



**HAL**  
open science

# Reconnaissance spécifique d'ADN double brins : développement de sondes luminescentes

Thibault Charnay

► **To cite this version:**

Thibault Charnay. Reconnaissance spécifique d'ADN double brins : développement de sondes luminescentes. Chimie inorganique. Université Grenoble Alpes [2020-..], 2022. Français. NNT : 2022GRALV091 . tel-04086764

**HAL Id: tel-04086764**

**<https://theses.hal.science/tel-04086764>**

Submitted on 2 May 2023

**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.

## THÈSE

Pour obtenir le grade de

### DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : CSV- Chimie et Sciences du Vivant

Spécialité : Chimie inorganique et Bio inorganique

Unité de recherche : Laboratoire de Chimie et Biologie des Métaux

## Reconnaissance spécifique d'ADN double brins : développement de sondes luminescentes

## Specific recognition of double-stranded DNA: development of luminescent probes

Présentée par :

**Thibault CHARNAY**

### Direction de thèse :

**OLIVIER SENEQUE**

Directeur de Recherche, Université Grenoble Alpes

Directeur de thèse

**MURIEL JOURDAN**

Maître de Conférences, Université Grenoble Alpes

Co-directrice de thèse

### Rapporteurs :

**Célia Bonnet**

DIRECTEUR DE RECHERCHE, CNRS

**Anton Granzhan**

CHARGE DE RECHERCHE, CNRS

### Thèse soutenue publiquement le **25 novembre 2022**, devant le jury composé de :

**Olivier Sénèque**

DIRECTEUR DE RECHERCHE, Université Grenoble Alpes

Directeur de thèse

**Muriel Jourdan**

MAITRE DE CONFERENCES, Université Grenoble Alpes

Co-directrice de thèse

**Gilles Subra**

PROFESSEUR DES UNIVERSITES, Université de Montpellier

Examinateur

**Eric Peyrin**

PROFESSEUR, Université Grenoble Alpes

Président

**Célia Bonnet**

DIRECTEUR DE RECHERCHE, CNRS

Rapporteuse

**Anton Granzhan**

CHARGE DE RECHERCHE, CNRS

Rapporteur

### Invités :

**Jean-François Constant**

CHARGE DE RECHERCHE, CNRS







*À mes grands-pères partis trop tôt,  
À ma famille,  
À mes amis qui désormais font partie de ma famille,*

---

## Remerciements

J'ai vécu pendant pas loin de quatre ans cette formidable aventure qu'est la thèse. Elle a été pour moi extrêmement enrichissante d'un point personnel, mais aussi professionnel, parsemée par de chaleureuses rencontres et d'échanges plus ou moins scientifiques. Ce simple mot que je m'appête à dire à toutes ces personnes, je le dis avec une profonde sincérité. Il ne s'agit pas d'un simple « Merci » : c'est pour moi une reconnaissance de l'aide que j'ai reçue durant cette belle aventure. Elle se matérialise de différentes façons : des échanges scientifiques, des conseils avisés, mais aussi un soutien au quotidien, un fou rire, de longues discussions autour d'un verre (plus ou moins quelques autres) ... Mon directeur de thèse m'a toujours dit que la thèse devait être à mon image. Alors cette partie je vais la rédiger avec les valeurs humaines qui sont les miennes.

Parce qu'elle est la chose qui a le plus de valeurs à mes yeux, je vais démarrer ces remerciements en me tournant vers ma famille. Je mesure la chance d'avoir une famille telle que l'ont construite mes parents, avec mon frère et moi. C'est grâce aux efforts consentis par mes parents et grands-parents que j'ai pu aller au bout de mes études, me laissant guider par ce dont j'avais envie de faire, à savoir de la chimie. Ma famille est une force qui m'a permis de devenir docteur. On ne peut pas dire que j'étais prédestiné à de longues études, compte tenu de mon côté un peu bête, un peu Toto. Malgré cela, vous m'avez toujours encouragé et soutenu dans cette invraisemblable aventure. J'ai beaucoup parlé de ma thèse, mais je n'ai peut-être pas réussi à vous montrer l'investissement, le travail acharné que j'ai fournis au laboratoire comme à la rédaction. Je vous ai visiblement étonné lors de cette soutenance tant elle contraste avec le Toto que vous connaissez. Mais je garderai dans ma mémoire cette image de vous et de vos larmes de joie qui exprimaient toute la fierté que vous avez à mon égard.

Je vais à présent remercier chaleureusement les membres de mon jury, Dr. Célia BONNET, Dr. Anton GRANZHAN, Pr. Gilles SUBRA et Pr. Éric PEYRIN pour avoir rapporté ou examiné mes travaux de thèse ainsi que pour la riche discussion scientifique que nous avons eue lors de ma soutenance. Je vous remercie également pour vos retours très positifs sur la qualité du manuscrit/soutenance. Les rapports ainsi que le procès-verbal que vous avez rédigés seront un véritable atout pour la suite de mon projet.

Je veux ensuite remercier mon directeur de thèse, Olivier. Depuis notre première rencontre il y a près de 6 ans pour un stage de M1, tu m'as permis d'apprendre et d'évoluer pour être le docteur que je suis devenu. Avant cette thèse, j'ai connu une période sombre. A l'issue d'un Master de chimie, diverses personnes m'avaient bien fait comprendre que je n'avais pas les capacités, compétences et qualités nécessaires pour réaliser une thèse, quand bien même Olivier Hamelin, mon encadrant de stage en M2, en vantait les mérites. J'ai réussi malgré tout à convaincre deux personnes d'avoir confiance en moi pour faire une thèse mais les écoles doctorales de Grenoble puis de Bordeaux me l'ont refusée. Je suis alors venu vers toi, avec un dernier espoir, et tu n'as pas hésité à me donner ma chance pour que je puisse avoir le droit de réaliser une thèse. C'est un geste que je n'oublierai pas et dont je ne sais comment te

---

remercier. Pourtant, la thèse a très mal commencé entre nous... Nous n'entendions pas d'une même oreille les chansons du grandiose Michel Sardou !! J'avais d'autres Michel, Georges et Joe à te chanter, mais visiblement ma musique de vieux ne faisait pas l'unanimité. Je n'ose pas parler du chanteur « dont on ne doit pas entendre les chansons » ... même si l'on s'est permis avec Jean-François un soir au téléphone de t'apporter une formidable nouvelle en chanson !! (Je ne dirai bien évidemment pas le titre) Heureusement, le vieux rock nous a réconciliés, merci entre autres aux Pixies et Smashing Pumpkins. Je garde d'excellents souvenirs de notre relation, de nos longues discussions dans le bureau tôt le matin à la suite de l'habituel « ça va M'sieur Charnay ? », ou de celles le soir, dans le labo, à réfléchir et comprendre pourquoi ça ne brillait toujours pas ... Tout le temps, tu as su être disponible, pour m'expliquer la spectro, discuter des résultats ou lorsque je venais tout simplement dans le bureau juste pour dire que ça marchait (enfin quand ça voulait bien marcher). Etant d'un naturel peu confiant et anxieux, tu m'as fait grandir sur bien des aspects. Après une audition Arcane complètement manquée et un rapport de M2 noté 9/20, tu as malgré tout cru que je pouvais apprendre et m'améliorer ... La peur est le chemin qui mène vers le côté obscur, mais le Maître Jedi qui est en toi m'a permis d'arriver à maîtriser la Force (même si parfois il t'a fallu répéter/expliciter à plusieurs reprises). Les retours des rapporteurs ainsi que les compliments sur la qualité de la soutenance montrent tout le chemin que nous avons parcouru. Je dis souvent « nous » parce qu'avec ton implication et ton soutien toutes ces années, je ne conçois pas de m'accorder tout le mérite des résultats et des compliments que j'ai reçus. Ils sont l'illustration de la formidable relation que j'ai eue avec toi. Te voir aussi souvent avec la blouse et investi dans tous les projets montre ta bienveillance et le plaisir que tu as de faire de la recherche, mais surtout de la partager. Je finirai en parlant d'un SMS que tu m'as envoyé qui m'a rempli de joie à tel point que je n'ai pas pu m'empêcher de le dire avec insistance à tout le monde au labo, notamment à Geneviève. C'est la larme à l'œil que je te remercie profondément Olivier pour ces formidables années que j'ai vécues !! Je te souhaite bien évidemment plein de réussites pour les différents projets que tu mènes. Je ne me fais que peu de soucis compte tenu du nombre d'étudiants, doctorants, pots-doc qui sont au labo. Et je sais que je pourrai toujours compter sur toi.

Je m'adresse maintenant à Muriel et Jean-François avec qui j'ai souffert pour comprendre les techniques et notions de biologie qui m'étaient alors totalement étrangères. Nous sommes dans deux mondes différents, ce qui a donné de longues discussions où l'on n'arrivait pas à se comprendre, chacun utilisant son vocabulaire. Mais après une heure de débat, on se rendait compte ... que l'on disait la même chose ! Je voudrais m'excuser Jean-François, pour t'avoir apporté autant de travail à un moment aussi proche de la retraite, mais je sais que tu l'as fait avec plaisir. Tu as manifesté beaucoup d'enthousiasme pour élucider les problèmes rencontrés sur nos gels et sur l'analyse de cette fameuse bande retardée. Je m'excuse encore, mais cette fois auprès de ta femme, pour toutes les fois où je t'ai monopolisé jusqu'à la tombée de la nuit, à en louper l'heure du souper. C'était un réel plaisir de venir vous voir, Muriel et toi, pour faire nos gels en espérant y voir ce que l'on attendait, même si on a eu quelques gels qui nous

---

ont laissé assez perplexes. Je sais Muriel que tu aurais aimé être plus souvent présente au labo, mais le devoir t'appellait et je sais que l'enseignement prend beaucoup de temps. Je te souhaite le meilleur pour la suite, aussi bien professionnel que personnel. Jean-François, merci pour ta personnalité très ouverte et riche d'enseignements. A présent, je ne peux que te souhaiter une bonne retraite : elle est bien méritée ! Je te donne rendez-vous à l'ouverture de la pêche, en mars, pour une sortie en float tube.

Je veux maintenant rendre hommage à Dame Geneviève, à ne pas confondre avec la dame de fer, bien qu'elle soit une célébrité dans le monde du fer. C'était avant tout ma voisine de bureau, mais aussi le Grand Chef de nous. Ce titre impose le respect et le vouvoiement était de rigueur ! Mais aujourd'hui, c'est à mon amie Geneviève que j'écris. Nous avons très vite eu une relation d'amitié, sans doute parce que l'on partage des valeurs communes, mais surtout les mêmes défauts (je n'en fais pas la liste sinon je n'aurai plus de papier). Nous avons cette capacité un peu étrange de nous quereller dans les couloirs du CEA comme un vieux couple. Je ne te cache pas que ces scènes de ménage quotidiennes m'ont beaucoup fait rire et je pense ne pas être le seul à en avoir ri. J'ai longtemps cherché un surnom à te donner : Cléopâtre, Queen Geneviève I, Grand Schtroumpf ou encore grand-mère (il m'a échappé celui-là). Si de mon côté je cherchais des femmes de pouvoir, de ton côté tu as dit de moi que j'étais un enfant, alternant entre 5 et 8 ans selon les moments, un garnement ou alors un petit cochon grillé (il n'y a pas le contexte, mais ces mots ont été prononcés). Après plus de 2 ans à te demander « gentiment » de venir travailler au labo le week-end, j'ai enfin réussi à avoir ce privilège. Entre nous, l'excuse de faire le plein du Mössbauer le week-end était un bon prétexte pour chanter et parfois même esquisser un pas de danse au labo. C'était aussi la réunion des potins, des confidences de la semaine, un moment simple de partage entre joie, colère et quelques fois tristesse. J'ai beaucoup de respect pour toi et même si parfois j'étais agaçant, un peu borné ou vulgaire parce que « ça me casse les c\*\*\*\*\* », ta porte m'était toujours ouverte. Durant la thèse, j'ai eu des moments de doutes, de remise en question où j'avais peur de ne pas être à la hauteur ou pire, de décevoir Olivier. À chacun de ces moments, j'ai pu venir partager tout cela avec toi, trouver une oreille attentive et entendre des mots qui me rassuraient. On devait se botter mutuellement les fesses pour rentrer chez nous. On a souvent fermé la boutique tard et je suis persuadé que si le CEA ne fermait pas, nous serions encore tous les deux à tailler le bout de gras (ou une feuille de salade) tard dans la nuit. On s'est souvent pris la tête, quand il s'agissait de prendre des vacances. Un cercle vicieux où je refusais de les prendre, tu attendais que je parte pour en faire de même, j'insistais pour que tu prennes du repos, mais comme je restais, tu restais aussi. J'ai partagé avec toi ma « passion » pour les vieilles chansons : on est sur un registre de chansons entre 1940-1980 tout de même ! Nous avons connu aussi des bas, non pas dans notre relation, mais sur le plan personnel, mais on peut compter l'un sur l'autre. Oui Geneviève je vais parler de ce dimanche où le téléphone a retenti. Equipée de ma cape de super héros, je suis venu te chercher à Valence après une vilaine fracture de la cheville. Je sais qu'il t'est difficile de demander de l'aide ; toi et moi, on se débrouille tout seul. Mais ce fut un plaisir de t'aider. Je dois avouer que j'ai pris un malin plaisir à te balader dans les couloirs du CEA ou au

---

supermarché dans ton bolide, même s'il serait bien d'y rajouter une ceinture car, parfois, ça secoue dans les virages. Plus sérieusement, tu es la personne à qui je me suis confié sur beaucoup de choses parce que l'on a des valeurs communes et des expériences similaires. De nouvelles larmes me viennent quand je relis ton message où tu soulignes les qualités humaines que j'ai, mais ses qualités sont tout autant les tiennes.

Tu m'as toujours dit que c'était nous, les jeunes, qui faisons la bonne ambiance au laboratoire. Je ne suis pas tout à fait d'accord avec toi. Je pense que c'est vous, les permanents, qui par vos personnalités, instaurez un endroit où l'on se sent à l'aise, où l'on a plaisir à venir. On parle d'équipe, mais à mes yeux, nous sommes une petite famille.

Je voudrais à présent remercier Amélie, qui, dans cette famille, s'apparenterait à la grande sœur avec qui on rigole à coup sûr, mais qui, dans les moments sérieux, est de bon conseil. Je sais que tu n'es pas à l'aise avec les grands discours de remerciement alors je serai bref. Chaque journée commençait par un rayon de soleil que l'on s'offrait mutuellement. Ce rayon pour toi était de me voir débarquer pour te dire qu'il y avait du café et pour moi c'était de voir la joie que cette nouvelle t'apportait. J'ai particulièrement apprécié ce petit brin de folie qui dès le matin au café nous faisait commencer la journée par des sourires. Quelques-unes de tes punchlines ou allusions m'ont mis dans de beaux draps. Je ne savais plus où me mettre, mais ça me faisait rire. Bien que je ne sois pas venu souvent t'embêter dans ton bureau, j'ai ressenti cette facilité qu'il y a venir te parler. Cela qui m'a permis d'être rassuré et aiguillé lors de la longue et difficile période de rédaction : « tu réfléchis trop, faut pas se prendre la tête pour ça ». Je te souhaite beaucoup de réussites pour la suite, au laboratoire et surtout en dehors.

Cette partie-là, je la dédie à mes collègues et amis de travail. Je remercie Guillaume Frémy pour son accueil en M1 et qui, avec ses amis Guillaume et Jordan, m'ont donné envie de faire une thèse. J'ai été ravi de te retrouver et partager tout ton savoir pour reprendre ton travail et aller vers cet objectif commun de cibler l'ADN avec un système FRET. Avec Olivier, vous m'aviez promis de la couleur avec les lanthanides, même si en stage c'était 50 nuances de noir toujours plus noires. Donc je suis revenu en thèse pour enfin voir toutes ces couleurs. J'ai pris la suite de ton projet et j'espère l'avoir amené au bout, ou du moins presque au bout. À mon retour en thèse, j'ai fait la connaissance de Ji-Hyung Choi, doctorante arrivée quelques mois auparavant. C'est avec Patrick Malikidogo l'une des plus belles rencontres de ma thèse. Passionnée et investie tout autant que moi, on a travaillé intensément durant ces trois années. Pas toujours d'accord sur les musiques que je mettais dans le labo, mais on trouvait un terrain d'entente, sur les fins de journée, passé 18h, avec ta techno qui était parfaite pour venir à bout de la journée. J'admirais et enviais cette capacité de compréhension et assimilation que tu as : Jiji, t'es trop forte ! Échanger avec toi sur la spectro des lanthanides ou la pénétration cellulaire était très enrichissant, mais aussi sur les sujets d'actualité, bien que je ne sois que peu au courant. Tes points forts étaient mes points faibles et inversement. Ça a été l'objet de longues discussions où l'on n'était pas d'accord : lequel

---

de nous deux avait les meilleures compétences pour réaliser une thèse de qualité ? Par contre quand il s'agissait de faire comprendre à Olivier qu'il était parfois un peu trop optimiste sur les résultats de luminescence, on était sur la même longueur d'onde (je te laisse apprécier le jeu de mots). J'ai adoré être l'arbitre de votre différent pour savoir si le carbazole est cytotoxique ou pas, un peu comme dans Astérix et Obélix pour savoir qui a le poisson le plus frais. Merci pour tous ces moments. Bonne chance pour la suite, même si je sais que tu vas gérer, quoi que tu fasses. Vient le tour de Monsieur Patrick Malikidogo, ou pour les plus intimes, Monsieur le Président ou Grand Maître. Cette joie de vivre qui t'anime est contagieuse, à tel point que je suis prêt à prendre un ministère quand tu seras Président du Congo. Ton parcours de vie et tes valeurs sont exemplaires et inspirants. Tu es parti pour Toulouse dans les débuts de ma thèse, mais on a eu le temps de très vite sympathiser puisque que tu as vu en moi le zigouilleur que je suis (attention pas d'amalgame). Tu m'as dit : « Je serai là pour ta thèse ». Et je ne sais pas comment tu as zigouillé, mais tu as réussi à revenir zigouiller dans l'équipe pour m'aider à avancer le projet et être au premier rang pour ma soutenance. Notre dernier congrès où l'on a zigouillé une chambre ensemble restera dans les mémoires. Je nous revois sur la plage en train de zigouiller avec nos nouveaux amis au lever du soleil. J'ai appris, grâce à toi, à voir la vie différemment, prendre conscience des choses simples, comme chanter du Whitney Houston à s'en casser la voix dans le labo ou quand on zoukait à nous entendre à l'autre bout du bâtiment. Tu as de nombreuses qualités que tu transmets maintenant à ton fils. Je vous souhaite à vous trois une vie de famille des plus heureuses et de la réussite pour ta conquête du pouvoir. J'ai aussi vu de nombreux p'tits jeunes au labo avec qui j'ai beaucoup rigolé et j'ai mis un point d'honneur à les intégrer dans l'équipe. Je les ai aidés du mieux que je pouvais, même si je les taquinais souvent. Je suis ravi de voir que chacun a réussi à trouver sa voie. Ces remerciements sont pour Junio, Adam, Tom et Éloïse, que j'ai eu le plaisir d'encadrer durant leur stage. Qui aime bien châtie bien ! Cette expression résume bien les relations que j'ai eues avec vous et même si je ne suis plus au labo, j'arrive encore à trouver les moyens de venir vous embêter (grâce à l'aide d'Olivier). Il y eut aussi Alicia, Guillaume, Morgan, Klara et Gaëtan qui ont fait partie de l'équipe. Des personnalités différentes avec qui j'ai pu longuement échanger et, malheureusement pour Guillaume et Morgan, convertir à Dofus (le confinement était plus simple à vivre grâce à vous). Pour clôturer cette partie, je vais remercier Noëlle pour sa gaité de cœur. Chaque jour, elle venait toute souriante et étincelante me voir pour rigoler avec Geneviève. Elle forme avec Geneviève un formidable duo capable dans toutes les circonstances de vous redonner le sourire. Je pense que nos fous rires résonnent encore dans les couloirs du C5. Je n'oublie pas ta promesse de me trouver une femme, sinon je vais finir à « Mariés au premier regard ». J'espère de tout cœur que ta santé va s'améliorer pour que tu brilles à nouveau de mille feux.

Il y a un dicton qui dit qu'on ne choisit pas sa famille, mais ses amis. Que se passe-t-il quand ses amis deviennent une famille ? Avec Richard, Baptiste, Damien et Romain, on a réussi depuis les bancs de la fac de chimie à aller jusqu'à la thèse. Bon avec un petit contre temps puisqu'après une sortie de trois

---

jours au Morétan pour fêter nos admissions en thèse, la mienne s'était envolée en revenant dans la vallée. On a vécu ces trois années difficiles ensemble. J'ai l'impression d'avoir vécu vos thèses autant que la mienne parce qu'on a partagé tous les hauts et les bas qui rythment une thèse. C'est sans doute pour ça que, pour chacune de vos soutenances, j'ai eu ma p'tite larme de joie de vous voir arriver au bout du voyage (bon d'accord je suis un peu sensible aussi). La thèse de Clarisse m'a rappelé que même si on est seul sur sa thèse c'est ensemble qu'on a tous réussi, entouré de ses amis pour se soutenir. L'épisode Covid nous a finalement encore plus soudés. C'est grâce au laissez-passer de Clarisse que les mercredis soirs au 22 ont pris de l'ampleur. C'était des moments très conviviaux et parfois des séances de doléances pour partager nos fardeaux et trouver du soutien. On ne compte plus le nombre d'histoires farfelues que l'on a eues lors de nos escapades/aventures pêche, un moyen de souffler en n'ayant qu'une seule chose en tête : « POISSON ! ». Je suis le p'tit dernier de la bande à finir sa thèse, même si j'en suis le doyen. Bien que la BaRiMou team soit en route pour la création d'une start-up (on croise les doigts), j'aurais aimé que vous ayiez une fin de thèse moins difficile, avec une soutenance ouverte au public pour Baptiste et des remerciements mérités, à la hauteur de votre investissement. Je pense que les liens que nous avons tissés toutes ces années et les souvenirs de vacances avec Clarisse, mais aussi Cathie, Justine, Louison et Pascaline ont créé un lien qui nous gardera tous unis, chacun de nous trouvera son bonheur que ce soit ici ou ailleurs.



---

## Table des matières

<b>Abréviations</b> .....	<b>1</b>
<b>Chapitre I - Introduction</b> .....	<b>5</b>
<b>I - Contexte</b> .....	<b>7</b>
<b>II - Reconnaissance d'ADN</b> .....	<b>8</b>
1 - Structure de l'ADN .....	8
2 - Principales méthodes de reconnaissance de l'ADN .....	10
2.1 - Reconnaissance de l'ADN par des oligonucléotides .....	11
2.1.a    Molecular Beacon .....	11
2.1.b    PNA .....	13
2.1.c    OCN .....	15
2.2 - Reconnaissance de l'ADN par des polymères .....	16
2.3 - Reconnaissance de l'ADN par des protéines .....	22
2.3.a    Protéines TALE .....	22
2.3.b    Doigts de zinc .....	23
2.3.c    Protéines ZFN .....	26
2.3.d    Protéine Zif12 .....	27
<b>III - La détection par luminescence</b> .....	<b>31</b>
1 - La luminescence .....	31
2 - Transfert d'énergie .....	34
2.1 - Définition et mécanisme .....	34
2.2 - Un exemple de système FRET entre fluorophores organiques .....	37
<b>IV - Les lanthanides</b> .....	<b>38</b>
1 - Définition .....	38
1.1 - Généralités .....	38
1.2 - Configuration électronique .....	39
1.3 - Coordination des lanthanides .....	40
2 - La luminescence des lanthanides .....	42
2.1 - Emission des lanthanides .....	42
2.2 - Effet d'antenne .....	42
3 - Complexes luminescents de lanthanides pour des applications en biologie .....	43
4 - Des exemples de sondes à base de lanthanides .....	46
4.1 - Sondes reposant sur du FRET à l'aide de complexes de lanthanides .....	46
4.2 - Détection d'oligonucléotides à l'aide de complexes de lanthanides .....	48
4.3 - Détection d'ADN simple brin .....	50
<b>V - Le projet de thèse</b> .....	<b>51</b>

1 - Contexte .....	51
2 - Travaux précédents .....	52
2.1 - Stratégie de synthèse des peptides fonctionnalisés .....	52
2.2 - La preuve de concept .....	53
3 - Les objectifs du projet de thèse .....	55
<b>Chapitre II - Synthèse par voie chimique de protéines Zif12 non fonctionnalisées pour la reconnaissance d'ADN non palindromique .....</b>	<b>57</b>
Préambule - Design de la protéine Zif12-B .....	59
I - Synthèse par voie chimique de Zif12 .....	60
1 - Ligation chimique native .....	60
1.1 - Ligation chimique native .....	62
1.1.a Ligation SEA .....	63
1.1.b Ligation hydrazide .....	64
1.2 - Segmentation des protéines Zif12-A et Zif12-B .....	65
1.3 - Stratégie d'assemblage d'une protéine par fragments .....	67
1.4 - Stratégie d'assemblage de la protéine Zif12 .....	69
2 - Mise au point de l'assemblage de la partie reconnaissance d'ADN (F2+F3) .....	71
2.1 - Ligation des fragments Thz-F2 et F3 par ligation SEA, protection Thz en N-ter .....	72
2.1.a Synthèse des fragments <i>Thz-F2-SEA<sup>off</sup></i> et <i>H-F3-NH<sub>2</sub></i> .....	72
2.1.b Ligation des fragments <i>Thz-F2-SEA<sup>off</sup></i> et <i>H-F3-NH<sub>2</sub></i> .....	73
2.2 - Ligation des fragments F2 et F3 par ligation hydrazide, protection N <sub>3</sub> en N-ter .....	74
2.2.a Synthèse de <i>N<sub>3</sub>-F2-NHNH<sub>2</sub></i> .....	74
2.2.b Ligation avec le fragment <i>H-F3-NH<sub>2</sub></i> .....	75
2.3 - Ligation des fragments F2 et F3 par ligation SEA, protection Acn en N-ter .....	76
2.4 - Ligation du fragment F2 et F3 par ligation hydrazide, protection Acn en N-ter .....	78
3 - Assemblage de la partie reconnaissance d'ADN avec l'extension N-terminale .....	79
3.1 - Synthèse du fragment F1-SEA .....	79
3.2 - Synthèse du fragment H-F1-Nle-SEA .....	80
3.3 - Assemblage de l'extension N-terminale et du domaine de reconnaissance .....	82
4 - Formation des doigts de zinc .....	84
II - Étude des interactions des protéines Zif12-A et Zif12-B avec l'ADN .....	84
1 - La technique EMSA .....	85
1.1 - Principe de la technique EMSA .....	85
1.2 - Interaction des protéines avec le duplex d'ADN .....	86
2 - Rappel des travaux de Wang et Pabo .....	87
3 - Interaction des protéines Zif avec les séquences palindromiques d'ADN .....	88

<b>Chapitre III - Fonctionnalisation des protéines Zif12.....</b>	<b>93</b>
I - Choix des chromophores .....	95
II - Fonctionnalisation du fragment F1 par les chromophores .....	98
1 - Couplage des chromophores sur le fragment F1 .....	98
2 - Ligation des fragments fonctionnalisés par les chromophores avec le fragment F2-F3 .....	100
III - Fonctionnalisation de la protéine par le complexe d'euporium [ $L_I$ -Eu] .....	102
1 - Synthèse du pro-ligand $PL_I^{CC}$ .....	102
1.1 - Description du ligand $L_1^{CC}$ .....	102
1.2 - Synthèse du macrocycle iodé 6.....	104
1.3 - Synthèse de la partie électrodonneuse de l'antenne de $PL_1^{CC}$ .....	105
1.4 - Synthèse du pro-ligand $PL_1^{CC}$ .....	106
2 - Synthèse du fragment $Ac$ -F1- $[L_I^{CC}$ -Eu]- $SEA$ .....	107
3 - Synthèse de la protéine Zif12- $[L_I^{CC}$ -Eu]-A.....	110
IV - Synthèse de complexes avec des nouvelles antennes .....	111
1 - Nouvelle voie de synthèse.....	111
2 - Antennes dérivées de l'anisole.....	113
2.1 - Synthèse du pro-ligand $PL_2$ .....	113
2.2 - Caractérisation spectroscopique du complexe TAT- $[L_2$ -Eu] .....	114
3 - Antennes dérivées du carbazole.....	117
3.1 - Synthèse des pro-ligands $PL_3$ et $PL_4$ .....	118
3.2 - Caractéristiques spectroscopiques des complexes TAT- $[L_3$ -Eu] et TAT- $[L_4$ -Eu].....	119
4 - Antennes dérivées de l'aniline .....	123
4.1 - Synthèse des pro-ligands $PL_5$ à $PL_{10}$ .....	126
4.2 - Synthèse et caractérisation spectroscopique des complexes $[L_i$ -Eu] ( $i=5-10$ ).....	128
V - Synthèse des protéines fonctionnalisées par les complexes de lanthanide.....	132
<b>Chapitre IV - Interaction des protéines Zif12 fonctionnalisées avec les duplex d'ADN.....</b>	<b>137</b>
I - Métallation de protéines Zif12 décorées de luminophores et caractérisations spectroscopiques	139
1 - Protéines Zif12*-( $SiR^{Me}$ )-B et Zif12*-(Lissa)-B .....	139
2 - Protéines Zif12*-[ $L_2$ -Tb]-A, Zif12*-[ $L_2$ -Eu]-A et Zif12*-[ $L_{10}$ -Eu]-A.....	141
2.1 - Protéines Zif12*-[ $L_2$ -Tb]-A.....	141
2.2 - Protéines Zif12*-[ $L_2$ -Eu]-A et Zif12*-[ $L_{10}$ -Eu]-A .....	143
II - Etude de l'interaction de Zif12*-[ $L_2$ -Eu]-A / Zif12*-[ $L_{10}$ -Eu]-A et Zif12*-( $SiR^{Me}$ )-B avec l'ADN	144
1 - Expériences préliminaires .....	144
2 - Stabilité des protéines .....	148
3 - Etude de la sélectivité avec le couple Zif12*-[ $L_2$ -Eu]-A / Zif12*-( $SiR^{Me}$ )-B .....	153

---

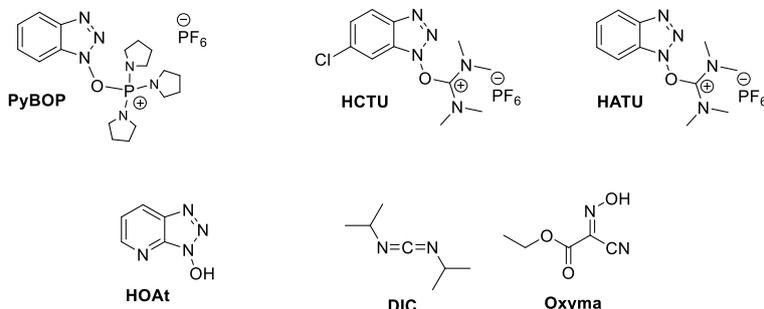
4 - Etude du couple Zif12*-[L <sub>10</sub> -Eu]-A et Zif12*-(SiR <sup>Me</sup> )-B .....	155
5 - Etude du couple Zif12*-[L <sub>10</sub> -Eu]-A et Zif12*-(Lissa)-B .....	159
III - Conclusion .....	160
<b>Conclusion et perspectives .....</b>	<b>163</b>
Conclusion.....	165
Perspectives .....	167
Fonctionnalisation des protéines .....	167
Le complexe d'euporium.....	169
Interaction des protéines décorées avec l'ADN .....	169
Interaction protéines / ADN .....	169
Transfert FRET.....	170
<b>Partie Expérimentale .....</b>	<b>173</b>
Organic Synthesis.....	178
Metallation of Ln <sup>3+</sup> after tert-butyl déprotection of PL <sub>i</sub> .....	201
Coupling of PL <sub>i</sub> on TAT.....	201
Peptidic synthesis .....	202
Peptidic sequences on resin.....	202
Peptides .....	202
Fragment F1, F2 and F3 .....	202
Peptide F2-F3 .....	203
Peptide F1-F2-F3.....	203
General procedure: .....	204
Peptide characterisation.....	207
Fragment F1, F2 and F3 .....	207
Peptide F2-F3 .....	210
Peptide F1-F2-F3.....	212
<b>Bibliographie.....</b>	<b>215</b>

## Abréviations

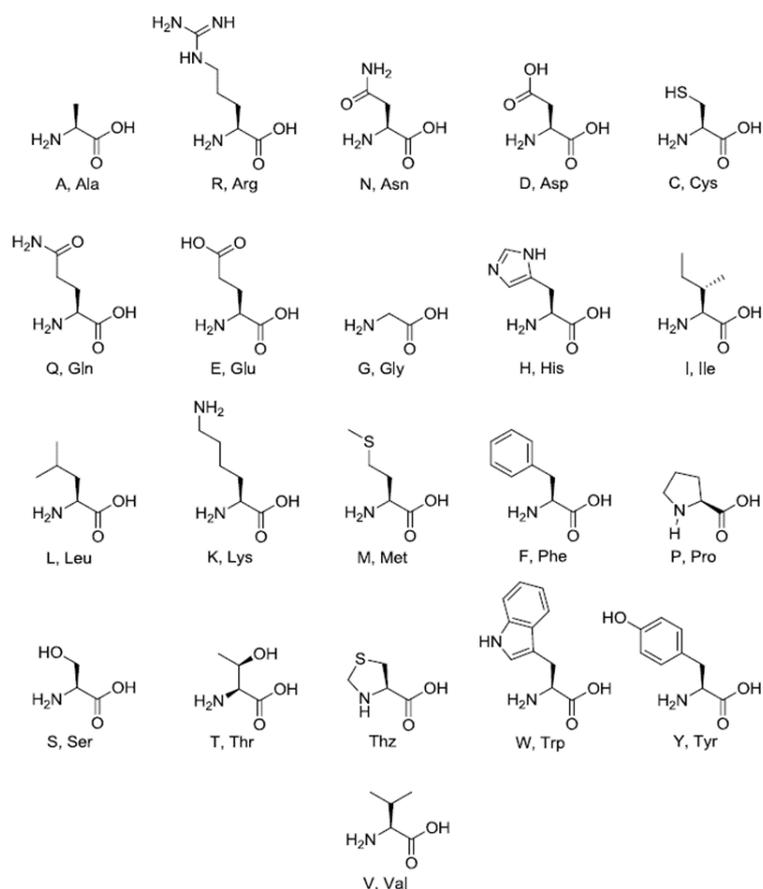
<b>A</b> : Adénine	<b>G</b> : Guanine
<b>Å</b> : Angström ( $10^{-10}$ m)	<b>Gn.HCl</b> : Guanidine hydrochloride
<b>aa</b> : Acide aminé	<b>GP</b> : Groupement Protecteur
<b>AcOEt</b> : Acétate d'éthyle	<b>HEPES</b> : Acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique
<b>AcM</b> : Acétamidométhyl	<b>HNB</b> : 2-hydroxy-5-nitrobenzyl
<b>ADN</b> : Acide Désoxyribonucléique	<b>Hp</b> : <i>N</i> -methyl-3-hydroxypyrrole
<b>Ar</b> : Aryle	<b>HPLC</b> : High Pressure Liquid Chromatography / Chromatographie liquide haute pression
<b>ARN</b> : Acide Ribonucléique	<b>HRMS</b> : High Resolution Mass Spectrometry / Spectroscopie de masse à haute résolution
<b>ARNm</b> : Acide Ribonucléique messenger	<b>ILCT</b> : Transfert de Charge Intra-Ligand
<b>ATP</b> : Adénosine-Triphosphate	<b>Im</b> : Imidazole
<b>B</b> : Brilliance	<b>IRM</b> : Imagerie par résonance magnétique
<b>Boc</b> : <i>tert</i> -butoxycarbonyl	<b>J</b> (Hz) : Constante de couplage
<b>bp</b> : paire de base	$\lambda$ : Longueur d'onde
<b>br</b> : broad	<b>LCMS</b> : Liquid Chromatography/Mass Spectrometry
<b>BSA</b> : Bovin serum albumin	<b>LE</b> : Locally excited state
<b>C</b> : Cytosine	<b>LMCT</b> : Transfert de charge ligand-métal
<b>CFP</b> : Cyan Fluorescent Protein	<b>Ln</b> : Lanthanide
<b>CIS</b> : Croisement Inter-Système	<b>MALDI</b> : Matrix Assisted Laser Desorption Ionisation
<b>CT</b> : Transfert de charge	<b>MB</b> : Molecular Beacon
$\delta$ : Déplacement chimique RMN	<b>MLCT</b> : Transfert de charge métal-ligand
$\Delta M$ : Différence de masse	<b>MPA</b> : Acide 2-mercaptoacétique
<b>Da</b> : Dalton	<b>MPAA</b> : Acide mercaptopropionique
<b>DCM</b> : Dichlorométhane	<b>MS</b> : Mass Spectrometry / Spectroscopie de Masse
<b>DFT</b> : Density Functional Theory	<b>NCL</b> : Native Chemical Ligation / Ligation chimique native
<b>DIEA</b> : Diisopropyléthylamine	<b>NLS</b> : Nuclear Localisation Signal
<b>DMAc</b> : N,N-Diméthylacetamide	<b>OCN</b> : Oligodésoxyribonucléotide circulaire
<b>DMAP</b> : 4-Diméthylaminopyridine	$\Phi$ : Rendement quantique
<b>DMF</b> : Diméthylformamide	<b>PBS</b> : Phosphate buffer saline
<b>DMSO</b> : Diméthylsulfoxyde	<b>Pbf</b> : 2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl
<b>DO3A</b> : Acide 1,4,7,10-tétraazacyclododécane-1,4,7,10-triacétique	<b>PCR</b> : Polymerase Chain Reaction
<b>DOTA</b> : Acide 1,4,7,10-tétraazacyclododécane-1,4,7,10-tétraacétique	<b>PDT</b> : Photo Dynamic Therapy
<b>DTPA</b> : Acide diéthylène triamine pentaacétique	<b>PEG</b> : Polyéthylène glycol
<b>DTT</b> : Dithiothréitol	<b>pH</b> : Potentiel hydrogène
$\epsilon$ : coefficient d'absorption molaire	<b>PNA</b> : Peptide nucleic acid / Acide nucléique peptidique
<b>EDTA</b> : Ethylenediamine tetraacétique acid	<b>PPI</b> : Pyrole Imidazole Polyamide
<b>EGR-1</b> : Early Growth Response gene 1	
<b>EMSA</b> : Electrophoretic Mobility Shift Assay/Décalage de mobilité électrophorétique	
<b>ESI</b> : Electrospray Ionisation/ Ionisation par électronébuliseur	
<b>FRET</b> : Transfert d'énergie résonnant de Förster	

<b>ppm</b> : Partie par million	<b>TBAF</b> : Fluorure de N-tétrabutylammonium
<b>PS</b> : Photosensibilisateur	<b>TCEP</b> : <i>tris</i> (2-carboxyéthyl)phosphine
<b>Py</b> : Pyrole	<b>TEA</b> : triéthylamine
<b>R<sub>0</sub></b> : Rayon de Förster	<b>TEMED</b> : Tétraméthyléthylènediamine
<b>RCA</b> : Rolling-circle amplification	<b>TFA</b> : Acide trifluoroacétique
<b>RMN</b> : Résonance Magnétique Nucléaire	<b>THF</b> : Tétrahydrofurane
<b>r.t</b> : Room temperature / Température ambiante	<b>TFO</b> : Triplex Forming Oligonucleotide / Oligonucléotide pour la formation de triplex
<b>ROS</b> : Espèce réactive de l'oxygène (Reactive Oxygen Species)	<b>TIS</b> : Triisopropylsilane
<b>RX</b> : Rayon X	<b>Tris</b> : <i>Tris</i> (hydroxyméthyl)aminométhane
<b>SEA</b> : <i>bis</i> (2-sulfanyléthyl)amido	<b>TICT</b> : Twisted Intramolecular Charge Transfer
<b>SPPS</b> : Solid Phase Peptide Synthesis / Synthèse peptidique sur support solide	<b>TMR</b> : Tetramethylrhodamine
<b>T</b> : Thymin	<b>TO</b> : Thiazole Orange
<b>TACN</b> : 1,4,7, Triazacyclononane	<b>U</b> : Uracile
<b>TALE</b> : Transcription activator-like effector / Effecteur TAL	<b>UV</b> : Ultra-Violet
<b>TALLEN</b> : Transcription activator-like effector nuclease / Nucléase effectrice de type activateur de transcription	<b>YFP</b> : Yellow Fluorescent Protein
	<b>ZFN</b> : Zinc Finger Nuclease

## Activateurs



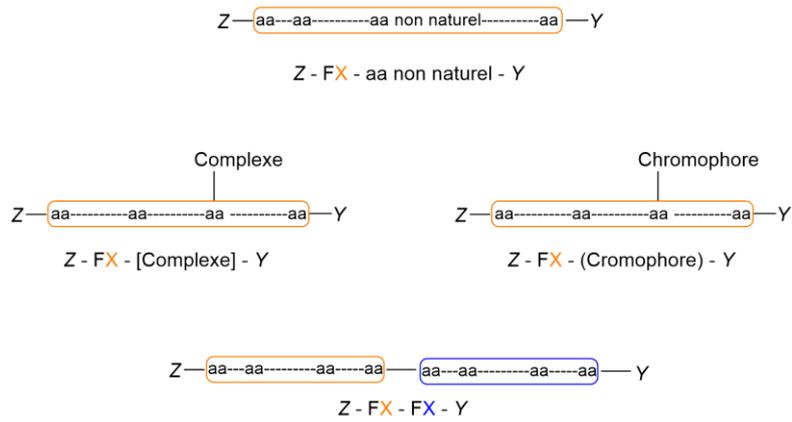
## Abréviation des acides aminés



## Notation des peptides

Les peptides représentés dans ce manuscrit possèdent la notation suivante :

La partie N-terminale (notée *Z* dans **Figure 1**) est d'abord indiquée en italique (si elle n'est pas composée d'un groupement  $\text{NH}_2$  classique, auquel cas elle est noté *H*), séparée par un tiret du nom du fragment (FX dans l'exemple avec  $X = 1, 2$  ou  $3$ ). Si un acide aminé non naturel est présent dans le peptide, il est mentionné entre deux tirets. La fonctionnalisation par un complexe de lanthanide sera indiquée entre crochets et la fonctionnalisation par un chromophore organique entre parenthèses. Finalement la partie C-terminale (*Y* dans l'exemple) est indiquée en italique, séparée par un tiret. Lorsque deux fragments sont reliés ensemble les noms des deux fragments sont notés. La notation est détaillée **Figure 1**.



**Figure 1** : Exemple de notation d'un fragment selon s'il possède un aa non naturel, un complexe/chromophore organique ou relié à un autre fragment.

### Composition des tampons

Tampon A : solution aqueuse à 0.1 % de TFA.

Tampon B : solution à 0.1 % de TFA dans l'acétonitrile.

Tampon C : HEPES (150 mM, pH 7.5), Tween (0.5%), NaCl (1 M), BSA (1 g.L<sup>-1</sup>)

Tampon D : HEPES (150 mM, pH 7.5), Tween (0.5%), NaCl (1 M)

# Chapitre I - Introduction

## Sommaire

I - Contexte.....	7
II - Reconnaissance d'ADN .....	8
1 - Structure de l'ADN .....	8
2 - Principales méthodes de reconnaissance de l'ADN .....	10
2.1 - Reconnaissance de l'ADN par des oligonucléotides.....	11
2.1.a    Molecular Beacon.....	11
2.1.b    PNA .....	13
2.1.c    TFO/OCN .....	15
2.2 - Reconnaissance de l'ADN par des polymères .....	16
2.3 - Reconnaissance de l'ADN par des protéines .....	22
2.3.a    Protéines TALE.....	22
2.3.b    Doigts de zinc .....	23
2.3.c    Protéines ZFN.....	26
2.3.d    Protéine Zif12.....	27
III - La détection par luminescence .....	31
1 - La luminescence.....	31
2 - Transfert d'énergie.....	34
2.1 - Définition et mécanisme .....	34
2.2 - Un exemple de système FRET entre fluorophores organiques .....	37
IV - Les lanthanides .....	38
1 - Définition .....	38
1.1 - Généralités .....	38
1.2 - Configuration électronique.....	39
1.3 - Coordination des lanthanides .....	40
2 - La luminescence des lanthanides .....	42
2.1 - Emission des lanthanides .....	42
2.2 - Effet d'antenne .....	42
3 - Complexes luminescents de lanthanides pour des applications en biologie .....	43
4 - Des exemples de sondes à base de lanthanides .....	46
4.1 - Sondes reposant sur du FRET à l'aide de complexes de lanthanides .....	46
4.2 - Détection d'oligonucléotides à l'aide de complexes de lanthanides .....	48
4.3 - Détection d'ADN simple brin .....	50
V - Le projet de thèse .....	51
1 - Contexte .....	51
2 - Travaux précédents .....	52
2.1 - Stratégie de synthèse des peptides fonctionnalisés .....	52
2.2 - La preuve de concept .....	53
3 - Les objectifs du projet de thèse.....	55

## I - Contexte

Un grand nombre de recherches, durant les 20 dernières années, se sont dirigées vers le développement d'outils capables de reconnaître une séquence précise d'ADN. Ces outils sont essentiels dans la compréhension de processus biologiques, mais aussi d'un point de vue thérapeutique. Une mutation au sein d'une séquence précise d'un gène peut avoir pour conséquence un dysfonctionnement du métabolisme. Avoir des outils capables de détecter ces mutations pourrait permettre de commencer un traitement avant l'apparition des premiers symptômes. L'un des outils génétiques les plus connus et les plus prometteurs est une enzyme bactérienne permettant de faire une coupure du double brin d'ADN avec deux zones de coupe actives, une pour chaque brin de la double hélice. La découverte de cette enzyme nommée CRISPR-Cas9 et de son mécanisme d'action valut le prix Nobel de Chimie en 2020 aux professeurs Emmanuelle Charpentier et Jennifer Doudna. Ce prix Nobel symbolise une avancée majeure dans le domaine de la thérapie génique et reflète l'importance du développement de systèmes capables de cibler spécifiquement une séquence d'ADN. En 2013, le système CRISPR-Cas9 a été associé avec une protéine fluorescente EGFP pour l'étude de l'élongation des télomères dans des cellules vivantes. Ainsi les images obtenues grâce à la fluorescence émise ont permis de faire un pas en avant sur la compréhension de ce processus d'élongation<sup>1</sup>. Plus récemment, des chercheurs aux USA ont prélevé des lymphocytes T du système immunitaire des patients et les ont modifiés génétiquement à l'aide de CRISPR-Cas9 pour les aider à reconnaître et à combattre le cancer, avec des effets secondaires minimes et contrôlables. L'objectif était de supprimer trois gènes qui empêchent les traitements d'agir efficacement contre les cellules cancéreuses. L'ajout d'une nouvelle fonctionnalité visant à améliorer leur efficacité est aussi au cœur du projet<sup>2</sup>. Il s'agit là d'un projet très ambitieux déclarait le Dr Edward Stadtmauer, directeur de cette étude : « *C'est l'ingénierie cellulaire et génétique la plus compliquée qui ait été tentée jusqu'à présent* ».

Il existe cependant un autre courant de recherche pour le traitement de maladies, cancéreuses entre autres, c'est le développement de thérapie photo-dynamique (PDT). C'est une méthode peu invasive qui permet la destruction des cellules cancéreuses par irradiation d'un photosensibilisateur (PS). Après irradiation, le PS va générer des espèces réactives de l'oxygène, connues sous l'abréviation ROS (Reactive Oxygen Species), qui vont détruire les cellules ciblées. Cependant, pour que cette méthode soit peu invasive il faut avoir un système capable d'être très spécifique de la cible. Auquel cas le PS utilisé pourra s'accumuler dans des tumeurs, mais aussi dans des tissus sains, ce qui pendant l'irradiation peut causer une phototoxicité et des réactions diverses chez le patient. Dans le traitement de cancers, il est donc primordial de pouvoir distinguer les cellules tumorales des cellules saines. Il s'agit là de cibler des protéines exprimées par les cellules cancéreuses. Une alternative pourrait être le ciblage non plus du produit de l'expression du gène, mais le ciblage direct de la séquence d'ADN codant pour cette protéine. Ceci pourrait alors accroître la spécificité de reconnaissance des cellules cancéreuses. Un tel système de reconnaissance, couplé à un système luminescent générateur de ROS

pourrait être un premier pas vers une nouvelle technique de PDT basée sur la reconnaissance d'une séquence spécifique d'ADN. Le génome humain contient environ 3 milliards de paires de bases et le ciblage spécifique d'une séquence est statistiquement possible si le système en question est capable de reconnaître une séquence longue de 17 paires de bases <sup>3</sup>. Bien que ce soit l'objet de nombreux projets de recherche, cela reste néanmoins un grand challenge.

## II - Reconnaissance d'ADN

L'acide désoxyribonucléique (ADN) est une macromolécule qui est omniprésente chez tous les êtres vivants, mais aussi chez un grand nombre de virus. Il contient l'information génétique aussi appelée génome (ensemble des gènes). Cette information permet à tout organisme de se développer, se reproduire. L'ADN est situé dans le cytoplasme des cellules procaryotes (bactéries) et dans le noyau des cellules eucaryotes. Cet ADN peut adopter plusieurs structures, primaires et secondaires.

### 1 - Structure de l'ADN

L'ADN est un polymère composé de trois briques élémentaires. Le point de départ est une base nucléique qui est reliée à un ose (sucre), lui-même relié à un groupement phosphate, le tout formant un nucléotide (**Figure 2**). Plus précisément, il existe 4 bases nucléiques, l'adénine (A), la guanine (G), la cytosine (C) et la thymine (T). Deux d'entre elles, l'adénine et la guanine sont des bases dites puriques. La cytosine et la thymine sont de nature pyrimidique. Concernant l'ose, c'est un ose à 5 carbones, le 2-désoxy-D-ribose. Il va être lié en position 1' à la base nucléique et au groupement phosphate *via* l'oxygène en position 5' et 3'.

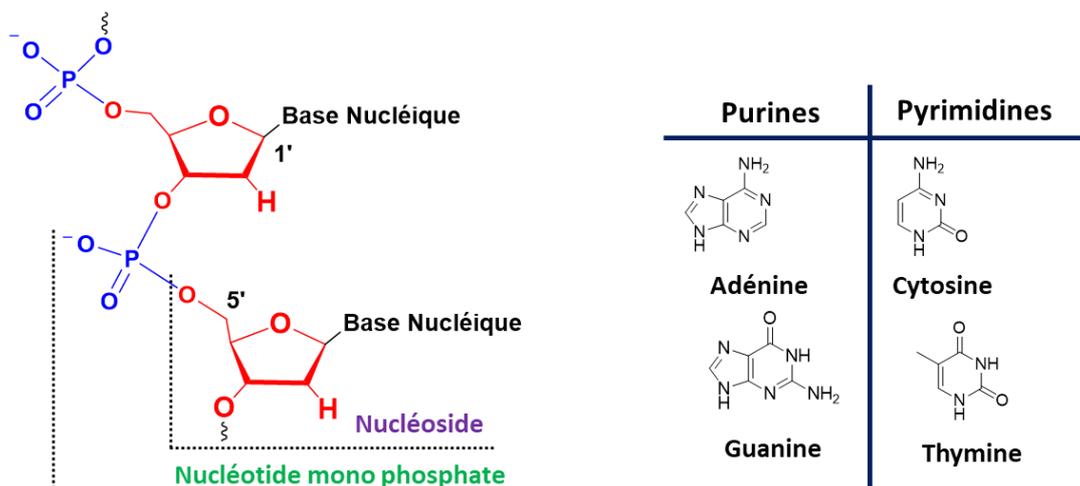
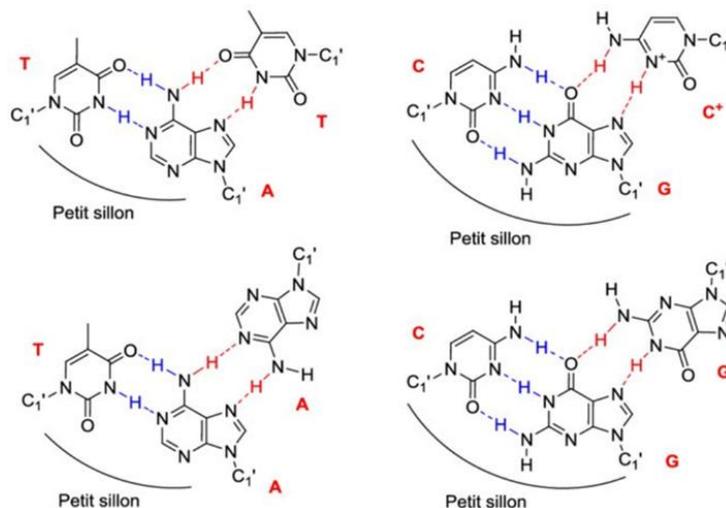


Figure 2 : Présentation de la structure primaire de l'ADN et des bases nucléiques.

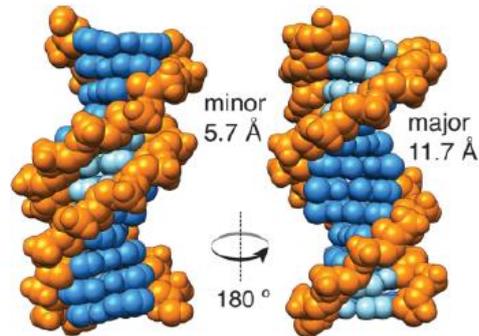
La polymérisation de nucléotides par des liaisons phosphate-ose et l'organisation en double hélice de deux brins complémentaires conduit à la formation d'un brin d'ADN. En 1953, Watson et Crick ont montré que l'ADN possède une structure secondaire sous la forme d'une hélice composée de deux brins

antiparallèles enroulés l'un autour de l'autre <sup>4</sup>. La formation de cette hélice, ou duplex, est possible grâce à des interactions de type liaison hydrogène (liaison H) entre les bases nucléiques. On distingue deux types de liaison hydrogène ; les liaisons de Watson et Crick qui sont établies entre deux bases azotées orientés *syn* (dans le petit sillon de l'ADN) et les liaisons Hoogsteen entre bases d'orientation opposées (*syn/anti*) qui sont présentes dans le grand sillon (**Figure 3**). Il existe plusieurs formes pour le duplex d'ADN, mais la plus répandue est la forme B, noté ADN B <sup>5</sup>, où les paires de bases sont toutes orientées *anti*. Les liaisons sont donc uniquement du type Watson et Crick. Les liaisons Hydrogène s'établissent entre deux bases se faisant face au sein du duplex. Il existe deux liaisons H entre l'adénine et la thymine, et trois liaisons H entre la guanine et la cytosine. C'est à la suite de ces interactions que l'on définit des paires de bases, la paire A-T et la paire C-G (et inversement). Cependant, une autre interaction permettant de stabiliser la formation du duplex vient en complément. Il s'agit d'interactions de type hydrophobe entre les cycles aromatiques des bases azotées au sein d'un même brin d'ADN, on parle de  $\pi$ -stacking <sup>6</sup>. La nomenclature d'écriture d'un brin d'ADN repose sur la notation des positions libres des sucres aux deux extrémités 5' ou 3' ainsi que les bases nucléiques présentes dans la séquence, exemple : 5'-ACTG-3'.



**Figure 3 :** Structure secondaire de l'ADN. Les liaisons bleues représentent les liaisons Watson et Crick et les liaisons en rouge représentent les liaisons Hoogsteen.

L'ADN B sous la forme d'un double brin possède plusieurs structures selon les conditions physiologiques (pH, température...). Sous cette forme, on peut voir que les deux brins s'enroulent en hélice autour d'un axe. La disposition des paires de bases par rapport à cet axe conduit à l'apparition d'un petit sillon et d'un grand sillon (**Figure 4**). Ce sont ces espaces qui vont permettre à des protéines ou divers ligands d'interagir avec l'ADN.



**Figure 4** : Structure 3D de l'ADN B laissant percevoir le petit sillon ainsi que le grand sillon <sup>7</sup>. En orange, le squelette sucre-phosphate, en bleu foncé les paires de bases exposées dans le grand sillon et en bleu clair les paires de bases exposées dans le petit sillon.

Beaucoup d'efforts sont effectués en biologie et dans le domaine du médical pour développer des outils permettant de reconnaître une séquence spécifique de l'ADN. Cela peut être dans le but d'isoler un gène afin d'en comprendre le fonctionnement ou tout simplement de détecter une séquence d'intérêt biologique. D'autres systèmes cherchent à aller plus loin en cherchant à générer des dégradations au niveau de l'ADN, mais uniquement sur une partie bien précise de l'ADN. D'où la nécessité de pouvoir reconnaître une séquence bien spécifique.

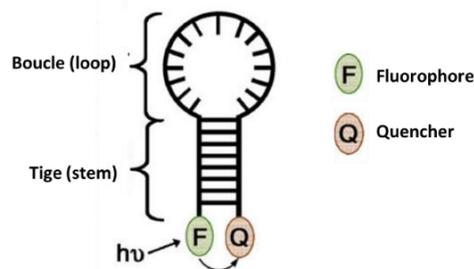
## 2 - Principales méthodes de reconnaissance de l'ADN

Il existe deux types de sondes en biologie, les sondes par affinité et par réaction. Comme son nom l'indique, les sondes par réaction émettent un signal uniquement lors d'une réaction entre la sonde et la cible aboutissant à une liaison covalente, non réversible. Par opposition, les sondes d'affinités viennent former des liaisons réversibles avec la cible, par exemple des liaisons H, Van der Waals, électrostatique. Pouvoir cibler de manière spécifique une séquence d'ADN reste un challenge. Il y a deux grandes stratégies pour effectuer cette reconnaissance : (i) reconnaître une suite de bases en interagissant avec une séquence simple brin de l'ADN cible, (ii) reconnaître une suite de paires de bases sur le double brin d'ADN. Dans le premier cas, il faut pouvoir ouvrir la double hélice afin que la sonde puisse par complémentarité se lier à l'un des deux brins. Dans le second cas, on va chercher à interagir au niveau du petit ou du grand sillon de l'ADN en formant un complexe avec la double hélice. En vue d'une application, il faut pouvoir rendre ce complexe « visible » et donc avoir un système qui permet de révéler cette interaction. La technique la plus répandue est la fluorescence de composés organiques ou de protéines fluorescentes. Cette technique présente une grande sensibilité, mais surtout requiert des équipements peu coûteux, un simple spectrofluoromètre étant généralement suffisant.

## 2.1 - Reconnaissance de l'ADN par des oligonucléotides

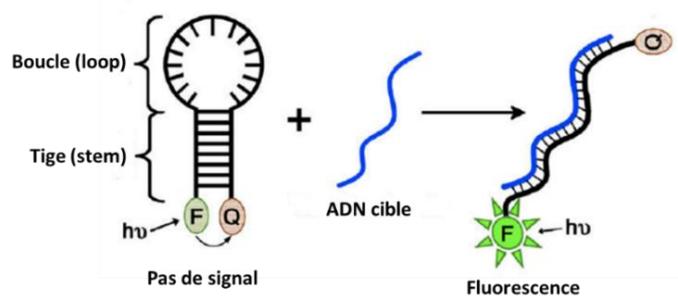
### 2.1.a Molecular Beacon

Les Molecular Beacons (MB) ont été décrits pour la première fois en 1996 par l'équipe de Kramer<sup>8</sup>. Le principe repose sur un unique oligonucléotide qui possède une structure particulière avec une tige en double brin (stem) et une boucle en simple brin (loop) (**Figure 5**). La tige est composée de 5 à 8 nucléotides présents à chaque extrémité du MB. Les séquences des deux extrémités sont complémentaires pour qu'elles puissent s'apparier. La boucle contient 15 à 30 nucléotides et sa forme correspond à une épingle à cheveux. C'est là que l'on retrouve la partie reconnaissance d'ADN *via* une séquence spécifique d'acides nucléiques. Après ouverture du double brin d'ADN cible, cette séquence spécifique va pouvoir former un duplex avec un simple brin d'ADN, on parle d'hybridation. La détection est assurée au moyen d'un fluorophore greffé au niveau de l'extrémité 5' du MB et d'un chromophore en 3' chargé d'éteindre le rayonnement émis par le fluorophore (quencher) (**Figure 5**).



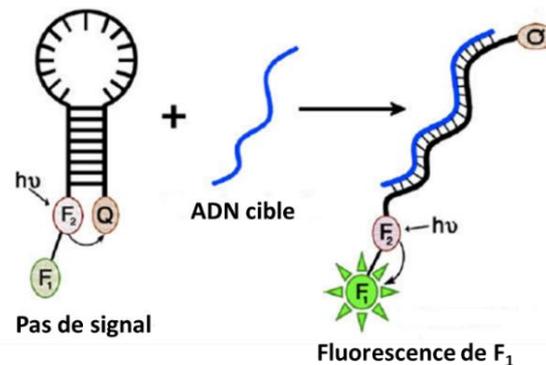
**Figure 5 :** Structure d'un Molecular Beacon. Figure adaptée de la publication <sup>9</sup>.

Le MB en absence de la séquence cible d'ADN n'est pas luminescent. En effet, lorsqu'il a cette structure boucle-tige, le quencher est positionné très proche de l'entité fluorescente. Après excitation du fluorophore, celui-ci est désexcité par un transfert d'énergie vers le quencher. Ainsi la sonde est « éteinte ». En présence de la séquence cible d'ADN, un changement de conformation a lieu : l'hybridation de la séquence d'acides nucléiques de la boucle avec sa cible induit la séparation du double brin de la tige. Ainsi les deux extrémités du MB s'éloignent, séparant ainsi l'entité fluorescente et son quencher, ce qui conduit à une restauration de la fluorescence : le MB est « allumé » <sup>9 10</sup> (**Figure 6**).



**Figure 6 :** Principe général du Molecular Beacon. En absence de la cible il n'y a aucune fluorescence, en présence de la cible un changement de conformation a lieu permettant d'éloigner le fluorophore du quencher. Figure adaptée de la publication <sup>9</sup>.

Une notion très importante dans l'élaboration d'un MB est la notion de constante d'affinité. Initialement, la structure boucle-tige est rendue possible par l'appariement des extrémités du brin. Le changement de conformation provient lui aussi de l'hybridation de la séquence d'acides nucléiques avec la cible. Il y a plusieurs paramètres importants à maîtriser lors de l'élaboration d'un MB. Le premier est primordial et concerne la constante d'affinité de l'hybridation avec la cible qui doit être nettement supérieure à celle permettant la formation de la tige. Auquel cas, aucun changement de conformation n'aura lieu et le MB restera « éteint ». Il existe une relation entre la longueur de la tige et la stabilité de celle-ci. Plus elle est courte et plus elle sera instable, le risque étant d'observer de faux positifs si la tige est trop instable. Il en est de même si elle est trop longue et donc trop stable, aboutissant à des faux négatifs. D'autres paramètres influençant les différentes constantes de stabilité (hybridation et formation de la tige) viennent s'ajouter, la température et le pH<sup>11</sup>. Enfin, la distance entre l'entité fluorescente et le quencher doit permettre en absence de la cible d'ADN d'avoir un transfert d'énergie des plus efficaces pour avoir le moins de luminescence possible. Tyagi et son équipe ont développé des systèmes comportant deux entités lumineuses qui permettent de révéler l'interaction par un déplacement (shift) de luminescence<sup>12</sup> (**Figure 7**). Le signal obtenu après reconnaissance de la cible ne provient pas du fluorophore, noté  $F_2$  à la longueur d'onde  $\lambda_{F2}$ , mais il va permettre de sensibiliser le second fluorophore  $F_1$  qui va émettre à la longueur d'onde  $\lambda_{F1}$  ( $\lambda_{F2} < \lambda_{F1}$ ).



**Figure 7 :** Principe du Molecular Beacon comportant deux fluorophores dans un système de FRET. En absence de la molécule cible, la luminescence de  $F_2$  est quenchée. En présence de la cible d'ADN le fluorophore  $F_2$  va pouvoir transférer son énergie vers le fluorophore  $F_1$  qui va luminescer. Figure adaptée de la publication<sup>9</sup>.

Durant ces vingt dernières années les MB ont été utilisés en biotechnologie pour différentes applications : (i) détection d'acides nucléiques<sup>13</sup>, (ii) l'étude de protéines<sup>14</sup>, (iii) détection d'ARN messager ou d'ADN simple brin<sup>15</sup>. Il est important de signaler leur capacité à pouvoir différencier deux séquences d'ADN qui ne diffèrent que par un nucléotide.

En couplant ce système à des techniques de luminescence et/ou de Polymerase Chain Reaction (PCR), il est alors possible d'étudier, d'identifier un gène au sein d'une cellule que ce soit *in vitro* ou *in vivo*. Cela offre aussi la possibilité de développer des systèmes de détection en temps réel grâce à des techniques de microscopie à fluorescence. Une utilisation maintenant très courante est la détection

d'ADN ou d'ARN en temps réel en surexprimant la séquence cible *via* une méthode de PCR. En quelques mots, la PCR est une technique d'amplification d'ADN *in vitro*. Elle comprend trois étapes (i) dénaturation du duplex d'ADN par voie thermique, (ii) hybridation des amorces aux extrémités de la séquence, (iii) élongation grâce à une ADN polymérase thermostable. Ces trois étapes forment un cycle que l'on peut répéter vingt à cinquante fois. Le MB sous sa forme non hybridé ayant une très faible luminescence, chaque cycle de la PCR va permettre d'augmenter la concentration en séquence cible et donc d'amplifier le signal du MB une fois hybridé<sup>16</sup>.

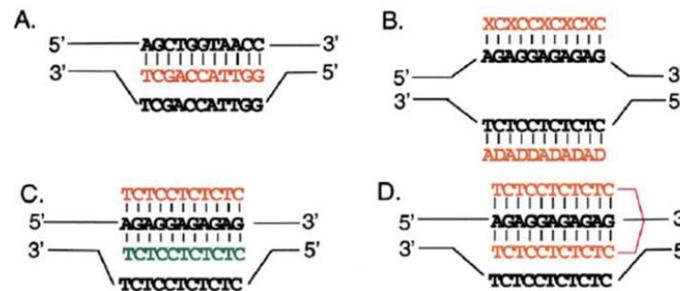
Pour une application *in vivo*<sup>17</sup>, Gewirtz et son équipe ont cherché à cibler les cellules tumorales par le biais d'un ARNm codant pour une protéine *vav* proto-oncogène (VAV1) qui est surexprimée dans les cellules cancéreuses. Dans un premier temps, ils ont montré que leur système était très sélectif, en testant plusieurs séquences d'ARNm en présence du MB. Ils ont pu extraire l'ARNm contenu dans des cellules tumorales et ont mesuré une augmentation significative de la luminescence, en comparaison avec le contrôle négatif. Les résultats observés, après 15 min, lors de l'injection du MB dans des cellules tumorales *in vivo* ont permis de valider cette approche pour la détection de cellules tumorales *via* la reconnaissance spécifique d'un ARNm. Cependant, ils ont pu observer qu'au-delà de 45 min, après injection du MB, le contrôle négatif commençait à luminescer. Une des hypothèses étant la dégradation du MB par des nucléases cellulaires.

La détection d'ADN utilisant les MB dans l'objectif de faire du monitoring *in vivo* est possible. Pour cela, plusieurs conditions doivent être remplies. Il faut que la cible soit en concentration suffisante pour que le rapport entre la fluorescence de la sonde et la fluorescence du milieu soit le plus grand possible. Dans le cas de cellules tumorales, certains ARNm vont être surexprimés, ainsi une détection *in vivo* est envisageable. Le ciblage d'une séquence d'ADN ne permet pas d'avoir une concentration suffisante, il est cependant possible de sur-exprimer cette séquence. Pour cela, il faut faire appel à des techniques d'amplification de l'ADN. Le problème est qu'il est nécessaire de passer par une étape de dénaturation de l'ADN à travers l'ouverture du duplex d'ADN. Une application *in vivo* n'est donc pas envisageable. Pour s'affranchir de cette dégradation du duplex d'ADN, le recours à un agent permettant l'ouverture du duplex est donc nécessaire. C'est ainsi que l'on a vu émerger l'utilisation de Peptide Nucleic Acids, en complément de l'utilisation des MB.

### 2.1.b PNA

Les Peptide Nucleic Acid (PNA) se composent d'oligomères de bases nucléiques, allant de 20 à 30 monomères. Contrairement à l'ADN, le squelette phosphate-ose est remplacé par un squelette de type peptidique qui permet une synthèse plus facile, mais surtout de pouvoir modifier plus aisément la séquence d'acide nucléique du PNA<sup>18</sup> et d'adapter le système. L'intérêt d'utiliser un PNA est d'avoir un système non chargé qui présente une très forte interaction avec son brin complémentaire dans le duplex d'ADN et facilite son ouverture. Il va donc y avoir une compétition entre l'homoduplex

ADN/ADN et l'hétéroduplex PNA/ADN. Si la formation de l'hétéroduplex est bien plus stable que le duplex d'ADN, le PNA peut s'insérer entre les deux brins d'ADN<sup>19</sup> (**Figure 8**). Des études théoriques suggèrent que la première étape d'insertion du PNA est liée à un intermédiaire de courte durée de vie où le PNA est complexé à l'ADN *via* des liaisons de type Hoogsteen et que c'est dans un second temps que les liaisons du type Watson et Crick interviennent<sup>20</sup>. En ayant recours à l'utilisation de deux PNA, l'ouverture du duplex d'ADN est possible puisque la stabilité du complexe (PNA)<sub>2</sub>/ADN est encore plus grande. Deux complexes sont possibles : (i) chaque PNA cible un brin d'ADN, (ii) les deux PNAs ciblent le même brin d'ADN. Pour augmenter significativement la stabilité du complexe (PNA)<sub>2</sub>/ADN formé (triplex), il a été démontré qu'un lien covalent flexible entre les deux PNA était primordial, on parle alors de *bis*-PNA. Un autre facteur permettant d'accélérer l'insertion du *bis*-PNA dans le duplex d'ADN est l'ajout d'une charge positive sur un des PNA. Une interaction électrostatique supplémentaire avec l'ADN est ajoutée du fait que l'ADN est chargé négativement. La conséquence directe de la formation du triplex est l'exposition du brin d'ADN complémentaire à celui impliqué dans le triplex.

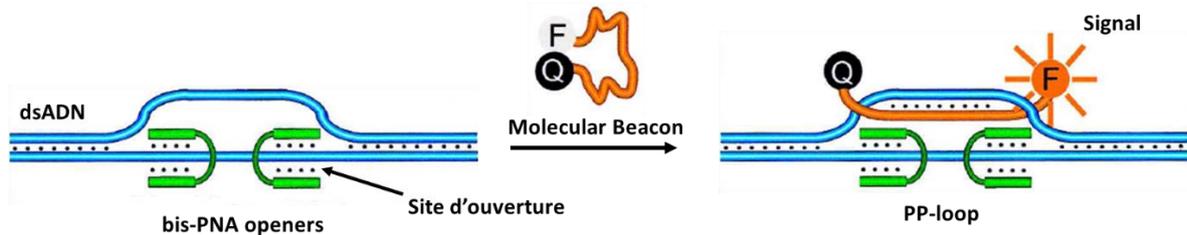


**Figure 8 :** A) Insertion d'un PNA par reconnaissance d'un brin B) Insertion de deux PNA pseudocomplémentaires ciblant chacun des brins. X = 2-thiouracil; D = diaminopurine C) Insertion de deux PNA autour d'un même brin D) Insertion d'un *bis*-PNA reliés par un lien flexible<sup>21</sup>.

L'utilisation de deux *bis*-PNA dans des régions proches va permettre d'augmenter la longueur de la séquence du brin exposé. Par exemple avec un seul *bis*-PNA, seul le brin complémentaire correspondant à la longueur de la séquence reconnue par le *bis*-PNA est exposé. En associant deux *bis*-PNA on peut obtenir une plus grande ouverture du double brin d'ADN qui ne se restreint pas aux deux séquences que ciblent les deux *bis*-PNA. Il existe deux structures différentes selon les séquences des deux PNA et de leur cible : soit la liaison se fait sur le même brin d'ADN, soit elle se fait sur les deux brins d'ADN<sup>22</sup>. L'accès à ce brin d'ADN permet d'envisager l'utilisation de systèmes de reconnaissance basés sur la complémentarité avec un simple brin d'ADN tel que le MB<sup>23</sup> ou par des oligodésoxyribonucléotides (ODN)<sup>24</sup>.

Le principe de l'association de deux *bis*-PNA avec un MB est d'ouvrir le duplex au niveau de la séquence cible du MB pour permettre à celui-ci de s'hybrider avec la séquence cible sans devoir dénaturer le duplex d'ADN au préalable (**Figure 9**). Kuhn et son équipe ont pu démontrer et valider cette approche en travaillant sur la reconnaissance d'un plasmide recombinant contenant une partie du

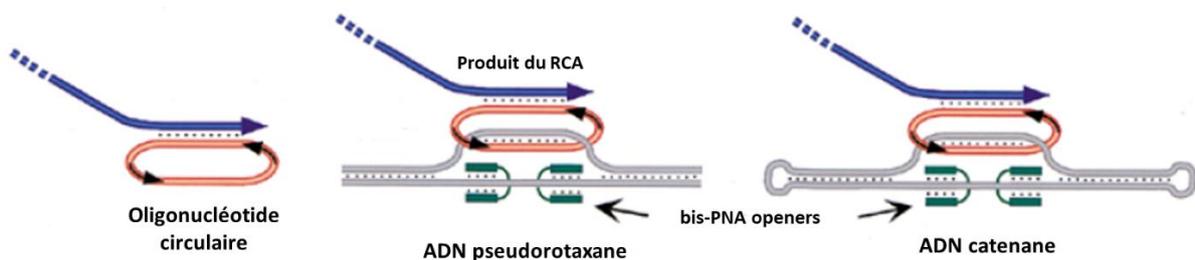
génomique du virus HIV-1<sup>23</sup>. Bien que les *bis*-PNA permettent d'ouvrir le double brin d'ADN ils souffrent d'une limitation. Les sites d'ouverture ne peuvent être que des séquences d'homopurine/homopyrimidine. Ce qui limite fortement les séquences d'ADN que le molecular beacon peut cibler. La séquence cible devant impérativement se trouver entre deux sites que les *bis*-PNA peuvent ouvrir.



**Figure 9 :** Principe de l'association des *bis*-PNA openers avec un Molecular Beacon utilisé par Kuhn<sup>23</sup>. Les PNA openers vont permettre l'ouverture du double brin d'ADN pour exposer le brin cible au MB.

### 2.1.c OCN

Un système dérivé de celui du MB est l'oligodésoxyribonucléotide circulaire (OCN) dont la structure est cyclique. Il présente deux avantages : (i) la formation d'hybrides stables avec l'ADN, (ii) la possibilité d'amplifier la séquence d'ADN. Toujours dans l'optique d'augmenter le signal qui émane de l'hybridation de la sonde, ici l'OCN, plusieurs équipes se sont intéressées à associer des méthodes de réplication de l'ADN. Ceci pour augmenter la concentration de la séquence ciblée et par conséquent augmenter la quantité de signal. Son utilisation s'associe aisément avec une méthode d'amplification nommée rolling-circle amplification (RCA)<sup>24</sup>. C'est un processus basé sur l'utilisation d'ADN polymérase qui permet de créer de multiples copies de la séquence cible. Ceci générant donc un concatémère, un ADN simple brin avec une répétition d'un seul et même monomère. Ainsi on peut augmenter significativement la quantité d'ADN cible pour qu'une sonde luminescente puisse s'hybrider autour de cette séquence (**Figure 10**). C'est en cherchant à cibler la séquence d'un plasmide recombinant au sein du gène HIV-1 que l'équipe de Frank-Kamenetskii a pu obtenir des résultats prometteurs grâce à cette méthodologie PNA-OCN-RCA.



**Figure 10 :** Principe général de l'association PNA-OCN-RCA. Les *bis*-PNA (vert) ouvrent le double brin d'ADN dont un des brins sert de support à l'OCN (rouge). L'ADN polymérase permet de former un concatémère (en bleu) dont la séquence est identique au brin ciblé. Figure reprise de la publication<sup>24</sup>.

L'utilisation de PNA permet de s'affranchir des étapes de dénaturation de l'ADN pour accéder à de l'ADN simple brin. Cependant, pour des applications *in vivo*, celui-ci présente certaines limites. La limitation majeure réside dans la nécessité de vectorisation pour traverser les membranes cellulaires. Ensuite, pour que l'OCN puisse reconnaître sa séquence cible, il faut que les deux *bis*-PNA interagissent de part et d'autre de la séquence ciblée. Les *bis*-PNA ne fonctionnant pas en tandem, l'insertion de l'un n'influe pas celle de l'autre et ne les empêche donc pas de s'insérer ailleurs que dans la région ciblée, provoquant de multiples ouvertures.

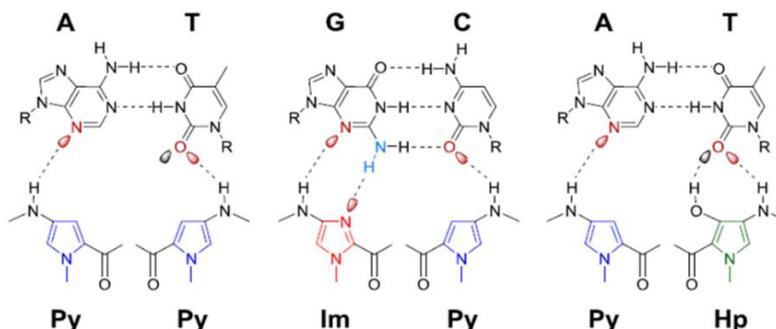
Les méthodes de reconnaissance d'une séquence d'ADN décrites précédemment reposent toutes sur l'ouverture, en conditions dénaturantes ou non dénaturantes, du duplex. Cela permet de développer différentes sondes capables de s'hybrider autour d'un ADN simple brin par complémentarité avec celui-ci. Il existe toutefois une méthode alternative qui permet de faire de la détection d'ADN double brin sans ouvrir celui-ci. Elle repose sur la reconnaissance des bases nucléiques par l'intermédiaire des groupes d'atomes exposés dans le grand ou le petit sillon. Un des exemples les mieux documentés repose sur la mise au point de ligands du petit sillon basés sur des oligomères de type Polyamide Pyrrole-Imidazole.

## 2.2 - Reconnaissance de l'ADN par des polymères

L'origine du développement de Polyamide Pyrrole-Imidazole (PPI) pour la reconnaissance d'ADN double brin découle de la découverte de deux antibiotiques naturels capables de bloquer les voies de réplifications de bactéries en interagissant avec l'ADN. Ces antibiotiques sont la Netropsine et la Distamycine A. Tous deux sont basés sur une structure ayant deux à trois pyrroles consécutifs. Ils interagissent avec l'ADN au niveau du petit sillon avec une préférence pour les séquences riches en paires AT. La structure RX a révélé la formation d'un complexe 1:1 avec l'ADN<sup>25</sup> mais la structure établie par RMN a montré que le rapport est de 2:1 dans le cas de la Distamycine<sup>26</sup>. C'est ainsi que dans les années 1990, un premier système de PPI, avec la séquence Im-Py-Py, fut synthétisé par l'équipe de P. Dervan<sup>27</sup>. Ce PPI étant capable de reconnaître la séquence de 5 paires de bases 5'-WGWCW-3' (W= A ou T) en formant un complexe 2:1 et de la distinguer de la séquence 5'-WGWWW-3'. Ceci offrit la perspective de pouvoir distinguer à l'aide de deux brins polyamides une séquence d'ADN par rapport à une autre. Dans ce complexe 2:1 la différenciation des quatre paires de bases met en jeu des paires d'hétérocycles (Im ou Py), chacun étant porté par un des deux polyamides (**Figure 11**).

La reconnaissance d'une base à une autre repose sur les liaisons hydrogène que les hétérocycles forment avec les bases nucléiques. L'imidazole cible sélectivement la guanine grâce à la formation de deux liaisons hydrogène. Le pyrrole par contre qui ne forme qu'une seule liaison hydrogène avec A, T et C ne permet pas de les distinguer. Dans l'optique de développer un PPI spécifique d'une séquence, il semblait vital de trouver un autre hétérocycle pour faire la distinction entre A et T. En ayant recours au N-méthyl-3-hydroxypyrrole (Hp), associé avec un pyrrole, la distinction entre A et T devient alors

possible, car Hp reconnaît spécifiquement la thymine grâce à une nouvelle liaison hydrogène (**Figure 11**) et la formation d'une paire Im-Py peut servir à distinguer la paire GC (et inversement). La paire Py-Py sera donc utilisée pour reconnaître AT ou TA sans pouvoir toutefois les distinguer.



**Figure 11** : Liaisons hydrogène impliquées dans la reconnaissance des bases nucléiques par les PPI <sup>28</sup>.

Cependant, l'utilisation de Hp à la place de Py diminue fortement l'affinité pour l'ADN, *in vitro* ou *in cellulo*. Des problèmes de stabilité ont été observés au niveau de Hp en présence d'acide ou d'espèces radicalaires ce qui rend difficile son utilisation pour des applications biologiques<sup>29</sup>.

Bien que ce système permette de distinguer les paires de bases AT et CG, certaines séquences telles que 5'-GAT-3' ou 5'-GCG-3' sont difficilement reconnues. Une alternative peut être envisagée en utilisant une  $\beta$ -alanine ( $\beta$ -ala) au niveau de l'extrémité du PPI, qui appariée avec un pyrrole, permet de reconnaître AT ou TA. Cela apporte aussi de la flexibilité au système et permet une meilleure affinité, les hétérocycles formant plus facilement des interactions de type Van der Waals avec les bases nucléiques<sup>27</sup>. Cependant ce PP modifié ne peut les distinguer et son utilisation n'est possible que si la séquence ciblée contient T ou A en fin de séquence. D'autre part, cibler de manière sélective et spécifique une séquence de plus de 5 paires de bases en utilisant des paires ( $\beta$ -ala)-Py reste difficile, l'affinité du PPI avec sa cible diminuant fortement<sup>30</sup>.

De manière à augmenter la sélectivité du PPI avec sa cible, un lien covalent entre les deux brins polyamides a été ajouté, montrant une augmentation significative de l'affinité en comparaison des brins non reliés. Plusieurs liens ont été utilisés pour relier la partie C-terminale de l'un avec le N-terminale de l'autre, tous étant basés sur un motif amine – alkyle – acide carboxylique. Celui qui présente les meilleurs résultats est l'acide aminobutyrique ( $\gamma$ -turn). Il permet aussi de mieux reconnaître la paire AT au niveau de la jonction. Le duplex d'ADN étant chargé négativement, l'ajout d'une charge positive au sein du PPI permettrait de former un complexe encore plus stable. C'est en ajoutant une amine tertiaire à la fin d'un des brins polyamides que la formation du complexe avec l'ADN est la plus favorisée <sup>31</sup> (**Figure 12**).

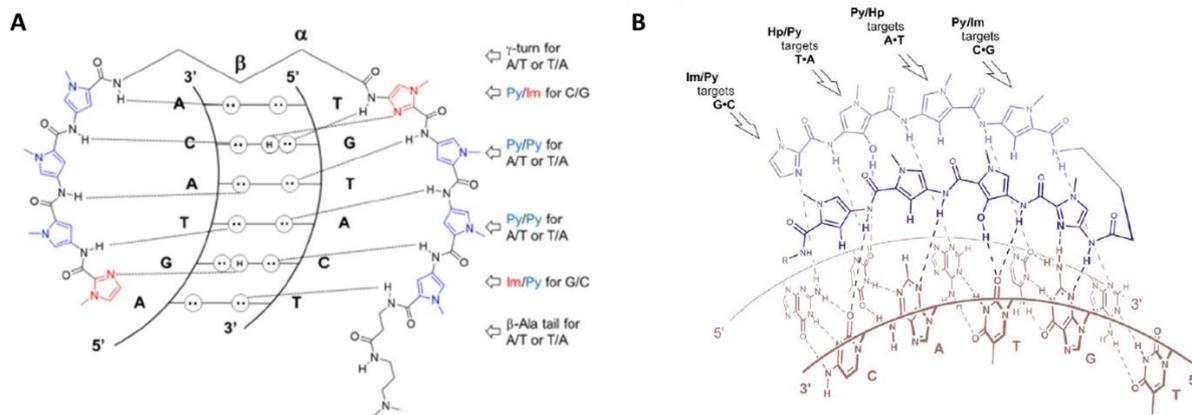


Figure 12 : A) Modèle de reconnaissance d'ADN double brin (5'-TGTAC-3') par un PPI<sup>28</sup>. B) Vue schématique d'un complexe PPI-ADN vu en perspective<sup>31</sup>.

On distingue plusieurs structures pour le PPI suivant le positionnement du lien  $\gamma$ -turn (**Figure 13**). Soit aux extrémités des deux brins polyamides (U-pin ou Hairpin) soit *via* les azotes des hétérocycles et dans ce cas, la chaîne alkyle varie entre 4 et 6 atomes de carbone. Ceci définit une structure dite H-pin. On peut aussi lier les deux brins aux deux extrémités et obtenir un PPI cyclique. Cependant, la forme H-pin présente certains désavantages vis-à-vis de la forme U-pin. La synthèse en phase solide d'H-pin avec une stratégie Boc ou Fmoc semble plus difficile à réaliser. En effet, il s'agit d'une synthèse d'un polypeptide branché comparé à une synthèse linéaire. Cependant, des méthodes de synthèses programmables ont été développées et permettent de les synthétiser de manière efficace<sup>32 33</sup>.

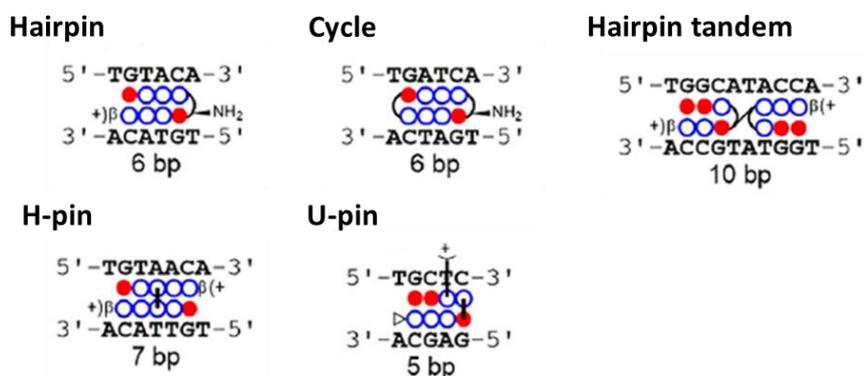
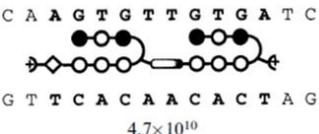
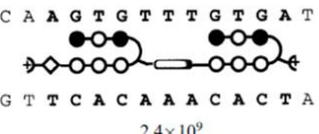
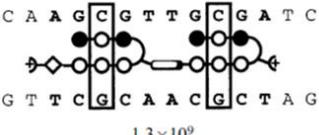
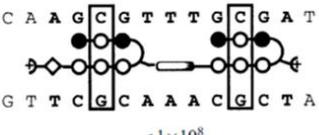


Figure 13 : Représentation schématique des différentes structures de PPI<sup>28</sup>.

Si l'on compare les affinités d'un U-pin et H-pin, pour une même cible, l'affinité d'un U-pin à huit hétérocycliques est comparable à celle d'un H-pin à six hétérocycles. Les PPI cycliques quant à eux montrent une sélectivité moindre par rapport aux H-pin, mais une meilleure affinité<sup>28</sup>. En ayant recours à un PPI du type Hairpin tandem l'équipe de Dervan en 2002 a démontré qu'il était possible de cibler une séquence d'ADN de 10 paires de bases (bp)<sup>34</sup>. Pour cela il a eu recours à deux hairpins à 6 cycles ( $\text{ImPyIm-}\gamma\text{-PyPyPy-}\beta\text{Ala}$ ) dont il a fait varier la longueur et la nature du linker. Celui présentant la plus forte affinité pour la séquence cible d'ADN palindromique 5'-AGTGTGTGA-3' est le composé

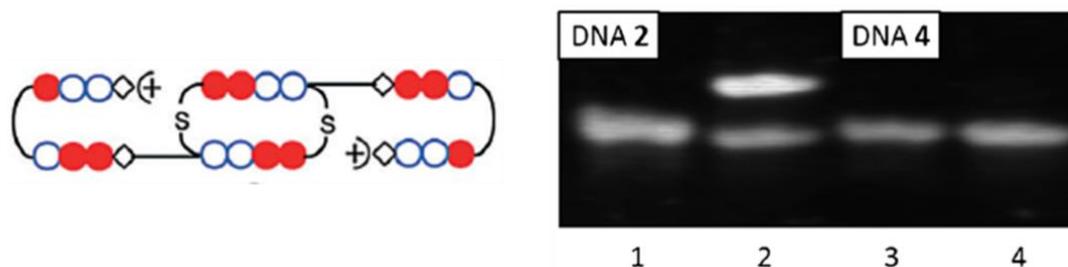
1b pIK3 dont le linker contient 4 atomes de carbones (**Tableau 1**). Sa constante de dissociation a été déterminée par foot printing et calculée à  $2.1 \times 10^{-11} \text{ M}$  ( $K_a = 4.7 \times 10^{10} \text{ M}^{-1}$ ). En présence d'un ADN d'une longueur de 11 bp (5'-AGTGTTTGTGA-3'), avec l'insertion d'une thymine au centre de la séquence, sa constante de dissociation est multipliée d'un facteur 20, mais reste élevée autour de  $4.2 \times 10^{-10} \text{ M}$  ( $K_a = 2.4 \times 10^9 \text{ M}^{-1}$ ). Ceci pouvant s'expliquer par un linker qui est devenu trop court vis-à-vis d'une séquence d'ADN qui s'allonge. L'utilisation d'un linker plus long pourrait permettre de retrouver une affinité plus forte. Lorsque ce même PPI est mis en présence d'une séquence « mismatch » d'ADN de 10 bp (5'-AGCGTTGCGA-3') l'affinité ne diminue que d'un facteur 40. Elle diminue encore d'un facteur 10 lorsque la séquence est de 11 bp (5'-AGCGTTTGCGGA-3'). Bien que l'affinité soit grande, cela montre que la sélectivité des PPI peut poser problème. Une modification de deux bp dans une séquence qui n'en possède que dix n'affecte finalement que peu son affinité. Un ciblage d'une séquence spécifique semble donc difficile avec les PPI.

**Tableau 1 :** Constantes d'association (en  $\text{M}^{-1}$ ) du PPI avec les cibles différents ADN ciblés<sup>34</sup>.

Polyamide and plasmid	10 bp	11 bp	Specificity, $K_A(10 \text{ bp})/K_A(11 \text{ bp})$
1b pIK3	 <p style="text-align: center;"><math>4.7 \times 10^{10}</math></p>	 <p style="text-align: center;"><math>2.4 \times 10^9</math></p>	20
1b pIK4	 <p style="text-align: center;"><math>1.3 \times 10^9</math></p>	 <p style="text-align: center;"><math>&lt; 1 \times 10^8</math></p>	>13

Un autre exemple de PPI est celui développé par Yamamoto<sup>35</sup> capable de reconnaître une séquence de 14 bp. Ce PPI est en réalité l'assemblage de 3 unités dont la partie centrale est un PPI cyclique auquel sont reliés deux hairpins de part et d'autre (**Figure 14**). La séquence cible d'ADN est la suivante 5'-AGCCTTGGCCTTGGCA-3' (ADN 2), les bp soulignées ont été insérées au sein de la séquence cible comme « espaceur » et se situent au niveau des deux liens hairpin-PPI cyclique. La constante de dissociation a été mesurée à  $3.4 \mu\text{M}$  et l'interaction du PPI en présence d'un ADN ciblé/non ciblé a été étudiée par une technique de gel retard en condition dénaturantes (Native PAGE). L'ADN non ciblé possédant deux bp modifiées, 5'-AGTCTTGGCCTTGTICA-3' (ADN 4), au niveau des cibles des deux hairpins. Sur le gel retard présenté ci-dessous on peut voir une bande retardée (entrée 2) qui correspondant à l'interaction du PPI avec la cible. Aucune bande retardée n'est visible pour l'ADN 4 (entrée 4) ce qui atteste de la sélectivité du système. Dans les conditions expérimentales utilisées pour réaliser ce gel retard il faut noter que les concentrations utilisées sont très loin de celles que l'on retrouve *in vivo*. La concentration d'ADN a été fixée à  $4 \mu\text{M}$  et celle du PPI à  $40 \mu\text{M}$ , les auteurs ont réalisé

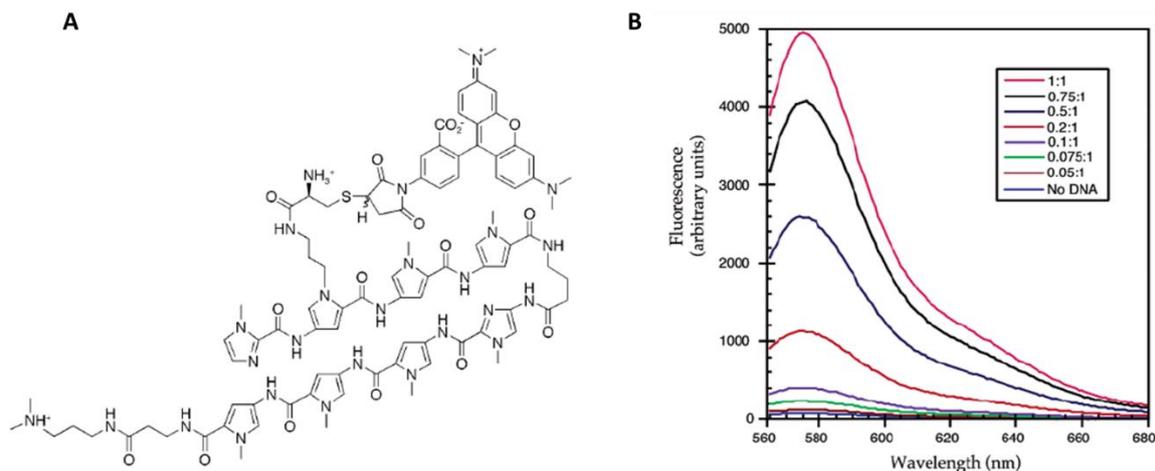
plusieurs expériences en essayant de diminuer le ratio PPI/ADN ainsi que la concentration en ADN mais aucune bande ne pouvait être observée.



**Figure 14 :** À gauche : schéma du PPI et des trois unités qui le compose, les Im sont représentés en rouge et les Py avec les cercles bleus. À droite : l'analyse PAGE obtenue avec les ADN 2 et 4 respectivement cible et non-cible. Les entrées 1 et 3 correspondent aux ADN seuls et les entrées 2 et 4 à l'interaction du PPI avec les ADN. La concentration en ADN a été fixée à 4  $\mu\text{M}$  et à 40  $\mu\text{M}$  pour le PPI.

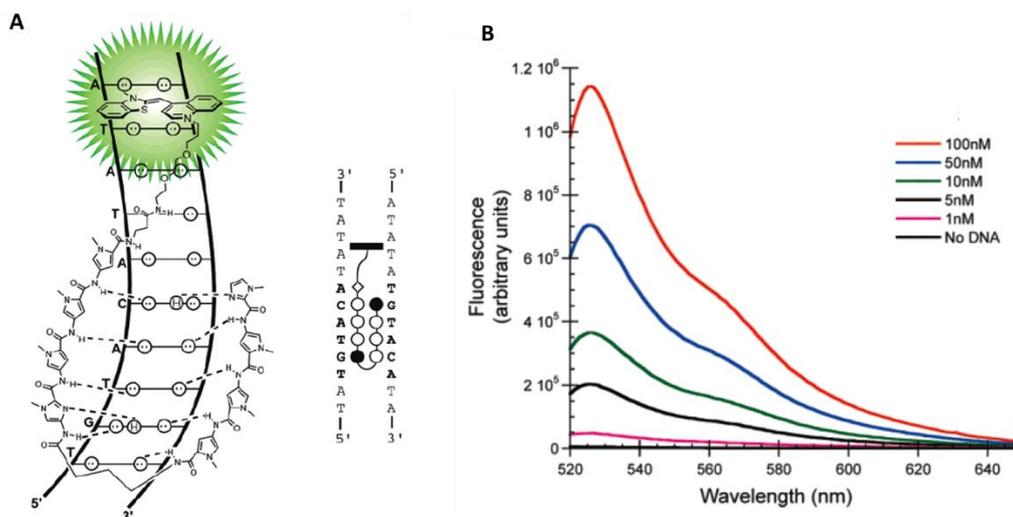
En comparaison avec le système de PPI présenté précédemment, la séquence ciblée est de 14 bp contre 10 bp. Cependant, la constante de dissociation est divisée par un facteur  $10^4$ . En effet, dans le cas du hairpin tandem la constante de dissociation est de l'ordre du nano molaire ( $4.2 \cdot 10^{-10} \text{ M}$ ) alors que celle avec le PPI cyclique est de l'ordre du micro molaire ( $3.4 \cdot 10^{-6} \text{ M}$ ). Ceci illustre bien la difficulté lorsque l'on souhaite cibler sélectivement une séquence d'ADN de plus de 5 bp avec un système basé sur les PPI.

Toutefois l'équipe de Rucker a commencé le développement d'une sonde fluorescente ciblant le duplex d'ADN en incorporant un fluorophore organique à un PPI<sup>36</sup>. L'un des PPI utilisés est un H-pin composé de huit hétérocycles ImPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ Ala dont la cible d'ADN est une séquence de 6 paires de bases 5'-AGT ACT-3' au sein d'un 17-mer d'ADN. Le fluorophore utilisé est la tétramethylrhodamine (TMR) greffé sur le pyrrole en position 2. En absence d'ADN, il présente une faible fluorescence due à la proximité du TMR avec les hétérocycles, cependant une fois le PPI lié à l'ADN l'intensité d'émission est augmentée. La constante de dissociation de ce PPI-TMR avec la cible, comparée avec celle du PPI nu autour du nano molaire, a montré une baisse d'affinité pouvant aller jusqu'à un facteur 10, la constante d'association ayant été mesurée à  $1.7 \cdot 10^9 \text{ M}^{-1}$  ( $K_d = 0.59 \text{ nM}$ ). La formation d'un complexe 1 : 1 a été vérifiée par titration avec des équivalents croissants de PPI-TMR pour une concentration fixe en ADN de 1  $\mu\text{M}$  (**Figure 15**). Une forte augmentation de l'intensité d'émission du fluorophore a été observée validant l'utilisation des PPI comme sonde à ADN.



**Figure 15 :** A) PPI utilisé par Rucker pour la reconnaissance de 5'GTAC-3' au sein d'un ADN long de 17 bp. B) Évolution de l'émission du PPI après ajout croissant du PPI-TMR, la concentration d'ADN est fixée à 1  $\mu$ M<sup>36</sup>.

Dans cette même logique, l'équipe de Dervan a utilisé un système de fluorophore différent du TMR. Il s'agit d'un thiazole orange (TO, **Figure 16**) qui est capable de s'intercaler entre les paires de bases de l'ADN<sup>37</sup>. Naturellement peu luminescent dû à la libre rotation entre le noyau benzothiazole et la quinoline, son intercalation va permettre de bloquer cette rotation pour une très nette augmentation du signal luminescent. Fonctionnalisée sur un PPI de séquence ImPyPyPy- $\gamma$ -ImPyPyPy- $\beta$  via une chaîne PEG, l'équipe de Dervan a pu démontrer une augmentation du signal par un facteur 1000 après ajout progressif de la cible d'ADN jusqu'à 1 équivalent (**Figure 16**). Cette séquence cible 5'-TGTACA-3' est contenue au sein d'un 32-mer.



**Figure 16 :** A) PPI utilisé par Dervan pour la reconnaissance de 5'TGTACA-3' au sein d'un ADN long de 32 bp. B) Évolution de l'émission du PPI-TO après ajout croissant d'ADN, la concentration initiale en protéine étant de 100 nM<sup>37</sup>.

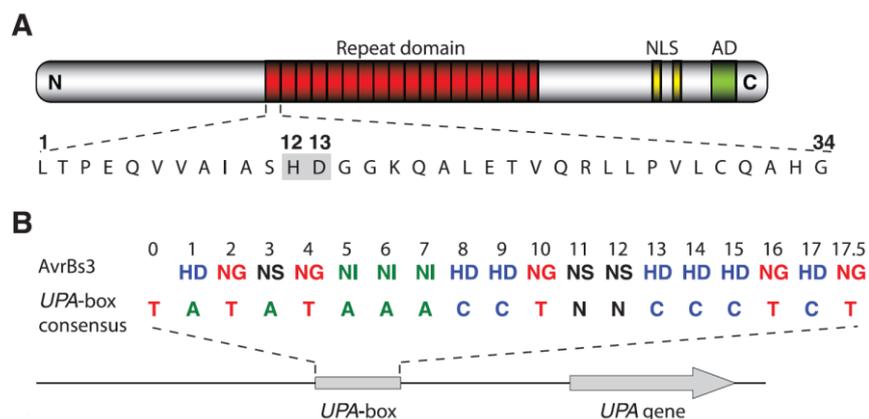
Les systèmes présentés jusqu'ici sont capables de reconnaître une séquence d'ADN, soit par appariement avec un brin d'ADN (MB), soit par la formation d'un complexe avec le duplex d'ADN. Les principaux problèmes rencontrés sont un manque d'affinité lorsque la séquence ciblée va au-delà des 10 bp dans le cas des PPI et dans certains cas un manque de sélectivité lorsqu'il est en présence

d'une séquence non cible. Les MB, quant à eux, nécessitent de mettre en œuvre des techniques impliquant plusieurs acteurs (PNA) pour l'ouverture du duplex d'ADN ce qui rend son utilisation délicate. Cependant ces systèmes ont montré d'intéressantes capacités en tant que régulateurs de l'expression des gènes, entrant en compétition avec les enzymes qui permettent la lecture de l'ADN (traduction, transcription). Il existe au sein du métabolisme des systèmes qui sont capables de reconnaître et décoder l'ADN. Ces protéines sont des facteurs de transcription dont le mécanisme de reconnaissance est basé sur l'utilisation de doigts de zinc. Ceci nous amène donc à un nouveau type de sonde à ADN qui cette fois-ci repose sur une reconnaissance par une protéine.

### 2.3 - Reconnaissance de l'ADN par des protéines

#### 2.3.a Protéines TALE

La protéine TALE (Transcription Activator-Like Effectors), identifiée dans des protéo-bactéries du genre *Xanthomonas*, est une protéine capable de reconnaître sélectivement une séquence d'ADN. C'est grâce à ces protéines que la bactérie peut se développer puisqu'elle se fixe au niveau des promoteurs des gènes impliqués dans la défense immunitaire. Les protéines TALE sont composées d'un domaine de répétition dans lequel plusieurs unités de répétition se succèdent, de deux domaines d'adressage au noyau (NLS = nuclear localization signal) et d'un domaine d'activation (AD) (**Figure 17**). Chacune des unités de répétition va permettre la reconnaissance d'une paire de bases de l'ADN<sup>38</sup>. Plus la longueur du domaine de répétition sera grande et plus la séquence d'ADN reconnue sera longue. Une unité de répétition est définie par 34 acides aminés dont la séquence est quasi identique d'une unité à l'autre, seuls les aa en position 12 et 13 diffèrent et permettent la différenciation des bases de l'ADN<sup>38</sup>.



**Figure 17 :** A) Séquence de la protéine TALE avec la séquence détaillée d'un des domaines de répétition. B) Sont présentées les paires de bases reconnues selon la séquence des acides aminés des positions 12 et 13 des doigts de zinc<sup>38</sup>.

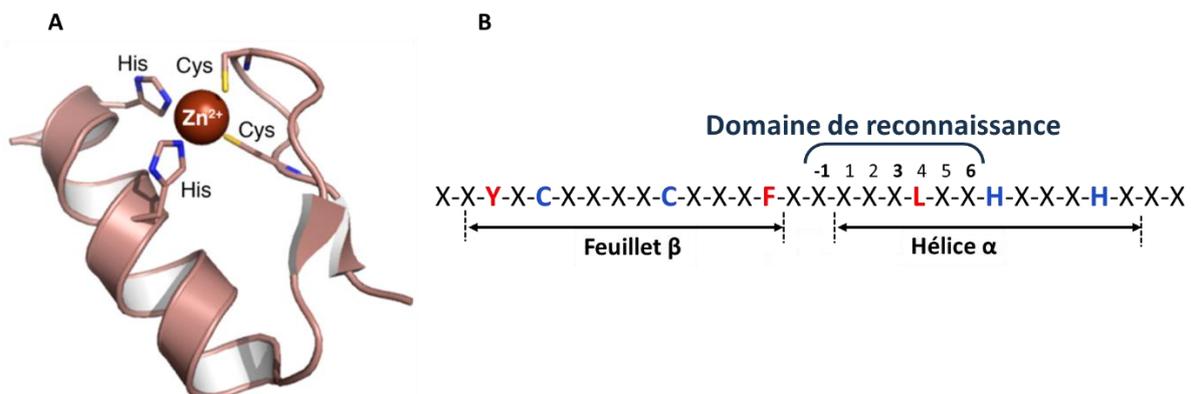
Le groupe de Pederson a travaillé sur le ciblage des télomères en utilisant des protéines TALE couplées à des protéines fluorescentes (jaune et rouge)<sup>39</sup>. Ces sondes fonctionnant en conditions non dénaturantes par dimérisation de deux protéines autour de la séquence cible, la protéine fluorescente jaune servant de

fluorophore donneur dans un système de transfert d'énergie. Le télomère ciblé étant une répétition de la séquence d'ADN 5'-TTA-GGG-3' ils ont pu corrélérer l'intensité de luminescence émise par la protéine fluorescente rouge avec la longueur du télomère. Le principal inconvénient des sondes basées sur des protéines TALE est que ce sont des systèmes très imposants et peu versatiles dans lesquels il est impossible d'incorporer un système de luminescence autre que des protéines fluorescentes. Leur synthèse ne pouvant être réalisée par voie chimique la fonctionnalisation par un fluorophore organique semble impossible.

### 2.3.b Doigts de zinc

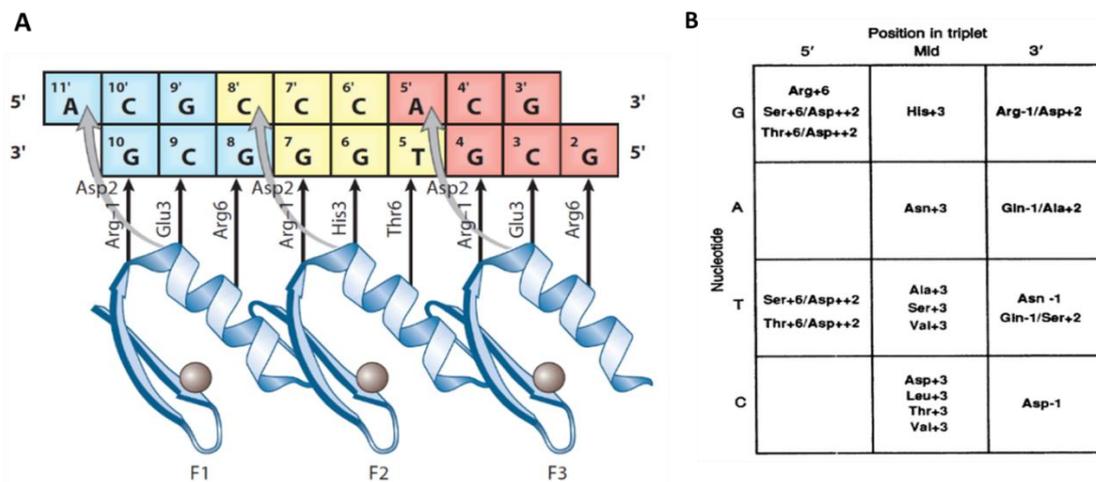
L'origine de la découverte des doigts de zinc repose sur les études réalisées par Miller, McLachlan et Klug en 1985 sur un facteur de transcription nommé TFIIIA chez *Xenopus Laevis*<sup>40</sup>. Ils ont remarqué qu'au sein de celui-ci un motif de 30 acides aminés venait à se répéter. Parmi les 412 acides aminés qui composent TFIIIA, ils ont en effet identifié 9 fois cette séquence. Sur chacun de ces motifs, ils ont pu mettre en évidence la présence d'un atome de zinc conférant une structure bien identique à chacun de ces petits domaines, qui furent appelés « finger ».

Sur la base d'études RMN, Berg a proposé une première structure de ce domaine en 1988 avec une boucle formée par un complexe autour d'un ion zinc au sein d'une hélice  $\alpha$  et d'un feuillet  $\beta$  antiparallèle<sup>41</sup>. D'autres études sur le facteur de transcription ADR1, menées par Parraga, ont montré des similarités avec le modèle de Berg ce qui a permis d'affiner la structure et de proposer un premier modèle 2-D du doigt de zinc<sup>42</sup>. Mais c'est en 1989 que le premier modèle 3-D a été publié par l'équipe de Lee<sup>43</sup>. Dans ce modèle, l'atome de zinc présente une géométrie de coordination tétraédrique, coordination assurée par deux cystéines et deux histidines. Plusieurs autres acides aminés de nature hydrophobe, tels que la tyrosine, phénylalanine et leucine, à des positions précises suggèrent la formation d'un cœur hydrophobe permettant la stabilisation du doigt de zinc (**Figure 18**).



**Figure 18 :** A) Structure 3D d'un doigt de zinc<sup>44</sup> comportant un feuillet  $\beta$  et une hélice  $\alpha$ , la coordination du zinc est assurée par deux cystéines et deux histidines. B) La séquence consensus des doigts de zinc comportant les 7 acides aminés essentiels à la reconnaissance.

Ce n'est qu'en 1991 que Pavletich et Pabo<sup>45</sup> ont pu résoudre la première structure cristalline d'un complexe entre un duplex d'ADN de séquence, 5'-GCG TGG GCG-3', et une protéine à trois doigts de zinc, Zif268. Il en résulte que l'interaction principale se situe au niveau du grand sillon de l'ADN *via* l'hélice  $\alpha$ . Plus précisément, ce sont les acides aminés en position -1,3 et 6 qui forment des liaisons hydrogène avec trois bases successives de l'ADN (triplet). La seconde interaction provient de l'acide aminé en position 2 qui lui interagit avec l'autre brin d'ADN (**Figure 19**).



**Figure 19** : A) Interaction de trois doigts de zinc avec l'ADN double brin. Chacun des doigts interagit avec un triplet de paires de bases de l'ADN. Issu de la structure cristalline de Zif268. B) Code de reconnaissance établi par Choo et Klug selon le triplet de bases ciblé<sup>44</sup>.

Dès lors, une séquence consensus fut établie, avec seulement 7 acides aminés essentiels, dont 4 participent à la reconnaissance d'un triplet d'ADN. Un premier code de reconnaissance a été décrit par Choo et Klug en 1994<sup>46</sup> (**Figure 19**). Les résultats ont été obtenus en mettant en évidence un doigt de zinc avec un triplet d'ADN cible dont une des trois bases nucléiques était fixe et les deux autres générées aléatoirement parmi les 4 bases. Chacun des doigts de zinc utilisés a révélé une séquence d'ADN cible où l'affinité se démarquait d'autres séquences avec une constante de dissociation de l'ordre de 10 nM, appelée séquence signature. C'est en se basant sur ces séquences signatures que le premier code a été décrit. Sur la **Figure 19**, pour une thymine au centre du triplet, plusieurs aa en position 3 forment des interactions avec celle-ci. En position 3' du triplet, seule l'utilisation d'une asparagine en position -1 permet une interaction. Cependant si à cette même position une glutamine était couplée à une alanine en position 2, alors les interactions de cette paire suffiraient à reconnaître la thymine. Bien qu'il reste quelques zones blanches dans lesquelles il est impossible de définir des corrélations, ce premier code permet néanmoins d'établir les premières règles de reconnaissance d'un triplet de bases en se basant non seulement sur les aa des positions -1, 3 et 6, mais aussi sur certaines associations d'aa avec ceux en position 2.

Un nouveau code plus fourni fut publié vingt ans plus tard<sup>44</sup> (**Figure 20**). En utilisant une technique de phage display pour l'élaboration de milliers de doigts de zinc, ceci a permis d'affiner le code de

reconnaissance. Néanmoins, il manque des séquences d'aa qui présentent une spécificité pour les séquences d'ADN comportant une thymine en position 5' (ex : TTT, TTC).

		Second position				
		T	C	A	G	
T	TTT -	TCT - <u>RL</u> <u>DIQ</u> F	TAT - <u>ARG</u> <u>N</u> <u>LRT</u>	TGT -	T	
	TTC -	TCC - <u>RS</u> <u>DER</u> <u>KR</u>	TAC - <u>SR</u> <u>G</u> <u>N</u> <u>LKS</u>	TGC - <u>AP</u> <u>K</u> <u>AL</u> <u>GW</u>	C	
	TTA -	TCA - <u>RS</u> <u>D</u> <u>HL</u> <u>TT</u>	TAA - <u>Q</u> <u>AS</u> <u>N</u> <u>LIS</u>	TGA - <u>Q</u> <u>AG</u> <u>H</u> <u>LAS</u>	A	
	TTG - <u>R</u> <u>K</u> <u>D</u> <u>AL</u> <u>RG</u>	TCG - <u>RL</u> <u>R</u> <u>AL</u> <u>DR</u>	TAG - <u>R</u> <u>E</u> <u>D</u> <u>N</u> <u>LHT</u>	TGG - <u>RS</u> <u>D</u> <u>H</u> <u>LTT</u>	G	
C	CTT - <u>T</u> <u>T</u> <u>G</u> <u>AL</u> <u>TE</u>	CCT - <u>T</u> <u>K</u> <u>N</u> <u>SL</u> <u>TE</u>	CAT - <u>T</u> <u>S</u> <u>G</u> <u>N</u> <u>LTE</u>	CGT - <u>S</u> <u>R</u> <u>R</u> <u>T</u> <u>C</u> <u>RA</u>	T	
	CTC - <u>Q</u> <u>R</u> <u>H</u> <u>HL</u> <u>VE</u>	CCC - <u>S</u> <u>K</u> <u>K</u> <u>H</u> <u>LAE</u>	CAC - <u>S</u> <u>K</u> <u>K</u> <u>AL</u> <u>TE</u>	CGC - <u>H</u> <u>T</u> <u>G</u> <u>H</u> <u>LLE</u>	C	
	CTA - <u>Q</u> <u>N</u> <u>ST</u> <u>LTE</u>	CCA - <u>T</u> <u>S</u> <u>H</u> <u>SL</u> <u>TE</u>	CAA - <u>Q</u> <u>S</u> <u>G</u> <u>N</u> <u>LTE</u>	CGA - <u>Q</u> <u>S</u> <u>G</u> <u>H</u> <u>LTE</u>	A	
	CTG - <u>R</u> <u>N</u> <u>D</u> <u>AL</u> <u>TE</u>	CCG - <u>R</u> <u>N</u> <u>D</u> <u>L</u> <u>TE</u>	CAG - <u>R</u> <u>A</u> <u>D</u> <u>N</u> <u>LTE</u>	CGG - <u>R</u> <u>S</u> <u>D</u> <u>K</u> <u>LTE</u>	G	
A	ATT - <u>H</u> <u>K</u> <u>N</u> <u>AL</u> <u>QN</u>	ACT - <u>T</u> <u>H</u> <u>L</u> <u>D</u> <u>LIR</u>	AAT - <u>T</u> <u>T</u> <u>G</u> <u>N</u> <u>LTV</u>	AGT - <u>H</u> <u>R</u> <u>T</u> <u>L</u> <u>TN</u>	T	
	ATC - <u>R</u> <u>R</u> <u>S</u> <u>A</u> <u>C</u> <u>R</u> <u>R</u>	ACC - <u>D</u> <u>K</u> <u>K</u> <u>D</u> <u>L</u> <u>T</u> <u>R</u>	AAC - <u>D</u> <u>S</u> <u>G</u> <u>N</u> <u>L</u> <u>R</u> <u>V</u>	AGC - <u>E</u> <u>R</u> <u>S</u> <u>H</u> <u>L</u> <u>R</u> <u>E</u>	C	
	ATA - <u>Q</u> <u>K</u> <u>S</u> <u>S</u> <u>L</u> <u>I</u> <u>A</u>	ACA - <u>S</u> <u>P</u> <u>A</u> <u>D</u> <u>L</u> <u>T</u> <u>R</u>	AAA - <u>Q</u> <u>R</u> <u>A</u> <u>N</u> <u>L</u> <u>R</u> <u>A</u>	AGA - <u>Q</u> <u>L</u> <u>A</u> <u>H</u> <u>L</u> <u>R</u> <u>A</u>	A	
	ATG - <u>R</u> <u>R</u> <u>D</u> <u>E</u> <u>L</u> <u>N</u> <u>V</u>	ACG - <u>R</u> <u>T</u> <u>D</u> <u>T</u> <u>L</u> <u>R</u> <u>D</u>	AAG - <u>R</u> <u>K</u> <u>D</u> <u>N</u> <u>L</u> <u>K</u> <u>N</u>	AGG - <u>R</u> <u>S</u> <u>D</u> <u>H</u> <u>L</u> <u>T</u> <u>N</u>	G	
G	GTT - <u>T</u> <u>S</u> <u>G</u> <u>S</u> <u>L</u> <u>V</u> <u>R</u>	GCT - <u>T</u> <u>S</u> <u>G</u> <u>E</u> <u>L</u> <u>V</u> <u>R</u>	GAT - <u>T</u> <u>S</u> <u>G</u> <u>N</u> <u>L</u> <u>V</u> <u>R</u>	GGT - <u>T</u> <u>S</u> <u>G</u> <u>H</u> <u>L</u> <u>V</u> <u>R</u>	T	
	GTC - <u>D</u> <u>P</u> <u>G</u> <u>A</u> <u>L</u> <u>V</u> <u>R</u>	GCC - <u>D</u> <u>C</u> <u>R</u> <u>D</u> <u>L</u> <u>A</u> <u>R</u>	GAC - <u>D</u> <u>P</u> <u>G</u> <u>N</u> <u>L</u> <u>V</u> <u>R</u>	GGC - <u>D</u> <u>P</u> <u>G</u> <u>H</u> <u>L</u> <u>V</u> <u>R</u>	C	
	GTA - <u>Q</u> <u>S</u> <u>S</u> <u>S</u> <u>L</u> <u>V</u> <u>R</u>	GCA - <u>Q</u> <u>S</u> <u>G</u> <u>D</u> <u>L</u> <u>R</u> <u>R</u>	GAA - <u>Q</u> <u>S</u> <u>S</u> <u>N</u> <u>L</u> <u>V</u> <u>R</u>	GGA - <u>Q</u> <u>R</u> <u>A</u> <u>H</u> <u>L</u> <u>E</u> <u>R</u>	A	
	GTG - <u>R</u> <u>S</u> <u>D</u> <u>E</u> <u>L</u> <u>V</u> <u>R</u>	GCG - <u>R</u> <u>S</u> <u>D</u> <u>D</u> <u>L</u> <u>V</u> <u>R</u>	GAG - <u>R</u> <u>S</u> <u>D</u> <u>N</u> <u>L</u> <u>V</u> <u>R</u>	GGG - <u>R</u> <u>S</u> <u>D</u> <u>K</u> <u>L</u> <u>V</u> <u>R</u>	G	

Figure 20 : Corrélation entre le domaine de reconnaissance de l'ADN et le triplet de paires de bases ciblé. Les acides aminés des positions -1, 3 et 6 sont soulignés.

Pour envisager une utilisation des doigts de zinc dans un système de reconnaissance, on ne peut se limiter à l'utilisation d'un seul doigt de zinc. Dans l'objectif de cibler 17 paires de bases afin d'être sélectif d'un seul gène parmi tout le génome, il faut donc plusieurs doigts de zinc qui se succèdent. C'est ainsi que Gupta et son équipe ont cherché à optimiser les séquences de reconnaissance de systèmes à deux doigts successifs, partant du principe qu'il existe des interactions à l'interface des deux doigts permettant d'augmenter la reconnaissance des triplets<sup>47</sup> (Figure 21).

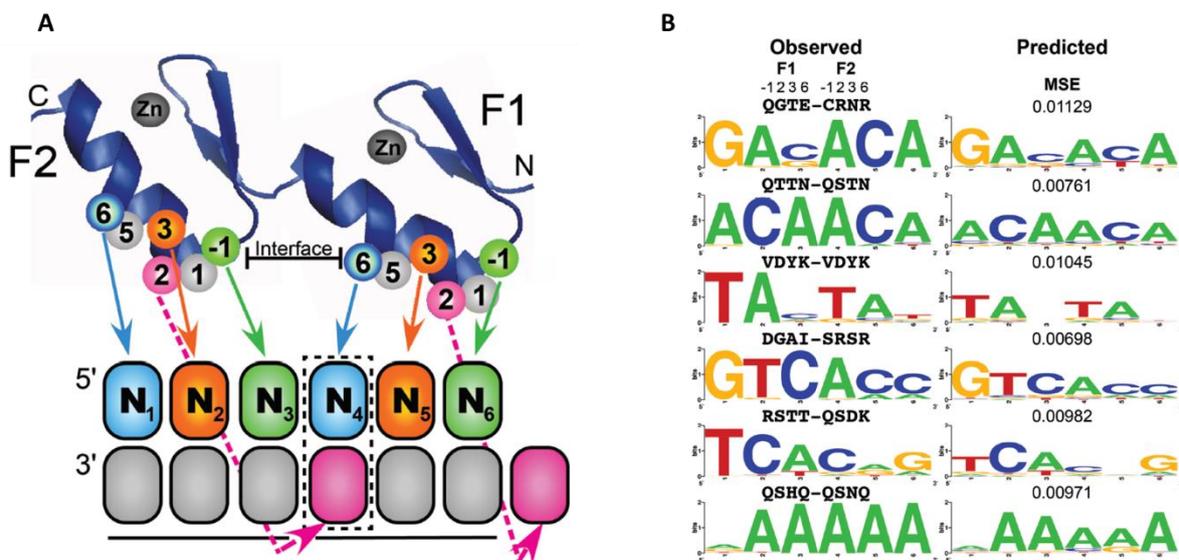


Figure 21 : A) Représentation schématique de la reconnaissance de six paires de bases par deux doigts de zinc adjacents. La paire de bases encadrée situe la contribution de l'acide aminé 2 du second doigt de zinc sur la reconnaissance du triplet du

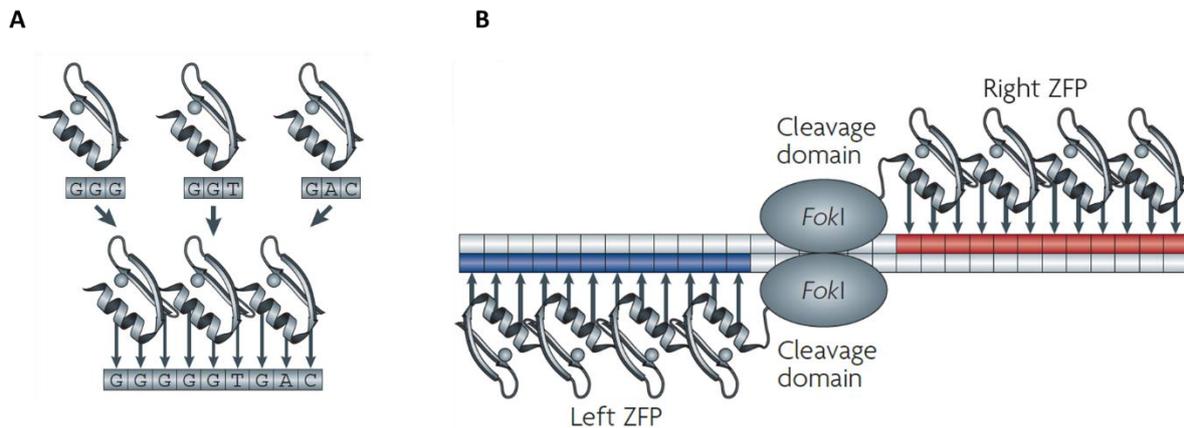
premier doigt de zinc<sup>48</sup>. B) Exemple de motifs à deux doigts de zinc avec le modèle de prédilection et les données de la base de données.

À partir des données existantes sur les doigts de zinc, ils ont développé des outils de modélisation qui prennent en compte davantage d'interactions, la séquence de reconnaissance d'un doigt influençant la reconnaissance de l'autre triplet. Les prédictions que donnent ces outils correspondent bien aux données recueillies sur des protéines à deux doigts de zinc. Pour des séquences d'ADN avec une guanine en position 5' les prédictions sont correctes, mais la reconnaissance de la thymine sur différentes positions, entre autres en début de séquence, reste difficile à modéliser.

La reconnaissance d'un triplet de bases de l'ADN repose donc sur les interactions que les acides aminés des positions -1, 2, 3 et 6 de l'hélice  $\alpha$  forment avec les bases nucléiques. Cependant, des études comparatives ont montré que les autres aa ont un rôle, non pas dans la reconnaissance, mais dans la formation du complexe doigt de zinc/ADN. C'est en prenant comme référence la protéine Sp1, un facteur de transcription connu pour sa forte affinité avec les régions de l'ADN riches en guanine, que Shi et son équipe ont comparé ses constantes d'affinités avec deux protéines synthétiques (Sp1C et Sp1C-2M)<sup>49</sup>. La protéine Sp1 est une protéine de 105 aa comportant 3 doigts de zinc, quant à Sp1C c'est une protéine ayant les mêmes aa essentiels à la reconnaissance, mais avec pas moins de 23 aa qui diffèrent de Sp1. Les constantes de dissociation mesurées pour ces deux protéines vis-à-vis de plusieurs cibles d'ADN ont montré qu'elles avaient des constantes de dissociation très similaires autour des 10 nM surtout sur les séquences du type 5'-NGGGNGGGN-3' (avec N= T, A, C, G). Pour savoir si cette affinité était seulement gouvernée pas les trois aa essentiels en position -1, 3 et 6, la protéine SP1C-2M a été synthétisée de manière à minimiser le second doigt de zinc en remplaçant les aa non essentiels (tous les aa sauf les deux Cys, les deux His, les 3 aa constituant le cœur hydrophobe et les 3 aa en position -1, 3 et 6 de l'hélice) par une alanine. L'affinité de cette protéine a été mesurée comme étant 400 fois plus faible que celle de Sp1. Bien que la cible soit reconnue, il faut prendre aussi en compte les interactions des autres acides aminés avec le duplex d'ADN. Sur cette même idée, Wang et Pabo se sont intéressés aux interactions protéine-protéine pour la formation d'un dimère autour de l'ADN par des protéines à doigt de zinc. Ceci sera développé dans le paragraphe suivant.

### 2.3.c Protéines ZFN

Une des applications des doigts de zinc en ingénierie moléculaire sont les Zinc Finger Nucleases (ZFN)<sup>50</sup>. Ces systèmes sont utilisés pour modifier un génome par insertion d'une séquence d'ADN double brin. Toutefois, l'insertion d'une nouvelle séquence a besoin d'être au niveau d'une position bien précise du brin d'ADN original. Il faut aussi pouvoir ouvrir et dégrader le double brin pour que le gène puisse être inséré. Les ZFN sont construites de la manière suivante : c'est un enchaînement de 3 à 6 doigts de zinc reconnaissant chacun un triplet de paires de bases au bout duquel une enzyme de restriction *FokI* capable de couper le double brin d'ADN est ajouté à une des extrémités. Les doigts de zinc vont permettre de cibler la séquence de l'ADN parent à couper (**Figure 22**).



**Figure 22 :** A) Association de 3 doigts de zinc reconnaissant 9 paires de bases du double brin d'ADN. B) Principe de l'utilisation de deux Zinc Finger Nuclease (ZFN) pour induire une coupure par l'enzyme *FokI* au niveau du domaine contenant l'activité nucléase au sein d'une séquence de 18 paires de bases (12 paires de bases reconnues par le doigt de zinc et 6 paires représentant la zone de coupure)<sup>50</sup>.

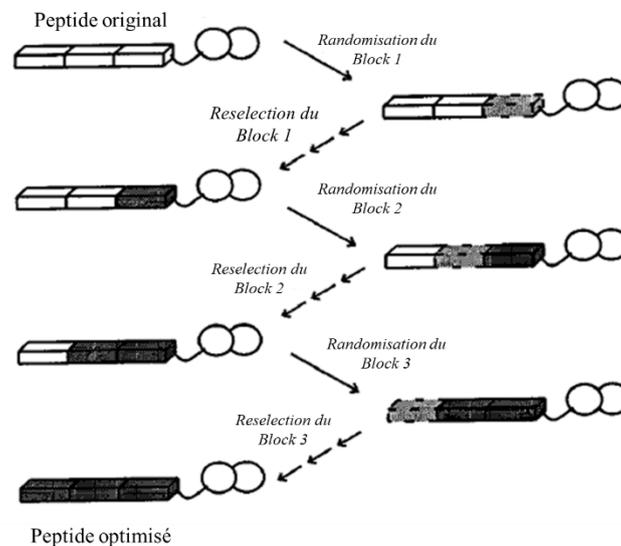
*FokI* est active uniquement sous la forme d'un dimère. Il faut donc avoir recours à deux ZFN qui s'associent pour réaliser la coupure, comme illustrée sur la **Figure 22**. Le premier ZFN à  $n$  doigts de zinc va reconnaître une séquence de  $3n$  paires de bases sur un brin en amont du site de coupure, le second reconnaîtra  $3n$  paires de bases, mais cette fois-ci sur le second brin d'ADN et ce, en aval du site de coupure. De cette manière les deux enzymes *FokI* vont être positionnés exactement au niveau du site de coupure au sein d'une séquence longue de 18 paires de bases. Le premier exemple de ZFN en ingénierie moléculaire a été publié en 2008 par Y. Santiago. Les ZFN ont été architecturées pour générer une coupure du gène *Dhfr*, codant pour une réductase dihydrofolate, dans les cellules des ovaires chez le hamster Chinois<sup>51</sup>.

#### 2.3.d Protéine Zif12

L'un des plus célèbres facteurs de transcription est le EGR-1 (Early growth response protein 1) aussi connu sous le nom de Zif268 (zinc finger protein 268) qui est composé de trois doigts de zinc ayant pour cible la séquence 5'-GCGTGGGCG-3'. Il est d'autant plus intéressant qu'il joue un rôle dans la lutte de l'organisme contre certains types de cancers et peut donc être une source d'inspiration pour le développement de nouveaux traitements<sup>52</sup>.

Wang et Pabo se sont intéressés à ce facteur de transcription il y a maintenant une vingtaine d'années dans le cadre de la compréhension des interactions protéine/protéine<sup>53</sup>. Ils ont cherché à concevoir une protéine à deux doigts de zinc capable de dimériser autour du duplex d'ADN cible formant un complexe 2:1. Cette protéine comprend les deux premiers doigts de zinc de Zif268 d'où provient son nom Zif12, ciblant la séquence palindromique 5'-TGGGCGCGCCCA-3'. Pour former un complexe sur cette séquence, il faut donc utiliser deux protéines Zif12, la séquence cible étant palindromique. Afin d'avoir un point de comparaison avec les systèmes 1:1, on ne s'intéresse pas à la constante de dissociation  $K_d$  dont l'unité est en  $M^2$ , mais à la concentration de protéines pour laquelle 50% de l'ADN a été complexé

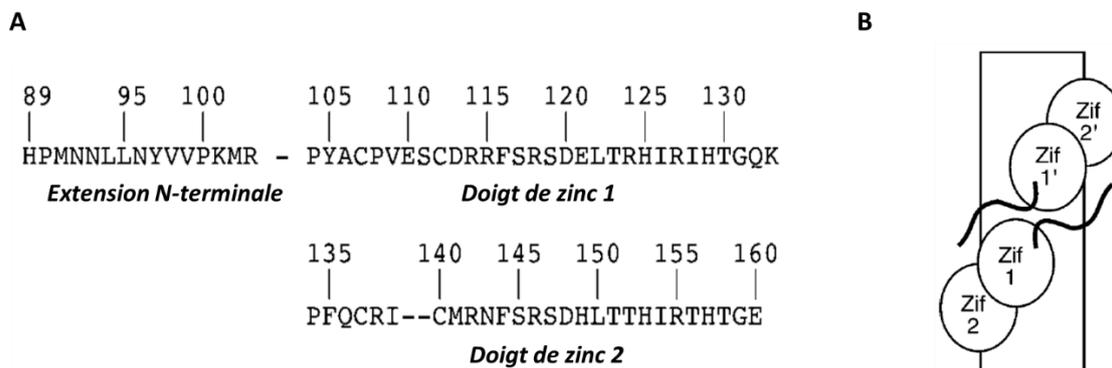
par les deux protéines, notée  $C_{50}$ . Dans le cas de la protéine Zif12 en interaction avec sa cible d'ADN, cette valeur de  $C_{50}$  est de  $9.6 \mu\text{M}$ . Malgré cette forte affinité, elles ne sont pas capables de former un dimère autour du duplex d'ADN pour des concentrations physiologiques de l'ordre de la dizaine de nM. Wang et Pabo ont alors cherché à ajouter aux doigts de zinc un domaine protéique permettant une reconnaissance mutuelle des monomères grâce à des interactions protéine/protéine. Pour cela, ils ont décidé d'ajouter une extension N-terminale de 15 acides aminés et de modifier sa séquence par bloc de 5 acides aminés. Chacun des trois blocs a été optimisé suivant le processus décrit **Figure 23** à l'aide de la technique de phage display. Brièvement, l'ADN codant pour les deux doigts de zinc a été inséré dans le gène codant pour une protéine de surface du phage. L'obtention de la séquence aléatoire pour le premier bloc est réalisée en faisant appel à des cassettes de mutagenèse. Par cette technique, les doigts de zinc sont présentés (display) à la surface des phages et la sélection des meilleurs candidats est alors réalisée par immobilisation de la cible (ici un ADN palindromique 5'-TGGGCGCGCCCA-3') sur un support solide. Les phages présentant les meilleures affinités sont retenus lors des différents lavages. Une fois décrochés du support, les phages sont cultivés en présence de bactéries pour produire suffisamment d'ADN en vue de son séquençage pour déterminer la séquence de la cassette correspondant au block 1. Ceci permet d'avoir un nouvel ADN à insérer dans le phage comportant les deux doigts de zinc avec le block 1 et une nouvelle cassette pour l'optimisation du second bloc. L'optimisation du dernier bloc suit la même procédure.



**Figure 23** : Schéma décrivant la procédure pour l'optimisation des trois blocs qui composent l'extension N-terminale de Zif12<sup>53</sup>.

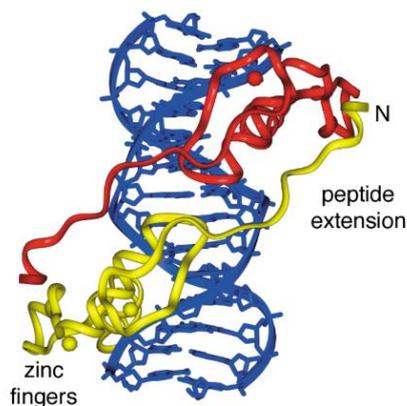
La séquence complète de cette extension en position N-terminale est : HPMMLLN<sub>Y</sub>VVPKMR. La nouvelle protéine Zif12 est donc composée de deux doigts de zinc issus des acides aminés 2-59 de Zif268 et de cette extension N-terminale (**Figure 24**). Le complexe formé, ADN-(Zif12)<sub>2</sub>, a pu être identifié par gel retard. La concentration  $C_{50}$  de Zif12 pour la séquence d'ADN 5'-TGGGCGCGCCCA-3' a été estimée à de 15 nM. L'ajout de cette extension N-terminale permet un

immense gain d'affinité d'un facteur 1000 environ. Pour rappel la concentration  $C_{50}$  sans cette extension est de  $9.6 \mu\text{M}$ .



**Figure 24 :** A) Séquence d'acides aminés composant les trois parties Zif12. B) Dessin simplifié du complexe ADN-(Zif12)<sub>2</sub> selon une orientation « tête à tête »<sup>54</sup>.

La résolution de la structure cristalline de Zif12 lié au duplex d'ADN 5'- ATGGGCGCGC CCAT-3' (**Figure 25**) apporte de nombreuses informations sur le rôle que l'extension de 15 aa joue sur la dimérisation<sup>54</sup>. Premièrement, les deux monomères sont disposés selon une orientation « tête à tête » sur l'ADN. Le complexe formé avec l'ADN présente lui un axe de symétrie C2 passant par le centre de l'ADN (perpendiculaire au plan de la feuille sur la **Figure 25**). Au sein de l'extension N-terminale deux acides aminés Lys<sup>101</sup> et Arg<sup>103</sup> présentent une interaction de type électrostatique avec les groupements phosphates de l'ADN qui améliore l'interaction du doigt de zinc avec l'ADN.

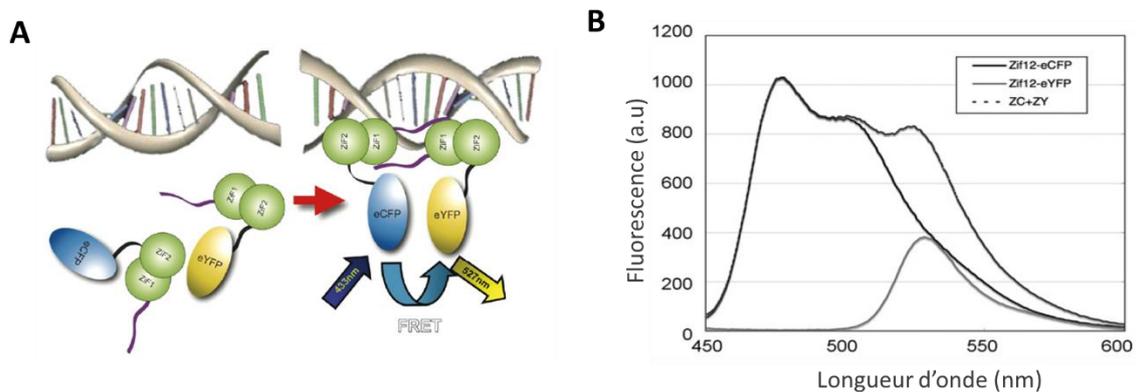


**Figure 25 :** Représentation 3D du complexe ADN-(Zif12)<sub>2</sub> publié par Wang et Pabo en 2001<sup>54</sup>.

Une information intéressante observée dans cette structure est que l'extension N-terminale d'un monomère s'éloigne de l'ADN pour interagir avec le second monomère. De nombreuses interactions entre chaînes hydrophobes sont observées entre l'extension N-terminale (ZF<sup>N</sup>) et le doigt de zinc 1, au niveau de la région où l'hélice  $\alpha$  et le premier feuillet  $\beta$  forment un cœur hydrophobe. Il y a aussi plusieurs liaisons hydrogène le long de l'extension avec les chaînes latérales des acides aminés qui composent le doigt de zinc 1. Tout ceci participe donc à la formation d'un complexe protéine/ADN 2:1

très stable, d'affinité plus forte que le complexe protéine/ADN 1:1. Cela ouvre aussi la possibilité d'utiliser l'extension N-terminale pour développer un système de deux protéines à doigts de zinc ciblant d'autres duplex de séquences différentes, palindromique ou non, dans l'optique de reconnaître sélectivement un gène. Cette extension n'ayant aucun rôle dans la reconnaissance de l'ADN cela signifie qu'en se basant sur la protéine Zif12 on peut modifier les acides aminés en position -1, 2, 3 et 6 de l'hélice  $\alpha$  du doigt de zinc pour changer la séquence d'ADN cible sans perturber l'agencement du complexe avec l'ADN.

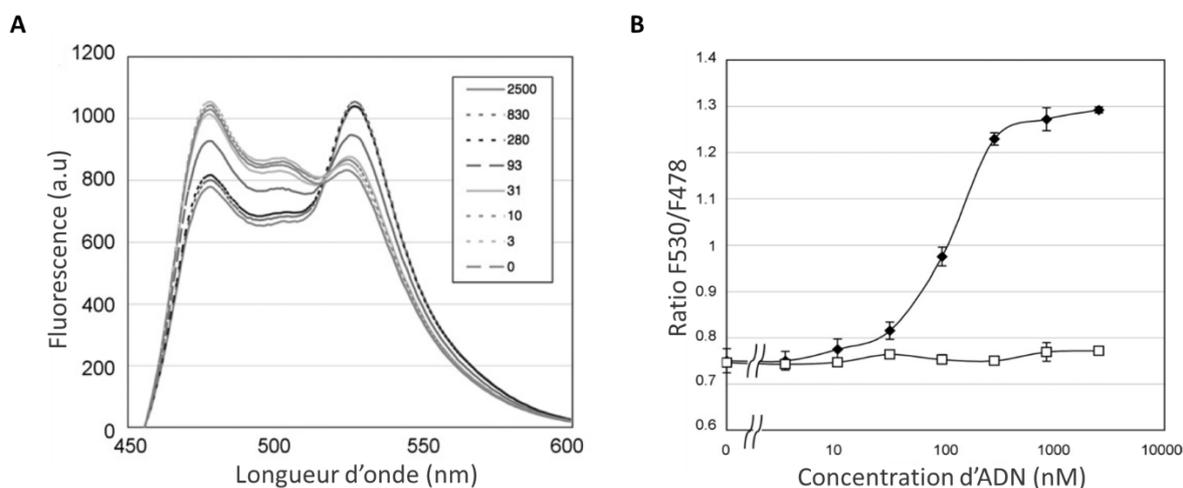
Cette protéine Zif12 a été la base de travail de l'équipe d'Hiroshi Ueda pour le développement d'une sonde fluorescente pour l'ADN<sup>55</sup>. Le système de fluorescence choisi repose sur un système de FRET entre deux protéines fluorescentes, plus précisément, entre une protéine eCFP (Cyan) et eYFP (Yellow). La séquence cible d'ADN utilisée est la séquence palindromique 5'-TGGGCGCGCCCA-3' reconnue par Zif12. Sur la base de deux protéines Zif12, un monomère possède à son extrémité C-terminale la protéine eCFP et l'autre monomère la protéine eYFP, elle aussi en position C-terminale. Ainsi lors de la dimérisation un changement dans le spectre de fluorescence provenant du transfert d'énergie de la protéine eCFP vers eYFP doit avoir lieu (**Figure 26 A**).



**Figure 26** : A) Principe de la sonde (Zif12-eCFP/Zif12-eYFP) décrite par l'équipe Hiroshi Ueda pour la détection d'ADN double brin<sup>55</sup>. B) Spectres d'émission de fluorescence des protéines Zif12-eCFP et Zif12-eYFP.

C'est à partir d'un mélange de ces deux protéines Zif12-eCFP/ Zif12-eYFP, avec des concentrations respectives de 120 nM et 600 nM, soit un ratio 1:5, que les auteurs ont pu démontrer un changement du spectre d'émission en présence de la cible d'ADN. Le ratio d'émission de fluorescence des deux protéines (Zif12-eYFP/ Zif12-eCFP) en absence d'ADN cible est de 0.75, après ajout de plus de 1 000 nM ( $\approx 8$  éq) d'ADN cible il passe à 1.27. Ce changement est lié au transfert d'énergie entre les deux partenaires devenus proches en interagissant avec l'ADN cible. L'émission du donneur (Zif12-eCFP) à 478 nm diminue et l'émission de l'accepteur (Zif12-eYFP) à 530 nm augmente (**Figure 27**). Cependant plusieurs choses sont à noter. La première est l'utilisation d'un mélange non stœchiométrique de ces deux protéines pour optimiser la réponse en fluorescence. La cible d'ADN est une séquence palindromique, ce qui signifie que la formation d'un homodimère autour de l'ADN est possible.

Avec cinq équivalents de la protéine Zif12-eYFP la formation d'un complexe 2:1 avec deux protéines Zif12-eYFP devrait être le produit majoritaire. Ceci doit avoir une influence non négligeable sur le spectre d'émission obtenu. La seconde chose à noter est le fort recouvrement spectral entre les deux protéines et la faible variation du rapport d'émission de fluorescence (YFP/CFP). Cependant la sélectivité de ce système a été démontré lors de l'ajout d'une séquence d'ADN ne comportant que la demi-cible, où une seule des deux protéines Zif12 peut se fixer (**Figure 27**). Le rapport d'émission reste constant alors qu'en présence la cible d'ADN celui-ci augmente. Cette étude semble montrer que la reconnaissance par FRET du palindrome 5'-TGGGCGCGCCA-3' est possible mais la détection doit fortement être améliorée pour envisager une application dans un milieu biologique (extrait cellulaire ou cellule, bien différent d'une solution où les concentrations de protéines et leurs concentrations peuvent être maîtrisées parfaitement pour optimiser le signal.



**Figure 27:** A) Spectre de fluorescence d'un mélange 1:5 entre le donneur (120 nM Zif12-eCFP) et l'accepteur (600 nM Zif12-eYFP) en présence de différents ajouts d'ADN cible (noté en nM). B) Ratio d'émission de fluorescence de Zif12-eCFP (12 nM) et Zif12-eYFP (15 nM) en fonction de la concentration en ADN. En diamant noir la courbe qui correspond à l'ADN cible, en carré blanc la courbe correspond à l'utilisation d'un ADN non-cible<sup>55</sup>.

### III - La détection par luminescence

#### 1 - La luminescence

Le terme de luminescence a pour la première fois été employé en 1888 par un physicien allemand, Eilhardt Wiedemann, bien que les premières observations du phénomène de fluorescence aient été reportées en 1845 par John Herschel à propos d'une solution de quinine placée sous la lumière du soleil<sup>56</sup>. La luminescence est définie par l'émission de photons (ou lumière) provenant d'une substance (cristal, molécule, protéine ...). Cette émission résulte d'une excitation de la molécule, à travers les états électroniques du composé suite à une exposition à la lumière<sup>57</sup>. Le terme luminescence en réalité comprend deux catégories, la fluorescence et la phosphorescence. Cette distinction repose sur l'état excité d'où provient l'émission de lumière. Si elle provient d'un état singulet alors on parle de fluorescence. Si elle provient d'un état triplet, c'est alors de la phosphorescence.

Les temps de vie des émissions en fluorescence et phosphorescence sont très différents. Ils sont de l'ordre de la nanoseconde (1 à 10 ns) pour la fluorescence alors que pour la phosphorescence, ils se situent autour de la microseconde jusqu'à la milliseconde. Ces différences de temps de vie s'expliquent par les états de spin de l'état fondamental et celui de l'état excité : les transitions entre fonction d'onde de spins différents sont interdites. L'état fondamental d'une molécule appelé  $S_0$ , est un état singulet. Un état excité singulet ( $S_1$ ) est un état dans lequel l'électron qui a été promu dans cet état est apparié au second électron de l'état fondamental. Autrement dit, lors de la promotion de cet électron dans l'état excité il y a conservation de l'état de spin. La conséquence de ceci est un retour de cet électron, dans son état fondamental qui se fait sans changement de spin et est donc rapide puisqu'il s'agit d'une transition autorisée par l'état de spin. Un état triplet, quant à lui, correspond à un état excité où l'électron qui a été promu n'est plus apparié à celui de l'état fondamental. Il présente le même état de spin ce qui implique que le retour à l'état fondamental devra passer par une inversion de spin, qui est une transition théoriquement interdite. C'est de là que provient le temps de vie particulièrement long pour la phosphorescence.

L'ensemble des processus et états de la molécule provenant de l'absorption d'énergie à l'émission de lumière sont illustrés plus en détail, dans le diagramme de Jablonski-Perrin (**Figure 28**).

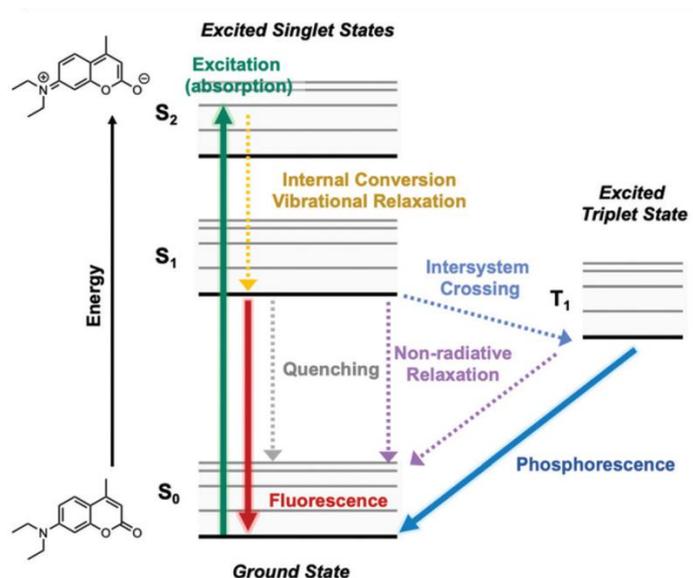


Figure 28 : Diagramme de Jablonski-Perrin détaillé <sup>58</sup>.

Partant de l'état fondamental  $S_0$ , l'absorption d'énergie lumineuse par la molécule amène celle-ci dans un état excité  $S_1$ . Cet état singulet peut revenir à son état fondamental et générer de la fluorescence. Cependant une transition non radiative vers un état triplet noté  $T_1$  est possible *via* un phénomène de transition de spin nommé Croisement Inter-Système (CIS), le retour à l'état fondamental  $S_0$  générant une émission de lumière de type phosphorescence. Si l'on s'intéresse aux spectres d'émission d'un fluorophore, celui-ci présentera toujours un décalage avec le spectre d'absorption vers des longueurs d'onde de plus basse énergie. Cela s'explique par le fait que l'énergie d'émission est inférieure à

l'énergie d'absorption. L'excitation peut amener à peupler des états vibrationnels supérieurs du niveau  $S_1$  mais ceux-ci sont dépeuplés au profit du plus bas état vibrationnel de  $S_1$ . La transition de fluorescence ramène la molécule dans un état vibrationnel supérieur du niveau  $S_0$ . Ainsi, l'énergie de la transition de fluorescence est plus faible que celle de l'absorption. Ce décalage entre la longueur d'onde maximale d'absorption et d'émission est appelé décalage de Stokes.

Le temps de vie de l'émission d'un fluorophore est un des paramètres importants qui définit un fluorophore. Un autre paramètre tout aussi important est le rendement quantique. Ce rendement ( $\Phi$ ) est le rapport de la quantité de photons qui a été émis (fluorescence ou phosphorescence) à la quantité de photons absorbés (1). Un troisième paramètre important est la brillance, produit du coefficient d'absorption molaire par le rendement quantique (2). Plus le rendement quantique est élevé et plus la brillance du système sera élevée.

$$\Phi = \frac{\text{nombre de photons émis}}{\text{nombre de photons absorbés}} \quad (1)$$

$$B = \Phi * \epsilon \quad (2)$$

Il existe également des voies de désexcitation non radiative (ou transition non radiative) n'impliquant pas d'émission de photon et qui se traduisent par une perte d'énergie. Cette énergie peut aussi être dissipée thermiquement par des vibrations, des rotations (isomérisation) ou des collisions entre molécules mais elle peut être aussi dissipée par transfert d'énergie vers une autre molécule, appelée accepteur, qui passe alors à son tour dans un état excité.

Une des grandes applications actuelles de la fluorescence pour la biologie est la microscopie de fluorescence, qui permet d'étudier des cellules ou des organismes entiers. Son utilisation *in vivo* est cependant limitée par les longueurs d'onde d'excitation et d'émission des composés et l'épaisseur de matière traverser. Le milieu biologique absorbe fortement dans la zone de l'UV et dans le visible jusqu'à 600 nm. Il existe cependant une fenêtre dans laquelle cette absorption est moindre<sup>59</sup>. On l'appelle la fenêtre de diagnostique, elle s'étend de 600 nm à 1300 nm (**Figure 29**). Mesurer une émission dans cette zone va permettre de minimiser l'absorption par le milieu, ainsi on obtient un signal plus intense. Le milieu absorbe moins dans cette zone ce qui permet au rayonnement d'aller plus en profondeur, entre 1000 et 1500  $\mu\text{M}$ .

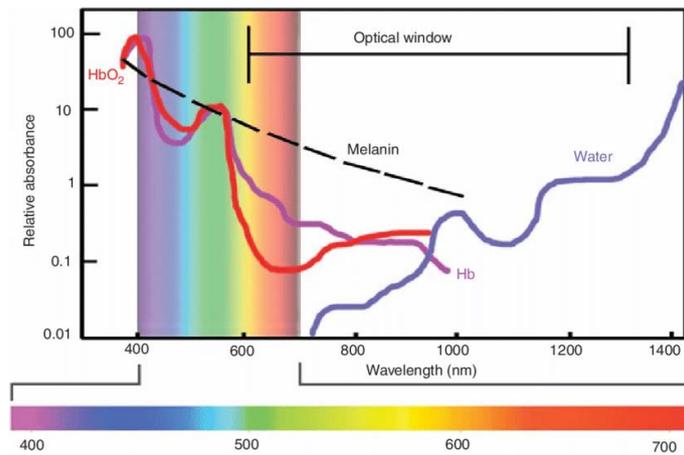


Figure 29 : Absorption du milieu biologique et représentation de la fenêtre de diagnostique entre 600 et 1300 nm <sup>59</sup>.

## 2 - Transfert d'énergie

### 2.1 - Définition et mécanisme

Comme nous venons de le dire, le processus de désexcitation peut se produire par transfert d'énergie vers un accepteur. Les premières études sur un système comportant un donneur et un accepteur d'énergie ont été réalisées par Théodore Förster en 1948. Le terme FRET désigne un processus de transfert d'énergie non radiatif entre deux partenaires, donneur et accepteur, qui peut se faire à partir d'un état excité singulet ou triplet du donneur et qui a lieu à travers l'espace et s'appuie sur des interactions dipôle-dipôle entre les partenaires. Plusieurs prérequis sont nécessaires pour qu'un système soit considéré comme un système FRET : (i) la paire donneur-accepteur doit être proche dans l'espace ( $<100 \text{ \AA}$ ) ; (ii) le spectre d'émission du donneur doit recouvrir en partie le spectre d'absorption de l'accepteur<sup>60</sup>. En opposition au transfert de type Förster, David L. Dexter a étudié un transfert qui repose sur de fortes interactions orbitales. Par un échange concerté d'électrons, un transfert d'énergie a lieu. Cependant l'efficacité de ce transfert diminue de manière exponentielle avec la distance entre les partenaires. Il n'est possible que pour de très courtes distances, inférieures à  $5 \text{ \AA}$ .

De manière générale, le donneur (D) possède deux électrons dans l'orbitale moléculaire de plus haute énergie (HOMO). L'énergie générée par une excitation UV va être absorbée permettant de promouvoir un électron dans l'orbitale moléculaire non peuplée de plus basse énergie (LUMO). C'est lorsque l'électron revient à son état fondamental que le transfert d'énergie s'effectue. Dans le cas d'un transfert de type Förster, de manière simultanée un électron provenant de l'état fondamental de l'accepteur (A) est promu dans sa LUMO<sup>61</sup> par couplage dipolaire. A contrario dans le cas d'un transfert de type Dexter, c'est un échange d'électron qui permet le transfert d'énergie. La résultante de tous ces transferts est que le donneur n'émet pas de photon lors du retour à l'état fondamental. C'est l'accepteur qui *via* l'énergie qui lui a été transmise en revenant à son état fondamental génère un photon (**Figure 30**).

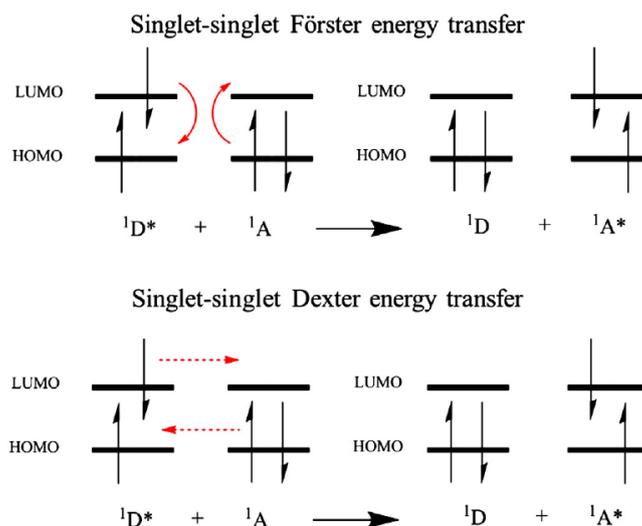


Figure 30 : Principe du transfert d'énergie de type FRET et Dexter <sup>62</sup>.

La condition pour qu'un tel transfert puisse avoir lieu est que le spectre d'émission du donneur recouvre en partie le spectre d'excitation de l'accepteur (**Figure 31**). Plus le recouvrement, noté  $J$ , sera important, plus le transfert d'énergie sera efficace.  $J$  est donné par l'équation (3) :

$$J(\lambda) = \frac{\int_0^\infty F_D(\lambda) \varepsilon_A(\lambda) \lambda^4 d\lambda}{\int_0^\infty F_D(\lambda) d\lambda} \quad (3)$$

avec  $F_D$  le spectre d'émission du donneur corrigé normalisé,  $\varepsilon_A$  le coefficient d'absorption molaire de A (accepteur) à la longueur d'onde  $\lambda$  (nm).

Étant donné qu'il s'agit d'un transfert intermoléculaire dans l'espace, la distance  $r$  qui sépare le donneur et l'accepteur est un facteur crucial. On définit le rayon de Förster, noté  $R_0$ , comme la distance à laquelle le transfert d'énergie est efficace à 50%. Il est spécifique d'une paire donneur-accepteur dans un solvant donné. Il peut être calculé selon l'équation (4).

$$R_0 = 0.211 [k^2 n^{-4} \Phi_{DO} J(\lambda)]^{\frac{1}{6}} \quad (4)$$

avec  $k$  le facteur d'orientation (souvent considéré à  $2/3$ ),  $n$ , l'indice de réfraction du milieu et  $\Phi_{DO}$ , le rendement quantique du donneur en l'absence d'accepteur.

L'efficacité du transfert d'énergie, notée  $E$ , est fonction de  $r$ , est définie par :

$$E = \frac{R_0}{R_0 + r^6} \quad (5)$$

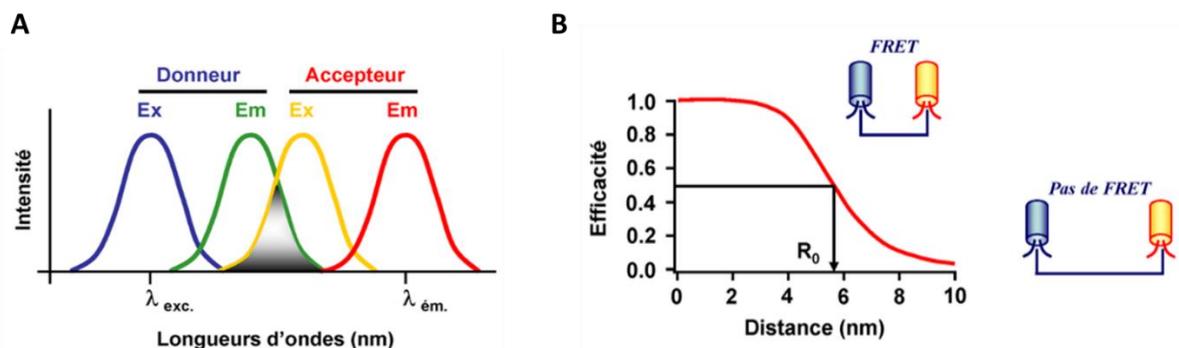


Figure 31 : A) Principe du recouvrement entre les spectres d'émission et d'excitation des partenaires FRET. B) Influence de l'efficacité du transfert en fonction de la distance donneur/accepteur.

Il existe plusieurs grandes stratégies pour développer un système de détection d'un analyte basé sur le FRET. Elles sont regroupées en quatre catégories (Figure 32). La première est un système « Turn-on » : l'accepteur a un niveau excité trop haut pour que le transfert d'énergie soit possible mais sa réaction avec l'analyte induit un changement de structure, qui permet d'abaisser suffisamment son niveau excité pour que le transfert d'énergie provenant du donneur devienne possible. Sur ce même principe un système « Turn-off » correspond au processus inverse où l'état excité de l'accepteur, initialement assez bas pour accepter l'énergie du donneur, est augmenté par réaction avec l'analyte, rendant impossible le transfert d'énergie. Les deux autres catégories reposent sur la modulation de la distance qui sépare les deux partenaires. Dans un cas, la distance est augmentée en présence de la cible, diminuant son émission. Dans l'autre cas, une diminution de la distance va permettre d'augmenter l'efficacité du transfert d'énergie et par conséquent l'émission de l'accepteur.

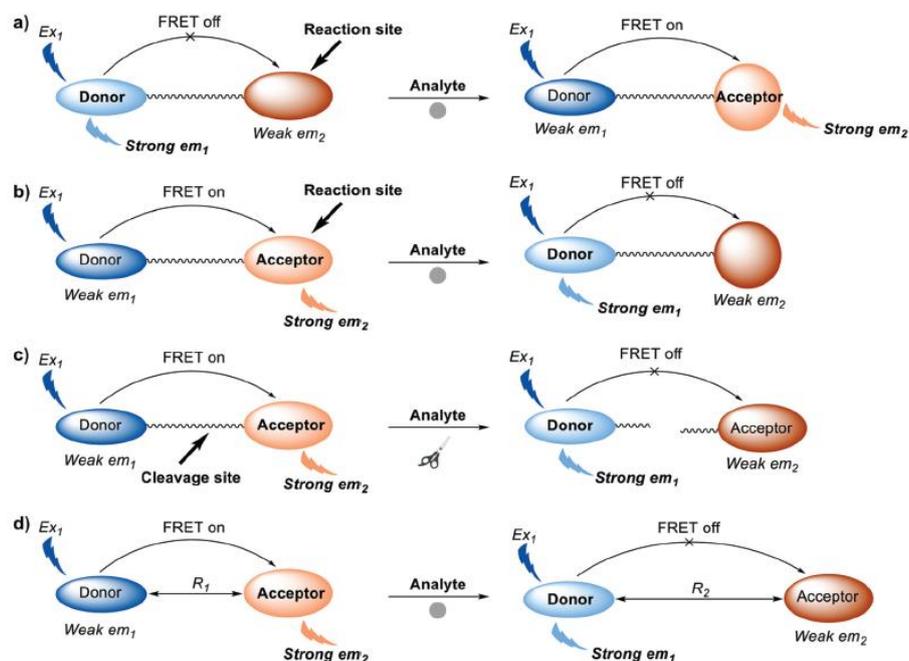
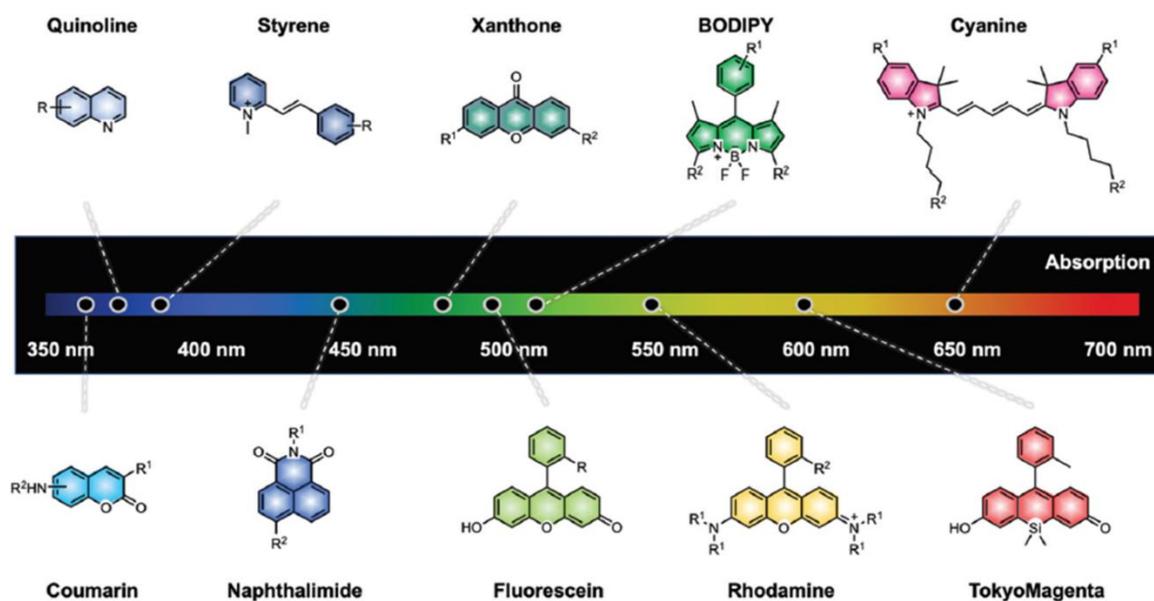


Figure 32 : Représentation des quatre stratégies les plus communes pour l'élaboration de sonde basée sur un système de FRET. a) Système turn-on par altération de la structure. b) Système turn-off par altération de la structure. c) Système turn-off par

coupure du lien entre les partenaires FRET. d) Système turn-off par agrandissement de la distance entre les partenaires FRET. Ex représente la longueur d'onde d'excitation et em la longueur d'onde maximale d'émission. R1 et R2 sont les distances qui séparent les deux fluorophores <sup>60</sup>.

Le nombre de fluorophores organiques utilisés pour l'imagerie ne cesse de grandir de par la facilité de synthèse que certains présentent. Sur la **Figure 33** sont représentées les grandes familles de fluorophores organiques les plus utilisées. Le groupement R correspondant aux positions sur lesquelles la plupart des différentes fonctionnalisations sont effectuées. La famille qui fait l'objet du plus grand nombre d'utilisations est la famille des fluorescéines puisqu'elle a été approuvée par la FDA (Food and Drug Administration), ce qui permet d'envisager son utilisation chez l'Homme. Les dérivés du BODIPY sont généralement les composés qui possèdent la plus forte brillance. Cependant le faible « déplacement de Stokes » (décalage entre l'absorbance et l'émission) observé pour cette famille rend ces composés sujets au self-quenching. Les rhodamines semblent être de bons candidats pour une émission dans la fenêtre de diagnostique. Leur principal avantage réside dans le fait de pouvoir facilement moduler leurs propriétés photo-physiques en introduisant des groupements électrodonneur / électroattracteur dans leur structure.



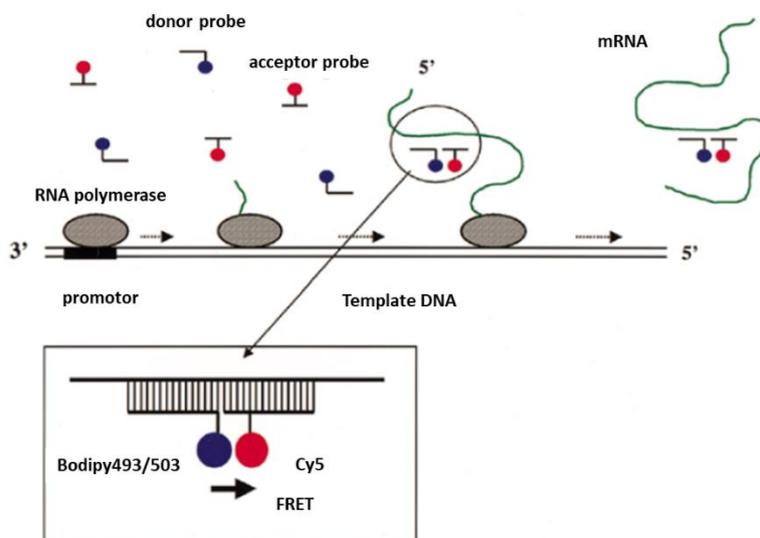
**Figure 33** : Liste des principales familles de fluorophores classées selon la structure et la longueur d'onde d'absorption croissante. Le groupement R représente les positions de fonctionnalisations les plus courantes <sup>58</sup>.

## 2.2 - Un exemple de système FRET entre fluorophores organiques

Le recours à un système de FRET a permis le développement de sondes ratiométriques. C'est un moyen simple à mettre en œuvre ne nécessitant pas de spectrophotomètres ou microscopes onéreux pour la détection. Pour qu'une sonde soit ratiométrique, il faut que son spectre d'émission évolue suite à la liaison / réaction avec l'analyte <sup>63</sup>. Il est alors possible de suivre la détection en enregistrant l'émission à deux longueurs d'onde différentes. Plus ces deux longueurs d'onde sont différentes, plus la résolution est grande. Il existe deux moyens pour cela. Le premier consiste à ajouter une référence interne

fluorescente qui est inerte vis-à-vis de la présence de la cible. Ainsi on obtient un ratio entre l'intensité du signal émis par la sonde et celle de la référence interne. Le second moyen est de disposer de deux entités fluorescentes ( $F_1$  et  $F_2$ ) émettant à deux longueurs différentes ( $\lambda_1 \neq \lambda_2$ ) au sein de la sonde. Dans le cas d'un transfert d'énergie de  $F_1$  à  $F_2$ , on s'intéresse à l'évolution des deux émissions. L'émission du donneur  $F_1$  va diminuer au profit d'une augmentation de l'émission de l'accepteur  $F_2$ .

Il existe de nombreux exemples de système FRET entre deux fluorophores organiques reliés l'un à l'autre conçus pour détecter un analyte. Les travaux de Tsuji sur le développement d'un système de détection en temps réel d'ARN par un transfert FRET entre deux chromophores est un parfait exemple<sup>64</sup>. Ce système repose sur l'utilisation de deux oligonucléotides (15 mers) pour la reconnaissance de l'ARN, par hybridation avec la séquence complémentaire au brin. Pour la détection, un des brins est fonctionnalisé par la Bodipy493/503 qui joue le rôle de donneur d'énergie et le second brin porte la Cy5, accepteur dans ce système. Lors de la transcription de l'ADN par l'ARN polymérase *in vivo*, ces deux oligonucléotides vont s'hybrider à l'ARNm sur deux sites adjacents. L'hybridation des deux oligonucléotides permet de rapprocher les deux chromophores et de réaliser un transfert FRET. Par cette méthode, l'équipe de Tian a été en mesure de suivre en temps réel par fluorescence l'élongation de cet ARNm (**Figure 34**).



**Figure 34** : Principe du transfert FRET entre deux chromophores greffés sur deux oligonucléotides pour la détection en temps réel de la synthèse d'ARNm.

## IV - Les lanthanides

### 1 - Définition

#### 1.1 - Généralités

Le terme de « lanthanide » désigne une famille de métaux aux numéros atomiques croissants allant de  $Z=57$  à  $Z=71$ , appartenant au bloc f du tableau périodique des éléments (**Figure 35**). Cette famille est

composée de 15 éléments allant du lanthane au lutécium et souvent nommée « terres rares » lorsque l'on inclut le scandium et l'yttrium. Ce terme ne se réfère pas à la rareté des éléments puisque leur abondance est proche de celle des métaux de transition, mais fait écho de leur disparité sur Terre. Principalement sous le degré d'oxydation +III, ils présentent une configuration électronique  $[\text{Xe}]4f^n$ . Par la suite, la notation générique  $\text{Ln}^{3+}$  sera utilisée pour désigner les cations lanthanides (III). Le lanthane n'a aucun électron dans la sous-couche 4f et le lutécium (III) ( $\text{Lu}^{3+}$ ), quant à lui, en possède 14, ce qui sature cette sous-couche.

Le tableau périodique des éléments est présenté avec les lanthanides (éléments Ce à Lu) insérés dans une boîte séparée en dessous des éléments de la 6ème période. Les lanthanides sont colorés en vert foncé. Les autres éléments sont colorés par groupes : métaux alcalins (orange), alcalin-terreux (rouge), métaux de transition (jaune/vert), métalloïdes (vert clair), non-métaux (bleu/bleu foncé), et gaz rares (bleu clair).

**Figure 35 :** Tableau périodique des éléments selon la classification de Mendeleïev. Les lanthanides sont représentés en vert foncé.

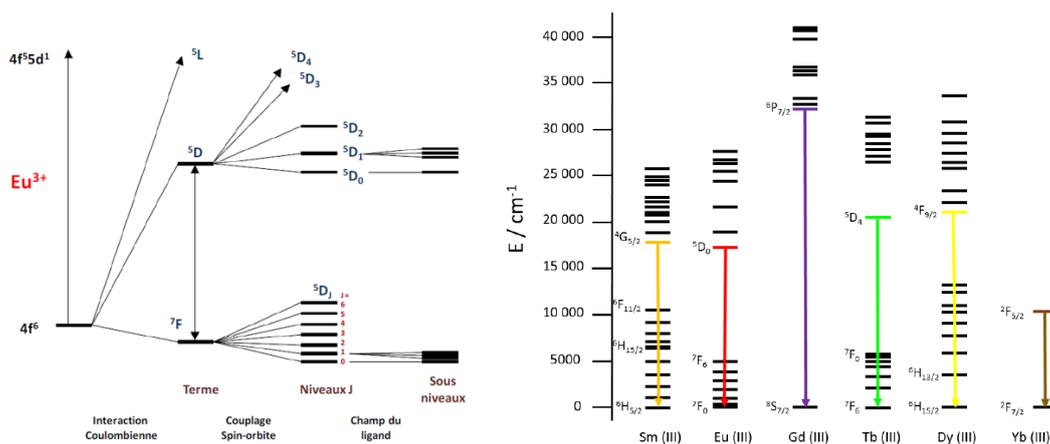
## 1.2 - Configuration électronique

L'un des aspects les plus importants de la structure électronique des  $\text{Ln}^{3+}$  est la localisation des électrons de valence dans les orbitales 4f qui sont moins étendues que les orbitales 5s et 5p qui sont pleines. Ainsi, ces dernières, écrantent les électrons 4f, contrairement à ce qui est observé avec les métaux d, dont les orbitales de valences sont les plus étendues. Par conséquent les propriétés électroniques des  $\text{Ln}^{3+}$  ne sont que très peu influencées par leur environnement et le champ cristallin n'aura qu'un effet mineur sur l'éclatement des niveaux 4f. Un autre phénomène associé à leur structure électronique est la contraction des lanthanides. Il se définit par la modification du rayon ionique en fonction de la charge nucléaire. Plus elle est forte, plus le rayon ionique est faible, cela va de 102 pm pour le cérium ( $Z = 58$ ) à 86.1 pm pour le lutécium ( $Z = 71$ ).

Les niveaux électroniques des  $\text{Ln}^{3+}$  sont très nombreux. Cette multiplicité est cruciale pour le développement de systèmes photoluminescents, chaque ion lanthanide possédant des niveaux d'énergies qui lui sont propres. À partir de la configuration électronique ( $4f^n$ ) et sous l'influence de la répulsion électronique, ces orbitales sont divisées en terme spectroscopique noté  $(2S+1)\Gamma$ . Le terme  $2S+1$  correspondant à la multiplicité de spin et  $\Gamma$  est la lettre qui correspond au moment angulaire orbitalaire

total, ici L. Chacun de ces termes vont être dégénérés sous l'effet du couplage spin-orbite aboutissant à des niveaux spectroscopiques notés  $^{(2S+1)}\Gamma_J$  (avec  $J = L+S$ ).

L'exemple de l'ion Eu(III) est présenté **Figure 36**. Sa structure électronique lui confère 6 électrons dans l'orbitale 4f. La répulsion électronique entre les électrons de l'orbitale 4f (interactions coulombiennes) divise l'orbitale 4f en deux termes notés  $^5D$  et  $^7F$ . Le couplage spin-orbite éclate ensuite chacun de ces termes pour donner les niveaux spectroscopiques  $^7F_0$  à  $^7F_6$  et  $^5D_0$  à  $^5D_5$ . L'ensemble des niveaux énergétiques des principaux lanthanides utilisés en imagerie est présenté **Figure 36**.



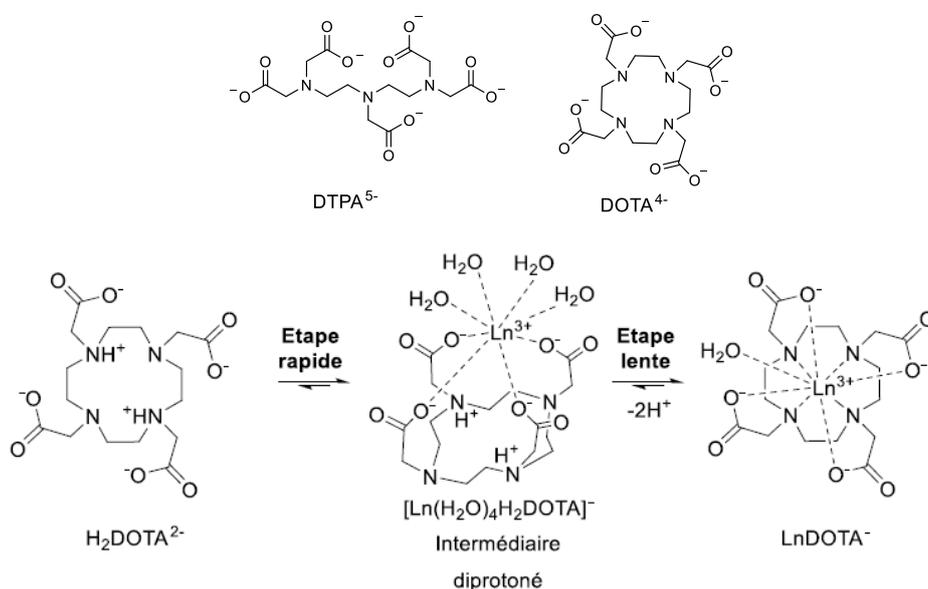
**Figure 36** : Gauche: Diagramme d'énergie partiel de l'ion  $\text{Eu}^{3+}$ <sup>65</sup>. Droite: Diagramme d'énergie d'ions lanthanides luminescents et leurs niveaux d'énergies, image adaptée de J. C. Bünzli et C. Piguet<sup>66</sup>.

### 1.3 - Coordination des lanthanides

De par leur configuration électronique, les  $\text{Ln}^{3+}$  présentent une chimie de coordination quasi identique. Elle se caractérise par la formation de liaisons purement électrostatiques avec les ligands qui s'organisent autour du Ln de manière à minimiser les interactions stériques répulsives. Les lanthanides sont définis comme des espèces dites « dures » selon la classification de Pearson, formant des liaisons avec des ligands durs tels que l'eau, des ligands azotés ou oxygénés<sup>67</sup>. La sphère de coordination varie entre 8 et 10 selon la taille du  $\text{Ln}^{3+}$ , de manière générale le nombre de coordination est de 9. Le degré d'oxydation le plus stable pour l'ensemble de la série est le degré +III, mais  $\text{Eu}^{3+}$  et  $\text{Yb}^{3+}$  peuvent être réduits sous la forme +II ( $E^\circ_{\text{Eu}^{3+}/\text{Eu}^{2+}} \approx -0.35 \text{ V}$  et  $E^\circ_{\text{Yb}^{3+}/\text{Yb}^{2+}} \approx -1.05 \text{ V}$  dans l'eau)<sup>68</sup>. Les autres  $\text{Ln}^{3+}$  sont plus difficilement réductibles.

Pour une application en milieu biologique la coordination des  $\text{Ln}^{3+}$  est un facteur primordial pour limiter leur toxicité. Citons en exemple le gadolinium, lanthanide amplement utilisé comme agent de contraste pour l'imagerie IRM. Celui-ci présente une forte toxicité à l'état libre<sup>69</sup>. Sa dose létale médiane ( $\text{LD}_{50}$ ) n'est que de 0.35 mmol/kg pour du  $\text{GdCl}_3$ . Son action est principalement sur la balance calcique puisque c'est un bio-isostère de l'ion calcium. Présent en faible quantité il peut induire rapidement des perturbations de la signalisation cellulaire, troubles musculaires, etc. La toxicologie des autres lanthanides semble être similaire puisqu'ils ont des propriétés chimiques semblables. Par conséquent les

lanthanides sont considérés comme toxiques à l'état libre. Pour la complexation des lanthanides, on distingue deux grandes familles, les ligands linéaires et les ligands macrocycliques de type polyazacycloalcane. En considérant comme exemple le ligand DTPA, linéaire, et le DOTA, macrocyclique, présentés **Figure 37**, la cinétique de complexation du gadolinium est plus élevée pour le ligand DTPA dont sa flexibilité lui permet de facilement s'enrouler autour du  $Gd^{3+}$  pour former le produit thermodynamique en quelques secondes/minutes alors qu'il faut plusieurs heures pour le DOTA<sup>70</sup>. Ceci s'explique par la formation d'un intermédiaire déprotoné (produit cinétique) où l'ion  $Gd^{3+}$  se retrouve au-dessus de la cavité macrocyclique *via* la complexation avec les quatre atomes d'oxygène des bras acétates. En milieu suffisamment basique les amines du cycle vont se déprotoner pour permettre au Ln de s'internaliser, c'est précisément cette étape qui est lente à cause de la rigidité du macrocycle (**Figure 37**).



**Figure 37** : Ligand DTPA<sup>5-</sup> et DOTA<sup>4-</sup>. Mécanisme de formation d'un complexe Ln-DOTA<sup>70</sup>.

Concernant la constante de stabilité thermodynamique des complexes<sup>71</sup>, bien que le ligand DOTA bénéficie d'une contribution entropique favorable grâce à sa pré-organisation (effet macrocyclique), elle est similaire à celle du LnDTPA ( $\log K_{GdI} \approx 24$  pour le DOTA et 22 pour le DTPA)<sup>72</sup>. Le dernier paramètre déterminant à prendre en compte est l'inertie cinétique qui résulte de la dissociation du complexe sous l'influence de différents facteurs, essentiellement une compétition avec d'autres métaux ou chélatants (transmétallation ou transchélation). En milieu acide (HCl 0.1M) la dissociation du complexe LnDTPA est quasi immédiate, la protonation d'un des carboxylates entraînant la protonation d'une des amines ce qui libère le cation  $Ln^{3+}$ . La pré-organisation du DOTA avec cette cavité où se situe l'ion  $Ln^{3+}$  rend les amines du macrocycle plus difficilement protonables. Les complexes du DOTA présentent une inertie très grande, rendant l'échange du  $Ln^{3+}$  avec d'autres chélateurs difficile<sup>73</sup>. C'est donc pour ces raisons qu'il est intéressant pour une application *in vivo* d'utiliser des dérivés cycliques inspirés du DOTA pour former des complexes de lanthanides stables en tout point.

## 2 - La luminescence des lanthanides

### 2.1 - Emission des lanthanides

La plupart des  $\text{Ln}^{3+}$  sont luminescents. Les transitions électroniques au sein des lanthanides sont différenciées en trois types : (i) transition 4f-4f (ii) transition 4f-5d (iii) un transfert de charge (MLCT ou LMCT). L'émission selon les transitions f-f est la plus généralement observée. Les règles de Laporte indiquent que les transitions de mêmes orbitales sont interdites, les transitions 4f-4f sont donc interdites. En conséquence, celles-ci sont de faibles intensités et possèdent un coefficient d'absorption molaire très faible de l'ordre de 1 à 10  $\text{L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$ . Tous les lanthanides ne présentent pas la même intensité de luminescence. Le rendement intrinsèque d'un lanthanide est directement lié à la différence d'énergie entre le niveau fondamental et les niveaux excités. Plus cette différence est élevée plus les phénomènes de désexcitation non radiative sont faibles ce qui augmente son intensité de luminescence. L'émission des lanthanides se répartit de l'UV au proche infrarouge (**Figure 38**). Le terbium ( $\text{Tb}^{3+}$ ) émet dans le vert et l'euprimum ( $\text{Eu}^{3+}$ ) émet dans le rouge. Ce sont ceux que l'on retrouve le plus souvent pour des applications de microscopie cellulaire. L'ytterbium ( $\text{Yb}^{3+}$ ) et le néodinium ( $\text{Nd}^{3+}$ ) ont une émission dans l'infrarouge à 978 nm et 1 060 nm, respectivement. Ils suscitent de plus en plus d'intérêt pour l'imagerie médicale puisque leur émission se situe où les photons ont la plus grande pénétration dans les tissus<sup>74</sup>. La configuration électronique des lanthanides, et plus particulièrement l'écran que forment les électrons de la couche 5s et 5p, permet d'avoir des transitions f-f à longueur d'onde fixe, quel que soit le ligand. De plus, du fait de l'interdiction des transitions f-f par les règles de Laporte, la désexcitation des  $\text{Ln}^{3+}$  est lente. Il en résulte des temps de vie de luminescence de l'ordre de la microseconde à la milliseconde, qui sont particulièrement longs en comparaison de la fluorescence organique ayant des temps de vie de l'ordre de la nanoseconde.

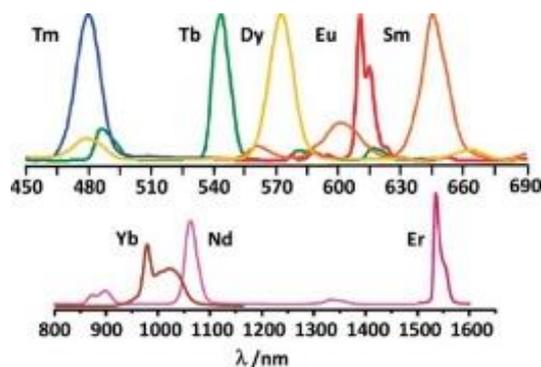
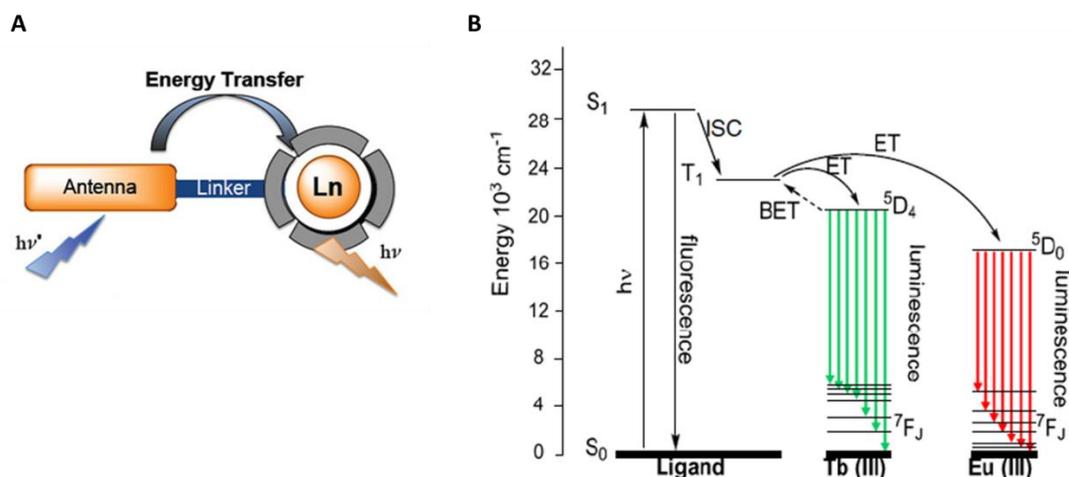


Figure 38 : Spectre d'émission des lanthanides luminescents.

### 2.2 - Effet d'antenne

L'excitation directe d'un  $\text{Ln}^{3+}$  est difficile car les transitions f-f ont de faibles coefficients d'absorption molaire ( $\epsilon < 10 \text{ M}^{-1}\text{cm}^{-1}$ ). Un des moyens pour sensibiliser efficacement un lanthanide est de pouvoir transférer de l'énergie de manière indirecte *via* un composé organique qui lui en revanche possède un

fort coefficient d'absorption molaire. Ce processus d'excitation, découvert en 1942 par Weissman<sup>75</sup>, est appelé « effet antenne », il est présenté **Figure 39**. Le chromophore va pouvoir absorber la lumière d'excitation, promouvant un électron de son état fondamental  $S_0$  vers un état excité  $S_1$ . Cet électron va alors être transféré vers un état donneur de la molécule (triplet  $T_1$ , singulet transfert de charge). Deux processus rentrent alors en compétition : (i) une désexcitation radiative du chromophore (phosphorescence) (ii) un transfert d'énergie vers un état accepteur (A). La proximité de ce chromophore avec le lanthanide va permettre de sensibiliser le lanthanide qui joue le rôle d'accepteur final d'énergie. Une désexcitation radiative lumineuse va amener le lanthanide à son état fondamental. Dans le cas de  $\text{Eu}^{3+}$ , les niveaux  $^5\text{D}_1$  et  $^5\text{D}_2$  sont les états accepteurs et une désexcitation par conversion interne peuple le niveau  $^5\text{D}_0$  qui est le niveau émetteur. Pour le  $\text{Tb}^{3+}$ , seul le niveau émissif  $^5\text{D}_4$  est peuplé par transfert d'énergie.



**Figure 39** : A) Principe de l'effet antenne figure tiré de la publication <sup>76</sup>. B) Diagramme de Jablonski illustrant l'effet d'antenne <sup>76</sup>.

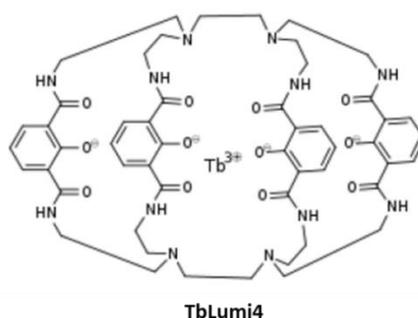
Le transfert d'énergie du ligand vers le lanthanide peut être de type Dexter ou Förster (I.2-1). Lorsque le chromophore se trouve dans la sphère de coordination du lanthanide, celui-ci présente une bien meilleure sensibilisation. Il existe cependant plusieurs phénomènes qui viennent en compétition avec la luminescence du lanthanide tels que : (i) des désexcitations non radiatives (par vibration OH, CH, NH ou de collisions), (ii) un transfert retour du lanthanide vers l'antenne, si le niveau donneur de l'antenne est trop proche du niveau émissif de  $\text{Ln}^{3+}$  ( $\Delta E < 2\,000 \text{ cm}^{-1}$ ) et (iii) un transfert d'énergie vers un chromophore.

### 3 - Complexes luminescents de lanthanides pour des applications en biologie

Les propriétés photophysiques des complexes de  $\text{Ln}^{3+}$  les rendent très intéressants pour les applications biologiques.

- (1) Du fait de leur spectre d'émission avec des bandes fines à des longueurs d'onde fixes, ils sont facilement identifiables.
- (2) L'émission de certains lanthanides dans le vert, le rouge ou le proche infrarouge les place dans une fenêtre spectrale adaptée.
- (3) A cause de l'effet antenne, la différence entre la longueur d'onde d'excitation et l'émission est grande, ce qui évite les phénomènes de ré-absorption.
- (4) Leur long temps de vie permet d'éliminer aisément la fluorescence du milieu biologique en utilisant une détection résolue en temps.

Tout ceci a poussé au développement d'une chimie de coordination très riche afin d'optimiser la stabilité des complexes de  $\text{Ln}^{3+}$  et leur propriétés de luminescence via l'antenne pour des applications en milieu biologique<sup>77</sup>. Dans beaucoup de cas, la sensibilisation du  $\text{Ln}^{3+}$  a été décrite comme passant par l'état triplet de l'antenne. Le groupe de Parker a été très actif dans le domaine en se basant sur des ligands macrocycliques de la famille DOTA. On peut citer également le complexe TbLumi4 de Raymond<sup>78</sup> qui présente un rendement quantique de 50% avec un coefficient d'absorption de  $27\,300\text{ M}^{-1}\text{cm}^{-1}$  à 340 nm, soit une brillance de plus de  $13\,000\text{ M}^{-1}\text{cm}^{-1}$  tout en ayant un temps de vie très élevé mesuré à 2.45 ms (**Figure 40**). Celui-ci est utilisé dans de nombreux test biologiques commercialisés par la société CisBio®.

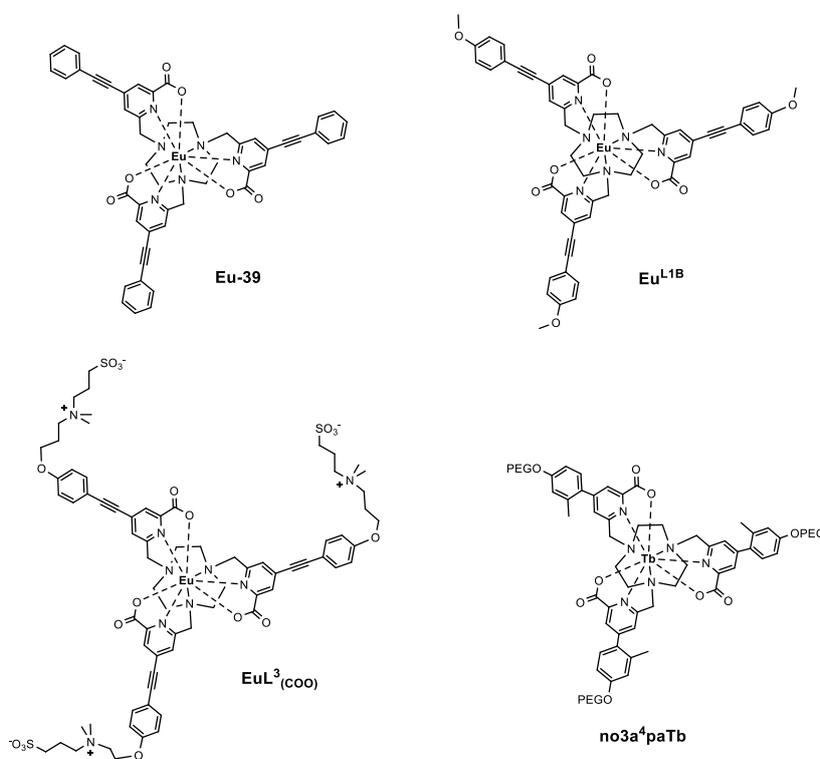


**Figure 40 :** Complexe de  $\text{Tb}^{3+}$ , TbLumi4<sup>78</sup> commercialisé par la société CisBio®.

Le groupe de Maury a également contribué au domaine en développant des complexes avec des antennes à transfert de charge (CT)<sup>79</sup>. Contrairement à ce qui avait été proposé pour d'autres antennes, la sensibilisation du  $\text{Ln}^{3+}$  ne provient pas d'un transfert d'énergie depuis l'état triplet de l'antenne mais depuis l'état CT (singulet) de l'antenne. De manière générale un état CT est présent dans une molécule lorsqu'elle présente de part et d'autre un groupement électroattracteur et électrodonneur reliés par un système  $\pi$  conjugué. Il y a donc une polarisation de la molécule qui présente un moment dipolaire. Ce type d'antenne est couramment appelé antenne « push-pull ». Cette délocalisation permet de déplacer l'absorption du système vers le visible. Les modifications de ces antennes reposent sur les groupements électrodonneur/attracteur et la longueur du système  $\pi$  les séparant. Ainsi lorsque le groupement électrodonneur est plus fort ( $\text{OR} < \text{SR} < \text{NR}_2$ ) l'absorption de l'antenne va être décalée vers des

longueurs d'ondes de plus basses énergies (vers le rouge). Ce phénomène est identique lorsque la force du groupement électroattracteur augmente ( $\text{COOR} < \text{SO}_2\text{R} < \text{NO}_2$ ). Quant au système  $\pi$  celui-ci peut être allongé pour décaler, là encore, la longueur d'onde d'absorption vers le rouge. En jouant sur ces trois paramètres, on peut donc moduler le niveau énergétique du CT et par extension les propriétés spectroscopiques de l'antenne.

Takalo a décrit en 1997 un complexe d'euprium basé sur une structure macrocycle triazacyclononane (TACN) avec trois antennes picolinate substitués par un groupement phénylalcynyl<sup>80</sup> (**Figure 41**). Le ligand permet une coordination *in fine* de 9 autour du Ln. Les équipes de O. Maury et D. Parker s'en sont servi de modèle pour développer d'autres complexes avec des antennes « push-pull » qui possèdent différents groupements donneurs en bout d'antenne (alkyl, alkoxy, alkylthio ou encore dialkylamino, parfois en supprimant aussi la fonction alcyne). Ils ont été utilisés pour différentes applications biologiques, notamment pour la microscopie biphotonique car ces antennes « push-pull » peuvent présenter d'intéressantes propriétés d'absorption biphotonique<sup>81,82,83</sup>. L'avantage que confère l'utilisation de trois antennes est de considérablement augmenter le coefficient d'absorption du complexe et donc une augmentation de la brillance comparée à une simple antenne. Le complexe de terbium  $\text{no3a}^4\text{paTb}^{84}$  quant à lui n'utilise pas d'espaceur, mais la présence d'un groupement méthyle en position méta du groupement donneur (OPEG) permet une sensibilisation efficace du Ln avec un rendement quantique de plus de 70%. La brillance de ce complexe est supérieure à celle de TbLumi4, les différentes caractéristiques de ces complexes sont présentées **Tableau 2**.



**Figure 41** : Différents exemples d'antennes « push-pull » pour la sensibilisation du terbium et de l'euprium.

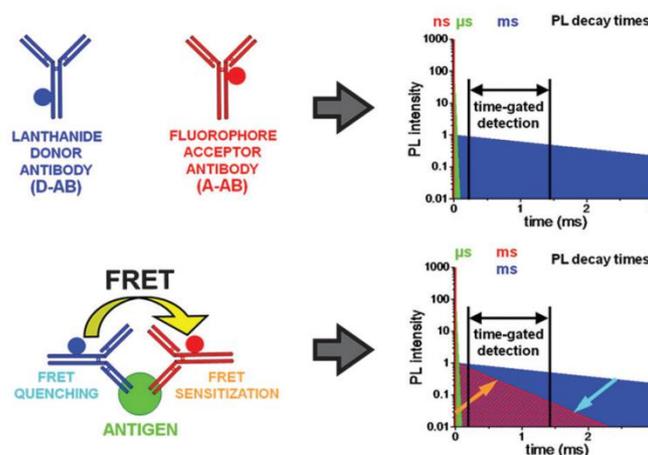
**Tableau 2 :** Caractéristiques des antennes « push-pull » présentées **Figure 41**.

	Brillance ( $B$ ) ( $M^{-1}cm^{-1}$ )	Epsilon ( $\epsilon$ ) ( $M^{-1}cm^{-1}$ )	$\lambda_{(ex)}$ (nm)	Rendement quantique ( $Q$ )
Eu-39 <sup>80</sup>	10 700	48 400	315	22% (H <sub>2</sub> O)
Eu <sup>L1B</sup> <sup>82</sup>	24 360	58 000	339	42 % (MeOH)
EuL <sup>3(COO)</sup> <sup>83</sup>	5 760	32 000	337	18% (H <sub>2</sub> O)
no3a <sup>4</sup> paTb <sup>84</sup>	22 500	30 400	302	74% (H <sub>2</sub> O, MeOH)

## 4 - Des exemples de sondes à base de lanthanides

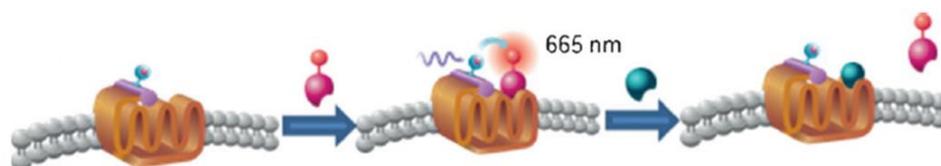
### 4.1 - Sondes reposant sur du FRET à l'aide de complexes de lanthanides

Les systèmes de fluorescence par FRET impliquant des lanthanides sont maintenant de plus en plus répandus. Dorénavant de nombreux kits sont commercialisés, notamment par la société française CisBio, qui utilise la détection en temps résolu pour divers applications<sup>85</sup>. Parmi ces applications on trouve le marquage de biomolécules<sup>86</sup> ou l'immunofluorescence en temps résolu. Le principe de l'immunofluorescence repose sur l'utilisation de deux anticorps ciblant un même antigène sur deux positions différentes. L'un des anticorps comporte un complexe de lanthanide comme donneur FRET, et le second porte un fluorophore (accepteur FRET). En présence de l'antigène, les deux partenaires du système FRET se rapprochent ce qui permet un transfert d'énergie du lanthanide vers le fluorophore (**Figure 42**). Cela permet d'observer la fluorescence de l'accepteur lors de la détection en temps résolu. Ce signal atteste de la présence de l'antigène et permet de le quantifier en fonction du rapport d'intensité de fluorescence du fluorophore et du lanthanide. L'équipe de L. J Charbonnière s'est basée sur ce système pour faire la détection d'antigène carcinoembryonnaire<sup>87</sup>, une glycoprotéine surexprimée dans plusieurs cancers.



**Figure 42 :** Principe de l'immunofluorescence en temps résolu. Deux anticorps fonctionnalisés par un complexe de lanthanide (donneur) et un fluorophore (accepteur) dont le transfert d'énergie est possible par la reconnaissance de l'antigène cible<sup>86</sup>.

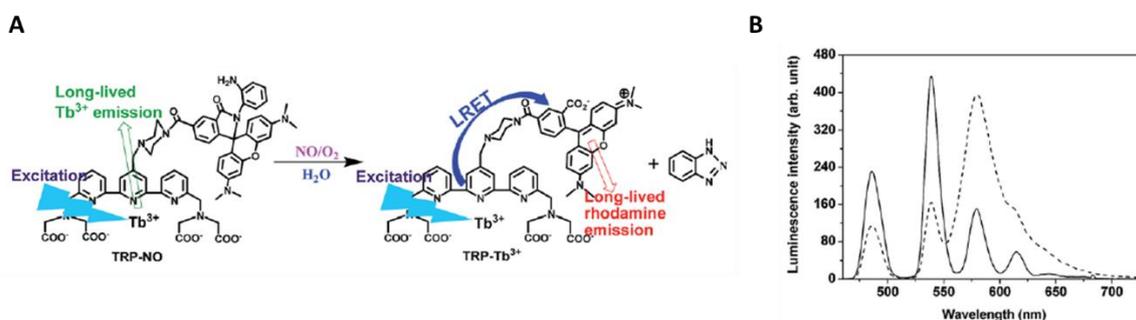
Un autre exemple de détection en temps résolu provient des travaux de Delbianco et Parker<sup>88</sup> pour l'étude de l'interaction du récepteur CCK2 (un GPCR, récepteur couplé aux protéines G) avec des molécules ligands (**Figure 43**). Cette équipe a utilisé la technologie SNAP-tag afin de conjuguer un complexe d' $\text{Eu}^{3+}$  au récepteur. Un agoniste CCK2 a été synthétisé avec un fluorophore accepteur d'énergie de  $\text{Eu}^{3+}$ . Le suivi du FRET de la paire  $\text{Eu}^{3+}$  / fluorophore permet de suivre le déplacement de l'agoniste par différents compétiteurs et de trouver et caractériser des ligands de CCK2.



**Figure 43** : Représentation du FRET en temps résolu de la fixation de l'agoniste CCK (rouge) sur le récepteur SNAP-tag CCK2 (orange). Le transfert FRET diminue en fonction de la compétition entre l'agoniste et la protéine d'intérêt CCK (turquoise)<sup>88</sup>.

Toutes ces sondes reposant sur un FRET avec un complexe de  $\text{Ln}^{3+}$  comme donneur sont particulièrement sensibles, notamment grâce à la détection résolue en temps.

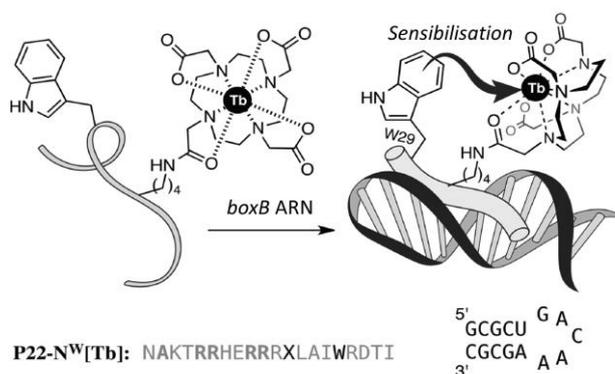
Un dernier exemple de détection en temps résolu, réalisé par l'équipe de Yuan, sur la détection d'oxyde nitrique présent dans les lysosomes<sup>89</sup>. La sonde utilisée, TRP-NO, est un complexe de  $\text{Tb}^{3+}$  relié à une forme non luminescente de la rhodamine (sous une forme spirolactame). La présence de NO induit une réaction au niveau de la rhodamine et permet de retrouver sa luminescence (ouverture du cycle spirolactame). L'excitation de la sonde, plus précisément le complexe de  $\text{Tb}^{3+}$ , génère un transfert d'énergie intramolécule entre le  $\text{Tb}^{3+}$  et la rhodamine. Lors de la détection en temps résolu, l'émission du  $\text{Tb}^{3+}$  diminue au profit de l'émission de la rhodamine, cela signifie que la rhodamine a acquis un temps de vie long, mais son émission provient du transfert d'énergie du  $\text{Tb}^{3+}$  qui lui possède un temps de vie long. Le rapport d'émission,  $I_{\text{rhodamine}}/I_{\text{Tb}^{3+}}$  obtenu, permet de quantifier la concentration en oxyde nitrique (**Figure 44**).



**Figure 44** : Principe de la sonde TRP-NO en présence d'oxyde nitrique. B) Spectre d'émission en temps résolu de TRP-NO avant (ligne pleine) et après ajout de NO (ligne pointillé)<sup>89</sup>.

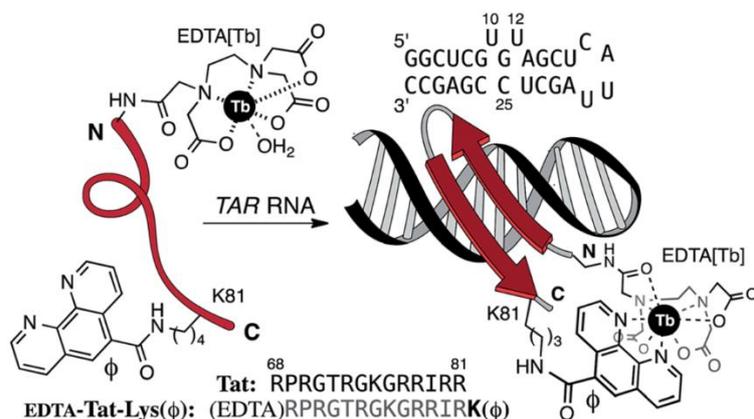
#### 4.2 - Détection d'oligonucléotides à l'aide de complexes de lanthanides

Il existe quelques exemples de ciblage d'ARNm, notamment dans l'équipe de M. E. Vázquez qui tire parti du repliement d'ARNm en épingle à cheveux pour pouvoir le cibler<sup>90</sup>. Ce système se compose d'un peptide de 20 acides aminés riche en arginine sur lequel un simple complexe de terbium, DOTA[Tb] est greffé. La sonde, noté P22-N<sup>W</sup>[Tb], comporte un tryptophane qui joue le rôle d'antenne. Elle est positionnée quelques acides aminés plus loin du complexe de terbium et absorbe dans l'UV (300 nm). Lorsque la sonde peptidique s'intercale dans l'ARNm cible (*boxB* RNA), cela va modifier la conformation de la sonde, ce qui rapproche le tryptophane du complexe de lanthanide (**Figure 45**). La conséquence est une augmentation de l'émission provenant du terbium lors du titrage en temps résolu, validant la formation d'un complexe 1:1 (sonde/ARN) avec un  $K_d$  calculé à 12 nM.



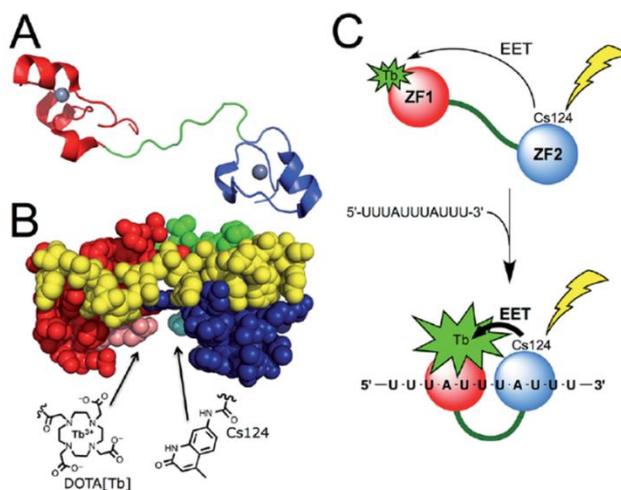
**Figure 45** : Mécanisme de reconnaissance. Le changement de conformation induit par la reconnaissance de l'ARN messager rapproche l'antenne du complexe de lanthanide. La séquence P22-N<sup>W</sup>[Tb] correspond à la sonde. X= Lys(DOTA-[Tb<sup>3+</sup>], *boxB* ARN étant la cible<sup>90</sup>.

Basé sur un principe similaire, une nouvelle sonde (EDTA[Eu]-Tat-Lys( $\phi$ )) ciblant un ARNm viral (*TAR* RNA) a été développée quelques années plus tard<sup>91</sup> par cette même équipe. Contrairement au système P22-N<sup>W</sup>[Tb] le macrocycle utilisé pour complexer le lanthanide est de l'EDTA (EDTA-[Tb]) qui offre six atomes coordinants au Ln<sup>3+</sup>, la sphère étant complétée par deux molécules d'eau. L'antenne est une phénanthroline qui absorbe à 300 nm. En absence de cible, les molécules d'eau coordonnées éteignent la luminescence du lanthanide. Lorsque l'ARN cible est complexé par la sonde, un changement de conformation a lieu et la sphère de coordination du lanthanide est modifiée : les molécules d'eau sont remplacées par la phénanthroline et la luminescence du Tb<sup>3+</sup> augmente, pour deux raisons : le rapprochement de l'antenne qui se retrouve directement coordonné au Tb<sup>3+</sup> et la suppression de la désactivation par les molécules d'eau (**Figure 46**).



**Figure 46** : Principe de la reconnaissance de EDTA[Ln]-Tat-Lys( $\phi$ ). Le repliement du peptide autour de la cible TAR ARN induit un changement dans la sphère de coordination du lanthanide<sup>91</sup>.

Enfin un dernier exemple de sonde à ARN, qui a été réalisé au sein de notre équipe, est celle développée par L. Raibaut<sup>92</sup> pour la détection sélective de séquences d'ARNm riche en AU. La sonde, LTIS<sup>Tb</sup>, est une protéine de 70 acides aminés qui comporte deux doigts de zinc inspirés de la protéine TIS11d impliquée dans la régulation d'ARNm. La sonde a été synthétisée par voie chimique, par assemblage de trois fragments par ligation native. Elle est fonctionnalisée de part et d'autre par un complexe de terbium DOTA[Tb] et par une antenne Cs124 absorbant à 330 nm. En présence de la cible d'ARN, 5'-UUUAUUUAUUU-3' la sonde se replie autour de l'ARN rapprochant l'antenne Cs124 du complexe de terbium ce qui permet une sensibilisation plus efficace et une augmentation de l'émission du Tb<sup>3+</sup> (**Figure 47**). La formation d'un complexe 1:1 a été vérifiée par titrage par luminescence en temps résolu et la constante d'affinité a été mesurée à 5.6 nM. D'autres séquences d'ARNm qui diffèrent d'une base nucléique (ex : 5'-UUUGUUUAUUU-3') ont aussi été testées mais aucune augmentation de luminescence n'a été observée, validant le caractère très spécifique de cette sonde.

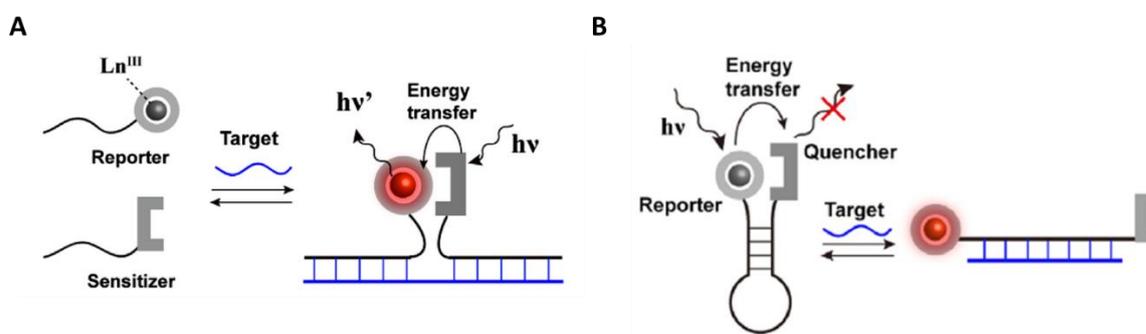


**Figure 47** : Structure de TIS11d. Les deux doigts de zinc (ZF1 et ZF2) sont représentés en rouge et bleu, l'ion Zn<sup>2+</sup> en gris et le lien les reliant en vert. B) Structure RMN en solution de TIS11d lié avec l'ARN UUAUUUAUU (jaune), en rose le complexe de terbium et en bleu l'antenne Cs124. C) Principe de la sonde LTIS<sup>Tb</sup> ciblant un ARN riche en AU<sup>92</sup>.

Les systèmes développés précédemment présentent de fortes affinités pour les séquences d'ARNm ciblées. Des exemples de systèmes FRET pour cibler cette fois-ci une séquence d'ADN simple brin où une seule base diffère de l'ARNm (la thymine remplace l'uracile) seront discutés dans le prochain paragraphe.

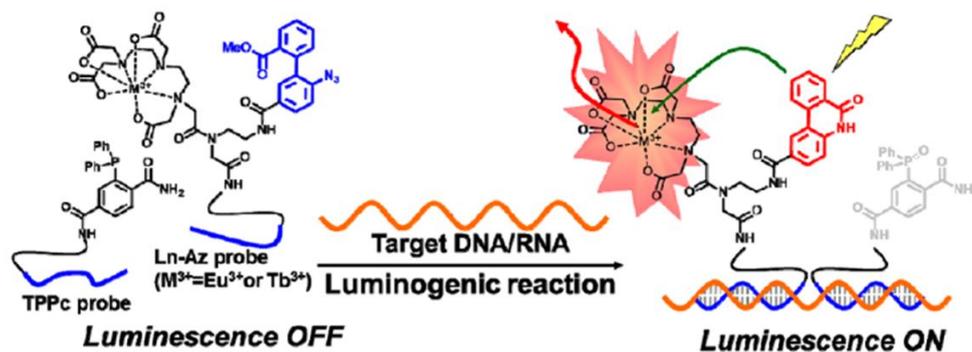
#### 4.3 - Détection d'ADN simple brin

Il existe plusieurs sondes luminescentes à base de  $\text{Ln}^{3+}$  pour la détection d'ADN<sup>93</sup>. Ces sondes peuvent être regroupées en deux catégories : les sondes qui reconnaissent des simples brins d'ADN de manière séquence-spécifique et les sondes qui reconnaissent un double brin par intercalation de l'antenne mais sans reconnaissance de séquence. La première catégorie de sonde est basée sur l'hybridation au simple brin cible (i) d'un brin portant un complexe de  $\text{Ln}^{3+}$  ( $\text{Tb}^{3+}$  ou  $\text{Eu}^{3+}$ ) et un autre portant une antenne ou (ii) d'un MB où le fluorophore organique a été remplacé par un complexe de  $\text{Ln}^{3+}$ <sup>94-97</sup> (**Figure 48**).



**Figure 48** : A) Principe du mécanisme de formation d'un complexe 2:1 ciblant un simple brin d'ADN. L'hybridation des deux oligonucléotides permet au transfert d'énergie d'avoir lieu. B) Représentation schématique d'un MB comportant un complexe de lanthanide ciblant un simple brin d'ADN<sup>93</sup>.

Concernant les sondes par hybridation, un exemple élégant a été décrit par Abe<sup>98</sup>. Dans ce système c'est une sonde par réaction qui est utilisée. Le premier oligonucléotide composant la sonde, Ln-Az, comporte un complexe de lanthanide EDTA[Ln] et à proximité une forme inactive de l'antenne qui comporte un azoture. Le second oligonucléotide sonde porte une phosphine (TPPc = triphénylphosphine carboxamide). Lors de l'hybridation des deux oligonucléotides avec l'ADN cible simple brin, la TPPc va être proche de l'azoture et la réaction de Staudinger peut avoir lieu entre la phosphine et l'azoture. L'amine obtenue est impliquée dans un processus de cyclisation intramoléculaire pour générer une phénanthridinone, qui est la forme active de l'antenne. Dès lors, un transfert d'énergie est possible entre l'antenne et le complexe de lanthanide (**Figure 49**). C'est un système « turn-on », totalement non-luminescent en l'absence de la cible.

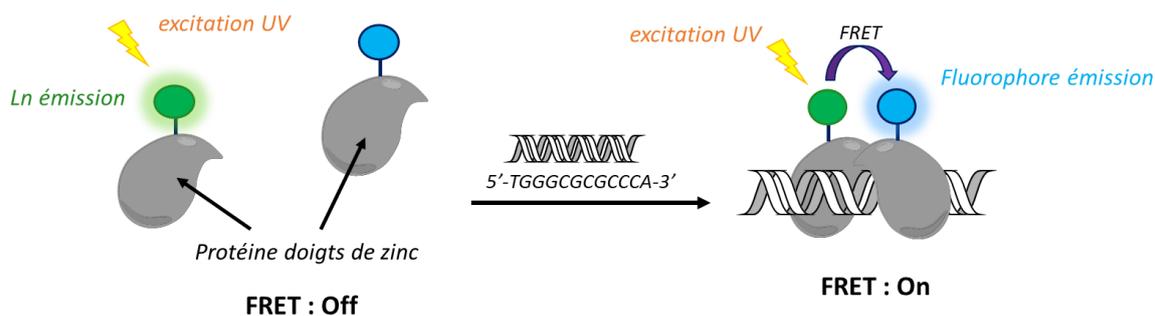


**Figure 49** : Principe de reconnaissance du simple brin d'ADN par les deux sondes Ln-Az et TPPc. La reconnaissance de l'ADN/ARN cible permet à la réaction de Staudinger d'avoir ce lieu qui induit une réponse lumineuse<sup>98</sup>.

## V - Le projet de thèse

### 1 - Contexte

L'objectif de ce projet de thèse est de développer un système de sonde lumineuse pour la reconnaissance spécifique d'une séquence d'ADN. On considère qu'une longueur d'au moins 16 paires de bases est unique à l'échelle du génome humain constitué de 3 milliards de paires de bases. Dans un premier temps, notre objectif est de reconnaître 12 paires de bases. Notre système va reposer sur l'utilisation de protéines à doigts de zinc pour leurs propriétés de reconnaissance de l'ADN double brin et sur un système FRET pour la détection. Nous avons retenu le système décrit par Pabo où deux protéines à deux doigts de zinc sont capables de dimériser autour de la séquence cible d'ADN qui a déjà été utilisée par Ueda dans une approche de sonde FRET codée génétiquement. Le système de Ueda reconnaît seulement un palindrome et les propriétés spectroscopique liées au FRET sont décevantes. La sonde est difficilement utilisable (la moitié des complexes formés sont des homodimères improductifs). Il nous semble qu'il serait possible d'améliorer ces propriétés optiques en utilisant un FRET basé sur les Ln<sup>3+</sup>, généralement très sensibles. La reconnaissance de la séquence d'ADN pourra être mise en avant grâce à la diminution d'intensité de luminescence du complexe de Ln<sup>3+</sup> donneur au profit de l'émission du fluorophore organique accepteur ou par observation du temps de vie du donneur (**Figure 50**). Enfin, il conviendra de modifier le système de reconnaissance afin qu'il soit capable de reconnaître autre chose qu'un palindrome, trop limité en ce qui concerne les applications possibles.



**Figure 50** : Principe général du système de sonde par deux protéines à doigts de zinc. Un monomère est fonctionnalisé par le complexe de lanthanide (vert) et l'autre par le fluorophore accepteur (bleu).

Les premiers travaux effectués sur ce projet ont été réalisés par le Dr Guillaume Frémy lors de sa thèse soutenue en 2019. Son objectif principal était de valider l'approche pour la détection du palindrome utilisé par Wang et Pabo 5'-TGGGCGCGCCA-3'. Une grosse partie du travail de Guillaume a été de synthétiser et caractériser des complexes de  $Tb^{3+}$  et  $Eu^{3+}$  aux propriétés optiques intéressantes (excitation au-delà de 300 nm et bon rendement quantique) et surtout conjugables à des peptides ou protéines. Guillaume a exploré la synthèse par voie chimique de la protéine Zif12 décrite par Pabo, fonctionnalisée par ces complexes ou des fluorophores organiques. A la toute fin de sa thèse, une preuve de concept a été obtenue sur le palindrome d'ADN. Ces résultats sont décrits dans le paragraphe suivant.

## 2 - Travaux précédents

### 2.1 - Stratégie de synthèse des peptides fonctionnalisés

Le premier travail de Guillaume a été de choisir une position dans la séquence de Zif12 où conjuguer le complexe de  $Ln^{3+}$  ou le chromophore. La **Figure 51** présente la séquence de Zif12. Plusieurs acides aminés devaient être conservés :

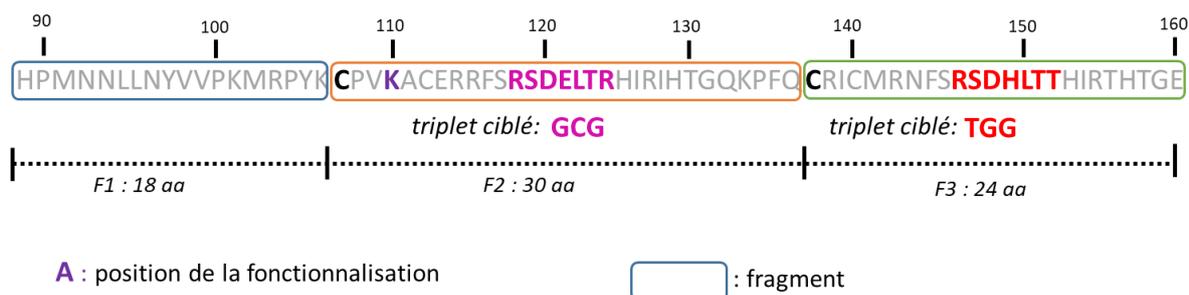
- Ceux marqués d'un triangle sont essentiels pour le repliement du doigt de zinc ;
- Ceux marqués d'un carré interviennent dans la reconnaissance des bases de l'ADN ;
- Ceux marqués d'un point participent à des interactions protéines/protéines entre les deux Zif12 au sein du dimère formé autour de l'ADN, comme le révèle la structure cristallographique obtenue par Wang et Pabo<sup>54</sup>.

La position 110 a été retenue pour la conjugaison, correspondant à une distance d'environ 40 Å entre les chromophores si le dimère de protéines fonctionnalisées se forme bien sur l'ADN.

90 95 100 105 110 115 120 125 130 135 140 145 150 155 160  
 HPMNLLLN<sup>.</sup>YVVPKMRPYACP<sup>Δ</sup>VESC<sup>Δ</sup>DRRF<sup>.</sup>SR<sup>Δ</sup>SDELTRHIRIHTGQKPFQCRICMRNFSRSDHLTTHIRHTGE<sup>Δ</sup>

**Figure 51** : Recensement des acides aminés utiles pour la formation du dimère. Les acides aminés caractérisés par un triangle sont impliqués dans la formation du double feuillet β, les carrés sont ceux interagissant avec l'ADN et les points représentent les acides aminés jouant un rôle dans la dimérisation. La numérotation des acides aminés est celle de la protéine Zif268.

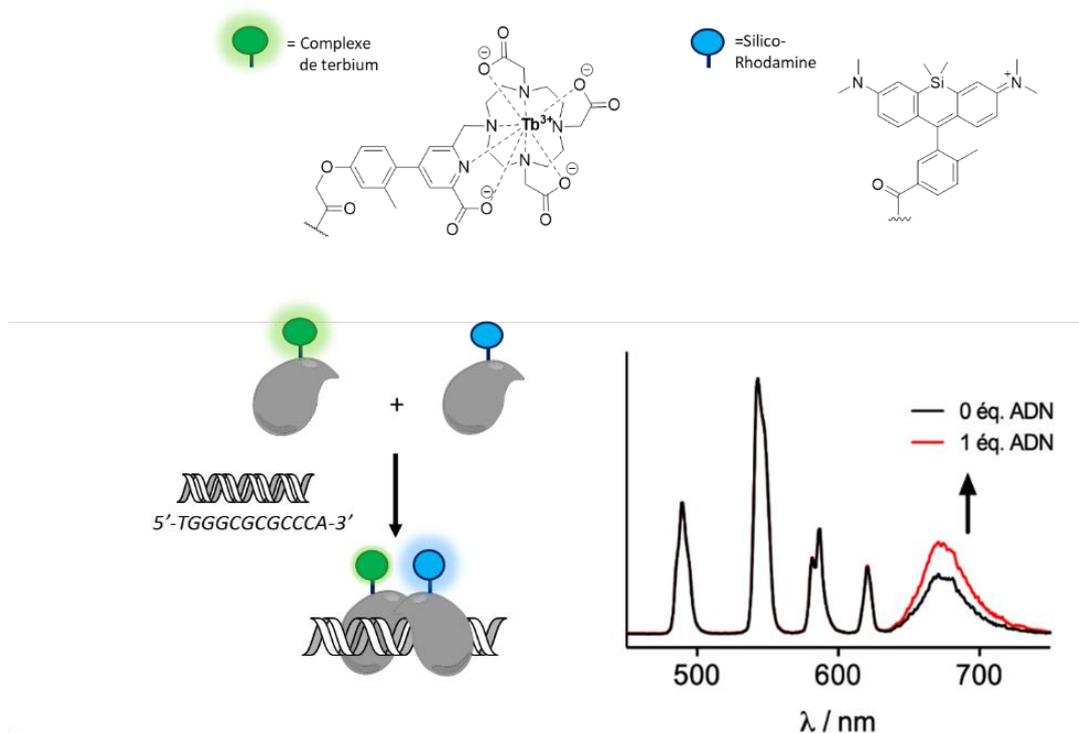
La fonctionnalisation d'une protéine par des chromophores organiques implique une synthèse par voie chimique de cette protéine. Une méthode de choix est la synthèse peptidique supportée sur résine (SPPS). Celle-ci offre de nombreuses possibilités pour la synthèse de protéines fonctionnalisées en permettant l'insertion aisée de molécules sur les chaînes latérales en jouant avec des groupements protecteurs orthogonaux à ceux utilisés classiquement en SPPS ou des acides aminés non naturels. Il existe cependant une limitation à la synthèse chimique des peptides par SPPS. Le peptide ne peut excéder une certaine longueur allant de 30 à 45 acides aminés car au-delà, les couplages d'acides d'aminés deviennent trop difficiles, les purifications aussi et les rendements chutent dramatiquement. La protéine Zif12 contenant 72 acides aminés, la synthèse doit être réalisée par assemblage de plusieurs fragments. En ayant recours à la réaction de ligation chimique native (NCL) pour l'assemblage de ces fragments (la NCL sera décrite dans le chapitre suivant), cette protéine peut être fragmentée en trois au niveau de ses cystéines<sup>99</sup>. La présence d'une cystéine en position N-terminale du fragment à lier est essentielle à la NCL. Ainsi, la protéine a été découpée en deux premiers fragments comportant *grosso modo* chacun un doigt de zinc (noté F2 et F3) et un dernier correspondant à l'extension N-terminale de la protéine (F1) (**Figure 52**).



**Figure 52** : Découpage de la protéine Zif12 par Guillaume Frémy. Les aa de couleur en gras correspondent au domaine de reconnaissance du doigt de zinc.

## 2.2 - La preuve de concept

Guillaume Frémy a synthétisé une paire de protéines Zif12 fonctionnalisées pour l'une par un complexe de  $Tb^{3+}$  et l'autre par une silico-rhodamine (**Figure 53**). Le complexe de  $Tb^{3+}$  est inspiré des travaux de Maury<sup>84</sup> pour l'antenne en y insérant un carboxylate pour la conjugaison sur le groupement électrodonneur du système push-pull. Notons que la synthèse de ces protéines a été difficile, du fait de problèmes rencontrés dans la synthèse du fragment central fonctionnalisé (mauvais couplages des acides aminés lors de la SPPS en fin de peptide, réaction parasite inattendue sur la cystéine N-terminale, déprotection du pro-ligand de  $Ln^{3+}$  peu aisée avec la voie de synthèse élaborée par Guillaume).

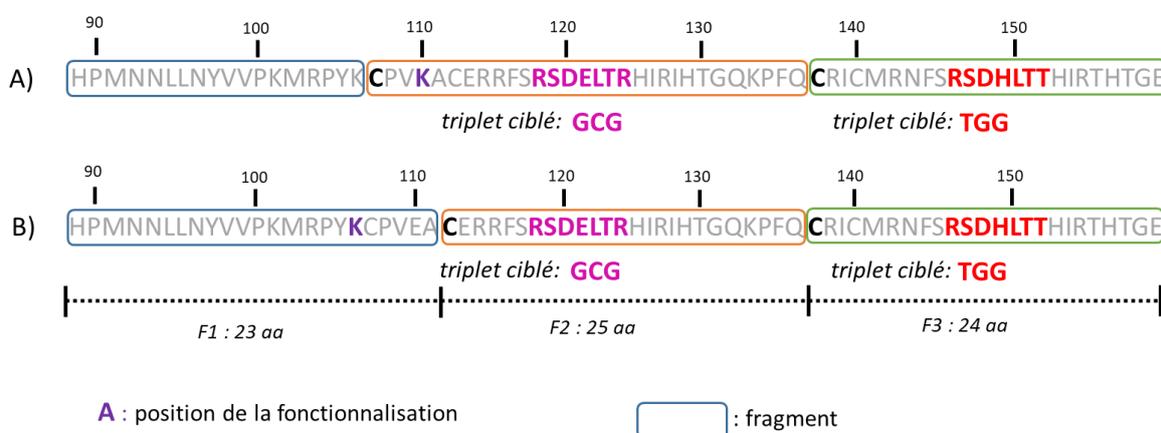


**Figure 53** : Preuve de concept obtenue par Guillaume Frémy sur une séquence palindromique d'ADN. En haut sont représentés les deux partenaires du système FRET, en bas à gauche le concept utilisé et en bas à droite les spectres d'émission de fluorescence normalisés obtenus avec et sans ADN cible.

Le spectre d'émission de fluorescence des deux protéines fonctionnalisées est présenté ci-dessus (courbe noire). On peut voir les raies d'émission caractéristiques du terbium, mais aussi l'émission de la Silico-Rhodamine (SiR<sup>Me</sup>). Cette dernière est due à la longueur d'onde d'excitation du complexe de terbium ( $\lambda_{ex} = 305$  nm) qui excite aussi le fluorophore. Cependant, en présence d'un équivalent d'ADN cible dont la séquence palindromique est 5'-TGGGCGCGCCA-3', on observe une augmentation du signal provenant du fluorophore lorsque l'on normalise par rapport à l'intensité du signal du terbium (courbe rouge). Des caractérisations spectroscopiques ont permis de mettre en évidence le transfert d'énergie du Tb<sup>3+</sup> à la SiR<sup>Me</sup>. Néanmoins celui-ci est modeste, indiquant que la distance entre les deux chromophores est peut-être trop grande et le système souffre d'un inconvénient : à la longueur d'onde où s'excite le complexe de Tb<sup>3+</sup> ( $\lambda_{max} = 302$  nm), la SiR<sup>Me</sup> est excitée également. De manière à accentuer cette augmentation du signal, la fonctionnalisation par un complexe avec une nouvelle antenne sensibilisant l'euporium a été entreprise, pour décaler l'excitation vers 340 nm où l'absorption de la SiR<sup>Me</sup> est négligeable. Par manque de temps et de composé, ce système n'a pu être testé en interaction avec l'ADN. Mais sans vouloir dévoiler la suite, nous nous sommes aperçus qu'une réaction parasite avait lieu lors de la synthèse de la protéine au niveau du complexe d'euporium, la protéine obtenue ne correspondant ainsi pas à la protéine voulue.

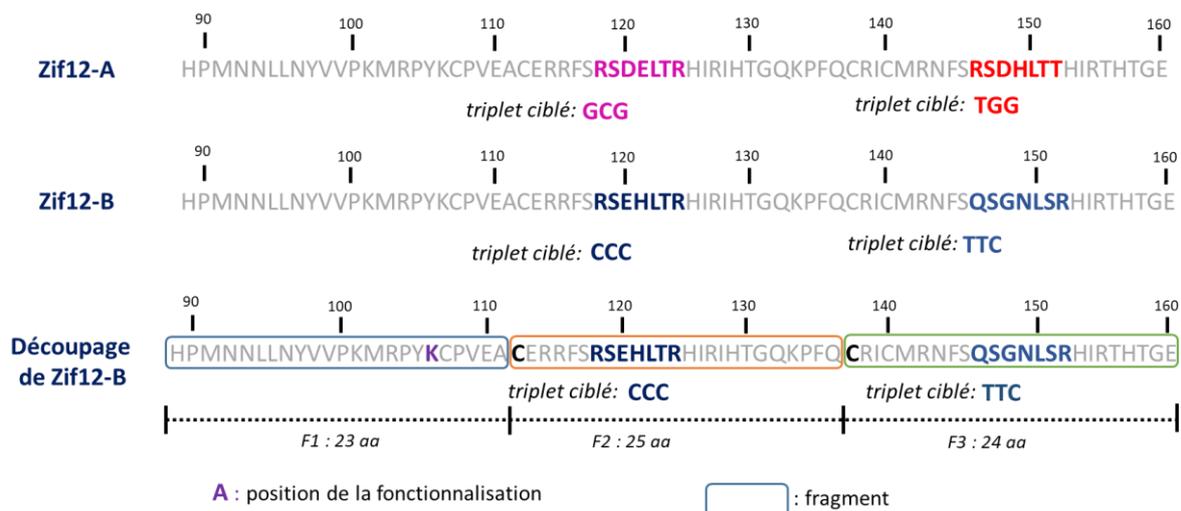
### 3 - Les objectifs du projet de thèse

Dans la continuité des travaux de thèse du Dr Guillaume Frémy, un objectif était d'améliorer le transfert d'énergie entre le complexe de lanthanide et la Silico-Rhodamine. Nous devons reprendre la synthèse de la protéine comportant le complexe d'euprium en apportant quelques modifications. Un nouveau découpage de la protéine a été choisi pour faciliter sa synthèse et apporter plus de versatilité (**Figure 54**). L'allongement du fragment F1 permettra de séparer la partie reconnaissance d'ADN (fragments F2 et F3) de la partie FRET (fragment F1). Compte tenu de la faible augmentation du signal du fluorophore en présence d'ADN, nous avons envisagé également de fonctionnaliser la protéine sur une autre position.



**Figure 54** : Présentation des modifications apportées sur le découpage de la protéine. A) Le précédent découpage de la protéine. B) Le nouveau découpage de la protéine Zif12-A.

L'autre objectif est de nous tourner vers le ciblage d'une séquence non palindromique d'ADN. Le principal problème que pose le ciblage d'une séquence palindromique est la formation d'homodimère. Statistiquement, sur un mélange contenant les deux protéines et la cible, on obtient seulement 50% d'hétérodimère. Le reste correspondant à la formation des deux homodimères. Ces homodimères, qu'ils soient faits avec le complexe de lanthanide ou le fluorophore, participent à l'émission totale du système en présence d'ADN. Par conséquent, la variation des intensités de luminescence des deux partenaires est influencée ce qui représente un biais quant à la réponse mesurée. La synthèse d'une nouvelle protéine, notée Zif12-B, comportant deux doigts de zinc nous permettra d'aller vers le ciblage d'une séquence non palindromique de 12 paires de bases en combinaison avec la protéine développée par Guillaume Frémy (Zif12-A). Les séquences des deux protéines notées Zif12-A et Zif12-B, ainsi que les cibles d'ADN de chacun des doigts de zinc sont présentés **Figure 55**.



**Figure 55** : Séquences des protéines Zif12-A et Zif12-B impliquées dans la reconnaissance d'une séquence non palindromique d'ADN 5'-TGGGCGCCCTTC.

La protéine Zif12-A correspond à celle utilisée pour la preuve de concept menée par Guillaume Frémy ciblant la séquence 5'-TGGGCG-3'. La seconde protéine, Zif12-B, est construite sur ce même modèle, mais en modifiant uniquement les acides aminés impliqués dans la reconnaissance de l'ADN afin de cibler la séquence 5'-CCCTTC-3', que nous avons choisie de manière arbitraire. L'utilisation de ces deux protéines nous amène donc au ciblage de cette nouvelle séquence d'ADN : 5'-TGGGCGCCCTTC-3'.

Dans un premier temps, il conviendra de valider la reconnaissance de l'ADN par ces nouvelles protéines à l'aide de protéines non fonctionnalisées. C'est dans un second temps que le système de FRET sera incrémenté pour mesurer la réponse lumineuse obtenue lors de la formation de l'hétérodimère obtenu.

# **Chapitre II - Synthèse par voie chimique de protéines Zif12 non fonctionnalisées pour la reconnaissance d'ADN non palindromique**

## Sommaire

Préambule - Design de la protéine Zif12-B.....	59
I - Synthèse par voie chimique de Zif12 .....	60
1 - Ligation chimique native .....	60
1.1 - Ligation chimique native par la méthode SEA .....	62
1.1.a Ligation SEA.....	63
1.1.b Ligation hydrazide.....	64
1.2 - Segmentation des protéines Zif12-A et Zif12-B .....	65
1.3 - Stratégie d'assemblage d'une protéine par fragments.....	67
1.4 - Stratégie d'assemblage de la protéine Zif12 .....	69
2 - Mise au point de l'assemblage de la partie reconnaissance d'ADN (F2+F3) .....	71
2.1 - Ligation des fragments Thz-F2 et F3 par ligation SEA, protection Thz en N-ter.....	72
2.1.a Synthèse des fragments $Thz-F2-SEA^{off}$ et $H-F3-NH_2$ .....	72
2.1.b Ligation des fragments $Thz-F2-SEA^{off}$ et $H-F3-NH_2$ .....	73
2.2 - Ligation des fragments F2 et F3 par ligation hydrazide, protection N <sub>3</sub> en N-ter .....	74
2.2.a Synthèse de $N_3-F2-NHNH_2$ .....	74
2.2.b Ligation avec le fragment $H-F3-NH_2$ .....	75
2.3 - Ligation des fragments F2 et F3 par ligation SEA, protection Acn en N-ter.....	76
2.4 - Ligation du fragment F2 et F3 par ligation hydrazide, protection Acn en N-ter .....	78
3 - Assemblage de la partie reconnaissance d'ADN avec l'extension N-terminale .....	79
3.1 - Synthèse du fragment F1-SEA.....	79
3.2 - Synthèse du fragment H-F1-Nle-SEA.....	80
3.3 - Assemblage de l'extension N-terminale et du domaine de reconnaissance .....	82
4 - Formation des doigts de zinc .....	84
II - Étude des interactions des protéines Zif12-A et Zif12-B avec l'ADN.....	84
1 - La technique EMSA.....	85
1.1 - Principe de la technique EMSA .....	85
1.2 - Interaction des protéines avec le duplex d'ADN .....	86
2 - Rappel des travaux de Wang et Pabo.....	87
3 - Interaction des protéines Zif avec les séquences palindromiques d'ADN .....	88

Ce premier chapitre de thèse est consacré à la synthèse des protéines Zif12-A et Zif12-B ne comportant aucun chromophore et l'étude des interactions avec différentes séquences d'ADN double brins. Le but est de vérifier que les deux protéines sont capables de former un dimère autour d'une séquence d'ADN cible non palindromique avant d'incorporer le système FRET.

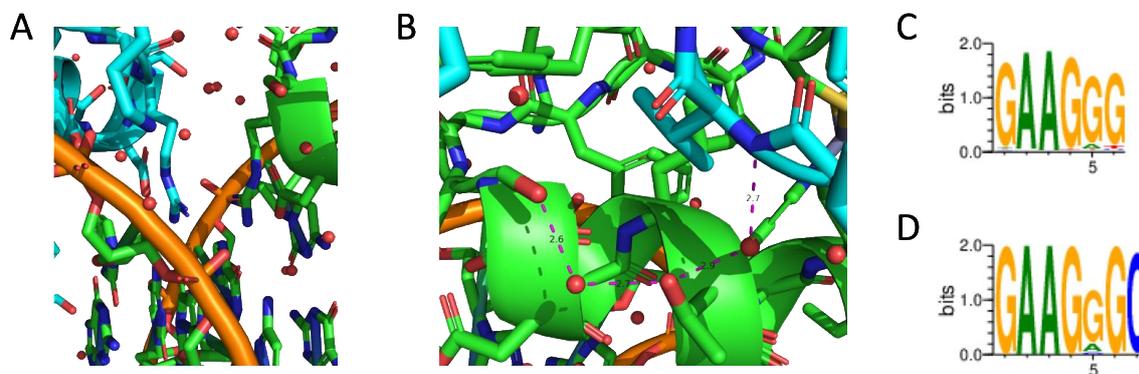
### Préambule - Design de la protéine Zif12-B

Le design de la séquence de la protéine Zif12-B qui doit reconnaître le duplex 5'-GGAGGG-3' a été réalisé par Olivier Sénèque avant mon arrivée au laboratoire. Celui-ci est décrit dans cette partie.

La protéine Zif12 reconnaît le duplex 5'-TGGGCGGCCCA-3' en formant un dimère sur cette séquence. Le dimère présente un axe de symétrie C<sub>2</sub>. La protéine Zif12 comporte deux doigts de zinc. Le doigt de zinc N-terminale (ZF<sup>N</sup>) reconnaît le triplet 5'-GCG-3' grâce à la séquence RSDELTR de son hélice et le doigt de zinc C-terminale (ZF<sup>C</sup>) reconnaît le triplet 5'-TGG-3' grâce à la séquence RSDHLTT de son hélice  $\alpha$ . Pour choisir la nouvelle séquence d'ADN reconnue, Olivier Sénèque a d'abord cherché s'il y avait des contraintes imposées par (i) des interactions entre l'extension N-terminale d'une des protéines avec l'hélice de reconnaissance de l'ADN de l'autre protéine du dimère et (ii) des interactions entre les deux hélices de reconnaissance des deux ZF<sup>N</sup>. L'atome d'oxygène de la chaîne latérale de la thréonine de l'hélice de ZF<sup>N</sup> est impliqué dans un réseau de liaison hydrogène avec l'extension N-ter de l'autre protéine. On peut noter aussi que la sérine de l'hélice  $\alpha$  forme une liaison hydrogène avec la thréonine (**Figure 56 A**). Il a été décidé de les conserver, ce qui impose une séquence pour l'hélice de reconnaissance de type XSXXLTX (la leucine est un des sept acides aminés canoniques du doigt de zinc classique  $\beta\beta\alpha$  et doit être conservé également). De plus, les deux domaines de reconnaissance des ZF<sup>N</sup> sont en contact l'une avec l'autre, au centre de la structure, par la chaîne latérale des arginines N-ter du motif de reconnaissance (**Figure 56 B**). Chacune de ces arginines reconnaît la guanine en 3' du triplet 5'-GCG-3'. Il semble donc important de conserver cette arginine et d'avoir un motif central GG au cœur du 12-mer reconnu par les deux Zif12. Ceci impose donc la séquence RSXXLTX pour ZF<sup>N</sup>, qui peut cibler à ce stade un triplet 5'-NNG-3'. Aucune autre contrainte n'a été identifiée pour ZF<sup>C</sup>. Il est connu que les doigts de zinc qui reconnaissent les triplets 5'-GNN-3' présentent la meilleure affinité de liaison. Afin de s'assurer d'une affinité suffisante pour la reconnaissance, il a été choisi de reconnaître deux triplets commençant par une guanine, d'où un motif 5'-GNNG-3'. Olivier Sénèque a opté finalement pour le motif 5'-GAAGGG-3'.

La base de données de doigts de zinc et le site web « DNA-binding Specificities of Cys<sub>2</sub>His<sub>2</sub> Zinc Finger Proteins » (<http://zf.princeton.edu/index.php>) décrits par Persikov<sup>100-102</sup> ont ensuite été utilisés pour déterminer la séquence peptidique la plus adaptée pour chacun des motifs de reconnaissance. Le site web/logiciel de Persikov permet de trouver une séquence de doigt de zinc (i) pour la cible d'ADN correspondante ou (ii) pour déterminer, à partir d'une séquence d'ADN, la séquence des doigts de zinc (ou plutôt les séquences d'aa des motifs de reconnaissance) qu'il faut utiliser. Le logiciel de Persikov tient compte des effets de voisinages entre les doigts de zinc : un doigt de zinc peut influencer la

reconnaissance du voisin. L'intérêt des travaux de Persikov réside dans la taille de la base de données (>160000 séquences d'ADN). Ainsi, Olivier Sénèque a cherché les motifs d'hélice qui permettent de reconnaître la séquence 5'-GAAGGG-3' avec ce logiciel. Parmi les motifs proposés, suite à une recherche avec la « Target analysis » et la base B1H de Persikov (sur F3)<sup>100,102</sup> pour la reconnaissance de 5'-GGG-3' avec un bon score d'affinité, aucune séquence ne correspondait à la séquence que l'on souhaite RSXXLTX. Aucun motif ne présentait de sérine en position 1, mais certains avaient une thréonine en position 6. La séquence RNEHLTR est proposée dans l'article de Persikov paru en 2015 pour ce motif<sup>100</sup>. Finalement, la séquence RSEHLTR a été retenue pour la reconnaissance de 5'-GGG-3'. Pour le triplet 5'-GAA-3', la recherche sur F3 et F2, a donné un motif consensus QXXNLXR et c'est le motif QSGNLSR, qui présente le meilleur score. C'est donc ce motif qui a été retenu. Finalement, une prédiction du motif d'ADN reconnu avec la séquence de la protéine Zif12-B, incluant les deux nouveaux motifs de reconnaissance a été faite pour vérifier que la séquence choisie pour reconnaître 5'-GAAGGG-3' convenait, notamment le motif RSEHLTR qui n'était pas sorti de l'analyse. Avec le module « interactive PWM predictor » du site de Persikov<sup>101</sup>, une prédiction de reconnaissance en utilisant le modèle « RF regression on B1H » donne un bon accord pour la séquence ciblée (**Figure 56 C**). La prédiction avec le modèle « Expanded linear SVM », qui permet de prédire aussi la base la plus favorable en position 3' du motif, a donné également un bon accord avec la séquence cible et a prédit la reconnaissance d'une C en position 3', ce qui sera le cas si le dimère se forme. Nous avons donc retenu la séquence de Zif12B donnée dans le chapitre précédent, avec les motifs RSEHLTR pour le doigt de zinc N-terminale et le motif QSGNLSR pour le doigt de zinc C-terminal.



**Figure 56 :** (A) Zoom sur la structure de Zif12 lié à l'ADN, montrant que les deux chaînes d'arginine se font face pour reconnaître le motif GG<sup>54</sup> (B) Zoom sur l'hélice du doigt de zinc N-terminale qui montre le réseau de liaisons hydrogène entre la sérine et la thréonine. (C,D) Score de prédiction de reconnaissance pour la protéine Zif12-B avec le modèle « RF regression on B1H » (C) et avec le modèle « Expanded linear SWM » (D).

## I - Synthèse par voie chimique de Zif12

### 1 - Ligation chimique native

Le système de reconnaissance de l'ADN que nous avons choisi est celui développé par Wang et Pabo (**Chapitre I, II, 2.3.d**). Pour rappel il s'agit d'une protéine longue de 72 aa dont la synthèse supportée sur résine est difficile du fait de sa longueur (voir plus loin). Il faut donc faire appel à une méthode qui

permet d'assembler plusieurs fragments peptidiques, en formant un lien amide entre deux peptides<sup>99</sup>. De nombreuses méthodes de ce genre ont été développées durant les vingt dernières années et continuent de susciter de l'intérêt. Sur la **Figure 57** sont représentés les grands types de ligation. Le premier (type I) correspond aux réactions qui impliquent la chaîne latérale d'un acide aminé. On la dénomme ligation chimique native (NCL, pour l'anglais « Native Chemical Ligation). C'est une réaction entre un peptide comportant un thioester C-terminale et un second peptide comportant une cystéine N-terminale. La fonction thiol, en chaîne latérale de la position N-terminale, joue un rôle clé. Le type II rassemble les réactions qui se situent au niveau de l'azote N-terminale de la chaîne principale du second peptide. Il se divise en deux sous-catégories. La première est la ligation de Staudinger qui a lieu entre un groupement phosphine et un azote. Elle conduit à la formation d'une iminophosphorane, qui par transfert d'acyle, permet la formation du lien natif<sup>103</sup>. La seconde est la ligation KAHA (Ketoacid-hydroxylamine ligation) entre un cétoacide et une hydroxylamine aboutissant à la formation d'un intermédiaire nitrone, qui par un réarrangement suivi d'une décarboxylation forme le lien natif<sup>104</sup>. Le type III, quant à lui, correspond à la réaction d'un peptide avec un aldéhyde en C-terminale avec un second peptide comportant un β-amino alcool (une sérine ou un de ces analogues) en N-terminale. Elle est appelée ligation sérine/thréonine (STL) et passe par un intermédiaire oxazolidine qui donne par transfert d'acyle, suivi d'une hydrolyse, le lien amide<sup>105</sup>.

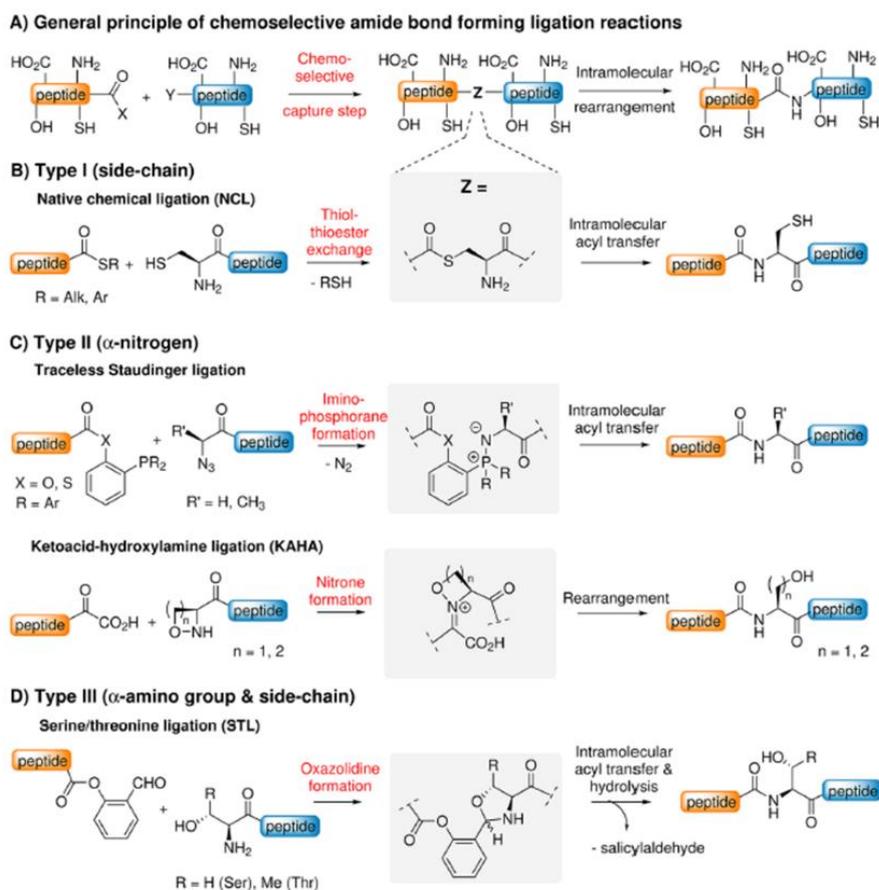
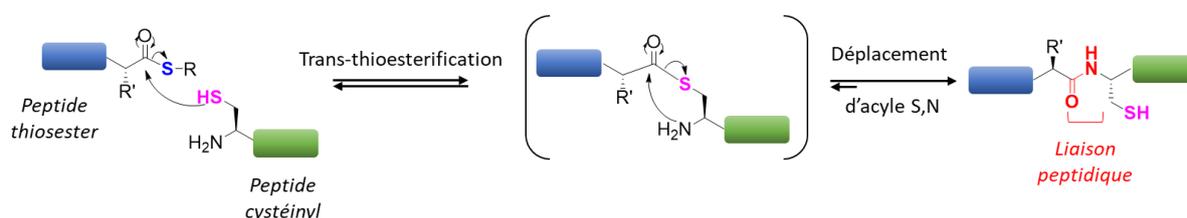


Figure 57 : Principales méthodes qui permettent la formation d'un lien natif entre deux fragments protéiques<sup>99</sup>.

### 1.1 - Ligation chimique native

Parmi ces différentes méthodes de ligation, la plus facile à mettre en œuvre est la NCL, de type I. C'est l'équipe de S. Kent qui, en 1994, développa cette réaction<sup>106</sup>. Le principe de la réaction est représenté **Figure 58**. C'est une réaction régiosélective entre un peptide comportant un thioester, en position C-terminale, et un peptide avec un groupement cystéinyl en position N-terminale. La première étape est une réaction de trans-thioesterification qui aboutit à la formation d'un intermédiaire. C'est une réaction réversible, le groupement SR<sup>-</sup> libéré pouvant réagir sur le thioester formé. La seconde étape est un transfert d'acyle de S vers N. Elle est en principe réversible, mais la forme amide est bien plus stable que la forme thioester ce qui constitue un puits thermodynamique : la réaction est donc tirée vers la formation d'un lien peptidique dit « natif », stable, entre les deux peptides tout en recouvrant la fonction thiol impliquée dans la réaction.



**Figure 58** : Mécanisme de la ligation chimique native (NCL). Réaction entre un peptide comportant un thioester en position C-terminale et un peptide cystéinyl en position N-terminale via une réaction de transthoesterification<sup>107</sup>.

L'espèce clé dans cette réaction de ligation est le thioester formé. Selon sa nature, il est plus ou moins réactif vis-à-vis de l'attaque du thiol. Ce qui est en fait l'étape cinétiquement déterminante dans le processus. Il est donc important d'avoir recours à un thioester réactif, que l'on qualifie généralement d'actif. La synthèse d'un peptide comportant un thioester en position C-terminale peut être réalisée par différents moyens. Certains thioesters étant stables en présence de TFA permettent la synthèse de peptides en phase solide (SPPS) avec la stratégie Boc, à l'aide d'une résine fonctionnalisée par un linker thiol adéquat<sup>99</sup>. Mais la stratégie Boc est de moins en moins utilisée, car elle nécessite l'utilisation d'acide fluorhydrique, très dangereux, pour décrocher le peptide de la résine. Ceci demande des précautions extrêmes, difficiles à mettre en place en routine dans un laboratoire de recherche. Cependant ce thioester peut aussi être généré après l'élongation du peptide en stratégie Fmoc. L'avantage ici est d'éviter son exposition avec les différents réactifs utilisés lors de la synthèse peptidique, entre autres la pipéridine (utilisée pour déprotéger le groupement Fmoc)<sup>108</sup>. Pour ce faire, il existe plusieurs méthodes pour générer la fonction thioester<sup>99</sup>, cela peut être à partir de différents groupements (i) N-alkyle, N-acyle (ii) sulfonamides<sup>109</sup> (iii) imines<sup>110</sup>. Il peut aussi être obtenu après une étape d'oxydation (ex: benzotriazole)<sup>111</sup> ou de condensation (ex : pyrazole)<sup>112</sup>. Le second moyen consiste à introduire ce thioester sous une forme protégée, avant l'élongation du peptide, on parle alors de crypto-thioester. Parmi eux, deux exemples qui ont été utilisés au sein de notre équipe, le N-2-hydroxybenzyl-cysteine (HNB) développé par V. Aucagne<sup>113</sup>, mais le plus couramment utilisé est le bis(2-sulfanylethyl)amido (SEA) développé par O. Melnyk<sup>114</sup> (**Figure 59**).

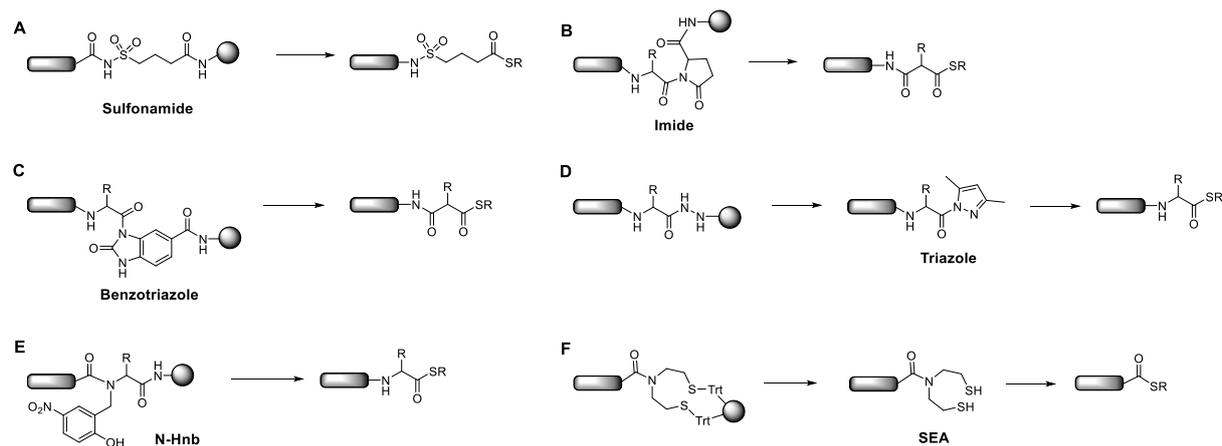


Figure 59 : Présentation des différentes méthodes pour obtenir un thioester activé.

### 1.1.a Ligation SEA

Le crypto-thioester SEA est un amide secondaire avec deux substituants thioéthyle. Les peptides avec un crypto-thioester SEA peuvent être synthétisés par SPPS en stratégie Fmoc en utilisant la résine 2Cl-Trt-PS commerciale, qui comporte un groupement *bis*(2-sulfamylethyl)amine greffé sur une résine 2-chlorotriptyl (**Figure 60**). La première étape consiste en l'acylation du groupement *bis*(2-sulfamylethyl)amine, suivie par une élongation classique du peptide selon la stratégie Fmoc<sup>114,115</sup>. Une fois la synthèse du peptide terminée, le peptide est décroché de la résine en condition acide, à l'aide de TFA, ce qui permet de libérer les deux fonctions thiols du groupement SEA. Sous cette forme le crypto-thioester est nommé SEA<sup>on</sup>. Cependant ce dithiol est en équilibre avec une autre forme, la forme thioester linéaire. Il peut également être oxydé sous forme d'un disulfure, noté SEA<sup>off</sup>. Pour des raisons de conservation du peptide, de facilité de purification et de rendement, il est préférable d'oxyder les peptides SEA pour n'obtenir que l'espèce cyclique SEA<sup>off</sup>. Sous cette forme le crypto-thioester est inactif vis-à-vis de la NCL. L'ajout d'un réducteur tel que le TCEP ou le DTT permet de le réduire et de le remettre sous sa forme active SEA<sup>on</sup>.

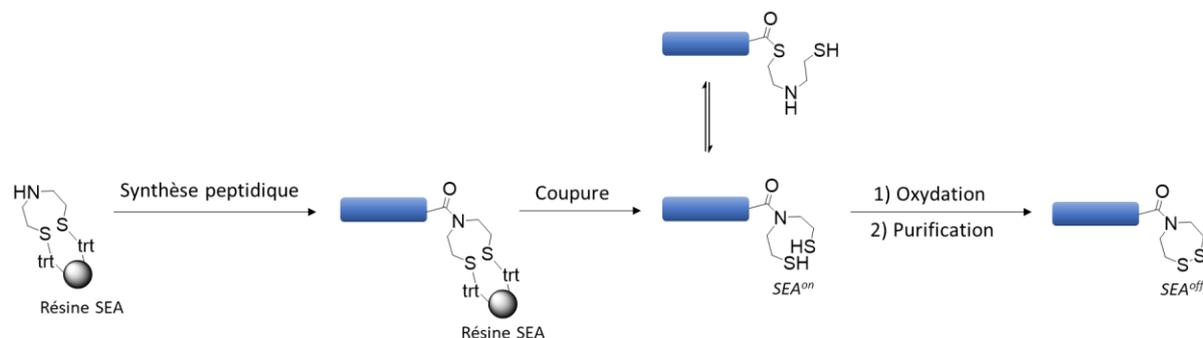


Figure 60 : Présentation de la synthèse d'un peptide sur résine SEA.

L'étape d'oxydation, réalisée au moyen de I<sub>2</sub> implique de protéger les acides aminés qui présentent des fonctions facilement oxydables, telles que les cystéines, pour éviter la formation de disulfure ou d'espèces sur-oxydées (sulfinates, sulfonates) au niveau de la chaîne latérale. Les cystéines seront donc protégées par un groupement *S*tBu qui résiste aux conditions de coupure et d'oxydation du peptide.

Il pourra être déprotégé lors de la réaction de ligation chimique native *via* l'ajout d'un réducteur comme le TCEP.

Bien que l'on dispose d'un thioester qui est facilement introduit sur le peptide, dont on peut aisément passer d'une forme active à inactive, celui-ci présente un inconvénient. Il n'est que peu réactif vis-à-vis de la trans-thioesterification avec le peptide cystéinyl. La solution est de passer par un thioester activé, c'est-à-dire plus réactif.

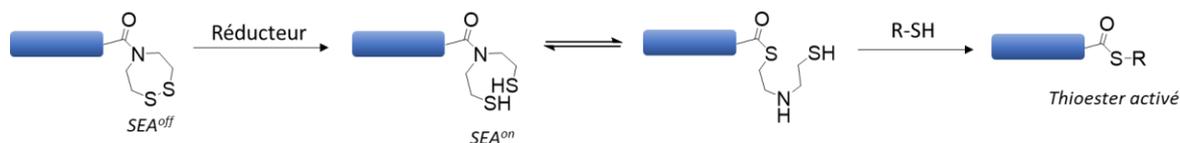
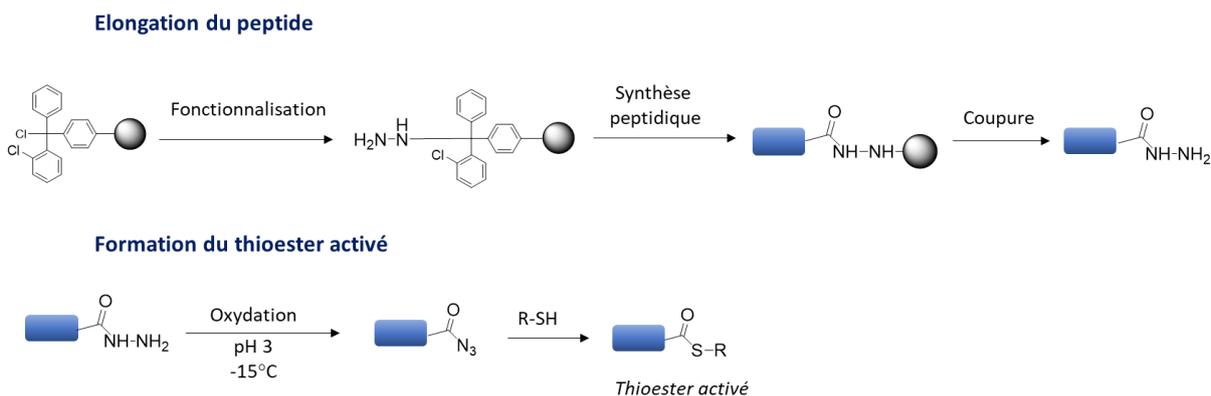


Figure 61 : Formation d'un thioester activé à partir d'un peptide SEA<sup>off</sup>.

Pour ce faire plusieurs catalyseurs pour la NCL ont été étudiés par le passé, basés sur l'ajout *in situ* de thioaryle ou thioalkyle pour faire une première réaction d'échange thiol-thioester en présence d'un large excès de thiol pour augmenter la vitesse de réaction<sup>116</sup> (**Figure 61**). Parmi ceux qui présentent le meilleur compromis, entre l'échange thiol-thioester et la réactivité lors de la trans-thioestérification avec le peptide cystéinyl, on retrouve le MPAA (4-mercaptophénylacetic acid), le MPA (3-mercaptopropionic acid) ou encore le MesNa (sodium 2-mercaptoethanesulfonate). De ces trois-là, le MPAA est le plus utilisé. D'autres paramètres rentrent aussi en compte lors de cette réaction de ligation chimique native. Le pH de la réaction joue un rôle important, notamment lors de l'échange thiol-thioester. La réaction est classiquement réalisée entre pH 5 et 8, mais la cinétique de ligation augmente quand le pH décroît (la précipitation du MPAA généralement utilisé à pH < 5 empêche de descendre en dessous de cette limite)<sup>117</sup>. Enfin il est d'usage d'utiliser un agent dénaturant qui permet de maintenir en solution les peptides durant la réaction de ligation, à savoir de l'urée ou du chlorure de guanidinium.

### 1.1.b Ligation hydrazide

Une autre méthode de ligation maintenant très répandue est la méthode de ligation hydrazide. Cette méthode a été décrite par l'équipe de Liu, en 2011 et appliquée à la synthèse séquentielle d'une protéine de 140 acides aminés ( $\alpha$ -synuclein)<sup>118</sup>. Elle consiste à générer le thioester à partir d'un peptide comportant une fonction hydrazide en C-terminale. Le peptide hydrazide est obtenu simplement par SPPS en stratégie Fmoc à partir d'une résine 2-chlorotriptyl préalablement fonctionnalisée avec de l'hydrazine (**Figure 62**). A noter que la fonction hydrazide est compatible avec l'ensemble des groupements protecteurs classiquement utilisés en SPPS.



**Figure 62** : Présentation de l'élongation d'un peptide sur une résine 2-chlorotrityl fonctionnalisée par une hydrazine et la formation du thioester activé à partir d'une fonction hydrazine.

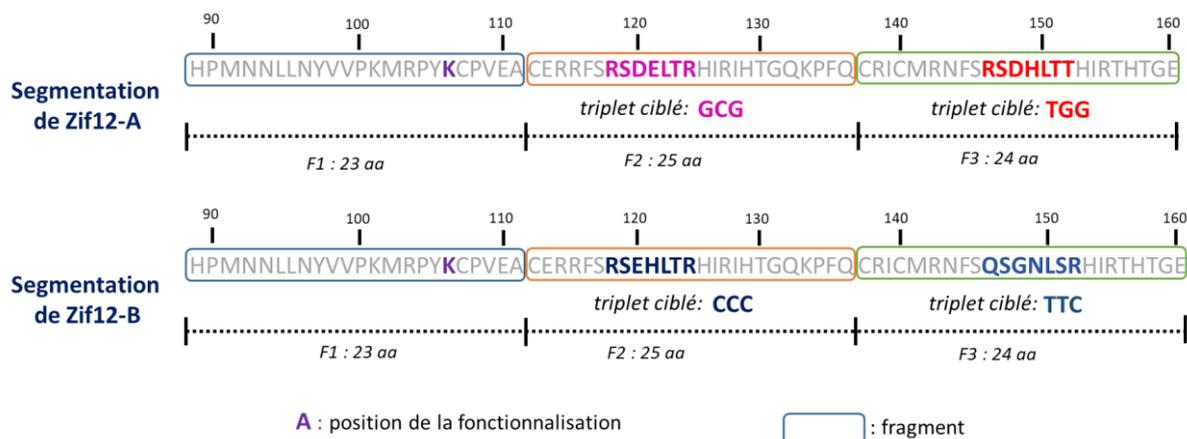
La première étape de ligation est une oxydation du peptide hydrazide par  $\text{NaNO}_2$  pour obtenir un azoture d'acyle. Il est particulièrement réactif, notamment vis-à-vis des thiols utilisés (MPAA) pour former un thioester activé. Cette oxydation est réalisée dans une solution de chlorure de guanidinium (6 M) à pH 3 à froid ( $-15^\circ\text{C}$ ). Ces conditions permettent de contrôler l'oxydation et de s'affranchir de l'utilisation de groupement protecteur, pour les acides aminés sensibles à l'oxydation tels que les cystéines. L'oxydation par  $\text{NaNO}_2$  peut générer des nitrosothiols ou des disulfures, mais ceux-ci sont réduits par le thiol en excès. L'ajout de ce thiol induit la formation du thioester activé en position C-terminale du peptide, avant de réagir avec un peptide cystéinyl. La réaction de trans-thioestérification aboutit au produit de couplage des deux peptides. Cette méthode est simple à mettre en œuvre et peu coûteuse, la résine 2Cl-Trt-PS étant plus accessible que la résine SEA-2Cl-Trt-PS. Mais le principal avantage que cette méthode offre est un temps de réaction particulièrement rapide, d'une dizaine de minutes à quelques heures.

### 1.2 - Segmentation des protéines Zif12-A et Zif12-B

La synthèse des protéines Zif12-A et Zif12-B, par assemblage de plusieurs fragments, est donc dictée par la méthode de ligation choisie. Dans le cas d'une ligation chimique native de type I, nous devons avoir des fragments qui comportent une cystéine en position N-terminale. Pour le découpage de la protéine Zif12, nous avons plusieurs possibilités puisque la séquence contient quatre cystéines (**Figure 63**).



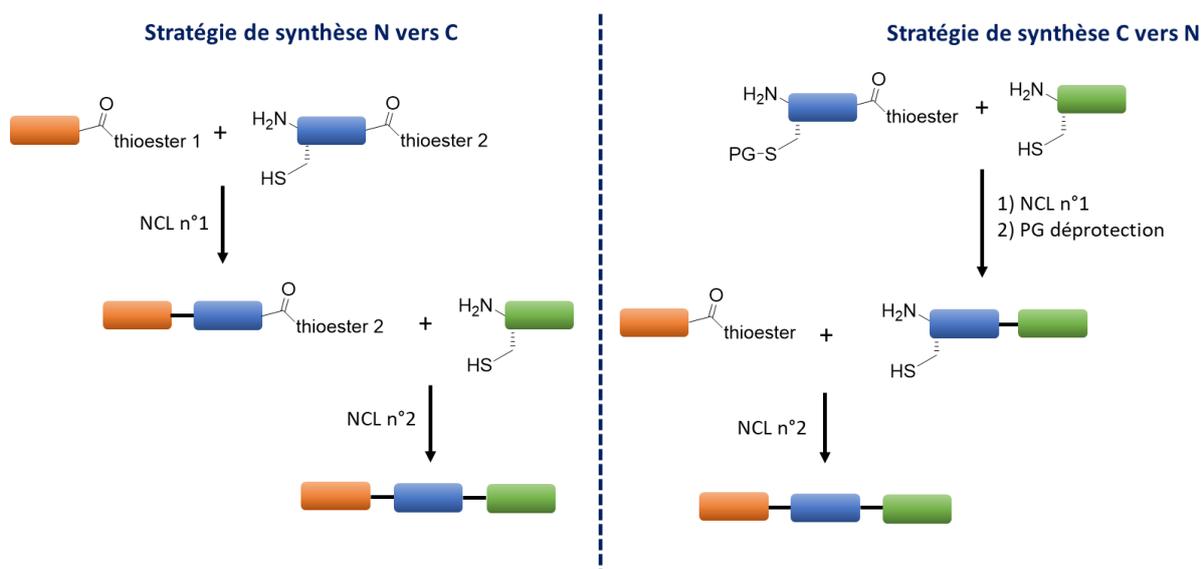
la reconnaissance d'un triplet de paires de bases. La segmentation en trois fragments des protéines Zif12-A et Zif12-B est présentée **Figure 65**.



**Figure 65** : Segmentation des protéines Zif12-A et Zif12-B en trois fragments. Le fragment F1 correspond à l'extension N-terminale de Zif12, les fragments F2 et F3 comportant chacun un doigt de zinc. Le domaine de reconnaissance de chacun des doigts de zinc est noté en gras et en couleurs.

### 1.3 - Stratégie d'assemblage d'une protéine par fragments

La synthèse d'une protéine par assemblage de plusieurs fragments peut s'effectuer selon deux voies. La première voie s'effectue dans le sens N-terminale vers le C-terminale. Dans ce cas, la difficulté réside dans le cryptothioester utilisé en position C-terminale du fragment central. Il ne doit pas être activé dans les conditions utilisées lors de la première étape de ligation. La seconde voie est un assemblage dans le sens inverse, du C-terminale vers le N-terminale. La difficulté se situe à présent sur le groupement protecteur de la cystéine en position N-terminale du fragment central. Il doit résister aux conditions de ligation native pour éviter la polymérisation. Ces deux voies de synthèse sont résumées **Figure 66**. Ces deux méthodes peuvent s'effectuer de manière séquentielle, en purifiant le produit obtenu après chaque ligation, ou de manière « one-pot », avec une seule étape de purification en toute fin de synthèse.



**Figure 66** : Présentation des stratégies de synthèse de N vers C et C vers N.

En 2004, Kent et son équipe ont réalisé la synthèse d'une protéine de 46 acides aminés (Crambin) par un assemblage séquentiel de trois fragments de N vers C par ligation chimique native<sup>119</sup>. C'est autour du groupement protecteur acétamidométhyle (Acm) de la cystéine du fragment central que cette stratégie s'articule. Après la première étape de ligation, la déprotection de ce groupement est réalisée en présence d'acétate d'argent dont l'excès est quenché par du DTT, en fin de réaction pour un rendement de 46%. Le fragment obtenu est ensuite impliqué dans la dernière réaction de ligation avec le fragment N-terminale comportant un thioester, dérivé du MPA en position C-terminale. Par cette méthode la protéine Crambin a été synthétisée avec un rendement global de 25%. Quelques années plus tard, l'assemblage de des trois fragments par une stratégie convergente de N vers C a été réalisé par Kent<sup>120</sup>. La protéine [Lys<sup>24</sup>, <sup>38</sup>, <sup>83</sup>] Erythropoïétine (EPO), dont la partie protéique est longue de 166 acides aminés, a été découpée dans un premier temps en deux. La partie N-terminale de la protéine EPO est obtenue par la ligation native de deux fragments de 29 et 38 acides aminés. La partie C-terminale quant à elle a été synthétisée par un assemblage séquentiel de trois fragments, respectivement de 29, 29 et 37 acides aminés (**Figure 67**). Le groupement protecteur utilisé pour protéger les cystéines du fragment central et N-terminale est une thiazolidine.

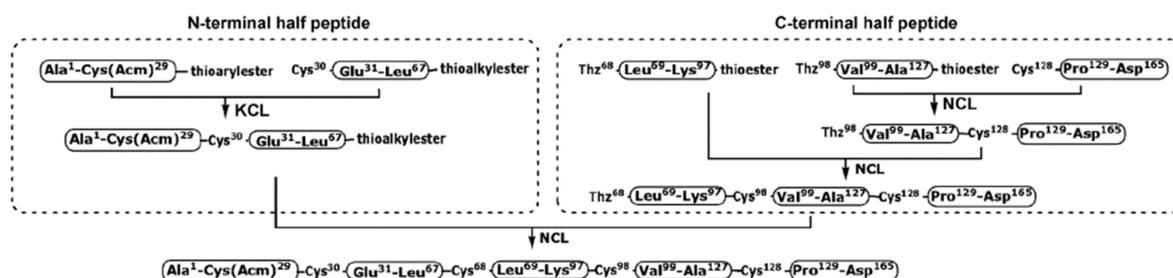


Figure 67 : Présentation de la stratégie de synthèse convergente de [Lys<sup>24</sup>, <sup>38</sup>, <sup>83</sup>] Erythropoïétine par Kent.

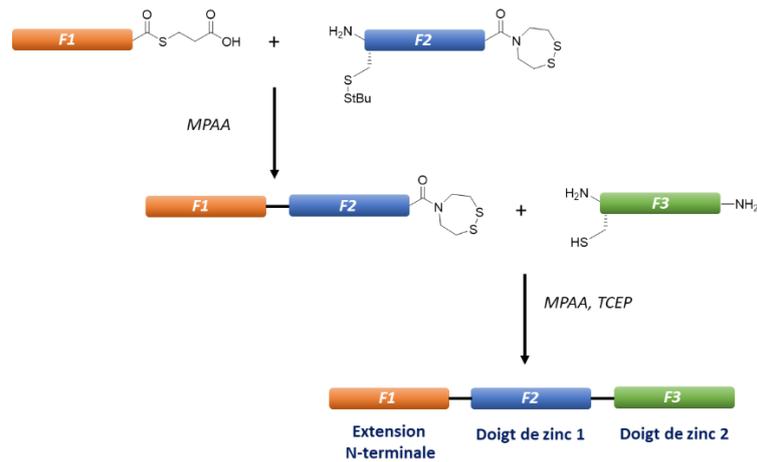
Des efforts d'optimisation pour la synthèse de la protéine Crambin dans l'équipe de Kent ont abouti à la synthèse de cette protéine par une stratégie d'assemblage « one-pot »<sup>121</sup>. À partir des mêmes fragments utilisés pour l'assemblage séquentiel, seul le groupement protecteur de la cystéine a été modifié au profit d'une thiazolidine. Contrairement au groupement Acm, celui-ci peut être déprotégé dans le milieu réactionnel, à la fin de la première ligation, par l'ajout de méthoxylamine à pH 4. Après deux heures de réaction, le dernier fragment N-terminale de la Crambin est ajouté et la réaction de ligation est laissée durant 20 heures à 37°C à pH 7.3. Une purification du milieu réactionnel a permis d'isoler la protéine Crambin avec un rendement global de 40%, supérieur à celui obtenu de manière séquentielle. Cependant A. Brick a remis l'utilisation du groupement Acm au centre des stratégies d'assemblage « one-pot » par fragments, en développant des conditions qui permettent sa déprotection en milieu aqueux. C'est en ayant recours à du chlorure de palladium (II) que la déprotection du groupement Acm est effectuée<sup>122,123</sup>. L'excès de palladium utilisé est retiré par l'ajout en fin de réaction de DTT. Grâce à cette nouvelle stratégie de synthèse, plusieurs protéines comportant entre autre des ponts disulfures ont pu être synthétisées<sup>124-126</sup>.

Un tout autre exemple d'assemblage « one-pot » nous vient de l'équipe de Tian sur la synthèse de la toxine Mambalgin-1 longue de 57 acides aminés, réalisée par assemblage de trois fragments peptidiques de N vers C<sup>127</sup>. La méthode de ligation employée est, cette fois-ci, la méthode des hydrazides. La cystéine présente sur la position N-terminale du fragment central est protégée au niveau de la chaîne principale par un azoture, la fonction thiol étant libre. Une première ligation de deux fragments peptidiques est réalisée en présence de MPAA. L'ajout d'un réducteur *in situ*, tel que le TCEP, permet de réduire l'azoture pour obtenir le peptide cystéinyl correspondant, qui réagit avec le dernier fragment. La protéine a été isolée après purification sur HPLC avec un rendement global de 35%.

La méthode des hydrazides a été utilisée par Liu en 2013 pour décrire une élégante synthèse d'une protéine de 140 acides aminés, la  $\alpha$ -synuclein<sup>128</sup>. L'assemblage est réalisé à partir de quatre fragments, mais dans le sens N vers C. En synthétisant les trois premiers fragments avec une fonction hydrazide en position C-terminale, il a pu faire un assemblage séquentiel. Cette stratégie a l'avantage de ne pas requérir de protection des cystéines. C'est par un simple enchaînement de ligations hydrazide/purification HPLC que la protéine a été synthétisée.

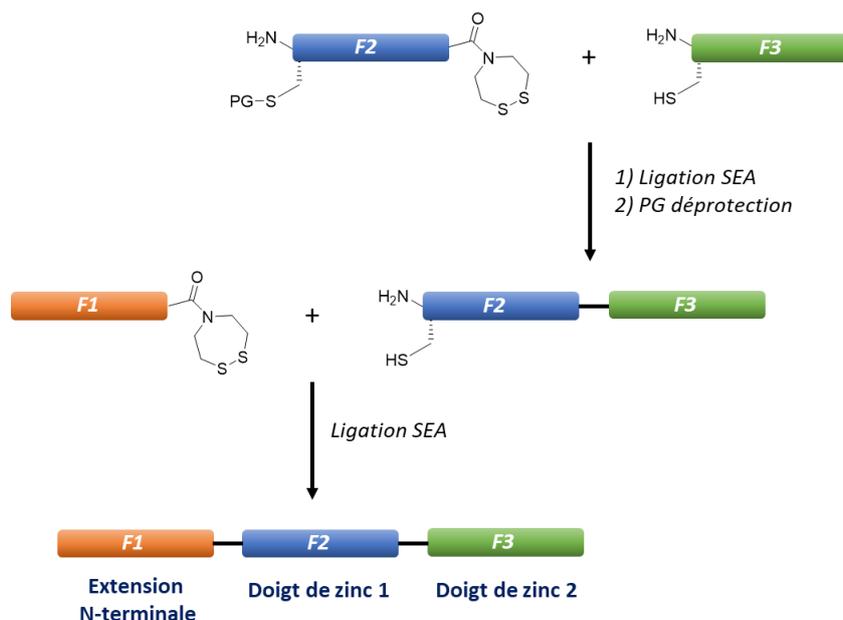
#### 1.4 - Stratégie d'assemblage de la protéine Zif12

En s'inspirant des travaux d'O. Melnyk<sup>129</sup> sur le développement d'une stratégie de synthèse d'une protéine par assemblage de trois fragments par ligation chimique native, G. Frémy a réalisé la synthèse de la protéine Zif12 de N vers C par une méthode d'assemblage « one-pot » (**Figure 68**). Le fragment F1 possède un thioester de MPA en position C-terminale, le second fragment F2 possède une fonction SEA<sup>off</sup> à son extrémité C-terminale et une cystéine N-terminale dont la chaîne latérale a été protégée par un groupement *StBu*. Lors de la première réaction de ligation chimique native en présence de MPAA, le thioester de MPA est converti *in situ* en un thioester de MPAA et la cystéine(*StBu*) est déprotégée. À la fin de la réaction, le dernier fragment F3 ainsi que du TCEP sont ajoutés dans le milieu réactionnel. La fonction SEA<sup>off</sup> est alors activée par la présence du réducteur (SEA<sup>on</sup>) et la formation du thioester de MPAA est possible. Une purification à la fin de cette seconde ligation a permis d'isoler la protéine Zif12 (**Figure 68**).



**Figure 68** : Stratégie d'assemblage de trois fragments pour la synthèse « one-pot » dans le sens N-terminale vers le C-terminale de la protéine Zif12 réalisée par Guillaume Frémy.

Lorsque Guillaume Frémy a appliqué cette méthode pour assembler la protéine fonctionnalisée par des chromophores, il a observé une réaction parasite sur le fragment central F2. En spectrométrie de masse, ce sous-produit possède un  $\Delta M$  de -121 dont la seule explication est une dégradation de la cystéine *S*tBu, qui aboutit à la formation d'une alanine, soit une perte d'un groupement S-*S*tBu. Ceci pose un réel problème pour la fonctionnalisation de la protéine Zif12. La segmentation de la protéine utilisée par G. Frémy, présentée sur la **Figure 63 B**, implique le greffage des chromophores au niveau du fragment central F2. Sa synthèse ayant posé plusieurs problèmes (évoqués précédemment) ce découpage ne semble pas adapté et nous conforte dans la nouvelle segmentation que nous avons réalisée **Figure 63 C**. La synthèse « one-pot » de la protéine Zif12 dans le sens N vers C est donc possible, cependant il a été décidé de faire l'assemblage dans le sens inverse, de C vers N. En effet, dans la suite du projet, nous allons fonctionnaliser la protéine Zif12 par des chromophores au niveau du fragment F1, qui devient donc le fragment à plus hautes valeurs ajoutées. Il semble donc logique d'introduire ce fragment dans les toutes dernières étapes de synthèses de la protéine. La stratégie globale est représentée **Figure 69**.



**Figure 69** : Stratégie d'assemblage de trois fragments pour la synthèse dans le sens C-terminale vers le N-terminale de la protéine Zif12.

La première étape est la réaction de ligation chimique native entre les fragments F2 et F3 par la méthode SEA. Au cours de cette étape, il est nécessaire de protéger la cystéine N-terminale du fragment central afin d'éviter la polymérisation de ce fragment. Cette protection peut être effectuée au niveau de la chaîne latérale, comme illustré sur la **Figure 69** ou de l'amine de la chaîne principale. Il s'en suit une étape de déprotection du groupement protecteur de la cystéine pour obtenir le peptide cystéinyl *H*-F2-F3, nécessaire pour la dernière étape de ligation SEA avec le fragment F1. Cette synthèse peut se faire de manière séquentielle. Cependant de nombreux exemples d'assemblages « one-pot » ont été rapportés dans la littérature, ce qui permet de s'affranchir des étapes de purifications, améliorant grandement les rendements de synthèses. Pour cela, nous avons choisi d'utiliser une thiazolidine (Thz) en position N-terminale du fragment central. Celle-ci se déprotège efficacement en milieu acide à pH 4 en présence de méthoxylamine.

La suite de ce chapitre est consacrée à l'assemblage des protéines Zif12-A et Zif12-B. Étant donné qu'elles ont été réalisées suivant les mêmes méthodes, pour des raisons de clarté dans le discours, les fragments notés F1 à F3 ne seront donc pas distingués selon qu'ils correspondent à la protéine Zif12-A ou Zif12-B.

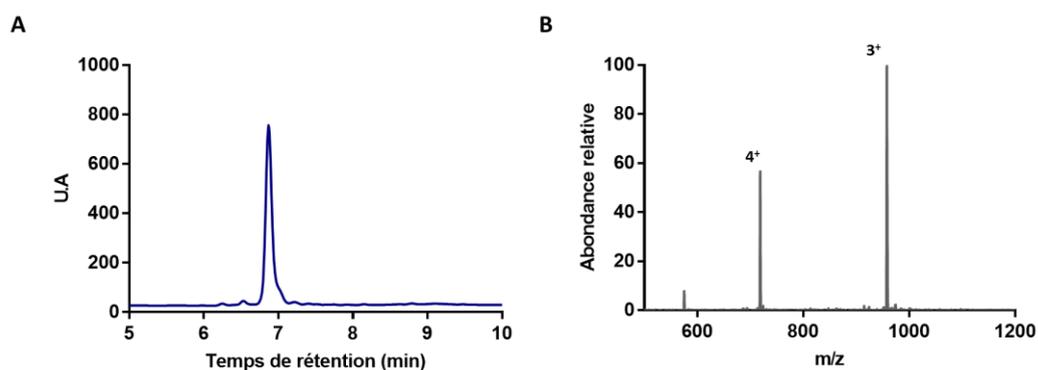
## 2 - Mise au point de l'assemblage de la partie reconnaissance d'ADN (F2+F3)

Notre objectif était de trouver une méthode efficace (peu ou pas de sous-produits avec de bons rendements) pour l'assemblage des fragments F2 et F3 en ouvrant si possible la porte à un assemblage « one-pot » des trois fragments.

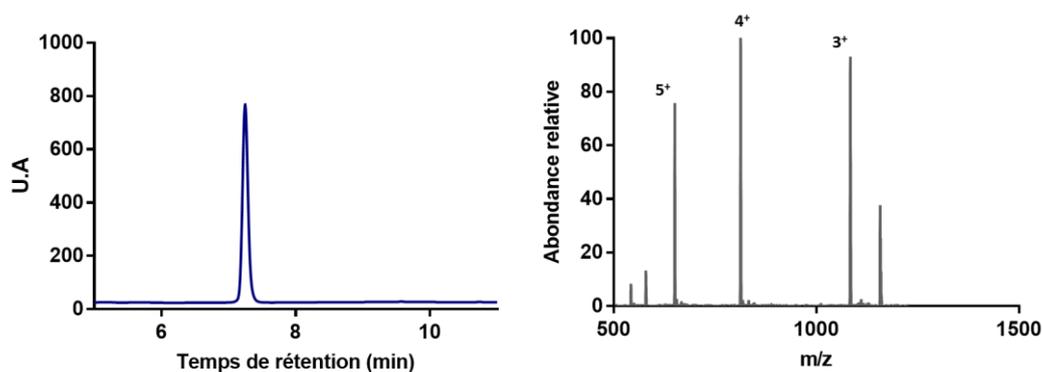
## 2.1 - Ligation des fragments *Thz-F2* et *F3* par ligation SEA, protection *Thz* en N-ter

### 2.1.a Synthèse des fragments *Thz-F2-SEA<sup>off</sup>* et *H-F3-NH<sub>2</sub>*

Nous avons commencé par réaliser la synthèse des fragments comportant les doigts de zinc. Le fragment F3 a été synthétisé sur une résine Rink Amide PEG afin d'obtenir une fonction amide en position C-terminale du peptide. Elle a été réalisée suivant une stratégie de synthèse Fmoc/*t*Bu. Les doigts de zinc peuvent parfois être difficiles à synthétiser avec des rendements de couplage diminuant dans les zones riches en histidines. Au laboratoire, nous avons pris l'habitude d'utiliser systématiquement la pseudoproline Fmoc-Phe-Ser( $\Psi^{\text{Me,Me}}\text{Pro}$ ) au niveau des motifs FS au milieu de la séquence pour casser les interactions entre les chaînes en croissance et pour améliorer les rendements de synthèse. Après traitement au TFA, cette pseudoproline redonne un motif FS naturel. Le premier acide aminé a été couplé à la main et la suite de l'élongation a été réalisée par le robot de synthèse. Nous avons utilisé une méthode de couplage sous microonde pour diminuer les temps de réaction, la résine Rink Amide PEG n'étant pas sensible à son usage. L'activateur utilisé pour le couplage des acides aminés est l'HCTU et la base la DIEA (détails de synthèses décrits dans la partie expérimentale). Après décrochage du peptide de la résine et déprotection des chaînes latérales dans du TFA avec des pièges à cations, la purification par HPLC nous a permis d'obtenir le fragment *H-F3-NH<sub>2</sub>*, caractérisé par HPLC et MS (**Figure 70**). Le fragment F2, quant à lui, a été synthétisé à partir d'une résine SEA. Le premier couplage est aussi réalisé à la main et l'élongation du peptide s'est effectuée dans les mêmes conditions de couplage que le fragment F3, mais à température ambiante et sans microonde, car la résine SEA est dérivée d'une résine chlorotriptyl qui peut être sensible aux microondes<sup>130</sup>. La cystéine en position N-terminale est remplacée par une thioproline (cystéines protégées sous forme thiazolidine). Le fragment *Thz-F2-SEA<sup>off</sup>* a pu être obtenu et caractérisé par HPLC et MS (**Figure 71**).



**Figure 70** : A) Chromatogramme HPLC à 214 nm du fragment *H-F3-NH<sub>2</sub>* purifié. B) Le spectre ESI-MS du fragment *H-F3-NH<sub>2</sub>* purifié (masse théorique calculée : 2870.27, masse déconvoluée obtenue : 2870.51).



**Figure 71** : A) Chromatogramme HPLC à 214 nm du fragment *Thz-F2-SEA<sup>off</sup>* purifié. B) Le spectre ESI-MS du fragment *Thz-F2-SEA<sup>off</sup>* purifié (masse théorique calculée : 3249.80, masse déconvoluée obtenue : 3248.97).

### 2.1.b Ligation des fragments *Thz-F2-SEA<sup>off</sup>* et *H-F3-NH<sub>2</sub>*

La réaction de ligation a été réalisée en boîte à gants afin d'éviter l'oxydation des cystéines lors de la réaction. La première ligation selon la méthode SEA, entre le peptide *Thz-F2-SEA<sup>off</sup>* et le peptide *H-F3-NH<sub>2</sub>*, a été effectuée à une concentration de 3.5 mM. La solution utilisée pour solubiliser les deux fragments contient du chlorure de guanidinium (6 M) dans du tampon phosphate (100 mM, pH 7.0). Du MPAA et du TCEP ont été ajoutés afin d'avoir des concentrations respectives de 0.1 mM. Le pH du milieu réactionnel est ajusté à 6.5 avant d'être dégazé sous argon. La dissolution des deux fragments est réalisée en boîte à gants, le volume de solution à ajouter est calculé pour ajuster la concentration du fragment *Thz-F2-SEA<sup>off</sup>* à 3.5 mM, le fragment F3 étant en léger excès de (1.1 éq). Le milieu réactionnel est ensuite placé sous agitation à 37°C. L'avancement de la réaction est suivi par LCMS. La réaction est lente et incomplète après 72h, de plus, le suivi par LCMS nous a révélé la formation d'un sous-produit que nous avons identifié comme étant un fragment *Thz-F2-F2-F3-NH<sub>2</sub>* (**Figure 72**). Il provient de l'hydrolyse partielle du Thz durant la ligation. La difficulté d'une synthèse par assemblage de trois fragments en « one-pot » est que la première réaction de ligation doit avoir un rendement très élevé avant d'ajouter le troisième fragment, pour éviter la formation du produit de ligation entre les deux fragments extrêmes. Après 72h de réaction, le peptide *Thz-F2-SEA<sup>on</sup>* est encore présent et le sous-produit *Thz-F2-F2-F3-NH<sub>2</sub>* est de mauvais augure pour la ligation avec le dernier fragment. Nous avons donc essayé de faire varier plusieurs paramètres pour améliorer la cinétique de réaction de ligation, afin que cette déprotection ne puisse s'effectuer. La réaction peut être accélérée (i) en augmentant la concentration des peptides (ii) en diminuant la concentration de guanidinium et en abaissant le pH. En jouant sur ces paramètres, nous n'avons cependant pas été en mesure d'empêcher la formation du sous-produit *Thz-F2-F2-F3-NH<sub>2</sub>*. Nous avons donc décidé d'abandonner l'assemblage « one-pot » de la protéine. Le nombre de sous-produits après ajout du dernier fragment ne peut que compliquer la purification de la protéine. Nous nous sommes donc résolus à faire la déprotection *in situ* de la thiazolidine pour réaliser une purification sur HPLC. Pour cela le pH du milieu réactionnel a été abaissé à pH 4 et 15 équivalents de méthoxylamine ont été ajoutés à partir d'une solution 1 M. Après deux heures de réaction, il reste encore une faible quantité du peptide *Thz-F2-F3-NH<sub>2</sub>*, mais il co-élue avec

le produit de la déprotection. Un nouvel ajout de méthoxylamine n'a pas permis de rendre la réaction de déprotection quantitative. Nous avons donc décidé de tenter la déprotection de la thiazolidine après une étape de purification du milieu réactionnel. L'utilisation d'un large excès de méthoxylamine a permis une déprotection quantitative, mais le sous-produit  $H-F2-F2-F3-NH_2$  co-éluait toujours avec le produit désiré.

Cette réaction de ligation soulève deux problèmes majeurs. Le premier concerne le temps de réaction, que ce soit pour la ligation ou la déprotection de la thiazolidine, qui sont particulièrement longs. Le second est lié au groupement protecteur de la cystéine du fragment F2. La déprotection partielle de la thiazolidine, lors de la réaction de ligation, a conduit à la formation d'un sous-produit qui ne peut être séparé du produit de ligation. La perspective d'une synthèse « one pot » de la protéine semble ainsi difficile par cette méthode. Compte tenu de toutes ces difficultés, nous avons décidé de nous tourner vers une ligation par la méthode des hydrazides, beaucoup plus rapide que la ligation SEA. Concernant le groupement protecteur de la cystéine, cette méthode synergise très bien avec l'utilisation d'un azoture de cystéine en position N-terminale du fragment F2 pour une stratégie d'assemblage « one pot » de trois fragments.

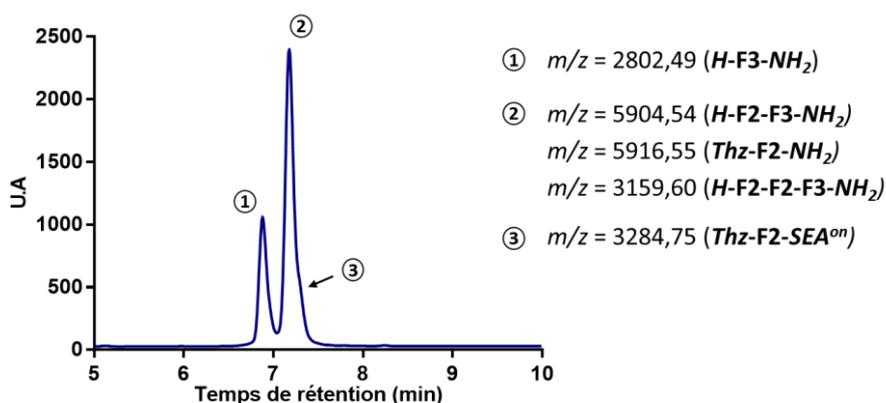


Figure 72 : Suivi LCMS à 214 nm obtenu après 48h de réaction de ligation entre les fragments  $Thz-F2-SEA^{off}$  et  $H-F3-NH_2$ .

## 2.2 - Ligation des fragments F2 et F3 par ligation hydrazide, protection $N_3$ en N-ter

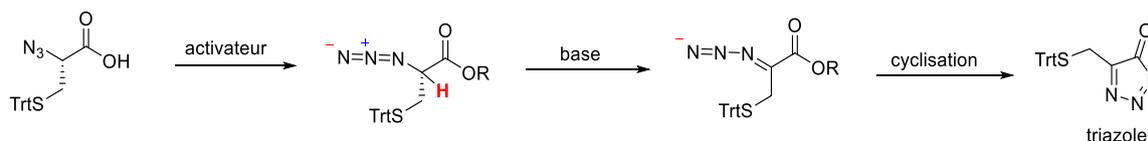
Nous avons décidé de reprendre la méthode d'assemblage de Tian, décrite plus haut pour résoudre nos deux problèmes : utilisation d'un azoture de cystéine ( $N_3$ -Cys) comme acide aminé protégé en position N-terminale du fragment F2 et l'utilisation de la ligation hydrazide pour accélérer la réaction.

### 2.2.a Synthèse de $N_3-F2-NHNH_2$

Nous avons dans un premier temps réalisé la synthèse du fragment  $N_3-F2-NHNH_2$ . La résine chlorotriptyl est traitée avec une solution d'hydrazine à 5% dans le DMF et le premier acide aminé est couplé à la main en utilisant de l'HCTU comme activateur. La suite de l'élongation est réalisée par le robot de synthèse avec un couplage à température ambiante sans microonde. Le couplage de la  $N_3$ -Cys(Trt)-OH (préalablement synthétisée<sup>131</sup>) a été réalisé selon les conditions décrites par la publication de l'équipe

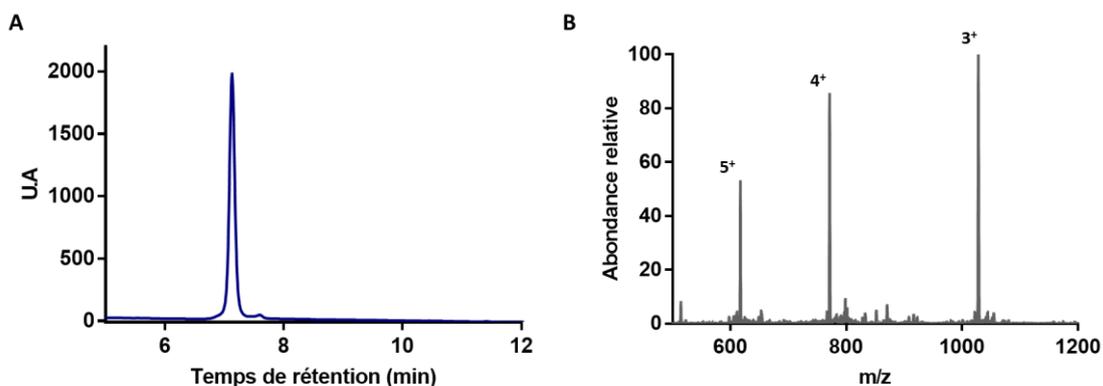
de Tian, à savoir 3.8 éq d'HCTU, 4 éq d'acide aminé et 8 éq de DIEA. Cependant aucun produit de couplage n'a été observé. Après avoir testé plusieurs activateurs en milieu basique (HOAT, PyBOP), et changé la DIEA pour une base moins forte (collidine), nous avons utilisé le mélange DIC/Oxyma qui ne requiert pas de base pour le couplage.

Le suivi de la réaction par LCMS a permis de montrer le couplage quantitatif de la  $N_3$ -Cys(Trt)-OH dans ces conditions. L'hypothèse que nous pouvons avancer est la formation d'un triazole en milieu basique, après l'étape d'activation de l'acide carboxylique (**Figure 73**).



**Figure 73** : Schéma de la cyclisation intramoléculaire de la  $N_3$ -cystéine après activation en condition basique.

Ce problème résolu, nous avons pu isoler la protéine  $N_3$ -F2-NH $NH_2$  (**Figure 74**) et réaliser la ligation chimique native entre ce fragment et le fragment  $H$ -F3-NH $_2$ .



**Figure 74** : A) Chromatogramme HPLC à 214 nm du fragment  $N_3$ -F2-NH $NH_2$  purifié. B) Le spectre ESI-MS du fragment  $N_3$ -F2-NH $NH_2$  purifié (masse théorique calculée : 3081.48, masse déconvoluée obtenue : 3080.98).

### 2.2.b Ligation avec le fragment $H$ -F3-NH $_2$

Le fragment  $N_3$ -F2-NH $NH_2$  a été dissous dans une solution de chlorure de guanidinium (6 M) dans du tampon phosphate (100 mM, pH 7.0) dont le pH a été ajusté à 3. La concentration du peptide a été ajustée à 3 mM. L'ajout de 1.5 éq de  $NaNO_2$  à  $-15^\circ C$  conduit à la conversion quantitative de l'hydrazine en azoture d'acyle en 10 min. Le MPAA (40 éq) a ensuite été ajouté pour former l'ester activé et détruire l'excès de  $NaNO_2$ . La conversion est totale après 45 min à température ambiante. Le fragment  $H$ -F3-NH $_2$  (1.1 éq) a ensuite été ajouté et la réaction est laissée à agiter à température ambiante à pH 6.5. Lors des premières réactions de ligation, nous avons observé une réduction partielle de l'azoture N-terminale (celui de la cystéine) qui aboutit à la formation d'un produit avec  $m/z = 3022$ . Il correspond probablement au produit de cyclisation intramoléculaire du fragment F2. Cette réactivité n'avait pas été mentionnée par l'équipe de Tian. Sur le suivi de la réaction de ligation après 30 min présenté **Figure 75** la proportion du produit de cyclisation avec le produit du couplage offre un faible rendement lors de

cette étape. Cela ne permet pas d'envisager la synthèse de la protéine avec une stratégie « one-pot ». Malgré la rapidité de la réaction de ligation, le sous-produit se confond avec le fragment  $N_3$ -F2-F3-NH<sub>2</sub>, la purification de ce fragment semble laborieuse et la synthèse séquentielle à plus grande échelle avec ce groupement protecteur semble difficile.

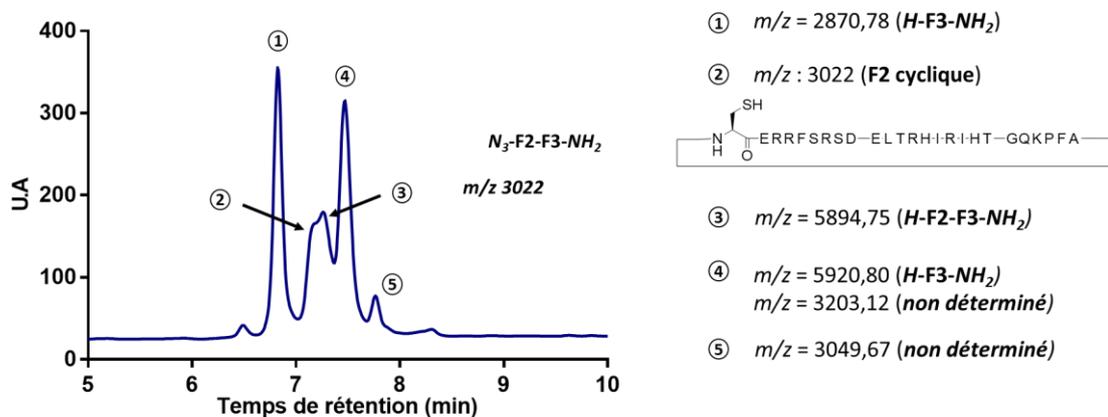


Figure 75 : Suivi de la réaction de ligation LCMS à 214 nm du fragment  $N_3$ -F2-NHNH<sub>2</sub> et  $H$ -F3-NH<sub>2</sub> après 30 min de réaction.

### 2.3 - Ligation des fragments F2 et F3 par ligation SEA, protection Acm en N-ter

Parmi les autres groupements protecteurs largement utilisés en ligation chimique native pour l'assemblage « one-pot » de plusieurs fragments, on trouve le groupement acétamidométhyle (Acm). Une déprotection quantitative est assurée par 30 équivalents de chlorure de palladium (II) avec 50 équivalents de chlorure de magnésium en seulement une dizaine de minutes. L'ajout de DTT à la fin de la réaction permet de complexer l'excès de palladium utilisé. Nous avons tenté cette stratégie de synthèse pour l'assemblage « one-pot » de la protéine Zif12 (Figure 76).

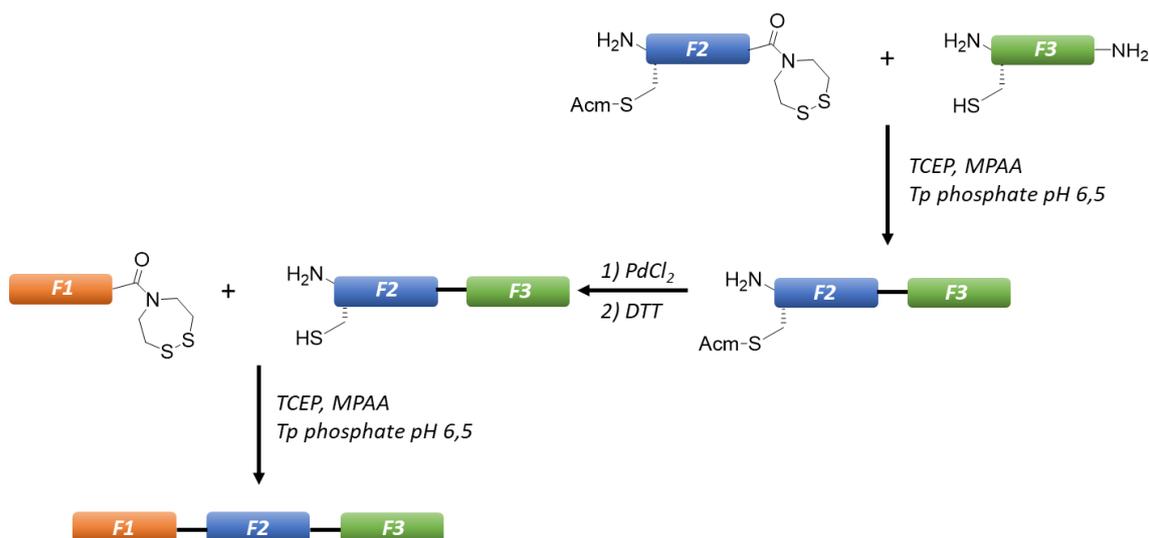
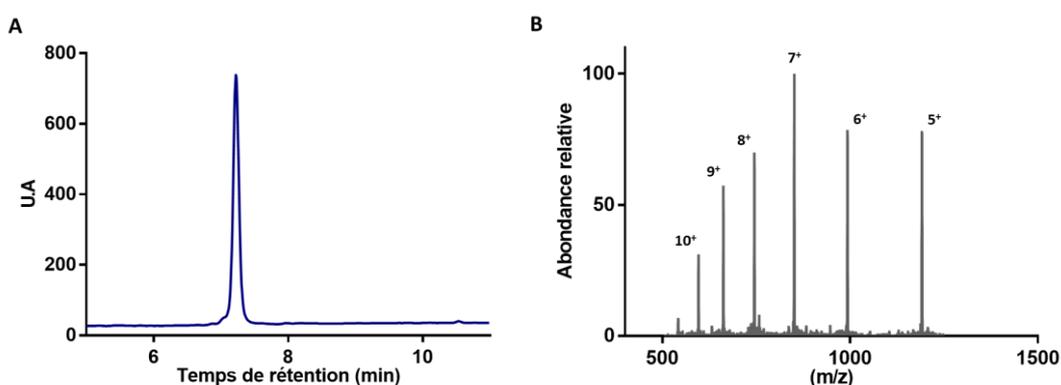


Figure 76 : Présentation de la stratégie de synthèse séquentielle à partir d'un fragment  $Acm$ -F2-SEA<sup>off</sup>.

Nous avons donc réalisé la synthèse du nouveau fragment  $Acm$ -F2-SEA<sup>off</sup>. La réaction de ligation chimique native entre ce fragment et le fragment  $H$ -F3-NH<sub>2</sub> a été effectuée dans les mêmes conditions

que précédemment. Bien que la réaction ne soit pas totale après 48h, l'analyse du milieu réactionnel par LCMS nous a montré qu'aucun sous-produit n'avait été formé et que le rendement est de 74% pour cette ligation (estimé par LCMS). Bien que la réaction ne soit pas totale, nous avons décidé de faire la déprotection du groupement AcM dans les conditions décrites par Brick<sup>122,123</sup>. Une solution contenant du chlorure de guanidinium (6 M) et du chlorure de palladium (II) (0.2 M) dans du tampon phosphate (100 mM, pH 7.0) a été préparée. L'ajout de 25  $\mu$ L de cette solution (15  $\mu$ eq de PdCl<sub>2</sub>) au milieu réactionnel a entraîné la formation quasi instantanée d'un précipité de couleur orange. Les ajouts de chlorure de magnésium et de DTT n'ont pas permis d'observer le produit désiré par LCMS. Malgré plusieurs tentatives, en variant la quantité de palladium (II) utilisée, l'ordre d'addition des réactifs que ce soit *in situ* ou après purification du milieu réactionnel, nous n'avons pu éviter ce phénomène. L'hypothèse que nous avons formulée est que la formation du palladium (II) par la protéine entraîne sa précipitation. Nous avons donc décidé de reprendre une synthèse séquentielle de la protéine Zif12. Une solution a été trouvée pour déprotéger ce groupement AcM en utilisant un sel d'argent<sup>120,132</sup>. Cependant cette méthode est incompatible avec le tampon phosphate utilisé lors de la ligation (formation d'un précipité de phosphate d'argent). Nous avons donc entrepris une déprotection après purification du milieu réactionnel. Après lyophilisation, le peptide *AcM-F2-F3-NH<sub>2</sub>* a été repris dans une solution aqueuse à 0.1% en TFA auquel a été ajoutée une solution de AgNO<sub>3</sub> (10 mM). Pour cette étape de déprotection, l'avancement de la réaction est suivi par LCMS en suivant le profil du spectre de masse, car les produits *H-F2-F3-NH<sub>2</sub>* et *AcM-F2-F3-NH<sub>2</sub>* co-éluent. Nous avons pu obtenir une déprotection quantitative à l'aide de 20 équivalents de nitrate d'argent après 12h de réaction. Le fragment *H-F2-F3-NH<sub>2</sub>* a été purifié et isolé (**Figure 77**). En augmentant la concentration d'AgNO<sub>3</sub> (une pointe de spatule), cette déprotection a été réalisée en l'espace d'une heure.



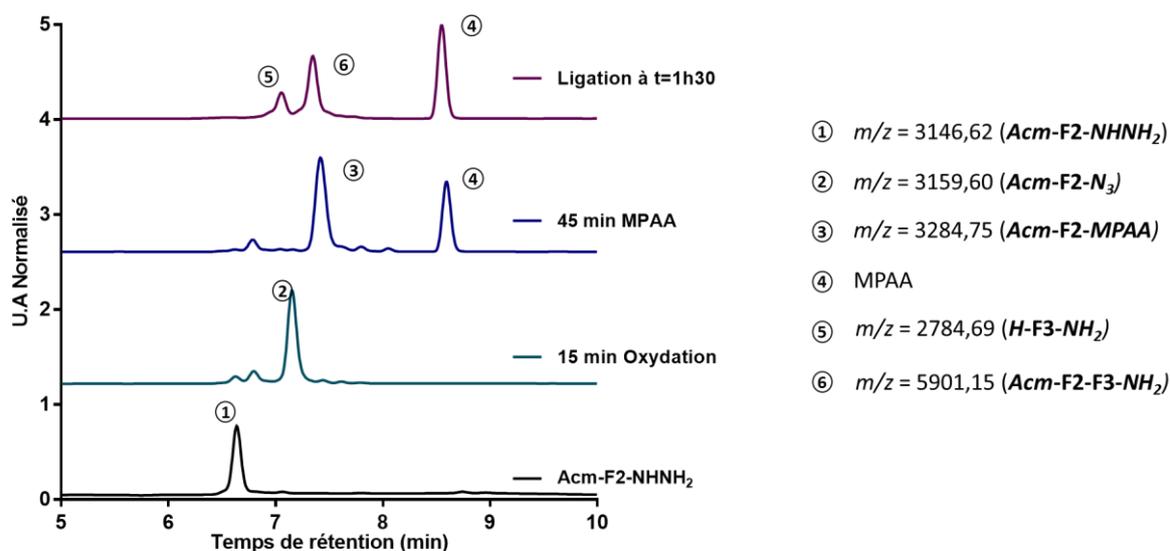
**Figure 77** : A) Chromatogramme HPLC à 214 nm du fragment *H-F2-F3-NH<sub>2</sub>* purifié. B) Le spectre ESI-MS du fragment *H-F2-F3-NH<sub>2</sub>* purifié (masse théorique calculée : 5951.77, masse déconvoluée obtenue : 5950.79).

La purification du milieu réactionnel est simplement réalisée à l'aide d'une cartouche d'extraction en phase solide Waters OASIS HLB, ce qui est plus rapide qu'une purification HPLC et permet de perdre moins de produit. Ceci nous a permis d'avoir suffisamment de produit pour la suite de la synthèse avec un rendement sur les deux étapes de 59%. Cependant le temps de réaction de ligation reste long pour des réactions de ligations chimiques natives.

De nombreuses publications portent sur le développement et l'amélioration de méthodes de ligation chimique native avec pour objectifs d'accélérer la cinétique de réaction, jouant notamment sur la nature du crypto-thioester. À l'image des travaux de l'équipe de Melnyk qui travaille sur des crypto-sélénoester<sup>117</sup> où les atomes de soufre sont remplacés par deux atomes de sélénium. À titre d'exemple, le crypto-thioester SEA que nous utilisons est moins réactif que son homologue sélénié (SeEA) d'un facteur 14. Cependant les rendements de synthèse de cet homologue sont peu élevés, 21% pour la synthèse du crypto-sélénoester (deux étapes) et 34% pour son couplage en position C-terminale. Cette stratégie rajoute plusieurs étapes de synthèses dans un projet qui en comporte déjà un certain nombre.

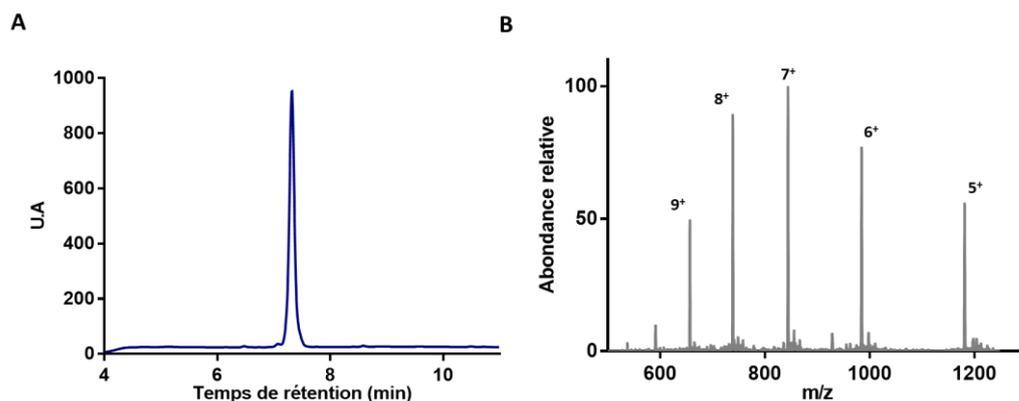
#### 2.4 - Ligation du fragment F2 et F3 par ligation hydrazide, protection Acm en N-ter

Nous avons pris le parti de reprendre la ligation par la méthode des hydrazides avec laquelle nous nous sommes familiarisés auparavant. La synthèse d'un fragment *Acm-F2-NHNH<sub>2</sub>* ainsi que la réaction de ligation avec le fragment F3 ont été réalisées selon les mêmes conditions que précédemment. Après 1h30 de réaction, le milieu réactionnel est purifié par HPLC. Nous avons pu synthétiser et purifier le fragment *Acm-F2-F3-NH<sub>2</sub>* en trois heures, à compter de la dissolution du fragment *Acm-F2-NHNH<sub>2</sub>*, au lieu de 48h avec la méthode de ligation SEA (**Figure 78**).



**Figure 78** : Suivi de réaction de ligation selon la méthode des hydrazides par HPLC à 214 nm des fragments *Acm-F2-NHNH<sub>2</sub>* et *H-F3-NH<sub>2</sub>*.

La déprotection du groupement Acm a été réalisée avec un large excès de nitrate d'argent et le fragment peptidique *H-F2-F3-NH<sub>2</sub>* a été purifié et caractérisé par HPLC et spectrométrie de masse. Cette méthode a été appliquée pour la synthèse de la partie reconnaissance d'ADN de la protéine Zif12-B dont le profil HPLC et le spectre de masse sont présentés **Figure 79**.



**Figure 79** : A) Chromatogramme HPLC à 214 nm du fragment  $H-F2-F3-NH_2$  (Zif12-B) purifié. B) Le spectre ESI-MS du fragment  $H-F2-F3-NH_2$  (Zif12-B) purifié (masse théorique calculée : 5830.66, masse déconvoluée obtenue : 5829.25).

Après plusieurs étapes d'optimisation, nous avons été en mesure de trouver une méthode efficace pour la synthèse du fragment  $H-F2-F3-NH_2$ . Le recours à la méthode de ligation hydrazide a considérablement réduit le temps de réaction et le groupement protecteur AcM de la cystéine nous a permis, après purification du milieu réactionnel, d'obtenir ce fragment avec une grande pureté. Le seul inconvénient de cette stratégie est que nous sommes obligés d'effectuer la synthèse de la protéine de manière séquentielle. Cependant notre méthode permet de synthétiser le fragment  $H-F2-F3-NH_2$  sur une échelle de synthèse comprise entre 5 et 10 mg.

### 3 - Assemblage de la partie reconnaissance d'ADN avec l'extension N-terminale

#### 3.1 - Synthèse du fragment F1-SEA

La stratégie initiale que nous avons choisie d'utiliser fait intervenir une ligation de type SEA entre le fragment F1 et le peptide  $H-F2-F3-NH_2$ . Le fragment F1 a donc été synthétisé par un robot de synthèse sur une résine SEA-2Cl-Trt-PS. La cystéine<sup>107</sup> présente dans la séquence doit avoir sa chaîne latérale protégée par un groupement *StBu* afin d'éviter l'oxydation de la fonction thiol lors de l'oxydation de la résine SEA<sup>on</sup>. En vue d'une fonctionnalisation ultérieure par les chromophores, le fragment F1 a été synthétisé en utilisant une lysine protégée par un groupement Alloc en position 106 (**Chapitre I, Figure 55**). À la fin de l'élongation de la protéine la position N-terminale est acétylée et le groupement Alloc déprotégé avant de faire une coupure du peptide et purifier le fragment  $Ac-F1-SEA^{off}$  (**Figure 80**).

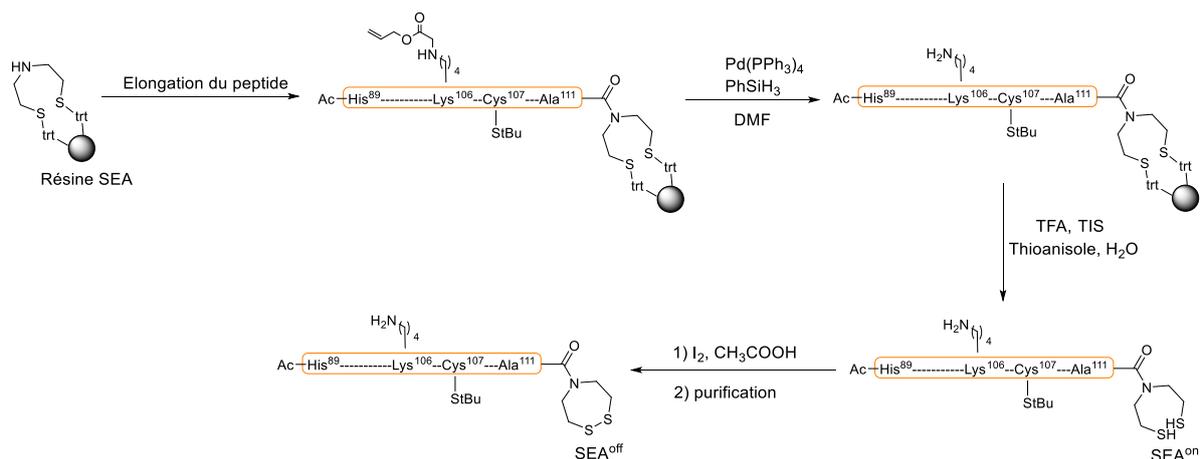


Figure 80 : Schéma de la synthèse du fragment Ac-F1-SEA<sup>off</sup>.

La déprotection sélective du groupement Alloc est réalisée sur la résine en présence de Pd(0) (1 éq, dans ce cas, Pd(PPh<sub>3</sub>)<sub>4</sub> est utilisé) et de phénylsilane (25 éq) dans du DCM anhydre sous argon. La déprotection des autres groupements ainsi que la coupure du peptide de la résine sont assurées par une solution de TFA comportant des capteurs de cations : TIS, thioanisole et eau (91.7/3.3/3.3/1.7). Après 2h de réaction, le TFA est évaporé et le peptide est précipité dans de l'éther glacé. La fonction SEA<sup>on</sup> est oxydée à l'aide d'une solution aqueuse d'acide acétique à 10% et par une solution de diiode (200 mM, DMSO). Le peptide Ac-F1-SEA<sup>off</sup> est obtenu pur après HPLC et lyophilisation, il a été caractérisé par spectrométrie de masse (Figure 81).

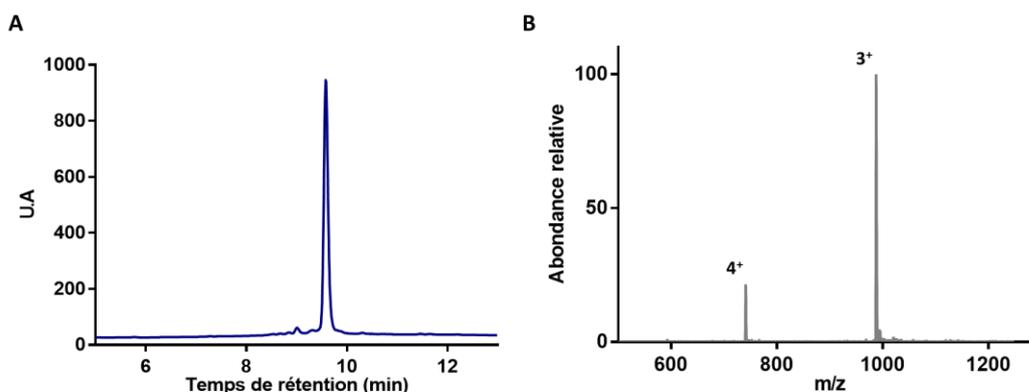
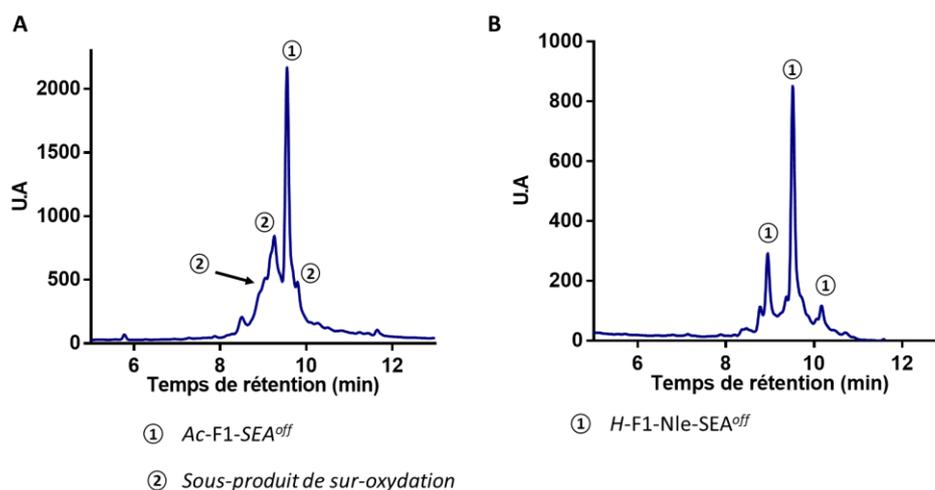


Figure 81 : A) Chromatogramme HPLC à 214 nm du fragment Ac-F1-SEA<sup>off</sup> purifié. B) Le spectre ESI-MS du fragment Ac-F1-SEA<sup>off</sup> purifié (masse théorique calculée : 2959.73, masse déconvoluée obtenue : 2960.90).

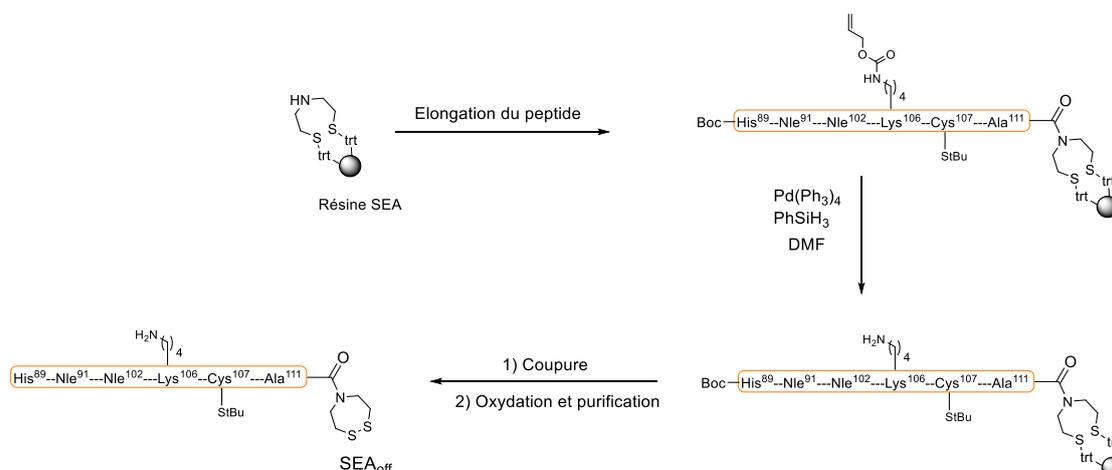
### 3.2 - Synthèse du fragment H-F1-Nle-SEA

Une remarque doit être faite concernant la fonctionnalisation de l'extension N-terminale discutée paragraphe Chapitre III, III-3. Nous avons rencontré de grosses difficultés lors du greffage des partenaires du FRET, entre autres le pro-ligand PL<sub>I</sub><sup>CC</sup> sur le fragment F1. Ce problème apparaît déjà lors de la synthèse du fragment Ac-F1-SEA<sup>off</sup>, présenté Figure 82. L'analyse par LCMS du produit brut a révélé la présence de nombreux produits d'oxydation, essentiellement liée à l'oxydation de la fonction thioéther des méthionines en sulfoxyde. Les rendements obtenus sont ainsi très faibles.



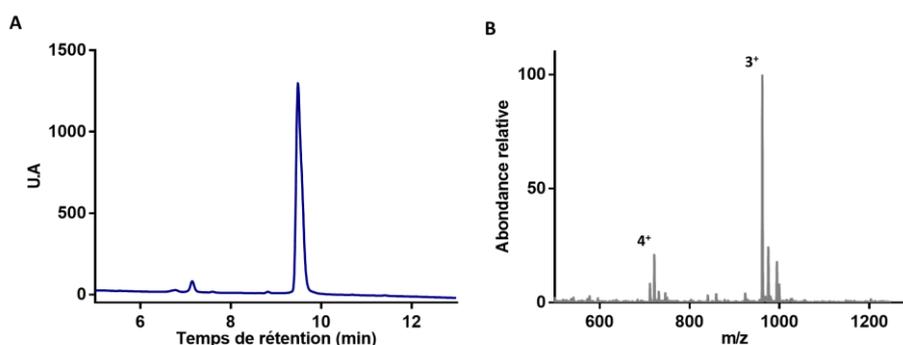
**Figure 82** : Chromatogramme HPLC à 214 nm des fragments  $H-F1-SEA^{off}$  (A) et  $H-F1-Nle-SEA^{off}$  (B) après oxydation de la résine SEA.

Nous avons décidé de remplacer les deux méthionines présentes dans la séquence du fragment F1 aux positions 91 et 102 par un acide aminé non naturel, la norleucine. L'unique différence est que la norleucine ne possède pas de fonction thioether mais une simple chaîne alkyle. On passe donc d'un groupement  $R-S-Me$  à  $R-CH_2-Me$  dont le volume et la forme sont comparables (isostère). Le chromatogramme obtenu après cette modification présente donc bien moins de produits de sur-oxydation (**Figure 82**). Ce changement se situe sur la partie N-terminale de la protéine dont la séquence a été optimisée pour obtenir un maximum d'interaction protéine/protéine, cette modification peut donc déstabiliser la formation du dimère autour de l'ADN et devra faire l'objet d'une vérification. Afin de nous rapprocher un peu plus de la protéine Zif12 nous avons décidé pour cette nouvelle synthèse du fragment F1 d'utiliser une Boc-His-OH en fin de synthèse pour obtenir une amine libre après coupure au TFA du peptide et déprotection du groupement Boc. La synthèse de ce nouveau fragment  $H-F1-Nle-SEA^{off}$  est présentée **Figure 83**.



**Figure 83** : Synthèse du fragment  $H-F1-Nle-SEA^{off}$ .

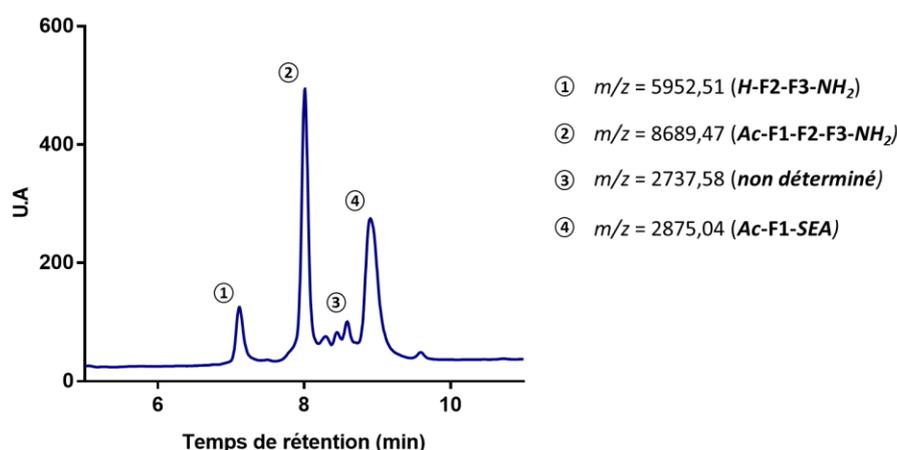
Ce nouveau fragment a été synthétisé et purifié exactement de la même manière que le fragment  $Ac-F1-SEA^{off}$  et sera noté  $H-F1-Nle-SEA^{off}$ . Son spectre de masse ainsi que le chromatogramme obtenu après purification sont présentés **Figure 84**.



**Figure 84** : A) Chromatogramme HPLC à 214 nm du fragment  $H-F1-Nle-SEA^{off}$  purifié. B) Le spectre ESI-MS du fragment  $H-F1-Nle-SEA^{off}$  purifié (masse théorique calculée : 2883.60, masse déconvoluée obtenue : 2883.01).

### 3.3 - Assemblage de l'extension N-terminale et du domaine de reconnaissance

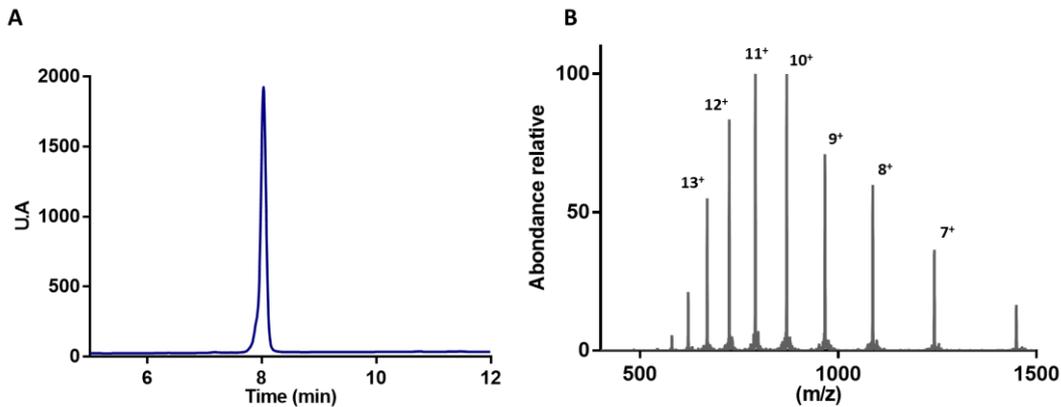
La réaction de ligation chimique native entre les fragments  $Ac-F1-SEA^{off}$  et  $H-F2-F3-NH_2$  a été réalisée à pH 6.5 dans une solution de tampon phosphate (100 mM) contenant du chlorure de guanidinium (6 M), du TCEP (10 éq) et du MPAA (10 éq). La concentration du fragment  $H-F2-F3-NH_2$  a été fixée à 7 mM et un léger excès du fragment F1 a été ajouté (1.3 éq). La réaction est laissée sous agitation à 37°C. Le chromatogramme obtenu après 30 heures de réaction est présenté **Figure 85**. Le suivi effectué après 48 heures de réaction ne présentant aucune évolution dans la réaction, une purification par HPLC a été réalisée.



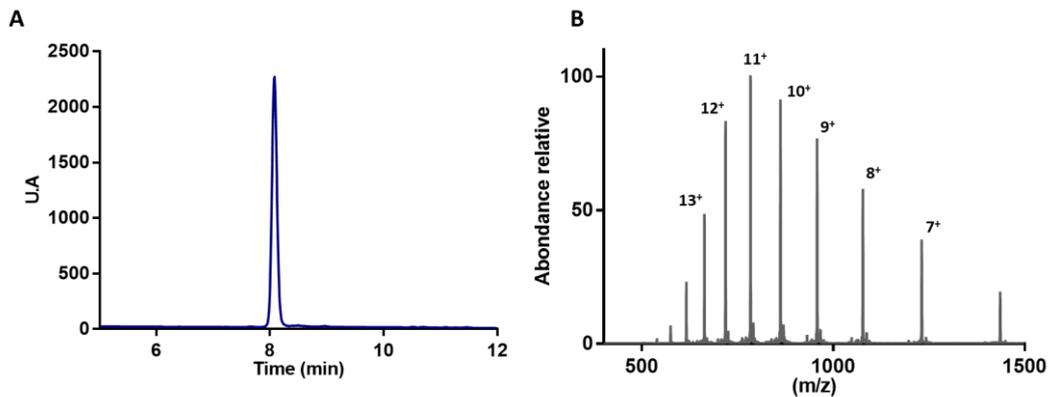
**Figure 85** : Suivi LCMS de la réaction de ligation à 214 nm entre  $Ac-F1-SEA^{off}$  et  $H-F2-F3-NH_2$  après 48h de réaction.

Cette réaction de ligation chimique native a été effectuée pour la synthèse des protéines Zif12-A, sa version avec les norleucines Zif12-Nle-A et Zif12-B avec des rendements respectifs de 77%, 85% et 88%. Le chromatogramme et le spectre de masse des protéines isolées et lyophilisées sont présentés respectivement **Figure 86**, **Figure 87** et **Figure 88**. La protéine Zif12-B présente un sous-produit avec

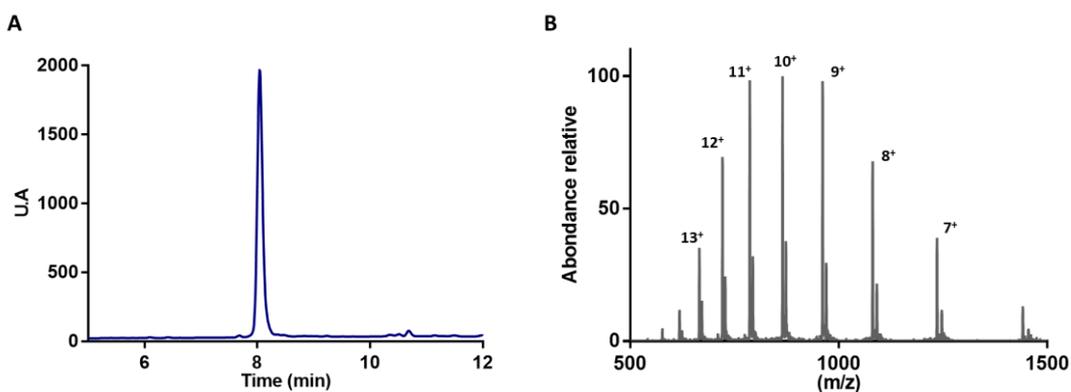
un  $\Delta+80$  qui n'a pas pu être séparé, la déprotection du groupement Alloc peut induire un  $\Delta+40$  par allylation, cet écart de masse pourrait donc provenir d'une double allylation.



**Figure 86** : A) Chromatogramme HPLC à 214 nm de la protéine Zif12-A purifiée. B) Le spectre ESI-MS de la protéine Zif12-A purifié (masse théorique calculée : 8690.06, masse déconvoluée obtenue : 8689.26).



**Figure 87** : A) Chromatogramme HPLC à 214 nm de la protéine Zif12-Nle-A purifiée. B) Le spectre ESI-MS de la protéine Zif12-Nle-A purifié (masse théorique calculée : 8611.96, masse déconvoluée obtenue : 8611.06).



**Figure 88** : A) Chromatogramme HPLC à 214 nm de la protéine Zif12-B purifiée. B) Le spectre ESI-MS de la protéine Zif12-B purifié (masse théorique calculée : 8642.06, masse déconvoluée obtenue : 8643.25).

#### 4 - Formation des doigts de zinc

Pour reconnaître leur séquence cible d'ADN double brin, les protéines doivent être actives et donc sous la forme doigts de zinc. Les protéines métallées au zinc seront pour la suite de ce manuscrit notées sous la forme Zif12\*-A.

La métallation des trois protéines a été réalisée par dissolution des protéines purifiées et lyophilisées dans de l'eau (**Figure 89**). Pour le dosage de ces protéines par spectroscopie UV nous allons utiliser l'absorbance à 280 nm provenant des deux tyrosines présentes dans la séquence. L'absorbance à 280 nm est corrigée par l'absorbance à 330 nm pour éliminer les contributions de diffusion de la lumière (équation<sup>133</sup> (6)) et la concentration en protéine a été calculée selon l'équation (7) avec  $\epsilon_{\text{Tyr}280} = 1\,490 \text{ L}\cdot\text{M}^{-1}\cdot\text{cm}^{-1}$ .

$$A_{280}Cor = A_{280} - 2 \times A_{330} \quad (6)$$

$$[protéine] = \frac{A_{280}Cor}{2 \times \epsilon_{\text{Tyr}280}} \quad (7)$$

Du tampon HEPES (100 mM, pH 7.0) est ajouté pour obtenir une solution à 10% (v/v) ainsi que du TCEP (250  $\mu\text{M}$ ) pour réduire d'éventuels ponts disulfure intramoléculaires. La métallation est assurée par l'ajout de 5 éq de  $\text{Zn}(\text{ClO}_4)_2$  et la protéine est aliquotée et lyophilisée pour être conservée à  $-20^\circ\text{C}$ . Les concentrations des protéines Zif12\*-A, Zif12\*-Nle-A et Zif12\*-B avant lyophilisation sont comprises entre 15 et 35  $\mu\text{M}$ .

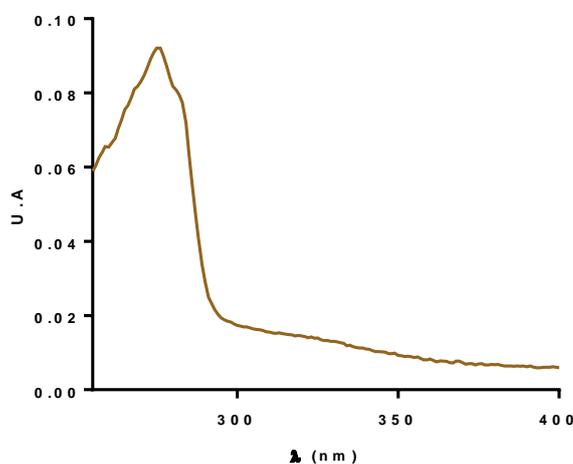


Figure 89 : Spectres d'absorbance de la protéine Zif12\*-B.

## II - Étude des interactions des protéines Zif12-A et Zif12-B avec l'ADN

Pour mettre en évidence l'interaction des deux protéines à doigts de zinc qui ne comportent aucun chromophore, avec une séquence d'ADN double brin, il a été choisi d'utiliser la technique EMSA (Electrophoretic Mobility Shift Assay), classiquement utilisée en biologie moléculaire. Ceci nous permettra par ailleurs d'avoir un point de comparaison avec les travaux de Wang et Pabo<sup>54</sup>. Nous avons effectué plusieurs modifications de la protéine LRZif initialement décrite par Wang et Pabo :

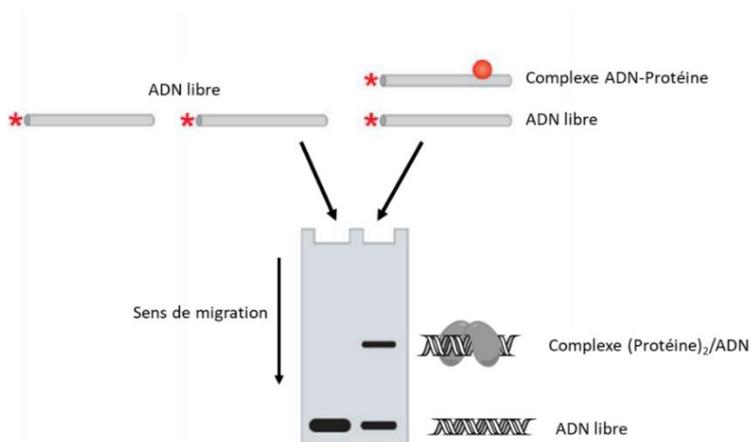
- Remplacement de la Ser<sup>111</sup> par Ala
- Remplacement de la Ala<sup>106</sup> par Lys
- Remplacement de Asp<sup>113</sup> par Glu

Nous devons donc nous assurer, dans un premier temps, que ces modifications n'affectent pas les propriétés de dimérisation des protéines. Dans un second temps nous vérifierons que le remplacement des méthionines aux positions 91 et 102 et l'amine secondaire N-terminale de la protéine Zif12\*-Nle-A permet toujours la fixation des protéines sur la séquence d'ADN décrite par Wang et Pabo avec une affinité similaire à son homologue Zif12\*-A.

## 1 - La technique EMSA

### *1.1 - Principe de la technique EMSA*

La technique EMSA est une technique basée sur la différence de mobilité électrophorétique d'un composé ionisé à travers un gel de polyacrylamide<sup>134</sup> lorsqu'il est libre ou en interaction avec un autre composé. Pour favoriser le maintien du complexe pendant l'analyse, la séparation doit être réalisée à basse température en conditions non dénaturantes. Cette technique simple à mettre en œuvre, est à la fois qualitative et quantitative, très répandue pour l'étude d'interactions de protéines ou de ligands avec l'ADN. Dans notre cas elle nous permettra de facilement distinguer la formation d'un dimère autour de la séquence d'ADN cible. Sa migration sera ralentie par rapport à l'ADN libre, la formation du dimère augmentant considérablement la masse du système (**Figure 90**). L'expérience se divise en trois étapes. La première est l'incubation de la protéine avec le double brin d'ADN, suivie de la migration sur le gel d'électrophorèse et enfin l'étape de révélation du complexe formé basée sur la détection de l'ADN. L'ADN utilisé à concentration fixe sera incubé à 4°C en présence de quantités variables de protéines. L'utilisation d'un marquage fluorescent de l'ADN a été écartée en raison du risque d'interactions parasites avec les chromophores utilisés pour décorer les protéines pour les applications spectroscopiques. Nous avons donc choisi d'utiliser la radioactivité du <sup>32</sup>P incorporé à l'extrémité 5'OH d'un oligonucléotide simple brin par transfert du phosphate de l'ATP  $\gamma^{32}\text{P}$  à l'aide d'une polynucléotide kinase. Ce marquage à l'aide d'un phosphate terminal ne modifie pas les propriétés de reconnaissance de l'ADN. Ce brin d'ADN marqué est ensuite hybridé avec le brin complémentaire en léger excès (1.1 éq). L'étape de migration requiert quelques précautions. Le courant imposé pour la migration du gel d'électrophorèse est une source de chaleur susceptible de dissocier le complexe formé lors de l'incubation. La migration se fera à une puissance modérée et utilisera un appareil doté d'un système de circulation réfrigérant permettant de maintenir une température d'environ 10°C au niveau de la zone de migration dans le gel. Enfin, le tampon utilisé pour former le gel et pour la migration ne contiendra pas de composés susceptibles de dénaturer le complexe (conditions natives). Le tampon Tris-Glycine est couramment utilisé pour cette application.



**Figure 90** : Principe du gel retard EMSA, à gauche l'échantillon témoin d'ADN seul et à droite l'échantillon d'ADN en présence d'une protéine. Le complexe protéine/ADN migre plus lentement que l'ADN seul.

### 1.2 - Interaction des protéines avec le duplex d'ADN

Afin de mettre au point les conditions expérimentales et d'avoir un point de comparaison, nous avons utilisé la protéine Zif12 correspondant à la séquence de Wang et Pabo, disponible au laboratoire. Cette protéine nous avait été gracieusement fournie par la société Genepep®, où elle avait été synthétisée par le Dr Laurent Raibaut, ancien post-doc de l'équipe. Elle sera notée LRZif. Plusieurs séquences d'ADN double brin longues de 38 paires de bases ont été synthétisées au laboratoire pour étudier dans un premier temps l'interaction des protéines avec une séquence palindromique d'ADN (AA ou BB). Dans un second temps un mélange des deux protéines Zif12\*-A et Zif12\*-B sera effectué pour le ciblage d'une séquence non palindromique d'ADN double brin (ADN-AB). Enfin un dernier ADN sera utilisé, noté ADN-rdm, qui possède les mêmes paires de bases externes que l'ADN-AB mais dont la partie centrale est organisée de manière à ne contenir aucune cible pour les doigts de zinc. Sur la **Figure 91** est représenté l'ensemble des séquences d'ADN double brins que nous allons utiliser, ainsi que les triplets de reconnaissance des différents doigts de zinc.

**ADN-AA** : 5' -GGT TGC AGT **GGG CGC GCC CAC** AGT ACT TGA ACG TAA CG-3'  
 3' -CCA ACG TCA **CCC GCG CGG GTG** TCA TGA ACT TGC ATT GC-5'

**ADN-BB** : 5' -GGT TGC AGG **AAG GGC CCT TCC** AGT ACT TGA ACG TAA CG-3'  
 3' -CCA ACG TCC **TTC CCG GGA AGG** TCA TGA ACT TGC ATT GC-5'

**ADN-AB** : 5' -GGT TGC AGG **AAG GGC GCC CAC** AGT ACT TGA ACG TAA CG-3'  
 3' -CCA ACG TCC **TTC CCG CGG GTG** TCA TGA ACT TGC ATT GC-5'

**ADN-rdm** : 5' -GGT TGC AGC **AGC GAC CGA GCC** AGT ACT TGA ACG TAA CG-3'  
 3' -CCA ACG TCG **ACG CTG GCT CGG** TCA TGA ACT TGC ATT GC-5'

Domaine de reconnaissance des doigts de zinc

Zif12-A :	(F2) <b>RSDELTR</b>	(F3) <b>RSDHLTT</b>
Zif12-B :	(F2) <b>RSEHLTR</b>	(F3) <b>QSGNLSR</b>

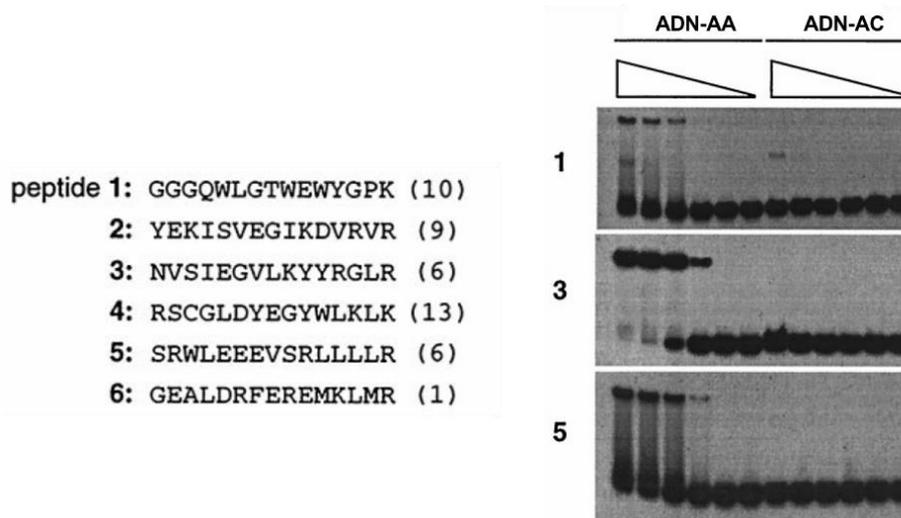
**Figure 91** : Séquences d'ADN double brin synthétisées. En couleurs sont représentées les séquences des triplets ciblés par les différents doigts de zinc de Zif12-A et Zif12-B.

## 2 - Rappel des travaux de Wang et Pabo

Pour mieux appréhender les résultats que nous avons obtenus lors de l'interaction des protéines avec les différents ADN par gel retard, un retour sur les travaux effectués par Wang et Pabo sur la protéine Zif12 est nécessaire. Le point de départ de leur étude est la protéine Zif268, mais plus précisément les deux premiers doigts de zinc. Pour former un dimère autour de la séquence cible 5'-TGGGCGCGCCCA-3', ils ont ajouté une extension N-terminale de 15 aa dont la séquence a été optimisée par phage display. Au total ce sont six protéines qui ont été sélectionnées. Les extensions N-terminale de ces protéines, présentées **Figure 92**, ne montrent aucune homologie entre elles. La formation du dimère autour de l'ADN par les protéines 1, 3 et 5 a été analysée par gel retard, deux séquences d'ADN ont été utilisées. Un premier ADN contient deux fois la cible notée ADN-AA (5'-TGGGCGCGCCCA-3') et le second ne contient qu'une seule cible (5'-TGGGCGTATGCT-3'). L'incubation des protéines avec l'ADN a été réalisée à 4°C dans une solution tamponnée complexe (15 mM de tampon HEPES pH 7.8, 60 mM AcOK, 60 mM glutamate K, 5 mM MgCl<sub>2</sub>, 20 mM ZnSO<sub>4</sub>, 5% glycérol, 0.1% nonidet P-40, 1 mM DTT, 0.1 mg/mL de BSA acétylée) pendant 1.5 à 16 heures (selon le peptide) à différentes concentrations en protéine. La concentration d'ADN a été fixée à 1/100 de la valeur à laquelle la protéine se fixe à 50%. Le gel d'électrophorèse a été effectué à 4°C avec un gel à 8% d'acrylamide contenant 2.5% de glycérol, le tampon de migration étant composé de 25 mM Tris, 190 mM de glycine et 1 mM d'EDTA. Les gels obtenus après l'étape de révélation sont présentés **Figure 92**. Pour chacune des protéines, on ne voit pas l'apparition d'une bande retardée lorsqu'une seule cible est présente sur la séquence d'ADN 5'-TGGGCGTATGCT-3'. Cependant, lorsque les deux cibles sont présentes, on peut voir l'apparition d'une bande retardée. Cette bande a été interprétée par les auteurs comme étant la bande correspondant à la formation du dimère autour de l'ADN, bien que la formation d'un complexe 2:1 n'ait pas été démontrée.

Confortés dans l'idée que les interactions protéine/protéine permettent la formation d'un dimère ils ont optimisé cette extension N-terminale par blocs de 5 aa pour obtenir la protéine Zif12 (**Chapitre I, II-2-3d**) présentant un gain d'affinité de l'ordre de 1000. La séquence optimisée de cette extension

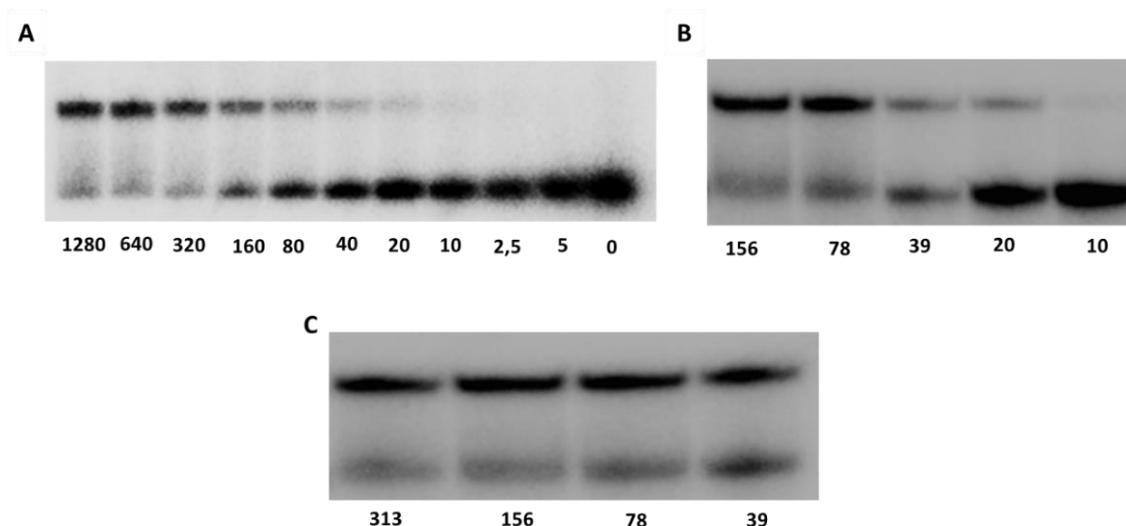
N-terminale est HPMNLLNYVVPKMR. Ce n'est que deux années plus tard que la formation d'un complexe 2:1 a pu être vérifiée par l'obtention de la structure RX du complexe (Zif12)<sub>2</sub>/ADN<sup>54</sup>.



**Figure 92 :** (Gauche) Séquences de l'extension N-terminale des protéines isolées par phage display. (Droite) Gels d'électrophorèses obtenus après interaction des protéines 1, 3 et 5 avec les ADN-AA et ADN-AC. La concentration des protéines est de 2.5  $\mu$ M, 250 nM, 25 nM, 2.5 nM, 250 pM et le dernier puit correspond à l'ADN seul.

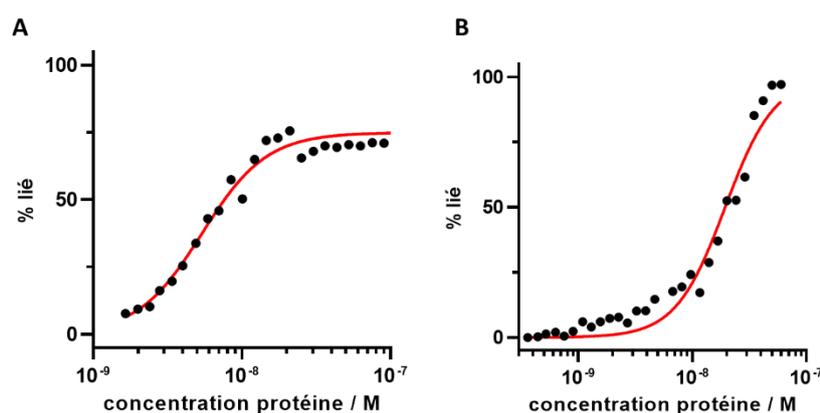
### 3 - Interaction des protéines Zif avec les séquences palindromiques d'ADN

Nous avons commencé par l'étude de la protéine notée LRZif dont la séquence correspond à celle de Zif12 décrite par Wang et Pabo. Nous avons réalisé un gel retard avec la protéine LRZif en interaction avec la séquence d'ADN-AA pour nous servir de point de comparaison vis-à-vis des modifications que nous avons apportées dans la séquence de la protéine Zif12\*-A. Pour l'étape d'incubation des protéines avec l'ADN, la concentration du duplex d'ADN a été fixée à 10 nM dans une solution à 10% de tampon C (la solution contient de l'HEPES (15 mM, pH 7.5) en présence de BSA (0.1 g.L<sup>-1</sup>), d'agent Tween (0.05%) et de NaCl (100 mM)). La composition de ce tampon est une version simplifiée de celui utilisé par Wang et Pabo. Une concentration croissante de protéine a été utilisée. Les différents mélanges protéine/ADN ont été placés à 4°C durant 12 heures avant d'être analysés sur le gel de polyacrylamide (gel à 8%). La migration du gel a été effectuée à 3 Watts durant 2h avec une température de gel n'excédant pas les 15°C au cours de la migration (le tampon de migration est identique à celui de Wang et Pabo). Sur la **Figure 93** sont présentés les gels d'électrophorèse obtenus après révélation. Nous avons observé dans chacun des deux gels une bande retardée que nous avons assignée à la formation du dimère autour de l'ADN, par analogie avec les résultats obtenus par Wang et Pabo. La formation d'une bande retardée avec la protéine Zif12\*-A nous indique que les changements apportés n'empêchent pas la formation du complexe (Zif12\*-A)<sub>2</sub>/ADN. Sur le gel obtenu avec la protéine Zif12\*-Nle-A nous avons encore cette bande retardée, il semble que celle-ci apparaisse à de plus faibles concentrations que la protéine LRZif et Zif12\*-A.



**Figure 93** : Gels retard obtenu après 12h d'incubation à 4°C, migration 2h à 10W à moins de 15°C des protéines LRZif (A), Zif12\*-A (B) et Zif12\*-Nle-A (C). La concentration en ADN-AA est fixée à 10 nM et la concentration en protéine est notée en nM sous le gel.

Nous avons ensuite déterminé pour la protéine Zif12\*-Nle-A la constante de dissociation et la valeur du  $C_{50}$  à l'aide d'une gamme de concentration en protéine de 1.87 à 40 nM pour une concentration fixe en ADN-AA de 10 nM. La quantification des bandes correspondant à la formation du dimère (lié) et de l'ADN seul (non lié) a été réalisée. Nous avons obtenu la valeur de la constante de dissociation grâce à la courbe représentant l'évolution du pourcentage d'ADN lié en fonction de la concentration (**Figure 94 A**). Elle a été estimée à  $2.9 \times 10^{-17} \text{ M}^2$  et la valeur  $C_{50}$  à 7.6 nM ( $\pm 0.6 \text{ nM}$ ), celle publiée pour la protéine LRZif étant de 15 nM<sup>53</sup>. Nous avons aussi calculé la constante de dissociation pour la protéine Zif12\*-B en interaction avec l'ADN-BB, la valeur obtenue est de  $3.8 \times 10^{-16} \text{ M}^2$  et la valeur  $C_{50}$  de 19.4 nM ( $\pm 0.2 \text{ nM}$ ) (**Figure 94 B**). Il faut cependant confirmer ce résultat en réalisant ces expériences 3 fois.

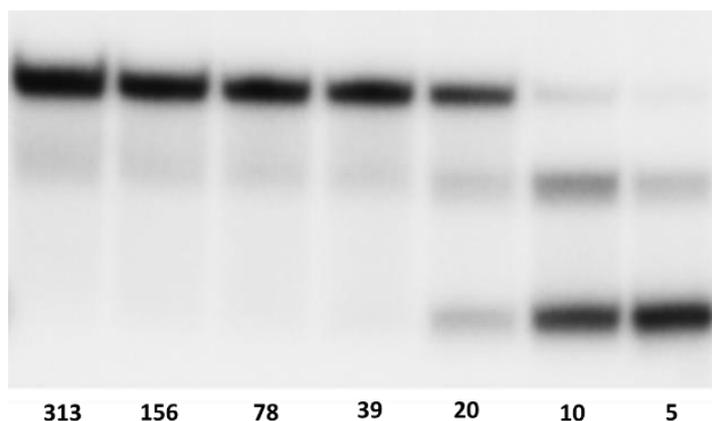


**Figure 94** : A) Courbe de  $K_d$  de la protéine Zif12\*-Nle-A avec l'ADN-AA. B) Courbe de  $K_d$  de la protéine Zif12\*-B avec l'ADN-BB

L'ensemble des modifications énumérées précédemment, comme le remplacement des Met<sup>91</sup> et Met<sup>102</sup> par des norleucines et le passage d'une amine acétylée (Zif12\*-A) à une amine primaire sur la partie N-terminale de la protéine n'ont pas perturbé la dimérisation. Elles semblent même avoir amélioré

l'affinité du système pour la cible d'ADN-AA. Cependant, nous sommes dans le même ordre de grandeur, avec les incertitudes liées à l'expérimentation, dire qu'il y a une véritable amélioration de l'affinité serait sur-interpréter ce résultat. La protéine Zif12\*-B que nous avons conçue a été étudiée en interaction avec le duplex d'ADN-BB dans les mêmes conditions. Une bande retardée correspondant à la formation du dimère autour de l'ADN a été observée. Pour cette protéine la constante de dissociation a été mesurée à 20 nM. Cette valeur se situe dans le même ordre de grandeur que les protéines LRZif et Zif12\*-Nle-A.

Pour le ciblage de la séquence d'ADN-AB nous avons choisi d'utiliser une solution équimolaire des protéines Zif12\*-Nle-A et Zif12\*-B. Les conditions d'incubation et de migration sont identiques aux précédents gels. Sur la **Figure 95**, nous pouvons observer la bande retardée la plus haute qui atteste de la formation d'un dimère autour de la séquence cible d'ADN. L'apparition de cette bande arrive très tôt dans la gamme ce qui suggère une forte affinité du couple de protéines pour la cible de 12 nucléotides. La valeur de la  $C_{50}$  de cet hétérodimère autour de la séquence d'ADN-AB a été calculée à 4.9 nM ( $\pm 0.5$  nM) ( $K_d = 2.4 \times 10^{-17}$  M<sup>2</sup>). À noter la présence d'une bande retardée intermédiaire à faible concentration qui peut être attribuée à la fixation d'une protéine sur sa séquence cible de 6 nucléotides.

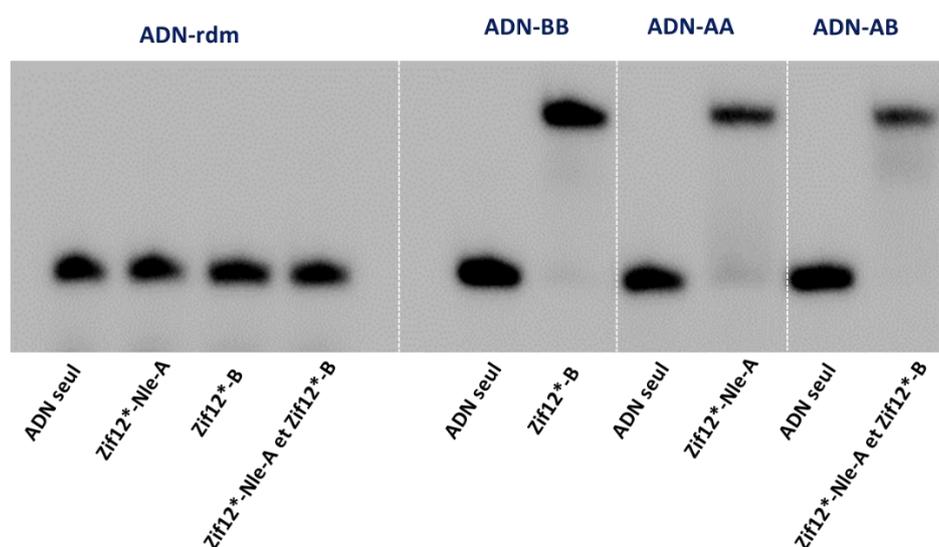


**Figure 95** : Gels retard obtenu après 12h d'incubation à 4°C, migration 2h à 10W à moins de 15°C des protéines Zif12\*-Nle-A et Zif12\*-B incubées en quantité équimolaire avec l'ADN-AB.

Nous avons voulu caractériser cette bande retardée par spectrométrie de masse. Pour cela nous avons réalisé plusieurs gels retard dans des conditions dans lesquelles la formation du dimère est quasi quantitative. À noter que pour cette expérience nous avons utilisé de l'ADN-AB non radioactif, à la fin de la migration, la coloration du gel est assurée par du GelRed, un intercalant fluorescent de l'ADN. Nous avons découpé la bande retardée pour en extraire le complexe Zif12\*-Nle-A/Zif12\*-B/ADN par agitation dans l'eau à 4°C pendant 48 heures. Une dialyse de la solution a été ensuite réalisée avec un cut-off de 3.5 kDa. Cependant cette méthodologie ne nous a pas permis d'avoir suffisamment de composés pour avoir une réponse claire lors de l'analyse par spectrométrie de masse MALDI. Des traces de protéines sont détectées mais l'intensité du signal est trop faible et peu résolue, la masse ne peut être mesurée avec précision ce qui ne permet pas de dire s'il y a deux protéines distinctes.

Nous avons en parallèle procédé à d'autres analyses sur gels afin de vérifier que l'on forme bien un hétérodimère. Nous avons mis en interaction d'une part la protéine Zif12\*-Nle-A seule avec la séquence d'ADN-AB, nous avons ensuite réalisé la même expérience avec la protéine Zif12\*-B. Dans les deux cas, aucune bande retardée n'a été observée pour la formation d'un dimère, nous confortant dans l'idée de la formation d'un hétérodimère. Nous avons cependant observé, dans le cas de la protéine Zif12\*-B, une bande retardée à mi-hauteur que nous avons attribuée à la fixation de la protéine sur la demi-cible d'ADN-AB formant un complexe 1:1. Ceci est en adéquation avec la bande retardée intermédiaire observée **Figure 95**. La protéine Zif12\*-Nle-A n'est pas capable de s'y fixer sur sa séquence cible en monomère contrairement à la protéine Zif12\*-B, bien que les  $K_d$  des deux protéines pour leur séquence palindromique soient très proches. La protéine Zif12\*-B cible une séquence qui est riche en G (5'-GAAGGG-3') dont la reconnaissance est particulièrement efficace lorsque que plusieurs arginines sont présentes dans la séquence de reconnaissance (3 dans le cas de Zif12\*-B), de plus la reconnaissance de triplets qui commencent par G (en comparaison avec les autres bases nucléiques) sont des séquences qui sont, une fois encore, reconnues plus efficacement.

Le système de reconnaissance que nous avons décidé d'utiliser, basé sur l'utilisation de deux protéines à doigts de zinc capables de former un dimère autour de la séquence cible d'ADN double brin, nous a permis de cibler sélectivement une séquence d'ADN longue de 12 paires de bases. Seule l'association des deux protéines (Zif12\*-Nle-A et Zif12\*-B) permet de former le dimère autour de la séquence d'ADN-AB. Nous avons aussi regardé l'interaction des protéines sur la séquence d'ADN-rdm (**Figure 96**), aucune bande retardée n'est observée. Tous ces résultats illustrent la sélectivité de notre système.



**Figure 96** : Gels retard obtenus après incubation des protéines Zif12\*-Nle-A et Zif12\*-B avec différentes séquences d'ADN.

Les valeurs de  $C_{50}$  des protéines Zif12\*-Nle-A et Zif12\*-B sur leur cible palindromique d'ADN sont respectivement de 7.6 nM et 19.4 nM. Lorsque l'on cible une séquence non palindromique d'ADN double brin, comportant les cibles de ces deux protéines (A et B), la  $C_{50}$  est de 4.9 nM. Nous pouvons

dès lors nous tourner vers la fonctionnalisation de ces protéines par notre système de FRET entre un complexe de lanthanide et un chromophore organique. La protéine Zif12\*-Nle-A sera fonctionnalisée par le complexe de lanthanide et la protéine Zif12\*-B par le chromophore organique.

# **Chapitre III - Fonctionnalisation des protéines Zif12**

## Sommaire

I - Choix des chromophores .....	95
II - Fonctionnalisation du fragment F1 par les chromophores .....	98
1 - Couplage des chromophores sur le fragment F1 .....	98
2 - Ligation des fragments fonctionnalisés par les chromophores avec le fragment F2-F3 .....	100
III - Fonctionnalisation de la protéine par le complexe d'euporium [L <sub>i</sub> -Eu] .....	102
1 - Synthèse du pro-ligand PL <sub>i</sub> <sup>CC</sup> .....	102
1.1 - Description du ligand L <sub>1</sub> <sup>CC</sup> .....	102
1.2 - Synthèse du macrocycle iodé 6 .....	104
1.3 - Synthèse de la partie électrodonneuse de l'antenne de PL <sub>1</sub> <sup>CC</sup> .....	105
1.4 - Synthèse du pro-ligand PL <sub>1</sub> <sup>CC</sup> .....	106
2 - Synthèse du fragment Ac-F1-[L <sub>i</sub> <sup>CC</sup> -Eu]-SEA .....	107
3 - Synthèse de la protéine Zif12-[L <sub>i</sub> <sup>CC</sup> -Eu]-A .....	110
IV - Synthèse de complexes avec des nouvelles antennes .....	111
1 - Nouvelle voie de synthèse .....	111
2 - Antennes dérivées de l'anisole .....	113
2.1 - Synthèse du pro-ligand PL <sub>2</sub> .....	115
2.2 - Caractérisation spectroscopique du complexe TAT-[L <sub>2</sub> -Eu] .....	114
3 - Antennes dérivées du carbazole .....	117
3.1 - Synthèse des pro-ligands PL <sub>3</sub> et PL <sub>4</sub> .....	118
3.2 - Caractéristiques spectroscopiques des complexes TAT-[L <sub>3</sub> -Eu] et TAT-[L <sub>4</sub> -Eu] .....	119
4 - Antennes dérivées de l'aniline .....	123
4.1 - Synthèse des pro-ligands PL <sub>5</sub> à PL <sub>10</sub> .....	126
4.2 - Synthèse et caractérisation spectroscopique des complexes [L <sub>i</sub> -Eu] (i=5-10) .....	128
V - Synthèse des protéines fonctionnalisées par les complexes de lanthanide .....	132

## I - Choix des chromophores

Pour rappel une première preuve de concept d'un transfert d'énergie induit par la dimérisation de deux protéines à doigts de zinc a été obtenue par Guillaume Frémy durant ses travaux de thèse. Ce transfert a été réalisé entre un complexe de terbium et un chromophore organique (Lissamine ou SiR<sup>Me</sup>). Le complexe utilisé est dérivé du complexe de Maury no3a<sup>4</sup>paTb présenté dans le chapitre d'introduction (**Chapitre I, IV-3**), noté [ $L_1$ -Tb] avec un macrocycle de type DO3A équipé d'un quatrième bras picolinate avec un système  $\pi$  étendu « push-pull », auquel une fonction de conjugaison (acide carboxylique) a été introduite au niveau du groupement électro-donneur (**Figure 97**). Ce complexe présente une intense émission du terbium ( $\Phi_{Tb} = 62\%$ ), et son spectre d'absorbance présente une bande localisée à 306 nm attribuée à la bande de transfert de charge ILCT et un épaulement à 280 nm correspondant à une transition vers un état localement excité de l'antenne, noté LE. Le coefficient d'absorption molaire de ce complexe, au maximum d'absorption à 306 nm est de 12 050 M<sup>-1</sup>cm<sup>-1</sup>.

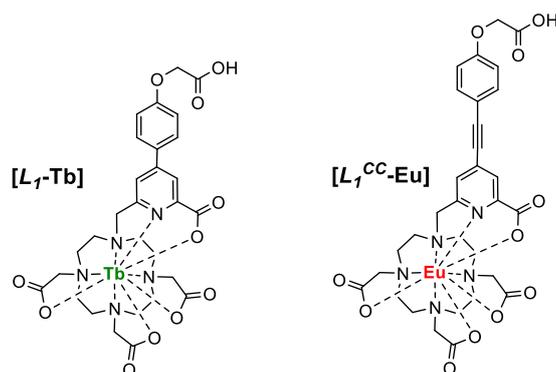
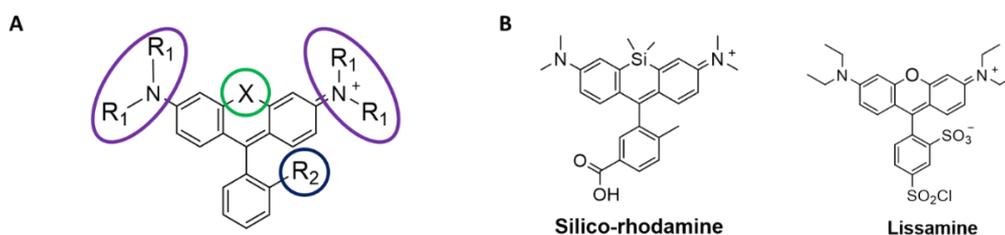


Figure 97 : Le complexe [ $L_1$ -Tb] développé par G. Frémy.

Avec ce complexe de terbium, Guillaume a étudié le transfert d'énergie vers deux fluorophores de la famille des rhodamines, la lissamine et une silico-rhodamine. Les rhodamines ont une émission vers 500 nm et le remplacement de l'atome d'oxygène du cœur xanthène par un autre atome permet de décaler l'émission vers le rouge. Malgré une émission intéressante au-delà de 650 nm les cyanines n'avaient pas été retenues car elles souffrent de photobleaching et sont peu solubles dans l'eau, ce qui pour notre système représente un problème. Les rhodamines font partie des fluorophores les plus photostable, ce qui a conduit à la synthèse de nombreux dérivés qui sont maintenant utilisés en médecine et pour l'imagerie in vivo (animaux ou essais cliniques)<sup>135-137</sup>. Il y a trois positions sur lesquelles des modifications des rhodamines sont possibles : (i) la nature des substituants des amines ( $R_1$  sur la **Figure 98**), (ii) la position en  $\alpha$  de l'aryle en position méso du cœur ( $R_2$ ) et (iii) la nature de l'atome de cœur xanthène, où l'oxygène peut être substitué par Si, Ge, Sn, P ou C. La première rhodamine retenue pour le projet était la lissamine, dont le spectre d'absorbance présente une large bande avec un maximum d'émission à 560 nm et un coefficient d'absorption molaire de 71 500 M<sup>-1</sup>cm<sup>-1</sup> à ce maximum.

Son maximum d'émission est à 580 nm ( $\Phi_{\text{Liss}} = 30\%$ ). Elle présente l'avantage d'être commercialisée (200€/g) sous la forme d'un chlorure de sulfonyle pour une conjugaison facile à une chaîne latérale de lysine. La seconde rhodamine utilisée est une Silico-rhodamine ( $\text{SiR}^{\text{Me}}$ ) synthétisée par notre collaborateur Nathan Mc. Clenaghan (Univ. Bordeaux) (**Figure 98**). Le changement de l'atome d'oxygène par un  $\text{Si}(\text{Me})_2$  permet un shift d'émission d'environ 100 nm vers le proche infrarouge par rapport aux rhodamines, l'alkylation des deux amines par des groupements méthyles vient aussi décaler la longueur dans le proche infrarouge ce qui permet d'avoir *in fine* une émission qui se situe au-delà de 650 nm<sup>138,139</sup>. Enfin le groupement méthyle en position ( $\text{R}_2$ ) permet d'avoir une brillance plus élevée, par rapport à l'utilisation d'un atome d'hydrogène, en empêchant la libre rotation entre les cycles aromatiques. Les Si-rhodamines sont particulièrement photostables<sup>5</sup>. L'émission de  $\text{SiR}^{\text{Me}}$  est décalée vers le rouge, avec un maximum d'excitation à 646 nm avec un coefficient d'absorption molaire de 110 000  $\text{M}^{-1}\text{cm}^{-1}$  et une bande d'émission large avec un maximum à 660 nm ( $\Phi_{\text{SiR}} = 30\%$ ) (**Figure 99**). Les propriétés spectroscopiques de ces deux chromophores sont rassemblées dans le **Tableau 3**.

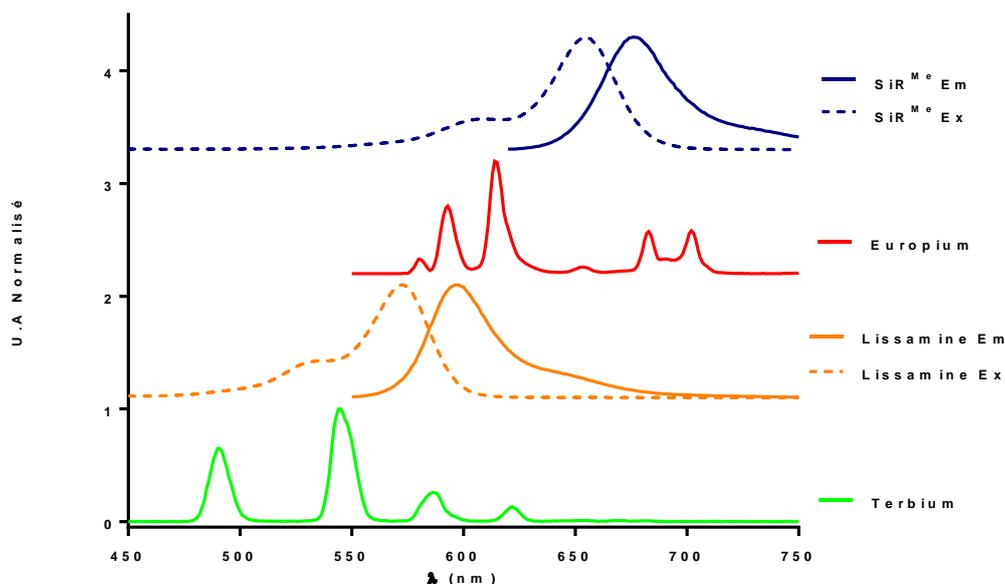


**Figure 98** : A) Schéma des différentes modifications sur une rhodamine pour modifier ses propriétés spectroscopiques. B) Structure de la Silico-Rhodamine ( $\text{SiR}^{\text{Me}}$ ) synthétisée par N. Mc. Clenaghan et de la lissamine commerciale.

**Tableau 3** : Propriétés spectroscopiques de la lissamine et de la  $\text{SiR}^{\text{Me}}$ .

	$\lambda_{\text{ex}}$ (nm)	$\lambda_{\text{em}}$ (nm)	$\epsilon_{\text{max}}$ ( $\text{M}^{-1}\text{cm}^{-1}$ )	$\Phi$ (%)
<b>Lissamine</b>	564	583	71 500	33 ( $\text{H}_2\text{O}$ )
<b><math>\text{SiR}^{\text{Me}}</math></b>	646	660	110 000	31 (PBS)

Les spectres d'absorption et d'émission des chromophores, ainsi que les spectres d'émission des deux lanthanides sont présentés **Figure 99**. Le spectre d'absorption de la lissamine s'étend de 520 à 600 nm ce qui en fait un parfait accepteur FRET en association avec un complexe de terbium. La silico-rhodamine présente un spectre d'excitation qui s'étend de 600 à 680 nm qui recouvre en partie avec les raies d'émission de l'euprium. Un transfert d'énergie de l'euprium vers la lissamine est aussi possible, mais semble moins efficace compte tenu du faible recouvrement des spectres d'émission/excitation des deux partenaires.



**Figure 99** : Spectres d'émissions et d'excitations de la lissamine et SiR<sup>Me</sup> superposés avec le spectre d'émission de l'euporium et du terbium. Spectres obtenus en solution aqueuse avec 10% de tampon HEPES (10 mM, pH 7.0).

Le rayon de Förster des couples Tb/Lissa et Tb/SiR<sup>Me</sup> ont été déterminés par l'équation (4) et sont respectivement de 53 et 48 Å. La fonctionnalisation des protéines Zif12 sur la position 110 par une lysine permet d'avoir une distance théorique de 35 Å, distance qui sépare les deux carbones  $\alpha$  des aa en position 110 des deux protéines. Cependant la fonctionnalisation des chromophores se situe sur les chaînes latérales des lysines, qui sur la structure RX tendent à s'orienter vers l'extérieur du dimère. La distance qui sépare donc les partenaires de FRET est sans doute plus longue et probablement proche ou au-delà de 45 Å en moyenne. Guillaume a néanmoins observé un transfert d'énergie entre le complexe de terbium et la lissamine. Mais l'ensemble des spectres enregistrés présente une forte émission de la lissamine, même en absence d'ADN. L'inconvénient est que la lissamine absorbe à la longueur d'onde d'excitation du complexe de Tb<sup>3+</sup>, à 306 nm les coefficients d'absorptions sont quasi équivalents,  $\epsilon_{\text{Lissa}} = 14\ 000\ \text{M}^{-1}\text{cm}^{-1}$  et  $\epsilon_{[\text{Tb}^{3+}] } = 12\ 050\ \text{M}^{-1}\text{cm}^{-1}$ . L'utilisation de la SiR<sup>Me</sup> comme accepteur FRET ne peut pas résoudre ce problème, car elle aussi absorbe à 306 nm, avec un epsilon de  $14\ 000\ \text{M}^{-1}\text{cm}^{-1}$  et les rendements quantiques  $\Phi_{\text{Lissa}}$  et  $\Phi_{\text{SiR}}$  sont de 30%. Cette excitation directe des fluorophores contribue beaucoup à l'émission globale d'où la nécessité de faire une détection en temps résolu qui est possible grâce au complexe de lanthanide. Guillaume s'est alors tourné vers l'utilisation d'un complexe d'euporium [ $L_I^{\text{CC}}$ -Eu] pour décaler la longueur d'excitation vers 350 nm (grâce à la fonction alcyne insérée dans l'antenne en comparaison de  $L_I^{\text{CC}}$ ), où les coefficients d'absorptions molaires de la lissamine et SiR<sup>Me</sup> sont respectivement de  $8\ 000\ \text{M}^{-1}\text{cm}^{-1}$  et  $3\ 600\ \text{M}^{-1}\text{cm}^{-1}$ . À cette longueur d'onde, c'est la SiR<sup>Me</sup> qui présentera le moins d'émission (par rapport à la lissamine). Ainsi il semble pertinent de vouloir déplacer l'excitation et d'utiliser, pour avoir une faible absorption directe par le fluorophore, un complexe d'euporium dont l'émission présente un fort recouvrement avec la SiR<sup>Me</sup> favorisant le transfert d'énergie  $\text{Eu}^{3+}/\text{SiR}^{\text{Me}}$ .

## II - Fonctionnalisation du fragment F1 par les chromophores

Notre système FRET tel que nous l'avons conçu, repose sur le transfert d'énergie d'un complexe de lanthanide vers un chromophore organique. Pour rappel, nous avons décidé de décaler la position de greffage de la Lys<sup>110</sup> à la Lys<sup>106</sup>. La distance mesurée entre les deux carbones  $\alpha$  des Lys<sup>106</sup> sur la structure cristallographique du dimère, autour de l'ADN, obtenue par Wang et Pabo est de 18 Å. Cette distance est suffisamment courte pour espérer un transfert FRET très efficace entre le complexe de lanthanide et le fluorophore accepteur. Nous avons choisi de travailler principalement avec l'euprium comme donneur dans ce système, mais l'utilisation du terbium reste une alternative compte tenu des travaux précédents sur ce sujet.

### 1 - Couplage des chromophores sur le fragment F1

Comme indiqué lors de la présentation du projet, nous souhaitons fonctionnaliser la partie N-terminale de la protéine de Zif12. C'est au niveau de la lysine 106 que seront couplés les chromophores *via* la fonction amine de la chaîne latérale. Compte tenu des résultats observés de l'interaction de la protéine Zif12\*-A et Zif12\*-Nle-A avec l'ADN-AA, nous avons choisi de ne travailler qu'avec le fragment F1 dont les méthionines 91 et 102 ont été remplacées par les norleucines. Le groupement protecteur orthogonal que nous avons employé pour la Lys<sup>106</sup> lors de la synthèse du fragment est le groupement Alloc. Après l'élongation du peptide sur résine, celui-ci peut être sélectivement déprotégé à l'aide de palladium(0) en présence de phénylsilane.

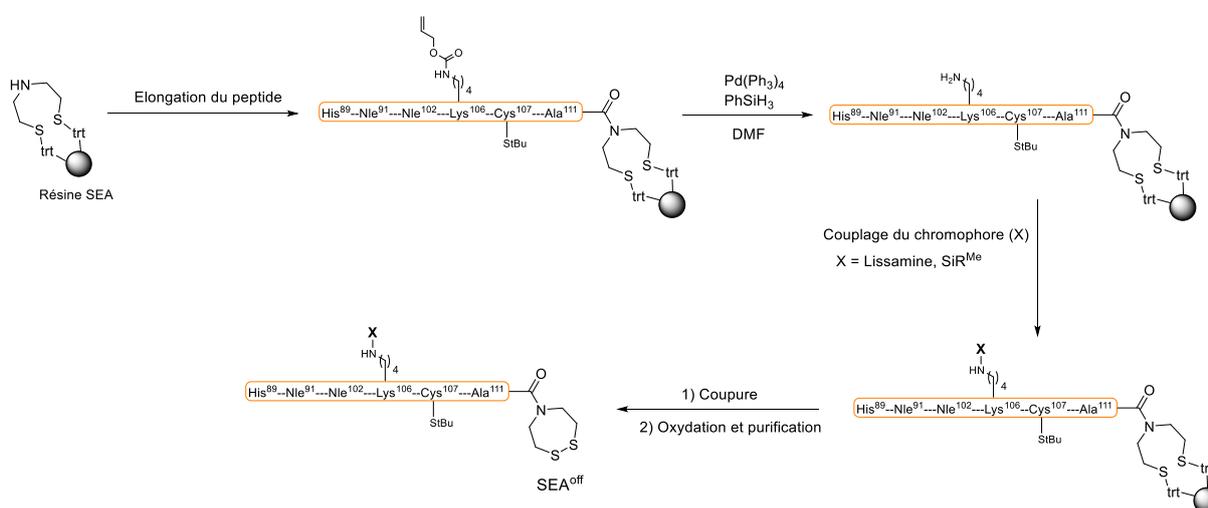
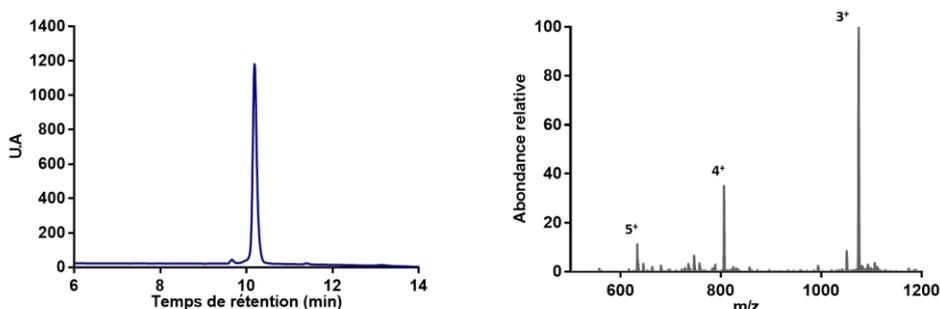


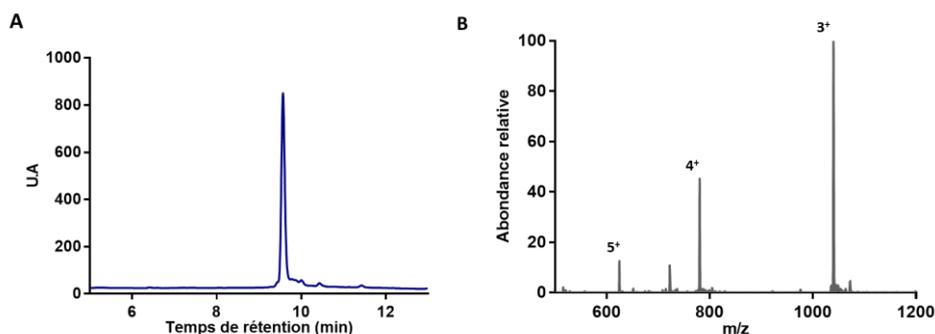
Figure 100 : Schéma de synthèse du fragment F1 fonctionnalisé par les chromophores organiques.

Le couplage de la SiR<sup>Me</sup> est effectué sur la résine en condition basique (DIEA, 4 éq) en utilisant comme activateur du PyBOP (2 éq) avec un léger excès de chromophore (1.1 éq). Après 12h de réaction à température ambiante, le suivi par LCMS révèle un couplage quantitatif du chromophore. La coupure du peptide a été effectuée par une solution de TFA comportant des scavengers : TIS, thioanisole et eau

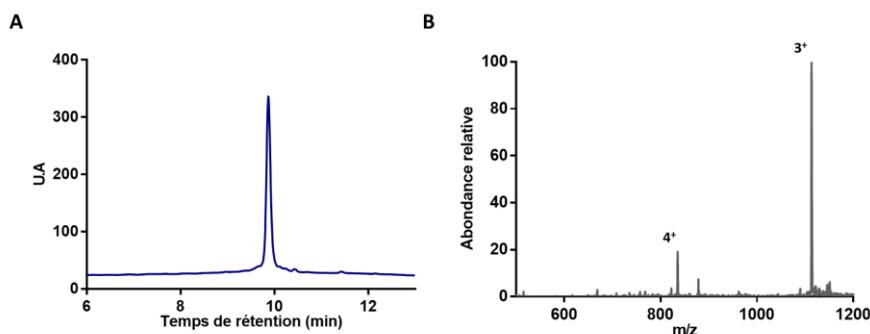
(91.7/3.3/3.3/1.7). Après 2h de réaction, le TFA est évaporé et le peptide est précipité dans de l'éther glacé. La fonction SEA<sup>on</sup> est oxydée par dissolution dans une solution aqueuse d'acide acétique à 10% et l'ajout d'une solution de diiode (200 mM, DMSO). L'excès de I<sub>2</sub> est quenché par l'ajout de DTT avant d'être purifiée par HPLC. La fonctionnalisation du fragment F1 par la SiR<sup>Me</sup> a été réalisée sur deux résines pour obtenir les protéines *H*-F1-(SiR<sup>Me</sup>)-SEA<sup>off</sup> et *H*-F1-(SiR<sup>Me</sup>)-NHNH<sub>2</sub> (**Figure 101** et **Figure 102**). Le couplage de la lissamine quant à lui ne requiert pas d'activateur puisqu'on dispose d'un chlorure de sulfonyle et d'une amine. L'ajout de 4 éq de DIEA avec 1.1 éq de lissamine est suffisant pour avoir, après 12h de réaction un couplage quantitatif. La protéine *H*-F1-(Lissa)-SEA<sup>off</sup> a pu être synthétisée et purifiée (**Figure 103**).



**Figure 101** : A) Chromatogramme HPLC, détection UV à 214 nm du fragment *H*-F1-(SiR<sup>Me</sup>)-SEA<sup>off</sup> purifié. B) Le spectre ESI-MS du fragment *H*-F1-(SiR<sup>Me</sup>)-SEA<sup>off</sup> purifié (masse théorique calculée : 3221.05, masse déconvoluée obtenue : 3221.36).



**Figure 102** : A) Chromatogramme HPLC, détection UV à 214 nm du fragment *H*-F1-(SiR<sup>Me</sup>)-NHNH<sub>2</sub> purifié. B) Le spectre ESI-MS du fragment *H*-F1-(SiR<sup>Me</sup>)-NHNH<sub>2</sub> purifié (masse théorique calculée : 3117.86, masse déconvoluée obtenue : 3116.78).

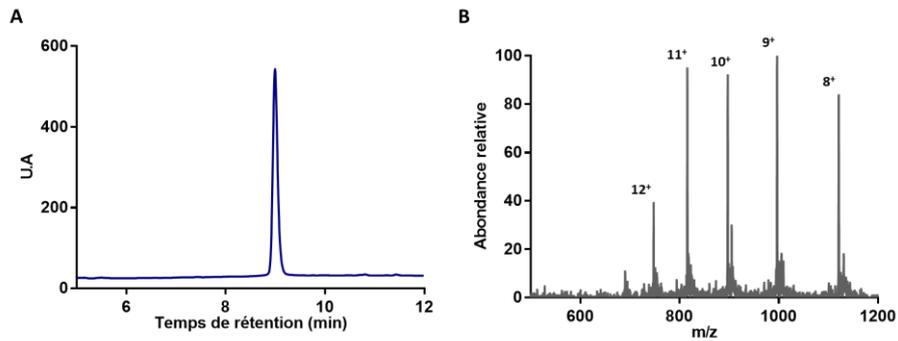


**Figure 103** : A) Chromatogramme HPLC, détection UV à 214 nm du fragment *H*-F1-(Lissa)-SEA<sup>off</sup> purifié. B) Le spectre ESI-MS du fragment *H*-F1-(Lissa)-SEA<sup>off</sup> purifié (masse théorique calculée : 3336.08, masse déconvoluée obtenue : 3337.41).

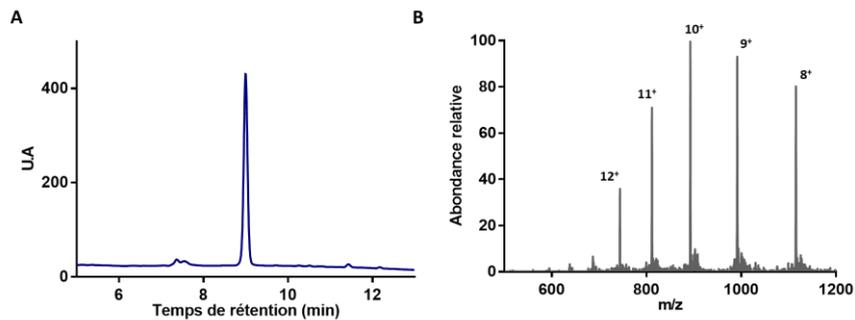
## 2 - Ligation des fragments fonctionnalisés par les chromophores avec le fragment F2-F3

Pour cette dernière étape de ligation, nous avons commencé par utiliser la méthode des hydrazides entre les fragments  $H-F1-(SiR^{Me})-NHNH_2$  et  $H-F2-F3-NH_2$ , puisque la ligation hydrazide est plus rapide que la ligation SEA.

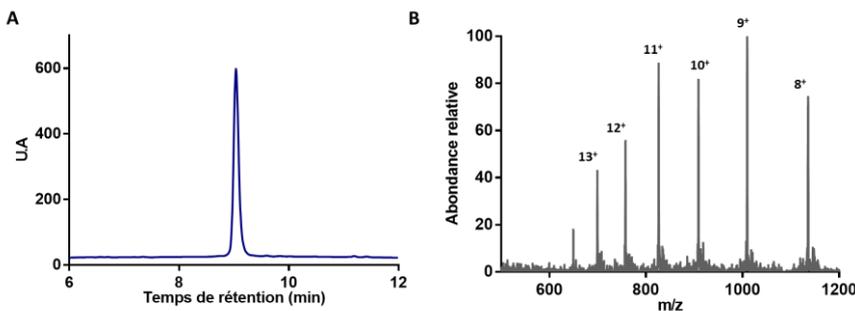
Le fragment  $H-F1-(SiR^{Me})-NHNH_2$  a été dissout dans une solution de chlorure de guanidinium (6 M) dans du tampon phosphate (100 mM, pH 7.0) dont le pH a été ajusté à 3. La concentration du peptide a été fixée à 3 mM. L'étape d'oxydation par 1.5 éq de  $NaNO_2$  à  $-15^\circ C$  durant 10 min a révélé, après analyse par LCMS, aucune présence du produit désiré, mais la présence d'un produit majoritaire non identifié avec un  $\Delta M = +28$ . La ligation hydrazide a été utilisée par de nombreux groupes de recherche et aucun n'a décrit de réactions parasites avec les acides aminés naturels. Il est donc logique de penser que c'est au niveau du chromophore que ces sous-produits proviennent et que la  $SiR^{Me}$  a été endommagée par une réaction avec le  $NaNO_2$ . Le groupe de Dawson a décrit une méthode d'activation douce de l'hydrazide, basée sur l'utilisation d'acétylacétone pour former un intermédiaire acétyl pyrazole<sup>112</sup>. La formation de cet intermédiaire s'effectue avec 2.5 éq d'acétylacétone dans une solution aqueuse contenant du chlorure de guanidinium (6 M) dans du tampon phosphate (100 mM, pH 7.5). Cependant nous n'avons pu obtenir le pyrazole correspondant ce qui nous a forcés à abandonner la ligation par la méthode des hydrazides. Nous n'avons pas persévéré et essayé d'optimiser notre protocole et nous sommes donc revenus sur la méthode de ligation SEA pour l'assemblage final des fragments F1 fonctionnalisés par les chromophores, même si cette ligation est plus lente. Les ligations des protéines  $H-F1-(SiR^{Me})-SEA^{off}$  et  $H-F1-(Lissa)-SEA^{off}$  avec le fragment  $H-F2-F3-NH_2$  ont été réalisées dans les mêmes conditions. Une solution contenant du chlorure de guanidinium (6 M), du MPAA (0.1 M) et du TCEP (0.1 M) dans du tampon phosphate (100 mM, pH 7.0) dont le pH a été ajusté à 6.5 est préparée. La dissolution des deux fragments est réalisée en boîte à gants et le volume de solution à ajouter est calculé pour ajuster la concentration du fragment F1 à 3 mM (le fragment F1 étant le réactif limitant). Le milieu réactionnel est ensuite placé sous agitation à  $37^\circ C$ . Après 24h de réaction, le milieu réactionnel est purifié par HPLC pour obtenir les protéines isolées. Au total ce sont quatre protéines qui ont été synthétisées : Zif12- $(SiR^{Me})$ -A (**Figure 104**), Zif12- $(SiR^{Me})$ -B (**Figure 105**), Zif12- $(Lissa)$ -A (**Figure 106**) et Zif12- $(Lissa)$ -B (**Figure 107**).



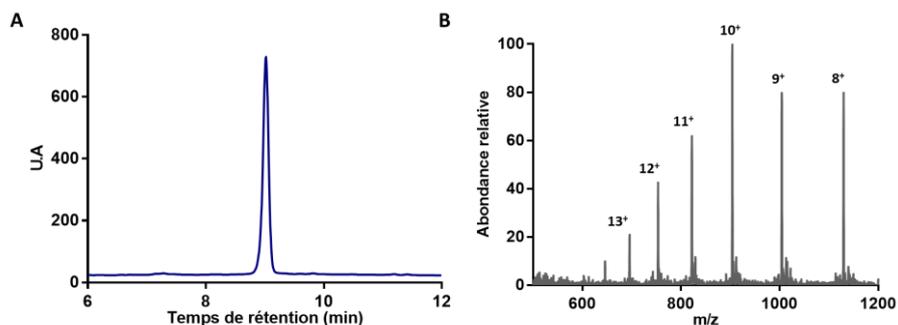
**Figure 104 :** A) Chromatogramme HPLC, détection UV à 214 nm de la protéine Zif12-(SiR<sup>Me</sup>)-A purifiée. B) Le spectre ESI-MS de la protéine Zif12-(SiR<sup>Me</sup>)-A purifiée (masse théorique calculée : 8962.49, masse déconvoluée obtenue : 8960.63).



**Figure 105 :** A) Chromatogramme HPLC, détection UV à 214 nm de la protéine Zif12-(SiR<sup>Me</sup>)-B purifiée. B) Le spectre ESI-MS de la protéine Zif12-(SiR<sup>Me</sup>)-B purifiée (masse théorique calculée : 8916.47, masse déconvoluée obtenue : 8914.73).



**Figure 106 :** A) Chromatogramme HPLC, détection UV à 214 nm de la protéine Zif12-(Lissa)-A purifiée. B) Le spectre ESI-MS de la protéine Zif12-(Lissa)-A purifiée (masse théorique calculée : 9077.52, masse déconvoluée obtenue : 9076.69).



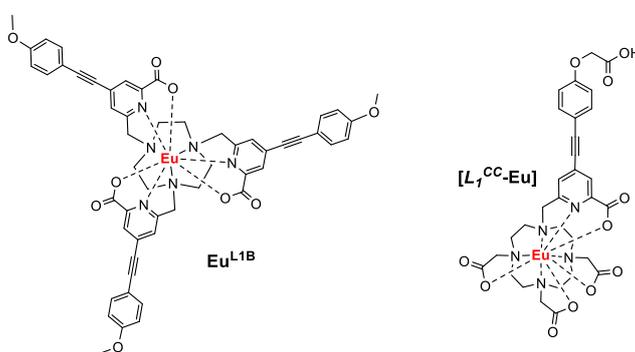
**Figure 107 :** A) Chromatogramme HPLC, détection UV à 214 nm de la protéine Zif12-(Lissa)-B purifiée. B) Le spectre ESI-MS de la protéine Zif12-(Lissa)-B purifiée (masse théorique calculée : 9030.49, masse déconvoluée obtenue : 9030.54).

### III - Fonctionnalisation de la protéine par le complexe d'euporium [ $L_1$ -Eu]

#### 1 - Synthèse du pro-ligand $PL_1^{CC}$

##### 1.1 - Description du ligand $L_1^{CC}$

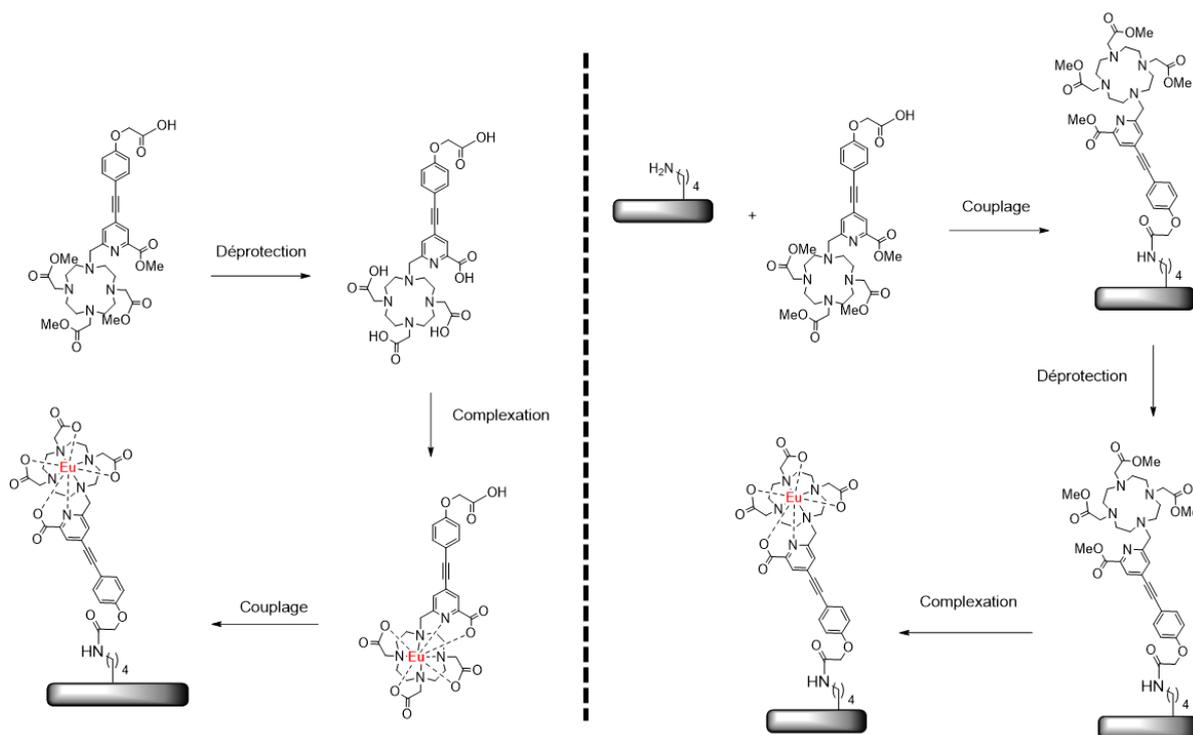
À partir des différents complexes de lanthanide présentés en introduction (**Chapitre I, IV-3**) c'est le complexe  $Eu^{L1B}$  développé par O. Maury, qui présente pour notre projet des propriétés très intéressantes<sup>140</sup>. Il est composé d'un macrocycle de type TACN dont chacune des amines est fonctionnalisée par une antenne « push-pull ». Cette antenne se compose (i) d'une unité picolinate qui permet de compléter la sphère de coordination du lanthanide *via* l'atome d'azote et l'acide carboxylique, et joue aussi le rôle de groupement électro-accepteur, (ii) d'un groupement *p*-méthoxyphényle qui sert de groupement électrodonneur et (iii) d'un groupement alcyne entre les deux cycles aromatiques pour augmenter la délocalisation électronique et déplacer l'absorption vers le rouge en comparaison du système 4-(*p*-méthoxyaryle)-picolinate utilisé par Guillaume Frémy ( $L_1^{CC}$ ). Ce complexe d'euporium présente un rendement quantique élevé de 25% dans un mélange  $H_2O/MeOH$  (3:1) avec une longueur d'onde d'absorption maximale à 339 nm et un coefficient d'absorption molaire mesuré à  $\epsilon_{339}L_1^{CC} = 58\,000\,M^{-1}cm^{-1}$ . Plusieurs complexes d'euporium ont été développés avec divers groupements électrodonneurs O-alkyle à la place du méthoxy. Cela n'affecte que peu le rendement quantique du complexe ce qui est très intéressant pour l'insertion d'une fonction de bioconjugaison à cette position de l'antenne. La synthèse d'un complexe d'euporium  $L_1^{CC}$  inspiré de celui de O. Maury a été réalisée durant les travaux de thèse de Guillaume Frémy. La fonction de bioconjugaison utilisée est un acide carboxylique espacée par un  $CH_2$  sur la partie électrodonneuse de l'antenne (**Figure 108**). J'avais d'ailleurs participé à la mise au point de cette synthèse lors de mon stage de Master 1 dans l'équipe.



**Figure 108** : Le complexe  $Eu^{L1B}$  développé par O. Maury et le complexe [ $L_1^{CC}$ -Eu] synthétisé par G. Frémy.

La conjugaison de ce complexe sur la partie N-terminale de la protéine, *via* la chaîne latérale de la lysine, sur support solide peut se faire de deux manières. La première est le couplage du complexe d'euporium [ $L_1^{CC}$ -Eu] dans lequel la seule fonction acide carboxylique disponible est celle située sur le groupement

électrodonneur de l'antenne<sup>141</sup>. L'atome d'euprium sert alors de groupement protecteur des acides carboxyliques du macrocycle/picolinate. Guillaume Frémy avait exploré cette voie de synthèse, mais le complexe est trop peu soluble dans les solvants organiques utilisés pour la SPPS et le couplage n'est pas efficace. La seconde méthode, plus classique, vise à protéger la fonction de bioconjugaison par un groupement orthogonal à ceux des acides carboxyliques du macrocycle/picolinate pour effectuer le greffage d'un pro-ligand ( $PL_I^{CC}$ ). Une déprotection des acides carboxyliques du macrocycle suivie d'une complexation permet d'obtenir le fragment F1 fonctionnalisé par le complexe (**Figure 109**).



**Figure 109** : Schéma des deux voies de couplage possible pour la fonctionnalisation du complexe [L1<sup>CC</sup>-Eu].

Pour la synthèse de ce pro-ligand, il a été décidé pendant la thèse de Guillaume Frémy de protéger les fonctions acides carboxyliques du macrocycle/picolinate par des esters méthyliques. La fonction de bioconjugaison quant à elle sera protégée par un ester *tert*-butylique. Ce choix avait été fait, car il avait été observé dans l'équipe que l'acidolyse des bras acétates protégés sous forme d'ester de *t*Bu était lente lorsque le macrocycle était greffé sur le peptide<sup>142</sup>. Le pro-ligand  $PL_I^{CC}$  est synthétisé en deux blocs. Le premier est le macrocycle DO3A fonctionnalisé par trois bras acétates et un bras picolinate substitué par un atome d'iode au niveau de sa position 4 (composé **6**, **Figure 110**). Le second bloc est la partie qui comprend l'espaceur et la partie électrodonneuse de l'antenne qui peut être couplée au picolinate en utilisant une réaction pallado-catalysée sur la position portant l'atome d'iode. L'acidolyse du groupement *t*Bu de la fonction de bioconjugaison sur la partie électrodonneuse de l'antenne aboutit à la formation du pro-ligand  $PL_I^{CC}$  (**Figure 110**).

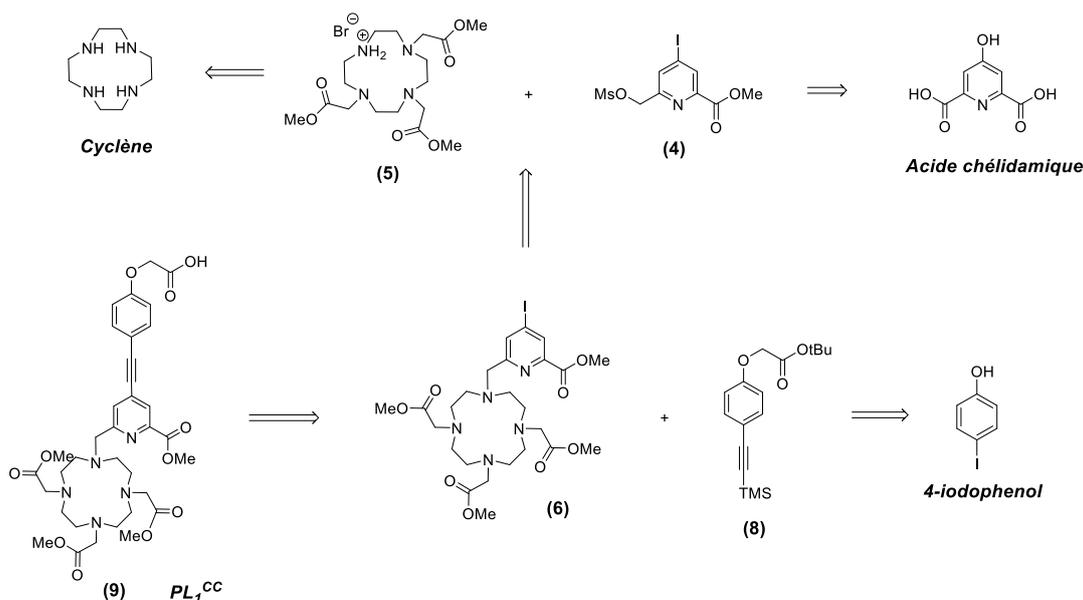


Figure 110 : Schéma global de la voie de synthèse du pro-ligand  $PL_1^{CC}$ .

### 1.2 - Synthèse du macrocycle iodé 6

Le composé **4** a été obtenu en quatre étapes avec un rendement global de 51% (**Figure 111**). Une réaction de Vilsmyer-Haack sur l'acide chélidamique à l'aide de chlorure de thionyle permet la chloration de la fonction alcool ainsi que la formation de deux esters méthyliques après addition de méthanol (composé **1**). L'atome de chlore en position 4 de la pyridine est ensuite substitué par un iode dans une réaction de substitution nucléophile aromatique en présence d'iodure de sodium, de chlorure d'acétyle dans l'acétonitrile et sous ultrasons pour former le composé **2**. L'étape suivante est la réduction sélective d'un des deux esters méthyliques en alcool par du borohydrure de sodium réalisé dans un mélange DCM/MeOH 2:3 à 0 °C (70%). Enfin le composé **4** est obtenu par activation au chlorure de mésyle de la fonction alcool (98%). Le macrocycle DO3A-*tris*(OMe)ester (**5**, **Figure 111**) a été obtenu avec un rendement de 85% en suivant le protocole décrit par Jagadish<sup>143</sup> par réaction du cyclène avec trois équivalents de bromoacétate de méthyle dans le DMF en présence d'acétate de sodium. Le macrocycle iodé **6** a été obtenu après une réaction de substitution nucléophile entre les composés **4** et **5** dans l'acétonitrile en présence de  $Na_2CO_3$ , le composé **6** a ensuite été purifié par cristallisation (99%). Le rendement global de synthèse du composé **6**, en 6 étapes, est de 43%.

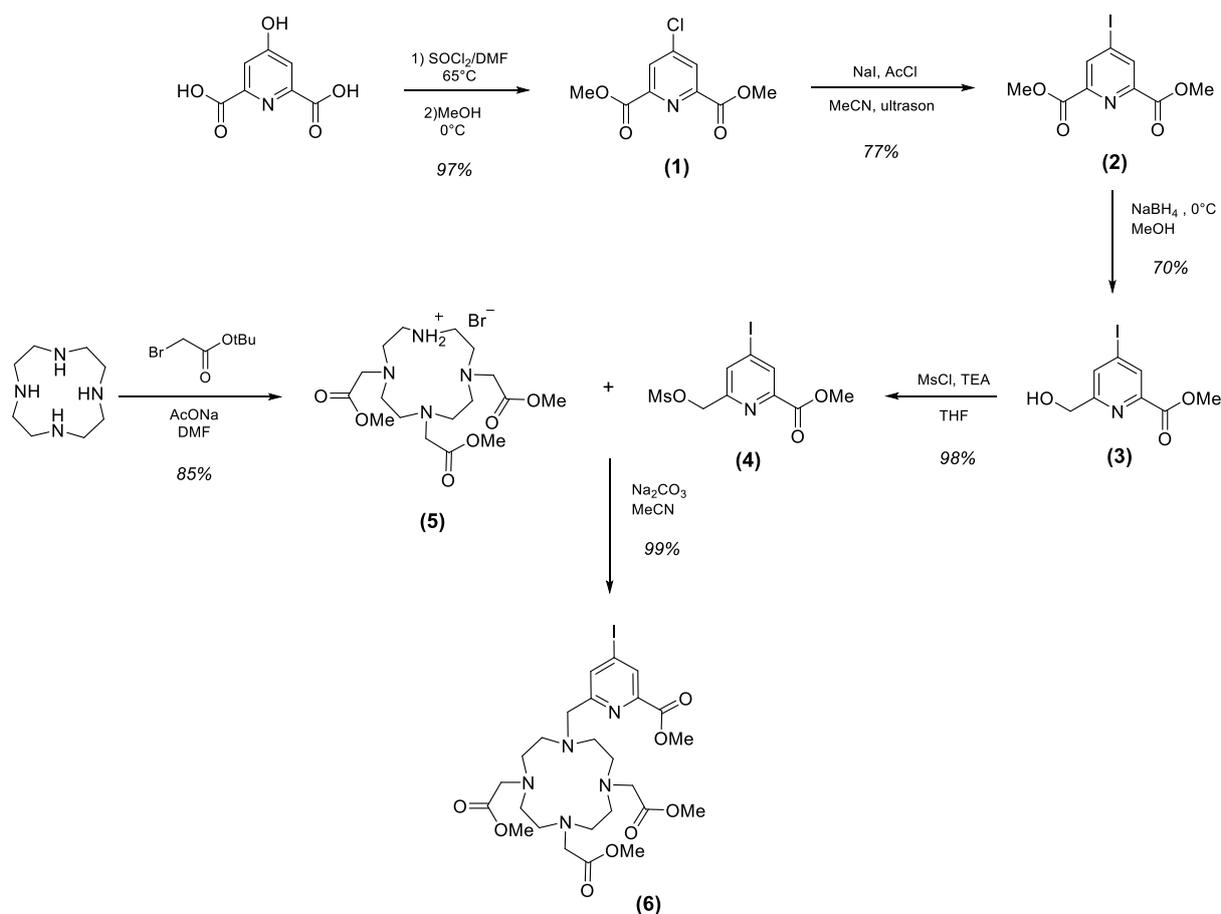


Figure 111 : Schéma de synthèse du composé iodé **6** à partir de l'acide chélidamique et du cyclène.

### 1.3 - Synthèse de la partie électrodonneuse de l'antenne de PL<sub>I</sub><sup>CC</sup>

La synthèse du composé **8** est réalisée en deux étapes à partir du 4-iodophénol (**Figure 113**). La première étape est une alkylation de Williamson à l'aide de bromoacétate de *tert*-butyle en milieu basique pour obtenir le composé **7** avec un rendement de 75%. La seconde étape est un couplage de Sonogashira entre un halogénure d'aryle et un alcyne. La réaction de Sonogashira, tel que l'a développé en 1975 Kenkichi Sonogashira<sup>144</sup>, permet de former une liaison entre deux carbones sp<sup>2</sup>. Le mécanisme proposé de cette réaction appliqué dans notre cas, à savoir le couplage d'un halogénure d'aryle avec un alcyne est présenté **Figure 112** et comprend deux cycles imbriqués. Le premier est une pré-catalyse grâce à l'utilisation d'un sel de cuivre (I) pour activer l'alcyne et rendre son proton suffisamment acide pour être déprotoné par la triéthylamine et former un acétylure de cuivre. Le cycle catalytique, à proprement dit, commence par une addition oxydante de l'halogénure d'aryle au sein de la sphère de coordination du palladium. Par une réaction de transmétallation avec l'acétylène de cuivre, un nouvel intermédiaire est formé qui par une élimination réductrice permet de régénérer le Pd(0) et de libérer le produit du couplage.

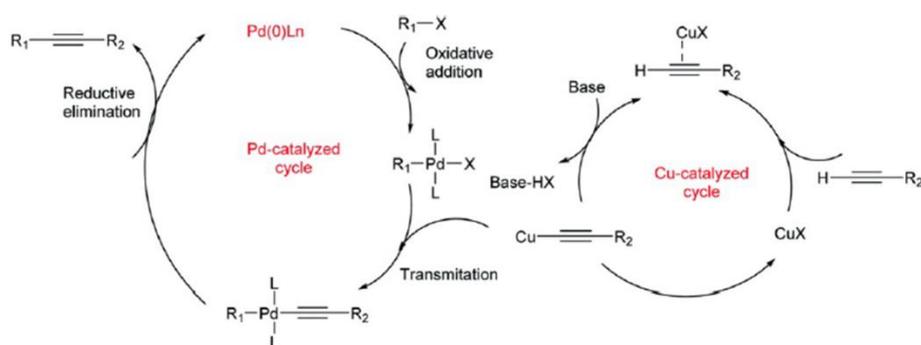


Figure 112 : Mécanisme proposé du cycle catalytique de la réaction de couplage de Sonogashira tiré de la littérature <sup>145</sup>.

Il est effectué en présence de palladium(0), ici le chlorure de *bis*(triphénylphosphine)palladium(II), de triéthylamine, d'iodure de cuivre et d'éthynyltriméthylsilane. Après 15 heures de réaction à température ambiante, la formation du produit de couplage est quasi quantitative. Ces deux étapes nous ont permis d'obtenir le composé **8** avec un rendement de 74%.

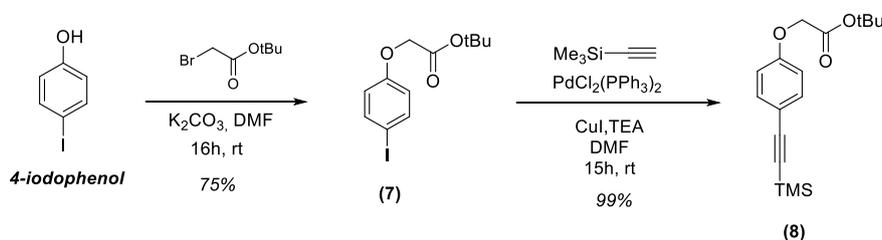
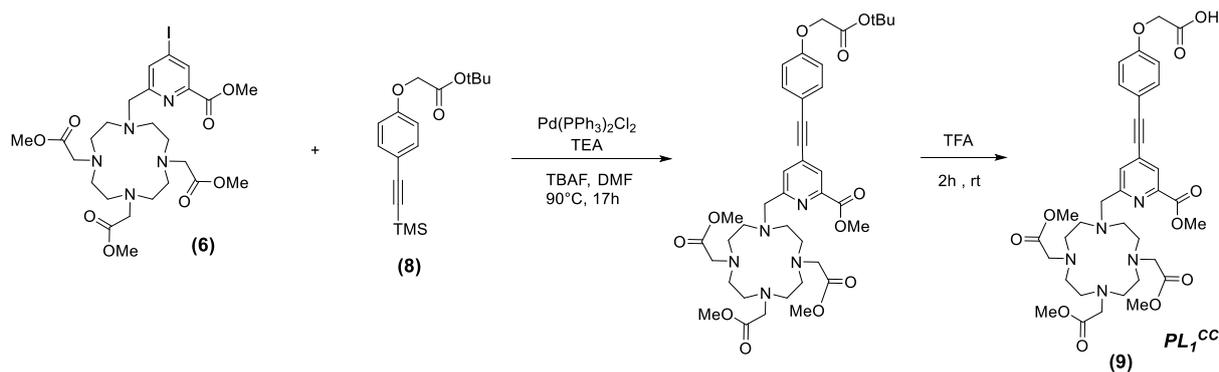


Figure 113 : Schéma de synthèse du composé **8**.

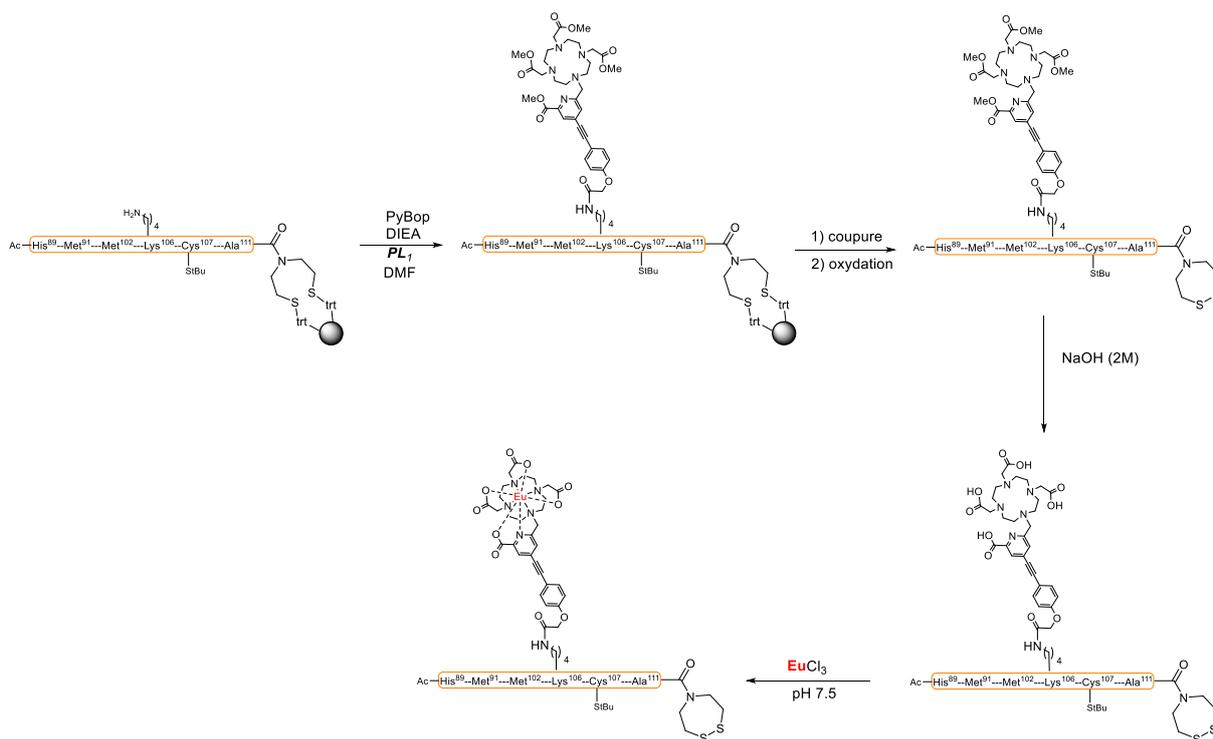
#### 1.4 - Synthèse du pro-ligand $PL_1^{CC}$

La dernière étape de synthèse du pro-ligand  $PL_1^{CC}$  est un couplage de Sonogashira catalysé au palladium(0) entre le composé **6** et le composé **8**. Il a été réalisé en présence de  $Pd(PPh_3)_2Cl_2$ , de triéthylamine et de fluorure de tétrabutylammonium (TBAF) dans le DMF. Le TBAF permet la déprotection du groupement  $SiMe_3$  de l'alcyne *in situ* pour obtenir l'alcyne vrai. Contrairement au couplage de Sonogashira précédent pour former le composé **7**, il est important de ne pas utiliser de sel de cuivre(I), car le cuivre se retrouve sous forme  $Cu^{2+}$  et complexé au macrocycle, il n'est pas possible de l'enlever ni de séparer le complexe du ligand libre. Dans ce cas,  $Cu^+$  n'est pas requis pour obtenir un bon rendement, le proton de cet alcyne vrai devant être suffisamment acide pour être déprotoné par la triéthylamine et rentrer dans le cycle catalytique présenté précédemment<sup>145</sup>. Le produit de la réaction est ensuite traité par du TFA pour libérer le groupement *tert*-butyle, il est ensuite purifié par HPLC et lyophilisé avec un rendement de 50% sur ces deux étapes (Figure 114).


 Figure 114 : Schéma de synthèse du pro-ligand  $PL_1^{CC}$ .

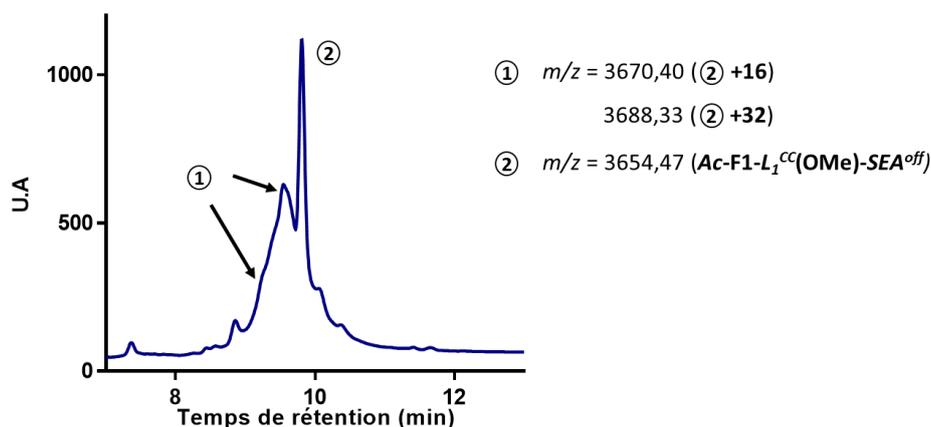
## 2 - Synthèse du fragment $Ac$ -F1- $[L_1^{CC}$ -Eu]-SEA

Le couplage du pro-ligand  $PL_1^{CC}$  a été effectué sur le fragment  $Ac$ -F1-SEA comportant des méthionines. Il est réalisé sur support solide après la déprotection sélective du groupement alloc de la lysine 106 (**Figure 115**). Un léger excès du pro-ligand  $PL_1^{CC}$  (1.1 éq) est utilisé avec de la DIEA (4 éq) et du PyBOP (2 éq) comme agent de couplage dans le DMF. Après 12 heures de réaction, le suivi par LCMS révèle un couplage quantitatif (pas de trace de peptide non acylé sur Lys<sup>106</sup>) du pro-ligand (**Figure 116**).


 Figure 115 : Schéma de synthèse du fragment  $Ac$ -F1- $[L_1$ -Eu]-SEA<sup>off</sup>.

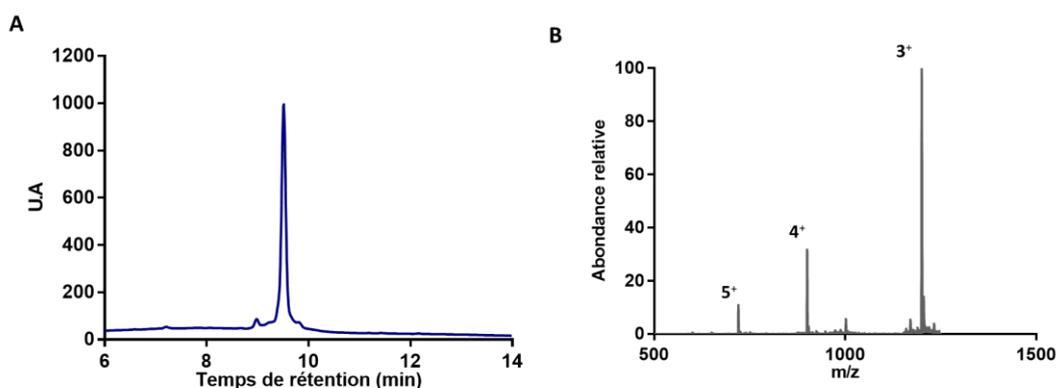
Ce suivi LCMS révèle aussi la présence de produits d'oxygénation ( $m/z = M+16$ ,  $M+32$ ) qui viennent fortement contaminer le chromatogramme. Ce problème est d'autant plus marqué après l'étape d'oxydation pour le passage de la fonction SEA<sup>on</sup> à SEA<sup>off</sup>. Il s'agit très probablement de produit

d'oxygénation au niveau des méthionines comme évoqué dans le chapitre précédent. Mais au moment de cette synthèse, nous n'avons pas encore testé l'emploi de norleucine à la place des méthionines.



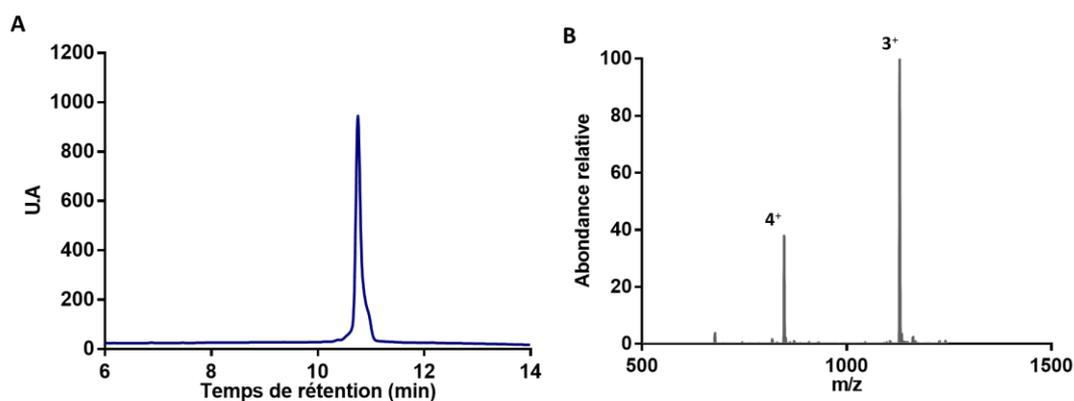
**Figure 116** : Suivi par LCMS de la réaction de couplage du pro-ligand  $PL_1^{CC}$  sur le fragment Ac-F1-SEA<sup>off</sup> après 12 heures de réaction.

En répétant plusieurs fois cette réaction, nous avons réussi à obtenir suffisamment de ce fragment pour faire l'hydrolyse des esters méthyliques des bras acétate et picolinate. Elle a été réalisée en milieu basique en dissolvant le fragment dans une solution de soude à 0.5 M durant 5 min à température ambiante. Le fragment Ac-F1- $L_1^{CC}$ (OH)-SEA<sup>off</sup> a été obtenu pur après séparation par HPLC (**Figure 117**).



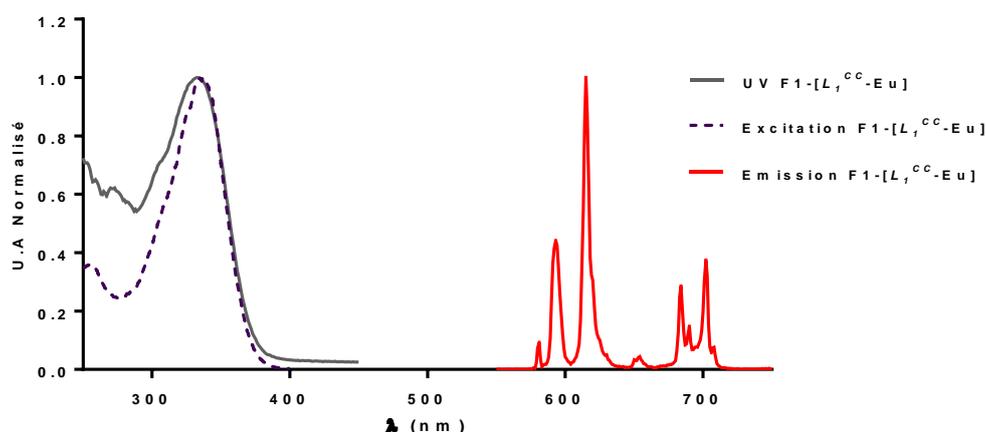
**Figure 117** : A) Chromatogramme HPLC, détection UV à 214 nm du fragment Ac-F1- $L_1^{CC}$ (OH)-SEA<sup>off</sup> purifié. B) Le spectre ESI-MS du fragment Ac-F1- $L_1^{CC}$ (OH)-SEA<sup>off</sup> purifié (masse théorique calculée : 3599.35, masse déconvoluée obtenue : 3598.38).

La dernière étape est la complexation du ligand avec l'euprium. Celle-ci a été réalisée par dissolution de 1 mg de Ac-F1- $L_1^{CC}$ (OH)-SEA<sup>off</sup> dans l'eau (concentration en peptide 0.5 M) en présence d'un excès de chlorure d'euprium(III). Le pH de la solution a été ajusté à 9 pendant 4 heures puis la solution a été agitée à température ambiante pendant 12 heures à pH 7.5. Malgré la formation d'oxyde d'euprium à ce pH, la complexation du ligand est complète et le fragment Ac-F1- $[L_1^{CC}-Eu]$ -SEA<sup>off</sup> a pu être isolé et caractérisé par LCMS (**Figure 118**).



**Figure 118** : A) Chromatogramme HPLC, détection UV à 214 nm du fragment Ac-F1-[L<sub>1</sub><sup>CC</sup>-Eu]-SEA<sup>off</sup> purifié. B) Le spectre ESI-MS du fragment Ac-F1-[L<sub>1</sub><sup>CC</sup>-Eu]-SEA<sup>off</sup> purifié (masse théorique calculée : 3747.28, masse déconvoluée obtenue : 3748.8).

La **Figure 119** présente les spectres normalisés d'absorption, d'émission résolue en temps et d'excitation du fragment Ac-F1-[L<sub>1</sub><sup>CC</sup>-Eu]-SEA<sup>off</sup> dans le tampon HEPES 10 mM pH 7.5. Le spectre d'émission confirme la présence d'euporium. Le complexe d'euporium présente une longueur d'onde maximale d'absorption et d'excitation à 332 nm comme attendu pour cette antenne en comparaison du complexe Eu<sup>L1B</sup> de Maury et [L<sub>1</sub><sup>CC</sup>-Eu] de Guillaume Frémy (335 nm).



**Figure 119** : Spectre d'absorbance (gris), d'émission (rouge) obtenus après excitation à  $\lambda_{ex} = 332$  nm et spectre d'excitation (pointillé gris) obtenue après détection à  $\lambda_{em} = 615$  nm du fragment Ac-F1-[L<sub>1</sub><sup>CC</sup>-Eu]-SEA<sup>off</sup> dans une solution aqueuse de tampon HEPES (10 mM, pH 7.0).

Le déclin de luminescence de l'euporium à 615 nm a été enregistré pour obtenir le temps de vie du lanthanide, il a été mesuré à 1.04 ms, ce qui est identique à celui du complexe [L<sub>1</sub><sup>CC</sup>-Eu]. Le rendement quantique et le coefficient d'absorption molaire du complexe [L<sub>1</sub><sup>CC</sup>-Eu] non greffé calculés par G. Frémy et valent respectivement  $\Phi_{Eu}^{L1CC} = 17\%$  et  $\epsilon_{332}^{L1CC} = 28200 \text{ M}^{-1}\text{cm}^{-1}$ . Nous n'avons pas remesuré ces valeurs avec le complexe greffé sur le peptide, car il y a peu de raisons qu'elles changent du fait d'avoir couplé le carboxylate du groupement électrodonneur au peptide : le coefficient d'absorption molaire ne doit pas varier si le solvant est le même (eau) et le rendement quantique doit être assez proche dans la mesure où le temps de vie est identique au complexe [L<sub>1</sub><sup>CC</sup>-Eu].

Ce complexe répond à tous nos critères d'exigences avec une longueur d'onde d'excitation dans la zone où la SiR<sup>Me</sup> n'absorbe pas (à 350 nm, l'absorption du complexe est encore forte, i.e. environ 22 000 M<sup>-1</sup>cm<sup>-1</sup>) et une forte brillance (4794 M<sup>-1</sup>cm<sup>-1</sup>).

### 3 - Synthèse de la protéine Zif12-[L<sub>I</sub><sup>CC</sup>-Eu]-A

Malgré les difficultés de synthèse du fragment Ac-F1-[L<sub>I</sub><sup>CC</sup>-Eu]-SEA<sup>off</sup> nous avons été en mesure de faire la dernière étape de ligation à partir de 0.3 mg de ce fragment avec le fragment H-F2-F3-NH<sub>2</sub>. La réaction de ligation chimique native a été réalisée dans les mêmes conditions que la synthèse de la protéine Zif12-(SiR<sup>Me</sup>)-A (Chapitre III : I-3). Sur le suivi par LCMS après 24 heures de réaction, présenté **Figure 120**, on observe la formation de trois produits mais aucun ne correspond à la masse attendue. Un premier, noté (3) sur la **Figure 120**, élue vers 8 minutes et présente un *m/z* de 9726.39 qui correspond à un incrément de 250 Da par rapport à la protéine Ac-F1-[L<sub>I</sub>-Eu]-F2-F3-NH<sub>2</sub>. Deux autres produits notés (4), vers 9 minutes, présente le même *m/z* (3910.16), ce qui correspond à +250 Da par rapport au fragment Ac-F1-[L<sub>I</sub>-Eu]-SEA<sup>on</sup>. Le chromatogramme obtenu par détection à 331 nm ne montre aucun pic au temps de rétention correspondant à ce produit, suggérant la modification de l'antenne pendant la réaction. En éclairant la solution avec une lampe UV (354 nm), l'émission rouge de l'euporium n'est plus visible. Cette addition de 250 Da peut correspondre à l'addition d'une molécule de TCEP sur la triple liaison alcyne. En parallèle de ces essais, Ji-Hyung Choi, autre doctorante au laboratoire a essayé de préparer un autre peptide avec ce complexe conjugué et un phénomène similaire a été observé. Après analyse du produit obtenu, elle a aussi observé l'addition du TCEP sur cette antenne. Elle a aussi observé qu'une addition de la cystéine (thiol) sur l'alcyne peut aussi se produire (réaction thiol-yne) en remplaçant le TCEP par du DTT (dithiotréitol, un dithiol). Elle a également observé qu'une cystéine du peptide ou un thioether (thioanisole) pouvaient s'ajouter sur cette fonction alcyne, très susceptible à la réaction avec des nucléophiles. La disparition de l'absorption à 331 nm plaide en faveur d'une addition de TCEP sur la triple liaison, conduisant à une protéine non fonctionnelle. Bien que le complexe [L<sub>I</sub><sup>CC</sup>-Eu] soit très prometteur cette réaction parasite, nous a conduits à abandonner l'idée d'utiliser ce complexe.

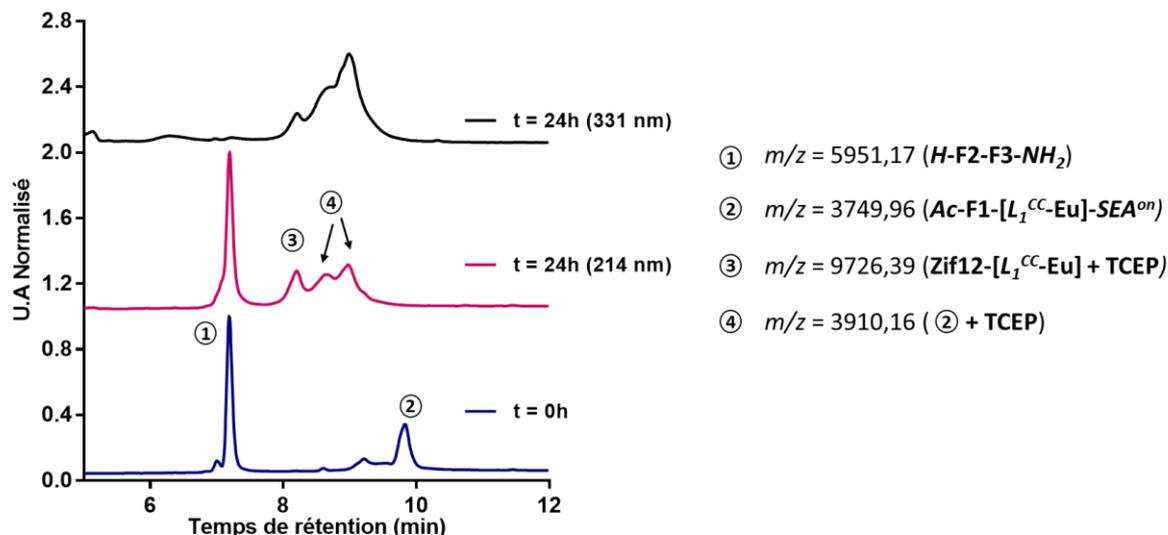


Figure 120 : Suivi par LCMS de la réaction de ligation du fragment  $Ac-F1-[L_1^{CC}-Eu]-SEA^{off}$  et  $H-F2-F3-NH_2$  à 214 et 331 nm.

## IV - Synthèse de complexes avec des nouvelles antennes

### 1 - Nouvelle voie de synthèse

Pour la recherche d'une nouvelle antenne pour l'euporium, nous avons décidé d'envisager une autre stratégie de synthèse. La déprotection des esters méthyliques, bien qu'efficace, rajoute une étape et une purification. De plus, les conditions d'hydrolyse basique (concentration de soude, temps de réaction) doivent être optimisées pour chaque peptide en fonction de sa solubilité en milieu très basique. Nous avons envisagé deux choses. La première est l'utilisation d'un macrocycle dont les bras carboxyliques sont protégés par des groupements *tert*-butyle et la fonction de bioconjugaison par un groupement O-méthyle. Ainsi lors de la coupure du peptide de sa résine en condition acide (TFA), ces groupements seront retirés, ce qui permet de s'affranchir de l'étape de déprotection après greffage. Même si cette déprotection peut être lente (8h)<sup>142</sup>, ce n'est pas un problème si la réaction est propre. La seconde chose est de greffer le complexe non plus au niveau de la partie électrodonneuse de l'antenne, mais au niveau de la fonction acide carboxylique du picolinate. Ce changement permettra de rapidement obtenir différentes antennes en faisant varier le donneur à partir de produits commerciaux d'halogénures d'aryle fonctionnalisés. L'espaceur de type alcyne présent sur le complexe  $[L_1^{CC}-Eu]$  est susceptible de subir l'addition de thiol, nous allons donc nous tourner vers la synthèse de systèmes de types biaryle, comportant un groupement aryle sur la position 4 du picolinate. Cette nouvelle stratégie est présentée **Figure 121**. La partie donneuse de l'antenne sera rattachée en position *para* de l'azote du picolinate par une réaction de couplage de Suzuki-Miyaura entre un halogénure d'aryle et un ester boronique<sup>146</sup>.

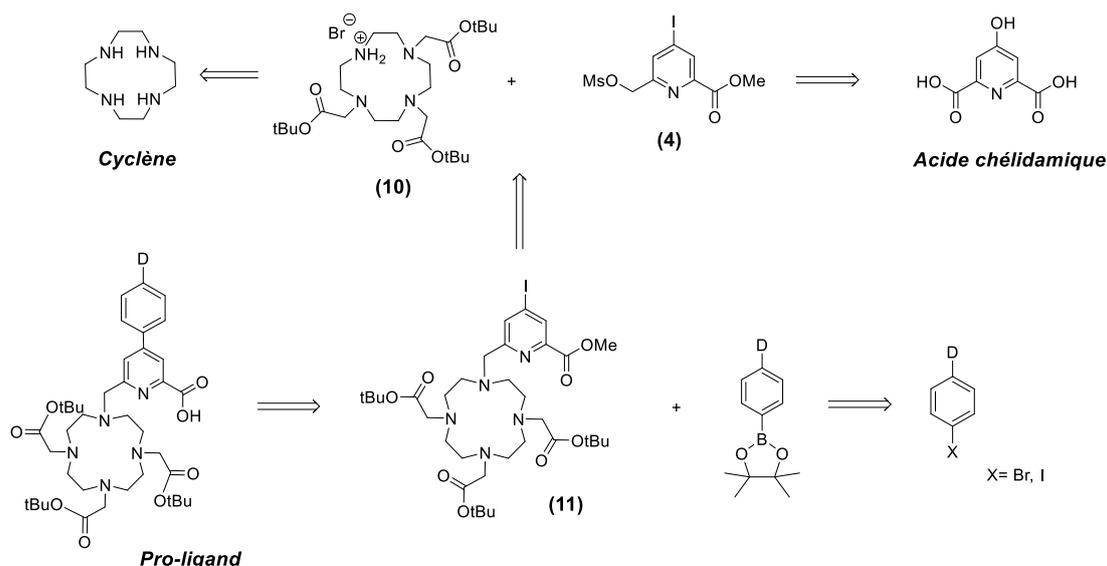


Figure 121 : Nouvelle stratégie de synthèse des pro-ligands, D représente le groupement donneur de l'antenne « push-pull ».

La réaction de Suzuki-Miyaura est le point clé de cette nouvelle stratégie de synthèse. Elle a été publiée en 1979 par Akira Suzuki et Norio Miyaura<sup>146</sup>. Cette nouvelle réaction aujourd'hui amplement utilisée a fait l'objet du prix Nobel de chimie en 2010. Elle permet de former une liaison carbone-carbone entre deux aryles. C'est une réaction catalysée par du Pd(0) en présence d'une part d'un halogénure d'aryle et d'autre part d'un acide boronique d'aryle. Le mécanisme de cette réaction de couplage est présenté Figure 122.

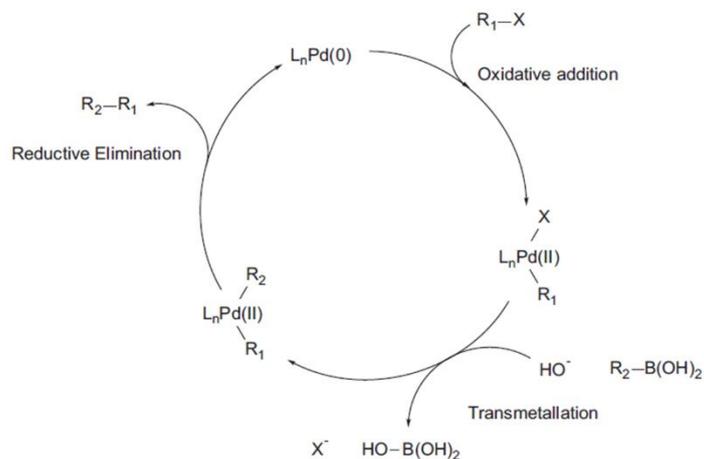


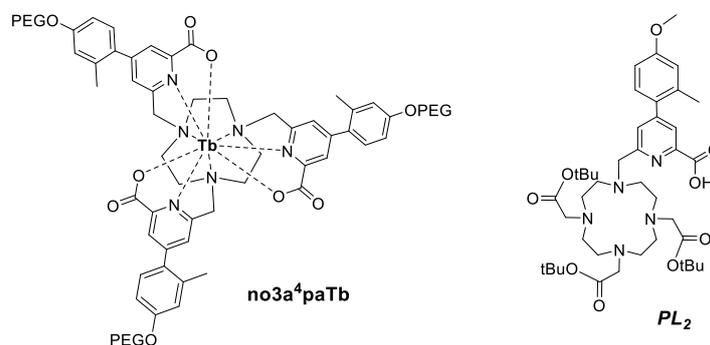
Figure 122 : Schéma du cycle catalytique du couplage de Suzuki-Miyaura<sup>146</sup>.

La première étape est une addition oxydante du palladium dans la liaison C-X de l'halogénure d'aryle pour donner l'espèce  $R_1Pd(II)X$ . La transmétallation avec l'acide boronique en présence de base permet la formation de l'espèce  $R_1Pd(II)R_2$ . Enfin l'élimination réductrice aboutit à la formation du composé  $R_1-R_2$  tout en recouvrant le Pd(0). Plusieurs paramètres pour ce couplage sont à prendre en compte. Pour l'étape d'addition oxydative, plus la liaison Ar-X est polarisée plus l'insertion du palladium est efficace ( $Ar-I > Ar-Br \geq Ar-OTf > Ar-Cl$ )<sup>147</sup>. La transmétallation avec l'acide boronique d'aryle est plus

difficile lorsque l'aryle est déficient en électrons. Cependant le facteur le plus important est la nature des ligands du palladium qui peuvent améliorer l'efficacité de ce processus catalytique<sup>148</sup>. La nature du composé boré est aussi un facteur à prendre en considération, si les acides boroniques sont fréquemment utilisés, ils souffrent d'un manque de stabilité. Pour pallier à cela la formation d'ester boronique est maintenant d'usage à partir de 4,4,5,5-tetraméthyl-1,3,2-dioxa-borolane ou de *bis*(pinacolato)biorane<sup>149</sup>.

## 2 - Antennes dérivées de l'anisole

En premier lieu, pour sécuriser nos résultats et commencer à travailler avec des protéines fonctionnalisées rapidement, nous avons commencé par utiliser un nouveau ligand avec une antenne anisole-picolinate, inspiré des travaux de O. Maury<sup>9</sup>, avec un groupement donneur de type méthoxy (**Figure 123**). Ce ligand est l'analogue de celui développé par Guillaume Frémy (**Chapitre I, Figure 53**), mais avec la fonction de conjugaison située au niveau du picolinate plutôt qu'au niveau du groupement électrodonneur. Il nous permettra aussi d'évaluer l'influence sur les propriétés spectroscopiques de l'utilisation du picolinate pour la conjugaison (picolinate vs picolinamide). Bien qu'il ait été conçu pour sensibiliser efficacement le terbium, cette antenne anisole-picolinate est aussi capable de sensibiliser l'euprium<sup>150</sup>.



**Figure 123** : Schéma du complexe *no3a<sup>4</sup>paTb* développé par O. Maury et le pro-ligand *PL<sub>2</sub>*.

### 2.1 - Synthèse du pro-ligand *PL<sub>2</sub>*

À partir du produit commercial, la 4-bromo-3-méthylanisole, la formation de l'ester boronique est réalisée en présence de 4,4,5,5-tetraméthyl-1,3,2-dioxa-borolane (1.5 éq) associé avec du chlorure de *bis*(triphénylphosphine)palladium(II) et de la triéthylamine (3 éq), le tout dissout dans du dioxane anhydre. La réaction est laissée 48 heures à 90°C. Après traitement de la réaction et purification sur colonne de gel de silice le composé **12** est isolé avec un rendement de 88%. Le macrocycle DO3A-*tris*(tBu)ester (composé **10**) a été obtenu avec un rendement de 79% par réaction du cyclène avec 3.3 éq de bromoacétate de *tert*-butyle dans le DMAc à -20°C en présence d'acétate de sodium 3.3 éq. Le macrocycle iodé **11** a été obtenu après une réaction de **4** avec **10** dans l'acétonitrile en présence de Na<sub>2</sub>CO<sub>3</sub>, il a été purifié par cristallisation (80%). Le couplage de Suzuki-Miyaura du composé **12** est

réalisé en présence de 0.8 éq du composé **11**, de *tetrakis*(triphenylphosphine) palladium(0), de fluorure de césium, le tout dans du DMF sous argon. Après 24h de réaction à 90°C, le milieu réactionnel est traité et le produit brut est repris dans de l'éthanol. Une solution concentrée de soude est ajoutée pour obtenir une concentration en soude de 2 M, afin de réaliser la déprotection de l'ester méthylique du picolinate. Après 10 minutes de réaction à température ambiante, le milieu réactionnel est extrait avec un mélange AcOEt/H<sub>2</sub>O. Après évaporation du solvant, le produit brut est purifié sur colonne HPLC, le rendement sur ces deux étapes est de 67% (**Figure 124**).

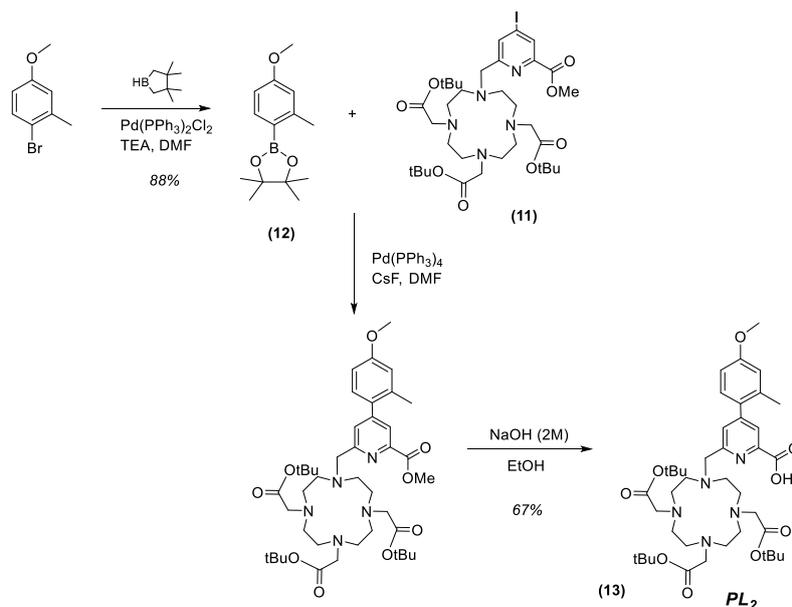
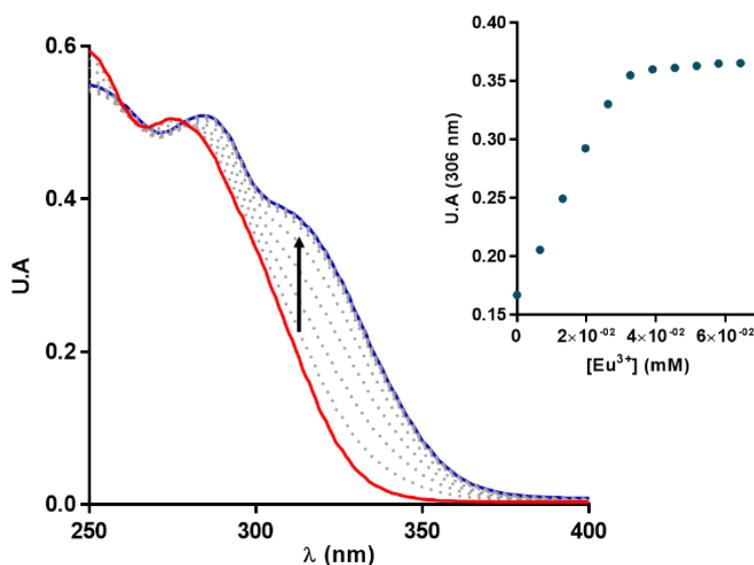


Figure 124 : Schéma de synthèse du pro-ligand  $PL_2$ .

## 2.2 - Caractérisation spectroscopique du complexe TAT-[ $L_2$ -Eu]

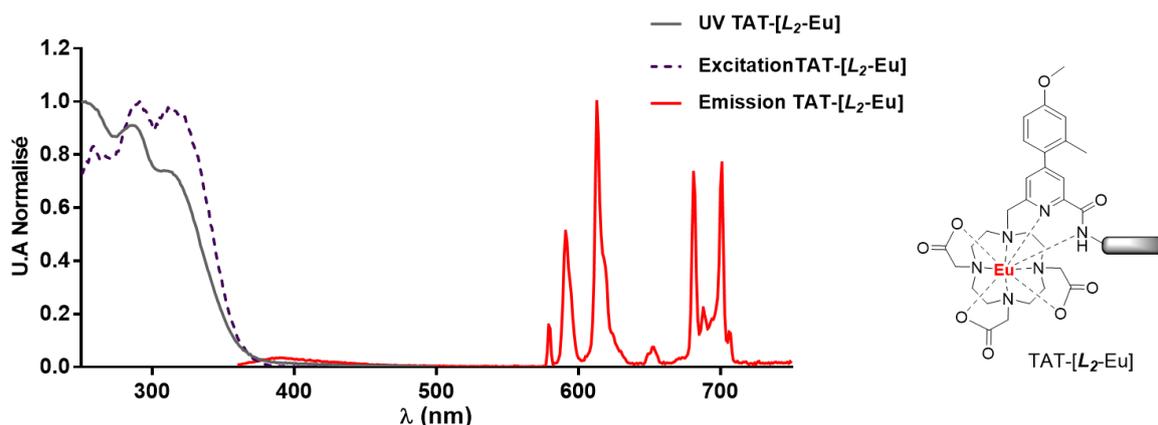
La caractérisation du complexe [ $L_2$ -Eu] a été réalisée après conjugaison sur le peptide TAT. C'est un peptide de 11 acides aminés, de séquence  $Ac$ -KRRKRRRQRRRG- $NH_2$ , avec l'amine N-terminale acétylée et le C-terminale sous la forme d'un amide. La lysine en N-terminale est celle dont la chaîne latérale est utilisée pour la conjugaison. Ce peptide fait partie de la classe des peptides de pénétration cellulaire et dérive du virus du SIDA<sup>151</sup>. Il est utilisé dans l'équipe pour cette application, mais il sert aussi de peptide modèle pour comparer les propriétés photophysiques des différents complexes de  $Ln^{3+}$ , indépendamment du peptide. Son caractère cationique important (8 charge positives) permet de solubiliser des complexes avec des antennes très hydrophobes telles que les antennes carbazoles décrites plus loin. Ce peptide est synthétisé sur résine Rink Amide (chimie Fmoc) avec une lysine(alloc) comme acide aminé N-terminale. Après couplage de  $PL_2$  (PyBOP/DIEA), le traitement de la résine par une solution de TFA en présence de scavengers : TIS et eau (94/3/3, v/v/v) permet le décrochage du peptide et la déprotection des esters *tert*-butyliques. Après purification du peptide, la complexation avec un sel d'euporium ( $EuCl_3$ , 1.2 éq) est réalisée à pH 7.5 durant 12h pour donner le peptide TAT-[ $L_2$ -Eu] après purification.

Le spectre d'absorbance du complexe TAT-[L<sub>2</sub>-Eu] a été enregistré dans une solution aqueuse tamponnée par du phosphate (10 mM, pH 7.0) (**Figure 126**). Il présente un maximum d'absorption à 285 nm avec un épaulement marqué vers 315 nm. La bande située à 285 nm est attribuée à une transition vers un état localement excité (LE) de l'antenne (transition entre deux orbitales qui ont un bon recouvrement) et la bande à 305 nm correspond à une transition vers un état à transfert de charge noté ILCT pour « intraligand charge transfert » (transition d'un électron vers une orbitale centrée sur le donneur – aryle substitué dans notre cas - vers une orbitale centrée sur l'accepteur – picolinate dans notre cas – avec peu de recouvrement entre ces orbitales). Le coefficient d'absorption molaire du complexe TAT-[L<sub>2</sub>-Eu] a été déterminé par titration d'une solution aqueuse du ligand par une solution de chlorure d'euprium (III) de concentration connue. L'évolution du spectre d'absorbance en fonction des ajouts d'euprium permet de remonter à la concentration du complexe TAT-[L<sub>2</sub>-Eu]. Le point de rupture de la courbe correspond à la formation d'un complexe 1:1. À l'aide de la loi de Beer-Lambert, on peut déterminer le coefficient d'absorption molaire du ligand à partir de la valeur de l'absorbance et de la concentration à l'équivalence. Ce coefficient est de 11800 M<sup>-1</sup>cm<sup>-1</sup> à 315 nm pour le complexe TAT-[L<sub>2</sub>-Eu] (**Figure 125**).



**Figure 125** : Détermination du coefficient d'absorption molaire du complexe [L<sub>2</sub>-Eu] à 315 nm.

La **Figure 126** montre le spectre d'émission en excitant à 315 nm dans la bande de transfert de charge. Il présente les raies caractéristiques des transitions <sup>5</sup>D<sub>4</sub> → <sup>7</sup>F<sub>J</sub> (J = 0 - 4) de l'euprium à 580 nm, 585, 615, 650 et 700 nm, respectivement. Le spectre d'excitation, obtenu par détection au maximum d'émission de l'euprium à 615 nm, comporte deux maxima à 290 nm et 308 nm, en accord avec le spectre d'absorption, ce qui confirme la sensibilisation de l'euprium par l'antenne anisole-picolinamide. Le temps de vie de l'euprium a été mesuré à 1.05 ms. Ce temps de vie est similaire à ceux obtenus pour le complexe [L<sub>1</sub><sup>CC</sup>-Eu] et [L<sub>2</sub>-Eu] non conjugué, il est caractéristique d'un Eu<sup>3+</sup> sans molécules d'eau dans sa sphère de coordination avec le ligand DO3Apic<sup>142,152</sup>.

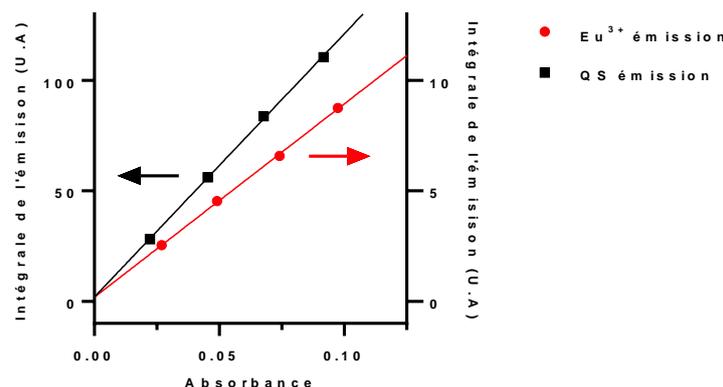


**Figure 126** : Spectres normalisés d'absorption (gris), d'émission (rouge) obtenus après excitation à  $\lambda_{ex} = 315$  nm et d'excitation (pointillé gris) obtenus après détection à  $\lambda_{em} = 615$  nm du conjugué TAT-[L<sub>2</sub>-Eu] dans une solution aqueuse de tampon phosphate (10 mM, pH 7.0).

La détermination du rendement quantique du complexe TAT-[L<sub>2</sub>-Eu] a été réalisée selon la méthode comparative<sup>153,154</sup>. Une solution du complexe TAT-[L<sub>2</sub>-Eu] dans une solution aqueuse tamponnée par du tampon phosphate (10 mM, pH 7.0) a été préparée pour obtenir une valeur d'absorbance proche de 0.1, une solution de référence a aussi été préparée. Il existe plusieurs références dont le rendement quantique est connu<sup>153</sup>, nous avons choisi comme référence la quinine sulfate dont le rendement quantique est de 54.6 % avec des maxima d'absorption et d'émission à 345 et 451 nm respectivement. Une solution de cette quinine, dans une solution d'acide sulfurique 0.5 M, est préparée avec une absorbance proche de 0.1. L'enregistrement des spectres d'émission de fluorescence de ces deux solutions (référence et TAT-[L<sub>2</sub>-Eu]) à la longueur d'onde d'excitation du complexe est réalisé. À partir de ces solutions, des spectres d'absorption et d'émission sont enregistrés après trois dilutions successives pour obtenir des absorbances respectives proches de 0.075, 0.05 et 0.0025. L'intégration du signal émis par fluorescence en fonction de l'absorbance de la solution permet de déterminer le rendement quantique selon l'équation (6) grâce au coefficient directeur des deux droites obtenues (**Figure 127**).

$$\Phi_{\text{complexe}} = \frac{\text{coef directeur}_{\text{complexe}}}{\text{coef directeur}_{\text{référence}}} \times \Phi_{\text{référence}} \quad (6)$$

Le rendement quantique obtenu est de 4 %. Bien que le rendement quantique du complexe soit faible, c'est un point de départ dans notre recherche d'une antenne plus efficace pour la sensibilisation de l'euporium.

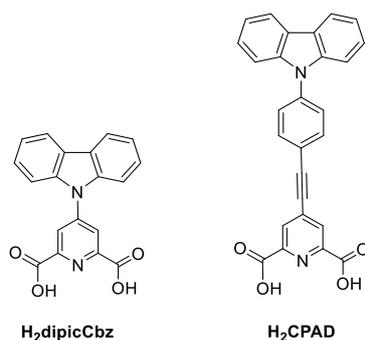


**Figure 127 :** Courbes représentant l'intégrale du signal émis par le complexe TAT-[L<sub>2</sub>-Eu] et le sulfate de quinine (référence) en fonction de l'absorbance des solutions. Intensité obtenue après excitation du complexe et de la référence à 315 nm.

Compte tenu du rendement quantique du complexe TAT-[L<sub>2</sub>-Eu], bien qu'il semble suffisant pour avoir un signal correct, nous avons décidé de synthétiser le complexe TAT-[L<sub>2</sub>-Tb] dont le rendement quantique est très élevé. Il a été calculé à 65% avec un coefficient d'absorption molaire de 11 800 M<sup>-1</sup>cm<sup>-1</sup> sa brillance est de 7670 M<sup>-1</sup>cm<sup>-1</sup>, bien supérieure à celle du complexe TAT-[L<sub>2</sub>-Eu] ( $B = 472$  M<sup>-1</sup>cm<sup>-1</sup>). Le complexe TAT-[L<sub>2</sub>-Tb] a été synthétisé et caractérisé suivant la méthode décrite précédemment pour le complexe TAT-[L<sub>2</sub>-Eu]. La fonctionnalisation de la protéine par ce complexe [L<sub>2</sub>-Tb] nous permettra d'avoir une première approche sur le transfert d'énergie de type FRET à partir d'un complexe de terbium à haute brillance. L'objectif est de pouvoir valider notre approche avant d'envisager l'utilisation du complexe d'euporium dont le rendement quantique est bien plus faible.

### 3 - Antennes dérivées du carbazole

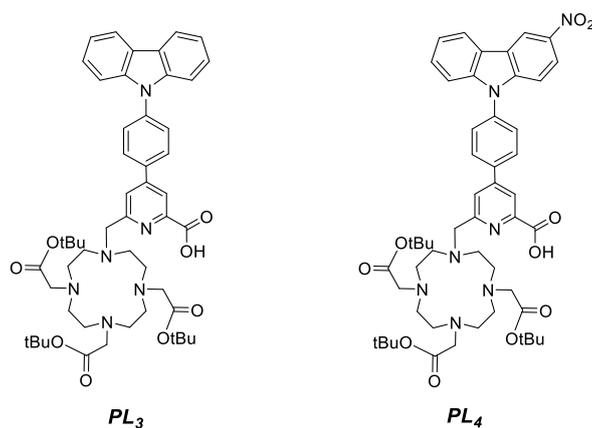
Au cours de nos recherches pour trouver des exemples d'antennes capables de sensibiliser efficacement l'euporium, avec une excitation vers 350 nm, nous avons trouvé les travaux d'Ana de Bettencourt-Dias. Ces travaux portent sur le développement de complexe de lanthanide pour l'imagerie mono et bi-photonique (**Figure 128**). Un premier exemple de complexe H<sub>2</sub>dipicCbz dont l'antenne est composée d'une unité carbazole (Cz) reliée sur un dipicolinate en position *para* de l'atome d'azote a été publié en 2017<sup>155</sup>. Ce ligand forme un complexe avec le terbium avec une stœchiométrie 3:1. Son spectre d'excitation enregistré dans un tampon Tris-HCl contenant 4% de DMSO (v/v) présente une large bande qui s'étend de 250 nm à 400 nm avec un rendement quantique de 3.8%. Quelques années plus tard, un autre complexe sensibilisant efficacement l'euporium a été reporté dans cette même équipe<sup>156</sup>. Il s'agit d'une antenne où un espaceur de type alcyne-aryle a été ajouté entre l'unité carbazole et le dipicolinate pour une meilleure absorption bi-photonique. On retrouve le même schéma d'antenne « push-pull » décrit par O. Maury et D. Parker. (**Chapitre I, IV-3**). Ce ligand H<sub>2</sub>CPAD forme un complexe 3:1 avec l'euporium dont le rendement quantique a été mesuré à 31.7%, dans le DMSO, car il n'est pas soluble dans l'eau du fait du fort caractère hydrophobe de l'antenne.



**Figure 128** : Antennes synthétisées au sein de l'équipe de A. Bettencourt-Dias<sup>155,156</sup>.

### 3.1 - Synthèse des pro-ligands **PL<sub>3</sub>** et **PL<sub>4</sub>**

Nous avons décidé de synthétiser deux antennes comportant une unité carbazole comme groupement électrodonneur pour notre système d'antenne « push-pull ». La première antenne est composée d'une unité picolinate fonctionnalisée en position *para* de l'atome d'azote par un groupement phényle-carbazole. La seconde antenne est un dérivé de la première où un groupement électroattracteur nitro est ajouté en position 9 du carbazole (**Figure 129**). Nous justifierons par la suite l'introduction de ce groupement nitro.



**Figure 129** : Schéma des pro-ligands **PL<sub>3</sub>** et **PL<sub>4</sub>** comportant une unité carbazole comme groupement électrodonneur de l'antenne « push-pull ».

La synthèse du pro-ligand **PL<sub>3</sub>** se divise en trois étapes. La première est la formation d'un ester boronique à partir du 9-(4-bromophényl)-9H-carbazole par le couplage de Suzuki-Miyaura. Un second couplage du même type est réalisé avec le composé macrocyclique iodé **11**, enfin une étape de déprotection de l'ester méthylique du picolinate aboutit au composé **14**. Le pro-ligand **PL<sub>4</sub>** est obtenu de la même manière, mais une étape supplémentaire de nitration est réalisée après la formation de l'ester boronique (**Figure 130**).

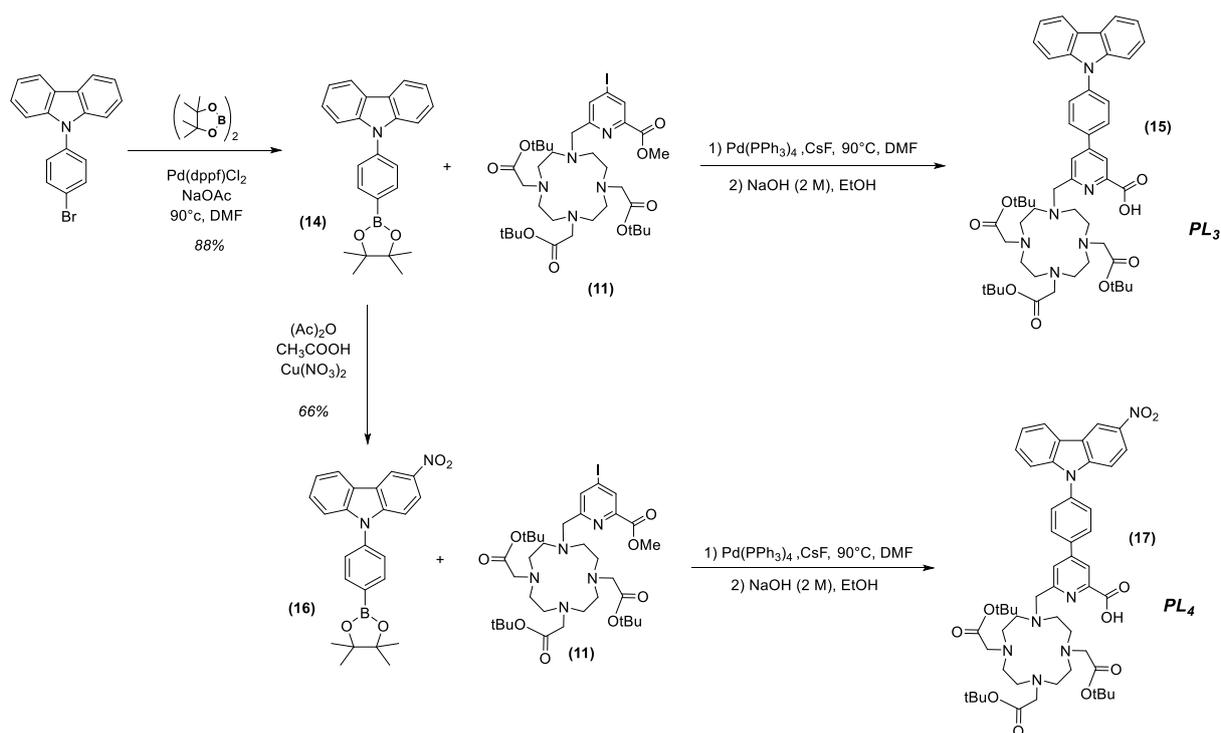


Figure 130 : Schéma de synthèse des pro-ligands  $PL_3$  et  $PL_4$ .

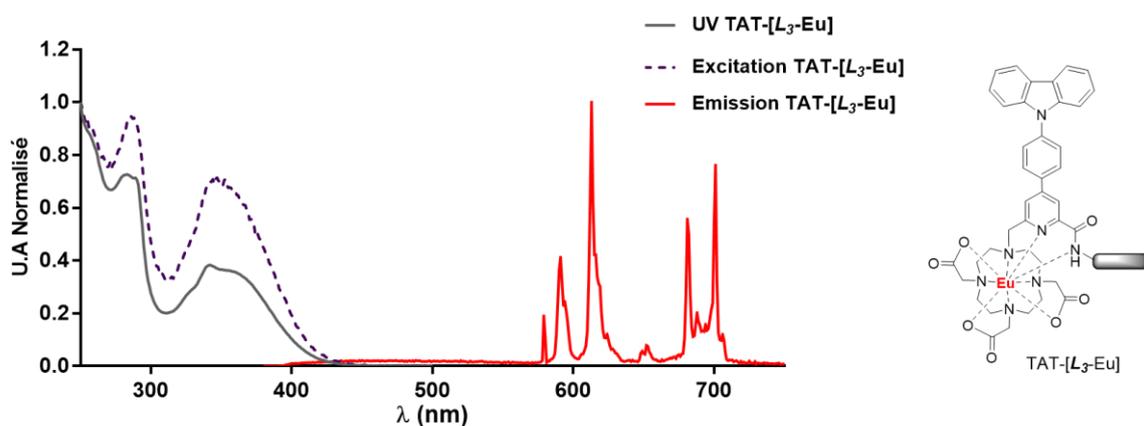
La première étape de couplage de Suzuki-Miyaura est réalisée à partir du 9-(4-bromophényl)-9H-carbazole en présence de [1,1'-bis(diphénylphosphino)ferrocène]palladium(II), de bis(pinacoloto) diboron et d'acétate de sodium dans le DMF. Après 48 heures de réaction à  $90^\circ C$  le composé **14** a été purifié sur colonne de gel de silice avec un rendement de 88%. Le composé **16** a été obtenu après avoir fait réagir le composé **14** en présence de nitrate de cuivre(II) (0.5 éq) dans un mélange acide acétique/anhydride acétique (0.7/1.2, v/v)<sup>157</sup>. Le rendement après purification sur colonne de gel de silice est de 66%.

Le couplage des composés **14** et **16** sur le composé **11** et l'étape de déprotection de l'ester méthylique du picolinate ont été réalisés de la même manière que précédemment (pour  $PL_2$ ) à l'exception du catalyseur, pour lequel nous avons substitué le  $Pd(PPh_3)_4$  par son homologue supporté sur résine afin de faciliter l'élimination du catalyseur (par filtration) lors du traitement. Les pro-ligands  $PL_3$  et  $PL_4$  ont été obtenus avec suffisamment de pureté pour être utilisés sans purification. Seule une partie du produit brut obtenu a été purifiée pour les caractérisations par HRMS et RMN.

### 3.2 - Caractéristiques spectroscopiques des complexes TAT-[ $L_3$ -Eu] et TAT-[ $L_4$ -Eu]

Les deux pro-ligands synthétisés présentent une faible solubilité dans l'eau. L'étude des caractéristiques spectroscopiques a été réalisée après greffage du complexe sur le peptide TAT tel que décrit pour le complexe TAT-[ $L_2$ -Eu]. Cette synthèse et les caractérisations spectroscopiques de TAT-[ $L_3$ -Eu] ont été réalisés par Ji-Hyung Choi pendant sa thèse et sont résumées ci-dessous. Le spectre d'absorbance du complexe TAT-[ $L_3$ -Eu] dans une solution aqueuse tamponnée par du phosphate (10 mM, pH 7.5)

présente une large bande qui s'étend de 310 à 430 nm. Le coefficient d'absorption molaire au maximum d'absorption à 345 nm du complexe a été mesuré à  $14500 \text{ M}^{-1}\text{cm}^{-1}$ . L'excitation de ce complexe à 345 nm permet d'observer l'émission caractéristique de l'euprium et le spectre d'excitation, obtenu en détectant l'émission de l'euprium à 615 nm, confirme la sensibilisation de l'euprium par l'antenne (**Figure 131**). Le temps de vie de luminescence de l'euprium est de 0.74 ms, plus bas qu'habituellement observé avec ce type de ligand DO3Apic ( $\approx 1.05 \text{ ms}$ ). Il est probable qu'un transfert d'énergie retour de l' $\text{Eu}^{3+}$  vers un état excité de basse énergie de l'antenne ait lieu. L'absence de molécule d'eau coordonnée à l'euprium a été vérifiée par comparaison des temps de vie de luminescence dans  $\text{H}_2\text{O}$  et  $\text{D}_2\text{O}$ . Le rendement quantique du complexe est de 0.2%, ce qui est beaucoup plus faible que celui obtenu pour TAT- $[\text{L}_2\text{-Eu}]$  (4%). Le rendement quantique est trop faible pour envisager une utilisation du complexe  $[\text{L}_3\text{-Eu}]$  dans notre système FRET.



**Figure 131** : Spectres normalisés d'absorption (gris), d'émission (rouge) obtenus après excitation à  $\lambda_{ex} = 340 \text{ nm}$  et d'excitation (pointillé violet) obtenu après détection à  $\lambda_{em} = 615 \text{ nm}$  du complexe TAT- $[\text{L}_3\text{-Eu}]$  dans une solution aqueuse tamponnée par du tampon phosphate (10 mM, pH 7.0).

Le faible rendement quantique du peptide TAT- $[\text{L}_3\text{-Eu}]$  en tampon phosphate a été attribué à une compétition entre (i) le transfert d'énergie entre l'antenne excitée et l' $\text{Eu}^{3+}$  et (ii) un transfert d'électron photo-induit du carbazole à  $\text{Eu}^{3+}$  qui conduit à la réduction de l'euprium en  $\text{Eu}^{2+}$ , tuant la luminescence. Ce dernier processus est majoritaire dans l'eau. Néanmoins, il est apparu pendant cette étude que ce complexe avait un comportement solvatochromique assez marqué. Lorsque la polarité du solvant diminue, le rendement quantique augmente : la polarité du solvant favorise le transfert d'électron par rapport au transfert d'énergie. Dans le DMF, un rendement quantique de 5.8 % a été mesuré. Lors de la synthèse de TAT- $[\text{L}_3\text{-Eu}]$ , une réaction parasite inattendue s'est produite et un deuxième composé a pu être isolé, en proportion égale à TAT- $[\text{L}_3\text{-Eu}]$ . Il s'agit du peptide TAT conjugué au complexe  $[\text{L}_3\text{-Eu}]$ , mais sur lequel un groupement Pbf (2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl), groupement protecteur des arginines, est venu s'ajouter sur la position 9 du carbazole (**Figure 132**).

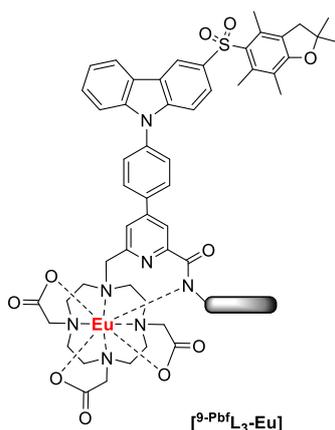
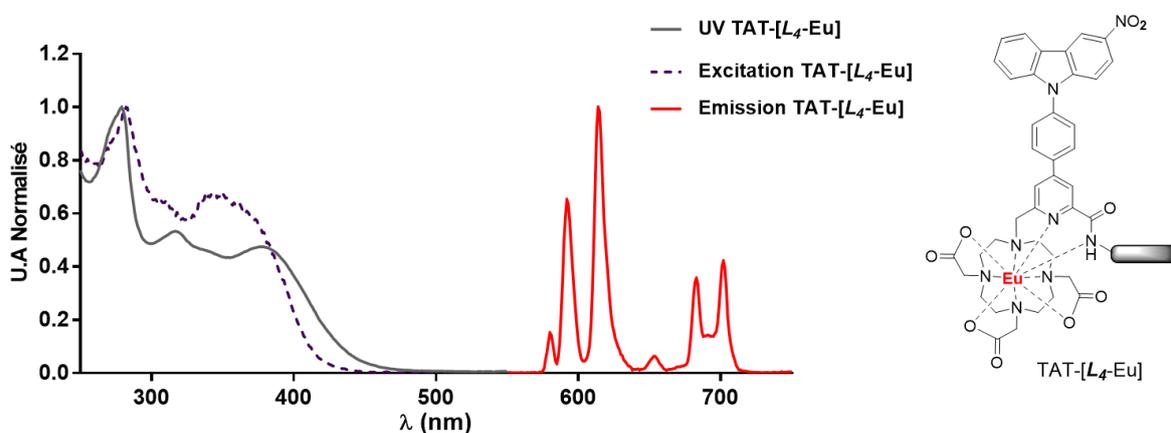


Figure 132 : Schéma du composé TAT- $[{}^9\text{-PbfL}_3\text{-Eu}]$ .

Ce complexe TAT- $[{}^9\text{-PbfL}_3\text{-Eu}]$  est beaucoup plus luminescent avec un rendement quantique de 4.7 % dans le tampon phosphate (10 mM, pH 7.5). Le même phénomène a été observé dans l'équipe par Olivier Sénèque avec un analogue comportant un alcyne entre le groupement carbazole-aryle et le picolinate. Le rendement quantique dans le tampon phosphate du composé sulfoné (avec Pbf en position 9) était de 11.8% alors que l'émission de l'Eu<sup>3+</sup> était indétectable pour le composé non sulfoné (sans Pbf). La réaction parasite étant relativement efficace, (TAT- $[L_3\text{-Eu}]$  et TAT- $[{}^9\text{-PbfL}_3\text{-Eu}]$  sont formés en proportions égales), nous avons décidé de voir si elle pouvait se produire sur le peptide F1, qui comporte aussi des arginines. Nous avons donc entrepris la synthèse du peptide *H*-F1- $[L_3\text{-Eu}]$ -SEA<sup>off</sup> dans l'espoir d'obtenir l'analogue avec le carbazole sulfoné. Le pro-ligand  $PL_3$  (2 éq) a donc été conjugué sur le fragment Ac-F1-SEA (1 éq) en utilisant du PyBOP (5 éq) et de la DIEA (10 éq) pendant 12 heures à température ambiante. La coupure et déprotection des esters *tert*-butylique par une solution de TFA nous a permis d'isoler et purifier le fragment Ac-F1- $L_3(OH)$ -SEA<sup>off</sup>. Nous n'avons pas observé le produit d'addition du groupement Pbf comme espéré. L'addition de Fmoc-Arg(Pbf)-OH dans le milieu de déprotection (TFA) n'a pas permis cette réaction non plus. Nous avons cependant poursuivi la synthèse et réalisé la métallation de ce fragment avec du chlorure d'euporium(III) à pH 7.5 durant 12 heures. La caractérisation du fragment Ac-F1- $[L_3\text{-Eu}]$ -SEA<sup>off</sup> a été effectuée, après purification sur HPLC. De par sa séquence d'acides aminés, le fragment F1 est peu soluble dans l'eau, et l'ajout du complexe  $[L_3\text{-Eu}]$  ne fait que l'accentuer. Nous avons tout de même réussi à réaliser les caractérisations spectroscopiques du fragment Ac-F1- $[L_3\text{-Eu}]$ -SEA<sup>off</sup> dans une solution de tampon HEPES (10 mM, pH 7.5). Les résultats que nous avons obtenus sont similaires à ceux du complexe TAT- $[L_3\text{-Eu}]$  mais le rendement quantique est un peu plus grand (2%), probablement à cause de la présence d'aa hydrophobes (Y, L, V, P) capable d'interagir avec le carbazole dans la séquence de F1. Bien que l'absorption du complexe soit décalée vers 340 nm la sensibilisation de l'euporium n'est pas suffisante pour qu'il soit utilisé dans notre système FRET. Le complexe  $[L_2\text{-Eu}]$  présentant une meilleure sensibilisation de l'euporium que le complexe  $[L_3\text{-Eu}]$ , nous avons abandonné celui-ci.

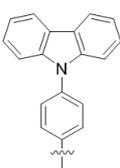
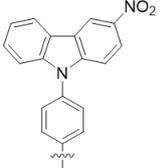
Nous avons vu qu'un groupement Pbf en position 9 défavorise le transfert électronique photo-induit au profit du transfert d'énergie qui peut s'expliquer de la façon suivante. Le groupement Pbf-SO<sub>2</sub>Ar est un groupement électro attracteur. Sa présence doit donc rendre le carbazole moins facilement oxydable en l'appauvrissant. Nous avons donc cherché à introduire un groupement électro-attracteur sur la position 9 du carbazole. Une recherche bibliographique nous a permis de voir qu'il était aisé de réaliser une nitration à cette position et nous avons synthétisé le composé **16** puis *PL*<sub>4</sub>, décrit plus haut. Celui-ci a été conjugué au peptide TAT puis métallé à l'euporium et les propriétés spectroscopiques du complexe TAT-[L<sub>4</sub>-Eu] ont ensuite été caractérisées. La fonctionnalisation du carbazole par un groupement nitro a permis de décaler le maximum d'absorption du complexe TAT-[L<sub>4</sub>-Eu] vers de plus hautes longueurs d'onde : le maximum est à 380 nm et la bande d'absorption s'étend jusqu'à 440 nm (**Figure 133** et **Tableau 4**). Cela conviendrait parfaitement pour notre système FRET puisque cette bande se situe dans la zone où la SiR<sup>Me</sup> n'absorbe pas. Le spectre d'émission de l'euporium, enregistré dans une solution aqueuse tamponnée par du tampon phosphate (10 mM, pH 7.0) montre les raies caractéristiques de l'euporium. Le spectre d'excitation a un maximum à 349 nm et ne s'étend que jusqu'à 418 nm. La bande d'absorption contient donc plusieurs composantes et celle de plus basse énergie ne permet pas la sensibilisation de l'Eu<sup>3+</sup>. Le temps de vie de l'euporium est de 0.705 ms, proche de celui mesuré pour [L<sub>3</sub>-Eu]. L'émission de l'euporium est très faible (il a fallu ouvrir largement les fentes du spectromètre pour l'observer). Nous n'avons donc pas jugé utile de mesurer son rendement quantique qui ne doit pas dépasser celui de TAT-[L<sub>3</sub>-Eu].



**Figure 133** : Spectres normalisés d'absorption (gris), d'émission (rouge) obtenus après excitation à  $\lambda_{ex} = 340$  nm et d'excitation (pointillé violet) obtenu après détection à  $\lambda_{em} = 615$  nm du complexe TAT-[L<sub>4</sub>-Eu] dans une solution aqueuse tamponnée par du tampon phosphate (10 mM, pH 7.0).

Au sein de cette série d'antennes dérivée de la fonction carbazole, le complexe TAT-[L<sub>3</sub>-Eu] semble peu intéressant. Malgré sa longueur d'onde d'excitation qui se situe dans la zone où la SiR<sup>Me</sup> n'absorbe pas, il possède un rendement quantique en solution aqueuse tamponnée faible, inférieur à celui de [L<sub>2</sub>-Eu]. Les mêmes constatations s'appliquent pour le complexe [L<sub>4</sub>-Eu].

**Tableau 4** : Caractéristiques spectroscopiques des complexes [L<sub>3</sub>-Eu] et [L<sub>4</sub>-Eu].

		
	L <sub>3</sub>	L <sub>4</sub>
$\lambda_{\max}$ UV (nm)	389	380
$\lambda_{\max}$ Excitation (nm)	349	349
Epsilon (M <sup>-1</sup> cm <sup>-1</sup> )	14 500	-
$\Phi$ dans H <sub>2</sub> O(%)	0,2 (TAT) 2(F1)	-
Temps de vie à $\lambda_{\max}$ Excitation	0,74	0,705

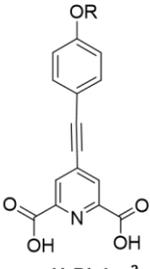
Pour notre projet, l'utilisation d'une unité carbazole comme groupement donneur n'est donc pas adaptée et nous nous sommes donc tournés vers une autre famille d'antenne dont le groupement donneur est dérivé de l'aniline. Néanmoins, ces complexes ont pu être utilisés dans l'équipe pour réaliser des images de cellules par microscopie biphotonique. Les antennes « push-pull » à base de carbazole ont d'excellentes propriétés d'absorption biphotonique, comme l'ont démontré Bettencourt-Dias et d'autres, qui compensent la faible émission du complexe. Je serai associé à la publication décrivant ces résultats. Le complexe [L<sub>3</sub>-Eu] a été conjugué par Ji-Hyung Choi à des peptides de pénétration cellulaire (TAT(voir ci-dessus) et ZF<sup>5.3</sup>) et testé pour la microscopie biphotonique de cellules vivantes. Malgré son rendement quantique très faible, des images en microscopie biphotonique *in cellulo* ont pu être obtenues grâce aux excellentes propriétés d'absorption biphotonique de ces antennes avec un groupement aryle-carbazole. Ces travaux feront bientôt l'objet d'une publication à laquelle je serai associé.

#### 4 - Antennes dérivées de l'aniline

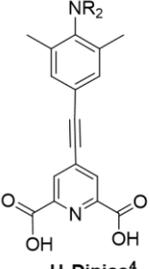
Dans l'élaboration d'une antenne « push-pull » la force du groupement donneur est un facteur important pour faire varier la longueur d'onde maximale d'absorbance de l'antenne. Plus le groupement donneur est fort, plus cette longueur d'onde va être décalée vers l'infrarouge. À titre d'exemple, le passage d'un groupement OR (R = alkyle) à un groupement N(R)<sub>2</sub> provoque un décalage de 50 nm. Le remplacement de la fonction éther par une fonction amine semble être une voie à examiner. Maury a décrit une série de dipicolinates fonctionnalisés en position 4 de l'azote pyridinique par divers groupements  $\pi$ -conjugués avec des donneurs oxygénés ou azotés et a étudié leurs propriétés de sensibilisation de l'euprium<sup>158,159</sup>. Ces dipicolinates sont présentés avec certaines de leurs propriétés spectroscopiques dans le **Tableau 5**. Ils forment des complexes 3:1 avec le lanthanide.

**Tableau 5** : Ligands Dicipico<sup>4-7</sup> développés par O. Maury et leurs caractéristiques spectroscopiques au sein d'un complexe Eu(Dipico<sup>x</sup>)<sub>3</sub> mesurées dans le DCM<sup>158,159</sup>

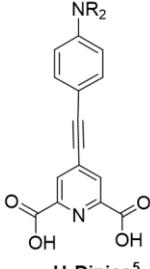
$$R = \text{Me-O-CH}_2\text{-CH}_2\text{-O-CH}_2\text{-CH}_2\text{-O-Me}$$



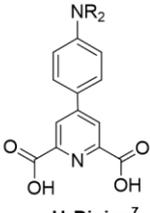
**H<sub>2</sub>Dipico<sup>2</sup>**



**H<sub>2</sub>Dipico<sup>4</sup>**



**H<sub>2</sub>Dipico<sup>5</sup>**



**H<sub>2</sub>Dipico<sup>7</sup>**

$\lambda_{\text{max abs}}$ (nm)	321	318	370	364
$\epsilon$ (M <sup>-1</sup> cm <sup>-1</sup> )	92 000	58 500	89 200	63 600
$\phi_{\text{Eu}}$ (%)	15 (DCM)	27 (DCM)	7 (DCM)	28(DCM)

L'absorption est dominée par une transition ILCT. Dans le cas de H<sub>2</sub>Dipico<sup>2</sup>, le maximum d'absorption est à 321 nm et pour le H<sub>2</sub>Dipico<sup>5</sup> le maximum est à 370 nm. L'épsilon est près de 90 000 M<sup>-1</sup>cm<sup>-1</sup> pour les deux complexes. Cependant [Eu(Dipico<sup>5</sup>)<sub>3</sub>] présente un rendement quantique plus faible que [Eu(Dipico<sup>2</sup>)<sub>3</sub>] (7% vs 15%) et un temps de vie plus faible (0.85 ms vs 1.90 ms). Ceci peut indiquer qu'un transfert d'énergie retour de l'Eu<sup>3+</sup> vers l'antenne est plus important pour Dipico<sup>5</sup> : le transfert de charge doit être trop bas en énergie. Des petites variations dans la structure des antennes permettent cependant de récupérer des propriétés d'émission intéressantes. Comme nous l'avons vu, il est impératif pour nous d'éviter les fonctions alcynes. La suppression de la fonction alcyne du Dipico<sup>5</sup> conduit à l'antenne de type biaryle H<sub>2</sub>Dipico<sup>7</sup> et permet de retrouver des propriétés spectroscopiques intéressantes, proches de celles avec le donneur oxygéné. Les mesures des propriétés spectroscopiques réalisées sur le complexe [Eu(Dicipico<sup>7</sup>)<sub>3</sub>] montrent un rendement quantique de 28 % avec un temps de vie de l'euporium mesuré à 1.75 ms, mais la longueur d'onde maximale d'absorbance reste décalée vers le rouge par rapport au donneur oxygéné, à 364 nm. Le coefficient d'absorption molaire de ce complexe à cette longueur d'onde a été déterminé autour de 64 000 M<sup>-1</sup>cm<sup>-1</sup> qui permet de l'estimer à environ 21 000 M<sup>-1</sup>cm<sup>-1</sup> par antenne. L'antenne dialkylamino-aryle-picolinate pourrait donc respecter notre cahier des charges pour le complexe d'euporium. Nous proposons donc l'utilisation du ligand **L<sub>5</sub>** pour notre projet qui nécessite la synthèse du pro-ligand **PL<sub>5</sub>** (**Figure 134**). Il reste à évaluer cependant l'effet d'un passage du dipicolinate au monopicolinate, qui dans le cas des donneurs oxygénés provoque un déplacement vers le rouge du transfert de charge de 10 à 30 nm.

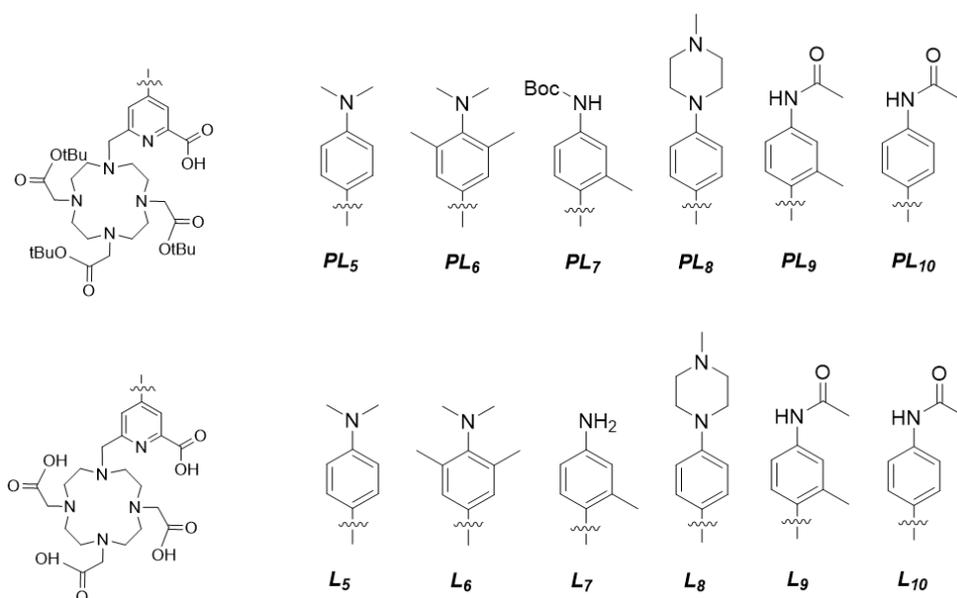
Lorsque deux groupements méthyles sont ajoutés en ortho du groupement dialkylamine (ligand H<sub>2</sub>Dipico<sup>4</sup>), on retrouve un haut rendement quantique, passant de 7 à 27 % et un temps de vie de 1.81 ms<sup>159</sup>. Il a été démontré que les deux méthyles du ligand H<sub>2</sub>Dipico<sup>4</sup> induisent un encombrement

stérique qui contraint le doublet non liant de l'azote à être quasi perpendiculaire au système  $\pi$ -conjugué (angle de  $89^\circ$  calculé par DFT). La conséquence est une moins bonne conjugaison entre le groupement donneur (dialkylamine) et l'aryle résultant en une transition de transfert de charge de plus haute énergie : le maximum d'absorption passe de 370 nm à 320 nm, valeur proche de celle de H<sub>2</sub>Dipico<sup>2</sup>. Il sera intéressant d'étudier cette modification si l'antenne de type biaryle avec donneur dialkylamino ne donnait pas satisfaction pour la sensibilisation de l'euporium. Ceci correspond au ligand **L<sub>6</sub>** et au pro-ligand **PL<sub>6</sub>** (**Figure 134**).

Si les résultats avec **L<sub>5</sub>** n'étaient pas satisfaisants, à cause d'un transfert de charge trop bas en énergie, nous pourrions le remonter en passant d'un groupement électrodonneur NR<sub>2</sub> (R = alkyle) à NH<sub>2</sub>. Un déplacement vers le bleu d'une trentaine de nm peut être espéré avec cette modification<sup>160</sup>. Ceci correspond au ligand **L<sub>7</sub>** et au pro-ligand **PL<sub>7</sub>** (**Figure 134**).

Un autre moyen de déplacer le transfert de charge vers le bleu est d'utiliser un groupement électrodonneur de type pipérazine. L'effet est d'autant plus marqué si la pipérazine est protonnée ou

N-alkylée (ammonium)<sup>160,161</sup>. Dans ce cas nous pouvons proposer d'évaluer le ligand **L<sub>8</sub>** (pro-ligand **PL<sub>8</sub>**, **Figure 134**). Enfin si les anilines devaient ne pas convenir, il a été envisagé d'utiliser un groupement acétamide comme électrodonneur. Du fait de la délocalisation du doublet électronique de l'azote avec le carbonyle de l'amide, le groupement acétamide est un moins bon groupement électrodonneur que NR<sub>2</sub> (R = alkyle ou H). Nous avons retenu deux ligands, **L<sub>9</sub>** et **L<sub>10</sub>**, le premier possédant un méthyle sur l'aryle, comme **L<sub>2</sub>**, et l'autre sans ce méthyle (**Figure 134**). Ce méthyle empêche le biaryle d'être plan et ceci a des conséquences sur le transfert de charge<sup>160,162</sup>.



**Figure 134** : Schéma des différents pro-ligands (**PL<sub>i</sub>**) et ligand (**L<sub>i</sub>**) à synthétiser.

4.1 - Synthèse des pro-ligands *PL*<sub>5</sub> à *PL*<sub>10</sub>

La synthèse des pro-ligands *PL*<sub>5</sub> à *PL*<sub>10</sub> suit le schéma décrit pour *PL*<sub>2</sub> avec le couplage d'un ester arylboronique sur le macrocycle iodé **11**. L'ester boronique est obtenu à partir d'un halogénure d'aryle bromé, ou iodé. Pour *PL*<sub>5</sub>, *PL*<sub>8</sub> et *PL*<sub>10</sub>, ceux-ci sont commercialisés. Pour les autres, la synthèse des bromures d'aryles correspondants doit être effectuée. Le composé bromé **20**, précurseur de *PL*<sub>6</sub>, a été obtenu après une première étape d'alkylation en présence d'iodométhane et de carbonate de potassium avec un rendement de 85%<sup>163</sup>. Pour *PL*<sub>7</sub>, nous avons opté pour une protection de l'aniline par un groupement Boc, que l'on peut enlever au TFA en même temps que les *t*Bu des bras acétate du macrocycle. La protection de la 4-bromo-3-méthylaniline par le groupement Boc a été réalisée avec l'anhydride Boc dans l'eau pour obtenir le composé bromé **24** avec un rendement de 99%<sup>164</sup>. La formation du composé bromé **28**, précurseur de *PL*<sub>9</sub>, a été réalisée dans l'acétate d'éthyle en présence d'anhydride acétique, de DMAP et de TEA<sup>165</sup> pour un rendement de 84% (**Figure 135**).

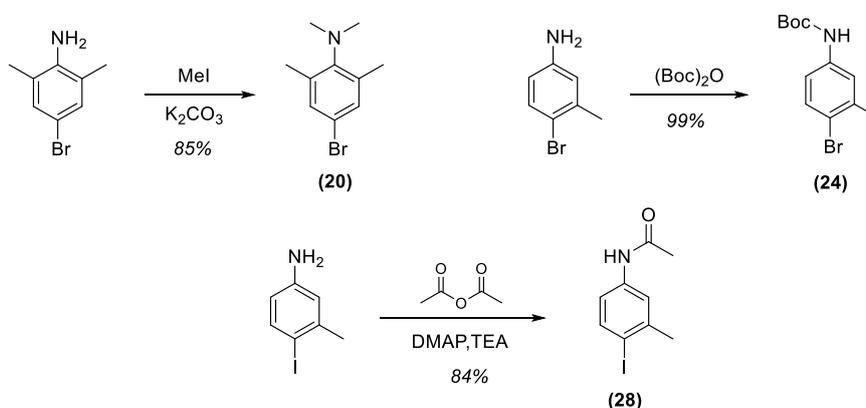
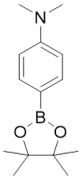
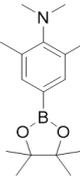
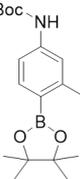
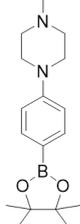
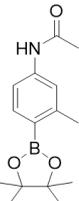
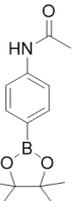


Figure 135 : Schéma de synthèse des composés **20**, **24** et **28**.

La formation de l'ester boronique sur la 4-bromo-*N,N*-diméthylaniline est réalisée en présence de 4,4,5,5-tétraméthyl-1,3,2-dioxa-borolane (3 éq) associé avec du chlorure de *bis*(triphénylphosphine) palladium(II) en quantité catalytique. La base utilisée est de la triéthylamine (6 éq) et l'ensemble des réactifs sont dissouts dans le solvant (dioxane, DMF ou DMSO), le tout sous atmosphère inerte. La réaction terminée, le milieu réactionnel est extrait et purifié sur colonne de gel de silice. Ces conditions ont été appliquées pour la synthèse de quelques-uns des autres esters boroniques, mais les faibles rendements que nous avons observés nous ont poussé à utiliser d'autres conditions. Les esters boroniques **18**, **21**, **24**, **26**, **29** et **31** ont pu être obtenus avec de bons rendements à partir de *bis*(pinacolato)diboron et de l'acétate de sodium dans le DMF en utilisant le [1,1'-*bis*(diphénylphosphino)ferrocène] palladium (II) comme catalyseur (**Tableau 6**).

Tableau 6 : Conditions et rendements de synthèse des composés **18**, **21**, **24**, **26**, **29** et **31**.

						
	<b>18</b>	<b>21</b>	<b>24</b>	<b>26</b>	<b>29</b>	<b>31</b>
<b>Pd(PPh<sub>3</sub>)<sub>2</sub>Cl<sub>2</sub></b>	x					
<b>Pd(dppf)Cl<sub>2</sub></b>		x	x	x	x	x
<b>4,4,5,5-tetramethyl-1,3,2-dioxaborolane</b>	x					
<b>Bis(pinacolato)diboron</b>		x	x	x	x	x
<b>Rendement</b>	<b>77 %</b>	<b>79 %</b>	<b>41 %</b>	<b>82 %</b>	<b>91 %</b>	<b>78 %</b>

Les dernières étapes de synthèses des pro-ligands sont le couplage de Suzuki-Miyaura entre l'ester boronique et le macrocycle iodé **11**, et la déprotection de l'ester méthylique porté par le picolinate. Pour le couplage de Suzuki-Miyaura nous avons testé les conditions suivantes ; nous avons utilisé un léger excès d'ester boronique (1.2 éq) par rapport à **11**, en présence de fluorure de césium (4 éq) et de palladium (0), ici le *tetrakis*(triphenylphosphine) palladium(0), dans le DMF. Après 20h d'agitation à 90°C sous argon, l'analyse par LCMS du brut de réaction a montré systématiquement la présence d'oxyde de triphénylphosphine souvent difficile à séparer du produit. Si l'on réalise l'hydrolyse de l'ester méthylique, la séparation ne s'améliore pas forcément. Nous avons alors décidé d'utiliser un catalyseur supporté sur microbille de polymère. À la fin de la réaction, une filtration sur célite est réalisée avant le traitement du milieu réactionnel. Le produit brut obtenu est repris dans l'éthanol et une solution concentrée de soude est ajoutée lentement pour obtenir une concentration finale de 2 M. La réaction est laissée sous agitation 10 min avant que le pH ne soit remonté autour de pH 7. Le précipité formé est séparé de la solution par centrifugation et le surnageant est extrait par un mélange AcOEt/H<sub>2</sub>O. L'analyse par LCMS du produit brut obtenu après évaporation de la phase organique montre que les sous-produits observés précédemment ne sont plus présents et que le composé est suffisamment pur pour être couplé sur un peptide sans purification supplémentaire. Une partie du produit brut a cependant été purifiée afin d'obtenir les caractérisations des pro-ligands **PL<sub>5</sub>** à **PL<sub>10</sub>** par spectrométrie de masse et par résonance magnétique nucléaire. L'ensemble des pro-ligands synthétisés est présenté **Figure 136**.

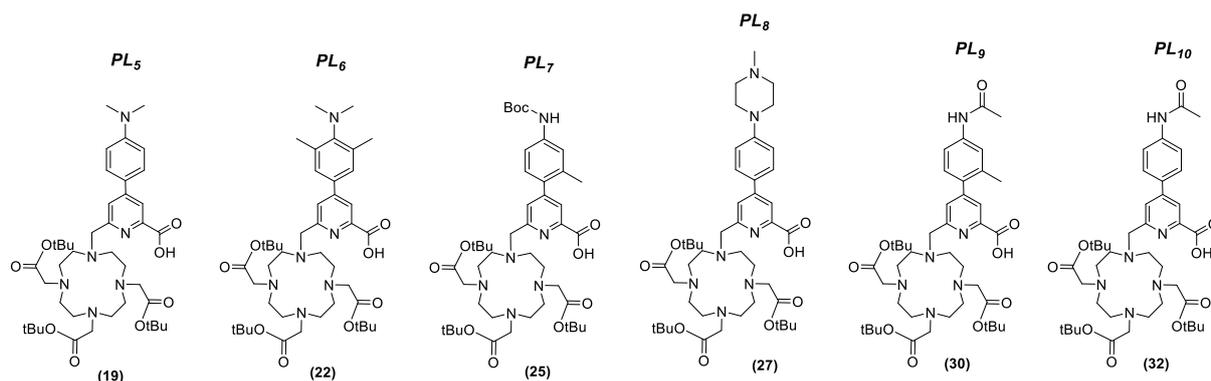


Figure 136 : Structures des pro-ligands  $PL_5$  à  $PL_{10}$  synthétisés.

#### 4.2 - Synthèse et caractérisation spectroscopique des complexes $[L_i-Eu]$ ( $i=5-10$ )

Pour des questions d'efficacité, les caractérisations spectroscopiques des complexes  $[L_{5-8}-Eu]$  ont été effectuées après avoir formé le complexe de lanthanide libre, sans être conjugué au peptide TAT. En effet contrairement à ceux avec les antennes carbazole, il n'y a pas de problème de solubilité dans l'eau, la conjugaison avec TAT perd de son intérêt. Pour cela, la déprotection des esters *tert*-butyliques du macrocycle est réalisée dans le TFA durant 30 min. Après évaporation du solvant, le solide obtenu est repris dans une solution aqueuse afin de réaliser la complexation avec un sel d'euprium ( $EuCl_3$ , 1.2 éq) durant 12h à pH 7.5. Le précipité est ensuite lavé avec de l'acétone et de l'éther pour obtenir le complexe isolé. Les spectres d'absorption, d'émission et d'excitation obtenus pour les complexes  $[L_5-Eu]$ ,  $[L_6-Eu]$ ,  $[L_7-Eu]$  et  $[L_8-Eu]$  sont présentés sur la **Figure 137**. L'ensemble des caractéristiques spectroscopiques est rassemblé **Tableau 7**.

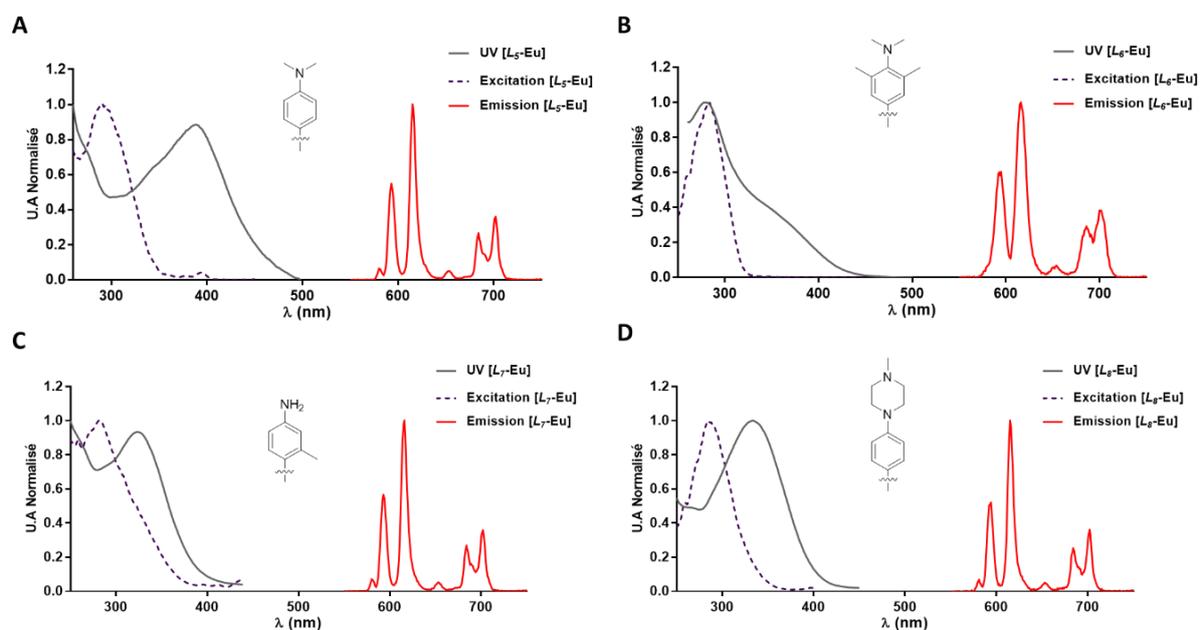


Figure 137 : Spectres normalisés d'absorption (gris), d'émission (rouge) et d'excitation (pointillés violets) obtenus après détection à  $\lambda_{em} = 615$  nm des complexes  $[L_i-Eu]$  dans une solution aqueuse tamponnée par du tampon phosphate (10 mM, pH 7.0). A) Excitation à  $\lambda_{ex} = 290$  nm. B) Excitation à  $\lambda_{ex} = 320$  nm. C) Excitation à  $\lambda_{ex} = 350$  nm. D) Excitation à  $\lambda_{ex} = 333$  nm.

Le complexe [ $L_5$ -Eu] présente une bande de transfert de charge centrée à 390 nm. En comparaison du complexe [Eu(Dipico<sup>5</sup>)<sub>3</sub>] présenté précédemment, on observe un décalage de l'absorbance de 20 nm vers le rouge, en accord avec le passage d'un ligand dipicolinate à un monopicolinate. Le spectre d'émission de l'euporium a pu être enregistré après excitation à cette longueur d'onde, mais l'intensité du signal obtenu semble très faible. Le spectre d'excitation, obtenu à partir du maximum d'émission de l'euporium à 615 nm, ne correspondant pas au spectre d'absorbance du complexe. En effet, il présente un maximum à 280 nm au lieu de 390 nm. Cette bande à 280 nm peut être attribuée à une transition vers un état localement excité (LE) de l'antenne. Il semble donc que l'excitation à travers la bande ILCT ne soit pas efficace. Le rendement quantique du complexe est trop faible pour être mesuré. Il serait tentant de conclure que le passage du dipicolinate (complexe [Eu(Dipico<sup>5</sup>)<sub>3</sub>]) au monopicolinate ([ $L_5$ -Eu]) impacte dramatiquement les propriétés de luminescence, puisque le complexe de Maury à un fort rendement quantique (28%) alors que [ $L_5$ -Eu] n'émet presque pas. Il faut cependant être prudent, car les mesures sur [Eu(Dipico<sup>5</sup>)<sub>3</sub>] ont été faites dans le DCM alors que nous travaillons dans l'eau. La polarité du milieu peut avoir un effet dramatique sur le transfert de charge. En comparaison du DCM, l'eau, plus polaire, peut favoriser le transfert d'électron photo-induit avec ce type d'antennes « push-pull ». Une des hypothèses avancées pour expliquer la faible émission de l'Eu<sup>3+</sup> est une réduction de l'ion Eu(III) en ion Eu(II) par transfert d'électron photo-induit depuis l'antenne excitée. Il apparaît donc que [ $L_5$ -Eu] n'est pas utilisable pour notre projet. Il pourrait cependant être intéressant pour sensibiliser l'Yb<sup>3+</sup>, ce qu'un stagiaire de M2 va étudier prochainement au laboratoire.

L'ajout des deux méthyles en position *ortho* de l'amine doit permettre de décaler vers le bleu la bande ILCT par rapport à [ $L_5$ -Eu] en affaiblissant la conjugaison électronique du groupement électrodonneur. Le maximum d'absorption pour le complexe [ $L_6$ -Eu] est à 280 nm avec un épaulement vers 350 nm, qui doit correspondre à la bande ILCT, comme attendu décalée vers le bleu, à 350 nm. L'émission de l'euporium est détectable, mais toujours très faible (le rendement quantique n'a pas été mesuré). Encore une fois, le spectre d'excitation montre que la transition ILCT ne contribue pas à la sensibilisation de l'Eu<sup>3+</sup>. Le complexe [ $L_8$ -Eu] (donneur = pipérazine) se comporte comme [ $L_5$ -Eu] avec une transition ILCT nettement déplacée vers le bleu ( $\lambda_{\max} = 334$  nm), mais qui ne sensibilise toujours pas l'Eu<sup>3+</sup> puisque le spectre d'excitation fait apparaître essentiellement la transition LE à 280 nm.

Enfin, pour le complexe [ $L_7$ -Eu] (donneur NH<sub>2</sub>), le spectre d'absorption est dominé par une bande ILCT à 325 nm alors que le spectre d'excitation est dominé par la bande LE à 280 nm, avec un ILCT qui apparaît comme un épaulement. La sensibilisation de la luminescence de l'Eu<sup>3+</sup> en excitant dans la bande ILCT reste très faible.

La mesure du temps de vie des complexes [ $L_5$ -Eu] à [ $L_8$ -Eu] en solution ont tous un temps de vie proche de 1.05 ms. Ce temps de vie long de l'euporium nous indique que le transfert (d'énergie ou d'électron) retour du lanthanide vers l'antenne n'a pas lieu. Néanmoins, pour tous ces complexes, le rendement

quantique est très faible. Comme pour  $[L_5\text{-Eu}]$ , nous pouvons proposer qu'un transfert d'électron photoinduit de l'antenne vers l' $\text{Eu}^{3+}$  réduit l'euporium, empêchant son émission.

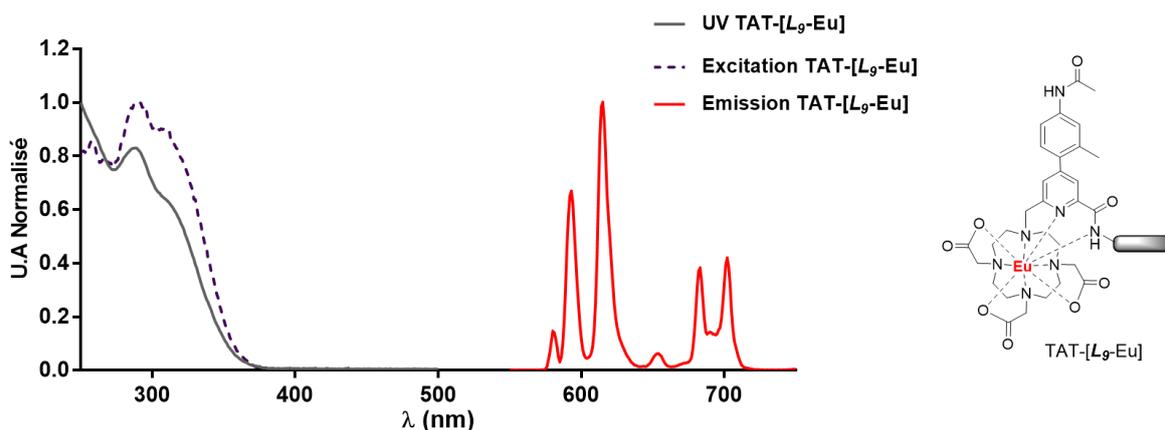
**Tableau 7** : Caractéristiques spectroscopiques des ligands  $L_5$  à  $L_8$ .

	$L_5$	$L_6$	$L_7$	$L_8$
$\lambda_{\text{max UV}}$ (nm)	389	350	325	334
Epsilon ( $\text{M}^{-1}\text{cm}^{-1}$ )	-	-	10 200	23 840
$\lambda_{\text{max Excitation}}$ (nm)	290	283	283	287
Temps de vie à 280 nm (ms)	1,06	1,10	1,05	1,01
Temps de vie à $\lambda_{\text{max}}$ (ms)	1,06	1,04	1,01	1,01

Les donneurs anilines étant facilement oxydable, nous avons alors choisi d'étudier les complexes  $[L_9\text{-Eu}]$  et  $[L_{10}\text{-Eu}]$  dont le groupement acétamide est un moins bon donneur, ce qui doit défavoriser le transfert d'électrons et la réduction du lanthanide.

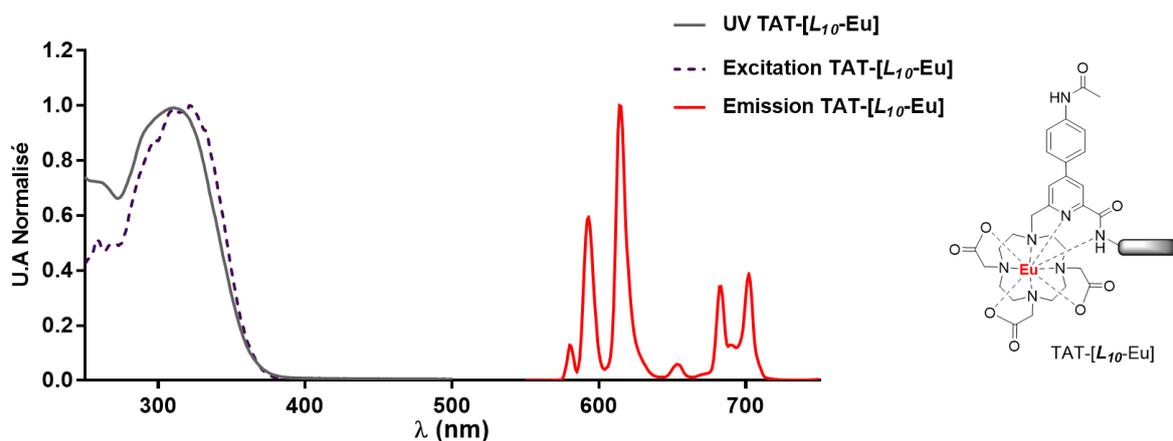
Un test rapide de synthèse du complexe  $[L_{10}\text{-Eu}]$  nous a montré qu'il émettait fortement. Contrairement aux précédents complexes, il a été choisi de procéder à la caractérisation de  $[L_9\text{-Eu}]$  et  $[L_{10}\text{-Eu}]$  conjugué sur peptide TAT, en vue d'une application en imagerie cellulaire pour un autre projet en cours dans l'équipe. Les conjugués TAT- $[L_9\text{-Eu}]$  et TAT- $[L_{10}\text{-Eu}]$  ont été préparés comme décrit ci-dessus pour TAT- $[L_2\text{-Eu}]$  (voir partie expérimentale). L'ensemble des caractérisations spectroscopiques a été réalisée dans une solution aqueuse tamponnée par du phosphate (10 mM, pH 7.0). Le spectre d'absorbance de TAT- $[L_9\text{-Eu}]$  présente une transition LE à 285 nm et une transition ILCT à 315 nm sous la forme d'un épaulement. Le spectre d'émission obtenu en excitant à 315 nm présente les transitions  ${}^5D_4 \rightarrow {}^7F_J$  ( $J = 0-4$ ) de l'euporium (**Figure 138**). Le temps de vie de luminescence de l' $\text{Eu}^{3+}$  (1.05 ms) est classique pour ce type de complexe. Cette fois, le spectre d'excitation coïncide parfaitement avec le spectre d'absorption, indiquant que l'on peut sensibiliser l' $\text{Eu}^{3+}$  par la transition ILCT qui s'étend jusqu'à 350 nm. Le coefficient d'absorption molaire ainsi que le rendement quantique ont été déterminés :  $\epsilon_{315} L_9 = 14800 \text{ M}^{-1}\text{cm}^{-1}$  et  $\Phi_{\text{Eu}} L_9 = 5\%$ .

Les caractéristiques spectroscopiques de ce complexe sont similaires avec celles du complexe TAT- $[L_2\text{-Eu}]$  dont le rendement quantique est de 4% et l'épsilon à 315 nm vaut  $11\,800 \text{ M}^{-1}\text{cm}^{-1}$ . Il possède cependant une brillance plus élevée,  $B_{315}([L_9\text{-Eu}]) = 740 \text{ M}^{-1}\text{cm}^{-1}$  et  $B_{315}([L_2\text{-Eu}]) = 472 \text{ M}^{-1}\text{cm}^{-1}$ .



**Figure 138** : Spectres normalisés d'absorption (gris), d'émission (rouge) obtenus après excitation à  $\lambda_{ex} = 315$  nm et d'excitation (pointillé violet) obtenu après détection à  $\lambda_{em} = 615$  nm du complexe TAT-[L<sub>9</sub>-Eu] dans une solution aqueuse tamponnée par du tampon phosphate (10 mM, pH 7.0).

Sur la **Figure 139** sont présentés les spectres d'absorbance, d'émission et d'excitation du complexe TAT-[L<sub>10</sub>-Eu]. Le spectre d'absorption présente une bande large avec un maximum à 315 nm, sur lequel on peut voir un épaulement vers 280 nm. Il semble donc y avoir, comme pour L<sub>9</sub> une transition LE vers 280 nm et une transition ILCT à 315 nm, mais cette dernière domine l'absorption cette fois. Le coefficient d'absorption molaire à 315 nm est de 20 500 M<sup>-1</sup>cm<sup>-1</sup>, ce qui est sensiblement supérieur à celui de TAT-[L<sub>9</sub>-Eu]. En excitant à 315 nm, une émission intense d'euporium est détectée avec un temps de vie de 1.05 ms, identique à celui de TAT-[L<sub>9</sub>-Eu]. Le spectre d'excitation montre que la sensibilisation de l'euporium provient essentiellement de la transition ILCT avec un maximum à 315 nm. Enfin le rendement quantique du complexe a été mesuré à 11%, plus du double de celui de TAT-[L<sub>9</sub>-Eu]. En conséquence de tout ceci, sa brillance à 315 nm, de 2250 M<sup>-1</sup>cm<sup>-1</sup>, est trois fois plus élevée que celle de TAT-[L<sub>9</sub>-Eu]. Le groupe de Maury a étudié en détail les complexes de Tb<sup>3+</sup> et Eu<sup>3+</sup> possédant des antennes 4-(p-methoxy)phényl-picolinate. L'insertion d'un groupement méthyle sur le phényle en ortho du picolinate bloque la rotation entre les deux cycles aromatiques. Ce blocage permet d'avoir un complexe de terbium plus luminescent qu'en l'absence de ce méthyle pour laquelle la rotation entre les deux aryles est possible ( $\Phi_{Tb} L_9 = 74\%$  et  $15\%$  avec et sans méthyle<sup>84,166</sup>). Cela a été observé aussi avec les complexes TAT-[L<sub>9</sub>-Tb] et TAT-[L<sub>10</sub>-Tb], non décrit dans cette thèse :  $\Phi_{Tb} L_{10} = 22\%$  avec le méthyle et  $2\%$  sans le méthyle). Pour les complexes d'euporium, il a été observé dans le groupe de Maury qu'avec les substituants méthoxy les rendements quantiques sont identiques avec ou sans méthyle. Ce n'est pas le cas avec le substituant acétamide, pour lequel l'Eu<sup>3+</sup> émet plus sans le méthyle qu'avec le méthyle.



**Figure 139** : Spectres normalisés d'absorption (gris), d'émission (rouge) obtenus après excitation à  $\lambda_{ex} = 315$  nm et d'excitation (pointillé violet) obtenu après détection à  $\lambda_{em} = 615$  nm du complexe TAT[L<sub>10</sub>-Eu] dans une solution aqueuse tamponnée par du tampon phosphate (10 mM, pH 7.0).

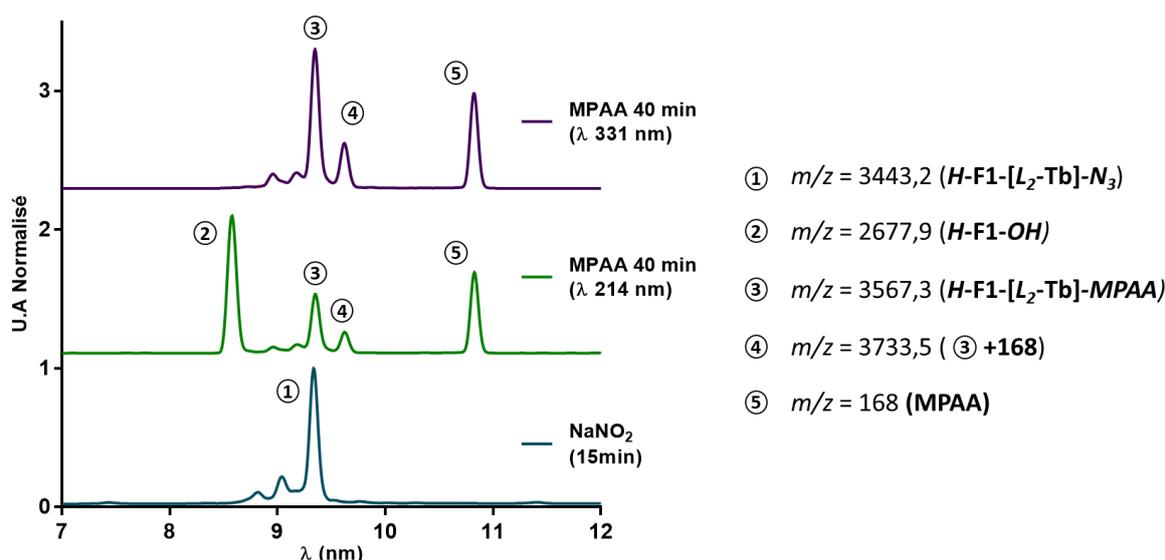
Le complexe TAT-[L<sub>10</sub>-Eu] présente une excellente sensibilisation de l'euprimum avec un rendement quantique dans l'eau de 11%. Par rapport à TAT-[L<sub>2</sub>-Eu], qui était notre complexe avec une antenne de type biaryle et un groupement électrodonneur méthoxy, nous avons décalé le maximum d'absorption de la bande ILCT de 10 nm environ. Cependant nous n'avons pas encore atteint les performances du complexe d'euprimum avec l'antenne *p*-méthoxyphényle-alcyne-picolinate, [L<sub>I</sub><sup>CC</sup>-Eu] dont la longueur d'onde d'excitation (331 nm), l'absorption ( $\epsilon_{331} = 28\,000$  M<sup>-1</sup>cm<sup>-1</sup>), le rendement quantique (17 %) et la brillance (4750 M<sup>-1</sup>cm<sup>-1</sup>) sont supérieurs. C'est cependant notre meilleur complexe compatible chimiquement avec les protéines contenant des cystéines et nous avons décidé de l'utiliser dans le système FRET, tout comme [L<sub>2</sub>-Eu] que nous avons eu plus tôt dans la thèse et qui nous a permis de démarrer la synthèse et l'étude des protéines fonctionnalisées.

Notons que les complexes TAT-[L<sub>9</sub>-Tb] et TAT-[L<sub>10</sub>-Eu] ont été testés récemment pour une application en microscopie biphotonique de cellules vivantes et ont donné d'excellents résultats qui seront publiés prochainement. Je serai associé à ces publications.

## V - Synthèse des protéines fonctionnalisées par les complexes de lanthanide

Nous avons à disposition deux méthodes de ligation pour réaliser l'assemblage du fragment F1 fonctionnalisé avec la partie reconnaissance d'ADN (fragment F2-F3). La ligation par la méthode des hydrazides, qui est bien plus rapide et efficace que la ligation SEA, n'est cependant pas compatible avec le fragment F1 fonctionnalisé par la SiR<sup>Me</sup>. Nous devons tout de même tester cette méthode avec un fragment F1 sur lequel est conjugué un complexe de lanthanide. Nous avons commencé par synthétiser le fragment *H*-F1-NHNH<sub>2</sub> fonctionnalisé par le complexe de terbium, [L<sub>2</sub>-Tb]. Le couplage sur résine a été réalisé avec un léger excès du pro-ligand PL<sub>2</sub> (1.1 éq) en présence de DIEA (4 éq) et de PyBOP (2 éq) dans le DMF. Après 12 heures de réaction, le peptide est décroché de la résine et les esters

*tert*-butyliques sont déprotégés à l'aide d'une solution de TFA en présence de scavengers : TIS et eau (93.4/3.3/3.3). Après purification du peptide, la métallation du complexe a été réalisée par un sel de terbium(III) en excès à pH 7.5 durant 12 heures. La protéine *H*-F1-[L<sub>2</sub>-Tb]-NHNH<sub>2</sub> a pu être isolée par HPLC. La première étape de ligation de ce fragment avec le fragment *H*-F2-F3-NH<sub>2</sub> est l'oxydation de l'hydrazide par NaNO<sub>2</sub> à -15°C durant 10 min dans une solution de chlorure de guanidinium (6 M) dans du tampon phosphate (100 mM, pH 7.0) dont le pH est ajusté à 3. Après ajout de 40 éq de MPAA, le suivi par LCMS après 40 minutes de réaction révèle l'apparition d'un pic majoritaire à 8.6 minutes (pic(2) sur la **Figure 140**) dont la masse ne correspond pas à la formation de l'ester activé de MPAA. Sa masse ( $m/z = 2677.9$ ) est inférieure au produit de départ *H*-F1-[L<sub>2</sub>-Tb]-NHNH<sub>2</sub> et correspond vraisemblablement au fragment *H*-F1-OH où le complexe n'est plus conjugué. L'absence d'absorbance sur le spectre à 331 nm confirme la perte du complexe de terbium. La ligation selon la méthode hydrazine ne semble pas compatible avec l'utilisation de protéines fonctionnalisées. Nous sommes donc retournés vers l'utilisation de la méthode de ligation SEA.



**Figure 140** : Suivi LCMS de la ligation entre les fragments *H*-F1-[L<sub>2</sub>-Tb]-NHNH<sub>2</sub> et *H*-F2-F3-NH<sub>2</sub>.

Le couplage des pro-ligands *PL*<sub>2</sub> et *PL*<sub>10</sub> sur le fragment F1-SEA a été réalisé sur support solide après la déprotection sélective du groupement Alloc de la lysine 106 (**Figure 141**). Un léger excès du pro-ligand (1.1 éq) est utilisé en présence de DIEA (4 éq) et de PyBOP (2 éq) dans le DMF. Après 12 heures de réaction, le peptide est décroché de la résine et les esters *tert*-butyliques sont déprotégés à l'aide d'une solution de TFA en présence de scavengers : TIS et eau (93.4/3.3/3.3). Après purification du peptide, la complexation avec un sel de terbium/europium (LnCl<sub>3</sub>, 1.2 éq) est réalisée à pH 7.5 durant 12h. Une dernière purification par HPLC a permis d'isoler les peptides *H*-F1-[L<sub>2</sub>-Tb]-SEA<sup>off</sup>, *H*-F1-[L<sub>2</sub>-Eu]-SEA<sup>off</sup> et *H*-F1-[L<sub>10</sub>-Eu]-SEA<sup>off</sup>.

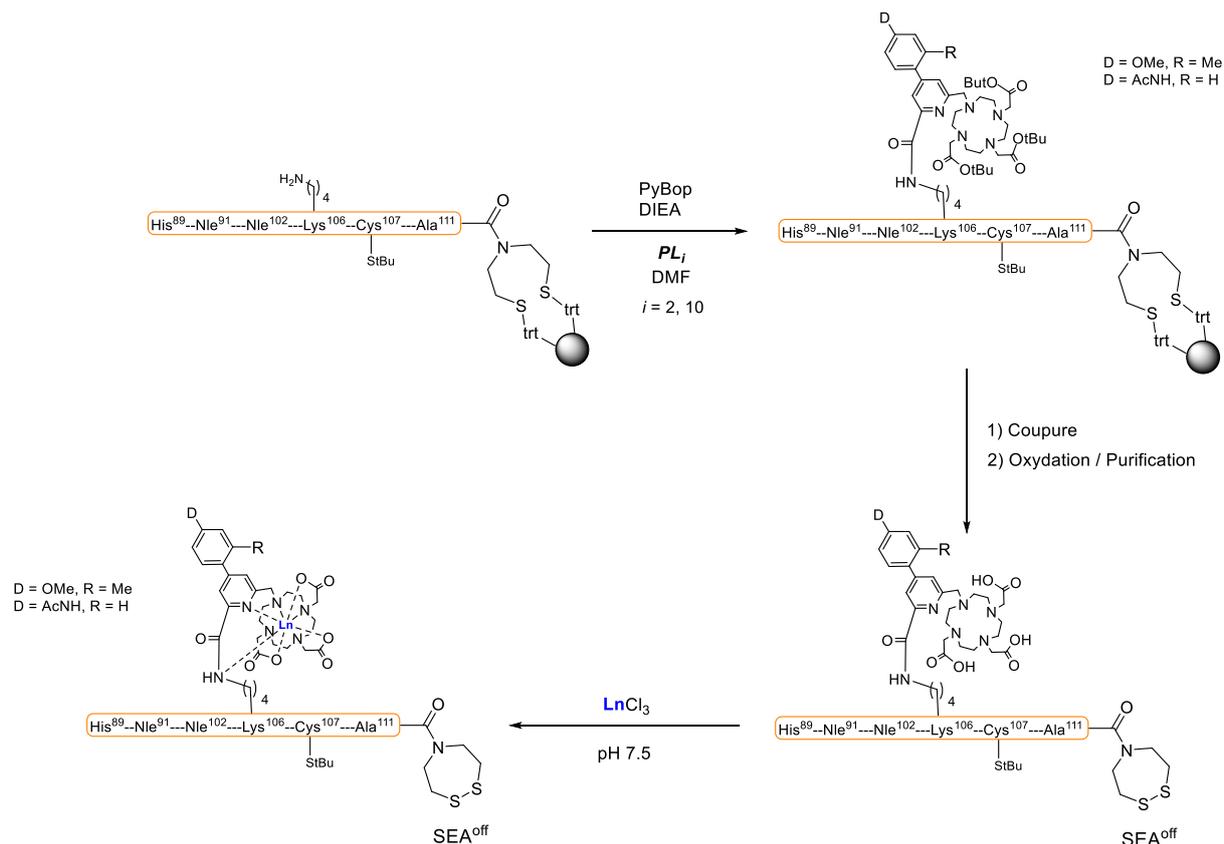


Figure 141 : Schéma de greffage des complexes  $[L_2-Tb]$ ,  $[L_2-Eu]$  et  $[L_{10}-Eu]$  sur le fragment H-F1-SEA.

Nous avons réalisé la ligation des fragments F1 avec F2-F3 selon la méthode SEA. Elle a été réalisée par dissolution des protéines  $H-F1-[L_2-Tb]-SEA^{off}$  (1 éq) et  $H-F2-F3-NH_2$  (1.5 éq) dans une solution contenant du chlorure de guanidinium (6 M), du MPAA (0.1 M) et du TCEP (0.1 M) dans du tampon phosphate (100 mM, pH 7.0) dont le pH a été ajusté à 6.5. Le suivi par LCMS après 24h de réaction est présenté Figure 142.

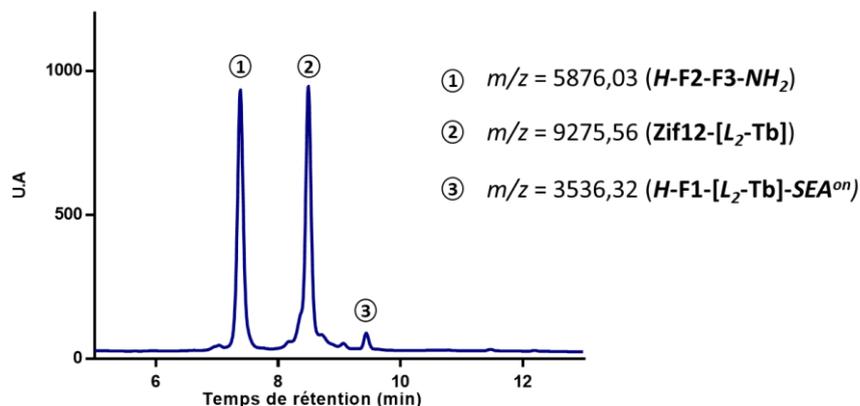
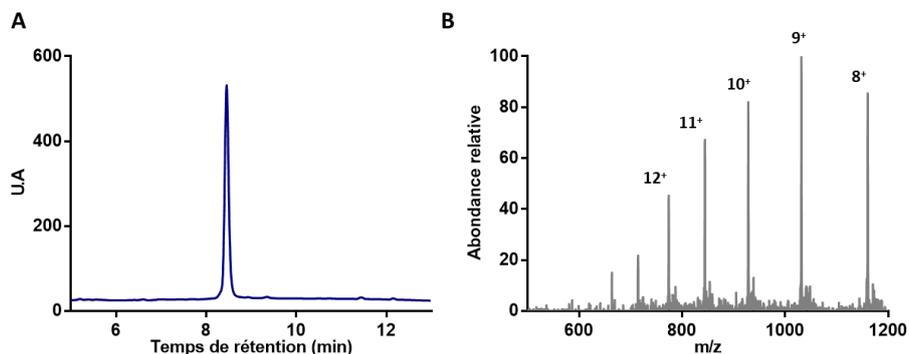


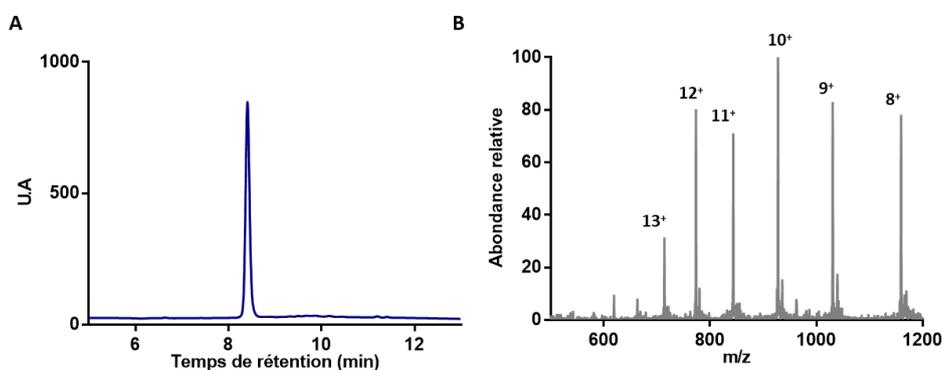
Figure 142 : Suivi LCMS de la réaction de ligation à 214 nm entre  $H-F1-[L_2-Tb]-SEA^{off}$  et  $H-F2-F3-NH_2$  après 24h de réaction.

Le milieu réactionnel est traité et purifié sur colonne HPLC pour obtenir la protéine isolée. En suivant le même protocole pour la ligation chimique native, les protéines  $Zif12-[L_2-Eu]-A$  et  $Zif12-[L_{10}-Eu]-A$

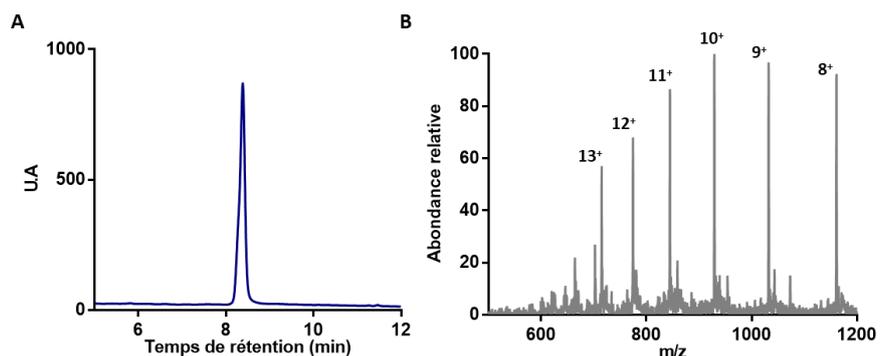
ont été synthétisées. Les protéines Zif12-[L<sub>2</sub>-Tb]-A, Zif12-[L<sub>2</sub>-Eu]-A et Zif12-[L<sub>10</sub>-Eu]-A ont été respectivement caractérisées **Figure 143**, **Figure 144** et **Figure 145**.



**Figure 143** : A) Chromatogramme HPLC, détection UV à 214 nm de la protéine Zif12-[L<sub>2</sub>-Tb]-A purifiée. B) Le spectre ESI-MS de la protéine Zif12-[L<sub>2</sub>-Tb]-A purifiée (masse théorique calculée : 9276.41, masse déconvoluée obtenue : 9275.43).



**Figure 144** : A) Chromatogramme HPLC, détection UV à 214 nm de la protéine Zif12-[L<sub>2</sub>-Eu]-A purifiée. B) Le spectre ESI-MS de la protéine Zif12-[L<sub>2</sub>-Eu]-A purifiée (masse théorique calculée : 9269.45, masse déconvoluée obtenue : 9268.71).



**Figure 145** : A) Chromatogramme HPLC, détection UV à 214 nm de la protéine Zif12-[L<sub>10</sub>-Eu]-A purifiée. B) Le spectre ESI-MS de la protéine Zif12-[L<sub>10</sub>-Eu]-A purifiée (masse théorique calculée : 9281.46, masse déconvoluée obtenue : 9281.42).

En conclusion de ce chapitre, nous avons réalisé la synthèse de plusieurs protéines fonctionnalisées par des chromophores organiques (SiR<sup>Me</sup> et Lissamine) qui jouent le rôle d'accepteur d'énergie dans notre système FRET à deux protéines. La synthèse de la protéine fonctionnalisée par le très prometteur complexe [L<sub>1</sub><sup>CC</sup>-Eu] n'a pu être réalisée à cause d'une réaction parasite thiol-yne entre les cystéines de

la protéine et la fonction alcyne de l'antenne lors de la dernière étape de ligation chimique native. En conséquence, plusieurs autres ligands avec des antennes de type biaryle et leurs complexes de lanthanide ont été synthétisés. Parmi eux, deux ligands ont été retenus et les complexes [ $L_2$ -Tb], [ $L_2$ -Eu] et [ $L_{10}$ -Eu] ont été greffés sur une protéine. Nous avons donc à disposition trois protéines comportant un complexe de lanthanide pour générer le transfert d'énergie après dimérisation autour de la séquence d'ADN ainsi que deux protéines comportant un chromophore (SiR<sup>Me</sup> ou Lissa).

# **Chapitre IV - Interaction des protéines Zif12 fonctionnalisées avec les duplex d'ADN**

## Sommaire

<b>I - Métallation de protéines Zif12 décorées de luminophores et caractérisations spectroscopiques</b>	<b>139</b>
1 - Protéines Zif12 <sup>*</sup> -(SiR <sup>Me</sup> )-B et Zif12 <sup>*</sup> -(Lissa)-B .....	139
2 - Protéines Zif12 <sup>*</sup> -[L <sub>2</sub> -Tb]-A, Zif12 <sup>*</sup> -[L <sub>2</sub> -Eu]-A et Zif12 <sup>*</sup> -[L <sub>10</sub> -Eu]-A.....	141
2.1 - Protéines Zif12 <sup>*</sup> -[L <sub>2</sub> -Tb]-A .....	141
2.2 - Protéines Zif12 <sup>*</sup> -[L <sub>2</sub> -Eu]-A et Zif12 <sup>*</sup> -[L <sub>10</sub> -Eu]-A .....	143
<b>II - Etude de l'interaction de Zif12<sup>*</sup>-[L<sub>2</sub>-Eu]-A / Zif12<sup>*</sup>-[L<sub>10</sub>-Eu]-A et Zif12<sup>*</sup>-(SiR<sup>Me</sup>)-B avec l'ADN</b>	<b>144</b>
1 - Expériences préliminaires .....	144
2 - Stabilité des protéines .....	148
3 - Etude de la sélectivité avec le couple Zif12 <sup>*</sup> -[L <sub>2</sub> -Eu]-A / Zif12 <sup>*</sup> -(SiR <sup>Me</sup> )-B .....	153
4 - Etude du couple Zif12 <sup>*</sup> -[L <sub>10</sub> -Eu]-A et Zif12 <sup>*</sup> -(SiR <sup>Me</sup> )-B .....	155
5 - Etude du couple Zif12 <sup>*</sup> -[L <sub>10</sub> -Eu]-A et Zif12 <sup>*</sup> -(Lissa)-B .....	159
<b>III - Conclusion .....</b>	<b>160</b>

Lors du second chapitre nous avons démontré la sélectivité et l'affinité de notre système à deux protéines doigts de zinc pour la reconnaissance d'une séquence d'ADN longue de 12 paires de bases et nous avons vu qu'il était possible de détourner le système décrit par Wang et Pabo pour reconnaître un double brin non palindromique. Au sein de ce dernier chapitre, nous allons nous intéresser à l'interaction des protéines fonctionnalisées avec les différents duplexes d'ADN par luminescence et dans certains cas par gel retard. La fonctionnalisation des protéines par le système FRET peut perturber la formation du dimère.

Pour rappel, dans le précédent chapitre, nous avons décrit la synthèse de plusieurs protéines Zif12 comportant un complexe de lanthanide comme donneur FRET ou un fluorophore organique qui joue le rôle d'accepteur d'énergie. L'ensemble des protéines fonctionnalisées obtenues et leurs cibles d'ADN sont rassemblées dans le **Tableau 8**.

*Tableau 8 : Protéines Zif12 fonctionnalisées synthétisées et la séquence d'ADN ciblée.*

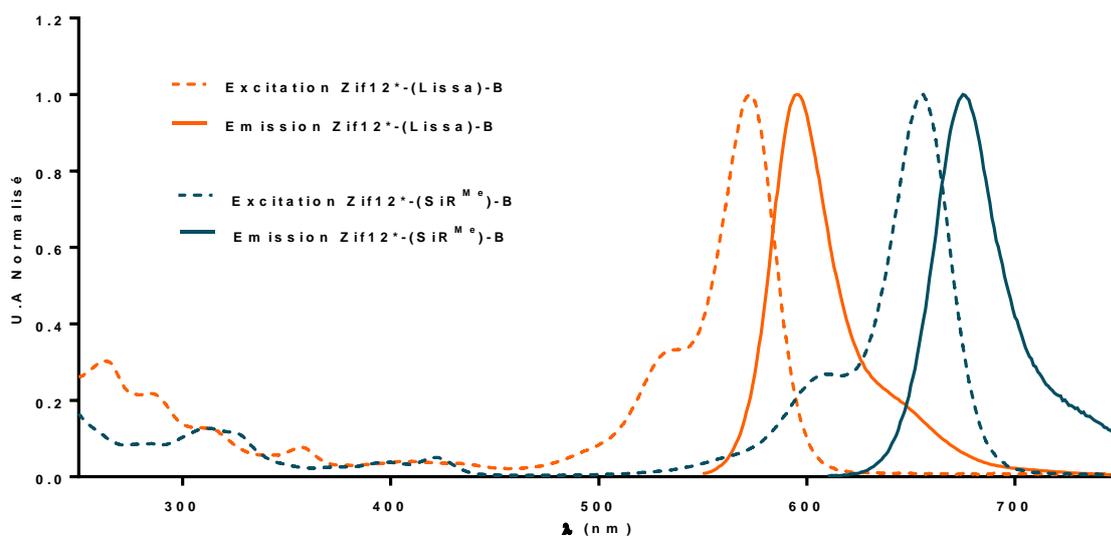
	Zif12-(SiR <sup>Me</sup> )	Zif12-(Lissa)	Zif12-[L <sub>2</sub> -Tb]	Zif12-[L <sub>2</sub> -Eu]	Zif12-[L <sub>10</sub> -Eu]
<b>ADN Ciblé (A) :</b> 5'-TGGGCG-3'	x	x	x	x	x
<b>ADN Ciblé (B) :</b> 5'-GAAGGG-3'	x	x			

Pour être fonctionnelles, ces protéines doivent d'abord être métallées avec du Zn<sup>2+</sup> afin de replier correctement les domaines en doigts de zinc. Dans une première partie, nous présenterons la métallation au zinc de ces protéines et leur caractérisation spectroscopique. Puis dans une seconde partie, nous étudierons leur interaction avec divers ADN double brin.

## I - Métallation de protéines Zif12 décorées de luminophores et caractérisations spectroscopiques

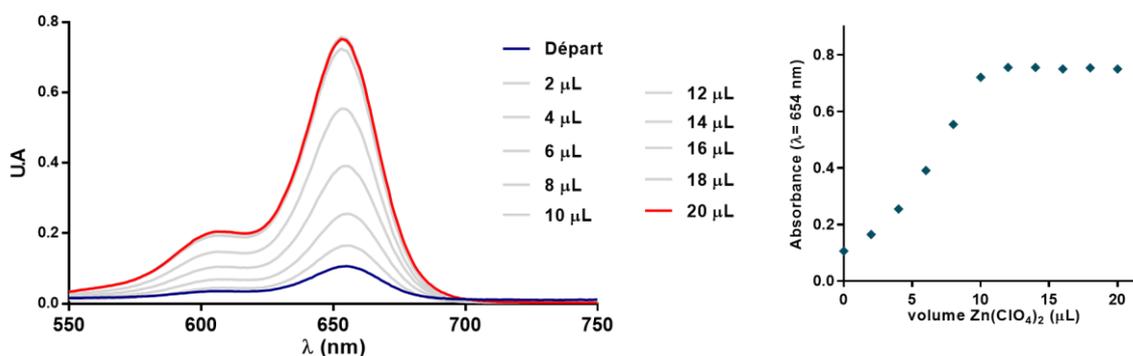
### 1 - Protéines Zif12<sup>\*</sup>-(SiR<sup>Me</sup>)-B et Zif12<sup>\*</sup>-(Lissa)-B

La métallation des protéines Zif12-(SiR<sup>Me</sup>)-B et Zif12-(Lissa)-B est réalisée dans une solution tamponnée par de l'HEPES (10 mM, pH 7.0), auquel est ajouté du TCEP (250 μM) pour réduire d'éventuels ponts disulfure intramoléculaires. Les coefficients d'absorption molaire de la SiR<sup>Me</sup><sup>167</sup> et de la Lissamine<sup>168</sup> trouvés dans la littérature sont respectivement 100 000 M<sup>-1</sup>cm<sup>-1</sup> et 71 500 M<sup>-1</sup>cm<sup>-1</sup>. La métallation est assurée par l'ajout de 5 éq de Zn(ClO<sub>4</sub>)<sub>2</sub> et la protéine est aliquotée et lyophilisée pour être conservée à -20°C. Les spectres d'émission et d'excitation de fluorescence des protéines Zif12<sup>\*</sup>-(SiR<sup>Me</sup>)-B et Zif12<sup>\*</sup>-(Lissa)-B sont présentés **Figure 146**. Les longueurs d'onde maximales d'émission de la SiR<sup>Me</sup> et de la Lissamine sont respectivement de 675 et 595 nm avec une excitation maximale à 658 nm pour la SiR<sup>Me</sup> et 573 nm pour la Lissamine.



**Figure 146** : Spectres d'émission et d'excitation normalisés des protéines et Zif12\*(SiR<sup>Me</sup>)-B et Zif12\*(Lissa)-B (orange). En orange le spectre d'émission ( $\lambda_{ex} = 525$  nm) et en pointillé orange l'excitation ( $\lambda_{em} = 600$  nm) normalisé de la protéine Zif12\*(Lissa)-B. En turquoise le spectre d'émission ( $\lambda_{ex} = 600$  nm) et en pointillé turquoise l'excitation ( $\lambda_{em} = 710$  nm) normalisé de la protéine Zif12\*(SiR<sup>Me</sup>)-B.

Notons qu'après l'ajout de  $Zn^{2+}$  à Zif12-(SiR<sup>Me</sup>)-B, nous avons noté que la couleur bleue de la solution devenait plus foncée. Nous avons suivi l'évolution de l'absorption de la protéine au cours d'un titrage par  $Zn^{2+}$  ( $Zn(ClO_4)_2$  à 0.5 mM dans  $H_2O$ ). Les spectres sont représentés sur la **Figure 147**, tout comme l'évolution de l'absorbance à 675 nm en fonction du volume de solution de zinc ajouté. L'absorbance augmente de manière non linéaire (une légère incurvation est visible) jusqu'à atteindre un maximum pour environ 10  $\mu L$  de solution de  $Zn^{2+}$  ajouté, puis elle reste stable pour les ajouts suivants de  $Zn^{2+}$ . Ce type d'évolution est attendu pour une protéine liant deux anions  $Zn^{2+}$  dans des sites d'affinité différentes. La fluorescence évolue de manière similaire.



**Figure 147** : Évolution du spectre d'absorption de la protéine Zif12-(SiR<sup>Me</sup>)-B en fonction des ajouts d'une solution de  $Zn(ClO_4)_2$  à 0.5 mM.

Cette forte augmentation de l'absorption (et de la fluorescence) provoquée par  $Zn^{2+}$  peut s'expliquer par la sensibilité de composés de la famille de la SiR<sup>Me</sup> décrite par le groupe d'Urano<sup>169</sup>. Les thiols peuvent s'additionner sur la position *meso* du cœur silico-xanthène (**Figure 148**). Dans le cas de la SiR<sup>Me</sup>,

la constante de formation de l'adduit avec la cystéine ou le glutathion est très faible et celui-ci ne peut pas exister à la concentration à laquelle nous travaillons. Mais dans notre cas, il s'agirait d'une réaction intramoléculaire. Ce titrage nous a cependant permis de déterminer précisément la valeur du coefficient d'absorption molaire de la protéine Zif12<sup>\*</sup>-(SiR<sup>Me</sup>)-B (métallée avec Zn<sup>2+</sup>) dans nos conditions. Nous avons trouvé une valeur de 135 000 M<sup>-1</sup>cm<sup>-1</sup>.

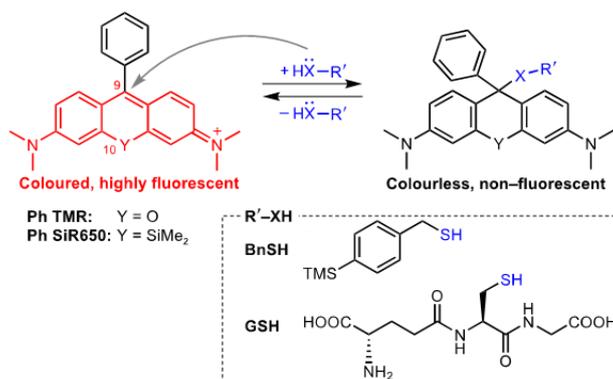


Figure 148 : Réaction d'addition nucléophile sur la rhodamine<sup>169</sup>.

## 2 - Protéines Zif12<sup>\*</sup>-[L<sub>2</sub>-Tb]-A, Zif12<sup>\*</sup>-[L<sub>2</sub>-Eu]-A et Zif12<sup>\*</sup>-[L<sub>10</sub>-Eu]-A

### 2.1 - Protéines Zif12<sup>\*</sup>-[L<sub>2</sub>-Tb]-A

Pour rappel notre objectif est d'utiliser un système FRET basé sur un complexe d'euprium. Cependant, afin d'explorer le transfert d'énergie nous avons synthétisé la protéine Zif12<sup>\*</sup>-[L<sub>2</sub>-Tb]-A basée sur le complexe [L<sub>2</sub>-Tb], dont le rendement quantique est de 55% (mesuré pour son conjugué au peptide TAT). La protéine Zif12<sup>\*</sup>-[L<sub>2</sub>-Tb]-A a été métallée dans les conditions décrites précédemment, avec 5 équivalents de Zn<sup>2+</sup> dans l'HEPES en présence de TCEP. Ses spectres d'émission et d'excitation sont présentés **Figure 149**.

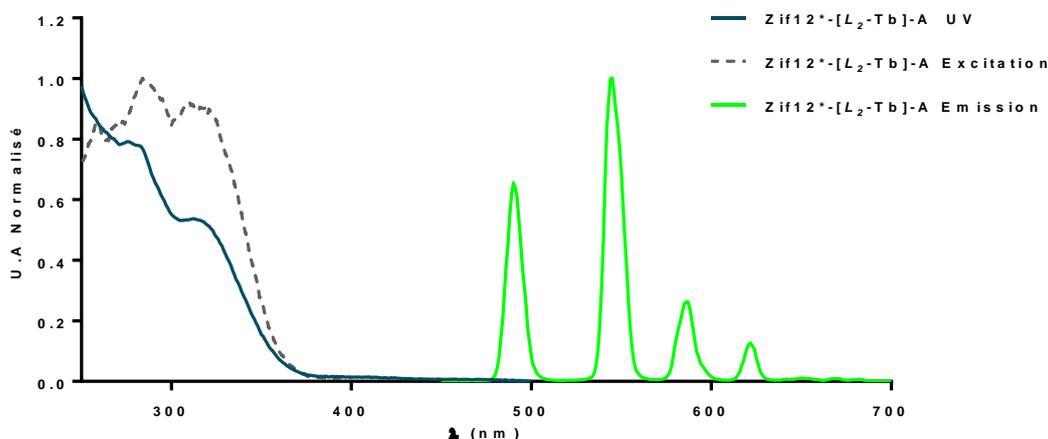


Figure 149 : Spectres d'absorbance (turquoise), d'émission (vert) obtenu après excitation à  $\lambda_{ex} = 315$  nm et le spectre d'excitation (pointillé gris) obtenu après détection à  $\lambda_{em} = 545$  nm du fragment Zif12<sup>\*</sup>-[L<sub>2</sub>-Tb]-A dans une solution aqueuse de tampon HEPES (10 mM, pH 7.0).

Le spectre d'émission enregistré après excitation du complexe à 315 nm présente les raies d'émission caractéristiques  $^5D_4 \rightarrow ^7F_J$  ( $J = 3-6$ ) du  $Tb^{3+}$  respectivement à 490, 545, 585 et 623 nm. Cependant le temps de vie du  $Tb^{3+}$ , mesuré à son maximum d'émission ( $\lambda_{max} Tb^{3+} = 545$  nm) n'est que de 0.32 ms. Ce temps de vie est extrêmement faible en comparaison du conjugué TAT-[L<sub>2</sub>-Eu] qui possède un temps de vie de 2.11 ms. Nous savons qu'avec ce type de complexe un transfert d'énergie retour du  $Tb^{3+}$  ( $E(^5D_4) = 20\,400\text{ cm}^{-1}$ ) vers l'état triplet excité T<sub>1</sub> de l'antenne est possible ( $E(T_1) = 22\,600\text{ cm}^{-1}$ ), mais que cet état T<sub>1</sub> excité n'est pas sensible à l'oxygène et n'offre pas de voie de désexcitation pour le  $Tb^{3+}$  (l'énergie revient du T<sub>1</sub> au  $^5D_4$ ). En parallèle de mes observations, Ji-Hyung Choi au laboratoire a également observé un temps de vie très court avec ce complexe conjugué à un autre peptide doigt de zinc, alors que pour le conjugué à TAT, le temps de vie est long. Le peptide doigt de zinc utilisé par Ji-Hyung Choi contenait une tyrosine et la mutation de cette tyrosine en phénylalanine a permis de retrouver un temps de vie long. L'explication que nous donnons à ce phénomène est un transfert d'énergie, thermiquement activé, de l'état excité  $^5D_4$  du terbium au T<sub>1</sub> de l'antenne puis du T<sub>1</sub> de l'antenne vers le T<sub>1</sub> de la tyrosine voisine ( $E(T_1^{Tyr}) = 26\,500\text{ cm}^{-1}$ ), qui lui est sensible à la présence d'O<sub>2</sub>, offrant cette fois une voie de désexcitation pour l'état excité  $^5D_4$  du  $Tb^{3+}$ , via les états T<sub>1</sub> de l'antenne et T<sub>1</sub> de la tyrosine (**Figure 150**). Dans la séquence de l'extension N-terminale le complexe de lanthanide est fixé sur la lysine 106 et une tyrosine se trouve en position

97.

Ces tyrosines peuvent faire office d'accepteur d'énergie. Nous avons songé dans un premier temps à synthétiser un nouvel analogue de la protéine Zif12-Nle-A, en remplaçant ces tyrosines par une phénylalanine dont le triplet est plus haut en énergie. Cependant sur la structure RX du dimère de (Zif12)<sub>2</sub>/ADN<sup>54</sup>, il y a de nombreuses liaisons hydrogène qui sont formées à partir du phénol de la chaîne latérale de la tyrosine, une modification de cet acide aminé ne semble donc pas judicieuse. Le faible temps de vie de Zif12\*-[L<sub>2</sub>-Tb]-A, et très probablement le faible rendement quantique associé, ne permet pas d'envisager raisonnablement cette protéine dans notre système FRET. Son étude n'a pas été poursuivie. L'état excité  $^5D_0$  de l'Eu<sup>3+</sup> étant plus bas en énergie ( $17\,500\text{ cm}^{-1}$ ), ce problème ne doit pas être rencontré avec Zif12-[L<sub>2</sub>-Eu]-A.

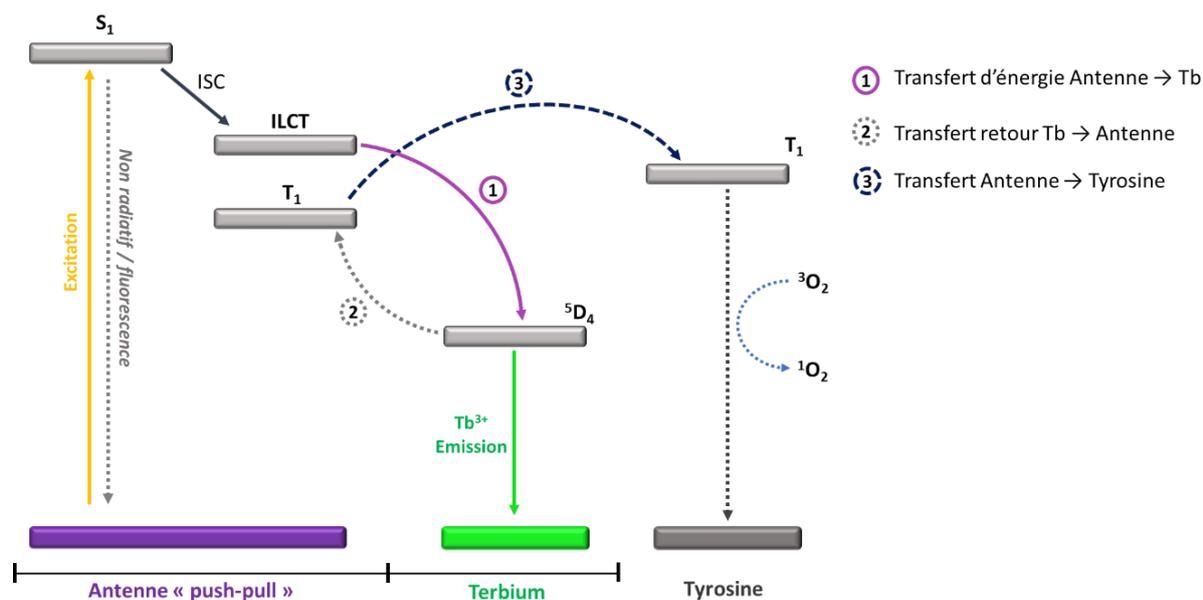
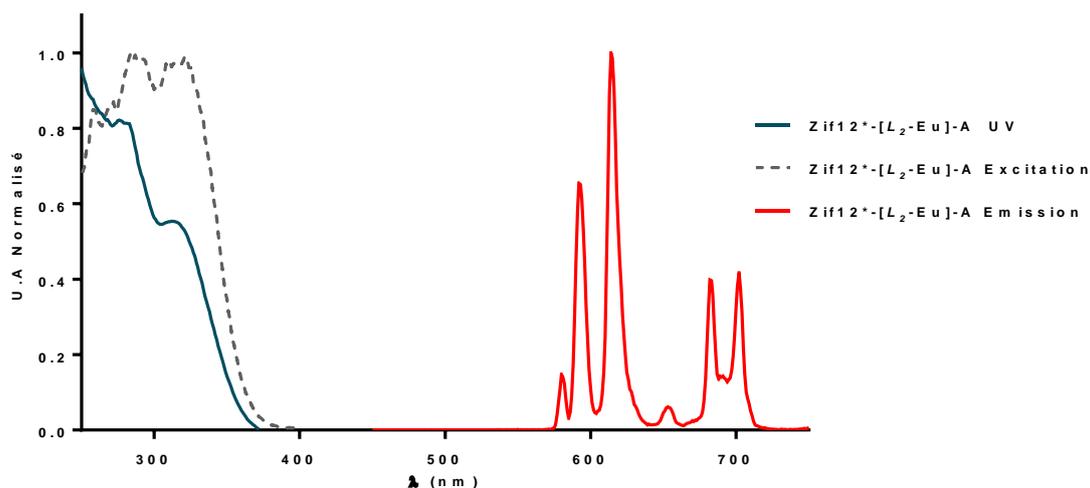


Figure 150 : Diagramme de Jablonski-Perrin simplifié du transfert d'énergie du complexe  $[L_2-Tb]$  vers la tyrosine.

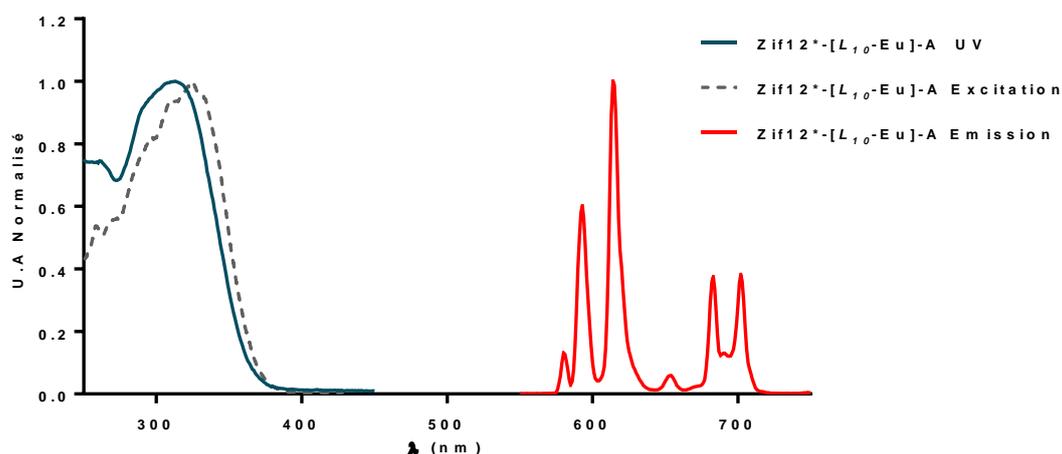
## 2.2 - Protéines Zif12\*-[L<sub>2</sub>-Eu]-A et Zif12\*-[L<sub>10</sub>-Eu]-A

Nous avons réalisé la métallation des protéines Zif12-[L<sub>2</sub>-Eu]-A et Zif12-[L<sub>10</sub>-Eu]-A selon la procédure décrite dans la partie expérimentale. Les spectres d'absorption, d'émission et d'excitation des Zif12\*-[L<sub>2</sub>-Eu]-A et Zif12\*-[L<sub>10</sub>-Eu]-A sont présentés respectivement **Figure 151** et **Figure 152**. L'ensemble des spectres coïncident avec ceux obtenus lors de la caractérisation des complexes TAT-[L<sub>2</sub>-Eu] et TAT-[L<sub>10</sub>-Eu] (**Chapitre III, III-4.2**). Sur le spectre d'absorbance de la protéine Zif12\*-[L<sub>2</sub>-Eu]-A on retrouve la transition LE autour de 280 nm et la transition ILCT à 315 nm. Cette dernière semble peu intense vis-à-vis de la transition LE, mais il faut prendre en compte que les tyrosines de la protéine absorbent dans cette zone, avec un maximum à 280 nm. L'excitation de la protéine à 315 nm a permis d'enregistrer une forte émission de l'euprium dont le temps de vie a été mesuré à 1.07 ms. Ce temps de vie est en adéquation avec celui mesuré sur la protéine TAT-[L<sub>2</sub>-Eu], ce qui atteste qu'il n'y a pas de quenching par transfert d'énergie vers la tyrosine. Le spectre d'excitation mesuré à 615 nm montre que la sensibilisation de l'euprium provient essentiellement de la bande de transfert de charge ILCT de l'antenne. Pour la protéine Zif12\*-[L<sub>10</sub>-Eu]-A, les spectres enregistrés coïncident avec ceux obtenus lors de la caractérisation de la protéine TAT-[L<sub>10</sub>-Eu]. Le temps de vie mesuré pour la protéine Zif12\*-[L<sub>10</sub>-Eu]-A est légèrement plus bas que celui de Zif12\*-[L<sub>2</sub>-Eu]-A à 0.96 ms, mais reste voisin de 1 ms.

La protéine Zif12\*-[L<sub>2</sub>-Eu]-A a été obtenue plusieurs mois avant Zif12\*-[L<sub>10</sub>-Eu]-A, et c'est donc avec elle que nous avons débuté les études avec l'ADN en luminescence en l'associant avec la protéine Zif12\*-(SiR<sup>Me</sup>)-B.



**Figure 151** : Spectres d'absorbance (turquoise), d'émission (rouge) obtenu après excitation à  $\lambda_{ex} = 315$  nm et le spectre d'excitation (pointillé gris) obtenu après détection à  $\lambda_{em} = 615$  nm du fragment Zif12\*-[L<sub>2</sub>-Eu]-A dans une solution aqueuse de tampon HEPES (10 mM, pH 7.0).



**Figure 152** : Spectres d'absorbance (turquoise), d'émission (rouge) obtenu après excitation à  $\lambda_{ex} = 315$  nm et le spectre d'excitation (pointillé gris) obtenu après détection à  $\lambda_{em} = 615$  nm du fragment Zif12\*-[L<sub>10</sub>-Eu]-A dans une solution aqueuse de tampon HEPES (10 mM, pH 7.0).

## II - Etude de l'interaction de Zif12\*-[L<sub>2</sub>-Eu]-A / Zif12\*-[L<sub>10</sub>-Eu]-A et Zif12\*-(SiR<sup>Me</sup>)-B avec l'ADN

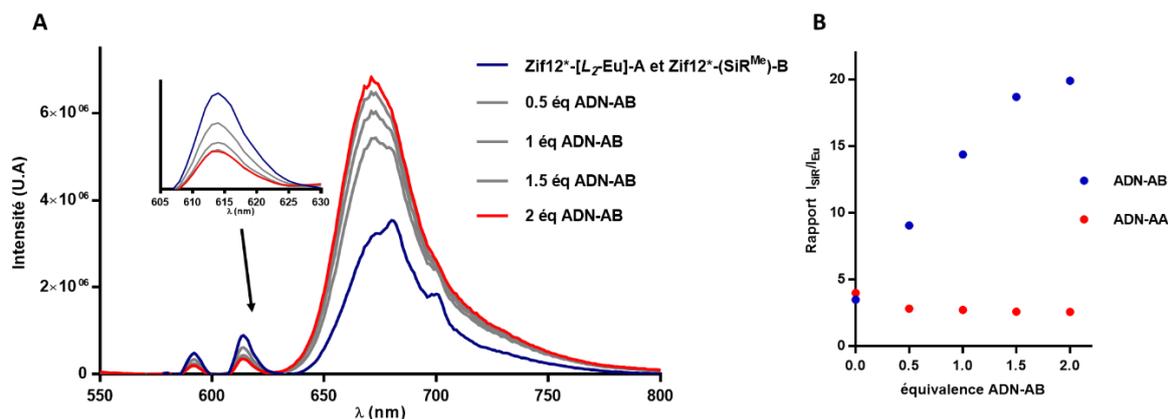
### 1 - Expériences préliminaires

Pour l'étude spectroscopique de l'interaction des protéines fonctionnalisées avec les luminophores (SiR<sup>Me</sup> ou complexe de lanthanide) avec l'ADN, nous avons choisi de travailler dans des conditions similaires à celles utilisées lors des gels retards. Nous avons travaillé dans une solution aqueuse de tampon C (10% v/v) dont la composition est la suivante : 0.1 mg/mL de BSA (agent de solubilisation),

0.05% (v/v) de Tween (détergent), de NaCl (100 mM) et de tampon HEPES (15mM) à pH 7.6. À partir des solutions mères des protéines Zif12\*-[L<sub>2</sub>-Eu]-A et Zif12\*-(SiR<sup>Me</sup>)-B, de concentration respective 10.4 μM et 10.2 μM, un volume de chacune des solutions est ajouté pour ajuster la concentration de chacune des protéines à 500 nM. Le duplex d'ADN-AB a été ajouté par la suite pour avoir respectivement 0.5, 1, 1.5 et 2 équivalents d'ADN à partir d'une solution mère de duplex dans l'eau de concentration 25 μM. Ce titrage avait deux buts : (i) évaluer si la sonde pouvait répondre à la présence de l'ADN via un FRET et (ii) valider la stœchiométrie un ADN pour une paire de protéines.

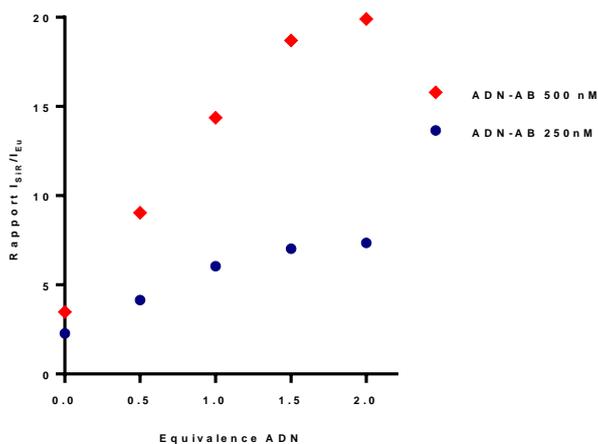
Les différents spectres d'émission ont été enregistrés en excitant à 315 nm. Chaque ajout d'ADN est espacé d'un délai de 5 min qui correspondant au temps d'acquisition du spectre. Lors de l'ajout de la protéine Zif12\*-(SiR<sup>Me</sup>)-B on observe une forte émission de la SiR<sup>Me</sup> vis-à-vis de l'émission de l'euporium. Il n'est pas surprenant de voir l'émission de la SiR<sup>Me</sup> en absence de la cible d'ADN puisqu'elle peut être sensibilisée à cette longueur d'onde d'excitation. Le coefficient d'extinction molaire  $\epsilon_{315\text{SiR}^{\text{Me}}}$  à 315 nm est de 16700 M<sup>-1</sup>cm<sup>-1</sup>, supérieur à celui de [L<sub>2</sub>-Eu] qui est de 11800 M<sup>-1</sup>cm<sup>-1</sup>. Le rendement quantique de SiR<sup>Me</sup> est de 30% contre 4% pour [L<sub>2</sub>-Eu]. Il n'est donc pas étonnant de voir l'émission de la silico-rhodamine dominer le spectre. Au fur et à mesure que les ajouts du duplex d'ADN-AB sont effectués, on observe une diminution de l'intensité d'émission de l'euporium (**Figure 153 A**).

L'émission de la SiR<sup>Me</sup> quant à elle augmente. Ceci va dans le sens d'un transfert d'énergie entre l'Eu<sup>3+</sup> et la SiR<sup>Me</sup> de plus en plus important au fur et à mesure des ajouts d'ADN et donc d'une fixation des deux protéines sur l'ADN-AB. Afin de mieux distinguer les émissions de Eu<sup>3+</sup> et SiR<sup>Me</sup> nous avons décidé d'intégrer le signal autour de la transition  $^5\text{D}_0 \rightarrow ^7\text{F}_1$  à 615 nm ( $\pm 2$  nm) pour Eu<sup>3+</sup> et l'émission autour de 668 nm ( $\pm 2$  nm) pour la SiR<sup>Me</sup> (l'euporium ne présente aucune émission dans cette zone) et de s'intéresser aux rapports des deux émissions  $I_{\text{SiR}}/I_{\text{Eu}}$  présentés **Figure 153 B**. On peut observer que la saturation n'est pas atteinte à 1 éq d'ADN-AB, ce qui semble indiquer que l'affinité est plus faible que mesurée avec les protéines nues (avec un C<sub>50</sub> de 75 nM pour le complexe ternaire et une concentration de 500 nM pour chaque protéine et 500 nM pour l'ADN, la formation du complexe ternaire devrait être totale). Lors des ajouts d'ADN-AB on peut voir que le rapport  $I_{\text{SiR}}/I_{\text{Eu}}$  initialement de 3.5 augmente jusqu'à presque 20.



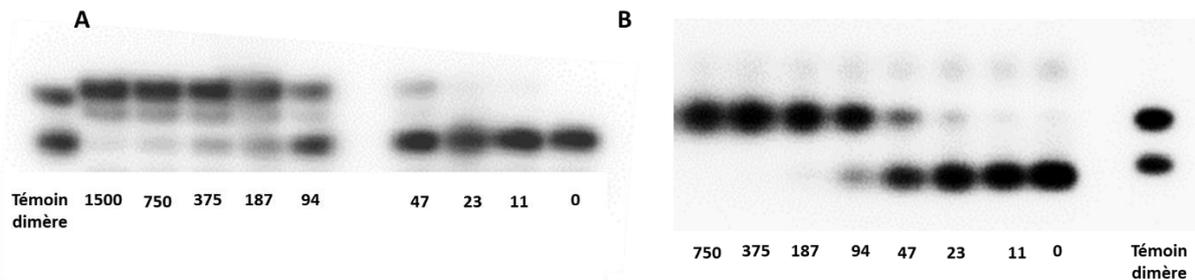
**Figure 153 :** A) Spectres d'émission totale à  $\lambda_{ex}=315$  nm des protéines Zif12\*-[L<sub>2</sub>-Eu]-A et Zif12\*-(SiR<sup>Me</sup>)-B à 500 nM en présence du duplex d'ADN-AB. B) Évolution du rapport des intensités d'émission  $I_{SiR}/I_{Eu}$  des protéines Zif12\*-[L<sub>2</sub>-Eu]-A et Zif12\*-(SiR<sup>Me</sup>)-B en présence d'équivalents d'ADN-AA d'ADN-AB.

Cette même expérience a été réalisée en utilisant l'ADN-AA sur lequel seule la protéine Zif12\*-[L<sub>2</sub>-Eu]-A peut théoriquement se fixer, le ratio  $I_{SiR}/I_{Eu}$  reste constant tout au long de l'expérience. L'augmentation du rapport  $I_{SiR}/I_{Eu}$  est donc liée à la reconnaissance de l'ADN cible. Ces premiers résultats semblent indiquer que notre système est capable de discriminer les ADN-AA et AB sur la base d'un FRET provoqué par la fixation des protéines sur le double brin d'ADN-AB, mais pas sur l'ADN-AA. Nous avons réitéré le titrage des protéines par l'ADN-AB en abaissant la concentration des protéines à 250 nM. Le rapport d'intensité  $I_{SiR}/I_{Eu}$  augmente, mais cette fois-ci, la valeur initiale est de 2.3 qui est inférieure à celle de 3.5 obtenue précédemment. En présence de 2 éq d'ADN, il atteint une valeur de 7.4 au lieu de 20 (**Figure 154**). On peut également voir une inflexion plus précoce. Ces deux derniers éléments laissent penser que la complexation de l'ADN aux protéines n'est pas totale. Ceci va dans le sens d'une affinité plus faible que celle mesurée avec les protéines fonctionnalisées avec les luminophores. C'est ce que nous avons cherché à vérifier par gel-retard.

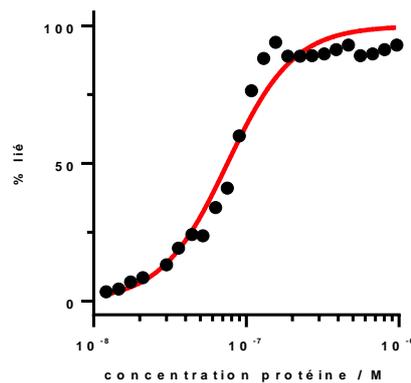


**Figure 154 :** Évolution du rapport  $I_{SiR}/I_{Eu}$  de l'interaction des protéines Zif12\*-[L<sub>2</sub>-Eu]-A et Zif12\*-(SiR<sup>Me</sup>)-B avec la séquence d'ADN-AB à 500 nM et 250 nM en protéine.

L'incubation d'un mélange 1:1 des protéines Zif12\*-[L<sub>2</sub>-Eu]-A et Zif12\*-(SiR<sup>Me</sup>)-B, à différentes concentrations avec la séquence d'ADN-AB (10 nM) marqué au <sup>32</sup>P dans le tampon C a été réalisée à 4°C durant 12 heures et la migration a été réalisée durant 2 heures à 5 Watts (**Figure 155 A**). On voit nettement sur le gel l'apparition d'une bande retardée qui indique la formation du complexe tertiaire avec deux protéines fixées sur l'ADN-AB ainsi qu'une bande intermédiaire attribuée à la fixation d'une seule des deux protéines sur l'ADN-AB. Visuellement, on peut estimer la concentration de demi-liaison C<sub>50</sub> de l'ADN à environ 100 nM. La même expérience a été réalisée en remplaçant Zif12\*-[L<sub>2</sub>-Eu]-A par Zif12\*-[L<sub>10</sub>-Eu]-A (**Figure 155 B**). On ne voit cette fois-ci qu'une seule bande retardée pour le complexe ternaire. La concentration en protéine conduisant à la formation de 50% du dimère est C<sub>50</sub> = 75 ± 8 nM et K<sub>d</sub> = (5.7 ± 1.2) × 10<sup>-15</sup> M<sup>2</sup> (**Figure 156**). Ce qui correspond à une perte d'affinité d'un facteur 100 par rapport aux protéines non décorées décrites dans le **Chapitre II, II-3** et une concentration de demi-liaison 10 fois plus faible.



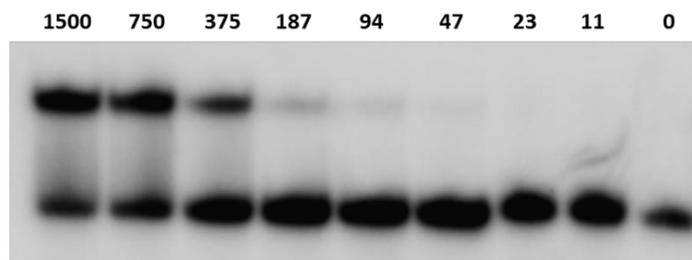
**Figure 155** : A) Gel retard obtenu des protéines Zif12\*-[L<sub>2</sub>-Eu]-A et Zif12\*-(SiR<sup>Me</sup>)-B en interaction avec l'ADN-AB. B) Gel retard obtenu des protéines Zif12\*-[L<sub>10</sub>-Eu]-A et Zif12\*-(SiR<sup>Me</sup>)-B en interaction avec l'ADN-AB. Pour chacun des gels : incubation durant 12 heures à 4°C, migration à 10 Watts durant 2 heures. La concentration des protéines est indiquée en nM, le témoin dimère est obtenu de l'interaction de la protéine LRZif sur l'ADN-AA.



**Figure 156** : Dosage de l'ADN-AB par les protéines Zif12\*-[L<sub>10</sub>-Eu]-A et Zif12\*-(SiR<sup>Me</sup>)-B (noir) et ajustement (rouge) permettant la détermination de C<sub>50</sub> et K<sub>d</sub>.

Nous avons ensuite étudié l'interaction des protéines Zif12\*-[L<sub>2</sub>-Eu]-A et Zif12\*-(SiR<sup>Me</sup>)-B avec les ADN palindromiques respectifs ADN-AA et ADN-BB, dans les mêmes conditions que précédemment. Le gel obtenu pour Zif12\*-[L<sub>2</sub>-Eu]-A, présenté **Figure 157** montre une bande retardée correspondant à la formation du dimère. Cependant cette bande n'est visible qu'à de très hautes concentrations en

protéine, c'est seulement à partir d'une concentration de 375 nM en protéine que nous voyons clairement l'apparition de cette bande retardée et le  $C_{50}$  est estimé à 750 nM. En comparaison avec le gel obtenu de la protéine non décorée, il est évident que la fonctionnalisation avec le complexe d'euporium a eu un impact négatif sur la formation du dimère. Pour rappel la valeur du  $C_{50}$  de la protéine Zif12\*-Nle-A sur la séquence d'ADN-AA est de  $7.6 \pm 0.6$  nM, soit environ une perte de deux ordres de grandeur.

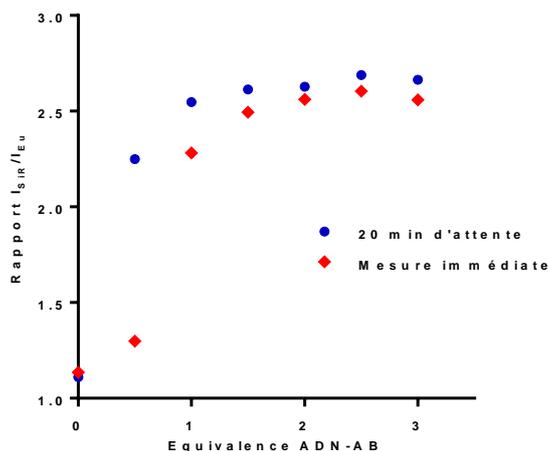


**Figure 157** : Gel retard obtenu de l'interaction de la protéine Zif12\*-[L<sub>2</sub>-Eu]-A avec la séquence d'ADN-AA. Incubation durant 12 heures à 4°C, migration à 10 Watts durant 2 heures. La concentration des protéines est indiquée en nM.

Pour la protéine Zif12\*-(SiR<sup>Me</sup>)-B avec l'ADN-BB, l'effet est encore plus dramatique puisqu'aucune bande retardée n'a été observée malgré une gamme de concentration allant jusqu'à 1500 nM en protéine. Il semble donc que la fonctionnalisation des protéines par les luminophores ait un impact important sur la stabilité du complexe ternaire en le déstabilisant de plusieurs ordres de grandeur. Pour expliquer cela, on peut proposer que la SiR<sup>Me</sup> ou l'antenne du complexe d'Eu<sup>3+</sup>, de nature hydrophobe, interagissent fortement avec les acides aminés hydrophobes de l'extension N-terminale (L, V, Y) lorsque les protéines ne sont pas fixées à l'ADN. La fixation à l'ADN détruirait ces interactions entre groupements hydrophobes pour exposer la SiR<sup>Me</sup> ou l'antenne à l'eau avec un coût énergétique important.

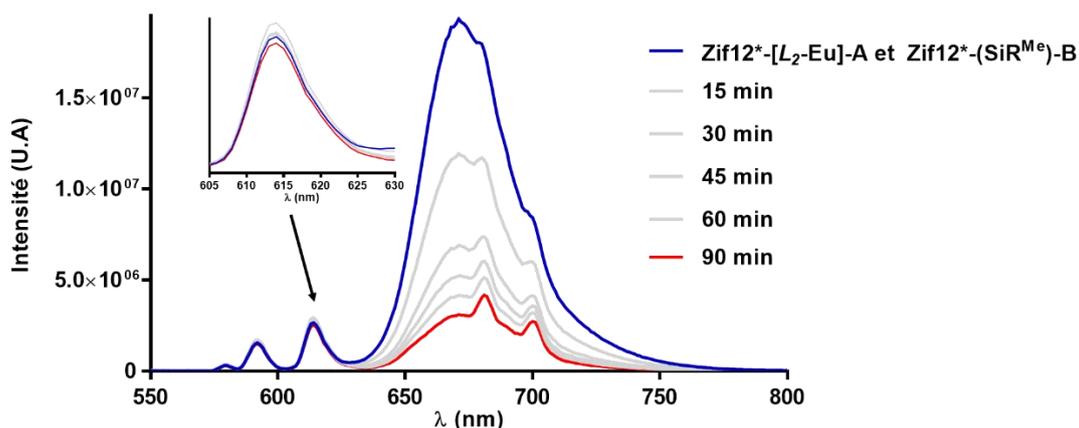
## 2 - Stabilité des protéines

Nous avons repris l'étude par luminescence et avons été confrontés à des problèmes de reproductibilité, notamment le ratio  $I_{SiR}/I_{Eu}$  initial et en excès d'ADN ne sont pas les mêmes d'un jour à l'autre. Nous nous sommes donc intéressés à l'évolution du spectre d'émission au cours du temps après chaque ajout d'ADN-AB lors du titrage des protéines Zif12\*-[L<sub>2</sub>-Eu]-A et Zif12\*-(SiR<sup>Me</sup>)-B à une concentration en protéine de 250 nM. Nous avons décidé de faire des mesures immédiatement après chaque ajout et après 20 minutes d'attente. Sur la **Figure 158**, qui représente l'évolution du rapport  $I_{SiR}/I_{Eu}$  en fonction des équivalents d'ADN-AB, l'écart entre les deux ratios à 0.5 éq (avec ou sans temps d'attente) nous indique que le système évolue au cours du temps et qu'un délai est nécessaire pour stabiliser le système lorsque la concentration en ADN est faible. Nous devrions voir une augmentation linéaire entre 0 et 1 équivalent et l'apparition d'un plateau à partir de 1 équivalent d'ADN. Cependant cette équivalence est atteinte en deçà de 1.



**Figure 158 :** Variation du rapport  $I_{SiR}/I_{Eu}$  des protéines Zif12\*-[L<sub>2</sub>-Eu]-A et Zif12\*-(SiR<sup>Me</sup>)-B en fonction des ajouts d'ADN-AB avec et sans temps d'attente. La concentration des protéines est de 250 nM.

Nous avons suivi l'évolution au cours du temps d'un mélange équimolaire de Zif12\*-[L<sub>2</sub>-Eu]-A et Zif12\*-(SiR<sup>Me</sup>)-B à une concentration de 250 nM. L'émission de Eu<sup>3+</sup> évolue peu dans le temps alors que l'émission de la SiR<sup>Me</sup> chute fortement après 15 minutes, puis continue de diminuer (**Figure 159**). Il est clair que la protéine Zif12\*-(SiR<sup>Me</sup>)-B présente un problème de stabilité. Plusieurs causes peuvent être envisagées pour expliquer ce phénomène : (i) un photoblanchiment de la SiR<sup>Me</sup>, mais ceci est peu probable puisque les silico-rhodamine sont connues pour être très photostables, (ii) une réaction de la SiR<sup>Me</sup> avec des thiols, notamment la cystéine de la BSA qui est utilisée dans le tampon ou (iii) une précipitation de la protéine, mais qui se produirait uniquement avec Zif12\*-(SiR<sup>Me</sup>)-B, mais pas avec Zif12\*-[L<sub>2</sub>-Eu]-A.

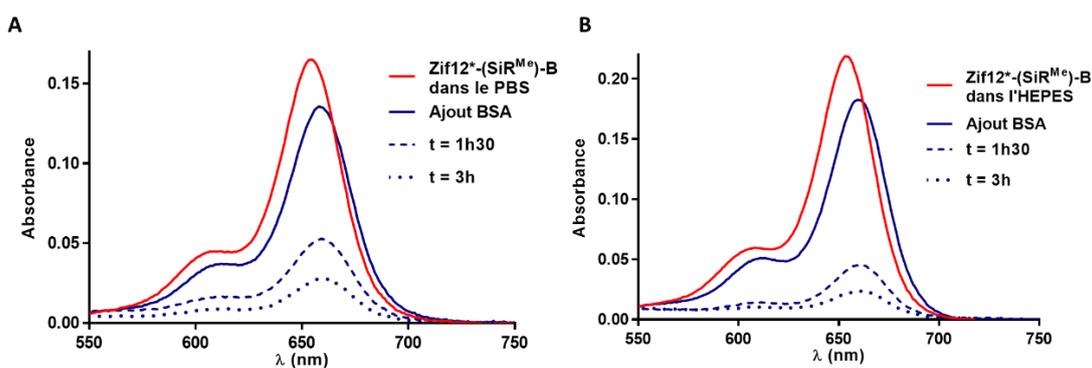


**Figure 159 :** Suivi dans le temps du spectre d'émission ( $\lambda_{ex}=315$  nm) de fluorescence des protéines Zif12\*-[L<sub>2</sub>-Eu]-A et Zif12\*-(SiR<sup>Me</sup>)-B dans le tampon C à une concentration de 250 nM à température ambiante.

Nous avons d'abord évalué la photostabilité de Zif12\*-(SiR<sup>Me</sup>)-B. Nous avons disposé la protéine dans le tampon C dans différentes conditions : (i) sous irradiation à 315 nm (ii) dans l'obscurité (iii) à la

lumière ambiante. Dans les trois cas, nous avons observé une diminution de l'émission de la SiR<sup>Me</sup> comparable, ce qui indique que le problème ne vient pas du photoblanchiment de la SiR<sup>Me</sup>.

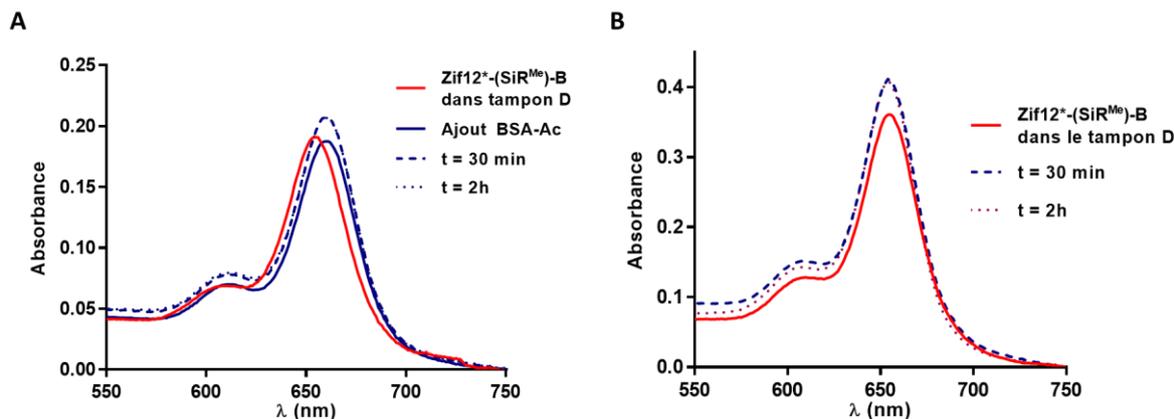
Ensuite nous avons testé l'effet de la BSA et du tampon. La BSA est souvent utilisée comme adjuvant dans les tampons pour empêcher la précipitation des protéines, la formation d'agrégats ou leur adhésion aux parois des contenants, notamment les cuves en quartz que nous utilisons pour la spectroscopie. Nous avons commencé par mettre en solution la protéine Zif12<sup>\*</sup>-(SiR<sup>Me</sup>)-B dans plusieurs solutions avec différents tampons et enregistrer le spectre d'absorbance (**Figure 160**). Nous avons utilisé plusieurs solutions tampons : le tampon C, HEPES et PBS. La SiR<sup>Me</sup> est sensible à la présence de thiol, or la BSA possède 35 cystéines, dont 34 sous la forme de pont disulfure, ce qui laisse une cystéine libre<sup>170</sup>. Nous avons donc enregistré un spectre d'absorbance de la protéine Zif12<sup>\*</sup>-(SiR<sup>Me</sup>)-B en solution dans l'HEPES (10 mM, pH 7.5) et dans le PBS (10 mM, pH 7.0) avant d'ajouter en excès de la BSA. À l'ajout de la protéine BSA (tampon HEPES ou PBS) on observe un déplacement du maximum d'absorbance suivi par la forte diminution de son absorbance après 1h30.



**Figure 160** : Spectre d'absorbance de la protéine Zif12<sup>\*</sup>-(SiR<sup>Me</sup>)-B dans différentes solutions tamponnées. A) Tampon PBS avec ajout de BSA, concentration en protéine de 1.4  $\mu$ M. B) Tampon HEPES avec ajout de BSA, concentration en protéine de 1.8  $\mu$ M.

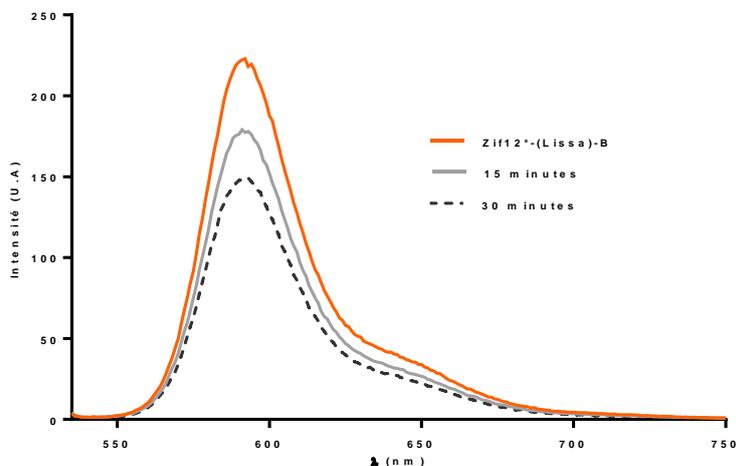
Nous avons par la suite comparé les absorbances de la protéine Zif12<sup>\*</sup>-(SiR<sup>Me</sup>)-B dans le tampon C' identique au tampon C où la BSA a été remplacée par de la BSA acétylée (BSA-Ac) dont la fonction thiol libre a été acétylée pour éviter les réactions parasites. Nous avons observé le même déplacement du maximum d'absorbance qu'en présence de la BSA, mais cette fois l'absorption ne diminue pas au cours du temps et reste stable. Ceci suggère que la cystéine de la BSA réagit avec la SiR<sup>Me</sup>. Nous avons enregistré un spectre d'absorption de la protéine dans l'eau avant d'ajouter du tampon D de même composition que le tampon C, mais sans BSA (**Figure 161**). Cette fois, aucun décalage du maximum n'est observé ni de perte d'absorption au cours du temps. Il semble donc que la SiR<sup>Me</sup> soit stable dans le tampon D, c'est-à-dire sans BSA. Il sera préférable d'utiliser le tampon D. Le décalage du maximum d'absorption observé avec la BSA acétylée peut être attribué à une interaction de la SiR<sup>Me</sup> avec une des poches de la BSA. Si l'on en revient à l'étude par gel-retard pour laquelle le tampon C avait été utilisé, la fixation de la BSA sur la SiR<sup>Me</sup> (par liaison covalente avec la cystéine ou

par complexation dans une de poches de la BSA) peut expliquer la perte d'affinité observée avec Zif12<sup>\*</sup>-(SiR<sup>Me</sup>)-B, plus importante que celle observée avec le complexe d'euporium.



**Figure 161** : Spectre d'absorbance de la protéine Zif12<sup>\*</sup>-(SiR<sup>Me</sup>)-B dans différentes solutions tamponnées. A) Tampon D avec ajout de BSA-Ac, concentration en protéine de 1.6 µM. B) Tampon D, concentration en protéine de 3 µM.

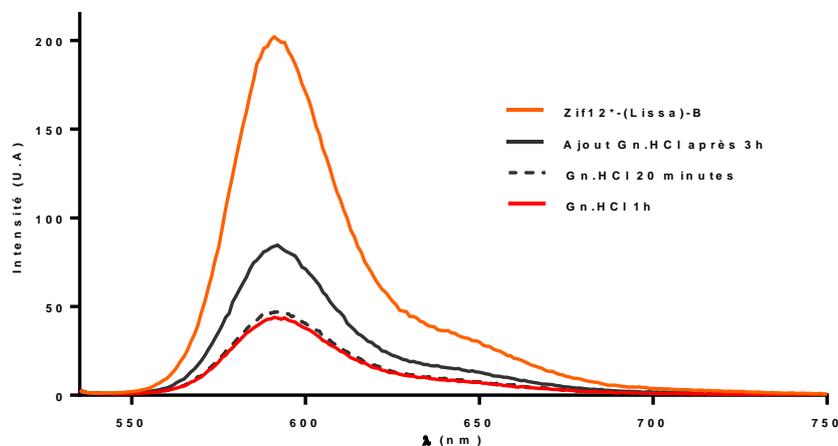
Nous avons ensuite étudié la stabilité de la protéine Zif12<sup>\*</sup>-(Lissa)-B. Nous n'avons pas vu dans la littérature l'évocation d'une possible réactivité de la lissamine avec les thiols. Nous avons testé cependant la stabilité de Zif12<sup>\*</sup>-(Lissa)-B dans le tampon C dans un premier temps. La **Figure 162** montre une perte d'absorption d'environ 35% en 30 minutes dans ces conditions.



**Figure 162** : Suivi par fluorescence de l'émission de Zif12<sup>\*</sup>-(Lissa)-B après excitation à 525 nm dans une solution de tampon C à concentration en protéine de 1.6 µM.

Dans le tampon D (sans BSA), le même comportement est observé, ce qui montre que le phénomène n'est pas lié aux thiols dans le cas de Zif12<sup>\*</sup>-(Lissa)-B. Les hypothèses les plus probables sont la précipitation de la protéine ou son adsorption sur les parois de la cuve mais nous n'avons pas testé la photostabilité de la Lissamine dans ces conditions. La solution étant relativement peu concentrée, la précipitation est difficile à identifier. L'ajout de chlorure de guanidinium dans la cuve semble

cependant ralentir voire stopper la perte d'absorbance (**Figure 163**), ce qui va dans le sens d'une précipitation puisque le guanidinium aide à solubiliser les protéines en les dénaturant toutefois.



**Figure 163** : Suivi par fluorescence de l'émission de Zif12\*(Lissa)-B après excitation à 525 nm dans une solution de tampon D (10%, v/v), concentration en protéine de 1.4 µM. Ajout de Gn.HCl après 3 heures puis mesure à 20 minutes et 1 heures.

Enfin, au cours de certaines expériences, nous avons observé l'adsorption des protéines sur les parois en quartz des cuves de fluorescence : après avoir laissé plusieurs minutes les protéines dans la cuve, celle-ci a été rincée plusieurs fois à l'eau et remplie d'eau. L'enregistrement du spectre d'émission montrait alors un signal correspondant aux différents luminophores non négligeable alors que la cuve n'est censée contenir que de l'eau, démontrant une adsorption des protéines sur les parois.

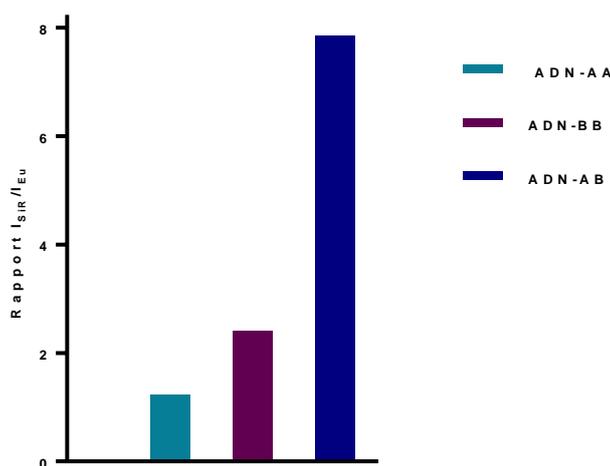
Pour récapituler, nous avons mis en évidence une interaction entre la BSA et la SiR<sup>Me</sup> : il y a probablement une réaction entre la cystéine libre de la BSA et SiR<sup>Me</sup>, mais possiblement aussi une complexation au sein d'une poche de la BSA puisqu'aucune interaction n'a été détectée avec la BSA acétylée. Ceci nous force à travailler dans un tampon sans BSA. Il est possible que le complexe d'euporium dont l'antenne est hydrophobe interagisse aussi avec une des poches de la BSA. Ces interactions avec la BSA pourraient expliquer la baisse d'affinité pour l'ADN observée par gel-retard. Il sera important d'évaluer l'affinité des protéines pour les ADN AB, AA et BB en réalisant des expériences de gel-retard en absence de BSA.

Nous avons également acquis la conviction qu'il n'était pas envisageable de mener des expériences comme un titrage où les protéines restent longtemps en solution sans ADN du fait de la précipitation et/ou adsorption des protéines sur les parois de la cuve en quartz. Il semble impératif de préparer les solutions dans des tubes eppendorf « low-binding » et de les transférer au dernier moment dans les cuves de fluorescence. Nous avons donc décidé d'évaluer par la suite la réponse de la sonde en préparant un mélange de protéines qui est rapidement ajouté à divers ADN pour voir si une réponse était observée et si l'ADN cible était sélectivement reconnu.

### 3 - Etude de la sélectivité avec le couple Zif12\*-[L<sub>2</sub>-Eu]-A / Zif12\*-(SiR<sup>Me</sup>)-B

Nous avons vu lors de la première partie de ce chapitre que notre système de deux protéines fonctionnalisées par notre système FRET est capable en présence de la séquence cible d'ADN-AB de former un dimère autour de l'ADN. Nous avons dans un premier temps observé le transfert d'énergie entre le complexe d'euporium et la SiR<sup>Me</sup>, nous devons maintenant nous assurer que la réponse que nous avons obtenue est due uniquement à la présence de l'ADN-AB. Nous avons donc réalisé des expériences de sélectivité avec diverses séquences d'ADN double brin et mesuré les rapports  $I_{SiR}/I_{Eu}$  obtenus. Les échantillons ont été préparés de la manière suivante. Des solutions comportant les différents duplexes d'ADN (AA, BB et AB), à une concentration de 500 nM (2 éq), dans le tampon D sont préalablement préparées. Une solution concentrée contenant les deux protéines Zif12\*-[L<sub>2</sub>-Eu]-A et Zif12\*-(SiR<sup>Me</sup>)-B est préparée en parallèle. Cette solution est ensuite ajoutée afin d'obtenir des quantités équivalentes des deux protéines dans chacune des expériences, la concentration finale de chacune des deux protéines est de 250 nM (1 éq). Des eppendorfs « low-binding » ont été utilisés pour préparer ces solutions. Les solutions sont laissées sous agitation pendant une heure à température ambiante avant d'être transférées dans une cuve de fluorescence micro (60 µL) pour enregistrer les spectres d'émission (avec une excitation à 315 nm). Les rapports d'intensité  $I_{SiR}/I_{Eu}$  obtenus après une heure d'incubation sont présentés **Figure 164**.

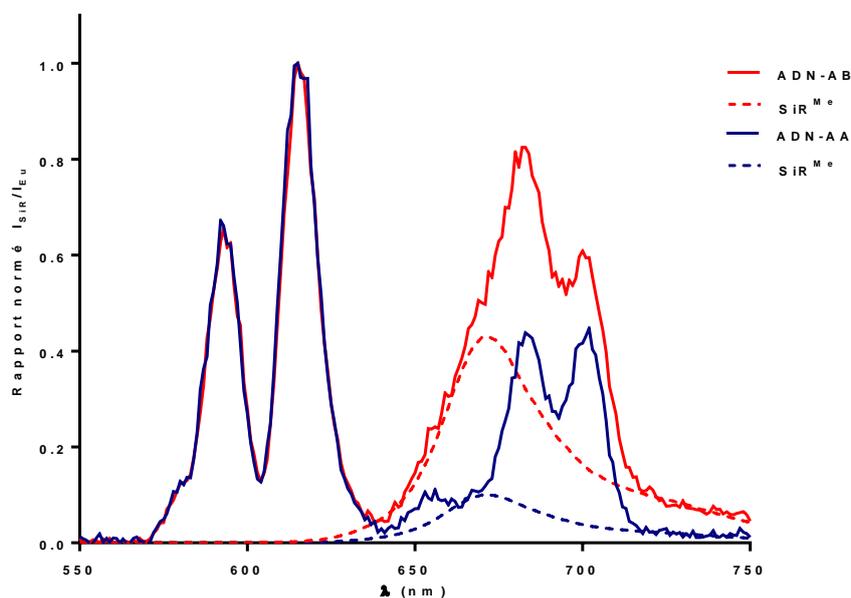
Le rapport  $I_{SiR}/I_{Eu}$  avec les séquences d'ADN-AA et ADN-BB est de 1.2 et 2.5 respectivement. Dans le cas de l'ADN-AB, il est proche de 8.3. On peut donc observer une discrimination de l'ADN-AB, cible attendue pour le système vis-à-vis des ADN-AA et BB.



**Figure 164** : Rapports d'émission de fluorescence  $I_{SiR}/I_{Eu}$  des protéines Zif12\*-[L<sub>2</sub>-Eu]-A et Zif12\*-(SiR<sup>Me</sup>)-B en présence de diverses séquences d'ADN double brin après une heure d'incubation.

Nous avons également enregistré les spectres d'émission en temps résolu avec un délai de 100 µs entre l'excitation et l'enregistrement du signal afin de filtrer les émissions à temps de vie court et d'éliminer ainsi l'émission de la SiR<sup>Me</sup> provenant de l'excitation directe du fluorophore. La **Figure 165** montre le

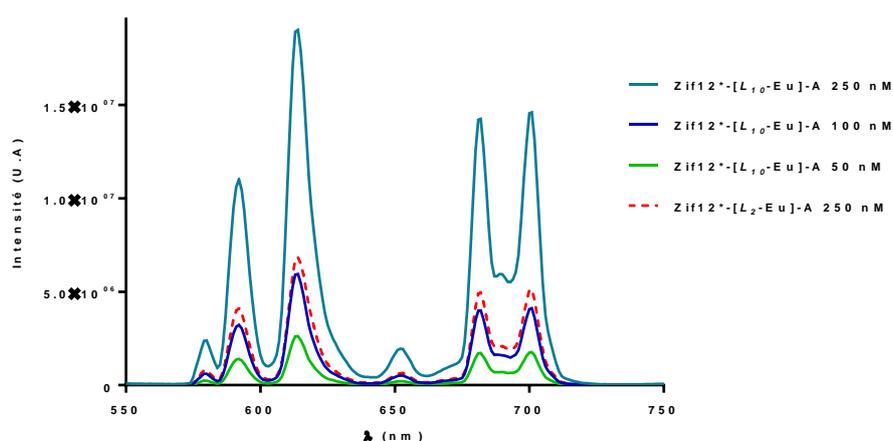
spectre d'émission du mélange Zif12\*-[L<sub>2</sub>-Eu]-A / Zif12\*-(SiR<sup>Me</sup>)-B / ADN-AB 1:1:1 (250 nM) enregistré. On peut clairement voir l'émission de l'Eu<sup>3+</sup> avec des raies fines caractéristiques des transitions <sup>5</sup>D<sub>0</sub> → <sup>7</sup>F<sub>J</sub> (J = 0, 1, 2, 3 et 4) avec, sous la bande J = 4, une émission large correspondant à l'émission de la SiR<sup>Me</sup>. Cela montre que le temps de vie d'émission de la SiR<sup>Me</sup> détecté est supérieur au délai de 100 μs. Un tel temps de vie ne peut résulter que d'un FRET entre l'Eu<sup>3+</sup> et le SiR<sup>Me</sup>. Le temps de vie de l'Eu<sup>3+</sup> est de 0.54 ms, très inférieur à celui de la protéine Zif12\*-[L<sub>2</sub>-Eu]-A seule en solution (1.07 ms). Ceci confirme donc le FRET provenant de Eu<sup>3+</sup> vers la SiR<sup>Me</sup>. À titre de comparaison, le spectre obtenu avec l'ADN-AA dans les mêmes conditions est représenté sur la **Figure 165**. Pour l'ADN-AA, on distingue également une contribution de la SiR<sup>Me</sup>, mais celle-ci est beaucoup plus faible qu'avec l'ADN-AB. Le temps de vie de Eu<sup>3+</sup> est de 0.84 ms, plus bas que celui de la protéine Zif12\*-[L<sub>2</sub>-Eu]-A seule, mais plus haut que celui mesuré avec ADN-AB. Pour l'ADN-BB, la situation est intermédiaire avec un peu plus de contribution de la SiR<sup>Me</sup> que l'ADN-AA, mais bien moins qu'avec l'ADN-AB et un temps de vie de Eu<sup>3+</sup> de 0.71 ms. Le fait d'observer du FRET pour l'ADN-AA et BB suggère que les protéines peuvent communiquer en solution, soit à cause de leur concentration élevée (250 nM), soit parce qu'elles interagissent de manière non-spécifique avec l'ADN-AA, mais moins fortement que dans le cas de l'interaction spécifique avec l'ADN-AB. Il serait intéressant de pouvoir abaisser la concentration pour voir si le FRET disparaît avec les ADN-AA et BB.



**Figure 165** : Spectres d'émission résolus en temps (délai 100 μs) normalisés des protéines Zif12-[L<sub>2</sub>-Eu]-A (250 nM) et Zif12\*-(SiR<sup>Me</sup>)-B (250 nM) en présence de ADN-AB (rouge) ou de ADN-AA (bleu) dans le tampon D. Le spectre d'émission de fluorescence de SiR<sup>Me</sup> (pointillé) est superposé dans chacun des cas.

#### 4 - Etude du couple Zif12\*-[L<sub>10</sub>-Eu]-A et Zif12\*-(SiR<sup>Me</sup>)-B

Nous avons vu dans le chapitre précédent que le complexe [L<sub>10</sub>-Eu] (conjugué sur TAT) présentait une brillance presque 5 fois plus élevée que [L<sub>2</sub>-Eu] en excitant à 315 nm. Son utilisation doit permettre de rééquilibrer les émissions de l'euporium et de la SiR<sup>Me</sup> dans un mélange 1:1 de protéines. La **Figure 166** montre les spectres d'émission de Zif12\*-[L<sub>2</sub>-Eu]-A et Zif12\*-[L<sub>10</sub>-Eu]-A dans le tampon D à différentes concentrations. On trouve finalement un rapport d'environ 2.5 entre les émissions des deux protéines excitées à 315 nm : l'intensité d'euporium pour Zif12\*-[L<sub>10</sub>-Eu]-A à 100 nM est quasi identique à celle de Zif12\*-[L<sub>2</sub>-Eu]-A à 250 nM. Pour la suite des études avec Zif12\*-[L<sub>10</sub>-Eu]-A nous avons travaillé avec une concentration de protéine fixée à 100 nM.

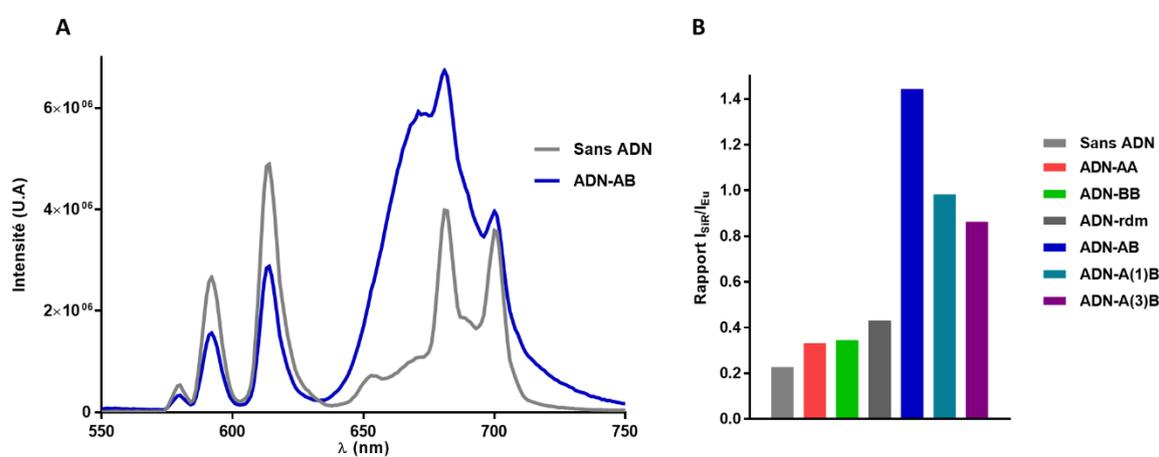


**Figure 166** : Comparaison des intensités d'émission de fluorescence de Eu<sup>3+</sup> des protéines Zif12\*-[L<sub>10</sub>-Eu]-A et Zif12\*-[L<sub>2</sub>-Eu]-A à différentes concentrations dans le tampon D ( $\lambda_{ex} = 315$  nm).

Nous avons réalisé une étude de sélectivité avec le système Zif12\*-[L<sub>10</sub>-Eu]-A / Zif12\*-(SiR<sup>Me</sup>)-B (100 nM) en présence de diverses séquences d'ADN double brin (200 nM). Par rapport à l'étude menée avec

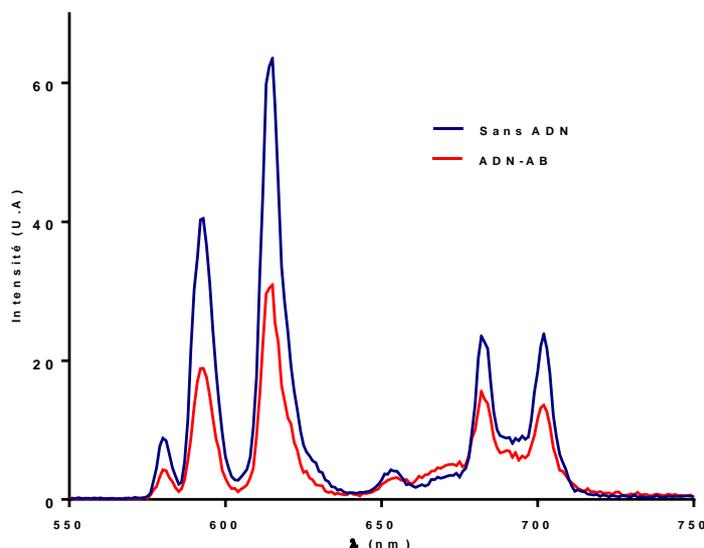
Zif12\*-[L<sub>2</sub>-Eu]-A, nous avons ajouté trois autres ADN : ADN-rdm, qui comprend exactement les mêmes paires de bases que le motif de reconnaissance AB (5'-TGGGCGCCCTTC-3'), mais mélangées pour qu'il ne soit pas reconnu ; ADN-A(1)B où une base est insérée entre les deux sextuplets reconnus par chacune des protéines (motif 5'-TGGGCGACCCTTC-3') et l'ADN-A(3)B (trois bases insérées) de motif 5'-TGGGCGATACCCTTC-3'. La **Figure 167 A** montre le spectre d'émission totale des deux protéines sans ADN et avec ADN-AB. On peut remarquer que par rapport à la **Figure 153** avec Zif12\*-[L<sub>2</sub>-Eu]-A, l'émission de l'euporium est plus importante en comparaison de la SiR<sup>Me</sup>. Pour une émission équivalente de Eu<sup>3+</sup> nous avons divisé la concentration de la protéine Zif12\*-(SiR<sup>Me</sup>)-B d'un facteur 2.5 grâce à l'utilisation du complexe [L<sub>10</sub>-Eu] d'où sa plus faible contribution au spectre d'émission total. Le rapport  $I_{SiR}/I_{Eu}$  en présence des différentes séquences d'ADN est représenté sur la **Figure 167 B**. On peut voir que le rapport  $I_{SiR}/I_{Eu}$  en absence d'ADN est de 0.22, pour les séquences d'ADN-AA et BB il est de 0.33 et 0.34 respectivement. Cela traduit un comportement similaire des deux

protéines en solution, avec une faible différence comparée à celui obtenu sans ADN. Nous pouvons en conclure qu'il n'y a aucun transfert d'énergie. Cette conclusion est identique avec l'ADN-rdm dont le rapport  $I_{SiR}/I_{Eu}$  est de 0.43. En présence de la cible d'ADN-AB ce rapport est de 1.45, ce qui représente une augmentation d'un facteur 7 vis-à-vis du rapport sans ADN. Cette augmentation traduit donc un transfert du complexe d' $Eu^{3+}$  vers la  $SiR^{Me}$ . L'interaction des deux protéines avec les ADN comportant une et trois insertions de bases présente des rapports  $I_{SiR}/I_{Eu}$  respectifs de 0.98 et 0.86. De tels rapports indiquent qu'un transfert d'énergie a lieu également. Il est cependant moins efficace qu'en présence de l'ADN-AB. Ces insertions de bases peuvent éloigner le complexe d'euporium de la  $SiR^{Me}$ , expliquant une diminution des rapports lorsque l'on insère une ou plusieurs bases entre les deux sextuplets.



**Figure 167 :** A) Spectres d'émission de fluorescence des protéines  $Zif12^*-[L_{10}-Eu]-A$  et  $Zif12^*-(SiR^{Me})-B$  avec et sans ADN-AB. B) Rapports bruts d'émission de fluorescence normalisés  $I_{SiR}/I_{Eu}$  des protéines  $Zif12^*-[L_{10}-Eu]-A$  et  $Zif12^*-(SiR^{Me})-B$  en présence de diverses séquences d'ADN double brin après une heure de réaction.

Nous avons également enregistré le spectre d'émission en temps résolu (délai de 100  $\mu s$ ) de l'interaction des protéines  $Zif12^*-[L_{10}-Eu]-A$  et  $Zif12^*-(SiR^{Me})-B$  avec les ADN-AB et rdm. Sur la **Figure 168**, en absence d'ADN, on observe très nettement l'ensemble des transitions  $^5D_0 \rightarrow ^7F_J$  ( $J = 0, 1, 2, 3$  et  $4$ ) de l' $Eu^{3+}$ , avec une faible émission de la  $SiR^{Me}$ . Les protéines sont donc capables de communiquer en solution avec établissement d'un transfert d'énergie mais semble-t-il peu efficace. Le temps de vie de l' $Eu^{3+}$  dans ce mélange est de 0.92 ms, légèrement inférieur à celui de la protéine  $Zif12^*-[L_{10}-Eu]-A$  seule en solution (0.96 ms), en accord avec un transfert d'énergie de  $Eu^{3+}$  vers la  $SiR^{Me}$  peu efficace. En présence de la séquence cible d'ADN-AB on observe une forte diminution de l'émission de l' $Eu^{3+}$  d'environ 50% ainsi qu'un temps de vie qui diminue un peu plus, mesuré à 0.82 ms. Cependant il n'y a qu'une faible augmentation de l'émission de la  $SiR^{Me}$  en temps résolu, en comparaison avec celle enregistrée sans ADN. Ceci va dans le sens d'un transfert d'énergie peu efficace de l' $Eu^{3+}$  vers la  $SiR^{Me}$ , malgré la distance courte (environ 17-20 Å) entre les deux luminophores au sein du dimère.



**Figure 168** : Spectres d'émission résolus en temps (délai 100  $\mu$ s) normalisés des protéines Zif12-[L<sub>10</sub>-Eu]-A (100 nM) et Zif12\*(SiR<sup>Me</sup>)-B (100 nM) en absence (bleu) et présence d'ADN-AB (200 nM) dans le tampon D.

Que ce soit pour la protéine Zif12\*-[L<sub>10</sub>-Eu]-A ou la protéine Zif12\*-[L<sub>2</sub>-Eu]-A nous avons observé la diminution de l'intensité d'émission de Eu<sup>3+</sup> au profit d'une forte augmentation de la SiR<sup>Me</sup> d'un facteur 6 à 7 lors de l'enregistrement des spectres d'émission totale (excitation continue, sans délai entre l'excitation et la détection). Cependant cette augmentation ne se vérifie pas avec les spectres obtenus en temps résolu. Dans le cas de la protéine Zif12\*-[L<sub>2</sub>-Eu]-A, l'augmentation de la SiR<sup>Me</sup> est cependant plus intense que pour la protéine Zif12\*-[L<sub>10</sub>-Eu]-A dont le spectre en temps résolu ne présente qu'une très faible augmentation de la SiR<sup>Me</sup>. Ce transfert d'énergie représente une voie de désexcitation de l'euporium qui devrait fortement diminuer son temps de vie, ce qui n'est pas le cas, une diminution de seulement 0.2 ms est observée (de 0.96 à 0.82 ms).

Cela pose de réelles questions quant à la sensibilisation de la SiR<sup>Me</sup>. Au vu de l'ensemble des expériences que nous avons menées, nous pouvons avancer plusieurs hypothèses pour expliquer ces résultats (**Figure 169**). La voie de sensibilisation notée 1, est celle attendue avec un transfert depuis le niveau <sup>5</sup>D<sub>0</sub> de l'euporium. Ce niveau d'énergie possède un temps de vie de l'ordre de la milliseconde ce qui permet d'observer ce transfert en temps résolu, dont le délai d'acquisition est de 100  $\mu$ s. Le temps de vie de l'Eu<sup>3+</sup> (qui est mesuré à partir du niveau <sup>5</sup>D<sub>0</sub>) est directement affecté car c'est une voie de désexcitation. L'émission de la SiR<sup>Me</sup> que nous avons observée en temps résolu est donc celle provenant de cette voie. La seconde voie de sensibilisation possible est celle qui provient du niveau <sup>5</sup>D<sub>1</sub> de l'euporium, celui-ci a un temps de vie plus court, généralement autour de la  $\mu$ s. Par cette voie là nous observons un transfert d'énergie vers la SiR<sup>Me</sup> lors de l'enregistrement du spectre total, mais pas en temps résolu, le temps de vie du niveau <sup>5</sup>D<sub>1</sub> étant beaucoup plus court que les 100  $\mu$ s de délai. Enfin la dernière voie pour ce transfert d'énergie est celle provenant du niveau ILCT de l'antenne. Le transfert d'énergie de l'état ILCT aux niveaux excités de l'Eu<sup>3+</sup> est très rapide, de type Dexter (par les orbitales). Un transfert de type

Förster à travers l'espace ne peut pas rentrer en compétition avec ce transfert ILCT vers  $^5D_1$ . Avoir le transfert ILCT vers SiR<sup>Me</sup>, cela sous-entend que la SiR<sup>Me</sup> vienne au contact de l'antenne pour qu'un transfert de type Dexter ait lieu. Un tel transfert ILCT (antenne) vers le  $S_1$  (SiR<sup>Me</sup>) peut se produire à l'échelle de temps de la nanoseconde et peut expliquer l'augmentation de l'émission de la SiR<sup>Me</sup> sur le spectre d'émission total mais pas en temps résolu.

Il est logique de penser que ces trois voies de sensibilisation sont présentes avec des contributions plus ou moins importantes. L'enregistrement du temps de vie n'est pas suffisant pour affirmer la provenance du transfert d'énergie puisque c'est le temps de vie global de  $\text{Eu}^{3+}$  qui est mesuré. Il est possible que l'on ait plusieurs population d' $\text{Eu}^{3+}$ , une qui ne transfère pas ou peu d'énergie à la SiR<sup>Me</sup> avec un temps de vie long et une autre avec un transfert efficace, avec un temps de vie très court (de l'ordre de la centaine de  $\mu\text{s}$ ). En temps résolu nous ne voyons que la population à temps long, qui peut par exemple correspondre à des protéines non fixées à l'ADN ou faiblement associées à l'ADN (car pas positionnées ailleurs que sur le site d'interaction). Le passage d'une voie par rapport à une autre doit reposer sur la vitesse de transfert des différents processus. Si le transfert depuis le niveau ILCT est plus rapide que celui provenant du niveau  $^5D_0$  alors c'est cette voie qui va contribuer le plus pour la sensibilisation de la SiR<sup>Me</sup>. En passant du complexe  $[\text{L}_2\text{-Eu}]$  à  $[\text{L}_{10}\text{-Eu}]$  on n'observe pas les mêmes augmentations du signal en temps résolu ce qui sous-entend que la contribution de chacune des voies est différente et que l'antenne intrinsèquement modifie la vitesse de transfert de ces trois voies. Afin de résoudre cette question il faudrait mesurer le temps de vie de l'état excité de la SiR<sup>Me</sup> : selon qu'il appartienne à la nano, micro ou milliseconde alors nous pourrions être plus affirmatif sur la voie de sensibilisation.

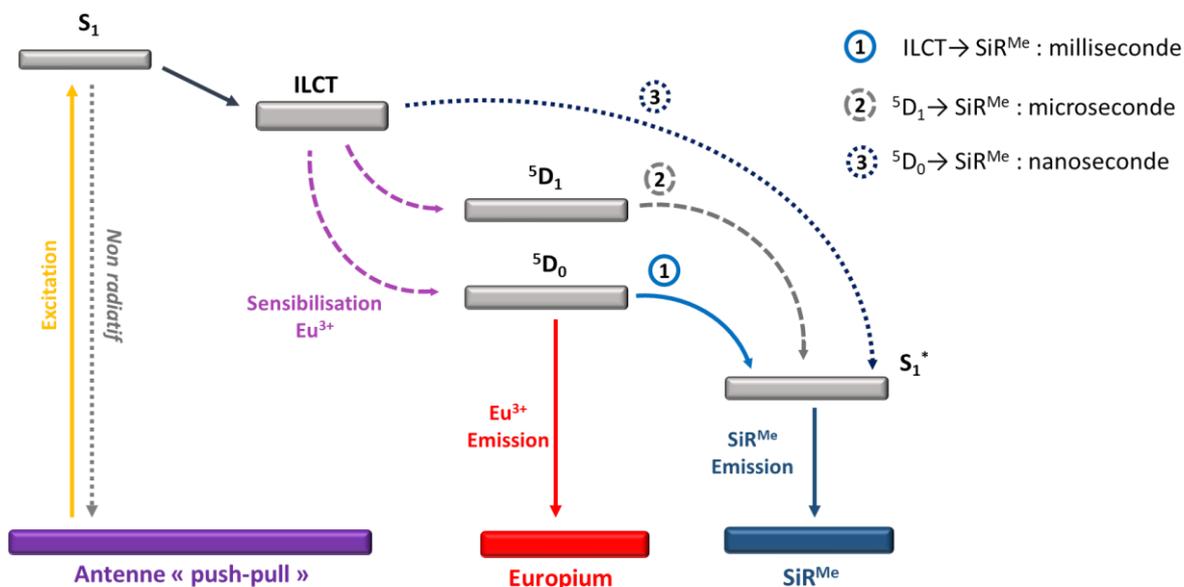
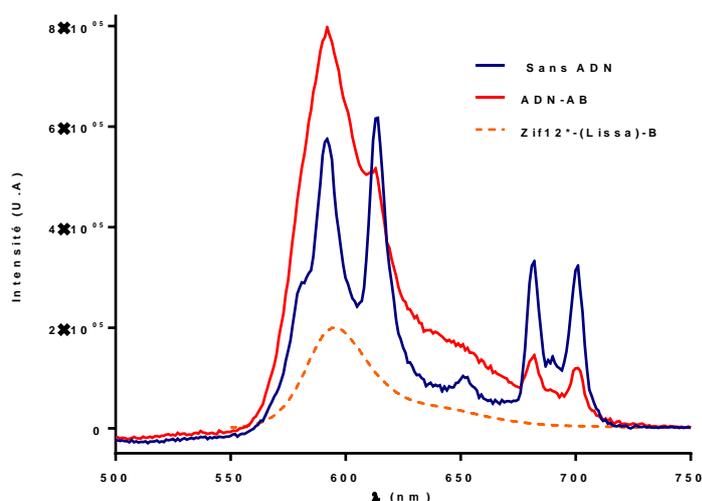


Figure 169 : Diagramme simplifié présentant les voies de sensibilisation possibles de la SiR<sup>Me</sup>.

## 5 - Etude du couple Zif12\*-[L<sub>10</sub>-Eu]-A et Zif12\*-(Lissa)-B

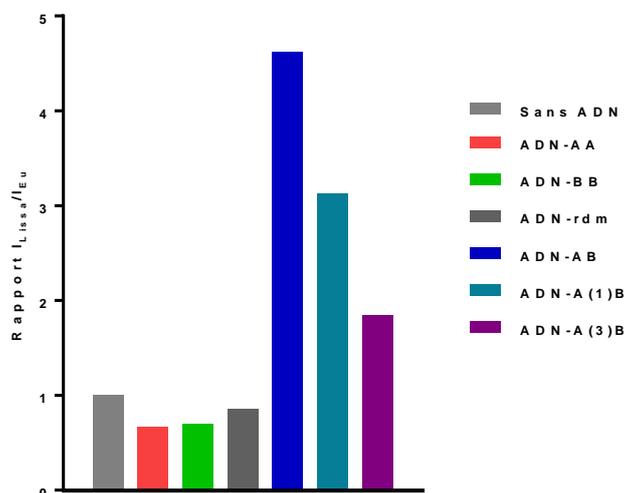
Pour terminer, nous avons évalué la réponse du couple de protéine Zif12\*-[L<sub>10</sub>-Eu]-A / Zif12\*-(Lissa)-B en présence de divers ADN. La lissamine présente un spectre d'absorbance qui recouvre très peu avec l'émission d'Eu<sup>3+</sup>, au niveau des transitions  $^5D_0 \rightarrow ^7F_{0/1}$ . Le transfert d'énergie doit donc être moins efficace qu'avec la SiR<sup>Me</sup>, mais il reste théoriquement possible. Sur la base du rendement quantique déterminé pour TAT-[L<sub>10</sub>-Eu] (11%), on peut estimer le rayon de Förster R<sub>0</sub> à 32 Å pour le couple Zif12\*-[L<sub>10</sub>-Eu]-A / Zif12\*-(Lissa)-B contre 45 Å pour le couple Zif12\*-[L<sub>10</sub>-Eu]-A / Zif12\*-(SiR<sup>Me</sup>)-B. Une distance de 32 Å entre les luminophores est supérieur à ce qui est attendu si le complexe ternaire se forme. Nous avons donc préparé, comme précédemment, plusieurs solutions contenant diverses séquences d'ADN (AA, BB, rdm et AB) auxquelles a été ajoutée une solution concentrée de la protéine Zif12\*-[L<sub>10</sub>-Eu]-A et la protéine Zif12\*-(Lissa)-B. La concentration de chaque protéine est de 100 nM et la concentration en ADN est de 200 nM. Le spectre d'émission total sans ADN et pour la séquence d'ADN-AB sont présentés **Figure 170**. En absence d'ADN on observe à la fois l'émission de l'Eu<sup>3+</sup> et l'émission directe de la lissamine, qui pour rappel, peut également être excitée à 315 nm. En présence de la séquence cible d'ADN-AB on observe une forte émission de la lissamine et une diminution de l'intensité d'émission de l'euporium. Il y a donc un transfert d'énergie de l'Eu<sup>3+</sup> vers la Lissamine.



**Figure 170** : Spectres d'émission total des protéines Zif12\*-[L<sub>10</sub>-Eu]-A et Zif12\*-(Lissa)-B après excitation à  $\lambda_{ex} = 315$  nm. La concentration des protéines est de 100 nM, la concentration d'ADN est de 200 nM.

Nous avons vérifié que cette réponse est spécifique de la séquence d'ADN-AB en calculant le rapport  $I_{Lissa}/I_{Eu}$  (**Figure 171**). Compte tenu de l'émission de la lissamine qui couvre l'émission des transitions  $^5D_0 \rightarrow ^7F_J$  ( $J = 1, 2$ ) de l'euporium nous avons intégré le signal de l'euporium à 700 nm ( $\pm 2$  nm) et le signal de la lissamine a été intégré dans une zone où l'Eu<sup>3+</sup> n'émet pas, à 583 nm ( $\pm 2$  nm). Pour les séquences d'ADN-AA, BB et rdm ce rapport est très légèrement inférieur à celui obtenu en absence d'ADN. Il n'y a donc aucun transfert d'énergie. En présence de la séquence cible d'ADN-AB

ce rapport atteint une valeur de 4.6. Cette sonde basée sur le couple Zif12\*-[L<sub>10</sub>-Eu]-A et la protéine Zif12\*-(Lissa)-B est donc également capable de détecter la séquence cible, mais répond un peu moins bien, comme précédemment avec les ADN présentant une ou trois bases insérées au milieu de la séquence cible.



**Figure 171** : Rapports bruts d'émission de fluorescence normalisés  $I_{Lissa}/I_{Eu}$  des protéines Zif12\*-[L<sub>10</sub>-Eu]-A et Zif12\*-(Lissa)-B en présence de diverses séquences d'ADN double brin après une heure de réaction. Mesures effectuées sans délai d'acquisition.

### III - Conclusion

Tout au long de ce chapitre, nous nous sommes intéressés à l'étude de l'interaction des protéines Zif12 fonctionnalisées par notre système FRET. Un premier couple de protéine Zif12\*-[L<sub>2</sub>-Tb]-A et Zif12\*-(Lissa)-B devait nous permettre d'étudier le transfert d'énergie à partir d'un complexe avec une brillance élevée. Cependant la présence d'une tyrosine proche du complexe de terbium n'a pas permis l'utilisation de la protéine Zif12\*-[L<sub>2</sub>-Tb]-A. En utilisant le couple Zif12\*-[L<sub>2</sub>-Eu]-A et Zif12\*-(SiR<sup>Mc</sup>)-B nous avons pu mettre en évidence une augmentation du rapport d'émission  $I_{SiR}/I_{Eu}$  lors d'un titrage avec la séquence d'ADN-AB. La formation du dimère autour de la séquence d'ADN-AB a ensuite été confirmé par gel-retard et le système Zif12\*-[L<sub>2</sub>-Eu]-A / Zif12\*-(SiR<sup>Mc</sup>)-B / ADN-AB a une constante de dissociation de  $(5.7 \pm 1.2) \times 10^{-15} \text{ M}^2$  et une valeur de C<sub>50</sub> de 75 nM. Nous avons pu grâce à notre système de deux protéines fonctionnalisées discriminer les ADN-AA, BB et rdm dont les ratios  $I_{Fluo}/I_{Eu}$  varient peu par rapport à celui obtenu avec les deux protéines en solution sans ADN, indiquant qu'aucun transfert d'énergie n'a lieu. L'augmentation de ce rapport pouvant aller jusqu'à un facteur 7 est une preuve du transfert FRET induit par la reconnaissance de la cible. Lorsque que nous avons étudié l'interaction de ce système avec les séquences d'ADN-A(1)B et A(3)B nous avons obtenu des ratios  $I_{SiR}/I_{Eu}$  qui laissent à penser qu'un transfert FRET a lieu. Toutefois, plus l'insertion entre les deux sextuplets est grand et plus le ratio diminue ( $I_{Fluo}/I_{Eu}$  : AB > A(1)B > A(3)B). Le positionnement de notre système FRET permet d'avoir un transfert efficace, mais ne permet pas de distinguer l'ADN-AB de ceux ayant une/des insertions de bases. En éloignant les deux partenaires du

système FRET l'efficacité du transfert sera plus faible, mais cela pourrait grandement diminuer le transfert observé avec les ADN-A(1)B et A(3)B. Distinguer l'ADN-AB de l'ADN-A(1)B serait là une nouvelle preuve de sélectivité.

Durant l'étude spectroscopique de notre système, nous avons été confrontés à un problème de reproductibilité. Le ratio  $I_{SiR}/I_{Eu}$  initial et en excès d'ADN ne coïncide pas d'une expérience à une autre bien que les solutions utilisées soient identiques. Ceci nous a amené à étudier la stabilité de nos protéines et plus particulièrement la protéine Zif12<sup>\*</sup>-(SiR<sup>Me</sup>)-B. Nous nous sommes rendu compte que l'émission de la protéine Zif12<sup>\*</sup>-(SiR<sup>Me</sup>)-B chute au cours du temps. Après avoir testé plusieurs paramètres (nature et composition des tampons) nous avons constaté que la présence de BSA pouvait être la cause de cette perte d'émission en formant des interactions avec la SiR<sup>Me</sup>. La présence de BSA dans le tampon utilisé pour les expériences de gels-retards pourrait donc influencer la formation du dimère, il est donc nécessaire de réitérer ces expériences dans le tampon D et de déterminer de nouvelles valeurs de  $K_d$  et  $C_{50}$ . Nous sommes en droit de penser que la valeur de  $C_{50}$  sera inférieure à celle mesurée à 75 nM, ce qui pourrait permettre de nous rapprocher de celui obtenu pour les protéines non-décorées calculées à 4.9 nM. Au-delà de ce problème lié à la BSA, nous avons remarqué une adsorption de nos protéines sur les parois de la cuve en quartz. Cela présente un biais qui pourrait expliquer les différences d'intensités du complexe d'euporium et de l'accepteur observées dans nos expériences. L'ajout au dernier moment de la solution à analyser dans la cuve en quartz nous permet d'essayer de minimiser ce phénomène. Une modification du tampon avec l'ajout d'un agent pouvant empêcher cette adsorption est à envisager, sous réserve qu'il ne présente aucune influence sur les émissions des partenaires FRET. Une autre solution serait de s'affranchir des cuves en quartz et d'utiliser des micro-plaques traitées « low-binding » pour les protéines et un lecteur de micro-plaques avec détection de fluorescence. Quant au transfert d'énergie du complexe vers la SiR<sup>Me</sup> celui-ci n'est effectif qu'en présence de la cible d'ADN-AB avec une augmentation significative du rapport  $I_{Fluo}/I_{Eu}$ . Cependant la sensibilisation de la SiR<sup>Me</sup> pose de nombreuses questions quant à son origine, des niveaux  $^5D_1$  ou  $^5D_0$  de  $Eu^{3+}$  ou encore de l'antenne « push-pull ». La mesure du temps de vie de la SiR<sup>Me</sup> pourra nous permettre de faire cette distinction selon le temps de vie que nous obtiendrons. Toutefois l'excitation directe de la SiR<sup>Me</sup> est un problème pour la mesure du temps de vie et nécessite l'utilisation d'un autre complexe d'euporium.



# Conclusion et perspectives

## Sommaire

Conclusion.....	165
Perspectives.....	167
Fonctionnalisation des protéines .....	167
Le complexe d'euporium.....	169
Interaction des protéines décorées avec l'ADN .....	169
Interaction protéines / ADN .....	169
Transfert FRET.....	170

## Conclusion

Le travail de cette thèse avait pour but de développer une sonde luminescente pour la reconnaissance spécifique d'ADN double brin, la réponse luminescente devant être apportée par le transfert d'énergie de type FRET entre un complexe d'euporium (donneur) et un chromophore accepteur. La reconnaissance de l'ADN double brin est rendue possible grâce à l'utilisation de deux protéines à doigts de zinc capables de former un dimère autour de sa séquence cible d'ADN. Chaque protéine comporte deux doigts de zinc qui permet de cibler une séquence de 6 paires de bases et l'association de ces deux protéines sur l'ADN nous permet donc de cibler une séquence d'ADN double brin longue de 12 paires de bases. Pour établir un système luminescent basé sur un FRET, une protéine est fonctionnalisée par le complexe d'euporium donneur et la seconde protéine par un fluorophore organique accepteur. En nous inspirant de la protéine Zif12 de Wang et Pabo (noté Zif12-A) qui ciblait la séquence 5'-TGGGCG-3' nous avons élaboré une nouvelle protéine, notée Zif12-B, où seuls les domaines de reconnaissance des doigts de zinc ont été modifiés pour cibler la séquence 5'-GAAGGG-3'. Compte tenu de la longueur de la protéine (72 acides aminés) c'est une méthode d'assemblage par ligation chimique native de trois fragments peptidiques, synthétisés chimiquement, qui a été utilisée. Le découpage comprend un premier fragment F1 correspondant à l'extension N-terminale, qui permet la formation du dimère autour de l'ADN, et les deux derniers fragments, F2 et F3, qui comportent chacun un doigt de zinc. Une stratégie de synthèse divergente, avec d'une part la synthèse de la partie reconnaissance d'ADN avec le fragment F2-F3 et d'autre part l'extension N-terminale (fragment F1) qui comporte le système FRET a été élaborée. L'optimisation des différentes réactions de ligation chimique native et des groupements protecteurs utilisés nous a permis d'avoir une synthèse efficace dans un temps raisonnable et relativement propre, en utilisant deux méthodes de ligation chimique natives différentes. Une première ligation par la méthode des hydrazides permet une synthèse très rapide de la partie reconnaissance d'ADN et la méthode de ligation SEA permet d'assembler le dernier fragment F1. Cette méthode est plus lente et plus douce, mais elle est compatible avec les luminophores contrairement à la méthode des hydrazides. Grâce à la synthèse de deux protéines Zif12, ne comportant aucun fluorophore, nous avons pu mettre en évidence, par une technique d'électrophorèse (gel-retard) la formation d'un hétérodimère autour de la séquence d'ADN-AB avec les protéines Zif12\*-Nle-A et Zif12\*-B. La constante de dissociation a été calculée ( $K_d = 2.4 \times 10^{-17} \text{ M}^2$ ) ainsi que la concentration en protéine conduisant à la formation de 50% de complexe ( $C_{50} = 4.9 \text{ nM}$  ( $\pm 0.5 \text{ nM}$ )). La validation du système de reconnaissance d'une séquence d'ADN double brin longue de 12 paires de bases nous a permis d'aller un cran plus loin que les précédents travaux réalisés par Guillaume Frémy portant sur le ciblage d'une séquence palindromique d'ADN double brin de 12 paires de bases.

La synthèse de la protéine Zif12 fonctionnalisée par le complexe d'euporium [ $L_I^{CC}$ -Eu], retenu au début de cette thèse, comprend un luminophore qui possède un rendement quantique dans l'eau de 17% avec un maximum d'excitation à 339 nm. Malheureusement, la synthèse n'a pas abouti car la triple liaison contenue dans le ligand s'est avérée sensible aux groupements thiols ou au TCEP utilisé lors de la dernière étape d'assemblage de la protéine. Nous avons donc cherché un complexe d'euporium bioconjuguable dont les propriétés s'approchent de celles du complexe [ $L_I^{CC}$ -Eu]. Un premier complexe [ $L_2$ -Eu] a été rapidement synthétisé ce qui a permis de nous assurer d'une brillance suffisante pour être greffée sur la protéine Zif12-A bien que nous soyons loin des caractéristiques du complexe [ $L_I^{CC}$ -Eu] ( $\lambda_{ex}L_2 = 315$  nm et  $\Phi_{Eu}L_2 = 4\%$ ). Toutefois nous avons poursuivi nos efforts d'optimisation avec deux nouveaux ligands  $L_3$  et  $L_4$  dont la partie donneuse de l'antenne « push-pull » est basée sur l'utilisation d'un carbazole. Malheureusement, ces deux complexes en milieu aqueux ne permettent pas d'obtenir une luminescence suffisante de l'euporium, avec des rendements quantiques inférieur à 2%. Nous nous sommes donc tournés vers la synthèse d'une série de complexes qui reposent sur un groupement donneur azoté de type  $NH_2$ ,  $NHR$  ou  $NR_2$ . Plusieurs complexes [ $L_5$ -Eu] à [ $L_8$ -Eu], ont présenté une faible émission de l'euporium due à un transfert d'électron photo-induit, au détriment du transfert d'énergie attendu, ce qui tue la luminescence du lanthanide. Nous avons finalement synthétisé les complexes [ $L_9$ -Eu] et [ $L_{10}$ -Eu] comportant un groupement électro-donneur acétanilide. C'est le complexe [ $L_{10}$ -Eu] qui présente les meilleures caractéristiques spectroscopiques avec un rendement quantique dans l'eau de 11% et une excitation maximale à 315 nm. Nous sommes encore loin des caractéristiques du complexe [ $L_I^{CC}$ -Eu], mais celui-ci sera notre meilleur complexe. Malgré la synthèse de nombreux complexes ne pouvant être utilisés dans ce projet de thèse, ils ont pu être utilisés dans d'autres projets de l'équipe, notamment pour l'imagerie cellulaire. Ils feront l'objet de publications dans lesquelles je serai associé, et pourront initier le développement d'autres sondes luminescentes.

Nous avons cependant réussi à synthétiser deux protéines comportant un complexe de lanthanide Zif12\*-[ $L_i$ -Eu] -A ( $i = 2, 10$ ) et une protéine comportant notre chromophore accepteur Zif12\*-(SiR<sup>Me</sup>)-B. L'étude de l'interaction des protéines Zif12\*-[ $L_2$ -Eu] -A et Zif12\*-(SiR<sup>Me</sup>)-B avec la séquence cible d'ADN-AB a montré la formation d'un dimère autour du duplex d'ADN mis en évidence par électrophorèse (EMSA). Dans les mêmes conditions que pour les protéines non-fonctionnalisées nous avons une perte d'affinité ( $K_d = (5.7 \pm 1.2) \times 10^{-15}$  M<sup>2</sup>,  $C_{50} = 75$  nM ( $\pm 8$ nM)), elle reste de l'ordre des dizaines de nanomolaire. Toutefois nous avons mis en évidence une interaction de la protéine Zif12\*-(SiR<sup>Me</sup>)-B avec la BSA qui peut sans doute affecter la liaison de la protéine à l'ADN. L'affinité réelle, sans BSA, est peut-être plus forte mais ceci doit être vérifié. Cette affinité pour une séquence aussi longue se démarque largement de ce que nous avons pu observer dans la littérature vis-à-vis d'autres systèmes de sonde luminescente pour l'ADN double brin.

La reconnaissance de la séquence cible a ensuite été étudiée par spectroscopie et la preuve d'un transfert d'énergie provenant de l' $\text{Eu}^{3+}$  vers la  $\text{SiR}^{\text{Me}}$  a été mise en évidence avec un rapport  $I_{\text{SiR}}/I_{\text{Eu}}$  qui augmente d'un facteur 7. En présence d'autres séquences d'ADN non cible (ADN-AA, BB et rdm), ce rapport n'évolue presque pas notablement, validant ainsi la sélectivité de la reconnaissance et du transfert d'énergie FRET. L'enregistrement de spectres en temps résolu a montré une émission croissante de l'émission de la  $\text{SiR}^{\text{Me}}$ , preuve d'un transfert d'énergie depuis un état excité ayant un temps de vie long, c'est-à-dire le niveau  $^5\text{D}_0$  de l'euporium. L'utilisation de la protéine  $\text{Zif12}^*-[L_{10}\text{-Eu}]\text{-A}$  laisse à penser que nous aurons un transfert d'énergie plus efficace, compte tenu des propriétés photophysiques du complexe  $[L_{10}\text{-Eu}]$ . Cependant l'utilisation de  $\text{Zif12}^*-[L_{10}\text{-Eu}]\text{-A}$  va nous permettre de diminuer la concentration et ainsi diminuer les interactions non-spécifique. Malgré une augmentation du rapport  $I_{\text{SiR}}/I_{\text{Eu}}$  d'un facteur 6, l'enregistrement des spectres d'émission résolus en temps n'a montré qu'une faible augmentation de la  $\text{SiR}^{\text{Me}}$ . Les différents temps de vie que nous avons mesurés pour l' $\text{Eu}^{3+}$ , que ce soit pour  $[L_2\text{-Eu}]$  ou  $[L_{10}\text{-Eu}]$  n'ont que faiblement diminué, ce qui sous-entend que la forte augmentation du signal de la  $\text{SiR}^{\text{Me}}$  ne provient pas seulement d'un transfert depuis le niveau  $^5\text{D}_0$  de l'euporium. Pour expliquer ce phénomène, d'autres voies de sensibilisation sont à envisager, notamment depuis le niveau  $^5\text{D}_1$  de l' $\text{Eu}^{3+}$  ou encore un transfert de type Dexter depuis l'antenne vers la  $\text{SiR}^{\text{Me}}$ . Le moyen le plus efficace de répondre à cette question est de mesurer le temps de vie de la  $\text{SiR}^{\text{Me}}$ , qui selon l'ordre de grandeur (nano, micro, milliseconde) nous permettra de discriminer selon les voies de sensibilisation. Ces expériences pourront être réalisées en collaboration avec Nathan Mc.Clenaghan, mais nous devons pour cela utiliser un autre complexe de lanthanide. Nous pouvons envisager de remplacer l'euporium par un lanthanide non luminescent tel que le gadolinium, si on enregistre une augmentation de la  $\text{SiR}^{\text{Me}}$  cela nous indiquera que l'état donneur pour le FRET est le niveau ILCT de l'antenne.

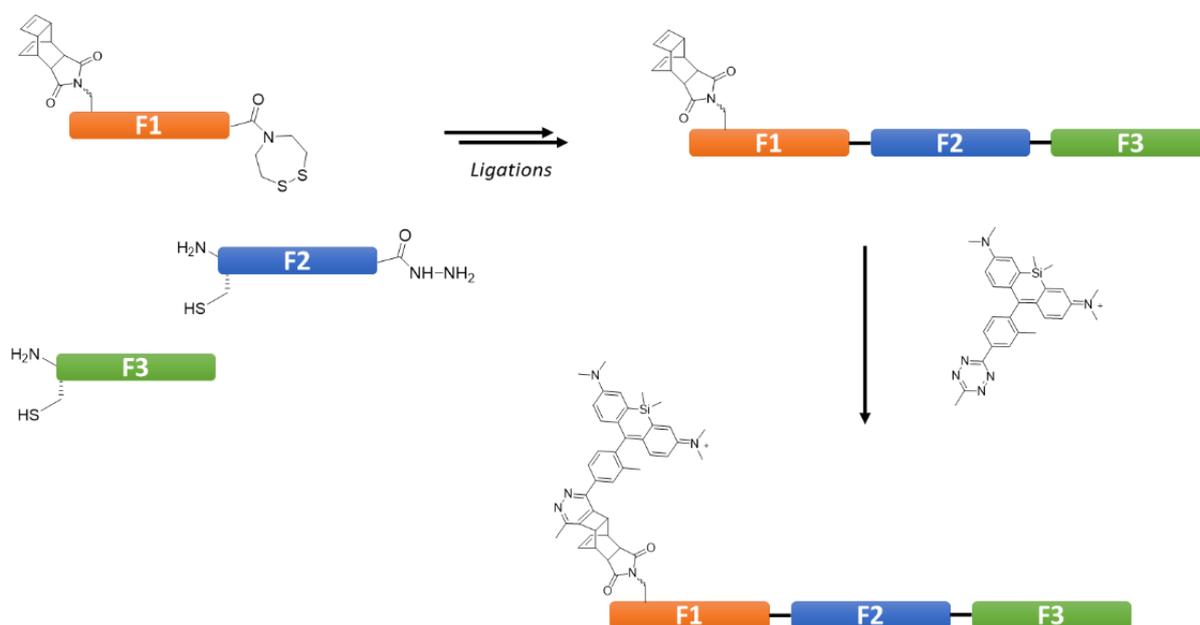
En résumé nous avons été capables d'élaborer un système de deux protéines comportant chacune deux doigts de zinc et une extension N-terminale capable de former sélectivement un dimère autour de la séquence d'ADN ciblé. La fonctionnalisation de la protéine par un système de FRET entre un complexe de lanthanide et un chromophore a permis d'avoir une réponse luminescente de la formation de ce dimère. Nous avons donc un système capable de reconnaître spécifiquement une séquence d'ADN double brin longue de 12 paires de bases et induire une réponse luminescente lorsque le complexe est formé correctement.

## Perspectives

### Fonctionnalisation des protéines

Nous avons eu lors de la fonctionnalisation des protéines par le complexe de lanthanide / chromophores, des problèmes de compatibilité entre la méthode de ligation utilisée et la réactivité des fluorophores. Un axe d'amélioration pour la synthèse de protéines fonctionnalisées est de réaliser une

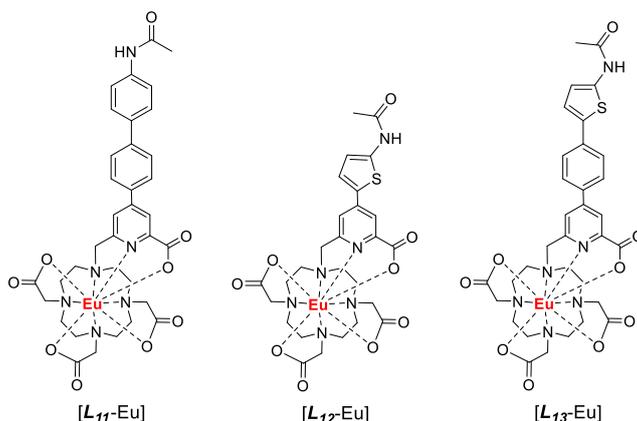
post-fonctionnalisation, après l'assemblage des trois fragments qui composent la protéine. Pour cela le recours à une réaction qui est à la fois rapide et très sélective est nécessaire. Il s'agit d'utiliser une méthode de chimie click qui soit compatible avec les méthodes de ligation native que nous utilisons. Parmi celles qui sont couramment utilisées (cycloaddition de Huisgen<sup>171</sup>, thiol-maléimide<sup>172</sup>...) la réaction entre une oléfine et une tétrazine semble très intéressante<sup>173</sup>, bien qu'aucune synthèse de protéine par assemblage de plusieurs fragments n'ait été décrite avec cette chimie click. Nous avons décidé de fonctionnaliser la protéine avec une oléfine puisque la tétrazine n'est pas compatible avec la synthèse peptidique supportée. Nous avons choisi une oléfine dérivée de l'imide de Reppe, plutôt que l'omniprésent trans-cyclooctène à cause de son coût élevé. Nous avons pu faire la synthèse d'une protéine fonctionnalisée par cette oléfine de manière efficace avec une première ligation par les hydrazides et la seconde par la méthode SEA. Notre collaborateur, Nathan McClenaghan, nous a fourni des silico-rhodamines comportant une unité tétrazine pour réaliser cette post-fonctionnalisation (**Figure 172**). Nous avons été en mesure de réaliser cette réaction de chimie click sur la protéine F1-F2-F3 et cette réaction est à la fois rapide et très efficace. Cette méthode de post-fonctionnalisation est donc compatible et pourrait permettre à l'avenir de greffer facilement différents chromophores organiques et à terme les complexes de lanthanide. Il faut cependant finir la mise au point de cette méthode de post-fonctionnalisation, cela implique aussi de revoir la stratégie de synthèse des complexes pour insérer cette unité tétrazine.



**Figure 172** : Schéma de post-fonctionnalisation de la protéine Zif12 par la SiR<sup>Me</sup> par une méthode de chimie click (oléfine/tétrazine).

## Le complexe d'euporium

Bien que notre système FRET basé sur le transfert d'énergie entre le complexe  $[L_2\text{-Eu}]$  ou  $[L_{10}\text{-Eu}]$  et la  $\text{SiR}^{\text{Me}}$  soit efficace, le complexe d'euporium ne répond pas aux exigences que nous nous étions fixées au début de ce projet de thèse. Le complexe d'euporium  $[L_{10}\text{-Eu}]$  est très brillant avec un rendement quantique de 11%, mais son excitation doit être décalée, vers de plus hautes longueurs d'onde pour éviter l'émission directe du chromophore accepteur en absence de la cible d'ADN. Afin d'augmenter la longueur d'onde d'excitation il faut étendre le système  $\pi$  conjugué de l'antenne en introduisant un espaceur entre la partie électro-donneuse et la partie électro-acceptrice (**Figure 173**). Un espaceur de type alcyne n'est pas envisageable, mais le recours à un groupement aryle peut permettre de décaler la longueur vers le rouge d'environ 20 nm pour se situer dans la zone de non-absorbance de la  $\text{SiR}^{\text{Me}}$  (**Figure 173**,  $[L_{11}\text{-Eu}]$ ). Le groupement thiophène peut aussi être intéressant pour décaler le transfert de charge vers le rouge car il offre une aromaticité et une délocalisation plus faible qu'un phényle. Il serait intéressant, dans un premier temps, de synthétiser le complexe  $[L_{12}\text{-Eu}]$  et de le comparer au complexe  $[L_{10}\text{-Eu}]$  pour évaluer l'influence du thiophène sur l'absorption (longueur d'onde et coefficient d'absorption) et le rendement quantique de l'euporium. Nous pouvons également envisager la synthèse du complexe noté  $[L_{13}\text{-Eu}]$  dans lequel un groupement aryle est inséré.



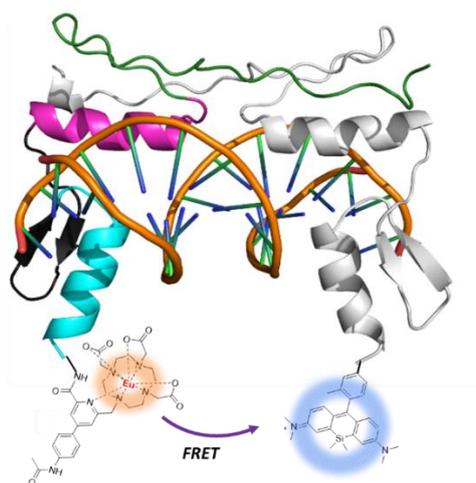
**Figure 173** : Exemple de nouveaux complexes de lanthanide qui pourraient être synthétisés.

## Interaction des protéines décorées avec l'ADN

### *Interaction protéines / ADN*

Au-delà du système FRET en lui-même, la fonctionnalisation des protéines par les chromophores a entraîné une perte d'affinité d'un facteur 15. Mais ce résultat doit être nuancé puisque les études spectroscopiques ont démontré que la présence de la BSA présente dans le tampon C interagit sans doute avec la  $\text{SiR}^{\text{Me}}$ . Ceci se traduit par une chute de luminescence et suggère des modifications empêchant la formation correcte du dimère. Il est donc nécessaire de refaire cette expérience dans un tampon sans

BSA (tampon D). Si la valeur de  $C_{50}$  est similaire à celle obtenue avec la BSA (75 nM), un positionnement différent des luminophores pourrait permettre de retrouver une affinité plus proche de celle des protéines non décorées. Notons enfin que les chromophores sont fixés sur les extensions N-terminales (fragment F1) chargées de stabiliser le dimère autour de l'ADN par la formation d'interactions protéine/protéine complémentaires. Selon leur position, les chromophores pourraient donc diminuer cette contribution des extensions N-terminales. Dans ce cas, une autre position de greffage, loin de l'extension N-terminale et de la zone de reconnaissance de l'ADN serait une option pour regagner en affinité. La fonctionnalisation à l'extrémité de l'hélice  $\alpha$  formée par le doigt de zinc, au niveau C-terminale de la protéine, est un bon compromis entre la distance qui sépare les luminophores et les influences qu'ils peuvent avoir sur la reconnaissance ou les interactions protéine/protéine (**Figure 174**). La distance qui sépare les deux luminophores est estimée à  $17 \pm 5$  Å, avec ce nouveau positionnement cette distance serait aux environs de  $35 \pm 5$  Å. Le rayon de Förster ( $R_0$ ) de notre système FRET est d'environ 40 Å, nous serons donc en mesure de générer le transfert d'énergie depuis le complexe d'euporium vers le fluorophore organique accepteur.



**Figure 174** : Illustration du nouveau positionnement de greffage du complexe d'euporium et de la  $SiR^{Me}$  à l'extrémité C-terminale de l'hélice  $\alpha$  lors de l'interaction avec le duplex d'ADN.

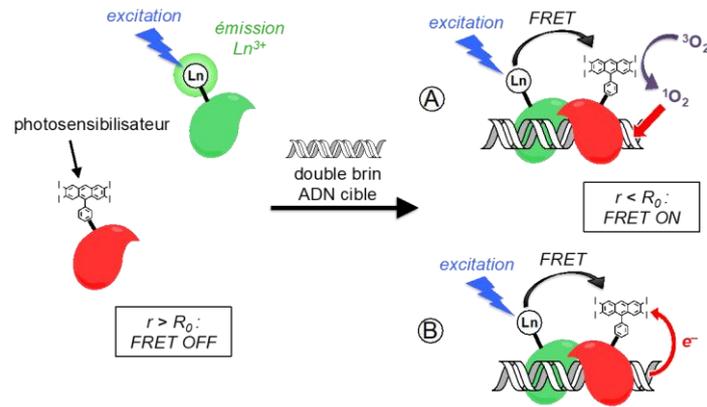
### Transfert FRET

Nous avons durant ces travaux de thèse, grâce aux protéines  $Zif12^*-[L_{10}-Eu]-A$  et  $Zif12^*-(SiR^{Me})-B$ , nous avons mis en évidence l'augmentation du rapport  $I_{SiR}/I_{Eu}$  en présence de la séquence cible non palindromique de 12 paires de bases d'ADN-AB. Cette augmentation n'est pas observée lorsque ces protéines sont incubées avec les séquences non cibles ADN-AA, BB et rdm ce qui permet de discriminer ses ADN de la cible. Nous avons cependant observé une réponse lumineuse lors de l'utilisation de deux séquences d'ADN comprenant une insertion d'une ou trois paires de bases entre les sextets ciblés (ADN-A(1)B et ADN-A(3)B). Bien que cette réponse soit inférieure à celle obtenue avec l'ADN-AB, elle est suffisante pour attester d'un transfert FRET, nous ne pouvons donc pas discriminer l'ADN-AB

des ADN-A(1)B et A(3)B. Cependant nous avons observé que le rapport  $I_{SiR}/I_{Eu}$  est plus faible lorsque l'insert de paires de bases est plus long,  $I_{SiR}/I_{Eu} AB > A(1)B > A(3)B$ . Ce qui laisse à penser que les luminophores ne sont pas suffisamment éloignés l'un de l'autre pour ne plus observer de transfert FRET. Le nouveau positionnement des luminophores évoqué précédemment augmentera la distance qui sépare les luminophores et diminuera le transfert d'énergie lors de l'interaction des protéines avec les ADN-A(1)B et A(3)B. Nous pourrions ainsi avoir un système dont la réponse luminescente, soit non seulement, sélective de la séquence cible d'ADN-AB vis-à-vis des séquences AA, BB et rdm, mais aussi des séquences d'ADN-A(1)B et A(3)B. Il serait intéressant de faire l'étude de l'interaction des protéines Zif12\*-[L<sub>10</sub>-Eu]-A et Zif12\*-(SiR<sup>Me</sup>)-B avec les séquences d'ADN-A(1)B et A(3)B par gel-retard et déterminer les valeurs de C<sub>50</sub> correspondantes. Pour cette étude, les concentrations utilisées sont 10 à 25 fois plus faibles que celles utilisées pour les études spectroscopiques et ces conditions pourraient minimiser la formation des complexes avec les séquences contenant des inserts de paires de bases. L'insertion des bases laisse à penser que les interactions protéine/protéine ne sont pas optimales compte tenu de l'augmentation de la distance qui sépare les deux sextets ciblés. Si les valeurs de C<sub>50</sub> confirment cette diminution de l'affinité, nous pourrions aussi envisager de diminuer la concentration des protéines fonctionnalisées en passant de 100 nM à 10nM, de tel sorte à ce que les protéines ne puissent former un dimère avec les ADN-A(1)B et A(3)B, mais seulement pour la cible d'ADN-AB. Mais cela implique d'utiliser un complexe d'euporium bien plus brillant.

Nous avons remarqué que la réalisation des études spectroscopiques des protéines dans les cuves en quartz induit l'adsorption de ses protéines sur les parois, ce qui influe sur les intensités d'émissions et peut « polluer » les mesures réalisées sur une même cuve malgré de multiples rinçages à l'eau. L'utilisation de micro-plaques traitées « low-binding » pourrait permettre à la fois d'éviter cette adsorption des protéines et de réaliser plus facilement les mesures de luminescence ainsi que la réalisation d'une gamme de concentration afin d'obtenir un titrage qui valide la formation d'un complexe 1:1:1 (Zif12\*-[L<sub>10</sub>-Eu]-A / Zif12\*-(SiR<sup>Me</sup>)-B / ADN-AB).

Enfin nous pouvons envisager de changer l'accepteur d'énergie dans le système FRET, en passant d'un chromophore luminescence à un photosensibilisateur capable de générer une dégradation au niveau de l'ADN. L'idée sous-jacente est d'utiliser notre système de reconnaissance à deux protéines pour générer une dégradation photo-induite au niveau de la séquence d'ADN ciblée. Cette dégradation peut avoir lieu par un transfert électronique depuis le photosensibilisateur ou par réaction avec l'oxygène triplet pour former de l'oxygène singulet (**Figure 175**). Le nouveau positionnement des luminophores proposé précédemment **Figure 174** pourra permettre de rapprocher au maximum le photosensibilisateur du duplex d'ADN.



**Figure 175 :** Principe de la photo-dégradation de l'ADN par le transfert d'énergie de type FRET entre un complexe de lanthanide et un photosensibilisateur. Le transfert d'énergie étant induit par la reconnaissance et la formation du dimère autour de la séquence cible.

# Partie Expérimentale

## Summary

Organic Synthesis.....	178
Metallation of Ln <sup>3+</sup> after tert-butyl déprotection of PL <sub>i</sub> .....	201
Coupling of PL <sub>i</sub> on TAT.....	201
Peptidic synthesis .....	202
Peptidic sequences on resin.....	202
Peptides .....	202
Fragment F1, F2 and F3 .....	202
Peptide F2-F3 .....	203
Peptide F1-F2-F3.....	203
General procedure: .....	204
Peptide characterisation.....	207
Fragment F1, F2 and F3 .....	207
Peptide F2-F3 .....	210
Peptide F1-F2-F3.....	212

Ce chapitre présente le matériel, les méthodes et techniques utilisées pour mener à bien les travaux de ce manuscrit. Un mode opératoire détaillé accompagné d'une analyse complète de chaque nouvelle molécule sont fournis. Afin de rendre cette partie accessible à un plus grand nombre de personnes et pour faciliter toute reproduction de manipulation, ce chapitre sera rédigé en anglais.

## Materials and methods

**Solvent and reagents:** N- $\alpha$ -Fmoc protected amino acids for peptide synthesis, PyBOP and HCTU coupling reagents were obtained from Novabiochem or Iris Biotech. SEA-PS, NovaPEG Rink Amide and Chlorotriptyl resin were purchased from X'prochem and Novabiochem. DOTA-*tris*(*t*Bu) ester was purchased from CheMatech. Other reagents for peptide synthesis, solvents, buffers and metal salts were purchased from Sigma-Aldrich. All buffer or zinc solution were prepared with MilliQ water (Millipore).

**Analytical HPLC:** analytical separations were performed on an Agilent Infinity 1200 system using Water XBridge BEH130 C18 (2.5 $\mu$ m, 75mm x 4.6 mm) column at 1 mL/min.

Mobile phase consisted in a gradient of solvent A (0.1% TFA in H<sub>2</sub>O) and B (0.1% TFA in MeCN/H<sub>2</sub>O 9:1). Method Xb100 consisted in a 5% B during 1 min followed by a 5 to 100% b gradient in 14 min at 1mL/min on the Waters XBridge BEH130 C18 column. Eluate was monitored by electronic absorption at 214, 254 280, 365, 560, 650 nm.

**Analytical HPLC-MS:** analytical separations followed by mass spectroscopy were performed on an Agilent 1260 Infinity II HPLC system equipped with a 6125 ESI-MS detector.

**Preparative HPLC:** separations for purification were performed on a VWR LaPrep $\Sigma$  system using a Waters XBridge Peptide BEH130 C18 (5  $\mu$ m, 150 mm, 19 mm) column at 14mL/min, or on a Waters XBridge Peptide BEH130 C18 (5  $\mu$ m, 150 mm, 10 mm) column at 6mL/min. Mobile phase consisted in a gradient of solvent A (0.1% TFA in H<sub>2</sub>O) and B (0.1% TFA in MeCN/H<sub>2</sub>O 9:1).

**Optical spectroscopy:** UV-Vis absorption spectra were recorded on a Perkin-Elmer Lambda 35 spectrophotometer. Luminescence spectra were recorded on a Varian Cary Eclipse spectrometer or on a modularFluorolog FL3-22 spectrometer from Horiba-Jobin Yvon-Spex equipped with a double-grating excitation monochromator and an iHR320 imaging spectrometer coupled to an R928P Hamamatsu photomultiplier. Quantum yields were determined by a relative method with quinine sulphate in 0.5M H<sub>2</sub>SO<sub>4</sub> as standard<sup>153</sup>. Estimated experimental error for the quantum yield determination is ~10%.

**NMR:** <sup>1</sup>H, <sup>13</sup>C and DEPT spectra were recorded at 400 MHz on a Varian Avance III 400 spectrometer, at 298 K if no temperature is specified. All shifts for <sup>1</sup>H and <sup>13</sup>C spectra were referenced to the residual solvent peak (CDCl<sub>3</sub>  $\delta$ <sub>H</sub> = 7.26 ppm and  $\delta$ <sub>C</sub> = 77.2 ppm; D<sub>2</sub>O  $\delta$ <sub>H</sub> = 4.79 ppm at 298 K and 4.20 ppm at 353 K; (CD<sub>3</sub>)<sub>2</sub>SO  $\delta$ <sub>H</sub> = 2.50 ppm and  $\delta$ <sub>C</sub> = 39.52 ppm; MeOD  $\delta$ <sub>H</sub> = 3.31 ppm and  $\delta$ <sub>C</sub> = 49.0 ppm). The

following abbreviations were used to explain the multiplicities: s= singlet, d= doublet, t= triplet, q= quarter, dd= doublet of doublet, m= multiplet. Coupling constant ( $J$ ) are measured in Hz. DEPT experiments was used to further characterize organic molecules.

**Mass spectrometry:** ESI-MS analyses were performed on a thermo Scientific LXQ spectrometer. HMRS analyses were performed on a Thermo Scientific LTQ Orbitrap XL spectrometer.

**Automated peptide synthesis:** Peptide synthesis was performed on a CEM Liberty 1 peptide synthesizer.

Peptide elongation: Peptide elongation was performed using standard SPPS protocols using Fmoc/*t*Bu chemistry either manually or on an automated peptide synthesizer (CEM Liberty 1 Microwave peptide Synthesizer). Double couplings (30 min) were performed using 4-fold molar excess of Fmoc-L-amino acid, 4-fold molar excess of HCTU and 8-fold molar excess of DIEA at room temperature. A capping step was performed after each coupling with Ac<sub>2</sub>O/DIEA in DMF (5min). Fmoc removal was performed using 20% piperidine in DMF (2x10 min) with Fmoc UV monitoring.

**Peptide elongation :** Peptide elongation was performed using standard SPPS protocols using Fmoc/*t*Bu chemistry either manually or with the automated peptide synthesizer (CEM Liberty1 Microwave Peptide Synthesizer). Double couplings (30 min) needed 4-fold molar excess of Fmoc-L-amino acid, 4-fold molar excess of HCTU and 8-fold molar excess of DIEA at room temperature. A capping step was performed after coupling with Ac<sub>2</sub>O/DIEA in DMF (5 min). Fmoc protecting group was removed with 20% piperidine in DMF (2x10 min) while Fmoc UV monitoring.

**Chromatography:** reactions were monitored by thin layer chromatography (TLC) using commercial Pre-coated TLC sheet ALUGRAM SIL G/UV<sub>254</sub> (from Macherey-Nagel). TLC spots were viewed under UV light (254 nm). Products were purified by silica gel (Merck silicagel 60; 0.063-0.200 mm) for flash column chromatography.

**Oligonucleotide synthesis & purification:** Oligonucleotide synthesis was performed by an automated 3400 Applied Biosystem synthesizer (reagents provided by Link Technology) using the phosphoramidite synthesis method. Removing from the solid support was performed by treatment with a 15 N solution of ammonia at room temperature. Total deprotection of the sequence was accomplished by heating at 55°C for 48 hours. Purification of oligonucleotide was performed using 500 µg of compound, dissolved in a 100 µL of a 7 M urea aqueous solution and purified by polyacrylamide gel electrophoresis (PAGE) under denaturing conditions in TBE buffer (10.8 g/L Tris, 5.5 g/L or boric acid and 2mM of EDTA pH8) at a power of 55 W for 10 hours. The PAGE gel was composed by a mixture of: 125g of urea; 125 mL of acrylamide/*bis*-acrylamide 40%; 25mL of TBE 10x; 8.75 mL H<sub>2</sub>O; 1.675 mL of ammonium persulfate (APS) at 5%; 100 µL of N,N,N',N'-tetramethylethylenediamine (TEMED). After separation, the bands containing the full-length oligonucleotides were detected by UV shadowing: the gel was placed

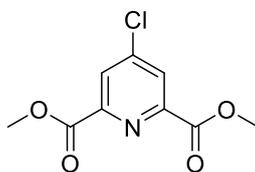
on a pre-coated TLC sheet ALUGRAM SIL G/UV<sub>254</sub> (from Macherey-Nagel) and illuminated with a 254 nm UV light. The gel slices containing the purified product were crushed and placed in 4 mL of water overnight at 4°C to extract the oligonucleotide from acrylamide. After centrifugation, oligonucleotide was desalted on C18 column using water, followed by elution with methanol/H<sub>2</sub>O (6/4), and then lyophilised.

**Radioactive labelling:** Oligonucleotide was labelled in 5' using the [ $\gamma$ -<sup>32</sup>P] ATP triphosphate (specific activity of 300 Ci/mmol; Perkin Elmer supplier) and the T4 polynucleotide kinase (approximately 10 units per 20 pmol of oligonucleotide, MBI supplier Fermentas) at 37°C for 30 min. Activity of the T4 polynucleotide kinase was stopped by incubating the reaction medium at 70°C for 5 min.

**Hybridization:** The <sup>32</sup>P-labeled oligonucleotide was hybridized with a 10% excess of unlabelled complementary strand in the appropriate incubation buffer by heating at 95°C for 10 min followed by a return to room temperature for 5 hours.

**Electrophoretic Mobility Shift Assay:** The 5'-labeled target duplex (10 nM) was incubated in the presence of various protein concentrations overnight at 4°C in a final solution volume of 20  $\mu$ L of HEPES pH 7.5/BSA/Tween/NaCl (15 mM, 0.1g/L, 0.05 %, 100 mM). After addition of 10  $\mu$ L of loading buffer (50% glycerol, 50% water) 15  $\mu$ L of the mixture were loaded onto an 8% acrylamide gel: *bis*-acrylamide EMSA gel (29:1) in Tris/glycine migration buffer (25 mM Tris, 190 mM glycine). The electrophoresis apparatus could be refrigerated by circulating a cooling bath for the duration of the migration (2h at a power not exceeding 10 Watts). In these conditions, the gel temperature did not exceed 10°C. The gels were finally quantified using a GE Healthcare Typhoon 9410 phosphorimager. The EMSA gel was composed by a solution of: 6 mL of 40% acrylamide/*bis*-acrylamide (29:1); 6 mL of 5x Tris/glycine buffer; 18 mL of water. This solution was polymerized by addition of a mixture of 300  $\mu$ L of APS 10% and 30  $\mu$ L of TEMED.

## Organic Synthesis



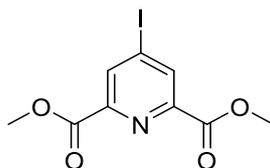
### Compound n°1: Dimethyl 4-chloropyridine-2,6-dicarboxylate

To a suspension of chelidamic acid (11.2 g, 56 mmol) in thionyl chloride (81 mL) was added DMF (11 mL). The mixture was heated to reflux for 16 h and solvents were removed under vacuum. Residues were dissolved in DCM (50 mL) and methanol (50 mL) was added dropwise at 0 °C. After 1h, solvents were removed under vacuum. The residue was dissolved in DCM and washed with saturated aq. NaHCO<sub>3</sub> and water. The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent removed under reduced pressure to afford slightly brown powder (12.4 g, 97 %).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ (ppm) = 8.29 (s, 2H), 4.03 (s, 6H).

<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>): δ (ppm) = 164.2 (C<sub>quat</sub>), 149.6 (C<sub>quat</sub>), 147.0 (C<sub>quat</sub>), 128.4 (CH), 53.6 (CH<sub>3</sub>).

LRMS (ESI<sup>+</sup>): monoisotopic m/z = 252.0 [M+Na]<sup>+</sup>, 268.0 [M+K]<sup>+</sup>, 496.6 [2M+K]<sup>+</sup> (calculated monoisotopic m/z = 252.0 [M+Na]<sup>+</sup>, 267.98 [M+K]<sup>+</sup>, 496.99 [2M+K]<sup>+</sup> for M = C<sub>9</sub>H<sub>8</sub>ClNO<sub>4</sub>).



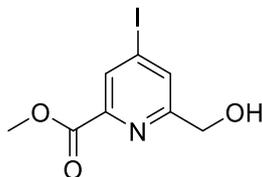
### Compound n°2: Dimethyl 4-iodopyridine-2,6-dicarboxylate

To a solution of **1** (12 g, 52 mmol) in acetonitrile was added sodium iodide (157 g, 1,05 mol) and acetyl chloride (11.2 mL, 157 mmol). The mixture was sonicated for 2 h then diluted with DCM and washed successively by saturated aq. NaHCO<sub>3</sub> and aq. sodium thiosulfate (0.1 M) until no colour change happens and, finally, with brine. The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent removed under reduced pressure. The crude product was dissolved in a minimal amount of DCM, and crystallized by diffusion in methanol to afford white crystals (12.9 g, 77 %).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ (ppm) = 8.62 (s, 2H), 3.98 (s, 6H).

$^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 163.9 ( $\text{C}_{\text{quat}}$ ), 148.3 ( $\text{C}_{\text{quat}}$ ), 137.2 (CH), 107.0 ( $\text{C}_{\text{quat}}$ ), 53.5 ( $\text{CH}_3$ ).

LRMS (ESI+): monoisotopic  $m/z$  = 344.0  $[\text{M}+\text{Na}]^+$ , 360.0  $[\text{M}+\text{K}]^+$ , 680.4  $[\text{2M}+\text{K}]^+$  (calculated monoisotopic  $m/z$  = 343.94  $[\text{M}+\text{Na}]^+$ , 359.92  $[\text{M}+\text{K}]^+$ , 680.86  $[\text{2M}+\text{H}]^+$  for  $\text{M} = \text{C}_9\text{H}_8\text{INO}_4$ ).



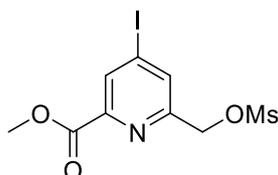
**Compound n°3:** Methyl 6-(hydroxymethyl)-4-iodopicolinate

Compound **2** (700 mg, 2.2 mmol) was dissolved in a solution of DCM/MeOH 2:3 (10 mL) cooled at  $0^\circ\text{C}$ .  $\text{NaBH}_4$  (41 mg, 1.1 mmol) was added by portions and the mixture was stirred at  $0^\circ\text{C}$  for 45 min. A second amount of  $\text{NaBH}_4$  (41 mg, 1.1 mmol) was added and the advancement of the reaction was monitored by TLC. When the di-reduced compound was detected by TLC, the reaction was quenched by addition of 0.1 M aq. HCl (4 mL). The organic solvents were evaporated under reduced pressure and the resulting aq. phase was extracted with AcOEt. The organic phase was dried over  $\text{Na}_2\text{SO}_4$  and the solvent removed under reduced pressure. The crude product was purified by flash chromatography (silica gel, DCM/MeOH 95:5) to give a white powder (445 mg, 70 %).

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 8.37 (d,  $^4J = 1.5$  Hz, 1H), 7.97 (d,  $^4J = 1.5$  Hz, 1H), 4.81 (s, 2H), 3.98 (s, 3H).

$^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 164.4 ( $\text{C}_{\text{quat}}$ ), 161.3 ( $\text{C}_{\text{quat}}$ ), 147.4 ( $\text{C}_{\text{quat}}$ ), 133.3 (CH), 133.1 (CH), 106.9 ( $\text{C}_{\text{quat}}$ ), 64.3 ( $\text{CH}_2$ ), 53.3 ( $\text{CH}_3$ ).

LRMS (ESI+): monoisotopic  $m/z$  = 294.2  $[\text{M}+\text{H}]^+$ , 316.1  $[\text{M}+\text{Na}]^+$ , 332.1  $[\text{M}+\text{K}]^+$  (calculated monoisotopic  $m/z$  = 293.96  $[\text{M}+\text{H}]^+$ , 315.94  $[\text{M}+\text{H}]^+$ , 331.92  $[\text{M}+\text{K}]^+$  for  $\text{M} = \text{C}_8\text{H}_8\text{INO}_3$ ).



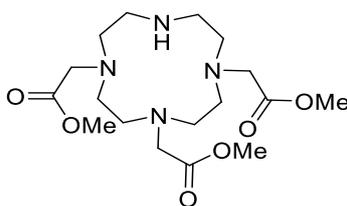
**Compound n°4:** Methyl 4-iodo-6-(((methylsulfonyl)oxy)methyl)picolinate

To a solution of **3** (1.35 g, 4.37 mmol) in THF (44 mL) was added triethylamine (2.1 mL, 13.1 mmol) and mesylchloride (570  $\mu\text{L}$ , 7.0 mmol). The mixture was stirred for 25 min, and solvents were

evaporated. The residue was dissolved in DCM and washed by saturated aq.  $\text{NaHCO}_3$  and water. The organic phase was dried over  $\text{Na}_2\text{SO}_4$  and the solvent was removed under reduced pressure to give a white solid (1.59 g, 98 %).

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 8.38 (d,  $^4J = 1.5$  Hz, 1H), 7.98 (d,  $^4J = 1.5$  Hz, 1H), 5.30 (s, 2H), 3.93 (s, 3H), 3.12 (s, 3H).

$^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 163.8 ( $\text{C}_{\text{quat}}$ ), 155.0 ( $\text{C}_{\text{quat}}$ ), 147.7 ( $\text{C}_{\text{quat}}$ ), 134.0 (CH), 107.1 ( $\text{C}_{\text{quat}}$ ), 69.8 ( $\text{CH}_2$ ), 53.2 ( $\text{CH}_3$ ), 38.0 ( $\text{CH}_3$ ).



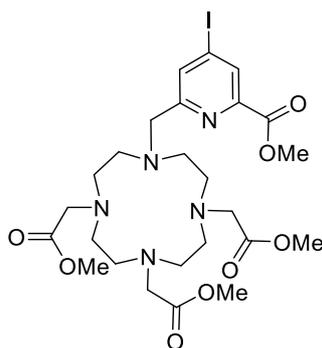
**Compound n°5:** 4,7,10-*tris*(2-methoxy-2-oxoethyl)-1,4,7,10-tetraazacyclododecane hydrobromide

To a suspension of sodium acetate (1.58 g, 19.3 mmol) and cyclen (1.0 g, 5.8 mmol, in DMF (8 mL) at  $-20$  °C, was added a solution of methyl bromoacetate (1.7 mL, 17.4 mmol) in DMF (4 mL). The reaction was stirred for 6 h at room temperature. The solution was diluted with DCM (100 mL) and the precipitate was removed by filtration over Celite. Solvents were evaporated under vacuum. The residue was precipitated in AcOEt and centrifuged. The supernatant was evaporated and purified by flash chromatography (silica gel, DCM/MeOH 95:5) to give an orange oil once sonicated on  $\text{Et}_2\text{O}$  give a white powder (2.31 g, 85 %).

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 9.98 (s, 1H), 3.70 (s, 9H), 3.49 (s, 4H), 3.43 (s, 2H), 3.15-2.81 (m, 16H).

$^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 171.6 ( $\text{C}_{\text{quat}}$ ), 170.8 ( $\text{C}_{\text{quat}}$ ), 57.1 ( $\text{CH}_2$ ), 51.7 ( $\text{CH}_3$ ), 51.5 ( $\text{CH}_2$ ), 49.4 ( $\text{CH}_2$ ), 48.5 ( $\text{CH}_2$ ) 47.3 ( $\text{CH}_2$ ).

LRMS (ESI $^+$ ): monoisotopic  $m/z$  = 389.4 [ $\text{M}+\text{H}$ ] $^+$ , 411.4 [ $\text{M}+\text{Na}$ ] $^+$  (calculated monoisotopic  $m/z$  = 389.24 [ $\text{M}+\text{H}$ ] $^+$ , 411.22 [ $\text{M}+\text{Na}$ ] $^+$  for  $\text{M} = \text{C}_{17}\text{H}_{32}\text{N}_4\text{O}_6$ ).



**Compound n°6:** Pico(OMe)-DO3A(Me)<sub>3</sub>

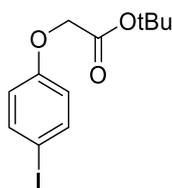
To a solution of compound **5** (1.3 g, 3.3 mmol) in acetonitrile (20 mL), was added Na<sub>2</sub>CO<sub>3</sub> (1.5 g, 14.0 mmol) and then a solution of **3** (1.04 g, 2.8 mmol) in acetonitrile (10 mL) was added. The mixture was heated to reflux for 22 h. The mixture was cooled down to room temperature and filtered. The filtrate was evaporated and the residue was triturated in AcOEt to give a white powder (3.29 g, 99%).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ (ppm) = 8.33 (d, <sup>4</sup>J = 1.4 Hz, 1H), 7.91 (d, <sup>4</sup>J = 1.4 Hz, 1H), 3.96 (s, 3H), 3.76 (s, 3H), 3.62, (s, 6H), 3.37-2.26 (m, 24H).

<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>): δ (ppm) = 173.8 (C<sub>quat</sub>), 172.7 (C<sub>quat</sub>), 164.4 (C<sub>quat</sub>), 160.0 (C<sub>quat</sub>), 147.8 (C<sub>quat</sub>), 136.7 (CH), 133.0 (CH), 107.9 (C<sub>quat</sub>), 58.9 (CH<sub>2</sub>), 55.4 (CH<sub>2</sub>), 55.2 (CH<sub>2</sub>), 53.2 (CH<sub>3</sub>), 52.2 (CH<sub>3</sub>), 52.1 (CH<sub>3</sub>), 50.6 (CH<sub>2</sub>).

LRMS (ESI+): monoisotopic m/z = 686.3 [M+Na]<sup>+</sup>, 702.3 [M+K]<sup>+</sup>, (calculated monoisotopic m/z = 686.17 [M+Na]<sup>+</sup>, 702.14 [M+K]<sup>+</sup> for M = C<sub>25</sub>H<sub>38</sub>IN<sub>5</sub>O<sub>8</sub>).

HRMS (ESI+): m/z = 664.1827 (+) (calculated m/z = 664.1838 [M+H]<sup>+</sup> for M = C<sub>25</sub>H<sub>39</sub>N<sub>5</sub>O<sub>8</sub>I).



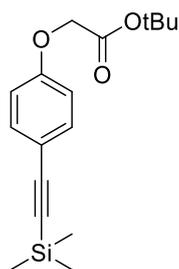
**Compound 7:** tert-butyl 2-(4-iodophenoxy)acetate

To a solution of 4-iodophenol (3.0 g, 13.6 mmol) in DMF (30 mL) was added potassium carbonate (10 g, 106 mmol), and tert-butyl bromoacetate (6 mL, 40.5 mmol). The mixture was stirred at room temperature for 16 h and then filtrated. The solvent was removed under reduced pressure. The oily residue was dissolved in DCM and washed with water then brine. The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent removed under reduced pressure. The residue was purified by flash chromatography (silica gel, DCM) to give a pale red liquid (3.43 g, 75 %).

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 7.51 (d,  $^3J = 9$  Hz, 2H), 6.63 (d,  $^3J = 9$  Hz, 2H), 4.44 (s, 2H), 1.45 (s, 9H).

$^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 167.6 ( $\text{C}_{\text{quat}}$ ), 157.8 ( $\text{C}_{\text{quat}}$ ), 138.3 (CH), 117.0 (CH), 83.8 ( $\text{C}_{\text{quat}}$ ), 82.6 ( $\text{C}_{\text{quat}}$ ), 65.7 ( $\text{CH}_2$ ), 28.0 ( $\text{CH}_3$ ).

LRMS (ESI+): monoisotopic  $m/z = 357.0$  [ $\text{M}+\text{Na}$ ] $^+$ , 373.0 [ $\text{M}+\text{K}$ ] $^+$  (calculated monoisotopic  $m/z = 357.00$  [ $\text{M}+\text{Na}$ ] $^+$ , 372.97 [ $\text{M}+\text{K}$ ] $^+$  for  $\text{M} = \text{C}_{12}\text{H}_{15}\text{IO}_3$ ).

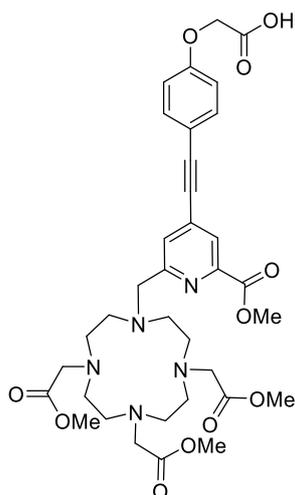


**Compound 8:** *tert*-butyl 2-(4-((trimethylsilyl)ethynyl)phenoxy)acetate

To a degassed solution of **7** (1.0 g, 3.0 mmol) and triethylamine (1.3 mL, 9.3 mmol) in DMF (1.3 mL) was added under an inert atmosphere *bis*(triphenylphosphine)palladium(II) dichloride (84 mg, 0.12 mmol), copper iodide (114 mg, 0.6 mmol) and ethynyltrimethylsilane (1.5 mL, 10.5 mmol). The mixture was stirred at room temperature for 15 h. Solvents were removed under reduced pressure. The residue was dissolved in DCM and washed with saturated aq.  $\text{NH}_4\text{Cl}$  and water. The organic layer was dried over  $\text{Na}_2\text{SO}_4$  and the solvent was removed under reduced pressure to give a dark oil. The crude product was purified by flash chromatography (silica gel, Cyclohexane/AcOEt 95:5) to give brown crystals (900 mg, 99 %).

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 7.39 (d,  $^3J = 9.0$  Hz, 2H), 6.80 (d,  $^3J = 9.0$  Hz, 2H), 4.50 (s, 2H), 1.47 (s, 9H), 0.23 (s, 9H).

$^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm): 167.8 ( $\text{C}_{\text{quat}}$ ), 158.2 ( $\text{C}_{\text{quat}}$ ), 133.6 (CH), 116.3 ( $\text{C}_{\text{quat}}$ ), 114.6 (CH), 105.0 ( $\text{C}_{\text{quat}}$ ), 92.9 ( $\text{C}_{\text{quat}}$ ), 82.7 ( $\text{C}_{\text{quat}}$ ), 65.8 ( $\text{CH}_2$ ), 28.1 ( $\text{CH}_3$ ), 0.16 ( $\text{CH}_3$ ).



**Compound n°9** →  $PL_1^{CC}$ : COOH-CH<sub>2</sub>-O-Ar-CC-Pico(OMe)-DO3A(OMe)<sub>3</sub>

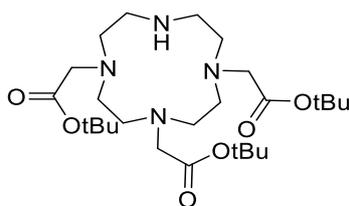
To a degassed solution of compound **6** (100 mg, 130 μmol) and compound **8** (33 mg, 108 μmol) in DMF (1.1 mL) was added triethylamine (425 μL, 3.05 mmol), *bis*(triphenylphosphin)palladium(II) dichloride (8 mg, 11.4 μmol), and tetrabutylammonium fluoride 1M (130 μL, 130 μmol). The mixture was stirred at 90 °C for 17 h. Solvent was removed under reduced pressure. The oily residue was dissolved in DCM and washed with water. The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and evaporated under reduced pressure. Water (16 mL) was added to the crude and the suspension was sonicated for 20 min, then centrifuged. The supernatant was lyophilized to give DO3A(Me)<sub>3</sub>pic(Me)Ar<sub>2</sub>(*t*Bu) as a brown oil (121 mg). Removal of tert-butyl protecting group was performed in TFA (5 mL) at room temperature for 2 h. TFA was evaporated under reduced pressure and the crude was dissolved in water and purified on HPLC to give a brown oil (38.6 mg, 50 %) after lyophilization.

<sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O, 353K): δ (ppm) = 8.16 (d, <sup>4</sup>J = 1.1 Hz, 1H), 7.87 (d, <sup>4</sup>J = 1.0 Hz, 1H), 7.54 (d, <sup>3</sup>J = 8.8 Hz, 2H), 7.02 (d, <sup>3</sup>J = 8.8 Hz, 2H), 4.76 (s, 2H), 4.45 (s, 2H), 4.03 (s, 3H), 4.01 (s, 2H), 3.74 (s, 3H), 3.66 (s, 4H), 3.61 (s, 6H), 3.54-3.02 (m, 16H).

<sup>13</sup>C NMR (100 MHz, D<sub>2</sub>O, 353K): δ (ppm) = 173.3 (C<sub>quat</sub>), 172.0 (C<sub>quat</sub>), 169.5 (C<sub>quat</sub>), 166.4 (C<sub>quat</sub>), 159.5 (C<sub>quat</sub>), 148.2 (C<sub>quat</sub>), 135.3 (C<sub>quat</sub>), 134.5 (C<sub>quat</sub>), 134.4 (CH), 130.9 (CH), 127.4 (CH), 115.9 (CH), 114.7 (C<sub>quat</sub>), 97.4 (C<sub>quat</sub>), 85.5 (C<sub>quat</sub>), 65.9 (CH<sub>2</sub>), 58.5 (CH<sub>2</sub>), 54.8 (CH<sub>2</sub>), 54.1 (CH<sub>3</sub>), 53.9 (CH<sub>3</sub>), 53.3 (CH<sub>2</sub>), 53.2 (CH<sub>3</sub>), 52.0 (CH<sub>2</sub>), 51.8 (CH<sub>2</sub>), 50.0 (CH<sub>2</sub>), 49.6 (CH<sub>2</sub>).

LRMS (ESI<sup>+</sup>): monoisotopic  $m/z$  = 712.6 [M+H]<sup>+</sup>, 734.6 [M+Na]<sup>+</sup>, 750.6 [M+K]<sup>+</sup> (calculated monoisotopic  $m/z$  = 712.32 [M+H]<sup>+</sup>, 734.30 [M+Na]<sup>+</sup>, 750.27 [M+K]<sup>+</sup> for M = C<sub>35</sub>H<sub>45</sub>N<sub>5</sub>O<sub>11</sub>).

HRMS (ESI<sup>+</sup>):  $m/z$  = 712.3174 (+) (calculated  $m/z$  = 712.3188 [M+H]<sup>+</sup> for M = C<sub>35</sub>H<sub>45</sub>N<sub>5</sub>O<sub>11</sub>).

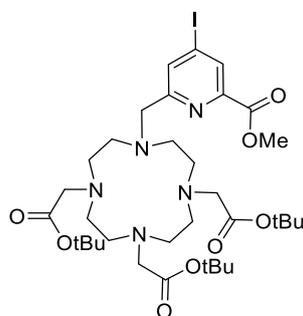


**Compound n°10:** tri-tert-butyl 2,2',2''-(1,4,7,10-tetraazacyclododecane-1,4,7-triyl) triacetate

To a suspension of sodium acetate (7.86 g, 96 mmol) and cyclen (5.0 g, 29 mmol, in DMAc (60 mL) at  $-20^{\circ}\text{C}$ , was added a solution of tert-butyl bromoacetate (14.1 mL, 96 mmol) in DMAc (20 mL). The reaction was stirred for 25h at room temperature. The solution was poured in water to give a clear solution.  $\text{KHCO}_3$  was added until precipitation. The mixture was filtrated and then dissolved in chloroform. The mixture was washed with water and the organic layer was dried over  $\text{Na}_2\text{SO}_4$  and the solvent was partially removed, diethyl ether was added to precipitate the product as a white solid (11.7g, 79%).

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 10.03 (br, 1H), 3.38 (s, 4H), 3.29 (s, 2H), 3.10 (m, 4H), 2.93 (m, 8H), 2.87 (m, 4H), 1.46 (s, 9H), 1.45 (s, 18H).

$^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 170.4, 169.5, 81.8, 81.6, 58.1, 51.3, 51.2, 49.1, 48.7, 47.4, 28.2, 28.1



**Compound n°11:** Pico(OMe)-DO3A(tBu)<sub>3</sub>

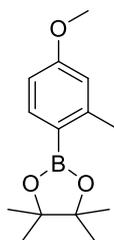
and  $\text{Na}_2\text{CO}_3$  (1.58 g, 15.0 mmol) were dissolved in acetonitrile (20 mL). Compound 3 was added (1.1 g, 3.0 mmol) and the mixture was heated to reflux for 20 h. The mixture was cooled to room temperature and filtered, the filtrate was evaporated and the yellow oil obtained triturated several times in AcOEt. The solution was filtered to give a white powder (2.34 g, 80 %).

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 8.44 (d,  $^4J = 1.35$  Hz, 1H), 8.03 (d,  $^4J = 1.35$  Hz, 1H), 3.98 (s, 3H), 3.20-2.27, (m, 24H), 1.46 (s, 9H), 1.41 (s, 18H).

$^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 169.26 ( $\text{C}_{\text{quat}}$ ), 163.94 ( $\text{C}_{\text{quat}}$ ), 161.06, ( $\text{C}_{\text{quat}}$ ), 160.70 ( $\text{C}_{\text{quat}}$ ), 160.33 ( $\text{C}_{\text{quat}}$ ), 154.86 ( $\text{C}_{\text{quat}}$ ), 137.10 (CH), 134.23 (CH), 117.74 ( $\text{C}_{\text{quat}}$ ), 114.85 ( $\text{C}_{\text{quat}}$ ), 107.42 ( $\text{C}_{\text{quat}}$ ), 83.30 ( $\text{C}_{\text{quat}}$ ), 82.92 ( $\text{C}_{\text{quat}}$ ), 57.32 ( $\text{CH}_2$ ), 55.62 ( $\text{CH}_2$ ), 55.45 ( $\text{CH}_2$ ), 53.26 ( $\text{CH}_3$ ), 51.22 ( $\text{CH}_2$ ), 51.12 ( $\text{CH}_2$ ), 50.41 ( $\text{CH}_2$ ), 50.11 ( $\text{CH}_2$ ), 29.84 ( $\text{CH}_2$ ), 28.15 ( $\text{CH}_3$ ).

LRMS (ESI+): monoisotopic  $m/z = 790.3$   $[\text{M}+\text{H}]^+$  (calculated monoisotopic  $m/z = 790.3$   $[\text{M}+\text{H}]^+$  for  $\text{M} = \text{C}_{34}\text{H}_{56}\text{N}_5\text{O}_8$ ).

HRMS (ESI+):  $m/z = 790.3239$  (+) (calculated  $m/z = 790.3246$   $[\text{M}+\text{H}]^+$  for  $\text{M} = \text{C}_{34}\text{H}_{57}\text{N}_5\text{O}_8$ ).



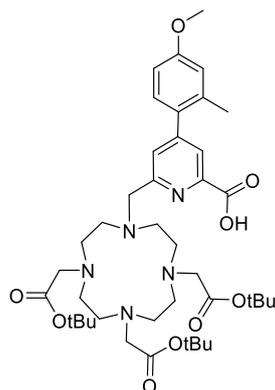
**Compound n°12:** 2-(4-methoxy-2-methylphenyl)-4,4,5,5-tetramethyl-1,3,2-dioxaborolane

Protocol adapted from: *Org. Lett.* **2004**, 6, 24, 4419–4422

To a solution of 4-bromo-3-methyl-anisole (157  $\mu\text{L}$ , 1.2 mmol) and 4,4,5,5-tetramethyl-1,3,2-dioxaborolane (260  $\mu\text{L}$ , 1.8 mmol) in dry dioxane (15 mL) were added *bis*(triphenylphosphine)palladium(II) dichloride (25 mg, 0.03 mmol) and triethylamine (500  $\mu\text{L}$ , 3.6 mmol). The reaction was stirred at 90°C for 48 hours under argon. The solvent was evaporated and the solid obtained was extracted using DCM and water. The organic layer was dried over  $\text{Na}_2\text{SO}_4$  and the solvent was removed under reduced pressure. The crude product was purified by flash chromatography (silica gel, Cyclohexane/DCM 1:1) to give a white powder (349 mg, 88%).

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm): 7.72 (dd,  $^3J = 9.07$  Hz,  $^4J = 2.89$  Hz, 1H), 6.70 (dd,  $^3J = 7.42$  Hz,  $^4J = 2.41$  Hz, 2H), 3.81 (s, 3H), 2.53 (s, 3H), 1.33 (s, 12H).

$^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm): 161.8 ( $\text{C}_{\text{quat}}$ ), 147.4 ( $\text{C}_{\text{quat}}$ ), 138.0 (CH), 115.7 (CH), 110.3 (CH), 83.3 ( $\text{C}_{\text{quat}}$ ), 55.1 ( $\text{CH}_3$ ), 25.0 ( $\text{CH}_3$ ), 22.5 ( $\text{CH}_3$ ).



**Compound n°13** → **PL<sub>2</sub>**: OMe-Ar(Me)-Pico(OH)-DO3A(tBu)<sub>3</sub>

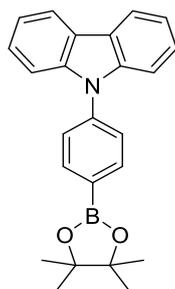
Compound **12** (61 mg, 0.25 mmol) and compound **11** (147 mg, 0.19 mmol) were dissolved in DMF (5 mL). Cesium fluoride (109 mg, 0.72 mmol) and *tetrakis*(triphenylphosphine)palladium(0) (23 mg, 0.02 mmol, 0.1 eq) were added. The reaction was stirred at 90 °C for 24 hours. The solvent was removed under reduce pressure. DCM was added and the solution was washed with water. The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent was removed under reduced pressure. The crude product was dissolved in EtOH (20 mL), NaOH 6M (4 mL) was added dropwise and the reaction was stirred for 20 min. The reaction was neutralized using HCl (6M). After extraction with AcOEt, the organic layer was washed with water and dried over Na<sub>2</sub>SO<sub>4</sub>. The solvent was removed under reduce pressure. The crude product was purified on HPLC to give a yellow powder (98 mg, 67%).

<sup>1</sup>H NMR in CDCl<sub>3</sub>, 298K, 400 MHz : δ (ppm) = 8.13 (s, 1H), 7.50 (s, 1H), 7.15 (d, <sup>3</sup>J = 8.81 Hz, 1H), 6.83 (s, 1H), 6.82 (dd, <sup>3</sup>J = 8.81 Hz, <sup>4</sup>J = 2.45 Hz, 1H), 4.52 (s, 2H), 3.89 (s, 2H), 3.84 (s, 3H), 3.58 – 3.27 (m, 20H), 2.29 (s, 3H) 1.43 (s, 9H), 1.38 (s, 18H).

<sup>13</sup>C NMR in CDCl<sub>3</sub>, 298K, 100 MHz : δ (ppm) = 168.3 (C<sub>quat</sub>), 167.3 (C<sub>quat</sub>), 166.1 (C<sub>quat</sub>), 161.1 (q, J = 37 Hz, TFA), 160.5 (C<sub>quat</sub>), 153.2 (C<sub>quat</sub>), 148.0 (C<sub>quat</sub>), 136.9 (C<sub>quat</sub>), 131.0 (CH), 129.8 (C<sub>quat</sub>), 127.4 (CH), 125.8 (CH), 116.6 (CH), 116.1 (q, J = 291 Hz, TFA), 112.1 (CH), 84.3 (C<sub>quat</sub>), 83.6 (C<sub>quat</sub>), 57.6 (CH<sub>2</sub>), 55.5 (CH<sub>3</sub>), 55.3 (CH<sub>2</sub>), 54.6 (CH<sub>2</sub>), 50.9 (CH<sub>2</sub>), 50.6 (CH<sub>2</sub>), 49.5 (CH<sub>2</sub>), 28.1 (CH<sub>3</sub>), 28.0 (CH<sub>3</sub>), 20.7 (CH<sub>3</sub>).

LRMS (ESI+): monoisotopic *m/z* = 770.7 [M+H]<sup>+</sup>, 792.7 [M+Na]<sup>+</sup> (calculated monoisotopic *m/z* = 770.5 [M+H]<sup>+</sup>, 792.7 [M+Na]<sup>+</sup>, for M = C<sub>41</sub>H<sub>63</sub>N<sub>5</sub>O<sub>9</sub>).

HRMS (ESI+): *m/z* = 770.4703 (+) (calculated *m/z* = 770.4699 [M+H]<sup>+</sup> for M = C<sub>41</sub>H<sub>64</sub>N<sub>5</sub>O<sub>9</sub>).



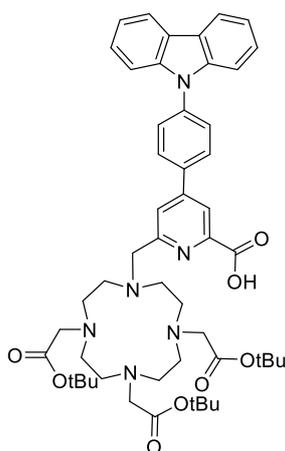
**Compound n°14:** 9-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)-9H-carbazole

Protocol adapted from: *Org. Lett.* **2004**, 6, 24, 4419–4422

To a solution of 9-(4-bromophenyl)-9H-carbazole (600 mg, 1.9 mmol) and *bis*(pinacolato)diboron (526 mg, 2.1 mmol) in dioxane (15 mL) were added [1,1'-*bis*(diphenylphosphino) ferrocene]palladium(II) dichloride (25 mg, 0.03 mmol) and sodium acetate (550 mg, 6 mmol). Argon was bubbled into the solution for 30min then the reaction was stirred at 90 °C during 48 hours under argon. The solvent was removed under reduce pressure. The solid residue was dissolved in DCM and the organic solution was washed with water and dried over Na<sub>2</sub>SO<sub>4</sub>. The solvent was removed under reduced pressure. The crude product was purified by flash chromatography (silica gel, Cyclohexane/DCM 1:1) to give a white powder (349 mg, 88%).

<sup>1</sup>H NMR in CDCl<sub>3</sub>, 298K, 400 MHz :  $\delta$  (ppm) = 8.14 (d, <sup>3</sup>J = 7.7 Hz, 2H), 8.06 (d, <sup>3</sup>J = 8.1 Hz, 2H), 7.60 (d, <sup>3</sup>J = 8.3 Hz, 2H), 7.46 - 7.38 (m, 4H), 7.29 (td, <sup>3</sup>J = 6.9 Hz, <sup>4</sup>J = 1.5 Hz, 2H), 1.40 (s, 12H).

<sup>13</sup>C NMR in CDCl<sub>3</sub>, 298K, 100 MHz :  $\delta$  (ppm) = 140.7 (C<sub>quat</sub>), 140.5 (C<sub>quat</sub>), 136.5 (CH), 126.2 (CH), 126.1 (C<sub>quat</sub>), 123.7 (CH), 120.4 (CH), 120.2 (CH), 110.0 (CH), 84.2 (C<sub>quat</sub>), 25.1 (CH<sub>3</sub>).



**Compound n°15** → **PL<sub>3</sub>**: Cz-Ar-Pico(OH)-DO3A(tBu)<sub>3</sub>

Compound **14** (44 mg, 0.1 mmol), compound **11** (79 mg, 0.12 mmol), cesium fluoride (55 mg, 0.36 mmol) and *tetrakis*(triphenylphosphine)palladium(0) (polymer-bound, 25 mg, 0.01 mmol) were

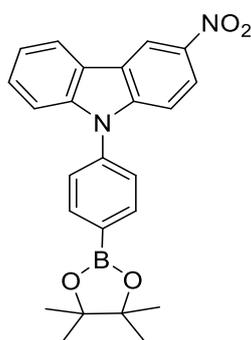
dissolved in anhydrous DMF (2.5 mL). The reaction was stirred at 90 °C for 24 hours under argon. The mixture was filtered on celite and the solvent was removed under reduce pressure. AcOEt was added and the solution was washed with saturated solution of NaHCO<sub>3</sub> then water. The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent was removed under reduced pressure. The crude product was dissolved in EtOH (20 mL), NaOH 6M (4 mL) was added dropwise and the reaction was stirred for 20 min. The reaction was neutralized HCl (6M). A precipitate was formed. The solution was centrifugated, the supernatant was diluted in AcOEt and the organic solution was washed with brine then water and dried over Na<sub>2</sub>SO<sub>4</sub>. The solvent was removed under reduced pressure. The product obtained was pure enough to be used without further purification (74 mg, 83%). A part of the crude product was purified on HPLC for characterisation.

<sup>1</sup>H NMR in CDCl<sub>3</sub>, 298K, 400 MHz :  $\delta$  (ppm) = 8.45 (d, <sup>4</sup>J = 1,3 Hz, 1H), 8.12 (d, <sup>3</sup>J = 7.7 Hz, 2H), 7.99 (d, <sup>3</sup>J = 8.5 Hz, 2H), 7.92 (s, 1H), 7.72 (d, <sup>3</sup>J = 8.4 Hz, 2H), 7.46 - 7.36 (m, 4H), 7.31 (td, <sup>3</sup>J = 8.5 Hz, <sup>3</sup>J = 7.3 Hz, <sup>4</sup>J = 1.3 Hz, 2H), 4.63 (s, 2H), 3.90 (s, 2H), 3.8 – 2.8 (br, 20H), 1.44 (s, 9H), 1.36 (s, 18H).

<sup>13</sup>C NMR in CDCl<sub>3</sub>, 298K, 100 MHz :  $\delta$  (ppm) = 168.9 (C<sub>quat</sub>), 165.9 (C<sub>quat</sub>), 161.1 (q, J = 37 Hz, TFA), 150.5 (C<sub>quat</sub>), 149.1 (C<sub>quat</sub>), 140.4 (C<sub>quat</sub>), 139.6 (C<sub>quat</sub>), 134.9 (C<sub>quat</sub>), 128.8 (CH), 127.5 (CH), 126.2 (CH), 125.1 (CH), 123.7 (C<sub>quat</sub>), 122.6 (C<sub>quat</sub>), 120.5 (CH), 116.4 (q, J = 291 Hz, TFA), 109.7 (CH), 84.2 (C<sub>quat</sub>), 83.2 (C<sub>quat</sub>), 57.6 (CH<sub>2</sub>), 55.2 (CH<sub>2</sub>), 54.6 (CH<sub>2</sub>), 51.0 (CH<sub>2</sub>), 50.7 (CH<sub>2</sub>), 49.5 (CH<sub>2</sub>), 49.2 (CH<sub>2</sub>), 27.9 (CH<sub>3</sub>).

LRMS (ESI+): monoisotopic  $m/z$  = 913.6 [M+Na]<sup>+</sup> (calculated monoisotopic  $m/z$  = 913.5 [M+Na]<sup>+</sup>).

HRMS (ESI+):  $m/z$  = 891.5021 (+) (calculated  $m/z$  = 891.5015 [M+H]<sup>+</sup> for M = C<sub>51</sub>H<sub>67</sub>N<sub>6</sub>O<sub>8</sub>).



**Compound n°16:** 3-nitro-9-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)-9H-carbazole

Protocol adapted from: *Org. Lett.* **2004**, 6, 24, 4419–4422

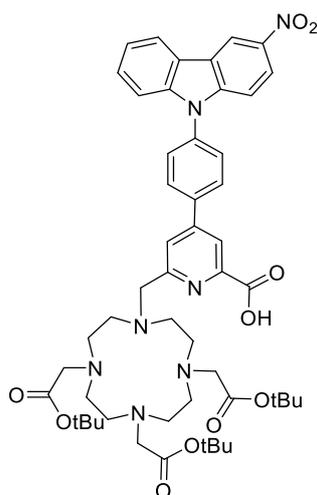
Copper nitrate trihydrate (82 mg, 0.34 mmol) was dissolved in 0.7 mL acetic acid then acetic anhydride was added (1.2 mL). After 10 min, 9-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)-9H-

carbazole (250 mg, 0.68 mmol) was added and the reaction was stirred 30 min at room temperature. In an Erlenmeyer containing 10 mL of water was added the reaction mixture. A brown precipitate was formed and the mixture was filtered and the solid recovered in DCM. The organic phase was dried on  $\text{Na}_2\text{SO}_4$  and the solvent was removed under reduced pressure. The crude product was purified by flash chromatography (silica gel, MeOH/DCM 2.5%) to give an oil (189 mg, 66%).

$^1\text{H}$  NMR in  $\text{CDCl}_3$ , 298K, 400 MHz :  $\delta$  (ppm) = 9.07 (d,  $^4J = 2.31$  Hz, 1H), 8.32 (dd,  $^3J = 9.03$  Hz,  $^4J = 2.31$  Hz, 1H), 8.20 (d,  $^3J = 7.92$  Hz, 1H), 8.09 (d,  $^3J = 8.26$  Hz, 2H), 7.56 (d,  $^3J = 8.37$  Hz, 2H), 7.51 (td,  $^3J = 7.67$  Hz,  $^3J = 6.97$  Hz,  $^4J = 1.22$  Hz, 2H), 7.43 - 7.38 (m, 3H), 1.41 (s, 12H)

$^{13}\text{C}$  NMR in  $\text{CDCl}_3$ , 298K, 100 MHz :  $\delta$  (ppm) = 143.9 ( $\text{C}_{\text{quat}}$ ), 142.3 ( $\text{C}_{\text{quat}}$ ), 141.6 ( $\text{C}_{\text{quat}}$ ), 139.0 ( $\text{C}_{\text{quat}}$ ), 136.8 (CH), 127.8 (CH), 126.3 (CH), 123.4 ( $\text{C}_{\text{quat}}$ ), 123.2 ( $\text{C}_{\text{quat}}$ ), 122.0 (CH), 121.8 (CH), 121.1 (CH), 117.4 (CH), 110.9 (CH), 109.7 (CH), 84.4 ( $\text{C}_{\text{quat}}$ ), 25.1 ( $\text{CH}_3$ ).

LRMS (ESI+): monoisotopic  $m/z = 415.2$   $[\text{M}+\text{H}]^+$  (calculated monoisotopic  $m/z = 415.2$   $[\text{M}+\text{H}]^+$ , for  $\text{M} = \text{C}_{24}\text{H}_{24}\text{N}_2\text{O}_4\text{B}$ ).



**Compound n°17**  $\rightarrow$  **PL4**: Cz(NO<sub>2</sub>)-Ar-Pico(OH)-DO3A(tBu)<sub>3</sub>

Compound **16** (94.5 mg, 0.23 mmol), compound **11** (150 mg, 0.19 mmol), cesium fluoride (87 mg, 0.57 mmol) and supported *tetrakis*(triphenylphosphine)palladium(0) (50 mg, 0.02 mmol) were dissolved in anhydrous DMF (5 mL). The reaction was stirred at 90 °C for 24 hours under argon. The mixture was filtered on celite and the solvent was removed under reduced pressure. AcOEt was added and the solution was extracted with a saturated solution of  $\text{NaHCO}_3$  and water. The organic layer was dried over  $\text{Na}_2\text{SO}_4$  and the solvent was removed under reduced pressure. The crude product was dissolved in EtOH (20 mL), NaOH 6M (4 mL) was added dropwise and the reaction was stirred for 20 min. The reaction was neutralized using HCl (6M). A precipitate was formed. The solution was centrifugated, the supernatant was diluted in AcOEt and the organic solution was washed with brine then water and dried

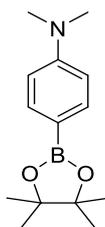
over Na<sub>2</sub>SO<sub>4</sub>. The solvent was removed under reduce pressure. The product obtained was pure enough to be used without further purification (155 mg, 89%). A part of the crude product was purified on HPLC for characterisation.

<sup>1</sup>H NMR in CD<sub>3</sub>OD, 298K, 400 MHz :  $\delta$  (ppm) = 9.17 (td, <sup>4</sup>J = 2.64 Hz, <sup>4</sup>J = 1.7 Hz, <sup>4</sup>J = 0.7 Hz, 1H), 8.57 (s, 1H), 8.36 (m, 2H), 8.15 (t, <sup>4</sup>J = 1.9 Hz, <sup>4</sup>J = 1.4 Hz, 2H), 8.14 (s, 1H), 7.84 (dd, <sup>3</sup>J = 7.83 Hz, <sup>4</sup>J = 1.1 Hz, 2H), 7.59-7.42 (m, 4H), 4.76 (s, 2H), 4.42-2.9 (br, 22H), 1.50 (s, 9H), 1.36 (s, 18H).

<sup>13</sup>C NMR in CD<sub>3</sub>OD, 298K, 100 MHz :  $\delta$  (ppm) = 145.0 (C<sub>quat</sub>), 143.4 (C<sub>quat</sub>), 143.2 (C<sub>quat</sub>), 130.4 (CH), 129.1 (CH), 129.0 (CH), 124.8 (C<sub>quat</sub>), 124.6 (C<sub>quat</sub>), 123.1 (C<sub>quat</sub>), 122.9 (CH), 122.6 (CH), 122.3 (CH), 118.2 (CH), 111.5 (CH), 110.7 (CH), 55.2 (CH<sub>2</sub>), 28.4 (CH<sub>3</sub>), 28.3 (CH<sub>3</sub>).

LRMS (ESI+): monoisotopic  $m/z$  = 936.5 [M+H]<sup>+</sup>, 974.4 [M+K]<sup>+</sup> (calculated monoisotopic  $m/z$  = 936.5 [M+H]<sup>+</sup>, 974.4 [M+K]<sup>+</sup> for M = C<sub>51</sub>H<sub>65</sub>N<sub>7</sub>O<sub>10</sub>).

HRMS (ESI+):  $m/z$  = 936.4870 (+) (calculated  $m/z$  = 936.4866 [M+H]<sup>+</sup> for M = C<sub>51</sub>H<sub>66</sub>N<sub>7</sub>O<sub>10</sub>).



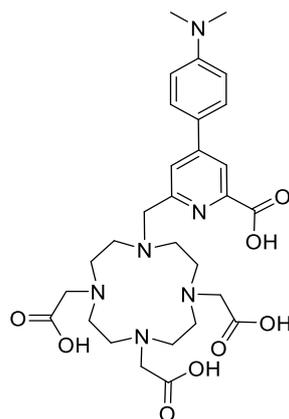
**Compound n°18:** N,N-dimethyl-4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)aniline

Protocol adapted from: *Org. Lett.* **2004**, 6, 24, 4419–4422

4-bromo-N,N-dimethylaniline (200 mg, 1 mmol), *bis*(pinacolato)diboron (435  $\mu$ L, 3 mmol) and *bis*(triphenylphosphine)palladium(II) dichloride (7 mg, 0.1 mmol) were dissolved in dioxane (3 mL). The solution was degassed using argon for 20 min and 800  $\mu$ L of degassed triethylamine (6 mmol) was added. The reaction was stirred at 90 °C for 15 hours under argon. The solvent was evaporated and the solid obtained was recovered using AcOEt and then washed with water. The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent was removed under reduced pressure. The crude product was purified by flash chromatography (silica gel, Petroleum ether/DCM 2:1) to give a white powder (190 mg, 77 %).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  (ppm): 7.71 (d, <sup>3</sup>J = 8.67 Hz, 2H), 6.71 (d, <sup>3</sup>J = 8.81 Hz, 2H), 3.01 (s, 6H), 1.35 (s, 12H).

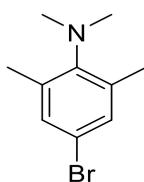
<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>):  $\delta$  (ppm): 152.7 (C<sub>quat</sub>), 136.3 (CH), 111.4 (CH), 83.3 (C<sub>quat</sub>), 40.3 (CH<sub>3</sub>), 25.0 (CH<sub>3</sub>).



**Compound n°19** → *L5*: N(Me)<sub>2</sub>-Ar-Pico(OH)-DO3A(OH)<sub>3</sub>

Compound **18** (89 mg, 0.36 mmol) and compound **6** (264 mg, 0.4 mmol) were dissolved in DMF (3.5 mL). Cesium fluoride (165 mg, 1.1 mmol) and *tetrakis*(triphenylphosphine)palladium(0) (42 mg, 0.04 mmol) were added. The reaction was stirred during 15 hours at 90 °C. The solvent was removed under reduced pressure. DCM was added and the solution was washed with water. The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent was removed under reduced pressure. The crude product was dissolved in EtOH (20 mL), NaOH 6M (4 mL) was added dropwise and the reaction was stirred for 20 min. The reaction was neutralized with HCl (6M). After extraction with AcOEt, the organic layer was washed with water and dried over Na<sub>2</sub>SO<sub>4</sub>. The solvent was removed under reduce pressure. The crude product was purified on HPLC to give a red solid (145 mg, 61%)

<sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O): δ (ppm) = 8.52 (d, <sup>4</sup>J = 1.37 Hz, 1H), 8.43 (d, <sup>4</sup>J = 1.47 Hz, 1H), 8.10 (d, <sup>3</sup>J = 8.98 Hz, 2H), 7.68 (d, <sup>3</sup>J = 9.07 Hz, 2H), 3.56-3.10 (m, 24H), 3.33 (s, 6H).



**Compound n°20**: 4-bromo-N,N,2,6-tetramethylaniline

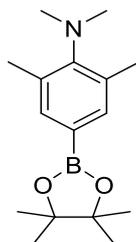
Protocol adapted from: *ChemBioChem* **2014**, 15, 399-412

To a solution of 4-bromo-2,6-dimethylaniline (2.06 g, 10 mmol) and potassium carbonate (4.14 g, 30 mmol) in DMF (10 mL) was added iodomethane (1.25 mL, 20 mmol). The reaction was stirred at room temperature for 15 hours. A solution of potassium hydroxide (5 eq) was added and the reaction was stirred 20 min. The solution was extracted using AcOEt and a saturated solution of NH<sub>4</sub>Cl. The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent was removed under reduced pressure. The crude product

was purified by flash chromatography (silica gel, Cyclohexane/DCM 7:3) to give a white powder (1.93g, 85 %).

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm): 7.14 (s, 2H), 2.80 (s, 6H), 2.27 (s, 6H).

$^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm): 149.0 ( $\text{C}_{\text{quat}}$ ), 139.5 (CH), 131.5 ( $\text{C}_{\text{quat}}$ ), 117.8 ( $\text{C}_{\text{quat}}$ ), 42.5 ( $\text{CH}_3$ ), 19.1 ( $\text{CH}_3$ ).

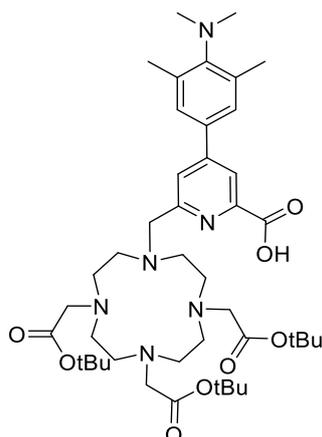


**Compound n°21:** N,N,2,6-tetramethyl-4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)aniline

Protocol adapted from: *Org. Lett.* **2004**, 6, 24, 4419–4422

To a solution of compound **20** (509 mg, 2.23 mmol) and *bis*(pinacolato)diboron (610 mg, 2.4 mmol) in dioxane (15 mL) were added [1,1'-*bis*(diphenylphosphino) ferrocene]palladium(II) dichloride (160 mg, 0.22 mmol) and sodium acetate (540 mg, 6.6 mmol). The reaction was stirred at 90 °C for 20 hours under argon. The solvent was evaporated and the solid obtained was extracted using DCM and water. The organic layer was dried over  $\text{Na}_2\text{SO}_4$  and the solvent was removed under reduced pressure. The crude product was purified by flash chromatography (silica gel, Cyclohexane/DCM 2:1) to give a white powder (476 mg, 79%).

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm): 7.50 (s, 2H), 2.84 (s, 6H), 2.31 (s, 6H), 1.35 (s, 12H).



**Compound n°22** → **PL6**: N(Me)<sub>2</sub>-Ar(Me)<sub>2</sub>-Pico(OH)-DO3A(tBu)<sub>3</sub>

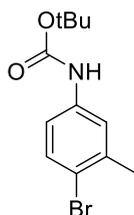
Compound **21** (53 mg, 0.19 mmol) and compound **11** (170 mg, 0.22 mmol) were dissolved in DMF (5 mL). Cesium fluoride (82 mg, 0.54 mmol) and *tetrakis*(triphenylphosphine)palladium(0) (21 mg, 0.02 mmol) were added. The reaction was stirred at 90 °C for 15 hours. The solvent was removed under reduced pressure. DCM was added and the solution was washed with water. The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent was removed under reduced pressure. The crude product was dissolved in EtOH (20 mL), NaOH 6M (4 mL) was added dropwise and the reaction was stirred for 20 min. The reaction was neutralized with HCl (6M). After extraction with AcOEt, the organic layer was washed with water and dried over Na<sub>2</sub>SO<sub>4</sub>. The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent was removed under reduced pressure to give a yellow raw product (108 mg, 74%). A part of the crude product was purified on HPLC for characterisation.

<sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD): δ (ppm) = 8.38 (s, 1H), 7.97 (s, 1H), 7.48 (s, 2H), 4.69 (s, 1H), 4.02-3.11 (m, 24H), 3.03 (s, 6H), 2.47 (s, 6H), 1.49 (s, 9H), 1.32 (s, 18H).

<sup>13</sup>C NMR (100 MHz, CD<sub>3</sub>OD): δ (ppm) = 167.3 (C<sub>quat</sub>), 161.7 (q, *J* = 37 Hz, TFA), 151.7 (C<sub>quat</sub>), 149.8 (C<sub>quat</sub>), 137.7 (C<sub>quat</sub>), 129.4 (CH), 126.4 (CH), 123.7 (CH), 119.1 (C<sub>quat</sub>), 116.1 (q, *J* = 291 Hz, TFA), 83.9 (C<sub>quat</sub>), 58.3 (CH<sub>2</sub>), 56.2 (CH<sub>2</sub>), 55.1 (CH<sub>2</sub>), 50.7 (CH<sub>2</sub>), 43.7 (CH<sub>3</sub>), 28.5 (CH<sub>3</sub>), 28.4 (CH<sub>3</sub>), 19.6 (CH<sub>3</sub>).

LRMS (ESI<sup>+</sup>): monoisotopic *m/z* = 797.7 [M+H]<sup>+</sup>, 819.7 [M+Na]<sup>+</sup> (calculated monoisotopic *m/z* = 797.7 [M+H]<sup>+</sup>, 819.7 [M+Na]<sup>+</sup> for M = C<sub>43</sub>H<sub>68</sub>N<sub>6</sub>O<sub>8</sub>).

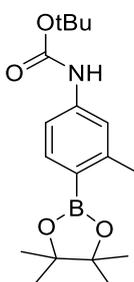
HRMS (ESI<sup>+</sup>): *m/z* = 797.5172 (+) (calculated *m/z* = 797.5171 [M+H]<sup>+</sup> for M = C<sub>43</sub>H<sub>69</sub>N<sub>6</sub>O<sub>8</sub>).

**Compound n°23:** tert-butyl (4-bromo-3-methylphenyl)carbamate

4-bromo-3-methylaniline (462 mg, 2.5 mmol) was dissolved in water (2.5 mL). Di-*tert*-butyl decarbonate (600 mg, 2.75 mmol) was added with 1.5 mL of water. The reaction was stirred for 90 minutes at room temperature and a white precipitate was formed. The solution was extracted using AcOEt. The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent was removed under reduced pressure to give a white powder (705 mg, 99%).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ (ppm) = 7.40 (d, <sup>3</sup>J = 8.54 Hz, 1H), 7.32 (d, <sup>4</sup>J = 2.48 Hz, 1H), 7.02 (dd, 7.71 (d, <sup>3</sup>J = 8.55 Hz, <sup>4</sup>J = 2.71 Hz, 1H), 6.43 (s, 1H), 2.36 (s, 3H), 1.51 (s, 9H).

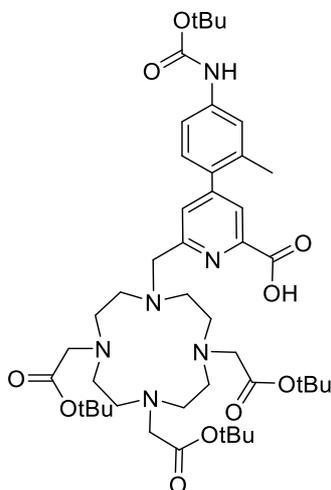
<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>): δ (ppm) = 152.7 (C<sub>quat</sub>), 138.6 (C<sub>quat</sub>), 137.7 (C<sub>quat</sub>), 132.7 (CH), 120.8 (CH), 118.1 (CH), 117.6 (CH), 80.9 (C<sub>quat</sub>), 28.5 (CH<sub>3</sub>), 27.6 (CH<sub>3</sub>), 23.2 (CH<sub>3</sub>).

**Compound n°24:** tert-butyl (3-methyl-4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)carbamate

Protocol adapted from: *Org. Lett.* **2004**, 6, 24, 4419–4422

To a solution of compound **23** (286 mg, 1 mmol, 1 eq) and *bis*(pinacolato)diboron (610 mg, 2.4 mmol) in DMF (15 mL) were added [1,1'-*bis*(diphenylphosphino)ferrocene]palladium(II) dichloride (25 mg, 0.03 mmol) and sodium acetate (265 mg, 3.2 mmol). The reaction was stirred at 90 °C for 48 hours under argon. The solvent was evaporated and the solid obtained was recovered using DCM and then washed with water. The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent was removed under reduced pressure. The crude product was purified by flash chromatography (silica gel, Cyclohexane/DCM 1:1) to give a white powder (138 mg, 41%).

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm): 7.68 (d,  $^3J = 8.03$  Hz, 1H), 7.21, (s, 1H), 7.11 (dd,  $^3J = 8.34$  Hz,  $^4J = 2.05$  Hz, 1H), 6.45 (s, 1H), 2.51 (s, 3H), 1.51 (s, 9H), 1.32 (s, 12H).



**Compound n°25**  $\rightarrow$  **PL<sub>7</sub>**: NH(Boc)-Ar(Me)-Pico(OH)-DO3A(tBu)<sub>3</sub>

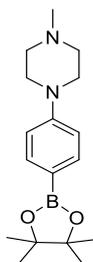
Compound **24** (62 mg, 0.19 mmol) and compound **11** (181 mg, 0.23 mmol) were dissolved in DMF (5 mL). Cesium fluoride (87 mg, 0.57 mmol) and *tetrakis*(triphenylphosphine)palladium(0) (22 mg, 0.02 mmol) were added. The reaction was stirred at 90 °C for 24 hours. The solvent was removed under reduced pressure. DCM was added and the solution was extracted with water. The organic layer was dried over  $\text{Na}_2\text{SO}_4$  and the solvent was removed under reduced pressure. The crude product was dissolved in EtOH (20 mL), NaOH 6M (4 mL) was added dropwise and the reaction was stirred for 20 min. The reaction was neutralized with HCl (6M). After extraction with AcOEt, the organic layer was washed with water and dried over  $\text{Na}_2\text{SO}_4$ . The solvent was removed under reduce pressure. The crude product was purified on HPLC to give a yellow powder (81 mg, 50%).

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm): 8.12 (d,  $^4J = 1.29$  Hz, 1H), 7.48 (s, 1H), 7.38 (s, 1H), 7.13 (d,  $^3J = 8.2$ Hz, 1H), 6.62 (s, 1H), 4.49 (s, 2H), 3.84 (s, 2H), 3.54-3.02 (m, 22H), 2.27 (s, 3H), 1.54 (s, 9H), 1.43 (s, 9H), 1.38 (s, 18H).

$^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm): 168.1 ( $\text{C}_{\text{quat}}$ ), 166.0 ( $\text{C}_{\text{quat}}$ ), 161.4 (q,  $J = 37$  Hz, TFA), 153.0 ( $\text{C}_{\text{quat}}$ ), 152.8 ( $\text{C}_{\text{quat}}$ ), 148.1 ( $\text{C}_{\text{quat}}$ ), 142.7, 139.6 ( $\text{C}_{\text{quat}}$ ), 136.2 ( $\text{C}_{\text{quat}}$ ), 131.9 ( $\text{C}_{\text{quat}}$ ), 130.3 (CH), 127.2 (CH), 125.7 (CH), 120.7 (CH), 117.7 (q,  $J = 291$  Hz, TFA), 116.6 (CH), 83.7 ( $\text{C}_{\text{quat}}$ ), 81.2 ( $\text{C}_{\text{quat}}$ ), 55.4 ( $\text{CH}_2$ ), 54.7 ( $\text{CH}_2$ ), 50.9 ( $\text{CH}_2$ ), 50.6 ( $\text{CH}_2$ ) 49.6 ( $\text{CH}_2$ ), 28.5 ( $\text{CH}_3$ ), 28.1 ( $\text{CH}_3$ ), 27.9 ( $\text{CH}_3$ ), 20.6 ( $\text{CH}_3$ ).

LRMS (ESI<sup>+</sup>): monoisotopic  $m/z = 855.7$   $[\text{M}+\text{H}]^+$ , 877.7  $[\text{M}+\text{Na}]^+$  (calculated monoisotopic  $m/z = 855.6$   $[\text{M}+\text{H}]^+$ , 877.6  $[\text{M}+\text{Na}]^+$  for  $\text{M} = \text{C}_{45}\text{H}_{71}\text{N}_6\text{O}_{10}$ ).

HRMS (ESI<sup>+</sup>):  $m/z = 855.5237$  (+) (calculated  $m/z = 855.5226$   $[\text{M}+\text{H}]^+$  for  $\text{M} = \text{C}_{45}\text{H}_{71}\text{N}_6\text{O}_{10}$ ).



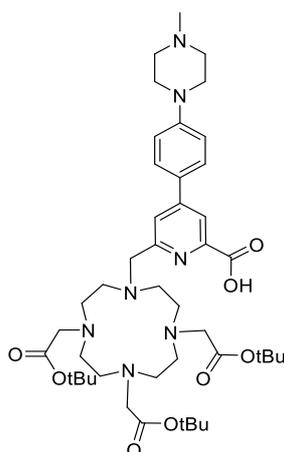
**Compound n°26:** 1-methyl-4-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)piperazine

Protocol adapted from: *Org. Lett.* **2004**, 6, 24, 4419–4422

To a solution of 1-(4-iodophenyl)-4-methylpiperazine (150 mg, 0.5 mmol) and *bis*(pinacolato)diboron (133 mg, 0.52 mmol) in dry DMSO (2 mL) were added [1,1'-*bis*(diphenylphosphino)ferrocene]palladium(II) dichloride (18.3 mg, 0.03 mmol) and sodium acetate (98 mg, 1 mmol). The reaction was stirred at 80 °C for 6 hours under argon. The reaction mixture was diluted in AcOEt. The solution was washed three times using a basic solution (pH 12), a saturated solution of NaCl and then water. The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent was removed under reduced pressure. The crude product was purified by flash chromatography (silica gel, MeOH/DCM, 5%) to give a powder (125 mg, 82%).

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ (ppm): 7.71 (d, <sup>3</sup>J = 8.66 Hz, 2H), 6.89 (d, <sup>3</sup>J = 8.63 Hz, 2H), 3.28 (t, <sup>3</sup>J = 6.84 Hz, <sup>3</sup>J = 5.03 Hz, 4H), 2.56 (t, <sup>3</sup>J = 6.55 Hz, <sup>3</sup>J = 5.13 Hz, 4H), 2.34 (s, 3H), 1.32 (s, 12H).

<sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>): δ (ppm): 152.4 (C<sub>quat</sub>), 136.2 (CH), 129.2 (CH), 119.8 (C<sub>quat</sub>), 116.2 (CH), 114.5 (CH), 83.5 (C<sub>quat</sub>), 55.3 (CH<sub>2</sub>), 55.1 (CH<sub>2</sub>), 49.2 (CH<sub>2</sub>), 48.2 (CH<sub>2</sub>), 46.3 (CH<sub>3</sub>), 25.0 (CH<sub>3</sub>).



**Compound n°27** → **PL3**: Piperazine-Ar-Pico(OH)-DO3A(tBu)<sub>3</sub>

Compound **26** (60 mg, 0.2 mmol) and compound **11** (174 mg, 0.22 mmol) were dissolved in DMF (2.5 mL). Cesium fluoride (91 mg, 0.6 mmol) and *tetrakis*(triphenylphosphine)palladium(0) (22 mg, 0.02

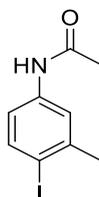
mmol) were added. The reaction was stirred at 90 °C for 24 hours. The solvent was removed under reduced pressure. DCM was added and the solution was washed with water. The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent was removed under reduced pressure. The crude product was dissolved in EtOH (20 mL), NaOH 6M (4 mL) was added dropwise and the reaction was stirred for 20 min. The reaction was neutralized with HCl (6M). After extraction with AcOEt, the organic layer was washed with water and dried over Na<sub>2</sub>SO<sub>4</sub>. The solvent was removed under reduce pressure. The crude product was purified on HPLC to give a yellow powder (58 mg, 35%).

<sup>1</sup>H NMR in (CD<sub>3</sub>)<sub>2</sub>CO, 298K, 400 MHz : δ (ppm) 8.37 (d, <sup>4</sup>J = 1.7 Hz, 1H), 8.11 (d, <sup>4</sup>J = 1.7 Hz, 1H), 7.86 (dt, <sup>3</sup>J = 8.9 Hz, <sup>4</sup>J = 2.15 Hz, 2H), 7.23 (td, <sup>3</sup>J = 9.02 Hz, <sup>4</sup>J = 2.09 Hz, 2H), 4.77 (s, 2H), 4.02-3.22 (m, 30H), 2.97 (s, 3H), 1.48 (s, 9H), 1.41 (s, 18H).

<sup>13</sup>C NMR in (CD<sub>3</sub>)<sub>2</sub>CO, 298K, 100 MHz : δ (ppm) = 152.4 (C<sub>quat</sub>), 151.3 (C<sub>quat</sub>), 149.7 (C<sub>quat</sub>), 129.0 (CH), 127.7 (C<sub>quat</sub>), 126.1 (C<sub>quat</sub>) 125.3 (CH), 121.7 (CH), 117.1 (CH), , 82.4 (C<sub>quat</sub>), 58.5 (CH<sub>2</sub>), 55.8 (CH<sub>2</sub>), 53.4 (CH<sub>2</sub>), 52.5 (CH<sub>2</sub>), 52.3 (CH<sub>2</sub>), 50.4 (CH<sub>2</sub>), 49.9 (CH<sub>2</sub>), 46.2 (CH<sub>3</sub>), 43.1 (CH<sub>3</sub>), 28.3 (CH<sub>3</sub>).

LRMS (ESI+): monoisotopic  $m/z$  = 824.7 [M+H]<sup>+</sup>, 846.7 [M+Na]<sup>+</sup> (calculated monoisotopic  $m/z$  = 824.5 [M+H]<sup>+</sup>, 846.5 [M+Na]<sup>+</sup> for M = C<sub>44</sub>H<sub>69</sub>N<sub>7</sub>O<sub>8</sub>).

HRMS (ESI+):  $m/z$  = 824.5287 (+) (calculated  $m/z$  = 824.5280 [M+H]<sup>+</sup> for M = C<sub>44</sub>H<sub>69</sub>N<sub>7</sub>O<sub>8</sub>).



**Compound n°28:** N-(4-iodo-3-methylphenyl)acetamide

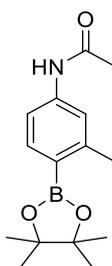
Protocol adapted from: *Adv. Synth. Catal.* **2005**, 347,1921-1924 and *Chem.Eur.J.* **2020**,26,10411-10416

4-Iodo-3-methylaniline (500 mg, 2.1 mmol) was dissolved in AcOEt (5 mL) with trimethylamine (800 μL, 6 mmol) and *N,N*-dimethylaminopyridine (24 mg, 0.2 mmol). The solution was stirred 20 min at 0 °C. A solution of acetic anhydride (200 μL, 2.1 mmol) in AcOEt (800 μL) was added dropwise. The solution was stirred 1 hour at 0 °C and the solution was warmed to room temperature and the reaction was stirred during 18 h. Water was added (2.5 mL). The organic layer was then washed tree times with brine, 2 M solution of HCl and water. The organic layer was dried using Na<sub>2</sub>SO<sub>4</sub>, filtered and the solvent was removed under reduced pressure. The solid obtained was sonicated in hexane (15 mL) for 20 min. The solution was filtered and the product was isolated as a grey solid used without purification (488 mg, 84%).

$^1\text{H}$  NMR in  $\text{CDCl}_3$ , 298K, 400 MHz :  $\delta$  (ppm) = 7.70 (d,  $^3J = 8,14$  Hz, 1H), 7.44 (s, 1H), 7.28 (br, 1H), 7.04 (d,  $^3J = 7.84$  Hz, 1H), 3.29 (s, 3H), 2.16 (s, 3H).

$^{13}\text{C}$  NMR in  $\text{CDCl}_3$ , 298K, 100 MHz :  $\delta$  (ppm) = 162.2 ( $\text{C}_{\text{quat}}$ ), 142.0 ( $\text{C}_{\text{quat}}$ ), 139.1 (CH), 138.3 ( $\text{C}_{\text{quat}}$ ), 121.4 (CH), 119.4 (CH), 94.6 ( $\text{C}_{\text{quat}}$ ), 28.2 ( $\text{CH}_3$ ), 24.5 ( $\text{CH}_3$ ).

LRMS (ESI+): monoisotopic  $m/z = 276.0$   $[\text{M}+\text{H}]^+$ , 298.0  $[\text{M}+\text{Na}]^+$  (calculated monoisotopic  $m/z = 276.0$   $[\text{M}+\text{H}]^+$ , 298.0  $[\text{M}+\text{Na}]^+$ , for  $\text{M} = \text{C}_9\text{H}_{10}\text{NOI}$ ).



**Compound n°29:** N-(3-methyl-4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)acetamide

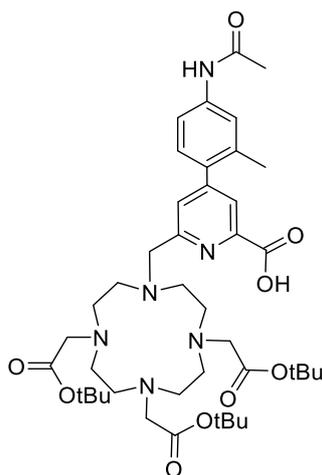
Protocol adapted from: *Org. Lett.* **2004**, 6, 24, 4419–4422

Compound **28** (300 mg, 1.1 mmol), *bis*(pinacolato)diboron (305 mg, 1.2 mmol), [1,1'-*bis*(diphenylphosphino)ferrocene]palladium(II) dichloride (40 mg, 0.05 mmol) and sodium acetate (214 mg, 2.2 mmol) were dissolved in dried DMSO (2 mL). The solution was degassed using argon during 30 min. The reaction was stirred at 90 °C for 16 hours under argon. The reaction mixture was diluted in AcOEt and filtrated under Celite. The organic layer was washed with a saturated solution of  $\text{NH}_4\text{Cl}$  twice and then with water. The organic layer was dried over  $\text{Na}_2\text{SO}_4$  and the solvent was removed under reduced pressure. The crude product was purified by flash chromatography (silica gel, MeOH/DCM, 2%) to give a colourless oil (275 mg, 91%).

$^1\text{H}$  NMR in  $\text{CDCl}_3$ , 298K, 400 MHz :  $\delta$  (ppm) = 7.71 (d,  $^3J = 8,10$  Hz, 1H), 7.33 (s, 1H), 7.29 (d,  $^3J = 8,26$  Hz, 1H), 2.52 (s, 3H), 2.16 (s, 3H), 1.33 (s, 12H).

$^{13}\text{C}$  NMR in  $\text{CDCl}_3$ , 298K, 100 MHz :  $\delta$  (ppm) = 168.4 ( $\text{C}_{\text{quat}}$ ), 146.5 ( $\text{C}_{\text{quat}}$ ), 140.2 ( $\text{C}_{\text{quat}}$ ), 137.2 (CH), 120.5 (CH), 115.8 (CH), 83.5 ( $\text{C}_{\text{quat}}$ ), 53.6 ( $\text{CH}_2$ ), 25.0 ( $\text{CH}_3$ ), 22.5 ( $\text{CH}_3$ ).

LRMS (ESI+): monoisotopic  $m/z = 276.2$   $[\text{M}+\text{H}]^+$  (calculated monoisotopic  $m/z = 276.2$   $[\text{M}+\text{H}]^+$ , for  $\text{M} = \text{C}_{15}\text{H}_{23}\text{NO}_3\text{B}$ ).



**Compound n°30** → **PL9**: AcNH-Ar(Me)-Pico(OH)-DO3A(tBu)<sub>3</sub>

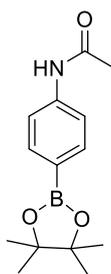
Compound **29** (60 mg, 0.23 mmol), compound **11** (160 mg, 0.2 mmol), cesium fluoride (90 mg, 0.59 mmol) and *tetrakis*(triphenylphosphine)palladium(0) (20 mg, 0.02 mmol) were dissolved in anhydrous DMF (2.5 mL). The reaction was stirred at 90 °C for 12 hours under argon. The reaction mixture was diluted using AcOEt and the solution was washed with saturated NaHCO<sub>3</sub> then water. The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent was removed under reduced pressure. The crude product was dissolved in EtOH (20 mL), NaOH 6M (4 mL) was added dropwise and the reaction was stirred for 20 min. The reaction was neutralized with HCl (6M). A precipitate was formed. The solution was centrifugated, the supernatant was diluted in AcOEt and the organic solution was washed with brine then water and dried over Na<sub>2</sub>SO<sub>4</sub>. The product obtained was purified on HPLC to give a white powder (84 mg, 55%).

<sup>1</sup>H NMR in (CD<sub>3</sub>)<sub>2</sub>CO, 298K, 400 MHz : δ (ppm) = 8.15 (s, 1H), 7.72 (d, <sup>4</sup>J = 1.1 Hz, 1H), 7.57 (m, 2H), 7.24 (d, <sup>3</sup>J = 8.1 Hz, 1H), 4.67 (s, 2H), 3.99-3.12 (m, 22H), 2.34 (s, 3H), 2.15 (s, 3H), 1.49 (s, 9H), 1.34 (s, 18H).

<sup>13</sup>C NMR in (CD<sub>3</sub>)<sub>2</sub>CO, 298K, 100 MHz : δ (ppm) = 171.9 (C<sub>quat</sub>), 167.4 (C<sub>quat</sub>), 162.5 (q, J = 38Hz, TFA) 153.3 (C<sub>quat</sub>), 149.0 (C<sub>quat</sub>), 141.0 (C<sub>quat</sub>), 137.1 (C<sub>quat</sub>), 134.3 (C<sub>quat</sub>), 131.2 (CH), 129.4 (CH), 126.8 (CH), 123.1 (CH), 119.5 (C<sub>quat</sub>), 119.0 (CH), 118.2 (q, J = 291 Hz, TFA), 85.4 (C<sub>quat</sub>), 83.9 (C<sub>quat</sub>), 58.1(CH<sub>2</sub>), 56.0 (CH<sub>2</sub>), 55.2 (CH<sub>2</sub>), 52.2 (CH<sub>2</sub>), 50.7 (CH<sub>2</sub>), 28.4 (CH<sub>3</sub>), 23.9 (CH<sub>3</sub>), 20.8 (CH<sub>3</sub>).

LRMS (ESI<sup>+</sup>): monoisotopic m/z = 797.5 [M+H]<sup>+</sup> (calculated monoisotopic m/z = 797.5 [M+H]<sup>+</sup>, for M = C<sub>42</sub>H<sub>64</sub>N<sub>6</sub>O<sub>9</sub>).

HRMS (ESI<sup>+</sup>): m/z = 797.4809 (+) (calculated m/z = 797.4808 [M+H]<sup>+</sup> for M = C<sub>42</sub>H<sub>64</sub>N<sub>6</sub>O<sub>9</sub>).



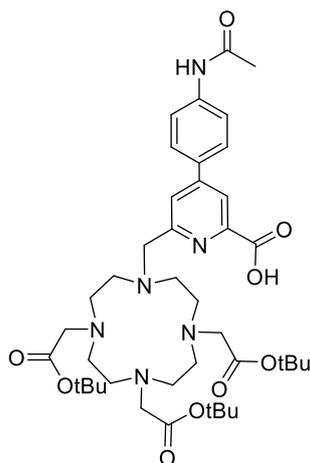
**Compound n°31:** N-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)acetamide

Protocol adapted from: *Org. Lett.* **2004**, 6, 24, 4419–4422

4-bromoacetanilide (300 mg, 1.4 mmol), *bis*(pinacolato)diboron (391 mg, 1.5 mmol), [1,1'-*bis*(diphenylphosphino)ferrocene]palladium(II) dichloride (102 mg, 0.1 mmol) and sodium acetate (275 mg, 2.8 mmol) were dissolved in DMSO (2 mL). The solution was degassed using argon during 30 min. The reaction was stirred at 90 °C for 24 hours under argon. The reaction mixture was diluted in AcOEt and washed with brine and water. The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent was removed under reduced pressure. The crude product was purified by flash chromatography (silica gel, AcOEt/Cyclohexane 30:70) to give a yellow oil (114 mg, 78%).

<sup>1</sup>H NMR in CDCl<sub>3</sub>, 298K, 400 MHz : δ (ppm) = 7.76 (d, <sup>3</sup>J = 8.48 Hz, 2H), 7.50 (d, <sup>3</sup>J = 8.50 Hz, 2H), 7.16 (s, 1H), 2.19 (s, 3H), 1.35 (s, 12H).

<sup>13</sup>C NMR in CDCl<sub>3</sub>, 298K, 100 MHz : δ (ppm) = 168.4 (C<sub>quat</sub>), 140.6 (C<sub>quat</sub>), 136.0 (CH), 118.6 (CH), 83.9 (C<sub>quat</sub>), 29.9 (CH<sub>3</sub>), 25.0 (CH<sub>3</sub>).



**Compound n°32** → *PL*<sub>10</sub>: AcNH-Ar-Pico(OH)-DO3A(tBu)<sub>3</sub>

Compound **31** (68 mg, 0.26 mmol), compound **11** (133 mg, 0.17 mmol), cesium fluoride (83 mg, 0.55 mmol) and supported *tetrakis*(triphenylphosphine)palladium(0) (25 mg, 0.01 mmol) were dissolved in

anhydrous DMF (2.5 mL). The reaction was stirred at 90 °C for 24 hours under argon. The mixture was filtered on celite and the solvent was removed under reduced pressure. AcOEt was added and the solution was extracted with a saturated solution of NaHCO<sub>3</sub> and water. The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and the solvent was removed under reduced pressure. The crude product was dissolved in EtOH (20 mL), NaOH 6M (4 mL) was added dropwise and the reaction was stirred for 20 min. The reaction was neutralized with HCl (6M). A precipitate was formed. The solution was centrifugated, the supernatant was diluted in AcOEt and the organic solution was washed with brine then water and dried over Na<sub>2</sub>SO<sub>4</sub>. The product was purified on HPLC to give a white powder (88 mg, 48%).

<sup>1</sup>H NMR in CD<sub>3</sub>OD, 298K, 400 MHz :  $\delta$  (ppm) = 8.41 (s, 1H), 8.01 (d, <sup>4</sup>J = 1.54 Hz, 1H), 7.76 (s, 4H), 4.67 (s, 2H), 4.20-3.10 (m, 22H), 2.17 (s, 3H), 1.48 (s, 9H), 1.32 (s, 18H).

<sup>13</sup>C NMR in CD<sub>3</sub>OD, 298K, 100 MHz :  $\delta$  (ppm) = 171.9 (C<sub>quat</sub>), 167.4 (C<sub>quat</sub>), 162.5 (q, J = 38 Hz, TFA), 151.6 (C<sub>quat</sub>), 149.8 (C<sub>quat</sub>), 142.1 (C<sub>quat</sub>), 132.7 (C<sub>quat</sub>), 128.8 (CH), 126.1 (CH), 123.3 (CH), 121.4 (CH), 118.0 (q, J = 38 Hz, TFA), 85.5 (C<sub>quat</sub>), 84.2 (C<sub>quat</sub>), 58.4 (CH<sub>2</sub>), 56.0 (CH<sub>2</sub>), 55.2 (CH<sub>2</sub>), 52.6 (CH<sub>2</sub>), 50.7 (CH<sub>2</sub>), 28.4 (CH<sub>3</sub>), 24.0 (CH<sub>3</sub>).

HRMS (ESI+):  $m/z$  = 783.4655 (+) (calculated  $m/z$  = 783.4651 [M+H]<sup>+</sup> for M = C<sub>41</sub>H<sub>63</sub>N<sub>6</sub>O<sub>9</sub>).

### **Metallation of Ln<sup>3+</sup> after tert-butyl déprotection of PL<sub>i</sub>**

The *tert*-butyl protecting group of the pro-ligand was removed using TFA during 30 min. The TFA was removed and the oil obtained was recovered in water. An excess (1.2 eq) of lanthanide salt (EuCl<sub>3</sub> or TbCl<sub>3</sub>) was added on the pH was adjusted at 9 using triethylamine. The reaction was stirred at room temperature overnight. The solution was centrifuged and the supernatant was treated with Et<sub>2</sub>O until the formation of a precipitate and the solution was centrifuged again. The solid obtained was then washed using acetone and diethyl ether. The complex obtained was used without purification.

### **Coupling of PL<sub>i</sub> on TAT**

The pro-ligand was coupled on 10  $\mu$ mol of peptide-resin F1 using classical peptidic conditions. A slight excess of the PL<sub>i</sub> (1.2 eq) was dissolved with PyBOP (2.4 eq) and DIEA (5 eq) in DMF. This solution was added on the resin and the reaction was stirred overnight at room temperature. The resin was washed with DMF, DCM and Et<sub>2</sub>O before to be dried.

## Peptidic synthesis

### Peptidic sequences on resin

All the amino acids are protected following the Fmoc/tBu synthesis strategy. For other protecting group they will be detailed in bracket. Non-natural amino acid will be in square bracket.

**F1-SEA:** *Fmoc – PMNLLNYVVPKMRPYK(alloc)C(StBu)PVEA – Sea Resin*

**F1-Nle-SEA:** *Fmoc – P[Nle]NLLNYVVPK[Nle]RPYK(alloc)C(StBu)PVEA – Sea Resin*

-----  
**Acm-F2-SEA-A:** *Fmoc – C(Acm)ERRFSRDELTRHIRIHTGQKPFA – Sea Resin*

**Thz-F2-SEA-A:** *Fmoc – [Thz]ERRFSRDELTRHIRIHTGQKPFA – Sea Resin*

**Acm-F2-SEA-B:** *Fmoc – C(Acm)ERRFSRSEHLTRHIRIHTGQKPFA – Sea Resin*

**Thz-F2-SEA-B:** *Fmoc – [Thz]ERRFSRSEHLTRHIRIHTGQKPFA – Sea Resin*

**N<sub>3</sub>-F2-NHNH<sub>2</sub>-A:** *Fmoc – C(N<sub>3</sub>)ERRFSRDELTRHIRIHTGQKPFA – NHNH<sub>2</sub>*

**Acm-F2-NHNH<sub>2</sub>-A:** *Fmoc – C(Acm)ERRFSRDELTRHIRIHTGQKPFA – NHNH<sub>2</sub>*

**Acm-F2-NHNH<sub>2</sub>-B:** *Fmoc – C(Acm)ERRFSRSEHLTRHIRIHTGQKPFA – NHNH<sub>2</sub>*

-----  
**F3-NH<sub>2</sub>-A:** *Fmoc – CRICMRNFSRSDHLTTHIRTHTGE – Rink Amid Resin*

**F3-Nle-NH<sub>2</sub>-A:** *Fmoc – CRIC[Nle]RNFSRSDHLTTHIRTHTGE – Rink Amid Resin*

**F3-NH<sub>2</sub>-B:** *Fmoc – CRICMRNFSQSGNLSRHIRTHTGE – Rink Amid Resin*

**F3-Nle-NH<sub>2</sub>-B:** *Fmoc – CRIC[Nle]RNFSQSGNLSRHIRTHTGE – Rink Amid Resin*

### Peptides

#### Fragment F1, F2 and F3

**H-F1-SEA<sup>off</sup>:** H-HPMNNLLNYVVPKMRPYKCPVEA-SEA<sup>off</sup>

**Ac-F1-SEA<sup>off</sup>:** H-HPMNNLLNYVVPKMRPYKCPVEA-SEA<sup>off</sup>

**H-F1-Nle-SEA<sup>off</sup>:** H-HP[Nle]NLLNYVVPK[Nle]RPYK(StBu)PVEA-SEA<sup>off</sup>

**H-F1-Nle-SiR-SEA<sup>off</sup>:** H-HP[Nle]NLLNYVVPK[Nle]RPYK(SiR)C(StBu)PVEA-SEA<sup>off</sup>

**H-F1-Nle-SiR-NHNH<sub>2</sub>:** H-HP[Nle]NLLNYVVPK[Nle]RPYK(SiR)C(StBu)PVEA-NHNH<sub>2</sub>

**H-F1-Nle-Lissa-SEA<sup>off</sup>:** H-HP[Nle]NLLNYVVPK[Nle]RPYK(Lissa)C(StBu)PVEA-SEA<sup>off</sup>

**Ac-F1-[L<sub>1</sub>-Eu]-SEA<sup>off</sup>:** Ac-H PMNNLLNYVVPKMRPYK([L<sub>1</sub>-Eu])CPVEA-SEA<sup>off</sup>

**H-F1-Nle-[L<sub>2</sub>-Tb]-SEA<sup>off</sup>:** H-HP[Nle]NLLNYVVPK[Nle]RPYK([L<sub>2</sub>-Tb])C(StBu)PVEA-SEA<sup>off</sup>

**H-F1-Nle-[L<sub>2</sub>-Eu]-SEA<sup>off</sup>**: H-HP[Nle]NLLNYVVPK[Nle]RPYK([L<sub>2</sub>-Tb])C(StBu)PVEA-SEA<sup>off</sup>

**H-F1-Nle-[L<sub>2</sub>-Tb]-NHNH<sub>2</sub>**: H-HP[Nle]NLLNYVVPK[Nle]RPYK([L<sub>2</sub>-Eu])C(StBu)PVEA-NHNH<sub>2</sub>

**H-F1-Nle-[L<sub>10</sub>-Eu]-SEA<sup>off</sup>**: H-HP[Nle]NLLNYVVPK[Nle]RPYK([L<sub>10</sub>-Eu])C(StBu)PVEA-SEA<sup>off</sup>

-----  
**Acm-F2-SEA<sup>off</sup>-A**: H-C(Acm)ERRFSRDELTRHIRIHTGQKPFA-SEA<sup>off</sup>

**Thz-F2-SEA<sup>off</sup>-A**: H-[Thz]ERRFSRDELTRHIRIHTGQKPFA-SEA<sup>off</sup>

**N<sub>3</sub>-F2-NHNH<sub>2</sub>-A**: H-C(N<sub>3</sub>)ERRFSRDELTRHIRIHTGQKPFA-NHNH<sub>2</sub>

**Acm-F2-NHNH<sub>2</sub>-A**: H-C(Acm)ERRFSRDELTRHIRIHTGQKPFA-NHNH<sub>2</sub>

**Acm-F2-SEA<sup>off</sup>-B**: H-C(Acm)ERRFSRSEHLTRHIRIHTGQKPFA-SEA<sup>off</sup>

**Thz-F2-SEA<sup>off</sup>-B**: H-[Thz]ERRFSRSEHLTRHIRIHTGQKPFA-SEA<sup>off</sup>

**Acm-F2-NHNH<sub>2</sub>-B**: H-C(Acm)ERRFSRSEHLTRHIRIHTGQKPFA-NHNH<sub>2</sub>

-----  
**H-F3-NH<sub>2</sub>-A**: H-CRICMRNFSRSDHLTTHIRHTHTGE-NH<sub>2</sub>

**H-F3-Nle-NH<sub>2</sub>-A**: H-CRIC[Nle]RNFSRSDHLTTHIRHTHTGE-NH<sub>2</sub>

**H-F3-NH<sub>2</sub>-B**: H-CRICMRNFSQSGNLSRHIRHTHTGE-NH<sub>2</sub>

**H-F3-Nle-NH<sub>2</sub>-B**: H-CRIC[Nle]RNFSQSGNLSRHIRHTHTGE-NH<sub>2</sub>

*Peptide F2-F3*

**Acm-F2-F3-NH<sub>2</sub>-A**:

H-C(Acm)ERRFSRDELTRHIRIHTGQKPFACRICMRNFSRSDHLTTHIRHTHTGE-NH<sub>2</sub>

**Acm-F2-F3-Nle-NH<sub>2</sub>-A**:

H-C(Acm)ERRFSRDELTRHIRIHTGQKPFACRIC[Nle]RNFSRSDHLTTHIRHTHTGE-NH<sub>2</sub>

**Acm-F2-F3-NH<sub>2</sub>-B**:

H-C(Acm)ERRFSRSEHLTRHIRIHTGQKPFACRICMRNFSQSGNLSRHIRHTHTGE-NH<sub>2</sub>

**Acm-F2-F3-Nle-NH<sub>2</sub>-B**:

H-C(Acm)ERRFSRSEHLTRHIRIHTGQKPFACRIC[Nle]RNFSQSGNLSRHIRHTHTGE-NH<sub>2</sub>

**Thz-F2-F3-NH<sub>2</sub>-B**:

H-[ThioP]ERRFSRSEHLTRHIRIHTGQKPFACRICMRNFSQSGNLSRHIRHTHTGE-NH<sub>2</sub>

*Peptide F1-F2-F3*

**H-F1-F2-F3-NH<sub>2</sub>-A: Zif12-A**

H-HPMNNLLNYVVPKMRPYKCPVEACERRFSRDELTRHIRIHTGQKPFA  
 CRICMRNFSRSDHLTTHIRHTHTGE-NH<sub>2</sub>

**H-F1-Nle-F2-F3-NH<sub>2</sub>-A: Zif12-A-Nle**

H-HP[Nle]NNLLNYVVPK[Nle]RPYKCPVEACERRFSRDELTRHIRIHTGQKPFA  
CRIC[Nle]RNFSRSDHLTTHIRTHTGE-NH<sub>2</sub>

**H-F1-SiR-F2-F3-NH<sub>2</sub>-A: Zif12-SiR-A**

H-HP[Nle]NNLLNYVVPK[Nle]RPYK(SiR)CPVEACERRFSRDELTRHIRIHTGQKPFA  
CRIC[Nle]RNFSRSDHLTTHIRTHTGE-NH<sub>2</sub>

**H-F1-Lissa-F2-F3-NH<sub>2</sub>-A: Zif12-Lissa-A**

H-HPNleNNLLNYVVPK[Nle]RPYK(Lissa)CPVEACERRFSRDELTRHIRIHTGQKPFA  
CRIC[Nle]RNFSRSDHLTTHIRTHTGE-NH<sub>2</sub>

**H-F1-[L<sub>2</sub>-Tb]-F2-F3-Nle-NH<sub>2</sub>-A: Zif12-[L<sub>2</sub>-Tb]-A**

H-HP[Nle]NNLLNYVVPK[Nle]RPYK-[L<sub>2</sub>-Eu]-CPVEACERRFSRDELTRHIRIHTGQKPFA  
CRIC[Nle]RNFSRSDHLTTHIRTHTGE-NH<sub>2</sub>

**H-F1-[L<sub>2</sub>-Eu]-F2-F3-Nle-NH<sub>2</sub>-A: Zif12-[L<sub>2</sub>-Eu]-A**

H-HP[Nle]NNLLNYVVPK[Nle]RPYK-[L<sub>2</sub>-Eu]-CPVEACERRFSRDELTRHIRIHTGQKPFA  
CRIC[Nle]RNFSRSDHLTTHIRTHTGE-NH<sub>2</sub>

**H-F1-[L<sub>10</sub>-Eu]-F2-F3-Nle-NH<sub>2</sub>-A: Zif12-[L<sub>10</sub>-Eu]-A**

H-HP[Nle]NNLLNYVVPK[Nle]RPYK-[L<sub>10</sub>-Eu]-CPVEACERRFSRDELTRHIRIHTGQKPFA  
CRIC[Nle]RNFSRSDHLTTHIRTHTGE-NH<sub>2</sub>

-----  
**H-F1-F2-F3-NH<sub>2</sub>-B: Zif12-B**

Ac-HPMNNLLNYVVPKMRPYKCPVEACERRFSRSEHLTRHIRIHTGQKPFA  
CRICMRNFSQSGNLSRHIRTHTGE-NH<sub>2</sub>

**H-F1-SiR-F2-F3-NH<sub>2</sub>-B: Zif12-SiR-B**

H-HP[Nle]NNLLNYVVPK[Nle]RPYK(SiR)CPVEACERRFSRSEHLTRHIRIHTGQKPFA  
CRIC[Nle]RNFSQSGNLSRHIRTHTGE-NH<sub>2</sub>

**H-F1-Lissa-F2-F3-NH<sub>2</sub>-B: Zif12-Lissa-B**

H-HP[Nle]NNLLNYVVPK[Nle]RPYK(Lissa)CPVEACERRFSRSEHLTRHIRIHTGQKPFA  
CRIC[Nle]RNFSQSGNLSRHIRTHTGE-NH<sub>2</sub>

**General procedure:**

**(1) Protein synthesis on SEA resin with alloc deprotection**

Peptide elongation was performed as described above by a peptide synthesizer on a SEA-PS resin (0.12 mmol/g) after introduction of the first amino acid by double manual coupling (2x30 min) using 8 eq. of Fmoc-Ala-OH, 7.5 eq. of HCTU and 16 eq. of DIEA in DMF with pre-activation followed by a capping step using Ac<sub>2</sub>O/pyridine in DMF (2:1:17 by vol., 10 mL, 10 min) and Fmoc removal using 20%

piperidine in DMF (3x3min each) with UV monitoring. The resin was washed with DMF, DCM and Et<sub>2</sub>O then dried.

Removal of acid-labile side chain protecting groups and cleavage of the peptidyl resin was performed with a mixture of TFA/H<sub>2</sub>O/TIS/thioanisole (91.7/3.3/3.3/1.7 by vol., 10 mL) during 4h. TFA was evaporated, and the peptide was then precipitated in ice-cold Et<sub>2</sub>O, dissolved in deionized water, and lyophilized. The peptide was dissolved in acetic acid solution (10 %) and treated with 200 mM I<sub>2</sub> in DMSO to oxidize the C-terminal SEA<sub>on</sub> group into SEA<sub>off</sub> group. DTT was finally added to quench the excess of iodine, and the oxidized peptide was immediately purified by HPLC.

#### (2) Protein synthesis on a 2-Cl-trityl resin functionalised by a hydrazine

The resin 2-Cl-trityl resin (1.60 mmol/g) was treated with a 5% solution of hydrazine (v/v) in DMF during 30 min. This step was made a second time. The resin was washed using DMF, DCM, and DMF. The resin was next capped using MeOH in DMF (5%, v/v) for 10 min and washed using DMF, DCM, and DMF. Peptide elongation was performed as described above by a peptide synthesizer on a SEA-PS resin (0.12 mmol/g) after introduction of the first amino acid by double manual coupling (2x30 min) using 8 eq. of Fmoc-AA -OH, 7.5 eq. of HCTU and 16 eq. of DIEA in DMF with pre-activation followed by a capping step using Ac<sub>2</sub>O/pyridine in DMF (2:1:17 by vol., 10 mL, 10 min) and Fmoc removal using 20% piperidine in DMF (3x3min each) with UV monitoring. The resin was washed with DMF, DCM and Et<sub>2</sub>O then dried.

Removal of acid-labile side chain protecting groups and cleavage of the peptidyl resin was performed with a mixture of TFA/H<sub>2</sub>O/TIS/thioanisole (91.7/3.3/3.3/1.7 by vol., 10 mL) during 4h. TFA was evaporated, and the peptide was then precipitated in ice-cold Et<sub>2</sub>O, dissolved in deionized water and purified by HPLC.

#### (3) Removal of the Lys(Alloc) protecting group on the resin

Alloc protecting group of Lys(Alloc) was removed using Pd(PPh<sub>3</sub>)<sub>4</sub> (0.5 eq) and phenylsilane (25 eq) in degassed DCM (10 mL) for 1h. The process was repeated twice. The resin was then washed successively with DCM (x2), DMF (x2), 1% H<sub>2</sub>O in DMF (3min), DMF (x2), 1% DIEA in DMF (3min), DMF (x2), 0.12 M sodium diethyldithiocarbamate in DMF (3min).

#### (4) Coupling of the pro-ligand **PL<sub>i</sub>** and protected cysteine

The pro-ligand was coupled on 10 μmol of peptide-resin F1 using classical peptidic conditions. A slight excess of the **PL<sub>i</sub>** (1.2 eq) was dissolved with PyBOP (2.4 eq) and DIEA (5 eq) in DMF. This solution was added on the resin and the reaction was stirred overnight at room temperature. The resin was washed with DMF, DCM and Et<sub>2</sub>O before dried.

For the coupling of the protected cysteine for the F2 fragment, the protocol is the same as the one described previously. This solution was added on the resin and the reaction was stirred overnight at room temperature. The resin was washed with DMF, DCM and Et<sub>2</sub>O before dried.

**(5) Methyl ester hydrolysis**

Methyl ester of the *PL<sub>I</sub>* have been hydrolysed on the purified protein. The protein was placed in a 0.5 M solution of NaOH du ring 10 min. The reaction was monitored on LCMS. Then the solution was acidified using a 0.1% TFA water solution before purification on HPLC. The purified product was then lyophilised.

**(6) Complexation of the ligand by the lanthanide on the peptide**

The peptide was dissolved in deionized water (1 mL) and the pH was adjusted around 7. A slight excess of hydrated lanthanide salt (1.2 eq of EuCl<sub>3</sub> or TbCl<sub>3</sub>) was dissolved in the solution. The pH of the solution was adjusted at 7.5 using a NaOH solution. The reaction was stirred overnight at room temperature. The solution was then recovered in 0.1% TFA water solution before purification on HPLC. The purified product was then lyophilised.

**(7) Native Chemical Ligation using the SEA method**

A degassed solution of TCEP (29 mg) and MPAA (17 mg) in 0.9 mL of a 6 M solution of guanidinium hydrochloride in a phosphate solution (100 mM, pH 7.0) was prepared. The pH of the solution was adjusted at 6.5 and degassed 30 min using argon. The two peptides are placed in a glove box, a slight excess of one peptide was used (from 1.1 to 1.3 eq) and then dissolved in the previous solution. The volume of the added solution was calculated in order to have a 3 mM solution of the limiting peptide. The reaction was placed and stirred at 37 °C in the glove box. The reaction was monitored by LCMS after treatment of the aliquot by a 10% TFA solution and extraction of the MPAA using diethyl ether. When the ligation was completed the reaction mixture was treated using a 10% TFA solution and extracted using Et<sub>2</sub>O. The crude product was directly purified on HPLC.

**(8) Native Chemical Ligation using the Hydrazone method**

Each step of the ligation was done separately and monitored by LCMS analysis. Three solutions were prepared: (1) 1.15 g of Gn.HCl was dissolved in a phosphate buffer solution (100 mM, pH 7.0) and the pH was adjusted at 7.0 (2) A 0.05 mM solution of NaNO<sub>2</sub> was prepared in deionised water (3) 21 mg of MPAA was dissolved in solution n°1 and the pH was roughly adjusted at 7.0. The hydrazine peptide was dissolved in solution n°1 in order to have a peptide concentration around 3 mM and placed at -15°C. Then 1.2 eq of NaNO<sub>2</sub> was added using the solution n°2 and the reaction was stirred at -15°C during 10 min, monitored by LCMS. 40 eq of MPAA was added using solution n°3 and the reaction was allowed to warm to room temperature and stirred for 45 min. Lastly the second peptide was added to the reaction

mixture and the pH was adjusted at 6.5 and stirred at room temperature until the reaction was finished. The reaction was dissolved in a 0.1% TFA solution and purified on HPLC.

#### (9) Acn protecting group removal

The protected fragment *Acn*-F2-F3-*NH*<sub>2</sub> (1 mg) was dissolved in 200 μL of water. A slight excess of silver nitrate was added and the reaction was stirred at room temperature. The reaction was followed by LCMS analysis until the reaction was finished. A 0.1% TFA solution was added and the deprotected *H*-F2-F3-*NH*<sub>2</sub> fragment was purified on HPLC.

#### (10) Determination of the molar extinction coefficient of a lanthanide complex

The sample (ligand *L*<sub>*i*</sub> or graphed on TAT peptide) was dissolved in water and the buffer was added (10% v/v) in a 1 mL quartz cell. Then a solution of known concentration of Eu<sup>3+</sup> was added. After each addition the UV spectra (or the emission spectra) was recorded until we obtain a plateau. The absorbance at the desired wavelength was plotted against the concentration of Eu<sup>3+</sup>. The inflection point corresponds to the concentration of the sample and using the Beer-Lambert law we obtain the molar extinction coefficient.

$$A = \varepsilon \times l \times c$$

#### (11) Quantum yield measurement

The quantum yield was determined by a comparative method using quinine sulphate in 0.5 M H<sub>2</sub>SO<sub>4</sub> solution as standard. The *L*<sub>*i*</sub> ligand (or the supported ligand on TAT peptide) was dissolved in a 10% HEPES buffer (10 mM, pH 7.5). Four different concentrations are prepared for the standard and the sample. The absorbance of the samples and the standards are recorded between 0 and 0.1 a.u at the desired wavelength. The emission spectra was recorded using the same excitation wavelength, with the same parameters. The integrated emission intensity was plotted against the absorbance to determine the slope for the standard and the sample solution. The quantum yield was calculated using the following equation:

$$\Phi_{sample} = \Phi_{standard} \times \frac{Slope_{sample}}{Slope_{standard}}$$

## Peptide characterisation

### Fragment F1, F2 and F3

#### *Ac*-F1-*SEA*<sup>off</sup>:

LR-MS (ESI<sup>+</sup>): average *m/z* = 988.3(3+), 741.9 (4+) / calculated average *m/z* = 988.2 [M+3H]<sup>3+</sup>, 741.4 [M+4H]<sup>4+</sup> for M = C<sub>132</sub>H<sub>210</sub>N<sub>34</sub>O<sub>31</sub>S<sub>6</sub>; deconvoluted mass found = 2960.90 / expected mass = 2961.70 (average isotopic composition)

***H-F1-Nle-SEA<sup>off</sup>***

LR-MS (ESI+): average  $m/z$  = 962.2 (3+), 721.3 (4+) / calculated average  $m/z$  = 962.2 [M+3H]<sup>3+</sup>, 721.9 [M+4H]<sup>4+</sup> for M = C<sub>132</sub>H<sub>212</sub>N<sub>34</sub>O<sub>30</sub>S<sub>4</sub>; deconvoluted mass found = 2883.01 / expected mass = 2883.60 (average isotopic composition)

***H-F1-Nle-SiR-SEA<sup>off</sup>***

LR-MS (ESI+): average  $m/z$  = 1075.1 (3+), 806.7 (4+), 633.5 (5+) / calculated average  $m/z$  = 1074.7 [M+3H]<sup>3+</sup>, 806.3 [M+4H]<sup>4+</sup>, 645.2 [M+5H]<sup>5+</sup> for M = C<sub>155</sub>H<sub>233</sub>N<sub>36</sub>O<sub>31</sub>S<sub>3</sub>Si<sup>+</sup>; deconvoluted mass found = 3221.36 / expected mass = 3221.05 (average isotopic composition)

***H-F1-Nle-SiR-NHNH<sub>2</sub>***

LR-MS (ESI+): average  $m/z$  = 1039.6 (3+), 780.6 (4+), 624.1 (5+) / calculated average  $m/z$  = 1039.5 [M+3H]<sup>3+</sup>, 780.2 [M+4H]<sup>4+</sup>, 624.4 [M+5H]<sup>5+</sup> for M = C<sub>152</sub>H<sub>229</sub>N<sub>36</sub>O<sub>31</sub>Si<sup>+</sup>; deconvoluted mass found = 3116.78 / expected mass = 3116.87 (average isotopic composition)

***H-F1-Nle-Lissa-SEA<sup>off</sup>***

LR-MS (ESI+): average  $m/z$  = 1113.1 (3+), 835.4 (4+) / calculated average  $m/z$  = 1113.0 [M+3H]<sup>3+</sup>, 835.0 [M+4H]<sup>4+</sup> for M = C<sub>155</sub>H<sub>232</sub>N<sub>36</sub>O<sub>36</sub>S<sub>5</sub>; deconvoluted mass found = 3336.08 / expected mass = 3337.41 (average isotopic composition)

***Ac-F1-[L<sub>1</sub>-Eu]-SEA<sup>off</sup>***

LR-MS (ESI+): average  $m/z$  = 1249.8 (3+), 938.6 (4+) / calculated average  $m/z$  = 1250.1 [M+3H]<sup>3+</sup>, 937.8 [M+4H]<sup>4+</sup> for M = C<sub>163</sub>H<sub>241</sub>N<sub>39</sub>O<sub>41</sub>S<sub>6</sub>Eu; deconvoluted mass found = 3748.8 / expected mass = 3747.28 (average isotopic composition)

***H-F1-Nle-[L<sub>2</sub>-Tb]-SEA<sup>off</sup>***

LR-MS (ESI+): average  $m/z$  = 1208.2 (3+), 906.2 (4+) / calculated average  $m/z$  = 1207.8 [M+3H]<sup>3+</sup>, 906.8 [M+4H]<sup>4+</sup> for M = C<sub>161</sub>H<sub>246</sub>N<sub>39</sub>O<sub>38</sub>S<sub>4</sub>Tb; deconvoluted mass found = 3622.57 / expected mass = 3623.14 (average isotopic composition)

***H-F1-Nle-[L<sub>2</sub>-Eu]-SEA<sup>off</sup>***

LR-MS (ESI+): average  $m/z$  = 1206.2 (3+), 905.1 (4+) / calculated average  $m/z$  = 1206.4 [M+3H]<sup>3+</sup>, 905.0 [M+4H]<sup>4+</sup> for M = C<sub>161</sub>H<sub>246</sub>N<sub>39</sub>O<sub>38</sub>S<sub>4</sub>Eu; deconvoluted mass found = 3615.73 / expected mass = 3616.18 (average isotopic composition)

***H-F1-Nle-[L<sub>2</sub>-Tb]-NHNH<sub>2</sub>***

LR-MS (ESI+): average  $m/z$  = 1144.5 (3+), 858.2 (4+) / calculated average  $m/z$  = 1144.9 [M+3H]<sup>3+</sup>, 858.9 [M+4H]<sup>4+</sup> for M = C<sub>153</sub>H<sub>233</sub>N<sub>40</sub>O<sub>38</sub>STb; deconvoluted mass found = 3431.22 / expected mass = 3431.77 (average isotopic composition)

***H-F1-Nle-[L<sub>10</sub>-Eu]-SEA<sup>off</sup>:***

LR-MS (ESI+): average  $m/z$  = 1211.3 (3+), 907.6 (4+) / calculated average  $m/z$  = 1210.7 [M+3H]<sup>3+</sup>, 908.3 [M+4H]<sup>4+</sup> for M = C<sub>161</sub>H<sub>245</sub>N<sub>40</sub>O<sub>38</sub>S<sub>4</sub>Eu; deconvoluted mass found = 3628.50/ expected mass = 3629.18 (average isotopic composition)

-----  
***Acm-F2-SEA<sup>off</sup>-A:***

LR-MS (ESI+): average  $m/z$  = 1096.7 (3+), 822.8 (4+), 658.4 (5+) / calculated average  $m/z$  = 1096.6 [M+3H]<sup>3+</sup>, 822.7 [M+4H]<sup>4+</sup>, 658.4 [M+5H]<sup>5+</sup> for M = C<sub>139</sub>H<sub>225</sub>N<sub>49</sub>O<sub>38</sub>S<sub>3</sub>; deconvoluted mass found = 3286.81 / expected mass = 3286.13 (average isotopic composition)

***Thz-F2-SEA<sup>off</sup>-A:***

LR-MS (ESI+): average  $m/z$  = 1083.7 (3+), 813.5 (4+) 650.5 (5+) / calculated average  $m/z$  = 1084.3 [M+3H]<sup>3+</sup>, 813.4 [M+4H]<sup>4+</sup>, 651.0 [M+5H]<sup>5+</sup> for M = C<sub>139</sub>H<sub>222</sub>N<sub>50</sub>O<sub>35</sub>S<sub>3</sub>; deconvoluted mass found = 3248.97 / expected mass = 3249.80 (average isotopic composition)

***N<sub>3</sub>-F2-NHNH<sub>2</sub>-A:***

LR-MS (ESI+): average  $m/z$  = 1028.2 (3+), 770.7 (4+), 617.4 (5+) / calculated average  $m/z$  = 1028.2 [M+3H]<sup>3+</sup>, 771.4 [M+4H]<sup>4+</sup>, 617.3 [M+5H]<sup>5+</sup> for M = C<sub>130</sub>H<sub>210</sub>N<sub>50</sub>O<sub>36</sub>S; deconvoluted mass found = 3081.48 / expected mass = 3080.98 (average isotopic composition)

***Acm-F2-NHNH<sub>2</sub>-A:***

LR-MS (ESI+): average  $m/z$  = 1042.4 (3+), 784.3 (4+), 627.5 (5+) / calculated average  $m/z$  = 1043.2 [M+3H]<sup>3+</sup>, 782.6 [M+4H]<sup>4+</sup>, 626.3 [M+5H]<sup>5+</sup> for M = C<sub>133</sub>H<sub>217</sub>N<sub>49</sub>O<sub>37</sub>S; deconvoluted mass found = 3125.84 / expected mass = 3126.57 (average isotopic composition)

-----  
***Acm-F2-SEA<sup>off</sup>-B:***

LR-MS (ESI+): average  $m/z$  = 1103.4 (3+), 829.6 (4+), 662.2 (5+) / calculated average  $m/z$  = 1104.0 [M+3H]<sup>3+</sup>, 828.2 [M+4H]<sup>4+</sup>, 662.8 [M+5H]<sup>5+</sup> for M = C<sub>141</sub>H<sub>227</sub>N<sub>51</sub>O<sub>36</sub>S<sub>3</sub>; deconvoluted mass found = 3308.37 / expected mass = 3308.87 (average isotopic composition)

***Thz-F2-SEA<sup>off</sup>-B:***

LR-MS (ESI+): average  $m/z$  = 1084.3 (3+), 813.5 (4+), 651.1 (5+) / calculated average  $m/z$  = 1084.3 [M+3H]<sup>3+</sup>, 813.5 [M+4H]<sup>4+</sup>, 651.0 [M+5H]<sup>5+</sup> for M = C<sub>139</sub>H<sub>222</sub>N<sub>50</sub>O<sub>35</sub>S<sub>3</sub>; deconvoluted mass found = 3249.00 / expected mass = 3249.80 (average isotopic composition)

***Acm-F2-NHNH<sub>2</sub>-B:***

LR-MS (ESI+): average  $m/z$  = 1051.3 (3+), 788.8 (4+), 630.6 (5+) / calculated average  $m/z$  = 1050.5 [M+3H]<sup>3+</sup>, 788.2 [M+4H]<sup>4+</sup>, 630.7 [M+5H]<sup>5+</sup> for M = C<sub>135</sub>H<sub>219</sub>N<sub>51</sub>O<sub>35</sub>S; deconvoluted mass found = 3148.28 / expected mass = 3148.62 (average isotopic composition)

-----

***H-F3-NH<sub>2</sub>-A:***

LR-MS (ESI+): average  $m/z$  = 959.6 (3+), 716.7 (4+), 576.0 (5+) / calculated average  $m/z$  = 958.1 [M+3H]<sup>3+</sup>, 718.8[M+4H]<sup>4+</sup>, 575.2 [M+5H]<sup>5+</sup>, for M = C<sub>117</sub>H<sub>192</sub>N<sub>44</sub>O<sub>35</sub>S<sub>3</sub>; deconvoluted mass found = 2870.44 / expected mass = 2871.28 (average isotopic composition)

***H-F3-Nle-NH<sub>2</sub>-A:***

LR-MS (ESI+): average  $m/z$  = 952.1 (3+), 713.6 (4+), 571.3 (5+) / calculated average  $m/z$  = 952.1 [M+3H]<sup>3+</sup>, 714.3 [M+4H]<sup>4+</sup>, 571.7 [M+5H]<sup>5+</sup> for M = C<sub>118</sub>H<sub>194</sub>N<sub>44</sub>O<sub>35</sub>S<sub>2</sub>; deconvoluted mass found = 2852.28 / expected mass = 2853.24 (average isotopic composition)

***H-F3-NH<sub>2</sub>-B:***

LR-MS (ESI+): average  $m/z$  = 934.8 (3+), 703.0 (4+), 561.2 (5+) / calculated average  $m/z$  = 935.4 [M+3H]<sup>3+</sup>, 701.8 [M+4H]<sup>4+</sup>, 561.6[M+5H]<sup>5+</sup> for M = C<sub>113</sub>H<sub>188</sub>N<sub>44</sub>O<sub>34</sub>S<sub>3</sub>; deconvoluted mass found = 2802.84 / expected mass = 2803.18 (average isotopic composition)

***H-F3-Nle-NH<sub>2</sub>-B:***

LR-MS (ESI+): average  $m/z$  = 930.3 (3+), 698.6 (4+), 558.1 (5+) / calculated average  $m/z$  = 929.4 [M+3H]<sup>3+</sup>, 697.3 [M+4H]<sup>4+</sup>, 558.0 [M+5H]<sup>5+</sup> for M = C<sub>114</sub>H<sub>190</sub>N<sub>44</sub>O<sub>34</sub>S<sub>2</sub>; deconvoluted mass found = 2784.60 / expected mass = 2785.17 (average isotopic composition)

**Peptide F2-F3**

***Acm-F2-F3-NH<sub>2</sub>-A:***

LR-MS (ESI+): average  $m/z$  = 1205.8 (5+), 1005.1 (6+), 860.9 (7+), 754.1 (8+), 670.5 (9+) / calculated average  $m/z$  = 1205.6 [M+5H]<sup>5+</sup>, 1004.8 [M+6H]<sup>6+</sup>, 861.4 [M+7H]<sup>7+</sup>, 753.9 [M+8H]<sup>8+</sup>, 670.2 [M+9H]<sup>9+</sup> for M = C<sub>252</sub>H<sub>408</sub>N<sub>92</sub>O<sub>73</sub>S<sub>4</sub>; deconvoluted mass found = 6021.70 / expected mass = 6022.85 (average isotopic composition)

**H-F2-F3-NH<sub>2</sub>-A:**

LR-MS (ESI+): average  $m/z$  = 1190.7 (5+), 992.3 (6+), 850.6 (7+) 745.3 (8+) / calculated average  $m/z$  = 1191.3 [M+5H]<sup>5+</sup>, 993.0 [M+6H]<sup>6+</sup>, 851.2 [M+7H]<sup>7+</sup>, 745.0 [M+8H]<sup>8+</sup> for M = C<sub>249</sub>H<sub>403</sub>N<sub>91</sub>O<sub>72</sub>S<sub>4</sub>; deconvoluted mass found = 5950.40 / expected mass = 5951.77 (average isotopic composition)

**Acm-F2-F3-Nle-NH<sub>2</sub>-A:**

LR-MS (ESI+): average  $m/z$  = 1191.2 (5+), 992.3 (6+), 850.8 (7+) 744.4 (8+), 661.4 (9+) / calculated average  $m/z$  = 1191.3 [M+5H]<sup>5+</sup>, 993.0 [M+6H]<sup>6+</sup>, 851.2 [M+7H]<sup>7+</sup>, 744.9 [M+8H]<sup>8+</sup>, 662.3 [M+9H]<sup>9+</sup> for M = C<sub>249</sub>H<sub>403</sub>N<sub>91</sub>O<sub>72</sub>S<sub>4</sub>; deconvoluted mass found = 5947.13 / expected mass = 5946.76 (average isotopic composition)

**Acm-F2-F3-NH<sub>2</sub>-B:**

LR-MS (ESI+): average  $m/z$  = 1194.6 (5+), 996.1 (6+), 853.7 (7+), 748.4 (8+), 664.1 (9+) / calculated average  $m/z$  = 1193.3 [M+5H]<sup>5+</sup>, 997.1 [M+6H]<sup>6+</sup>, 854.8 [M+7H]<sup>7+</sup>, 748.1 [M+8H]<sup>8+</sup>, 665.0 [M+9H]<sup>9+</sup> for M = C<sub>250</sub>H<sub>406</sub>N<sub>94</sub>O<sub>70</sub>S<sub>4</sub>; deconvoluted mass found = 5976.83 / expected mass = 5976.20 (average isotopic composition)

**Acm-F2-F3-Nle-NH<sub>2</sub>-B:**

LR-MS (ESI+): average  $m/z$  = 1179.8 (5+), 984.2 (6+), 844.6 (7+), 738.4 (8+)/ calculated average  $m/z$  = 1181.3 [M+5H]<sup>5+</sup>, 984.6 [M+6H]<sup>6+</sup>, 844.1 [M+7H]<sup>7+</sup>, 738.7 [M+8H]<sup>8+</sup> for M = C<sub>249</sub>H<sub>405</sub>N<sub>93</sub>O<sub>69</sub>S<sub>3</sub>; deconvoluted mass found = 5901.10 / expected mass = 5901.74 (average isotopic composition)

**Thz-F2-F3-NH<sub>2</sub>-B:**

LR-MS (ESI+): average  $m/z$  = 1184.7 (5+), 987.4 (6+), 846.0 (7+) 740.4 (8+), 658.2 (9+) / calculated average  $m/z$  = 1184.6 [M+5H]<sup>5+</sup>, 987.3 [M+6H]<sup>6+</sup>, 846.4 [M+7H]<sup>7+</sup>, 740.7 [M+8H]<sup>8+</sup>, 658.5 [M+9H]<sup>9+</sup> for M = C<sub>248</sub>H<sub>401</sub>N<sub>93</sub>O<sub>69</sub>S<sub>4</sub>; deconvoluted mass found = 5917.76 / expected mass = 5916.88 (average isotopic composition)

**H-F2-F3-Nle-NH<sub>2</sub>-A:**

LR-MS (ESI+): average  $m/z$  = 1176.7 (5+), 980.8 (6+), 838.3 (7+), 736.4 (8+), 654.3 (9+) / calculated average  $m/z$  = 1176.3 [M+5H]<sup>5+</sup>, 980.4 [M+6H]<sup>6+</sup>, 840.5 [M+7H]<sup>7+</sup>, 735.6 [M+8H]<sup>8+</sup>, 654.0 [M+9H]<sup>9+</sup> for M = C<sub>248</sub>H<sub>402</sub>N<sub>90</sub>O<sub>71</sub>S<sub>3</sub>; deconvoluted mass found = 5875.96 / expected mass = 5876.68 (average isotopic composition)

**H-F2-F3-NH<sub>2</sub>-B:**

LR-MS (ESI+): average  $m/z$  = 1181.7 (5+), 984.1 (6+), 844.6 (7+), 739.2 (8+), 658.5(9+) / calculated average  $m/z$  = 1182.1 [M+5H]<sup>5+</sup>, 985.2 [M+6H]<sup>6+</sup>, 844.7 [M+7H]<sup>7+</sup>, 739.2 [M+8H]<sup>8+</sup>, 657.2 [M+9H]<sup>9+</sup>

for M = C<sub>247</sub>H<sub>401</sub>N<sub>93</sub>O<sub>69</sub>S<sub>4</sub>; deconvoluted mass found = 5904.96 / expected mass = 5905.75 (average isotopic composition)

**H-F2-F3-Nle-NH<sub>2</sub>-B:**

LR-MS (ESI+): average  $m/z$  = 1167.3 (5+), 972.9 (6+), 833.7 (7+), 729.6 (8+), 684.9 (9+) / calculated average  $m/z$  = 1167.1 [M+5H]<sup>5+</sup>, 972.8 [M+6H]<sup>6+</sup>, 833.9 [M+7H]<sup>7+</sup>, 729.8 [M+8H]<sup>8+</sup>, 648.9 [M+9H]<sup>9+</sup> for M = C<sub>246</sub>H<sub>400</sub>N<sub>92</sub>O<sub>68</sub>S<sub>3</sub>; deconvoluted mass found = 5830.12 / expected mass = 5830.66 (average isotopic composition)

**Peptide F1-F2-F3****Ac-F1-F2-F3-NH<sub>2</sub>-A: Zif12-A**

LR-MS (ESI+): average  $m/z$  = 1241.9 (7+), 1087.3 (8+), 966.4 (9+), 869.9 (10+), 790.8 (11+), 725.0 (12+), 669.6 (13+) / calculated average  $m/z$  = 1242.4 [M+7H]<sup>4+</sup>, 1087.3 [M+8H]<sup>8+</sup>, 966.6 [M+9H]<sup>9+</sup>, 870.0 [M+10H]<sup>10+</sup>, 791.0 [M+11H]<sup>11+</sup>, 725.2 [M+12H]<sup>9+</sup>, 669.5 [M+13H]<sup>13+</sup> for M = C<sub>373</sub>H<sub>596</sub>N<sub>124</sub>O<sub>103</sub>S<sub>7</sub>; deconvoluted mass found = 8689.26 / expected mass = 8690.06 (average isotopic composition)

**H-F1-F2-F3-Nle-NH<sub>2</sub>-A: Zif12-Nle-A**

LR-MS (ESI+): average  $m/z$  = 1230.8 (7+), 1077.6 (8+), 957.6 (9+), 862.2 (10+), 783.9 (11+), 718.5 (12+), 663.3 (13+) / for M = C<sub>373</sub>H<sub>598</sub>N<sub>124</sub>O<sub>102</sub>S<sub>5</sub>; deconvoluted mass found = 8611.06 / expected mass = 8611.96 (average isotopic composition)

**H-F1-SiR-F2-F3-Nle-NH<sub>2</sub>-A: Zif12-SiR-A**

LR-MS (ESI+): average  $m/z$  = 1121.2 (8+), 996.9 (9+), 897.8 (10+), 815.6 (11+), 747.9 (12+) / calculated average  $m/z$  = 1121.3 [M+8H]<sup>8+</sup>, 996.8 [M+9H]<sup>9+</sup>, 897.3 [M+10H]<sup>10+</sup>, 815.8 [M+11H]<sup>11+</sup>, 747.9 [M+12H]<sup>12+</sup> for M = C<sub>399</sub>H<sub>626</sub>N<sub>125</sub>O<sub>102</sub>S<sub>4</sub>Si<sup>+</sup>; deconvoluted mass found = 8962.49 / expected mass = 8960.63 (average isotopic composition)

**H-F1-Lissa-F2-F3-Nle-NH<sub>2</sub>-A: Zif12-Lissa-A**

LR-MS (ESI+): average  $m/z$  = 1136.6 (8+), 1009.4 (9+), 908.4 (10+), 826.8 (11+), 757.2 (12+) / calculated average  $m/z$  = 1135.7 [M+8H]<sup>8+</sup>, 1009.6 [M+9H]<sup>9+</sup>, 908.7 [M+10H]<sup>10+</sup>, 826.2 [M+11H]<sup>11+</sup>, 757.5 [M+12H]<sup>12+</sup> for M = C<sub>399</sub>H<sub>625</sub>N<sub>125</sub>O<sub>107</sub>S<sub>6</sub>; deconvoluted mass found = 9077.52 / expected mass = 9076.69 (average isotopic composition)

**H-F1-[L<sub>2</sub>-Tb]-F2-F3-Nle-NH<sub>2</sub>-A: Zif12-[L<sub>2</sub>-Tb]-A**

LR-MS (ESI+): average  $m/z$  = 1160.1 (8+), 1031.3 (9+), 928.9 (10+), 843.8 (11+), 773.7 (12+) / calculated average  $m/z$  = 1160.5 [M+8H]<sup>8+</sup>, 1031.7 [M+9H]<sup>9+</sup>, 928.6 [M+10H]<sup>10+</sup>, 844.3 [M+11H]<sup>11+</sup>,

774.0 [M+12H]<sup>12+</sup> for M = C<sub>401</sub>H<sub>631</sub>N<sub>128</sub>O<sub>109</sub>S<sub>4</sub>Tb; deconvoluted mass found = 9275.43 / expected mass = 9276.41 (average isotopic composition)

**H-F1-[L<sub>2</sub>-Eu]-F2-F3-Nle-NH<sub>2</sub>-A: Zif12-[L<sub>2</sub>-Eu]-A**

LR-MS (ESI+): average  $m/z$  = 1159.4 (8+), 1030.6 (9+), 927.7 (10+), 843.5 (11+), 773.2 (12+) / calculated average  $m/z$  = 1159.7 [M+8H]<sup>8+</sup>, 1030.9 [M+9H]<sup>9+</sup>, 927.9 [M+10H]<sup>10+</sup>, 843.7 [M+11H]<sup>11+</sup>, 773.5 [M+12H]<sup>12+</sup> for M = C<sub>401</sub>H<sub>631</sub>N<sub>128</sub>O<sub>109</sub>S<sub>4</sub>Eu; deconvoluted mass found = 9268.7 / expected mass = 9269.5 (average isotopic composition)

**H-F1-[L<sub>10</sub>-Eu]-F2-F3-Nle-NH<sub>2</sub>-A: Zif12-[L<sub>10</sub>-Eu]-A**

LR-MS (ESI+): average  $m/z$  = 1160.1 (8+), 1031.3 (9+), 928.9 (10+), 843.8 (11+), 773.7 (12+) / calculated average  $m/z$  = 1161.3 [M+8H]<sup>8+</sup>, 1032.4 [M+9H]<sup>9+</sup>, 929.2 [M+10H]<sup>10+</sup>, 844.9 [M+11H]<sup>11+</sup>, 774.5 [M+12H]<sup>12+</sup> for M = C<sub>401</sub>H<sub>630</sub>EuN<sub>129</sub>O<sub>109</sub>S<sub>4</sub>; deconvoluted mass found = 9281.42 / expected mass = 9282.45 (average isotopic composition)

**Ac-F1-F2-F3-NH<sub>2</sub>-B: Zif12-B**

LR-MS (ESI+): average  $m/z$  = 1234.5 (7+), 1081.5 (8+), 961.2 (9+), 865.3 (10+), 786.7 (11+), 721.2 (12+) / calculated average  $m/z$  = 1235.9 [M+7H]<sup>7+</sup>, 1081.5 [M+8H]<sup>8+</sup>, 961.4 [M+9H]<sup>9+</sup>, 865.4 [M+10H]<sup>10+</sup>, 786.8 [M+11H]<sup>11+</sup>, 721.3 [M+12H]<sup>12+</sup> for M = C<sub>371</sub>H<sub>594</sub>N<sub>126</sub>O<sub>100</sub>S<sub>7</sub>; deconvoluted mass found = 8643.25 / expected mass = 8642.06 (average isotopic composition)

**H-F1-SiR-F2-F3-NH<sub>2</sub>-B: Zif12-SiR-B**

LR-MS (ESI+): average  $m/z$  = 1115.1 (8+), 991.4 (9+), 892.4 (10+), 811.6 (11+), 743.7 (12+) / calculated average  $m/z$  = 1115.6 [M+8H]<sup>8+</sup>, 991.7 [M+9H]<sup>9+</sup>, 892.6 [M+10H]<sup>10+</sup>, 811.6 [M+11H]<sup>11+</sup>, 744.0 [M+12H]<sup>12+</sup> for M = C<sub>397</sub>H<sub>624</sub>N<sub>127</sub>O<sub>99</sub>S<sub>4</sub>Si<sup>+</sup>; deconvoluted mass found = 8914.73 / expected mass = 8916.47 (average isotopic composition)

**H-F1-Lissa-F2-F3-NH<sub>2</sub>-B: Zif12-Lissa-B**

LR-MS (ESI+): average  $m/z$  = 1130.6 (8+), 1004.6 (9+), 903.8 (10+), 822.3 (11+), 753.7 (12+), 695.4 (13+) / calculated average  $m/z$  = 1129.9 [M+8H]<sup>8+</sup>, 1004.5 [M+9H]<sup>9+</sup>, 904.2 [M+10H]<sup>10+</sup>, 822.0 [M+11H]<sup>11+</sup>, 753.6 [M+12H]<sup>12+</sup>, 695.7 [M+13H]<sup>13+</sup> for M = C<sub>397</sub>H<sub>623</sub>N<sub>127</sub>O<sub>104</sub>S<sub>6</sub>; deconvoluted mass found = 9031.50 / expected mass = 9030.49 (average isotopic composition)



## Bibliographie

1. Chen, B. *et al.* Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell* **155**, 1479–1491 (2013).
2. Stadtmauer, E. A. *et al.* CRISPR-engineered T cells in patients with refractory cancer. *Science* (2020) doi:10.1126/science.aba7365.
3. Hélène, C. & Toulmé, J.-J. Control of Gene Expression by Oligodeoxynucleotides Covalently Linked to Intercalating Agents and Nucleic Acid-cleaving Reagents. in *Oligodeoxynucleotides: Antisense Inhibitors of Gene Expression* (ed. Cohen, J. S.) 137–172 (Macmillan Education UK, 1989). doi:10.1007/978-1-349-10869-5\_8.
4. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).
5. Ghosh, A. & Bansal, M. A glossary of DNA structures from A to Z. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 620–626 (2003).
6. Yakovchuk, P., Protozanova, E. & Frank-Kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* **34**, 564–574 (2006).
7. Boga, S., Bouzada, D., García Peña, D., Vázquez López, M. & Vázquez, M. E. Sequence-Specific DNA Recognition with Designed Peptides. *Eur. J. Org. Chem.* **2018**, 249–261 (2018).
8. Tyagi, S. & Kramer, F. R. Molecular beacons: probes that fluoresce upon hybridization. *Nat. Biotechnol.* **14**, 303–308 (1996).
9. Bioanalytical applications of aptamer and molecular-beacon probes in fluorescence-affinity assays. *TrAC Trends Anal. Chem.* **28**, 878–892 (2009).
10. Tan, W., Wang, K. & Drake, T. J. Molecular beacons. *Curr. Opin. Chem. Biol.* **8**, 547–553 (2004).
11. Li, Y., Zhou, X. & Ye, D. Molecular beacons: an optimal multifunctional biological probe. *Biochem. Biophys. Res. Commun.* **373**, 457–461 (2008).
12. Tyagi, S., Marras, S. A. & Kramer, F. R. Wavelength-shifting molecular beacons. *Nat. Biotechnol.* **18**, 1191–1196 (2000).

13. Lee, D.-H. *et al.* Individual Donor Nucleic Acid Amplification Testing for Detection of West Nile Virus. *J. Clin. Microbiol.* (2005) doi:10.1128/JCM.43.10.5111-5116.2005.
14. Tan, W., Fang, X., Li, J. & Liu, X. Molecular Beacons: A Novel DNA Probe for Nucleic Acid and Protein Studies. *Chem. – Eur. J.* **6**, 1107–1111 (2000).
15. Kuhn, H. *et al.* Hybridization of DNA and PNA Molecular Beacons to Single-Stranded and Double-Stranded DNA Targets. *J. Am. Chem. Soc.* **124**, 1097–1103 (2002).
16. Li, J. J. & Tan, W. A real-time assay for DNA sticky-end pairing using molecular beacons. *Anal. Biochem.* **312**, 251–254 (2003).
17. Sokol, D. L., Zhang, X., Lu, P. & Gewirtz, A. M. Real time detection of DNA:RNA hybridization in living cells. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 11538–11543 (1998).
18. Nielsen, P. E., Egholm, M., Berg, R. H. & Buchardt, O. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* **254**, 1497–1500 (1991).
19. Frank-Kamenetskii, M. D. & Mirkin, S. M. Triplex DNA structures. *Annu. Rev. Biochem.* **64**, 65–95 (1995).
20. Kuhn, H., Demidov, V. V., Nielsen, P. E. & Frank-Kamenetskii, M. D. An experimental study of mechanism and specificity of peptide nucleic acid (PNA) binding to duplex DNA. *J. Mol. Biol.* **286**, 1337–1345 (1999).
21. Kaihatsu, K., Janowski, B. A. & Corey, D. R. Recognition of Chromosomal DNA by PNAs. *Chem. Biol.* **11**, 749–758 (2004).
22. Demidov, V. & Frank-Kamenetskii, M. PNA Openers and Their Applications. *Methods Mol. Biol. Clifton NJ* **208**, 119–30 (2002).
23. Kuhn, H. *et al.* PNA beacons for duplex DNA. *Antisense Nucleic Acid Drug Dev.* **11**, 265–270 (2001).
24. Kuhn, H., Demidov, V. V. & Frank-Kamenetskii, M. D. Rolling-circle amplification under topological constraints. *Nucleic Acids Res.* **30**, 574–580 (2002).
25. Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P. & Dickerson, R. E. The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proc. Natl. Acad. Sci.* **82**, 1376–1380 (1985).

26. Olsen, G. L., Louie, E. A., Drobny, G. P. & Sigurdsson, S. T. Determination of DNA minor groove width in distamycin-DNA complexes by solid-state NMR. *Nucleic Acids Res.* **31**, 5084–5089 (2003).
27. Wade, W. S., Mrksich, M. & Dervan, P. B. Design of peptides that bind in the minor groove of DNA at 5'-(A,T)G(A,T)C(A,T)-3' sequences by a dimeric side-by-side motif. *J. Am. Chem. Soc.* **114**, 8783–8794 (1992).
28. Sequence-specific DNA binding Pyrrole–imidazole polyamides and their applications. *Bioorg. Med. Chem.* **26**, 1393–1411 (2018).
29. Marques, M. A., Doss, R. M., Urbach, A. R. & Dervan, P. B. Toward an Understanding of the Chemical Etiology for DNA Minor-Groove Recognition by Polyamides. *Helv. Chim. Acta* **85**, 4485–4517 (2002).
30. Kelly, J. J., Baird, E. E. & Dervan, P. B. Binding site size limit of the 2:1 pyrrole-imidazole polyamide-DNA motif. *Proc. Natl. Acad. Sci.* **93**, 6981–6985 (1996).
31. Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr. Opin. Struct. Biol.* **13**, 284–299 (2003).
32. Wurtz, N. R., Turner, J. M., Baird, E. E. & Dervan, P. B. Fmoc Solid Phase Synthesis of Polyamides Containing Pyrrole and Imidazole Amino Acids. *Org. Lett.* **3**, 1201–1203 (2001).
33. Baird, E. E. & Dervan, P. B. Solid Phase Synthesis of Polyamides Containing Imidazole and Pyrrole Amino Acids. *J. Am. Chem. Soc.* **118**, 6141–6146 (1996).
34. Search for the Optimal Linker in Tandem Hairpin Polyamides. *Bioorg. Med. Chem.* **10**, 3339–3349 (2002).
35. Yamamoto, M. *et al.* Sequence-Specific DNA Recognition by Cyclic Pyrrole–Imidazole Cysteine-Derived Polyamide Dimers. *Chem. – Eur. J.* **20**, 752–759 (2014).
36. Rucker, V. C., Foister, S., Melander, C. & Dervan, P. B. Sequence Specific Fluorescence Detection of Double Strand DNA. *J. Am. Chem. Soc.* **125**, 1195–1202 (2003).
37. Fechter, E. J., Olenyuk, B. & Dervan, P. B. Sequence-Specific Fluorescence Detection of DNA by Polyamide–Thiazole Orange Conjugates. *J. Am. Chem. Soc.* **127**, 16685–16691 (2005).

38. Boch, J. *et al.* Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science* **326**, 1509–1512 (2009).
39. Ma, H., Reyes-Gutierrez, P. & Pederson, T. Visualization of repetitive DNA sequences in human chromosomes with transcription activator-like effectors. *Proc. Natl. Acad. Sci.* **110**, 21048–21053 (2013).
40. Miller, J., McLachlan, A. D. & Klug, A. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* **4**, 1609–1614 (1985).
41. Berg, J. M. Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 99–102 (1988).
42. Parraga, G. *et al.* Zinc-dependent structure of a single-finger domain of yeast ADR1. *Science* **241**, 1489–1492 (1988).
43. Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. & Wright, P. E. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science* **245**, 635–637 (1989).
44. Gersbach, C. A., Gaj, T. & Barbas, C. F. Synthetic Zinc Finger Proteins: The Advent of Targeted Gene Regulation and Genome Modification Technologies. *Acc. Chem. Res.* **47**, 2309–2318 (2014).
45. Pavletich, N. P. & Pabo, C. O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252**, 809–817 (1991).
46. Choo, Y. & Klug, A. Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 11168–11172 (1994).
47. Gupta, A. *et al.* An optimized two-finger archive for ZFN-mediated gene targeting. *Nat. Methods* **9**, 588–590 (2012).
48. Gupta, A. *et al.* An improved predictive recognition model for Cys2-His2 zinc finger proteins. *Nucleic Acids Res.* **42**, 4800–4812 (2014).
49. Shi, Y. & Berg, J. M. A direct comparison of the properties of natural and designed zinc-finger proteins. *Chem. Biol.* **2**, 83–89 (1995).
50. Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* **11**, 636–646 (2010).

51. Santiago, Y. *et al.* Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 5809–5814 (2008).
52. Li, L. *et al.* EGR1 regulates angiogenic and osteoclastogenic factors in prostate cancer and promotes metastasis. *Oncogene* **38**, 6241–6255 (2019).
53. Wang, B. S. & Pabo, C. O. Dimerization of zinc fingers mediated by peptides evolved in vitro from random sequences. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 9568–9573 (1999).
54. Wang, B. S., Grant, R. A. & Pabo, C. O. Selected peptide extension contacts hydrophobic patch on neighboring zinc finger and mediates dimerization on DNA. *Nat. Struct. Biol.* **8**, 589–593 (2001).
55. Yoshitake, K., Waki, S. & Ueda, H. Dimerization-based homogeneous fluorosensor proteins for the detection of specific dsDNA. *Biosens. Bioelectron.* **23**, 1266–1271 (2008).
56. Herschel, J. F. W.  $\rho\Phi\omega\tau\alpha$  No. I. On a Case of Superficial Colour Presented by a Homogeneous Liquid Internally Colourless. *Philos. Trans. R. Soc. Lond.* **135**, 143–145 (1845).
57. *Principles of Fluorescence Spectroscopy* / Joseph R. Lakowicz / Springer.
58. Jun, J. V., Chenoweth, D. M. & Petersson, E. J. Rational design of small molecule fluorescent probes for biological applications. *Org. Biomol. Chem.* **18**, 5747–5763 (2020).
59. Lyons, S. K., Patrick, P. S. & Brindle, K. M. Imaging Mouse Cancer Models In Vivo Using Reporter Transgenes. *Cold Spring Harb. Protoc.* **2013**, pdb.top069864 (2013).
60. Wu, L. *et al.* Förster resonance energy transfer (FRET)-based small-molecule sensors and imaging agents. *Chem. Soc. Rev.* **49**, 5110–5139 (2020).
61. Rowland, C. E., Brown, C. W., Medintz, I. L. & Delehanty, J. B. Intracellular FRET-based probes: a review. *Methods Appl. Fluoresc.* **3**, 042006 (2015).
62. Zhu, J., Maza, W. A. & Morris, A. J. Light-harvesting and energy transfer in ruthenium(II)-polypyridyl doped zirconium(IV) metal-organic frameworks: A look toward solar cell applications. *J. Photochem. Photobiol. Chem.* **344**, (2017).
63. Demchenko, A. P. The concept of  $\lambda$ -ratiometry in fluorescence sensing and imaging. *J. Fluoresc.* **20**, 1099–1128 (2010).

64. Sei-Iida, Y., Koshimoto, H., Kondo, S. & Tsuji, A. Real-time monitoring of in vitro transcriptional RNA synthesis using fluorescence resonance energy transfer. *Nucleic Acids Res.* **28**, e59–e59 (2000).
65. Werts, M. H. V. Making sense of Lanthanide Luminescence. *Sci. Prog.* **88**, 101–131 (2005).
66. Bünzli, J.-C. G. & Piguet, C. Taking advantage of luminescent lanthanide ions. *Chem. Soc. Rev.* **34**, 1048–1077 (2005).
67. Pearson, R. G. Hard and Soft Acids and Bases. in *Survey of Progress in Chemistry* (ed. Scott, A. F.) vol. 5 1–52 (Elsevier, 1969).
68. Bard, A. J., Parsons, R. & Jordan, J. *Standard Potentials in Aqueous Solution*. (Routledge, 2017). doi:10.1201/9780203738764.
69. Idée, J. M. *et al.* Haemodynamic effects of macrocyclic and linear gadolinium chelates in rats: role of calcium and transmetallation. *Biometals* **11**, 113–123 (1998).
70. Burai, L., Fábrián, I., Király, R., Szilágyi, E. & Brücher, E. Equilibrium and kinetic studies on the formation of the lanthanide(III) complexes, [Ce(dota)]<sup>-</sup> and [Yb(dota)]<sup>-</sup> (H<sub>4</sub>dota = 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid). *J. Chem. Soc. Dalton Trans.* 243–248 (1998) doi:10.1039/A705158A.
71. Martell, A. E., Hancock, R. D. & Motekaitis, R. J. Factors affecting stabilities of chelate, macrocyclic and macrobicyclic complexes in solution. *Coord. Chem. Rev.* **133**, 39–65 (1994).
72. Caravan, P., Ellison, J. J., McMurry, T. J. & Lauffer, R. B. Gadolinium(III) Chelates as MRI Contrast Agents: Structure, Dynamics, and Applications. *Chem. Rev.* **99**, 2293–2352 (1999).
73. Wedeking, P., Kumar, K. & Tweedle, M. F. Dissociation of gadolinium chelates in mice: Relationship to chemical characteristics. *Magn. Reson. Imaging* **10**, 641–648 (1992).
74. D'Aléo, A. *et al.* Ytterbium-Based Bioprobes for Near-Infrared Two-Photon Scanning Laser Microscopy Imaging. *Angew. Chem. Int. Ed.* **51**, 6622–6625 (2012).
75. Weissman, S. I. Intramolecular Energy Transfer The Fluorescence of Complexes of Europium. *J. Chem. Phys.* **10**, 214–217 (1942).
76. Thibon, A. & Pierre, V. C. Principles of responsive lanthanide-based luminescent probes for cellular imaging. *Anal. Bioanal. Chem.* **394**, 107–120 (2009).

77. Mathis, G. & Bazin, H. Stable Luminescent Chelates and Macrocyclic Compounds. in *Lanthanide Luminescence: Photophysical, Analytical and Biological Aspects* (eds. Hänninen, P. & Härmä, H.) 47–88 (Springer Berlin Heidelberg, 2011). doi:10.1007/4243\_2010\_5.
78. Xu, J. *et al.* Octadentate Cages of Tb(III) 2-Hydroxyisophthalamides: A New Standard for Luminescent Lanthanide Labels. *J. Am. Chem. Soc.* **133**, 19900–19910 (2011).
79. D'Aléo, A., Pointillart, F., Ouahab, L., Andraud, C. & Maury, O. Charge transfer excited states sensitization of lanthanide emitting from the visible to the near-infra-red. *Coord. Chem. Rev.* **256**, 1604–1620 (2012).
80. Latva, M. *et al.* Correlation between the lowest triplet state energy level of the ligand and lanthanide(III) luminescence quantum yield. *J. Lumin.* **75**, 149–169 (1997).
81. D'Aléo, A. *et al.* Sensitization of Eu(III) luminescence by donor-phenylethynyl-functionalized DTPA and DO3A macrocycles. *Comptes Rendus Chim.* **13**, 681–690 (2010).
82. Soulié, M. *et al.* Comparative Analysis of Conjugated Alkynyl Chromophore–Triazacyclononane Ligands for Sensitized Emission of Europium and Terbium. *Chem. – Eur. J.* **20**, 8636–8646 (2014).
83. Placide, V. *et al.* Design and synthesis of europium luminescent bio-probes featuring sulfobetaine moieties. *Tetrahedron Lett.* **55**, 1357–1361 (2014).
84. Bui, A. T. *et al.* Twisted Charge-Transfer Antennae for Ultra-Bright Terbium(III) and Dysprosium(III) Bioprobes. *Chem. – Eur. J.* **24**, 3408–3412 (2018).
85. Zwier, J. M., Bazin, H., Lamarque, L. & Mathis, G. Luminescent Lanthanide Cryptates: from the Bench to the Bedside. *Inorg. Chem.* **53**, 1854–1866 (2014).
86. Sy, M., Nonat, A., Hildebrandt, N. & Charbonnière, L. J. Lanthanide-based luminescence biolabelling. *Chem. Commun.* **52**, 5080–5095 (2016).
87. Starck, M. *et al.* Towards Libraries of Luminescent Lanthanide Complexes and Labels from Generic Synthons. *Chem. – Eur. J.* **17**, 9164–9179 (2011).
88. Delbianco, M. *et al.* Bright, Highly Water-Soluble Triazacyclononane Europium Complexes To Detect Ligand Binding with Time-Resolved FRET Microscopy. *Angew. Chem. Int. Ed.* **53**, 10718–10722 (2014).

89. Dai, Z., Tian, L., Song, B., Liu, X. & Yuan, J. Development of a novel lysosome-targetable time-gated luminescence probe for ratiometric and luminescence lifetime detection of nitric oxide in vivo. *Chem. Sci.* **8**, 1969–1976 (2017).
90. Penas, C., Pazos, E., Mascareñas, J. L. & Vázquez, M. E. A Folding-Based Approach for the Luminescent Detection of a Short RNA Hairpin. *J. Am. Chem. Soc.* **135**, 3812–3814 (2013).
91. Penas, C., Mascareñas, J. L. & Vázquez, M. E. Coupling the folding of a  $\beta$ -hairpin with chelation-enhanced luminescence of Tb(III) and Eu(III) ions for specific sensing of a viral RNA. *Chem. Sci.* **7**, 2674–2678 (2016).
92. Raibaut, L. *et al.* Design of a synthetic luminescent probe from a biomolecule binding domain: selective detection of AU-rich mRNA sequences. *Chem. Sci.* **8**, 1658–1664 (2017).
93. Zhao, C., Sun, Y., Ren, J. & Qu, X. Recent progress in lanthanide complexes for DNA sensing and targeting specific DNA structures. *Inorganica Chim. Acta* **452**, 50–61 (2016).
94. Oser, A. & Valet, G. Nonradioactive Assay of DNA Hybridization by DNA-Template-Mediated Formation of a Ternary TbIII Complex in Pure Liquid Phase. *Angew. Chem. Int. Ed. Engl.* **29**, 1167–1169 (1990).
95. Karhunen, U., Jaakkola, L., Wang, Q., Lamminmäki, U. & Soukka, T. Luminescence Switching by Hybridization-Directed Mixed Lanthanide Complex Formation. *Anal. Chem.* **82**, 751–754 (2010).
96. Kitamura, Y. *et al.* Colorimetric allele typing through cooperative binding of DNA probes carrying a metal chelator for luminescent lanthanide ions. *Anal. Biochem.* **359**, 259–261 (2006).
97. Lehmusvuori, A., Karhunen, U., Tapio, A.-H., Lamminmäki, U. & Soukka, T. High-performance closed-tube PCR based on switchable luminescence probes. *Anal. Chim. Acta* **731**, 88–92 (2012).
98. Saneyoshi, H., Ito, Y. & Abe, H. Long-Lived Luminogenic Probe for Detection of RNA in a Crude Solution of Living Bacterial Cells. *J. Am. Chem. Soc.* **135**, 13632–13635 (2013).
99. Agouridas, V. *et al.* Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations. *Chem. Rev.* **119**, (2019).
100. Persikov, A. V. *et al.* A systematic survey of the Cys2His2 zinc finger DNA-binding landscape. *Nucleic Acids Res.* **43**, 1965–1984 (2015).

101. Persikov, A. V. & Singh, M. De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins. *Nucleic Acids Res.* **42**, 97–108 (2014).
102. Persikov, A. V., Rowland, E. F., Oakes, B. L., Singh, M. & Noyes, M. B. Deep sequencing of large library selections allows computational discovery of diverse sets of zinc fingers that bind common targets. *Nucleic Acids Res.* **42**, 1497–1508 (2014).
103. Saxon, E., Armstrong, J. I. & Bertozzi, C. R. A “Traceless” Staudinger Ligation for the Chemoselective Synthesis of Amide Bonds. *Org. Lett.* **2**, 2141–2143 (2000).
104. Bode, J. W., Fox, R. M. & Baucom, K. D. Chemoselective Amide Ligations by Decarboxylative Condensations of N-Alkylhydroxylamines and  $\alpha$ -Ketoacids. *Angew. Chem. Int. Ed.* **45**, 1248–1252 (2006).
105. Zhang, Y., Xu, C., Lam, H. Y., Lee, C. L. & Li, X. Protein chemical synthesis by serine and threonine ligation. *Proc. Natl. Acad. Sci.* **110**, 6657–6662 (2013).
106. Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. H. Synthesis of Proteins by Native Chemical Ligation. *Science* **266**, 776–779 (1994).
107. Conibear, A. C., Watson, E. E., Payne, R. J. & Becker, C. F. W. Native chemical ligation in protein synthesis and semi-synthesis. *Chem. Soc. Rev.* **47**, 9046–9068 (2018).
108. Shelton, P. T. & Jensen, K. J. Synthesis of C-Terminal Peptide Thioesters Using Fmoc-Based Solid-Phase Peptide Chemistry. in *Peptide Synthesis and Applications* (eds. Jensen, K. J., Tofteng Shelton, P. & Pedersen, S. L.) 119–129 (Humana Press, 2013). doi:10.1007/978-1-62703-544-6\_8.
109. Ingenito, R., Bianchi, E., Fattori, D. & Pessi, A. Solid Phase Synthesis of Peptide C-Terminal Thioesters by Fmoc/t-Bu Chemistry. *J. Am. Chem. Soc.* **121**, 11369–11374 (1999).
110. Tofteng, A. P., Sørensen, K. K., Conde-Frieboes, K. W., Hoeg-Jensen, T. & Jensen, K. J. Fmoc Solid-Phase Synthesis of C-Terminal Peptide Thioesters by Formation of a Backbone Pyroglutamyl Imide Moiety. *Angew. Chem. Int. Ed.* **48**, 7411–7414 (2009).
111. Katritzky, A. R., Shestopalov, A. A. & Suzuki, K. A New Convenient Preparation of Thiol Esters Utilizing N-Acylbenzotriazoles. *Synthesis* **2004**, 1806–1813 (2004).

112. Flood, D. T. *et al.* Leveraging the Knorr Pyrazole Synthesis for the Facile Generation of Thioester Surrogates for use in Native Chemical Ligation. *Angew. Chem. Int. Ed Engl.* **57**, 11634–11639 (2018).
113. Terrier, V. P., Adihou, H., Arnould, M., Delmas, A. F. & Aucagne, V. A straightforward method for automated Fmoc-based synthesis of bio-inspired peptide crypto-thioesters. *Chem. Sci.* **7**, 339–345 (2015).
114. Ollivier, N., Dheur, J., Mhidia, R., Blanpain, A. & Melnyk, O. Bis(2-sulfanylethyl)amino Native Peptide Ligation. *Org. Lett.* **12**, 5238–5241 (2010).
115. Ollivier, N. *et al.* Tidbits for the synthesis of bis(2-sulfanylethyl)amido (SEA) polystyrene resin, SEA peptides and peptide thioesters. *J. Pept. Sci.* **20**, 92–97 (2014).
116. Johnson, E. C. B. & Kent, S. B. H. Insights into the Mechanism and Catalysis of the Native Chemical Ligation Reaction. *J. Am. Chem. Soc.* **128**, 6640–6646 (2006).
117. Raibaut, L. *et al.* Accelerating chemoselective peptide bond formation using bis(2-selenylethyl)amido peptide selenoester surrogates. *Chem. Sci.* **7**, 2657–2665 (2016).
118. Fang, G.-M. *et al.* Protein Chemical Synthesis by Ligation of Peptide Hydrazides. *Angew. Chem. Int. Ed.* **50**, 7645–7649 (2011).
119. Bang, D., Chopra, N. & Kent, S. B. H. Total chemical synthesis of crambin. *J. Am. Chem. Soc.* **126**, 1377–1383 (2004).
120. Liu, S., Pentelute, B. L. & Kent, S. B. H. Convergent Chemical Synthesis of [Lysine<sup>24, 38, 83</sup>] Human Erythropoietin. *Angew. Chem.* **124**, 1017–1023 (2012).
121. Bang, D. & Kent, S. B. H. A One-Pot Total Synthesis of Crambin. *Angew. Chem. Int. Ed.* **43**, 2534–2538 (2004).
122. Laps, S., Sun, H., Kamnesky, G. & Brik, A. Palladium-Mediated Direct Disulfide Bond Formation in Proteins Containing S-Acetamidomethyl-cysteine under Aqueous Conditions. *Angew. Chem. Int. Ed Engl.* **58**, 5729–5733 (2019).
123. Maity, S. K., Jbara, M., Laps, S. & Brik, A. Efficient Palladium-Assisted One-Pot Deprotection of (Acetamidomethyl)Cysteine Following Native Chemical Ligation and/or Desulfurization To Expedite Chemical Protein Synthesis. *Angew. Chem. Int. Ed.* **55**, 8108–8112 (2016).

124. Hossain, M. A. & Wade, J. D. Novel Methods for the Chemical Synthesis of Insulin Superfamily Peptides and of Analogues Containing Disulfide Isosteres. *Acc. Chem. Res.* **50**, 2116–2127 (2017).
125. Wu, F., Mayer, J. P., Gelfanov, V. M., Liu, F. & DiMarchi, R. D. Synthesis of Four-Disulfide Insulin Analogs via Sequential Disulfide Bond Formation. *J. Org. Chem.* **82**, 3506–3512 (2017).
126. Dekan, Z. *et al.* Total Synthesis of Human Hecpidin through Regioselective Disulfide-Bond Formation by using the Safety-Catch Cysteine Protecting Group 4,4'-Dimethylsulfinylbenzhydryl. *Angew. Chem.* **126**, 2975–2978 (2014).
127. Pan, M. *et al.* One-pot hydrazide-based native chemical ligation for efficient chemical synthesis and structure determination of toxin Mambalgin-1. *Chem. Commun.* **50**, 5837–5839 (2014).
128. Zheng, J.-S., Tang, S., Qi, Y.-K., Wang, Z.-P. & Liu, L. Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nat. Protoc.* **8**, 2483–2495 (2013).
129. Ollivier, N. *et al.* A One-Pot Three-Segment Ligation Strategy for Protein Chemical Synthesis. *Angew. Chem. Int. Ed.* **51**, 209–213 (2012).
130. Echalié, C. *et al.* Heating and microwave assisted SPPS of C-terminal acid peptides on trityl resin: the truth behind the yield. *Amino Acids* **45**, 1395–1403 (2013).
131. Titz, A., Radic, Z., Schwardt, O. & Ernst, B. A safe and convenient method for the preparation of triflyl azide, and its use in diazo transfer reactions to primary amines. *Tetrahedron Lett.* **47**, 2383–2385 (2006).
132. Lechner, C. C., Agashe, N. D. & Fierz, B. Traceless Synthesis of Asymmetrically Modified Bivalent Nucleosomes. *Angew. Chem. Int. Ed.* **55**, 2903–2906 (2016).
133. Biophysics Facility - Biozentrum. <https://www.biozentrum.unibas.ch/research/research-groups/research-groups-a-z/overview/unit/biophysics-facility>.
134. Hellman, L. M. & Fried, M. G. Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nat. Protoc.* **2**, 1849–1861 (2007).
135. Uddin, Md. J. & Marnett, L. J. Synthesis of 5- and 6-Carboxy-X-rhodamines. *Org. Lett.* **10**, 4799–4801 (2008).
136. Johnson, L. V., Walsh, M. L. & Chen, L. B. Localization of mitochondria in living cells with rhodamine 123. *Proc. Natl. Acad. Sci.* **77**, 990–994 (1980).

137. Belmont, L. D., Hyman, A. A., Sawin, K. E. & Mitchison, T. J. Real-time visualization of cell cycle-dependent changes in microtubule dynamics in cytoplasmic extracts. *Cell* **62**, 579–589 (1990).
138. Umezawa, K., Yoshida, M., Kamiya, M., Yamasoba, T. & Urano, Y. Rational design of reversible fluorescent probes for live-cell imaging and quantification of fast glutathione dynamics. *Nat. Chem.* **9**, 279–286 (2017).
139. Lukinavičius, G. *et al.* A near-infrared fluorophore for live-cell super-resolution microscopy of cellular proteins. *Nat. Chem.* **5**, 132–139 (2013).
140. Placide, V. *et al.* Two-photon multiplexing bio-imaging using a combination of Eu- and Tb-bioprobes. *Dalton Trans.* **44**, 4918–4924 (2015).
141. Barnhill, H. N., Claudel-Gillet, S., Ziesel, R., Charbonnière, L. J. & Wang, Q. Prototype Protein Assembly as Scaffold for Time-Resolved Fluoroimmuno Assays. *J. Am. Chem. Soc.* **129**, 7799–7806 (2007).
142. Fremy, G. *et al.* A novel DOTA-like building block with a picolinate arm for the synthesis of lanthanide complex-peptide conjugates with improved luminescence properties. *J. Inorg. Biochem.* **213**, 111257 (2020).
143. Jagadish, B. *et al.* Improved synthesis of 10-(2-alkylamino-2-oxoethyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triacetic acid derivatives bearing acid-sensitive linkers. *Synth. Commun.* **44**, 441–449 (2014).
144. Sonogashira, K., Tohda, Y. & Hagihara, N. A convenient synthesis of acetylenes: catalytic substitutions of acetylenic hydrogen with bromoalkenes, iodoarenes and bromopyridines. *Tetrahedron Lett.* **16**, 4467–4470 (1975).
145. Das, S. K. *et al.* Biogenic palladium nanostructures for Suzuki-Miyaura and Sonogashira cross-coupling reaction under mild reaction conditions. *Curr. Res. Green Sustain. Chem.* **5**, 100301 (2022).
146. Blakemore, D. Chapter 1:Suzuki–Miyaura Coupling. in *Synthetic Methods in Drug Discovery* 1–69 (2016). doi:10.1039/9781782622086-00001.

147. Espino, G., Kurbangalieva, A. & Brown, J. M. Aryl bromide/triflate selectivities reveal mechanistic divergence in palladium-catalysed couplings; the Suzuki–Miyaura anomaly. *Chem. Commun.* 1742–1744 (2007) doi:10.1039/B701517H.
148. van Leeuwen, P. W. N. M., Kamer, P. C. J., Reek, J. N. H. & Dierkes, P. Ligand Bite Angle Effects in Metal-catalyzed C–C Bond Formation. *Chem. Rev.* **100**, 2741–2770 (2000).
149. Ishiyama, T., Murata, M. & Miyaura, N. Palladium(0)-Catalyzed Cross-Coupling Reaction of Alkoxydiboron with Haloarenes: A Direct Procedure for Arylboronic Esters. *J. Org. Chem.* **60**, 7508–7510 (1995).
150. Frémy, G. Synthèse de sondes luminescentes pour la détection séquence spécifique d'ADN double brin. (Université Grenoble Alpes, 2019).
151. Vivès, E., Brodin, P. & Lebleu, B. A Truncated HIV-1 Tat Protein Basic Domain Rapidly Translocates through the Plasma Membrane and Accumulates in the Cell Nucleus\*. *J. Biol. Chem.* **272**, 16010–16017 (1997).
152. Regueiro-Figueroa, M. *et al.* Lanthanide dota-like Complexes Containing a Picolinate Pendant: Structural Entry for the Design of LnIII-Based Luminescent Probes. *Inorg. Chem.* **50**, 4125–4141 (2011).
153. Brouwer, A. M. Standards for photoluminescence quantum yield measurements in solution (IUPAC Technical Report). *Pure Appl. Chem.* **83**, 2213–2228 (2011).
154. Resch-Genger, U. & Rurack, K. Determination of the photoluminescence quantum yield of dilute dye solutions (IUPAC Technical Report). *Pure Appl. Chem.* **85**, 2005–2013 (2013).
155. Monteiro, J. H. S. K., Sigoli, F. A. & Bettencourt-Dias, A. de. A water-soluble TbIII complex as a temperature-sensitive luminescent probe. *Can. J. Chem.* (2017) doi:10.1139/cjc-2017-0436.
156. Monteiro, J. H. S. K., Fetto, N. R., Tucker, M. J. & de Bettencourt-Dias, A. Luminescent Carbazole-Based EuIII and YbIII Complexes with a High Two-Photon Absorption Cross-Section Enable Viscosity Sensing in the Visible and Near IR with One- and Two-Photon Excitation. *Inorg. Chem.* **59**, 3193–3199 (2020).
157. Huang, J.-J., Leung, M.-K., Chiang, Y.-L. & Lee, J.-H. Organic electroluminescent material, organic electroluminescent device and quantum dot electroluminescent unit. (2018).

158. D'Aléo, A., Picot, A., Baldeck, P. L., Andraud, C. & Maury, O. Design of Dipicolinic Acid Ligands for the Two-Photon Sensitized Luminescence of Europium Complexes with Optimized Cross-Sections. *Inorg. Chem.* **47**, 10269–10279 (2008).
159. D'Aléo, A. *et al.* Efficient Sensitization of Europium, Ytterbium, and Neodymium Functionalized Tris-Dipicolinate Lanthanide Complexes through Tunable Charge-Transfer Excited States. *Inorg. Chem.* **47**, 10258–10268 (2008).
160. Wang, C. *et al.* Twisted intramolecular charge transfer (TICT) and twists beyond TICT: from mechanisms to rational designs of bright and sensitive fluorophores. *Chem. Soc. Rev.* **50**, 12656–12678 (2021).
161. Lee, M. H. *et al.* Mitochondria-Immobilized pH-Sensitive Off–On Fluorescent Probe. *J. Am. Chem. Soc.* **136**, 14136–14142 (2014).
162. Sasaki, S., Drummen, G. P. C. & Konishi, G. Recent advances in twisted intramolecular charge transfer (TICT) fluorescence and related phenomena in materials chemistry. *J. Mater. Chem. C* **4**, 2731–2743 (2016).
163. Fabre, B. *et al.* An Integrated Computational and Experimental Approach to Gaining Selectivity for MMP-2 within the Gelatinase Subfamily. *ChemBioChem* **15**, 399–412 (2014).
164. Chankeshwara, S. V. & Chakraborti, A. K. Catalyst-Free Chemoselective N-tert-Butyloxycarbonylation of Amines in Water. *Org. Lett.* **8**, 3259–3262 (2006).
165. Lee, G. T., Jiang, X., Prasad, K., Repič, O. & Blacklock, T. J. Pd-Catalyzed ortho-Selective Oxidative Coupling of Halogenated Acetanilides with Acrylates. *Adv. Synth. Catal.* **347**, 1921–1924 (2005).
166. Bui, A. T. *et al.* Terbium(III) Luminescent Complexes as Millisecond-Scale Viscosity Probes for Lifetime Imaging. *J. Am. Chem. Soc.* **139**, 7693–7696 (2017).
167. Lukinavičius, G. *et al.* A near-infrared fluorophore for live-cell super-resolution microscopy of cellular proteins. *Nat. Chem.* **5**, 132–139 (2013).
168. Smith, S. N. & Steer, R. P. The photophysics of Lissamine rhodamine-B sulphonyl chloride in aqueous solution: implications for fluorescent protein–dye conjugates. *J. Photochem. Photobiol. Chem.* **139**, 151–156 (2001).

169. Umezawa, K., Yoshida, M., Kamiya, M., Yamasoba, T. & Urano, Y. Rational design of reversible fluorescent probes for live-cell imaging and quantification of fast glutathione dynamics. *Nat. Chem.* **9**, 279–286 (2017).
170. Rombouts, I. *et al.* Formation and reshuffling of disulfide bonds in bovine serum albumin demonstrated using tandem mass spectrometry with collision-induced and electron-transfer dissociation. *Sci. Rep.* **5**, 12210 (2015).
171. Breugst, M. & Reissig, H.-U. The Huisgen Reaction: Milestones of the 1,3-Dipolar Cycloaddition. *Angew. Chem. Int. Ed.* **59**, 12293–12307 (2020).
172. Ravasco, J. M. J. M., Faustino, H., Trindade, A. & Gois, P. M. P. Bioconjugation with Maleimides: A Useful Tool for Chemical Biology. *Chem. – Eur. J.* **25**, 43–59 (2019).
173. Blackman, M. L., Royzen, M. & Fox, J. M. Tetrazine Ligation: Fast Bioconjugation Based on Inverse-Electron-Demand Diels–Alder Reactivity. *J. Am. Chem. Soc.* **130**, 13518–13519 (2008).