV, a su trouver sa place comme vecteur de choix. Ses caractéristiques intrinsèques telles que sa faible immunogénicité et sa capacité à transduire un large panel de cellules ont mené au succès de nombreux essais de thérapie génique et à la mise sur le marché de six médicaments dont trois commercialisés en 2022.

Néanmoins, malgré cela, le transfert de gène à l'aide d'AAV présente encore de nombreuses limites. Récemment quatre jeunes patients sont décédés des suites d'un essai clinique de thérapie génique à base d'AAV pour le traitement de la myopathie tubulaire liée au chromosome X.

Améliorer l'efficacité des vecteurs est une priorité aujourd'hui pour réduire les doses et en conséquence, réduire les effets indésirables, les quantités à produire et les coûts.

Les principales techniques pour optimiser les AAV font intervenir des procédés de génie génétique visant à concevoir de nouvelles capsides, moins immunogènes et avec un meilleur ciblage de l'organe désiré. Notre équipe, des laboratoires CEISAM (UMR CNRS 6230) et TaRGeT (UMR INSERM 1089), a envisagé l'utilisation de la bioconjugaison comme alternative à l'ingénierie biologique pour concevoir de nouvelles capsides chimiquement modifiées. Ainsi, des ligands seront greffés par couplage covalent sur la capside de l'AAV afin de le protéger des anticorps circulants et de le guider vers les cellules à cibler. C'est dans ce contexte que s'inscrit ma thèse, utiliser la chimie organique pour développer une nouvelle génération d'AAV fonctionnalisés avec un ciblage plus efficace du foie.

Après une partie introductive, ce manuscrit décrira la synthèse de nouvelles molécules permettant de cibler le récepteur aux asialoglycoprotéines, protéines présentes à la surface des hépatocytes, ainsi que la caractérisation et l'étude des propriétés biologiques *in vitro* et *in vivo* des vecteurs chimiquement modifiés.



# Chapitre 1

## Contexte





# 1. La Thérapie génique

## A. Maladies génétiques

En Avril 2022 était publié une version complète de notre code génétique dans le journal Science.<sup>1</sup> Les auteurs ont notamment montré que le génome humain est composé d'environ 3 milliards de paires de bases nucléiques, présentes dans 19 969 gènes.<sup>1</sup> L'importante variabilité de ces gènes explique notre grande diversité. Ils sont le mode d'emploi qu'utilise la cellule pour produire des protéines qui assureront une multitude de fonctions au sein de l'organisme comme la glycolyse, la respiration cellulaire, l'immunité ou encore la production d'hormones et d'autre molécules signalétiques.

Chez certains individus, ce mode d'emploi est erroné ce qui perturbe fortement l'ingénierie des cellules et peut provoquer un dysfonctionnement au sein de l'organisme et induire une maladie génétique.

Aujourd'hui, entre 8 000 et 10 000 maladies génétiques sont répertoriées. La plupart d'entre elles sont rares (moins d'une personne sur 2 000) et on y retrouve aussi bien des affections peu invalidantes comme le daltonisme, qu'extrêmement graves comme l'amyotrophie spinale ou la mucoviscidose.<sup>2</sup>

L'OMS estime qu'environ 7,9 millions de nouveau-nés dans le monde sont atteints d'une anomalie génétique. En France, cela représente environ 3% des naissances. La drépanocytose est la maladie génétique la plus fréquente avec, chaque année, environ 310 000 cas dans le monde et 500 en France.<sup>3</sup>

Avec l'avancée de la recherche médicale, la liste des maladies génétiques s'allonge. Par exemple, en juin 2022, des chercheurs des Universités de Portsmouth, Southampton et Copenhague ont identifié le gène *grin1* qui aide à déplacer les signaux électriques autour du cerveau. Cette découverte leur a permis de diagnostiquer une maladie génétique provoquant un retard du développement intellectuel et des difficultés d'apprentissage.<sup>4</sup>

La nature génétique des maladies rend complexe l'élaboration d'un traitement curatif et les prises en charge actuelles concernent uniquement les symptômes. Néanmoins, grâce aux avancées récentes sur la compréhension du génome, les chercheurs ont maintenant à leur disposition des outils permettant de le modifier durablement et ainsi de soigner diverses pathologies. Cette nouvelle branche de la médecine est appelée thérapie génique.

## B. Histoire et principe

L'acide désoxyribonucléique (ADN) est une macromolécule biologique présente dans de nombreuses cellules ainsi que dans la plupart des virus. Isolé la première fois à la fin du 19<sup>ème</sup> siècle par Friedrich Miescher, l'ADN est un polymère de nucléotides composé d'un groupement phosphate, d'un sucre (le désoxyribose), et d'une base nucléique (adénine, guanine, thymine ou cytosine). Ainsi, il existe 4 nucléotides différents qui composent l'ensemble de l'ADN présent dans un organisme vivant. Les différentes séquences de cet ADN se répartissent dans l'organisme sous forme de gènes qui permettent par la suite l'expression d'une ou plusieurs protéines. L'ensemble de ces gènes est présent sous forme de chromosomes dans le noyau des cellules et constitue l'information génétique d'un individu, encore appelé code génétique. Il permet de fabriquer l'ensemble des protéines de l'organisme pour le bon fonctionnement des cellules (figure 1).<sup>5</sup>

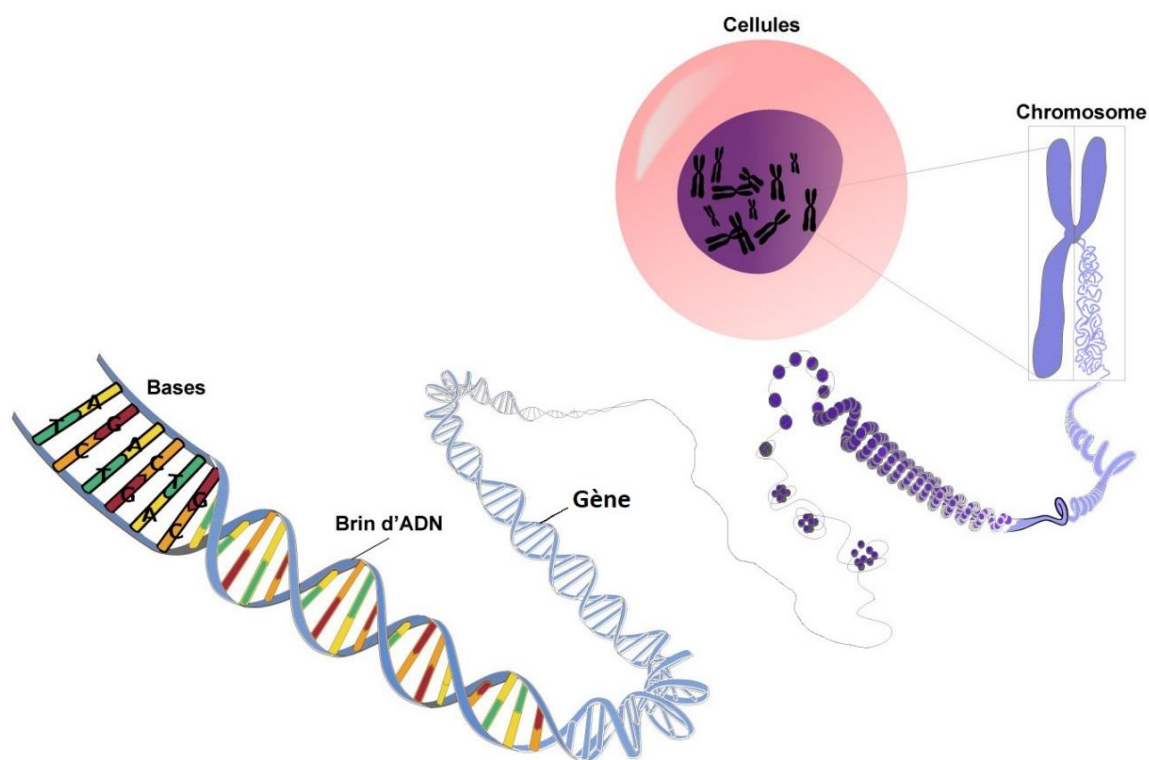


Figure 1 : Les chromosomes, formés des gènes, eux-mêmes formés de l'ADN, centre de l'information génétique.

Aujourd'hui, la thérapie génique est définie comme une stratégie thérapeutique visant à utiliser une séquence d'ADN ou d'ARN comme médicament.

### a) Histoire :

Dans les années 1950, Zinder et Lederberg montrent la possibilité de transférer de l'ADN d'une souche bactérienne à une autre en utilisant un phage (virus infectant exclusivement les bactéries). Ce procédé de transfert d'ADN par les virus dans les organismes hôtes qu'ils infectent est nommé transduction.<sup>6</sup>

Le premier transfert d'un gène dans des lignées cellulaires humaines est réalisé en 1962 par Szybalsky. Il a extrait le gène *hgpert* de cellules et l'a placé en présence d'autres cellules ne l'exprimant pas. Une fois isolées, ces cellules sont alors devenues capables de se développer dans un milieu spécifique nécessitant l'enzyme HGPRT ce qui montre qu'il y a eu transfert du gène *hgpert* entre deux lignées cellulaires différentes.<sup>7</sup>

C'est en juillet 1966, dans la publication du Dr Campbell "Reflections on Research and the Future of Medicine", et dans "the new genetic" (1967) d'Edward Tatum, que le concept de transfert de gènes de cellules humaines à des fins thérapeutiques a été évoqué pour la première fois.<sup>8</sup>

Le premier essai clinique de thérapie génique, dirigé par l'équipe du Professeur Rosenberg, a vu le jour dans les années 1990 afin de traiter des patients atteints de mélanomes avancés. Son travail a consisté à extraire des lymphocytes infiltrant des tumeurs (TIL) de patients, puis à les transduire *via* un rétrovirus contenant le gène du facteur de nécrose tumorale (TNF). Ce gène permet la production d'une protéine, le TNF $\alpha$ , qui induit la nécrose et/ou l'apoptose des cellules cancéreuses. Ensuite les TIL sont réinjectés aux patients puis se regroupent au niveau des tumeurs, les TNF $\alpha$  permettent alors de réduire l'activité cancéreuse. Cette étude a permis à l'équipe du Professeur Rosenberg d'ouvrir la voie pour le développement de cette nouvelle thérapie.<sup>9,10</sup>

Fort de cette réussite, les essais cliniques de thérapie génique ont explosé, dépassant les 480 essais entre 1990 et 1999,<sup>11</sup> date à laquelle la thérapie génique subit son premier revers avec le décès d'un patient lors d'un essai clinique aux Etats-Unis.<sup>12</sup> Par la suite, une plus grande attention fut portée aux possibles réactions immunitaires. Le nombre d'essais cliniques approuvés repartira finalement à la hausse dans les années 2000 après un succès très médiatisé : le traitement des bébés bulles à l'hôpital Necker de Paris. L'équipe des Professeurs Alain Fisher, Marina Cavazzana-Calvo et Salima Hacein-Bey-Abina s'intéresse au déficit immunitaire combiné sévère (DICS). Cette maladie génétique rare est caractérisée par une anomalie de certains lymphocytes qui prive les patients de système immunitaire, ils doivent alors évoluer dans un environnement stérile. Une copie saine du gène responsable de la maladie a été transduite dans des cellules CD34+ grâce à un vecteur rétroviral (c'est à dire que ces cellules sont ensuite capables de générer des lymphocytes corrigés et donc viables). Ces cellules ont ensuite été transfusées chez les patients et après une période de 10 mois les chercheurs ont pu détecter des lymphocytes fonctionnels. Cela a permis d'améliorer la vie de 7 enfants en leur rendant un système

immunitaire compétent. Cette étude très médiatisée a montré l'efficacité de cette thérapie au monde entier.<sup>13</sup>

Aujourd'hui, la thérapie génique représente plus de 3 000 essais cliniques, avec une petite vingtaine de médicaments sur le marché. Ainsi, c'est une stratégie thérapeutique à fort potentiel mais avec néanmoins de nombreuses barrières à lever pour augmenter de manière considérable les autorisations de mises sur le marché de nouveaux médicaments.<sup>14</sup>

### *b) Fonctionnement de la thérapie génique*

La thérapie génique est basée sur la notion de transfert de gènes qui consiste à apporter un acide nucléique exogène (ou transgène) dans une cellule porteuse d'une anomalie génétique. Ce transfert de gènes a pour objectif de permettre la correction de l'anomalie génétique et ainsi de rétablir une fonction cellulaire normale. Le transgène étant une macromolécule polyanionique, il lui est difficile de traverser les membranes cellulaires elles-mêmes chargées négativement et encore moins d'évoluer librement dans l'organisme du patient. C'est pourquoi le transgène est régulièrement vectorisé, c'est-à-dire introduit dans un transporteur, qui lui permettra de franchir les barrières biologiques sans dégradation. Le transporteur, aussi appelé vecteur, peut-être de différentes natures, ce point sera abordé plus en détail dans la suite du manuscrit.

La thérapie génique comprend un certain nombre d'étapes clés qui sont (i) d'identifier le gène thérapeutique, (ii) d'élaborer le produit de thérapie génique, (iii) de le vectoriser, (iv) de l'acheminer aux cellules cibles et (v) d'être pris en charge par la cellule (figure 2).



*Figure 2 : Les étapes clés de la thérapie génique.<sup>15</sup>*

Ce sont ces étapes qui doivent être optimisées au maximum par les chercheurs afin d'obtenir des produits de thérapie efficaces et utilisables comme candidat médicament dans les essais cliniques.

A l'heure actuelle, de nombreuses approches sont utilisées en thérapie génique. Il est possible de les diviser en 3 grandes catégories (figure 3).

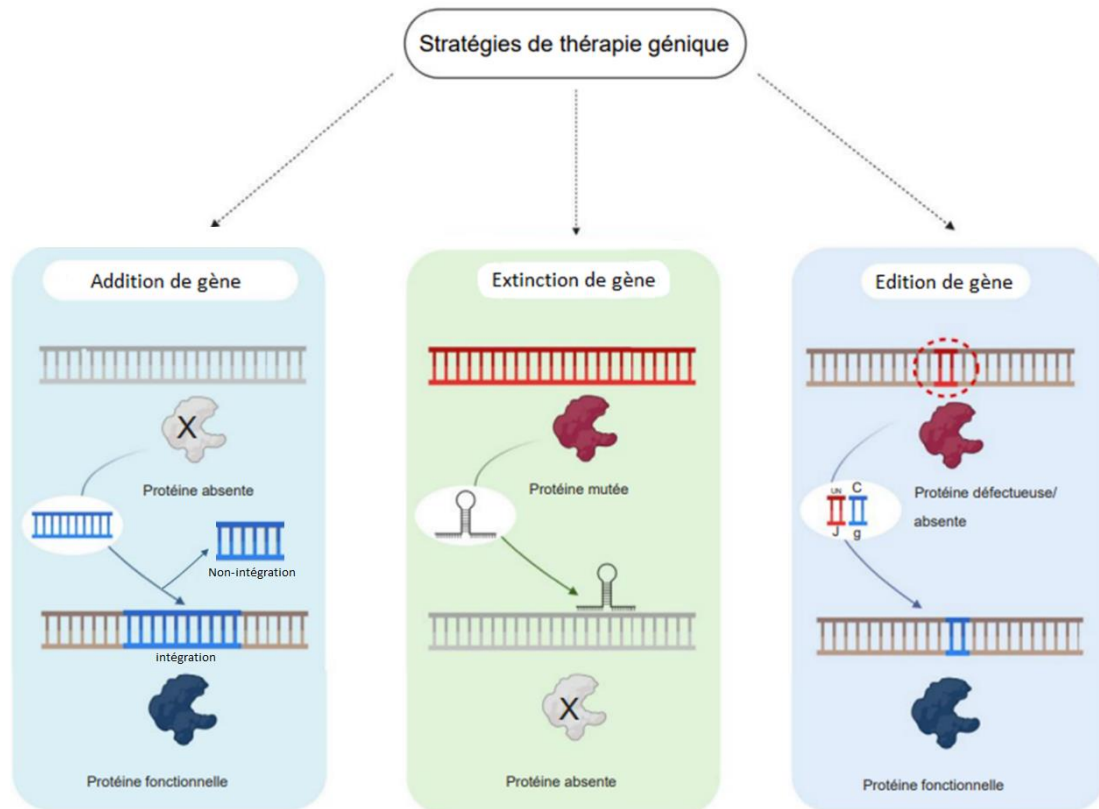


Figure 3 : Les différentes stratégies de thérapie génique<sup>16</sup>

(1) Addition de gène

Cette technique consiste à ajouter du matériel génétique, codant pour une protéine thérapeutique, dans les cellules cibles du patient. Ce matériel génétique peut être une version fonctionnelle d'un gène muté ou absent provoquant une maladie, ou bien un gène codant pour une protéine exogène par exemple un antigène de virus dans le cas de la vaccination. Le gène thérapeutique vient soit s'intégrer dans le génome du patient, soit s'ajouter sous forme circulaire extra chromosomique (épisode) sans remplacer le gène muté. La cellule produira alors la protéine thérapeutique pour restaurer la fonction biologique défectueuse. La protéine thérapeutique peut aussi être un facteur de nécrose tumorale comme, par exemple, dans le traitement de cancer.<sup>17</sup>

(2) L'extinction de gène

Parfois, un gène défectueux peut produire des protéines toxiques causant des maladies ou être sur activé et favoriser une croissance cellulaire incontrôlée, entraînant des maladies telles que le cancer. La thérapie génique d'extinction de gènes agit pour les inhiber en empêchant la production des

protéines toxiques. Selon l'approche thérapeutique, cela peut impliquer l'arrêt complet du gène ou la réduction de ses capacités, pour permettre à la cellule de recommencer à fonctionner correctement.<sup>18</sup>

Les ARN interférent régulent l'expression des protéines en venant interagir avec l'ARN messager. Une des techniques d'extinction de gènes est d'utiliser ce processus naturel, à partir d'ARN synthétique qui après interférence avec l'ARN messager conduira à sa dégradation et donc à une inhibition de sa traduction en protéine. D'autres techniques peuvent être utilisées, comme le saut d'exons qui consiste à venir modifier l'épissage en n'incluant pas un exon muté. Là aussi l'utilisation d'ARN spécial appelé « antisens » est nécessaire.<sup>19</sup>

### (3) Edition du génome

Le développement de techniques comme les nucléases à doigt de zinc, les TALENs et surtout l'outil CRISPR-Cas9 (**Box 1**)<sup>20</sup> a permis d'ouvrir de nouvelles voies en thérapie génique. En effet, ces techniques permettent d'éditer directement le génome du patient, c'est-à-dire de venir à l'aide de nucléases couper un endroit précis du gène. Une fois celui-ci coupé, les mécanismes de réparation de la cellule vont intervenir. Ces étapes peuvent engendrer des mutations que l'on appelle micro-insertion ou délétion et qui auront pour effet : (i) de restaurer le cadre de lecture d'un gène et ainsi son expression, ou au contraire, (ii) de décaler le cadre de lecture et empêcher l'expressions d'un gène. Il est également possible d'apporter un petit ADN d'intérêt (parfois quelques nucléotides) pour l'insérer dans le génome, cette séquence doit alors présenter une forte homologie à ses extrémités pour que la cellule puisse l'utiliser. Depuis la découverte de l'outil CRISPR-Cas9 l'édition de gène a connu un fort développement de sorte que dorénavant plusieurs essais cliniques utilisent cette technique pour le transfert de gène.<sup>21</sup>

### Box 1 : La technologie CRISPR-Cas9

La technologie CRISPR-Cas9, aussi appelée ciseaux moléculaires, développée par la française Emmanuelle Charpentier et l'américaine Jennifer Doudna (prix Nobel de chimie en 2020), a largement été utilisée dans le monde scientifique à partir de 2012. Il s'agit à la base d'un mécanisme de défense immunitaire chez les bactéries pour contrer l'infection par des virus en coupant l'ADN viral et en le conservant en vue d'autres infections. La technique repose sur la capacité de l'enzyme Cas9, utilisée conjointement à une molécule ARN guide (sgRNA), à couper spécifiquement une partie du génome. La coupure de l'ADN double brin active le système cellulaire de réparation de l'ADN et deux mécanismes peuvent alors s'opérer, soit il y a une recombinaison homologue (HDR), soit il y a jonction non homologue des extrémités (NHEJ). L'HDR nécessite l'apport d'une séquence d'ADN avec les extrémités pratiquement identiques à celles coupées pour venir s'insérer dans le génome. La NHEJ rassemble directement les extrémités coupées de l'ADN, cette technique est notamment utilisée pour rétablir le cadre de lecture d'un gène en réalisant des micros-insertions ou délétions (en fonction de la mutation) (figure 4).

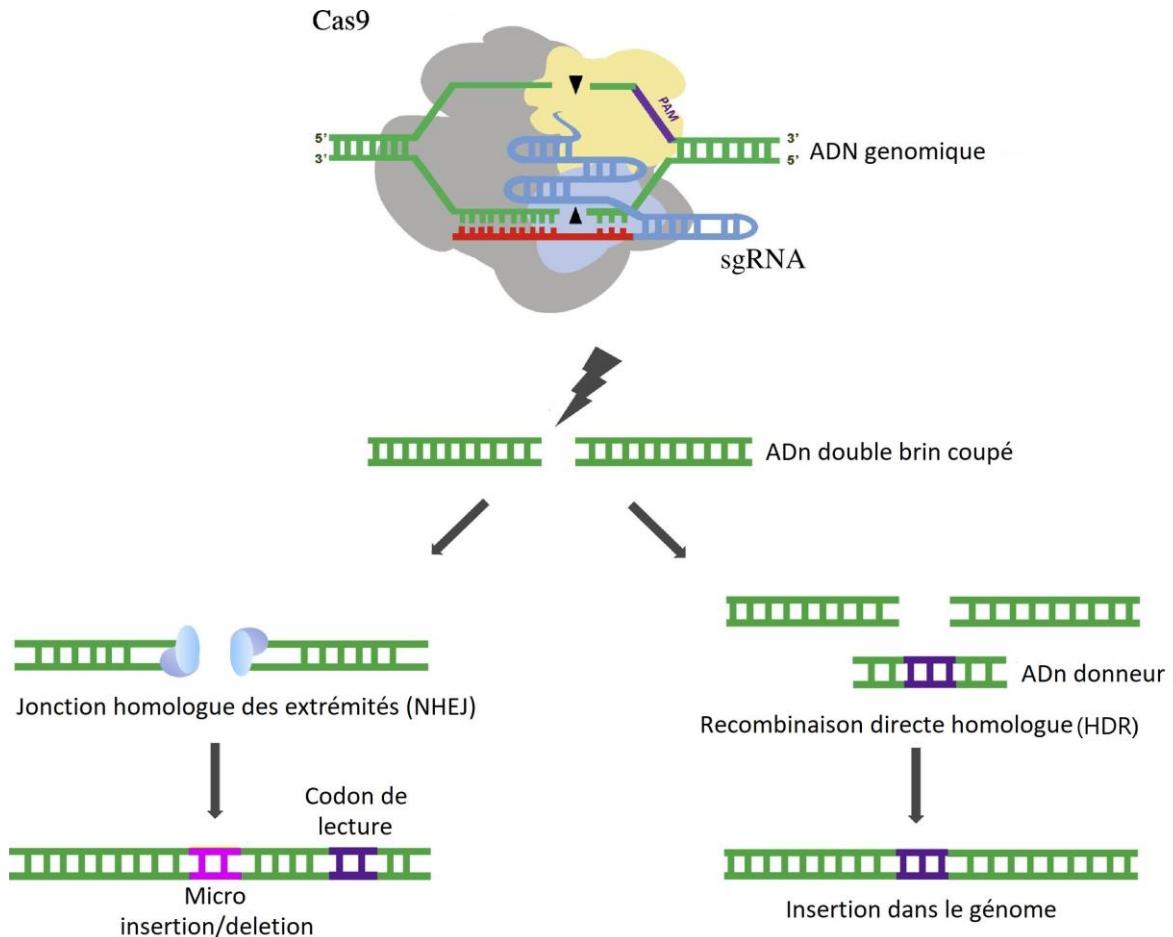


Figure 4 : Schéma représentant la technique d'édition de gène CRISPR CAS9, avec les deux techniques HDR et NHEJ.



En thérapie génique on distinguera deux méthodologies : la thérapie génique *in vivo*, qui consiste à injecter directement le vecteur dans l'organisme du patient, et la thérapie génique *ex vivo*, pour laquelle il faut d'abord prélever les cellules du patient pour ensuite les modifier génétiquement en laboratoire avant de les lui réinjecter. C'est cette approche qui est largement utilisée dans le traitement des cancers (figure 5).

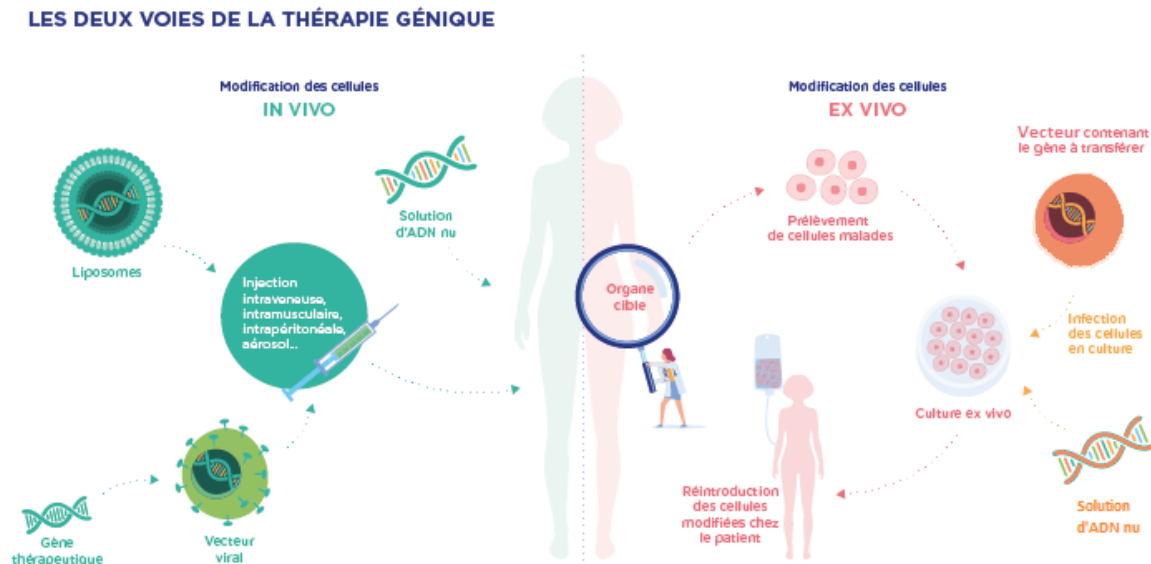


Figure 5 : Schéma détaillant les deux voies de thérapie génique.<sup>22</sup>

Ces deux méthodologies ne sont pas forcément utilisées de manière égale selon la stratégie mise en place. Par exemple, l'extinction de gène est une approche majoritairement explorée *in vivo* alors que l'édition sera majoritairement utilisée *ex vivo* (bien que de nombreuses études *in vivo* commencent à émerger). L'addition de gène quant à elle, selon les applications, concerne les deux méthodologies.

Toutes ces techniques (addition, édition, extinction de gène) permettent de couvrir une multitude de pathologies avec le cancer qui est la maladie la plus traitée et qui représente 67% des 3100 essais cliniques actuellement en cours. On trouve ensuite les maladies monogéniques à 12% puis les maladies infectieuses et cardiovasculaires avec 6% (figure 6).<sup>11,14,23</sup>

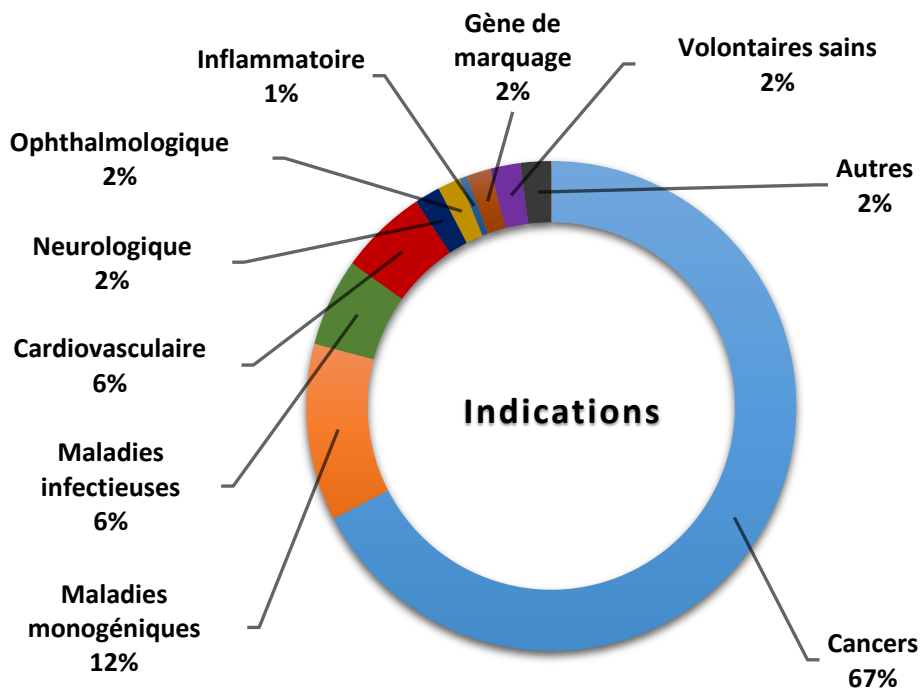


Figure 6 : Les différentes applications des essais cliniques de thérapie génique en cours.<sup>14,23</sup>

Ces différentes applications regroupent l'intégralité des techniques et approches de thérapie génique (*ex vivo*, *in vivo* / addition, édition, extinction de gène). Chaque indication peut être traitée par différentes méthodes. Par exemple, le cancer est souvent traité par méthode *ex vivo*, tant dis que pour les maladies monogéniques les chercheurs préfèrent une approche *in vivo*.

### C. Les différents vecteurs pour la thérapie génique

Comme nous l'avons évoqué auparavant, le transgène doit être vectorisé pour ne pas être dégradé et le choix du vecteur est un facteur important en thérapie génique. Outre le fait qu'il doit protéger l'ADN il doit aussi être capable de le transporter jusqu'aux cellules cibles et cela sans être toxique. Il existe deux grandes familles de vecteurs, les non-viraux et les viraux.

#### a) Les vecteurs non viraux

Plusieurs méthodes non virales existent pour introduire dans une cellule de l'ADN exogène. Nous distinguerons les méthodes physiques permettant « l'injection de l'ADN » aux méthodes chimiques utilisant des nanoparticules.<sup>24</sup>

## (1) ADN « nu » ou plasmidique

La vectorisation d'ADN nu repose sur des méthodes d'injection qui utilisent des techniques physiques. Ces dernières visent souvent la déstructuration de la membrane cellulaire afin de la rendre perméable facilitant alors le passage de l'ADN. Ces techniques sont généralement utilisées *ex vivo* et sont relativement lourdes à mettre en place (tableau 1).

Méthode	Principe	Avantages	Inconvénients
<b><i>Injection directe</i></b>	Micro-aiguille, gène gun	Simple, rapide et direct	Induit beaucoup de mort cellulaire, besoin d'instruments très spécifiques
<b><i>Electroporation</i></b>	Rend la membrane perméable par chocs électriques	Très efficace	Mort cellulaire
<b><i>Sonoporation</i></b>	Déstabilise la membrane par ultrason	Sécurité et large choix de cellules	Faible efficacité
<b><i>Photoporation</i></b>	Utilisation d'un laser pour générer des pores	Efficace, très précis	Peu d'études répertoriées
<b><i>Hydroporation</i></b>	Pression hydrodynamique	Faible toxicité	Utilisé principalement pour les cellules hépatiques et essentiellement <i>in vivo</i>
<b><i>Magnétofection</i></b>	Utilisation de nanoparticules magnétiques et d'un aimant	Peut transfecter des cellules primaires très difficiles à transfecter	Procédure difficile

Tableau 1 : Les différentes techniques de transfection physique, avantages et inconvénients.<sup>24</sup>

Un problème de ces techniques est la perte de l'expression du matériel génétique due aux divisions cellulaires. Plus récemment, d'autres méthodes innovantes ont permis l'intégration du transgène dans le matériel génétique de la cellule. C'est le cas des séquences transposons qui contiennent des répétitions inversées qui sont reconnues et clivées par une enzyme transposase permettant leur intégration dans le génome de l'hôte. Il existe aussi les séquences S/MAR utilisées dans le processus de réplication de l'ADN et qui ont montrées des résultats intéressants chez la souris et le porc.<sup>25,26</sup>

## (2) Vecteurs chimiques

Les vecteurs chimiques sont composés de molécules capables, grâce aux interactions électrostatiques, de protéger l'ADN en formant des particules. Il existe deux grands types de vecteurs chimiques, ceux

à base de lipides cationiques, et ceux issus de polymères cationiques. Les lipides cationiques sont des molécules possédant une tête hydrophile chargée positivement et une queue hydrophobe. La tête chargée positivement se lie aux groupements phosphates des acides nucléiques qui sont chargés négativement, formant une structure compacte unique appelée lipoplexe. Les polymères cationiques (protonables) se complexent de la même façon à l'ADN formant des polyplexes qui peuvent être de différentes natures (chimiques ou naturelles). Les plus connus étant le PEI, le chitosan, les dendrimères et le polyméthacrylate (figure 7).<sup>15</sup> Il est également rapporté l'utilisation de nanoparticules d'or avec un effet photo thermique, de phosphates de calcium et de silice.

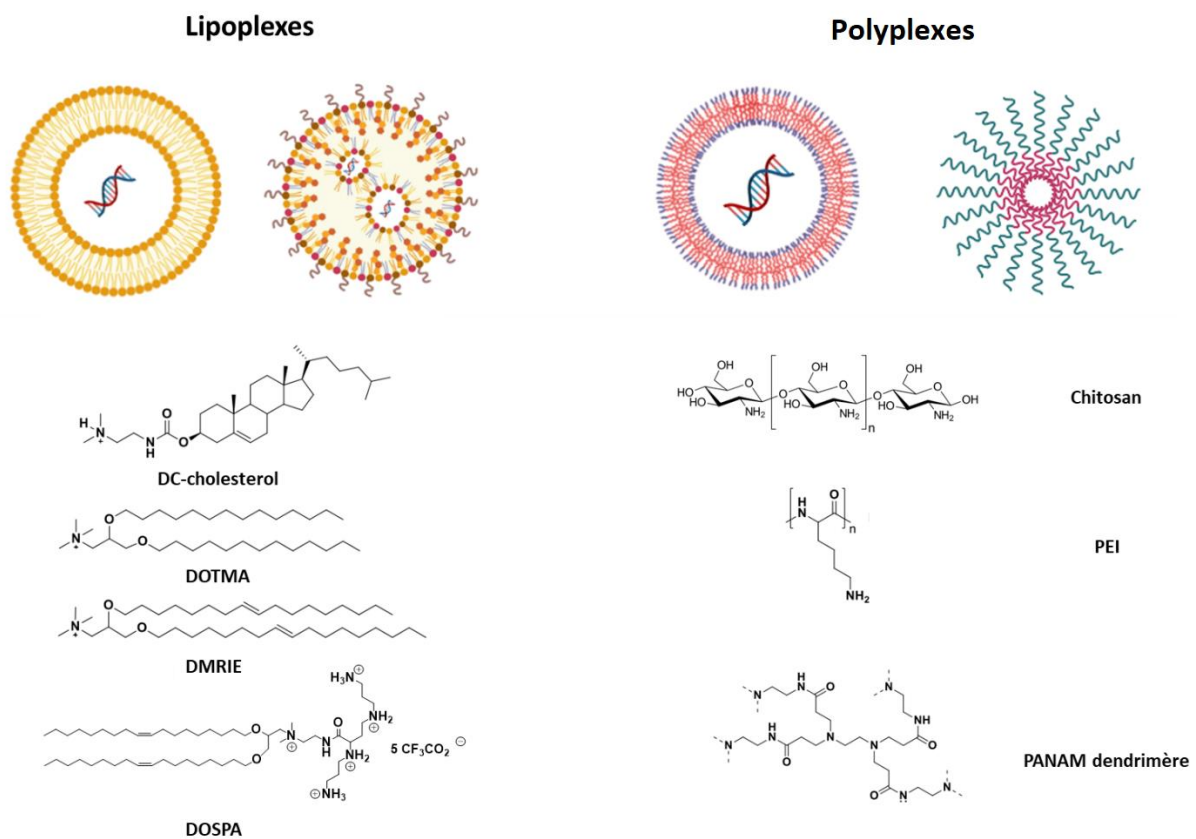


Figure 7 : Exemples de lipoplexes et de polyplexes actuellement utilisés.

Ces vecteurs ont l'avantage d'être faciles à synthétiser, d'avoir un coût de production compétitif et de pouvoir encapsuler un transgène de n'importe quelle taille. De plus, ils ne sont pas pathogènes pour l'organisme et ne provoquent que très peu de réaction du système immunitaire. Il est également possible de fonctionnaliser les lipides/polymères avec des molécules d'intérêt pour, par exemple, cibler spécifiquement un récepteur cellulaire.

Cependant, les vecteurs non-viraux possèdent plusieurs limitations liées à leur nature (souvent cationique) qui entraînent des interactions non-spécifiques avec les composants des fluides

biologiques. Ces interactions limitent, en termes d'efficacité, de spécificité et de biodisponibilité, l'accès des particules à la cible cellulaire et donc à la transfection du gène dans celle-ci. De plus, certains de ces vecteurs intègrent la cellule par fusion membranaire libérant dans le cytoplasme leur ADN thérapeutique. Ce dernier ne migre alors que difficilement jusqu'au noyau ce qui limite son utilisation.<sup>27</sup> Les polyplexes, par contre, rentrent dans la cellule par endocytose, il est alors nécessaire pour ces particules de s'en échapper auquel cas elles seront détruites.

Jusqu'à fin 2020, les vecteurs non viraux n'étaient pas représentés sur le marché des médicaments de thérapie génique et sur les 3000 essais de thérapie génique seulement 5% les concernaient. Cependant, suite à l'épidémie de SARS-COV-2, le laboratoire Moderna a développé le premier vaccin Spikevax<sup>®</sup> utilisant un ARN messager codant pour la protéine spike du virus.<sup>28</sup> Cet ARN est encapsulé dans un liposome, vecteur non viral nanolipidique cationique composé principalement de lipides aminés de type SM-102 et de PEG.<sup>29</sup> Ce vaccin est une avancée majeure dans le domaine de la thérapie génique contre les maladies infectieuses avec plus de 13 millions d'injection en France. Dans la même lignée, la société américaine Pfizer-BioNTech a développé le vaccin Comirnaty<sup>®</sup> composé aussi d'un lipoplexe et d'un ARN messager codant pour la protéine Spike, ici les types de lipides utilisés proviennent de la famille des ALC (0315 et 0159).<sup>30</sup> Ce vaccin a été le plus administré en France avec plus de 80 millions de doses.

## b) *Les vecteurs viraux*

Les virus sont des agents infectieux parasites qui ont acquis la capacité d'utiliser la machinerie cellulaire pour se multiplier. L'évolution leur a permis de développer des mécanismes spécifiques et particulièrement efficaces pour incorporer leur génome à l'intérieur des cellules qu'ils infectent. Ce mécanisme d'incorporation de matériel génétique (ADN ou ARN) s'appelle la transduction.<sup>31</sup>

En retirant le matériel génétique sauvage et pathogène du virus et en le remplaçant par un gène thérapeutique il est alors possible d'utiliser le mécanisme naturel de transduction des virus pour le délivrer aux cellules cibles. On parlera alors de virus recombinant ou de vecteur viral.

C'est en 1976 que Paul Berg réussit à hybrider deux séquences d'ADN de deux virus différents et ainsi produire le premier virus hybride, c'est-à-dire un virus contenant la capsidie d'un polyomavirus SV40 et une partie du génome d'un bactériophage. Ce virus hybride était capable de transduire et de répliquer son génome *in vitro* dans des cellules rénales de singes. Paul Berg a ainsi démontré qu'il est possible de produire des virus fonctionnels avec un génome muté. Cette découverte marqua le début de l'utilisation de virus comme vecteurs de transfert de gène.<sup>32</sup>

Il existe plus 5000 espèces de virus infectant les humains, de formes, de structures et de tailles très variables. On peut noter également des différences dans la composition de leur génome (ADN simple brin, double brin voire même ARN), avec une taille de génome pouvant aller d'environ 5000 paires de bases à 1000 kilobases. Les plus connus et utilisés pour le transfert de gène sont les adénovirus, les virus adéno-associés, les rétrovirus et les virus de l'herpès.<sup>31</sup>

De façon simpliste, les virus infectent les cellules après interaction entre leur capsid, ou leurs protéines membranaires, et des récepteurs cellulaires. Ils rentrent ensuite dans les cellules par endocytose ou par fusion membranaire (cas des virus enveloppés). Une fois à l'intérieur de la cellule le virus doit s'échapper de l'endosome (cas des virus non enveloppés) puis libérer son information génétique dans le cytoplasme ou le véhiculer et le libérer directement dans le noyau. La cellule utilise alors cette information génétique contre son gré pour synthétiser de nouveaux virions.

Les virus sont devenus au fil du temps des vecteurs très pertinents pour le transfert de gène. En plus de pouvoir s'internaliser dans les cellules pour utiliser leur machinerie cellulaire, les virus ont acquis la capacité de perdurer dans différents milieux biologiques ce qui en fait des vecteurs particulièrement efficaces. Le choix du vecteur est important et plusieurs critères sont donc à prendre en compte comme son mode de production, sa voie d'entrée dans la cellule, le type cellulaire qu'il peut infecter, son immunogénicité ou encore la taille de l'ADN thérapeutique qu'il peut transporter. Ce choix doit aussi prendre en compte la stratégie de thérapie génique utilisée ainsi que la méthode de traitement (*ex* ou *in vivo*).

Les vecteurs viraux bien que très efficaces possèdent certaines limites. Ils sont capables d'infecter de nombreux types de cellules mais ne sont pas forcément spécifiques pour l'un d'entre eux. Ce large tropisme induit un transfert du gène thérapeutique dans d'autres cellules non ciblées. Il est alors nécessaire d'injecter de fortes doses pour avoir une couverture optimale. La deuxième problématique majeure est leur immunogénicité. Injecter un vecteur viral dans l'organisme n'est pas anodin et si certains sont mieux tolérés que d'autres, ce paramètre peut aussi être patient dépendant. Certains essais se sont vus contraints d'être mis à l'arrêt à cause de complications immunologiques. Par exemple, en septembre 1999, un américain souffrant d'une maladie génétique provoquant une carence en ornithine transcarbamylase, participe à un essai clinique à l'université de Pennsylvanie. Il reçoit une injection d'adénovirus recombinant mais son organisme sur-réagit et active de manière trop importante le système immunitaire ce qui lui sera fatal. Cela montre la nécessité d'optimiser les vecteurs, d'une part en contrôlant mieux le transfert du gène et sa prise en charge par la cellule afin d'éviter sa génotoxicité et d'autre part en maîtrisant la réponse du système immunitaire du patient afin de limiter les réponses inflammatoires.<sup>33</sup>

Actuellement, selon les données référencées plus de 3100 essais cliniques utilisent des vecteurs viraux (figure 8).<sup>34</sup>

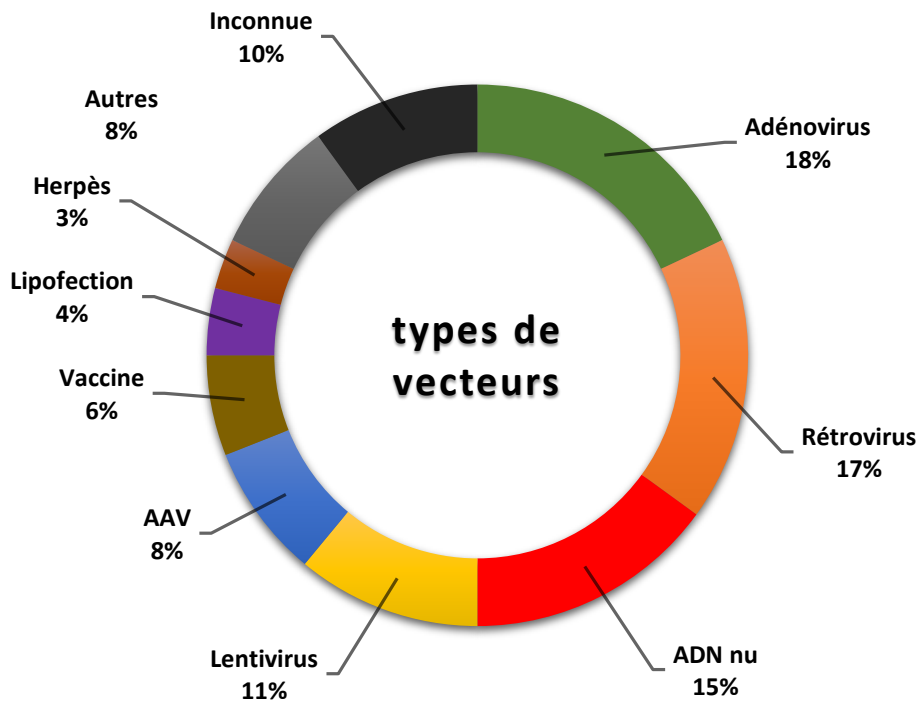


Figure 8 : Virus actuellement utilisés comme vecteur dans l'intégralité des essais cliniques de thérapie génique.<sup>34</sup>

Avec 66% des essais cliniques les virus sont les vecteurs les plus utilisés en thérapie génique, parmi eux les adénovirus, rétrovirus (dont les lentivirus font partie) et les virus adéno-associés représentent actuellement la très grande majorité (figure 9). Dans ce manuscrit nous allons nous focaliser sur les virus adéno-associés ou AAV.

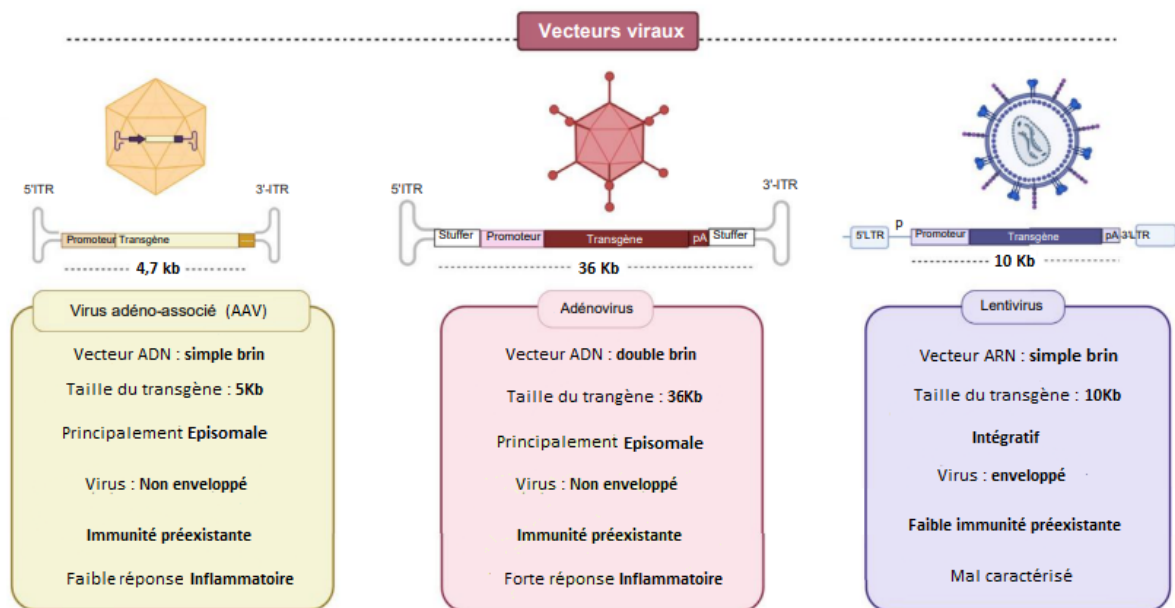


Figure 9 : Les trois vecteurs viraux les plus utilisés en thérapie génique : l'AAV, l'adénovirus et le lentivirus selon Maestro<sup>16</sup>

### (1) Lentivirus (rétrovirus)

Les lentivirus qui font partie de la famille des rétrovirus sont des vecteurs très utilisés en thérapie génique *ex vivo* pour le traitement des cancers. Le lentivirus mesure 80 à 100 nm et est entouré d'une enveloppe lipidique contenant une capsid sphérique. Son matériel génétique est sous forme d'ARN simple brin et est converti en ADN par une enzyme endogène. Il possède une autre enzyme, nommée intégrase, qui permet à l'ADN de s'intégrer durablement dans le patrimoine génétique de la cellule. Ce dernier point est particulièrement problématique car l'intégration est généralement aléatoire et peut engendrer le développement de cancer.<sup>35</sup> A ce jour, divers traitements *ex vivo* utilisent les lentivirus et rétrovirus. Strimvelis<sup>®</sup> (2016) traitant l'immunodéficience sévère, Zynteglo<sup>®</sup> (2019) pour la  $\beta$ -thalassémie (maladie du sang) ou encore le traitement de cancers avec la technologie des Car-T cell pour Kymriah<sup>®</sup> (2017) et Yescarta<sup>®</sup> (2018) ou plus récemment Abecma<sup>®</sup> (2021). Quelques essais cliniques utilisent cependant les vecteurs lentivirus/rétrovirus *in vivo* comme d'ailleurs le Rixin-G<sup>®</sup> (2007) utilisé pour traiter le Sarcome.

### (2) Les adénovirus

Découvert en 1953 par Wannace P. Rowe, les adénovirus sont des particules d'un diamètre de 60 à 90 nm, sans enveloppe, à capsid icosaédrique formée de 252 capsomères. Les protéines situées au sommet de l'icosaèdre sont des pentons de longueur variable et surplombés d'un antigène.<sup>36</sup> On dénombre actuellement 52 sérotypes d'adénovirus, les sérotypes Ad5 et Ad26 sont très souvent utilisés en thérapie génique. Ils sont non-intégratifs et délivrent leur ADN sous forme circulaire extra chromosomique, appelé épisome, dans le noyau des cellules transduites. Ce virus est très



immunogène, c'est pourquoi il est principalement utilisé en traitement du cancer et comme vecteur vaccinal pour les maladies infectieuses.<sup>37</sup> Plusieurs médicaments actuellement sur le marché utilisent un vecteur adénovirus, c'est le cas de la Gendicine<sup>®</sup> (2003) qui est le premier traitement de thérapie génique virale anticancéreux autorisé (en chine). L'adénovirus est aussi largement utilisé en vaccination notamment contre la COVID-19, AstraZeneca<sup>®</sup>, Johnson et Johnson<sup>®</sup> et Convidicea<sup>®</sup> (2021) produisent des vecteurs recombinants contenant un gène codant pour la protéine spike du SARS-COV-2.

### (3) Les virus adéno-associés

Les virus adéno-associés (AAV) sont les virus au cœur de ce manuscrit et seront présentés plus en détail dans le paragraphe suivant. Il est important de préciser que sur les 3100 essais cliniques actuellement recensés, bon nombre d'entre eux sont de la thérapie génique *ex vivo*. Par la suite nous nous intéresserons aux études *in vivo* et plus précisément aux études *in vivo* d'addition de gènes. En effet, l'AAV est un vecteur de choix pour traiter des maladies monogéniques par injection *in vivo*. L'adénovirus, lui, est plus globalement utilisé comme vecteur *in vivo* pour des applications en maladie infectieuse. Enfin les rétrovirus sont plus globalement utilisés dans des méthodes de thérapie génique *ex vivo* pour traiter le cancer.

## 2. Les virus adéno-associés

### A. L'utilisation des AVV comme vecteurs en thérapie génique

#### a) Biologie de l'AAV

Les virus adéno-associés (**figure 10**) appartiennent à la famille *Parvoviridae* du genre *Dependovirus*. Ils ont été découverts en 1965 par l'équipe d'Atchinson aux Etats-Unis comme contaminants d'une production d'adénovirus. L'AAV est un virus non pathogène et non-enveloppé possédant une capsidie protéique de forme icosaédrique de 25 nm et contenant un ADN simple brin d'environ 4700 paires de bases.<sup>38</sup>

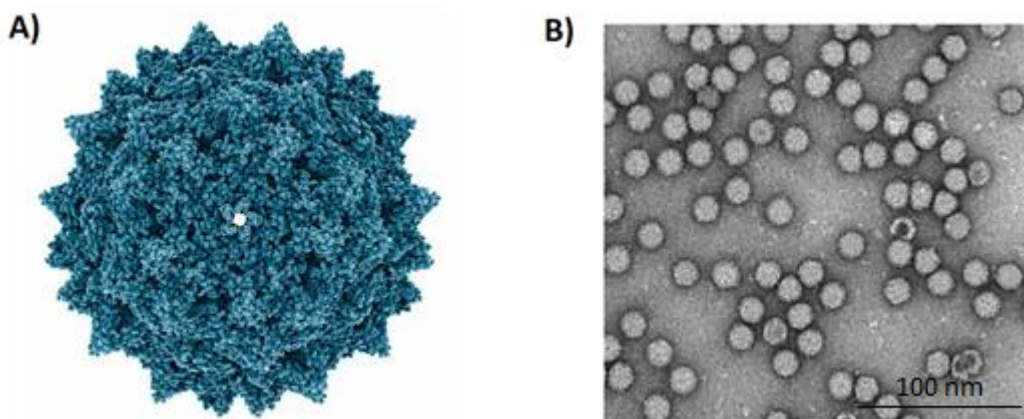


Figure 10 : A) Représentation par modélisation d'un AAV de sérotype 2 B) image TEM à coloration négative d'un batch d'AAV caractérisé par la société Vironova.<sup>39</sup>

L'ADN comporte deux gènes *rep* et *cap* possédant 3 trois ORF (Open Reading Frames). Ces deux gènes sont bordés par les séquences ITR (Inverted Terminal Repeats), qui sont des motifs palindromiques d'environ 145 paires de bases, et sont impliqués dans la réplication et l'encapsidation du virus (**Figure 11**).

- Le gène *rep* code pour quatre protéines de régulation non structurales nécessaires à la réplication du virus :

- Rep78 et Rep68 sous le contrôle du promoteur p5
- Rep52 et Rep40 sous le contrôle du promoteur p19

Les protéines rep78 et rep68 possèdent des fonctions régulatrices qui agissent dans toutes les phases du cycle de l'AAV. Elles permettent une régulation de l'expression des gènes, de la réplication du génome et jouent également un rôle dans l'intégration site-spécifique de l'AAV.<sup>40</sup> En l'état actuel des connaissances, l'AAV serait le seul virus de mammifère capable de s'intégrer sur un site-spécifique nommé AAVS1 se trouvant sur le locus 19q13.3-qter du chromosome 19.<sup>41</sup> Ces protéines sont intimement liées aux ITR. Les protéines rep52 et rep40 participent à la production et à l'accumulation du génome viral et interviennent également pour l'encapsidation de l'ADN simple brin dans la capsidie assemblée.

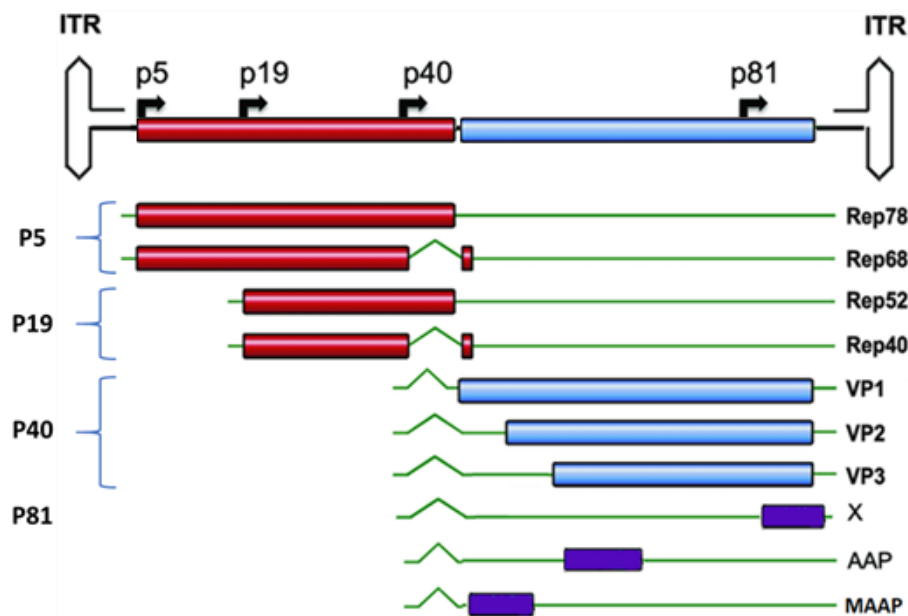


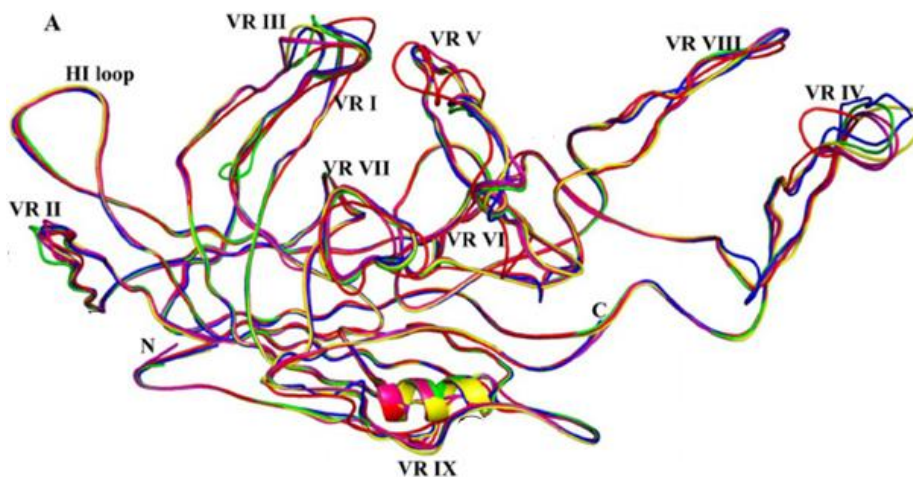
Figure 11 : Schéma du génome de l'AAV sauvage, trois ORF codent pour 9 protéines.<sup>42</sup>

- Le gène *cap*, initié par le promoteur P40, code pour trois protéines nommées VP1 (87 kDa), VP2 (72 kDa) et VP3 (62 kDa) qui constituent la capsidie des AAV. Elles possèdent la même séquence peptidique mais sont issues de codons start différents en position N-terminale ce qui explique leur différence de taille. Ces trois protéines s'assemblent dans une proportion 5/5/50 formant alors la capsidie composée de 60 unités protéiques. Un autre promoteur, le P81, produit *via* le gène *cap* une protéine nommée X pour la réplication de l'ADN viral.<sup>43</sup> Enfin, ce gène permet aussi la synthèse des protéines non structurales AAP et MAAP qui sont responsables de l'interaction entre les différentes VP pour aider à la formation de la capsidie.<sup>44,45</sup>

Pour se développer, l'AAV doit transférer son matériel génétique à une cellule pour permettre l'expression des protéines nécessaires à la production de nouveaux virions. L'AAV est incapable de se répliquer seul, c'est-à-dire qu'il a besoin d'un virus auxiliaire (virus « helper ») pour se multiplier. Une

fois la cellule infectée par un AAV, son ADN restera dans une phase latente sous forme d'épisome dans le noyau cellulaire ou bien sera intégré dans le chromosome 19.<sup>46</sup> Après infection de la cellule par le virus auxiliaire (généralement un adénovirus) la réplication peut commencer et la machinerie cellulaire permet alors la production de nouveaux AAV.<sup>47</sup>

A ce jour, il existe treize sérotypes naturels d'AAV issus d'humains et de primates, notés AAV1 à AAV13.<sup>48</sup> Les sérotypes sont discriminés selon leur propriété antigénique, si un sérotype n'est pas neutralisé par un sérum d'anticorps connu alors il est considéré comme nouveau. Des études structurales réalisées par cryo-microscopie ont permis d'étudier les différences entre les sérotypes d'AAV. Les protéines VP1/2 sont principalement regroupées à l'intérieur de la capside virale rendant leur étude difficile. La partie externe de la capside est composée en grande majorité de protéines VP3 présentant une homologie de structure comprise entre 53% et 99% en fonction des sérotypes.<sup>48</sup> Les différences d'homologie sont principalement regroupées dans 9 régions variables (VR) qui jouent un rôle dans l'association des VP entre elles et donc dans la conformation finale de la capside (**Figure 12**).



*Figure 12 : Alignement des sérotypes d'AAV, les 9 régions variables situées dans des boucles indiquées par VR I à VR IX sont parfaitement visibles.<sup>49</sup>*

Les régions variables interviennent également dans la reconnaissance de l'AAV avec son environnement. Concrètement, elles sont responsables des différentes interactions avec les récepteurs cellulaires impliqués lors de l'infection. Il est donc logique que les sérotypes puissent infecter différents tissus. Le tropisme se réfère aux différents tissus que le vecteur est capable de cibler. Il est admis que les sérotypes AAV1 et AAV6 transduisent efficacement le muscle, l'AAV5 le cerveau, l'AAV8 le foie, le muscle et le cœur, et que l'AAV9 peut surpasser d'autres sérotypes d'AAV dans le muscle, le cerveau et le cœur (**tableau 2**).<sup>50</sup>

<b>Sérotype</b>	<b>Tropisme</b>
AAV1	Foie, cœur, muscle, système nerveux central
AAV2	Cœur, muscle, système nerveux central, œil, foie
AAV3	Cœur, foie
AAV4	Cœur, poumon, foie, œil
AAV5	Foie, poumon, œil, cerveau
AAV6	Foie, muscle, cœur, système nerveux central
AAV7	Foie, muscle, système nerveux central
AAV8	Cœur, foie, cerveau, muscle, pancréas, système nerveux central
AAV9	Tous les tissus
AAV10	Muscle
AAV11	X
AAV12	Nez
AAV13	Cerveau

Tableau 2 : Les différents sérotypes et leur tropisme, prédominant, naturel chez l'homme.<sup>50,51</sup>

### b) Trafficking de l'AAV

Le trafficking désigne l'ensemble des étapes nécessaires pour qu'une cellule exprime le génome d'un AAV à la suite de son infection. Ce phénomène est encore mal connu mais, grossièrement, pour infecter une cellule l'AAV utilise les régions spécifiques des protéines VP3. Les récepteurs cellulaires impliqués sont très souvent des protéoglycanes et le premier d'entre eux, qui a été identifié en 1998, est un récepteur héparane sulfate. On peut également citer le galactose pour l'AAV9 ainsi que les acides sialiques pour les AAV1/4/5 et l'AAV6.<sup>52,53</sup>

L'interaction entre l'AAV et le récepteur membranaire va déclencher l'internalisation du virus. Aujourd'hui 23 protéoglycanes ont été identifiés comme récepteurs primaires de différents sérotypes d'AAV.<sup>54</sup>

Les AAV utilisent également des récepteurs secondaires qui permettent de stabiliser le virus pendant son entrée dans la cellule. Ce sont des protéines souvent *N* ou *O* glycosylées comme la laminine (LamR) ou l'AAVR.<sup>55</sup> L'AAVR, protéine transmembranaire, est un co-récepteur, décrit en 2016 par Pilay, qui a un impact dans le mécanisme d'infection mais aussi dans le tropisme de nombreux sérotypes d'AAV (tableau 3).<sup>56</sup>

<b>Sérotype</b>	<b>Récepteur</b>	<b>Corécepteur</b>
AAV1	N-Acide sialique	AAVR
AAV2	HSGP	HGFR, Intégrine, LamR, AAVR
AAV3	HSGP	HGFR, LamR, AAVR
AAV4	O-Acide sialique	X
AAV5	HSGP, N-Acide sialique	PDGFR, AAVR
AAV6	HSGP, N-Acide sialique	EGFR, AAVR
AAV7	X	X
AAV8	X	LamR, AAVR
AAV9	N-Galactose	LamR, AAVR
AAV10	X	LamR
AAV11	X	X
AAV12	X	X
AAV13	HSGP	X

Tableau 3 : Récepteurs permettant l'internalisation cellulaire des différents sérotype d'AAV<sup>52-54,56</sup>

Un fois à l'intérieur de la cellule, le virus se retrouve dans un endosome, vésicule lipidique permettant le transport et la prise en charge de composés extracellulaires au sein de la cellule. L'endosome va ensuite mûrir en lysosome en s'acidifiant ce qui va changer la conformation spatiale de la capsid du virus. Les domaines N-terminaux des protéines VP1 et VP2 vont alors être externalisés libérant une région phospholipidique. A l'aide d'une phospholipase l'AAV pourra s'échapper de l'endosome, dans le cas contraire le virus sera détruit.<sup>57</sup>

Une fois l'AAV libéré dans le cytoplasme il migre jusqu'au noyau afin d'y délivrer son génome. Cette étape est mal connue et fait surement intervenir différents mécanismes et partenaires utilisés par l'AAV pour se déplacer dans le cytoplasme. Toutefois, il est bien décrit que l'AAV peut subir une ubiquitination (modification post-traductionnelle qui permet la fixation d'une ou plusieurs protéines d'ubiquitine) le rendant vulnérable à un complexe protéique appelé le protéasome. Ce complexe est composé de protéases dont le rôle est de dégrader en peptides les protéines non fonctionnelles ou extérieures à la cellule. Il est donc primordial pour l'AAV d'échapper à ce phénomène.<sup>58</sup>

L'AAV pénètre ensuite dans le noyau *via* des pores nucléaires où il est décapsidé libérant son ADN simple brin.<sup>59</sup> Sous cette forme le génome du virus n'est pas encore fonctionnel, il doit d'abord être

converti en ADN double brin. Les ITR servent d'initiateurs à cette étape permettant à des enzymes de la cellule hôte, appelées ADN polymérasés, de synthétiser le brin d'ADN complémentaire.

Le matériel génétique du virus va alors rentrer dans un processus de circularisation entraînant une recombinaison intra ou inter moléculaire au niveau de ses ITR. Cette concatémérisation va permettre au génome de persister dans le noyau sous forme épisomale, conduisant son expression persistante dans le temps.<sup>60,61</sup> C'est à cette étape que l'AAV a besoin d'un virus helper qui va, en infectant la cellule, permettre à l'ADN circulaire d'être exprimé produisant ainsi de nouveaux virions.<sup>62</sup>

### c) Des AAV recombinants pour la thérapie génique

Pour utiliser les AAV comme vecteur il est nécessaire de les modifier génétiquement en remplaçant les séquences *rep* et *cap* par le gène d'intérêt appelé transgène. Le virus modifié est alors nommé AAV recombinant (AAVr) (figure 13).

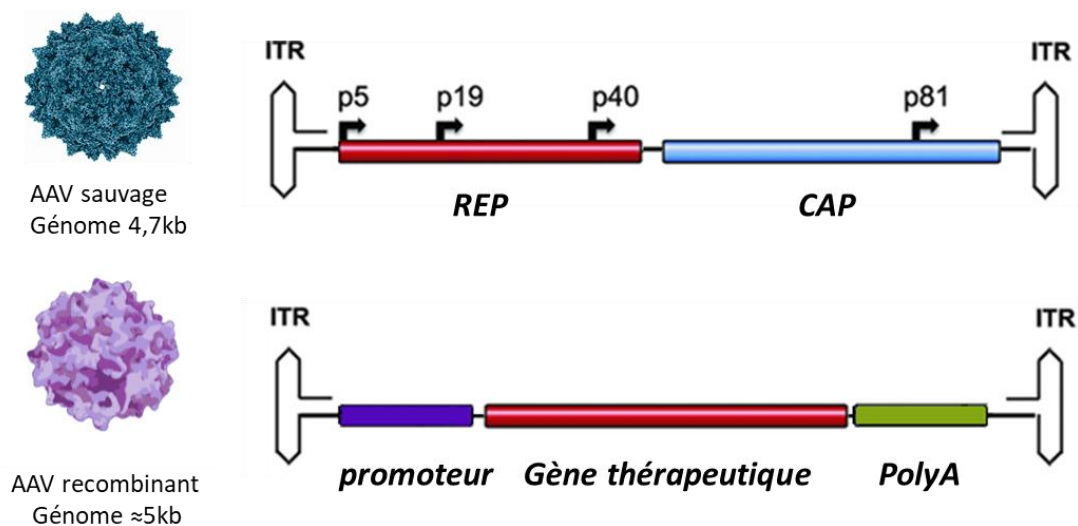


Figure 13 : Génome de l'AAVr. Les gènes *Rep* et *Cap* de l'AAV sauvage sont remplacés par une cassette d'expression contenant un promoteur, un transgène et une séquence de polyadénylation.

Pour construire un transgène il faut conserver les ITR de l'AAV sauvage car ils sont indispensables dans le processus d'encapsidation mais aussi de répllication.<sup>63</sup> Ils sont positionnés de part et d'autre du transgène lui-même composé de trois parties.

- La première est le promoteur qui permet de recruter les ARN polymérasés cellulaires pour enclencher la transcription. L'utilisation de virus helper n'est donc plus nécessaire à ce stade.
- La deuxième est la partie codante, elle peut être soit d'intérêt thérapeutique comme c'est le cas lors des différentes phases cliniques, par exemple pour la micro dystrophines,<sup>64</sup> soit un gène rapporteur qui permettra le suivi des expérimentations *in vivo* ou *in vitro* (notamment

dans le cadre de l'optimisation d'AAVr). Dans ce cas, le gène rapporteur peut permettre l'expression d'un facteur de résistance à un antibiotique (néomycine)<sup>65,66</sup>, d'une protéine fluorescente comme la GFP ou encore d'une protéine chimioluminescente telle que la luciférase.

- Pour finir, la troisième partie est composée d'un signal de polyadénylation qui empêche la dégradation de l'ARN résultant de la transcription.

Comme évoqué précédemment, il faut bien prendre en compte que l'ADN encapsidé dans un AAVr ne peut excéder 4,7 kb, dépasser cette limite la production du vecteur diminue fortement. Au delà de cette taille il est possible de générer des vecteurs qu'avec une partie du transgène ce qui les rend complètement inefficaces.

Une fois le transgène sélectionné, deux techniques majeures sont actuellement utilisées pour produire les AAVr. La première consiste à transfecter des cellules adhérentes, souvent des HEK293, avec plusieurs plasmides contenant les séquences d'intérêts. Concrètement, deux ou trois plasmides sont utilisés, l'un contient le transgène flanqué des ITR précédemment décrit, le deuxième contient les gènes *rep* et *cap* et un dernier ceux du virus helper permettant la réplication<sup>67</sup> (figure 14).

- Dans un premier temps, la cellule va produire les protéines REP et CAP grâce au plasmide adéquat.
- Puis, les protéines REP vont reconnaître et s'amarrer aux ITR du transgène initiant alors sa réplication.
- Les réplicats du transgène vont alors rentrer dans le processus d'encapsulation et former avec les protéines CAP les AAVr d'intérêt.

Dans le cadre de ma thèse seulement deux plasmides sont utilisés. Dans ce cas les gènes du virus helper sont directement insérés sur le plasmide contenant les gènes *rep* et *cap*.



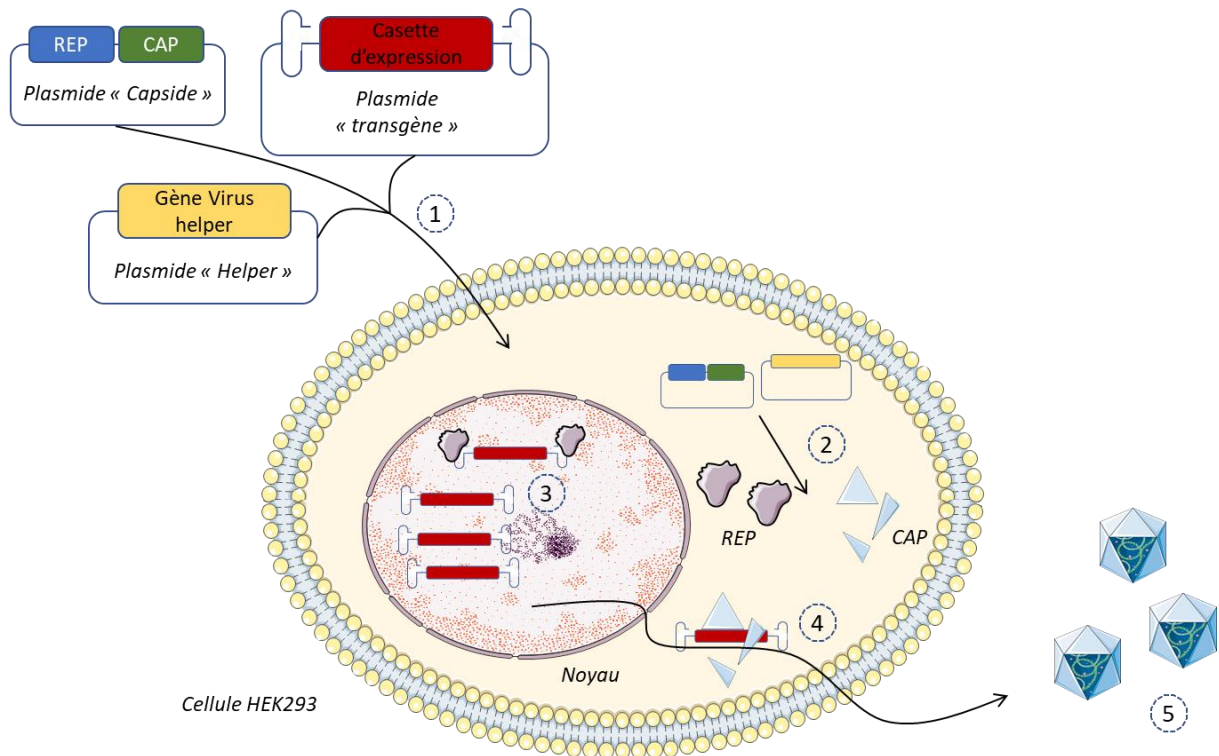


Figure 14 : Production d'AAVr dans des cellules HEK293 adhérentes. Ici trois plasmides sont utilisés, l'un codant pour la capside, le deuxième pour la cassette d'expression et le dernier pour les gènes du virus helper permettant la réplication.

La deuxième technique pour produire des AAVr utilise des Sf9, cellules non adhérentes d'insectes, qui sont transduites à l'aide d'un baculovirus (virus infectant exclusivement des cellules d'arthropodes). De la même manière que les plasmides, les baculovirus vont apporter les gènes *rep*, *cap* et le transgène. L'utilisation d'un ADN helper n'est pas nécessaire ici, la machinerie du baculovirus remplira naturellement cette fonction. La production d'AAVr en cellules d'insectes permet de travailler sur de plus grandes quantités par rapport aux cellules adhérentes (notamment *via* l'utilisation de bioréacteurs pouvant atteindre plusieurs centaines voire milliers de litres). Cette méthode est toutefois plus difficile à mettre en place du fait de la nécessité de produire les baculovirus contenant le transgène d'intérêt.<sup>68</sup>

L'étape commune des deux techniques est l'isolement des AAVr par lyse des cellules dans lesquelles ils sont produits. La purification des vecteurs devient alors une phase critique car il est nécessaire de s'affranchir des protéines cellulaires et contaminants potentiels tout en gardant l'intégrité de l'AAV recombinant. Plusieurs méthodes ont été développées en s'appuyant sur les propriétés physiques et chimiques des particules virales (densité, taille, charge, affinité).

Les lots de vecteurs utilisés dans ce manuscrit ont été purifiés par ultracentrifugation grâce à un gradient de densité au chlorure de césium. Cette technique permet la séparation, en fonction de leur masse, des débris et protéines cellulaires de l'AAVr. Une deuxième purification avec un gradient plus fin est ensuite réalisée pour séparer les particules pleines (c'est-à-dire contenant le transgène) des particules vides. D'autres méthodes utilisent par exemple des colonnes d'affinités qui contiennent des anticorps spécifiques des différents sérotypes d'AAV ou encore des molécules greffées (héparine ou acide sialique) qui interagissent avec les AAVr permettant uniquement sa rétention.<sup>69,70</sup>

La caractérisation des lots de vecteurs est primordiale. En effet, certains contaminants peuvent générer une toxicité ou favoriser une réponse immune de l'hôte diminuant ainsi l'efficacité de l'AAVr. Plusieurs techniques existent pour caractériser un lot d'AAV, les principales étant (i) une analyse par Polymérase Chain Reaction (PCR) quantitative pour connaître la concentration (appelé titre) d'AAV présent dans l'échantillon, (ii) une électrophorèse sur gel de polyacrylamide en condition dénaturante qui permet de séparer les VP pour les faire migrer en fonction de leur taille, (iii) un test d'infectiosité sur des lignées cellulaires comme HEK293. D'autres techniques telles qu'un test ELISA ou un Dot-blot peuvent également être effectuées.

#### d) Traitement utilisant des AAV

Les AAV comptent pour 8% de la totalité des essais cliniques mais, d'après la revue de Zhao, Anselmo et Mitragotri, ils représentent 48% des études cliniques de thérapie génique *in vivo* d'addition de gène. Ils sont ainsi devenus au fil des années les vecteurs de choix pour la thérapie génique *in vivo*. Cela peut être probablement attribué aux progrès récents dans l'ingénierie de divers variants d'AAV, à leur faible immunogénicité et leur expression génétique stable (figure 15).<sup>71</sup>

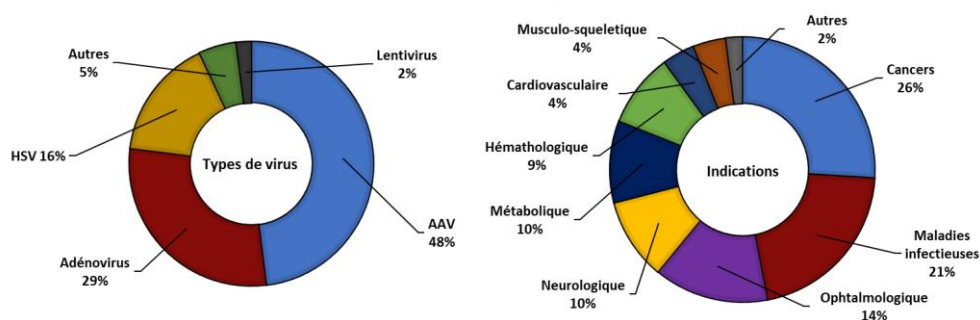


Figure 15 : A) Pourcentage des différents vecteurs viraux AAV utilisés dans les 284 essais cliniques *in vivo*. B) Listes des différentes indications en pourcentages.

Sur l'intégralité des essais cliniques qui utilisent un AAV comme vecteur, 7% sont en phase clinique 3, ce qui reflète son potentiel thérapeutique. Les indications se concentrent principalement sur des maladies hématologiques telles que les hémophilies A et B, des maladies neurologiques telles que l'atrophie musculaire spinale, des maladies ophtalmologiques telles que la rétinite pigmentaire et des troubles musculosquelettiques tels que la dystrophie musculaire de Duchenne.

Au cours des dernières années, six médicaments ont eu une autorisation de mise sur le marché (AMM). Le Glybéra<sup>®</sup>, qui a eu son AMM en 2012, est un AAV1 codant pour la lipoprotéine lipase. Il soigne les patients souffrant de crises de pancréatites sévères ou multiples malgré un régime pauvre en lipides. Depuis 2018, il a été retiré du marché faute de demandes suffisantes. Le Luxturna<sup>®</sup> (AMM en 2017), AAV2 qui contient une copie saine du gène *rpe65*, permet le traitement de la neuropathie optique de Leber qui touche des adultes et des enfants présentant une perte visuelle due à une dystrophie rétinienne héréditaire. Le Zolgensma<sup>®</sup> est un AAV9 codant pour le gène *smn1* qui soigne l'amyotrophie spinale infantile (affection grave des nerfs qui provoque une atrophie et une faiblesse musculaire). Il a provoqué de vives réactions auprès du grand public lors de sa commercialisation en 2019 car il était (tous domaines confondus) le médicament le plus cher du marché avec un coût de 2 millions d'euros l'injection.<sup>72</sup>

L'année 2022 fut marquée par une avancée importante dans l'utilisation de vecteurs AAV comme produit de thérapie génique avec la commercialisation de 3 médicaments. En mai 2022, Upstaza<sup>®</sup> reçoit une approbation de la commission européenne pour le traitement du déficit en dopa-décarboxylase, une enzyme catalysant la réaction de décarboxylation d'acides aminés aromatiques comme la formation de l'histamine par décarboxylation de la L-histidine.<sup>73</sup> Upstaza<sup>®</sup> est un AAV2 contenant le gène de la dopa-décarboxylase (DDC) humaine sous le contrôle du promoteur cytomégalovirus (CMV).<sup>74</sup> Les deux autres médicaments commercialisés en août et novembre 2022 traitent l'hémophilie. Roctavian<sup>®</sup> qui traite l'hémophilie A contient est un AAV5 avec un gène codant pour le facteur VIII humain, ainsi qu'un promoteur spécifique du foie humain qui encourage la traduction dans les hépatocytes. Hemgenix est utilisé dans le cas de l'hémophilie B, il s'agit d'un AAV5 contenant le gène du facteur de coagulation IX contrôlé par un promoteur spécifique du foie.<sup>75</sup> C'est ce dernier qui est maintenant le médicament le plus cher avec un coût de 3,3 millions d'euros pour une dose.

## B. Les limites des AAVr

Bien que l'AAVr soit devenu au fil des années la référence pour le transfert de gène *in vivo*, son utilisation croissante a mis en lumière un certain nombre de limites. Récemment, quatre jeunes patients inclus dans un essai clinique pour la myopathie tubulaire liée à l'X sont décédés des suites d'une injection d'un médicament à base d'AAV : tous présentaient une fonction hépatique dégradée<sup>76</sup>. Il est donc nécessaire d'optimiser ce vecteur viral afin de le rendre plus efficace et plus tolérogène.

### a) *La réponse immunitaire contre les AAV*

Sur la notice du Zolgensma<sup>®</sup> on peut lire qu'il est nécessaire de réaliser, préalablement au traitement, un test antigénique pour déterminer la quantité d'anticorps dirigée contre le vecteur.<sup>77</sup> Si ce test révèle une concentration trop importante d'anticorps alors le traitement n'est pas administré au patient. Cet exemple illustre bien le problème d'immunogénicité des AAV.

Le système immunitaire est un ensemble de processus complexes qui permet de discriminer le soi du non-soi. Il permet, par exemple, de prévenir les infections dues aux virus et/ou aux bactéries. De manière simple il se compose de deux grandes réponses :

- La réponse immunitaire innée comprend les cellules et mécanismes permettant la défense immédiate de l'organisme contre l'agent pathogène. Elle est classifiée d'immédiate car elle ne nécessite pas de division cellulaire. Certaines études montrent que l'AAV peut provoquer une réponse immunitaire innée.<sup>78</sup>
- La réponse immunitaire adaptative intervient dans un second temps et nécessite une division cellulaire (lymphocytes B et T). Elle est cependant bien plus spécifique et capable de garder en mémoire les agents pathogènes afin, lors d'une deuxième infection, d'être très efficace contre ceux-ci. Pour exemple, c'est cette voie qui permet la production des anticorps dépistés avant l'administration du Zolgensma<sup>®</sup>.<sup>79</sup>

Les AAVr, bien qu'ils ne soient plus pourvus d'ADN viral, sont composés d'une capsidie protéique qui va être la cible du système immunitaire et cela à différentes étapes :

Premièrement des anticorps neutralisants (NAb) dirigés contre des zones précises de la capsidie l'empêchent d'interagir avec les récepteurs membranaires bloquant ainsi l'accès à la cellule. Ces anticorps se retrouvent également dans le cytoplasme de la cellule et interviennent en bloquant la décapsidation du vecteur. Ensuite, dans les endosomes, les capsides des AAVr peuvent rencontrer des récepteurs appelé TLR (récepteurs de type Toll) qui vont déclencher une réponse immunitaire innée

qui pourra engendrer la mort de la cellule infectée par apoptose. Les AAVr arrivant à s'échapper de l'endosome peuvent également être ubiquitinylés et dégradés par le protéasome. Ce dernier va dénaturer le vecteur et fractionner les protéines composant sa capsid pour ne libérer que de petits peptides. Ceux-ci seront utilisés pour être chargés sur la protéine du complexe majeur d'histocompatibilité (CMH) de classe I et présentés à la surface de la cellule, les peptides vont alors être reconnus par des récepteurs spécifiques des lymphocytes T provoquant la reconnaissance de la cellule par ces derniers et donc son infection. Ce qui induira, finalement, son élimination par une réponse cytotoxique (réponse immunitaire adaptative). Par la suite, une fois dans le noyau, les ITR présents sur le transgène peuvent aussi donner lieu à une réponse immunitaire innée. Enfin, dans certain cas, les protéines thérapeutiques vont engendrer une réponse immunitaire adaptative *via* leur dégradation par le protéasome et l'exposition d'épitopes reconnus par les lymphocytes T (figure 16).<sup>52,80</sup>

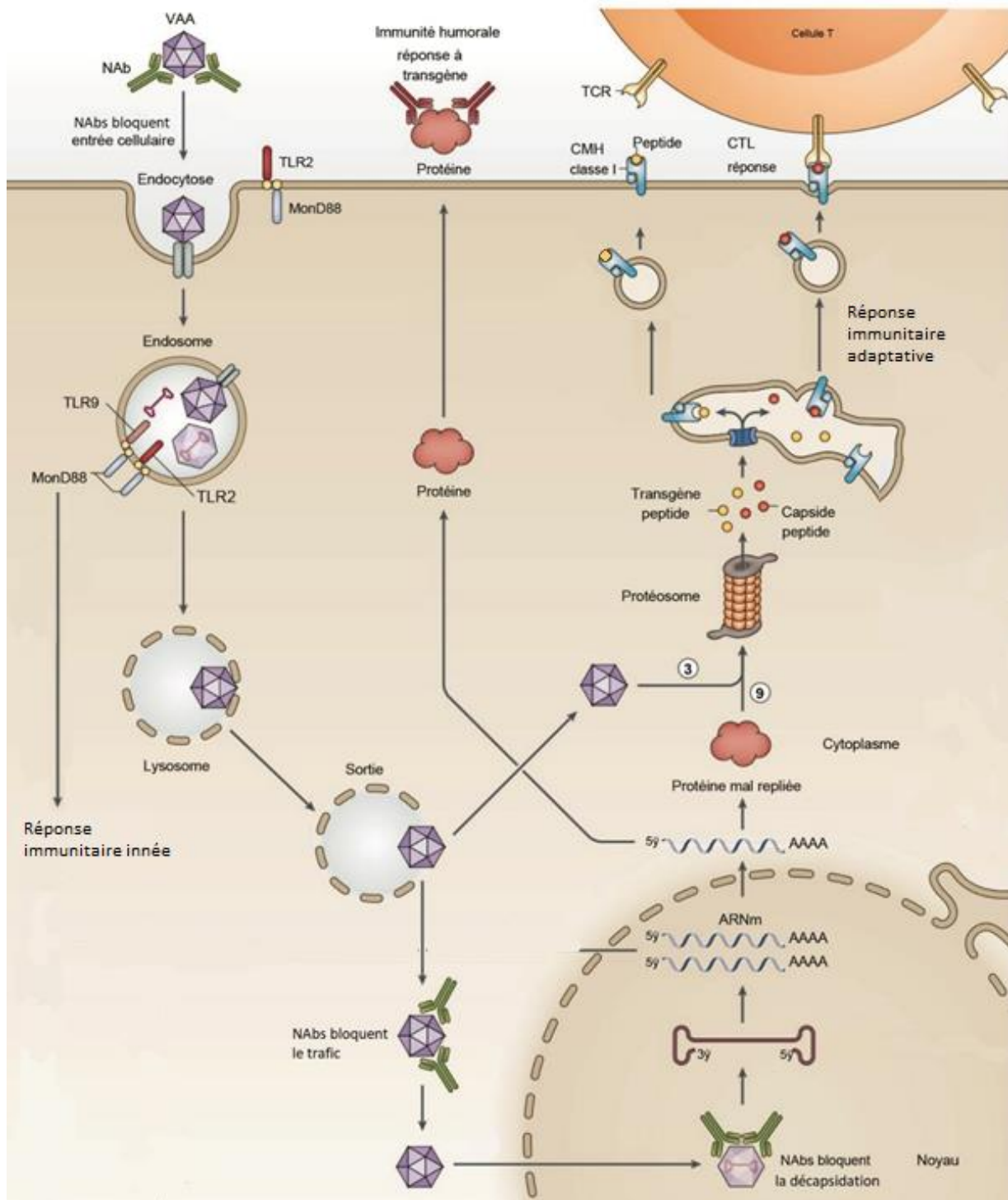


Figure 16 : Réaction immunitaire contre l'AAV.<sup>80</sup>

## *b) Biodistribution et production des AAVr*

Comme nous l'avons précédemment évoqué l'AAV possèdent un tropisme large, ce qui lui permet de transduire de nombreux organes ce qui rend difficile le ciblage spécifique de l'un d'entre eux. Ainsi, cela va induire une transduction non désirée de certaines cellules et donc une diminution de l'effet thérapeutique recherché. De plus, certains tissus non ciblés peuvent être particulièrement sensibles à une infection, comme dans le cas du traitement du cancer si un gène qui promeut la mort cellulaire est transduit dans des cellules saines. Cette biodistribution large des AAV implique alors l'injection de plus fortes doses afin d'avoir une couverture cellulaire maximale, pouvant engendrer une possible réponse immunitaire. A titre d'exemple, le Zolgensma<sup>®</sup> est injecté à une concentration de  $2^{E13}$  vecteurs par mL pour une posologie de  $1,1^{E14}$  vecteurs par kilos. Les plus fortes doses pouvant être injectées vont jusqu'à  $1^{E15}$  par kilo.<sup>81</sup> Bien sur, il faut également prendre en compte le nombre de cellules à transduire, qui peut varier en fonction de la maladie traitée.

Une autre limitation est le coût de production des AAVr sachant que pour obtenir un lot de vecteurs de grade clinique il est nécessaire de remplir un cahier des charges strict. Le principal problème réside dans la présence d'éléments de contamination issus principalement des cellules productrices, comme de l'ADN cellulaire ou diverses protéines. La présence de ces éléments peut induire une réponse immunitaire et donc baisser l'efficacité du vecteur produit. Les techniques de purification engendrent alors un coût important d'autant plus qu'elles sont forcément corrélées à des analyses de contrôle qualité pour vérifier la conformité du lot d'AAVr.

Un autre problème est la présence de vecteurs ne contenant pas ou qu'une partie du transgène. Ce phénomène se produit lors de l'encapsidation et la réplication de l'ADN. La présence de ces virus « vides » est problématique car ils pourront quand même induire une réponse immunitaire.

Ces différents paramètres peuvent expliquer, en partie, les coûts importants des médicaments actuellement sur le marché.

## C. Optimisation des vecteurs

Améliorer les AAVr est une priorité aussi bien scientifique qu'industrielle avec le double défi de cibler plus spécifiquement le tissu ou l'organe souhaité tout en diminuant la dose injectée, pour diminuer les coûts de production et les réactions immunitaires. C'est pourquoi cette thèse portera sur la modification de la capsid de l'AAVr afin d'augmenter son efficacité. Les travaux d'optimisations portant sur le transgène, la production à grande échelle d'AAVr ou encore l'expression du gène médicament ne seront pas abordés dans ce manuscrit.

Il existe cinq grandes approches pour concevoir de nouvelles capsides d'AAVr :

- Le design rationnel
- L'évolution dirigée
- L'approche bio-informatique
- L'insertion de peptide
- L'incubation de peptide

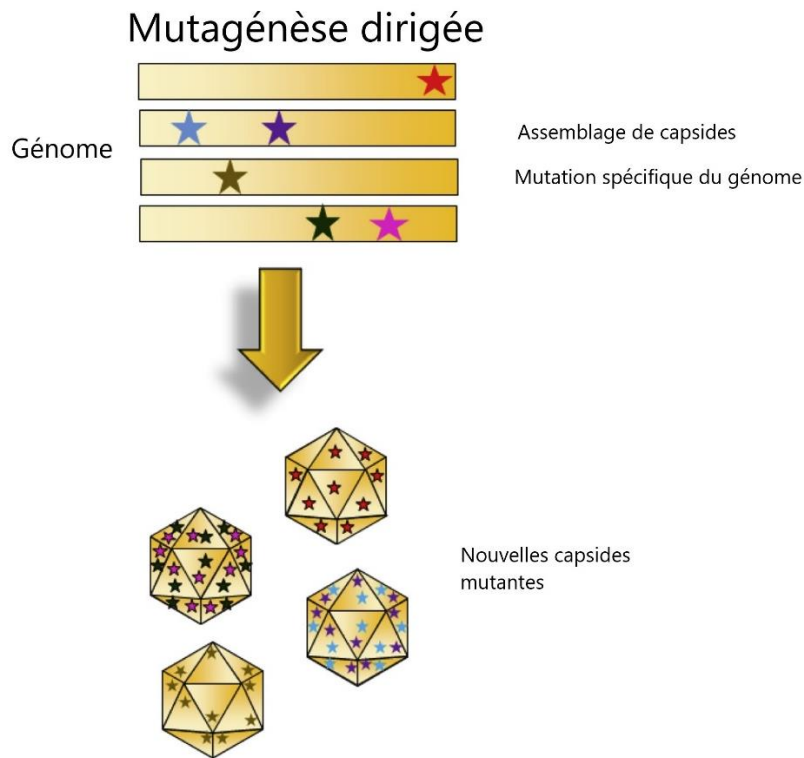
De manière générale, l'optimisation des capsides a pour but :

- De réduire les interactions avec les anticorps neutralisants ou/et les différents TLR.
- D'augmenter l'affinité avec les récepteurs spécifiques d'organes ou de tissus pour une meilleure internalisation ou un tropisme plus efficace.

### *a) Design rationnel*

Le design rationnel est une technique visant à combiner plusieurs capsides de sérotypes connus afin d'en obtenir une nouvelle présentant uniquement les avantages des capsides originelles. Les travaux de Samulski et son équipe, dans le cadre du traitement de la dystrophie musculaire de Duchenne (DMD), en sont un bon exemple. Les chercheurs ont préparé un sérotype optimisé l'AAV2.5 capable de transduire plus efficacement les cellules musculaires. Le nouveau vecteur chimérique combine la capacité de transduction musculaire améliorée de l'AAV1, tout en conservant la liaison aux récepteurs cellulaires qu'utilise l'AAV2.<sup>82</sup> Ce processus a aussi permis de masquer les épitopes qui reconnaissent les anticorps neutralisants et ainsi permettre à ce nouveau sérotype de pouvoir être administré à des patients présentant une réponse immunitaire adaptative à l'AAV2. Ce vecteur a prouvé son efficacité lors d'essais cliniques pour la myopathie de Duchenne, présentant un meilleur ciblage du tissu musculaire et une immunotoxicité plus faible que son homologue (**figure 17**).





*Figure 17 : Illustration du principe de design rationnel.*

### *b) Evolution dirigée*

L'évolution dirigée est basée sur la conception de nouvelles capsides issues de gènes *cap* préparés de façon aléatoire. Concrètement, il existe deux manières de préparer ces nouvelles séquences, la première étant la mutation aléatoire où un gène *cap* d'un sérotype connu est amplifié par méthode PCR (polymérase chain reaction). La particularité ici est l'utilisation d'une polymérase spéciale qui va induire des erreurs de réplication et ainsi générer de nombreuses versions mutées du gène initial. La librairie de séquences ainsi obtenue servira à produire de nouveaux vecteurs mutants pouvant acquérir des capacités uniques de ciblage de cellules ou de résistances aux anticorps/TLR.

La deuxième manière est le DNA shuffling, développée à la fin des années 90, qui consiste à mélanger de façon aléatoire des gènes. Pour cela plusieurs gènes *cap* provenant de sérotypes différents, ou parfois même d'autre parvovirus, sont incubés avec une enzyme de restriction qui va découper les gènes en une multitude de fragments. Une deuxième enzyme sera alors utilisée pour les recombinaison de manière aléatoire donnant accès à de nouveaux gènes *cap*. De la même façon que pour la mutation aléatoire, de nouvelles capsides seront ainsi obtenues. En 2019 Grimm et ses collaborateurs ont produit des virus recombinant par DNA shuffling en combinant l'AAV de sérotype 2 et un bocavirus

(parvovirus des voies respiratoires). Ces nouveaux vecteurs étaient capables d'encapsider un transgène de 5,8 kb et de transduire plus efficacement des types cellulaires pulmonaires.

Ces techniques impliquent le criblage d'un grand nombre d'AAVr qui sont pour la plupart pas ou peu efficaces. Dans les deux cas, la génération de nouvelles capsides est complètement aléatoire et ne repose pas sur une prédiction scientifique et le coût et l'effort investi pour obtenir des capsides d'intérêt est important. De nouvelles techniques de screening haut débit tendent à optimiser cette approche (figure 18).

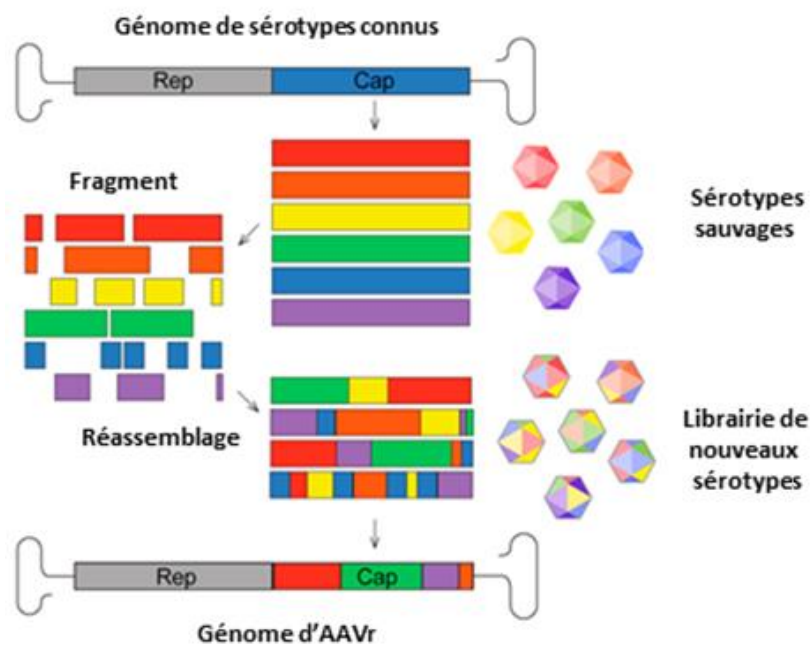


Figure 18 : Illustration du principe de production de nouveaux sérotypes par ingénierie de capsides, méthode DNA shuffling.<sup>83</sup>

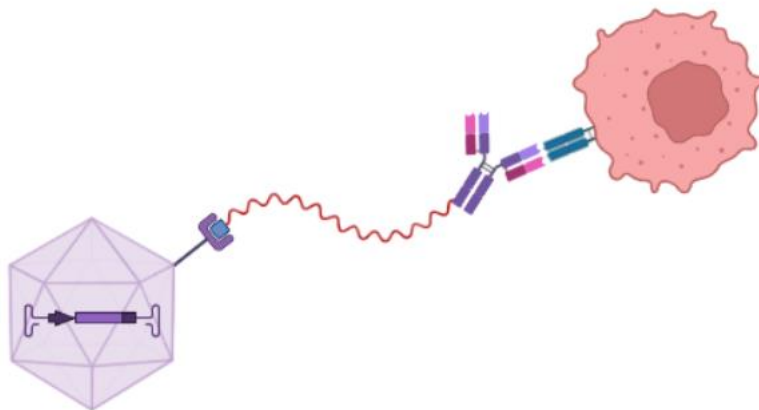
### c) Approche bio-informatique

Certains logiciels de prédiction peuvent permettre, par des méthodes de phylogénie, la génération de capsides ancestrales. A partir de l'ADN de différents sérotypes, il est possible de remonter l'arbre génétique des AAV et ainsi d'accéder à un large choix de capsides. Ces capsides étant ancestrales leurs interactions avec le système immunitaire devraient être limitées. Anc80, conçu en 2015 par l'équipe de Vandenberg, fait partie des vecteurs générés par cette technique. L'intérêt d'avoir de telles structures permet aussi de pouvoir étudier l'évolution des AAV au fil du temps et donc de comprendre certains processus de son infection.<sup>84</sup>

#### d) Peptide display

Le peptide display est une technique qui permet d'ajouter dans la structure de la capside une série d'acides aminés. Ces incorporations sont faites dans une région variable de VP3 (VIII) de manière à être exposées à la surface de la capside. Les peptides insérés sont généralement courts, de 5 à 35 acides aminés, et servent souvent à cibler des récepteurs spécifiques à la surface de cellules. L'avantage de cette approche est qu'elle ne nécessite pas la génération d'une librairie importante de capsides pour être efficace. Büning a décrit en 2022 deux nouveaux variant d'AAV2 présentant une insertion de peptide au niveau des sites de reconnaissance aux héparanes sulfates. Cette insertion provoque une diminution de l'interaction entre le récepteur et l'AAV modifiant ainsi son tropisme. Outre le fait que ces variants transduisent mieux le foie, ils sont également moins reconnus par les anticorps dirigés contre AAV2.<sup>85</sup>

Il est possible également d'insérer des motifs peptidiques plus grands, voir même des protéines. C'est ce qu'a montré l'équipe de Schmidt en encodant un motif protéique, appelé HUH tag, capable de reconnaître de l'ADN simple brin lui-même relié par liaison covalente à un anticorps qui guidera le vecteur vers la cellule cible. Cependant cette insertion est mal tolérée et il est nécessaire de fabriquer des capsides chimériques contenant certaines VP3 modifiées et d'autres non (**figure 19**).<sup>86</sup>



*Figure 19 : Représentation de l'AAVr modifié par Schmidt, l'insertion peptidique permet la reconnaissance avec un brin d'ADN lié à un anticorps permettant de guider le vecteur vers la cellule cible.*

L'inconvénient majeur de cette technique est la nécessité de produire un plasmide contenant le gène *cap* avec la séquence codant pour l'insertion peptidique. Un tel plasmide est parfois difficile à concevoir et peut être un frein.

### e) *Incubation de peptides*

Une autre méthode consiste à utiliser des peptides de pénétration cellulaire (CPP pour Cell-penetrating peptide) pour améliorer l'internalisation de la capsid d'AAV dans les cellules par le processus d'endocytose. Les CPP sont des peptides courts, capables de traverser les tissus et les membranes cellulaires par des mécanismes indépendants ou dépendants de l'énergie.<sup>87</sup> Les CPP sont considérés comme des vecteurs d'administration de médicaments parce qu'ils peuvent se lier à une grande variété de conjugués biologiquement actifs, y compris des protéines, des peptides, des ADN, des petits ARN interférents (siRNA) et des petites molécules d'ADN. Cette technique récente a été utilisée par l'équipe de Ying Liu en 2021 où le peptide LAH4 incubé avec de l'AAV9 a permis une meilleure transduction des astrocytes humains *in vivo*.<sup>88</sup>

### 3. Optimisation chimique des AVV

Les approches que nous venons de décrire utilisent des techniques de génie génétique afin de générer de nouvelles capsides aux propriétés inédites. Elles ont largement fait leurs preuves et permettent aujourd'hui de produire bons nombres d'AAVr utilisés dans différents essais cliniques. Cependant, elles présentent le désavantage de générer énormément d'échantillons et de devoir ensuite tous les évaluer avant de trouver la capside optimale. Néanmoins, l'insertion ou l'incubation de peptides apparaît comme des options intéressantes mais nécessitent encore des optimisations notamment au niveau de la production, de la purification et de la caractérisation.

Parallèlement, une nouvelle stratégie d'optimisation a peu à peu vu le jour qui combine la chimie et la vectorologie et qui consiste à greffer de façon covalente de petites molécules d'intérêts à la surface de la capside d'AAV.

#### A. Principe

La modification chimique des AAV consiste à ajouter sur la capside des molécules ayant une propriété particulière (biologique ou physico-chimique) qui permet de modifier le tropisme du vecteur. Cet ajout de composés chimiques se fait par réaction de bioconjugaison bioorthogonale grâce aux acides aminés composant les VP de la capside. Un des atouts majeurs de la modification chimique est la possibilité d'introduire sur une même capside diverses molécules telles que des polymères, des peptides, des dérivés glucidiques ou encore des lipides ne pouvant pas être incorporés génétiquement.

##### *a) La bioconjugaison*

La bioconjugaison est une approche chimique visant à lier de manière covalente deux molécules entre elles dont l'une au moins est une biomolécule. Cette technique trouve des applications dans de nombreux domaines comme la médecine ou les matériaux. Dans ce manuscrit de thèse nous allons nous concentrer sur la bioconjugaison de protéines dans un but thérapeutique.

Pour réaliser une réaction de bioconjugaison sur une protéine il est nécessaire de connaître sa structure primaire, c'est-à-dire les acides aminés qui la composent, car ce sont eux qui vont former une liaison covalente avec la molécule d'intérêt. Il existe 20 acides aminés naturels et tous ne présentent pas les mêmes caractéristiques, la même abondance ou encore la même localisation dans et sur la biomolécule (figure 20).

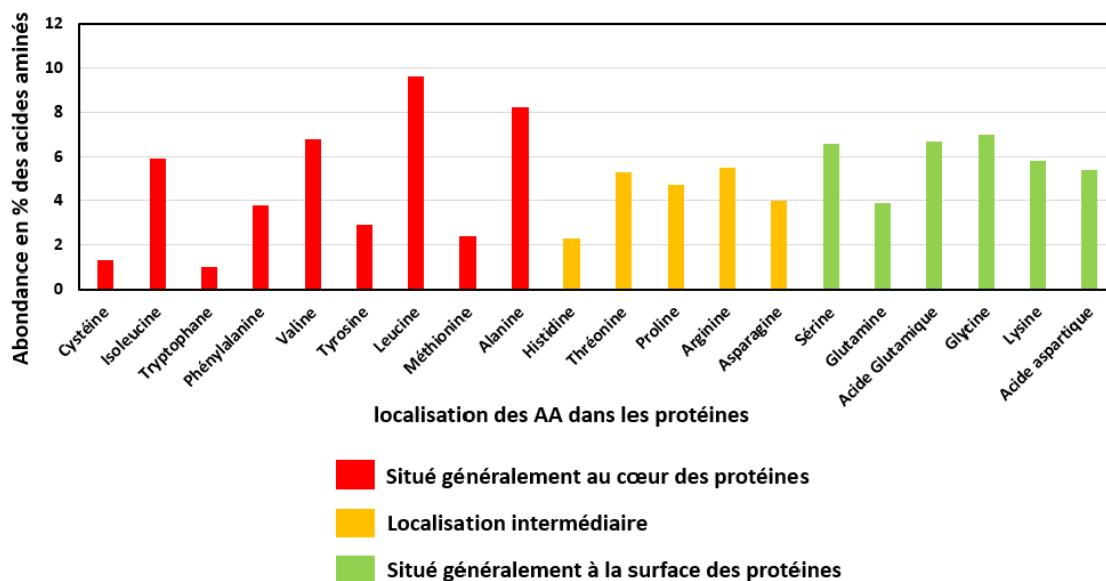


Figure 20 : Abondance naturelle des différents acides aminés composant les protéines ainsi que leur localisation au sein de cette dernière.<sup>89,90</sup>

Deux critères sont importants pour modifier efficacement un acide aminé. Il doit être disponible et donc ne pas être situé à l'intérieur de la structure tridimensionnelle de la protéine et il doit posséder une chaîne latérale fonctionnalisable dans des conditions compatibles avec la biomolécule. De manière générale les acides aminés aliphatiques/aromatiques sont difficilement modifiables, de même que les acides aminés anioniques. D'autre part, les conditions de la réaction de bioconjugaison devront être compatibles avec la biomolécule sans la dénaturer ou altérer ses fonctions et sa structure. Actuellement trois acides aminés sont largement utilisés pour des réactions de bioconjugaison. Il s'agit de la cystéine, de la tyrosine et de la lysine.

(1) Modification de la cystéine :

La cystéine est l'un des acides aminés le moins abondant et le moins disponible à la surface des protéines. Néanmoins, le groupement thiol est partiellement sous la forme thiolate à pH physiologique ce qui en fait un candidat de choix pour des réactions d'addition ou de substitution nucléophile avec comme partenaire de mauvais électrophiles limitant ainsi les réactions secondaires. A titre d'exemple, les maléimides sont largement utilisés comme accepteur de Michael pour venir sélectivement modifier les cystéines,<sup>91</sup> même si cette réaction présente des inconvénients. En effet, le produit formé peut notamment subir une réaction de rétro-Michael reformant ainsi un thiol libre et le maléimide correspondant.<sup>92</sup> Pour pallier ce problème d'autres électrophiles comme les dérivés carbonyles  $\alpha,\beta$ -insaturés fonctionnalisés développés par l'équipe du professeur Bernardim<sup>93</sup>, ou les vinyles sulfones peuvent être avantageusement utilisés (figure 21).<sup>94</sup>

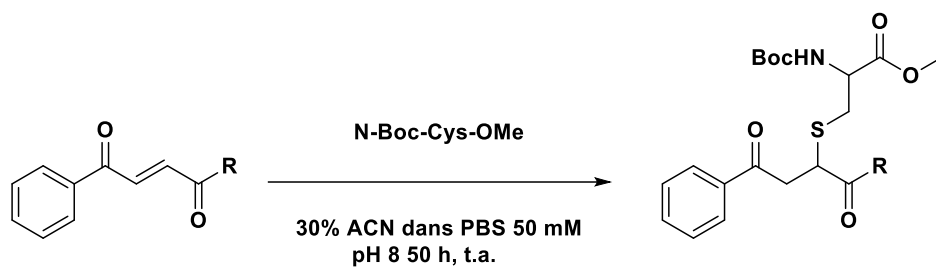


Figure 21 : Réaction d'un dérivé carbonyle  $\alpha,\beta$ -insaturé développé par l'équipe du professeur Bernardin<sup>93</sup>

## (2) Modification de la tyrosine :

Les résidus tyrosine sont relativement peu réactifs, cependant cet acide aminé voit son intérêt grandir au fil des années. Historiquement les sels de diazonium ont été largement utilisés en réagissant avec le noyau phénolique de la tyrosine par réaction de substitution électrophile aromatique.<sup>95</sup> Ces dernières années, le Professeur Barbas a décrit la réactivité de dérivés cycliques de diazodicarboxamide tels que les phényltriazolinediones (PTAD) très réactives vis-à-vis des phénols selon une réaction péricyclique.<sup>96</sup> L'inconvénient de cette approche est qu'il est nécessaire de partir d'urazole qu'il faut oxyder en PTAD en présence de la biomolécule. Très récemment, notre équipe a développé une approche de couplage bioorthogonale par électrochimie sans oxydant chimique directement à partir d'urazole. Cette stratégie innovante a été améliorée en remplaçant l'urazole par des dérivés du méthylluminol, ce qui a permis la fonctionnalisation d'anticorps, d'enzymes ou de cellules (figure 22).<sup>97</sup>

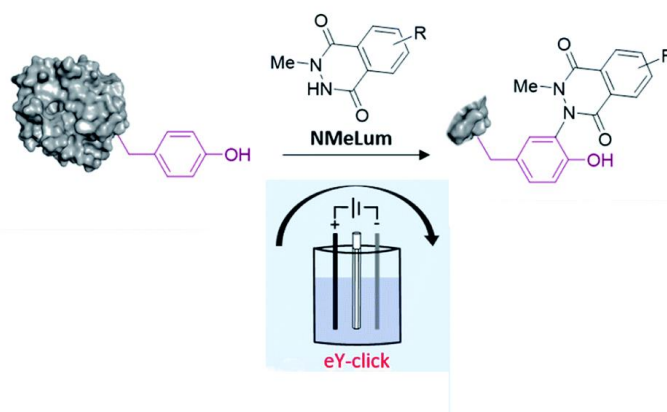


Figure 22 : Principe de la technique de bioconjugaison eY-click développée par l'équipe du Professeur Gouin<sup>98</sup>

### (3) Modification de la lysine :

La lysine est l'acide aminé le plus utilisé pour les réactions de bioconjugaison étant donné son abondance et la réactivité de la fonction amine primaire de sa chaîne latérale. Un point important avec les fonctions amine des lysines est le fait qu'elles soient protonées à pH physiologique ce qui diminue leur réactivité. Ainsi, les couplages se font principalement en milieu basique ce qui implique que la biomolécule ou le ligand ne soient pas sensibles à ce pH.<sup>99</sup>

Classiquement les résidus de lysine réagissent sur des esters de *N*-hydroxysuccinimides afin d'effectuer un couplage peptidique avec formation d'une fonction amide stable en milieu physiologique. On peut également citer l'utilisation d'azalactone comme électrophile vis-à-vis des amines en raison d'une cinétique de réaction très rapide et notamment pour la construction de plateformes multivalentes.<sup>100-</sup>  
<sup>102</sup> Les fonctions iso(thio)cyanate sont aussi très efficaces et forment un lien (thio)urée solide entre la biomolécule et le ligand greffé.<sup>103</sup>

## B. Bioconjugaison de capsid d'AAVr

Le principal avantage de la bioconjugaison pour obtenir de nouvelles capsides est la possibilité d'incorporer un large panel de molécules ce qui ne serait pas possible par ingénierie biologique. On peut citer les sucres, les acides gras, les sondes ou encore les agents chélatants de métaux. Pour simplifier, il est théoriquement possible de tout coupler à la surface de la capsid dès lors que la méthode de bioconjugaison n'altère pas la fonction première du vecteur.

La première étude utilisant la bioconjugaison sur un AAV2 a été réalisée en 2005 par l'équipe de Schaffer dans le but d'ajouter des PEG (polyéthylène glycol) à la surface du virus. Partant du constat qu'une grande partie de la population possède des anticorps anti AAV2, les chercheurs se sont intéressés à la pégylation de sa capsid afin de masquer les épitopes que reconnaissent les anticorps neutralisants. Trois PEG de différentes longueurs ont été utilisés dans des proportions comprises entre 1 et 10000 équivalents (eq) par AAV avec à l'une de leurs extrémités une fonction ester de succinimidyle pour un couplage sur les lysines (**figure 23**).<sup>104</sup>



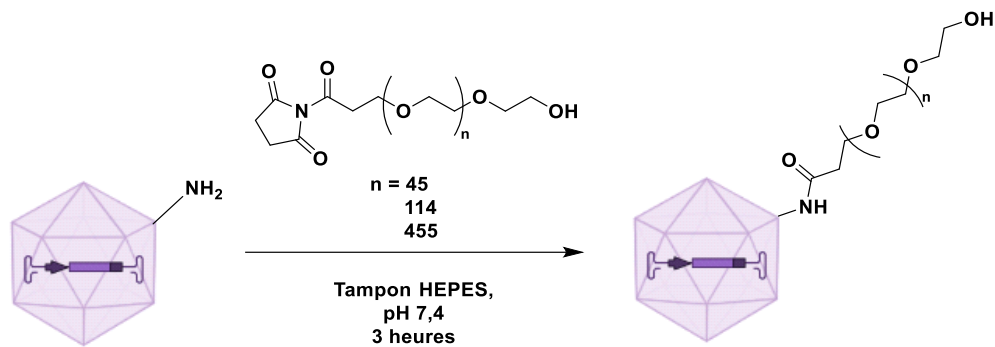


Figure 23 : Réaction de bioconjugaison entre le résidu amine des lysines de l'AAVr et un PEG activé.

Une fois le couplage réalisé, l'infectiosité du vecteur viral a été mesurée sur des cellules HEK293 et a montré l'influence du nombre de ligands sur la capsid. Au-dessus de 1000 eq de PEG, le vecteur perd sa capacité infectieuse et cela indépendamment de la longueur du PEG. En revanche, les chercheurs ont montré que c'est avec la plus grande quantité de motif PEG que l'on obtient le plus de résistance face aux anticorps neutralisants. Il y a donc une dualité entre le fait de trop fonctionnaliser la capsid, au risque de perdre le pouvoir infectieux du vecteur, et de suffisamment la modifier pour avoir une protection efficace envers les anticorps.

Ce travail précurseur indique qu'il est possible de modifier chimiquement un AAV tout en conservant son infectiosité, générant ainsi de nouveaux AAVr aux propriétés différentes du virus natif.

Un deuxième exemple de modification chimique, décrit en 2011 par l'équipe d'Asokan, porte sur le couplage des arginines, *via* leur groupement guanidine, par du méthylglyoxal et cela directement sur la capsid assemblée d'un AAV2 (MGO-AAV2). Cette modification entraîne une diminution de l'interaction entre l'AAV2 et les récepteurs aux héparanes sulfates qui sont impliqués pour infecter les cellules. Une étude *in vivo* a mis en avant la modification du tropisme de l'AAV2 après fonctionnalisation avec une diminution de l'infection du foie et une augmentation pour le muscle squelettique et le cœur (figure 24).<sup>105</sup>

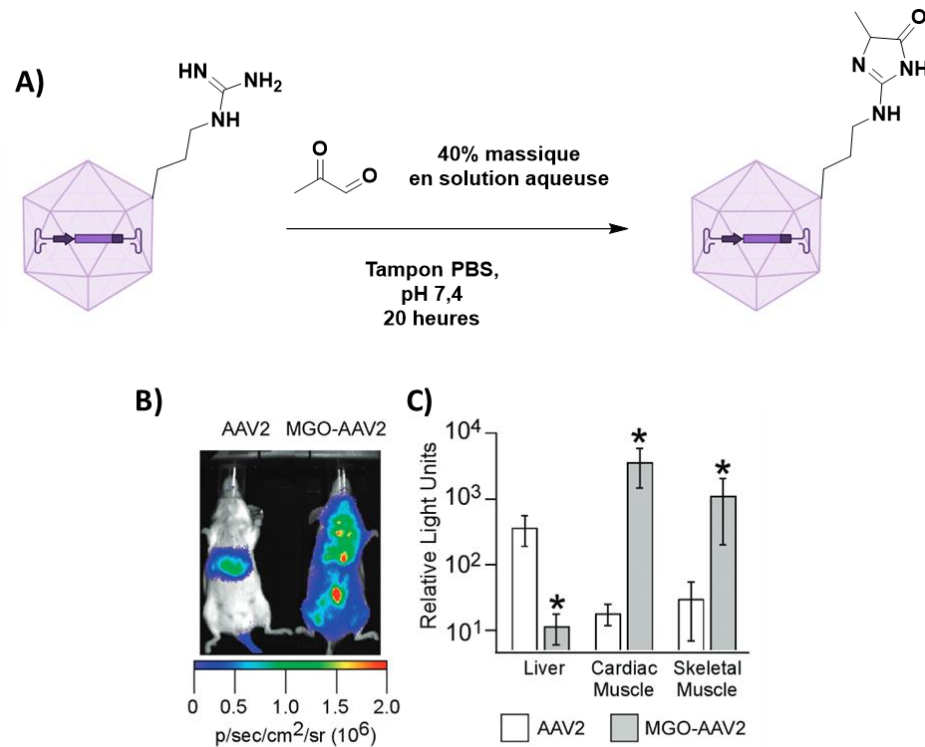


Figure 24 : a) Réaction de bioconjugaison entre le résidu guanidine des arginines de l'AAVr et le méthylglyoxal formant une dihydroimidazolone. B) image de bioluminescence de la luciférine sur souris injectées avec le vecteur AAV modifié MGO-AAV2 (à droite) et l'AAVr de sérotype 2 (à gauche) C) expression de la luciférase dans trois tissus prélevés avec diminution du tropisme pour le foie et augmentation de celui pour le cœur et le muscle squelettique.<sup>105</sup>

Plus récemment, l'équipe de Ferrara a utilisé la bioconjugaison comme outil afin de suivre, à l'aide d'un radionucléide, l'infection d'un AAVr dans le cerveau de souris. L'AAV9 PHP.eB est un vecteur AAV obtenu par évolution dirigée qui présente un tropisme accru pour le cerveau, *via* son interaction spécifique avec le récepteur LY6A présent à la surface des cellules endothéliales cérébrales. Dans cette étude cet AAVr a été modifié chimiquement avec le (NOTA)<sub>8</sub>-MI, un composé complexant le cuivre (figure 25).

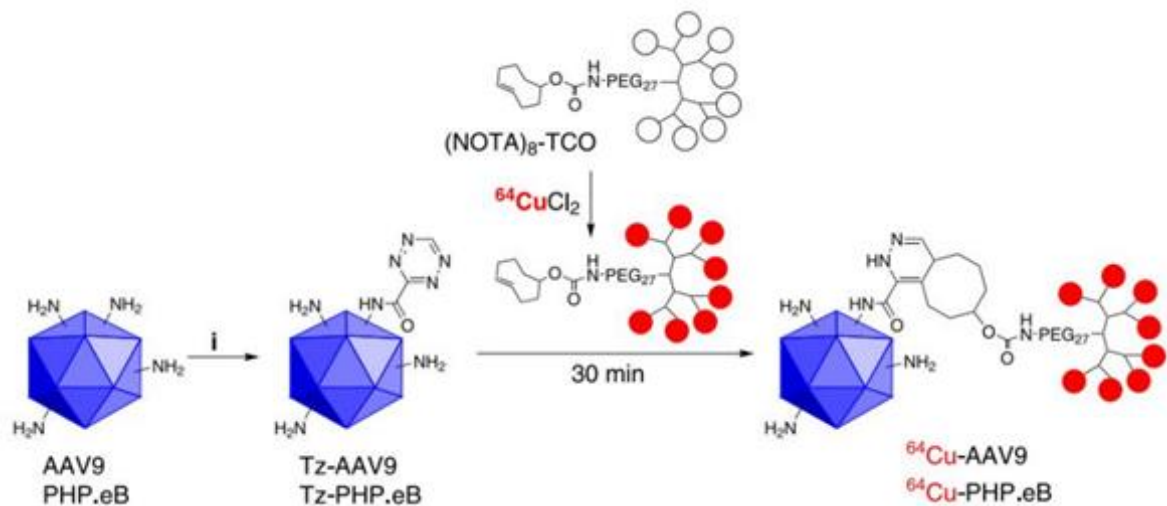


Figure 25 : Réaction de bioconjugaison sur l'AAV9 et l'AAV-PHP.eB des lysines par une tétrazine puis réaction IEDDA avec un alcène contraint porteur de complexants du cuivre. Les AAVr chimiquement modifiés possèdent à leurs surface des ligands du  $^{64}\text{Cu}$  pour de l'imagerie.<sup>106</sup>

Cette nouvelle génération d'AAVr a été obtenue dans un premier temps après modification des lysines par bioconjugaison avec un ester activé pour introduire une tétrazine. Dans une deuxième étape, la tétrazine réagit alors selon une réaction de Diels-Alder à demande inverse (IEDDA) avec un alcène contraint porteur de chélatants du cuivre. L'ajout de cuivre radioactif a permis de faire de l'imagerie par émission de positrons et ainsi de suivre les AAVr dans l'organisme des souris.<sup>106</sup> Les chercheurs ont ainsi mis en évidence l'accumulation dans le cerveaux du mutant PHP.eB.

Il est aussi possible de combiner la technique du peptide display pour introduire un acide aminé non naturel à la surface du vecteur et de venir par la suite réaliser une réaction de bioconjugaison sur ce dernier. En 2016 l'équipe de Chatterjee a produit un AAV2 contenant un acide aminé non naturel sur chacune de ses VP, l'AzK, qui présente une fonction azoture en fin de chaîne latérale. Les chercheurs ont ensuite procédé au couplage par réaction bioorthogonale entre l'azoture de l'AzK et un peptide fonctionnalisé avec un dibenzocyclooctyne. Ce peptide a permis à l'AAV2 de transduire plus efficacement les cellules contenant le domaine des intégrines  $\alpha_v\beta_3$ .<sup>107</sup>

### C. Travaux antérieurs à la thèse

Notre équipe travaille depuis plusieurs années sur la modification de capsides virales par bioconjugaison. Elle a commencé par l'étude de couplage des lysines d'AAV2 par un fluorophore porteur d'une fonction phénylisothiocyanate (FITC). La réaction de bioconjugaison se fait dans un tampon TBS à pH 9,3 pendant 4h. Deux concentrations de fluorophore ont été utilisées l'une de  $3^5$

équivalents de molécules par vecteur appelée « faible concentration » (**f**) et l'une de  $3^{E6}$  appelée « forte concentration » (**F**). Le mélange est ensuite dialysé afin d'éliminer le réactif en excès et de passer à un tampon dPBS à pH 7,4.

Après validation d'un couplage covalent par des techniques de Dot et Western-blot, l'efficacité de transduction *in vitro* des vecteurs modifiés a été effectuée sur des cellules HeLa. L'AAV contenant un gène rapporteur codant pour la GFP, il a été possible de déterminer par cytométrie de flux (analyse FACS) le pourcentage des cellules exprimant la GFP. L'expérience montre que les AAVr modifiés aux deux concentrations de fluorophore sont toujours infectieux après couplage. Par la suite, de l'imagerie confocale a été réalisée sur ces échantillons, la (**figure 26**) représente les images obtenues.<sup>108</sup>

L'intensité de la fluorescence de FITC a été directement détectée (vert) tandis que la capsid de l'AAV2 a été marquée par l'anticorps A20 suivi de l'anticorps secondaire AI647 (rouge). Comme prévu, la capsid assemblée de l'AAV2 n'est visible que par l'anticorps A20 (Fig. 26-B). La présence de FITC et de la capsid assemblée a été détectée pour les conditions de couplage les plus faibles (rapport molaire  $3^{E5}$ ), comme le montre la figure 26-F. La colocalisation des deux signaux fluorescents a confirmé l'intégrité de la capsid de l'AAV2, la présence de FITC à la surface de l'AAV, la reconnaissance de la capsid par l'anticorps A20 et son infectivité (c'est-à-dire l'entrée cellulaire) de ces nanoparticules. Pour les conditions de rapport molaire  $3^{E6}$ , la présence de FITC a été clairement montrée par analyse confocale (Fig. 26-G) indiquant un couplage chimique efficace. Cependant, la capsid assemblée n'a pas (ou peu) été reconnue par l'anticorps A20. Cela confirme que le masquage chimique par les molécules de FITC sur les domaines exposés à la surface peut modifier l'interaction des vecteurs AAV2-FITC avec un anticorps. Un tel AAV chimiquement fonctionnalisé pourrait donc potentiellement échapper aux anticorps neutralisants préexistants dans la population humaine tout en étant infectieux. Dans l'ensemble, ces résultats intéressants démontrent la possibilité de modifier les propriétés de ciblage des AAV en utilisant des modifications chimiques à différents rapports molaires pendant la conjugaison.

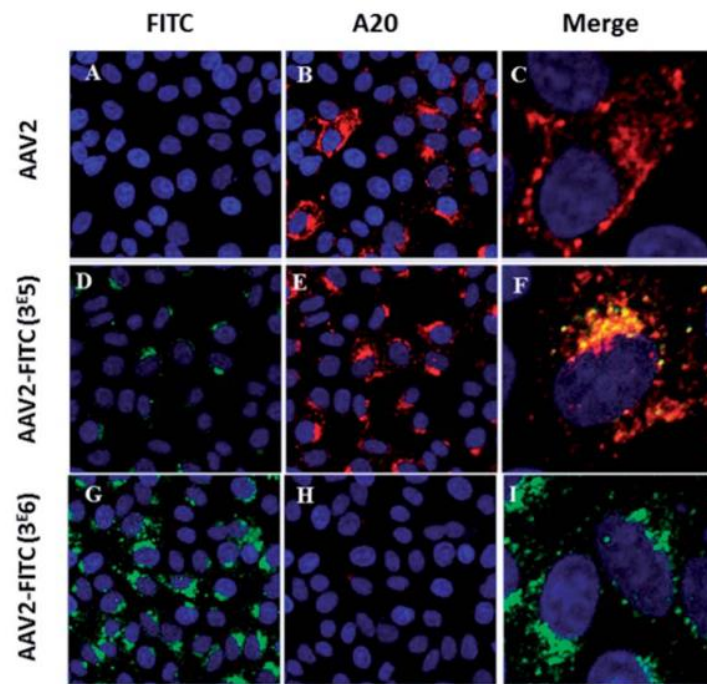


Figure 26 : Images confocales de cellules HeLa transduites avec les échantillons d'AAV2 modifiés ou non.

Après cette preuve de concept, des ligands bifonctionnels ont été synthétisés et greffés sur les lysines afin de cibler spécifiquement un type cellulaire. Concrètement les ligands sont formés d'un phénylisothiocyanate relié à un mannose par un espaceur triéthylène glycol. Le mannose a été choisi pour cibler les cellules RPE de la rétine qui possèdent des récepteurs aux mannoses (figure 27).<sup>109</sup>

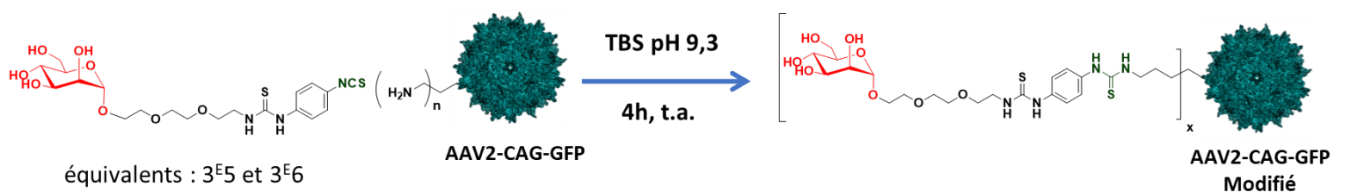


Figure 27 : Réaction de bioconjugaison entre le ligand bifonctionnel et l'AAV2 contenant un gène codant pour la protéine fluorescente GFP.

Comme pour l'expérience avec FITC, deux concentrations de ligands ont été utilisées pour le couplage, une faible et une forte. Après greffage, notre équipe a validé un couplage covalent, et non pas une adsorption des ligands à la surface de la capsid, et que ces vecteurs sont toujours infectieux *in vitro* sur des cellules HeLa. L'efficacité *in vivo* de ces particules dans la rétine de rats a alors été menée par injection sous rétinienne. Une imagerie du fond de l'œil montre, de façon spectaculaire, une augmentation de l'intensité et de la diffusion de la GFP au cours du temps dans le cas des rats qui ont

reçu l'AAV fonctionnalisé. Quatre semaines après l'injection, les rats sont sacrifiés puis l'ARN codant pour la GFP est dosé dans les différentes cellules de la rétine afin de confirmer l'analyse visuelle. Il a été montré que sa quantité est plus importante pour les souris traitées avec l'AAV2 modifié à 3<sup>E6</sup> eq qu'avec l'AAV2 non modifié (figure 28). Les résultats obtenus avec la bioconjugaison avec la faible concentration sont équivalents à l'AAV2.

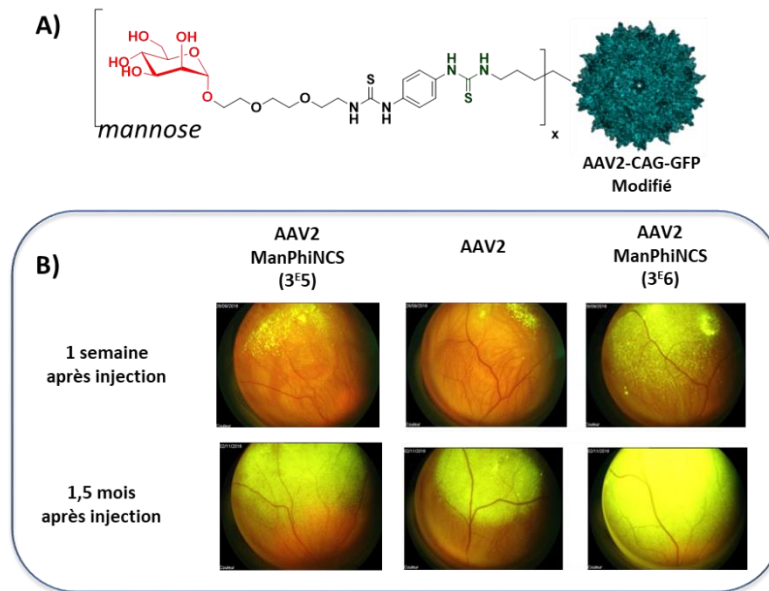


Figure 28 : Images de rétines de rats injectés avec le vecteur AAVr modifié et non modifié. A) structure du ligand lié à la capsid par bioconjugaison. B) Les clichés permettent d'observer une fluorescence due à la GFP bien plus importante chez les rats injectés avec vecteur fortement modifié.<sup>109</sup>

Ainsi, l'équipe a validé la faisabilité de modifier chimiquement les lysines de la capsid d'un AAV avec une transduction plus efficace qu'un AAV conventionnel.

## 4. Ciblage du foie

Après les résultats obtenus sur la rétine, l'équipe s'oriente maintenant sur la faisabilité de cibler d'autres organes en modifiant le ligand apporté par bioconjugaison. Le foie est une cible de choix car il est le foyer de plusieurs maladies génétiques comme les hémophilies, la maladie de Crigler-Najjar ou encore l'hémochromatose.<sup>110</sup> C'est également un organe facilement atteignable par injection systémique. Dans un premier temps, nous allons décrire cet organe ainsi que des exemples de maladies génétiques et leurs potentiels traitements. Dans un deuxième temps, un récepteur caractéristique, largement utilisé comme porte d'entrée pour le traitement des maladies hépatiques sera décrit. Enfin, le contexte général de la thèse sera introduit.

### A. Le foie

Le foie, qui appartient au système digestif, est l'organe le plus volumineux dans le corps humain (en moyenne 1,5 kg) (figure 29). De plus, il est celui qui effectue le plus grand nombre de transformations chimiques et qui présente des propriétés endocrines et exocrines.

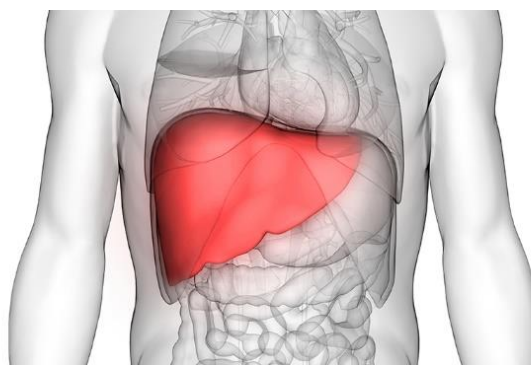


Figure 29 : Illustration montrant l'emplacement du foie (en rouge) dans le corp humain.

Il est impliqué dans plus de 300 fonctions vitales, parmi les plus connues on peut citer le métabolisme des glucides et lipides (glycogénèse, synthèse de cholestérol, production d'acides gras), la synthèse de protéines sanguines (albumine, globines, apoprotéines, facteurs de coagulation), la détoxification ou la production de la bile.<sup>111</sup>

D'un point de vue structural le foie est composé de plusieurs types de cellules, on retrouve les cellules de Küppfer (macrophage), les cellules de Ito qui peuvent se transformer en myofibroblastes, ou encore les cellules endothéliales et pluripotentes. Cependant, les principales fonctions effectuées par le foie

se réalisent au sein des hépatocytes qui sont de véritables usines biochimiques et qui représentent plus de 70% de l'ensemble des cellules.<sup>112</sup>

Les maladies rares du foie concernent plus de 10 000 personnes en France, aussi bien des adultes que des enfants chez qui le début est souvent avant l'âge de deux ans. En pédiatrie, le pronostic vital est en jeu dans presque la moitié des cas, expliquant 35% des décès avant l'âge d'un an, 10% entre 1 et 5 ans et 12% entre 5 et 15 ans. On dénombre aujourd'hui plus de 40 maladies rares du foie adultes et pédiatriques.<sup>113</sup>

Parmi certaines maladies génétiques du foie on retrouve, le syndrome de Crigler-Najjar, la maladie de Wilson ou encore la glyco-génose. Les plus répandus étant les hémophilies et l'hémochromatose.

### *a) Les hémophilies*

La coagulation est un phénomène complexe qui intervient dans l'hémostase. Elle met en œuvre des facteurs de coagulation qui s'activent en cascade pour former des caillots permettant de stopper une hémorragie. L'hémophilie est une anomalie constitutionnelle de la coagulation sanguine due à un déficit d'un des facteurs de coagulation qui se traduit par une impossibilité pour le sang de coaguler.<sup>114</sup>

On connaît à l'heure actuelle trois types d'hémophilie en rapport avec des facteurs de coagulation déficitaires :

- L'hémophilie A (« hémophilie classique ») correspond à un déficit du facteur de coagulation VIII et est provoquée par une mutation du gène F8 présent sur le chromosome X. Cette maladie représente 1 cas sur 5000 naissances de garçon (maladie rarissime chez les filles).
- L'hémophilie B (ou maladie de Christmas) est due quant à elle à un déficit du facteur de coagulation IX suite à une mutation du gène F9 du chromosome X. Les garçons sont également quasi les seuls touchés avec 1 cas pour 25 000 naissances.
- L'hémophilie C (ou maladie de Rosenthal) correspond à un déficit congénital en facteur de coagulation XI avec une prévalence de 1/1 000 000 et n'est pas liée au sexe.

L'hémophilie ne se guérit pas mais elle se contrôle bien par administration de facteurs de coagulation fonctionnels qui permettent d'obtenir une activité coagulante suffisante pour arrêter, voire prévenir, l'hémorragie.

Il existe plusieurs pistes de recherche pour soigner durablement l'hémophilie, parmi elles on peut citer l'Emcizumab qui est un anticorps monoclonal développé et commercialisé en 2017 par la filiale Roche, Ce dernier est capable de se lier au facteur de coagulation IX et X, ce rôle étant normalement effectué par le facteur VIII, Emcizumab est donc indiqué pour les patients souffrants d'hémophilie de



type A.<sup>115</sup> En 2022 deux nouveaux médicaments de thérapie génique utilisant des AAV ont obtenu une autorisation de mise sur le marché Roctavian<sup>®</sup> pour l'hémophile A et Hemgenix<sup>®</sup> pour le traitement de l'hémophilie B.<sup>116–118</sup>

## b) Hémochromatose

L'hémochromatose est la maladie génétique la plus fréquente du foie. Elle est causée par le dysfonctionnement ou l'absence de l'hepcidine, une hormone peptidique qui régule la concentration de fer dans le sang. Les patients atteints d'hémochromatose développent une hyperabsorption intestinale de fer provoquant des dépôts de fer dans l'organisme qui détruisent peu à peu les organes.<sup>119</sup>

Le dysfonctionnement de l'hepcidine est principalement dû à une mutation du gène *hfe* qui code principalement pour la protéine HFE qui permet l'absorption du fer. La mutation la plus courante, 90% des cas d'hémochromatose, est la substitution de la cystéine 282 par une tyrosine.<sup>120</sup>

On estime qu'une personne sur 300 est porteuse de la principale anomalie génétique prédisposant à l'apparition de cette pathologie, les hommes étant trois fois plus touchés que les femmes. De l'ordre de 200 000 personnes seraient porteuses de cette prédisposition génétique en France, 2 200 000 en Europe et 2 millions aux États-Unis.

Il existe, à l'heure actuelle, très peu de traitements pour l'hémochromatose, le principal étant la réalisation d'une saignée qui correspond habituellement à celui d'un don du sang (250 à 400 ml). Les autres traitements consistent en l'absorption de chélateur du fer comme la déféroxamine. Depuis la découverte dans les années 2000 de la structure de l'hepcidine, des travaux portent sur sa conception mais ces derniers sont peu avancés car le peptide de 25 acides aminés présente quatre ponts disulfure qui mettent en échec la plupart des synthèses (figure 30).<sup>121</sup>

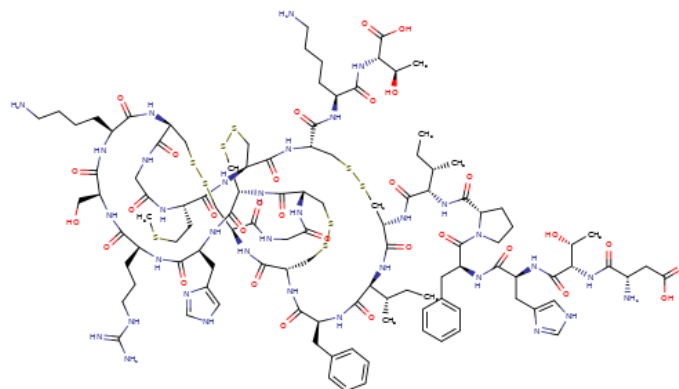


Figure 30 : Structure de l'hepcidine.<sup>122</sup>

## B. Les hépatocytes

### a) Généralités

Les hépatocytes ou cellules hépatiques sont des cellules parenchymateuses du foie assurant la quasi-totalité de ses fonctions métaboliques. Elles représentent 70% des cellules de l'organe.<sup>123</sup>

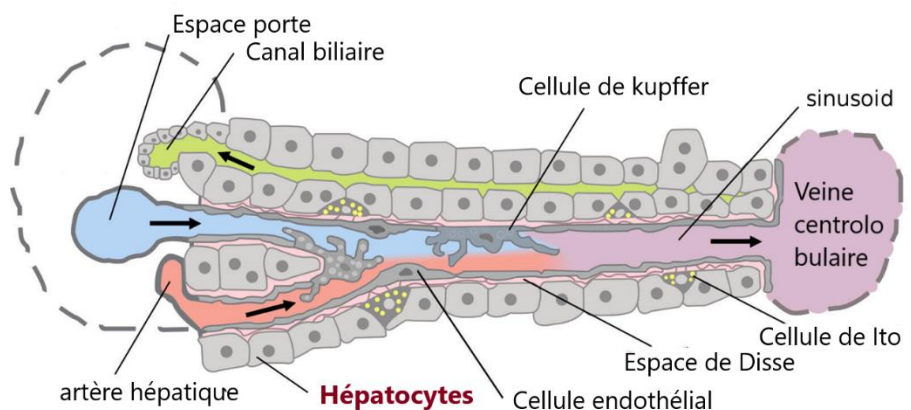


Figure 31 : Illustration représentant une travée de Remak d'un lobule du foie, les hépatocytes sont indiqués en rouge.<sup>124</sup>

Les hépatocytes sont organisés en travées (appelées travées de Remak) entre l'espace porte (comprenant une triade de vaisseaux, la veine porte, l'artère hépatique et un canal biliaire) et la veine centro-lobulaire. De forme polyédrique, les hépatocytes mesurent entre 20 et 30  $\mu\text{m}$  de côtés, ils sont exposés d'une part au sang et d'autre part aux réseaux de canicule biliaire. Cette architecture permet une irrigation maximale des hépatocytes, ce qui facilite le transfert des sécrétions dans la circulation sanguine (figure 31).

Une particularité des hépatocytes est leur aneuploïdie, c'est-à-dire qu'ils possèdent un nombre anormal de chromosomes (jusqu'à 8 jeux de chromosomes) ce qui facilite probablement la régénérescence du foie de plus ils sont souvent binucléés.<sup>125</sup>

Les hépatocytes possèdent un grand nombre de récepteurs transmembranaires qui captent les différents composés de métabolisation transportés par le sang. Parmi les récepteurs transmembranaires, on peut citer le récepteur aux asialoglycoprotéines massivement présent à sa surface et qui permet notamment de reconnaître des glycoprotéines du même nom.<sup>126,127</sup>

## C. Récepteurs ASGP

Le récepteur aux asialoglycoprotéines (ASGPr) a été identifié pour la première fois dans les années 1960 par Ashwell *et al.* Il a été isolé lors d'une étude sur le métabolisme des glycoprotéines du plasma de lapin et appartient au groupe des lectines de type C, protéines possédant un ion calcium dans leur site de reconnaissance (les lectines sont des récepteurs interagissant avec des sucres). Il est majoritairement présent à la surface des hépatocytes avec en moyenne  $1-5 \times 10^5$  copie.<sup>128,129</sup> L'ASGPr joue un rôle de médiateur dans l'endocytose des glycoprotéines du sérum au niveau du foie, il reconnaît des ligands de type Galactose (Gal) et *N*-acétylgalactosamine (GalNAc), avec une forte affinité pour ce dernier.<sup>130,131</sup>

Les fonctions de ce récepteur sont multiples, il permet notamment de maintenir l'homéostasie des glycoprotéines sériques et assure l'internalisation et l'élimination subséquente d'une large gamme de glycoprotéines portant des polysaccharides avec des motifs Gal ou GalNAc terminaux. Les glycoprotéines sont internalisées par endocytose puis le complexe glycoprotéine-ASGPr formé (*via* des puits recouverts de clathrine) se dissocie après acidification de l'endosome. Le récepteur est alors réacheminé à la surface de la cellule et la glycoprotéine est dégradée après la maturation (par acidification) de l'endosome en lysosome, ce cycle intervient en moyenne toute les 15 minutes.<sup>132-134</sup>

Il a également été démontré que l'ASGPr était impliqué dans l'élimination des IgA de la circulation sanguine, dans l'élimination des cellules apoptotiques, des lipoprotéines de basse densité (LDL), des restes de chylomicrons et de la fibronectine cellulaire. Des données récentes soutiennent l'hypothèse que l'ASGPr est utilisé par les virus hépatotropes pour infecter les hépatocytes. L'ASGP est donc un récepteur important des hépatocytes et une véritable porte d'entrée pour transférer un gène dans ces cellules.

D'un point de vue structurel l'ASGPr de mammifère est composé de deux sous-unités, une principale H1 et une secondaire H2. Chez l'homme, H1 et H2 ont des tailles respectives de 46 et 50 kDa et possèdent 58% de similitude, la principale différence s'effectue sur 18 acides aminés dans le domaine cytoplasmique. Chacune de ces sous-unités possède un domaine de reconnaissance des sucres (CDR) faisant intervenir un cation calcium, permettant une coordination avec les hydroxyles des glucides. Il existe un certain nombre de récepteurs avec des proportions H1 et H2 différentes. Parmi toutes les configurations des ASGPr identifiées, le trimère composé de deux sous-unités H1 et une sous-unité H2 est le plus abondant (**figure 32**).<sup>135,136</sup>

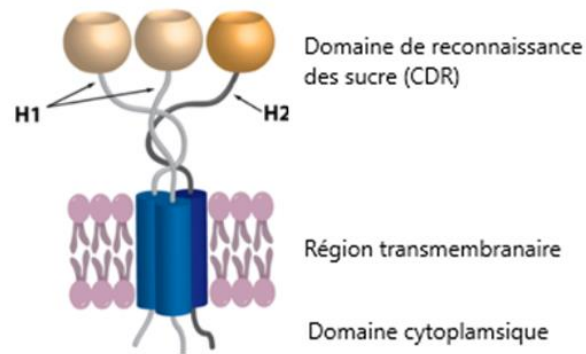


Figure 32 : Structure de l'ASGPr

L'une des premières investigations pour comprendre la nature de l'interaction de ce récepteur pour le galactose a été menée par Weis et Kolatar.<sup>137,138</sup> Ils ont démontré que le sucre *N*-acétylgalactosamine a une affinité de l'ordre de 40  $\mu$ M, plus élevée que le galactose qui est de l'ordre de 200 $\mu$ M. La poursuite de leurs travaux a permis de mieux comprendre l'interaction du GalNAc avec un mutant mimant l'ASGPr après co-cristallisation (figure 33 A). Il semblerait que le  $\text{Ca}^{2+}$  se lie au CRD par coordination des résidus Gln185, Asp187, Glu198, Asn210 et Asp211. Quant au GalNAc, il interagirait avec le  $\text{Ca}^{2+}$  par ses hydroxyles en position 3 et 4 en remplaçant deux molécules d'eau. Combiné à cela, le sucre forme aussi des liaisons hydrogène avec les résidus Gln185, Asp187, Glu198, Asn210 et Asp211. De plus, la face apolaire (C3, C4, C5 et C6) du GalNAc présente une interaction hydrophobe avec la chaîne latérale de Trp189 et enfin, le résidu histidine (His202) joue un rôle clé par coordination directe avec la fonction amide du GalNAc (figure 33 B).

La première structure cristalline du CRD de la sous unité H1 a été déterminée par Meier *et al.*<sup>139</sup> (figure 33 C). Elle montre deux sites de liaison au calcium (sphère verte) avec une coordination *via* les résidus aminés identiques à celle du mutant utilisé par Weis et Kolatar.

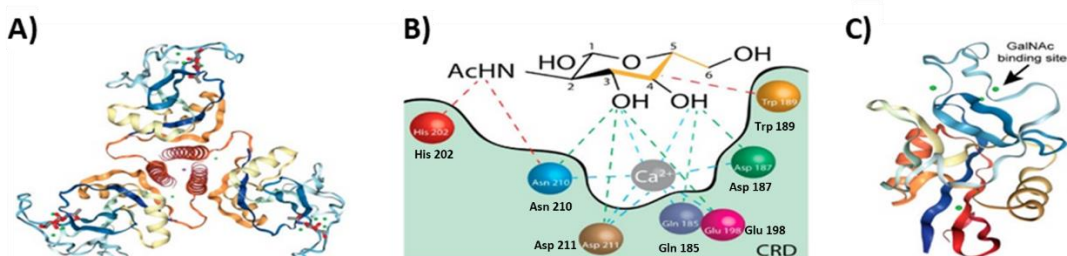


Figure 33 : A) Mutant de l'ASGPr cristallisé avec le GalNAc, le calcium est représenté en vert et le GalNAc en stick rouge et gris. B) interaction du GalNAc dans le site de liaison du mutant. C) structure cristalline du CDR de H1, le ligand GalNAc viendrait se lier au site de liaison.

## D. Mais quel ligand utiliser ?

Les études d'interactions entre le galactose, la *N*-acétylgalactosamine et la poche de reconnaissance du récepteur ont permis la conception de nouveaux ligands originaux présentant de meilleures affinités.

Les ligands naturels du récepteur ASGP sont les asialoglycoprotéines des protéines circulantes présentant à leur surface des polysaccharides.<sup>126</sup> Le terme « asialo » signifie que les polysaccharides de surface possédaient un ou plusieurs acides sialiques terminaux, ces derniers ont été clivés ce qui permet de révéler des motifs Gal ou GalNAc qui seront fortement reconnus par le récepteur. La famille principale regroupe par exemple l'asialo-orosomucoïde (ASOR) qui possède une vingtaine de résidus Gal au extrémité de ces polysaccharides ou l'asialofetuin avec des motifs Gal et GalNAc.<sup>140,141</sup> La nature multimérique de ces polysaccharides induit une affinité très élevée avec l'ASGPr.

En 2009, Ernst *et al* se sont intéressés à la modification du groupement acétamide en position C2 du sucre GalNAc. En effet, la zone de reconnaissance de l'ASGPr qui interagit avec l'acétamide est une cavité en forme d'haltère qui pourrait accueillir des groupements plus volumineux linéaires de nature plutôt hydrophobes. L'affinité de plusieurs pharmacophores avec un groupement triazole substitué par des chaînes aryles/alkyles à la place de l'acétamide en C2 a été évaluée. Les différents ligands testés présentent des affinités de l'ordre de 40 μM similaires à celle du GalNAc. La fonction acétamide étant absente, l'interaction avec l'histidine 256 et l'arginine 236 n'est plus possible ce qui semble indiquer que l'affinité est due à d'autres interactions. Les chercheurs ont donc montré qu'il était possible de moduler la position C2 du sucre GalNAc avec des substituants hydrophobes plus ou moins volumineux sans altérer l'affinité au récepteur.<sup>142</sup>

Plus tard, le groupe de M.G. Finn a synthétisé et évalué un grand nombre de molécules dérivées du galactose. Leur stratégie a été de modifier les positions C1, C2, C5 et C6 du sucre en laissant les hydroxyles en C3 et C4 intacts (**figure 34**). Ces différents ligands ont montré une affinité variable avec pour certains des constantes de l'ordre de la dizaine de micromolaire. Les ligands les plus affins présentaient des modifications de l'acétamide en C2 par des groupements trifluoroacétamides. Ces résultats sont complémentaires des travaux de Ernst et montrent que conserver un motif amide en C2 est important.<sup>143</sup>

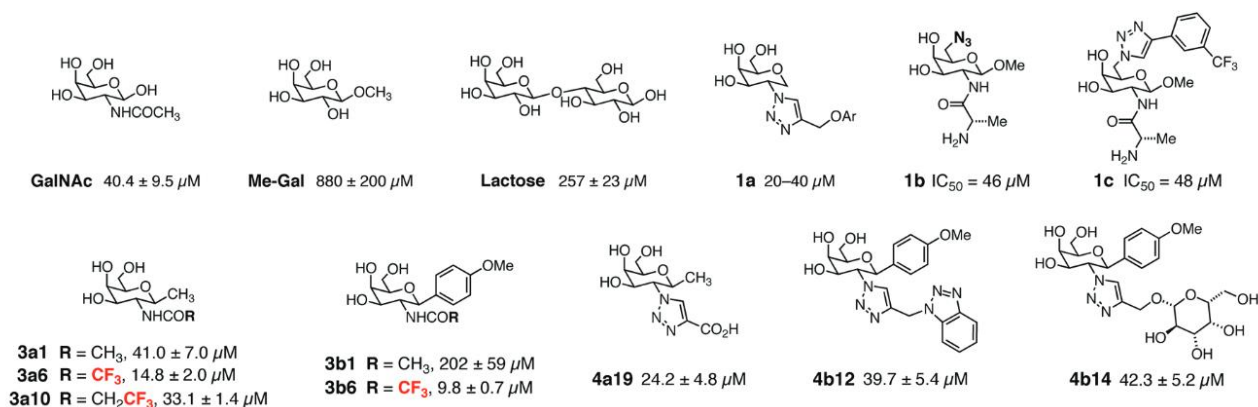


Figure 34 : Exemples de structures étudiées par M.G. Finn.<sup>143</sup>

Par la suite, cette même équipe a étudié l'interaction de la face hydrophobe du GalNAc avec le Trp243. Elle a imaginé une structure acétal pontée qui pourrait se verrouiller dans une conformation où la face  $\alpha$  (hydrophobe) serait plus accessible pour réaliser des interactions de van der Waals avec l'indole du tryptophane. Il y a alors conservation de l'orientation spatiale du substituant en C2 avec une interaction avec l'arginine 236 et de l'hydroxyle en C6 qui pointe en dehors de la poche. Une structure cristalline de cette molécule dans la partie lectinique de l'ASGPr a pu être obtenue. La formation du pont permet une meilleure interaction de la face  $\alpha$  hydroxylée en C5-C4 avec le tryptophane 243 (figure 35).

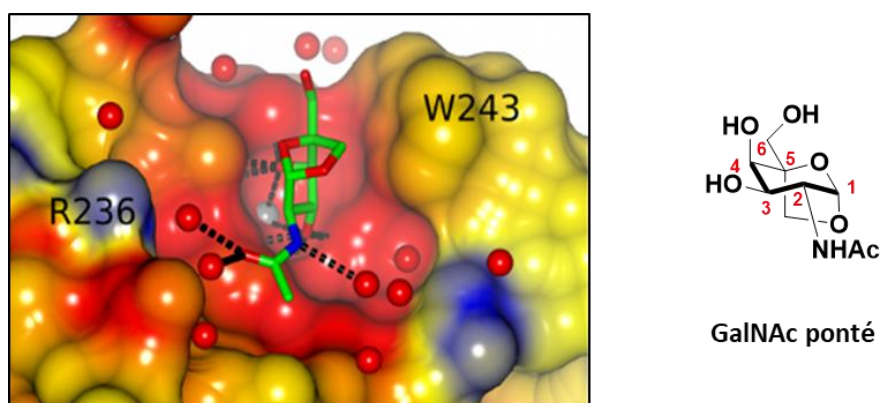


Figure 35 : Cliché du GalNAc ponté cristallisé dans le domaine lectinique de l'ASGPr.

Une augmentation de l'interaction de la face hydrophobe du sucre avec le Trp 243 et la conservation de l'interaction « clé » entre Arg 236 et l'acétamide permet d'avoir une constante d'affinité cinq fois plus forte pour le récepteur ASGP qu'avec le GalNAc. Il a également été observé que l'hydroxyle en C6 est en dehors du site de liaison, laissant envisageable sa fonctionnalisation sans altérer l'affinité du ligand. Dans cette même étude les chercheurs ont évalué la capacité de ce sucre à cibler les

hépatocytes. Pour cela, une sonde fluorescente a été couplée sur le sucre puis injectée *in vivo* en intraveineuse chez la souris. L'analyse du nombre d'hépatocytes positifs par détection de la fluorescence prouve que le sucre ponté est huit fois plus efficace que le GalNac (**figure 36**).<sup>144</sup>

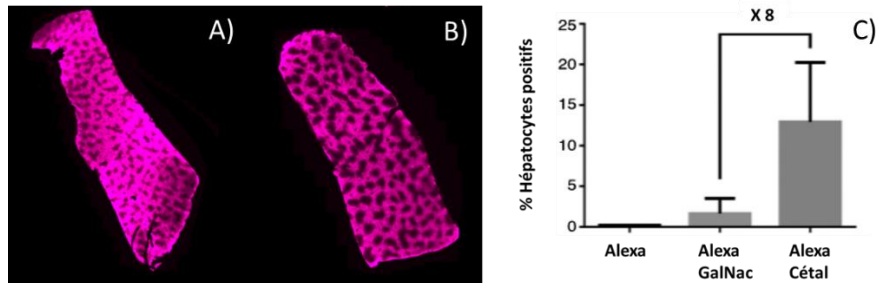


Figure 36 : Scan de coupes de foies de souris avec : A) Sonde fluorescente ALEXA fonctionnalisée avec le sucre ponté et B) Sonde fluorescente ALEXA fonctionnalisée avec le GalNac. C) Pourcentage d'hépatocytes positifs après exposition aux sondes fonctionnalisées quantifié par cytométrie de flux.<sup>143</sup>

Dans une autre étude les chercheurs Liras, Mascitti et A. Doudna ont bioconjugué un trimère du sucre ponté à la surface d'une protéine CAS9 fluorescente pour réaliser de l'édition de gène dans les hépatocytes. Cette construction a permis une meilleure internalisation dans des cellules HEPG2 ASGPr positive (lignée cellulaire cancéreuse du foie très utilisée pour mimer les hépatocytes) de la protéine CAS9 fluorescente, en comparaison avec la lignée cellulaire SKHEP ASGPr négative. Les chercheurs ont ainsi pu concevoir une protéine modifiée chimiquement capable de cibler les hépatocytes et d'y intégrer un gène en utilisant un ligand spécifique de l'ASGPr.<sup>145</sup>

## 5. Etude préliminaire et objectifs de la thèse

Aujourd'hui, même si l'AAV montre un réel potentiel en thérapie génique, la première partie de ce manuscrit a mis en évidence ses limites et la nécessité d'optimiser son efficacité. Pour remédier à ces différentes limitations nous avons vu qu'il est possible de modifier la capsidie selon diverses approches de génie génétique. Notre équipe a développé une stratégie alternative par bioconjugaison de ligands sur les résidus d'acides aminés présents à la surface de la capsidie du rAAV.

L'objectif de ma thèse s'inscrit dans cette dynamique et consiste à développer de nouveaux AAVr modifiés chimiquement et ayant un tropisme spécifique pour le foie, plus particulièrement pour les hépatocytes, ainsi qu'une forte capacité de transduction. Pour cela nous utiliserons des ligands bifonctionnels avec un motif qui sera reconnu par le récepteur ASGP et une fonction de couplage pour sa bioconjugaison à la surface de l'AAVr.

Au sein de l'équipe, un premier ligand, **(PMS-01)**, ciblant les hépatocytes et ayant eu des résultats biologiques intéressants a déjà vu le jour. Sa structure chimique présentée (**figure 37**) comprend un GalNAc comme motif de reconnaissance aux ASGP et un groupement phénylisothiocyanate pour la bioconjugaison des fonctions amines des résidus des acides aminés (essentiellement les lysines) présents sur la surface de la capsidie. Une petite chaîne triéthylène glycol entre le GalNAc et le phénylisothiocyanate est ajoutée afin d'induire de la mobilité à l'ensemble et d'espacer les deux motifs.

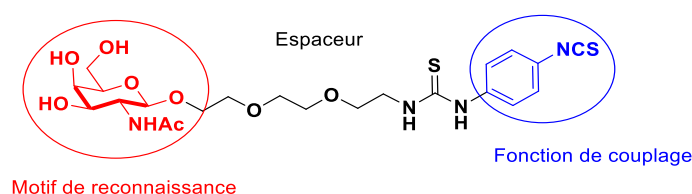


Figure 37 : Ligand **(PMS-01)** utilisé dans la première étude du ciblage des hépatocytes

Après couplage, les AAVr fonctionnalisés ont montré une efficacité de transduction sur des hépatocytes primaires murins significativement plus élevée que les vecteurs non-modifiés (**figure 38**). Le même constat est observé avec des AAVr modifiés avec un ligand mannose, non spécifique des ASGP, qui sont beaucoup moins efficaces que les AAV-GalNAc. Ces résultats nous ont conforté dans l'utilisation de ligands plus affins pour les récepteurs ASGP afin d'augmenter encore plus la transduction des hépatocytes.



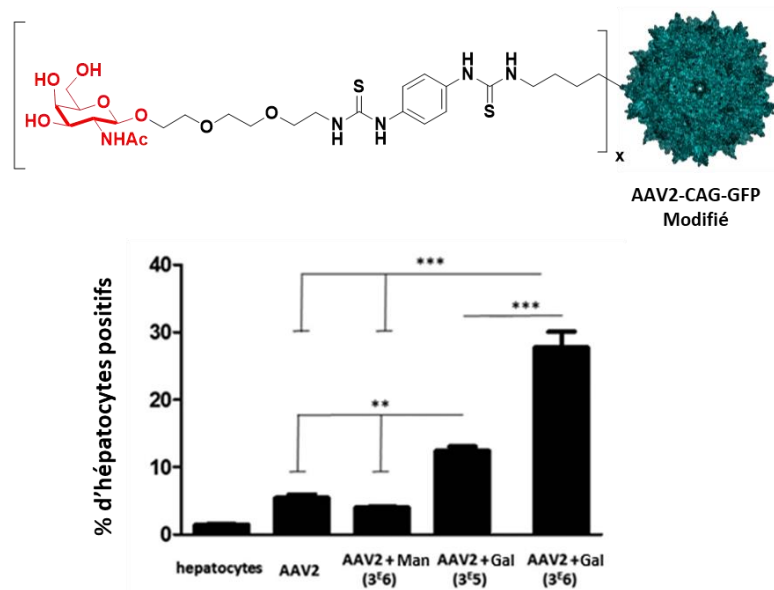


Figure 38 : Pourcentage d'hépatocytes primaire murins transduits grâce à l'AAVr modifié avec (PMS-01).

Ainsi, mon travail de thèse va se dérouler de la façon suivante :

Dans un premier temps il s'agira de synthétiser de nouveaux ligands, analogues de (PMS-01), avec des structures théoriquement mieux reconnues par le récepteur ASGP. Ces ligands seront ensuite couplés à la surface d'AAVr puis les vecteurs résultants seront caractérisés par des techniques de biologie moléculaire. Leur capacité de transduction au niveau des hépatocytes sera évaluée *in vitro* puis *in vivo* sur un modèle murin. L'idée est de voir si l'ajout d'un motif de reconnaissance présentant *a priori* une meilleure affinité avec le récepteur ASGP permet un ciblage amélioré des hépatocytes et une transduction plus efficace de ces derniers (figure 39).

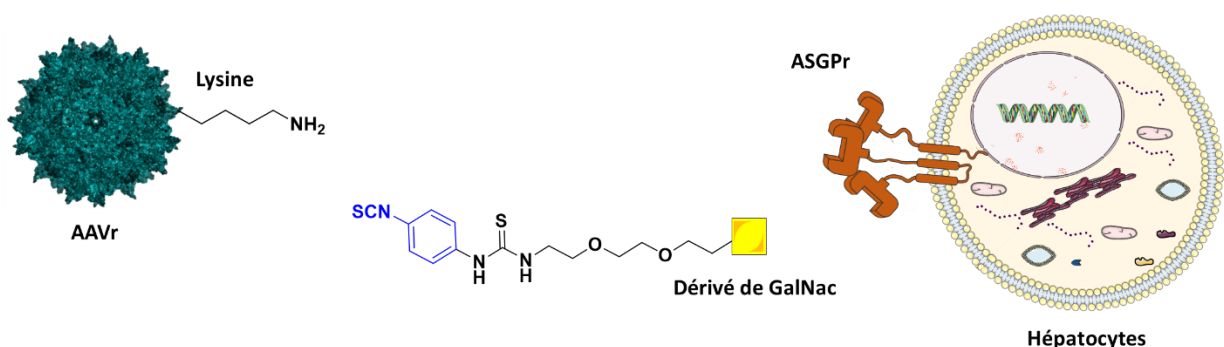


Figure 39 : Schéma général du projet montrant le ligand bifonctionnel avec d'une part la fonction phénylisothiocyanate pour modifier les lysines du virus et d'autre part un sucre dérivé du GalNac pour cibler les ASGP.

Ainsi nous commencerons par décrire la synthèse du ligand (**PMS-02**) qui possède le motif bicyclique dérivé du GalNAc développé par l'équipe du Professeur MG. Finn. En effet, l'étude cristallographique présenté précédemment montre que la position C6 du sucre n'intervient pas dans l'interaction avec le récepteur ASGP, ce qui laisse la possibilité d'y ajouter la chaîne éthylène glycol et la fonction de couplage. La synthèse de ce ligand sera entreprise en se basant sur la synthèse décrite du sucre ponté (figure 40).

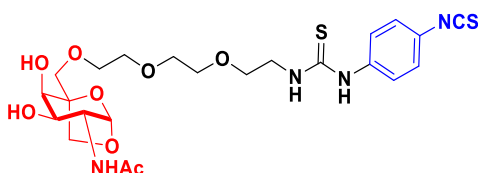


Figure 40 : Nouveaux ligand, (**PMS-02**).

Par ailleurs, nous effectuerons une étude par modélisation moléculaire afin d'explorer les possibilités de synthèse de nouveaux motifs *a priori* plus affins pour l'ASGPr. En s'appuyant sur ce travail et la littérature, nous sélectionnerons différentes structures potentiellement pertinentes pour cibler les récepteurs ASGP.

Il sera également possible d'imaginer un ligand qui conserverait le motif de reconnaissance GalNAc mais avec la chaîne triéthylène glycol liée à l'hydroxyle en C6 du sucre (et non en C1 comme pour **PMS-01**). Dans cette optique, deux ligands stéréoisomères seront synthétisés (composés **PMS-03** et **PMS-04**).

Dans le chapitre suivant sera décrit la bioconjugaison des ligands, précédemment développés, à la surface d'AAVr, ainsi que leurs analyses selon différentes techniques de biologie moléculaire (gels d'électrophorèses, Western et Dot-blot, ELISA etc...). Par la suite, l'efficacité du couplage sera évaluée *in vitro* par transduction de lignées cellulaires (HEK293, HUH7 et hépatocytes primaires). Pour finir, une étude *in vivo* chez la souris sera réalisée afin d'avoir une vue plus large de l'efficacité des vecteurs modifiés. Plusieurs analyses seront faites comme le dosage du transgène, la quantification de l'expression de la protéine reportrice ou encore l'observation de coupes de l'organe après marquage immunochimique. Ces analyses nous permettront de dresser un bilan sur la capacité des nouveaux vecteurs modifiés à transduire le foie.

Après une conclusion/perspective, la thèse se terminera par une partie expérimentale reprenant tous les protocoles de synthèse organique et les expérimentations biologiques puis par un listing des références bibliographiques support de ce travail.



# Chapitre II

## Synthèse des ligands



## 1. Introduction

Cette partie s'articulera autour de la synthèse des différents ligands envisagés lors de ma thèse et notamment sur les modifications apportées au motif de reconnaissance pour une meilleure interaction avec les récepteurs ASGP.

Dans un premier temps sera présentée la synthèse du ligand (**PMS-02**) qui comprend comme squelette de base le GalNAc ponté. Comme expliqué précédemment, ce motif est décrit pour être un des plus afin pour le récepteur ASGP. En effet, l'ajout d'un cycle, entre les carbones C1 et C5 du sucre, contraint sa structure tridimensionnelle et l'impose d'adopter une conformation « chaise  ${}^4C_1$  », en règle générale les sucres sont des structures mobiles qui peuvent dans l'environnement aqueux adopter des conformations cavalières multiples, ici la contrainte apportée par le deuxième cycle l'en empêche. L'orientation des substituants sous cette conformation favorise les différentes interactions du ligand dans le site lectinique du récepteur, comme visible sur la (figure 41) issue de la structure 5JQ1 observée avec le logiciel Pymol.

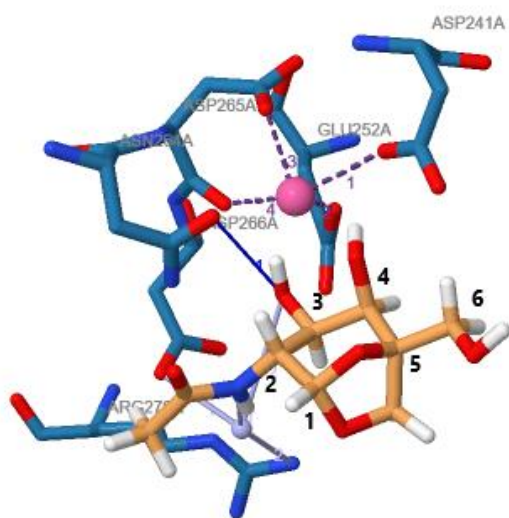


Figure 41 : Représentation Pymol du sucre ponté dans la poche lectinique de l'ASGPr.

De plus, sur cette structure on peut voir que l'hydroxyle en position 6 du sucre est localisé à l'extérieur de la poche lectinique et n'intervient pas ou peu dans la reconnaissance. Il est donc envisageable d'ajouter l'espaceur triéthylène glycol sur cette position sans trop de risques de provoquer une perte d'affinité.

Parallèlement à la synthèse du ligand **(PMS-02)**, une étude par modélisation moléculaire sera entreprise pour trouver de nouvelles structures potentiellement plus affines pour le récepteur ASGP. Dans la lignée de ce travail, une étude bibliographique portera sur la conception de sucres bicycliques analogues à celui utilisé pour le ligand **(PMS-02)** selon une réaction radicalaire de type 1,5 HAT, décrite par l'équipe du professeur Suarez.<sup>146</sup> Une partie de ce chapitre fera l'état de l'art de cette réaction radicalaire et les différentes voies envisagées pour l'élaboration d'un ligand **(PMS-02bis)**.

La clôture de ce chapitre portera sur la synthèse des deux ligands, **(PMS-03)** et **(PMS-04)**, qui possèdent comme structure de reconnaissance le sucre GalNAc fonctionnalisé par une chaîne triéthylène glycole en position C6. La position anomérique du sucre est figée par la présence d'un groupement méthoxy en position  $\beta$  pour **(PMS-03)** et en position  $\alpha$  pour **(PMS-04)** (figure 42).

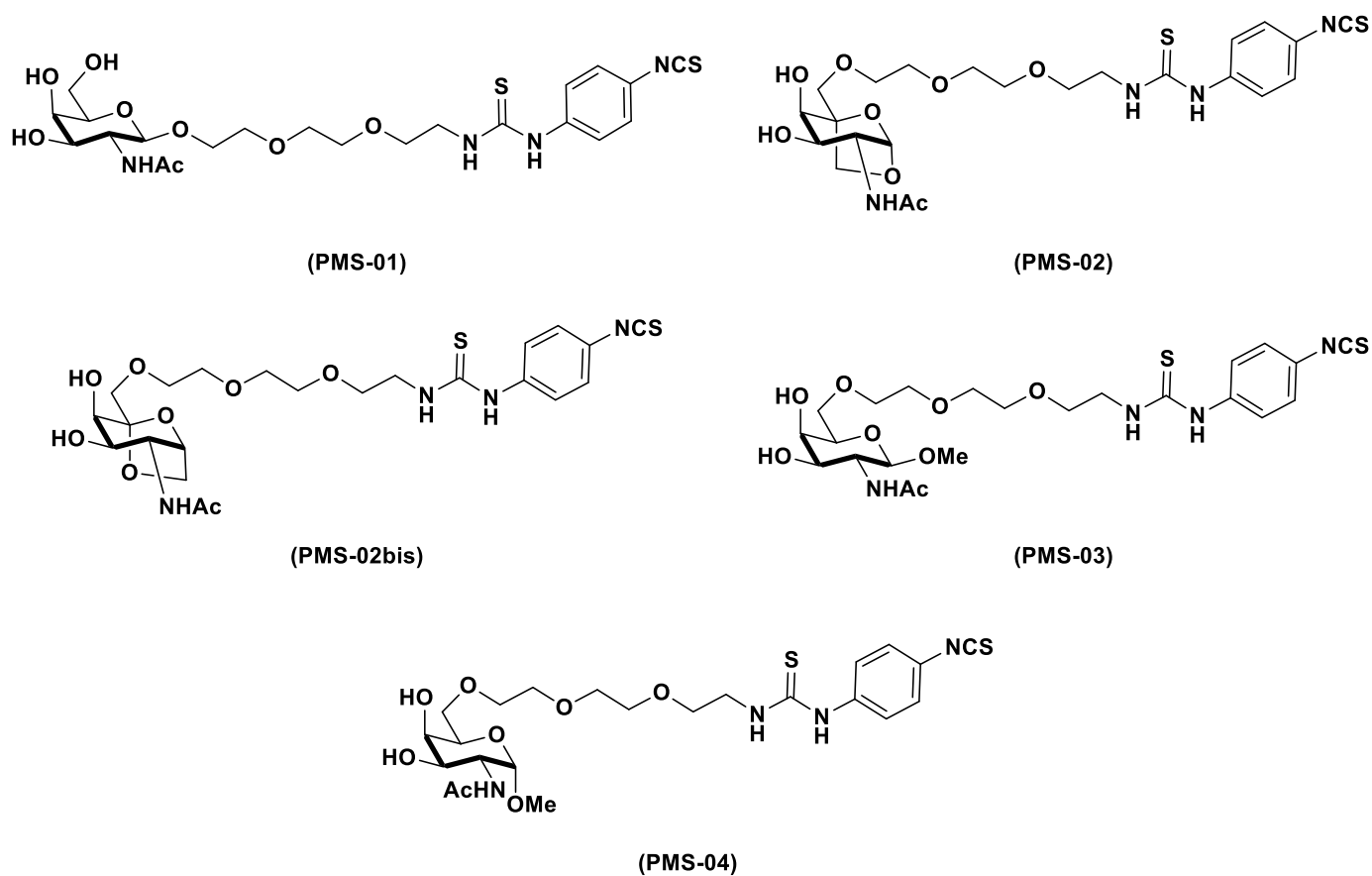


Figure 42 : Présentation des différents ligands.

## 2. Synthèse du ligand PMS-02

### A. Synthèse du motif bicyclique

Le ligand (**PMS-02**) est composé d'un motif glucidique bicyclique relié à la fonction de couplage phénylisothiocyanate par un espaceur triéthylène glycol. Pour commencer il est nécessaire de synthétiser le GalNAc ponté, noté composé (**10**), en s'inspirant de la procédure décrite par l'équipe de Carlos Sanhueza.<sup>144</sup> La synthèse comprend 12 étapes, présentées sur la (figure 43), et met en œuvre plusieurs réactions complexes comme une azidonitration, une glycosylation de Koenigs-Knorr, une aldolisation en tandem avec une réaction de Cannizzaro et une cyclisation intramoléculaire, le tout en contrôlant la stéréochimie du centre anomérique. Plusieurs étapes ont été optimisées notamment pour accéder au composé (**4**) et pour adapter les protocoles à l'échelle du laboratoire.

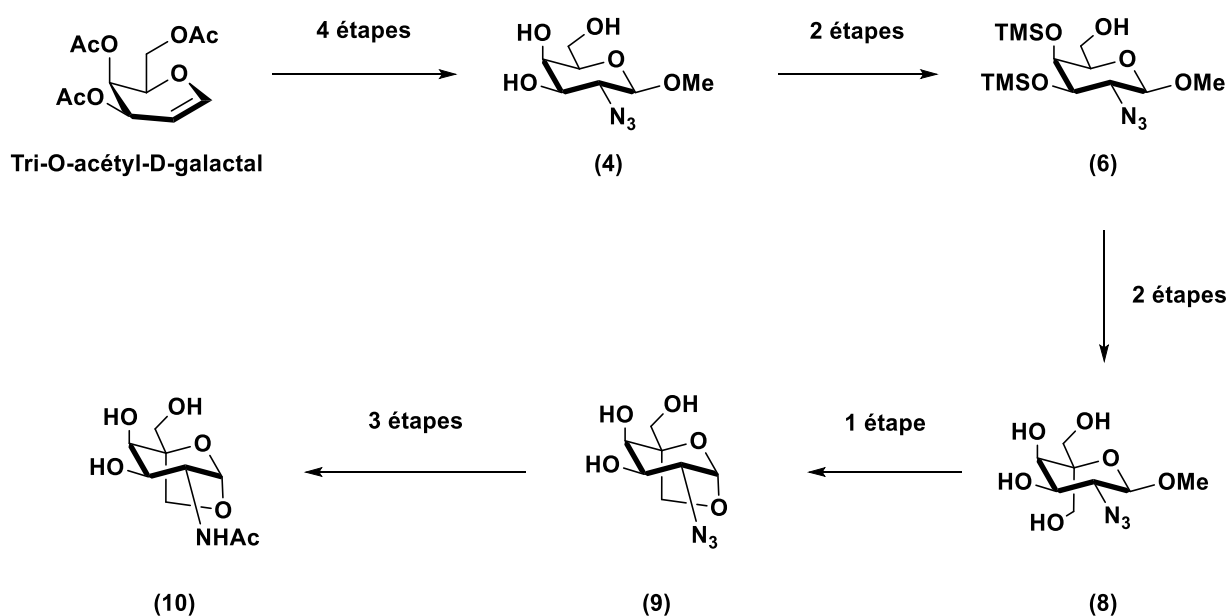


Figure 43 : Synthèse du sucre GalNAc ponté (**10**).

#### a) Réaction d'azidonitration et obtention du composé (**4**)

La synthèse débute par une réaction d'azidonitration du tri-O-acétyl-D-galactal, composé commercial, en présence d'azoture de sodium, de nitrate de cérium et d'ammonium (CAN) dans l'acétonitrile pendant 10 heures à -15°C. Le produit d'azidonitration (**1**) est isolé après purification sur colonne de



silice (Cy 80% - AcOEt 20%) avec un rendement de 75% sous la forme d'un mélange de stéréoisomères  $\alpha/\beta$  (figure 44).

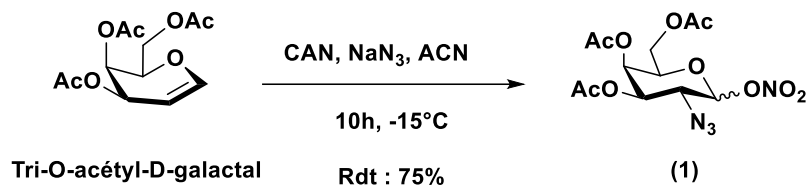


Figure 44 : Réaction d'azidonitration du tri-O-acétyl-D-galactal permettant d'obtenir le dérivé azoture (1).

Le mécanisme supposé de la réaction est présenté sur la (figure 45).

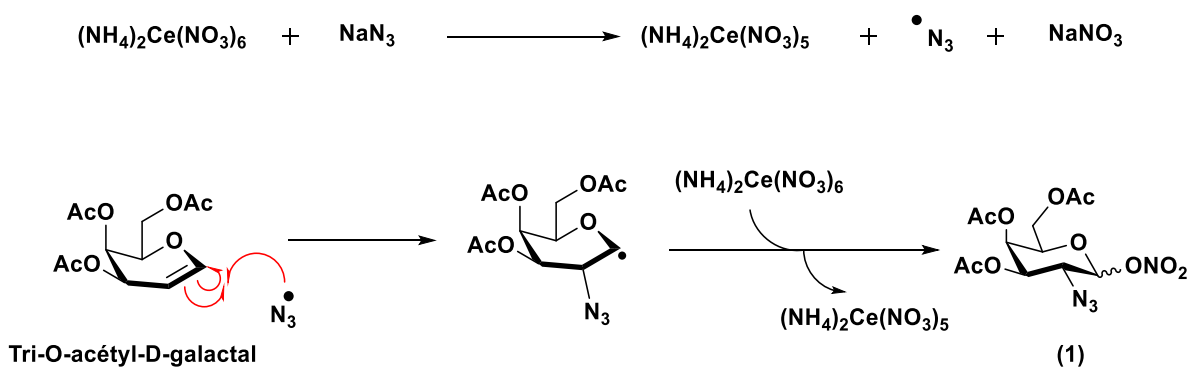


Figure 45 : Mécanisme de la réaction d'azidonitration.

Les propriétés oxydantes du nitrate de cérium et d'ammonium permettent de générer *in situ* un radical  $\text{N}_3^\bullet$ .<sup>147</sup> Ce radical ainsi formé est électrophile en raison de l'électronégativité de l'atome d'azote ainsi que sa délocalisation possible sur tous les atomes de l'azote. Il s'additionnera donc préférentiellement sur une oléfine nucléophile.

La délocalisation d'un des doublets non-liants de l'oxygène endo-cyclique en alpha de la double liaison augmente son caractère nucléophile et explique que le  $\text{N}_3^\bullet$  s'additionne en position 2 du galactal. De plus, la régiosélectivité est renforcée puisque le radical ainsi formé est stabilisé par le doublet non-liant de l'oxygène du sucre. En revanche, pour ce qui est de la stéréochimie de l'addition du  $\text{N}_3^\bullet$ , elle est régie par un contrôle stérique et le composé (1) est obtenu dans un rapport 90/10 en faveur de l'azoture en position équatoriale (galactose vs talose). En effet, la même réaction à partir du glucal

donne un mélange de dérivés de glucose et de mannose en proportion 40/60 comme le montrent les travaux de P.H. Seeberger (figure 46).<sup>148</sup>

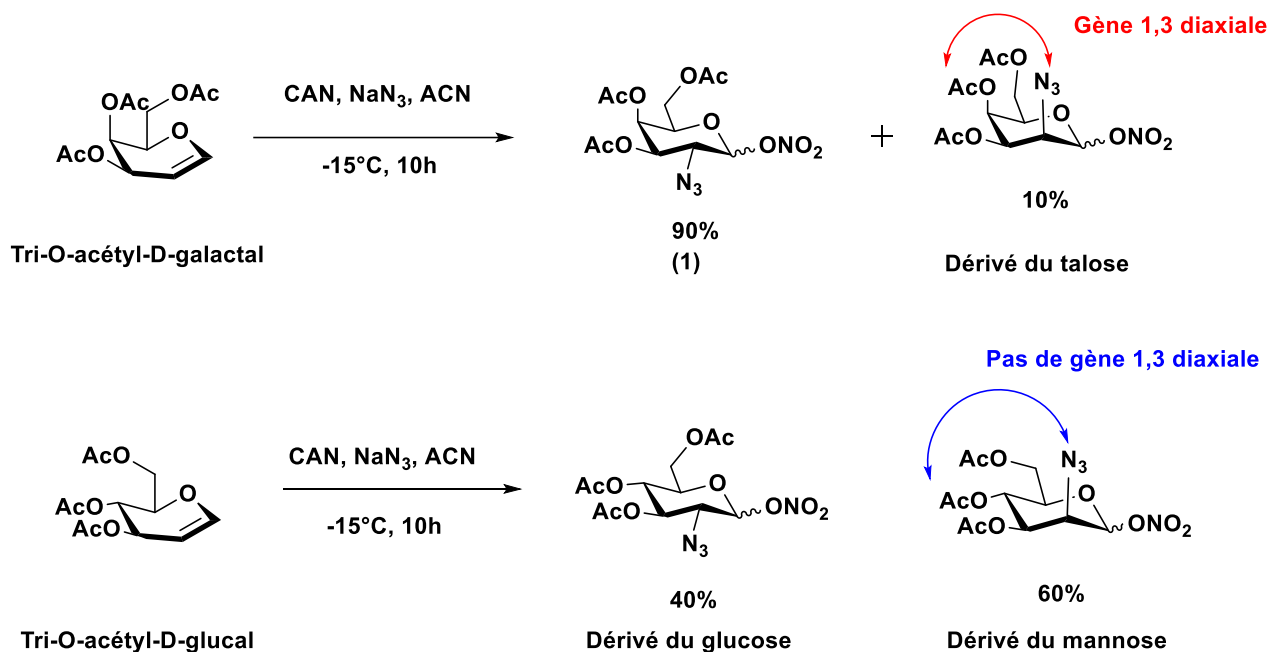


Figure 46 : Comparaison de la stéréosélectivité après l'addition du N<sub>3</sub> sur l'oléfine.

Cette différence de stéréosélectivité s'explique par l'orientation en position axiale du groupement acétate en C4 qui induit une interaction 1,3 diaxiale importante avec le N<sub>3</sub> du talose. Cette gêne stérique n'est pas présente dans le cas du dérivé du galactose où l'azoture est en position équatoriale. Pour les sucres mannose et glucose issus du D-glucal, la gêne stérique occasionnée par le groupement en C4 est moindre du fait de son orientation équatoriale, ce qui explique le rapport 40/60 entre ces composés (figure 46).

Après la formation de l'intermédiaire radicalaire, le nitrate de cérium et d'ammonium vient oxyder le radical afin de générer un carbocation anomérique, du cérium III et un ion nitrate NO<sub>3</sub><sup>-</sup>. Le carbocation stabilisé formé est à son tour attaqué par l'ion nitrate pour conduire au produit d'azidonitration (1). La stéréosélectivité de cette réaction n'est pas contrôlée en ce qui concerne l'attaque de l'ion nitrate sur le carbocation ce qui explique la formation du composé (1) sous la forme α / β en proportion 57% / 43%.

Lors de cette réaction d'azidonitration nous avons également observé un sous-produit, le 1-déoxy-1-acétamido-α-D-galactose (figure 47). Ce composé est obtenu par addition du doublet non-liant de l'azote de l'acétonitrile, utilisé comme solvant, sur le carbocation anomérique. L'intermédiaire activé

est alors hydrolysé pour former une fonction amide. Le sous-produit 1-déoxy-1-acétamido- $\alpha$ -D-galactose peut cependant être facilement éliminé par purification sur colonne de silice.

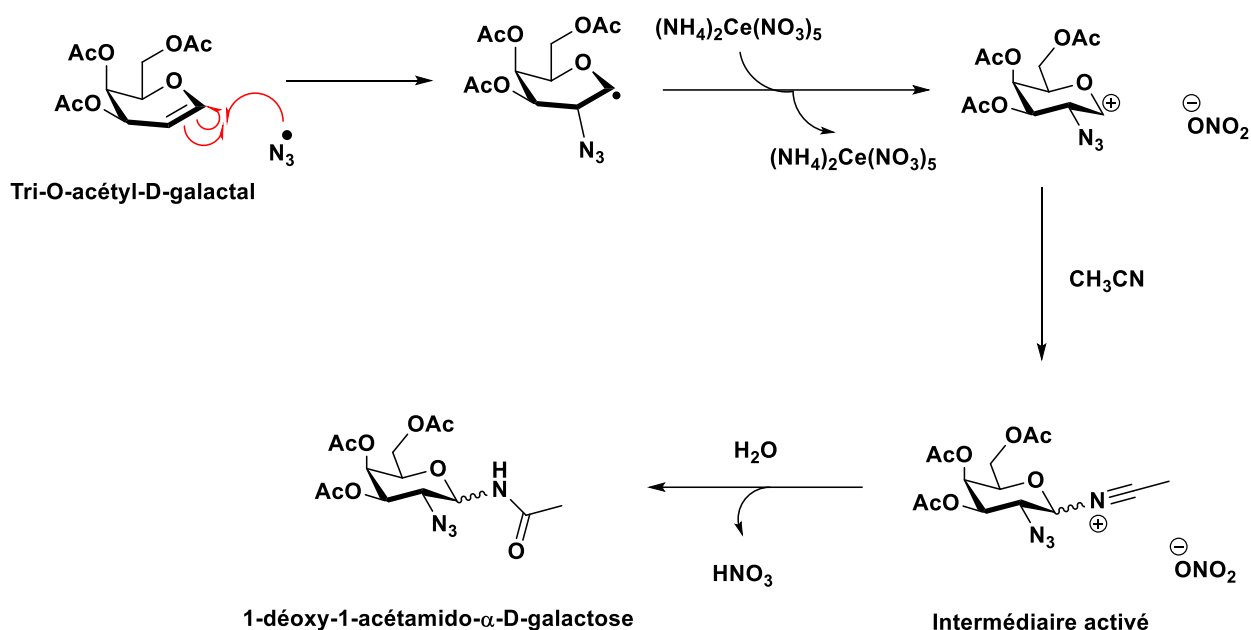


Figure 47 : Formation du 1-déoxy-1-acétamido- $\alpha$ -D-galactose.

Initialement la stratégie de synthèse développée par Sanhueza pour obtenir le composé **(4)** reprenait les travaux de Paulsen et consistait à substituer directement le groupement nitrate de **(1)** par un groupement méthoxy (figure 48).<sup>149</sup>

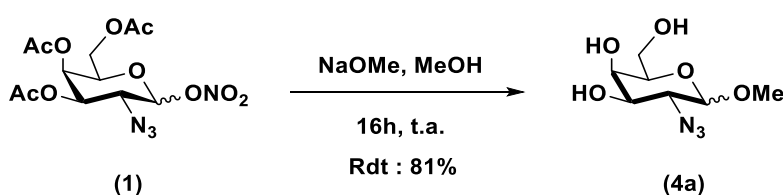


Figure 48 : Synthèse du composé **(4a)** à partir de **(1)** selon Paulsen.<sup>149</sup>

Cette réaction s'est avérée contraignante car le composé **(1)** est sous la forme d'un mélange  $\alpha/\beta$ . En effet, la réaction de substitution du nitrate par un groupement méthoxy a permis d'obtenir le bon produit avec un rendement de 81% mais également sous la forme d'un mélange d'anomères difficilement séparables par chromatographie. La quantité obtenue de composé **(4a)** avec la bonne configuration anomérique est trop faible pour poursuivre cette voie.

Pour résoudre ce problème et obtenir le composé de configuration  $\beta$  souhaité, le passage par un dérivé halogéné a été envisagé. L'introduction d'un atome d'halogène sur la position anomérique du composé **(1)** renforce l'effet anomère avec l'un des doublets non-liants de l'oxygène endocyclique qui vient stabiliser l'orbitale anti-liante de la liaison halogène-carbone. Cependant, cette stabilisation n'est possible que si l'halogène est en position axiale comme le montre la **(figure 49)**.

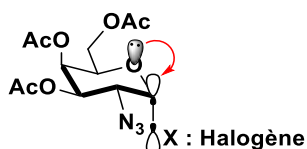


Figure 49 : Représentation orbitale montrant l'effet anomère.

De ce fait, l'introduction d'un halogène sur cette position entrainera, en théorie, uniquement l'obtention de l'anomère  $\alpha$ .

Le premier essai d'halogénéation a été effectué dans l'acétonitrile à l'abri de la lumière et à température ambiante durant 16 heures avec différentes sources d'iode (iodures de sodium et de lithium). Dans les deux cas le composé iodé est obtenu avec un rendement de 79% et seul l'anomère  $\alpha$  est présent **(figure 50)**. Il est cependant à noter que ce produit est instable et s'hydrolyse rapidement. En effet, la présence d'un atome d'iode qui est un très bon nucléofuge augmente énormément la réactivité du carbone anomérique. De ce fait, le composé **2'** est engagé sans purification dans l'étape suivante mais quelles que soient les conditions utilisées nous n'avons pas obtenu le produit **(3)** attendu. Les conditions basiques employées afin d'introduire le groupement méthoxy en position anomérique induisent la formation de sous-produits.

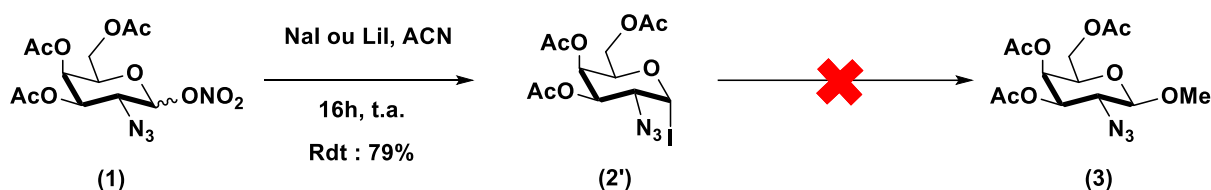


Figure 50 : Iodation de la position anomérique du sucre **(1)**.

Afin de pallier ce problème, la synthèse du dérivé chloré, composé a priori plus stable, a été envisagée **(figure 51)**.

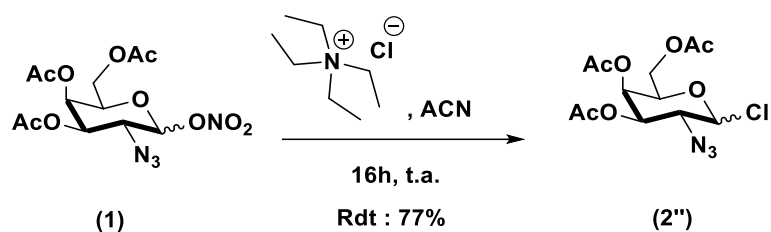


Figure 51 : Chloration de la position anomérique du sucre **(1)**.

Cette réaction est menée en présence de chlorure de tétraéthylammonium dans l'acétonitrile à température ambiante durant 16 heures. Le composé chloré (**2''**) est obtenu après purification sur gel de silice (AcOEt 15% - Cy 85%) avec un rendement de 77%. L'analyse RMN montre l'obtention du produit sous forme d'un mélange  $\alpha/\beta$  en proportion 90/10. L'effet anomère ne permettant pas une conversion totale en produit  $\alpha$ , ce composé est difficilement utilisable pour la suite de notre stratégie de synthèse.

Finalement, pour garder une stabilité convenable du dérivé halogéné et avoir un effet anomère assez fort, la synthèse du composé (**2**) bromé a été réalisée. Cette réaction est effectuée par action du bromure de lithium dans l'acétonitrile à température ambiante pendant 4,5 heures. Le produit d'halogénéation (**2**) est obtenu après un simple lavage de la phase organique avec un rendement de 87% et uniquement sous la forme  $\alpha$  (figure 52).

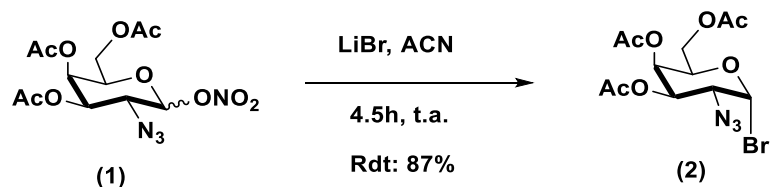


Figure 52 : Bromation de la position anomérique du sucre **(1)**.

L'étape suivante pour aboutir au produit (**3**) est une réaction de glycosylation en utilisant des conditions type Koenigs-Knorr. Les sels d'argent jouent le rôle d'acide de Lewis et sont chélatés par l'atome de brome en position axiale ce qui permet une attaque stéréosélective du méthanol en position anomérique par la face  $\beta$ . Après purification sur colonne de silice (AcOEt 70% - Cy 30%) seul le produit **3** de configuration  $\beta$  est obtenu avec un rendement de 51%.

La déprotection des hydroxyles se fait ensuite par trans-estérification des acétates à l'aide du méthanolate de sodium. L'ajout d'une résine acide DOWEX-50 à la fin de la réaction neutralise la base.

Le composé **(4)** est isolé par chromatographie sur gel de silice (DCM 95% - MeOH 5%) avec un rendement de 99% (figure 53).

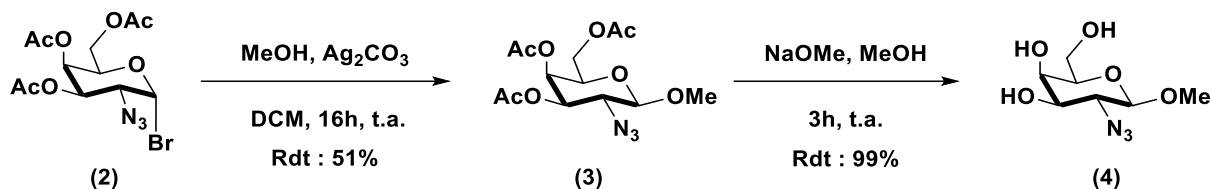


Figure 53 : Conditions de réaction de Koenigs-Knorr permettant d'obtenir le composé **(3)** puis hydrolyse des esters.

### b) Synthèse de l'alcool primaire **(6)**

La suite de la méthodologie de Sanhueza consiste à protéger les hydroxyles du triol **(4)** par action du chlorure de triméthylsilyle (TMSCl) dans la pyridine à température ambiante durant 16 heures. La pyridine active le chlorure de triméthylsilyle formant un complexe pyridinium réactif et piège l'acide chlorhydrique formé au cours de la réaction. Le brut réactionnel est directement engagé dans la réaction suivante. Il est ici important de protéger les hydroxyles du sucre avec des groupements TMS car il est plus facile par la suite de déprotéger sélectivement l'hydroxyle en C6. Cela n'aurait pas été possible à partir de **(3)** où les alcools sont protégés en acétates.

La déprotection sélective de l'hydroxyle primaire du composé persilylé **(5)** est effectuée en milieu faiblement basique en présence d'une très faible quantité de carbonate de potassium dans le méthanol. Le  $K_2CO_3$  génère *in situ* une quantité catalytique de méthanolate de potassium qui permet de déprotéger uniquement l'hydroxyle primaire plus réactif que ses homologues secondaires (figure 54).

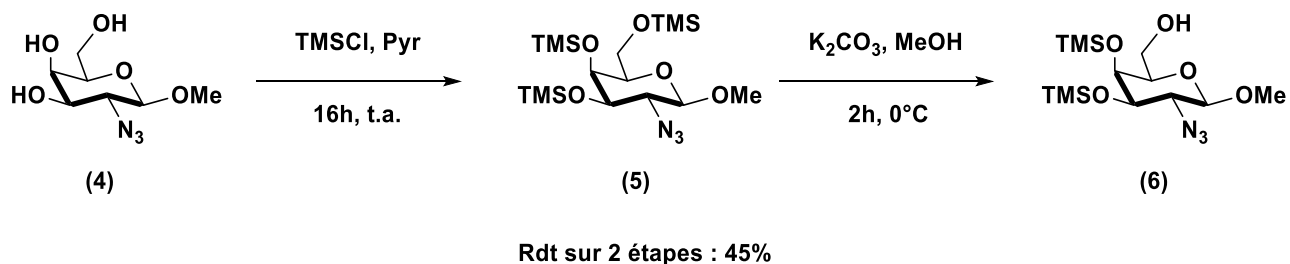


Figure 54 : Condition de réaction pour la protection des hydroxyles puis la déprotection sélective de l'alcool en C6.

Dans la publication, une quantité catalytique d'acide acétique est ajoutée afin de neutraliser la base et ainsi pouvoir traiter la réaction. Cependant, dans notre cas, cela est peu adapté car la quantité d'acide à ajouter est trop faible. Une alternative plus simple est d'ajouter directement de la silice dans le brut réactionnel afin que l'acidité de cette dernière neutralise la base. Il est alors directement possible de purifier le brut réactionnel sur colonne de silice (Cy 80% - AcOEt 20%) et d'obtenir le composé **(6)** avec un rendement sur deux étapes de 45% (figure 54).

### c) Formation du tétraol **(8)**

L'oxydation de l'alcool du substrat **(6)** en aldéhyde **(7)** est réalisée en utilisant la méthode de Parikh-Doering.<sup>150</sup> Cette réaction de type Swern fait intervenir le diméthylsulfoxyde, le complexe trioxyde de soufre/pyridine et de la triéthylamine. L'aldéhyde **(7)** étant un produit peu stable, il n'est pas purifié et le brut réactionnel est directement engagé dans une réaction d'aldolisation / Cannizzaro pour donner le produit **(8)** (figure 55).

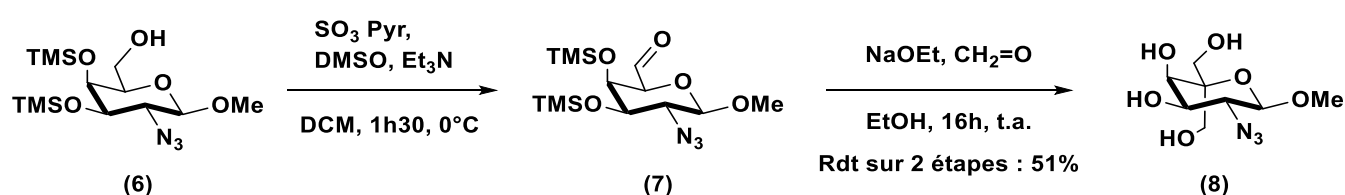


Figure 55 : Réaction d'oxydation de Parikh-Doering permettant d'obtenir l'aldéhyde **(7)**.

Le mécanisme de la réaction de Parikh-Doering<sup>150</sup> est présenté ci-dessous (figure 56).

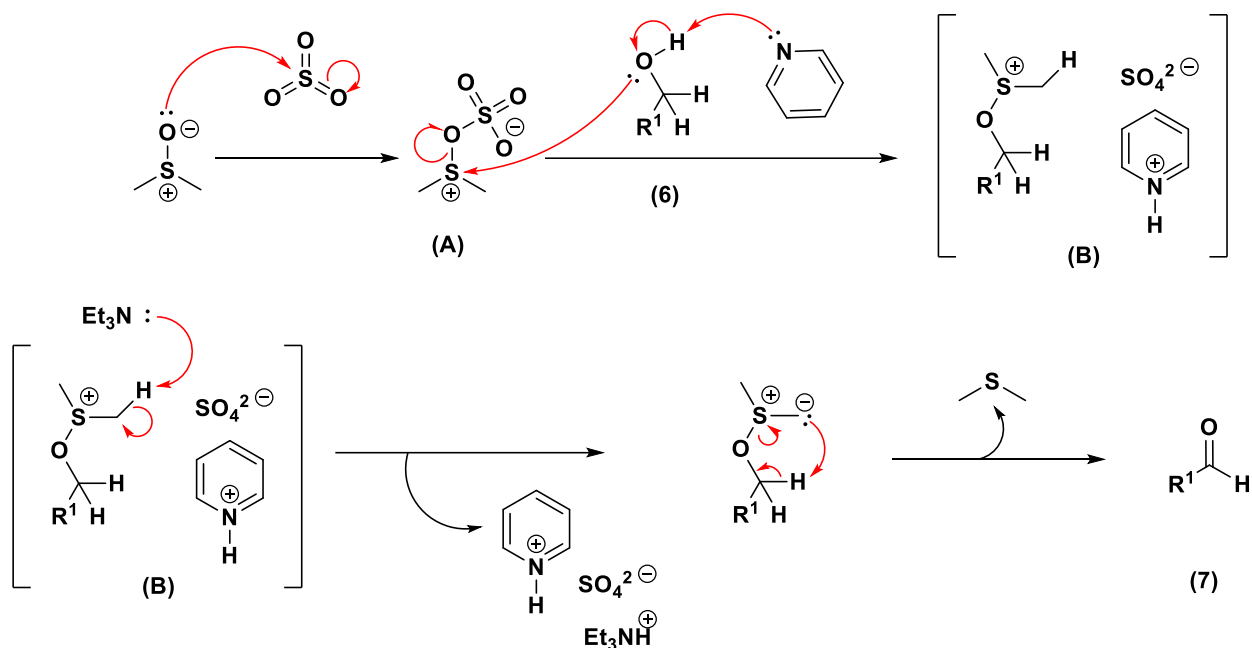


Figure 56 : Mécanisme de la réaction de Parikh-Doering.<sup>150</sup>

La première étape de la réaction est l'attaque de l'oxygène du DMSO sur le trioxyde de soufre pour former l'intermédiaire activé (A). L'alcool primaire du sucre (6) substitue alors l'ion sulfate  $\text{SO}_4^{2-}$  puis la pyridine déprotonne l'oxygène pour générer le complexe (B). Le  $\text{CH}_3$  en  $\alpha$  du soufre est ensuite déprotoné par la triéthylamine et le carbanion formé arrache à son tour l'hydrogène en alpha de l'oxygène générant ainsi l'aldéhyde désiré et libérant une molécule de diméthylsulfure. La réaction dure 1h30 à  $0^\circ\text{C}$  et l'aldéhyde (7) est isolé, après lavage de la phase organique par une solution saturée de NaCl, sans purification supplémentaire. La RMN  $^1\text{H}$  montre un signal caractéristique du proton aldéhydique à 9,5 ppm.

Pour la synthèse de (8), les conditions utilisées permettent de réaliser deux réactions simultanément. La première est une aldolisation qui introduit en position axiale du carbone C5 de (7) un groupement  $\text{CH}_2\text{OH}$ , par réaction entre la fonction aldéhyde du sucre et le paraformaldéhyde, en présence d'éthanolate de sodium. La deuxième réaction, appelée Cannizzaro, donne un alcool primaire par dismutation de l'aldéhyde. Lors de ces étapes les TMS sont également clivés par action de la base pour donner les hydroxyles correspondants. On obtient ainsi le produit (8) sous la forme d'un solide blanc avec un rendement de 51% après purification sur colonne de silice (AcOEt 90% - MeOH 10%). Classiquement, la réaction de Cannizzaro est une dismutation d'un aldéhyde, sans hydrogène en alpha, en acide carboxylique et en alcool correspondant. Dans notre cas, il s'agit d'une réaction croisée de Cannizzaro où le formaldéhyde est oxydé et l'aldéhyde réduit. Le mécanisme supposé est présenté (figure 57).



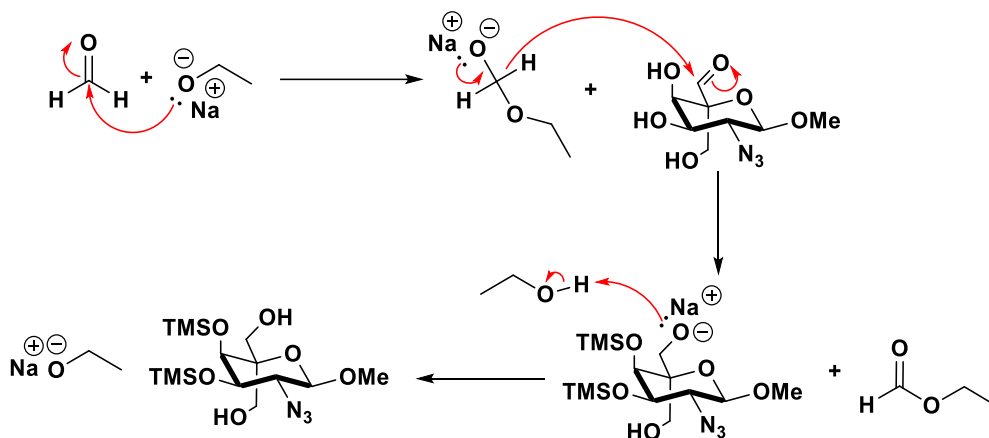


Figure 57 : Mécanisme de la réaction croisée de Cannizzaro.

L'éthanolate de sodium attaque uniquement le formaldéhyde, présent en large excès dans le milieu réactionnel et plus réactif que l'aldéhyde, pour former un intermédiaire qui libère un hydrure. Celui-ci attaque le carbone de l'aldéhyde pour donner l'alcoolate correspondant et libérer une molécule de formiate d'éthyle. Une réaction acido-basique avec l'éthanol (solvant de la réaction) permet alors la synthèse du sucre **(8)** avec un rendement de 51 % à partir de **(6)** après purification sur colonne de silice (AcOEt 90% - MeOH 10%).

#### d) Synthèse du GalNAc ponté **(10)**

L'étape suivante est la cyclisation intramoléculaire stéréosélective, en présence d'acide sulfurique concentré, du composé **(8)** par substitution nucléophile du méthoxy par l'hydroxyméthyle axial du carbone C5. Les conditions drastiques ( $\text{H}_2\text{SO}_4$  conc,  $100^\circ\text{C}$  et 40h) permettent la formation d'un cycle à 5 chaînons avec libération d'une molécule de méthanol. Une purification sur colonne de silice (AcOEt 90% - MeOH 10%) donne le produit **(9)** avec un rendement modéré de 48% (figure 58).

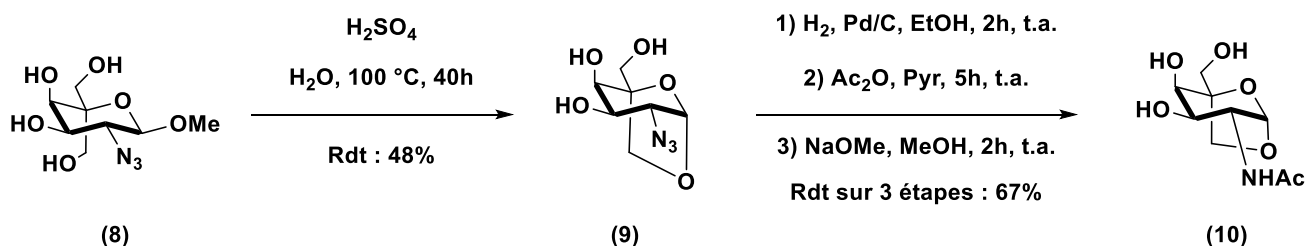


Figure 58 : Réaction de cyclisation intramoléculaire pour la synthèse du bicyclic **(9)**.

L'obtention du GalNAc ponté (**10**) s'effectue en trois étapes à partir du substrat (**9**). D'abord une réaction pallado-catalysée sous pression d'hydrogène réduit l'azoture en amine correspondante. Ensuite, l'amine est acétylée en présence d'anhydride acétique dans la pyridine pour former le motif *N*-acétamide du GalNAc ponté. Cette réaction conduit à l'ajout de groupements acétate sur les hydroxyles du sucre, un traitement par du méthanolate de sodium dans le méthanol régénère les fonctions alcools pour obtenir le composé (**10**) sous la forme d'un solide blanc sans nécessité de purification (rendement de 67%).

## B. Introduction de l'espaceur et synthèse du ligand final.

Une fois le GalNAc ponté (**10**) synthétisé, il est nécessaire d'ajouter une chaîne triéthylène glycol en position 6 du sucre et de réduire son extrémité azoture en amine primaire. Cela permet d'introduire la fonction de couplage phénylisothiocyanate et ainsi d'obtenir le ligand (**PMS-02**) (figure 59).

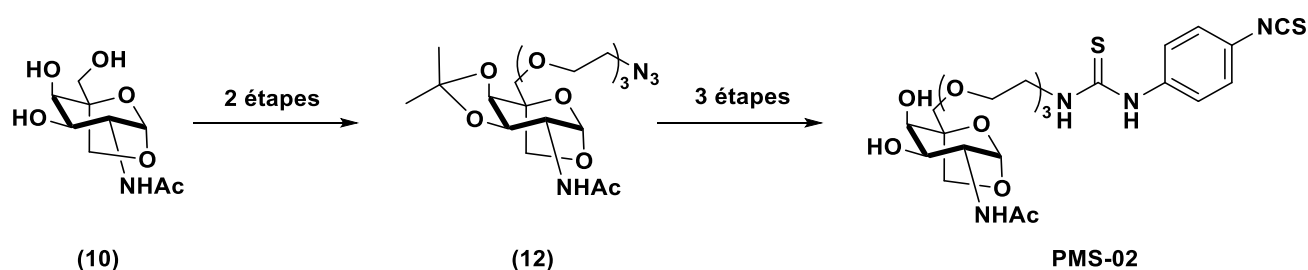


Figure 59 : Synthèse du ligand (**PMS-02**).

Dans un premier temps il est indispensable de protéger les hydroxyles en positions 3 et 4 du GalNAc (**10**). La réaction se fait en présence de 2,2-diméthoxypropane et d'acide camphre sulfonique (CSA) dans la DMF pendant 24h à 70°C afin de favoriser la formation du produit thermodynamique (**11**) et non pas celle du produit cinétique avec l'acétal en position 4 et 6. Une purification sur gel de silice (AcOEt 90% - MeOH 10%) donne le produit (**11**) avec un rendement de 93% (figure 60).

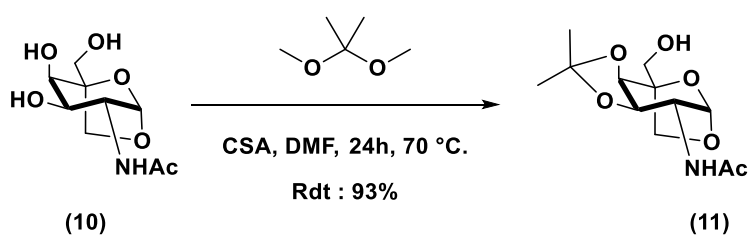


Figure 60 : Réaction de protection sélective des hydroxyles en position 3 et 4.

L'introduction de la chaîne triéthylène glycol s'effectue *via* une réaction d'O-alkylation sur l'hydroxyle libre en position 6 du sucre **(11)** par substitution d'un atome d'iode. Un mécanisme S<sub>N</sub>2 est favorable dans cette réaction mais nous pouvons également envisager une substitution intramoléculaire (S<sub>N</sub>i). En effet, le doublet non liant de l'oxygène peut attaquer le carbone portant l'iode pour former ainsi l'ion oxiranium, très réactif, qui se fait lui-même attaquer par l'alcool primaire du composé **(11)**. Il est à noter que les conditions opératoires sont très particulières. En effet, la réaction a lieu en milieu biphasique dichlorométhane/solution aqueuse de soude concentrée à 12,5 mol.L<sup>-1</sup> en présence d'un éther couronne pendant 24h à 55°C. Le composé **(12)** est isolé après purification sur colonne de silice (AcOEt 90% - MeOH 10%) avec un rendement de 80% (figure 61).

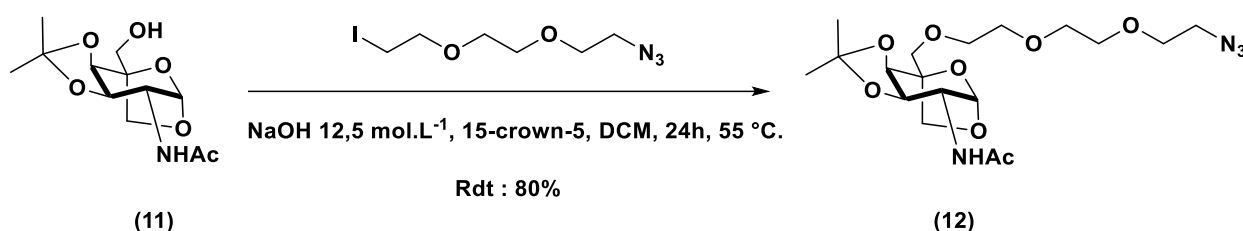


Figure 61 : Réaction d'O-alkylation introduisant la chaîne triéthylène glycol en position 6 du sucre.

Le composé **(13)** est obtenu en deux étapes. La première est une réaction de déprotection en milieu acide des hydroxyles en position 3 et 4 du composé **(12)** par action d'acide acétique dans un mélange eau/méthanol à 70°C. Aucune purification n'est nécessaire et l'ensemble des sous-produits formés est éliminé par lyophilisation. La deuxième étape est une réduction pallado-catalysée sous pression d'hydrogène dans l'éthanol anhydre afin de transformer l'azoture en amine correspondante. Pour éviter toutes réactions secondaires, de l'acide *para*-toluènesulfonique (APTS) est ajouté au milieu réactionnel afin de protoner l'amine formée. Le composé **(13)** est ainsi obtenu sans purification supplémentaire avec un rendement global de 90% (figure 62).

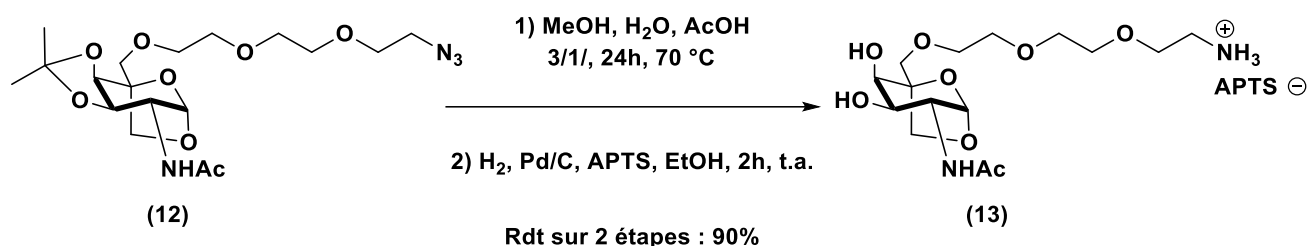


Figure 62 : Réaction de déprotection des hydroxyles et réduction de l'amine.

La (figure 63) présente les deux dernières étapes de la synthèse. Dans un premier temps, l'amine primaire du composé (13) est libérée de son sel par passage sur une résine basique, l'Amberlist IRN 78, dans le méthanol. Dans un deuxième temps, après filtration de la résine, l'amine primaire est mise en réaction avec le bitoscanate dans la DMF afin d'obtenir le ligand final (PMS-02). Pour éviter la formation d'un dimère, l'amine est préalablement dissoute dans la DMF et ajoutée goutte à goutte dans une solution contenant cinq équivalents de bitoscanate également dans la DMF. La purification du composé (PMS-02), très polaire, est difficile et plusieurs colonnes de silice (DMC 80% - MeOH 20%) sont nécessaires pour finalement l'isoler avec un rendement de 37%.

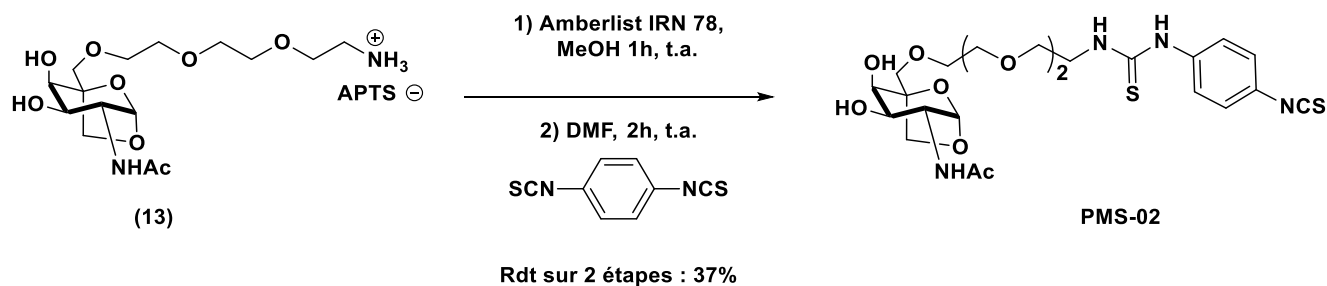


Figure 63 : Synthèse du ligand (PMS-02).

Pour conclure, le ligand (PMS-02) est obtenu après 17 étapes de synthèse pour un rendement global inférieur à 0,005%. Ce ligand est relativement difficile à produire et il serait plus intéressant pour l'évolution du projet de concevoir de nouvelles structures plus accessibles qui demanderaient, d'une part, moins d'étapes de synthèse et, d'autres part, des rendements de réactions plus élevés.

C'est dans cette optique que nous avons réalisé une étude par modélisation moléculaire afin d'identifier de nouvelles structures pertinentes pour le ciblage de l'ASGPr et *a priori* plus simple à synthétiser.

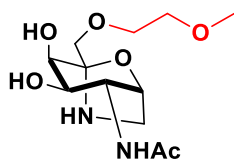
## 3. Modélisation

### A. Introduction

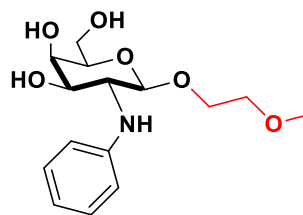
Une étude par modélisation moléculaire a été menée dans le but de mieux appréhender les interactions entre le sucre et le domaine de reconnaissance de l'ASGPr et de pouvoir éventuellement en sortir de nouvelles molécules plus faciles à synthétiser que le ligand (**PMS-02**). Ne possédant pas de compétence dans ce domaine, ce travail a été réalisé par Zakaria Bouchouireb, doctorant au laboratoire Ceisam. Pour commencer nous nous sommes basés sur les différents travaux précédemment cités (publication sur le GalNac ponté de Sanhueza<sup>144</sup>, étude de l'interaction entre le galactose<sup>137</sup> et le GalNac<sup>138</sup> avec l'ASGPr) afin de construire une bibliothèque de composés (**annexe 1**) se décomposant en deux séries : les sucres classiques et les sucres pontés. Nous avons procédé à une étude de docking moléculaire sur l'ensemble des composés et à une étude de dynamique moléculaire entre l'ASGPr, la *N*-acétylgalactosamine et le sucre ponté afin de mieux décrire le comportement de chaque série de sucres en interaction avec le récepteur. La structure 5JQ1 issue de la PDB est le domaine de reconnaissance des sucres de l'ASGPr utilisé dans cette étude. Cette structure est la plus pertinente car elle possède une résolution de 1,83Å.

### B. Chimiothèque et docking

Afin d'explorer les corrélations existantes entre structures et activités de différents ligands de l'ASGPr nous avons élaboré une chimiothèque de 70 composés en partant des deux séries évoquées précédemment (sucre classique / sucre ponté). Nous nous sommes concentrés sur quelques propriétés spécifiques : la configuration du carbone anomérique du sucre ( $\alpha$  et  $\beta$ ), la taille du cycle, le type d'hétéroatome endocyclique et la nature des substituants en position C1, C2 et C6. Afin d'avoir une vision plus représentative de l'interaction du ligand final sans pour autant devoir le modéliser en intégralité, nous avons étudié uniquement le motif sucre en ajoutant un petit espaceur éthylène glycol méthylé sur la position appropriée afin de mimer la chaîne (**figure 64**).



(PA7\_1)



(PA39)

Figure 64 : Exemple de composé **(PA7\_1)** avec l'éthylène glycol méthylé sur la position 6 et **(PA39)** avec la chaîne sur la position 1.

Pour comparer les 70 composés de la chimiothèque entre eux et avec la littérature, un certain nombre de ligands connus y sont incorporés. Ainsi :

- L'étude comporte le galactose et le mannose (sans chaîne éthylène glycol méthylé)
- Le composé **(PA1)** est le GalNAc (avec chaîne éthylène glycol méthylé)
- Le composé **(PA4)** est le GalNAc ponté utilisé dans la confection du ligand **(PMS-02)** (avec chaîne éthylène glycol méthylé)

Les molécules ont été dockées en utilisant le mode extra précision (XP) de Glide présent dans la version 2019-3 de Schrödinger Suite. Les poses obtenues ont été évaluées par GlideScore (GScore), une fonction de scoring du logiciel Schrodinger, le GScore est calculé par l'équation suivante :

$$\text{GScore} = (a * \text{VdW}) + (b * \text{Coul}) + \text{Lipo} + \text{Hbond} + \text{Metal} + \text{Site} + \text{BuryP} + \text{RotB}$$

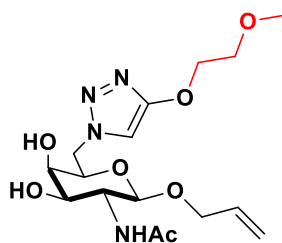
où les différents termes de l'équation sont donnés dans le tableau suivant (tableau 4).

<i>Terme</i>	<i>Description</i>
<i>VdW</i>	Énergies de Van der Waals (hydrophobe)
<i>Coul</i>	Énergies coulombiennes (polaire)
<i>Lipo</i>	Contacts lipophiles
<i>Hbond</i>	Énergies des liaisons hydrogène
<i>Metal</i>	Chélation métalliques
<i>Site</i>	Interactions polaires dans le site de fixation
<i>BuryP</i>	Pénalités pour les groupements polaires enfuis
<i>RotB</i>	Pénalités pour les liaisons flexibles figées
<i>a</i>	Facteur de poids pour VdW = 0.065
<i>b</i>	Facteur de poids pour Coul = 0.130

Tableau 4 : Liste des termes de l'équation et leurs descriptions.

## C. Résultats du docking

Les résultats de docking en mode XP pour tous les ligands sont donnés sous forme de graphique de sucettes présenté (figure 66). Pour chaque position obtenue par docking, le score de docking (Gscore) est donné sous forme globale en kcal/mol, il est composé de 3 termes : polaire, Van der Waals (VdW) et pénalités. L'étoile bleue représente la valeur du score de docking globale, le carré orange le terme de VdW et le cercle rouge le terme de pénalité qui regroupe les différentes pénalités de torsion et d'encombrement stérique. Le trait bleu correspond à la soustraction des termes de VdW, pénalités et score de docking globale permettant ainsi d'exprimer la composante polaire. Les ligands colorés en rouge correspondent à la série des sucres classiques, tandis que ceux en noir correspondent à la série des sucres pontés. En nous intéressant plus précisément aux scores obtenus, nous pouvons observer une tendance globale montrant que la majorité de la chimiothèque (68 composés sur 70) interagit favorablement avec le récepteur puisque nous avons un score de docking négatif. Nous observons également que le terme polaire est le terme prépondérant dans les composantes du score de docking, indiquant une majorité d'interactions polaires entre le sucre et la poche de l'ASGPr. La pose avec le meilleur score de docking concerne le composé (**PA26**) qui appartient à la série des sucres non pontés, avec une valeur de -6,503 kcal/mol, cette valeur devrait correspondre à une affinité de l'ordre de la dizaine de millimolaire (figure 65).



(PA26)

Figure 65 : Structure du composé (PA26).

Concernant une éventuelle ségrégation de l'interaction des ligands en fonction de leur structure, nous observons une répartition quasi aléatoire des composés pontés et non pontés. Ceci laisse à penser que la rigidification du sucre ne semble pas participer de manière importante à l'interaction avec le récepteur. Or, nous savons que le contraire a été démontré par les travaux de Sanhueza, notamment pour les composés (**PA1**) et (**PA4**). De ce fait, nous émettons l'hypothèse que le docking, avec la fonction de scoring Gscore, n'arrive pas complètement à prendre en compte certains effets

conformationnels expliquant les différences d'affinité entre la série des composés pontés ou non. De plus, l'écart-type de score de docking pour les 22 premiers composés (**PA26** → **PA43**) n'est que de 0,142 kcal/mol, ce qui est très faible. Il est donc difficile avec cette méthode de correctement classer les ligands et donc de les discriminer.



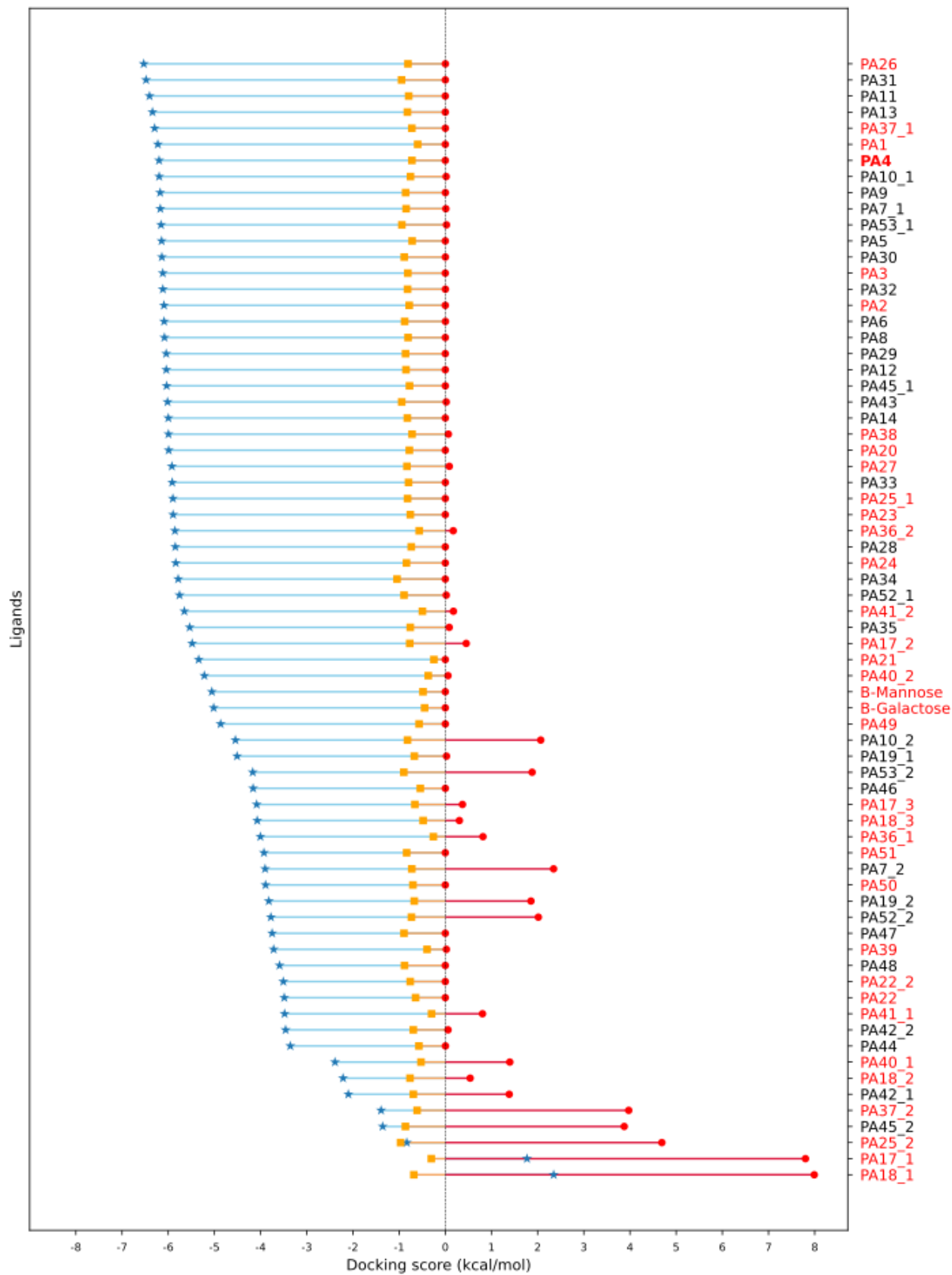


Figure 66 : Graphique en sucette représentant le score de docking (Gscore) pour chaque ligand, (PA4) est la référence.

## D. Mode d'interaction et optimisation

Afin de mieux comprendre les scores de docking obtenus, nous nous intéressons dans cette section à l'élucidation du mode de fixation des ligands. La (figure 67), représente le mode de liaison de (PA4). Les liaisons bleues représentent les liaisons hydrogène, les liaisons violettes les chélation métalliques. Ce ligand est stabilisé dans le site de reconnaissance par une multitude d'interactions polaire qui est présentée dans le (tableau 5). Ce tableau met également en évidence la présence des mêmes interactions pour un échantillon représentatif de la chimiothèque soit (PA26) le meilleur ligand non cyclique, (PA31) le meilleur ligand cyclique, (PA34) un ligand avec un score moyen et (PA53\_2) avec un score faible et une pénalité élevée.

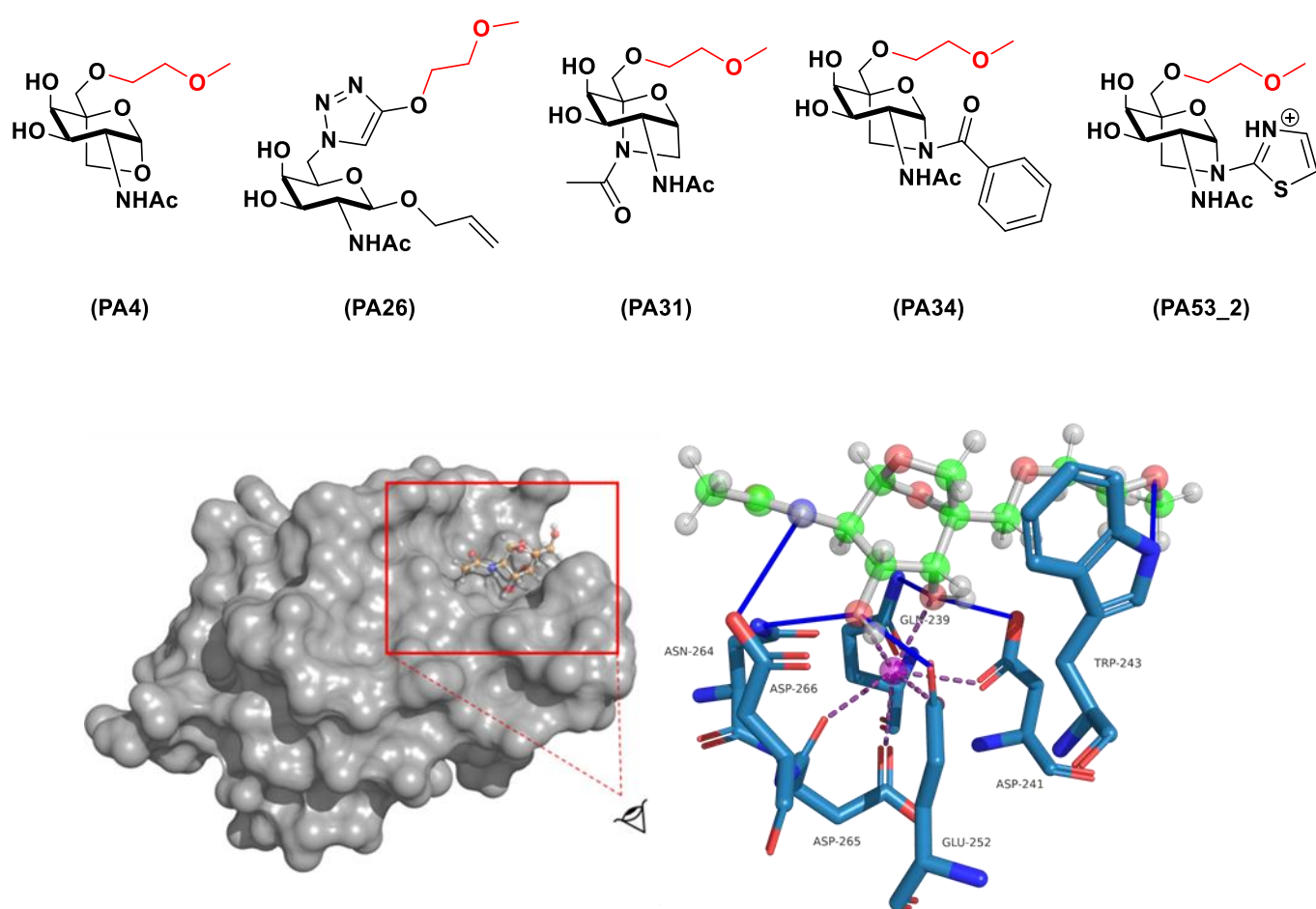
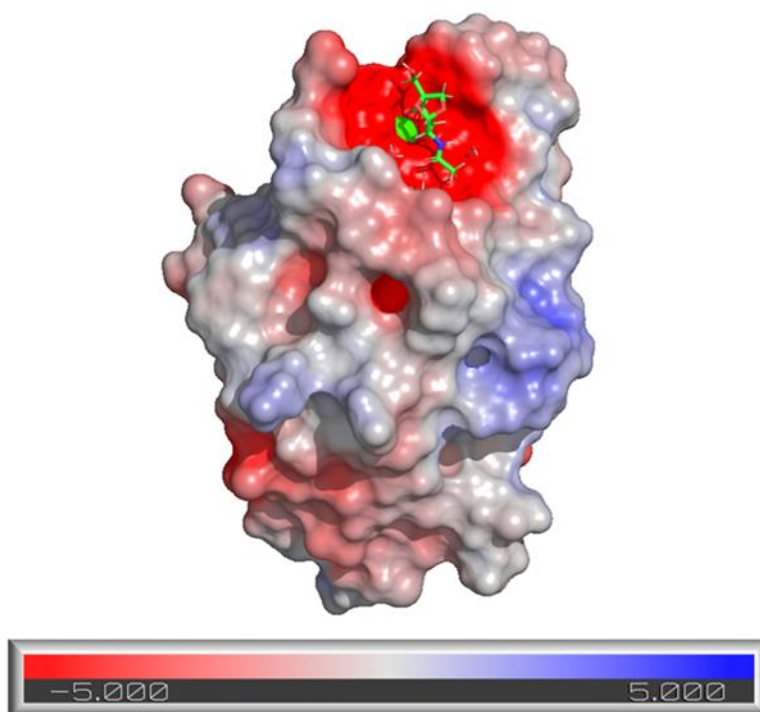


Figure 67 : Structure des différents ligand et mode d'interaction de (PA4) dans le site de fixation de l'ASGPr.

## Erreur ! Liaison incorrecte.

*Tableau 5 : listes des interactions du mode de fixation pour différents ligands.*

Au travers de ces interactions, il est important de retenir deux ensembles définissant le schéma de base d'interactions dans le site de fixation : (i) la chélation de l'ion calcium par les hydroxyles en position C3 et C4, puis (ii) les liaisons hydrogène impliquant les résidus Gln-239, Asp-241, Glu-252, Asn-264 et Asp-266. Il est à noter que ce sont les mêmes interactions clés décrites par l'équipe de Meier.<sup>139</sup> La présence de cette multitude d'interactions polaire est attendue étant donné que le site de fixation possède un potentiel électrostatique bien plus négatif que le reste du domaine protéique (**figure 68**). Ceci explique la participation importante du terme polaire aux scores de docking observés.



*Figure 68 : Isosurface représentant le potentiel électrostatique du domaine de fixation de l'ASGPr issue de la structure 5JQ1. Les valeurs sont exprimées en (kT/e).*

Au vu des modes de fixation sensiblement similaires entre les différents ligands, nous avons essayé un docking avec ajustement induit aussi appelé semi-flexible ou (IFD) pour mieux décrire les effets conformationnels. Ce type de docking permet une liberté de mouvement des chaînes principales de la protéine au voisinage du site d'interaction. Cette technique décrit mieux les éventuels changements conformationnels mineurs de la protéine mais demande beaucoup de ressources de calcul. Suite à des limitations de ressources informatiques, nous avons réduit l'IFD au composé (**PA4**) dont les résultats des modes d'interactions sont également décrits dans le (**tableau 5**), dernière ligne **PA4\_IFD**.

Malheureusement nous observons que les modes de fixation obtenue avec le docking en mode XP et l'IFD sont identiques. Ainsi, sachant que l'essentiel du mode d'interaction est dépendant de la chélation du calcium par les hydroxyles et que ce dernier interagit avec une multitude de résidus, il est nécessaire d'utiliser des méthodes prenant en compte toute la flexibilité du récepteur. Cet aspect a également été investigué par une étude pilote de dynamique moléculaire mais malheureusement il n'existe aucun champ de force permettant de correctement décrire l'interaction entre ce type de composé et le cation calcium. Cela est un réel problème car cette interaction est décrite comme primordiale pour ce type de lectine. Les résultats de l'étude pilote par dynamique moléculaire ne sont pas concluants et montreraient presque une mauvaise affinité du GalNAc ponté pour le récepteur. Ceci est notamment dû à des limitations techniques et au choix du champ de force peu pertinent.

## E. Conclusion

En conclusion, les résultats obtenus après le docking moléculaire ne permettent pas de ségréguer pertinemment les différents composés de la chimiothèque. En effet, bien que le docking reproduise le bon mode de fixation des composés modèles, il ne prend pas suffisamment en compte certains aspects conformationnels essentiels et notamment l'interaction entre les ligands et le cation calcium qui est pourtant décrite comme essentielle pour des lectines de types C comme l'ASGPr. Il a donc été nécessaire d'utiliser des méthodes incluant différents degrés de flexibilité partielle (IFD) et complète

(MD). Cela n'a toujours pas permis d'obtenir des résultats concluants mais a mis en évidence l'absence d'un champ de force approprié pour une étude concernant ce système. Il serait donc nécessaire d'investir du temps et des ressources afin de développer et tester un champ de force adéquat avant de pouvoir reprendre ou poursuivre une étude sur le même système.

Cependant, malgré des résultats décevants, le travail de docking peut nous permettre d'orienter notre choix vers la synthèse de nouveaux analogues du ligand (**PMS-02**). Ainsi, il ressort que nous pourrions nous intéresser à des sucres pontés possédant :

- Différentes tailles du cycle avec 2 ou 3 carbones (**PA11** et **13**).
- Un atome d'azote endocyclique à la place de l'oxygène (**PA10\_1** et **PA9**).
- Un hétéroatome endocyclique en position 5 et non en position 1 (**PA5** et **PA7\_1**).

Nous pourrions aussi effectuer la synthèse de sucres classiques fonctionnalisés par une espaceur en position 6 et par une fonction apolaire en position anomérique comme le suggère les composés (**PA2**, **PA3** ou **PA26**) (figure 69).

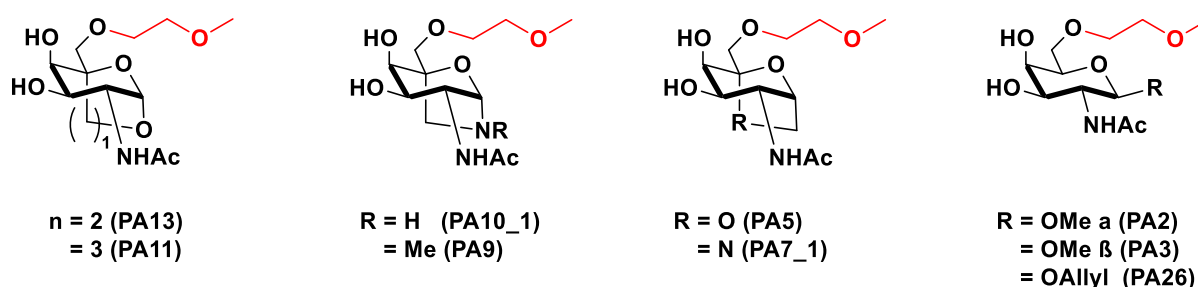


Figure 69 : Structure des différents sucres.

Dans la partie suivante nous allons nous intéresser à la modification de la partie pontée. Nous avons fait le choix de partir sur l'inversion de la position de l'oxygène endocyclique du second cycle plutôt que de changer sa nature et cela pour des raisons de simplicité de synthèse. Cette possibilité n'est toutefois pas à écarter et pourrait faire l'objet de travaux ultérieurs.

## 4. Synthèse du composé PMS-02bis

Le ligand (**PMS-02**) contient un motif innovant qui a montré un intérêt pour le ciblage des ASGPr. Néanmoins, sa synthèse est délicate et nécessite 17 étapes pour un rendement global inférieur à 0,005%. La complexité de ce ligand rend difficile sa production à grande échelle et une potentielle utilisation clinique. C'est pourquoi nous avons orienté nos travaux vers la synthèse de composés *a priori* plus simple à obtenir et toujours affins pour le récepteur ASGP. La structure pontée du GalNAc de (**PMS-02**) permet une meilleure interaction avec la poche lectinique. L'idée est donc de conserver

cette structure bicyclique tout en y apportant des modifications simples, comme la nature et la position de l'hétéroatome endocyclique ou encore la taille du cycle. L'étude par modélisation nous a montré la possibilité que de telles structures interagissent effectivement avec le récepteur.

Une réaction décrite dans la littérature par l'équipe d'Ernesto Suarez permet d'accéder à des motifs bicycliques de type dioxabicyclo[3.2.1]octane, analogue du GalNAc ponté mais avec la position de l'oxygène qui diffère.<sup>151</sup> L'idée est donc d'utiliser cette stratégie pour aboutir à un ligand proche de **(PMS-02)** mais en réduisant le nombre d'étapes de synthèse. La structure du composé cible appelé **(PMS-02bis)** est présentée (figure 70). Cette stratégie est cohérente avec l'étude par modélisation moléculaire où le ligand **(PMS-02bis)** présente un sucre équivalent au composé **(PA5)**. Bien qu'il ne s'agisse pas du meilleur hit, sa synthèse s'avère moins difficile que celle d'autres composés révélés par le docking.

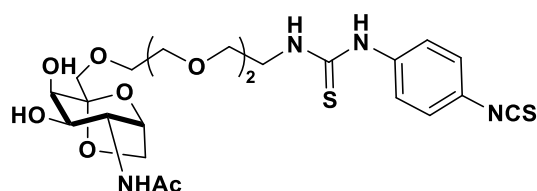


Figure 70 : Structure du ligand **(PMS-02bis)**.

Pour bien comprendre la conception du ligand (**PMS-02bis**), il est nécessaire de s'arrêter sur la réaction radicalaire de types 1,5 HAT (réarrangement type Surzur-Tanner<sup>152</sup>) qui permet la formation du bicyclic. L'équipe d'Ernesto Suarez l'a décrite dans les années 2000 en faisant réagir des dérivés de sucres de types  $\alpha$ -D-C-glycopyranosyl méthanol (c'est-à-dire possédant un groupement  $\text{CH}_2\text{OH}$  de stéréochimie  $\alpha$  en position anomérique) avec du (diacétoxyiodo)benzène (DIB) ou de l'iodosylbenzène et de l'iode en présence de lumière. (**figure 71**).<sup>151,153</sup>

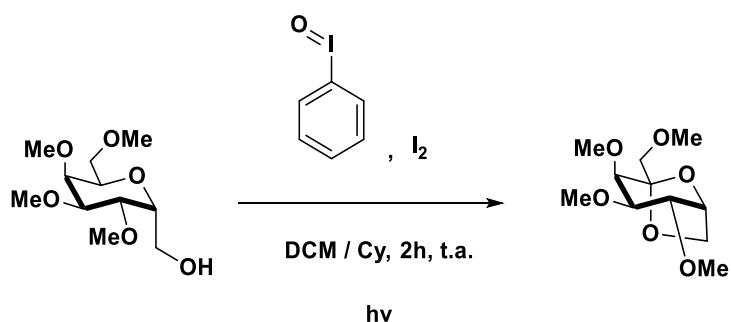


Figure 71 : Exemple de réaction de Suarez avec un dérivé  $\alpha$ -D-C-galactopyranosyl méthanol.<sup>153</sup>

Cette réaction relativement douce peut être considérée comme une substitution intramoléculaire qui commence par un transfert d'hydrogène favorisé par un radical alkoxy puis une oxydation de l'intermédiaire radicalaire en un ion oxacarbénium. Le mécanisme détaillé de cette réaction est présenté (**figure 72**).

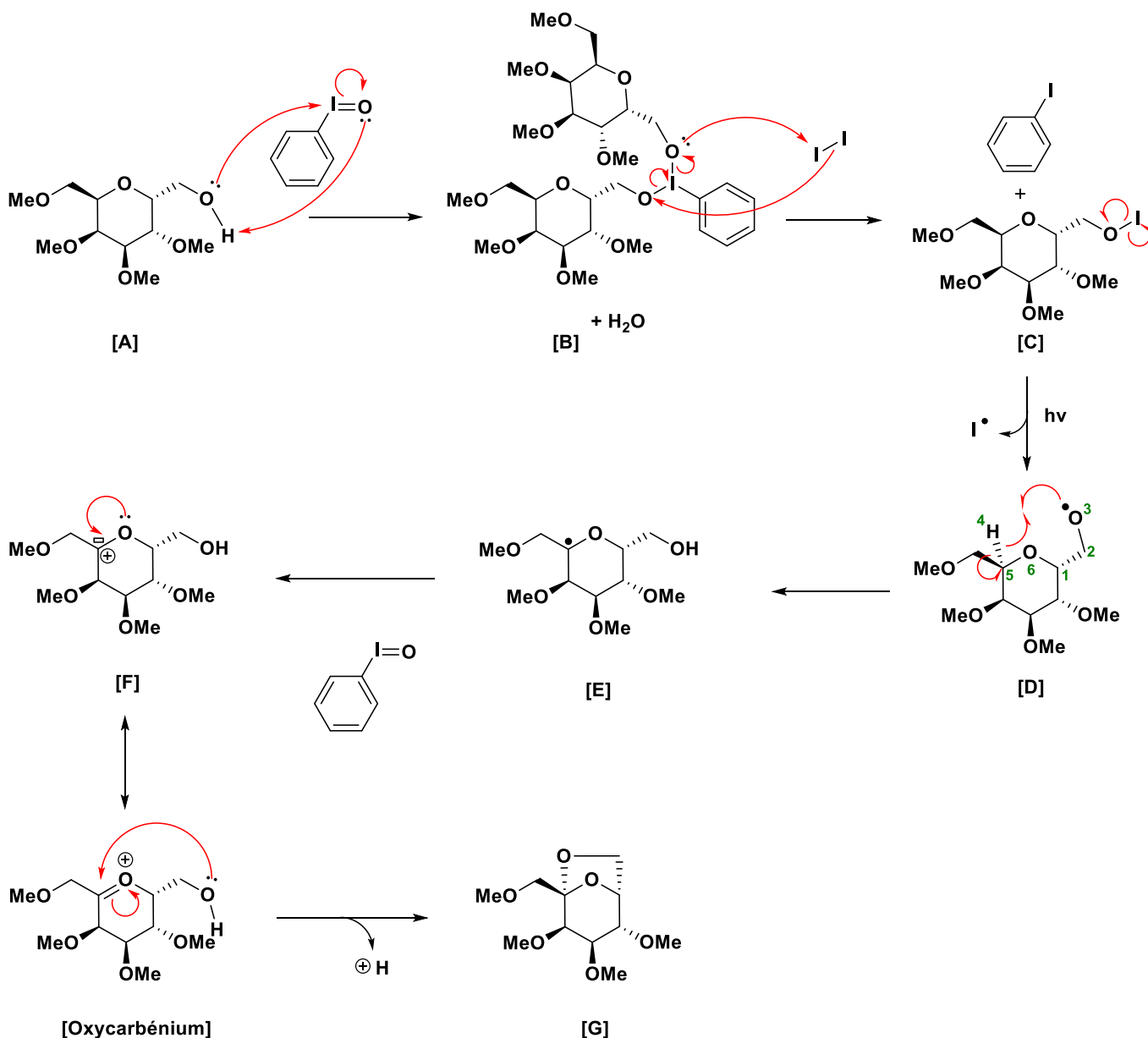


Figure 72 : Mécanisme de la réaction de cyclisation radicalaire selon Suarez.

Pour débiter, l'iodosylbenzène réagit avec l'hydroxyle libre de deux molécules de C-galactopyranosyl méthanol [A] pour donner l'intermédiaire périodé [B].<sup>154</sup> Cet intermédiaire réagit avec le diiode pour former du iodobenzène et une liaison hypoiodite [C] qui se clive à la lumière en radical alcoyle et iode [D]. Ce dernier se dimérise en diiode tandis que l'alcoyle effectue une réaction de transfert d'hydrogène intramoléculaire pour aboutir à l'intermédiaire radicalaire [E]. Cette étape est régiosélective car elle passe par un intermédiaire à six centres favorables [D']. Le radical [E] est ensuite oxydé par une molécule d'iodosylbenzène pour former un carbocation stabilisé [F] qui subit une attaque intramoléculaire de l'alcool primaire pour aboutir au composé bicyclique désiré [G].



Pour la confection du ligand (**PMS-02bis**), il faudra réaliser une réaction de cyclisation radicalaire similaire mais en partant d'un dérivé C-galactopyranosyl méthanol avec une fonction acétamide en position 2. Cette réaction n'ayant jamais été décrite avec un sucre N-acétylé, si elle ne fonctionne pas il sera possible de remplacer la fonction acétamide par un précurseur azoture. D'autre part, les groupements protecteurs des hydroxyles du sucre seront choisis en fonction de leur synergie avec la voie de synthèse (**figure 73**).

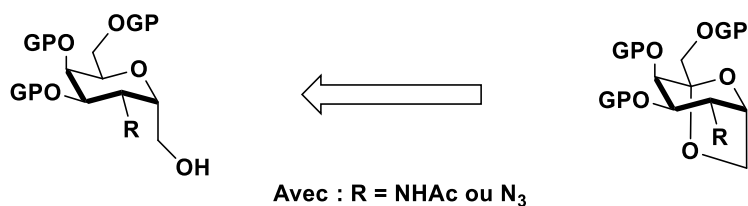


Figure 73 : Rétro synthèse montrant un exemple de composé de type C-glycopyranosyl méthanol permettant d'obtenir le bicyclic désiré.

Pour la synthèse du sucre de départ nous avons envisagé plusieurs stratégies regroupées sous forme d'un schéma rétro synthétique (**figure 74**) et détaillées ci-dessous.

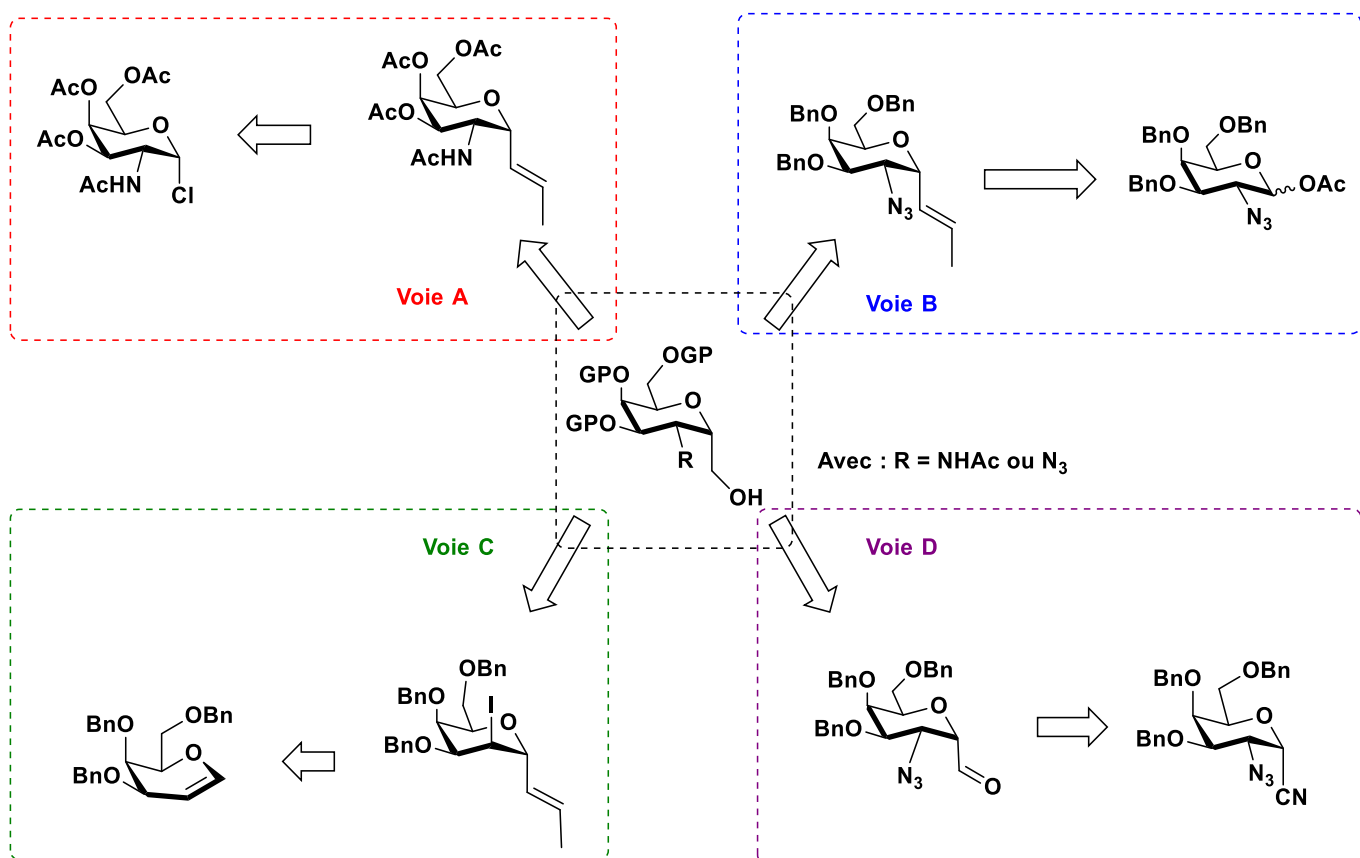


Figure 74 : Présentation des différentes voies de synthèse pour le dérivé  $\alpha$ -D-C-galactopyranosyl méthanol.

## A. Voie A

La première voie envisagée, présentée ci-dessous, débute par la protection des hydroxyles du GalNAc, suivie d'une chloration de la position anomérique qui permet une réaction de C-glycosylation pour introduire un groupement allyle, précurseur du motif  $\text{CH}_2\text{OH}$  de l' $\alpha$ -D-C-galactopyranosyl méthanol (figure 75).

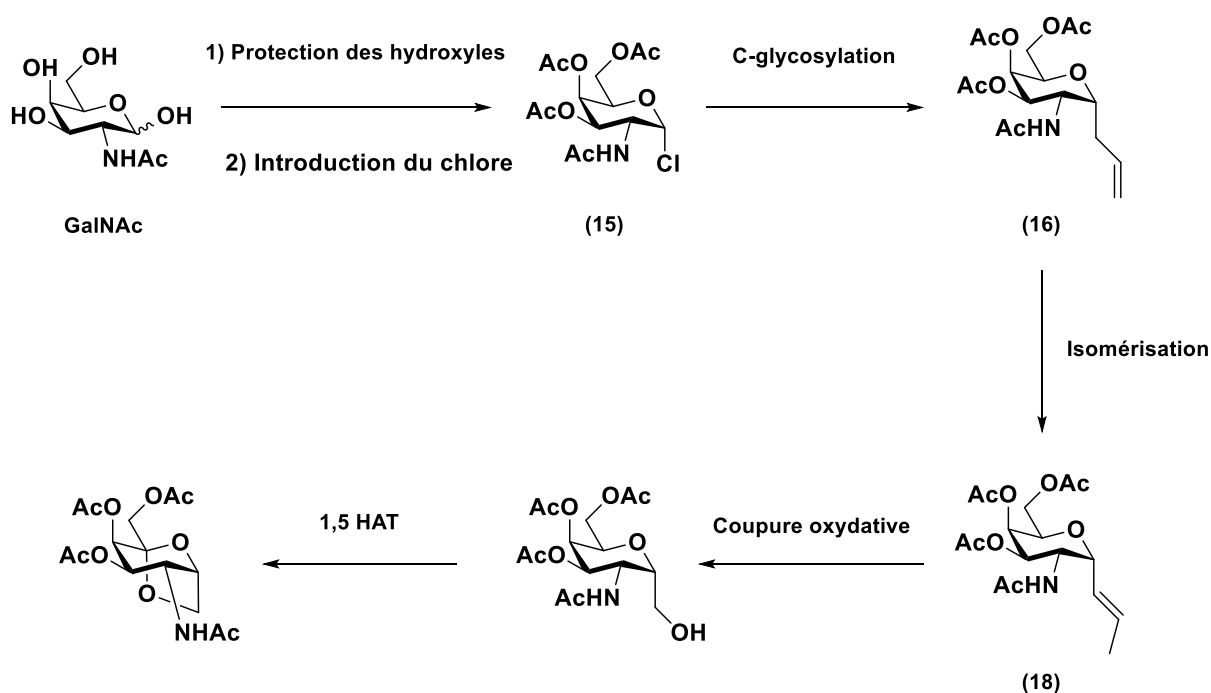


Figure 75 : Obtention de l' $\alpha$ -D-C-galactopyranosyl méthanol d'intérêt selon la voie A et réaction d'1,5 HAT.

La première étape de la synthèse est la protection des fonctions hydroxyles du sucre par des groupements acétates, classiquement utilisés en chimie des sucres. Cette étape se fait avec de la pyridine et de l'anhydride acétique, comme solvants de réaction, en condition anhydre sous  $\text{N}_2$  pendant 35h à température ambiante. De la DMAP est ajoutée en quantité catalytique afin d'activer l'anhydride acétique. Le composé (14) est obtenu avec un rendement de 98%. Cette réaction ne demande pas de purification car le sous-produit formé est l'acide acétique qui est évaporé sous pression réduite (figure 76).

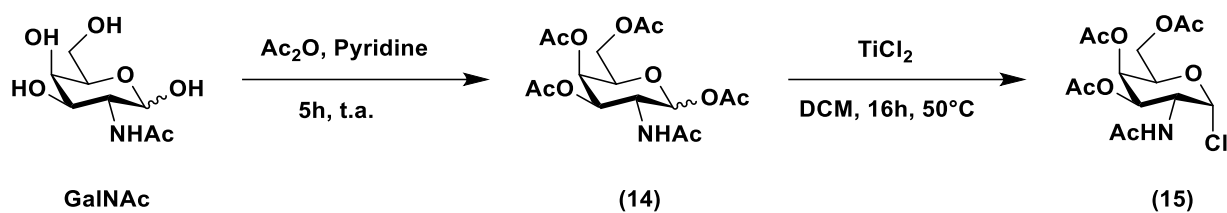


Figure 76 : Réaction de protection des alcools et activation de la position anomérique du sucre.

La seconde étape est l'activation de la position anomérique du sucre par substitution du groupement acétate par un atome de chlore. Cette étape se réalise dans le dichlorométhane en présence de chlorure de titane (II) à une température de 50°C. Le composé **(15)** est obtenu après purification sur gel de silice (AcOEt 100%) avec un rendement moyen de 53% étant donné la formation d'un produit secondaire **(15')** résultant de l'élimination du chlore et de l'hydrogène en C2 (figure 77).

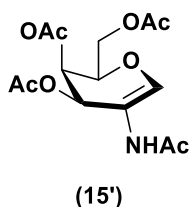


Figure 77 : Produit d'élimination obtenu durant la réaction de chloration.

L'étape suivante est une réaction de C-glycosylation pour introduire une fonction allyle en position anomérique et donner le composé **(16)**.<sup>155-157</sup> L'isomérisation du groupement allyle conduit à un alcène stable qui peut facilement être clivé par oxydation en condition réductrice pour former le motif CH<sub>2</sub>OH souhaité.<sup>155</sup> Plusieurs essais ont été nécessaires afin d'obtenir le produit souhaité avec un rendement viable pour la poursuite de la synthèse. Parmi les optimisations présentées (figure 78) on peut notamment citer le changement d'acide de Lewis et l'utilisation de la chimie radicalaire.

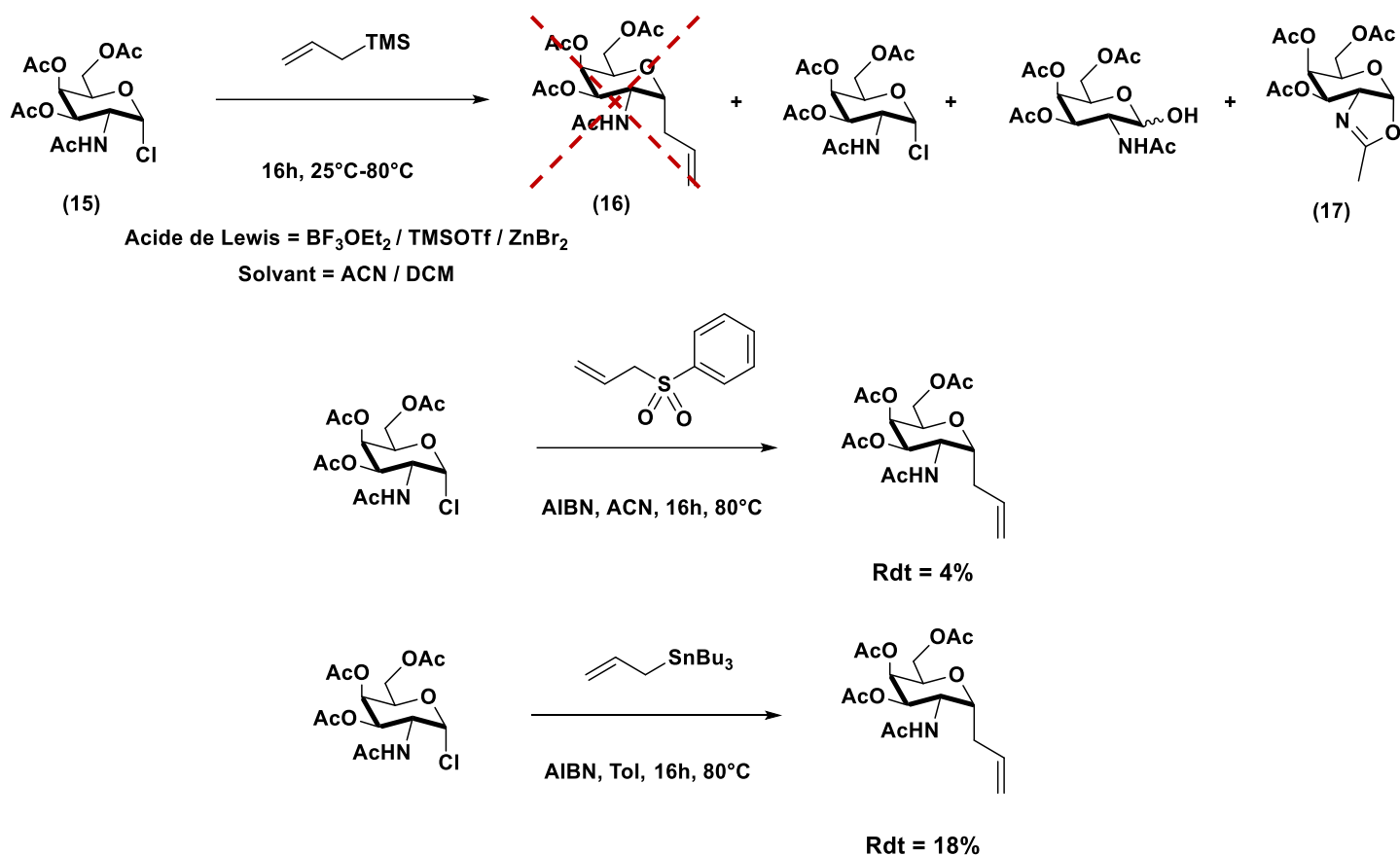


Figure 78 : Différents essais de réaction de C-glycosylation permettant d'obtenir le composé (16).

Quels que soient les acides de Lewis et les conditions opératoires utilisées, nous n'avons jamais observé la formation du composé (16). L'utilisation de  $\text{BF}_3\text{OEt}_2$  ou de  $\text{TMSOTf}$  à température ambiante a conduit à un mélange de produits de départ et d'hydrolyse de la position anomérique. Par contre, l'emploi de  $\text{TMSOTf}$  à reflux du  $\text{DCM}$  donne bien l'intermédiaire de glycosylation (17) mais sans la formation du composé désiré. L'utilisation de bromure de zinc (II) anhydre n'a pas permis non plus de former le composé (16).

Les réactions radicalaires ont eu une réussite modeste : l'action d'allylphényle sulfone en présence d'azobisisobutyronitrile comme initiateur radicalaire donne le composé (16) avec un rendement faible de 4%. Par contre, en présence d'allyltributylstannane dans le toluène à  $80^\circ\text{C}$  toujours en utilisant l'AIBN comme initiateur radicalaire nous isolons le composé (16) après purification sur gel de silice (Cy 25% -  $\text{AcOEt}$  75%) avec un rendement de 18%. Les rendements faibles peuvent s'expliquer par la présence de la fonction amide en position 2 du sucre. Outre le fait d'engendrer le bicyclic (17), cette fonction n'est pas compatible avec une étape radicalaire car elle peut former un radical amidyle. Les équipes de Horton et de Schmidt ont rencontré les mêmes obstacles en montrant qu'une réaction de polymérisation a également lieu.<sup>158,159</sup>



## B. Voie B

La stratégie de synthèse du ligands (**PMS-02**) comprend plusieurs intermédiaires avec un azoture en position C2 du sucre. Cet azoture peut être dans notre cas un bon précurseur du groupement acétamide et ne devrait pas interférer dans la réaction de C-glycosylation. La voie B reprend les bases synthétiques de la voie A mais avec des intermédiaires qui possèdent un azoture en position C2. Comme le montre la (figure 80), le dérivé  $\alpha$ -D-C-galactopyranosyl méthanol est obtenu de la même manière que la voie A par une suite réactionnelle de réduction, coupure oxydative et d'isomérisation à partir de l'allyle (**22**). Pour obtenir ce dernier il est nécessaire d'accéder à la molécule (**20**) qui est le produit de la réaction d'azidonitration du D-galactal, composé commercial, après protection de ses hydroxyles par des groupements benzyles. Le composé (**22**) est alors facilement atteignable *via* une réaction de C-glycosylation sur le sucre (**21**), lui-même obtenu par substitution nucléophile du groupement nitrate de (**20**) par un acétate.

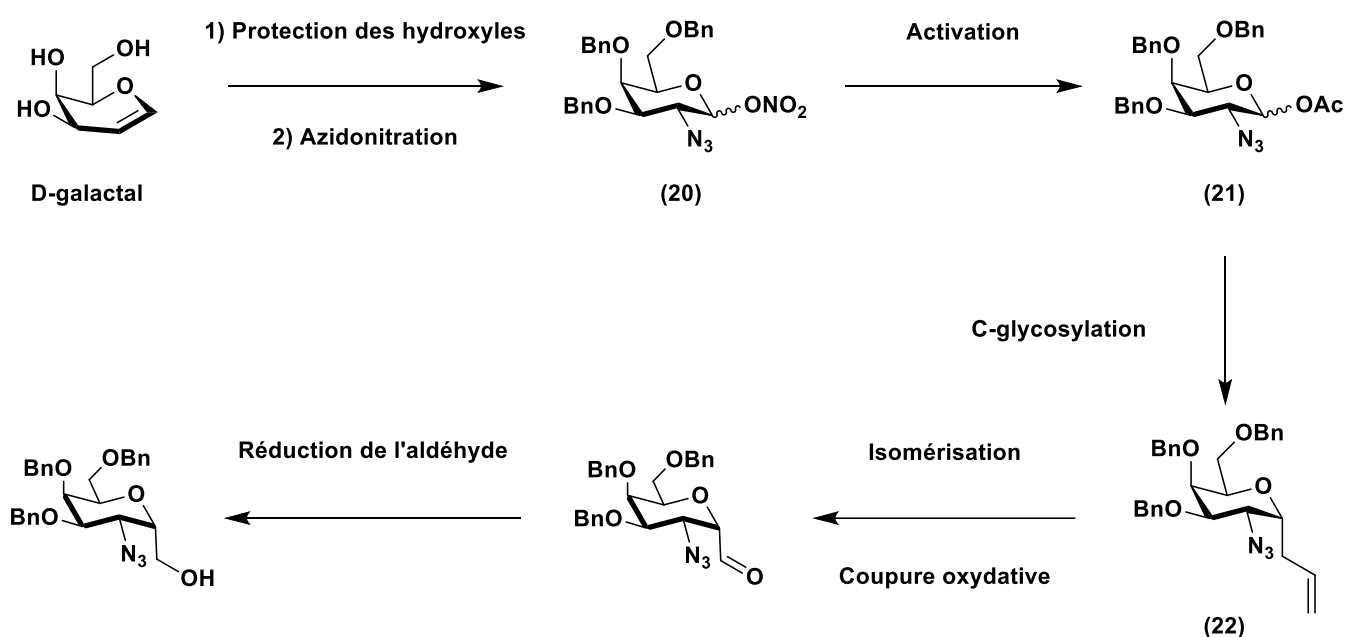


Figure 80 : Obtention de l' $\alpha$ -D-C-galactopyranosyl méthanol d'intérêt selon la voie B.

La première étape de cette synthèse est la protection des hydroxyles du D-galactal, produit commercial, par l'action de bromure de benzyle et de NaH dans la DMF. Le produit (**19**), obtenu après purification sur gel de silice avec un rendement de 91% est traité par du nitrate de cérium et d'ammonium, et de l'azoture de sodium dans l'acétonitrile à  $-20^{\circ}\text{C}$  afin d'effectuer la réaction

d'azidonitration (**figure 81**). Le produit (**20**) est isolé sous la forme d'une huile incolore très visqueuse après purification avec un rendement faible de 40% et une forte dégradation observée. Ici, les groupements benzyles sont choisis pour leur propriété à favoriser les réactions de glycosylation en facilitant l'apparition de l'intermédiaire oxacarbénium. Cela peut s'expliquer par l'approche armée/désarmée dont un des effets montre, notamment, le fait qu'un groupement protecteur des hydroxyles du sucre fortement attracteur (comme l'acétate) déstabilise l'ion oxacarbénium ralentissant, voir inhibant, alors sa formation. A contrario, un groupement moins attracteur comme le benzyle favorise mieux l'apparition de l'intermédiaire.<sup>160,161</sup>

Contrairement à son homologue le composé (**1**) présenté dans la synthèse du ligand (**PMS-02**) (**figure 46**), cette réaction d'azidonitration ne conduit pas à la formation d'un second stéréoisomère en C2. Ceci est facilement explicable par la présence des groupements benzyles plus volumineux que les groupements acétates, induisant donc une gêne stérique 1,3 diaxiale qui est alors trop importante pour permettre la formation du stéréoisomère en C2.

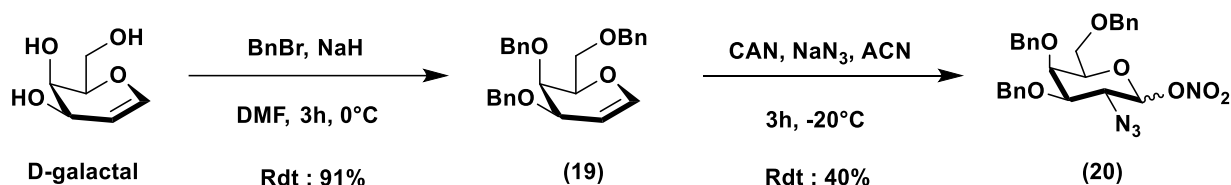


Figure 81 : Protection des hydroxyles du D-Galactal suivi d'une réaction d'azidonitration.

La réaction de C-glycosylation pour obtenir le composé (**22**) a directement été tentée sur le nitrate (**20**). Différentes conditions ont été explorées, notamment en variant le solvant (ACN / DCM) ainsi que la nature et le nombre d'équivalent d'acide de Lewis, mais aucune n'a donné le produit souhaité. Seul le produit de départ est retrouvé après réaction avec, selon les conditions opératoires, plus ou moins de produits de dégradation (**figure 82**).

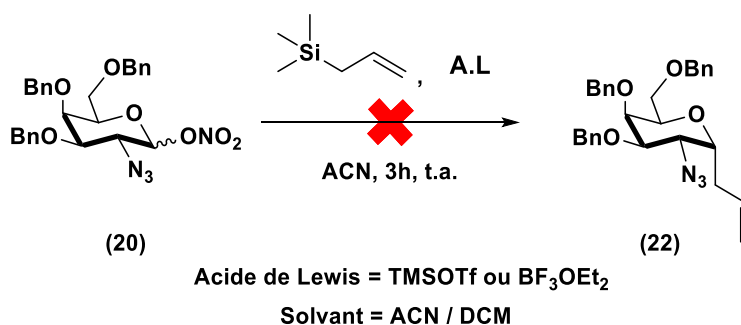


Figure 82 : Tentative de C-glycosylation sur le composé (**20**).

Il peut être conclu que le précurseur **(20)** n'est pas un bon donneur de glycosyle donneur, même s'il a parfois été utilisé dans la littérature comme tel.<sup>162,163</sup> Le groupement nitrate ne permet pas d'activer suffisamment la position anomérique du sucre mais peut cependant être facilement substitué par un brome ou un acétate qui sont de meilleurs groupements partants et bien plus utilisés dans les réactions de glycosylation<sup>164</sup>.

Afin de pouvoir réaliser la réaction de C-glycosylation, le sucre **(20)** doit être préalablement activé par substitution du groupement nitrate, en position anomérique, par un acétate. Ce dernier induit un effet inductif attracteur plus fort ce qui va exalter le caractère électrophile du carbone anomérique. Cette étape a lieu par action d'acide acétique et de triéthylamine à 100°C durant 1h. Le produit initial étant un mélange d'anomères  $\alpha$  et  $\beta$ , le produit obtenu est également un mélange de ces deux anomères. Le composé **(21)** est isolé par purification sur gel de silice (Cy 90% - AcOEt 10%) avec un rendement de 78% (figure 83).

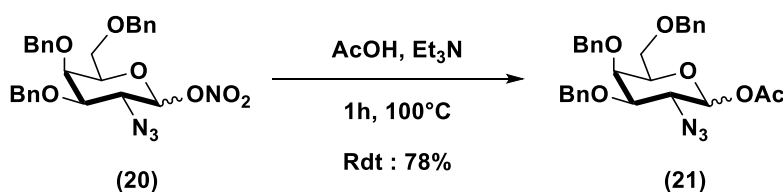


Figure 83 : Activation de la position anomérique du sucre **(20)**.

Le sucre **(21)** est traité avec de l'allyltriméthylsilane dans l'acétonitrile à température ambiante durant 3h en présence de TMSOTf qui permet la réaction de C-glycosylation. Lors de cette étape, il y a formation d'un produit secondaire, le composé **(23)**, que l'équipe de Bednarski a décrit comme étant le résultat de la cyclisation 1,3 dipolaire entre l'azoture en position 2 et l'allyle anomérique du composé **(22)**.<sup>162</sup> Cette réaction est très rapide si bien que la concentration du brut réactionnel à l'évaporateur rotatif à 40°C suffit à consommer l'intégralité du produit **(22)**. Ce dernier a quand même réussi à être isolé avec un rendement faible de 27% après purification sur gel de silice et en évaporant le solvant à température ambiante (Cy 70% - AcOEt 30%) (figure 84).



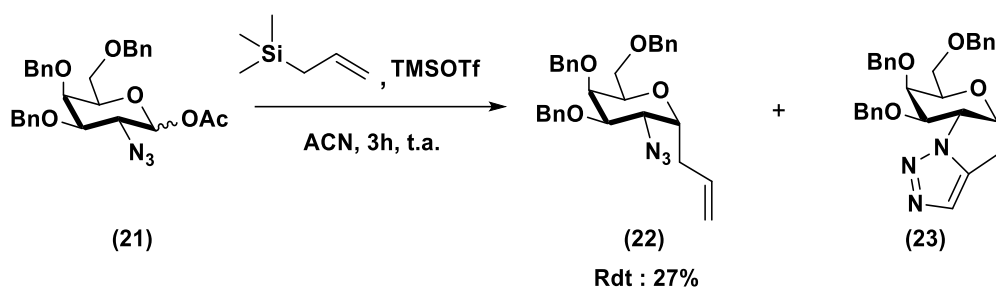


Figure 84 : Réaction de C-glycosylation permettant d'obtenir le composé allylé.

L'étape suivante d'isomérisation de la double liaison du composé **(22)** a été tentée, comme pour la voie A, dans du toluène en présence de catalyseur au palladium durant 20h à 80°C (figure 85). L'analyse du brut réactionnel par RMN du proton, montre une multitude de signaux et ne permet pas d'identifier les produits obtenus. Cependant, l'hypothèse la plus probable est la formation du triazole **(23)** ou/et la formation de plusieurs isomères du composé souhaité. Dans tous les cas, l'utilisation d'un dérivé comportant une fonction azoture et allyle s'avère compromise et il est maintenant nécessaire de trouver une autre alternative.

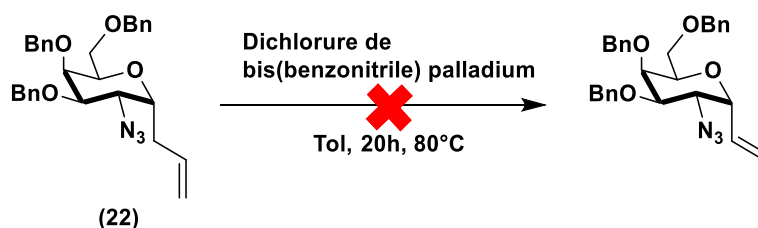


Figure 85 : Isomérisation de la double liaison du composé **(22)**.

## A. Voie C

Le problème majeur des deux premières voies est la fonction présente en position 2 du sucre. Celle-ci, nous l'avons vu, interfère lors de la réaction de C-glycosylation pour introduire le groupement allyle, utilisé comme précurseur du motif CH<sub>2</sub>OH. Une alternative serait d'utiliser, au début de la synthèse, un groupement partant en position C2 que l'on pourrait au moment opportun substituer par un motif précurseur du groupement acétamide, comme par exemple un azoture.

La voie C implique d'avoir un atome d'iode en position 2 du sucre obtenu par réaction d'iodation régio et énantiosélective sur le composé **(19)**. Après la réaction de C-glycosylation et l'isomérisation de la

double liaison, une coupure oxydative permet l'obtention du groument CH<sub>2</sub>OH. Enfin, c'est ici qu'une réaction de substitution nucléophile d'ordre 2 permet d'obtenir le dérivé α-D-C-galactopyranosyl méthanol d'intérêt. Il est important que le sucre iodé soit un dérivé du talose, avec l'iode en position axiale, afin d'avoir un sucre de configuration galactose après la réaction d'azoturation. En effet, cette dernière inverse la stéréochimie du carbone concerné (figure 86).

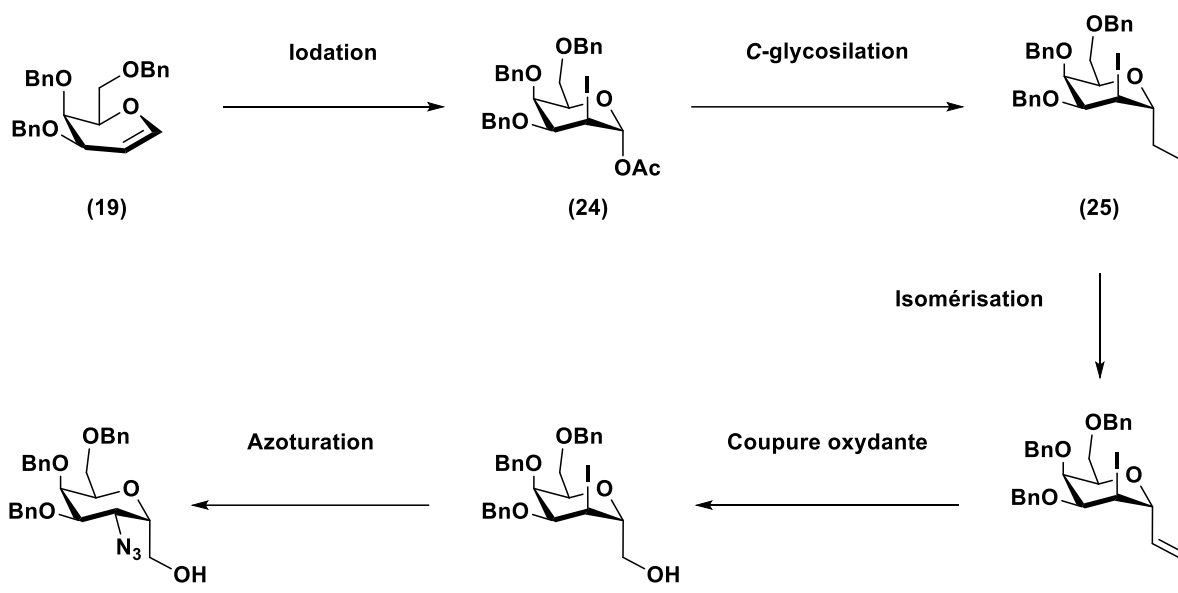


Figure 86 : Obtention de l'α-D-C-galactopyranosyl méthanol d'intérêt selon la voie C.

La première étape de cette synthèse permet de fonctionnaliser la double liaison du composé (19) afin d'introduire un atome d'iode en position 2 du sucre et un acétate en position anomérique, par action du diiode et du (diacétoxyido)benzène (DIB) dans le DCM à température ambiante.<sup>165</sup> Seul l'anomère α du produit (24) est isolé avec un rendement de 78% après purification sur gel de silice (Cy 85% – AcOEt 15%) (figure 87). De plus, le stéréoisomère en C2 (iode en position équatoriale) en configuration anomérique β est également obtenu dans une proportion bien plus modeste (rapport 1/9).

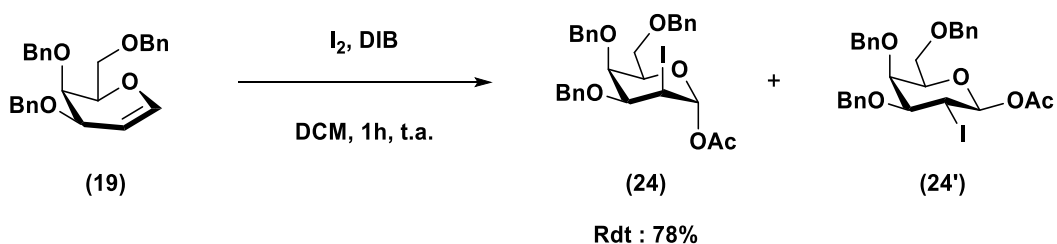


Figure 87 : Iodation du composé (19) permettant d'accéder à la structure (24).<sup>165</sup>

Pour expliquer la régio et énantiosélectivité de cette réaction, le mécanisme est présenté (figure 88).

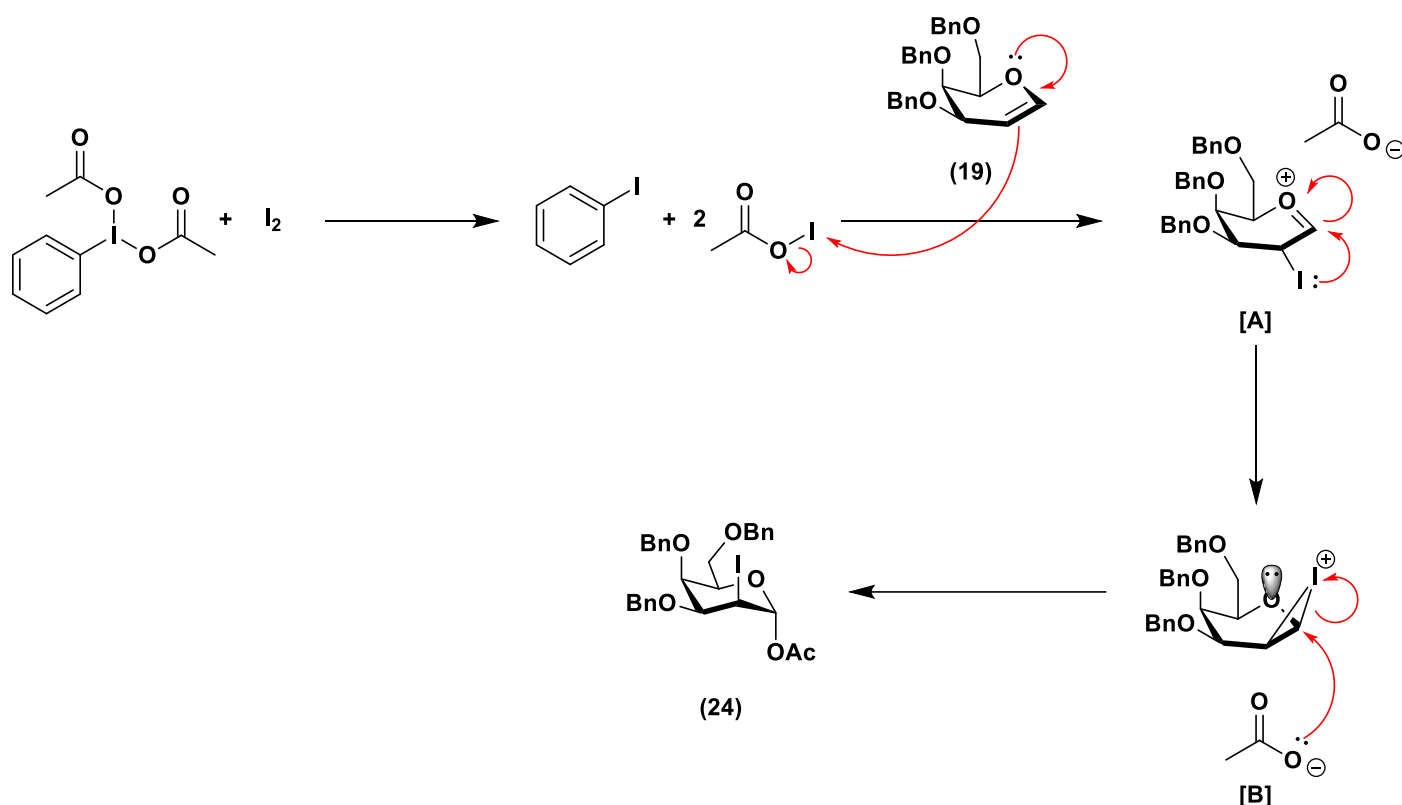


Figure 88 : Mécanisme réactionnel de la formation du produit **(24)** selon l'équipe de Shaw.<sup>165,166</sup>

La première étape est la formation du monoacétate d'iode par interaction entre le DIB et le diiode.<sup>166</sup> Ce réactif réalise l'iodation du galactal par addition électrophile permettant d'obtenir l'intermédiaire **[A]**. Cette réaction est régiosélective, en effet, il est possible d'écrire une forme limite de résonance entre le doublet non-liant de l'oxygène endocyclique et la double liaison du sucre. Cet effet mésomère a pour avantage de contribuer à augmenter le caractère nucléophile du carbone 2, ce qui explique l'addition de l'iode sur cette position. La formation du pont iodonium, intermédiaire **[B]**, s'effectue par l'attaque du doublet non-liant de l'iode sur la position 1 du sucre. Ici, le pont se forme préférentiellement au-dessus du plan car la charge positive est stabilisée par le doublet non-liant en position axiale de l'oxygène endocyclique. En revanche, cette stabilisation ne pourra pas avoir lieu si le pont est vers le bas ce qui explique que le composé **(24')** qui en découle soit ultra minoritaire. Pour finir le pont iodonium est ouvert après attaque par la face opposée de l'acétate présent dans le milieu sur la position anomérique (position la plus électrophile) aboutissant au composé **(24)**.

La réaction suivante est une C-glycosylation permettant d'apporter le groupement allyle en position anomérique. Pour cela, le composé **(24)** est traité par de l'allyltriméthylsilane dans l'acétonitrile à

température ambiante. Le diéthyléthérate de trifluorure de bore est utilisé comme acide Lewis afin d'exalter le caractère électrophile du carbone anomérique. Cette réaction n'a pas conduit à la formation du produit souhaité, le composé **(25)**. Cependant un autre produit **(26)** a été isolé et caractérisé, il est présenté (figure 89). Ce dernier est obtenu avec un rendement de 60% après purification sur colonne de silice (Cy 80% - AcOEt 20%).

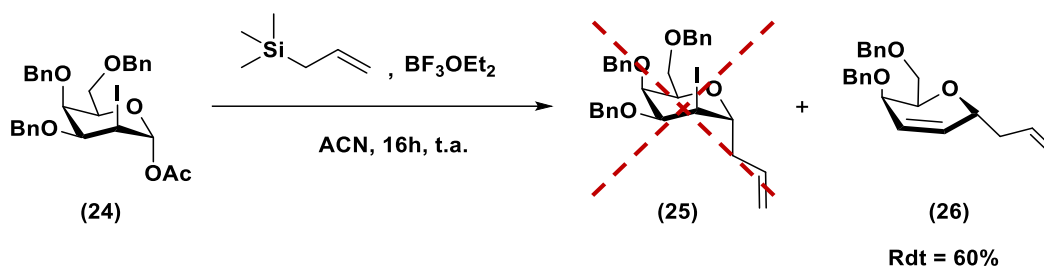


Figure 89 : Synthèse du composé **(25)** à partir de **(24)** par une réaction de C-glycosylation.

Le mécanisme supposé de formation de ce composé est inspiré de la réaction de Ferrier décrit par le chercheur Robert J. Ferrier (figure 90).<sup>167</sup> Il implique une réaction de substitution nucléophile combinée à un déplacement allylique dans un glycal (un glycoside-2,3-insaturé).

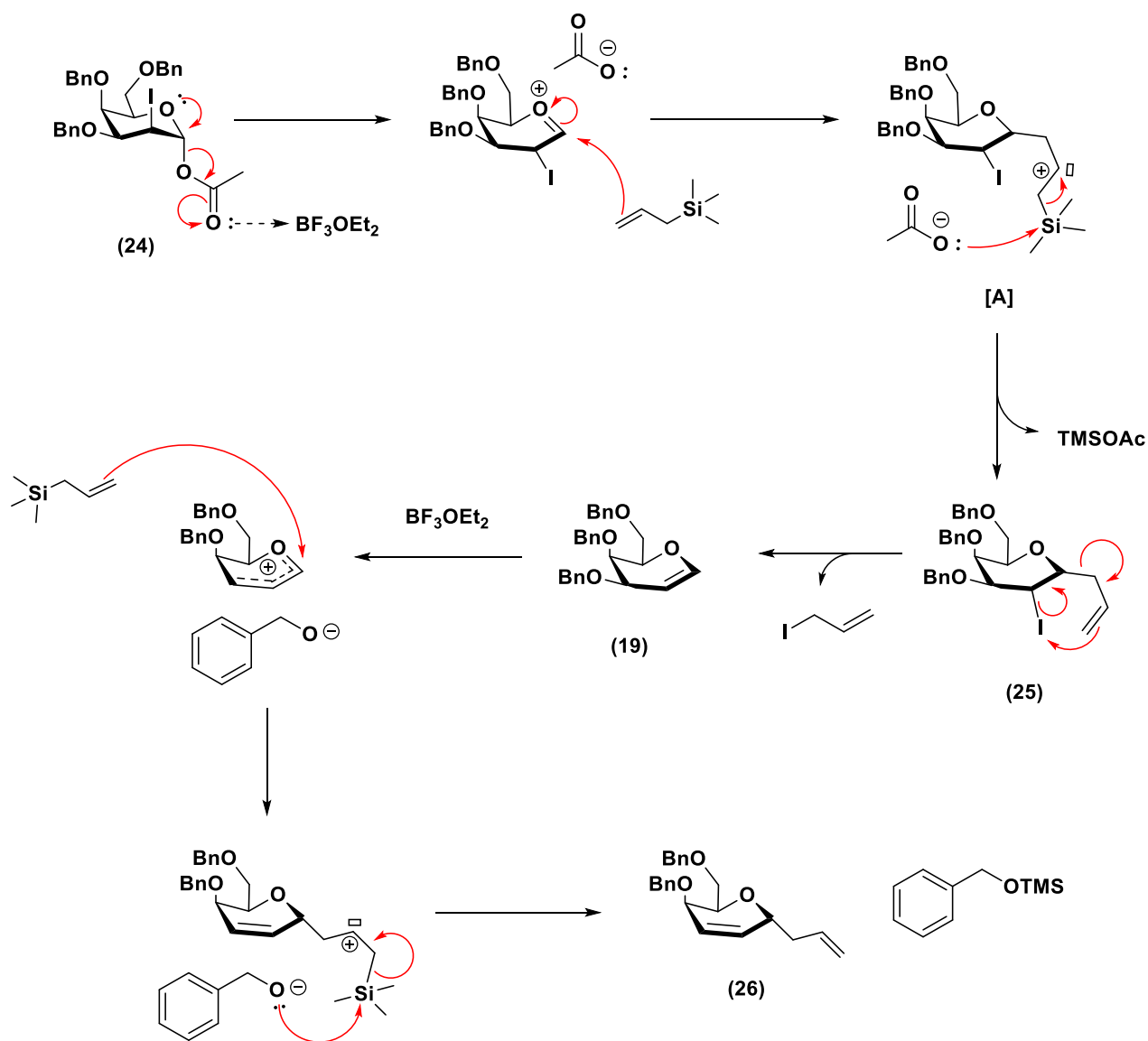


Figure 90 : Mécanisme réactionnel pouvant expliquer l'obtention du composé (26).

Dans un premier temps, le mécanisme débute par la formation de l'intermédiaire de glycosylation oxcarbénium par action du  $\text{BF}_3\text{OEt}_2$  sur la molécule (24). L'allyltriméthylsilane attaque ensuite la position anomérique du sucre conduisant à l'intermédiaire [A], qui donne le composé (26) désiré par élimination de  $\text{TMSOAc}$ . Cependant, la réaction ne s'arrête pas là et le composé (25) subit un réarrangement sigmatropique, favorisé par un intermédiaire à 6 centres, pour redonner le produit (19). Celui-ci subit alors une réaction de Ferrier grâce au  $\text{BF}_3\text{OEt}_2$  pour permettre d'obtenir le composé (26).

Le composé (25) n'a jamais été isolé malgré les différents essais (changement d'acide de Lewis et de température). Le composé (26) est observé à chaque fois avec des rendements relativement élevés,

ce qui marque l'arrêt de cette voie C. Encore une fois le groupement en position C2 du sucre pose un problème dans le déroulement de la synthèse.

D'autres alternatives ont été imaginées comme utiliser des groupements protecteurs d'amine en position 2 du sucre, comme par exemple un phtalimide, un dibenzyle ou encore un tétrachlorophtalimide. L'idée est de s'affranchir des problèmes liés à l'azoture et de pouvoir facilement déprotéger l'amine pour obtenir par la suite la fonction acétamide en position 2. Cependant, d'après la bibliographie, du fait de leur encombrement stérique, les groupements protecteurs benzyle bloquent la face  $\alpha$  du sucre ce qui conduit uniquement à des sucres de configuration  $\beta$ .<sup>158,168-170</sup> Cette configuration ne permettant pas la formation du bicyclic désiré cette possibilité n'a pas été investiguée.

## B. Voie D

La stratégie opérée jusqu'alors était de moduler le groupement présent en position C2 du sucre pour que celui-ci soit compatible avec une réaction de C-glycosylation. Cette dernière est primordiale car elle permet d'ajouter un groupement allyle en position anomérique, précurseur du motif CH<sub>2</sub>OH, essentiel pour la formation du bicyclic désiré. Malheureusement, les 3 premières voies de synthèse n'ont pas permis d'arriver au dérivé  $\alpha$ -D-C-galactopyranosyl méthanol d'intérêt. Pour cette voie de synthèse D nous allons introduire, selon une réaction de C-glycosylation, un groupement cyano en position anomérique du sucre.

Ainsi, la synthèse de la voie D débute par une protection des hydroxyles du D-galactal, suivi d'une réaction d'azidohydratation. Une réaction de C-glycosylation permet d'avoir en position anomérique un groupement cyano précurseur d'une fonction aldéhyde qui sera réduite en alcool pour former le composé recherché (**figure 91**).

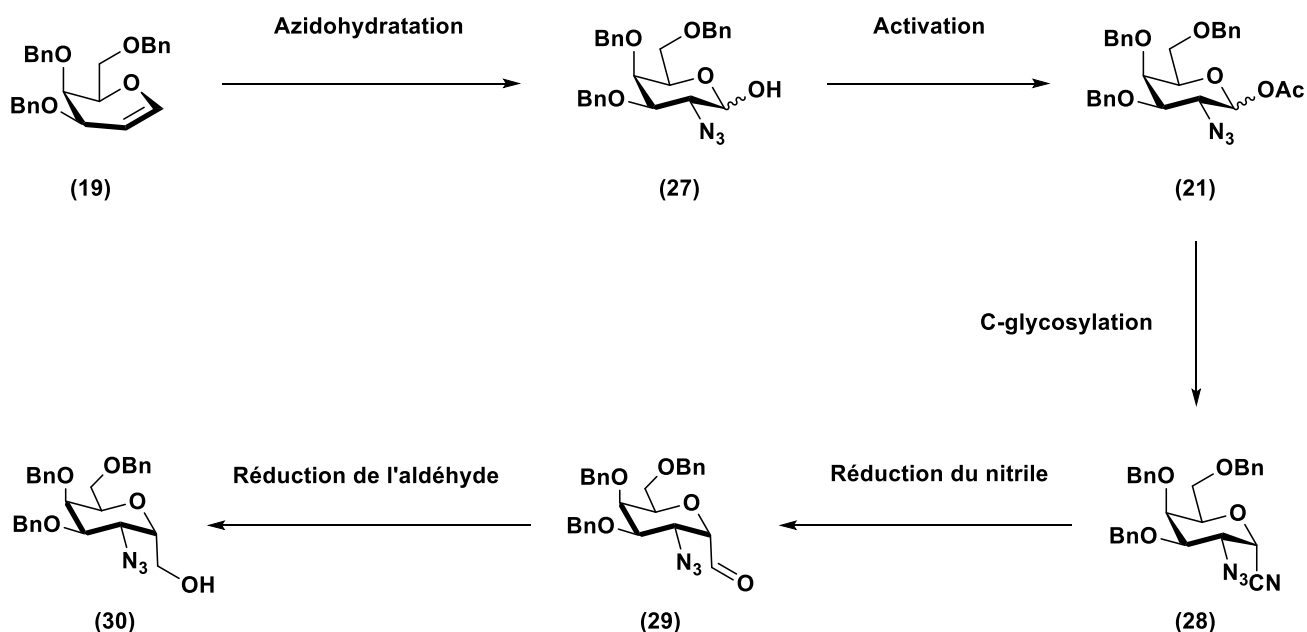


Figure 91 : Obtention de l' $\alpha$ -D-C-galactopyranosyl méthanol d'intérêt selon la voie D.

Pour cette synthèse nous partons du composé **(19)** précédemment préparé qui est mis en réaction avec de l'azoture de triméthylsilyle, du bis(trifluoroacétoxy)iodobenzène (PIFA), de la (2,2,6,6-tétraméthylpipéridin-1-yl)oxy (TEMPO), de l'hydrogènesulfate de tétrabutylammonium et de l'eau dans du dichlorométhane à 0°C pendant 1h30 pour donner le composé **(27)**. Le produit **(27)** est purifié par chromatographie sur colonne de silice (Cy 85% - AcOEt 15%) avec un rendement de 70% (figure 92). Cette réaction appelée azidohydratation permet de fonctionnaliser la double liaison du galactal **(19)** par un azoture en position C2 et un hydroxyle en position anomérique.<sup>171</sup> Elle est proche de l'azidonitration avec néanmoins la particularité d'être plus facile à mettre en œuvre au laboratoire et de posséder un rendement nettement meilleur.

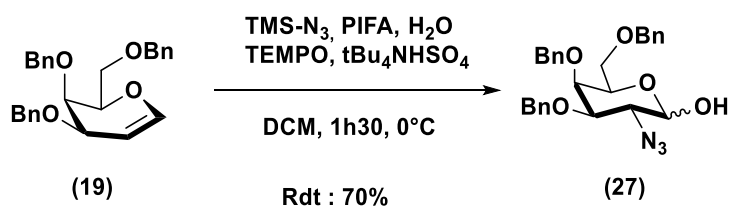


Figure 92 : Réaction d'azidohydratation.

son mécanisme est détaillé (figure 93) selon l'équipe de Vankar.<sup>171</sup>

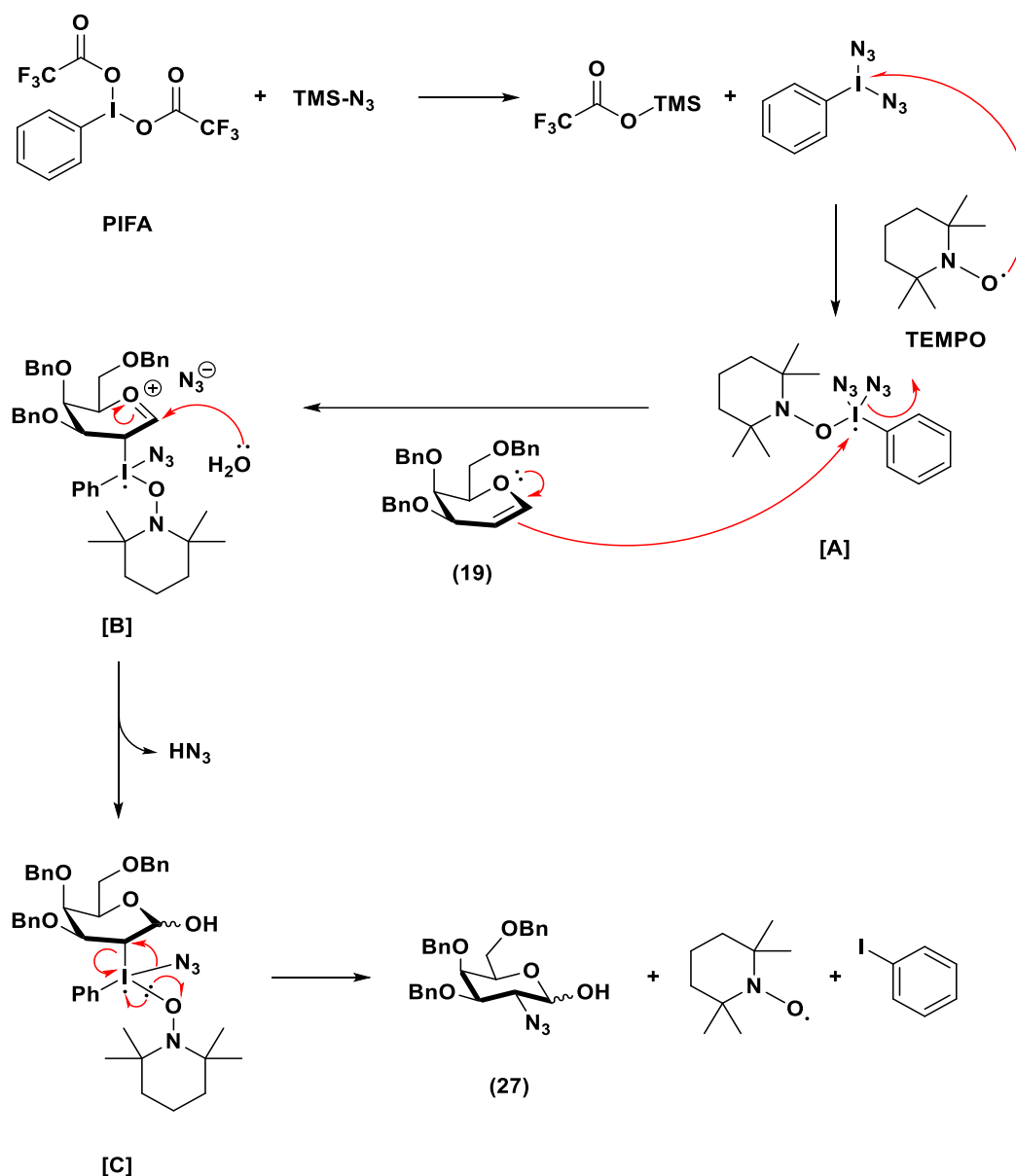


Figure 93 : Mécanisme de la réaction d'azidohydratation.

Le mécanisme commence par la formation de diazidophényliodane par l'action de l'azoture de triméthylsilane sur le PIFA (étapes décrite par l'équipe de Antonchick).<sup>172</sup> L'attaque du TEMPO sur le diazidophényliodane donne l'intermédiaire [A]<sup>173</sup> qui réagit avec le composé (19) par attaque nucléophile de la double liaison. L'oxcarbénium [B] ainsi formé est suffisamment réactif pour réaliser une réaction de glycosylation avec l'eau présente dans le milieu conduisant à l'intermédiaire [C]. Il est à noter que l'attaque de H<sub>2</sub>O peut se faire sur les deux faces du carbone anomérique et c'est pour cela que l'on obtient un mélange d'anomère en proportion 70/30 en faveur de l'anomère  $\alpha$ , ce qui paraît étonnant au vu de l'encombrement de la face  $\alpha$ . Cependant, cette proportion peut facilement être expliquée par l'effet de mutarotation possible du carbone anomérique grâce à la présence d'H<sub>2</sub>O. Enfin



un réarrangement de **[C]** permet de libérer de l'iodobenzène, le composé **(27)** et aussi le tempo qui servira dans un nouveau cycle réactionnel.

Pour la réaction de C-glycosylation, le sucre **(27)** est activé par un groupement acétate en position anomérique par action d'acide acétique et de triméthylamine à 100°C pendant 1h. Le produit initial étant un mélange d'anomères  $\alpha$  et  $\beta$ , le produit **(21)** obtenu est également un mélange de deux anomères. Celui-ci est isolé par purification sur gel de silice (Cy 90% - AcOEt 10%) avec un rendement de 88% (figure 94).

Le sucre **(21)** est mis en réaction avec du TMS-CN dans le dichlorométhane à température ambiante durant 3h en présence de TMSOTf pour la C-glycosylation. Après purification sur gel de silice (Cy 70% - EtOAc 30%) l'anomère  $\alpha$  du produit **(28)** est obtenu avec un rendement de 43%. L'anomère  $\beta$  est également isolé dans une proportion plus faible (15% de rendement).

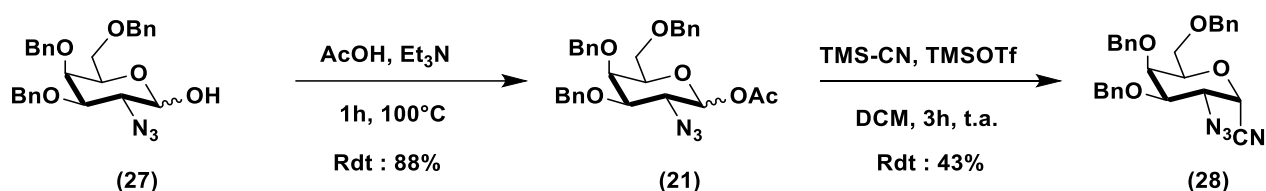


Figure 94 : Activation du sucre par un groupement acétate et réaction de C-glycosylation.

L'étape suivante consiste à hydrolyser la fonction cyano en acide carboxylique par action d'une solution aqueuse d'acide chlorhydrique dans un mélange dioxane/H<sub>2</sub>O en proportion 10/1. Plusieurs conditions de réaction ont été testées notamment en faisant varier la concentration en acide chlorhydrique (1, 4 et 12 mol.L<sup>-1</sup>) et la température (25 et 100°C) mais dans tous les cas nous n'obtenons pas le produit attendu et récupérons quasiment que le réactif de départ. Pour la condition la plus drastique (12 mol.L<sup>-1</sup> à 80-100°C) une dégradation complète du sucre **(28)** est observée. Nous avons également tenté une hydrolyse basique mais sans plus de succès (NaOH à 4 mol.L<sup>-1</sup>) (figure 95). A l'heure actuelle, nous n'avons pas d'explication de l'absence d'hydrolyse du groupement cyano en acide carboxylique.

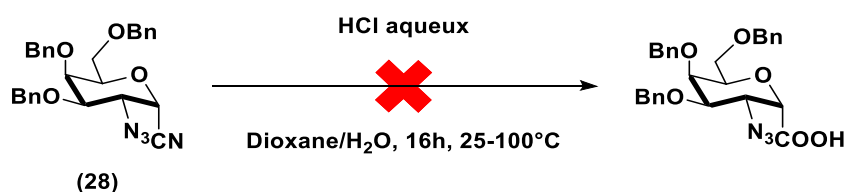


Figure 95 : Hydrolyse de la fonction nitrile du composé **(28)** pour donner l'acide carboxylique correspondant.

Pour pallier le problème de l'hydrolyse, une réaction de réduction du groupement cyano en aldéhyde est envisagée. Cette dernière fait intervenir l'hydrure de diisobutylaluminium (DIBAL-H) comme réducteur doux afin de former *in situ* l'imine correspondante qui est par la suite hydrolysée en aldéhyde lors du traitement acide de la réaction. Le composé **(28)** est mis en réaction avec 1,8 équivalents de DIBAL-H dans le dichlorométhane à -78°C pendant 2h. L'aldéhyde **(29)**, peu stable, est directement réduit en alcool **(30)** par action de tétrahydruroborate de sodium à -20°C dans le méthanol durant 1h. Après purification par chromatographie sur gel de silice (Cy 80% - AcOEt 20%) nous isolons le composé **(30)** avec un rendement faible de 10% sur les deux étapes (figure 96).

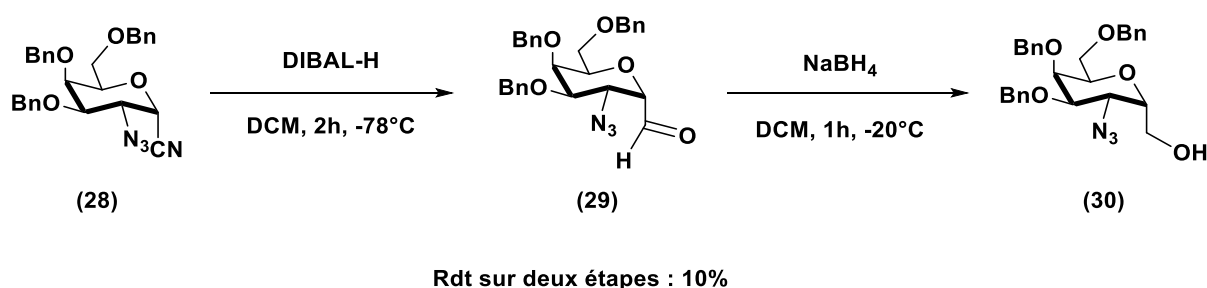


Figure 96 : Synthèse de l'alcool **(30)** à partir du dérivé cyano **(28)**.

Ce faible rendement s'explique notamment par un taux de conversion de 60 à 80% du composé **(29)** et la formation d'un produit secondaire **(29')** résultant d'une réaction d'élimination de l'azoture (figure 97). De plus, lors de la deuxième étape de la synthèse, **(29)** est également réduit en alcool correspondant qui est difficilement séparable du produit **(30)**. Deux purifications par colonne de silice sont nécessaires ce qui diminue encore le rendement global de la réaction.

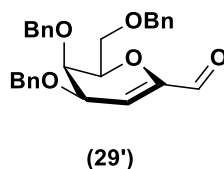


Figure 97 : Produit d'élimination **(29')** obtenu à la suite de la réduction avec le DIBAL.

Ainsi, contrairement aux voies A, B et C, l'objectif que nous nous étions fixés a été atteint à savoir la synthèse de l' $\alpha$ -D-C-galactopyranosyl méthanol **(30)**. Néanmoins, devant les difficultés de la synthèse,

notamment les dernières étapes, cette stratégie n'est pas viable pour une montée en échelle. Toutefois nous avons voulu tester la réaction radicalaire de 1,5-HAT à partir de **(28)** afin d'obtenir le bicycle cible **(31)**. Pour cela, le substrat **(30)** est dissous dans le DCM à température ambiante sous irradiation à 480 nm en présence d'iodosylbenzène et de diiode. Le composé **(31)** est isolé après purification sur gel de silice (Cy 90% - AcOEt 10%) avec un rendement de 41% (figure 98).

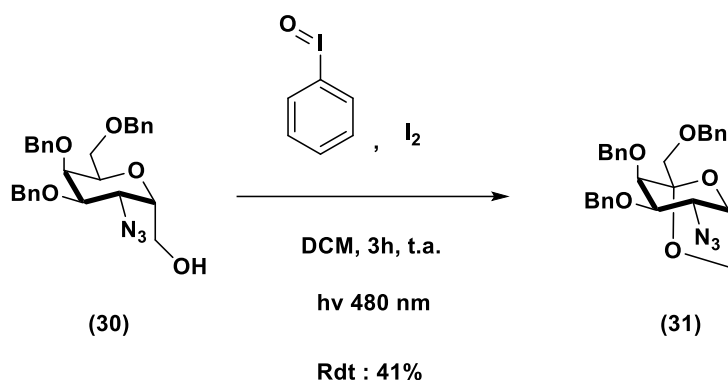


Figure 98 : Réaction de 1,5 HAT permettant d'obtenir le bicycle d'intérêt **(31)** à partir de l'alcool **(30)**.

En conclusion, nous avons obtenu le motif  $\alpha$ -D-C-galactopyranosyl méthanol **(30)** en 7 étapes avec un rendement global de 0,01%. Nous avons également réalisé l'étape de cyclisation radicalaire pour accéder au bicycle **(31)**. Malgré cela, les nombreux échecs ainsi que certains rendements faibles et le nombre d'étapes restant pour atteindre la molécule cible **(PMS-02bis)** nous ont incité à stopper sa synthèse.

C'est pourquoi nous allons maintenant nous intéresser à la synthèse des ligands **(PMS-03)** et **(PMS-04)**.

## 5. Synthèse des ligands PMS-03 et PMS-04

Les deux ligands (**PMS-03**) et (**PMS-04**) sont en quelque sorte un mélange des molécules (**PMS-01**) et (**PMS-02**). En effet, l'idée a été de conserver le motif GalNAc du ligand (**PMS-01**) mais avec en position 6 du sucre, comme pour le ligand (**PMS-02**), la chaîne triéthylène glycol et la fonction d'accroche. L'intérêt de cette stratégie repose sur le fait que la position 6 du sucre est plus disponible car elle se trouve à l'extérieur de la poche lectinique de l'ASGPr. De plus, les études de modélisation moléculaire ont montré qu'une telle stratégie peut s'avérer payante, avec comme exemple les composés (**PA2**), (**PA3**) ou (**PA26**) (figure 99) qui présentent un score de docking élevé.

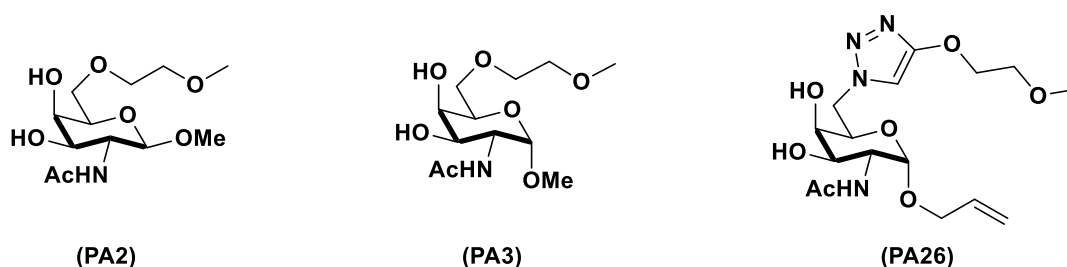


Figure 99 : Composés isolés du docking moléculaire.

Le composé (**PA26**) a la particularité de posséder un groupement allyle en position anomérique, ce qui d'un point de vue de la synthèse est compliqué à obtenir (cf voie de synthèse du ligand **PMS-02**) c'est pourquoi il a été écarté des possibilités.

Pour la synthèse de (**PMS-03**) et (**PMS-04**), nous nous sommes donc inspiré des composés (**PA2**) et (**PA3**) qui possèdent un groupement méthoxy en position anomérique. Le groupement méthoxy possède l'avantage d'être peu volumineux ce qui permettra de limiter au maximum la formation possible d'interactions et/ou d'encombrements dans la poche lectinique. (**PMS-03**) aura le groupement méthoxy anomérique en position  $\beta$  et (**PMS-04**) en position  $\alpha$  (figure 100).

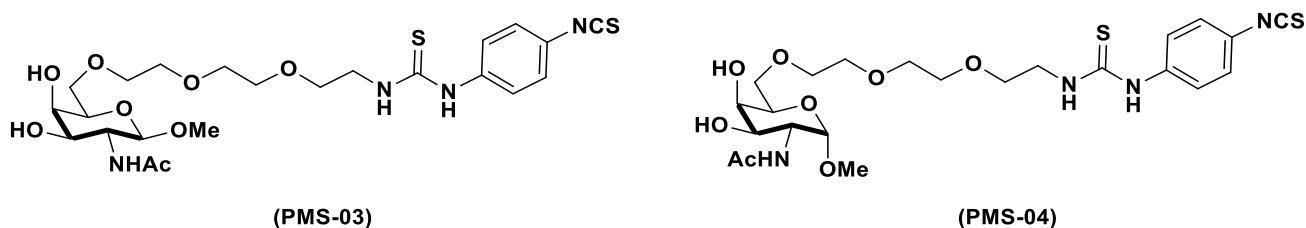


Figure 100 : Structure des ligands (**PMS-03**) et (**PMS-04**).

L'avantage de synthétiser les deux stéréoisomères (**PMS-03**) et (**PMS-04**) est de pouvoir regarder l'influence de la stéréochimie de la position anomérique sur l'affinité pour les ASGPr. En effet, les deux configurations pourraient avoir une constante d'affinité différente pour le récepteur. De plus, il est possible de comparer ces ligands à (**PMS-01**) et (**PMS-02**) qui possèdent également deux centres anomériques différents en plus de l'emplacement de la chaîne triéthylène glycol.

La synthèse des ligands (**PMS-03**) et (**PMS-04**) étant très similaire, nous décrivons dans ce manuscrit uniquement celle de la molécule (**PMS-03**) qui est obtenue, après 9 étapes, avec un rendement global de 0,05%. Comme le montre la (figure 101), la synthèse débute par un enchainement de réactions de protection des hydroxyles, activation du sucre et C-glycosylation pour obtenir l'intermédiaire (**32**) avec un méthoxy en position anomérique de configuration  $\beta$ . Puis, un jeu de déprotection et protection sélective permettra d'obtenir l'alcool primaire (**34**) qui pourra par la suite être fonctionnalisé par l'espacer triéthylène glycol. L'azoture du produit (**35**) sera réduit en amine correspondante qui permettra après déprotection des hydroxyles l'obtention du ligands (**PMS-03**) après réaction sur le bitoscanate.

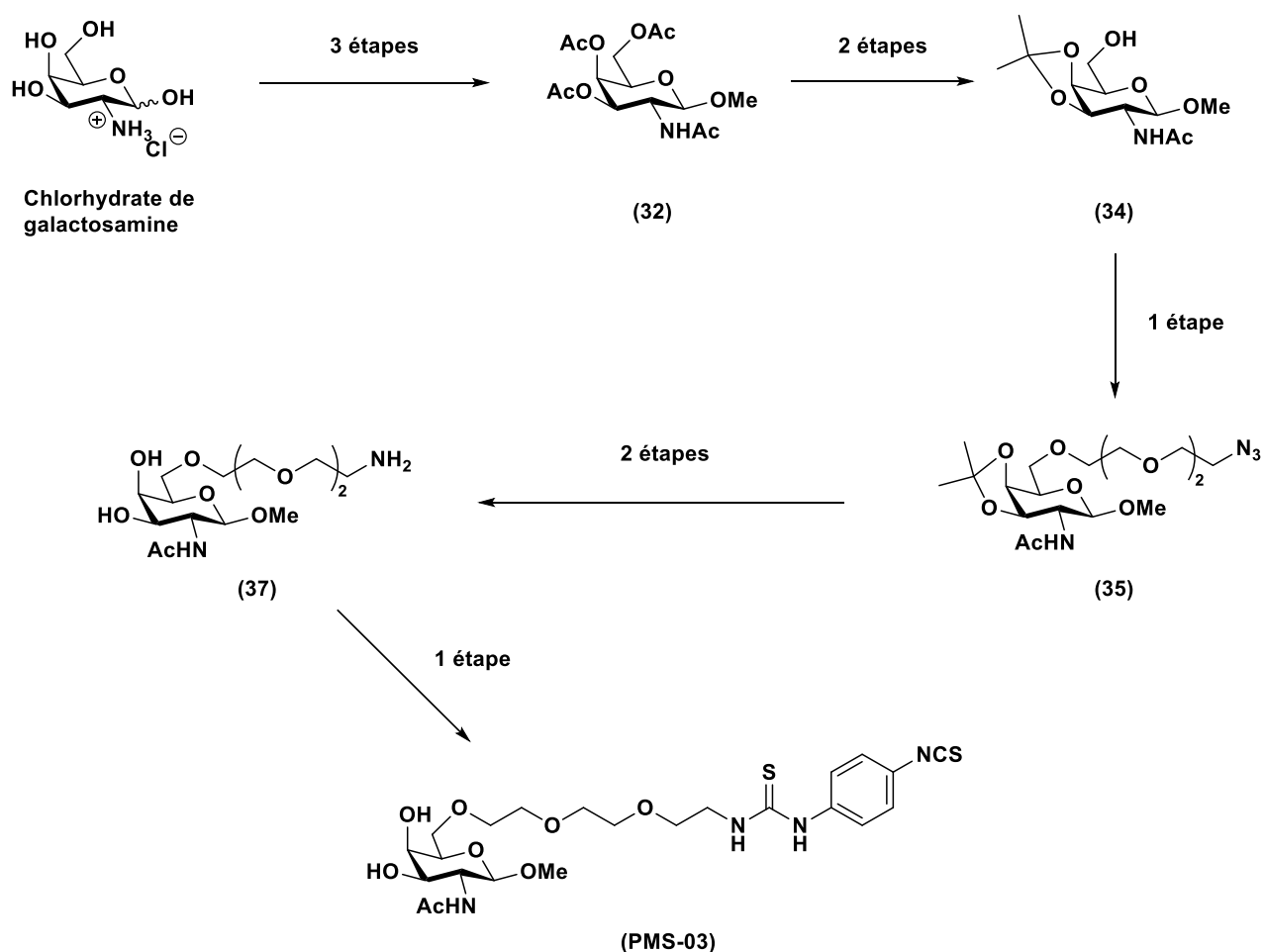


Figure 101 : Synthèse du ligand (**PMS-03**) à partir du chlorhydrate de galactosamine.

La première étape de cette synthèse est une réaction d'acétylation des hydroxyles et de l'amine du produit commercial, le chlorure d'ammonium du D-(+)-galactosamine. Cette étape se fait dans un mélange de pyridine et d'anhydride acétique, en condition anhydre, pendant 5h à température ambiante. Le composé **(14)** est obtenu avec un rendement de 72% après cristallisation dans le MeOH où seul l'anomère  $\beta$  cristallise. La deuxième étape est une cyclisation intra moléculaire qui active la position anomérique du sucre pour la glycosylation. La formation du dihydrooxazole **(17)** se fait en présence d'un acide de Lewis, le triflate de triméthylsilyle (TMSOTf), à reflux du dichlorométhane. En fin de réaction, de la triéthylamine est ajoutée afin de neutraliser le milieu. L'intermédiaire **(17)** est obtenu après lavage sans purification, avec un rendement brut de 95% (figure 102).

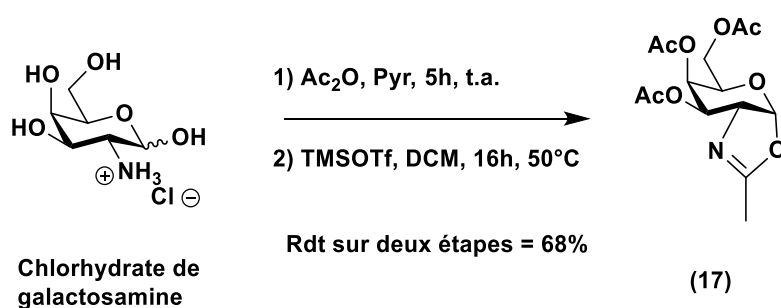


Figure 102 : Protection et activation de la galactosamine en dihydrooxazole.

La suite de la synthèse est une réaction de glycosylation de l'oxazole **(17)** qui se fait dans le méthanol en présence de  $\text{CuCl}_2$  utilisé comme acide de Lewis afin d'exalter le caractère électrophile du carbone anomérique. Le brut réactionnel est lavé par une solution aqueuse d'EDTA afin de s'affranchir du cuivre puis purifié par chromatographie sur colonne de silice (DCM 95% - MeOH 5%) afin d'obtenir le produit **(32)** avec un rendement de 52% (figure 103).

Cette réaction stéréosélective permet d'isoler uniquement l'anomère  $\beta$  étant donné l'encombrement stérique important en position  $\alpha$  dû à la présence du pont oxazole. Des essais préalables réalisés en présence d'acides de Lewis plus conventionnels comme le  $\text{BF}_3\text{OEt}_2$  ou le TMSOTf n'ont rien donné (absence de réaction, dégradation ou hydrolyse du produit de départ).

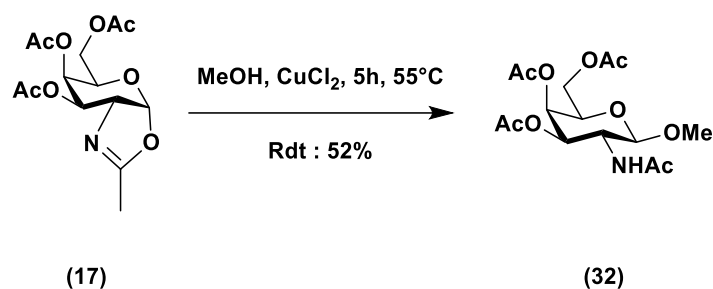
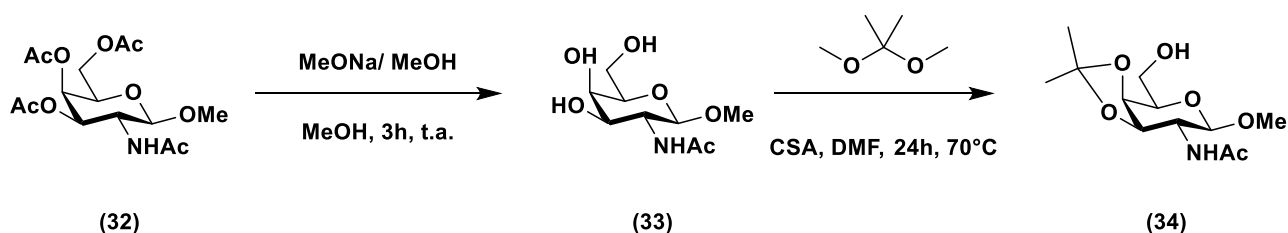


Figure 103 : Réaction de glycosylation permettant d'obtenir le groupement méthoxy en position 8.

De la même manière que pour le ligand **(PMS-02)**, il est nécessaire de venir fonctionnaliser le motif GalNAc avec la chaîne triéthylène glycol en position 6. Pour cela, les hydroxyles du composé **(32)** sont déprotégés par trans-estérification des acétates à l'aide du méthanolate de sodium pour donner le triol **(33)**. Dans un second temps, les hydroxyles en position C3 et C4 sont protégés sous forme d'acétal par l'action du 2,2-diméthoxypropane en condition acide. Le produit **(34)** est finalement isolé après purification sur colonne de silice (AcOEt 90% - MeOH 10%) avec un rendement de 67% (figure 104).



Rdt sur deux étapes : 67%

Figure 104 : Synthèse de l'alcool **(34)** après déprotection et protection sélective du composé **(32)**.

La réaction suivante est une réaction de O-alkylation pour introduire la chaîne triéthylène glycol, fonctionnalisée avec un azoture, sur l'hydroxyle primaire du sucre. Les conditions de réaction utilisées pour l'obtention du composé **(35)** sont présentées (figure 105).

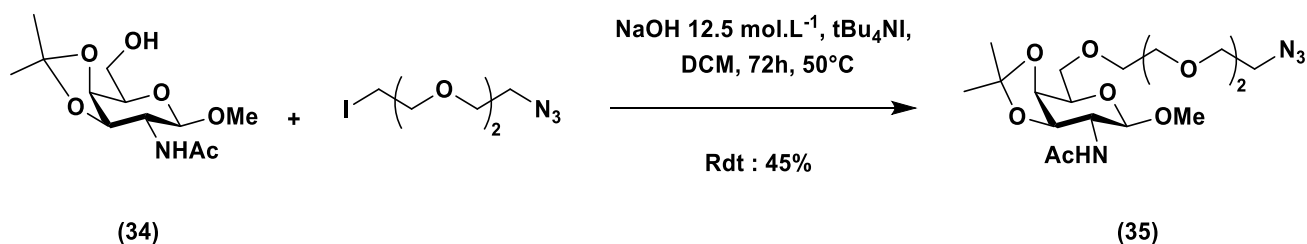


Figure 105 : Synthèse du composé **(35)** à partir de l'alcool **(34)**.

Il a d'abord été testé les conditions de réactions identiques à celles déjà utilisées pour **(PMS-02)** (milieu biphasique, soude très concentré). En utilisant ce même protocole expérimental la réaction n'est pas complète et une proportion importante de produit de départ est récupérée après purification. De plus, un produit secondaire se forme au fur et à mesure de l'avancée de la réaction. Ainsi, après 24h on récupère 25% du bon produit **(35)** et 25% du produit secondaire qui est l'acétal **(36)**, composé des deux unités de sucre **(34)** liées l'une à l'autre par un carbone *via* leur hydroxyle en position 6 (**figure 106**).

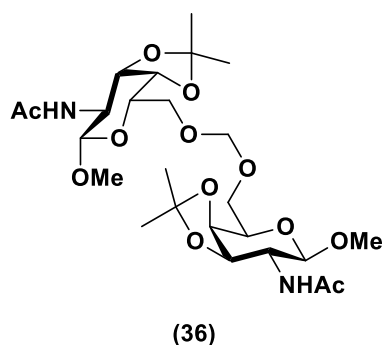


Figure 106: Structure du composé secondaire **(36)** obtenu durant l'étape de O-alkylation.

Le seul moyen d'avoir un CH<sub>2</sub> supplémentaire entre les deux sucres est d'avoir réalisé une alkylation sur le carbone du dichlorométhane et non sur celui portant l'iode de la chaîne triéthylène glycol. Pour obtenir un taux de conversion complet et d'autre part pour s'affranchir du produit secondaire plusieurs conditions opératoires sont testées.

Technique	Solvant	Temps	Base	(34)	(35)	(36)	Commentaire
Batch	CH <sub>2</sub> Cl <sub>2</sub>	24h	NaOH	30%	25%	25%	Condition <b>(PMS-02)</b>
Batch	CH <sub>2</sub> Cl <sub>2</sub>	48h	NaOH	25%	32%	30%	
Batch	CH <sub>2</sub> Cl <sub>2</sub>	72h	NaOH	24%	31%	31%	
Batch	DMF	30h	NaH	>40%	25%	n/a	
Batch	DMF	30h	tBuOK	>40%	21%	n/a	
Batch	DMF	48h	NaOH	85%	0%	0%	dégradation chaîne
Batch	ACN	48h	NaOH	80%	0%	0%	dégradation chaîne
Micro-onde	CH <sub>2</sub> Cl <sub>2</sub> (100 W)	4 x 2h	NaOH	45%	25%	15%	
Micro-onde	CH <sub>2</sub> Cl <sub>2</sub> (150 W)	16h	NaOH	45%	20%	15%	
Batch	CH <sub>2</sub> Cl <sub>2</sub>	48h	NaOH	15%	32%	30%	tBu <sub>4</sub> NI
Batch	CH <sub>2</sub> Cl <sub>2</sub>	72h	NaOH	0%	45%	40%	tBu <sub>4</sub> NI

Tableau 6 : Les différentes conditions opératoires réalisées pour l'optimisation de la réaction de O-alkylation.

Le tableau ci-dessus (**tableau 6**) illustre les différentes conditions opératoires retenues pour optimiser la réaction d'O-alkylation :



- L'utilisation d'une base autre que la soude n'a pas donné de meilleur rendement et ce bien qu'il ne soit pas observé de produits secondaires.
- Le changement de solvant pour passer à un milieu miscible (DMF, ACN) conduit à une forte dégradation de la chaîne sans obtenir le produit **(35)**.
- L'utilisation du micro-onde a donné des résultats très similaires aux synthèses en batch et plusieurs cycles de 45 minutes sont utilisés pour obtenir le produit **(35)** avec un rendement médiocre de 20 à 25%.
- L'éther couronne utilisé précédemment a été remplacé par un catalyseur de transfert de phase, le  $t\text{Bu}_4\text{NI}$ , cela a donné une conversion complète et donc une amélioration significative du rendement de la réaction (45%).

Ces résultats montrent que la meilleure condition est l'utilisation du  $\text{CH}_2\text{Cl}_2$  à  $55^\circ\text{C}$  pendant 72h en présence de soude 12 M et de  $t\text{Bu}_4\text{NI}$ . Malgré la formation importante de produits secondaires, la réaction est totale et permet de s'approcher au mieux de 50% de rendement après purification sur gel de silice (AcOEt 90% - MeOH 10%).

Une fois le produit alkylé **(35)** obtenu, les hydroxyles en position 3 et 4 sont déprotégés en condition acide par l'action d'acide acétique dans l'eau à  $50^\circ\text{C}$  durant 1 heure. Dans un premier temps, nous avons fait cette réaction dans un mélange ternaire AcOH/ $\text{H}_2\text{O}$ /MeOH dans des proportions 1/1/3 à  $70^\circ\text{C}$ . Cependant, ces conditions favorisaient l'attaque du méthanol sur la position anomérique du sucre conduisant à la formation du même produit mais de configuration  $\alpha$ . Il a donc été décidé de s'affranchir du méthanol et de réduire la température pour éviter toute réaction secondaire sur la position anomérique. Le composé déprotégé est ensuite traité par du palladium sur charbon sous pression d'hydrogène dans l'isopropanol afin de réduire l'azoture en amine primaire correspondante. Une simple filtration sur célite donne le composé **(37)** avec un rendement sur les deux étapes de 90% (figure 107).

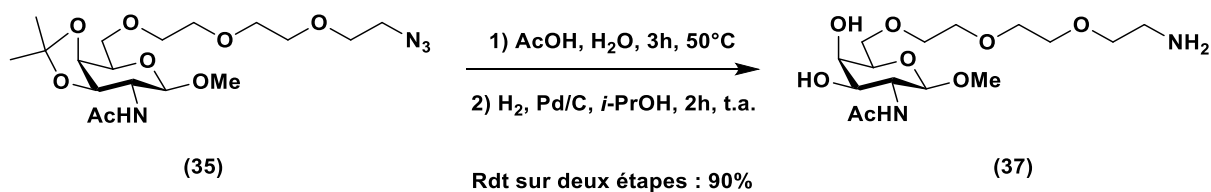


Figure 107 : Déprotection des hydroxyles et réduction de l'azoture.

Enfin, le ligand **(PMS-03)** est obtenu après ajout goutte à goutte d'une solution contenant l'amine **(37)**, solubilisée dans un mélange dioxane/H<sub>2</sub>O, sur une solution contenant cinq équivalents de bitoscanate dissous dans le dioxane. Après deux heures d'agitation à température ambiante, le composé final **(PMS-03)** est obtenu après purification sur colonne à polarité inverse C-18 (ACN 20% - H<sub>2</sub>O 80%) avec un rendement de 55% (figure 108).

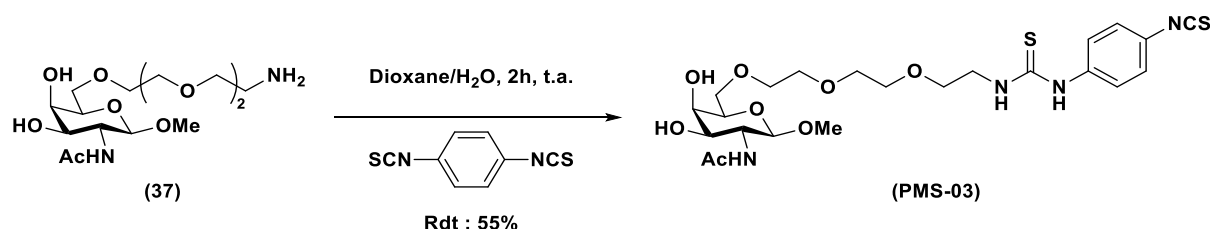


Figure 108 : Synthèse du ligand **(PMS-03)**.

Le ligand **(PMS-04)** est synthétisé de façon similaire à **(PMS-03)**. En effet, le composé **(38)** est commercial et est donc utilisé directement comme produit de départ. Les réactions de protection sélective, O-alkylation, déprotection/réduction et ajout de la fonction de couplage se déroule de la même manière que dans la synthèse de **(PMS-03)** avec des rendements semblables. Le ligand **(PMS-04)** est obtenu après 5 étapes de synthèse avec un rendement total de 0.11% (figure 109).

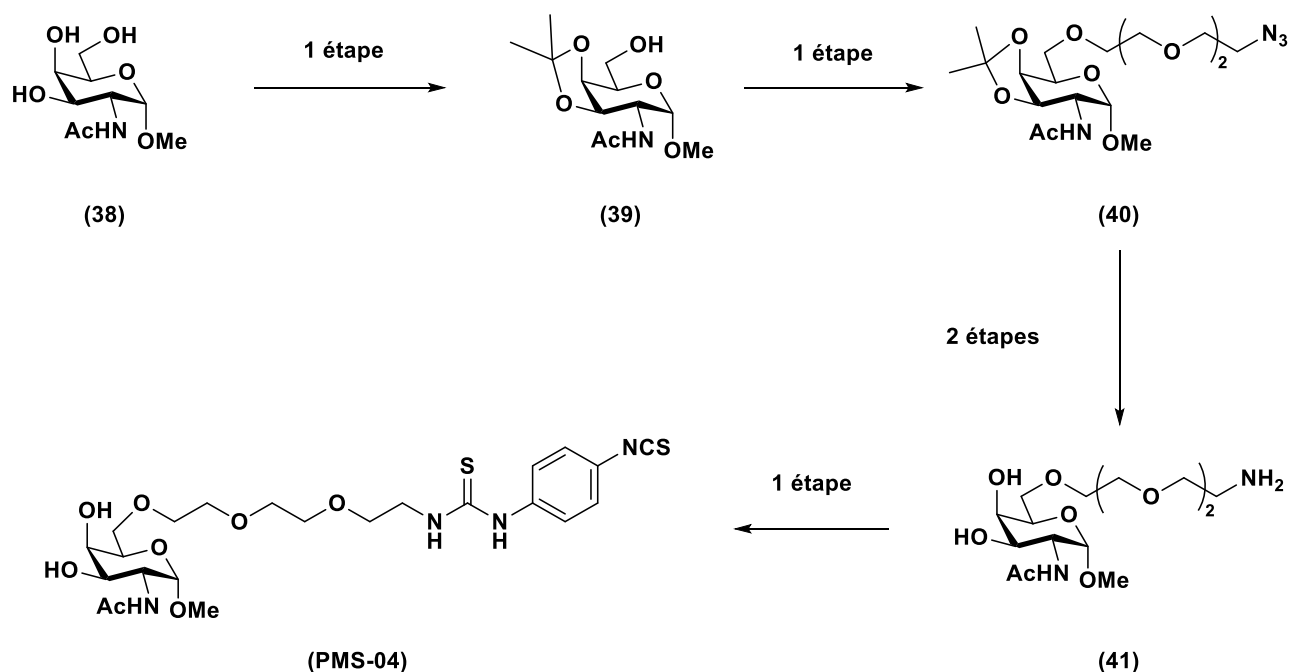


Figure 109 : Synthèse du ligand **(PMS-04)**.

## 6. Conclusion

Au cours de ce chapitre nous avons vu la synthèse de différents ligands qui ont pour vocation d'être bioconjugués à la surface d'un vecteur viral recombinant, l'AAVr. Une fois greffés, ils permettront de guider la particule à travers le milieu biologique vers l'organe d'intérêt. Pour ce faire les ligands possèdent un motif spécifiquement reconnu par le récepteur ASGPr présent à la surface des hépatocytes qui composent le foie.

Les trois ligands synthétisés possèdent trois types de motif de reconnaissance que l'on pourrait schématiser ici comme des clés pour le récepteur ASGPr ainsi qu'une fonction permettant la réaction de bioconjugaison des lysines. Le premier d'entre eux, le ligand **(PMS-02)** possède un motif de reconnaissance récemment décrit dans la littérature pour être 7 fois plus afin que le ligand classiquement utilisé comme référence, la *N*-acétylgalactosamine. Les deux autres ligands, **(PMS-03)** et **(PMS-04)**, ont quant à eux le sucre GalNAc comme motif de reconnaissance, leur particularité est la position de la chaîne possédant la fonction de couplage. Cette dernière est greffée en position 6 du sucre et non en position anomérique comme les ligands classiquement utilisés pour cibler l'ASGPr. Ces ligands sont synthétisés respectivement en 17, 9 et 5 étapes pour des rendements de 0.005, 0,052 et 0.11% témoignant de nombreux échecs et difficultés tout le long de la thèse. Parallèlement, le ligand bicyclique **(PMS-02bis)** n'a jamais pu être préparé, même si l'une des voies de synthèse a toutefois permis d'obtenir un intermédiaire clef, le motif  $\alpha$ -D-C-galactopyranosyl méthanol **(30)** en 7 étapes avec un rendement global de 0,01%. Une réaction de cyclisation radicalaire de type 1,5-HAT a pu être réalisée sur cet intermédiaire aboutissant au composé **(31)**. Cette structure inédite pourrait ouvrir la voie vers une nouvelle série de sucre bicyclique, *via* sa fonction azoture en position 2 permettant la fonctionnalisation d'un grand nombre de substituant. Il sera néanmoins nécessaire d'optimiser la voie de synthèse afin d'obtenir un rendement compatible pour des applications biologiques futures.

Par ailleurs, le criblage d'une chimiothèque de composés par docking moléculaire a été entrepris mais n'a pas donné les résultats escomptés. La raison principale est la présence dans le site de reconnaissance de l'ASGPr d'un cation calcium qu'il a été difficile de correctement modéliser quel que soit le type de docking (normal, semi-flexible et dynamique). En effet, comme toutes les lectines de types-C, l'interaction protéine-ligand est principalement guidée par la coordination entre le ligand et le calcium. L'élaboration d'un champ de force adéquat permettrait vraisemblablement d'être plus pertinent mais les ressources de calcul et de temps nécessaires ne nous ont pas donné l'occasion de le faire, cela constituerait un travail de thèse à part entière.

Pour finir, les trois ligands obtenus feront l'objet d'une étude biologique détaillée dans le chapitre suivant qui sera réalisée au laboratoire TaRGeT INSERM UMR-1089. Cette étude regroupe (i) la bioconjugaison des ligands sur les lysines présentes à la surface des capsides des AAVr, (ii) la caractérisation de ces nouveaux vecteurs et (iii) une étude *in vivo* visant à analyser leur comportement et leur potentiel pertinence pour le transfert de gène dans le foie.



# Chapitre III

*Bioconjugaison,  
caractérisations des vecteurs  
chimiquement modifiés et étude  
in vivo*



## 1. Introduction

Suite à la partie synthèse organique trois nouveaux ligands avec une fonction phénylthiocyanate ont été obtenus. Le premier (**PMS-02**) présente un motif de reconnaissance GalNAc ponté décrit dans la littérature pour avoir une affinité cinq fois plus forte que le GalNAc (actuellement la référence). Ce ligand complexe a été obtenu après 17 étapes de synthèse et un rendement très faible. Les deux autres ligands (**PMS-03** et **PMS-04**) sont des dérivés du GalNAc, l'un est de série alpha et l'autre de série bêta, relié à un espaceur de type triéthylène glycol par la position C6 du sucre afin d'optimiser l'interaction dans la poche lectinique. Enfin, un quatrième ligand (**PMS-01**) a été synthétisé. Il possède un GalNAc fonctionnalisé avec le même espaceur *via* la position anomérique et a déjà été utilisé lors d'une étude précédente.<sup>108</sup> Il nous servira de référence afin de comparer l'efficacité des autres ligands (**figure 110**). Toutes ces molécules possèdent donc un motif présentant une affinité pour les récepteurs ASGP, un espaceur de type triéthylène glycol et une fonction de couplage phénylthiocyanate. Il sera alors possible d'effectuer une réaction de bioconjugaison par addition nucléophile des résidus amino des acides aminés présents à la surface des AAV avec formation d'une fonction thiourée. En plus de ces quatre ligands, un cinquième composé a été ajouté à l'étude. Ce dernier est similaire au ligand (**PMS-01**) mais sans fonction de couplage phénylthiocyanate permettant la bioconjugaison. Il nous servira de contrôle afin de valider qu'il y a bien un couplage chimique entre les différents ligands et l'AAV et non pas une adsorption des molécules à sa surface.



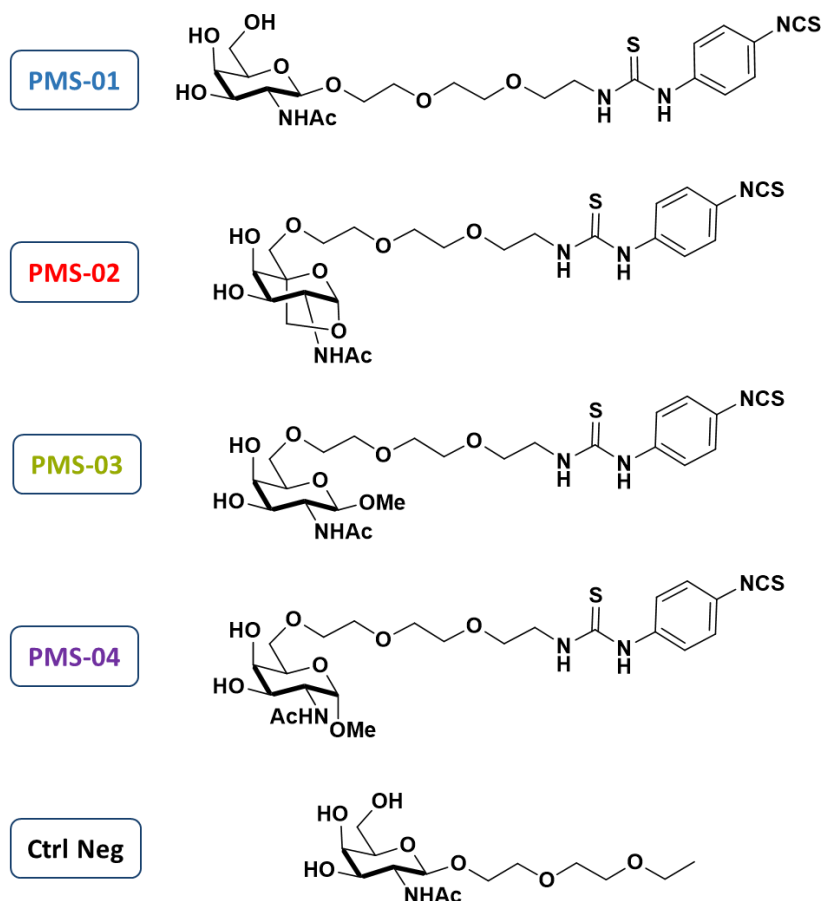


Figure 110 : Structures des différents ligands utilisés dans cette étude.

## 2. Couplage chimique et caractérisations des AAV

### A. Bioconjugaison des différents ligands

Les ligands préparés précédemment ont été couplés à la surface d'AAV2 qui contiennent un promoteur CAG et un gène rapporteur codant pour la protéine GFP.

La procédure de bioconjugaison consiste à incuber une solution d'AAV contenant  $1^{E12}$  vg (vg = viral génome, c'est-à-dire le nombre de particules infectieuses) pendant 4h, à température ambiante, en présence du ligand dans un tampon TBS à pH 9,3 (figure 111). Nous avons choisi de faire la réaction de couplage à ce pH afin d'avoir le résidu amino des lysines sous forme neutre et ainsi optimisé le couplage avec la fonction phényl isothiocyanate.<sup>174</sup> En ce qui concerne les ligands, deux protocoles sont utilisés, l'un avec  $3^{E5}$  équivalents de ligands par AAV appelé « faible concentration » et noté **(f)** et l'autre avec  $3^{E6}$  équivalents appelé « forte concentration » et noté **(F)**.

Afin d'avoir une solution parfaitement limpide, les ligands sont préalablement dissous dans un volume de diméthylsulfoxyde (DMSO) correspondant à 10% du volume final qui est de 1mL.

A la suite de cette étape de bioconjugaison, une faible quantité de Pluronic® est ajoutée (0,001% du volume final) afin de s'affranchir de possibles agglomérations ou adsorptions de l'AAV sur les membranes de la cassette de dialyse. En effet, le Pluronic®, qui est un poloxamère, c'est-à-dire un copolymère formé de polyéthylène glycol et de polypropylène glycol, est utilisé comme surfactant non ionique. Les échantillons sont alors dialysés contre du dPBS, tampon de formulation de l'AAV, contenant également du Pluronic®, afin de repasser à pH 7 et d'éliminer l'excès du réactif ainsi que le DMSO, cette étape est obligatoire car conserver l'AAV2 dans un tampon à pH 9,3 pourrait à long terme dénaturer sa structure quaternaire. Un volume d'environ 1 mL est obtenu à la fin du processus de couplage.

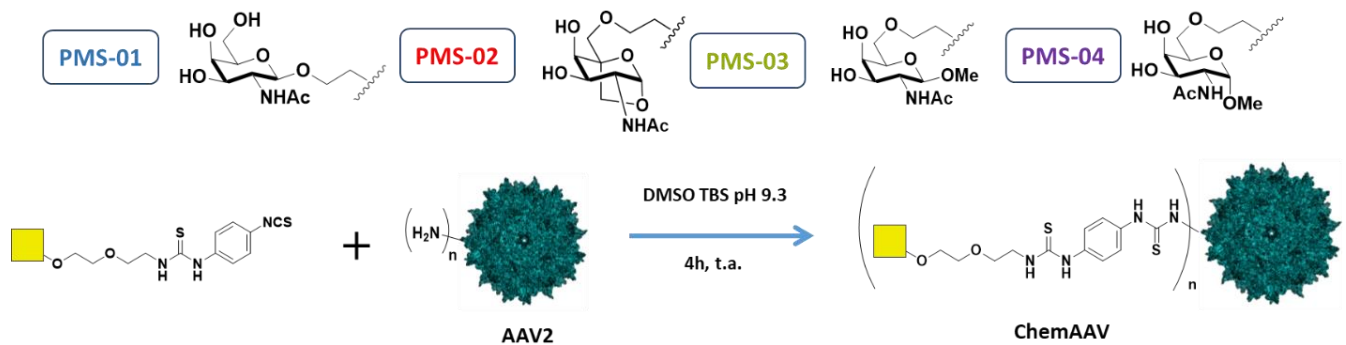


Figure 111 : Réaction de couplage entre les différents ligands et le vecteur AAV2-CAG-GFP.

Suite à la réaction de greffage, il est avant tout nécessaire de savoir si les conditions de couplage, notamment le fort excès de ligand, le DMSO et le pH basique, ne sont pas délétères pour la structure de l'AAV et du transgène. Ainsi une titration est réalisée pour connaître le titre en vg/mL (box 2), c'est-à-dire la quantité de vecteurs obtenue contenant l'intégralité du transgène.

#### Box 2 : Méthode de titration

La titration des AAV par qPCR va permettre de doser la quantité de transgène et ainsi de pouvoir calculer la quantité de vecteur contenant un transgène présents dans l'échantillon. Le titre obtenu s'exprime en vg/mL (vector génome/mL), où le terme vg représente le transgène encapsidé dans un vecteur. Cela représente donc par extension le nombre de capsides d'AAV, dites « pleines » car elles contiennent un transgène, par millilitre. La titration est une analyse de PCR quantitative (qPCR) qui permet de déterminer la quantité originelle d'ADN dans un échantillon en l'amplifiant. Avant d'être titré, l'échantillon va subir plusieurs traitements car les productions d'AAV peuvent contenir des contaminants comme des peptides, des protéines ou diverses molécules d'acides nucléiques. La première étape est un traitement par une enzyme, la DNase, qui détruit tous les résidus d'acides nucléiques contenus sur la surface de l'échantillon sans toucher au transgène qui est protégé par la capside. Ensuite, la DNase est inactivée et une protéase détruit les contaminants protéiques et désassemble la capside. Le mélange est alors purifié sur colonne d'affinité afin d'obtenir le transgène qui est dosé par qPCR. Cette technique visualise la quantité d'ADN amplifiée grâce à la présence d'une sonde fluorescente et permet ainsi de remonter à la quantité initiale et donner un titre en vg/mL.

Le tableau ci-dessous (**tableau 7**) présente le titre des différents vecteurs modifiés. Ces titres sont de l'ordre de  $10^{12}$  vg/mL et traduisent que la réaction de bioconjugaison n'a pas d'impact délétère sur le titre en AAV de l'échantillon. Il est important de noter ici que l'incertitude de la technique de titration peut amener à obtenir des titres après couplage plus élevés qu'avant la réaction. **Erreur !**

**Liaison incorrecte.**

Une fois les titres obtenus, différentes analyses biochimiques vont être indispensables.

## B. Caractérisations des nouveaux vecteurs modifiés chimiquement

Dans un premier temps, il est nécessaire de vérifier que la réaction de couplage, ou les différentes dialyses, n'ont pas entraîné une modification conformationnelle ou une dénaturation de la capside de l'AAV qui pourraient induire une perte d'infectiosité. Afin de contrôler cela nous avons effectué un Dot-blot (**Box 3, figure 112**).

### Box 3 : Le Dot-blot

Un Dot-blot (ou Slot blot) est une technique de biologie moléculaire utilisée pour détecter des protéines. Une fois les échantillons de protéines déposés sur une membrane de nitrocellulose, elle est placée dans une cuve et incubée séquentiellement avec un tampon de blocage (composé de gélatine ou de lait), une solution d'anticorps primaires (selon la nature des protéines voulant être révélées) et une solution d'anticorps secondaires couplés, classiquement, à une enzyme. Entre chaque étape la membrane est rincée à l'aide d'un tampon spécifique. Pour finir, la membrane est placée sur un film transparent et incubée avec une solution contenant le substrat de l'enzyme. Si la protéine d'intérêt est présente alors l'enzyme pourra convertir le substrat en produit chimioluminescent qui sera alors révélé en chambre noire à l'aide d'un film photographique.

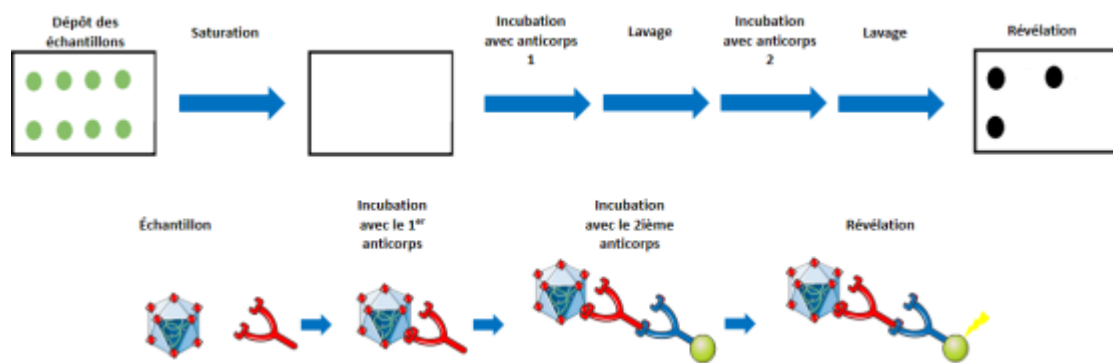


Figure 112 : Présentation de la technique de Dot-blot.

Les échantillons sont déposés sur une membrane de nitrocellulose qui est ensuite saturée puis incubée une nuit en présence d'un anticorps primaire (A20) qui reconnaît la capsidie assemblée du vecteur. Un second anticorps, couplé à une peroxydase de raifort (HRP pour horseradish peroxidase), est ajouté et s'associe au premier avant l'étape de révélation. L'enzyme permet, en présence de son substrat le peroxyde d'hydrogène, l'émission de lumière qui est captée par un film photographique. La présence d'un point (Dot) noir sur le film signifie alors que l'anticorps primaire reconnaît la capsidie du vecteur et donc que cette dernière est toujours dans sa conformation native.

Dans notre cas, les Dot positifs obtenus pour tous les échantillons, après l'utilisation de l'anticorps A20, montrent que l'étape de modification chimique du vecteur ne déstructure ou ne dénature pas la conformation 3D de la capsidie (figure 113).

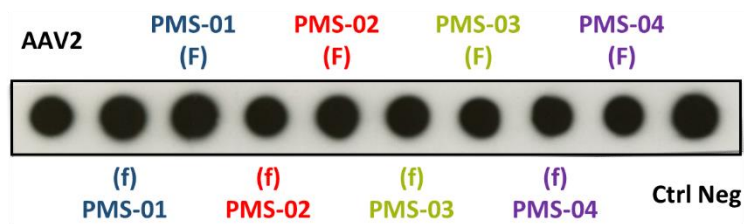


Figure 113 : Dot-blot réalisé avec l'anticorps A20 pour la détection des capsides assemblées.

Dans un deuxième temps, il est nécessaire de valider le fait que les différents ligands soient bien greffés de façon covalente à la surface du vecteur. Pour cela, un nouveau Dot-blot est réalisé en utilisant la lectine de soja SBA (Soybean agglutinin en anglais) qui reconnaît le sucre GalNAc. Cette lectine est couplée à une sonde fluorescente, FITC, qui s'associe à un anticorps secondaire anti-FITC pour la détection. L'alcaline phosphatase présente sur cet anticorps secondaire sera transformée par son substrat, au même titre que la HRP, entraînant l'émission de lumière et permettant la révélation sur le film photographique.

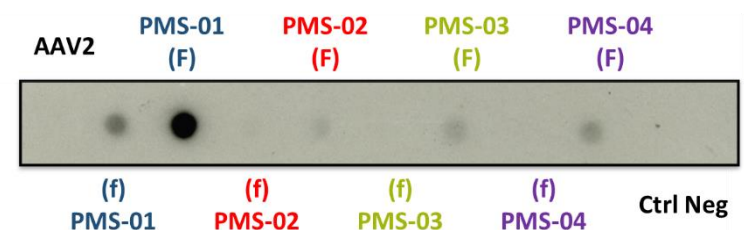


Figure 114 : Dot-blot réalisé avec la lectine Soybean « SBA » reconnaissant le sucre GalNAc.

Le Dot-blot, présenté sur la (figure 114), montre clairement que l'AAV2 ainsi que l'AAV2 incubé avec le ligand contrôle (sans fonction de couplage) ne sont pas reconnus par la lectine Soybean, prouvant qu'il n'y a pas de détection non spécifique ou d'adsorption sur la surface de la capside de l'AAV2. Les échantillons d'AAV modifiés avec (PMS-01) sont quant à eux positifs pour (f) et (F) avec une intensité plus importante pour la condition (F). Ce résultat confirme un couplage covalent de cette molécule sur les amines primaires de la capside de l'AAV2, mais également qu'il est possible de moduler le nombre de ligands greffés. Concernant les échantillons d'AAV modifiés avec (PMS-02, 03 et 04), l'intensité du Dot-blot est très faible pour les (F) et inexistante pour les (f). Ces résultats peuvent s'expliquer par une faible reconnaissance de sucres non-naturels, présents sur ces trois composés, par la lectine Soybean spécifique du GalNAc. En effet, la présence en position 6 de la chaîne triéthylène glycol pourrait provoquer un encombrement stérique voir un blocage au niveau de la poche lectinique. Pour (PMS-

**02)**, même s'il s'agit d'un dérivé du GalNAc, sa conformation et sa structure sont totalement différentes, ce résultat est donc plutôt attendu.

Plusieurs lectines sont décrites dans la littérature pour interagir avec les résidus GalNAc.<sup>175</sup> Ainsi, nous avons également effectué un Dot-blot avec la lectine « Wisteria floribunda » (WFA), aussi conjuguée à FITC (**figure 115**). Comme précédemment, l'AAV2 et l'AAV2 incubé avec le ligand contrôle sont toujours négatifs. Les échantillons d'AAV modifiés avec les ligands (**PMS-01**), (**PMS-02**), (**PMS-03**) et (**PMS-04**) sont quant à eux positifs pour les deux concentrations utilisées avec une intensité plus forte pour les conditions (**F**), validant une nouvelle fois la présence de ces ligands ainsi que la possibilité de moduler leur nombre lors de l'étape de bioconjugaison. Ce Dot valide aussi l'hypothèse que la fonctionnalisation en position 6 interfère dans la reconnaissance des ligands (**PMS-03**) et (**PMS-04**) avec la lectine Soybean. Concernant l'**AAV2-PMS-02** l'intensité du Dot est beaucoup plus faible. Ce sucre très différent d'un point de vue conformationnel comparé à un sucre naturel bien plus flexible, pourrait être moins bien reconnu par la lectine WL et donc expliquer cette observation.

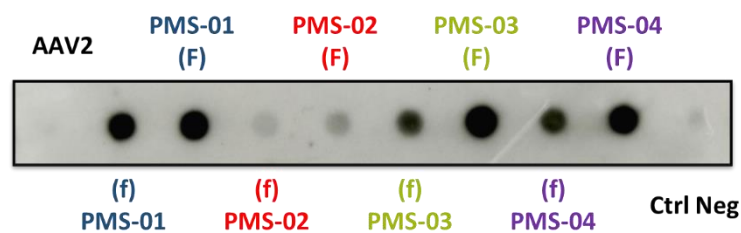
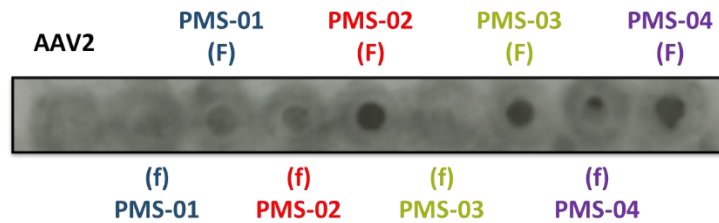


Figure 115 : Dot-blot utilisant la lectine « WFA » spécifique des dérivés du GalNAc.

Dans l'étude publiée par l'équipe de M.G. Finn<sup>143</sup>, il est fait mention de la lectine du domaine de reconnaissance des sucres de l'ASGPr pour l'évaluation des interactions entre le GalNAc ponté et ce récepteur. Cette lectine étant commerciale nous l'avons également utilisée afin de comparer l'interaction possible des différents ligands couplés sur la capsid de l'AAV2. Après dépôt des différents échantillons, saturation de la membrane, incubation avec l'ASGPr, puis avec l'anticorps anti-ASGPr et enfin avec un anticorps anti-souris-HRP, un nouveau Dot-blot a ainsi été obtenu (**figure 116**). Comme précédemment l'AAV2 natif et le contrôle sont bien négatifs. Les échantillons **AAV2-PMS-01**, **AAV2-PMS-03** et **AAV2-PMS-04** sont négatifs pour les faibles concentration (**f**) utilisées lors de la bioconjugaison contrairement à l'**AAV2-PMS-02** qui est lui faiblement positif. Pour les (**F**) tous les échantillons d'AAV2 modifiés chimiquement sont positifs avec une intensité beaucoup plus faible pour l'**AAV2-PMS-01**. Même si ce Dot-blot est à optimiser à la vue du bruit de fond, ces résultats très encourageants confirment une meilleure interaction des AAV2 modifiés par les sucres (**PMS-02**), (**PMS-**

**03)** et **(PMS-04)** avec l'ASGPr en comparaison avec le ligand GalNAc « natif » et montrent donc l'importance de la structure des différents ligands couplés sur la capsid de l'AAV2.



*Figure 116 : Dot-blot révélé avec le domaine lectinique de l'ASGPr.*

La capsid de l'AAV2 est composée de 60 protéines nommées VP1, VP2 et VP3 et présentes dans un rapport 1/1/10. Une fois validée la structure quaternaire des AAV modifiés ainsi que la présence de ligands à leur surface, il est également nécessaire de vérifier si la bioconjugaison a lieu sur les différentes protéines composant sa capsid. Pour cela, une analyse de Western-blot est réalisée ([figure 117](#) et [box 4](#)).

#### Box 4 : Le Western-blot

La technique de transfert de protéines, en anglais Western-blot, permet la détection et l'identification de protéines spécifiques dans un échantillon biologique. Comme le Dot-blot, il s'agit d'une identification *via* des anticorps dirigés contre la partie N-terminale des VP que l'on souhaite détecter. La principale différence est qu'ici l'échantillon sera préalablement traité à l'aide d'une solution contenant un détergent (le dodécylsulfate de sodium) et parfois un réducteur des ponts disulfures comme le  $\beta$ -mercaptoéthanol. Cela aura pour effet de dénaturer les protéines et de les charger négativement. Le mélange sera alors déposé sur un gel de polyacrylamide et placé dans une cuve contenant un tampon spécifique (SDS-PAGE), l'action d'un courant électrique au sein de la cuve fera migrer les protéines dans le gel en fonction de leur poids moléculaire. Une fois cette étape réalisée les protéines seront transférées du gel à une membrane de nitrocellulose qui sera traitée de la même façon qu'un Dot-blot.

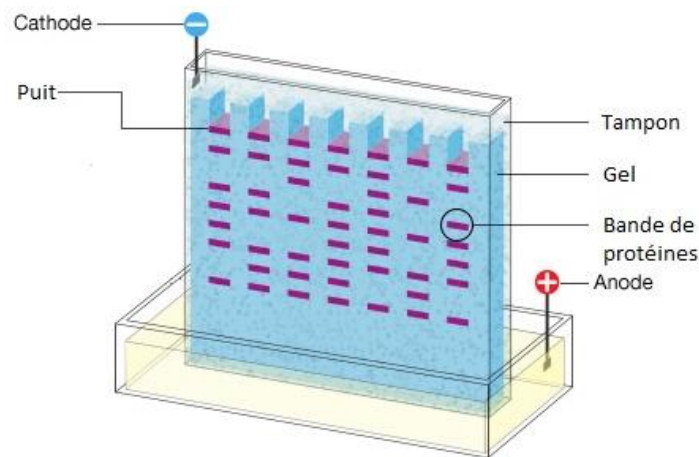


Figure 117 : Présentation de la technique de Western-blot.

Suite à la dénaturation de la capsid des différents AAV modifiés chimiquement ou non, le mélange est déposé sur un gel d'électrophorèse, transféré sur une membrane de nitrocellulose et les VP détectées par un anticorps polyclonal. L'utilisation d'une lectine spécifique du GalNAc va elle permettre la détection du sucre couplé sur les différentes protéines.

Comme le montre la (figure 118) l'utilisation d'un anticorps polyclonal permet de détecter VP1, VP2 et VP3 qui constituent la capsid des différents vecteurs avec un signal plus intense pour VP3 qui est la protéine la plus abondante (ratio 1/1/10). Cela confirme que la modification chimique n'est pas délétère pour les protéines qui conservent leur structure primaire. De plus, il est possible de voir une différence de migration entre les différentes VP de l'AAV2 non modifié et celles des échantillons à faible concentration (**f**) et surtout à forte concentration (**F**). Cette analyse met en évidence une augmentation de la masse molaire des différentes protéines surtout visible pour les protéines VP3.



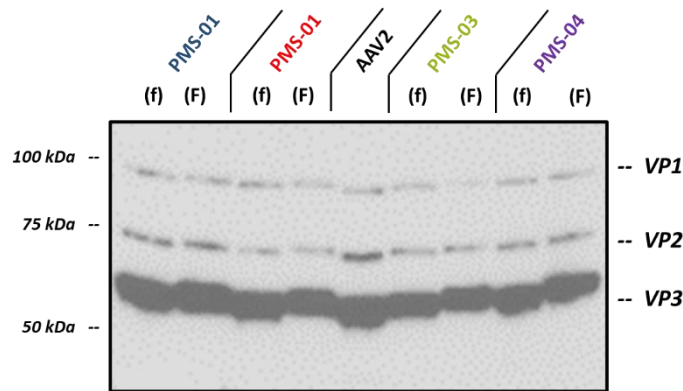


Figure 118 : Détection des protéines virales par Western-blot à l'aide de l'anticorps polyclonal.

Un nouveau Western-blot est réalisé avec la lectine WL pour valider la bioconjugaison des différents ligands sur les trois protéines de la capside (figure 119). Pour l'AAV2 natif on observe au niveau de la VP3 une bande de très faible intensité qui pourrait être due à une interaction non spécifique avec la lectine ou à la faible présence de sucre sur cette protéine. Concernant les différents échantillons d'AAV2-PMS-01, AAV2-PMS-03 et AAV2-PMS-04, on remarque sans ambiguïté une bande positive pour les protéines VP3 et cela pour les deux concentrations utilisées. Cette intensité est beaucoup plus forte pour les conditions (F) avec néanmoins les deux protéines VP1 et VP2 moins intenses car dix fois moins nombreuses que VP3. De plus, les protéines VP1 et VP2 sont principalement localisées à l'intérieur de la capside, elles ne sont donc que très peu exposées au milieu externe et donc couplées. Sans surprise les capsides d'AAV2-PMS-02 sont très peu reconnues par cette lectine comme déjà observé lors du Dot-blot.

L'utilisation de la lectine ASGPr n'a pas donné de résultats interprétables pour l'instant et l'optimisation de ce Western-blot est actuellement en cours.

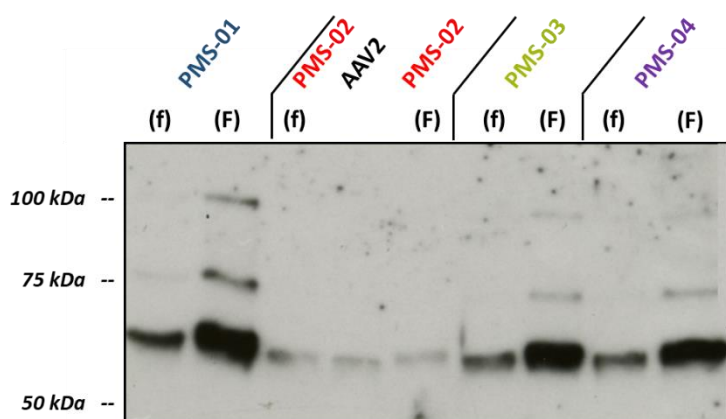


Figure 119 : Détection par Western-blot, à l'aide de la lectine WL, des ligands couplés chimiquement sur les protéines virales.

A ce stade, tous les résultats obtenus confirment la modification chimique de la capside de l'AAV2 sans dégradation et la possibilité de moduler le nombre de ligand bioconjugué. Il est cependant nécessaire de vérifier la composition protéique totale des échantillons et s'assurer qu'il n'y a pas eu de dégradation des VP qui ne seraient pas détectées par les différents anticorps ou lectines utilisés. Pour cela, un gel SDS-PAGE suivi d'une coloration au nitrate d'argent est effectuée. Cette technique dénature totalement la structure quaternaire de l'AAV pour finalement n'obtenir que les protéines le constituant qui seront ensuite révélés par l'action du nitrate d'argent. Ce dernier est réduit par les groupements sulfhydryle et carboxylique des chaînes latérales des acides aminés correspondant. Une fois réduit l'argent est sensible à la lumière et provoque l'apparition de bandes sur le gel. Comme le montre la (figure 120), le rapport VP1 / VP2 / VP3 est cohérent pour tous les échantillons ayant subi le processus de bioconjugaison en comparaison avec l'AAV2. Les résultats du Western-blot avec l'anticorps polyclonal sont également confirmés car pour tous les vecteurs le poids moléculaire de chaque VP augmente en fonction de la concentration en ligand (figure 120).

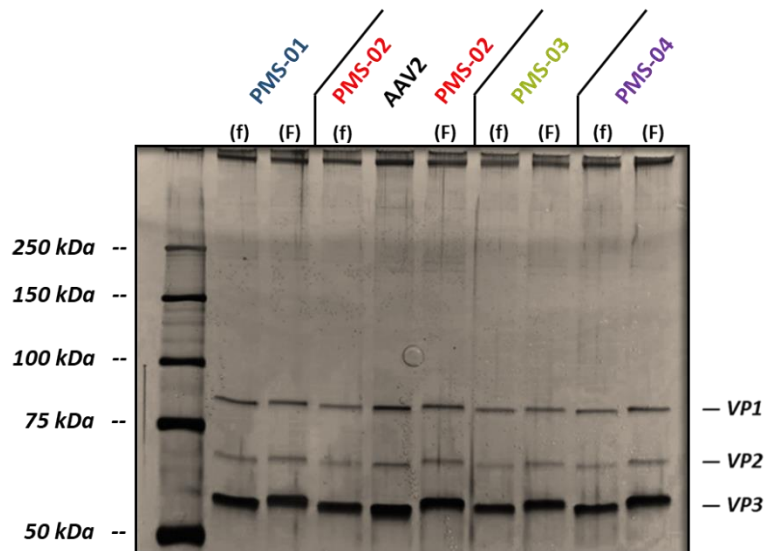


Figure 120 : Coloration au nitrate d'argent après SDS-PAGE afin de révéler les 3 protéines constitutives de la capsid des AAVs.

### C. Reconnaissance du vecteur

Après avoir caractérisé les vecteurs modifiés et validé le couplage covalent des ligands à leur surface, nous nous sommes intéressés à l'impact de la reconnaissance d'un anticorps avec la capsid assemblée. En effet, l'équipe de Patel<sup>176</sup> a identifié cinq peptides sur la capsid de l'AAV2 constituant chacun un épitope neutralisant unique, et trois de ces peptides contiennent au moins une lysine. Ainsi, la modification chimique de ces peptides fondamentaux pourrait avoir un impact sur les interactions des AAV2 avec les anticorps neutralisants préexistant dans la population.

Pour cela un test ELISA est effectué où l'anticorps A20 est immobilisé au fond des puits d'une plaque 96 puits, puis après incubation des différents AAV2, modifiés ou non, une lecture par absorbance à 450 nm est faite après plusieurs rinçages et après avoir ajouté un deuxième anticorps anti-A20 conjugué à l'HRP. La valeur de l'absorbance obtenue pour l'AAV2 a été normalisée à 100% et la (figure 121) regroupe l'ensemble des mesures pour chaque échantillon. L'étape de bioconjugaison ayant été effectuée à partir d'un même lot d'AAV2, une différence de signal sera due au couplage covalent des ligands sur la capsid des vecteurs.

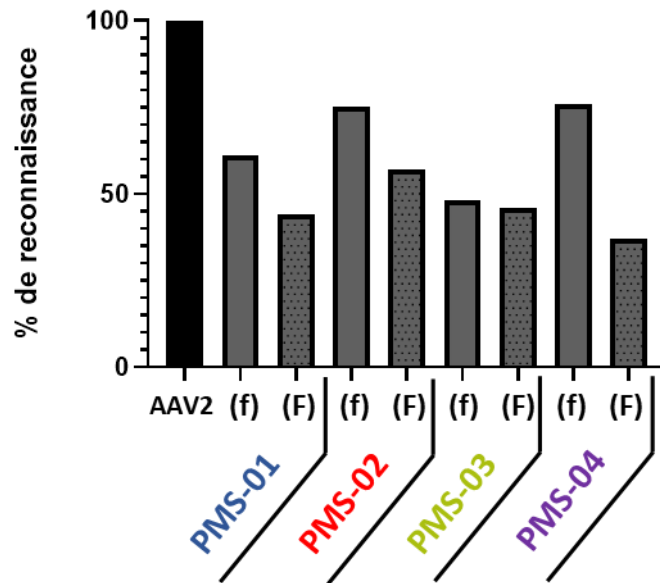


Figure 121 : Reconnaissance de l'anticorps A20 avec les différents échantillons.

Le test ELISA montre que tous les échantillons fonctionnalisés sont moins bien reconnus par l'anticorps A20. En effet, dans tous les cas on observe une diminution du pourcentage de reconnaissance. De manière logique les échantillons traités avec la faible concentration de ligands (**f**) sont mieux reconnus par l'anticorps que ceux traités avec la forte concentration (**F**). Ce résultat peut s'expliquer par la présence d'une plus grande quantité de ligands à la surface des AAV ce qui entrainerait possiblement un masquage plus important de ces épitopes. Ce constat est cependant différent pour l'**AAV2-PMS-03** qui présente un pourcentage de reconnaissance identique après greffage par une faible ou une forte dose de ligands. Cette analyse ayant été effectuée une seule fois, une deuxième est en cours de réalisation afin de confirmer ces résultats.

Pour aller plus loin dans la compréhension de l'interaction AAV modifiés / anticorps, nous avons entrepris un test avec des facteurs neutralisants comme précédemment décrit par l'équipe d'accueil.<sup>108</sup> Les AAV2, modifiés ou non, sont incubés avec un ensemble de sérums humains connu pour neutraliser la transduction des AAV2. Les titres de neutralisation ont été déterminés comme étant la plus haute dilution de sérum qui inhibe plus de 50 % du signal de transduction de l'AAV. Cependant, les résultats obtenus n'étant pas concluants ce test est actuellement en cours d'optimisation.

Après avoir évalué la capacité de reconnaissance de l'anticorps A20 pour ces nouvelles générations de vecteurs, il est maintenant intéressant de regarder leur profil d'affinité pour diverses lectines. Pour cela, nous avons fait appel à l'entreprise Glycodiag<sup>®</sup> spécialisée dans le criblage de glyco-composés. L'étude se divise en deux expériences, la première vise à étudier le comportement des vecteurs vis-à-vis de différentes lectines. La seconde vise à quantifier l'affinité des particules pour les récepteurs ASGP.

Les caractéristiques et le nom commun des neuf lectines examinées pour la première expérience sont répertoriés en (Annexe 2). Six d'entre elles reconnaissent les motifs GalNAc, dont la SBA, la WFA et l'ASGPr utilisées pour les analyses Dot-blot. Les lectines UEA-1, DSA et AIA reconnaissent les sucres galactose, glucosamine et fucose et ont été choisies comme contrôle négatif car a priori elles ne vont pas reconnaître les ligands PMS-01 à PMS-04.

Le protocole suivant a été utilisé :

Les vecteurs sont déposés à trois concentrations (*i.e.*  $5 \times 10^6$ ,  $2,5 \times 10^6$  et  $1,25 \times 10^6$  vg/mL) dans des puits contenant les différentes lectines et incubés 2h à température ambiante. Les plaques sont ensuite lavées avec du PBS puis l'anticorps A20 marqué à la biotine est ajouté et incubé pendant 30 min. Après lavage, le conjugué extravidine-peroxydase est additionné et laissé pendant 30 min. Après un dernier lavage des plaques, 100  $\mu$ L d'*ortho*-phénylènediamine sont ajoutés dans chaque puits afin d'être oxydés par la peroxydase en 2,3-diaminophénazine, un composé coloré qui émet à une longueur d'onde de 450 nm.<sup>130</sup> Après 15 min d'incubation, 100  $\mu$ L d'une solution d'HCl 1 M sont ajoutés pour dénaturer l'enzyme afin de stopper la réaction. Pour finir, l'absorbance émise par chaque puits est lue à 450 nm, valeur qui est proportionnelle au nombre de vecteurs interagissant avec la lectine, c'est-à-dire que plus la valeur de l'absorbance est élevée plus la lectine reconnaît le vecteur.

Le (tableau 8) rassemble les résultats des échantillons obtenus lors du screening sur les neuf lectines.

Vecteurs	Lectines								
	BPA	DBA	WFA	SBA	HPA	ASGPr	UEA-1	DSA	AIA
Sucre reconnu	GalNAc						Gal	GlcNac	Fuc
<b>AAV2</b>	-	-	-	-	-	-	-	-	-
<b>AAV2-PMS-01 (f)</b>	+	-	+	+	+	-	-	-	-
<b>AAV2-PMS-01 (F)</b>	+	-	+	+	+	-	-	-	-
<b>AAV2-PMS-02 (f)</b>	-	-	-	-	-	+	-	-	-
<b>AAV2-PMS-02 (F)</b>	+	-	-	-	-	+	-	-	-
<b>AAV2-PMS-03 (f)</b>	+	+	+	-	-	+	-	-	-
<b>AAV2-PMS-03 (F)</b>	+	+	+	-	-	+	-	-	-
<b>AAV2-PMS-04 (f)</b>	+	+	+	-	-	+	-	-	-
<b>AAV2-PMS-04 (F)</b>	+	+	+	-	-	+	-	-	-

Tableau 7 : Bilan des interactions entre vecteurs et lectines. + : Présence d'une interaction (valeur d'absorbance supérieur à 0,1); - : absence d'une interaction.

Comme attendu, il est important de noter que l'AAV2 ne présente aucune interaction avec les différentes lectines. Tous les vecteurs, excepté l'AAV-PMS-01 modifié aux deux concentrations, présentent une interaction pour l'ASGPr. Ce résultat est étonnant car le ligand (PMS-01) devrait présenter une affinité pour l'ASGPr, ce qui suggère qu'une fois greffé à l'AAV son interaction avec la lectine diminue. L'AAV-PMS-02 n'est reconnu que par l'ASGPr confirmant sa structure très sélective pour ce dernier. Il faut noter l'absence de reconnaissance des vecteurs AAV-PMS-03/04 pour les lectines SBA (comme dans le cas des analyses par Dot-blot) et HPA, la position de la chaîne triéthylène glycol en position 6 du sucre empêchant probablement son approche dans la poche lectinique. Enfin, les lectines AIA, UEA-1 et DSA ne reconnaissent aucun vecteur confirmant leur spécificité pour d'autres sucres.

Pour quantifier l'affinité des vecteurs pour le domaine lectinique du récepteur ASGP, la société Glycodiag® a étudié l'influence de la concentration (huit concentrations sériées de 0,5 en 0,5 à partir d'une concentration de vecteur de  $10^{11}$  vg/mL) tout en gardant le même protocole. Les résultats obtenus sont présentés (figure 122).

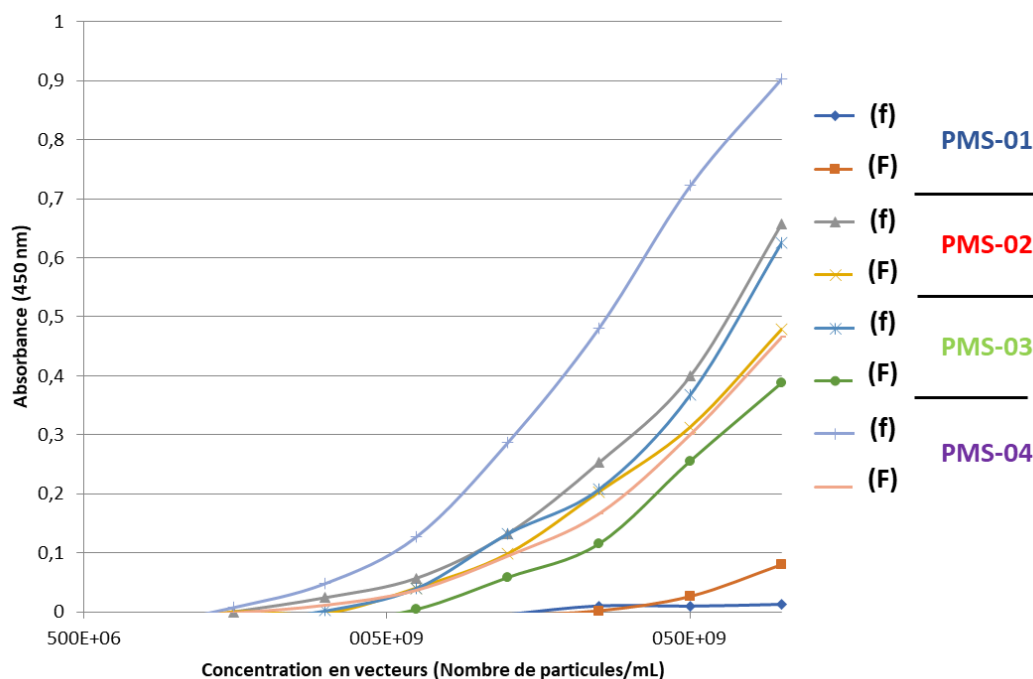


Figure 122 : Affinité pour le domaine lectinique de l'ASGPr des différents échantillons d'AAV2 modifiés chimiquement.

Tout d'abord, concernant l'AAV2-PMS-01 (F), on constate qu'il possède une affinité bien inférieure aux autres échantillons et une affinité quasiment nulle pour la condition (f) avec le domaine lectinique de l'ASGPr. Les résultats montrent aussi que les différents échantillons d'AAV2-PMS-02 (f), AAV2-PMS-03 (f) et AAV2-PMS-04 (f) sont plus affins pour l'ASGPr que leurs homologues modifiés à la forte concentration. Ce résultat peut paraître contre intuitif car il serait logique que les AAV avec à leur surface une plus grande quantité de ligands aient une interaction plus forte avec le récepteur et de ce fait une meilleure affinité. Néanmoins, il ne faut pas oublier que pour réaliser cette expérience il est nécessaire d'immobiliser les échantillons au fond des puits avec l'anticorps A20. Cependant, comme l'a montré l'analyse ELISA (figure 121), l'anticorps A20 reconnaît moins les vecteurs fortement modifiés (F) que ceux faiblement modifiés (f). De ce fait, il est possible que la quantité de vecteurs immobilisée dans l'expérience menée par Glycodiag® soit plus faible pour les vecteurs (F) que (f). Ce biais apporté par l'utilisation de l'anticorps A20 complique l'analyse comparative des différents groupes. Il est également important de préciser que lors de cette analyse nous ne sommes pas arrivés à un maximum d'absorbance pour les différents échantillons, il n'est donc pas possible de calculer les constantes d'affinités pour les vecteurs modifiés pour l'ASGPr. Pour cela, il aurait fallu utiliser des concentrations plus importantes de vecteurs ce qui aurait demandé la bioconjugaison d'un nouveau lot d'AAV et d'en modifier une quantité bien plus importante.

Etant conscient des possibles biais engendrés lors de cette expérience, nous avons tout de même voulu analyser les résultats en comparant les échantillons **(f)** et **(F)** entre eux.

Concernant les faibles conditions de modification, l'**AAV2-PMS-04** est l'échantillon présentant la meilleure affinité suivie de l'**AAV2-PMS-02**, très proche de l'**AAV2-PMS-03**. Ces résultats ne coïncident pas avec la littérature.<sup>96</sup> En effet, le sucre GalNAc ponté composant le ligand **(PMS-02)** a été décrit comme étant plus affiné pour le récepteur que celui du ligand **(PMS-04)**. La chaîne TEG est sur la même position pour ces deux sucres, il est donc peu probable qu'elle joue un rôle déterminant dans cette différence d'affinité. Concernant le sucre issu de **(PMS-03)** aucune donnée sur son affinité pour le récepteur n'a été révélée. Pour les échantillons modifiés à la forte concentration, la tendance, bien que moins marquée, reste la même.

Finalement, il est difficile de tirer de réelles conclusions de cette étude à cause des différents biais discutés précédemment. De plus, les tendances observées ne semblent pas être en accord avec celles de la littérature, suggérant une potentielle contribution de l'AAV sur l'interaction avec le récepteur.

Une façon d'améliorer cette expérience pourrait être de s'affranchir de l'étape avec l'anticorps A20 en venant directement immobiliser au fond du puits une quantité connue des différents AAV modifiés. Une seconde critique de l'analyse effectuée est que la quantité de vecteur n'est pas assez élevée, de ce fait les plafonds des absorbances ne sont jamais atteints ce qui est nécessaire pour pouvoir correctement définir une constante d'affinité.



### 3. Infectiosité des vecteurs modifiés *in vitro*

Cette partie concernera l'impact du ligand sur l'infectiosité de la particule sur des cellules HEK293 (lignée cellulaire non hépatocytaire), HUH7 et HepG2 (lignées cellulaires hépatocytaires), et sur des hépatocytes primaires de souris.

L'infectiosité a été calculée, sur les différentes lignées cellulaires, par détection de la protéine GFP par cytométrie de flux (FACS) (Box 5).<sup>177</sup>

#### Box 5 : L'analyse par FACS

FACS est l'abréviation de "Fluorescence Activated Cell Sorting" (tri cellulaire activé par fluorescence) et est une technique utilisée en cytométrie en flux pour séparer physiquement les cellules sur la base de leurs propriétés de fluorescence. Dans le FACS, les cellules sont d'abord colorées avec des colorants fluorescents ou des anticorps qui se lient à certaines molécules à la surface des cellules (dans notre cas c'est la protéine GFP qui permet la fluorescence de la cellule). Les cellules colorées sont ensuite passées dans une cellule à flux continu où elles sont éclairées par un faisceau laser. Pendant que les cellules traversent le laser, leur fluorescence est enregistrée et analysée en temps réel par un ordinateur. Sur la base des propriétés de fluorescence, les cellules sont triées et réparties dans différents tubes de collecte ou plaques. Il est aussi possible de directement quantifier le nombre de cellules fluorescentes.

#### A. Lignée cellulaire non hépatocytaire

Pour commencer, des cellules HEK293 ont été infectées à une MOI de  $10^5$ , où la MOI (de **M**ultiplicity **O**f **I**nfection) représente le nombre de particules virales par cellule. La phase d'incubation dure 48h puis les cellules sont récoltées et analysées par cytométrie de flux.

Comme le montre la (figure 123), le pourcentage de cellules GFP positives, à cette MOI, est de l'ordre de 100% pour tous les vecteurs utilisés. Ce résultat confirme que la réaction de bioconjugaison n'est pas délétère à l'infectiosité des différents échantillons d'AAV modifiés. Afin de mettre en évidence de possibles différences d'infectiosité, des MOI plus faibles auraient pu être utilisées afin d'éviter la saturation du signal. Cependant, comme à ce stade notre seul critère d'évaluation était un maintien de l'infectiosité, nous n'avons pas été plus loin dans cette investigation.

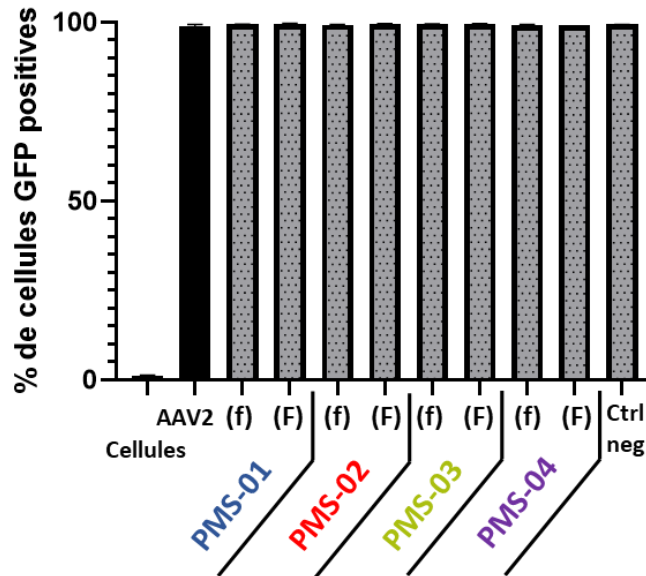


Figure 123 : Mesure de l'infectiosité des différents échantillons d'AAV2 sur des cellules HEK293. MOI=1<sup>E5</sup>, n=3.

## B. Lignées cellulaires hépatocytaires

### a) Transduction de cellules HUH7

Pour la suite, nous nous sommes intéressés à l'infectiosité de nos échantillons d'AAV modifiés sur des lignées cellulaires hépatocytaires. Les cellules HUH7 sont issues d'une lignée cellulaire immortalisée de carcinome dérivée d'hépatocytes primaires. Ces cellules possèdent donc à leur surface le récepteur ASGP. Les différents sucres couplés sur la surface de la capsid de l'AAV2 devraient permettre de modifier le tropisme du vecteur en interagissant avec le récepteur ASGP. Nous espérons ainsi voir une différence du nombre de cellules GFP positives entre l'AAV2 et les AAV modifiés mais également entre ces derniers en fonction des ligands greffés à leur surface.

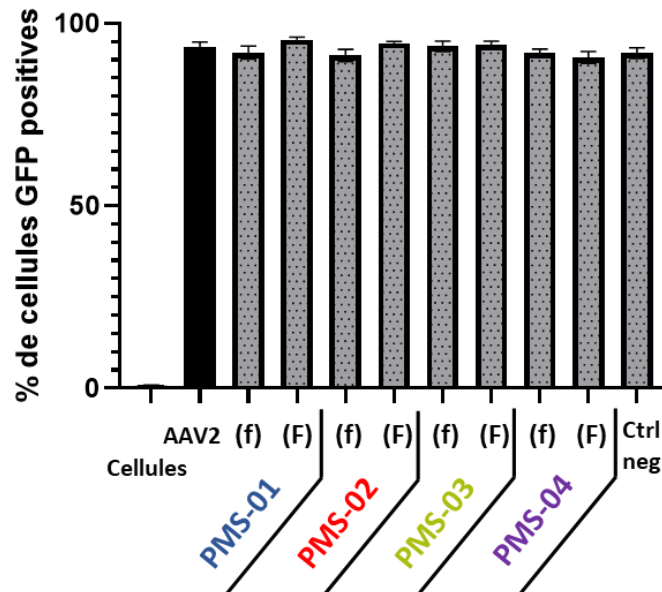


Figure 124 : Mesure de l'infectiosité des différents échantillons d'AAV2 sur des cellules HUH7. MOI=1<sup>E5</sup>, n=3.

Les cellules HUH7 ont été infectées à une MOI de 1<sup>E5</sup> avec le même protocole que les HEK293. La (figure 124) représente les résultats obtenus, 48h après transduction, par analyse FACS. Ici encore, le nombre de cellules GFP positives de chaque échantillon est similaire et proche des 95%. Il est difficile à cette MOI d'observer une différence significative entre les différents échantillons, c'est pourquoi une seconde mesure a été réalisée en diminuant la MOI à 1<sup>E4</sup> et 1<sup>E3</sup> (figure 125).

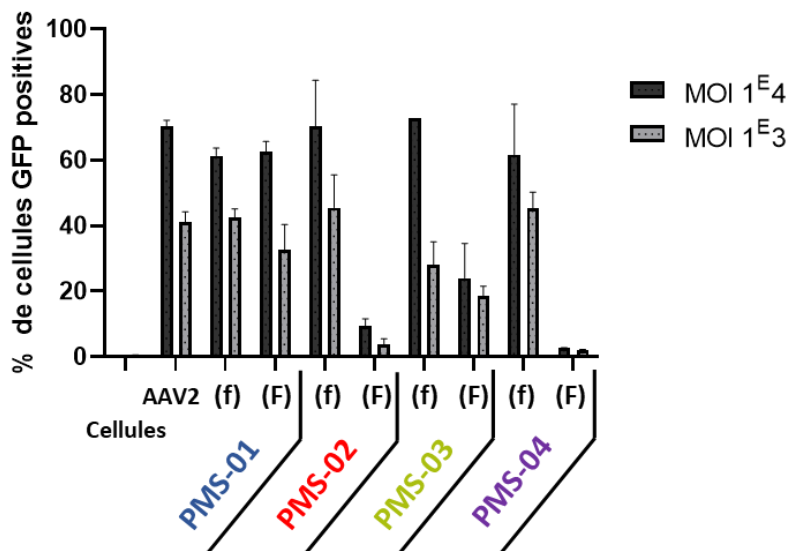


Figure 125 : Mesure de l'infectiosité des différents échantillons d'AAV2 sur des cellules HUH7 à différentes MOI. n=3.

Cette deuxième expérience montre que pour les deux MOI utilisées les vecteurs **AAV2-PMS-01**, **AAV2-PMS-02**, **AAV2-PMS-03** et **AAV2-PMS-04**, modifiés à une faible concentration, ont une efficacité de transduction du même ordre de grandeur que celle de l'AAV2. Ces résultats sont décevants car nous nous attendions au contraire à une augmentation de la transduction une fois les AAV greffés avec des ligands affins aux ASGPr. Pour les vecteurs modifiés à forte concentration, à part pour l'**AAV2-PMS-01**, l'efficacité est plus faible que celle obtenue avec l'AAV2. Une hypothèse par rapport à ces résultats aurait pu être qu'une trop forte proportion du ligand à la surface du vecteur l'empêche de correctement s'intérioriser dans la cellule. Cependant ce phénomène n'est pas mis en évidence avec l'**AAV2-PMS-01 (F)** invalidant cette hypothèse. Une deuxième possibilité pourrait être le devenir des vecteurs une fois internalisés et notamment lors de l'étape d'échappement endosomal. En effet, si l'interaction avec le récepteur ASGP est trop forte, la dissociation avec ce récepteur après internalisation pourrait empêcher l'échappement endosomal et conduire à une destruction de l'AAV modifié chimiquement. Des études de suivi du trafficking intra-cellulaire seraient intéressantes pour observer le cheminement des différents vecteurs et tenter de répondre aux hypothèses émises.

### b) Transduction de cellules HEPG2

Pour confirmer ou non ces résultats, une autre lignée hépatocytaire (cellules HEPG2) a été transduite avec une MOI de  $10^4$  (figure 126). Dans ce cas tous les échantillons présentent la même infectiosité que l'AAV2 à l'exception de l'**AAV2-PMS-04 (F)** qui induit un nombre de cellules GFP positives plus faible.

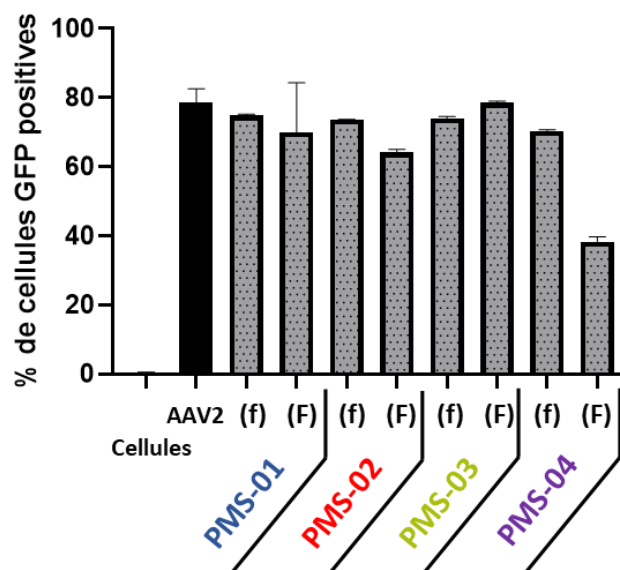


Figure 126 : Mesure de l'infectiosité des différents échantillons d'AAV2 sur des cellules HEPG2. MOI=10<sup>4</sup>, n=2.

En définitive, il nous est impossible de conclure sur l'efficacité des échantillons modifiés par rapport à l'AAV2 et encore moins sur une différence entre les différents ligands. Il faut cependant garder en tête que les cellules utilisées proviennent de lignées cellulaires immortelles et bien que la littérature indique la présence de récepteurs ASGP à leur surface, il est difficile d'en connaître le nombre. C'est pourquoi, étant donné les niveaux de transduction élevés et la difficulté de comparer les différents vecteurs entre eux, nous nous sommes orientés vers l'utilisation d'hépatocytes primaires murins pour lesquels l'efficacité de transduction de l'AAV2 est plus faible.<sup>108</sup>

### *c) Transduction d'hépatocytes primaires murins*

Les propriétés de transduction des différents vecteurs modifiés ont été évaluées *in vitro* sur des hépatocytes primaires murins qui sont connues pour ne pas être bien transduites par l'AAV2. Ainsi, il pourrait être plus facile de mettre en évidence un effet bénéfique de la bioconjugaison sur la transduction.

La (figure 127) représente le pourcentage d'hépatocytes primaires exprimant la GFP. Comme évoqué précédemment, l'AAV2 est peu efficace et ne transduit que 3% de ces cellules. Pour les échantillons d'**AAV2-PMS-01**, **AAV2-PMS-02**, **AAV2-PMS-03** fonctionnalisés avec la faible concentration en ligand (**f**) aucune différence d'infectiosité par rapport à l'AAV2 n'est observée. En revanche, ces mêmes échantillons modifiés à la forte concentration (**F**) montrent une différence significative d'efficacité de transduction avec une augmentation de plus de 100% par rapport à l'AAV2. En effet, les valeurs observées passent d'environ 3% avec l'AAV2 à des valeurs entre 6 et 8% avec les vecteurs **AAV2-PMS-01**, **AAV2-PMS-02**, **AAV2-PMS-03**. L'**AAV2-PMS-04** quant à lui se comporte différemment et on constate une infectiosité supérieure à l'AAV2 pour la faible concentration en ligand (**f**) et une plus faible pour la forte (**F**). Ce résultat est étonnant et va à l'encontre de la conclusion établie pour les autres échantillons. Il n'est cependant pas illogique et va dans le sens de ceux déjà montrés sur les lignées cellulaires HUH7 et HEPG2 où la condition (**F**) était déjà moins infectieuse que la condition (**f**) et que l'AAV2. Seule une étude sur le trafficking intra-cellulaire des particules pourrait nous permettre de comprendre ce résultat.

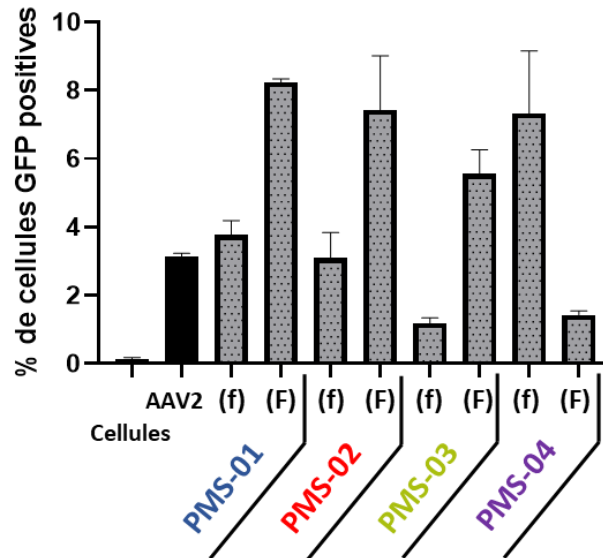


Figure 127 : Mesure de l'infectiosité des différents échantillons d'AAV2 sur des hépatocytes primaires murins.  $MOI=1^{E5}$ ,  $n=3$ .

Pour conclure, cette expérience est à prendre avec un certain recul car le taux de viabilité des cellules utilisées était très faible lors de leur arrivée dans le laboratoire, ce constat s'est avéré identique pour les différentes commandes réalisées. Les hépatocytes primaires sont des cellules difficiles à utiliser du fait de leur grande sensibilité. La viabilité de ces dernières n'était pas optimale dans cette manipulation pour tirer toutes les conclusions nécessaires. Ces premiers résultats sont néanmoins encourageants et ont été suivis d'une étude *in vivo*.

#### 4. Etude *in vivo* des nouveaux vecteurs chimiquement modifiés.

Après avoir caractérisé la réaction de bioconjugaison, testé l'infectiosité des vecteurs sur différentes lignées cellulaires et montré l'apport de la modification sur une lignée hépatocytaire primaire, une étude *in vivo* chez des souris C57Bl6 mâles a été réalisée.

Suite à la validation éthique du protocole de cette étude *in vivo* par le ministère de l'enseignement supérieur et de la recherche, 60 animaux divisés en 10 groupes de 6 souris ont été injectés par voie intraveineuse. Huit groupes de souris ont reçu les vecteurs modifiés **AAV2-PMS-01**, **AAV2-PMS-02**, **AAV2-PMS-03** et **AAV2-PMS-04**, chacun modifié à la faible (**f**) et la forte (**F**) concentration de ligands. Un groupe a été injecté avec le vecteur AAV2 non modifié et enfin, le dernier groupe a reçu uniquement du dPBS, le tampon de formulation des vecteurs AAVr. Les premiers animaux inclus dans l'étude (groupes 1, 2, 3, 4 et 7) ont été injectés au niveau de la veine caudale (veine de la queue) et les

suivants (groupes 5, 6, 8, 9 et 10) ont été injectés au niveau de la veine péniennne car ce geste était plus simple à réaliser et permettait d'améliorer la reproductibilité des injections. Par ailleurs, les précédentes données du laboratoire ne montraient pas de différence de biodistribution des vecteurs AAVr entre ces 2 voies d'administration (données non publiées). Les animaux de l'étude ont reçu une dose de vecteur de  $5 \times 10^{12}$  vg/kg et ont été euthanasiés 1 mois après l'injection. Nous avons fait le choix de cette dose intermédiaire à  $5 \times 10^{12}$  vg/kg afin de ne pas saturer la transduction du foie et ainsi mettre en évidence le potentiel impact positif de la bioconjugaison sur l'efficacité de transfert de gène dans le foie. C'est dans cette même logique que le sérotype 2 a été choisi, car son tropisme (ou ciblage) du foie n'est pas maximal, si on le compare avec l'AAV8 par exemple.<sup>178</sup> Il sera ainsi plus simple de voir l'impact de la modification chimique de la capsid.

Après l'euthanasie des différentes souris, les organes suivants ont été prélevés : foie, cœur, rate, triceps et quadriceps gauche, poumons et reins. Ces organes ont été découpés puis traités de différentes manières. Une partie a été directement congelée dans de l'azote liquide pour des analyses moléculaires et biochimiques (extractions d'ADN, d'ARN et de protéines). Une autre a été congelée dans de l'isopentane refroidi dans l'azote liquide pour des analyses immunohistochimiques (immunomarquages à l'aide d'anticorps sur coupes de tissus). Et enfin une dernière partie a été conservée dans du formol pour des analyses histologiques (analyses des structures tissulaires sur coupes de tissus), non présentées dans ce manuscrit.

## A. Quantification des génomes viraux dans le foie des souris injectées avec les différents vecteurs AAV2

Dans un premier temps nous nous sommes intéressés à la quantification du nombre de copies du génome viral, porté par les vecteurs AAV2 modifiés ou non, présents dans le foie des différentes souris de l'étude. Pour cela, l'ADN génomique total d'une partie du foie a été extrait, puis la quantification absolue du nombre de copies de génome viral a été réalisée par PCR quantitative (qPCR) en ciblant la séquence du transgène GFP. La qPCR permet de mesurer, en temps réel, l'amplification d'un gène que l'on veut quantifier (ou amplicon) par la détection d'un signal fluorescent. Une réaction de qPCR est composée de plusieurs cycles d'amplification. Au fur et à mesure des cycles d'amplification, l'intensité de fluorescence dépassera un seuil de fluorescence et déterminera une valeur de C<sub>q</sub> (quantification cycle) pour l'échantillon analysé. Autrement dit, la valeur de C<sub>q</sub> d'un échantillon correspond au nombre de cycle nécessaires pour que sa courbe d'amplification dépasse le seuil de fluorescence. Une gamme étalon présentant des nombres de copies connues de l'amplicon est analysée en parallèle des échantillons et permet de reporter les valeurs de C<sub>q</sub> de ces derniers afin de déterminer le nombre absolu de copie du gène. Le nombre de copies de génome viral GFP est ensuite normalisé par le

nombre de copie du gène endogène hypoxanthine phosphoribosyltransférase (*hprt1*), quantifié également par qPCR. Sachant que chaque noyau cellulaire contient 1 copie de ce gène endogène *hprt1*, il est alors possible de réaliser un rapport entre le nombre de copie de transgène GFP et le nombre de copie de gène *hprt1* pour obtenir le nombre de copie de transgène GFP par noyau cellulaire en vg/dg (vector genome/diploïd genome).

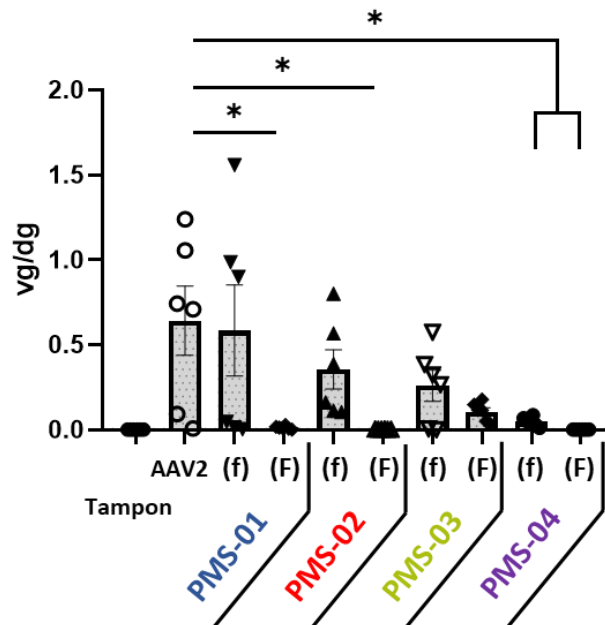


Figure 128 : Quantification par qPCR des génomes de vecteur (vg/dg) dans le foie des souris injectées (n =6) avec les différents vecteurs de l'étude in vivo. Les données sont présentées sous forme de moyenne  $\pm$  SEM avec une limite de quantification (LOQ = 0,006 vg/dg). Les statistiques ont été réalisées à l'aide d'un test Mann-withney non paramétrique. Les valeurs P \*p < 0,05 ont été considérées comme statistiquement significatives. Les valeurs respectives sont de 0,0156, 0,0133, 0,0292 et 0,0133.

Les résultats de quantification des génomes viraux dans le foie des souris injectées sont présentés sur la (figure 128). Comme attendu, aucun vg/dg n'a été détecté chez les animaux injectés avec le tampon de formulation dPBS alors que le transgène GFP a été détecté chez les animaux ayant reçu l'AAV2 non modifié, avec en moyenne 0,641 vg/dg ce qui est cohérent avec la littérature.<sup>179,180</sup> Concernant les groupes injectés avec les vecteurs modifiés, on observe une diminution significative du niveau de transduction par rapport au groupe AAV2 non modifié pour les groupes **AAV2-PMS-01 (F)**, **AAV2-PMS-02 (F)** et **AAV2-PMS-04 (F)**, avec respectivement une moyenne de 0,013, 0,001 et < 0,006 vg/dg. Une tendance à la diminution est également observée chez les animaux injectés avec les vecteurs **AAV2-PMS-02 (f)**, **AAV2-PMS-03 (f)**, **AAV2-PMS-03 (F)** et **AAV2-PMS-04 (f)**, avec respectivement une moyenne de 0,355, 0,259, 0,106 et 0,047 vg/dg, ces valeurs sont relativement faibles vis-à-vis de celle pour l'AAV2 qui est déjà considéré comme un vecteur moyen pour la transduction hépatique.<sup>180</sup> En revanche, les niveaux de transduction du foie semblent similaires entre les animaux injectés avec le vecteur AAV2 non modifié et le vecteur **AAV2-PMS-01 (f)** avec 0,584 vg/dg pour ce dernier. Ces



résultats mettent en évidence un effet de la bioconjugaison sur la quantité de génomes viraux retrouvés dans le foie des souris injectées. L'impact de cette bioconjugaison semble dépendre de la nature du ligand et du nombre de ligand conjugué à la capsid de l'AAV2. En effet, les fortes doses de ligand induisent une diminution du nombre de vg/dg dans le foie des souris injectées. De la même manière, les ligands **(PMS-03)** et **(PMS-04)** semblent d'avantage impacter l'efficacité de transduction de l'AAV2 dans le foie de souris.

Même si à première vue, la modification chimique n'améliore pas la présence du nombre de copies de génomes viraux dans le foie, cela ne permet pas de conclure sur leur présence dans le noyau cellulaire. En effet, l'analyse qPCR ne permet pas de localiser pas les vg au niveau du noyau cellulaire mais uniquement de les quantifier dans la globalité du tissu. Or, seuls les vg présents dans le noyau cellulaire seront transcriptionnellement actifs et permettront l'expression du transgène sous forme d'ARNm puis de protéine. C'est pourquoi nous avons ensuite évalué la présence de l'ARN messenger et de la protéine GFP dans le foie des animaux injectés.

## B. Quantification de l'ARN messenger GFP dans le foie des souris injectées avec les différents vecteurs AAV2

Afin de quantifier les ARN messenger GFP présents dans le foie des souris injectées, l'ARN total d'une partie du tissu a été extrait. L'ARN messenger de la GFP a ensuite été converti en ADN complémentaire (ADNc) selon un processus appelé « rétro-transcription » (RT), étape nécessaire car seul l'ADN peut être amplifié et quantifié en qPCR. A la suite de cette rétro-transcription, les ADNc de la GFP ont été quantifiés par qPCR. Afin de normaliser l'expression de la GFP, les ADNc endogène HPRT1 ont également été quantifié par cette technique. Cette fois-ci une quantification relative du messenger GFP a été effectuée. On établit un rapport RQ (Relative Quantification) grâce à la formule suivante :

$$\mathbf{RQ = 2^{-\Delta Cq} = 2^{-(Cq_{GFP} - Cq_{mHPRT1})}}$$

La **(figure 129)** représente la quantité d'ARN messenger de la GFP en quantité relative (RQ) dans le foie des souris injectées avec les différents vecteurs testés dans cette étude *in vivo*.

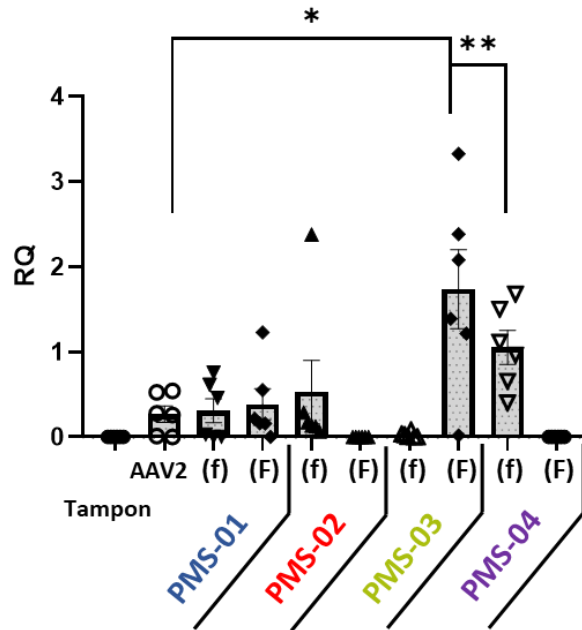


Figure 129 : Quantification relative par RT-qPCR des messagers GFP dans le foie des souris injectées (n = 6) avec les différents vecteurs de l'étude in vivo. Les données sont présentées sous forme de moyenne  $\pm$  SEM avec une limite de quantification (LOQ = 0,0004). Les statistiques ont été réalisées à l'aide d'un test Mann-whitney non paramétrique. Les valeurs P sont \*p < 0,05, \*\*p < 0,01 ont été considérées comme statistiquement significatives, les valeurs respectives sont de 0,0260 et 0,0087.

Comme attendu, aucun messager GFP n'a été détecté chez les animaux injectés avec le tampon de formulation dPBS alors que le messager GFP a été détecté chez les animaux ayant reçu l'AAV2 non modifié, avec un RQ d'en moyenne 0,2697. Concernant les groupes injectés avec les vecteurs modifiés, on observe une augmentation significative de l'expression du messager GFP pour les groupes **AAV2-PMS-03 (F)** et **AAV2-PMS-04 (f)** avec respectivement des valeurs de RQ d'environ 6 et 4 fois plus importantes que pour le groupe AAV2 non modifié. Pour les animaux injectés avec les vecteurs **AAV2-PMS-01 (f) et (F)** et **AAV2-PMS-02 (f)**, l'expression de la GFP est similaire au groupe ayant reçu l'AAV2 non modifié. Enfin, pour les groupes **AAV2-PMS-02 (F)**, **AAV2-PMS-03 (f)** et **AAV2-PMS-04 (F)**, le niveau d'expression de l'ARNm GFP est quasi indétectable avec des valeur RQ respectivement de 0,0008, 0,0246 et 0,0007, soit juste très proche de la LOQ (limite of quantification) qui est de 0,0004.

Cette analyse RT-qPCR nous indique que les souris injectées avec les vecteurs modifiés **AAV2-PMS-03 (F)** et **AAV2-PMS-04 (f)** expriment davantage l'ARNm GFP que les souris injectées avec l'AAV2 non modifié, alors même que le nombre de vg/dg détectés dans le foie des souris injectées avec ces vecteurs modifiés étaient inférieurs à ceux détectés dans les foies injectés avec l'AAV2 non modifié (**figure 129**). Nous avons donc voulu mettre en relation les résultats d'expression du messager GFP avec les résultats du nombre de copies du transgène GFP, en exprimant le rapport de la valeur de RQ sur la valeur vg/dg. En effet, nos résultats semblent indiquer que ce n'est pas parce que le nombre de vg/dg

dans un tissu est important que cela sera aussi le cas pour l'ARN messenger, voire pour la protéine, car rien n'indique la présence nucléaire du transgène, nécessaire pour sa prise en charge par la cellule et l'exécution des étapes de transcription et de traduction.

Ce rapport RQ sur vg/dg, que nous avons appelés « pouvoir transcriptionnel » est présenté sur la (figure 130). Les résultats montrent que les groupes **AAV2-PMS-01 (F)**, **AAV2-PMS-03 (F)** et l'**AAV2-PMS-04 (f)** présentent respectivement des rapports d'infectiosité environ 77, 48 et 84 fois plus importants que celui du groupe AAV2 non modifié. En d'autres termes, cela signifie qu'à nombre de copies équivalent dans un noyau cellulaire, le transgène GFP semble transcriptionnellement plus actif dans ces 3 groupes injectés avec les vecteurs modifiés. Il faut tout de même analyser ce rapport avec précaution. En effet, pour le groupe **AAV2-PMS-01 (F)**, le rapport RQ sur (vg/dg) est élevé mais ce n'est pas le groupe qui exprime le plus le messenger GFP. Néanmoins, ces résultats illustrent que la modification chimique des capsides de ces rAAV2 a un impact sur l'efficacité du transfert de gène dans le foie de souris injectées en intraveineuse à une dose de  $5 \times 10^{12}$  vg/kg. L'impact sur la particule d'AAVr pourrait intervenir dès son entrée dans la cellule, sur son cheminement intracellulaire, jusqu'à la libération du génome et son atteinte au noyau pour permettre une expression efficace du transgène.

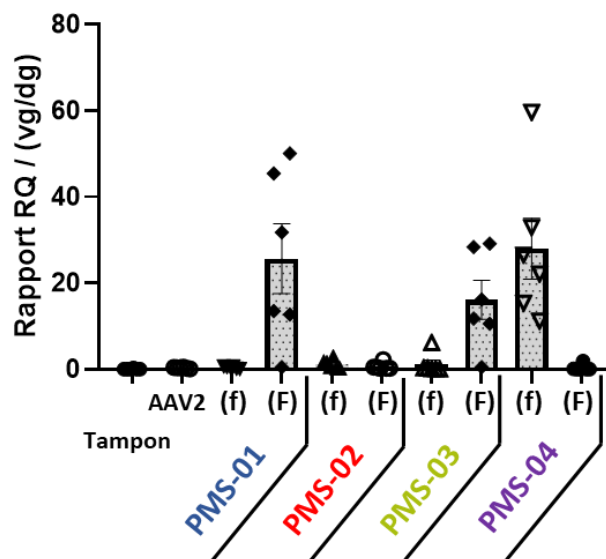


Figure 130 : Rapport RQ / (vg/dg) dans le foie des souris injectées (n = 6) avec les différents vecteurs de l'étude in vivo.

## C. Evaluation de l'expression de la protéine GFP dans le foie des souris injectées avec les différents vecteurs AAV2

### *d) Evaluation de l'expression de la protéine GFP et semi-quantification par Western-blot*

Les analyses RT-qPCR semblent nous indiquer que les vecteurs modifiés **AAV2-PMS-03 (F)** et **AAV2-PMS-04 (f)** présentent un réel potentiel pour augmenter l'expression d'un transgène thérapeutique dans le foie de souris après injection intraveineuse. Il restait à valider cette hypothèse par l'évaluation de la présence de la protéine GFP dans le tissu d'intérêt.

Pour cela, les protéines totales des différents échantillons de foie ont été extraites, puis la présence de la GFP a été évaluée par Western-blot. Dans cette étude, deux protéines ont été marquées : la protéine GFP et une protéine endogène, la glycéraldéhyde-3-phosphate déshydrogénase (GAPDH). La GAPDH est une protéine présente à des niveaux comparables dans différents tissus et a été utilisée pour normaliser l'expression de la GFP des différents échantillons, de façon semi-quantitative.

L'ensemble des images issues des analyses Western-blot et la semi-quantification du signal de la protéine GFP relative au signal GAPDH sont présentés (**figure 131**). Cette figure présente les 6 gels regroupant les 60 échantillons protéiques de foie des souris injectées lors de cette étude *in vivo*. Chaque gel contient une souris de chaque groupe ainsi que la souris S820 injectée avec l'**AAV2-PMS-01 (F)**, utilisée ici à la fois comme contrôle positif et comme référent inter-gel afin de pouvoir comparer les résultats de semi-quantification de la GFP d'échantillons déposés sur des gels différents.

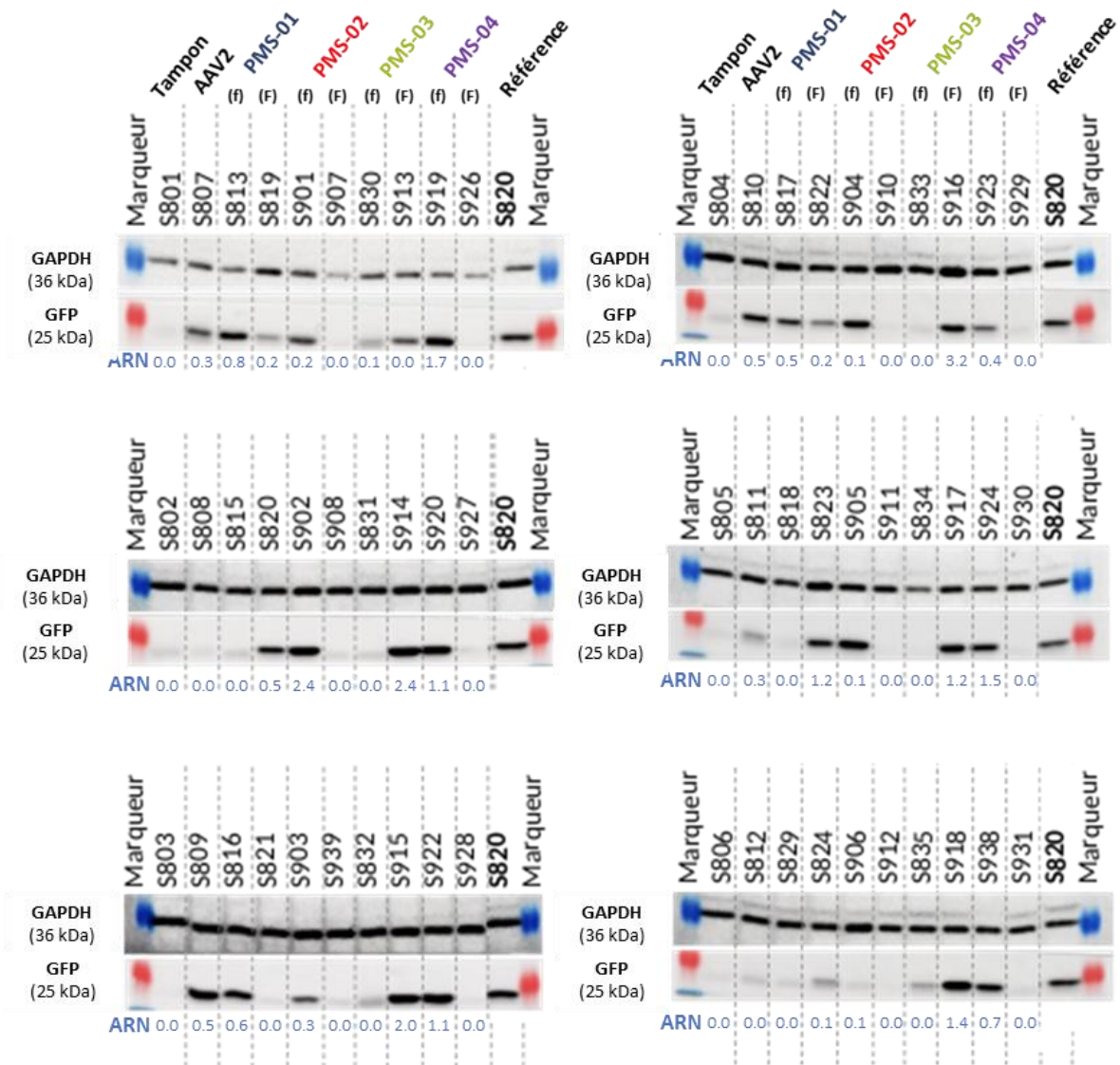


Figure 131 : Détection de la présence de GFP par Western-blot dans le foie des souris injectées avec les différents vecteurs de l'étude *in vivo*.

Cette figure indique dans un premier temps qu'aucun signal GFP n'est observé pour l'ensemble des animaux injectés avec le tampon de formulation dPBS et les vecteurs **AAV2-PMS-02 (F)** et **AAV2-PMS-04 (F)**. Il est intéressant de voir que les intensités de signal GFP sont en accord avec ceux quantifiant la présence d'ARN messager. En effet, plus la valeur de RQ est importante, plus le signal GFP détecté par Western-blot est fort. Ce constat est néanmoins discutable pour les souris injectées avec l'**AAV2-PMS-02 (f)** et plus précisément les souris S904 et S905 qui présentent une valeur de RQ faible mais un signal GFP à forte intensité sur le Western-blot. Ce résultat pourrait s'expliquer par une possible hétérogénéité de transduction du tissu avec des zones fortement transduites et d'autres beaucoup moins.

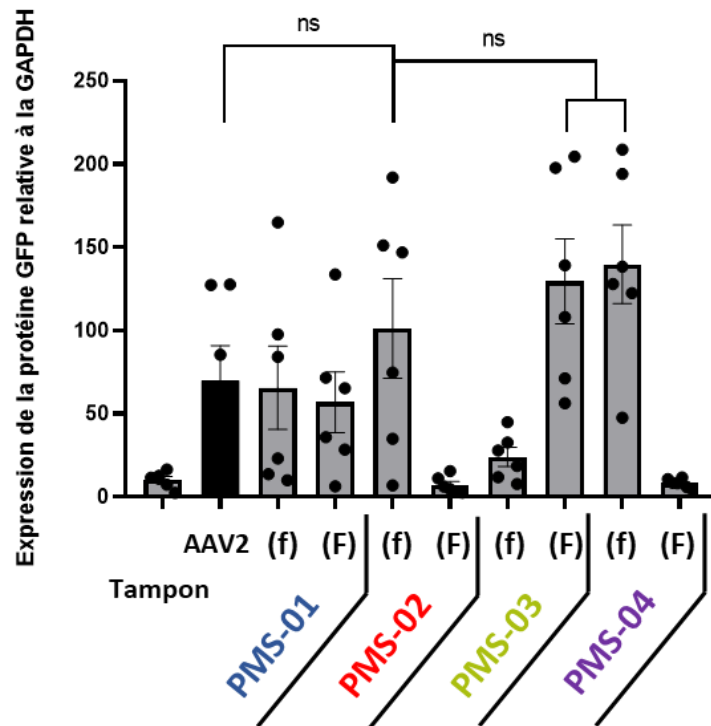


Figure 132 : Semi-quantification de la protéine GFP dans les différents échantillons ( $n = 6$  par groupe d'échantillons) par rapport à la GAPDH. Les données sont présentées sous forme de moyenne  $\pm$  SEM. Les statistiques ont été réalisées à l'aide d'un test Mann whitney non paramétrique. Les valeurs  $P$  sont  $*p < 0,05$  ont été considérées comme statistiquement significatives. Les valeurs respectives sont de 0,4848, 0,1320 et 0,0649.

Les résultats de semi-quantification (figure 132) indiquent dans un premier temps qu'aucune expression de la GFP n'est observée pour chez les animaux injectés avec le tampon de formulation dPBS et les vecteurs **AAV2-PMS-02 (F)** et **AAV2-PMS-04 (F)**. Ensuite, l'expression de la protéine GFP est 1,4 fois plus élevée chez les souris injectées avec l'**AAV2-PMS-02 (f)**, 1,9 fois plus élevée avec l'**AAV2-PMS-03 (F)** et 2 fois plus élevée avec l'**AAV2-PMS-04 (f)**, par rapport à l'AAV2 non modifié. Bien que ces différences ne soient pas statistiquement différentes en raison de l'hétérogénéité des données, les résultats suggèrent fortement un avantage des capsides modifiées par rapport à la capside non modifiée, pour permettre d'obtenir des niveaux d'expression plus élevés de la protéine GFP. Concernant les échantillons d'**AAV2-PMS-01 (f)** et **AAV2-PMS-01 (F)**, ils présentent quant à eux un niveau d'expression de la protéine similaire à celui obtenu avec l'AAV2 non modifié. Pour finir, le vecteurs **AAV2-PMS-03 (f)** a bien réussi à faire exprimer sont transgène mais à un niveau plus faible que l'AAV2 non modifié

Ces résultats sont encourageants et montrent que la quasi-totalité des vecteurs a permis une expression du transgène dans le foie des souris injectés. Ils concordent, de plus, avec ceux rapportant la quantité d'ARNm. Il semblerait donc qu'une fois le transgène libéré dans le noyau cellulaire (transcription et donc présence d'ARNm), son évolution ne soit plus dépendante de la configuration

initiale de la capsid. Il est donc possible de voir une certaine corrélation entre la présence de l'ARNm et l'expression protéique, à l'exception des souris S904 et S905 traitées avec l'**AAV2-PMS-02 (F)**.

Cette évaluation de l'expression de la GFP par Western-blot nous a ainsi permis de consolider l'hypothèse d'un fort potentiel de certains vecteurs modifiés, notamment les vecteurs **AAV2-PMS-03 (F)** et **AAV2-PMS-04 (f)**, pour augmenter l'expression d'un transgène thérapeutique dans le foie de souris après injection intraveineuse. Nous nous sommes par la suite intéressés à la localisation et la distribution de l'expression de la GFP au sein du foie par la réalisation de marquage immunohistochimique.

### *e) Evaluation de l'expression de la protéine GFP par immunohistochimie*

Afin d'avoir une vision de la répartition de l'expression de la GFP dans le foie, un immunomarquage de la protéine GFP, suivi d'une analyse par microscopie à fluorescence, ont été réalisés sur des coupes histologiques obtenues à partir des foies des souris injectées lors de cette étude. Cette analyse, effectuée par la plateforme APEX à Nantes, a permis d'évaluer l'homogénéité ou non de la transduction du foie par les différents vecteurs modifiés, mais également d'identifier le type de cellules transduites.

Cette technique consiste à détecter la présence de la GFP à l'aide d'un anticorps fluorescent après incubation des coupes des différents échantillons de foie. L'analyse des images permet ensuite de quantifier le pourcentage de cellules transduites exprimant la GFP et de renseigner la distribution de l'expression de la GFP dans le foie. Une partie des coupes de foie analysées est présentée ci-après (**figure 133**), une image caractéristique de chaque groupe expérimental ayant été sélectionnée.

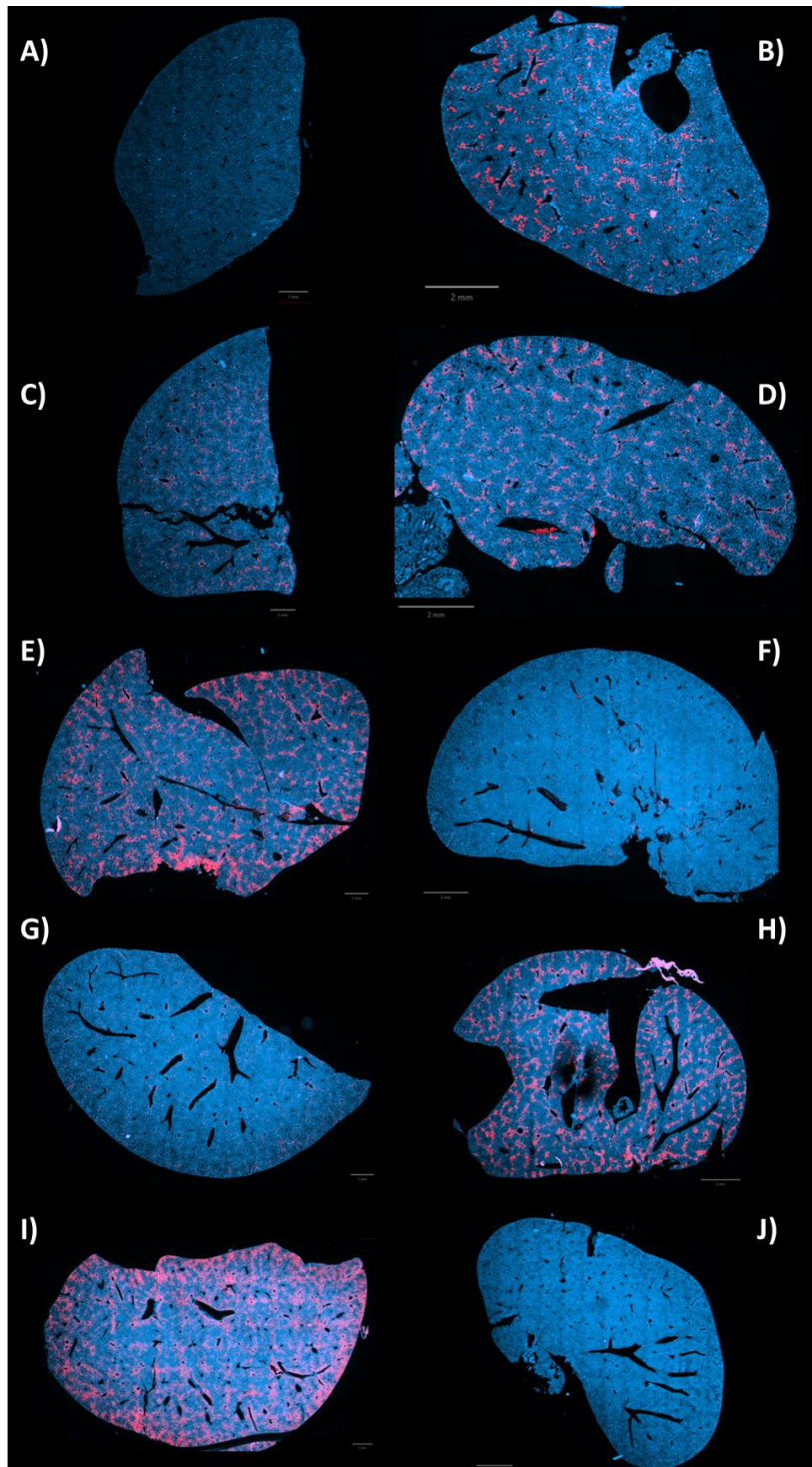


Figure 133 : Marquages immunohistochimiques sur coupe de foie, pour les souris injectées avec : A) Tampon, B) AAV2, C) AAV2-PMS-01 (f), D) AAV2-PMS-01 (F), E) AAV2-PMS-02 (f), F) AAV2-PMS-02 (F), G) AAV2-PMS-03 (f), H) AAV2-PMS-03 (F), I) AAV2-PMS-04 (f), J) AAV2-PMS-04 (F). Le noyau des cellules est coloré au DAPI (bleu), les cellules positives à la GFP apparaissent en rouge.



Il est facilement visible sur les coupes histologiques que les foies des souris injectées avec les vecteurs **AAV2-PMS-02 (f)** (identifié E sur la figure), **AAV2-PMS-03 (F)** (identifié H sur la figure) et **AAV2-PMS-04 (f)** (identifié I sur la figure) présentent une surface de marquage GFP positif supérieure aux autres groupes. Cette observation témoigne de la présence d'un plus grand nombre de cellules GFP positives, qui semblent également présenter un marquage GFP plus intense. Ces résultats sont en corrélation avec ceux de la semi-quantification de l'expression de la protéine GFP par Western-blot. On retrouve un résultat similaire entre l'AAV2 non modifié (identifié B sur la figure) et les vecteurs **AAV2-PMS-01 (f)** (identifié C sur la figure) et **AAV2-PMS-01 (F)** (identifié D sur la figure) par les deux analyses effectuées. Pour les souris injectées avec les vecteurs **AAV2-PMS-02 (F)** (identifié F sur la figure), **AAV2-PMS-03 (f)** (identifié G sur la figure) et **AAV2-PMS-04 (F)** (identifié J sur la figure) aucun marquage des cellules n'est détecté, ce qui est également cohérent avec les résultats précédents.

Pour analyser plus finement les images, un zoom a été effectué pour les coupes des souris injectées avec les vecteurs AAV2 non modifié, **AAV2-PMS-02 (f)**, **AAV2-PMS-03 (F)** et **AAV2-PMS-04 (f)** (figure 134). Les cellules transduites sont en très grande majorité des hépatocytes, reconnaissables par leur forme. De manière générale, les cellules exprimant la GFP sont observées à proximité des espaces portes. Pour les souris injectées avec l'AAV2 non modifié on remarque également quelques amas de cellules GFP positives au niveau des veines centro-lobulaires. En revanche, pour les souris injectées avec les vecteurs **AAV2-PMS-02 (f)**, **AAV2-PMS-03 (F)** et **AAV2-PMS-04 (f)** l'observation de ces images et de l'ensemble de l'échantillon marqué semble montrer une distribution plus large avec des cellules positives à la GFP regroupées autour des espaces portes, notamment la veine porte, mais également dans tout le parenchyme hépatique jusqu'à la veine centro-lobulaire, ce constat semble plus visible pour les échantillons traité avec l'**AAV2-PMS-02 (f)**, et dans une moindre mesure pour **AAV2-PMS-04 (F)**, (observations réalisées par Thibaut Larcher, vétérinaire anatomopathologiste au sein de la plateforme APEX). Ce résultat est encourageant, on pourrait envisager que le ligand induise une meilleure distribution de l'AAVr dans le parenchyme hépatique pouvant expliquer pourquoi l'expression protéique est plus importante au sein de cet organe avec les vecteurs modifiés.

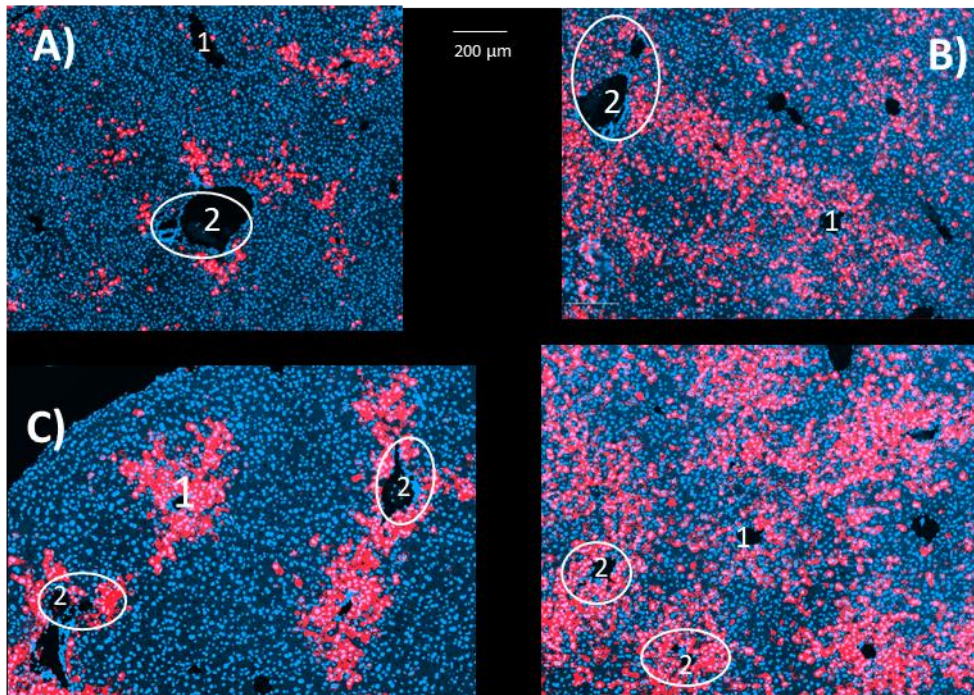


Figure 134 : Zoom de marquages immunohistochimiques d'une coupe de foie des souris injectées avec A) AAV2, B) **AAV2-PMS-02 (f)**, C) **AAV2-PMS-03 (F)**, D) **AAV2-PMS-04 (f)**. Le noyau des cellules est coloré au DAPI (bleu), les cellules positives à la GFP apparaissent en rouge. 1 veine centro-lobulaire, l'espace porte est entouré en blanc avec en 2 la veine porte.

L'utilisation du logiciel QuPath a établi une moyenne du pourcentage de cellules GFP positives pour une coupe de foie par animal. Ces valeurs, présentées en (figure 135), ont été obtenues après comptage de toute les cellules positives à la GFP dans trois zones aléatoires au sein d'une coupe, et pour toutes les souris de l'étude.

Ici encore, aucune cellule exprimant la GFP n'est observée pour l'ensemble des animaux injectés avec le tampon de formulation dPBS et les vecteurs **AAV2-PMS-02 (F)** et **AAV2-PMS-04 (F)**, confirmant l'absence de transduction. De la même manière, les animaux ayant reçu l'**AAV2-PMS-02 (f)**, l'**AAV2-PMS-03 (F)** et l'**AAV2-PMS-04 (f)** possèdent un pourcentage de cellules GFP positives environ 3 fois plus important que ceux traités avec l'AAV2 non modifié. Les échantillons traités avec l'**AAV2-PMS-01 (f)** et **(F)** montrent un résultat similaire à celui de l'AAV2. De la même manière que précédemment l'**AAV2-PMS-03 (f)** présente un pourcentage de cellules GFP plus faible que l'AAV2 non modifié.

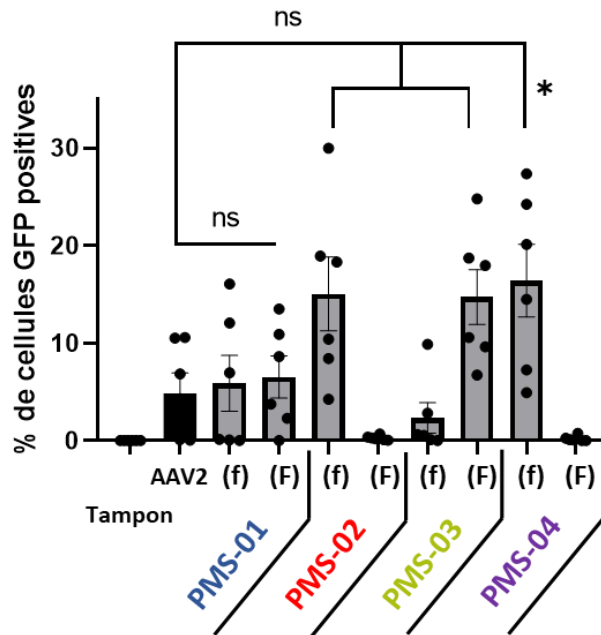


Figure 135 : Pourcentage de cellules hépatiques positives à la GFP dans le foie des souris injectées ( $n = 6$ ) avec les différents vecteurs de l'étude *in vivo*. Les données sont présentées sous forme de moyenne  $\pm$  SEM. Les statistiques ont été réalisées à l'aide d'un test Mann-withney non paramétrique. Les valeurs P sont \* $p < 0,05$  ont été considérées comme statistiquement significatives. Les valeurs respectives sont de ns : 0,927, 0,5887, 0,0931, 0,0649 et 0,0411.

#### D. Quantification de l'ARN messager GFP dans les organes non ciblés des souris injectées avec les différents vecteurs AAV2

Au regard de l'effet sur l'efficacité du transfert de gène avec certaines modifications chimiques dans le foie, il nous a paru important d'investiguer l'expression de l'ARNm GFP dans les autres organes que le foie. Notre choix d'analyse s'est porté sur la quantification de l'ARNm GFP par RT-qPCR et non pas sur une quantification des génomes viraux (vg/dg) par qPCR. En effet, l'analyse des ARNm par RT-qPCR s'est avérée être en corrélation avec les données des analyses protéiques par Western-blot et immunohistochimie pour les échantillons de foie.

Les analyses ont donc été réalisées, comme décrit précédemment, sur les échantillons de cœur, poumon, rein, quadriceps et rate de tous les animaux inclus dans l'étude.

L'ensemble des données, présenté sur la (figure 136), indique qu'il n'y a pas de différences significatives d'expression du messager GFP entre les souris injectées avec les vecteurs modifiés et non modifiés dans les échantillons analysés de cœur, de poumon, de rein, de quadriceps et de rate. Contrairement au foie, les modifications chimiques des AAV2 testées dans cette étude *in vivo* ne semblent donc pas

avoir impacté les propriétés de l'AAV2 pour le transfert de gène dans ces autres tissus. Nos résultats nous laissent donc supposer que les ligands sélectionnés dans cette étude sont bien spécifiques du foie et permettent de moduler l'efficacité du transfert de gène dans le foie, après bioconjugaison à l'AAV2. Pour s'assurer de cette spécificité, il faudrait renouveler l'expérience avec des doses de vecteur plus importantes. En effet, cela permettrait de savoir si la transduction des vecteurs modifiés pourrait s'étendre à d'autres organes que le foie, et s'ils présentent un profil de transduction différent de l'AAV2 non modifié. Il serait également important de prélever un panel plus large d'organes pour renseigner un profil d'expression plus exhaustif des vecteurs modifiés.

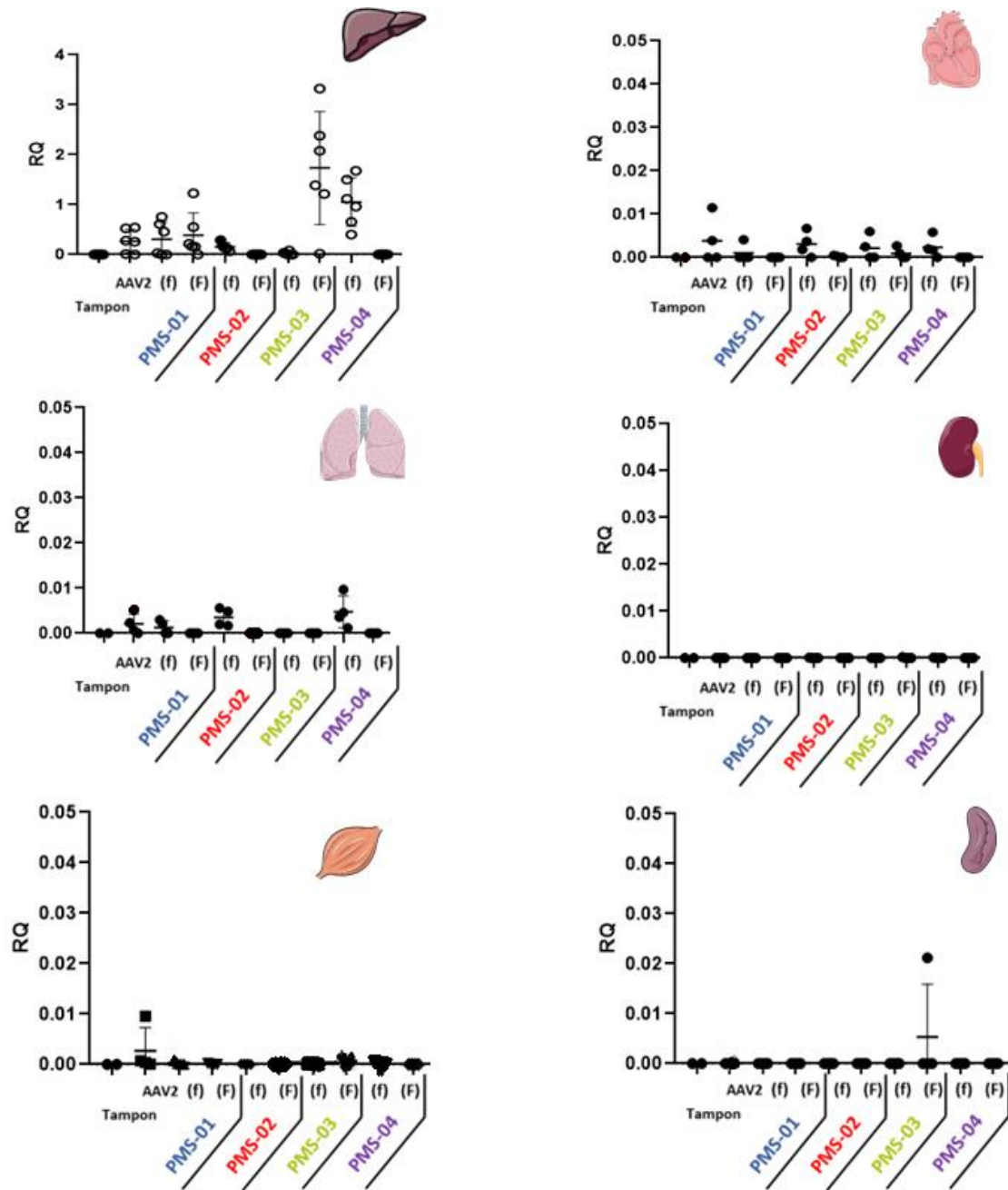


Figure 136 : Quantification relative par RT-qPCR des messagers GFP dans les différents organes prélevés (foie, cœur, poumons, reins, quadriceps et rate) des souris injectées ( $n = 6$ ) avec les différents vecteurs de l'étude in vivo. Les données sont présentées sous forme de moyenne  $\pm$  SEM avec une limite de quantification (LOQ = 0,0004).

## 5. Conclusion

Pour conclure ce chapitre, les ligands **(PMS-01)**, **(PMS-02)**, **(PMS-03)** et **(PMS-04)** ont été bioconjugués avec succès à la surface de l'AAV2. Les différentes caractérisations de ces nouvelles capsides ont montré qu'il s'agissait d'un couplage covalent et non pas d'une absorption du ligand à la surface de l'AAV2. Les analyses biochimiques ont montré qu'une fois greffés les dérivés du GalNAc sont toujours accessibles et peuvent donc être reconnus par des lectines ainsi que par le domaine de reconnaissance de l'ASGPr.

L'efficacité *in vitro* de ces vecteurs a été testée sur différentes lignées cellulaires. Dans un premier temps, des cellules HEK293 ont été transduites afin de vérifier l'infectiosité des particules, puis des tests de transductions ont été menés sur des lignées hépatocytaires, dont une primaire murin. Cette dernière analyse a montré une meilleure efficacité des vecteurs modifiés chimiquement avec un pourcentage de cellules GFP positives plus important que pour l'AAV2 non modifié. Ainsi, à la suite de ces résultats encourageants, une étude *in vivo* a été entreprise.

Durant l'étude *in vivo*, soixante souris ont été injectées avec six souris par groupe (tampon de formulation de l'AAV, l'AAV2, et les quatre nouveaux vecteurs modifiés aux deux concentrations). Un mois après l'injection, les animaux ont été euthanasiés pour réaliser les prélèvements nécessaires aux analyses post-mortem. Différentes techniques de biologie moléculaire ont été réalisées : (i) la quantification du nombre de copies de génomes viraux par cellule (vg/dg) dans le foie par qPCR, (ii) la quantification relative de l'ARNm GFP (RQ) dans tous les organes prélevés par RT-qPCR. Le rapport d'infectiosité a été établi pour chaque souris entre la quantification de l'ARNm GFP (RQ) et le nombre de vecteur par cellule (vg/dg), pour voir si les échantillons avec une forte quantité de transgène GFP présentent aussi une quantité importante d'ARNm. L'expression de la protéine GFP dans le foie a été étudiée par semi-quantification par Western-blot. Enfin, des coupes de foie ont été analysées à la suite d'un marquage immunohistochimique pour quantifier le pourcentage de cellules GFP positives dans le foie.

La mise en parallèle des résultats des différentes analyses nous a permis de tirer plusieurs conclusions de l'étude *in vivo*. Il semblerait, que le vecteur **AAV2-PMS-01** modifié aux deux concentrations se comporte de la même manière que l'AAV2 puisque les résultats d'expression de l'ARNm, de la protéine GFP et le pourcentage de cellules GFP positives sont similaires. La seule différence réside dans le nombre de copie de vecteur par cellule (vg/dg), bien plus faible pour les souris traitées avec **l'AAV2-PMS-01 (F)**. Ces résultats sont en accord avec les données obtenues par la société Glycodiag<sup>®</sup> qui montraient que l'affinité des vecteurs **AAV2-PMS-01 (f)** et **(F)** pour l'ASGPr était très faible (**figure 121,**

[page 132](#)). Cependant, ils sont en contradiction avec ceux obtenus après transduction des hépatocytes primaires murins. Il faut tout de même prendre un certain recul sur cette analyse au vu de la stabilité de cette lignée cellulaire. Il semblerait donc que le ligand **(PMS-01)** ne permette pas, *in vivo*, une meilleure efficacité de transduction de l'AAV2 pour le foie.

Concernant les vecteurs **AAV2-PMS-02 (F)** et **AAV2-PMS-04 (F)**, nos différentes analyses n'ont pas permis de détecter les vg, l'ARNm et la protéine GFP dans le foie et les six autres organes analysés. Fait intéressant, leurs homologues modifiés avec la faible concentration en ligand, les **AAV2-PMS-02 (f)** et **AAV2-PMS-04 (f)**, ont montré des taux d'expression de la protéine d'intérêt et des pourcentages de cellules positives à la GFP plus importants qu'avec l'AAV2 non modifié. La quantité de ligands couplés à la surface de l'AAV2 semble donc avoir un impact sur l'efficacité de transfert de gène de la particule, au moins pour les ligands **(PMS-02)** et **(PMS-04)**. En effet, le phénomène inverse est observé avec les vecteurs **AAV2-PMS-03 (f)** et **AAV2-PMS-03 (F)**. La quantité de ligand à coupler à la capsid de l'AAV dépendrait donc de la nature même du ligand, hypothétiquement de son affinité aux cellules ciblées et/ou de son immunogénicité. En effet, il est possible qu'une forte bioconjugaison des ligands **(PMS-02)** et **(PMS-04)** puissent conduire à une réponse immunitaire innée ou adaptative, induisant une neutralisation des vecteurs par l'organisme, une autre hypothèse serait l'incapacité de ces vecteurs à s'internaliser dans la cellule et donc à son élimination. Une analyse possible serait de réaliser un ELISPOT avec les splénocytes isolés des rates des différentes souris afin de voir une possible réponse immunitaire adaptative contre les vecteurs. Il serait également possible, dans une future étude *in vivo*, de sacrifier certaines souris 7 jours après l'injection des vecteurs afin d'observer une possible présence de génome dans le foie avant activation de la réponse immunitaire adaptative. A l'inverse, le ligand **(PMS-03)** pourrait être moins immunogène, mais également moins affiné pour le foie, ce qui expliquerait la nécessité d'une plus grande quantité pour obtenir des résultats similaires à **(PMS-02/04)**. Les résultats de la société Glycodiag<sup>®</sup> ([figure 121, page 132](#)) semblent aller dans ce sens et montrent que les vecteurs modifiés avec **(PMS-02)** et **(PMS-04)** sont plus affins que ceux modifiés avec **(PMS-03)**.

Au final, l'étude *in vivo* aura permis d'identifier 3 candidats de vecteurs chimiquement modifiés présentant un meilleur potentiel pour le transfert de gène dans le foie par rapport à l'AAV2 non modifié. Il s'agit des vecteurs **AAV2-PMS-02 (f)**, **AAV2-PMS-03 (F)** et **AAV2-PMS-04 (f)**. Le [tableau 9](#) regroupe les différents résultats obtenus pour ces nouveaux vecteurs ainsi que l'AAV2.

	AAV2 non modifié	AAV2-PMS-02 (f)	AAV2-PMS-03 (F)	AAV2-PMS-04 (f)
Quantification des génomes de vecteur dans le foie	0,641	0,355	0,106	0,047
Quantification relative de l'ARN transgénique dans le foie	0,270	0,528	1,735	1,052
Quantification de la GFP	70	101	129	140
% cellules exprimant la GFP	5	15	15	16
Analyse visuelle des coupes	Répartition globalement autour de la veine porte	Distribution forte entre espace porte et veine porte	Répartition globalement de la veine porte et veine centro-lobulaire	Distribution moyenne entre espace porte et veine porte

Tableau 8 : Comparaison des résultats des analyses de l'étude in vivo entre les différents vecteurs modifiés et l'AAV2.

Ce tableau illustre bien les différences entre les vecteurs, la première chose intéressante à souligner est le rapport entre le pourcentage de cellules GFP positives et la quantité de génome. Par exemple les échantillons traités avec l'**AAV-PMS-04 (f)** présentent une quantité de génome vecteur par noyaux très faible (0,047 vg/dg) alors que 16% des hépatocytes produisent de la GFP. Cette différence est frappante en comparaison avec l'AAV2 qui présente un taux de génome par noyaux de 0,641 vg/dg mais que 5% de cellules positives à la GFP. Il semblerait donc que l'**AAV-PMS-04 (f)** a plus de difficulté à s'internaliser dans la cellule mais qu'une fois à l'intérieur il a une efficacité de transduction bien plus efficace, ce constat est similaire pour l'**AAV-PMS-03 (F)** mais moins flagrant pour l'**AAV-PMS-02 (f)**. Il serait donc intéressant de réaliser une étude de « trafficking » intra cellulaire afin de voir le comportement des vecteurs modifiés en comparaison à l'AAV2. L'analyse des lames après le marquage immunohistochimique a semblé montrer une répartition homogène des cellules transduites au sein de l'organe. Cependant, les vecteurs **AAV-PMS-02 (f)** (mieux visible) et **AAV-PMS-04 (f)** semble avoir transduit des cellules aux alentours de la veine porte et la veine centro-lobulaire mais également dans le parenchyme hépatique, on peut émettre l'hypothèse que ces vecteurs puissent mieux évoluer au sein des lobules du foie possiblement dans les travées de Remak.



Enfin, ces trois nouveaux vecteurs valident la stratégie de modification chimique des capsides d'AAV en montrant que la bioconjugaison d'un ligand à la surface d'un AAV peut-être une alternative intéressante pour augmenter sa capacité de transfert de gène et ainsi répondre aux problématiques que rencontrent actuellement le vecteur AAVr. A l'avenir, une seconde étude *in vivo* pourrait être envisagée en augmentant le nombre de souris pour mieux apprécier la variabilité et en augmentant également la dose injectée pour observer la bio-distribution et l'immunogénicité. Enfin, il serait intéressant de générer de nouveaux vecteurs avec les ligands **(PMS-02)**, **(PMS-03)** et **(PMS-04)** mais avec un sérotype d'AAVr différent, le sérotype ayant le tropisme le plus fort pour le foie comme l'AAV8 ou l'AAV5 qui lui présente une immunogénicité plus faible que ses congénères.<sup>116</sup>

# *Conclusion et Perspective*



## Conclusion

Les travaux exposés dans ce manuscrit mettent en évidence la richesse et la complexité de la fonctionnalisation chimique des capsides de virus adéno-associés pour la thérapie génique.

Sur la base du savoir-faire et de l'expérience des deux laboratoires impliqués dans ce projet (laboratoires CEISAM et TaRGeT) l'objectif de la thèse était de synthétiser des ligands originaux afin de les greffer sur des AAV2 pour cibler le foie. Le but était d'augmenter la transduction des hépatocytes tout en préservant les AAV des anticorps circulants pour, *in fine*, avoir en main un nouveau vecteur capable à l'avenir de transférer, à dose moindre, un gène médicament à des patients atteints d'une maladie génétique du foie.

Le premier chapitre de la thèse décrit les principes et enjeux de la thérapie génique, l'utilisation de vecteurs AAVr ainsi que leurs limites et leurs optimisations notamment la modification chimique. Nous avons enchaîné sur une présentation des réactions de bioconjugaison et du choix de la fonction de couplage amino présente sur les lysines de la capsidie pour conclure sur une description du foie et des cellules hépatocytaires.

Le second chapitre porte sur la synthèse de ligands que l'on peut décomposer en trois parties : (i) une fonction phénylthiocyanate permettant le greffage sur les lysines ; (ii) un espaceur triéthylène glycol et (iii) un motif glycosidique afin de cibler les récepteurs aux asialoglycoprotéines présents sur les hépatocytes. C'est essentiellement sur cette dernière partie que nous avons porté nos efforts en préparant des sucres dérivés de la *N*-acétylgalactosamine (ligand de référence). Dans ce chapitre nous développons également des travaux préliminaires de docking moléculaire sur une librairie de 70 composés afin de sélectionner les sucres potentiellement les plus intéressants. Malheureusement, bien que les différentes interactions ligands/récepteur coïncident avec la littérature, le champ de force utilisé ne prend pas en compte le calcium présent dans la poche lectinique. Il est donc impossible de bien décrire l'affinité des composés testés envers le récepteur et de ressortir une structure glycosidique que nous pourrions synthétiser. Malgré plusieurs échecs, comme notamment la synthèse de **(PMS-02bis)**, trois nouveaux ligands ont été obtenus avec succès. Le ligand **(PMS-02)** qui possède comme motif de reconnaissance un sucre bicyclique ponté décrit comme étant particulièrement affiné pour l'ASGPr a été synthétisé en 17 étapes. Les ligands **(PMS-03)** et **(PMS-04)** sont similaires avec un motif GalNAc substitué en position anomérique par un méthoxy respectivement en configuration  $\beta$  et  $\alpha$ .

Pour finir, ces trois ligands ont fait l'objet d'une étude biologique détaillée dans le troisième chapitre et réalisée au laboratoire TaRGeT. Dans un premier temps, des couplages ont été effectués entre des

AAV2 et les ligands à pH basique dans un tampon TBS. Le greffage covalent, et non pas une adsorption, des ligands à la surface des capsides a alors été validé par différentes analyses de biologie moléculaire, comme le Western-blot ou le Dot-blot. Dans un deuxième temps, l'impact de la reconnaissance d'un anticorps A20 avec la capside assemblée a été menée par un test ELISA. Nous avons également regardé le profil d'affinité des AAV modifiés pour diverses lectines en faisant appel à l'entreprise Glycodiag® spécialisée dans le criblage de glyco-composés. Toujours dans ce chapitre, nous avons évalué l'impact des ligands sur l'infectiosité des vecteurs sur des cellules HEK293 (lignée cellulaire non hépatocytaire), puis sur les lignées cellulaires hépatocytaires HUH7 et HepG2 et enfin sur des hépatocytes primaires murin. Nous avons ainsi mis en évidence que les vecteurs modifiés sont plus efficaces que l'AAV2 natif et que le nombre d'équivalent de ligand utilisé lors de l'étape de bioconjugaison joue également un rôle. Après ces résultats très encourageants, une étude *in-vivo* a été menée pour évaluer l'efficacité de transduction après injection des vecteurs dans un système complexe et essayer de définir le meilleur ligand pour des applications cliniques possibles. Soixante souris ont été injectées, euthanasiées après 1 mois puis les organes, dont le foie, ont été prélevés. Les différentes analyses biomoléculaires et histologiques ont mis en évidence l'intérêt des capsides modifiés avec une meilleure transduction et une meilleure expression du gène dans le foie comparativement à l'AAV2. De ce travail, trois vecteurs ont pu ressortir, l'**AAV-PMS-02 (f)**, l'**AAV-PMS-03 (F)** et l'**AAV-PMS-04 (f)**.

## Perspectives :

Les perspectives de ce travail de thèse peuvent se diviser en 3 axes allant de la synthèse organique à une étude *in-vivo* en passant par la caractérisation. En ce qui concerne le volet chimie organique, nous pourrions poursuivre la synthèse de nouvelles molécules plus affines pour le récepteur ASGPr. On y retrouve une nouvelle voie de synthèse pour le ligand (**PMS-02bis**), ainsi que la confection de ligand multivalent à forte affinité pour le récepteur. Le deuxième axe d'amélioration est la caractérisation des vecteurs après bioconjugaison et notamment le nombre de ligands greffé, et le troisième concernera une seconde étude *in vivo* pour répondre aux questions d'immunogénicité et de biodistribution des vecteurs. L'étude de nouveaux sérotypes comme l'AAV8 et/ou l'AAV5 est aussi une perspective indispensable.

### A. Axe 1 : synthèse organique

#### a) *Poursuite de la synthèse de (PMS-02bis)*

Une perspective intéressante serait de continuer la synthèse du ligands (**PMS-02bis**) qui n'a pu être terminée. Nous avons isolé le composé (**31**) mais avec un rendement trop faible pour poursuivre la synthèse. Une alternative pour obtenir le composé l' $\alpha$ -D-C-galactopyranosyl méthanol (**30**) d'intérêt est présenté (figure 137). Celle-ci, démarrée en fin de thèse, commence par une réaction de C-glycosylation du composé (**27**) afin d'obtenir l'alcyne protégé (**B**) après activation de la position anomérique par un fluor (composé **A**). Après déprotection et oxydation de l'alcyne nous devrions générer l'acide carboxylique (**C**) qui, une fois réduit, donnera le composé d'intérêt (**30**) (figure 137).

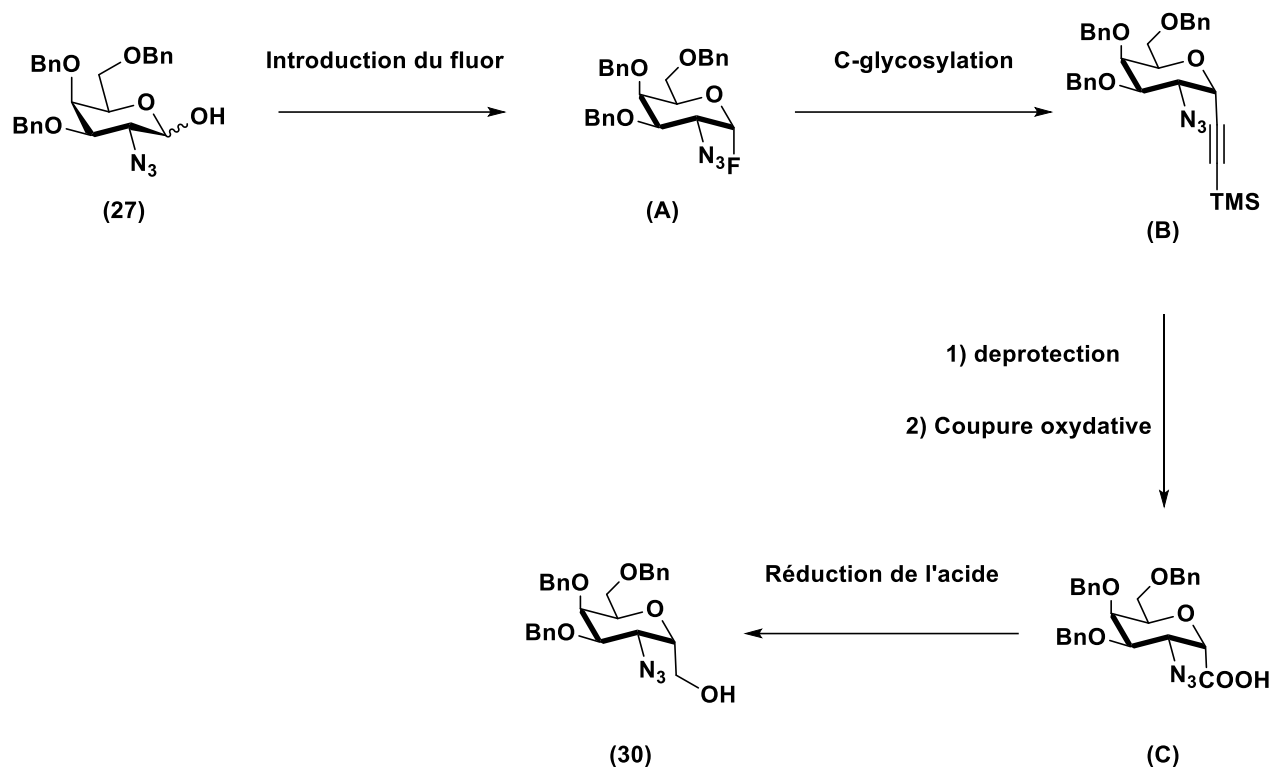


Figure 137 : Obtention de l' $\alpha$ -D-C-galactopyranosyl méthanol d'intérêt selon la voie E.

La première étape de cette voie consiste à traiter le sucre **(27)** par un donneur de fluor, le trifluoro diéthylaminosulfure dans le dichlorométhane à  $-40^{\circ}\text{C}$  pendant 2h. Le composé **(A)** est obtenu après purification sur gel de silice (Cy 20% - AcOEt 80%) avec un rendement de 93% (figure 138). La réaction de glycosylation se fait dans le dichlorométhane parfaitement anhydre par addition de triméthylsilyléthyne fraîchement distillé en présence de diéthyléthérate de trifluorure de bore à  $0^{\circ}\text{C}$ . Le composé **(B)** est récupéré avec un rendement moyen de 38%.

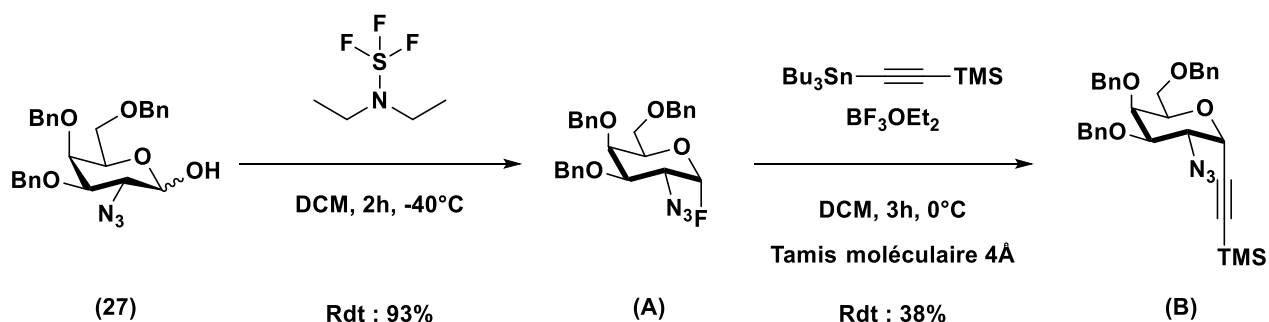


Figure 138 : Réaction d'activation par fluoration du composé **(27)** suivie d'une réaction de C-glycosylation.

Il reste maintenant à réaliser les réactions de déprotection puis de coupure oxydative qui devraient permettre d'obtenir le composé **(C)**. Il sera ensuite nécessaire de le réduire afin d'avoir le dérivé **(31)** d'intérêt, une activation préliminaire de l'acide carboxylique sera peut-être nécessaire pour mener à bien cette réaction.<sup>181</sup>

### b) Synthèse de ligands multivalents

Enfin, une perspective très intéressante serait l'exploitation de l'effet de multivalence qui nous apparaît comme une alternative thérapeutique très prometteuse en raison de sa simplicité et de son efficacité. En effet, le fait d'avoir de nombreuses copies d'un ligand sur une même charpente permet de générer des interactions simultanées avec les sites de reconnaissance et peut entraîner un gain considérable d'affinité qui peut être très largement supérieur à l'addition des interactions individuelles.<sup>182,183</sup> De plus, greffer des ligands multivalents et donc volumineux sur un AAVr peut induire une protection de la capsid vis-à-vis des anticorps neutralisants ou cellules phagocytaires présents dans l'organisme en jouant un rôle de bouclier. Nous pourrions envisager des ligands avec trois motifs GalNAc ou analogues monomérique générés lors de la thèse à partir d'une structure tris(hydroxyméthyl)aminométhane (Tris) (figure 139).

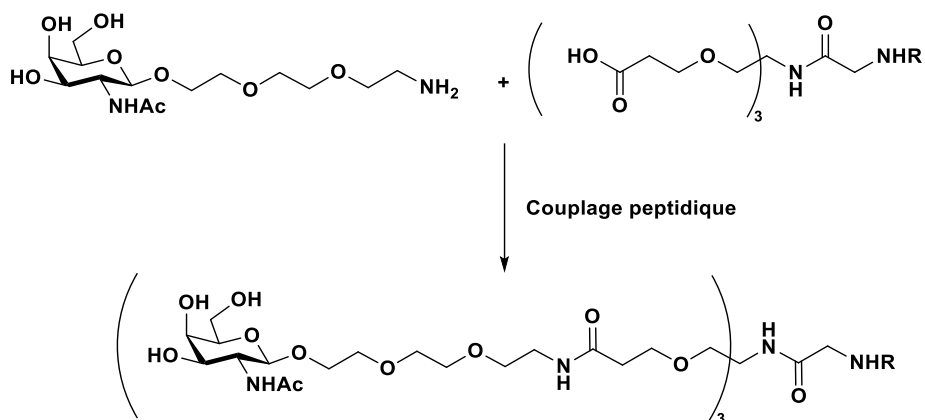


Figure 139 : Couplage peptidique entre le sucre et le dérivé tris aboutissant au ligand multivalent de Biessen.<sup>184</sup>

Comme évoqué dans le chapitre 1, cette stratégie est étayée par le fait que le l'ASGPr présente trois domaines de reconnaissance des sucres. De plus, dans la littérature, il est décrit que des molécules avec trois unités de GalNAc sont plus affines que leurs homologues di-ou mono-valentes.<sup>185</sup> L'enjeu serait donc de préparer des molécules multivalentes présentant un, deux ou trois motifs osidiques



avec exactement la même charpente et d'évaluer l'activité biologique des vecteurs d'AAV modifiés correspondant et l'éventuel effet multivalent (**figure 140**).

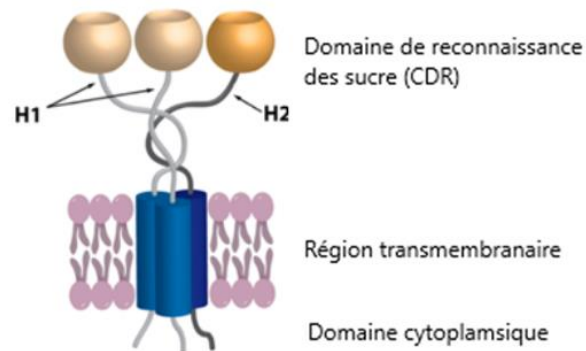


Figure 140 : Structure de l'ASGPr

L'étude et la conception de ligand multivalent pour l'ASGPr ont largement été étudiées dans la littérature notamment par l'équipe du professeur Castagner.<sup>186</sup> Plusieurs critères ont été définis tels que la longueur des bras espaceurs, la balance hydrophile/hydrophobe ou la géométrie spatiale de la charpente. La géométrie la mieux adaptée s'est avérée être un triangle avec des angles de  $108^\circ$  et des longueurs de côté compris entre 19 et 32 Å (**figure 141**).<sup>187</sup>

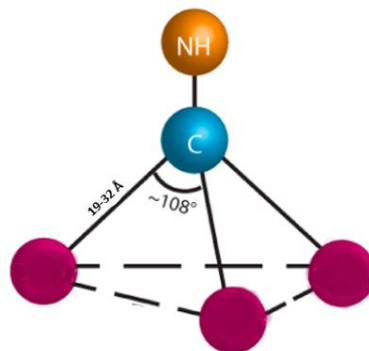


Figure 141 : Représentation d'un ligand trivalent optimisé pour l'ASGPr.

Au vu des résultats biologiques, il serait judicieux de concevoir des ligands multivalents avec la structure (**PMS-03**) car l'étude *in vivo* a montré que les AAV2 modifiés à la forte concentration en ligands (**PMS-02**) et (**PMS-04**) ne sont plus efficaces. Cela suggère qu'augmenter d'avantage le nombre de ligands ne devrait qu'amplifier ce phénomène. A contrario, les particules modifiées avec la forte

concentration (**PMS-03**) sont plus efficaces qu'avec la faible concentration. On peut donc imaginer qu'apporter un ligand multivalent pourrait rendre le vecteur encore plus efficace.

Notre stratégie pour concevoir les charpentes mono, di et trimériques est de partir de la molécule (**37**) que nous ferions réagir par couplage peptidique sur des analogues du Tris (**figure 142**) afin de les greffer ultérieurement sur les AAV.

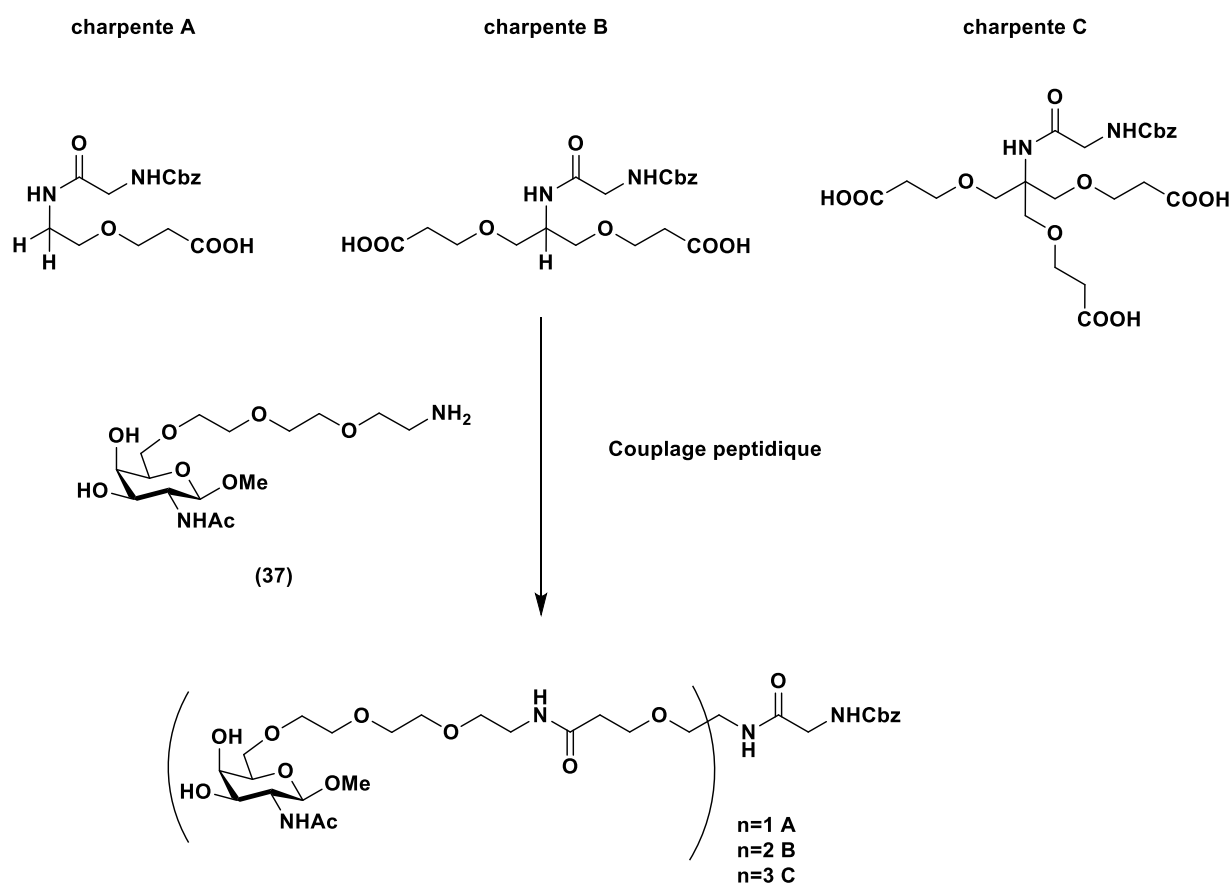


Figure 142 : Stratégie de synthèse des molécules multivalentes.

## B. Axe 2 : caractérisations

Tout d'abord il serait nécessaire de déterminer précisément le nombre de ligands greffés à la surface du vecteur. En effet, il a été montré durant la thèse que la concentration en ligands lors de l'étape de bioconjugaison influence les propriétés biologiques du vecteur et parfois même de façon importante. On peut citer l'exemple de l'**AAV2-PMS-04 (f)** qui est un des meilleurs vecteurs de l'étude *in vivo* et l'**AAV2-PMS-04 (F)** qui par contre ne présente aucune transduction dans le foie. Ainsi, la méthode de choix est la chromatographie HPLC couplée à la spectrométrie de masse (LC-HRMS) qui pourra

déterminer de façon précise le nombre de ligands couplés sur chacune des trois protéines constituant la capsid de l'AAV. Nous pourrions réaliser une cartographie peptidique *in silico* de chaque protéine virale composant la capsid afin d'étudier quel fragment peptidique est impacté par les modifications chimiques. Pour détecter avec précision quel résidu lysine a été modifié, nous pourrions faire de la protéolyse des protéines de la capsid des AAV par digestion enzymatique (trypsine ou autre). Il pourrait également être intéressant de réaliser de la cryo-microscopie électronique des vecteurs modifiés chimiquement pour déterminer la structure de la capsid et avoir des informations sur l'arrangement des ligands à sa surface.

Il serait également intéressant de poursuivre la collaboration avec la société Glycodiag<sup>®</sup>. Pour rappel, lors de la caractérisation de l'affinité des particules pour le récepteur, les plafonds des courbes d'absorbances n'ont pas été éteints ce qui rend impossible la détermination de la constante d'affinité ligand-récepteur. En outre, il serait aussi intéressant de s'affranchir de l'anticorps A20 qui induit un biais dans la comparaison entre les AAVr modifiés avec les fortes concentrations et les faibles concentrations. Ainsi, la résolution de ces deux problématiques nous permettrait d'obtenir des constantes d'affinités pour les vecteurs modifiés et pouvoir ainsi mieux corrélérer leur affinité pour l'ASGPr avec leur capacité de transduction des hépatocytes.

### C. Axe 3 : étude biologique

Une perspective possible concernant l'aspect biologique serait de générer de nouveaux vecteurs à partir du sérotype 8 de l'AAV qui est transcriptionnellement plus efficace que l'AAV2 dans le foie.<sup>188</sup> Il serait également intéressant d'utiliser le sérotype 5 qui est moins immunogène et qui est actuellement utilisé comme vecteur dans les deux médicaments récemment mis sur le marché (Roctavian<sup>®</sup> et Hemgenix<sup>®</sup>).<sup>75</sup>

Il serait également intéressant d'utiliser un autre transgène que celui de la GFP afin de simplifier l'étude *in-vivo*. Nous pourrions générer des AAV avec un transgène codant pour la luciférase qui catalyse une réaction de bioluminescence en présence de luciférine avec émission d'un photon dont la lumière résultante est jaune-vert.<sup>189</sup> Il sera alors possible de suivre la transduction des tissus chez les souris traités après injection en IV de luciférine. .

Pour finir, il faudrait également prendre en considération la nature du transgène. Au sein du groupe une équipe s'intéresse aux Itr de l'AAV et notamment à leur optimisation afin d'améliorer d'une part

la production de vecteur et d'autre part de limiter leur caractère immunogène. Il est probable que les futur vecteurs d'AAVr soit le résultat d'une optimisation globale incluant la capsid et le transgène.

En combinant la chimie et la vectorologie, il est ainsi possible de concevoir de nouvelles capsides chimiquement modifiées qui peuvent être une réponse aux problèmes de ciblage et d'immunogénicité, ou du moins être une des clés pour leurs résolutions. Il est évident que ces travaux préliminaires n'apportent que des premiers éléments de réponse à un problème complexe mais fondamentale pour aboutir à l'amélioration de l'efficacité de transduction de l'AAV.



# *Materials and methods*



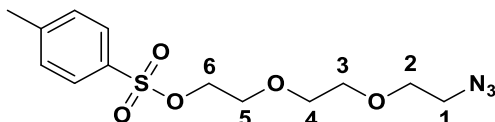
## 1. Chemistry

All chemicals reagents were purchased from Acros Organics, Sigma Aldrich or Carbosynth and were used without further purification. Reactions requiring anhydrous conditions were performed under nitrogen atmosphere. All chemically synthesized products were purified by column chromatography on silica gel Kieselgel SI60 (40–63 mm) from Merck or were purified by flash chromatography (GRACE REVELERIS Flash Chromatography System) equipped with UV and DLS detectors. All chemically synthesized compounds were characterized by  $^1\text{H}$  (400.133 or 300.135 MHz),  $^{13}\text{C}$  (125.773 or 75.480 MHz) NMR spectroscopy (Bruker Avance 300 Ultra Shield or Bruker Avance III 400 spectrometer). Chemical shifts are reported in parts per million (ppm); coupling constants are reported in units of Hertz [Hz]. The following abbreviations were used: s = singlet, d = doublet, t = triplet, q = quartet, quin = quintet, br = broad singlet, m = multiplet. When needed,  $^{13}\text{C}$  heteronuclear HMQC and HMBC were used to unambiguously establish structures. High resolution mass spectra (HRMS) were recorded with a Thermofisher hybrid LTQ-orbitrap spectrometer (ESI+) and a Bruker Autoflex III SmartBeam spectrometer (MALDI). HPLC analyses were performed on an HPLC autopurification system (WATERS) equipped with BGM 2545 binary pump, a 2767 Sample Manager, and an UV-visible diode array detector and evaporative light scattering detector in series (WATERS 996 PDA and WATERS 2424 ESLD). The  $\alpha/\beta$  ratio was determined by semi-quantification after integration of anomeric protons by  $^1\text{H}$  NMR.



## A. Peg linker

### Compound (PEG-1)



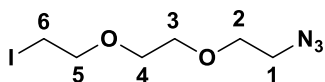
To a solution of 2-(2-(2-azidoethoxy)ethoxy)ethan-1-ol (3 g, 17.2 mmol) in THF (86 mL) was added, at 0°C under N<sub>2</sub> atmosphere, p-toluenesulfonyl chloride (13 g, 69 mmol), DMAP (210 mg, 1.72 mmol) and Et<sub>3</sub>N (20 ml, 137.6 mmol). After stirring for 16h at RT, the reaction mixture was diluted in H<sub>2</sub>O (50 ml) and extracted with DCM (3 X 30 mL), then the combined organic layers were washed with H<sub>2</sub>O (2 x 100 mL), brine, dried over anhydrous MgSO<sub>4</sub>, filtered, and concentrated under vacuum. The crude product was purified by flash chromatography on silica gel (Cy/AcOEt 80:20), to afford the desired compound (**PEG-1**) as a yellow liquid (4.40 g, 13.6 mmol, yield 79%).

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>, 298 K):  $\delta$  = 7.83-7.77 (m, 2H, CH aromatic), 7.36-7.30 (m, 2H, CH aromatic), 4.19-4.14 (m, 2H, H-6), 3.73-3.61 (m, 4H, H-2, H-5), 3.60 (s, 4H, H-3, H-4), 3.40-3.33 (m, 2H, H-1), 2.44 (s, 3H, CH<sub>3</sub>).

<sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>, 298K):  $\delta$  = 145.0 (C, aromatic), 133.1 (C, aromatic), 130.0 (2 x CH, aromatic), 128.1 (2\*CH, aromatic), 70.9, 70.7, 70.2, 69.3, 68.9 (5 x CH<sub>2</sub>, C-2, C-3, C-4, C-5, C-6), 50.8 (CH<sub>2</sub>, C-1), 21.8 (CH<sub>3</sub>).

HRMS ESI+ : m/z calcd for C<sub>13</sub>H<sub>2</sub>N<sub>3</sub>O<sub>5</sub> [M+H]<sup>+</sup>: 330.1124 found: 399.1121.

### Compound (PEG-2)



To a solution of (**PEG-1**) (4.4 g, 13.6 mmol) in DMF (82 mL) was added at RT, NaI (6 g, 40 mmol). After stirring for 16h at 50°C, the solvent was evaporated and the residue was co-evaporated 3 times with toluene. The crude product was purified by flash chromatography on silica gel (Cy/AcOEt 60:40) to afford compound (**PEG-2**) as a white solid (3.3 g, 11.8 mmol, yield 87%).

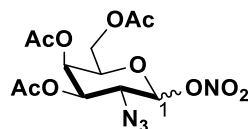
<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>, 298 K):  $\delta$  = 3.80-3.67 (m, 4H, H-2, H-5), 3.68 (s, 4H, H-3, H-4), 3.43-3.37 (m, 2H, H-1), 3.30-3.23 (m, 2H, H-6).

<sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>, 298K):  $\delta$  = 72.1, 70.8, 70.4, 70.3 (CH<sub>2</sub>, C-2, C-3, C-4, C-5), 50.8 (CH<sub>2</sub>, C-1), 3.03 (CH<sub>2</sub>, C-6).

HRMS ESI+ : m/z calcd for C<sub>6</sub>H<sub>12</sub>N<sub>3</sub>O<sub>2</sub>NaI [M+Na]<sup>+</sup>: 307.9872 found: 307.9868.

## B. Ligand PMS-02

### Compound (1)



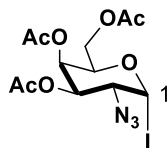
To a solution of tri-O-acetyl-D-galactal (5 g, 18.32 mmol) in dry ACN (20 mL) containing NaN<sub>3</sub> (1.8 g, 27.48 mmol) was added portionwise, at –15°C under N<sub>2</sub> atmosphere, CAN (36.3 g, 66.95 mmol) over 2.5 h. After 16h of stirring at –15°C, the solvent was removed under vacuum and the residue was taken up in 200 mL of AcOEt. The organic layer was washed with H<sub>2</sub>O (3 x 30 mL), brine, dried over anhydrous MgSO<sub>4</sub>, filtered, and concentrated under vacuum. The crude product was purified by flash chromatography on silica gel (Cy/AcOEt 80:20), to afford **(1)** as a yellow oil (5.11 g, 13.58 mmol, 75 % yield, α/β 6/4).

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>, 298K): δ = 6.33 (d, 1H, *J*<sub>1,2</sub> = 4.3 Hz, H-1, anomer α), 5.57 (d, 1H, *J*<sub>1,2</sub> = 8.6 Hz, H-1, anomer β), 5.49 (dd, 1H, *J*<sub>3,4</sub> = 3.2 Hz, *J*<sub>4,5</sub> = 1.4 Hz, H-4, anomer α), 5.37 (dd, 1H, *J*<sub>3,4</sub> = 3.3 Hz, *J*<sub>4,5</sub> = 0.9 Hz, H-4, anomer β), 5.23 (dd, 1H, *J*<sub>3,2</sub> = 11.38 Hz, *J*<sub>3,4</sub> = 3.15 Hz, H-3, anomer α), 4.95 (dd, 1H, *J*<sub>2,3</sub> = 10.56 Hz, *J*<sub>3,4</sub> = 3.31 Hz, H-3, anomer β), 4.34 (ddd, 1H, *J*<sub>5,6</sub> = 5.3 Hz, *J*<sub>5,6</sub> = 5.3 Hz, *J*<sub>4,5</sub> = 1.4 Hz, H-5, anomer α), 4.14-4.07 (m, 5H, H-2, H-6 anomer α, H-6 anomer β), 4.04 (ddd, 1H, *J*<sub>5,6</sub> = 6.3 Hz, *J*<sub>5,6</sub> = 6.3 Hz, *J*<sub>4,5</sub> = 0.9 Hz, H-5, anomer β), 3.82 (dd, 1H, *J*<sub>2,3</sub> = 10.6 Hz, *J*<sub>1,2</sub> = 8.6 Hz, H-2, anomer β), 2.16 (s, 6H, 2 x AcO, anomers α/β), 2.06 (s, 3H, AcO, anomer β), 2.05 (s, 3H, AcO, anomer α), 2.03 (s, 3H, AcO, anomer β), 2.02 (s, 3H, AcO, anomer α).

<sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>, 298K): δ = 170.4 (C, AcO), 169.8 (C, AcO), 169.6 (C, AcO), 98.1 (CH, C-1, anomer α), 96.9 (CH, C-1, anomer β), 71.9 (CH, C-5, anomer β), 71.7 (CH, C-3, anomer β), 69.5 (CH, C-5, anomer α), 68.6 (CH, C-3, anomer α), 66.7 (CH, C-4, anomer α), 65.9 (CH, C-4, anomer β), 61.0 (CH<sub>2</sub>, C-6, anomer α), 60.9 (CH<sub>2</sub>, C-6, anomer β), 57.5 (CH, C-2, anomer β), 56.0 (CH, C-2, anomer α), 20.6 (3 x CH<sub>3</sub>, 3 x AcO), 20.6 (3 x CH<sub>3</sub>, 3 x AcO).

HRMS ESI<sup>+</sup> : m/z calcd for C<sub>12</sub>H<sub>16</sub>N<sub>4</sub>O<sub>10</sub>Na [M+Na]<sup>+</sup>: 399.0759 found: 399.0761.

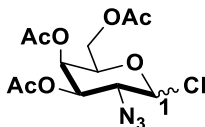
## Compound 2'



To a solution of **(1)** (250 mg, 0.665 mmol) in dry ACN (8 mL), NaI (700 mg, 4.65 mmol) was added at RT under N<sub>2</sub> atmosphere and in the dark. The reaction mixture was stirred for 16h. The solvent was removed by rotary evaporation and the residue was taken up in 200 mL of AcOEt, the organic layer was washed with saturated aqueous solution of Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub> (2 x 15 mL), H<sub>2</sub>O (1 x 30 mL), dried over anhydrous MgSO<sub>4</sub>, filtered, and concentrated under vacuum to yield the desired compound **(2')** as a yellow oil (231.8 mg, 0.525 mmol, yield 79% of the crude) used in the next step without further purification.

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>, 303K):  $\delta$  = 6.88 (d, 1H, H-1), 5.47 (m, 1H, H-4), 5.20 (dd, 1H, H-3), 4.22 (ddd, 1H, H-5), 4.1970-4.1143 (m, 2H, H-6), 4.40 (dd, 1H, H-2), 2.16 (s, 3H, AcO), 2.06 (s, 3H, AcO), 2.06 (s, 3H, AcO).

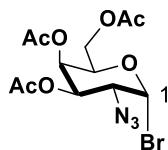
## Compound 2''



To a solution of **(1)** (300 mg, 0.80 mmol) in dry ACN (8 mL), was added, at RT under N<sub>2</sub> atmosphere, TBAC (845  $\mu$ L, 3.20 mmol). After 18 h of stirring at RT, the reaction was concentrated under vacuum and the crude product was purified by flash chromatography on silica gel (DCM/MeOH 90%:10%), to afford the desired compound **(2'')** as a white solid (221.8 mg, 0.616 mmol, yield 77%, mixture  $\alpha/\beta$  9/1).

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>, 298K):  $\delta$  = 6.43 (d, 1H, H-1 anomer  $\alpha$ ), 5.634 (d, 1H, H-1 anomer  $\beta$ ), 5.498 (dd, 1H, H-4 anomer  $\alpha$ ), 5.379 (dd, 1H, H-4 anomer  $\beta$ ), 3.24 (dd, 1H, H-3 anomer  $\alpha$ ), 4.93 (dd, 1H, H-3 anomer  $\beta$ ), 4.35 (ddd, 1H, H-5 anomer  $\alpha$ ), 4.22-4.17 (m, 5H, H-2, H-6 anomer  $\alpha/\beta$ ), 4.92 (ddd, 1H, H-5 anomer  $\beta$ ), 3.56 (dd, 1H, H-2 anomer  $\beta$ ), 2.17 (s, 6H, AcO anomer  $\alpha/\beta$ ), 2.13 (s, 3H, AcO anomer  $\beta$ ), 2.09 (s, 3H, AcO anomer  $\alpha$ ), 2.03 (s, 3H, AcO anomer  $\beta$ ), 2.01 (s, 3H, AcO anomer  $\alpha$ ).

## Compound (2)



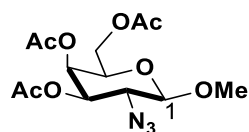
To a solution of **(1)** (2.5 g, 6.62 mmol) in dry ACN (20 mL) was added, at RT under N<sub>2</sub> atmosphere, LiBr (2.26 g, 26.64 mmol). After stirring for 4.5 h, the solvent was removed under reduced pressure and the residue was taken up in 200 mL of AcOEt. The organic layer was washed with H<sub>2</sub>O (3 x 30 mL), brine, dried over anhydrous MgSO<sub>4</sub>, filtered, and concentrated under vacuum to afford the desired crude compound **(2)** as a yellow syrup (2.312 g, 5.83 mmol, yield 87% as a crude), used in the next step without further purification.

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>, 298K):  $\delta$  = 6.46 (d, 1H,  $J_{1,2}$  = 3.7 Hz, H-1), 5.49 (dd, 1H,  $J_{4,3}$  = 3.3 Hz,  $J_{4,5}$  = 1.3 Hz, H-4), 5.35 (dd, 1H,  $J_{2,3}$  = 10.7 Hz,  $J_{3,4}$  = 3.3 Hz, H-3), 4.46 (ddd, 1H,  $J_{5,6}$  = 7.8 Hz,  $J_{5,6}$  = 7.8 Hz,  $J_{5,4}$  = 1.3 Hz, H-5), 4.16 (m, 2H, H-6 a/b) 3.98 (dd, 1H,  $J_{2,3}$  = 10.7 Hz,  $J_{1,2}$  = 3.7 Hz, H-2), 2.15 (s, 3H, AcO), 2.05 (s, 3H, AcO), 2.04 (s, 3H, AcO).

<sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>, 298K):  $\delta$  = 170.4 (C, AcO), 169.8 (C, AcO), 169.6 (C, AcO), 88.9 (CH, C-1), 71.5 (CH, C-5), 69.9 (CH, C-3), 66.6 (CH, C-4), 60.8 (CH, C-6), 58.7 (CH, C-2), 20.7 (CH<sub>3</sub>, AcO), 20.6 (2 x CH<sub>3</sub>, 2 x AcO).

HRMS ESI+: m/z calcd for C<sub>12</sub>H<sub>16</sub>N<sub>3</sub>O<sub>7</sub>NaBr [M+Na]<sup>+</sup>: 416.0064 found: 416.0057.

### Compound (3)



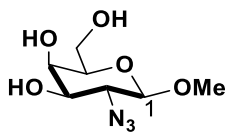
To a solution of **(2)** (2.30 g, 5.83 mmol) in dry MeOH (30 mL) was added, at 0°C under N<sub>2</sub> atmosphere, Ag<sub>2</sub>CO<sub>3</sub> (804 mg, 2.92 mmol). After stirring for 16h at RT in the dark, the reaction mixture was filtered through Celite<sup>®</sup>. The solvent was removed under vacuum and the residue was taken up in 200 mL of DCM. The organic layers were washed with saturated aqueous solution of NaHCO<sub>3</sub> (2 x 50 mL), H<sub>2</sub>O (2 x 50 mL), brine, dried over anhydrous MgSO<sub>4</sub>, filtered and concentrated under vacuum. The crude product was purified by flash chromatography on silica gel (Cy/AcOEt 70:30), to afford the desired compound **(3)** as a yellow solid (1.03 g, 2.97 mmol, yield 51%).

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>, 298K): δ = 5.31 (dd, 1H, J<sub>4,3</sub> = 3.3 Hz, J<sub>4,5</sub> = 0.9 Hz, H-4), 4.78 (dd, 1H, J<sub>2,3</sub> = 10.9 Hz, J<sub>3,4</sub> = 3.3 Hz, H-3), 4.26 (d, 1H, J<sub>1,2</sub> = 8 Hz, H-1), 4.16 (dd, 1H, J<sub>5,6</sub> = 6.7 Hz, J<sub>6a,6b</sub> = 11.3 Hz, H-6 a/b), 4.10 (dd, 1H, J<sub>5,6</sub> = 6.7 Hz, J<sub>6a,6b</sub> = 11.3 Hz, H-6 a/b), 3.84 (ddd, 1H, J<sub>5,6</sub> = 6.7 Hz, J<sub>5,6</sub> = 6.7 Hz, J<sub>5,4</sub> = 0.9 Hz, H-5), 3.64 (dd, 1H, J<sub>2,3</sub> = 10.9 Hz, J<sub>1,2</sub> = 8.0 Hz, H-2), 3.59 (s, 3H, MeO), 2.15 (s, 3H, AcO), 2.05 (s, 3H, AcO), 2.04 (s, 3H, AcO).

<sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>, 298K): δ = 170.5 (C, AcO), 170.2 (C, AcO), 169.9 (C, AcO), 103.2 (CH, C-1), 71.3 (CH, C-3), 70.7 (CH, C-5), 66.5 (CH, C-4), 61.3 (CH, C-2), 61.0 (CH, C-6), 57.5 (CH<sub>3</sub>, MeO), 20.7 (CH<sub>3</sub>, AcO), 20.7 (2 x CH<sub>3</sub>, 2 x AcO).

HRMS ESI+: m/z calcd for C<sub>13</sub>H<sub>19</sub>N<sub>3</sub>O<sub>8</sub>Na [M+Na]<sup>+</sup>: 345.1272 found: 345.1278.

### Compound (4)



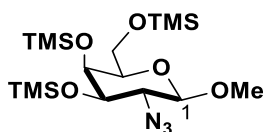
To a solution of **(3)** (1.03 g, 4.56 mmol) in dry MeOH (20 mL) was added, at 0°C under N<sub>2</sub> atmosphere, NaOMe (30 wt% solution in MeOH, 0.2 mL). After stirring for 16h at RT, the reaction mixture was neutralized with DOWEX-50 acid resin (30 mg). The mixture was stirred for 30 minutes under N<sub>2</sub>, filtered and concentrated under vacuum. The crude product was purified by flash chromatography on silica gel (DCM/MeOH 90:10), to afford the desired compound **(4)** as a yellow solid (644 mg, 2.94 mmol, yield 99%).

<sup>1</sup>H NMR (400 MHz, MeOH-D<sub>4</sub>, 303K):  $\delta$  = 4.18 (m, 1H, H-1), 3.79 (m, 1H, H-4), 3.74 (m, 2H, H-6), 3.54 (s, 3H, MeO), 3.47 (m, 1H, H-5), 3.44-3.40 (m, 2H, H-3, H-2).

<sup>13</sup>C NMR (100 MHz, MeOH-D<sub>4</sub>, 303K):  $\delta$  = 104.6 (CH, C-1), 76.6 (CH, C-5), 73.4 (CH, C-3), 69.6 (CH, C-4), 65.7 (CH, C-2), 62.3 (CH, C-6), 57.2 (CH<sub>3</sub>, MeO).

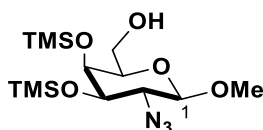
HRMS ESI+: m/z calcd for C<sub>7</sub>H<sub>13</sub>N<sub>3</sub>O<sub>5</sub>Na [M+Na]<sup>+</sup>: 242.0745 found: 242.0753.

### Compound (5)



To a solution of **(4)** (270 mg, 1.23 mmol) in dry pyridine (5 mL), was added, at 0°C under N<sub>2</sub> atmosphere, TMSCl (3.2 mL, 24.63 mmol). After stirring for 12h at RT, the solvent was removed under reduced pressure and the residue was taken up in 200 mL of AcOEt/H<sub>2</sub>O. The aqueous layer was extracted once with EtOAc. The combined organic layers were then washed with H<sub>2</sub>O (2 x 10 mL), brine, dried over anhydrous MgSO<sub>4</sub>, filtered, and concentrated under vacuum to afford the tri-silylated intermediate **(5)** as a yellow oil (530 mg, 1.22 mmol). The crude product was directly used in the next step without further purification.

### Compound (6)



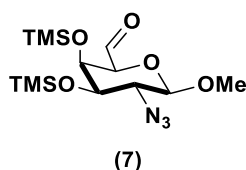
To a solution of the crude **(5)** (530 mg, 1.22 mmol) in dry MeOH (20 mL) was added, at 0°C under N<sub>2</sub> atmosphere, 30 mM K<sub>2</sub>CO<sub>3</sub> in MeOH (520 μL, 16.57 μmol). After stirring for 2.5h at 0°C, the mixture was neutralized by the addition of silica gel (2 g). The solvent was removed under vacuum and the silica pad was purified by flash chromatography on silica gel (PE/AcOEt 80:20) to afford the di-silylated derivative **(6)** as a yellow oil (200 mg, 0.55 mmol, 45% in two steps).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>, 303K): δ = 4.12 (d, 1H, J<sub>1,2</sub> = 8.8 Hz, H-1), 3.85 (dd, 1H, J<sub>5,6</sub> = 7.2 Hz, J<sub>6a,6b</sub> = 11.2 Hz, H-6 a/b), 3.75 (dd, 1H, J<sub>4,5</sub> = 0.6 Hz, J<sub>3,4</sub> = 2.7 Hz, H-4), 3.6 (dd, 1H, J<sub>5,6</sub> = 4.8 Hz, J<sub>6a,6b</sub> = 11.2 Hz, H-6 a/b), 3.57 (s, 3H, MeO), 3.52 (dd, 1H, J<sub>1,2</sub> = 8.8 Hz, J<sub>2,3</sub> = 10.0 Hz, H-2), 3.43 (ddd, 1H, J<sub>4,5</sub> = 0.6 Hz, J<sub>5,6</sub> = 4.8 Hz, J<sub>5,6</sub> = 4.8 Hz, H-5), 3.36 (dd, 1H, J<sub>3,4</sub> = 2.7 Hz, J<sub>2,3</sub> = 10.0 Hz, H-3), 0.18 (s, 9H, 3 x CH<sub>3</sub>, TMSO), 0.12 (s, 9H, 3 x CH<sub>3</sub>, TMSO).

<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>, 303K): δ = 103.5 (CH, C-1), 75.3 (CH, C-5), 73.8 (CH, C-3), 71.2 (CH, C-4), 64.1 (CH, C-2), 62.7 (CH<sub>2</sub>, C-6), 57.4 (CH<sub>3</sub>, MeO), 0.7 (3 x CH<sub>3</sub>, TMSO), 0.4 (3 x CH<sub>3</sub>, TMSO).

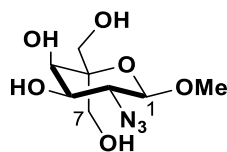
HRMS ESI+: m/z calcd for C<sub>13</sub>H<sub>29</sub>N<sub>3</sub>O<sub>5</sub>Si<sub>2</sub>Na. [M+Na]<sup>+</sup>: 386.1538 found: 386.1542.

### Compound (7)



To a solution of **(6)** (115 mg, 0.31 mmol) in dry DCM (2 mL) were added, at RT under N<sub>2</sub> atmosphere, DMSO (157 μL, 2.11 mmol) and Et<sub>3</sub>N (134 μL, 0.89 mmol). The mixture was cooled to 0 °C and SO<sub>3</sub>Py (152 mg, 0.95 mmol) was added. The mixture was stirred at 0 °C for 1h then, warmed to RT over 30 minutes. The reaction was quenched with brine and diluted with DCM (15 mL). The organic layer was washed with brine (3 x 30 mL), dried over anhydrous MgSO<sub>4</sub>, filtered and concentrated under vacuum to afford the aldehyde derivate **(7)** (113 mg, 0.31 mmol). Due to the instability of the aldehyde, the crude was used in the next step without further purification neither characterization.

### Compound (8)



To a solution of **(7)** (1.3 g, 3.66 mmol) in dry EtOH (18 mL) was treated with paraformaldehyde powder (6.8 g, 73.20 mmol, 20 eq) followed by NaOEt (21 wt% solution in EtOH, 2.7 mL, 7.320 mmol). After 16 h of stirring at RT, the solvent was evaporated. MeOH was added to the crude and the solid was filtered and thoroughly rinsed with MeOH. To the filtrate containing the desired product was added silica gel and the solvent was concentrated under vacuum. The resulting solid was dried under high vacuum and purified by flash chromatography on silica gel (AcOEt/MeOH 90:10) to afford **(7)** as a colorless oil (465 mg, 1.87 mmol, yield of 51% two steps).

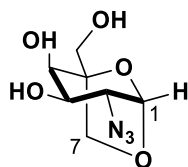
<sup>1</sup>H NMR (400 MHz, MeOH-D<sub>4</sub>, 303K): δ = 4.54 (d, 1H, J<sub>1-2</sub> = 8.2 Hz, H-1), 3.84-3.64 (m, 6H, H-3, H-4, H-6, H-7), 3.51 (s, 3H, MeO), 3.45 (dd, 1H, J<sub>2-3</sub> = 10.3 Hz, J<sub>1-2</sub> = 8.2 Hz, H-2).

<sup>13</sup>C NMR (100 MHz, MeOH-D<sub>4</sub>, 303K): δ = 101.1 (CH, C-1), 80.8 (C, C-5), 71.1 (CH, C-3), 69.9 (CH, C-4), 65.8 (CH<sub>2</sub>, C-6), 63.7 (CH, C-7), 61.28 (CH, C-2), 57.02 (MeO).

HRMS ESI+: m/z calcd for C<sub>8</sub>H<sub>15</sub>N<sub>3</sub>O<sub>6</sub>Na. [M+Na]<sup>+</sup>: 272.0853 found: 272.0859.



### Compound (9)



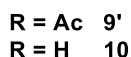
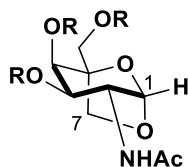
To a solution of **(8)** (520 mg, 2.08 mmol) in H<sub>2</sub>O (7 mL) was treated with concentrated H<sub>2</sub>SO<sub>4</sub> (1.16 mL). The reaction mixture was stirred at 100 °C for 48h and then neutralized by the addition of concentrated NH<sub>4</sub>OH. H<sub>2</sub>O was removed under vacuum and the resulting mixture of desired product and salts was treated with MeOH. The solid was filtered and thoroughly rinsed with MeOH. Silica gel was added to the filtrate and the solvent evaporated. The resulting silica pad was dried under high vacuum, loaded onto a silica gel chromatography column, and purified by flash chromatography on silica gel (AcOEt/MeOH 90:10) to afford **(9)** as a colorless oil (215 mg, 1.00 mmol, yield 48%).

<sup>1</sup>H NMR (400 MHz, MeOH-D<sub>4</sub>, 303K):  $\delta$  = 5.31 (d, 1H,  $J_{1-2}$  = 1.5 Hz, H-1), 3.91 (d, 1H,  $J_{6a-6b}$  = 11.4 Hz, H-6 a/b), 3.89-3.83 (m, 2H, H-3, H-4), 3.80 (d, 1H,  $J_{6a-6b}$  = 11.4 Hz, H-6 a/b), 3.76 (d, 1H,  $J_{7a-7b}$  = 7.9 Hz, H-7 a/b), 3.69 (d, 1H,  $J_{7a-7b}$  = 7.9 Hz, H-7 a/b ), 3.34 (dd, 1H,  $J_{2-3}$  = 9.15 Hz,  $J_{1-2}$  = 1.5 Hz, H-2).

<sup>13</sup>C NMR (100 MHz, MeOH-D<sub>4</sub>, 303K):  $\delta$  = 102.6 (CH, C-1), 85.2 (C, C-5), 71.0 (CH, C-3), 69.5 (CH, C-4), 69.4 (CH<sub>2</sub>, C-6), 66.0 (CH, C-2), 61.8 (CH<sub>2</sub>, C-7).

HRMS ESI+: m/z calcd for C<sub>8</sub>H<sub>14</sub>N<sub>3</sub>O<sub>6</sub>. [M-H]<sup>-</sup>: 216.0621 found: 216.0620.

### Compound (10)



A solution of compound **(9)** (90 mg, 0.41 mmol) in EtOH (2.5 mL) under N<sub>2</sub> atmosphere was treated with Lindlar catalyst (88.2 mg, 0.41 mmol). The mixture was stirred at RT under an atmosphere of H<sub>2</sub> (1 atm) for 24 h. The catalyst was filtered through Celite® and thoroughly rinsed with MeOH and then H<sub>2</sub>O. The solvent was evaporated and the residue dissolved in H<sub>2</sub>O and lyophilized. The resulting material was dissolved in Py (2.5 mL) and Ac<sub>2</sub>O was added (2.5ml). After 24 h of stirring at RT, the solvents were removed by rotary evaporation. The residue was dissolved in AcOEt, washed with saturated aqueous NaHCO<sub>3</sub>. The aqueous phase was extracted twice with AcOEt and the combined organic layers were washed with H<sub>2</sub>O and brine, dried over anhydrous MgSO<sub>4</sub>, filtered and concentrated under vacuum. The crude was purified by flash chromatography on silica gel (DCM/MeOH 97:3) to afford the peracetylated precursor **(9')** as white solid (100 mg, 0.28 mmol, yield 67 %).

**<sup>1</sup>H NMR** (400 MHz, MeOH-D<sub>4</sub>, 303K):  $\delta$  = 5.38 (d, 1H,  $J_{3,4}$  = 4.4 Hz, H-4), 5.35 (d, 1H,  $J_{1,2}$  = 1.5 Hz, H-1), 5.13 (dd, 1H,  $J_{2,3}$  = 10.5 Hz,  $J_{3,4}$  = 4.4 Hz, H-3), 4.45 (d, 1H,  $J_{7a,7b}$  = 11.4 Hz, H-7 a/b), 4.20 (dd, 1H,  $J_{2,3}$  = 10.5 Hz,  $J_{1,2}$  = 1.5 Hz, H-2), 4.13 (d, 1H,  $J_{7a,7b}$  = 11.4 Hz, H-7 a/b), 4.05 (d, 1H,  $J_{6a,6b}$  = 8.5 Hz, H-6 a/b), 3.74 (d, 1H,  $J_{6a,6b}$  = 8.5 Hz, H-6 a/b), 2.15 (s, 3H, AcNH), 2.03 (s, 3H, AcO), 1.95 (s, 3H, AcO), 1.94 (s, 3H, AcO).

**<sup>13</sup>C NMR** (100 MHz, MeOH-D<sub>4</sub>, 303K):  $\delta$  = 173.9 (C, NHAc), 172.0 (C, AcO), 171.8 (C, AcO), 171.7 (C, AcO), 102.92 (CH, C-1), 82.4 (C, C-5), 70.2 (CH<sub>2</sub>, C-7), 69.0 (CH, C-3), 68.8 (CH, C-4), 62.9 (CH<sub>2</sub>, C-6), 53.2 (CH, C-2), 22.4 (CH<sub>3</sub>, NHAc) 20.5 (CH<sub>3</sub>, AcO), 20.3 (CH<sub>3</sub>, AcO).

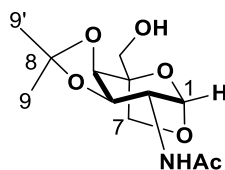
To a solution of **(9')** (100 mg, 0.28 mmol) in dry MeOH (3ml) was added, at 0°C under N<sub>2</sub> atmosphere, NaOMe in MeOH (30 wt% solution in MeOH, 40  $\mu$ L, 0.278 mmol). After 16 h of stirring at RT, DOWEX-50 resin (30 mg) was then added. The mixture was stirred for 30 minutes under N<sub>2</sub>, filtered and concentrated under vacuum to yield the desired compound **(10)** as a white solid (60 mg, 0.28 mmol, quantitative yield).

**<sup>1</sup>H NMR** (400 MHz, MeOH-D<sub>4</sub>, 303K):  $\delta$  = 5.22 (d, 1H,  $J_{1,2}$  = 1.4 Hz H-1), 3.95 (dd, 1H,  $J_{2,3}$  = 9.9 Hz,  $J_{1,2}$  = 1.4 Hz, H-2), 3.92 (d, 1H,  $J_{7a,7b}$  = 11.4 Hz, H-7 a/b), 3.87 (dd, 1H,  $J_{3,4}$  = 4.3 Hz, H-4), 3.80 (d, 1H,  $J_{7a,7b}$  = 11.4 Hz, H-7 a/b), 3.79 (d, 1H,  $J_{6a,6b}$  = 7.9 Hz, H-6 a/b), 3.72 (dd, 1H,  $J_{2,3}$  = 9.9 Hz,  $J_{3,4}$  = 4.3 Hz, H-3), 3.68 (d, 1H,  $J_{6a,6b}$  = 7.9 Hz, H-6a/b), 1.99 (s, 3H, AcNH).

**<sup>13</sup>C NMR** (100 MHz, MeOH-D<sub>4</sub>, 303K):  $\delta$  = 173.9 (C, NHAc), 102.6 (CH, C-1), 85.0 (C, C-5), 70.5 (CH, C-4), 69.2 (CH<sub>2</sub>, C-6), 69.1 (CH, C-3), 62.0 (CH<sub>2</sub>, C-7), 56.3 (CH, C-2), 22.6 (CH<sub>3</sub>, NHAc).

**HRMS** ESI-: m/z calcd for C<sub>9</sub>H<sub>14</sub>NO<sub>6</sub>. [M-H]<sup>-</sup> 232.0827 found: 232.0821.

### Compound (11)



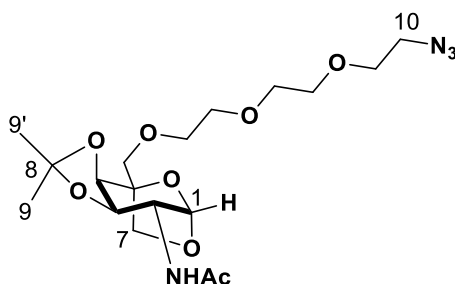
To a solution of compound **(10)** (60 mg, 0.26 mmol) in DMF (6 mL) was added 2,2-dimethoxypropane (190  $\mu$ L, 1.54 mmol) followed by (+/-)-camphor-10-sulphonic acid (30 mg, 0.13 mmol). After stirring at 70 °C for 24 h, the mixture was cooled to RT and treated with MeOH (313  $\mu$ L). The mixture was stirred at RT for 30 minutes and then neutralized with TEA (14,6  $\mu$ L). The solvent was concentrated under vacuum and the residue was co-evaporated 3 times with toluene. The resulting crude material was purified by flash chromatography on silica gel (AcOEt/MeOH 90:10) to afford compound **(11)** as a white solid (65 mg, 0.24 mmol, yield 93%).

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>, 303K):  $\delta$  = 5.56 (d, 1H,  $J_{\text{NH},2}$  = 8.5 Hz NH), 5.37 (d, 1H,  $J_{1,2}$  = 2.2 Hz, H-1), 4.17-4.11 (m, 3H, H-2, H-4, H-6 a/b), 4.06-4.03 (m, 1H, H-3), 3.89-3.83 (m, 1H, H-6 a/b), 3.86 (d, 1H,  $J_{7a,7b}$  = 7.9 Hz, H-7 a/b), 3.80 (d, 1H,  $J_{7a,7b}$  = 7.9 Hz, H-7 a/b), 2.02 (s, 3H, AcNH), 1.98 (Bs, 1H, OH), 1.56 (s, 3H, H-9), 1.35 (s, 3H, H-9').

**<sup>13</sup>C NMR** (100 MHz, CDCl<sub>3</sub>, 303K):  $\delta$  = 170.0 (C, AcNH), 112.0 (C, C-8), 101.4 (CH, C-1), 81.5 (C, C-5), 77.0 (CH, C-3), 75.5 (CH, C-4), 68.6 (CH<sub>2</sub>, C-7), 61.6 (CH<sub>2</sub>, C-6), 54.4 (C-2), 28.0 (CH<sub>3</sub> H-9'), 26.6 (CH<sub>3</sub>, H-9), 23.5 (CH<sub>3</sub>, NHAc).

**HRMS ESI+**: m/z calcd for C<sub>12</sub>H<sub>19</sub>NO<sub>6</sub>Na. [M+Na]<sup>+</sup>: 296.1110 found: 296.1110.

## Compound (12)



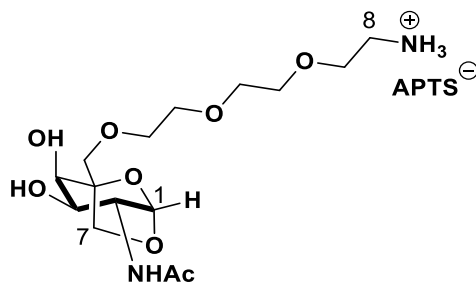
To a solution of **(11)** (50 mg, 0.18 mmol) in DCM (1 mL) containing aqueous NaOH (0.5 mL of 12.5M solution) and 15-crown-5 (5  $\mu$ L, 0.02 mmol) was added the compound **(PEG-2)** (260 mg, 0.92 mmol). The reaction mixture was stirred vigorously at 55 °C. After 24h of stirring, the mixture was cooled to RT, diluted with DCM (3 ml) and washed with brine (2mL). The aqueous layer was extracted three time with DCM. The combined organic layer was washed with H<sub>2</sub>O, brine, dried over anhydrous MgSO<sub>4</sub>, filtered and concentrated under vacuum. The crude product was purified by flash chromatography on silica gel (Cyclo/AcOEt 50:50) then (AcOEt/MeOH 95:5) to afford compound **(12)** as a colorless oil (62 mg, 0.15 mmol, 80% yield).

<sup>1</sup>H NMR (400 MHz, MeOH-D<sub>4</sub>, 303K):  $\delta$  = 5.23 (d, 1H,  $J_{1,2}$  = 2.0 Hz, H-1), 4.31 (d, 1H,  $J_{3,4}$  = 5.9 Hz, H-4), 4.17-4.14 (m, 1H, H-3), 3.95 (d, 1H,  $J_{6a,6b}$  = 10.1 Hz, H-6 a/b), 3.92 (dd, 1H,  $J_{1,2}$  = 2.0 Hz,  $J_{2,3}$  = 7.3 Hz, H-2), 3.86 (d, 1H,  $J_{7a,7b}$  = 7.9 Hz, H-7 a/b), 3.77 (d, 1H,  $J_{6a,6b}$  = 10.1 Hz, H-6 a/b), 3.76 (d, 1H,  $J_{7a,7b}$  = 7.9 Hz, H-7  $\alpha/\beta$ ), 3.71-3.63 (m, 10H, CH<sub>2</sub> PEG), 3.38-3.38 (t, 2H,  $J_{10-PEG}$  = 5.00 Hz, H-10), 1.98 (s, 3H, AcNH), 1.49 (s, 3H, CH<sub>3</sub>, H-9), 1.34 (s, 3H, CH<sub>3</sub>, H-9').

<sup>13</sup>C NMR (100 MHz, MeOH-D<sub>4</sub>, 303K):  $\delta$  = 173.4 (NHAc), 112.2 (C, C-8), 102.3 (CH, C-1), 81.9 (C, C-5), 76.37 (CH, C-3), 76.0 (CH, C-4), 72.5, 71.7, 71.6, 71.5, 71.2 (CH<sub>2</sub> PEG), 71.0 (CH<sub>2</sub>, C-6), 70.7 (CH<sub>2</sub>, C-7), 56.6 (CH, C-2), 51.8 (CH<sub>2</sub>, C-10), 28.3 (CH<sub>3</sub>, C-9), 26.7 (CH<sub>3</sub>, C-9'), 22.5 (CH<sub>3</sub>, NHAc).

HRMS ESI+: m/z calcd for C<sub>18</sub>H<sub>30</sub>N<sub>4</sub>O<sub>8</sub>Na. [M+Na]<sup>+</sup>: 453.1962 found: 453.1961.

### Compound (13)



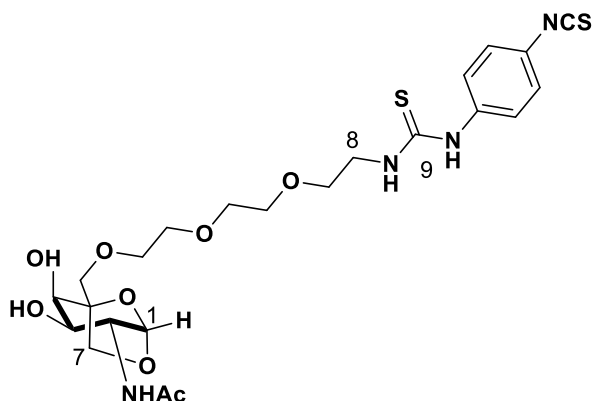
Compound **(12)** (62 mg, 0.14 mmol) was dissolved in 2.5 mL of AcOH/MeOH/H<sub>2</sub>O (3:1:1) mixture and the reaction mixture was stirred at 70°C for 36 h. After that period of time TLC showed complete consumption of the starting material and the solvents were evaporated under vacuum to give a deprotected intermediate. The crude was dissolved in MeOH (3 mL) and Pd/C (10% weight) was added. The mixture was stirred at RT under an atmosphere of H<sub>2</sub> (1 atm) for 16 h. The catalyst was filtered through a Celite<sup>®</sup> pad and thoroughly rinsed with MeOH and then H<sub>2</sub>O. The solvent was evaporated under vacuum to afford **(13)** as a white solid (70mg, 0.13 mmol, yield 90% as a crude).

<sup>1</sup>H NMR (400 MHz, MeOH-D<sub>4</sub>, 303K):  $\delta$  = 7.72 (m, 2H, aromatique), 7.24 (m, 2H, aromatique), 5.22 (d, 1H,  $J_{1,2}$  = 2.0 Hz, H-1), 3.97-3.91 (m, 3H, H-2, H-3, H-4), (m, 14H, CH<sub>2</sub> PEG, H-7, H-6), 3.00-2.96 (t, 2H,  $J_{8-PEG}$  = 5.00 Hz, H-8), 2.37 (s, 3H, CH<sub>3</sub> APTS), 1.99 (s, 3H, AcNH).

<sup>13</sup>C NMR (100 MHz, MeOH-D<sub>4</sub>, 303K):  $\delta$  = 174.1 (C, NHAc), 142.2, 141.9 (C, APTS), 129.8, 127.0 (CH<sub>2</sub>, APTS), 102.3 (CH, C-1), 84.2 (C, C-5), 72.1 (CH, C-3), 71.5 (CH, C-4), 71.4, 71.2, 70.8, 70.3 (CH<sub>2</sub> PEG), 69.8 (CH<sub>2</sub>, C-6), 68.9 (CH<sub>2</sub>, C-7), 56.1 (CH, C-2), 41.3 (CH<sub>2</sub>, C-8), 22.6 (NHAc), 21.3 (CH<sub>3</sub>, APTS).

HRMS ESI+: m/z calcd for C<sub>15</sub>H<sub>29</sub>N<sub>2</sub>O<sub>8</sub>. [M]<sup>+</sup>: found:365.1919 found: 365.1922.

## Compound PMS-02



To a solution of **(13)** (77 mg, 0.14 mmol) in MeOH (5 mL) was added Amberlite Resin basic (30 mg). After 1 h of stirring at RT, the resin was filtered off and the solvents were evaporated under vacuum to give the free amine intermediate as a crude.

A solution of this crude in DMF (2 mL) was added dropwise over 1h to a solution of p-Phenylene diisothiocyanate (131 mg, 0.685 mmol) in DMF (3mL). After 1h of stirring at RT, the solvent was evaporated under vacuum. The resulting crude material was purified by flash chromatography on silica gel (DCM/MeOH 85:15) to afford compound **(PMS-02)** as a white solid (28.3 mg, 0.05 mmol, yield 37 %).

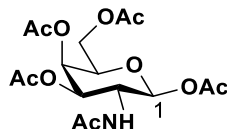
**<sup>1</sup>H NMR** (400 MHz, DMSO-D<sub>6</sub>, 298K):  $\delta$  = 9.82 (bs, 1H, H-9), 7.94 (bs, 1H, N<sub>H</sub>SNH), 7.90 (d, 1H,  $J_{\text{NH},2}$  = 7.58 Hz, NHAc), 7.35 (m, 2H, CH aromatic), 7.37 (m, 2H, CH aromatic), 5.08 (dd, 1H,  $J_{1,2}$  = 1.4 Hz, H-1), 4.85 (dd, 1H,  $J^3$  = 5.4 Hz, OH), 4.59 (dd, 1H,  $J^3$  = 7.4 Hz, OH), 3.83-3.47 (m, 19H, H-2, H-3, H-4, H-6, H-7, H-8, CH<sub>2</sub> PEG), 1.83 (s, 3H, NHAc).

**<sup>13</sup>C NMR** (100 MHz, DMSO-D<sub>6</sub>, 303K):  $\delta$  = 180.4 (C, C-9), 169.66 (C, NHAc), Not detected (C, NCS), not detected (C, aromatic), not detected (C, aromatic), 126.15 (2\*CH, aromatic), 123.31 (2\*CH, aromatic), 100.71 (CH, C-1), 82.8 (C, C-5), 72.2-66.09 (C-3, C-4, CH<sub>2</sub> PEG, CH<sub>2</sub> C-6, CH<sub>2</sub> C-7, C-8), 54.3 (CH, C-2), 43.9 (CH<sub>2</sub>, C-8), 22.93 (CH<sub>3</sub>, NHAc).

**HRMS** ESI+: m/z calcd for C<sub>23</sub>H<sub>32</sub>N<sub>4</sub>O<sub>8</sub>NaS<sub>2</sub>. [M+Na]<sup>+</sup>: 581.1550 found: 581.1559.

## C. Ligand PMS-02bis

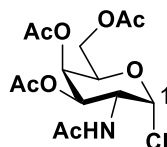
### Compound (14)



D-Galactosamine hydrochloride (5 g, 23.2 mmol) was dissolved in Py (140 mL) and stirred at 80°C for 4h. Then Ac<sub>2</sub>O (50 mL) was added at 0°C. After stirring for 16h at RT, the mixture was concentrated under reduced pressure and MeOH (75 mL) was added. The solution was stirred at 80°C for 30 min, then cooled at RT to lead to crystallize. The mixture was filtered using a Buchner funnel to afford the pure product **(14)** as a white solid (7.77 g, 19.9 mmol, yield 86%). Spectroscopic datas were in concordance with those described in the bibliography.<sup>190</sup>

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>, 303 K):  $\delta$  = 5.69 (d, 1H,  $J_{1,2}$  = 8.80 Hz, H-1), 5.38 (d, 1H,  $J_{2,NH}$  = 9.20 Hz, NH), 5.37 (dd, 1H,  $J_{4,5}$  = 0.84 Hz,  $J_{3,4}$  = 3.28 Hz, H-4), 5.08 (dd, 1H,  $J_{2,3}$  = 11.40 Hz,  $J_{3,4}$  = 3.28 Hz, H-3), 4.45 (ddd, 1H,  $J_{1,2}$  = 8.80 Hz,  $J_{2,3}$  = 11.40 Hz,  $J_{2,NH}$  = 9.20 Hz, H-2), 4.21-4.07 (m, 2H, H-6), 4.02 (m, 1H, H-5), 2.17, 2.13, 2.05, 2.02, 1.94 (s, 5 x 3H, AcO/NHAc).

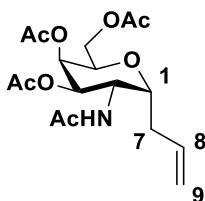
### Compound (15)



To a solution of **(14)** (2.00 g, 5.1 mmol) in dry DCM (50 mL) was added, at 0°C under N<sub>2</sub> atmosphere, TiCl<sub>4</sub> (1.95 g, 10.2 mmol). After stirring for 16h at 50°C, the reaction mixture was diluted in DCM (50 ml), washed with saturated aqueous solution of NaHCO<sub>3</sub> (2 x 100 mL), H<sub>2</sub>O (2 x 100 mL), brine, dried over anhydrous MgSO<sub>4</sub>, filtered, and concentrated under vacuum. The crude product was purified by flash chromatography on silica gel (AcOEt 100%), to afford the desired compound **(15)** as a yellow solid (1.00 mg, 2.7 mmol, yield 53%). Spectroscopic datas were in concordance with those described in the bibliography.<sup>191</sup>

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>, 298K):  $\delta$  = 6.20 (d, 1H,  $J_{1,2}$  = 3.6 Hz, H-1), 5.50 (d, 1H,  $J_{NH,2}$  = 9.2 Hz, NH), 5.41 (dd, 1H,  $J_{4,3}$  = 3.2 Hz,  $J_{4,5}$  = 1.2 Hz, H-4), 5.21 (dd, 1H,  $J_{2,3}$  = 11.5 Hz,  $J_{3,4}$  = 3.2 Hz, H-3), 4.71 (dd, 1H,  $J_{2,3}$  = 11.5 Hz,  $J_{1,2}$  = 3.6,  $J_{NH,2}$  = 9.2 Hz, H-2), 4.23 (ddd, 1H,  $J_{5,6}$  = 6.7 Hz,  $J_{5,6}$  = 7.6 Hz,  $J_{5,4}$  = 1.2 Hz, H-5), 4.15-4.00 (m, 2H, H-6), 2.16 (s, 3H, AcO), 2.02 (s, 3H, AcO), 2.01 (s, 3H, AcO), 1.94 (s, 3H, NHAc).

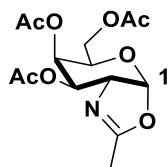
### Compound (16)



To a solution of **(15)** (700 g, 1.9 mmol) in dry toluene (10 mL) was added, at RT under N<sub>2</sub> atmosphere, allyltributylstannane (3.2 g, 9.6 mmol). The mixture was degassed by purging N<sub>2</sub> for 20 minutes and AIBN (31 mg, 0.2 mmol) was added. After stirring for 16h at 80°C, the reaction mixture was diluted in AcOEt (50 ml). The crude product was purified by flash chromatography on silica gel (Cy/AcOEt 70:30) to afford the desired compound **(16)** as a yellow solid (130 mg, 0.34 mmol, yield 18%). Spectroscopic data were in concordance with those described in the bibliography.<sup>158</sup>

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>, 298K):  $\delta$  = 5.84-5.66 (m, 3H, H-8, NH), 5.34-5.30 (m, 1H, H-4), 5.18-5.00 (m, 3H, H-3 H-9a, H-9b), 4.52-4.41 (m, 1H, H-2), 4.37-3.94 (m, 4H, H-1, H-5, H-6a, H-6b), 2.50-2.17 (m, 2H, H-7), 2.10 (s, 3H, AcO), 2.05 (s, 3H, AcO), 2.03 (s, 3H, AcO), 1.96 (s, 3H, NHAc).

### Compound (17)



To a solution of **(14)** (3 g, 7.7 mmol) in dry DCM (40 mL), was added, at 0°C and under N<sub>2</sub>, TMSOTf (4.89 mL, 26.26 mmol). After stirring for 24h at 50°C under N<sub>2</sub>, the reaction mixture was quenched by adding Et<sub>3</sub>N (1.1 mL) at 0°C, diluted in DCM, washed with saturated aqueous solution of NaHCO<sub>3</sub> (2 x 50 mL), H<sub>2</sub>O (2 x 50 mL), brine, dried over anhydrous MgSO<sub>4</sub>, filtered, and concentrated under vacuum to afford **(17)** as a brown oil (1.97g, 6.00 mmol, yield 78% as a crude).

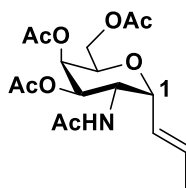
<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>, 303 K):  $\delta$  = 6.00 (d, 1H, J<sub>1-2</sub> = 6.9 Hz, H-1), 5.46 (dd, 1H, J<sub>4-3</sub> = 3.2 Hz, H-4), 4.90 (dd, 1H, J<sub>3-4</sub> = 3.2 Hz, J<sub>2-1</sub> = 7.3 Hz, H-3), 4.24 (dd, 1H, J = 7.3 Hz, J<sub>2-1</sub> = 6.9 Hz, H-2), 4.19 (dd, 1H, H<sub>6a</sub>, J<sub>6a-5</sub> = 6.9 Hz, J<sub>6a-6b</sub> = 11.2 Hz), 4.10 (dd, 1H, J<sub>6b-5</sub> = 5.9 Hz, J<sub>6b-6a</sub> = 11.2 Hz, H-6), 3.99 (m, 1H, H-5), 2.12 (s, 3H, CH<sub>3</sub>), 2.00 (m, 9H, 3 x CH<sub>3</sub>).

<sup>13</sup>C NMR (400 MHz, CDCl<sub>3</sub>, 303K):  $\delta$  = 170.5 (C, AcO), 170.2 (C, AcO), 169.8 (C, AcO), 166.4 (C, NHAc), 101.5 (CH, C-1), 71.9 (CH, C-3), 69.5 (CH, C-2), 65.3 (CH, C-4), (CH, C-5), 61.6 (CH<sub>2</sub>, C-6), 20.8 (CH<sub>3</sub>, AcO), 20.7 (CH<sub>3</sub>, AcO), 20.6 (CH<sub>3</sub>, AcO), 14.5 (CH<sub>3</sub>, NHAc).

HRMS ESI+ : m/z calc for C<sub>14</sub>H<sub>20</sub>NO<sub>8</sub>. [M+H]<sup>+</sup> 330.1189 found 330.1188.



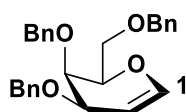
### Compound (18)



To a solution of **(16)** (100 mg, 0.27 mmol) in dry toluene (3 mL), was added, at RT and under N<sub>2</sub>, Bis(benzonitrile) Palladium (II) chloride (20.6 mg, 0.054 mmol). After stirring at RT for 16h, the reaction mixture was stirred at 80°C for 8 h then, cooled to RT and the catalyst was filtered through a Celite<sup>®</sup> pad and washed thoroughly with DCM. The crude product was purified by flash chromatography on silica gel (AcOEt 100%), to afford **(19)** as a white solid (33mg, 0.09 mmol, yield 33%). the mixture of compounds (probably isomers) makes the characterization difficult.

HRMS ESI<sup>+</sup> : m/z calc for C<sub>17</sub>H<sub>25</sub>NO<sub>8</sub>. [M+H]<sup>+</sup> 372.1658 found 372.1660.

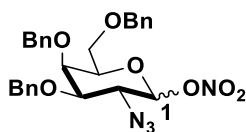
### Compound (19)



To a solution of D-galactal (2g, 13.69 mmol) in dry DMF (6.6 mL) was added portion wise, at 0°C under N<sub>2</sub> atmosphere, NaH (2.3 g, 54.74 mmol) over 30 minutes. BnBr (8 ml, 68.45 mmol) was then added dropwise at 0°C over 1 h and the reaction mixture was stirred for 3 h at 0°C. The reaction mixture was quenched with MeOH (25 ml), stirred 30 minutes and the solvent was removed under vacuum. The residue was taken up in 150 mL of AcOEt. The organic layer was washed with H<sub>2</sub>O (3 x 60 mL), brine, dried over anhydrous MgSO<sub>4</sub>, filtered and concentrated under vacuum. The crude product was purified by flash chromatography on silica gel (Cy/AcOEt 90:10), to afford **(19)** as a white solid (5.19 g, 12.45 mmol, yield 91 %). Spectroscopic datas were in concordance with those described in the bibliography.<sup>192</sup>

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>, 303 K): δ = 7.40–7.22 (m, 15H, CH BnO), 6.33 (dd, 1H, J = 6.1, 1.4 Hz, H-1), 4.88 (d, 1H, J<sub>3</sub> = 12.2 Hz, CH<sub>2</sub> BnO), 4.87-4.85 (m, 1H, H-2), 4.70-7.60 (m, 3H, CH<sub>2</sub> BnO), 4.52 (d, 1H, J<sub>3</sub> = 12.0 Hz, CH<sub>2</sub> BnO), 4.45 (d, 1H, J<sub>3</sub> = 12. Hz, CH<sub>2</sub> BnO), 4.23-4.17 (m, 2H, H-3, H-4), 3.98–3.93 (m, 1H, H-5), 3.80 (dd, 1H, J<sub>5,6</sub> = 7.2 Hz, J<sub>6,6</sub> = 10.3 Hz, H-6a), 3.80 (dd, 1H, J<sub>5,6</sub> = 5.2 Hz, J<sub>6,6</sub> = 10.3 Hz, H-6b).

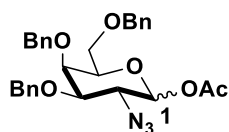
## Compound (20)



To a solution of **(19)** (5 g, 12.01 mmol) in dry ACN (60 mL) containing NaN<sub>3</sub> (1.18 g, 18.01 mmol) was added portion wise, at -20°C under N<sub>2</sub> atmosphere, CAN (19.76 g, 36.03 mmol) over 2.5 h. Then the reaction mixture was stirred for 3 h at -20°C. The mixture was diluted with 200 ml of AcOEt and washed with 160 ml of H<sub>2</sub>O. The aqueous layer was extracted once with EtOAc. The combined organic layers were then washed with H<sub>2</sub>O (3 x 30 mL), brine, dried over anhydrous MgSO<sub>4</sub>, filtered and concentrated under vacuum. The crude product was purified by flash chromatography on silica gel (Cy/AcOEt 90:10), to afford **(20)** as a yellow oil (2.50 g, 4.02 mmol, yield 40 %,  $\alpha/\beta$  8/2). Spectroscopic datas were in concordance with those described in the bibliography.<sup>193,194</sup>

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>, 298K):  $\delta$  = 7.40–7.22 (m, 17H, CH BnO, anomers  $\alpha/\beta$ ), 6.26 (d, 1H,  $J_{1,2}$  = 4.2 Hz, H-1, anomer  $\alpha$ ), 5.44 (d, 0.2H,  $J_{1,2}$  = 8.8 Hz, H-1, anomer  $\beta$ ), 4.93-4.86 (m, 1.2H, CH<sub>2</sub> BnO, anomers  $\alpha/\beta$ ), 4.80-4.70 (m, 2.4H, CH<sub>2</sub> BnO, anomers  $\alpha/\beta$ ), 4.62-4.40 (m, 3.6H, CH<sub>2</sub> BnO, anomers  $\alpha/\beta$ ), 4.3 (dd, 1H,  $J_{2,3}$  = 10.81 Hz,  $J_{2,1}$  = 4.2 Hz, H-2, anomer  $\alpha$ ), 4.13-4.06 (m, 1.4H, H-4, anomer  $\alpha$ , H-2, H-3, anomer  $\beta$ ), 4.00-3.97 (m, 0.2H, H-4, anomer  $\beta$ ), 3.87 (dd, 1H,  $J_{3,2}$  = 11.38 Hz,  $J_{3,4}$  = 2.15 Hz, H-3, anomer  $\alpha$ ), 3.73–7.49 (m, 3.6H, H-6a, H-6b, H-5 anomers  $\alpha/\beta$ ).

## Compound (21)



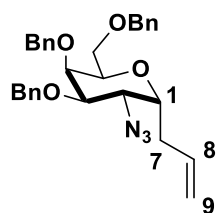
To a solution of **(20)** (1.8 g, 3.46 mmol) in AcOH (17.3 mL) was added Et<sub>3</sub>N (2.5 mL, 17.3 mmol) at RT. After 1h of stirring at RT, the solvent was removed under vacuum. The residue was taken up in 200 mL of AcOEt. The organic layers were washed with saturated aqueous solution of NaHCO<sub>3</sub> (2 x 50 mL), H<sub>2</sub>O (2 x 50 mL), brine, dried over anhydrous MgSO<sub>4</sub>, filtered, and concentrated under vacuum. The crude product was purified by flash chromatography on silica gel (Cy/AcOEt 90:10), to afford **(21)** as a white oil (1.40 g, 2.70 mmol, yield 78%,  $\alpha/\beta$  7/3).

**<sup>1</sup>H NMR** (300 MHz, CDCl<sub>3</sub>, 298K):  $\delta$  = 7.48-7.25 (m, 21H, BnO, anomer  $\alpha/\beta$ ), 6.26 (d, 1H,  $J_{1,2}$  = 3.8 Hz, H-1, anomer  $\alpha$ ), 5.57 (d, 0.4H,  $J_{1,2}$  = 8.6 Hz, H-1, anomer  $\beta$ ), 4.6-4.39 (m, 9H, BnO, anomer  $\alpha/\beta$ ), 4.13 (m, 1H, H-4 anomer  $\alpha$ ), 4.12 (dd, 1H,  $J_{2,3}$  = 10.5 Hz,  $J_{1,2}$  = 3.8 Hz, H-2, anomer  $\alpha$ ), 4.05-3.94 (m, 2.8H, H-5  $\alpha$ , H-2  $\beta$ , H-4  $\beta$ ), 3.93 (dd, 1H,  $J_{3,2}$  = 10.6 Hz,  $J_{3,4}$  = 2.6 Hz, H-3, anomer  $\alpha$ ), 3.71-3.50 (m, 3.2H, H-6 anomer  $\alpha/\beta$ , H-5  $\beta$ ), 3.46 (dd, 1H,  $J_{2,3}$  = 10.3 Hz,  $J_{3,4}$  = 2.7 Hz, H-3, anomer  $\beta$ ). (s, 3H, AcO, anomer  $\alpha$ ), (s, 3H, AcO, anomer  $\beta$ ).

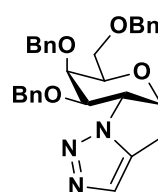
**<sup>13</sup>C NMR** (75 MHz, CDCl<sub>3</sub>, 298K):  $\delta$  = 169.2 (C, AcO, anomer  $\beta$ ), 169.2 (C, AcO, anomer  $\alpha$ ), 138.3 (C, BnO, anomer  $\alpha/\beta$ ), 137.8 (C, BnO, anomer  $\alpha$ ), 137.8 (C, BnO, anomer  $\beta$ ), 137.5 (C, BnO, anomer  $\alpha$ ), 137.5 (C, BnO, anomer  $\beta$ ), 129.0 – 127.7 (30 x CH, BnO, anomer  $\alpha/\beta$ ) 93.2 (CH, C-1, anomer  $\beta$ ), 91.3 (CH, C-1, anomer  $\alpha$ ), 81.0 (CH, C-3, anomer  $\beta$ ), 77.8 (CH, C-3, anomer  $\alpha$ ), 75.1 (CH<sub>2</sub>, BnO, anomer  $\alpha$ ), 75.0 (CH<sub>2</sub>, BnO, anomer  $\beta$ ), 74.4 (CH, C-5, anomer  $\beta$ ), 73.8 (CH<sub>2</sub>, BnO, anomer  $\alpha$ ), 73.7 (CH<sub>2</sub>, BnO, anomer  $\beta$ ), 72.9 (CH, C-4, anomer  $\alpha$ ), 72.7 (CH<sub>2</sub>, BnO, anomer  $\beta$ ), 72.3 (CH<sub>2</sub>, BnO, anomer  $\alpha$ ), 72.0 (CH, C-5, anomer  $\alpha$ ), 71.8 (CH, C-4, anomer  $\beta$ ) 68.2 (CH<sub>2</sub>, C-6, anomer  $\alpha$ ), 67.8 (CH<sub>2</sub>, C-6, anomer  $\beta$ ), 62.1 (CH, C-2, anomer  $\beta$ ), 59.1 (CH, C-2, anomer  $\alpha$ ), 22.2 (CH<sub>3</sub>, AcO, anomer  $\alpha$ ), 21.1 (CH<sub>3</sub>, AcO, anomer  $\beta$ ).

**HRMS ESI+** : m/z calcd for C<sub>29</sub>H<sub>31</sub>N<sub>3</sub>O<sub>6</sub>Na [M+Na]<sup>+</sup>: 540.2111, found 540.2118.

### Compound (22) and (23)



Desired product

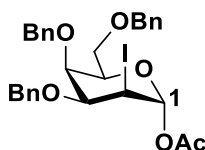


Side product

To a solution of **(21)** (920 mg, 1.779 mmol) in dry ACN (18 mL) containing oven-dried powdered 4 Å molecular sieves (1 g) was added ATMS (846  $\mu$ L, 5.338 mmol) and the mixture was stirred at 0°C under N<sub>2</sub> for 15 min. TMSOTf (642  $\mu$ L, 3.558 mmol) was then added dropwise and the stirring continued for 1 h at 0°C. The mixture was diluted with AcOEt and washed with an aqueous saturated solution of NaHCO<sub>3</sub>, H<sub>2</sub>O, brine, dried over anhydrous MgSO<sub>4</sub>, filtered, and concentrated under vacuum. The crude of the allyl C-glycoside **(22)** was immediately used in the next step without further purification in order to avoid its degradation via an intramolecular 1,3-dipolar cyclization, **(23)**.<sup>162</sup>

Compound **(22)**: <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>, 298K):  $\delta$  = 7.24-7.44 (m, 15H, BnO), 5.76 (dddd, 1H,  $J_3$  = 6.9 Hz,  $J_3$  = 6.9 Hz,  $J_3$  = 10.2 Hz,  $J_3$  = 17.2 Hz, H-8), 5.02-5.15 (m, 2H, H-9a, H-9b), 4.79 (d, 1H,  $J_3$  = 11.7 Hz, BnO), 4.75 (d, 1H,  $J_3$  = 11.7 Hz, BnO), 4.64 (d, 1H,  $J_3$  = 11.7 Hz, BnO), 4.55 (d, 1H,  $J_3$  = 11.7 Hz, BnO), 4.51 (d, 1H,  $J_3$  = 11.9 Hz, BnO), 4.45 (d, 1H,  $J_3$  = 11.9 Hz, BnO), 3.97-4.01 (m, 3H), 3.90 (m, 1H, H-5), 3.54-3.72 (m, 3H), 2.32 (m, 2H, H-7a, H-7b).

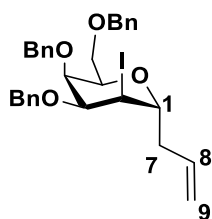
### Compound (24)



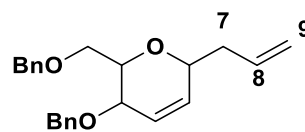
To a solution of **(23)** (560 mg, 1.34 mmol) in dry DCM (13 mL), were added BAIB (258 mg, 0.80 mmol), and I<sub>2</sub> (409 mg, 1.61 mmol) at RT. After 1 h of stirring, the reaction mixture was quenched with saturated solution of Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub>, (2 x 10 mL). The aqueous layer was extracted with DCM (3 X 20 mL), then the combined organic layers were washed with saturated aqueous solution of NaHCO<sub>3</sub> (2 x 30 mL), H<sub>2</sub>O (2 x 30 mL), brine, dried over anhydrous MgSO<sub>4</sub>, filtered, and concentrated under vacuum. The crude product was purified by flash chromatography on silica gel (Cy/AcOEt 85:15), to afford **(24)** as a white oil (626 mg, 1.04 mmol, yield 78%). Spectroscopic data were in concordance with those described in the bibliography.<sup>165</sup>

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>, 298K):  $\delta$  = 7.48–7.21 (m, 15H, CH BnO), 6.50 (d, 1H,  $J_{1,2}$  = 2.6 Hz, H-1), 5.02 (d, 1H,  $J_3$  = 11.8 Hz CH<sub>2</sub> BnO), 4.75 (d, 1H,  $J_3$  = 11.4 Hz CH<sub>2</sub> BnO), 4.57 (d, 1H,  $J_3$  = 11.8 Hz CH<sub>2</sub> BnO), 4.54 (d, 1H,  $J_3$  = 11.4 Hz CH<sub>2</sub> BnO), 4.50 (d, 1H,  $J_3$  = 11.7 Hz CH<sub>2</sub> BnO), 4.41 (d, 1H,  $J_3$  = 11.7 Hz CH<sub>2</sub> BnO), 4.33–4.28 (m, 1H), 4.21–4.15 (m, 1H), 4.02–3.98 (m, 1H), 3.73–3.68 (m, 1H), 3.52–3.47 (m, 1H). 2.05 (s, 3H, AcO).

### Compound (25) and (26)



Not detected



Detected

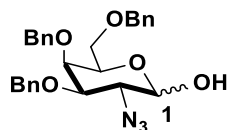
To a solution of compound **(24)** (450 mg, 0.75 mmol) in ACN (7.5 mL) and ATMS (356  $\mu$ L, 2.24 mmol) was added, at 0°C under N<sub>2</sub> atmosphere, BF<sub>3</sub>OEt<sub>2</sub> (280  $\mu$ L, 2.24 mmol). After 16h of stirring at RT, the mixture was quenched by the addition of 20 ml saturated aqueous solution of NaHCO<sub>3</sub> (2 x 50 mL). The aqueous layer was extracted with DCM (3 X 30 mL), then the combined organic layers were washed with saturated aqueous solution of NaHCO<sub>3</sub> (2 x 50 mL), H<sub>2</sub>O (2 x 50 mL), brine, dried over anhydrous MgSO<sub>4</sub>, filtered, and concentrated under vacuum. The crude product was purified by flash chromatography on silica gel (Cy/AcOEt 80:20), to afford **(26)** as a white oil (157 mg, 0.45 mmol, 60% yield). Not detection of the desired compound (25).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>, 298K):  $\delta$  = 7.40-7.24 (m, 10H, CH, BnO), 6.00-5.80 (m, 3H, H-2, H-3, H-8), 5.17-5.02 (m, 2H, H-9), 4.67-4.52 (m, 4H, CH<sub>2</sub> BnO), 4.34-4.29 (m, 1H, H-1), 4.16-4.08 (m, 1H, H-5), 3.92-3.87 (m, 1H, H-4), 3.82-3.70 (m, 2H, H-6), 3.50-3.38 (m, 1H, H-7a), 3.34-3.23 (m, 1H, H-7b).

<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>, 298K):  $\delta$  = 138.73 (C, BnO), 138.57 (C, BnO), 134.52 (CH, C-8), 133.42 (CH, C-3 or C-4), 128.73-127.4 (10 x CH, BnO), 124.20 (CH, C-3 or C-4), 117.46 (CH<sub>2</sub>, C-9), 73.59 (CH<sub>2</sub>, BnO), 71.92 (CH, C-1), 71.39 (CH<sub>2</sub>, C-5), 70.94 (CH<sub>2</sub>, OBn), 68.99 (CH<sub>2</sub>, C-6), 68.69 (CH<sub>2</sub>, C-4), 37.67 (CH<sub>2</sub>, C-7).

HRMS ESI+ : m/z calcd for C<sub>23</sub>H<sub>26</sub>O<sub>3</sub>Na [M+Na]<sup>+</sup>: 373.1777 found: 373.1780.

## Compound (27)



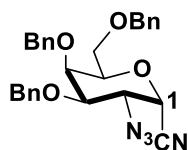
To a solution of compound (**19**) (250 mg, 0.60 mmol) and TMSN<sub>3</sub> (240 μL, 1.80 mmol) in DCM (9.6 mL) were added, at 0°C and sequentially, PIFA (430 mg, 1.20 mmol), TEMPO (19 mg, 0.12 mmol), Bu<sub>4</sub>NHSO<sub>4</sub> (40 mg, 0.12 mmol) and H<sub>2</sub>O (540 μL, 30.0 mmol). After stirring for 1.5 h at 0 °C, the reaction mixture was quenched by the addition of a saturated aqueous solution of NaHCO<sub>3</sub> (2 x 10 mL). The aqueous layer was extracted with DCM (3 X 20 mL), then the combined organic layers were washed with saturated aqueous solution of NaHCO<sub>3</sub> (2 x 30 mL), H<sub>2</sub>O (2 x 30 mL), brine, dried over anhydrous MgSO<sub>4</sub>, filtered, and concentrated under vacuum. The crude product was purified by flash chromatography on silica gel (Cy/AcOEt 85:15), to afford (**27**) as a white oil (200 mg, 0.42 mmol, yield 70%, α/β 75/25).

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>, 298K): δ = 7.46-7.20 (m, 23H, CH, BnO, anomer α/β), 5.32 (d, 1H, J<sub>1,2</sub> = 2.3 Hz, H-1, anomer α), 4.95-4.85 (m, 1.5H, BnO, anomer α/β), 4.79-4.67 (m, 5H, BnO, anomer α/β, H-1 anomer β), 4.20-4.12 (m, 1H, anomer α), 4.02-3.70 (m, 4H, anomer α/β), 3.64-3.31 (m, 4H, anomer α/β).

<sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>, 298K): δ = 138.3-137.5 (C, BnO, anomer α/β), 129.0-127.5 (CH, BnO, anomer α/β) 96.62 (CH, C-1, anomer β), 92.56 (CH, C-1, anomer α), 75.0-72.1 (CH<sub>2</sub>, BnO, CH carbohydrate, anomer α/β), 69.87, 69.31, 68.8, 64.8, 60.51 (CH, CH<sub>2</sub> carbohydrate, anomer α/β).

HRMS ESI+ : m/z calcd for C<sub>27</sub>H<sub>29</sub>N<sub>3</sub>O<sub>5</sub>Na [M+Na]<sup>+</sup>: 498.2005, found 499.2006.

### Compound (28)



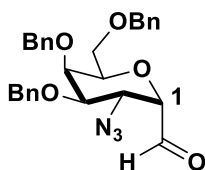
To a solution of compound **(21)** (4 g, 7.73 mmol) in DCM (80 mL) was added, at 0°C under N<sub>2</sub> atmosphere, TMSCN (2.90 mL, 23.19 mmol) followed by TMSOTf (4.90 mL, 27.05 mmol). After 16 h of stirring at RT, the mixture was diluted with 200 mL of AcOEt and the organic layers were washed with saturated aqueous solution of NaHCO<sub>3</sub> (2 x 50 mL), H<sub>2</sub>O (2 x 50 mL), brine, dried over anhydrous MgSO<sub>4</sub>, filtered, and concentrated under vacuum. The crude was purified by flash chromatography on silica gel (Cy/AcOEt 90:10), to afford **(28)** as a colorless oil (1.61 g, 3.32 mmol, yield 43%).

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>, 298K):  $\delta$  = 7.46-7.21 (m, 15H, CH, BnO), 4.86 (d, 1H,  $J_3 = 11.2$  Hz, CH<sub>2</sub> BnO), 4.84 (d, 1H,  $J_{1,2} = 5.9$  Hz, H-1), 4.75 (s, 2H, CH<sub>2</sub> BnO), 4.51 (d, 1H,  $J_3 = 11.2$  Hz, CH<sub>2</sub> BnO), 4.50 (d, 1H,  $J_3 = 11.8$  Hz, CH<sub>2</sub> BnO), 4.43 (d, 1H,  $J_3 = 11.8$  Hz, CH<sub>2</sub> BnO), 4.2 (dd, 1H,  $J_{2,3} = 10.3$  Hz,  $J_{1,2} = 5.95$  Hz, H-2), 4.05 (dd, 1H,  $J_{3,4} = 2.6$  Hz,  $J_{4,5} = 0.7$  Hz, H-4), 4.00 (ddd, 1H,  $J_{5,6} = 6.5$  Hz,  $J_{5,6} = 6.5$  Hz,  $J_{4,5} = 0.7$  Hz, H-5), 3.8 (dd, 1H,  $J_{3,2} = 10.3$  Hz,  $J_{3,4} = 2.6$  Hz, H-3), 3.57 (d, 2H,  $J_{5,6} = 6.5$  Hz, H-6).

<sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>, 298K):  $\delta$  = 137.9 (C, BnO), 137.5 (C, BnO), 137.0 (C, BnO), 128.7, 128.6, 128.5, 128.3, 128.2, 128.1, 128.0 (CH, BnO), 114.9 (C, CN), 80.1 (CH, C-3), 75.3 (CH, C-5), 75.2 (CH<sub>2</sub>, BnO), 73.7 (CH<sub>2</sub>, BnO), 72.7 (CH<sub>2</sub>, BnO), 72.3 (CH, C-4), 67.8 (CH, C-6), 67.3 (CH, C-1), 57.3 (CH, C-2).

HRMS ESI+ : m/z calcd for C<sub>28</sub>H<sub>28</sub>N<sub>4</sub>O<sub>4</sub>Na [M+Na]<sup>+</sup>: 507.2026 found: 507.2008.

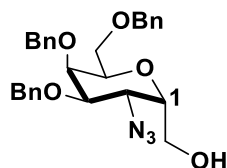
### Compound (29)



To a solution of **(28)** (250 mg, 0.52 mmol) in dry DCM (26 mL), was added, at -78°C under N<sub>2</sub> atmosphere, 516  $\mu$ L a solution of DIBAL-H (1M in DCM). The mixture was stirred at -78°C for 2h, then was quenched by aqueous solution of HCl (1 M) and stirred over 30 minutes. The mixture was warmed up at RT and diluted with DCM (15 mL). The organic layer was washed with H<sub>2</sub>O (3 x 30 mL), brine, dried over anhydrous MgSO<sub>4</sub>, filtered and concentrated under vacuum to afford the aldehyde **(29)** derivate without further purification (251 mg, 0.52 mmol). NMR was not recorded because the product was not stable.



### Compound (30)



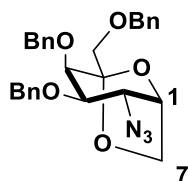
To a solution of **(29)** (250 mg, 0.52 mmol) in dry MeOH (13 mL) was added, at -20°C under N<sub>2</sub> atmosphere, NaBH<sub>4</sub> (29.3 mg 0.77 mmol). After 1 h of stirring at -20°C, the reaction mixture was quenched by saturated aqueous NH<sub>4</sub>Cl and stirred over 30 minutes. The mixture was warmed up at RT and diluted with DCM (15 mL). The aqueous layers were extracted with DCM (3 x 10 mL) then, the combined organic layer was washed with H<sub>2</sub>O (3 x 30 mL), brine, dried over anhydrous MgSO<sub>4</sub>, filtered and concentrated under vacuum. The crude was purified by flash chromatography on silica gel (Cy/AcOEt 60:40) to afford the pure product **(30)** as a white solid (25 mg, 0.05 mmol, yield 10% over two steps).

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>, 298K):  $\delta$  = 7.44-7.30 (m, 15H, CH, BnO), 4.92 (d, 1H,  $J_3$  = 11.9 Hz, CH<sub>2</sub> BnO), 4.80 (d, 1H,  $J_3$  = 11.4 Hz, CH<sub>2</sub> BnO), 4.74-4.64 (m, 2H, CH<sub>2</sub> BnO), 4.60 (d, 1H,  $J_3$  = 11.2 Hz, CH<sub>2</sub> BnO), 4.52 (d, 1H,  $J_3$  = 11.2 Hz, CH<sub>2</sub> BnO), 4.30-4.12 (m, 3H), 4.10-3.96 (m, 2H), 3.89-3.70 (m, 4H), (bs, 1H, OH).

<sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>, 298K):  $\delta$  = in progress

HRMS ESI+ : m/z calcd for C<sub>28</sub>H<sub>31</sub>N<sub>3</sub>O<sub>5</sub>Na [M+Na]<sup>+</sup>: 512.2161 found: 512.2166.

### Compound (31)



To a solution of **(30)** (22 mg, 0.045 mmol) in dry DCM (2.25 mL) were added, at RT under N<sub>2</sub> atmosphere, iodobenzene (20 mg, 0.09 mmol) and I<sub>2</sub> (11.4 mg, 0.045 mmol). The mixture was irradiated with a tungsten-filament lamp for 3 h at reflux. The reaction was cooled to RT, quenched with a saturated solution of Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub> (10 mL) and stirred 10 minutes. The aqueous layers were extracted with DCM (3 x 10 mL) and the combined organic layer was washed with H<sub>2</sub>O (3 x 30 mL), brine, dried over anhydrous MgSO<sub>4</sub>, filtered and concentrated under vacuum. The crude was purified by flash chromatography on silica gel (Cy/AcOEt 90:10) to afford the pure product **(31)** as a white solid (9 mg, 0.05 mmol, yield 41% over two steps).

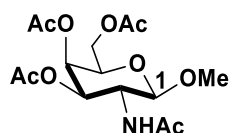
<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>, 298K):  $\delta$  = 7.36-7.10 (m, 15H, CH, BnO), 4.72 (d, 1H,  $J_3$  = 11.1 Hz, CH<sub>2</sub> BnO), 4.63 (d, 1H,  $J_3$  = 11.3 Hz, CH<sub>2</sub> BnO), 4.58 (d, 1H,  $J_3$  = 11.3 Hz, CH<sub>2</sub> BnO), 4.52 (d, 1H,  $J_3$  = 11.9 Hz, CH<sub>2</sub> BnO), 4.41 (d, 1H,  $J_3$  = 11.9 Hz, CH<sub>2</sub> BnO), 4.40 (d, 1H,  $J_3$  = 11.1 Hz, CH<sub>2</sub> BnO), 4.35-4.31 (m, 1H, H-4), 4.06-4.02 (m, 1H), 4.00-3.93 (m, 1H), 3.85-3.80 (m, 1H), 3.79 (s, 1H,  $J_{6a,6b}$  = 10.07 Hz, H-6a), 3.74-3.68 (m, 1H), 3.67-3.62 (m, 1H), 3.36 (s, 1H,  $J_{6a,6b}$  = 10.07 Hz, H-6b).

<sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>, 298K):  $\delta$  = in progress

HRMS ESI<sup>+</sup> : m/z calcd for C<sub>28</sub>H<sub>30</sub>N<sub>3</sub>O<sub>5</sub> [M+H]<sup>+</sup>: 488.2185 found: 488.2183.

## D. Ligand PMS-03

### Compound (32)



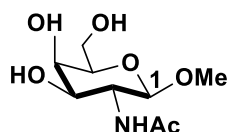
To a solution of **(17)** (1 g, 3.04 mmol) in dry MeOH (6 mL) containing 4Å MS, was added at RT under N<sub>2</sub> atmosphere CuCl<sub>2</sub> (408g, 3.04 mmol). The reaction mixture was stirred for 5 h at 55°C. The solvent was removed under vacuum and the residue was taken up in 100mL of DCM. The organic layers were washed with saturated aqueous NaHCO<sub>3</sub> (40 ml), H<sub>2</sub>O (2 x 40 mL), brine, dried over anhydrous MgSO<sub>4</sub>, filtered, concentrated under vacuum. The crude was purified by flash chromatography on silica gel (DCM/MeOH 95:05) to afford the pure product **(32)** as a white solid (547 mg, 1.52 mmol, yield 52%).

<sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>, 298K):  $\delta$  = 5.84 (d, 1H,  $J_{\text{NH},2}$  = 8.7 Hz, NH), 5.34 (dd, 1H,  $J_{3,4}$  = 3.2 Hz,  $J_{4,5}$  = 0.96 Hz, H-4), 5.26 (dd, 1H,  $J_{3,2}$  = 11.2 Hz,  $J_{3,4}$  = 3.2 Hz, H-3), 4.59 (d, 1H,  $J_{1,2}$  = 8.4 Hz, H-1), 4.14 (m, 2H, H-6), 3.96 (ddd, 1H,  $J_{2,3}$  = 11.2 Hz,  $J_{\text{NH},2}$  = 8.7 Hz,  $J_{2,1}$  = 8.4 Hz, H-2), 3.92 (m, 1H, H-5), 3.49 (s, 3H, MeO), 2.12 (s, 3H, AcO), 2.02 (s, 3H, AcO), 1.97 (s, 3H, AcO), 1.94 (s, 3H, NHAc).

<sup>13</sup>C NMR (75 MHz, MeOH-D<sub>4</sub>, 298K):  $\delta$  = 170.7-170.59-170.42 (C, NHAc-AcO), 101.9 (CH, C-1), 70.7 (CH, C-5), 70.0 (CH, C-3), 68.9 (CH, C-4), 61.6 (CH<sub>2</sub>, C-6), 56.9 (CH<sub>3</sub>, MeO), 51.5 (CH, C-2), 23.5 (CH<sub>3</sub>, NHAc) 20.8 (3xCH<sub>3</sub>, AcO).

HRMS ESI+:  $m/z$  calcd for C<sub>9</sub>H<sub>17</sub>NO<sub>6</sub>Na. [M+Na]<sup>+</sup>: 362.1447 found: 362.1451.

### Compound (33)



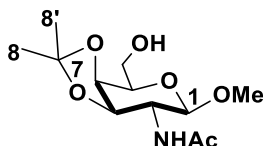
To a solution of **(32)** (500 mg, 1.38 mmol) in dry MeOH (10ml) was added, at 0°C under N<sub>2</sub> atmosphere, NaOMe in MeOH (30 wt% solution in MeOH, 40μL, 0,278 mmol). After 4 h of stirring at RT, DOWEX-50 acid resin (30 mg) was added to neutralize the mixture. The mixture was filtered and concentrated under vacuum to afford the desired compound **(33)** as a white solid (310 mg, 1.32 mmol, yield 96% as a crude).

<sup>1</sup>H NMR (300 MHz, MeOH-D<sub>4</sub>, 298K): δ = 4.28 (d, 1H, *J*<sub>1,2</sub> = 8.4 Hz, H-1), 3.91 (dd, 1H, *J*<sub>1,2</sub> = 8.4 Hz, *J*<sub>2,3</sub> = 10.8 Hz, H-2), 3.83 (dd, 1H, *J*<sub>4,5</sub> = 0.8 Hz, *J*<sub>3,4</sub> = 3.2 Hz, H-4), 3.76 (m, 2H, H-6), 3.55 (dd, 1H, *J*<sub>3,4</sub> = 3.2 Hz, *J*<sub>2,3</sub> = 10.8 Hz, H-3), 3.49 (m, 1H, H-5), 3.46 (s, 3H, MeO), 1.98 (s, 3H, NHAc).

<sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>, 298K): δ = 174.2 (C, NHAc), 103.9 (CH, C-1), 76.7 (CH, C-5), 73.5 (CH, C-3), 69.7 (CH, C-4), 62.5 (CH<sub>2</sub>, C-6), 56.9 (CH<sub>3</sub>, MeO), 54.2 (CH, C-2), 22.9 (CH<sub>3</sub>, NHAc).

HRMS ESI+: *m/z* calcd for C<sub>9</sub>H<sub>17</sub>NO<sub>6</sub>Na. [M+Na]<sup>+</sup>: 258.0956 found: 258.0954.

### Compound (34)



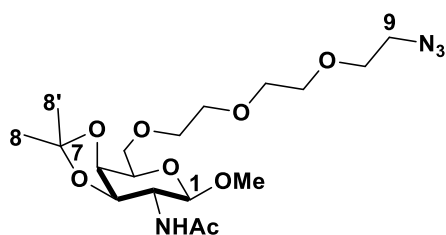
To a solution of compound **(33)** (200 mg, 0.85 mmol) in DMF (6 mL) was added 2,2-dimethoxypropane (634 μL mL, 5.10mmol) followed by (+/-)-camphor-10-sulphonic acid (100 mg, 0.42mmol). The reaction mixture was stirred at 70 °C for 24 h, cooled to RT and treated with MeOH (1 mL, 25.5 mmol). The mixture was stirred at RT for 30 minutes and then neutralized with TEA (60 μL, 0.425 mmol). The solvent was evaporated and the residue was co-evaporated 3 times with toluene. The resulting crude material was purified by flash chromatography on silica gel (AcOEt/MeOH 90:10) to afford compound **(34)** as a white solid (157 mg, 0.60 mmol, yield 67%).

<sup>1</sup>H NMR (400 MHz, MeOH-D<sub>4</sub>, 303K): δ = 4.30 (d, 1H, *J*<sub>1,2</sub> = 8.4 Hz, H-1), 4.21-4.13 (m, 2H, H-3, H-4), 3.89-3.71 (m, 4H, H-2,H-5,H-6), 3.45 (s, 3H, MeO), 1.96 (s, 3H, NHAc), 1.51 (s, 3H, H-8), 1.31 (s, 3H, H-8').

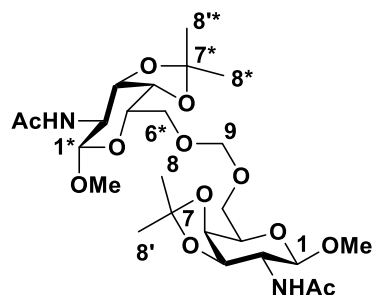
<sup>13</sup>C NMR (100 MHz, MeOH-D<sub>4</sub>, 303K): δ = 173.44 (C, NHAc), 111.14 (C, C-7), 103.1 (CH, C-1), 78.4 (CH, C-3), 75.1 (CH, C-5), 74.6 (CH, C-4), 62.6 (CH<sub>2</sub>, C-6), 56.9 (CH<sub>3</sub>, MeO), 55.6 (CH, C-2), 28.3 (CH<sub>3</sub>, C-8), 26.6 (CH<sub>3</sub>, C-8'), 22.9 (CH<sub>3</sub>, NHAc).

HRMS ESI+: *m/z* calcd for C<sub>12</sub>H<sub>22</sub>NO<sub>6</sub>. [M+H]<sup>+</sup>: 276.1441 found: 276.1447.

## Compound (35) & (36)



(35) (desired product)



(36) (by-product)

To a solution of **(34)** (180 mg, 0.66 mmol) in DCM (7 mL) containing aqueous NaOH (3.5 mL of 12.5 M solution) and 5-crow-15 (18  $\mu$ g, 0.07 mmol) was added the compound **(PEG-2)** (571 mg, 1.98 mmol). The reaction mixture was stirred vigorously at 55 °C. After 72h of stirring, the mixture was cooled to RT, diluted with DCM (30 ml) and washed with H<sub>2</sub>O (3 x 20mL). The combined organic layer was washed with H<sub>2</sub>O, brine, dried over anhydrous MgSO<sub>4</sub>, filtered and concentrated under vacuum. The crude product was purified by flash chromatography on silica gel (AcOEt 100% - AcOEt/MeOH 90:10) to afford the desired compound **(35)** as a sticky solid (136 mg, 0.32 mmol, yield 48%) and the by-product **(36)** as a sticky solid (146 mg, 0.26 mmol, yield 40%).

**Compound (35):** <sup>1</sup>H NMR (300 MHz, MeOH-D<sub>4</sub>, 298K):  $\delta$  = 4.30 (d, 1H,  $J_{1,2}$  = 8.7 Hz, H-1), 4.21-4.10 (m, 2H, H-3, H-4), 4.01 (m, 1H, H-5), 3.84-3.61 (m, 4H, H-2, H-5, H-6, CH<sub>2</sub> PEG), 3.43 (s, 3H, MeO), 3.37 (t, 2H,  $J_{\text{CH}_2 \text{ peg}, 9}$  = 5.0 Hz, H-9), 1.96 (s, 3H, NHAc), 1.50 (s, 3H, H-8), 1.31 (s, 3H, H-8').

<sup>13</sup>C NMR (75 MHz, MeOH-D<sub>4</sub>, 298K):  $\delta$  = 173.4 (C, NHAc), 111.1 (C, C-7), 103.0 (CH, C-1), 78.3 (CH, C-3), 74.7 (CH, C-4), 73.4 (CH, C-5), 72.0, 71.7, 71.6, 71.6, 71.5, 71.2 (CH<sub>2</sub>, peg, C-6), 56.9 (CH<sub>2</sub>, MeO), 55.5 (CH<sub>3</sub>, C-2), 51.8 (CH<sub>2</sub>, C-9), 28.3 (CH<sub>3</sub>, C-8), 26.7 (CH<sub>3</sub>, C-8'), 22.9 (CH<sub>3</sub>, NHAc).

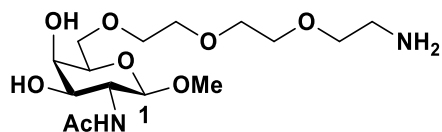
HRMS ESI+: m/z calcd for C<sub>18</sub>H<sub>32</sub>N<sub>4</sub>O<sub>8</sub>Na. [M+Na]<sup>+</sup>: 455.2132 found: 455.2118.

**Compound (36):** <sup>1</sup>H NMR (400 MHz, MeOH-D<sub>4</sub>, 303K):  $\delta$  = 4.70 (s, 2H, H-9), 4.20 (d, 2H,  $J_{1,2}$  = 8.9 Hz, H-1, H-1\*), 4.21-4.07 (m, 4H, H-3, H-3\*, H-4, H-4\*), 3.96 (ddd, 2H,  $J_{5,6}$  = 7.3 Hz,  $J_{5,6}$  = 5.4 Hz,  $J_{5,4}$  = 2.0 Hz, H-5, H-5\*), 3.81-3.62 (m, 6H, H-2, H-2\*, H-6, H-6\*), 3.34 (s, 6H, MeO, MeO\*), 1.86 (s, 6H, NHAc, NHAc\*), 1.41 (s, 6H, H-8), 1.21 (s, 6H, H-8', H-8'\*).

<sup>13</sup>C NMR (100 MHz, MeOH-D<sub>4</sub>, 303K):  $\delta$  = 173.47 (2 x C, NHAc, NHAc\*), 111.1 (2 x C, C-7, C-7\*), 103.0 (2 x CH, C-1, C-1'), 96.8 (CH<sub>2</sub>, C-9), 78.3 (2 x CH, C-3, C-3\*), 74.7 (2 x CH, C-4, C-4\*), 73.2 (2 x CH, C-5, C-5\*), 67.6 (2 x CH<sub>2</sub>, C-6, C-6\*), 56.9 (2 x CH<sub>3</sub>, MeO, MeO\*), 55.5 (2 x CH, C-2, C-2\*), 28.4 (2 x CH<sub>3</sub>, C-8, C-8\*), 26.7 (CH<sub>3</sub>, C-8', C-8'\*), 22.9 (2 x CH<sub>3</sub>, NHAc, NHAc\*).

HRMS ESI+: m/z calcd for C<sub>25</sub>H<sub>42</sub>N<sub>2</sub>O<sub>12</sub>Na. [M+Na]<sup>+</sup>: 585.2640 found: 585.2635.

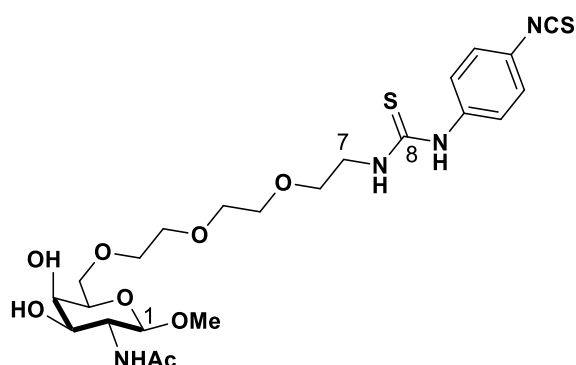
### Compound (37)



A solution of the compound **(35)** (60 mg, 0.14 mmol) in 2.5 mL of a mixture of AcOH/MeOH (1:1) was stirred at 50°C for 5 h. After that period of time TLC showed complete consumption of the starting material and the solvents were evaporated under vacuum to give the deprotected intermediate. To a solution of this crude in <sup>i</sup>PrOH (3 mL) was added Pd/C (10% weight). The mixture was stirred at RT under an atmosphere of H<sub>2</sub> (1 atm) for 2 h. The catalyst was filtered through a Celite® pad and thoroughly rinsed with MeOH and then H<sub>2</sub>O. The solvent was evaporated under vacuum to afford **(37)** as a white solid (44 mg, 0.12 mmol, yield 90% as a crude).

<sup>1</sup>H NMR (300 MHz, MeOH-D<sub>4</sub>, 298K): δ = 4.30 (d, 1H, *J*<sub>1,2</sub> = 8.4 Hz, H-1), 3.92 (dd, 1H, *J*<sub>1,2</sub> = 8.4 Hz, *J*<sub>2,3</sub> = 10.6 Hz, H-2), 3.84 (dd, 1H, *J*<sub>4,5</sub> = 0.6 Hz, *J*<sub>3,4</sub> = 3.2 Hz, H-4), 3.78-3.61 (m, 11H, H-5, H-6, CH<sub>2</sub> PEG), 3.58 (dd, 1H, *J*<sub>3,4</sub> = 3.2 Hz, *J*<sub>2,3</sub> = 10.6 Hz, H-3), 3.54 (t, 2H, *J*<sub>3</sub> = 5.2 Hz, CH<sub>2</sub> PEG), 3.46 (s, 3H, MeO), 2.8 (t, *J*<sub>3,4</sub> = 5.3 Hz, 2H, H-9), 1.98 (s, 3H, NHAc). Just <sup>1</sup>H NMR has been described.

### Ligand PMS-03



A solution of **(37)** (44 mg, 0.12 mmol) in 6 mL of 1,4-dioxane / H<sub>2</sub>O 5:1 was added dropwise over 1 h to a solution of p-Phenylene diisothiocyanate (118.5 mg, 0.615 mmol) in 1,4-dioxane (2mL) at RT. After 1h of stirring at RT, the solvent was evaporated under vacuum. The resulting crude material was purified by flash chromatography on silica gel (DCM/MeOH 85:15) to afford the compound **(PMS-03)** as a white solid (35.5 mg, 0.06 mmol, yield 53%).

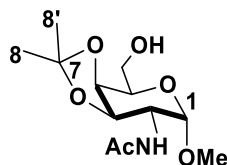
<sup>1</sup>H NMR (300 MHz, DMSO-D<sub>6</sub>, 303K):  $\delta$  = 9.81 (bs, 1H, H-8), 7.95 (bs, 1H, N<sub>8</sub>HNH), 6.65 (d, 1H, NHAc) 7.56 (m, 2H, CH aromatic), 7.38 (m, 2H, CH aromatic), 4.62 (m, 2H, OH), 4.16 (dd, 1H,  $J_{1,2}$  = 8.4 Hz, H-1), 7.79-7.37 (m, 18H, H-2, H-3, H-4, H-5, H-6, CH<sub>2</sub> PEG), 3.28 (s, 3H, MeO), 1.80 (s, 3H, NHAc).

<sup>13</sup>C NMR (75 MHz, DMSO-D<sub>6</sub>, 303K):  $\delta$  = 181.2 (C, C-8), 170.4 (C, NHAc), 151.7 (C, NCS), 140.2 (C, aromatic), 133.5 (C, aromatic), 127.1 (2\*CH, aromatic), 124.0 (2\*CH, aromatic), 103.0 (CH, C-1), 74.4, 72.4, 71.1, 71.0, 70.7, 70.6, 69.4, 68.9 (CH, C-3, C-4, C-5, CH<sub>2</sub> PEG, CH<sub>2</sub>, C-6), 56.2 (CH<sub>2</sub>, MeO), 52.5 (CH, C-2), 44.5 (CH<sub>2</sub>, C-7), 24.1 (CH<sub>3</sub>, NHAc).

HRMS ESI+: m/z calcd for C<sub>23</sub>H<sub>34</sub>N<sub>4</sub>O<sub>8</sub>NaS<sub>2</sub>. [M+Na]<sup>+</sup>: 581.1735 found: 581.1716.

## E. Ligand PMS-04

### Compound (39)



To a solution of **methyl 2-acetamido-2-deoxy-α-D-galactopyranoside** (200 mg, 0.85 mmol) in DMF (6 mL) was added 2,2-dimethoxypropane (634 μL, 5.10 mmol) followed by (+/-)-camphor-10-sulphonic acid (100 mg, 0.42 mmol). After 24 h of stirring at 70 °C, the mixture was cooled to RT and treated with MeOH (1 mL, 25.5 mmol). The mixture was stirred at RT for 30 minutes and then neutralized with TEA (60 μL, 0.425 mmol). The solvent was evaporated and the residue was co-evaporated 3 times with toluene. The resulting crude material was purified by flash chromatography on silica gel (AcOEt/MeOH 90:10) to afford the compound **(39)** as a white solid (180 mg, 0.65 mmol, yield 77%).

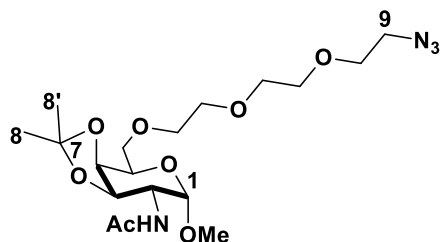
<sup>1</sup>H NMR (300 MHz, MeOH-D<sub>4</sub>, 298K): δ = 4.62 (d, 1H,  $J_{1,2} = 2.9$  Hz, H-1), 4.24 (dd, 1H,  $J_{4,5} = 2.6$  Hz,  $J_{3,4} = 4.5$  Hz, H-4), 4.20-4.07 (m, 2H, H-3, H-2), 4.01 (ddd, 1H,  $J_{4,5} = 2.6$  Hz,  $J_{5,6} = 6.2$  Hz, H-5), 3.78 (d, 2H,  $J_{5,6} = 6.2$  Hz, H-6), 3.39 (s, 3H, MeO), 1.98 (s, 3H, NHAc), 1.48 (s, 3H, H-8), 1.32 (s, 3H, H-8').

<sup>13</sup>C NMR (75 MHz, MeOH-D<sub>4</sub>, 298K): δ = 173.5 (C, NHAc), 110.5 (C, C-7), 100.1 (CH, C-1), 75.1 (CH, C-3), 74.2 (CH, C-4), 69.5 (CH, C-5), 62.7 (CH<sub>2</sub>, C-6), 55.8 (CH<sub>3</sub>, MeO), 52.5 (CH<sub>2</sub>, C-2), 28.4 (CH<sub>3</sub>, C-8), 26.8 (CH<sub>3</sub>, C-8'), 22.5 (CH<sub>3</sub>, NHAc).

HRMS ESI+: m/z calcd for C<sub>12</sub>H<sub>22</sub>NO<sub>6</sub>Na. [M+Na]<sup>+</sup>: 298.1269 found: 298.1267.



### Compound (40)



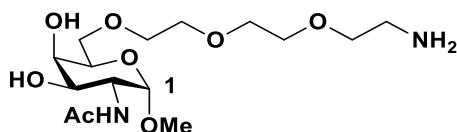
To a solution of **(39)** (200 mg, 0.726 mmol) in DCM (7 mL) containing aqueous NaOH (3.5 mL of 12.5M solution) and TBAI (130 mg, 0.726 mmol) was added the compound **(PEG-2)** (621 mg, 2.178 mmol). After 50 h of vigorous stirring at 55 °C, the reaction mixture was cooled RT. DCM (30 ml) and H<sub>2</sub>O (20mL) were added and the aqueous layers was extracted with DCM three time. The combined organic layer was washed with H<sub>2</sub>O, brine, dried over anhydrous MgSO<sub>4</sub>, filtered and concentrated under vacuum. The crude product was purified by flash chromatography on silica gel (AcOEt - AcOEt/MeOH 90:10) to afford compound **(40)** as a sticky solid (150 mg, 0.355 mmol, yield 55%).

<sup>1</sup>H NMR (300 MHz, MeOH-D<sub>4</sub>, 298K):  $\delta$  = 4.60 (d, 1H,  $J_{1,2}$  = 2.9 Hz, H-1), 4.24 (dd, 1H,  $J_{4,5}$  = 2.5 Hz,  $J_{3,4}$  = 4.4 Hz, H-4), 4.20-4.08 (m, 2H, H-2, H-3, H-5), 3.76 (m, 2H, H-6), 3.71.-3.61 (m, 10H, CH<sub>2</sub> PEG), 3.38 (brown triplet, 2H, H-9), 3.38 (s, 3H, MeO), 1.98 (s, 3H, NHAc), 1.48 (s, 3H, H-8), 1.32 (s, 3H, H-8').

<sup>13</sup>C NMR (75 MHz, MeOH-D<sub>4</sub>, 298K):  $\delta$  = 173.5 (C, NHAc), 110.5 (C, C-7), 100.2 (CH, C-1), 75.1 (CH, C-3), 74.4 (CH, C-4), 72.0, 71.7, 71.6, 71.6, 71.2 (CH<sub>2</sub>, peg, C-6), 67.7 (CH, C-5), 55.9 (CH<sub>3</sub>, MeO), 52.5 (CH<sub>2</sub>, C-2), 51.8 (CH<sub>2</sub>, C-9), 28.4 (CH<sub>3</sub>, C-8), 26.8 (CH<sub>3</sub>, C-8'), 22.5 (CH<sub>3</sub>, NHAc).

HRMS ESI+: m/z calcd for C<sub>18</sub>H<sub>32</sub>N<sub>4</sub>O<sub>8</sub>Na. [M+Na]<sup>+</sup>: 455.2126 found: 455.2118.

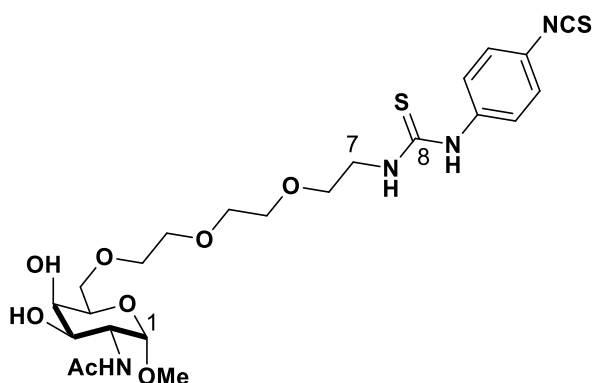
### Compound (41)



Compound **(40)** (50 mg, 0.12 mmol) in dissolved in 3 mL of an AcOH/MeOH (1:1) mixture and stirred at 50°C for 5 h. After that period of time TLC showed complete consumption of the starting material and the solvents were evaporated under vacuum to give a deprotected intermediate. To a solution of this crude in IPrOH (3 mL) was added Pd/C (10% weight). The mixture was stirred at RT under an atmosphere of H<sub>2</sub> (1 atm) for 2 h. The catalyst was filtered through a Celite<sup>®</sup> pad and thoroughly rinsed with MeOH and then H<sub>2</sub>O. The solvent was evaporated under vacuum to afford **(41)** as a white solid (40 mg, 0.11 mmol, yield 90% as a crude). Just <sup>1</sup>H NMR has been described.

<sup>1</sup>H NMR (300 MHz, MeOH-D<sub>4</sub>, 298K):  $\delta$  = 4.68 (d, 1H,  $J_{1,2}$  = 3.6 Hz, H-1), 4.27 (dd, 1H,  $J_{1,2}$  = 3.6 Hz,  $J_{2,3}$  = 11.0 Hz, H-2), 3.90 (ddd, 1H, , H-5), 3.89 (dd, 1H,  $J_{4,5}$  = 0.6 Hz,  $J_{3,4}$  = 3.2 Hz, H-4), 3.78-3.61 (m, 11H, H-3, H-6, CH<sub>2</sub> PEG), 3.53 (t, 2H,  $J_3$  = 5.2 Hz, CH<sub>2</sub> PEG), 3.37 (s, 3H, MeO), 2.8 (t,  $J_{3,4}$  = 5.2 Hz, 2H, H-9), 1.98 (s, 3H, NHAc).

### PMS-04



A solution of **(41)** (40 mg, 0.11 mmol) in 6 mL of 1,4-dioxane / H<sub>2</sub>O 5:1 was added dropwise over 1h to a solution of p-Phenylene diisothiocyanate (118.5 mg, 0.615 mmol) in 1,4-dioxane (2mL) at RT. After 1h of stirring at RT, the solvent was evaporated under vacuum. The resulting crude material was purified by flash chromatography on silica gel (DCM/MeOH 85:15) to afford compound **(PMS-04)** as a white solid (34 mg, 0.06 mmol, yield 55%).

<sup>1</sup>H NMR (300 MHz, DMSO-D<sub>6</sub>, 298K):  $\delta$  = 10.0 (bs, 1H, H-8), 8.13 (bs, 1H, NHSNH), 7.65 (d, 1H,  $J_{\text{NH},2}$  = 8.7 Hz, NHAc), 7.61 (m, 2H, CH aromatic), 7.37 (m, 2H, CH aromatic), 4.65 (m, 1H, OH), 4.53 (dd, 1H,  $J_{1,2}$  = 3.6 Hz, H-1), 4.50 (m, 1H, OH), 4.03 (ddd, 1H,  $J_{2,3}$  = 11.2 Hz,  $J_{\text{NH},2}$  = 8.4 Hz,  $J_{2,1}$  = 3.6 Hz, H-2), 3.73-3.45 (m, 17H, H-3, H-4, H-5, H-6, CH<sub>2</sub> PEG), 3.22 (s, 3H, MeO), 1.82 (s, 3H, NHAc).

<sup>13</sup>C NMR (75 MHz, DMSO-D<sub>6</sub>, 298K):  $\delta$  = 181.3 (C, C-8), 170.76 (C, NHAc), 153.4 (C, NCS), 140.3 (C, aromatic), 132.3 (C, aromatic), 127.3 (2\*CH, aromatic), 123.8 (2\*CH, aromatic), 99.3 (CH, C-1), 71.4, 71.0, 70.7, 70.5, 70.1, 69.4, 69.3, 68.2 (CH, C-3, C-4, C-5, CH<sub>2</sub> PEG, CH<sub>2</sub>, C-6), 55.4 (CH<sub>2</sub>, MeO), 50.6 (CH<sub>3</sub>, C-2), 44.5 (CH<sub>2</sub>, C-7), 23.6 (CH<sub>3</sub>, NHAc).

HRMS ESI+: m/z calcd for C<sub>23</sub>H<sub>34</sub>N<sub>4</sub>O<sub>8</sub>NaS<sub>2</sub>. [M+Na]<sup>+</sup>: 581.1727 found: 581.1716.

## F. Molecular modeling

### *a) Generation of the chemical library and preparation of the target*

The 70s were built using RDKit from their Lewis structures, then minimized to MP2/6-31G(d) using Gaussian16.[2] The compounds were treated with PROPKA3[3] and Ligprep[4] to define the different tautomers and protonation states at pH 7.4 using the OPLS3 force field implemented in Glide.<sup>195,196</sup>

We used the 5JQ1 structure from the PDB. The latter is the most relevant, because it has a resolution of 1.83 Å. The different protonation states of the structure were calculated using PROPKA3<sup>197</sup> using implicit solvation and the different missing hydrogens were added in accordance with these results. The system was minimized and then balanced for 50 ns under a harmonic stress of 20 kcal/mol\*Å<sup>2</sup> applied to all heavy atoms using NAMD 2.14 and the CHARMM36m protein force field.<sup>198</sup>

### *b) Docking*

The grids were defined taking as origin the center of the ZPF ligand (**PA4**) present in the 5JQ1 structure. A weight factor of 1.0 was applied to the Van der Waals radius (VdW) for atoms with a partial charge less than or equal to 0.25, while atoms with a partial charge of 0.15 were applied a factor of 0.8. Non-flat conformations of the amide bond were penalized while the rotation of the 5- and 6-member cycles was allowed. No hydrogen bond constraints were used. The prepared molecules were docked using Glide's extra precision mode (XP) present in the 2019-3 version of Schrödinger suite.

The exposures obtained were evaluated by GlideScore (GScore), a proprietary Schrodinger scoring function based on ChemScore, but incorporating several components including one describing steric clashes and penalties for polar fragments leaked into the protein. The GScore is given by the following equation:  $GScore = a * VdW + b * coul + Lipo + Hbond + Metal + Site + BuryP + RotB$

### *c) molecular dynamics*

We performed classical molecular dynamics calculations on the ZPF model compounds. We limited ourselves to this compound mainly due to lack of computing resources. Molecular Dynamics (MD) were performed with NAMD2.14<sup>199</sup> and the following force fields: CHARMM36m for the receiver[9] and a hybrid of CHARMM carbohydrate<sup>200</sup>/CGENFF for ZPF.<sup>201</sup> The systems were built using the best poses from the DFI as a starting point. The ligand-receptor complexes were solvated using the TIP3P water model<sup>202</sup> and a padding of 15 Å resulting in a 55x55x67 Å box. The systems were minimized, balanced for 1 ns and then launched in production for 200 ns in set NPT in triplicate. For each MD, unrelated interactions 1-3 and 1-4 were taken into account and were applied a weight factor of 1.0 with a switching distance of 12.0 Å. MDs were calculated at 310 K and with an integration factor of 2 fs.

## 2. Biology

Rabbit polyclonal anti-AAV capsid proteins antibody (cat N°61084), mouse monoclonal A20 antibody (cat N° 61055) were obtained from PROGEN Biotechnik. Anti-Fluorescein-AP Fab fragment antibody (cat N°11426338910) for the detection of fluorescein-labeled compounds was obtained from Sigma Aldrich. FITC-Soybean Agglutinin lectin (SBA), Wisteria Floribunda lectin (WFA), ASGPr lectin was purchased from Vector Laboratories. Nitrocellulose membrane were purchased and polyacrylamide gels purchased from Thermo Fisher. AAV preparations were run on polyacrylamide gel at 120 V for 3 hr 30 min.

### A. AAV2 production and purification

AAV2 vectors were produced from two plasmids: (i) pHelper, PDP2-KANA encoding AAV Rep2-Cap2 and adenovirus helper genes (E2A, VA RNA, and E4); and (ii) the pVector ssCAG-eGFP containing the ITRs. All vectors were produced by transient transfection of HEK293 cells using the calcium phosphate-HeBS method. AAV2 transfected cells were harvested 48h after transfection and treated with Triton-1% and benzonase (25 U/mL) for 1h at 37°C. The resulting bulk was subjected to freeze-thaw cycles to release vector particles. The cellular debris were removed by centrifugation at 2500 rpm for 15 min. Cell lysates were precipitated with PEG overnight and clarified by centrifugation at 4000 rpm for 1 h. The precipitates were then incubated with benzonase for 30 min at 37 °C and collected after centrifugation at 10,000 g for 10 min at 4°C. Vectors were purified by double cesium chloride (CsCl) gradient ultracentrifugation. The viral suspension was then subjected to four successive rounds of dialysis with mild stirring in a Slide-a-Lyzer cassette (Pierce) against dPBS (containing Ca<sup>++</sup> and Mg<sup>++</sup>).

### B. Coupling and purification

AAV2-GFP (1<sup>E</sup>13 vg, 2.49 nmol) were added to a solution of TBS buffer with 10% of DMSO, (pH 9.3) containing compounds **(PMS-01)**, **(PMS-02)**, **(PMS-03)** and **(PMS-04)** (3<sup>E</sup>5 and 3<sup>E</sup>6 equivalents), for final volume 1ml, then incubated for 4h at RT. The solutions containing the vectors were then dialyzed 4 times against dPBS + 0.001% Pluronic (1.5L) to remove free molecules that had not bound to the AAV

capsid. The vectors solution was aliquoted in 100, 50 and 10  $\mu\text{L}$  sample and store at  $-80^{\circ}\text{C}$ , a titration by qPCR was done after dialysis.

### C. Titration of AAV vector genomes

A total of 3  $\mu\text{L}$  of AAV was treated with 20 units of DNase I (Roche #04716728001) at  $37^{\circ}\text{C}$  for 45 min to remove residual DNA in vector samples. After treatment with DNase I, 20  $\mu\text{L}$  of proteinase K (20 mg/mL; MACHEREY-NAGEL # 740506) was added and the mixture incubated at  $70^{\circ}\text{C}$  for 20 min. An extraction column (NucleoSpin<sup>®</sup>RNA Virus) was then used to extract DNA from purified AAV vectors. Quantitative real time PCR (qPCR) was performed with a StepOnePlus<sup>™</sup> Real-Time PCR System Upgrade (Life Technologies). All PCRs were performed with a final volume of 20  $\mu\text{L}$ , including primers and probes targeting the ITR2 sequence, PCR Master Mix (TaKaRa), and 5 $\mu\text{L}$  of template DNA (plasmid standard or sample DNA). qPCR was carried out with an initial denaturation step at  $95^{\circ}\text{C}$  for 20 seconds, followed by 45 cycles of denaturation at  $95^{\circ}\text{C}$  for 1 second and annealing/extension at  $56^{\circ}\text{C}$  for 20 seconds. Plasmid standards were generated with seven serial dilutions (containing 10<sup>8</sup> to 10<sup>2</sup> plasmid copies), as described by D'Costa *et al.*<sup>203</sup>

### D. Western blot and silver staining

All vectors were denatured at  $100^{\circ}\text{C}$  for 5 min using Laemmli sample buffer and separated by SDS-PAGE on 10% Tris-glycine polyacrylamide gels (Life Technologies). Precision Plus Protein All Blue Standards (BioRad) were used as a molecular-weight size marker. After electrophoresis, gels were either silver stained (PlusOne Silver Staining Kit, Protein; GE Healthcare) or transferred onto nitrocellulose membranes for Western blot. After transferring the proteins to nitrocellulose membrane using a transfer buffer (25 mM Tris/192 mM glycine/0.1 (w/v) SDS/20% MeOH) for 1h at 150 mA in a Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad), the membrane was saturated for 2h at RT with 5% semi-skimmed milk in PBS-Tween (0.1%) or with 1% gelatin, 0.1% Igepal in PBS-Tween (0.01%). After saturation, the membrane was probed with the corresponding antibody (anti-capsid polyclonal, B1 monoclonal, or anti fluorescein-AP) and FITC-Soybean lectin, FITC-WL lectin or ASGPr lectin (Carbohydrate detection) overnight at  $4^{\circ}\text{C}$ . Three washes (15 min at RT) with PBS-Tween (0.1%) were performed between each stage to remove unbound reagents. Bands were visualized by chemiluminescence using alkaline phosphatase (AP) or horseradish peroxidase (HRP)-conjugated secondary antibodies and captured on X-ray film.

## E. Dot blot

AAV vectors were loaded at a dose of 1010 vg on a nitrocellulose paper soaked briefly in PBS prior to assembling the dot blot manifold (Bio-Rad). Nitrocellulose membrane containing the vectors was treated as for Western blotting.

## F. Transduction of HEK cells HEK293, HUH7 and HEPG2

Cells were seeded in a 24-well plate at a density of approximately  $1.5 \times 10^5$  cells/well. After 24h, the cells were transduced at MOI of  $10^3$  /  $10^4$  /  $10^5$  with AAV2 or chemically modified AAV2 vectors in 0.5 mL of culture medium. 6h after the transduction, 0.5 mL of fresh culture medium was added to each well. All AAV vectors encoded for GFP. The percentage of GFP positive cells was measured by FACS analysis 48h after the transduction. Cells were dissociated with Trypsin-EDTA (Sigma-Aldrich), fixed with 4% paraformaldehyde and analysed on a BD-LSRII Flow Cytometer (BD Bioscience). All data were processed by FlowJo (V10; Flowjo LLC, Ashland, OR).

## G. Transduction of primary mouse hepatocytes

Primary mouse hepatocytes and culture medium were purchased from BIOPREDIC international (Rennes, FRANCE). Mouse hepatocytes were seeded in a 24-well plate at a density of approximately  $2.5 \times 10^5$  cells/well. After reception, cell culture medium was removed and replaced with 1mL of basal medium (MIL600) with additives (ADD222) and incubated 2h at 37°C and 5% CO<sub>2</sub>. Primary mouse hepatocytes were transduced at MOI of  $10^5$  with **AAV2**, **AAV2-PMS-01 (f)** and **(F)**, **AAV2-PMS-02 (f)** and **(F)**, **AAV2-PMS-03 (f)** and **(F)**, **AAV2-PMS-04 (f)** and **(F)** vectors in 0.5 mL of culture medium. 6h after the transduction, 0.5 mL of fresh culture medium was added to each well. All AAV vectors encoded for GFP. The percentage of GFP positive cells was measured by FACS analysis 72h after the transduction. Cells were dissociated with Trypsin-EDTA (Sigma-Aldrich), fixed with 4% paraformaldehyde and analysed on a BD-LSRII Flow Cytometer (BD Bioscience). All data were processed by FlowJo (V10; Flowjo LLC, Ashland, OR).

## H. Glycodiag© experiment

The evaluation of AAV2 particle interactions was performed according to the GLYcoPROFILE methodology (GLYcoDiag, Orléans, France), which remains the company's trade secret, on a panel of lectins of plant and human recombinant origin ([Annexe 2](#)). Briefly, particles are deposited at three concentrations (i.e.  $5.0 \times 10^{10}$  particles/mL,  $2.5 \times 10^{10}$  particles/mL,  $1.25 \times 10^{10}$  particles/mL) in the lectin wells and incubated for 2 h at RT. After washing, extravidin-peroxidase conjugate is added and incubated for 30 min. After a final wash of the plates, 100  $\mu$ L of OPD reagent is added to each lectin well. After 15 min of incubation, 100  $\mu$ L of 1 M HCl solution is added before reading the plates ( $\lambda$ Abs = 450 nm) using a multiplate reader.

### A. ELISA Assay

The evaluation of AAV2 particle interactions with (A20) antibody was performed according to the Progen AAV2 titration ELISA (cat. no. PRATV; Progen Biotechnik).<sup>204</sup> AAV2 is used as a standard and the value obtained is normalized to 100%.

### B. Animal care and welfare

Experiments were performed on C57BL/6 (B6) mice male. Animals were euthanized 1 month after AAV injection. Research was conducted at the UTE IRS2 (University of Nantes, France. The Institutional Animal Care and Use Committee of the Région des Pays de la Loire as well as the French Ministry for National Education, Higher Education and Research approved the protocols (authorizations #29288). Animals were randomly assigned to the different experimental groups. They were sacrificed by beheading in accordance with approved protocols. We used only male mice to facilitate quantification of the vector genome using the endogenous hprt1 gene on the X chromosome (Procedure detailed below).

### C. Intravenous injection

Mice were injected intravenously (i.v.) with a dose of  $5 \times 10^{12}$  vg/kg of either **AAV2-PMS-01 (f) and (F)**, **AAV2-PMS-02 (f) and (F)**, **AAV2-PMS-03 (f) and (F)**, **AAV2-PMS-04 (f) and (F)** eGFP-expressing or unmodified AAV2 eGFP-expressing. Eight weeks old C57BL/6 (B6) mice (Charles River Laboratories)



were used in this study. 10 groups were formed, 6 mice per group. Administration of chemically modified AAV2 vectors (or formulation buffer) by IV injection through the caudal tail vein for groups 1, 2, 3, 4 and 7 and the penile vein for groups 5, 6, 8, 9, 10 under general anesthesia. The blood samples will be taken on the day of the injection, after every two weeks up to one month post injection. All animals will be euthanized one-month after injection, samples were collected blood, liver, spleen, quadriceps, triceps, heart, kidney, and lung.

#### D. Absolute quantification of vector genomes by qPCR in mice tissue samples

Liver samples were obtained just after sacrifice in conditions that minimized cross contamination and avoided qPCR inhibition, as described by Le Guiner *et al.*<sup>205</sup> Samples were snap-frozen in liquid nitrogen and stored at  $\leq -70^{\circ}\text{C}$  before DNA extraction. gDNA was extracted using Genra Puregene kit (Qiagen) and TissueLyser II (Qiagen), according to the manufacturer's instructions. qPCR analyses were conducted on a C1000 touch thermal cycler (Bio-Rad) using 50 ng of gDNA in duplicate and premix Ex taq (Ozyme). Vector genome copy number was determined using the following primer/probe combination, designed to amplify a specific region of the GFP sequence present in the transgene (Forward, 5'- 27 AGTCCGCCCTGAGCAAAGA-3'; Reverse, 5'- GCGGTCACGAACTCCAGC -3'; Probe, 5'- FAM- CAACGAGAAGCGCGATCACATGGTC-TAMRA-3'). Endogenous gDNA copy number was determined using a primer/probe combination designed to amplify the murine Hprt1 gene (Forward, 5'- TCTGTAAGAAGGATTTAAAGAGAAGCTA-3'; Reverse, 5'- ATCACATGTTTATTCCACTGAGCAA-3'; Probe, 5'- FAMAGCTCTCGATTCCTATCAGTAACAGC -TAMRA-3'). For each sample, cycle threshold (Ct) values were compared with those obtained with different dilutions of linearized standard plasmids (containing either the eGFP expression cassette or the murine Hprt1 gene). The absence of qPCR inhibition in the presence of gDNA was determined by analyzing 50 ng genomic DNA (gDNA) extracted from tissue samples from a control animal and spiked with different dilutions of standard plasmid. Results are expressed as vector genome copy number per diploid genome (vg/dg). The lower limit of quantification (LLOQ) of our test was 0.006 vg/dg. GraphPad Prism 9 software was used for statistical analysis. Non-parametric Mann Whitney test was used. Samples were considered significantly different if  $*p < 0.05$ .

#### E. Relative quantification of GFP expression by RT-qPCR in mice tissue samples

Total RNA was extracted from snap-frozen liver, heart, skeletal muscle, lung, spleen and kidney, using Qiazol reagent (Qiagen) and treated with ezDNAse kit (Thermo Fisher Scientific), according to the manufacturer's instructions. Then, 500 ng of this RNA was reverse transcribed using Superscript IV Vilo kit (Thermo Fisher Scientific), according to the manufacturer's instructions. qPCR analysis was then performed on cDNA (diluted 1/20) using the same GFP primers and probe as for the quantification of transgene copy number by qPCR. Hprt1 messenger was also amplified as an endogenous control. For each RNA sample, the absence of DNA contamination was confirmed by analysis of "cDNA-like samples" obtained without adding reverse transcriptase to the reaction mix. The absence of qPCR inhibition in the presence of cDNA was determined by analyzing cDNA obtained from tissue sample from a control animal, spiked with different dilutions of a standard plasmid. Results were expressed in relative quantification (RQ):  $RQ = 2^{-\Delta Ct} = 2^{-(Ct\ GFP - Ct\ HPRT1)}$ . The lower limit of quantification (LLOQ) of our test was 0.0004. GraphPad Prism 9 software was used for statistical analysis. Non-parametric Mann Whitney test was used. Samples were considered significantly different if \* $p < 0.05$ , \*\* $p < 0.01$ .

## F. Western blot of GFP expression

Total proteins were extracted from snap-frozen liver using RIPA buffer (Tris 10 mM pH 7.5; NaCl 150 mM; EDTA 1 mM; NP40 1%; sodium deoxycholate 0.5%; SDS 0.1%) containing protease inhibitor cocktail (Sigma-Aldrich). 25  $\mu$ g of total protein extract were prepared in Laemmli buffer + 200mM final of DTT, reduced 10min at 70°C and then loaded into a Novex 10% Tris-glycine gel (Thermo Fisher Scientific). Proteins were transferred on silica membrane using the trans-blot turbo kit (Biorad), according to the manufacturer's instructions. Membranes were blocked in 5% skim milk, 1% NP40 (Sigma-Aldrich) in TBST (Tris-buffered saline, 0.1% Tween-20) and hybridized with an anti-GFP antibody (1:8,000, JL-8, Takarabio) and with a secondary anti-mouse IgG HRP-conjugated antibody (1:5,000, Dako P0447). For protein loading control, the same membrane was also hybridized with an anti-mouse GAPDH antibody (1:2,000, Novus IMG3073) and with a secondary anti-goat IgG HRP-conjugated antibody (1:2,000, Dako P0449). Immunoblots were visualized by ECL Chemiluminescent analysis system. Semi-quantification of GFP protein expression relative to mouse GAPDH was obtained by densitometry analysis, using the ImageQuant software. GraphPad Prism 9 software was used for statistical analysis. The nonparametric Mann Whitney test was used.

## G. Immunohistochemistry GFP staining in mice

Formalin-fixed paraffin embedded liver samples were cut into 4  $\mu\text{m}$  thick slices. Sections were first deparaffinized in methylcyclohexan and then rehydrated using decreasing concentrations of ethanol. Antigen unmasking was further performed through incubation in citrate buffer pH=6 (40 min, 98°C). After a permeabilization step (10 min, RT) with Triton (0.2%; VWR International, Leuven, Belgium), and a saturation step (45 min, RT) with goat serum (2%; Abcam Inc., Cambridge, MA, USA) and bovine serum albumin (5%, Sigma, St. Louis, USA), sections were then incubated (over-night, 4°C) with a mouse monoclonal antibody 29 against GFP (1:100; JL8 clone, Thermo Fisher Scientific, Geel, Belgium). After PBS-washing, sections were incubated with an Alexa fluor 555-conjugated goat anti-murine antibody (1:300, 1 h, RT, InVitroGen, Carlsbad, CA, USA) before washing, nuclei counterstaining with DAPI (1:1000, InVitroGen) and mounting. Whole histological preparations were digitized using fluorescence detection (Zeiss Axioscan Z1, x10 magnification). Numbers of total nuclei and positive cells based on red fluorescence were counted using QuPath software (Bankhead, P. et al. Scientific Reports, 2017) on 3 randomly selected field (2 mm<sup>2</sup> each) enabling to observe a mean of 17882 nuclei per sample. GraphPad Prism 9 software was used for statistical analysis. The nonparametric Mann Whitney test was used. Samples were considered significantly different if \*p < 0.05.

# Bibliographie :

- (1) Nurk, S.; Koren, S.; Rhie, A.; Rautiainen, M.; Bzikadze, A. V.; Mikheenko, A.; Vollger, M. R.; Altemose, N.; Uralsky, L.; Gershman, A.; Aganezov, S.; Hoyt, S. J.; Diekhans, M.; Logsdon, G. A.; Alonge, M.; Antonarakis, S. E.; Borchers, M.; Bouffard, G. G.; Brooks, S. Y.; Caldas, G. V.; Chen, N.-C.; Cheng, H.; Chin, C.-S.; Chow, W.; de Lima, L. G.; Dishuck, P. C.; Durbin, R.; Dvorkina, T.; Fiddes, I. T.; Formenti, G.; Fulton, R. S.; Functammasan, A.; Garrison, E.; Grady, P. G. S.; Graves-Lindsay, T. A.; Hall, I. M.; Hansen, N. F.; Hartley, G. A.; Haukness, M.; Howe, K.; Hunkapiller, M. W.; Jain, C.; Jain, M.; Jarvis, E. D.; Kerpedjiev, P.; Kirsche, M.; Kolmogorov, M.; Korlach, J.; Kremitzki, M.; Li, H.; Maduro, V. V.; Marschall, T.; McCartney, A. M.; McDaniel, J.; Miller, D. E.; Mullikin, J. C.; Myers, E. W.; Olson, N. D.; Paten, B.; Peluso, P.; Pevzner, P. A.; Porubsky, D.; Potapova, T.; Rogaev, E. I.; Rosenfeld, J. A.; Salzberg, S. L.; Schneider, V. A.; Sedlazeck, F. J.; Shafin, K.; Shew, C. J.; Shumate, A.; Sims, Y.; Smit, A. F. A.; Soto, D. C.; Sović, I.; Storer, J. M.; Streets, A.; Sullivan, B. A.; Thibaud-Nissen, F.; Torrance, J.; Wagner, J.; Walenz, B. P.; Wenger, A.; Wood, J. M. D.; Xiao, C.; Yan, S. M.; Young, A. C.; Zarate, S.; Surti, U.; McCoy, R. C.; Dennis, M. Y.; Alexandrov, I. A.; Gerton, J. L.; O'Neill, R. J.; Timp, W.; Zook, J. M.; Schatz, M. C.; Eichler, E. E.; Miga, K. H.; Phillippy, A. M. The Complete Sequence of a Human Genome. *Science* **2022**, *376* (6588), 44–53. <https://doi.org/10.1126/science.abj6987>.
- (2) *Qu'est-ce qu'une maladie génétique ?* CHU de Liège. [https://www.chuliege.be/jcms/c\\_16593918/fr/centre-des-maladies-rares/qu-est-ce-qu-une-maladie-genetique](https://www.chuliege.be/jcms/c_16593918/fr/centre-des-maladies-rares/qu-est-ce-qu-une-maladie-genetique) (accessed 2022-11-16).
- (3) *Dépistage néonatal de la drépanocytose en France*. Haute Autorité de Santé. [https://www.has-sante.fr/jcms/c\\_1728538/fr/depistage-neonatal-de-la-drepanocytose-en-france](https://www.has-sante.fr/jcms/c_1728538/fr/depistage-neonatal-de-la-drepanocytose-en-france) (accessed 2022-11-16).
- (4) Ismail, V.; Zachariassen, L. G.; Godwin, A.; Sahakian, M.; Ellard, S.; Stals, K. L.; Baple, E.; Brown, K. T.; Foulds, N.; Whewey, G.; Parker, M. O.; Lyngby, S. M.; Pedersen, M. G.; Desir, J.; Bayat, A.; Musgaard, M.; Guille, M.; Kristensen, A. S.; Baralle, D. Identification and Functional Evaluation of GRIA1 Missense and Truncation Variants in Individuals with ID: An Emerging Neurodevelopmental Syndrome. *Am. J. Hum. Genet.* **2022**, *109* (7), 1217–1241. <https://doi.org/10.1016/j.ajhg.2022.05.009>.
- (5) Ghosh, A.; Bansal, M. A Glossary of DNA Structures from A to Z. *Acta Crystallogr. Sect. D* **2003**, *59* (4), 620–626. <https://doi.org/10.1107/S0907444903003251>.
- (6) Zinder, N. D.; Lederberg, J. GENETIC EXCHANGE IN SALMONELLA. *J. Bacteriol.* **1952**, *64* (5), 679–699. <https://doi.org/10.1128/jb.64.5.679-699.1952>.
- (7) Szybalska, E. H.; Szybalski, W. GENETICS OF HUMAN CELL LINES, IV. DNA-MEDIATED HERITABLE TRANSFORMATION OF A BIOCHEMICAL TRAIT. *Proc. Natl. Acad. Sci.* **1962**, *48* (12), 2026–2034. <https://doi.org/10.1073/pnas.48.12.2026>.
- (8) Strauss, B. S. Biochemical Genetics and Molecular Biology: The Contributions of George Beadle and Edward Tatum. *Genetics* **2016**, *203* (1), 13–20. <https://doi.org/10.1534/genetics.116.188995>.
- (9) Rosenberg, S. A.; Anderson, W. F.; Blaese, M.; Hwu, P.; Yannelli, J. R.; Yang, J. C.; Topalian, S. L.; Schwartzentruber, D. J.; Weber, J. S.; Ettinghausen, S. E. The Development of Gene Therapy for the Treatment of Cancer. *Ann. Surg.* **1993**, *218* (4), 455–464.
- (10) *Gene Transfer into Humans — Immunotherapy of Patients with Advanced Melanoma, Using Tumor-Infiltrating Lymphocytes Modified by Retroviral Gene Transduction | NEJM*. <https://www.nejm.org/doi/full/10.1056/NEJM199008303230904> (accessed 2022-11-17).
- (11) Ginn, S. L.; Amaya, A. K.; Alexander, I. E.; Edelstein, M.; Abedi, M. R. Gene Therapy Clinical Trials Worldwide to 2017: An Update. *J. Gene Med.* **2018**, *20* (5), e3015. <https://doi.org/10.1002/jgm.3015>.

- (12) Bensimon, C. *Coup d'arrêt à la thérapie génique aux Etats-Unis. Le décès d'un jeune homme révèle les risques encourus par les patients*. Libération.  
[https://www.liberation.fr/societe/2000/02/10/coup-d-arret-a-la-therapie-genique-aux-etats-unis-le-deces-d-un-jeune-homme-revele-les-risques-encou\\_316045/](https://www.liberation.fr/societe/2000/02/10/coup-d-arret-a-la-therapie-genique-aux-etats-unis-le-deces-d-un-jeune-homme-revele-les-risques-encou_316045/) (accessed 2022-11-17).
- (13) Hacein-Bey-Abina, S.; Le Deist, F.; Carlier, F.; Bouneaud, C.; Hue, C.; De Villartay, J.-P.; Thrasher, A. J.; Wulffraat, N.; Sorensen, R.; Dupuis-Girod, S.; Fischer, A.; Davies, E. G.; Kuis, W.; Leiva, L.; Cavazzana-Calvo, M. Sustained Correction of X-Linked Severe Combined Immunodeficiency by Ex Vivo Gene Therapy. *N. Engl. J. Med.* **2002**, *346* (16), 1185–1193.  
<https://doi.org/10.1056/NEJMoa012616>.
- (14) *GTCT (FMS19)*. <https://a873679.fmpost.com/fmi/webd/GTCT> (accessed 2022-11-17).
- (15) Ramamoorth, M.; Narvekar, A. Non Viral Vectors in Gene Therapy- An Overview. *J. Clin. Diagn. Res. JCDR* **2015**, *9* (1), GE01–GE06. <https://doi.org/10.7860/JCDR/2015/10443.5394>.
- (16) Maestro, S.; Weber, N. D.; Zabaleta, N.; Aldabe, R.; Gonzalez-Aseguinolaza, G. Novel Vectors and Approaches for Gene Therapy in Liver Diseases. *JHEP Rep.* **2021**, *3* (4), 100300.  
<https://doi.org/10.1016/j.jhepr.2021.100300>.
- (17) Hood, J. D.; Bednarski, M.; Frausto, R.; Guccione, S.; Reisfeld, R. A.; Xiang, R.; Cheresch, D. A. Tumor Regression by Targeted Gene Delivery to the Neovasculature. *Science* **2002**, *296* (5577), 2404–2407. <https://doi.org/10.1126/science.1070200>.
- (18) *Gene Inhibition Therapy | Explore Gene Therapy*. <https://www.exploregenetherapy.com/about-gene-therapy/gene-inhibition> (accessed 2022-11-18).
- (19) Di Blasi, C. Mild Muscular Dystrophy Due to a Nonsense Mutation in the LAMA2 Gene Resulting in Exon Skipping. *Brain* **2001**, *124* (4), 698–704. <https://doi.org/10.1093/brain/124.4.698>.
- (20) Ghosh, D.; Venkataramani, P.; Nandi, S.; Bhattacharjee, S. CRISPR–Cas9 a Boon or Bane: The Bumpy Road Ahead to Cancer Therapeutics. *Cancer Cell Int.* **2019**, *19* (1), 1–10.  
<https://doi.org/10.1186/s12935-019-0726-0>.
- (21) Liu, C.; Zhang, L.; Liu, H.; Cheng, K. Delivery Strategies of the CRISPR–Cas9 Gene-Editing System for Therapeutic Applications. *J. Controlled Release* **2017**, *266*, 17–26.  
<https://doi.org/10.1016/j.jconrel.2017.09.012>.
- (22) *Thérapie génique : les traitements du futur ?* <https://www.leem.org/100-questions/therapie-genique-les-traitements-du-futur> (accessed 2023-04-02).
- (23) Arabi, F.; Mansouri, V.; Ahmadbeigi, N. Gene Therapy Clinical Trials, Where Do We Go? An Overview. *Biomed. Pharmacother.* **2022**, *153*, 113324.  
<https://doi.org/10.1016/j.biopha.2022.113324>.
- (24) Kim, T. K.; Eberwine, J. H. Mammalian Cell Transfection: The Present and the Future. *Anal. Bioanal. Chem.* **2010**, *397* (8), 3173–3178. <https://doi.org/10.1007/s00216-010-3821-6>.
- (25) Heng, H. H. Q.; Goetze, S.; Ye, C. J.; Liu, G.; Stevens, J. B.; Bremer, S. W.; Wykes, S. M.; Bode, J.; Krawetz, S. A. Chromatin Loops Are Selectively Anchored Using Scaffold/Matrix-Attachment Regions. *J. Cell Sci.* **2004**, *117* (7), 999–1008. <https://doi.org/10.1242/jcs.00976>.
- (26) Giannakopoulos, A.; Quiviger, M.; Stavrou, E.; Verras, M.; Marie, C.; Scherman, D.; Athanassiadou, A. Efficient Episomal Gene Transfer to Human Hepatic Cells Using the pFAR4-S/MAR Vector. *Mol. Biol. Rep.* **2019**, *46* (3), 3203–3211. <https://doi.org/10.1007/s11033-019-04777-9>.
- (27) Jones, C. H.; Chen, C.-K.; Ravikrishnan, A.; Rane, S.; Pfeifer, B. A. Overcoming Nonviral Gene Delivery Barriers: Perspective and Future. *Mol. Pharm.* **2013**, *10* (11), 4082–4098.  
<https://doi.org/10.1021/mp400467x>.
- (28) Rueda-Fernández, M.; Melguizo-Rodríguez, L.; Costela-Ruiz, V. J.; González-Acedo, A.; Ramos-Torrecillas, J.; Illescas-Montes, R. The Current Status of COVID-19 Vaccines. A Scoping Review. *Drug Discov. Today* **2022**, *27* (11), 103336. <https://doi.org/10.1016/j.drudis.2022.08.004>.
- (29) *Vaccin - Spikevax (Moderna) - ANSM*. <https://ansm.sante.fr/tableau-vaccin/covid-19-vaccine-moderna> (accessed 2023-04-04).
- (30) *Vaccin - Comirnaty adulte à diluer (Pfizer & BioNTech) - ANSM*. <https://ansm.sante.fr/tableau-vaccin/comirnaty-pfizer-biontech> (accessed 2023-04-04).

- (31) Virus. *Wikipédia*; 2022.
- (32) PII: 0092-8674(76)90133-1 | Elsevier Enhanced Reader. [https://doi.org/10.1016/0092-8674\(76\)90133-1](https://doi.org/10.1016/0092-8674(76)90133-1).
- (33) *Teen Dies Undergoing Experimental Gene Therapy*. <https://www.washingtonpost.com/wp-srv/WPcap/1999-09/29/060r-092999-idx.html> (accessed 2023-05-09).
- (34) Lundstrom, K. Viral Vectors in Gene Therapy. *Diseases* **2018**, 6 (2), 42. <https://doi.org/10.3390/diseases6020042>.
- (35) Buchschacher, G. L., Jr; Wong-Staal, F. Development of Lentiviral Vectors for Gene Therapy for Human Diseases. *Blood* **2000**, 95 (8), 2499–2504. <https://doi.org/10.1182/blood.V95.8.2499>.
- (36) Valentine, R. C.; Pereira, H. G. Antigens and Structure of the Adenovirus. *J. Mol. Biol.* **1965**, 13 (1), 13-IN13. [https://doi.org/10.1016/S0022-2836\(65\)80076-6](https://doi.org/10.1016/S0022-2836(65)80076-6).
- (37) Wold, W. S. M.; Toth, K. Adenovirus Vectors for Gene Therapy, Vaccination and Cancer Gene Therapy. *Curr. Gene Ther.* **2013**, 13 (6), 421–433.
- (38) Roy-Chowdhury, J.; Horwitz, M. S. Evolution of Adenoviruses as Gene Therapy Vectors. *Mol. Ther.* **2002**, 5 (4), 340–344. <https://doi.org/10.1006/mthe.2001.0575>.
- (39) *Provider of adeno associated virus (AAV) vectors analyses*. <https://www.vironova.com/em-services/analyses/adeno-associated-virus-vectors> (accessed 2023-04-04).
- (40) Zarate-Perez, F.; Mansilla-Soto, J.; Bardelli, M.; Burgner, J. W.; Villamil-Jarauta, M.; Kekilli, D.; Samsó, M.; Linden, R. M.; Escalante, C. R. Oligomeric Properties of Adeno-Associated Virus Rep68 Reflect Its Multifunctionality. *J. Virol.* **2013**, 87 (2), 1232–1241. <https://doi.org/10.1128/JVI.02441-12>.
- (41) Kotin, R. M.; Siniscalco, M.; Samulski, R. J.; Zhu, X. D.; Hunter, L.; Laughlin, C. A.; McLaughlin, S.; Muzyczka, N.; Rocchi, M.; Berns, K. I. Site-Specific Integration by Adeno-Associated Virus. *Proc. Natl. Acad. Sci.* **1990**, 87 (6), 2211–2215. <https://doi.org/10.1073/pnas.87.6.2211>.
- (42) Büning, H.; Srivastava, A. Capsid Modifications for Targeting and Improving the Efficacy of AAV Vectors. *Mol. Ther. - Methods Clin. Dev.* **2019**, 12, 248–265. <https://doi.org/10.1016/j.omtm.2019.01.008>.
- (43) Cao, M.; You, H.; Hermonat, P. L. The X Gene of Adeno-Associated Virus 2 (AAV2) Is Involved in Viral DNA Replication. *PLoS ONE* **2014**, 9 (8), e104596. <https://doi.org/10.1371/journal.pone.0104596>.
- (44) Galibert, L.; Hyvönen, A.; Eriksson, R. A. E.; Mattola, S.; Aho, V.; Salminen, S.; Albers, J. D.; Peltola, S. K.; Weman, S.; Nieminen, T.; Ylä-Herttuala, S.; Lesch, H. P.; Vihinen-Ranta, M.; Airene, K. J. Functional Roles of the Membrane-Associated AAV Protein MAAP. *Sci. Rep.* **2021**, 11 (1), 21698. <https://doi.org/10.1038/s41598-021-01220-7>.
- (45) Earley, L. F.; Powers, J. M.; Adachi, K.; Baumgart, J. T.; Meyer, N. L.; Xie, Q.; Chapman, M. S.; Nakai, H. Adeno-Associated Virus (AAV) Assembly-Activating Protein Is Not an Essential Requirement for Capsid Assembly of AAV Serotypes 4, 5, and 11. *J. Virol.* **2017**, 91 (3), e01980-16. <https://doi.org/10.1128/JVI.01980-16>.
- (46) Kotin, R. M.; Siniscalco, M.; Samulski, R. J.; Zhu, X. D.; Hunter, L.; Laughlin, C. A.; McLaughlin, S.; Muzyczka, N.; Rocchi, M.; Berns, K. I. Site-Specific Integration by Adeno-Associated Virus. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, 87 (6), 2211–2215. <https://doi.org/10.1073/pnas.87.6.2211>.
- (47) Balakrishnan, B.; Jayandharan, G. R. Basic Biology of Adeno-Associated Virus (AAV) Vectors Used in Gene Therapy. *Curr. Gene Ther.* **2014**, 14 (2), 86–100. <https://doi.org/10.2174/1566523214666140302193709>.
- (48) Stagg, S. M.; Yoshioka, C.; Davulcu, O.; Chapman, M. S. Cryo-Electron Microscopy of Adeno-Associated Virus. *Chem. Rev.* **2022**, acs.chemrev.1c00936. <https://doi.org/10.1021/acs.chemrev.1c00936>.
- (49) Ng, R.; Govindasamy, L.; Gurda, B. L.; McKenna, R.; Kozyreva, O. G.; Samulski, R. J.; Parent, K. N.; Baker, T. S.; Agbandje-McKenna, M. Structural Characterization of the Dual Glycan Binding Adeno-Associated Virus Serotype 6. *J. Virol.* **2010**, 84 (24), 12945–12957. <https://doi.org/10.1128/JVI.01235-10>.

- (50) Srivastava, A. In Vivo Tissue-Tropism of Adeno-Associated Viral Vectors. *Curr. Opin. Virol.* **2016**, *21*, 75–80. <https://doi.org/10.1016/j.coviro.2016.08.003>.
- (51) Au, H. K. E.; Isalan, M.; Mielcarek, M. Gene Therapy Advances: A Meta-Analysis of AAV Usage in Clinical Settings. *Front. Med.* **2021**, *8*, 809118. <https://doi.org/10.3389/fmed.2021.809118>.
- (52) Wang, D.; Tai, P. W. L.; Gao, G. Adeno-Associated Virus Vector as a Platform for Gene Therapy Delivery. *Nat. Rev. Drug Discov.* **2019**, *18* (5), 358–378. <https://doi.org/10.1038/s41573-019-0012-9>.
- (53) Summerford, C.; Samulski, R. J. Membrane-Associated Heparan Sulfate Proteoglycan Is a Receptor for Adeno-Associated Virus Type 2 Virions. *J. Virol.* **1998**, *72* (2), 1438–1445.
- (54) Qing, K.; Mah, C.; Hansen, J.; Zhou, S.; Dwarki, V.; Srivastava, A. Human Fibroblast Growth Factor Receptor 1 Is a Co-Receptor for Infection by Adeno-Associated Virus 2. *Nat. Med.* **1999**, *5* (1), 71–77. <https://doi.org/10.1038/4758>.
- (55) Zengel, J.; Carette, J. E. Structural and Cellular Biology of Adeno-Associated Virus Attachment and Entry. In *Advances in Virus Research*; Elsevier, 2020; Vol. 106, pp 39–84. <https://doi.org/10.1016/bs.aivir.2020.01.002>.
- (56) Pillay, S.; Meyer, N. L.; Puschnik, A. S.; Davulcu, O.; Diep, J.; Ishikawa, Y.; Jae, L. T.; Wosen, J. E.; Nagamine, C. M.; Chapman, M. S.; Carette, J. E. An Essential Receptor for Adeno-Associated Virus Infection. *Nature* **2016**, *530* (7588), 108–112. <https://doi.org/10.1038/nature16465>.
- (57) Girod, A.; Wobus, C. E.; Zádori, Z.; Ried, M.; Leike, K.; Tijssen, P.; Kleinschmidt, J. A.; Hallek, M. The VP1 Capsid Protein of Adeno-Associated Virus Type 2 Is Carrying a Phospholipase A2 Domain Required for Virus Infectivity. *J. Gen. Virol.* **2002**, *83* (5), 973–978. <https://doi.org/10.1099/0022-1317-83-5-973>.
- (58) Finn, J. D.; Hui, D.; Downey, H. D.; Dunn, D.; Pien, G. C.; Mingozzi, F.; Zhou, S.; High, K. A. Proteasome Inhibitors Decrease AAV2 Capsid Derived Peptide Epitope Presentation on MHC Class I Following Transduction. *Mol. Ther.* **2010**, *18* (1), 135–142. <https://doi.org/10.1038/mt.2009.257>.
- (59) Riyad, J. M.; Weber, T. Intracellular Trafficking of Adeno-Associated Virus (AAV) Vectors: Challenges and Future Directions. *Gene Ther.* **2021**, *28* (12), 683–696. <https://doi.org/10.1038/s41434-021-00243-z>.
- (60) Ding, W.; Zhang, L.; Yan, Z.; Engelhardt, J. F. Intracellular Trafficking of Adeno-Associated Viral Vectors. *Gene Ther.* **2005**, *12* (11), 873–880. <https://doi.org/10.1038/sj.gt.3302527>.
- (61) Srivastava, A. Adeno-Associated Virus-Mediated Gene Transfer. *J. Cell. Biochem.* **2008**, *105* (1), 17–24. <https://doi.org/10.1002/jcb.21819>.
- (62) Samulski, R. J.; Chang, L. S.; Shenk, T. A Recombinant Plasmid from Which an Infectious Adeno-Associated Virus Genome Can Be Excised In Vitro and Its Use to Study Viral Replication. *J. Virol.* **1987**, *61* (10), 3096–3101. <https://doi.org/10.1128/JVI.61.10.3096-3101.1987>.
- (63) Xiao, X.; Li, J.; Samulski, R. J. Production of High-Titer Recombinant Adeno-Associated Virus Vectors in the Absence of Helper Adenovirus. *J. Virol.* **1998**, *72* (3), 2224–2232. <https://doi.org/10.1128/JVI.72.3.2224-2232.1998>.
- (64) Bourdon, A.; François, V.; Zhang, L.; Lafoux, A.; Fraysse, B.; Toumaniantz, G.; Larcher, T.; Girard, T.; Ledevin, M.; Lebreton, C.; Hivonnait, A.; Creisneas, A.; Allais, M.; Marie, B.; Guguin, J.; Blouin, V.; Remy, S.; Anegon, I.; Huchet, C.; Malerba, A.; Kao, B.; Heron, A. L.; Moullier, P.; Dickson, G.; Popplewell, L.; Adjali, O.; Montanaro, F.; Guiner, C. L. Evaluation of the Dystrophin Carboxy-Terminal Domain for Micro-Dystrophin Gene Therapy in Cardiac and Skeletal Muscles in the DMDmdx Rat Model. *Gene Ther.* **2022**, *29* (9), 520. <https://doi.org/10.1038/s41434-022-00317-6>.
- (65) Shelling, A. N.; Smith, M. G. Targeted Integration of Transfected and Infected Adeno-Associated Virus Vectors Containing the Neomycin Resistance Gene. *Gene Ther.* **1994**, *1* (3), 165–169.
- (66) Damdindorj, L.; Karnan, S.; Ota, A.; Hossain, E.; Konishi, Y.; Hosokawa, Y.; Konishi, H. A Comparative Analysis of Constitutive Promoters Located in Adeno-Associated Viral Vectors. *PLOS ONE* **2014**, *9* (8), e106472. <https://doi.org/10.1371/journal.pone.0106472>.

- (67) Matsushita, T.; Elliger, S.; Elliger, C.; Podsakoff, G.; Villarreal, L.; Kurtzman, G. J.; Iwaki, Y.; Colosi, P. Adeno-Associated Virus Vectors Can Be Efficiently Produced without Helper Virus. *Gene Ther.* **1998**, *5* (7), 938–945. <https://doi.org/10.1038/sj.gt.3300680>.
- (68) Galibert, L.; Jacob, A.; Savy, A.; Dickx, Y.; Bonnin, D.; Lecomte, C.; Rivollet, L.; Sanatine, P.; Boutin Fontaine, M.; Le Bec, C.; Merten, O.-W. Monobac System—A Single Baculovirus for the Production of rAAV. *Microorganisms* **2021**, *9* (9), 1799. <https://doi.org/10.3390/microorganisms9091799>.
- (69) Auricchio, A.; Kobinger, G.; Anand, V.; Hildinger, M.; O'Connor, E.; Maguire, A. M.; Wilson, J. M.; Bennett, J. Exchange of Surface Proteins Impacts on Viral Vector Cellular Specificity and Transduction Characteristics: The Retina as a Model. *Hum. Mol. Genet.* **2001**, *10* (26), 3075–3081. <https://doi.org/10.1093/hmg/10.26.3075>.
- (70) Zolotukhin, S.; Byrne, B. J.; Mason, E.; Zolotukhin, I.; Potter, M.; Chesnut, K.; Summerford, C.; Samulski, R. J.; Muzyczka, N. Recombinant Adeno-Associated Virus Purification Using Novel Methods Improves Infectious Titer and Yield. *Gene Ther.* **1999**, *6* (6), 973–985. <https://doi.org/10.1038/sj.gt.3300938>.
- (71) Zhao, Z.; Anselmo, A. C.; Mitragotri, S. Viral Vector-Based Gene Therapies in the Clinic. *Bioeng. Transl. Med.* **2022**, *7* (1), e10258. <https://doi.org/10.1002/btm2.10258>.
- (72) Mével, M.; Gouin, S.; Ayuso, E.; Deniaud, D. Chimie et Virus Adéno-Associés, Un Couple Prometteur Pour La Thérapie Génique. *Actual. Chim.* **2020**.
- (73) Zhu, M.-Y.; Juorio, A. V. Aromatic L-Amino Acid Decarboxylase: Biological Characterization and Functional Role. *Gen. Pharmacol. Vasc. Syst.* **1995**, *26* (4), 681–696. [https://doi.org/10.1016/0306-3623\(94\)00223-A](https://doi.org/10.1016/0306-3623(94)00223-A).
- (74) Keam, S. J. Eladocagene Exuparvovec: First Approval. *Drugs* **2022**, *82* (13), 1427–1432. <https://doi.org/10.1007/s40265-022-01775-3>.
- (75) Bhosale, K. Hemgenix as First Gene Therapy for Treatment of Haemophilia B. *Int. J. Adv. Res. Sci. Commun. Technol.* **2022**, 2581–9429. <https://doi.org/10.48175/IJARST-7657>.
- (76) YAOUANC, C. *Thérapie génique : quatrième décès dans l'essai d'Astellas*. Genethique. <https://www.genethique.org/therapie-genique-quatrieme-deces-dans-lessai-dastellas/> (accessed 2023-01-05).
- (77) Mendell, J. R.; Al-Zaidy, S.; Shell, R.; Arnold, W. D.; Rodino-Klapac, L. R.; Prior, T. W.; Lowes, L.; Alfano, L.; Berry, K.; Church, K.; Kissel, J. T.; Nagendran, S.; L'Italien, J.; Sproule, D. M.; Wells, C.; Cardenas, J. A.; Heitzer, M. D.; Kaspar, A.; Corcoran, S.; Braun, L.; Likhite, S.; Miranda, C.; Meyer, K.; Foust, K. D.; Burghes, A. H. M.; Kaspar, B. K. Single-Dose Gene-Replacement Therapy for Spinal Muscular Atrophy. *N. Engl. J. Med.* **2017**, *377* (18), 1713–1722. <https://doi.org/10.1056/NEJMoa1706198>.
- (78) Zaiss, A. K.; Cotter, M. J.; White, L. R.; Clark, S. A.; Wong, N. C. W.; Holers, V. M.; Bartlett, J. S.; Muruve, D. A. Complement Is an Essential Component of the Immune Response to Adeno-Associated Virus Vectors. *J. Virol.* **2008**, *82* (6), 2727–2740. <https://doi.org/10.1128/JVI.01990-07>.
- (79) Shirley, J. L.; de Jong, Y. P.; Terhorst, C.; Herzog, R. W. Immune Responses to Viral Gene Therapy Vectors. *Mol. Ther. J. Am. Soc. Gene Ther.* **2020**, *28* (3), 709–722. <https://doi.org/10.1016/j.ymthe.2020.01.001>.
- (80) Li, C.; Samulski, R. J. Engineering Adeno-Associated Virus Vectors for Gene Therapy. *Nat. Rev. Genet.* **2020**, *21* (4), 255–272. <https://doi.org/10.1038/s41576-019-0205-4>.
- (81) Mingozzi, F.; High, K. A. Therapeutic in Vivo Gene Transfer for Genetic Disease Using AAV: Progress and Challenges. *Nat. Rev. Genet.* **2011**, *12* (5), 341–355. <https://doi.org/10.1038/nrg2988>.
- (82) Mendell, J. R.; Campbell, K.; Rodino-Klapac, L.; Sahenk, Z.; Shilling, C.; Lewis, S.; Bowles, D.; Gray, S.; Li, C.; Galloway, G.; Malik, V.; Coley, B.; Clark, K. R.; Li, J.; Xiao, X.; Samulski, J.; McPhee, S. W.; Samulski, R. J.; Walker, C. M. Dystrophin Immunity in Duchenne's Muscular Dystrophy. *N. Engl. J. Med.* **2010**, *363* (15), 1429–1437. <https://doi.org/10.1056/NEJMoa1000228>.
- (83) Paulk, N. K.; Pekrun, K.; Zhu, E.; Nygaard, S.; Li, B.; Xu, J.; Chu, K.; Leborgne, C.; Dane, A. P.; Haft, A.; Zhang, Y.; Zhang, F.; Morton, C.; Valentine, M. B.; Davidoff, A. M.; Nathwani, A. C.; Mingozzi,



- F.; Grompe, M.; Alexander, I. E.; Lisowski, L.; Kay, M. A. Bioengineered AAV Capsids with Combined High Human Liver Transduction In Vivo and Unique Humoral Seroreactivity. *Mol. Ther.* **2018**, *26* (1), 289–303. <https://doi.org/10.1016/j.ymthe.2017.09.021>.
- (84) Zinn, E.; Pacouret, S.; Khaychuk, V.; Turunen, H. T.; Carvalho, L. S.; Andres-Mateos, E.; Shah, S.; Shelke, R.; Maurer, A. C.; Plovie, E.; Xiao, R.; Vandenberghe, L. H. In Silico Reconstruction of the Viral Evolutionary Lineage Yields a Potent Gene Therapy Vector. *Cell Rep.* **2015**, *12* (6), 1056–1068. <https://doi.org/10.1016/j.celrep.2015.07.019>.
- (85) Meumann, N.; Cabanes-Creus, M.; Ertelt, M.; Navarro, R. G.; Lucifora, J.; Yuan, Q.; Nien-Huber, K.; Abdelrahman, A.; Vu, X.-K.; Zhang, L.; Franke, A.-C.; Schmithals, C.; Piiper, A.; Vogt, A.; Gonzalez-Carmona, M.; Frueh, J. T.; Ullrich, E.; Meuleman, P.; Talbot, S. R.; Odenthal, M.; Ott, M.; Seifried, E.; Schoeder, C. T.; Schwäble, J.; Lisowski, L.; Büning, H. Novel Adeno-Associated Virus (AAV) Serotype 2 Capsid Variants for Improved Liver-Directed Gene Therapy. *Hepatology* n/a (n/a). <https://doi.org/10.1002/hep.32733>.
- (86) Zdechlik, A. C.; He, Y.; Aird, E. J.; Gordon, W. R.; Schmidt, D. Programmable Assembly of Adeno-Associated Virus–Antibody Composites for Receptor-Mediated Gene Delivery. *Bioconjug. Chem.* **2020**, *31* (4), 1093–1106. <https://doi.org/10.1021/acs.bioconjchem.9b00790>.
- (87) Stalmans, S.; Bracke, N.; Wynendaele, E.; Gevaert, B.; Peremans, K.; Burvenich, C.; Polis, I.; De Spiegeleer, B. Cell-Penetrating Peptides Selectively Cross the Blood-Brain Barrier In Vivo. *PLOS ONE* **2015**, *10* (10), e0139652. <https://doi.org/10.1371/journal.pone.0139652>.
- (88) Meng, Y.; Sun, D.; Qin, Y.; Dong, X.; Luo, G.; Liu, Y. Cell-Penetrating Peptides Enhance the Transduction of Adeno-Associated Virus Serotype 9 in the Central Nervous System. *Mol. Ther. - Methods Clin. Dev.* **2021**, *21*, 28–41. <https://doi.org/10.1016/j.omtm.2021.02.019>.
- (89) UniProt. <https://www.uniprot.org/statistics/Swiss-Prot> (accessed 2022-10-12).
- (90) Wang, Y.; Wu, C. Site-Specific Conjugation of Polymers to Proteins. *Biomacromolecules* **2018**, *19* (6), 1804–1825. <https://doi.org/10.1021/acs.biomac.8b00248>.
- (91) Ravasco, J. M. J. M.; Faustino, H.; Trindade, A.; Gois, P. M. P. Bioconjugation with Maleimides: A Useful Tool for Chemical Biology. *Chem. Weinh. Bergstr. Ger.* **2019**, *25* (1), 43–59. <https://doi.org/10.1002/chem.201803174>.
- (92) Szijj, P. A.; Bahou, C.; Chudasama, V. Minireview: Addressing the Retro-Michael Instability of Maleimide Bioconjugates. *Drug Discov. Today Technol.* **2018**, *30*, 27–34. <https://doi.org/10.1016/j.ddtec.2018.07.002>.
- (93) Bernardim, B.; Matos, M. J.; Ferhati, X.; Compañón, I.; Guerreiro, A.; Akkapeddi, P.; Burtoloso, A. C. B.; Jiménez-Osés, G.; Corzana, F.; Bernardes, G. J. L. Efficient and Irreversible Antibody–Cysteine Bioconjugation Using Carbonylacrylic Reagents. *Nat. Protoc.* **2019**, *14* (1), 86–99. <https://doi.org/10.1038/s41596-018-0083-9>.
- (94) Ochtrop, P.; Hackenberger, C. P. R. Recent Advances of Thiol-Selective Bioconjugation Reactions. *Curr. Opin. Chem. Biol.* **2020**, *58*, 28–36. <https://doi.org/10.1016/j.cbpa.2020.04.017>.
- (95) Leier, S.; Richter, S.; Bergmann, R.; Wuest, M.; Wuest, F. Radiometal-Containing Aryl Diazonium Salts for Chemoselective Bioconjugation of Tyrosine Residues. *ACS Omega* **2019**, *4* (26), 22101–22107. <https://doi.org/10.1021/acsomega.9b03248>.
- (96) Ban, H.; Nagano, M.; Gavriluk, J.; Hakamata, W.; Inokuma, T.; Barbas, C. F. I. Facile and Stable Linkages through Tyrosine: Bioconjugation Strategies with the Tyrosine-Click Reaction. *Bioconjug. Chem.* **2013**, *24* (4), 520–532. <https://doi.org/10.1021/bc300665t>.
- (97) Depienne, S.; Alvarez-Dorta, D.; Croyal, M.; T. Temgoua, R. C.; Charlier, C.; Deniaud, D.; Mével, M.; Boujtita, M.; G. Gouin, S. Luminol Anchors Improve the Electrochemical-Tyrosine-Click Labelling of Proteins. *Chem. Sci.* **2021**, *12* (46), 15374–15381. <https://doi.org/10.1039/D1SC04809K>.
- (98) Depienne, S.; Bouzelha, M.; Courtois, E.; Pavageau, K.; Lalys, P.-A.; Marchand, M.; Alvarez-Dorta, D.; Nedellec, S.; Marin-Fernandez, L.; Grandjean, C.; Boujtita, M.; Deniaud, D.; Mével, M.; Gouin, S. *Click-Electrochemistry for the Rapid Labeling of Virus, Bacteria and Cell Surfaces*; preprint; Chemistry, 2023. <https://doi.org/10.26434/chemrxiv-2023-q3sd8>.

- (99) deGruyter, J. N.; Malins, L. R.; Baran, P. S. Residue-Specific Peptide Modification: A Chemist's Guide. *Biochemistry* **2017**, *56* (30), 3863–3873. <https://doi.org/10.1021/acs.biochem.7b00536>.
- (100) Ho, H. T.; Bénard, A.; Forcher, G.; Bohec, M. L.; Montembault, V.; Pascual, S.; Fontaine, L. Azlactone-Based Heterobifunctional Linkers with Orthogonal Clickable Groups: Efficient Tools for Bioconjugation with Complete Atom Economy. *Org. Biomol. Chem.* **2018**, *16* (39), 7124–7128. <https://doi.org/10.1039/C8OB01807C>.
- (101) Ho, H. T.; Levere, M. E.; Fournier, D.; Montembault, V.; Pascual, S.; Fontaine, L.; Ho, H. T.; Levere, M. E.; Fournier, D.; Montembault, V.; Pascual, S.; Fontaine, L. Introducing the Azlactone Functionality into Polymers through Controlled Radical Polymerization: Strategies and Recent Developments. *Aust. J. Chem.* **2012**, *65* (8), 970–977. <https://doi.org/10.1071/CH12192>.
- (102) Kim, J. S.; Sirois, A. R.; Vazquez Cegla, A. J.; Jumai'an, E.; Murata, N.; Buck, M. E.; Moore, S. J. Protein–Polymer Conjugates Synthesized Using Water-Soluble Azlactone-Functionalized Polymers Enable Receptor-Specific Cellular Uptake toward Targeted Drug Delivery. *Bioconjug. Chem.* **2019**, *30* (4), 1220–1231. <https://doi.org/10.1021/acs.bioconjchem.9b00155>.
- (103) *Isothiocyanate - an overview | ScienceDirect Topics*. <https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/isothiocyanate> (accessed 2023-05-09).
- (104) Lee, G. K.; Maheshri, N.; Kaspar, B.; Schaffer, D. V. PEG Conjugation Moderately Protects Adeno-Associated Viral Vectors against Antibody Neutralization. *Biotechnol. Bioeng.* **2005**, *92* (1), 24–34. <https://doi.org/10.1002/bit.20562>.
- (105) Horowitz, E. D.; Weinberg, M. S.; Asokan, A. Glycated AAV Vectors: Chemical Redirection of Viral Tissue Tropism. *Bioconjug. Chem.* **2011**, *22* (4), 529–532. <https://doi.org/10.1021/bc100477g>.
- (106) Seo, J. W.; Ingham, E. S.; Mahakian, L.; Tumbale, S.; Wu, B.; Aghevlian, S.; Shams, S.; Baikoghli, M.; Jain, P.; Ding, X.; Goeden, N.; Dobрева, T.; Flytzanis, N. C.; Chavez, M.; Singhal, K.; Leib, R.; James, M. L.; Segal, D. J.; Cheng, R. H.; Silva, E. A.; Gradinaru, V.; Ferrara, K. W. Positron Emission Tomography Imaging of Novel AAV Capsids Maps Rapid Brain Accumulation. *Nat. Commun.* **2020**, *11* (1), 2102. <https://doi.org/10.1038/s41467-020-15818-4>.
- (107) Kelemen, R. E.; Mukherjee, R.; Cao, X.; Erickson, S. B.; Zheng, Y.; Chatterjee, A. A Precise Chemical Strategy To Alter the Receptor Specificity of the Adeno-Associated Virus. *Angew. Chem. Int. Ed.* **2016**, *55* (36), 10645–10649. <https://doi.org/10.1002/anie.201604067>.
- (108) Mével, M.; Bouzelha, M.; Leray, A.; Pacouret, S.; Guilbaud, M.; Penaud-Budloo, M.; Alvarez-Dorta, D.; Dubreil, L.; Gouin, S. G.; Combal, J. P.; Hommel, M.; Gonzalez-Aseguinolaza, G.; Blouin, V.; Moullier, P.; Adjali, O.; Deniaud, D.; Ayuso, E. Chemical Modification of the Adeno-Associated Virus Capsid to Improve Gene Delivery. *Chem. Sci.* **2020**, *11* (4), 1122–1131. <https://doi.org/10.1039/C9SC04189C>.
- (109) Shepherd, V. L.; Tarnowski, B. I.; McLaughlin, B. J. Isolation and Characterization of a Mannose Receptor from Human Pigment Epithelium. *Invest. Ophthalmol. Vis. Sci.* **1991**, *32* (6), 1779–1784.
- (110) Karlsen, T. H.; Lammert, F.; Thompson, R. J. Genetics of Liver Disease: From Pathophysiology to Clinical Practice. *J. Hepatol.* **2015**, *62* (1, Supplement), S6–S14. <https://doi.org/10.1016/j.jhep.2015.02.025>.
- (111) *Physiologie. Le foie, cet organe à tout faire*. Courrier international. <https://www.courrierinternational.com/article/physiologie-le-foie-cet-organe-tout-faire> (accessed 2023-05-09).
- (112) Ding, C.; Li, Y.; Guo, F.; Jiang, Y.; Ying, W.; Li, D.; Yang, D.; Xia, X.; Liu, W.; Zhao, Y.; He, Y.; Li, X.; Sun, W.; Liu, Q.; Song, L.; Zhen, B.; Zhang, P.; Qian, X.; Qin, J.; He, F. A Cell-Type-Resolved Liver Proteome. *Mol. Cell. Proteomics MCP* **2016**, *15* (10), 3190–3202. <https://doi.org/10.1074/mcp.M116.060145>.
- (113) *Le foie et les maladies rares*. Filfoie: tout savoir sur les maladies rares du foie, recherche, enseignement. <https://www.filfoie.com/maladies-rares-foie/foie-maladies-rares/> (accessed 2023-01-09).

- (114) <https://www.facebook.com/inserm.fr>. *Hémophilie · Inserm, La science pour la santé*. Inserm. <https://www.inserm.fr/dossier/hemophilie/> (accessed 2023-05-09).
- (115) Oldenburg, J.; Mahlangu, J. N.; Kim, B.; Schmitt, C.; Callaghan, M. U.; Young, G.; Santagostino, E.; Kruse-Jarres, R.; Negrier, C.; Kessler, C.; Valente, N.; Asikanius, E.; Levy, G. G.; Windyga, J.; Shima, M. Eficizumab Prophylaxis in Hemophilia A with Inhibitors. *N. Engl. J. Med.* **2017**, *377* (9), 809–818. <https://doi.org/10.1056/NEJMoa1703068>.
- (116) Long, B. R.; Veron, P.; Kuranda, K.; Hardet, R.; Mitchell, N.; Hayes, G. M.; Wong, W. Y.; Lau, K.; Li, M.; Hock, M. B.; Zoog, S. J.; Vettermann, C.; Mingozi, F.; Schweighardt, B. Early Phase Clinical Immunogenicity of Valoctocogene Roxaparvovec, an AAV5-Mediated Gene Therapy for Hemophilia A. *Mol. Ther.* **2021**, *29* (2), 597–610. <https://doi.org/10.1016/j.ymthe.2020.12.008>.
- (117) George, L. A.; Sullivan, S. K.; Giermasz, A.; Rasko, J. E. J.; Samelson-Jones, B. J.; Ducore, J.; Cuker, A.; Sullivan, L. M.; Majumdar, S.; Teitel, J.; McGuinn, C. E.; Ragni, M. V.; Luk, A. Y.; Hui, D.; Wright, J. F.; Chen, Y.; Liu, Y.; Wachtel, K.; Winters, A.; Tiefenbacher, S.; Arruda, V. R.; van der Loo, J. C. M.; Zeleniaia, O.; Takefman, D.; Carr, M. E.; Couto, L. B.; Anguela, X. M.; High, K. A. Hemophilia B Gene Therapy with a High-Specific-Activity Factor IX Variant. *N. Engl. J. Med.* **2017**, *377* (23), 2215–2227. <https://doi.org/10.1056/NEJMoa1708538>.
- (118) Mahlangu, J.; Kaczmarek, R.; von Drygalski, A.; Shapiro, S.; Chou, S.-C.; Ozelo, M. C.; Kenet, G.; Peyvandi, F.; Wang, M.; Madan, B.; Key, N. S.; Laffan, M.; Dunn, A. L.; Mason, J.; Quon, D. V.; Symington, E.; Leavitt, A. D.; Oldenburg, J.; Chambost, H.; Reding, M. T.; Jayaram, K.; Yu, H.; Mahajan, R.; Chavele, K.-M.; Reddy, D. B.; Henshaw, J.; Robinson, T. M.; Wong, W. Y.; Pipe, S. W. Two-Year Outcomes of Valoctocogene Roxaparvovec Therapy for Hemophilia A. *N. Engl. J. Med.* **2023**, *388* (8), 694–705. <https://doi.org/10.1056/NEJMoa2211075>.
- (119) Durupt, S.; Durieu, I.; Nové-Josserand, R.; Bencharif, L.; Rousset, H.; Vital Durand, D. L'hémochromatose génétique. *Rev. Médecine Interne* **2000**, *21* (11), 961–971. [https://doi.org/10.1016/S0248-8663\(00\)00252-6](https://doi.org/10.1016/S0248-8663(00)00252-6).
- (120) Feder, J. N.; Gnirke, A.; Thomas, W.; Tsuchihashi, Z.; Ruddy, D. A.; Basava, A.; Dormishian, F.; Domingo, R.; Ellis, M. C.; Fullan, A.; Hinton, L. M.; Jones, N. L.; Kimmel, B. E.; Kronmal, G. S.; Lauer, P.; Lee, V. K.; Loeb, D. B.; Mapa, F. A.; McClelland, E.; Meyer, N. C.; Mintier, G. A.; Moeller, N.; Moore, T.; Morikang, E.; Prass, C. E.; Quintana, L.; Starnes, S. M.; Schatzman, R. C.; Brunke, K. J.; Drayna, D. T.; Risch, N. J.; Bacon, B. R.; Wolff, R. K. A Novel MHC Class I-like Gene Is Mutated in Patients with Hereditary Haemochromatosis. *Nat. Genet.* **1996**, *13* (4), 399–408. <https://doi.org/10.1038/ng0896-399>.
- (121) Hémochromatose de type 1. *Wikipédia*; 2022.
- (122) FH109655 | 1356390-47-0 | Hecpidin-25 (human) trifluoroacetate salt. <https://www.biosynth.com/p/FH109655/1356390-47-0-hepcidin-25-human-trifluoroacetat> (accessed 2023-04-18).
- (123) Si-Tayeb, K.; Lemaigre, F. P.; Duncan, S. A. Organogenesis and Development of the Liver. *Dev. Cell* **2010**, *18* (2), 175–189. <https://doi.org/10.1016/j.devcel.2010.01.011>.
- (124) *Histologie des glandes digestives Cartes*. Quizlet. <https://quizlet.com/fr/792155964/histologie-des-glandes-digestives-flash-cards/> (accessed 2023-05-10).
- (125) Duncan, A. W.; Hanlon Newell, A. E.; Smith, L.; Wilson, E. M.; Olson, S. B.; Thayer, M. J.; Strom, S. C.; Grompe, M. Frequent Aneuploidy Among Normal Human Hepatocytes. *Gastroenterology* **2012**, *142* (1), 25–28. <https://doi.org/10.1053/j.gastro.2011.10.029>.
- (126) D'Souza, A. A.; Devarajan, P. V. Asialoglycoprotein Receptor Mediated Hepatocyte Targeting — Strategies and Applications. *J. Controlled Release* **2015**, *203*, 126–139. <https://doi.org/10.1016/j.jconrel.2015.02.022>.
- (127) Poelstra, K.; Prakash, J.; Beljaars, L. Drug Targeting to the Diseased Liver. *J. Control. Release Off. J. Control. Release Soc.* **2012**, *161* (2), 188–197. <https://doi.org/10.1016/j.jconrel.2012.02.011>.

- (128) Monroe, R. S.; Huber, B. E. The Major Form of the Murine Asialoglycoprotein Receptor: cDNA Sequence and Expression in Liver, Testis and Epididymis. *Gene* **1994**, *148* (2), 237–244. [https://doi.org/10.1016/0378-1119\(94\)90694-7](https://doi.org/10.1016/0378-1119(94)90694-7).
- (129) Weigel, P. H.; Oka, J. A. The Large Intracellular Pool of Asialoglycoprotein Receptors Functions during the Endocytosis of Asialoglycoproteins by Isolated Rat Hepatocytes. *J. Biol. Chem.* **1983**, *258* (8), 5095–5102. [https://doi.org/10.1016/S0021-9258\(18\)32543-2](https://doi.org/10.1016/S0021-9258(18)32543-2).
- (130) Ashwell, G.; Harford, J. Carbohydrate-Specific Receptors of the Liver. *Annu. Rev. Biochem.* **1982**, *51*, 531–554. <https://doi.org/10.1146/annurev.bi.51.070182.002531>.
- (131) Zelensky, A. N.; Gready, J. E. The C-Type Lectin-like Domain Superfamily. *FEBS J.* **2005**, *272* (24), 6179–6217. <https://doi.org/10.1111/j.1742-4658.2005.05031.x>.
- (132) Spiess, M. The Asialoglycoprotein Receptor: A Model for Endocytic Transport Receptors. *Biochemistry* **1990**, *29* (43), 10009–10018. <https://doi.org/10.1021/bi00495a001>.
- (133) Bider, M. D.; Spiess, M. Ligand-Induced Endocytosis of the Asialoglycoprotein Receptor: Evidence for Heterogeneity in Subunit Oligomerization. *FEBS Lett.* **1998**, *434* (1–2), 37–41. [https://doi.org/10.1016/S0014-5793\(98\)00947-8](https://doi.org/10.1016/S0014-5793(98)00947-8).
- (134) Schwartz, A. L.; Fridovich, S. E.; Lodish, H. F. Kinetics of Internalization and Recycling of the Asialoglycoprotein Receptor in a Hepatoma Cell Line. *J. Biol. Chem.* **1982**, *257* (8), 4230–4237.
- (135) Oligomeric Structure of the Human Asialoglycoprotein Receptor: Nature and Stoichiometry of Mutual Complexes Containing H1 and H2 Polypeptides Assessed by Fluorescence Photobleaching Recovery. *J. Cell Biol.* **1990**, *111* (4), 1409–1418.
- (136) Stockert, R. J. The Asialoglycoprotein Receptor: Relationships between Structure, Function, and Expression. *Physiol. Rev.* **1995**, *75* (3), 591–609. <https://doi.org/10.1152/physrev.1995.75.3.591>.
- (137) Kolatkar, A. R.; Weis, W. I. Structural Basis of Galactose Recognition by C-Type Animal Lectins. *J. Biol. Chem.* **1996**, *271* (12), 6679–6685. <https://doi.org/10.1074/jbc.271.12.6679>.
- (138) Kolatkar, A. R.; Leung, A. K.; Isecke, R.; Brossmer, R.; Drickamer, K.; Weis, W. I. Mechanism of N-Acetylgalactosamine Binding to a C-Type Animal Lectin Carbohydrate-Recognition Domain. *J. Biol. Chem.* **1998**, *273* (31), 19502–19508. <https://doi.org/10.1074/jbc.273.31.19502>.
- (139) Meier, M.; Bider, M. D.; Malashkevich, V. N.; Spiess, M.; Burkhard, P. Crystal Structure of the Carbohydrate Recognition Domain of the H1 Subunit of the Asialoglycoprotein Receptor. **9**.
- (140) McKee, T. D.; DeRome, M. E.; Wu, G. Y.; Findeis, M. A. Preparation of Asialoorosomucoid-Polylysine Conjugates. *Bioconjug. Chem.* **1994**, *5* (4), 306–311. <https://doi.org/10.1021/bc00028a004>.
- (141) Arangoa, M. A.; Düzgüneş, N.; Tros de Ilarduya, C. Increased Receptor-Mediated Gene Delivery to the Liver by Protamine-Enhanced-Asialofetuin-Lipoplexes. *Gene Ther.* **2003**, *10* (1), 5–14. <https://doi.org/10.1038/sj.gt.3301840>.
- (142) Stokmaier, D.; Khorev, O.; Cutting, B.; Born, R.; Ricklin, D.; Ernst, T. O. G.; Böni, F.; Schwingruber, K.; Gentner, M.; Wittwer, M.; Spreafico, M.; Vedani, A.; Rabbani, S.; Schwardt, O.; Ernst, B. Design, Synthesis and Evaluation of Monovalent Ligands for the Asialoglycoprotein Receptor (ASGP-R). *Bioorg. Med. Chem.* **2009**, *17* (20), 7254–7264. <https://doi.org/10.1016/j.bmc.2009.08.049>.
- (143) Mamidyala, S. K.; Dutta, S.; Chrnyk, B. A.; Préville, C.; Wang, H.; Withka, J. M.; McColl, A.; Subashi, T. A.; Hawrylik, S. J.; Griffor, M. C.; Kim, S.; Pfefferkorn, J. A.; Price, D. A.; Menhaji-Klotz, E.; Mascitti, V.; Finn, M. G. Glycomimetic Ligands for the Human Asialoglycoprotein Receptor. *J. Am. Chem. Soc.* **2012**, *134* (4), 1978–1981. <https://doi.org/10.1021/ja2104679>.
- (144) Sanhueza, C. A.; Baksh, M. M.; Thuma, B.; Roy, M. D.; Dutta, S.; Préville, C.; Chrnyk, B. A.; Beaumont, K.; Dullea, R.; Ammirati, M.; Liu, S.; Gebhard, D.; Finley, J. E.; Salatto, C. T.; King-Ahmad, A.; Stock, I.; Atkinson, K.; Reidich, B.; Lin, W.; Kumar, R.; Tu, M.; Menhaji-Klotz, E.; Price, D. A.; Liras, S.; Finn, M. G.; Mascitti, V. Efficient Liver Targeting by Polyvalent Display of a Compact Ligand for the Asialoglycoprotein Receptor. *J. Am. Chem. Soc.* **2017**, *139* (9), 3528–3536. <https://doi.org/10.1021/jacs.6b12964>.

- (145) Rouet, R.; Thuma, B. A.; Roy, M. D.; Lintner, N. G.; Rubitski, D. M.; Finley, J. E.; Wisniewska, H. M.; Mendonsa, R.; Hirsh, A.; de Oñate, L.; Compte Barrón, J.; McLellan, T. J.; Bellenger, J.; Feng, X.; Varghese, A.; Chrunchy, B. A.; Borzilleri, K.; Hesp, K. D.; Zhou, K.; Ma, N.; Tu, M.; Dullea, R.; McClure, K. F.; Wilson, R. C.; Liras, S.; Mascitti, V.; Doudna, J. A. Receptor-Mediated Delivery of CRISPR-Cas9 Endonuclease for Cell-Type-Specific Gene Editing. *J. Am. Chem. Soc.* **2018**, *140* (21), 6596–6603. <https://doi.org/10.1021/jacs.8b01551>.
- (146) Francisco, C. G.; Freire, R.; Herrera, A. J.; Pérez-Martín, I.; Suárez, E. Intramolecular 1,5-versus 1,6-Hydrogen Abstraction Reaction Promoted by Alkoxy Radicals in Carbohydrate Models. *Org. Lett.* **2002**, *4* (11), 1959–1961. <https://doi.org/10.1021/ol025981u>.
- (147) Trahanovsky, W. S.; Cramer, J. Oxidation of Organic Compounds with Cerium(IV). XII. Oxidative Cleavage and Ketone Formation of Alkylphenylcarbinols. *J. Org. Chem.* **1971**, *36* (14), 1890–1893. <https://doi.org/10.1021/jo00813a010>.
- (148) Seeberger, P. H.; Roehrig, S.; Schell, P.; Wang, Y.; Christ, W. J. Selective Formation of C-2 Azidodeoxy-d-Glucose Derivatives from d-Glucal Precursors Using the Azidonitration Reaction. *Carbohydr. Res.* **2000**, *328* (1), 61–69. [https://doi.org/10.1016/S0008-6215\(00\)00111-7](https://doi.org/10.1016/S0008-6215(00)00111-7).
- (149) Paulsen, H.; Paal, M. Lewissäure-katalysierte synthesen von di- und trisaccharid-sequenzen der O- und N-glycoproteine. Anwendung von trimethylsilyltrifluoromethanesulfonat. *Carbohydr. Res.* **1984**, *135* (1), 53–69. [https://doi.org/10.1016/0008-6215\(84\)85005-3](https://doi.org/10.1016/0008-6215(84)85005-3).
- (150) Parikh, J. R.; Doering, W. v. E. Sulfur Trioxide in the Oxidation of Alcohols by Dimethyl Sulfoxide. *J. Am. Chem. Soc.* **1967**, *89* (21), 5505–5507. <https://doi.org/10.1021/ja00997a067>.
- (151) Francisco\*, C. G.; Herrera, A. J.; Suárez, E. Intramolecular Hydrogen Abstraction Reaction in Carbohydrate Chemistry. Synthesis of Chiral 2,7-Dioxabicyclo[2.2.1]Heptane and 6,8-Dioxabicyclo[3.2.1]Octane Ring Systems. *Tetrahedron Lett.* **2000**, *41* (41), 7869–7873. [https://doi.org/10.1016/S0040-4039\(00\)01393-9](https://doi.org/10.1016/S0040-4039(00)01393-9).
- (152) León, E. I.; Martín, Á.; Montes, A. S.; Pérez-Martín, I.; del Sol Rodríguez, M.; Suárez, E. 1,5-Hydrogen Atom Transfer/Surzur–Tanner Rearrangement: A Radical Cascade Approach for the Synthesis of 1,6-Dioxaspiro[4.5]Decane and 6,8-Dioxabicyclo[3.2.1]Octane Scaffolds in Carbohydrate Systems. *J. Org. Chem.* **2021**, *86* (21), 14508–14552. <https://doi.org/10.1021/acs.joc.1c01376>.
- (153) Francisco, C. G.; Herrera, A. J.; Suárez, E. Intramolecular Hydrogen Abstraction Reaction Promoted by Alkoxy Radicals in Carbohydrates. Synthesis of Chiral 2,7-Dioxabicyclo[2.2.1]Heptane and 6,8-Dioxabicyclo[3.2.1]Octane Ring Systems. *J. Org. Chem.* **2002**, *67* (21), 7439–7445. <https://doi.org/10.1021/jo026004z>.
- (154) Concepcion, J. I.; Francisco, C. G.; Freire, R.; Hernandez, R.; Salazar, J. A.; Suarez, E. Iodosobenzene Diacetate, an Efficient Reagent for the Oxidative Decarboxylation of Carboxylic Acids. *J. Org. Chem.* **1986**, *51* (3), 402–404. <https://doi.org/10.1021/jo00353a026>.
- (155) Guillaume, J.; Seki, T.; Decruy, T.; Venken, K.; Elewaut, D.; Tsuji, M.; Calenbergh, S. V. Synthesis of C6''-Modified  $\alpha$ -C-GalCer Analogues as Mouse and Human iNKT Cell Agonists. *Org. Biomol. Chem.* **2017**, *15* (10), 2217–2225. <https://doi.org/10.1039/C7OB00081B>.
- (156) Lövyova, Z.; Parkan, K.; Kniežo, L. Stereoselective Preparation of Four 3-C-Mannosylated d- and l-Glucals from a Single Starting Compound. *Tetrahedron* **2011**, *67* (27), 4967–4979. <https://doi.org/10.1016/j.tet.2011.04.044>.
- (157) Kramer, J. R.; Deming, T. J. Glycopolypeptides via Living Polymerization of Glycosylated- L - Lysine N -Carboxyanhydrides. *J. Am. Chem. Soc.* **2010**, *132* (42), 15068–15071. <https://doi.org/10.1021/ja107425f>.
- (158) Cui, J.; Horton, D. Stereocontrolled Allylation of 2-Amino-2-Deoxy Sugar Derivatives by a Free-Radical procedure<sup>11</sup>For Preliminary Reports on This Work, See Ref[1]. *Carbohydr. Res.* **1998**, *309* (4), 319–330. [https://doi.org/10.1016/S0008-6215\(98\)00141-4](https://doi.org/10.1016/S0008-6215(98)00141-4).
- (159) Röhrig, C. H.; Takhi, M.; Schmidt, R. R. Synthesis of a C-Glycoside Analog of the Tumor-Associated Tn Antigen. *Synlett* **2001**, *2001* (07), 1170–1172. <https://doi.org/10.1055/s-2001-15144>.

- (160) Mootoo, D. R.; Konradsson, Peter.; Udodong, Uko.; Fraser-Reid, Bert. Armed and Disarmed N-Pentenyl Glycosides in Saccharide Couplings Leading to Oligosaccharides. *J. Am. Chem. Soc.* **1988**, *110* (16), 5583–5584. <https://doi.org/10.1021/ja00224a060>.
- (161) Zhang, Z.; Ollmann, I. R.; Ye, X.-S.; Wischnat, R.; Baasov, T.; Wong, C.-H. Programmable One-Pot Oligosaccharide Synthesis. *J. Am. Chem. Soc.* **1999**, *121* (4), 734–753. <https://doi.org/10.1021/ja982232s>.
- (162) Bertozzi, C. R.; Bednarski, M. D. The Synthesis of 2-Azido C-Glycosyl Sugars. *Tetrahedron Lett.* **1992**, *33* (22), 3109–3112. [https://doi.org/10.1016/S0040-4039\(00\)79826-1](https://doi.org/10.1016/S0040-4039(00)79826-1).
- (163) Keith, D. J.; Townsend, S. D. Direct, Microwave-Assisted Substitution of Anomeric Nitrate Esters. *Carbohydr. Res.* **2017**, *442*, 20–24. <https://doi.org/10.1016/j.carres.2017.02.005>.
- (164) Tam, R. Y.; Ferreira, S. S.; Czechura, P.; Chaytor, J. L.; Ben, R. N. Hydration Index—A Better Parameter for Explaining Small Molecule Hydration in Inhibition of Ice Recrystallization. *J. Am. Chem. Soc.* **2008**, *130* (51), 17494–17501. <https://doi.org/10.1021/ja806284x>.
- (165) Saidhareddy, P.; Ajay, S.; Shaw, A. K. Iodosobenzene Diacetate-Iodine and IBX-Iodine: Reagent Systems for the Synthesis of Diastereomerically Enriched 2-Deoxy-2-Iodoglycosyl Acetates and 2-Deoxy-2-Iodoglycosyl Ortho-Iodobenzoates from Protected Glycals. *Tetrahedron* **2017**, *73* (30), 4407–4417. <https://doi.org/10.1016/j.tet.2017.06.001>.
- (166) Priebbenow, D. L.; Gable, Robert. W.; Baell, J. Regio- and Stereoselective Iodoacyloxylation of Alkynes. *J. Org. Chem.* **2015**, *80* (9), 4412–4418. <https://doi.org/10.1021/acs.joc.5b00250>.
- (167) Ferrier, R. J.; Zubkov, O. A. Transformation of Glycals into 2,3-Unsaturated Glycosyl Derivatives. In *Organic Reactions*; John Wiley & Sons, Ltd, 2004; pp 569–736. <https://doi.org/10.1002/0471264180.or062.04>.
- (168) Fuchss, T.; Schmidt, R. R. Synthesis of the C-Analog of 2-Acetylamino-2-Deoxy- $\beta$ -d-Glucopyranosyl I- and d-Serine. *Synthesis* **1998**, *1998* (05), 753–758. <https://doi.org/10.1055/s-1998-4490>.
- (169) Reintjens, N. R. M.; Koemans, T. S.; Zilverschoon, N.; Castelli, R.; Cordfunke, R. A.; Drijfhout, J. W.; Meeuwenoord, N. J.; Overkleeft, H. S.; Filippov, D. V.; van der Marel, G. A.; Codée, J. D. C. Synthesis of C-Glycosyl Amino Acid Building Blocks Suitable for the Solid-Phase Synthesis of Multivalent Glycopeptide Mimics. *Eur. J. Org. Chem.* **2020**, *2020* (32), 5126–5139. <https://doi.org/10.1002/ejoc.202000587>.
- (170) Roe, B. A.; Boojamra, C. G.; Griggs, J. L.; Bertozzi, C. R. Synthesis of  $\beta$ -C-Glycosides of N-Acetylglucosamine via Keck Allylation Directed by Neighboring Phthalimide Groups. *J. Org. Chem.* **1996**, *61* (18), 6442–6445. <https://doi.org/10.1021/jo960819o>.
- (171) Chennaiah, A.; Vankar, Y. D. One-Step TEMPO-Catalyzed and Water-Mediated Stereoselective Conversion of Glycals into 2-Azido-2-Deoxysugars with a PIFA–Trimethylsilyl Azide Reagent System. *Org. Lett.* **2018**, *20* (9), 2611–2614. <https://doi.org/10.1021/acs.orglett.8b00814>.
- (172) Matcha, K.; Antonchick, A. P. Metal-Free Cross-Dehydrogenative Coupling of Heterocycles with Aldehydes. *Angew. Chem.* **2013**, *125* (7), 2136–2140. <https://doi.org/10.1002/ange.201208851>.
- (173) Magnus, P.; Lacour, J.; Evans, P. A.; Roe, M. B.; Hulme, C. Hypervalent Iodine Chemistry: New Oxidation Reactions Using the Iodosylbenzene-Trimethylsilyl Azide Reagent Combination. Direct R- and  $\beta$ -Azido Functionalization of Triisopropylsilyl Enol Ethers. *HyperValent Iodine Chem.*
- (174) Petri, L.; Szijj, P. A.; Kelemen, Á.; Imre, T.; Gömöry, Á.; Lee, M. T. W.; Hegedűs, K.; Ábrányi-Balogh, P.; Chudasama, V.; Keserű, G. M. Cysteine Specific Bioconjugation with Benzyl Isothiocyanates. *RSC Adv.* **2020**, *10* (25), 14928–14936. <https://doi.org/10.1039/D0RA02934C>.
- (175) Wu, A. M.; Song, S. C.; Sugii, S.; Herp, A. Differential Binding Properties of Gal/GalNAc Specific Lectins Available for Characterization of Glycoreceptors. *Indian J. Biochem. Biophys.* **1997**, *34* (1–2), 61–71.
- (176) Moskalenko, M.; Chen, L.; van Roey, M.; Donahue, B. A.; Snyder, R. O.; McArthur, J. G.; Patel, S. D. Epitope Mapping of Human Anti-Adeno-Associated Virus Type 2 Neutralizing Antibodies: Implications for Gene Therapy and Virus Structure. *J. Virol.* **2000**, *74* (4), 1761–1766. <https://doi.org/10.1128/jvi.74.4.1761-1766.2000>.

- (177) antibodies-online. *Qu'est-ce que la cytométrie de flux (analyse FACS) ?*  
<https://www.anticorps-enligne.fr/resources/17/1247/qu-est-ce-que-la-cytometrie-de-flux-analyse-facs/> (accessed 2023-05-06).
- (178) Pipe, S.; Leebeek, F. W. G.; Ferreira, V.; Sawyer, E. K.; Pasi, J. Clinical Considerations for Capsid Choice in the Development of Liver-Targeted AAV-Based Gene Transfer. *Mol. Ther. - Methods Clin. Dev.* **2019**, *15*, 170–178. <https://doi.org/10.1016/j.omtm.2019.08.015>.
- (179) Nakai, H.; Thomas, C. E.; Storm, T. A.; Fuess, S.; Powell, S.; Wright, J. F.; Kay, M. A. A Limited Number of Transducible Hepatocytes Restricts a Wide-Range Linear Vector Dose Response in Recombinant Adeno-Associated Virus-Mediated Liver Transduction. *J. Virol.* **2002**, *76* (22), 11343–11349. <https://doi.org/10.1128/JVI.76.22.11343-11349.2002>.
- (180) Thomas, C. E.; Storm, T. A.; Huang, Z.; Kay, M. A. Rapid Uncoating of Vector Genomes Is the Key to Efficient Liver Transduction with Pseudotyped Adeno-Associated Virus Vectors. *J. Virol.* **2004**, *78* (6), 3110–3122. <https://doi.org/10.1128/JVI.78.6.3110-3122.2004>.
- (181) Stolz, F.; Blume, A.; Hinderlich, S.; Reutter, W.; Schmidt, R. R. C-Glycosidic UDP-GlcNAc Analogues as Inhibitors of UDP-GlcNAc 2-Epimerase. *Eur. J. Org. Chem.* **2004**, *2004* (15), 3304–3312. <https://doi.org/10.1002/ejoc.200400197>.
- (182) Lehot, V.; Brissonnet, Y.; Dussouy, C.; Brument, S.; Cabanettes, A.; Gillon, E.; Deniaud, D.; Varrot, A.; Le Pape, P.; Gouin, S. G. Multivalent Fucosides with Nanomolar Affinity for the *Aspergillus Fumigatus* Lectin FleA Prevent Spore Adhesion to Pneumocytes. *Chem. – Eur. J.* **2018**, *24* (72), 19243–19249. <https://doi.org/10.1002/chem.201803602>.
- (183) Assailly, C.; Bridot, C.; Saumonneau, A.; Lottin, P.; Roubinet, B.; Krammer, E.-M.; François, F.; Vena, F.; Landemarre, L.; Alvarez Dorta, D.; Deniaud, D.; Grandjean, C.; Tellier, C.; Pascual, S.; Montembault, V.; Fontaine, L.; Daligault, F.; Bouckaert, J.; Gouin, S. G. Polyvalent Transition-State Analogues of Sialyl Substrates Strongly Inhibit Bacterial Sialidases\*\*. *Chem. – Eur. J.* **2021**, *27* (9), 3142–3150. <https://doi.org/10.1002/chem.202004672>.
- (184) Rensen, P. C. N.; Sliedregt, L. A. J. M.; Ferns, M.; Kieviet, E.; van Rossenberg, S. M. W.; van Leeuwen, S. H.; van Berkel, T. J. C.; Biessen, E. A. L. Determination of the Upper Size Limit for Uptake and Processing of Ligands by the Asialoglycoprotein Receptor on Hepatocytes in Vitro and in Vivo\*. *J. Biol. Chem.* **2001**, *276* (40), 37577–37584. <https://doi.org/10.1074/jbc.M101786200>.
- (185) Lee, Y. C.; Townsend, R. R.; Hardy, M. R.; Lönngren, J.; Arnarp, J.; Haraldsson, M.; Lönn, H. Binding of Synthetic Oligosaccharides to the Hepatic Gal/GalNAc Lectin. Dependence on Fine Structural Features. *J. Biol. Chem.* **1983**, *258* (1), 199–202. [https://doi.org/10.1016/S0021-9258\(18\)33240-X](https://doi.org/10.1016/S0021-9258(18)33240-X).
- (186) Huang, X.; Leroux, J.-C.; Castagner, B. Well-Defined Multivalent Ligands for Hepatocytes Targeting via Asialoglycoprotein Receptor. *Bioconj. Chem.* **2017**, *28* (2), 283–295. <https://doi.org/10.1021/acs.bioconjchem.6b00651>.
- (187) Khatri, V.; Parshad, B.; Prasad, A. K.; Bhatia, S. Design, Synthesis, and Biomedical Applications of Glycotripods for Targeting Trimeric Lectins. *Eur. J. Org. Chem.* **2023**, *26* (9), e202201360. <https://doi.org/10.1002/ejoc.202201360>.
- (188) Davidoff, A. M.; Gray, J. T.; Ng, C. Y. C.; Zhang, Y.; Zhou, J.; Spence, Y.; Bakar, Y.; Nathwani, A. C. Comparison of the Ability of Adeno-Associated Viral Vectors Pseudotyped with Serotype 2, 5, and 8 Capsid Proteins to Mediate Efficient Transduction of the Liver in Murine and Nonhuman Primate Models. *Mol. Ther. J. Am. Soc. Gene Ther.* **2005**, *11* (6), 875–888. <https://doi.org/10.1016/j.ymthe.2004.12.022>.
- (189) Oba, Y.; Ojika, M.; Inouye, S. Firefly Luciferase Is a Bifunctional Enzyme: ATP-Dependent Monooxygenase and a Long Chain Fatty Acyl-CoA Synthetase. *FEBS Lett.* **2003**, *540* (1), 251–254. [https://doi.org/10.1016/S0014-5793\(03\)00272-2](https://doi.org/10.1016/S0014-5793(03)00272-2).
- (190) Dowlut, M.; Hall, D. G.; Hindsgaul, O. Investigation of Nonspecific Effects of Different Dyes in the Screening of Labeled Carbohydrates against Immobilized Proteins. *J. Org. Chem.* **2005**, *70* (24), 9809–9813. <https://doi.org/10.1021/jo051503w>.

- (191) Bouvet, V. R.; Ben, R. N. A Short and Economical Synthesis of Orthogonally Protected C-Linked 2-Deoxy-2-Acetamido- $\alpha$ -D-Galactopyranose Derivatives. *J. Org. Chem.* **2006**, *71* (9), 3619–3622. <https://doi.org/10.1021/jo051938j>.
- (192) Jiang, F.; Liu, Q.; Zhao, D.; Luo, C.; Guo, C.; Ye, W.; Luo, C.; Chen, H. A Concise Synthesis of N-Substituted Fagomine Derivatives and the Systematic Exploration of Their  $\alpha$ -Glycosidase Inhibition. *Eur. J. Med. Chem.* **2014**, *77*, 211–222. <https://doi.org/10.1016/j.ejmech.2014.03.004>.
- (193) Singh, Y.; Wang, T.; Demchenko, A. V. Direct Glycosidation of 2-Azido-2-deoxyglycosyl Nitrates. *Eur. J. Org. Chem.* **2019**, *2019* (37), 6413–6416. <https://doi.org/10.1002/ejoc.201900748>.
- (194) Briner, K.; Vasella, A. Glycosylphosphonates of 2-Amino-2-Deoxy-Aldoses. Synthesis of a Phosphonate Analogue of Lipid X. *Helv. Chim. Acta* **1987**, *70* (5), 1341–1356. <https://doi.org/10.1002/hlca.19870700516>.
- (195) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47* (7), 1750–1759. <https://doi.org/10.1021/jm030644s>.
- (196) *Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes | Journal of Medicinal Chemistry.* <https://pubs.acs.org/doi/10.1021/jm051256o> (accessed 2023-05-10).
- (197) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7* (2), 525–537. <https://doi.org/10.1021/ct100578z>.
- (198) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2017**, *14* (1), 71–73. <https://doi.org/10.1038/nmeth.4067>.
- (199) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B. K.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. *J. Chem. Phys.* **2020**, *153* (4), 044130. <https://doi.org/10.1063/5.0014475>.
- (200) Guvench, O.; Mallajosyula, S. S.; Raman, E. P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T. J.; Jamison, F. W.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Carbohydrate Derivatives and Its Utility in Polysaccharide and Carbohydrate-Protein Modeling. *J. Chem. Theory Comput.* **2011**, *7* (10), 3162–3180. <https://doi.org/10.1021/ct200328p>.
- (201) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D. CHARMM General Force Field (CGenFF): A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31* (4), 671–690. <https://doi.org/10.1002/jcc.21367>.
- (202) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935. <https://doi.org/10.1063/1.445869>.
- (203) D’Costa, S.; Blouin, V.; Broucque, F.; Penaud-Budloo, M.; François, A.; Perez, I. C.; Le Bec, C.; Moullier, P.; Snyder, R. O.; Ayuso, E. Practical Utilization of Recombinant AAV Vector Reference Standards: Focus on Vector Genomes Titration by Free ITR qPCR. *Mol. Ther. - Methods Clin. Dev.* **2016**, *3*, 16019. <https://doi.org/10.1038/mtm.2016.19>.
- (204) Lock, M.; McGorray, S.; Auricchio, A.; Ayuso, E.; Beecham, E. J.; Blouin-Tavel, V.; Bosch, F.; Bose, M.; Byrne, B. J.; Caton, T.; Chiorini, J. A.; Chtarto, A.; Clark, K. R.; Conlon, T.; Darmon, C.; Doria, M.; Douar, A.; Flotte, T. R.; Francis, J. D.; Francois, A.; Giacca, M.; Korn, M. T.; Korytov, I.; Leon, X.; Leuchs, B.; Lux, G.; Melas, C.; Mizukami, H.; Moullier, P.; Müller, M.; Ozawa, K.; Philipsberg, T.; Poulard, K.; Raupp, C.; Rivière, C.; Roosendaal, S. D.; Samulski, R. J.; Soltys, S. M.; Surosky, R.; Tenenbaum, L.; Thomas, D. L.; van Montfort, B.; Veres, G.; Wright, J. F.; Xu, Y.; Zelenia, O.; Zentilin, L.; Snyder, R. O. Characterization of a Recombinant Adeno-Associated Virus

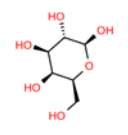
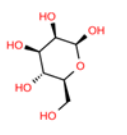
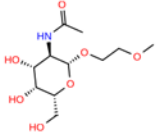
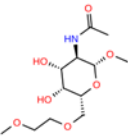
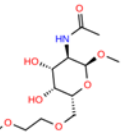
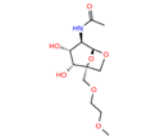
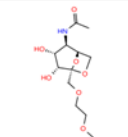
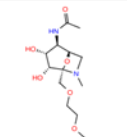
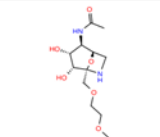
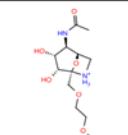
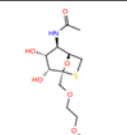
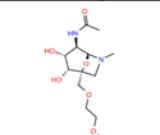
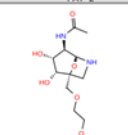
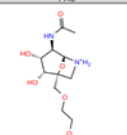
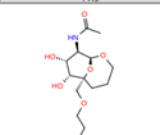
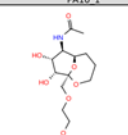
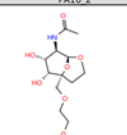
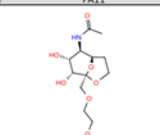
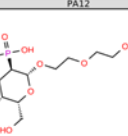
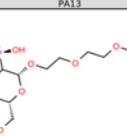
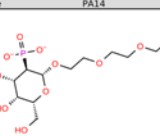
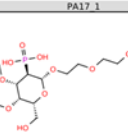
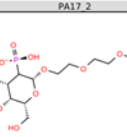
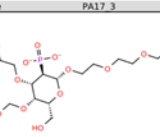


Type 2 Reference Standard Material. *Hum. Gene Ther.* **2010**, *21* (10), 1273–1285.  
<https://doi.org/10.1089/hum.2009.223>.

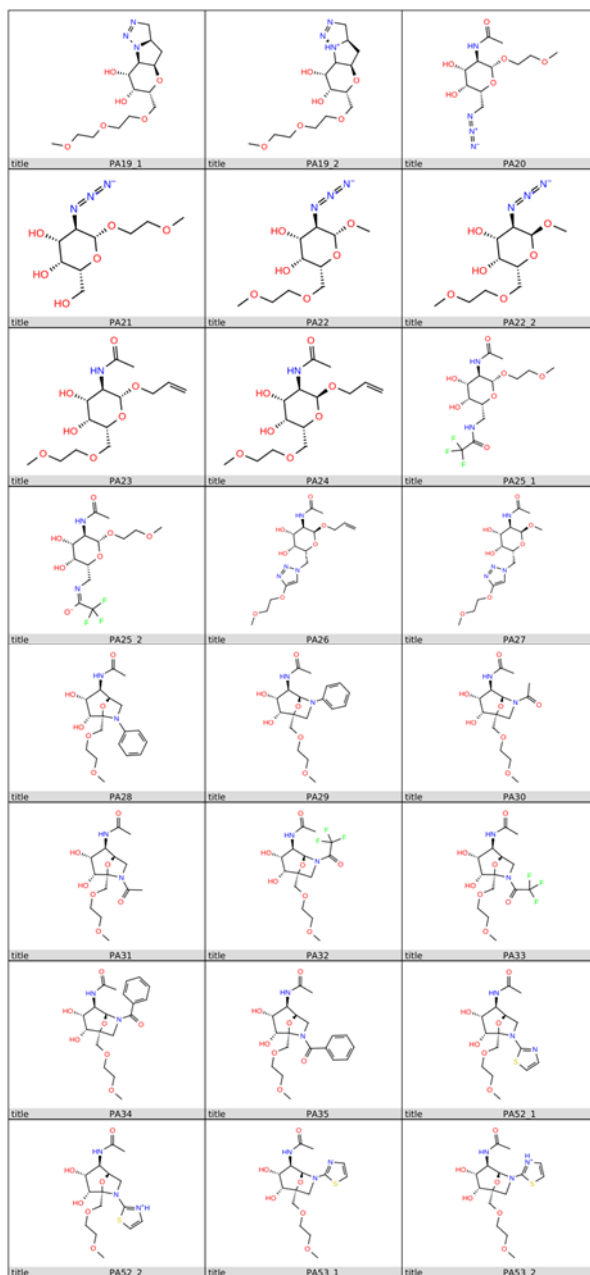
- (205) Guilbaud, M.; Devaux, M.; Couzinié, C.; Le Duff, J.; Toromanoff, A.; Vandamme, C.; Jaulin, N.; Gernoux, G.; Larcher, T.; Moullier, P.; Le Guiner, C.; Adjali, O. Five Years of Successful Inducible Transgene Expression Following Locoregional Adeno-Associated Virus Delivery in Nonhuman Primates with No Detectable Immunity. *Hum. Gene Ther.* **2019**, *30* (7), 802–813.  
<https://doi.org/10.1089/hum.2018.234>.



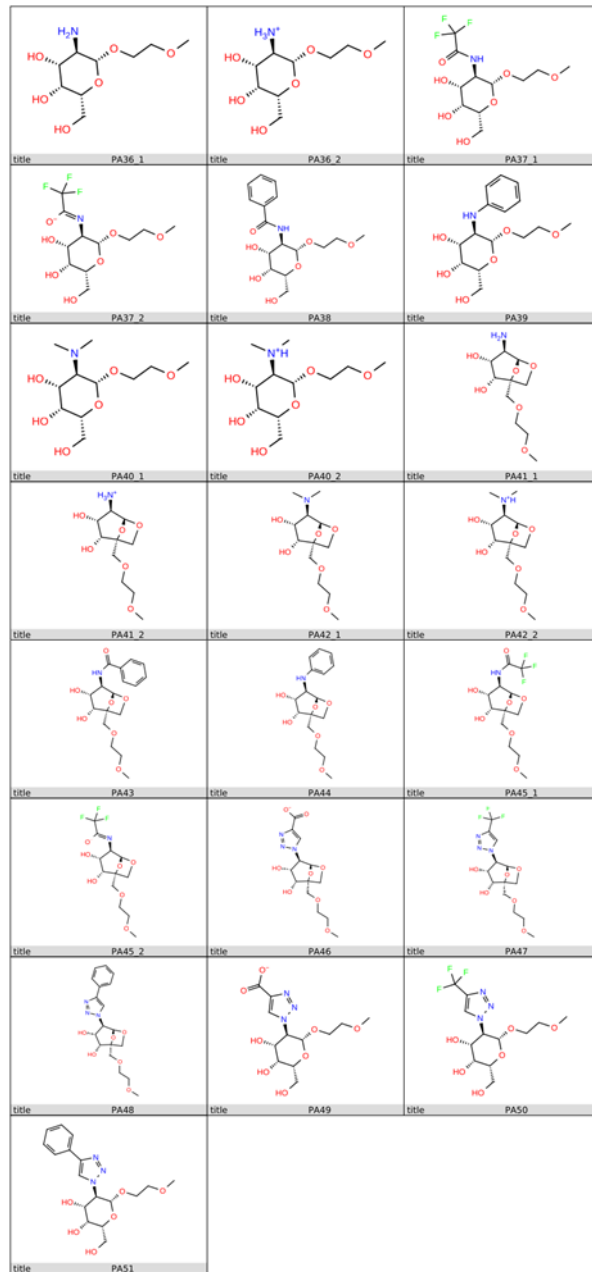
# Annexes

		
title B-Galactose	title B-Mannose	title PA1
		
title PA2	title PA3	title PA4
		
title PA5	title PA6	title PA7 1
		
title PA7 2	title PA8	title PA9
		
title PA10 1	title PA10 2	title PA11
		
title PA12	title PA13	title PA14
		
title PA17 1	title PA17 2	title PA17 3
		
title PA18 1	title PA18 2	title PA18 3

Annexe 1 : Chimiothèque des molécules utilisées pour le docking (1).



Annexe 2 : Chimiothèque des molécules utilisées pour le docking (2).



Annexe 3 : Chimiothèque des molécules utilisées pour le docking (3).

Abréviation	Nom commun	GalNAc	Gala	Galb	GlcNAc	Fuc	Glycans structures specificity
BPA	<i>Bauhinia purpurea</i>	+					Gal bêta3 GalNAc (T-antigen) > GalNAc > Gal
DBA	<i>Dolichos biflorus</i>	+					Terminal GalNAc alpha, GalNAc alpha 3GalNAc, trisaccharide
HPA	<i>Helix pomatia</i>	+					GalNAc alpha
SBA	<i>Glycine max</i>	+					préférence pour les lien bêta-glycosidique I
WFA	<i>Wisteria floribunda</i>	+					GalNAc alpha 6Gal > GalNAc alpha 3GalNAc > GalNAc
ASGPr	Recombinant human Asialoglycoprotein receptor 1	+					GalNAc
AIA / Jacalin	<i>Artocarpus integrifolia</i>		+	+			Gala6 or Galb3GalNAc (T-antigen)>> lactose, more specific for T-antigen than PNA
DSL (DSA) +	<i>Datura stramonium</i>				+		GlcNAcb4GlcNAc oligomers, Galb4GlcNAc,
UEA I	<i>Ulex europeus</i>					+	Fucose

Annexe 4 : Listes des lectines utilisées pour le GLYcoPROFILE. Gal, Man, GalNAc etc. sont les abréviations des motifs osidiques. (+) correspond aux principaux sucres inhibiteurs de l'interaction avec chaque lectine.

**Titre :** Nouvelle génération d'AAV : un mélange de chimie organique et de vectorologie.

**Mots clés :** AAV, Bioconjugaison, Chimie des sucres, Vectorisation, Ciblage, Thérapie génique.

**Résumé :** Les virus adéno-associés ou AAV sont utilisés comme vecteurs pour le transfert de gènes thérapeutiques chez l'homme. Leur efficacité a été démontrée pour le traitement de maladies génétiques avec cinq médicaments actuellement sur le marché. Cependant, améliorer leur action est une priorité aujourd'hui nécessaire pour réduire les doses injectées et en conséquence, diminuer les effets indésirables, les quantités à produire et donc les coûts.

Les principales techniques pour optimiser les AAV font intervenir des procédés de génie génétique. Dans le cadre de la thèse, nous avons opté pour une approche innovante, combinant la chimie organique et la vectorologie, pour concevoir de nouvelles capsides chimiquement modifiées. Ainsi, des ligands ont été greffés par couplage covalent sur la capside de l'AAV2 afin de le protéger des anticorps circulants et de le guider vers les cellules du foie.

Après une partie introductive, ce manuscrit décrit la synthèse de nouvelles molécules permettant de cibler le récepteur aux asialoglycoprotéines, protéines présentes à la surface des hépatocytes, ainsi que la caractérisation et l'étude des propriétés biologiques *in vitro* et *in vivo* des vecteurs chimiquement modifiés. Les ligands sont formés d'une ancre permettant la bioconjugaison des lysines, d'un motif de reconnaissance des récepteurs du foie à base de GalNAc, le tout séparé par un espaceur triéthylène glycol. L'étude *in-vivo* sur des souris, à travers différentes analyses biomoléculaires et histologiques, a mis en évidence l'intérêt des capsides modifiées avec une meilleure transduction et une meilleure expression du gène dans le foie comparativement à l'AAV2 pour certains de nos ligands.

**Title:** New generation of AAV: a mixture of organic chemistry and vectorology.

**Keywords:** AAV, Bioconjugation, Carbohydrate Chemistry, Vectorization, Targeting, Gene therapy.

**Abstract:** Adeno-associated viruses or AAVs are used as vectors for therapeutic gene transfer in humans. Their efficacy has been demonstrated for the treatment of genetic diseases with five drugs currently on the market. However, improving their action is a priority today in order to reduce the doses injected and, consequently, to reduce the undesirable effects, the quantities to be produced and therefore the costs.

The main techniques to optimize AAVs involve genetic engineering processes. In the framework of this thesis, we opted for an innovative approach, combining organic chemistry and vectorology, to design new chemically modified capsids. Thus, ligands were grafted by covalent coupling onto the AAV2 capsid in order to protect it from circulating antibodies and to guide it to liver cells.

After an introductory part, this manuscript describes the synthesis of new molecules allowing to target the receptor to asialoglycoproteins, proteins present on the surface of hepatocytes, as well as the characterization and the study of the biological properties *in vitro* and *in vivo* of the chemically modified vectors. The ligands consist of a lysine bioconjugation anchor, a GalNAc-based liver receptor recognition motif, and a triethylene glycol spacer. The *in-vivo* study on mice, through various biomolecular and histological analyses, has highlighted the interest of modified capsids with a better transduction and a better gene expression in the liver compared to AAV2 for some of our ligands.