



HAL
open science

Conception, construction et applications d'une bibliothèque de protéines artificielles ayant des propriétés allostériques

Adrien Fleck

► **To cite this version:**

Adrien Fleck. Conception, construction et applications d'une bibliothèque de protéines artificielles ayant des propriétés allostériques. Biochimie [q-bio.BM]. Université Paris-Saclay, 2021. Français. NNT : 2021UPASQ055 . tel-03627599

HAL Id: tel-03627599

<https://theses.hal.science/tel-03627599>

Submitted on 1 Apr 2022

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.

Conception, construction, et applications d'une bibliothèque de protéines artificielles ayant des propriétés allostériques

*Design, construction, and applications of artificial proteins library with
allosteric properties*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°569

Innovation thérapeutique : Du fondamental à l'appliqué (ITFA)

Spécialité de doctorat : Biochimie et Biologie Structurale

Graduate School : Santé et médicaments. Référent : Faculté de pharmacie

Thèse préparée dans l'unité de recherche **I2BC (Université Paris-Saclay, CEA, CNRS)**, sous la direction de **Philippe MINARD**, professeur des universités, le co-encadrement de **Marie VALERIO-LEPINIEC**, maîtresse de conférences, et **Agathe URVOAS**, professeure des universités

Thèse soutenue à Paris-Saclay, le 16 Décembre 2021, par

Adrien FLECK

Composition du Jury

Marc GRAILLE Directeur de recherche, Ecole Polytechnique (UMR7654 du CNRS)	Président
Magali REMAUD-SIMEON Professeure, Université de Toulouse (CNRS 5504 - INRA 792 - INSA)	Rapporteuse & Examinatrice
Pierre MARTINEAU Directeur de recherche, Université de Montpellier (Inserm U1194)	Rapporteur & Examinateur
Yannick RONDELEZ Directeur de recherche, ESPCI (UMR 7083)	Examinateur
Philippe MINARD Professeur, I2BC (UMR 9198)	Directeur de thèse

Titre : Conception, construction, et applications d'une bibliothèque de protéines artificielles ayant des propriétés allostériques

Mots clés : Biotechnologies, Techniques d'immobilisation à la surface cellulaire, Motifs et domaines d'interaction protéique

Résumé : Cette thèse porte sur la génération de protéines artificielles aux propriétés d'interactions modulables par l'ajout d'un modulateur protéique. L'objectif est de pouvoir capturer une cible protéique ou cellulaire au moyen d'une protéine sélectionnée au sein d'une bibliothèque à deux domaines, puis de la relâcher en présence d'une tierce protéine appelée modulateur allostérique. Le principe de ce projet repose donc sur la flexibilité de l'architecture des protéines de la bibliothèque qui doivent pouvoir passer d'une conformation compatible avec la fixation de leur cible à une conformation ne permettant pas leur interaction avec cette même cible en présence du modulateur. Ce modulateur est générique, c'est à dire qu'il doit pouvoir induire un changement de conformation incompatible avec la fixation de la cible pour tous les interacteurs de la bibliothèque.

Dans un premier temps, une bibliothèque de protéines artificielles à deux domaines a été conçue et construite à partir d'une nouvelle ossature protéique thermostable. Cette bibliothèque comporte 1.4 milliards de clones indépendants. Dans un second temps, un modulateur allostérique capable de reconnaître les deux domaines des protéines de la bibliothèque a été construit et sa structure a été résolue par cristallographie. Des sélections par phage display ont ensuite été réalisées sur plusieurs cibles protéiques. Des interacteurs ont été obtenus sur deux cibles : l'eGFP choisie comme protéine modèle et TROP 2, une protéine membranaire marqueur de cellules épithéliales et surexprimée dans les tumeurs. L'étude des complexes ainsi générés et des effets du modulateur a été initiée.

Title : Design, construction, and applications of artificial proteins library with allosteric properties

Keywords : Biotechnologies, Cell display techniques, protein interaction motifs and domains

Abstract : This thesis focuses on the generation of artificial proteins with interaction properties that can be modulated by the addition of a protein modulator. The objective is to be able to capture a protein or cell target by means of a protein selected from a two-domain library, and then release it in the presence of a third protein called an allosteric modulator. The principle of this project is therefore based on the flexibility of the architecture of the proteins in the library, which must be able to change from a conformation compatible with the binding of their target to a conformation that does not allow them to interact with this same target in the presence of the modulator. This modulator is generic, i.e. it must be able to induce a conformational change incompatible with target binding for all the interactors in the

library. As a first step, a library of artificial two-domain proteins was designed and constructed from a novel thermostable protein backbone. This library contains 1.4 billion independent clones. In a second step, an allosteric modulator capable of recognising both domains of the library proteins was constructed and its structure was resolved by crystallography. Selections by phage display were then made on several protein targets. Interactors were obtained on two targets: eGFP chosen as a model protein and TROP 2, a membrane protein marker of epithelial cells and overexpressed in tumours. The study of the complexes thus generated and the effects of the modulator has been initiated.

Remerciements	1
Liste des abréviations	4
Introduction.....	6
I. La reconnaissance moléculaire.....	8
II. Ingénierie modulaire	27
III. La purification cellulaire	36
IV. Stratégie générale	46
Bibliographie.....	50
Résultats et discussion	63
Partie 1 : Conception et construction de la bibliothèque de biP5H.....	65
I. Identification du <i>scaffold</i> de départ	66
II. Production des P5H et caractérisation de leurs propriétés de stabilité	74
II. Ingénierie des bi-domaines P5H.....	78
IV. Stratégie d'assemblage de la bibliothèque	89
Partie 2 : Conception du modulateur allostérique.....	104
I. Sélections par phage display d' α Reps reconnaissant les P5H.....	105
II. Caractérisation des α Reps reconnaissant les P5H.....	115
III. Conception du modulateur	118
IV. Caractérisation des modulateurs	125
Partie 3 : Applications de la bibliothèque de bi-P5H.....	136
I. Evaluation de la bibliothèque de bi-P5H sur l'eGFP	138
II. Evaluation de la bibliothèque de bi-P5H sur des protéines membranaires.....	145
V. Bilan et perspectives	158
Bibliographie.....	165
Conclusion	167
Bibliographie.....	172
Matériels et méthodes	173

Remerciements

Ce manuscrit vient clore 3 ans et demi de travaux, et écrire cette page est un grand soulagement.

Comme j'ai beaucoup de personnes à remercier, je vais essayer de faire ça par sections, n'y voyez aucune hiérarchisation !

Je tiens tout d'abord à remercier l'équipe MIP, qui m'a permis de réaliser cette thèse. J'ai entendu beaucoup d'histoires de la part d'autres thésards à propos d'encadrants non présents et de litiges, et je pense honnêtement que je n'aurais pas pu trouver un meilleur labo pour faire ma thèse !

Je remercie les anciens de MIP : Sébastien et Augustin, avec qui j'ai passé de très bons moments et qui m'ont beaucoup appris au début de ma thèse. Ils sont d'ailleurs devenus des amis avec le temps. Je remercie également Pierrette pour toutes les fois où un petit cadeau était posé sur mon bureau, je me sers toujours de mon verre personnalisé ! Je tiens également à remercier Jean, dont l'expertise m'a aidé à de nombreuses reprises et je n'ai pas oublié que je te dois une bouteille de champagne ! Je remercie aussi Agnès la sportive (parfois un peu trop) qui m'a toujours accompagné dans l'expérimentation puisque pendant un moment nous n'étions que deux à travailler à la paillasse ! Je tiens également à remercier les Magali avec qui j'ai passé de très bons moments et avec qui on a bien rigolé, même si ça pouvait faire peur aux gens qui venaient sur la plateforme. Merci pour les formations sur SPR et BLI et pour vos jolis rapports et figures ! Je remercie Agathe et Marielle, le duo de choc de l'équipe, qui en plus d'être d'une bienveillance extrême, ont eu à corriger un grand nombre de présentations/rapports qui n'étaient pas toujours rédigés avec le plus grand soin. Vous êtes les principales raisons pour lesquelles ce travail de thèse est arrivé au bout, et des encadrantes modèle. Je remercie également Philippe (que j'arriverai peut-être à tutoyer un jour) qui a toujours été présent lorsque j'avais des questions et qui s'est beaucoup impliqué à la fois dans mon projet et dans la correction du manuscrit. Je pense que peu de directeurs de thèse sont aussi présents pour leurs étudiants.

La conclusion des remerciements à l'équipe MIP m'offre une transition de choix pour élargir le champ des remerciements au bâtiment 430. J'ai en effet eu la possibilité d'interagir avec plusieurs équipes au cours de ces 3 années de thèse, et j'ai eu la chance de pouvoir être conseillé par des experts dans leur domaine. Je remercie Thierry et Dominique (Dodo) pour les sessions foot du vendredi qui étaient vivifiantes, même si je n'ai pas pu y participer la dernière année. Je remercie également Rodolphe avec qui j'ai à la fois passé de bons moments et qui a facilité ma thèse par son implication dans la résolution des problèmes du bâtiment. Je remercie Müge, Luis, et Nicolas d'avoir participé aux soirées jeux avec autant d'assiduité, vous avez rendu ces soirées encore plus joyeuses et je vous souhaite le meilleur

pour la suite ! Enfin, je remercie dans sa globalité l'équipe FAAM, avec laquelle j'ai eu beaucoup d'interactions professionnelles et amicales, et qui m'a formé en cristallographie. Je ne sais toujours pas faire le remplacement moléculaire mais je suis à peu près capable de faire l'affinement ! J'espère que les burgredis continueront même après mon départ. Le flambeau m'a été donné par Corentin et c'est maintenant à vous que je le transmets. Vous êtes trop nombreux pour que je vous cite tous, sinon la reprographie va me charger un extra, mais je pense que vous savez tous que j'ai apprécié travailler avec vous.

Maintenant que le tour des collègues est passé, j'ai la difficile tâche de remercier mes amis. Certains pensent que la rédaction d'une thèse est compliquée, mais faire attention à n'oublier personne l'est encore plus. Je vais donc botter en touche et remercier le CROU, ceux qui s'y identifient se reconnaîtront... En fait non ! Je remercie mes amis de longue date, Gaspard qui m'a soutenu même avant que décide de faire de la bio, tu as toujours été présent et je n'ai pas besoin de beaucoup écrire pour que tu saches ce que je pense. Matos le Destructeur, dont les photos envoyées en soum-soum m'ont fait rire ou pleurer. Rocco l'Ingénieur, dont les discours passionnés, l'humour, et les dessins de « moutons » m'ont accompagné depuis de longues années. Monch le Bricoleur, également connu sous le nom de Mr. Technologie, tu peux trouver un job comme rédacteur en chef de FrAndroid. Théal, le Pipoteur, dont l'humour fin et les histoires classieuses ont égayé beaucoup de soirée. Gratthias le Surfeur, qui est parti s'exiler à Bordeaux tellement il en pouvait plus de mes blagues. Clemzer le Sapeur, qui restera probablement indétronable sur le nombre d'heures passées à geeker à deux. Orane qui fait Peur, mais seulement quand elle se donne à 350% pour Halloween, sinon tu fais pas trop peur promis ; merci pour les petits messages d'encouragement ! Max le Tombeur, merci pour tous les barbecues que tu as organisé, je suis pas le plus actif pour les relances donc heureusement que tu es au taquet ! Alice la palpeuse de Tumeurs, qui est quand même plus que juste un « +1 », je te souhaite du courage pour ta propre thèse. Je remercie aussi Julie qui m'a soutenu depuis le début de la rédaction jusqu'à la fin, et qui a su cerner le TDAH que je suis, je vais maintenant manquer d'excuses pour les oublis de rendez-vous. Petite mention particulière pour Chetan, la seule rencontre que j'ai faite à la conférence d'Heidelberg, merci pour toutes les soirées qu'on a passé à disserter sur le futur et j'espère que tu vas trouver un post-doc qui te plaise et que ton chef sera sympa ! Je remercie aussi beaucoup Corentin (et Fantine !) pour m'avoir appris pas mal de choses au début de ma thèse, m'avoir fourni un support technique de qualité, et m'avoir initié à la descente (pas de verre) où le ciel n'est pas celui qu'on croit. J'ai aussi une pensée pour la team originelle, Erwan et Axel, même si on se voit moins vous savez ce que je pense. Je remercie aussi toute la clique du volley : Max, Marie, Yuri, Florian, vous êtes au top ! Je remercie également la team escalade qui s'est élargie avec le temps : Clément, Florianne, Julie,

Kanok, petite pensée pour Méghane qui serait probablement de la partie si elle n'était pas allée faire un post-doc dans la perfide Albion. Je dois maintenant remercier les gens qui ont contribué à la réussite de ma thèse ! Merci à toi Sophia, pour les soirées tisane et pour ton aide sur la résolution de ma structure ! Je te souhaite pleins de bonheur dans la suite de ta carrière. Je remercie également Claire (tu as vu tu es pas dans les collègues) qui, en plus d'avoir fourni un support émotionnel et humoristique de qualité, m'a permis de développer mon karaté de pitoune. Je remercie mille fois Morane, pour tout, mais surtout pour ton aide dans les moments compliqués de la mise en page du manuscrit et pour tout le temps que tu as donné. Je reste persuadé que je gagnerai un jour au clanck. Enfin je remercie Hugo, mon ami et colloc' pour toutes les fois où tu m'as sermonné pour que je rédige, et pour avoir toi-même sacrifié quelques soirées dans mon intérêt. Tu gères dude, et même si tu sais pas cuisiner t'es le colloc' parfait!

Je remercie enfin bien sûr ma famille, mes parents, mon frère et ma sœur, pour leur soutien indéfectible et pour avoir grandement participé à faire de moi qui je suis aujourd'hui.

Liste des abréviations

Å : Angström

mA : milliampère

mins : minutes

h : heures

mM : millimolaire

μM : micromolaire

nM : nanomolaire

pM : picomolaire

TBS : *tris buffer saline*

TBST : *tris buffer saline tween*

N-ter : N-terminale

C-ter : C-terminale

pb : paires de base

bp : *base pair*

MM : marqueur de masse (moléculaire)

mL : millilitres

μL : microlitres

ng : nanogramme

μg : microgramme

mg : milligramme

g : gramme

rpm : rotation par minute

°C : degré celsius

K : degré Kelvin

NaCl : chlorure de sodium

mol : mole

kcal : kilocalorie

kDa : kilodalton

Da : dalton

IgG : Immunoglobuline G

Fc : Fragment cristallisable

scFv : *single chain variable fragment*

Fab : *antigen binding fragment*

ET : *extrait total*

FS : *fraction soluble*

Introduction

Introduction

Table des matières

Introduction.....	6
I. La reconnaissance moléculaire.....	8
II. Ingénierie modulaire	27
III. La purification cellulaire	36
IV. Stratégie générale	46
Bibliographie.....	50

I. La reconnaissance moléculaire

A. Les interactions protéine-protéine

Les protéines sont des molécules essentielles du vivant qui interviennent dans une multitude de processus cellulaires tels que la signalisation, la catalyse de réactions chimiques, le transport de molécules, la réplication cellulaire, la migration, etc... Leur fonction est directement liée à leur structure tridimensionnelle, qui dépend de leur séquence en acides aminés et leur permet d'établir des interactions à l'échelle moléculaire. Il existe 20 acides aminés usuels, ce qui donne un nombre de séquences possibles de 20^{100} pour une séquence de 100 acides aminés. La longueur médiane d'une séquence protéique se situant entre 240 et 360 acides aminés selon le règne du vivant concerné (Brocchieri, L. & Karlin, S., 2005), l'espace des séquences qui en résulte est si vaste qu'il peut potentiellement fournir une infinité de structures tridimensionnelles, et donc de surfaces d'interaction.

Les différents processus cellulaires évoqués précédemment se font en grande partie via des interactions protéine-protéine, qui entraînent la formation de complexes protéiques. Ces interactions se font par une combinaison de liaisons intermoléculaires faibles et non covalentes, comme les liaisons hydrogènes, les forces de Van der Waals, et les interactions électrostatiques, qui visent à stabiliser le complexe. Ces complexes se forment le plus souvent entre deux protéines, mais il existe aussi des complexes qui impliquent plus de deux protéines différentes comme le complexe de réplication de l'ADN par exemple, et de nombreuses autres machineries moléculaires.

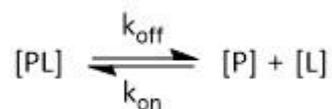
Ces complexes protéiques peuvent être décrits à travers la stabilité des monomères qui le composent. Ainsi, un complexe dit obligatoire fera intervenir des protéines qui ne sont pas stables sous leur forme monomérique comme le répresseur P22 Arc, ou d'autres répresseurs transcriptionnels. À l'inverse, un complexe non obligatoire sera composé de protéines stables indépendamment de la formation du complexe, comme le complexe RAF-1/MEK par exemple.

Un autre paramètre permettant la description des complexes protéiques est le temps de demi-vie du complexe. On peut distinguer les complexes permanents, dont l'interaction est très forte et conditionne l'existence du complexe (comme les complexes obligatoires), des complexes transitoires qui font intervenir un équilibre dynamique *in vivo* (Nooren, I. M. A., 2003). Les interactions impliquées dans les complexes transitoires peuvent être faibles, et le complexe concerné sera alors dans un équilibre dynamique en solution avec les monomères qui le composent, ou fortes, de sorte que le complexe transitoire formé nécessite un déclencheur moléculaire pour se dissocier. C'est le cas par exemple de la protéine G hétérotrimérique, qui se dissocie en deux sous-unités G_α et $G_\beta\gamma$ lorsqu'elle

est en interaction avec le GTP, mais forme un trimère stable en interaction avec le GDP. Globalement, la plupart des interactions sont transitoires et les complexes protéine-protéine sont donc dynamiques en solution.

La formation de tous ces types de complexes est dépendante de l'énergie libre du complexe (G_{complexe}), et plus particulièrement de la différence entre l'énergie libre du complexe et l'énergie libre des protéines qui le composent. Cette différence (notée $\Delta G_{\text{complexe}} = G_{\text{complexe}} - \Sigma G_{\text{protéines seules}}$) renseigne sur la spontanéité de la formation du complexe. En effet, un $\Delta G_{\text{complexe}}$ négatif signifie que l'énergie libre du complexe est plus faible que l'énergie libre des protéines seules, et la formation du complexe sera alors spontanée. A l'inverse, un $\Delta G_{\text{complexe}}$ positif indiquera que la formation du complexe n'est pas spontanée. Le ΔG du complexe peut s'exprimer par la formule $\Delta G_{\text{complexe}} = \Delta H_{\text{complexe}} - T\Delta S_{\text{complexe}}$, où $\Delta H_{\text{complexe}}$ et $\Delta S_{\text{complexe}}$ représentent respectivement la différence d'enthalpie et d'entropie entre le complexe et les protéines seules, et T représente la température. Le $\Delta H_{\text{complexe}}$ est directement liée à la surface d'interaction impliquée sur chaque partenaire qui forme le complexe. Le terme $T\Delta S_{\text{complexe}}$ correspond à la composante entropique du système, qui est une mesure de la variation du désordre des composants et du solvant. Dans une interaction entre deux protéines, ce terme peut également représenter un changement de conformation de la protéine. L'énergie d'activation de la formation du complexe, qui correspond à la différence d'enthalpie libre entre les protéines seules et l'intermédiaire de plus haute énergie, peut être affectée par plusieurs paramètres tels que l'environnement physico-chimique autour des partenaires protéiques, leur localisation dans le temps et l'espace, et leur concentration locale.

Le $\Delta G_{\text{complexe}}$ rend compte de l'affinité d'une protéine pour son partenaire. En effet, on peut exprimer le $\Delta G_{\text{complexe}}$ de la façon suivante : $\Delta G_{\text{complexe}} = -RT \ln(K_D)$ où R est la constante des gaz parfaits, T est la température du système, et K_D correspond à la constante d'équilibre de la réaction de dissociation du complexe protéique à l'équilibre. En écrivant cet équilibre,



où [PL] correspond au complexe protéique composé des protéines [P] et [L], on peut montrer que la constante K_D est égale au rapport entre la constante de vitesse de dissociation du complexe et sa constante de vitesse d'association ($k_{\text{off}}/k_{\text{on}}$). L'affinité d'une protéine pour ses partenaires est un paramètre crucial, puisqu'il influe sur la durée de vie du complexe et donc sa disponibilité biologique. Un des paramètres qu'on relie souvent à l'affinité d'une protéine pour une autre est la surface enfouie, c'est-à-dire, la surface qui n'est plus accessible au solvant après la formation du complexe. L'étude de la surface enfouie induite par la formation de plusieurs complexes a été étudiée dans les travaux de [Chen, J. et al., 2013](#). Globalement, ces travaux montrent une corrélation entre l'augmentation de la

surface enfouie d'un complexe et l'augmentation de l'affinité entre les protéines de ce complexe. De plus, les auteurs ont étudié l'impact de la taille de la surface enfouie sur l'enthalpie libre par Å² enfoui. Ils observent une différence de comportement entre les surfaces enfouies supérieures à ~2000 Å², dont l'enthalpie libre par Å² enfoui stagne autour de 4 cal.mol⁻¹.Å⁻², et les surfaces enfouies inférieures à ~2000 Å², dont l'enthalpie libre par Å² enfoui varie entre 4 cal.mol⁻¹.Å⁻² et 13 cal.mol⁻¹.Å⁻². Ces résultats tendent à soutenir l'hypothèse selon laquelle il existe des *hots spots* d'interaction, qui sont des résidus contribuant bien plus que les autres à l'énergie d'interaction. Les auteurs ont montré que les surfaces d'interactions plus faibles sont proportionnellement plus riches en *hots spots*. Les *hots spots* semblent donc permettre une forte affinité même avec un petit nombre de résidus impliqués dans l'interaction. Les *hots spots* sont généralement identifiés par *alanine scanning*. Cette méthode consiste à muter un par un les résidus d'une protéine, afin de déterminer l'implication de chaque résidu dans divers paramètres comme la stabilité de la protéine ou leur implication dans une interaction. Il est en effet possible d'identifier les résidus qui contribuent le plus à l'énergie de liaison, et on peut définir comme *hot spot* tout résidu contribuant à l'interaction à hauteur de 2 kcal/mol ou plus. (Bogan, A. A. & Thorn, K. S., 1998)

L'une des caractéristiques des interactions protéine-protéine est la spécificité des partenaires l'un pour l'autre. En effet, les protéines évoluent dans un milieu très encombré, où elles sont en contact permanent avec une multitude d'autres protéines, de petites molécules, etc... Leur proximité permanente avec d'autres surface protéique requiert que les protéines soient spécifiques pour leur cible. Il existe aussi des interactions multi-spécifiques, où une même protéine peut interagir avec plusieurs partenaires qui sont alors en compétition les uns avec les autres. Ce mécanisme est souvent utilisé dans des cascades de signalisation par exemple, où la fixation de l'un ou l'autre des partenaires entraîne une modification du signal transmis. Certaines protéines vont par exemple reconnaître un domaine phosphorylé, mais ne pas reconnaître le même domaine non phosphorylé, ce qui permet la régulation de la voie de signalisation impliquée. La spécificité dépend directement de la complémentarité de formes entre les protéines, et des interactions au niveau des chaînes latérales impliquées, ce qui détermine l'énergie libre du complexe.

B. Les anticorps, un outil de reconnaissance incontournable

Historiquement, les anticorps (IgGs) sont les molécules phares de la reconnaissance moléculaire. En effet, leur fonction biologique nécessite que les anticorps soient à la fois affins et spécifiques pour les cibles qu'ils reconnaissent. Les IgGs ont une structure en forme de Y, dont le bout des « bras » fixe l'antigène et la « tige » déclenche les fonctions effectrices qui servent à éliminer l'antigène. Ils sont constitués de 4 chaînes polypeptidiques, 2 chaînes dites « lourdes » (H) et 2 chaînes dites « légères » (L) (Figure 1). Sur chaque chaîne, lourde ou légère, on peut distinguer au moins un domaine constant et un seul domaine variable, qui tient son nom de l'hypervariabilité de 3 boucles exposées au solvant (également appelées CDRs pour *complementary determining region*). Une chaîne lourde et une chaîne légère s'assemblent pour former un bras, et 2 bras identiques s'assemblent par le biais de ponts disulfures et d'un *packing* latéral pour former l'anticorps. L'anticorps possède donc 2 « bras » qui portent chacun un domaine hypervariable identique, constitué par les 3 boucles hypervariables de la chaîne lourde et les 3 boucles hypervariables de la chaîne légère.

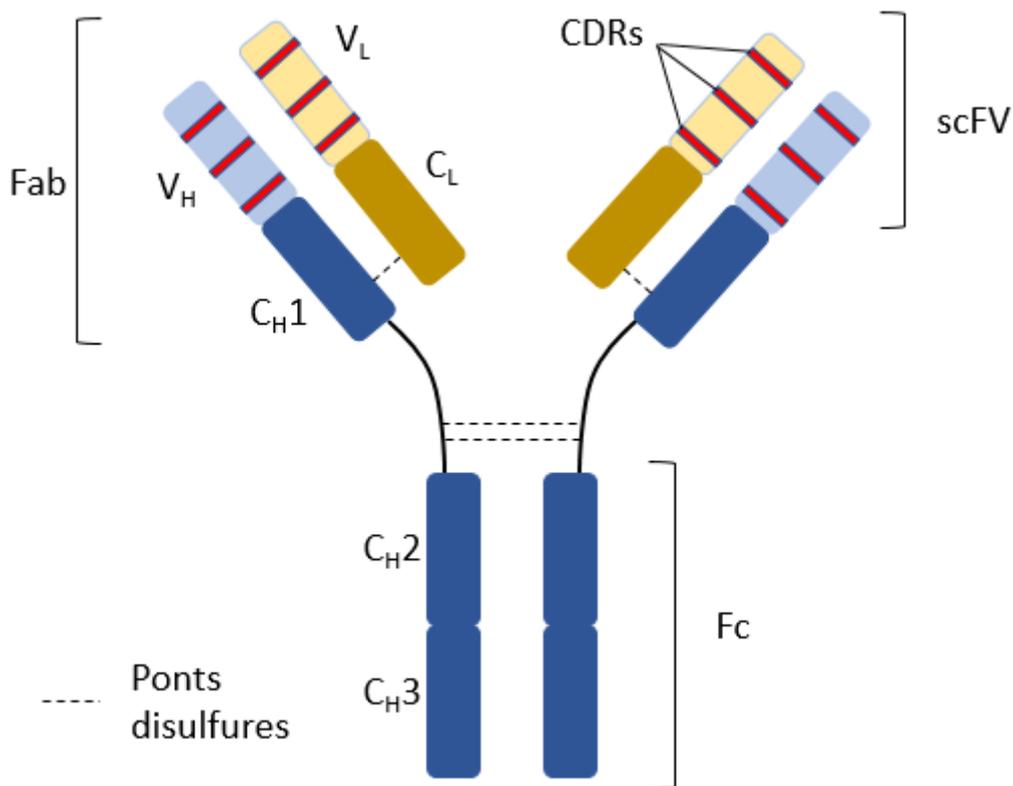


Figure 1 : Représentation schématique de la structure d'une immunoglobuline de type G1 humaine (IgG1)

La structure d'une IgG est composée de 4 chaînes polypeptidiques, 2 lourdes (H) et 2 légères (L) reliées par au moins un pont disulfure et des interactions faibles latérales. Les domaines variables (V) des chaînes lourdes et légères portent l'hypervariabilité au niveau de 3 boucles appelées CDRs (en rouge). Ce sont les CDRs qui sont impliquées dans l'interaction avec l'antigène. Les domaines CH2 (constant heavy 2) et CH3 (constant heavy 3) constituent le domaine effecteur (Fc) qui va être impliqué dans le recrutement des cellules de l'immunité (lymphocytes T, macrophages, etc...).

L'un des tournants dans l'utilisation des anticorps a probablement eu lieu en 1975, lorsque G. Köhler et C. Milstein ont montré qu'il était possible de fusionner des cellules de myélome avec des lymphocytes B afin de générer des hybridomes capables d'exprimer des anticorps monoclonaux spécifiques d'un épitope particulier sur une cible (Köhler, G. & Milstein, C., 1975). Les anticorps monoclonaux ont alors largement été utilisés dans le monde entier, dans des applications diverses telles que la détection et la quantification de protéines en recherche fondamentale ou dans le domaine de la biologie médicale, comme transporteur de principe actif, etc... Cependant, ces anticorps provenant des souris, leur utilisation dans le domaine thérapeutique et dans le diagnostic *in vivo* a posé plusieurs problèmes. D'une part, les anticorps monoclonaux issus d'hybridomes peuvent être à l'origine de réponses immunitaires dirigées contre les séquences propres aux anticorps de souris (Courtenay-Luck, N. S. et al., 1986), ce qui n'est pas compatible avec un usage chez le patient. D'autre part, leur production nécessite l'immunisation d'un animal, ce qui est un frein à l'obtention d'anticorps dirigés contre des cibles toxiques, ou très conservées entre les espèces.

Une stratégie a été mise en place afin d'obtenir des anticorps dits « chimérisés », qui consiste à mixer entre eux les domaines variables d'anticorps murins, et des fragments constants d'anticorps humains (Neuberger, M. S. et al., 1985). Cela permet d'allier l'obtention des anticorps chez la souris, plus facile que l'obtention d'anticorps humains, avec les propriétés des domaines constants des anticorps humains pour pallier aux problèmes évoqués précédemment (Winter, G. & Milstein, C., 1991). Pour aller plus loin dans l'ingénierie des anticorps, il est possible de greffer directement les boucles des anticorps murins spécifiques d'une cible sur l'architecture d'un anticorps humain pour obtenir des anticorps dits « humanisés » (Jones, P. T. et al., 1986).

D'autres méthodes que la technologie des hybridomes ont été développées, notamment pour faire face aux problèmes liés à l'immunisation des animaux. Ces techniques reposent sur l'exposition de bibliothèques de fragments d'anticorps et leur sélection grâce à des méthodes de *display in vitro*, permettant de sélectionner des interacteurs sans forcément passer par l'immunisation d'un animal. Le principe consiste à créer une bibliothèque de fragments d'anticorps qui comporte un très grand nombre de variants, et de présenter chacun de ces variants sur une plateforme de présentation. Ces plateformes de présentation, également appelées méthodes de *display* sont le phage *display*, le *yeast display*, ou le ribosome et mRNA *display* (Boder, E. T. & Wittrup, K. D., 1997; Feldhaus, M. J. et al., 2003; McCafferty, J. et al., 1990), et consistent à exposer les protéines que l'on souhaite cribler

respectivement à la surface d'un phage, d'une levure, d'un ribosome ou d'un ARN messager. Les protéines les plus affines pour une cible spécifique sont alors capturées par l'intermédiaire de cette cible. L'avantage de ces techniques est qu'il existe un lien physique direct entre la protéine sélectionnée et son ADN, ce qui permet de récupérer facilement l'ADN codant pour les protéines les plus affines pour la cible. Cela facilite l'ingénierie des interacteurs, puisqu'il est par exemple possible dans le cas des anticorps de recombinaison entre elles les séquences codantes pour les CDRs dans le but de faire de la maturation d'affinité (Yang, W.-P. et al., 1995). Un autre avantage de ces techniques est qu'il est possible d'obtenir des interacteurs de protéines homologues entre l'Homme et la souris, ou des cibles toxiques qui ne pourraient pas être administrées à l'animal dans un processus d'immunisation classique. En 2018, George Smith et Gregory Winter ont été récompensés pour leurs travaux sur le phage *display* par le prix Nobel de chimie, ce qui témoigne des avancées réalisées grâce à ces méthodes de *display*, en particulier dans la présentation de bibliothèques de fragments d'anticorps à la surface des phages.

Bien que l'humanisation des anticorps et les techniques de sélection *in vitro* aient permis de s'affranchir du problème de l'immunogénicité chez l'Homme, leur utilisation de plus en plus répandue a soulevé plusieurs problèmes inhérents à leur structure. D'une part, leur taille (environ 150 kDa) rend difficile leur utilisation dans l'imagerie *in vivo*. Leur taille explique également une élimination rénale lente, favorable à une action thérapeutique. Le récepteur aux Fc néonatal (FcRn) participe également à la longue demi-vie des anticorps puisqu'il inhibe la dégradation lysosomale de l'albumine et des IgGs dans les cellules endothéliales. Cette élimination lente est cependant défavorable aux applications d'imagerie, et nécessite de réaliser la prise d'image plusieurs jours après l'administration du produit pour minimiser le bruit de fond. D'autre part, les IgGs sont composés de 4 chaînes polypeptidiques, qui comportent toutes au moins un pont disulfure essentiel, et dont les chaînes lourdes sont glycosylées. Leur production doit donc être réalisée dans un organisme eucaryote, et l'optimisation de la production est à la fois laborieuse et coûteuse.

Pour toutes ces raisons, des alternatives aux IgGs classiques ont été développées, basées soit sur l'ingénierie des immunoglobulines elles-mêmes, soit sur le développement d'ossatures protéiques alternatives.

En 1959, G. M. Edelman a montré que les anticorps étaient composés de plusieurs sous-unités reliées entre elles notamment par des ponts disulfures, et qu'il était possible d'isoler ces domaines (Edelman, G. M., 1959). La même année, R. R. Porter observe que la digestion de la γ -globuline du lapin donne 3 fragments distincts, dont 2 portent encore la capacité de liaison à l'antigène (Porter, R. R., 1959). Ces travaux vont donner naissance aux Fab (*antigen binding fragment*), des fragments d'environ 50 kDa qui sont constitués du domaine variable des chaînes lourdes et légères ainsi que du domaine C_{H1} et du domaine C_L. Les Fabs portent donc le site de reconnaissance de l'antigène mais ne possèdent pas le

domaine Fc et n'ont donc pas les fonctions effectrices qui lui sont associées. Plus tard, le développement de l'ingénierie protéique via l'évolution des techniques de biologie moléculaire a permis d'obtenir des fragments encore plus petits (~27 kDa) qu'on appelle scFv (*single chain variable fragment*) et qui ne contiennent que les domaines variables des chaînes lourdes et légères, reliés entre eux par un *linker* flexible.

Ces fragments d'anticorps présentent plusieurs avantages par rapport aux IgGs entières, notamment dans le domaine thérapeutique. Tout d'abord, leur petite taille leur permet de mieux pénétrer les tumeurs et les tissus, et leur permet d'avoir accès à des épitopes plus enfouis (épitopes cryptiques) que ceux classiquement reconnus par les IgGs entières (Bates, A. & Power, C. A., 2019). D'autre part, la plupart d'entre eux peuvent être produits dans le périplasme d'organismes procaryotes, ce qui diminue grandement leur coût de production et améliore les rendements. Enfin, l'absence de domaine Fc permet d'éviter l'activation accidentelle de cellules de l'immunité. Les fragments d'anticorps ont également permis le développement des méthodes de sélection d'interacteurs *in vitro* qui ont été discutées précédemment (Geyer, C. R. et al., 2012). C'est d'ailleurs dans cette optique que des efforts ont été faits pour créer des bibliothèques de fragments d'anticorps naïves, dont certaines ont même été optimisées pour exprimer des scFvs fonctionnels dans le cytoplasme de bactéries (Philibert, P. et al., 2007).

Bien que les fragments d'anticorps présentent de nombreux avantages, ils ne sont pas exempts de défauts. Dans le cadre d'un usage thérapeutique, leur petite taille les rend facilement éliminables par les reins, ce qui nécessite d'administrer régulièrement de fortes doses pour atteindre et maintenir la dose efficace. Pour augmenter leur demi-vie dans l'organisme, plusieurs techniques existent comme fusionner les fragments d'anticorps avec du PEG ou avec des domaines interagissant avec l'albumine, mais ces procédés sont coûteux et lourds d'un point de vue technique, et viennent contrebalancer les avantages de la production en système procaryote. D'autre part, comme le fragment Fc participe grandement à la stabilité de la structure des IgGs, son absence entraîne une diminution de la stabilité des fragments d'anticorps, qui se traduit par une plus forte propension à l'agrégation. Les Fabs ont donc tendance à être moins stables que les anticorps monoclonaux classiques, et les scFvs ont tendance à être moins stables que les Fabs (Röthlisberger, D. et al., 2005). De plus, dans le cas des scFvs, il est nécessaire d'optimiser le *linker* qui relie les 2 domaines variables afin que celui-ci soit à la fois d'une longueur optimale, et peu immunogène. Enfin, on peut noter que peu importe le format, il subsiste toujours des ponts disulfures intra-domaines, nécessaires à leur bon repliement, ce qui complique leur utilisation. Bien que certaines équipes se soient penchées sur la résolution de ce problème, notamment avec les *intrabodies*, la plupart des formats utilisés aujourd'hui ne sont pas optimisés pour être produits ni actifs dans le cytoplasme des cellules.

La dernière génération de fragments d'anticorps repose sur la structure d'anticorps naturellement présents dans le règne animal. On retrouve ainsi chez les camélidés en plus des IgGs classiques, des anticorps qui sont uniquement composés d'une chaîne lourde constituée d'un domaine variable et 2 domaines constants. La capacité de liaison à l'antigène est portée uniquement par le domaine variable (VHH), qui peut être produit seul de façon soluble. Les VHH possèdent tous les avantages des fragments d'anticorps, mais sont plus stables que les Fab et les scFvs. Ils peuvent supporter une large gamme de pH, de température, et résistent même à certains solvants organiques. Ils sont solubles en solution, et ont moins tendance à s'agréger puisque contrairement aux Fab et aux scFvs, leur stabilité n'implique pas d'interactions inter-domaine. Bien que d'origine non-humaine et souvent humanisés, leur similarité avec la famille de gènes VH3 chez l'homme les rend peu immunogènes. En revanche, ils rencontrent le même problème que les fragments d'anticorps au niveau de l'élimination par les reins, puisque leur petite taille (12-15 kDa) leur confère un temps de demi-vie dans l'organisme de 2 heures, ce qui est défavorable lors d'un usage thérapeutique chronique. Assez similaires aux VHH, les V-NAR (11kDa) sont les domaines variables d'antigènes de requin qui portent seuls les 2 CDRs (au lieu de 3) impliquées dans la liaison de l'antigène. Ils peuvent être produits seuls, sont encore plus stables que les VHH et possèdent à peu près les mêmes propriétés. Enfin, il existe une autre classe de molécules de l'immunité chez les agnathes, appelés les VLRs (*variable lymphocytes receptors*), qui représente une alternative aux immunoglobulines. Cette classe de molécules se compose d'un domaine N-ter de type LRR (*Leucine rich repeat*), composé notamment d'un nombre variable de motifs LRRs de 24 acides aminés. La diversité du répertoire immunitaire est contenue dans différentes cassettes LRRs présentes dans le *locus* VLR qui seront intégrées au gène incomplet des VLRs pendant la maturation lymphocytaire afin de générer de la reconnaissance spécifique des antigènes (Pancer, Z. *et al.*, 2004). Ces architectures sont très prometteuses et sont la cible de nombreuses études (Bathula, N. V. *et al.*, 2021; Bezverbnaya, K. *et al.*, 2017).

L'autre alternative aux anticorps monoclonaux a été le développement d'ossatures artificielles qui possèdent une capacité de liaison à une cible d'intérêt sans présenter les défauts inhérents à la structure des IgGs et de leurs dérivés.

C. Les ossatures protéiques alternatives et les bibliothèques de protéines artificielles

Parallèlement au développement des anticorps et de façon indépendante, différentes équipes ont cherché à exploiter des structures non apparentées aux immunoglobulines et sur lesquelles il était possible de modifier ou d'introduire une surface d'interaction. Plusieurs architectures ont ainsi été obtenues, et ont été dénommées de façon générale comme les ossatures alternatives. Ces ossatures alternatives se prêtent tout à fait aux méthodes de *display* qui permettent la sélection *in vitro* d'interacteurs affins et spécifiques de cibles d'intérêt à partir de bibliothèques de protéines.

Plusieurs bibliothèques d'ossatures alternatives ont donc été développées à partir de la fin des années 1990, mettant fin au dogme selon lequel les anticorps seraient la seule classe de protéines utilisables comme outil de reconnaissance d'antigènes d'intérêt en biotechnologie et en médecine.

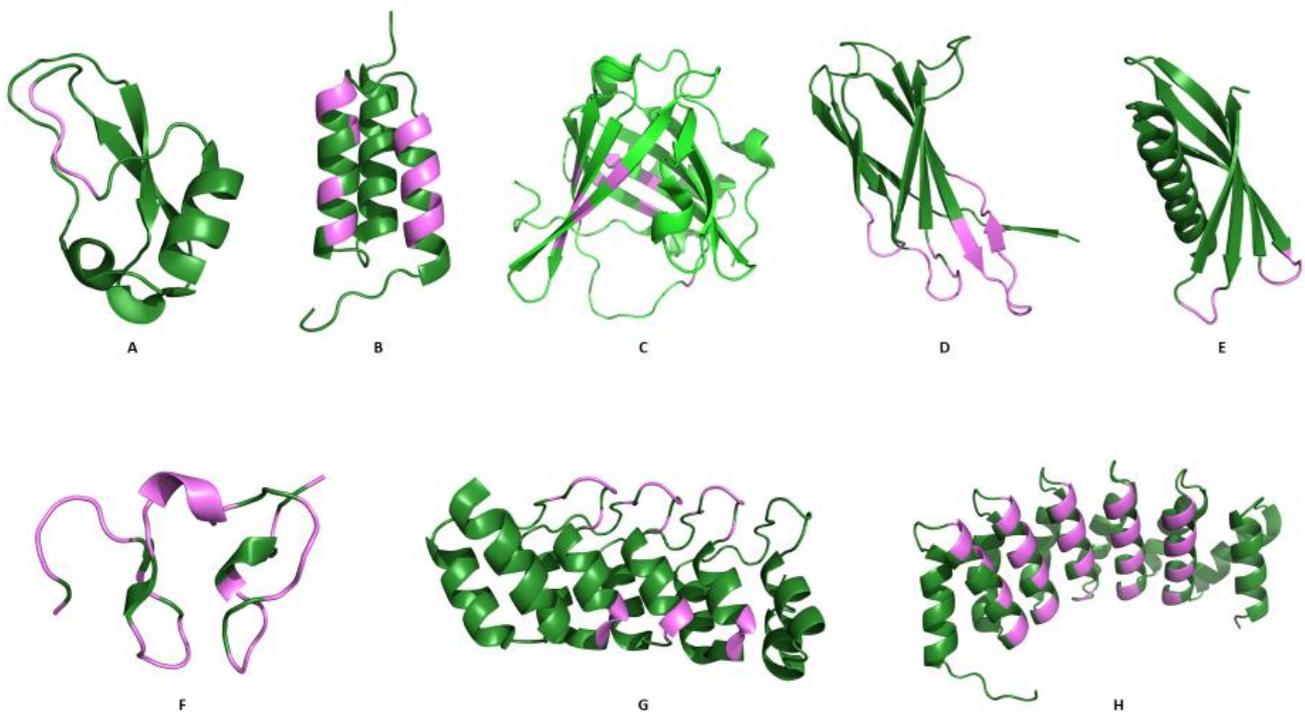


Figure 2 : Structures de différentes architectures alternatives aux anticorps représentée en *cartoon*

Les positions diversifiées sont représentées en rose, et le reste de la protéine est coloré en vert. (A) Domaine de Kunitz (code PDB : 2RA3) (B) Affibody (code PDB : 5DJT) (C) Anticaline (code PDB : 3DSZ) (D) Monobody (code PDB : 3K2M) (E) Adhiron (code PDB : 4N6T) (F) Avimer (code PDB : 1AJJ) (G) DARPins (code PDB : 1SVX) (H) α Rep (code PDB : 3LTJ)

L'une des premières ossatures alternatives aux anticorps identifiée a été la famille des inhibiteurs à protéases de type Kunitz. En effet, les protéines de cette famille sont généralement petites et stables, et leur domaine de liaison à la protéase consiste souvent en un court allongement peptidique présenté

sous la forme d'une ou plusieurs boucles, ce qui en fait de bons candidats pour être la cible d'ingénierie protéique. Un domaine de Kunitz fait généralement une soixantaine d'acides aminés et adopte un repliement en hélices α et brins β permettant de présenter 2 boucles étendues qui servent à lier le site actif d'une protéase (Figure 2). Plusieurs bibliothèques ont ainsi été créées, basées par exemple sur la structure du premier domaine Kunitz de l'inhibiteur de la coagulation associé aux lipoprotéines humain (LACI), ou sur les domaines actifs des inhibiteurs pancréatiques de trypsine bovins (BPTI) ou humains (PSTI). Des interacteurs spécifiques de cibles d'intérêt ont pu être sélectionnés à partir de ces bibliothèques avec des affinités allant jusqu'au picomolaire (Roberts, B. L. *et al.*, 1992), et un inhibiteur de l'inflammation médiée par l'élastase des neutrophiles a montré des effets significatifs (Dunlevy, F. K. *et al.*, 2012). Cependant, cette architecture ne semble pas avoir connu l'essor qu'ont connu les autres architectures présentées ci-après, notamment du fait des ponts disulfures nécessaires au repliement de la protéine qui limite la production cytoplasmique.

Les Affibodies sont une autre ossature alternative découverte vers la fin des années 1990. Leur structure est basée sur celle du domaine B de la protéine A des staphylocoques (SPA), qui est composé de 58 acides aminés repliés en 3 hélices α antiparallèles. Dans la nature, la protéine A des staphylocoques fixe la plupart des IgGs de mammifères par leur domaine Fc, et la structure du complexe entre le domaine B de la SPA et une IgG humaine a été obtenue par cristallographie à 2.8 Å (Deisenhofer, J., *s. d.*). À partir de cette structure et des données présentes dans la littérature, un domaine Z synthétique a été conçu dans le but de parvenir à un domaine résistant à des traitements chimiques utilisés en routine dans le domaine médical (Nilsson, B. *et al.*, 1987). Ce domaine Z a alors été utilisé comme base structurale pour le développement de la bibliothèque des Affibodies (Figure 2). Les positions qui ont été choisies pour servir de surface hypervariable au sein de la bibliothèque sont principalement les positions impliquées dans l'interaction entre le domaine B de SPA et l'IgG humaine, et des chaînes latérales exposées au solvant se trouvant sur la même « face » (Nord, K. *et al.*, 1995). Cette bibliothèque a donné des interacteurs de plusieurs cibles non apparentées (insuline, transferrine, TNF- α , IL-8, gp120, CD28, HER2, EGFR, etc...) avec des K_D de l'ordre du μ M au pM. De plus, si l'affinité obtenue à l'issue des sélections n'est pas assez forte pour le type d'application désiré, des étapes de maturation d'affinité sont possibles par recombinaison d'hélices (Gunneriusson, E. *et al.*, 1999) ou par alignement de séquences combiné à des mutations dirigées combinatoires (Orlova, A. *et al.*, 2006). Les Affibodies ont été utilisés dans des domaines tels que l'imagerie *in vivo*, la purification d'anticorps, ou dans le domaine thérapeutique comme protéine de ciblage. Pour cette dernière application, un Affibody spécifique d'une cible thérapeutique est fusionné à une molécule effectrice (souvent une toxine) dans le but d'adresser cet agent toxique directement au niveau des cellules responsables de la pathologie. Cette technique doit permettre de minimiser la toxicité du traitement, c'est-à-dire le ciblage aspécifique des tissus sains, grâce à la spécificité de l'Affibody pour sa cible. Ce

principe a été utilisé avec succès par R. Zielinski *et al.*, en fusionnant un Affibody spécifique de HER2 avec la toxine PE38KDEL, une version tronquée de l'endotoxine A de *Pseudomonas*, pour cibler les cellules tumorales HER2-positives (Zielinski, R. *et al.*, 2009).

Également développées vers la fin des années 90, les Anticalines sont une autre classe de protéines artificielles. Elles sont basées sur la structure des lipocalines, une classe de protéines généralement impliquées dans le transport ou le stockage de molécules hydrophobes ou de composés organiques sensibles. Les lipocalines ont une structure conservée composée de 8 brins β organisés en tonneau- β , et connectés entre eux par 4 boucles exposées sur la même face du tonneau (Figure 2). Si les brins β des différentes lipocalines dont on possède la structure s'alignent très bien entre eux, ce n'est le cas des 4 boucles qui varient par leur taille et leur conformation entre les différentes lipocalines. Ces boucles sont impliquées dans la spécificité de l'interaction avec le ligand, et elles ont donc été la cible d'ingénierie protéique visant à les modifier pour générer une bibliothèque de lipocalines (Beste, G. *et al.*, 1999). Pour cela, la structure d'une lipocaline, la BBP (bilin-binding protein), a été modifiée au niveau de 16 positions qui couvrent majoritairement la surface d'interaction avec le ligand naturel. Ces positions sont réparties entre les 4 boucles qui connectent les brins du tonneau β , et des positions adjacentes sur les brins β . Les positions ainsi mutées forment une poche peu profonde propice à la fixation de petites molécules. Une bibliothèque de 3.7×10^8 variants a ainsi été obtenue et a donné des interacteurs (appelés anticalines) de petites molécules comme la fluorescéine ou la digoxigénine avec des affinités de l'ordre du nanomolaire après maturation d'affinité. D'autres bibliothèques de lipocalines ont été conçues dans le but de reconnaître des protéines et d'être utilisées dans le domaine thérapeutique (Schonfeld, D. *et al.*, 2009; Vogt, M. & Skerra, A., 2004). Pour faciliter ce champ d'application et diminuer l'immunogénicité, ces nouvelles bibliothèques ont été basées sur la structure de lipocalines humaines comme l'apolipoprotéine D (ApoD) qui était la plus proche de BBP d'un point de vue structural, ou la lipocaline neutrophile associée à la gélatinase (NGAL). De la même manière que pour la première bibliothèque de lipocalines, 24 (bibliothèque ApoD) ou 20 (bibliothèque NGAL) positions au niveau des boucles et du site de fixation naturel ont été la cible de mutagenèse aléatoire afin de générer l'hypervariabilité nécessaire à la constitution de la bibliothèque. Les bibliothèques ainsi obtenues comportent 1.7×10^9 variants (ApoD) ou 2×10^{10} variants (NGAL), et ont donné des interacteurs de cibles variées telles que l'hémoglobine (Vogt, M. & Skerra, A., 2004), l'antigène-4 des lymphocytes T cytotoxiques (CTLA-4) (Schlehuber, S. & Skerra, A., 2005) ou le facteur de croissance endothélial vasculaire (VEGF) (Hohlbaum, A. M. & Skerra, A., 2007) avec des affinités subnanomolaires. Ces résultats valident le choix de l'architecture des lipocalines pour générer des bibliothèques pouvant reconnaître de petites molécules comme des protéines, et sont prometteurs pour des applications thérapeutiques.

Le 10^{ème} domaine de la fibronectine humaine de type III (¹⁰F_n3 ou FN) a servi de base structurale pour générer plusieurs bibliothèques de protéines artificielles. En effet, ¹⁰F_n3 appartient à la superfamille des IgGs en raison de sa structure organisée en sandwich β très proche de celle du domaine V_H des IgGs. Cependant, ¹⁰F_n3 ne possède pas de pont disulfures, et est plus stable que les IgGs, ce qui en fait un candidat idéal pour générer une bibliothèque de protéines artificielles. Pour générer une telle bibliothèque, la séquence de ¹⁰F_n3 a été mutée au niveau des boucles qui se situent à proximité de l'extrémité N-ter et qui sont analogues aux CDRs chez les IgGs (Figure 2). Une première bibliothèque de ~10⁸ clones indépendants a ainsi été obtenue en mutant de façon aléatoire 5 positions sur chacune des boucles BC et FG qui sont analogues aux CDRs 1 et 3 d'une IgG, respectivement (Koide, A. *et al.*, 1998). Cette bibliothèque a donné des interacteurs de l'ubiquitine de faible affinité, et sa diversité était trop faible devant l'espace des séquences théorique. D'autres bibliothèques basées sur ¹⁰F_n3 ou des domaines analogues aux fibronectines ont donc été créées selon le même principe, en modifiant 1, 2 ou les 3 boucles homologues aux CDRs des IgGs et dans certains cas les boucles non homologues aux CDRs présentes à l'autre extrémité de la protéine (Diem, M. D. *et al.*, 2014; Getmanova, E. V. *et al.*, 2006; Gilbreth, R. N. *et al.*, 2008; Hackel, B. J. *et al.*, 2008, 2010; Hackel, B. J. & Wittrup, K. D., 2010; Karatan, E. *et al.*, 2004; Koide, A. *et al.*, 2002, 2007; Liao, H.-I. *et al.*, 2009; Lipovsek, D., 2011; Lipovšek, D. *et al.*, 2007; Olson, C. A. *et al.*, 2008; Parker, M. H. *et al.*, 2005; Richards, J. *et al.*, 2003; Swers, J. S. *et al.*, 2013; Wojcik, J. *et al.*, 2010; Xu, L. *et al.*, 2002). Ces bibliothèques ont donné des interacteurs de cibles variées et non apparentées, et le nombre de bibliothèques créées ainsi que le nombre de groupes travaillant sur cette ossature depuis 1998 montre bien l'intérêt que peuvent avoir les protéines issues de ces bibliothèques. En effet, l'origine humaine de cette ossature la rend peu immunogène et son abondance dans le sang est un indicateur de sa non-toxicité, ce qui en fait un bon candidat pour des usages thérapeutiques (Goldberg, S. D. *et al.*, 2016; Gross, G. G. *et al.*, 2013; Nichols, A. L. *et al.*, 2018).

Les avimers sont une autre classe de protéines ayant servi à réaliser des bibliothèques d'interacteurs, et leur ossature est basée sur celle des domaines dits de la famille A. Cette famille de domaines est présente dans différents récepteurs tels que les LRP (low-density related proteins) ou les VLDLR (very low-density lipoprotein receptor) (North, C. L. & Blacklow, S. C., 1999). Chaque domaine A contient environ 35 acides aminés, et adopte une structure uniforme stabilisée par le calcium et 3 paires de ponts disulfures (Figure 2) (Koduri, V. & Blacklow, S. C., 2001). Si 12 résidus sont cruciaux pour la structure, le reste des positions peut être diversifié pour générer une bibliothèque d'avimers d'environ 10¹⁰ variants compatible avec le phage *display* (Silverman, J. *et al.*, 2005). Cette bibliothèque a donné des interacteurs de plusieurs cibles d'intérêt thérapeutique comme CD28, IL-6, cMET, CD40L, et BAFF. Certains des interacteurs obtenus ont des affinités de l'ordre du picomolaire pour leur cible, et leur stabilité dans du sérum humain a été démontré. Ces résultats sont encourageants pour la découverte

de nouveaux agents thérapeutiques, bien que la présence des nombreux ponts disulfures nécessaires au bon repliement des avimers soit un frein à certaines utilisations. Ainsi, des avimers reconnaissant le collagène ont été employées pour rallonger le temps de demi-vie de protéines thérapeutiques utilisées dans le traitement de l'arthrose (Hulme, J. T. *et al.*, 2017).

Une autre bibliothèque de protéines artificielles qui semble prometteuse est celle des affimers (initialement appelée adhiron). L'architecture de cette bibliothèque est basée sur celle des phycocystatines, une famille de protéines inhibitrices des protéases à cystéines dans le règne végétal. La bibliothèque a été construite à partir d'une séquence consensus obtenue par un alignement de 57 phycocystatines, et la structure correspondante comporte un feuillet β à 4 brins et une hélice α , ne possède pas de ponts disulfures (Figure 2). Son T_m est supérieur à 100°C, et cette protéine est monomérique en solution et se produit efficacement dans des systèmes bactériens. Dans cette structure, 2 boucles ont été mutées aléatoirement sur 9 positions chacune pour générer une diversité d'environ 10^{10} variants (Tiede, C. *et al.*, 2014). Pour des usages thérapeutiques, une autre architecture a servi de base à une nouvelle bibliothèque très similaire à la première. Il s'agit de l'inhibiteur de protéases humaine stefine A, qui est un homologue des phycocystatines chez l'Homme. La structure des protéines des deux types de bibliothèques sont identiques et les mêmes positions sont randomisées. Ces bibliothèques ont donné des interacteurs de nombreuses cibles non apparentées telles que SH2, VEGFR2, TRPV1, MDV-1, IL-37 mais aussi de petites molécules comme la TNT (Kyle, S., 2018). Cette architecture semble très prometteuse, tant dans le domaine diagnostique que dans le domaine thérapeutique. À ce titre, elle est d'ailleurs commercialisée par la société Avacta Life Sciences.

L'architecture protéique alternative aux anticorps la plus développée sur le plan biomédical est celle des DARPins. Cette ossature est basée sur des motifs ankyrin répétés. Dans la nature, les protéines à motifs ankyrin répétés sont souvent impliquées dans les interactions protéines-protéines. Elles consistent en la répétition d'un module d'environ 33 acides aminés, organisé en un coude β suivi de deux hélices α antiparallèles (Figure 2). Binz *et al* ont exploité la nature modulaire de ces protéines pour générer des bibliothèques combinatoires de protéines constituées de modules répétés auto-compatibles (Binz, H. K. *et al.*, 2003). Pour cela, ils ont développé une méthode appelée *consensus design* qui permet d'identifier les régions diversifiables dans les protéines à motifs répétés (Forrer, P. *et al.*, 2004). Le principe de cette méthode consiste à réaliser un alignement de séquences codant pour un motif répété, ici un motif ankyrine, et qui partage donc le même repliement. Il est alors possible de dégager une séquence consensus, dont la conservation des résidus renseigne sur l'importance des résidus à chaque position de la protéine. On peut faire l'hypothèse que les résidus impliqués dans le maintien de la structure de la protéine sont plus conservés que les autres, car même si sa séquence se modifie au cours de l'évolution une protéine doit toujours pouvoir se replier pour exercer sa fonction.

A l'inverse, des résidus impliqués dans la fonction de la protéine sont susceptibles de changer si la fonction n'est pas partagée par tous les membres de la famille de protéines étudiée. Grâce à la séquence consensus, il devient alors possible de déterminer les positions hypervariables qui sont diversifiables sans altérer l'intégrité structurale de la protéine. Dans la bibliothèque développée par Binz *et al.*, chaque module varie au niveau de 6 positions hypervariables qui sont situées dans le coude β et dans la première hélice α qui suit (Binz, H. K. *et al.*, 2004). Les modules répétés sont entourés par 2 modules « externes » spéciaux qui ne possèdent pas les mêmes résidus au niveau de leurs faces latérales, exposées au solvant. En effet, les modules dits « internes » possèdent des résidus hydrophobes au niveau de leurs faces latérales, ce qui leur permet de s'assembler en un solénoïde compact. De telles propriétés sur les modules externes pourraient conduire à une agrégation des protéines en solution via les résidus hydrophobes exposés au solvant, et c'est dans ce but que les modules externes ont été conçus de façon à ne pas exposer de résidus hydrophobes. Ces bibliothèques de DARPins comportent généralement entre 2 et 3 modules internes, et utilisent le ribosome *display* comme plateforme de présentation et de sélection, ce qui permet d'obtenir des bibliothèques de grande taille ($\sim 10^{12}$ variants). S'inspirant des interactions anticorps-antigènes, Schilling *et al* ont conçu un nouveau type de bibliothèque de DARPins appelée LoopDARPins. Cette nouvelle bibliothèque repose sur l'insertion d'une boucle de 19 acides aminés dans le coude β du deuxième module interne d'une DARPIn à 3 modules internes (Schilling, J. *et al.*, 2014). Cette boucle possède 10 positions mutables, et est spatialement entourée par les coudes β des deux modules adjacents, constituant ainsi une surface continue potentielle d'interaction. Dans cette bibliothèque, les modules extérieurs (en N-terminal et C-terminal) ont également été mutés aléatoirement afin d'étendre la surface potentielle d'interaction. L'ensemble de ces bibliothèques ont donné des interacteurs d'une multitude de cibles différentes comme des kinases, la tubuline, des caspases, HER2, ou la protéine gp120 du VIH (Boersma, Y. L., 2018). Les DARPins ont été utilisées comme chaperonnes de cristallisation (Batyuk, A. *et al.*, 2016; Seeger, M. A. *et al.*, 2013), biosenseurs (Kummer, L. *et al.*, 2013; Schütz, M. *et al.*, 2016), outil diagnostic (Gu, G. J. *et al.*, 2013; Theurillat, J.-P. *et al.*, 2010; Wolff, A. C. *et al.*, 2013), et même comme agent thérapeutique (Boersma, Y. L. *et al.*, 2011; Egel, A. *et al.*, 2014; Jost, C. *et al.*, 2013; Kim, B. *et al.*, 2012; Münch, R. C. *et al.*, 2015; Plückthun, A., 2015). Plusieurs DARPins sont en essais cliniques et d'autres suivront probablement puisqu'une entreprise pharmaceutique, Molecular Partners, a été créée en 2004 dans le but de commercialiser les DARPins comme agents thérapeutiques.

L'équipe MIP a également fait partie des premiers groupes à développer des bibliothèques de protéines basées sur une ossature alternative aux anticorps. En effet, une première bibliothèque de protéines basée sur la néocarzinostatine (NCS) a été développée au début des années 2000 pour reconnaître les petites molécules (Heyd, B. *et al.*, 2003). Plus tard, l'équipe a développé une bibliothèque basée sur une protéine à motifs répétés, car cela permettait de générer plus facilement

une grande surface hypervariable d'interaction sans déstabiliser l'ossature. Ces bibliothèques ont pour but de fournir des répertoires de protéines possédant des régions aléatoirement mutées, à partir desquels on pourra théoriquement sélectionner des interacteurs contre n'importe quelle cible protéique.

Les α Reps sont des protéines artificielles à motifs répétés constituées par la répétition en tandem d'un motif HEAT. Ce motif HEAT est constitué de 2 hélices α antiparallèles reliées par une boucle. La répétition en tandem de ce motif confère aux α Reps une forme solénoïdale, dont la face concave a été mutée aléatoirement afin de constituer la bibliothèque (Figure 2). Dans la nature, les protéines HEAT *repeat* constituent une famille de protéines basée sur la répétition de ce motif. La protéine Mth187, qui provient de *Methanobacterium thermoautotrophicum* est l'un des membres de cette famille. Sa structure, résolue en 2006, a révélé la présence d'un domaine N terminal non structuré, suivi d'un domaine C terminal constitué par la répétition de 3 motifs HEAT de 30-31 acides aminés chacun (Julien, O. et al., 2006). C'est cette protéine, et plus particulièrement son domaine C-terminal, qui a servi à déterminer la séquence consensus utilisée pour la création de la bibliothèque α Reps. Pour cela, un premier BLAST (*Basic Local Alignment Search Tool*) a été réalisé contre les 2 premiers motifs HEAT de Mth187, afin d'analyser la conservation des résidus entre les motifs HEAT des protéines les plus proches de Mth187 dans la PDB (*Protein Data Bank*). Le troisième motif HEAT a été exclu du BLAST car il possédait des propriétés particulières probablement liées à son exposition au solvant. Dans le but d'élargir le nombre de séquences de motifs auto-compatibles ayant adopté ce type de structure, une seconde recherche a alors été réalisée en prenant comme entrée la séquence consensus obtenue pour un motif HEAT à la sortie du premier BLAST, répétée 5 fois. Ce second BLAST a donné une seconde séquence consensus à partir des 100 protéines les plus proches et comptabilisant 500 motifs différents (Figure 3).

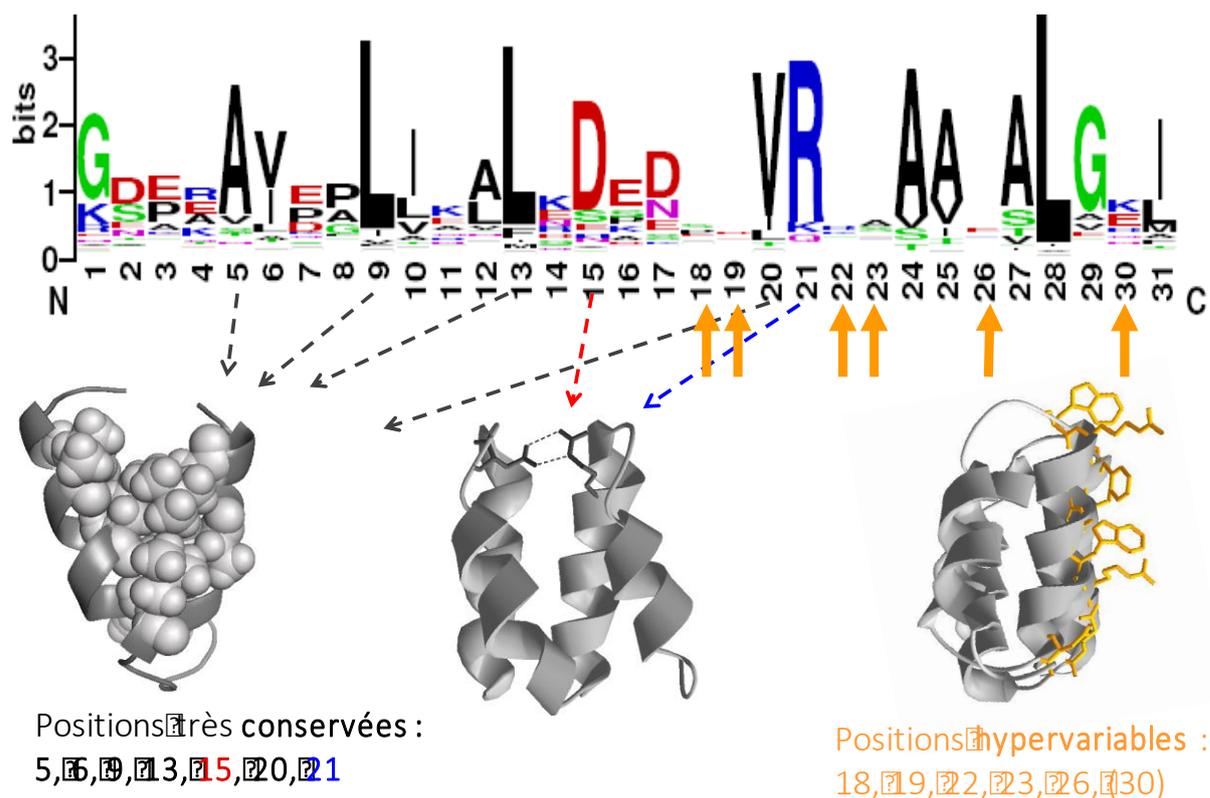


Figure 3 : Conception du motif HEAT de la bibliothèque α Reps

Un premier BLAST a été réalisé contre les 2 premiers motifs HEAT du domaine C terminal de la protéine Mth187. La séquence consensus obtenue a été répétée 5 fois en vue d'être utilisée lors d'un second BLAST qui a donné la séquence d'acides aminés représentée ci-dessus. La taille des lettres correspond à la conservation des résidus entre les 500 motifs différents alignés. Les flèches orange indiquent les 6 positions hypervariables identifiées grâce à ce BLAST qui seront les positions mutées aléatoirement dans la bibliothèque d' α Reps. (Urvoas, A. *et al.*, 2010)

Dans cette séquence consensus, les résidus sont représentés par leur code à une lettre et la taille des lettres est proportionnelle à leur niveau de conservation entre les différentes séquences alignées. Une fois cartographiée de la sorte, la séquence peut être subdivisée en 3 groupes de résidus conservés. Un premier groupe, constitué de résidus à chaînes latérales apolaires (A5, L9, L13, V20, A24, A25, A27 et L28), contribue à la compacité de la structure en réalisant des contacts hydrophobes entre les 2 hélices d'un même motif HEAT, ainsi qu'entre 2 motifs consécutifs. Un deuxième groupe est composé des résidus R21 et D15. En effet, l'arginine R21 d'un motif est impliquée dans un pont salin avec l'aspartate D15 du motif suivant, ce qui contribue à la stabilité de la structure. Enfin, un troisième groupe est constitué par les résidus G1 et G29, G1 étant localisés dans une boucle reliant 2 motifs consécutifs et G29 permettant une compacité étroite entre 2 motifs consécutifs. Les positions restantes sont plus variables et correspondent principalement aux positions accessibles au solvant sur les 2 hélices d'un motif. Les travaux de Julien, O. *et al.*, 2006 avaient déjà montré une asymétrie entre les positions variables de chaque hélice d'un motif. Les positions variables de la première hélice (positions 3, 4, 7,

et 11), localisées sur sa face externe, sont occupées par une forte proportion de résidus polaires ou stabilisateurs d'hélices α (K, R, E, Q, et A). Les positions les plus variables (18, 19, 22, 23, et 26) sont localisées sur la surface externe de la deuxième hélice du motif et sont occupées par une plus grande diversité de résidus, notamment une proportion non négligeable de résidus aromatiques. Cela suggère que ces dernières positions pourraient donc être utilisées pour prendre part à des processus de reconnaissance moléculaire. Ce sont ces positions (indiquées par les flèches orange [Figure 3](#)) qui ont été diversifiées dans la bibliothèque d' α Reps afin de générer une surface hypervariable sur leur face concave ([Figure 4](#)). Le premier et le dernier motif d'une α Rep, que l'on appellera N-cap et C-cap, diffèrent des motifs internes. En effet, leur exposition au solvant est incompatible avec les contacts hydrophobes qui stabilisent les motifs internes entre eux, car ceux-ci pourraient entraîner l'aggrégation des α Reps entre elles. La séquence du C-cap a été basée sur celle naturellement présente dans Mth187. En revanche, il était nécessaire de trouver une autre protéine de la même famille afin d'identifier une séquence qui pourrait servir de modèle pour le N-cap. Dans cette optique, une recherche a été menée dans la PDB afin de trouver une protéine à motif HEAT répété dont la partie N-terminale soit proche de la séquence consensus obtenue [Figure 3](#). La protéine retenue (identifiant UniProtKB : A0B7C6) possède une extension N-terminale de 34 acides aminés dont la séquence est assez proche du consensus obtenu pour un motif HEAT :

LXDXXXVRXXAAXALGXI

La longueur de la séquence (34 acides aminés) et sa composition suggèrent un repliement en 2 hélices similaire à celui des motifs HEAT internes. Pour cette raison, les positions de ce N-cap équivalentes aux positions hypervariables des motifs internes ont été aléatoirement mutées afin d'allonger la surface hypervariable des protéines de la bibliothèque.

Dans la bibliothèque d' α Reps, le nombre de motifs HEAT internes (c'est-à-dire, compris entre le N-cap et le C-cap) est variable. De façon générale, ce nombre est de 2.3 ± 1.9 ([Guellouz, A. et al., 2013](#)), bien qu'on puisse trouver des α Reps à plus de 10 motifs internes, de même que des α Reps à 0 motifs interne.

Le protocole d'assemblage de la bibliothèque d' α Reps ne sera pas détaillé ici car bien que j'aie utilisé de nombreuses techniques mises au point pour ce protocole, il a fallu l'adapter pour construire la bibliothèque de bi-pentahélines. Le protocole que j'ai utilisé sera, lui, détaillé ultérieurement.

La [Figure 4](#) représente une α Rep à 4 motifs HEAT internes (orange) avec son N-cap (vert) et son C-cap (-rouge). Les positions variables du N-cap sont également représentées en orange.

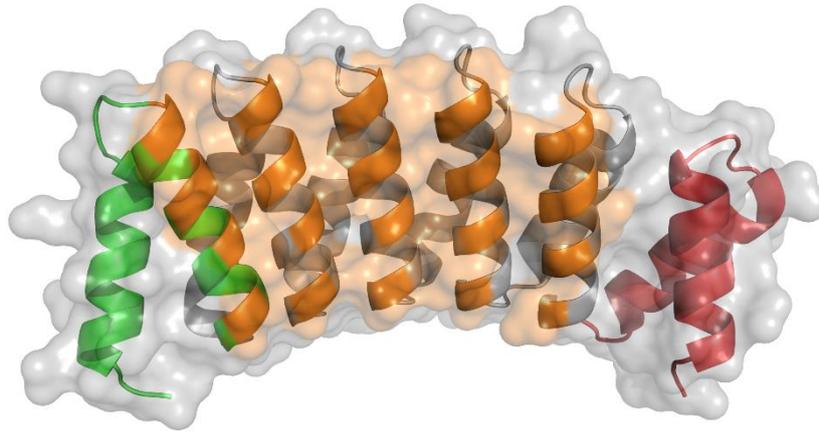


Figure 4 : Structure cristallographique d'une α Rep

Le N-cap est représenté en vert, le C-cap en rouge. Les positions variables sont représentées en orange.

Des interacteurs ont été obtenus par phage *display* à partir de la bibliothèque d' α Reps pour plus d'une quarantaine de cibles non apparentées, et 9 structures de complexes α Rep-cible ont déjà été obtenues (Figure 5), confirmant ainsi la robustesse de cette méthode. Le phage *display* étant également utilisé lors de la création de la bibliothèque de bi-pentahélines, son principe sera détaillé ultérieurement.

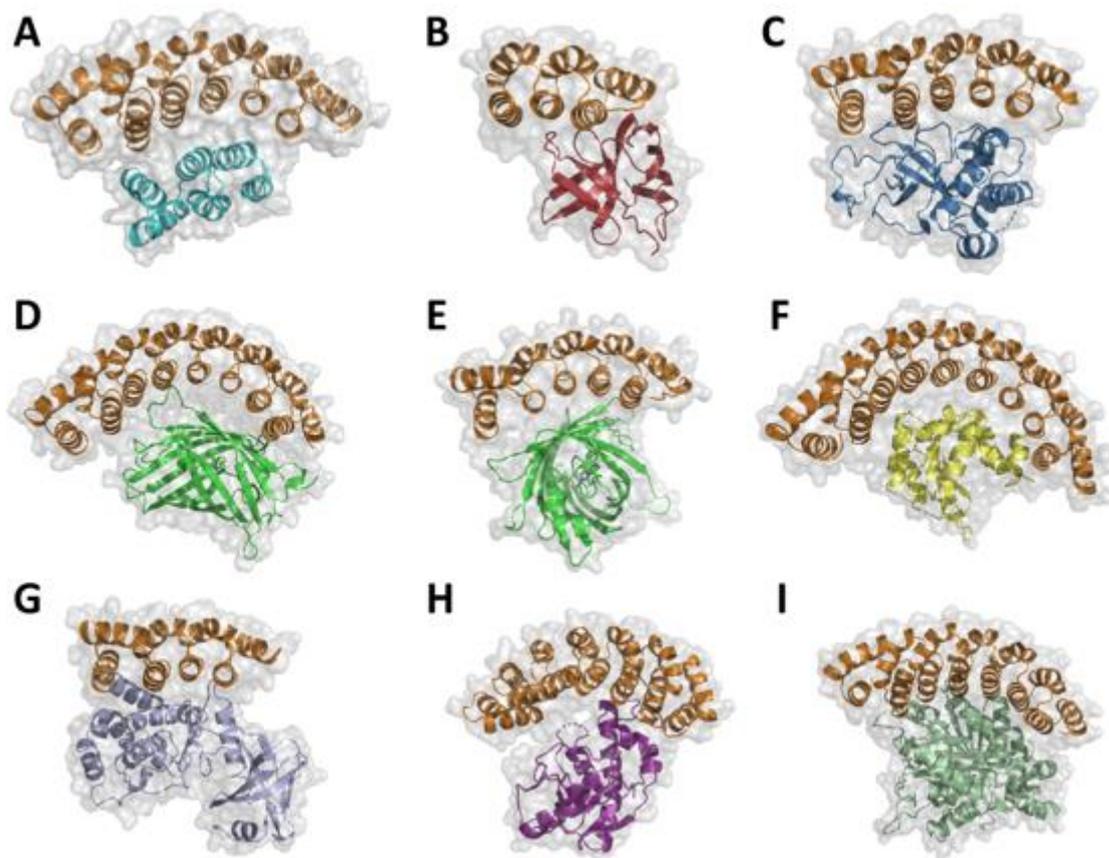


Figure 5 : Structures des différents complexes α Rep obtenus par cristallographie aux rayons X (collaborations H. van Tilbeurgh et B. Gigant, I2BC)

Les différentes structures de complexes sont représentées en *cartoon*, avec l' α Rep représentée en orange et sa cible. Les différentes cibles sont : A) bA3-2. B) NCS. (Guellouz, A. *et al.*, 2013) C) FNE. (Tiouajni, M. *et al.*, 2014) D et E) GFP. (Chevrel, A. *et al.*, 2015) F) Octarelline. (Figuroa, M. *et al.*, 2016) G) YabT. (Chevrel, A. *et al.*, 2018) H) HP1473. (Non publié) I) Tubuline. (Campanacci, V. *et al.*, 2019)

Si les bibliothèques de protéines artificielles et l'ingénierie des anticorps ont connu un tel essor à partir des années 2000, c'est parce-que le développement des techniques de biologie moléculaire et les connaissances en immunologie ont permis de réaliser des objets de plus en plus complexes qui peuvent répondre à des problématiques plus poussées. Ainsi, l'ingénierie modulaire a vu naître des objets moléculaires tels que les anticorps bi-spécifiques (bsAbs), les lymphocytes T à récepteur d'antigène chimérique (CAR-T cells), les engageurs de cellules-T bi-spécifiques (BiTEs), etc... Ces molécules sont notamment très utilisées pour répondre à des besoins dans le domaine de la cancérologie.

II. Ingénierie modulaire

Au moins un tiers des protéines procaryotes et plus de deux tiers des protéines eucaryotes sont multidomaines (Ekman, D. *et al.*, 2005). L'une des explications possibles à cette proportion de protéines multidomaines pourrait résider dans les processus souvent adoptés par l'évolution naturelle pour l'émergence de nouvelles fonctions comme la combinaison de plusieurs domaines protéiques existants. En effet, par des phénomènes de duplication, de spécialisation et de fusion de leur gène, les domaines protéiques déjà existants sont associés à d'autres domaines pour donner naissance à une nouvelle protéine (Magadum, S. *et al.*, 2013).

Les chercheurs se sont inspirés de ces processus naturels pour générer de nouvelles molécules en mettant au point une ingénierie multimodulaire dont les anticorps sont l'archétype. En effet, les anticorps sont d'excellents candidats thérapeutiques, du fait de leur faible immunogénicité et de leur capacité de fixation très spécifique, ce qui fait de leur ingénierie un réel enjeu. Cet intérêt pour les anticorps se matérialise d'ailleurs par de gros investissements de l'industrie pharmaceutique (Morrison, C., 2020; Morrison, C. & Lähteenmäki, R., 2017), ce qui a grandement motivé les recherches sur ce type de molécules. Beaucoup des efforts ont notamment été concentrés sur la multi-spécificité des objets créés, afin de dépasser la limite des anticorps que l'on trouve dans la nature et qui possèdent le même site de fixation sur les deux chaînes polypeptidiques.

La première évocation d'un anticorps possédant 2 sites de fixations à des antigènes différents remonte à 1960 (Nisonoff, A. *et al.*, 1960), et fait écho aux premières avancées dans la compréhension structurale des anticorps (Riethmüller, G., 2012). Pour coupler des fragments d'anticorps (Fabs) de lapins de spécificités différentes, des chercheurs ont mis au point une technique de ré-oxydation douce. Ils ont ainsi démontré que ces Fabs bispécifiques provoquaient l'agglutination de deux types de cellules différentes (Fudenberg, H. H. *et al.*, 1964). Aux innovations dans la conception d'anticorps bispécifiques (bsAbs) ont suivi tout un pan de l'ingénierie et de la biologie des anticorps, qui a permis d'aboutir à une vaste collection de plus de 100 formats de bsAbs disponibles aujourd'hui (Figure 6). Près d'un quart de ces formats ont été développés comme plateforme technologique et sont commercialisés par des industries pharmaceutiques ou biotechnologiques comme nouveaux agents thérapeutiques (Godar, M. *et al.*, 2018). Cependant, bien que ces nouveaux agents thérapeutiques soient très prometteurs (Chames, P. & Wurch, T., 2019), on observe un temps très long entre la découverte du concept et son approbation par les régulateurs du médicament. Par exemple, l'application type des bsAbs, la redirection et l'engagement des lymphocytes T, a été décrite pour la première fois en 1985 (Perez, P. *et al.*, 1985; Staerz, U. D. *et al.*, 1985). Or ce n'est qu'en 2009 que l'Union Européenne a approuvé le *Catumaxomab* pour le traitement des ascites malignes intrapéritonéales. On note toutefois un intérêt certain pour ce type de format d'anticorps puisqu'en

Juin 2019, 43 bsAbs « redirecteurs » de lymphocytes T étaient en développement clinique pour le traitement des tumeurs solides ou hématologiques.

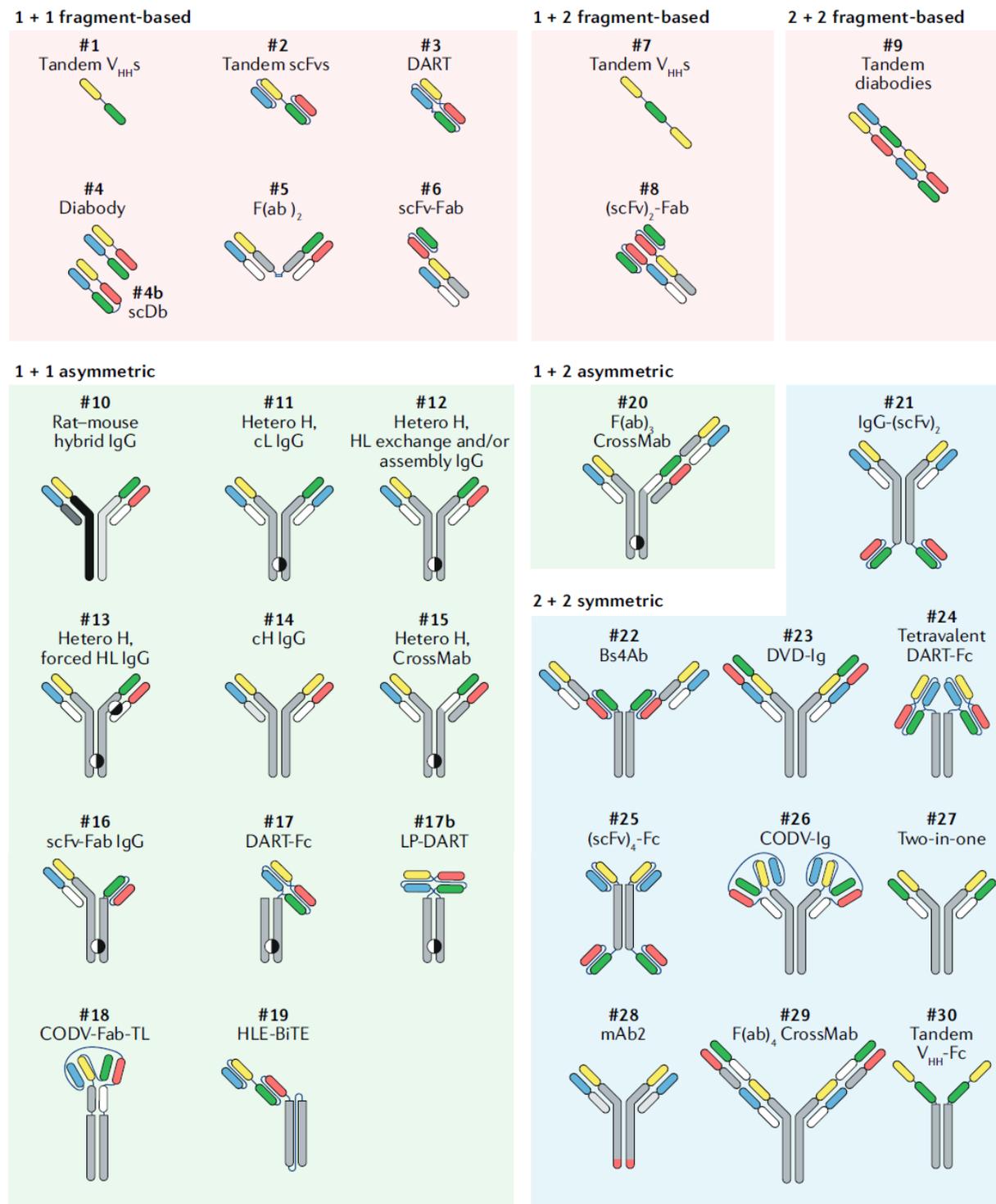


Figure 6 : Sélection de formats d'anticorps bispécifiques (bsAbs)

Les anticorps bispécifiques sont classés par valence (nombre de site de fixation) croissante pour chaque antigène reconnu de gauche à droite et par classe de format. Les formats d'anticorps sont classés ici en molécules basées sur des fragments d'anticorps (*fragment-based*) (sans domaine Fc) et en molécules comportant un domaine Fc asymétrique ou symétrique. Les

molécules peuvent contenir des mutations qui affectent l'appariement des chaînes et d'autres paramètres importants pour leur synthèse, les fonctions effectrices médiées par les domaines Fc et leur demi-vie.

Exemples de dérivés d'anticorps bi/multispécifiques : #1 et #7, nanobody ; #2, *bispecific T cell engager* (BiTE); #3, #17(b) et #24, *dual-affinity re-targeting* (DART); #6 et #8, *immunotherapy antibody* (ITab); #9, *tandem diabody* (TandAb); #10, *Triomab*; #11, *asymmetric reengineering technology immunoglobulin* (ART-Ig) et *biclonics*; #12, *DuoBody* et *knobs into holes* (KiH); #13, *DuetMab*; #14, *κλ body* ; #16, *Xmab*, *YBODY* et *bispecific engagement by antibodies based on the T cell receptor* (BEAT); #15, #20 et #29, *CrossMab*; #18 et #26, *crossover dual variable* (CODV); #23, *dual variable domain* (DVD); #25, *Adaptir*; et #27, *two-in-one and dual-action antigen-binding fragment* (Fab) (DAF). Les domaines des anticorps sont colorés par rapport à leur architecture : orange, variable lourde (H) spécifique pour 1; vert, variable H spécifique pour 2; bleu, variable légère (L) spécifique pour 1; rouge, variable L spécifique pour 2; gris, région constante de la chaîne H; blanc, région constante de la chaîne L; gris clair, région constante alternative de la chaîne L; format #10: gris foncé et noir, chaîne L du rat et chaîne H de l'immunoglobuline G2b (IgG2b); blanc et gris clair, chaîne L de la souris et chaîne H d'une IgG2a.cH lourde commune ; cL, légère commune; HLE, demi-vie prolongée; scDb, *single-chain diabody* ; scFv, *single-chain variable fragment*; VHH, *heavy chain-only variable domain*. (Labrijn, A. F. et al., 2019)

Le mode d'action des bsAbs consiste le plus souvent à rapprocher spatialement 2 types cellulaires différents. On peut par exemple concevoir un anticorps interagissant d'une part avec une cellule tumorale et d'autre part avec un lymphocyte T pour guider ce dernier vers la cellule tumorale à éliminer. Le lymphocyte T est alors activé le plus souvent grâce à une interaction au niveau de CD3ε dans le complexe TCR (*T-cell receptor*), ce qui permet de s'affranchir du complexe majeur d'histocompatibilité et d'activer le lymphocyte T peu importe l'épitope reconnu par le TCR. Ces anticorps sont appelés engageurs de cellules T bispécifiques (bsTCE ou BiTEs), et sont l'objet de nombreux développements cliniques. Dans un premier temps, les bsTCEs semblaient prometteurs mais les essais cliniques se sont révélés décevants (De Gast, G. C. et al., 1995; Tibben, J. G. et al., 1996). Les bsTCEs entraînaient en effet un relargage de cytokines médié par les lymphocytes T, même à faibles doses, en plus d'être difficiles à produire et de ne pas montrer d'effet cliniques durables. Ces problèmes ont freiné l'intérêt de la communauté scientifique pour ce format, et ce n'est qu'en 2008, lorsque le blinatumomab (un bsTCE CD19 x CD3) a présenté de bons résultats cliniques (Bargou, R. et al., 2008), que l'intérêt pour les bsTCEs a été relancé. Le blinatumomab est un petit bsAb d'environ 55 kDa basé sur un fragment d'anticorps qui ne possède pas de domaine Fc et il a été approuvé par la FDA (*Food and Drugs Administration*) en 2014 pour traiter les lymphomes non hodgkiniens.

L'anticorps monoclonal murin OKT3, utilisé comme immunosuppresseur dans les greffes, a présenté de lourds problèmes de relargage de cytokines et d'activation des lymphocytes T *in vivo* (Chatenoud, L. et al., 1990). Cette expérience a permis aux immunologistes d'établir des règles permettant la création de bsTCEs thérapeutiques. En effet, il est apparu que le mécanisme principal entraînant le relargage de cytokines était le regroupement de récepteurs CD3 sur les lymphocytes T via leur

interaction avec les récepteurs de Fc d'autres cellules immunitaires. Une seconde génération de d'anticorps anti-CD3 humanisés a donc vu le jour, avec une liaison aux FcγR (récepteurs aux domaines Fc des IgGs) réduite et ces nouveaux anticorps ont montré une réduction significative du relargage de cytokines (Xu, D. *et al.*, 2000). De façon similaire, les lourds effets indésirables du catumaxomab (un bsTCE CD3 x EpCAM) ont été attribués à la fixation aspécifique de son domaine Fc sur les récepteurs FcγR des cellules de Kupffer présentes dans le foie. Ces données ont suggéré qu'il était préférable pour les bsTCEs fixant le récepteur CD3 de supprimer les fonctions effectrices liées au domaine Fc afin de minimiser la toxicité du produit et d'optimiser son effet thérapeutique. De nouveaux bsTCEs ont alors pu être conçus, préférentiellement avec une bivalence pour l'antigène spécifique des tumeurs et une monovalence pour CD3 afin d'optimiser l'effet thérapeutique par rapport à la toxicité (Bacac, M. *et al.*, 2016).

Une autre utilisation possible des bsAbs est de créer des interacteurs biparatopiques qui interagissent simultanément avec 2 épitopes non chevauchants d'une même cible. Cela permet de créer de l'avidité, et d'avoir par exemple un inhibiteur plus efficace, comme le ZW25 qui inhibe la signalisation de HER2 et qui semble efficace dans le traitement des cancers HER2-dépendants (Weisser, N. *et al.*, 2017).

On peut également noter le développement de plus en plus important des anticorps multispécifiques, possédant une spécificité pour plus de 2 antigènes. Ces anticorps peuvent être générés en concevant simplement plus de fragments sur une architecture bsAbs basée sur des fragments ou asymétrique classique (Khan, S. N. *et al.*, 2018). Pour les architectures symétriques, il est possible de réaliser une hétérodimérisation asymétrique de la chaîne H ou alors de forcer un appariement HL permettant d'obtenir de nouvelles spécificités (Klein, C. *et al.*, 2016). Enfin, par de l'ingénierie plus « fine », des chercheurs ont réussi à favoriser les associations de chaînes HL d'intérêt au détriment des mésappariements, en mutant les résidus à l'interface CH1 – CL pour défavoriser les interactions entre les domaines wild type (Lewis, S. M. *et al.*, 2014; Wu, X. *et al.*, 2018). En effet, lorsque les chercheurs créent des anticorps bi ou multi spécifiques, ils sont confrontés au problème dit de l'association des chaînes. Lorsque l'on souhaite créer un bsAbs, on fusionne généralement deux cellules exprimant chacune un anticorps classique portant une spécificité pour un antigène au niveau de son déterminant, lequel est composé du déterminant de la chaîne lourde H et de la chaîne légère L. L'objectif est d'obtenir une molécule possédant sur chaque bras un déterminant composé de l'association HL de chaque anticorps monospécifique. Or, cette molécule ne représente qu'une possibilité sur les 10 associations H2L2 possibles lorsqu'on exprime dans une même cellule les chaînes lourdes et variables de 2 anticorps différents. Il est donc nécessaire de réaliser des étapes contraignantes de purification afin de trier uniquement les molécules d'intérêt, ce qui peut s'avérer compliqué et ce qui rend la production plus coûteuse.

Les CAR (*Chimeric Antigen Receptor*) T-cells sont un autre exemple d'ingénierie des anticorps, développé principalement dans les années 2000 et prometteur sur le plan thérapeutique. Le principe consiste à modifier des lymphocytes T en leur faisant exprimer un récepteur artificiel conçu en laboratoire, lequel est d'une part affiné pour l'antigène souhaité et d'autre part possède un domaine intracellulaire d'activation du lymphocyte T pour le rendre cytotoxique. Structurellement, le récepteur chimérique est composé de 3 parties distinctes : l'ectodomaine, le domaine transmembranaire, et l'endodomaine (Becker, M. L. B. *et al.*, 1989). L'ectodomaine est la partie extracellulaire du récepteur et celle responsable de la fixation de l'antigène. Ce domaine est le plus souvent basé sur une architecture de type fragments d'anticorps (scFvs par exemple). L'endodomaine est la partie intracellulaire du récepteur et porte sa fonction effectrice, à savoir l'activation du lymphocyte T, généralement via la chaîne CD3 ζ contenant 3 motifs ITAMs (*Immunoreceptor Tyrosine-Based Activation*). L'ingénierie des CAR T-cells a vu des évolutions à la fois au niveau de l'ectodomaine et de l'endodomaine (Figure 7).

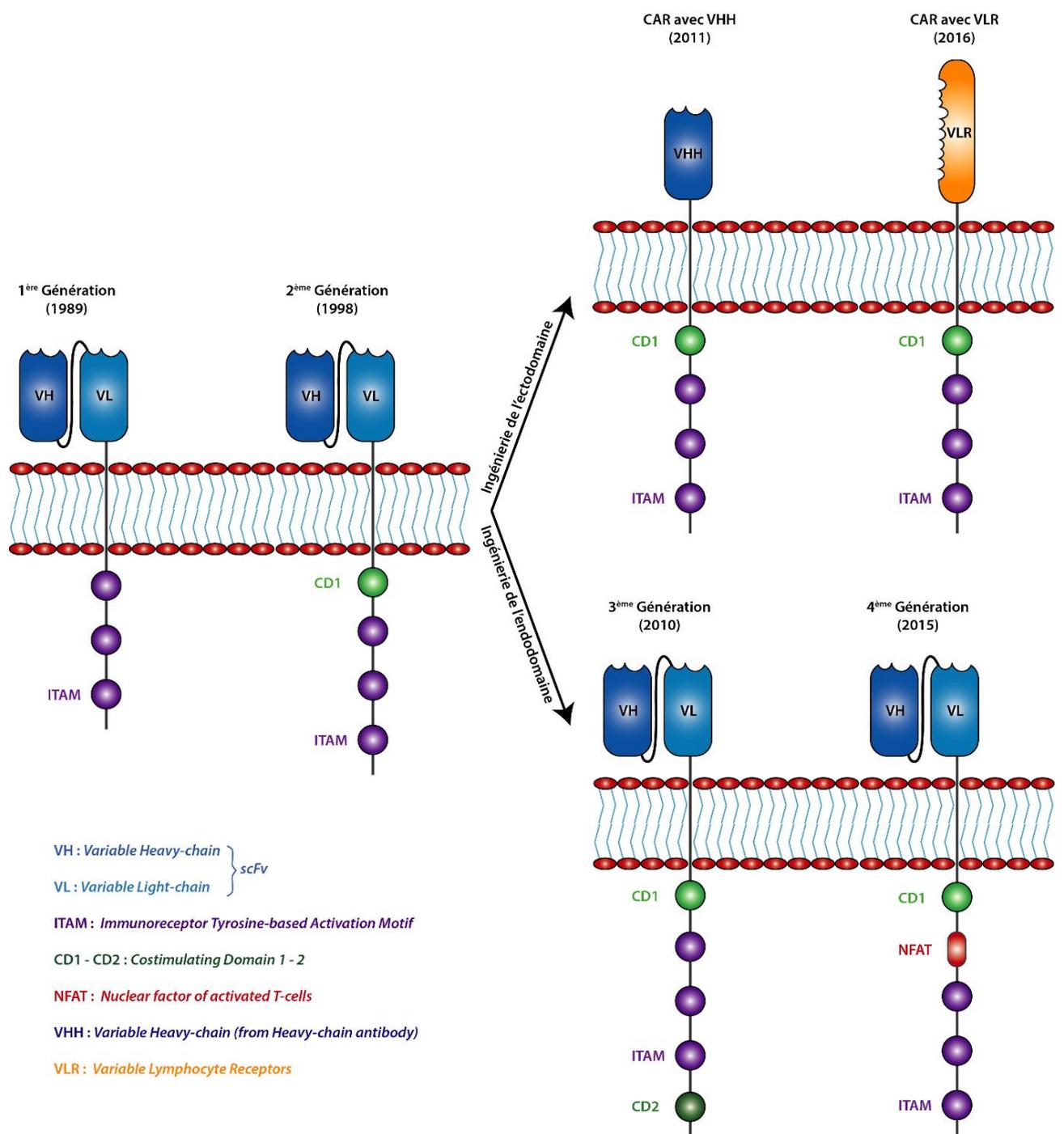


Figure 7 : : Évolution des récepteurs antigéniques chimériques (CAR) au cours des dernières années.

La 1^{ère} génération de CAR était constituée d'un ectodomaine (scFv + peptide signal), d'un domaine transmembranaire et d'un endodomaine (CD3 ζ avec trois ITAMs). La seconde génération de CAR a été générée par ajout d'un domaine de costimulation (CD1), souvent CD28. La 3^{ème} génération a été créée par ajout d'un second domaine de costimulation (CD2), généralement CD134 ou CD137. Enfin la 4^{ème} génération ne possède qu'un domaine de costimulation mais une séquence NFAT permettant la production d'interleukine 12 à la suite de l'activation du récepteur. Au niveau de l'ectodomaine, le scFv a été remplacé soit par un VHH soit par un VLR (Thèse de Léger, C.).

Au niveau de l'endodomaine, les efforts d'ingénierie se sont concentrés sur l'amélioration de l'activation des lymphocytes T, c'est-à-dire leur cytotoxicité et leur prolifération, ainsi que sur leur

durée de vie. Dans cette optique, la deuxième génération de CAR T-cells a été créée par ajout d'un premier domaine de co-stimulation (CD28, CD137 ou CD 134) en amont des ITAMs (Finney, H. M. *et al.*, 1998). La troisième génération de CAR T-cells a été obtenue en combinant deux domaines de co-stimulations en tandem (CD28 + CD137 ou CD28 + CD134), ce qui améliore significativement la cytotoxicité de ces CAR T-cells (Marin, V. *et al.*, 2010). Enfin, la quatrième génération de CAR T-cells fait intervenir une séquence dite NFAT (*Nuclear Factor of the Activated T-cell*) qui, une fois le récepteur activé, va induire la production de cytokines comme l'interleukine 12 (IL-12) (Chmielewski, M. *et al.*, 2011). L'IL-12 ainsi produite va servir à recruter d'autres cellules de l'immunité au contact de la tumeur, permettant de tuer des cellules cancéreuses qui ne seraient pas reconnues par la CAR T-cell. Ce domaine NFAT, en combinaison avec un unique domaine de co-stimulation, a grandement amélioré l'efficacité de cette génération de CAR T-cells, que l'on désigne généralement par TRUCK (*T-cell Redirected for Universal Cytokine-mediated Killing*).

L'ectodomaine a lui aussi été la cible de diverses améliorations, afin de dépasser les limites apportées par l'utilisation des scFvs. En 2011, les scFvs ont été remplacés par des VHH sur les CAR T-cells de deuxième génération (Iri-Sofla, F. J. *et al.*, 2011), et en 2016, ce sont les VLRs qui ont été préférés aux autres types d'architectures pour remplir la fonction de reconnaissance, encore une fois sur des CAR T-cells de deuxième génération (Moot, R. *et al.*, 2016).

Les anticorps multi-spécifiques et les CAR T-cells sont des exemples de réussite d'ingénierie modulaire, par leur utilisation dans le monde biomédical et par le nombre d'études qu'ils suscitent. Cependant, ces technologies basées sur les dérivés d'anticorps rencontrent les problèmes inhérents à ce type d'architecture et déjà discutés. Pour contrer ce type de limites, l'utilisation des architectures alternatives aux anticorps a également fait l'objet d'une ingénierie multi-modulaire de différentes architectures protéiques. Afin d'illustrer ce type de recherches, je me limiterai ici à la description de l'ingénierie multi-domaines basée sur les DARPins, représentant à elle seule l'étendue des développements possibles avec les protéines artificielles.

Les DARPins sont un bon exemple d'ingénierie modulaire sur des architectures alternatives aux anticorps. En effet, des DARPins ont été sélectionnées contre la chaîne α du récepteur aux immunoglobulines E humaines (IgE) (Fc ϵ R1 α) (Eggel, A. *et al.*, 2009). Si les DARPins monovalentes pour le site de fixation des IgE n'ont pas suffi à inhiber leur fixation, une fusion de deux DARPins spécifiques d'épitopes distincts sur Fc ϵ R1 α a permis une inhibition efficace de la fixation des IgE, allant de pair avec une inhibition du relargage de facteurs pro-inflammatoires. De façon similaire, Boersma *et al.* ont également réalisé une fusion de DARPins spécifiques pour des épitopes différents sur le récepteur aux facteurs épidermiques de croissance (EGFR). Le produit de cette fusion est plus efficace pour inhiber la prolifération cellulaire que les DARPins monovalentes seules, et possède surtout un effet beaucoup plus long grâce à l'avidité provoquée par sa multivalence (Boersma, Y. L. *et al.*, 2011). Le récepteur aux

facteurs épidermiques de croissance 2 (HER2), surexprimé dans de nombreux cancers, a lui aussi été la cible des DARPins multi-domaines. Plusieurs fusions de DARPins spécifiques d'épitopes différents ont été testées (Jost, C. *et al.*, 2013). Jost *et al.* ont alors déterminé que les fusions les plus efficaces pour lutter contre la prolifération cellulaire étaient les constructions faisant intervenir des interacteurs des domaines I et IV du récepteur HER2. Ces fusions possèdent un effet cytotoxique en plus d'un effet cytostatique, dus à par une déformation de l'ectodomaine de deux molécules de HER2 connectées via la fusion de DARPins, rendant incompatible l'association et la trans-activation des HER2. Une autre étude a montré que cette interaction croisée entre deux molécules de HER2 induit la déphosphorylation des récepteurs HER2 et HER3 (Tamaskovic, R. *et al.*, 2016), et que l'une des fusions de DARPins provoque l'apoptose des cellules reconnues en empêchant la formation d'homodimères de HER2 et d'hétérodimères HER2-EGFR et HER2-HER3. Une autre illustration possible d'une utilisation modulaire des DARPins est leur utilisation comme plateforme de construction. Il est possible de réaliser un couplage orthogonal sur les DARPins pour y greffer une toxine d'une part, et un composant augmentant le temps de demie-vie de l'assemblage d'autre part. En effet, l'absence de cystéines permet l'introduction d'une cystéine unique en C-ter fonctionnalisable par la chimie des thiols. Il est également possible de remplacer la méthionine en N-ter par une azidohomoalanine, afin de réaliser de la chimie click (Simon, M. *et al.*, 2012, 2013; Stefan, N. *et al.*, 2014). Enfin, Les DARPins ont également été utilisées comme protéines de ciblage pour adresser des particules virales à certains types cellulaires dans le cadre de la thérapie génique (Dreier, B. *et al.*, 2013). Pour cela, deux DARPins ont été fusionnées, l'une chargée de reconnaître l'adénovirus de sérotype 5 (Ad5), et l'autre chargée d'adresser cet adénovirus à des cellules cancéreuses. La DARPins reconnaissant Ad5 a été multimérisée afin de diminuer au maximum sa dissociation grâce à l'effet d'avidité ainsi créé. Ce multimère a été fusionné à une DARPins spécifique de HER2, et le complexe final (avec Ad5) a efficacement fixé sa cible dont il ne s'est pas décroché pendant 10 jours, démontrant ainsi la robustesse de l'interaction. De plus, aucune fixation aspécifique n'a été démontrée, ni *ex vivo*, ni *in vivo* (Münch, R. C. *et al.*, 2015).

Ces exemples montrent bien l'intérêt de fusionner des protéines pour obtenir des objets dont la fonction dépasse la simple addition des effets de chacun des domaines qui les constituent. En effet, la présence d'un lien intramoléculaire peut générer de l'avidité, ou rapprocher spatialement deux objets qui ne l'auraient pas été par la simple addition de deux interacteurs spécifiques de chacun de ces objets. De plus, l'évolution des méthodes de biologie moléculaire et le développement des méthodes de *screening* à haut débit permettent de tester beaucoup de combinaisons de protéines pour trouver la combinaison la plus adaptée à une problématique donnée.

Si la multi-spécificité proférée par l'ingénierie modulaire a été très utilisée en cancérologie, d'autres domaines bénéficient de ces avancées technologiques. Par exemple, dans le domaine de la purification cellulaire, un interacteur multi-spécifique peut permettre de séparer plus finement des populations

cellulaires qui ont des récepteurs en communs, en les discriminants sur un panel plus large de récepteurs.

III. La purification cellulaire

La purification cellulaire est un outil puissant en biologie médicale, qu'il s'agisse de purifier un type cellulaire en vue de réaliser un test diagnostique, pour réaliser de la thérapie cellulaire, ou comme étape préliminaire en vue de réaliser d'autres expériences (Orfao, A. & Ruiz-Argüelles, A., 1996). Ces méthodes apparaissent particulièrement importantes de nos jours, avec l'avènement des thérapies cellulaires personnalisées. Cet intérêt se manifeste par une utilisation croissante des méthodes de purification cellulaire dans l'ingénierie tissulaire, la médecine régénérative (Guo, K.-T. *et al.*, 2006), les thérapies anticancéreuses, et la recherche sur les pathologies associées aux maladies infectieuses (Terry, V. H. *et al.*, 2009). D'un point de vue clinique, ces techniques ont significativement contribué au diagnostic et au traitement de nombreuses maladies (Jamieson, C. H. M. *et al.*, 2004; Majeti, R. *et al.*, 2009; Will, B. & Steidl, U., 2010), à la transfusion de globules rouges, de plaquettes et de neutrophiles, et à la transplantation de populations spécifiques de cellules hématopoïétiques et immunes (Lanzoni, G. *et al.*, 2009). Ces multiples applications ont conduit au développement d'un large panel de techniques permettant la purification de cellules. On peut classiquement identifier quatre grands critères qui permettent de trier les cellules les unes des autres : leur adhérence, leur densité et leur taille, leurs caractéristiques morphologiques, et leurs propriétés immunophénotypiques (Almeida, M. *et al.*, 2015).

Ainsi, on peut distinguer d'une part les méthodes de séparation basées sur les propriétés physiques des cellules, et d'autre part les méthodes basées sur la reconnaissance moléculaire des marqueurs spécifiques d'une population cellulaire d'intérêt.

A. Les méthodes basées sur les propriétés physiques

Les méthodes de séparation basées sur les propriétés physiques des cellules permettent de trier de gros volumes de cellules et sont généralement très efficaces pour réaliser une première étape de purification.

Une des façons de trier les cellules est de se baser sur leurs propriétés d'adhésion. Cela consiste le plus souvent à récupérer un extrait de tissu primaire digéré ou explanté. L'extrait ainsi obtenu peut être filtré pour obtenir une suspension cellulaire qui sera mise en contact d'une surface plastique par exemple (Tomlinson, M. J. *et al.*, 2013). La vitesse et la force d'adhésion des différents types cellulaires peuvent alors permettre d'éliminer certaines cellules de l'échantillon. Cette méthode est simple d'utilisation et peu coûteuse, mais n'est pas spécifique, ce qui n'en fait pas une technique de choix pour obtenir un type cellulaire particulier.

On peut aussi trier les cellules selon leur taille, pour isoler des cellules tumorales circulantes à partir d'un échantillon sanguin par exemple (Hou, H. W. *et al.*, 2013). Dans cet exemple, les cellules qu'on souhaite isoler sont plus grosses (10-20 μm) que les cellules sanguines (érythrocytes : 8 μm ; leucocytes : 7-12 μm), ce qui permet de les séparer en utilisant leur différence d'inertie dans un système de microfluidique spécialement conçu pour récupérer ces cellules.

Les cellules peuvent également être triées grâce à leur densité, par centrifugation. Cette technique est très largement utilisée car elle permet rapidement de séparer différents types cellulaires comme les cellules du sang par exemple. Comme les autres techniques basées sur les propriétés physiques, cette technique n'est pas très spécifique, même si une meilleure séparation peut être obtenue en réalisant plusieurs centrifugations successives, ou en utilisant différents gradients de sucre. La centrifugation est par exemple utilisée comme première étape de purification pour éliminer les globules rouges et les plaquettes d'un échantillon sanguin.

Il existe une autre méthode qui mélange densité et fixation aux anticorps. Cette méthode, appelée *rosetting*, consiste à mettre en contact l'échantillon avec des anticorps dirigés contre les cellules qu'on souhaite éliminer. Ces anticorps vont ensuite former des complexes avec les érythrocytes, et ainsi augmenter la densité des types cellulaires à éliminer pour permettre leur séparation par centrifugation (Tomlinson, M. J. *et al.*, 2013).

Si les méthodes de séparation des cellules basées sur les propriétés physiques permettent rapidement de traiter de gros volumes et d'éliminer facilement des types cellulaires non désirés, c'est grâce aux méthodes de séparation basées sur la reconnaissance moléculaire qu'il est possible d'obtenir un type cellulaire donné sans contaminations par d'autres cellules. En effet, les techniques de purification basées sur les propriétés physiques des cellules permettent de récupérer un grand nombre de cellules

et de réaliser un premier tri, mais ne permettent pas d'obtenir des fractions très pures d'un type cellulaire en particulier. Cela est encore plus vrai lorsque le type cellulaire d'intérêt est en minorité dans l'échantillon. Des étapes supplémentaires sont donc souvent nécessaires à l'obtention de la population de cellules désirée. Ce manque de spécificité est dû au fait que les propriétés physiques utilisées pour trier les cellules sont partagées par plusieurs types cellulaires.

B. Les méthodes basées sur les anticorps/la reconnaissance moléculaire

Les méthodes de séparation basées sur la reconnaissance moléculaire font intervenir des anticorps dirigés contre des récepteurs spécifiques d'un type cellulaire donné. Les méthodes les plus fréquemment utilisées sont la cytométrie en flux (FACS : *fluorescent-activated cell sorting*) et le tri magnétique (MACS : *magnetic-activated cell sorting*). Dans ces deux techniques, les cellules d'intérêt sont marquées par un anticorps couplé à des billes magnétiques (MACS) ou à un fluorophore (FACS) avant d'être séparées du reste des cellules de l'échantillon.

La cytométrie en flux basée sur le FACS (*Fluorescent-activated cell sorting*) est un outil très puissant qui permet l'analyse d'un échantillon à l'échelle de la cellule isolée (Figure 8). Il est donc possible d'identifier et d'isoler des populations de cellules homogènes à partir d'une population hétérogène. L'échantillon à trier est marqué par des anticorps couplés à des fluorophores. Ces anticorps sont spécifiques d'une population cellulaire donnée. L'échantillon passe alors dans un flux qui permet de contrôler le passage des cellules une par une. Les cellules vont diffuser un rayon lumineux dans toutes les directions, et les rayons diffusés sont analysés à de faibles angles (0° à 15°), tandis que les rayons réfractés sont analysés à angle droit (90°). Les rayons diffusés à faibles angles renseignent sur la taille des cellules, tandis que les rayons réfractés à 90° renseignent sur la granulosité des cellules. Un laser excite alors les fluorophores présents sur les anticorps, et les intensités d'émission sont enregistrés pour chaque cellule. Le flux contenant les cellules est fractionné en gouttelettes homogènes, contenant environ 1 cellule par gouttelette. En fonction des paramètres souhaités par l'utilisateur (signal de fluorescence, taille des cellules, granulosité), une charge électrique ponctuelle est attribuée aux gouttelettes, qui peuvent être collectées dans différents tubes grâce à un champ électrostatique (Macey, M. G., s. d.). L'inconvénient principal du FACS est la lenteur du tri : une purification qui prendra plusieurs heures en FACS n'en prendra qu'une en MACS (Tomlinson, M. J. et al., 2013). Chaque cellule étant analysée individuellement, la capacité de tri reste limitée même dans des trieurs à débits élevés.

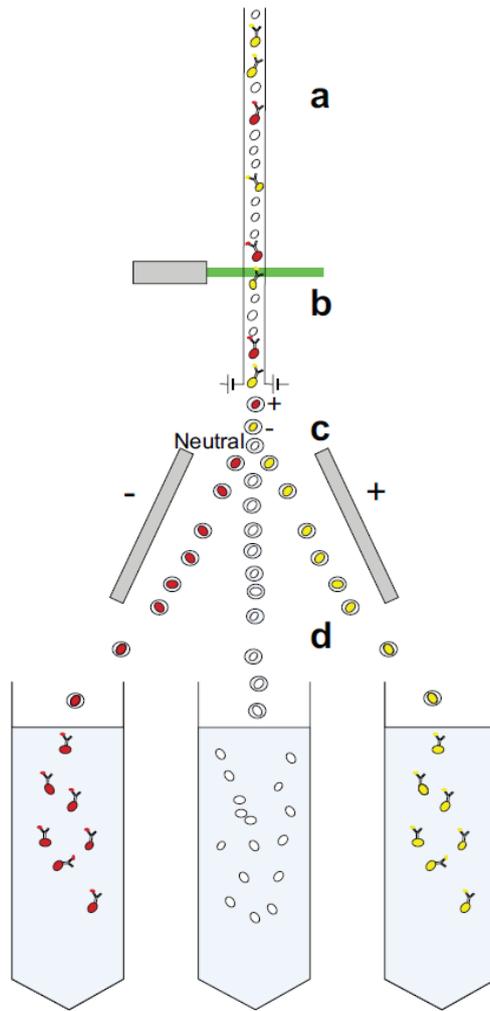


Figure 8 : : Schéma du principe de la cytométrie en flux (FACS)

Chaque type cellulaire à purifier est marqué par un anticorps couplé à un fluorophore différent. (a) Les cellules ainsi marquées sont canalisées pour donner un flux continu de cellules isolées. (b) Ces cellules passent à travers une source lumineuse ou un laser, et le signal émis par chaque cellule est détecté. Cette détection permet de déterminer si le signal émis par chaque cellule est supérieur ou inférieur à une valeur seuil donnée, et il est décidé de collecter ou non chaque cellule. (c) La collecte est réalisée en chargeant électriquement la gouttelette dans laquelle chaque cellule est contenue. (d) En faisant passer les gouttelettes à travers des plaques déflectrices chargées, il est possible de dévier les cellules vers les tubes de collecte appropriés (Tomlinson, M. J. *et al.*, 2013).

Le tri cellulaire par magnétisme (MACS) repose, comme le FACS, sur le marquage de l'échantillon à purifier par des anticorps spécifiques du type cellulaire que l'on souhaite trier. Cependant, contrairement au FACS, les anticorps sont ici couplés à des billes ou des nanoparticules magnétiques (Figure 9). Cela permet la rétention des cellules marquées dans un champ magnétique, tandis que les cellules non marquées sont éluées. Le retrait du champ magnétique permet ensuite de récupérer les

cellules précédemment retenues, c'est-à-dire le type cellulaire d'intérêt. Cette technique a comme avantage d'être très simple à mettre en œuvre, puisque contrairement au FACS elle ne demande pas beaucoup d'équipement et peut être réalisée rapidement sur des volumes importants. En revanche, un seul marquage peut être réalisé dans le cas du MACS alors que le FACS permet le marquage et le tri simultané de plusieurs populations de cellules.

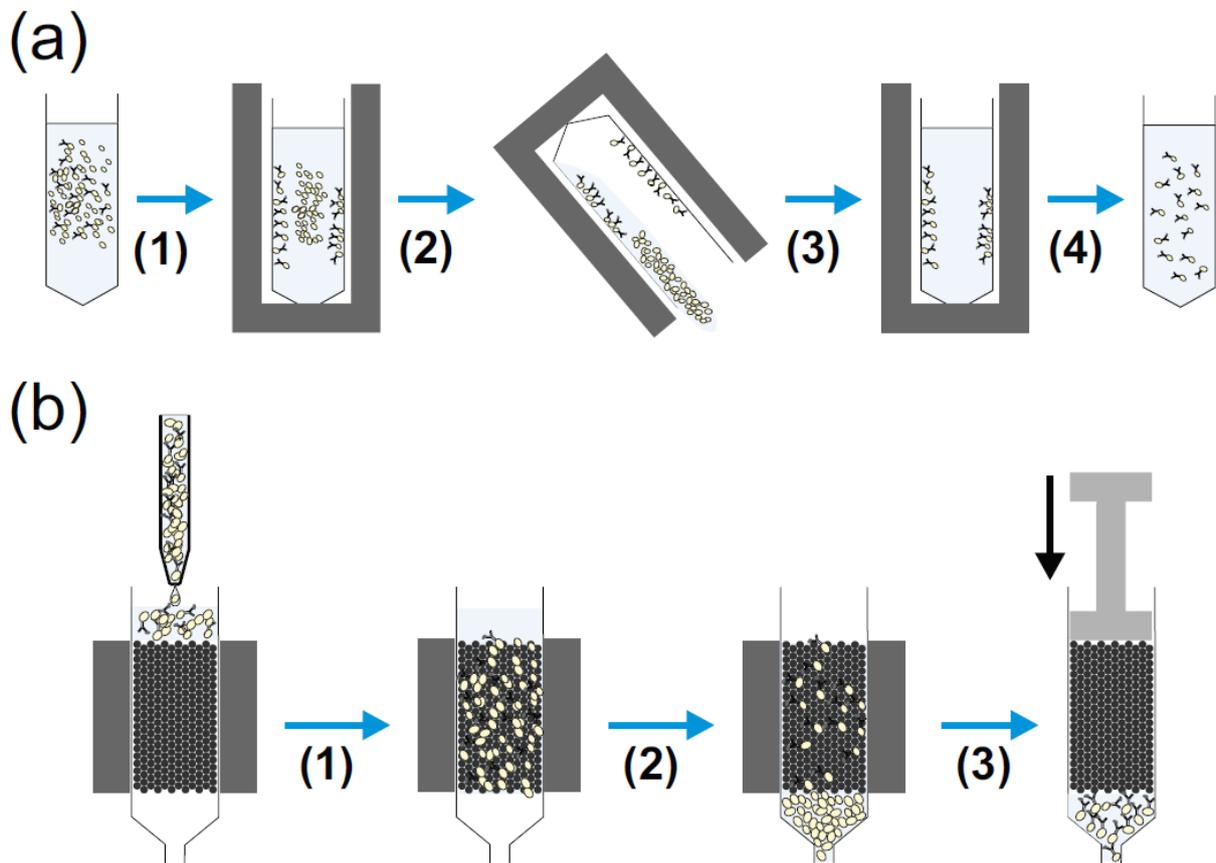


Figure 9 : Schéma du principe du tri magnétique (MACS)

(a) Séparation en tube : une suspension cellulaire marquée par des anticorps couplés à des billes magnétiques est placée dans un aimant. (1) Cela provoque le déplacement des cellules marquées vers les côtés du tube, en direction de l'aimant. Ce tube est ensuite (2) inversé (ou aspiré), ce qui permet de retirer les cellules non marquées avant (3) de remettre en suspension les cellules marquées et de les retirer de l'aimant, ce qui donne (4) une suspension dispersée de cellules cibles marquées. (b) Séparation sur colonne : une suspension cellulaire marquée par des anticorps couplés à des billes magnétiques est injectée dans une colonne maintenue à l'intérieur d'un aimant. (1) Les cellules traversent ensuite la colonne et (2) les cellules marquées sont retenues, tandis que les cellules non marquées sont éliminées par lavage. (3) Après l'élimination des cellules non marquées, la colonne est retirée de l'aimant, et un tampon de suspension est poussé à travers la colonne par un piston, ce qui permet de récupérer les cellules cibles marquées en suspension (Tomlinson, M. J. *et al.*, 2013).

Enfin, il existe d'autres méthodes de purification basées sur la fixation d'anticorps spécifiques d'un type cellulaire donné un peu moins utilisées que le FACS et le MACS. Ces techniques reposent sur

l'immobilisation des anticorps spécifiques du type cellulaire d'intérêt sur une surface, permettant la rétention du type cellulaire d'intérêt sur cette surface (Adams, A. A. *et al.*, 2008; Mahara, A. & Yamaoka, T., 2010). Dans ces articles, les cellules ont été libérées de la surface par une digestion enzymatique à la trypsine.

Ces techniques permettent de discriminer de façon précise des types cellulaires, mais reposent sur l'existence de marqueurs spécifiques du type cellulaire souhaité. De plus, ces techniques nécessitent d'avoir des anticorps dirigés contre ces marqueurs, ce qui limite souvent les possibilités de détection aux anticorps disponibles commercialement. Enfin, les anticorps utilisés pour marquer les cellules restent fixés à la surface de la cellule, et peuvent entraîner des réponses cellulaires comme des activations ou inhibitions de certaines voies de signalisation. Si l'activation de certaines voies de signalisation peut être volontaire pour certaines applications, comme l'activation des lymphocytes T par les anticorps bispécifiques CD3 x EpCAM, pour beaucoup d'autres applications on souhaite récupérer des cellules non altérées par la fixation d'anticorps à leur surface. Pour décrocher les anticorps de leur cible, il faut appliquer des conditions stressantes pour les cellules (pH acide, digestion à la trypsine) qui ne permettent pas forcément leur remise en culture dans le même état, ce qui est un problème pour certaines applications décrites plus en détail dans le chapitre suivant.

La compagnie Miltenyi Biotec a développé un outil de reconnaissance moléculaire basé sur des fragments d'anticorps qui permet de fixer puis de relâcher une cellule (Figure 10). Des fragments d'anticorps peu affins pour leur cible sont générés et fusionnés grâce à une molécule qu'on peut marquer avec un fluorochrome ou une nanoparticule magnétique. L'avidité générée par la fusion des interacteurs permet de marquer efficacement les cellules d'intérêt, et une purification par FACS ou par MACS est alors possible selon que l'interacteur a été marqué avec un fluorochrome ou une nanoparticule magnétique. Le principe de cette technologie est de pouvoir dissocier le complexe, probablement par l'ajout dans le milieu d'une protéase spécifique. Cette dissociation permet la perte du phénomène d'avidité des interacteurs, qui peuvent alors plus facilement libérer la cible puisqu'ils ont été sélectionnés pour avoir une faible affinité pour cette dernière. Les cellules ainsi libérées de tout marquage sont alors disponibles pour subir un nouveau marquage, ou pour réaliser d'autres expériences.

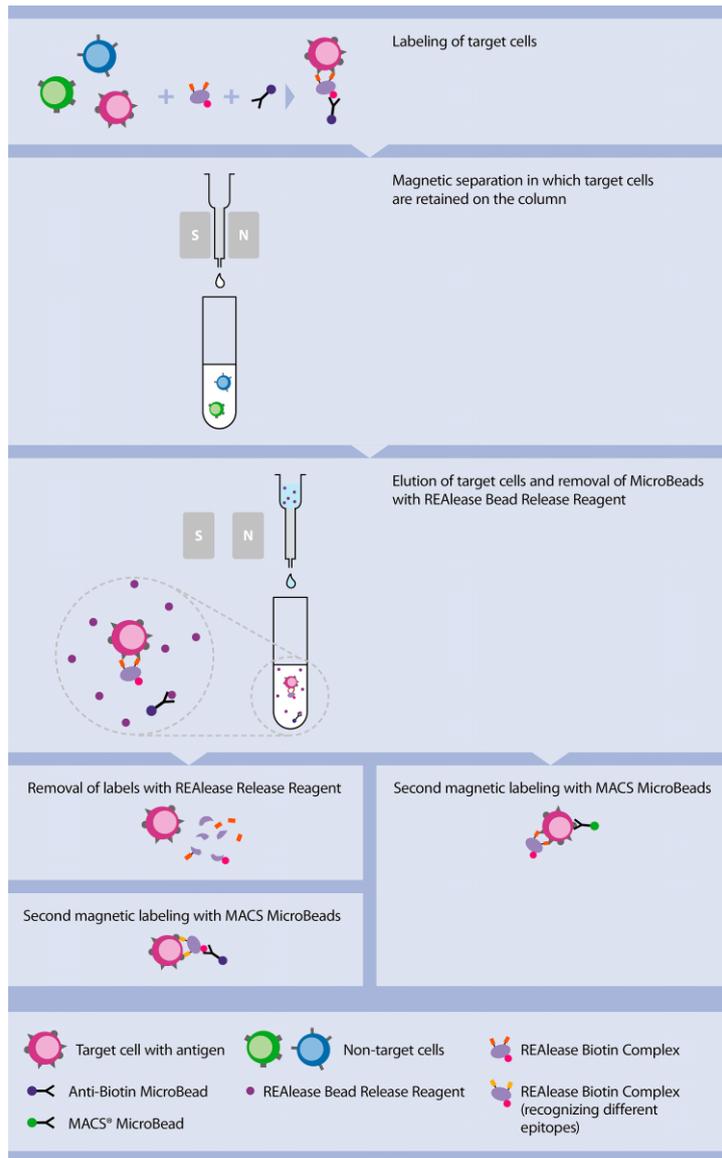


Figure 10 : schéma de fonctionnement de la technologie REAlease de Miltenyi Biotec

C. Les cellules tumorales circulantes

L'une des applications possibles des techniques de purification de cellule entière est la capture de cellules tumorales circulantes (CTCs). Ces cellules sont à l'origine des métastases qui sont la première cause de mortalité dans les cancers (Vila, A. *et al.*, 2016). Les cancers sont majoritairement d'origine épithéliale (Siegel, R. L. *et al.*, 2021), et pour se répandre dans l'organisme et former des tumeurs secondaires qu'on appelle métastases, les cellules tumorales doivent réaliser une transition épithélio-mésenchymateuse pour venir au contact des vaisseaux sanguins et lymphatiques. Une certaine proportion de ces cellules parvient alors à pénétrer ces vaisseaux par un phénomène qu'on appelle l'intravasation, et sont alors considérées comme des CTCs. Ce sont ces CTCs qui, par un phénomène d'extravasation suivi d'une transition mésenchymato-épithéliale, peuvent s'implanter dans d'autres tissus et donner les métastases (Vila, A. *et al.*, 2016).

La purification des CTCs apparaît particulièrement intéressante pour orienter la stratégie thérapeutique, ou pour étudier la résistance des cellules tumorales aux traitements par exemple. Ces applications peuvent nécessiter une remise en culture de ces cellules, ce qui n'est pas toujours compatible avec le marquage par des anticorps. Le Dr Catherine Alix-Panabières a d'ailleurs reçu une récompense de *l'American Association for Cancer Research Annual Meeting 2017*, le plus grand congrès annuel de cancérologie, pour un article de 2015 dans lequel son équipe est parvenu à obtenir une lignée stable à partir de cellules tumorales circulantes (Cayrefourcq, L. *et al.*, 2015). Cette distinction témoigne de l'intérêt de ce type de recherche en cancérologie, et des difficultés techniques que la remise en culture de CTCs représente.

Pour isoler des CTCs à partir d'échantillons sanguins de patients, Cayrefourcq *et al.* ont d'abord retiré les cellules hématopoïétiques des 10 mL de sang prélevés chez les patients par *rosetting* (cf partie III. B.) en utilisant le produit *RosetteSep™ CTC Enrichment Cocktail* (Stemcell technologies). Ce produit est un mélange d'anticorps dirigés contre les récepteurs CD2, CD16, CD19, CD36, CD38, CD45, CD66b et la glycophorine A, des marqueurs spécifiques des cellules sanguines. Cela permet d'enrichir l'échantillon en CTCs par tri négatif. Les cellules encore présentes dans l'échantillon ont alors été mises en cultures dans des conditions hypoxiques (2% O₂), favorables au développement des CTCs mais pas des autres types cellulaires. Au bout de plusieurs semaines, des lignées tumorales ont pu être obtenues.

Bien que cette méthode ait fonctionné, on peut quand même dégager plusieurs pistes d'amélioration. En réalité, l'étude de Cayrefourcq *et al.* a porté sur les échantillons sanguins de 71 patients atteints de cancers. Parmi ces 71 échantillons, 50 (70.4%) contenaient au moins 1 CTC, 38 (53.5%) contenaient au moins 3 CTCs, 18 (25.4%) contenaient plus de 10 CTCs, et 3 (4.2%) contenaient plus de 100 CTCs. Les lignées stables de CTCs ont pu être obtenues pour 2 patients qui possédaient plus de 300 CTCs

dans leur prélèvement sanguin, soit 2.8% du total des patients. Cette technique n'est donc pas forcément transposable à une majorité de patients atteints de cancers et pour lesquels il faudrait développer une nouvelle approche pour cultiver les CTCs. De plus, le principe du tri négatif impose d'utiliser un cocktail d'anticorps de façon à retenir l'ensemble des types cellulaires non recherchés, ce qui induit pour ce type d'approche un coût élevé lié aux anticorps utilisés. Il serait intéressant de pouvoir réaliser un tri positif en une étape plutôt que d'utiliser un mélange d'anticorps très coûteux. Réaliser une purification de CTCs par tri positif suppose d'avoir des marqueurs spécifiques des CTCs. Si on ne peut pas affirmer avec certitude qu'un marqueur est spécifique des CTCs, une très vaste majorité des CTCs surexpriment le récepteur EpCAM (Simon, M. *et al.*, 2013), un marqueur spécifique des cellules épithéliales. Or, les cellules épithéliales ne sont pas retrouvées dans la circulation sanguine en conditions normales, et ce marqueur peut donc être utilisé pour différencier les CTCs des cellules sanguines. Certaines CTCs expriment cependant le récepteur EpCAM à un faible niveau, voire pas du tout, et d'autres récepteurs homologues à EpCAM, comme TROP2 ou EGFR, peuvent donc être utilisés en complément d'EpCAM pour identifier le plus grand nombre de CTCs possible (Schneck, H. *et al.*, 2015).

Nous avons vu que la capture et la libération des cellules capturées dans des conditions permettant leur remise en culture en conditions natives représentait un intérêt certain, entre autres dans la recherche en cancérologie. Il existe déjà des anticorps qui reconnaissent EpCAM avec une très bonne affinité. Cependant, si les anticorps sont de formidables outils de reconnaissance moléculaire, l'avidité engendrée par leurs deux sites de fixation ne leur permet pas de se décrocher pas de leur cible en conditions normales. Leur dissociation requiert généralement une diminution du pH ou une digestion enzymatique ce qui n'est pas forcément compatible avec toutes les expériences qui pourraient être réalisées après la purification des cellules.

Mon projet de thèse propose de trouver une solution à cette problématique, en créant un outil dont l'une des applications serait de capturer des cibles d'intérêt avec une très bonne affinité, puis de les libérer en conditions natives.

IV. Stratégie générale

Objectifs de la thèse

Nous avons vu qu'il existait beaucoup d'outils de reconnaissance moléculaire dont les anticorps et leurs dérivés, et des bibliothèques d'architectures protéiques alternatives. Ces outils de reconnaissance sont la cible d'une ingénierie protéique qui vise souvent à améliorer l'affinité pour une cible, ou le nombre de cibles reconnues. Cependant, il n'existe pas à ce jour d'outil de reconnaissance moléculaire dont la fixation sur la cible puisse être modulée. C'est pourquoi nous nous sommes intéressés à créer un outil de reconnaissance moléculaire modulable. Comme la reconnaissance n'est plus limitée aux anticorps, nous nous sommes affranchis des limites de ces derniers en construisant une bibliothèque de protéines artificielles à partir d'une nouvelle ossature protéique alternative.

Pour développer un tel outil, nous avons décidé de concevoir une bibliothèque de protéines artificielles bi-domaines qui peuvent fixer une cible, mais également y associer un système qui permette de libérer cette cible dans un état natif et fonctionnel. En d'autres termes, il s'agirait de développer une bibliothèque de récepteurs allostérique bi-domaines pouvant fixer puis décrocher une cible en présence d'un modulateur générique. Ce modulateur générique sera généré par la sélection d' α Reps contre la surface du bi-domaine qui n'est pas en interaction avec la cible. De cette façon, l'interaction du modulateur avec la bi-P5H devrait induire un changement de conformation de cette dernière qui en conséquence, perdra sa capacité à interagir avec la cible. Ce projet repose donc sur l'hypothèse selon laquelle la conformation relative des 2 domaines composants les protéines de la bibliothèque ne sera pas la même lorsqu'elles seront en interaction avec une cible et lorsqu'elles seront en interaction avec le modulateur (Figure 11). De ce fait, la liaison de l'interacteur flexible avec le modulateur rigide devrait provoquer un changement de conformation chez l'interacteur issu de la bibliothèque. Ce changement de conformation ne doit pas permettre son interaction simultanée par ses deux domaines avec la cible, et donc provoquer une forte diminution de son énergie de liaison pour sa cible conduisant à sa dissociation.

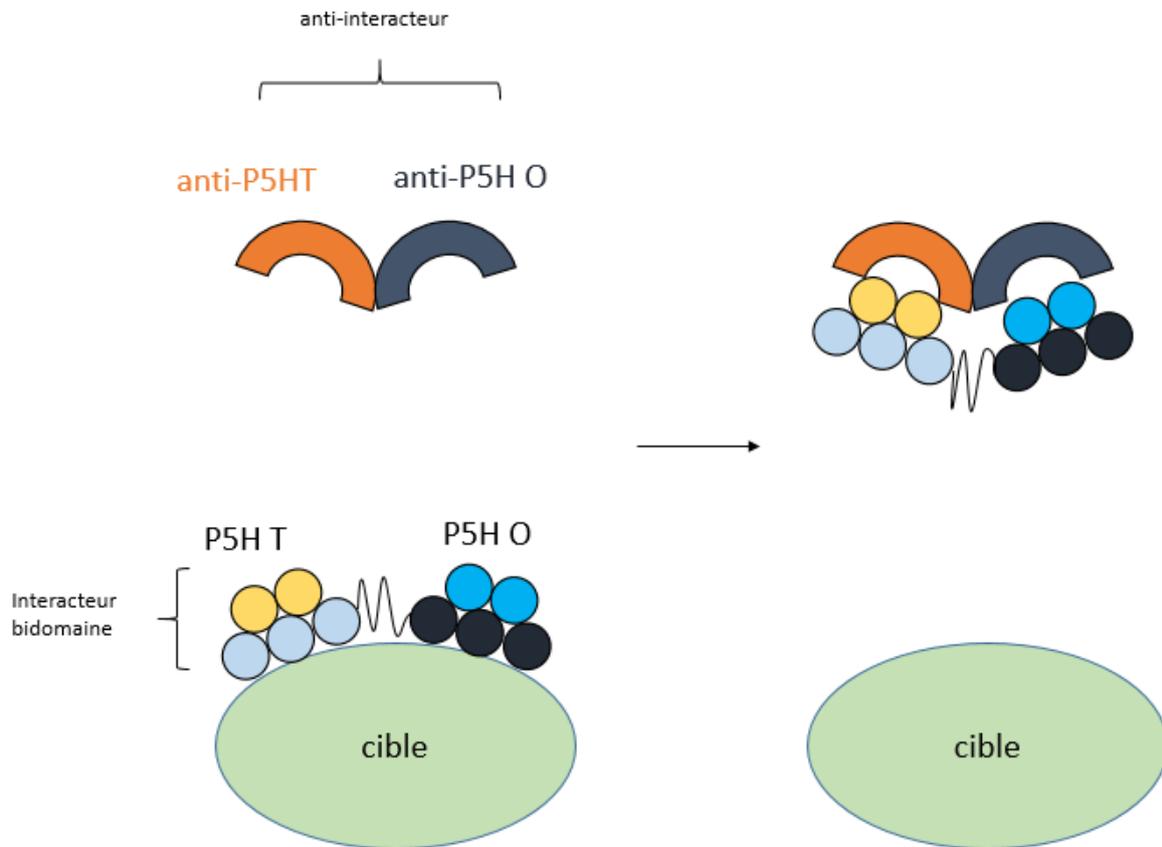


Figure 11 : Schéma du mode d'action d'un modulateur idéal issu de la bibliothèque

Dans un système où l'interacteur bi-modulaire est en interaction avec une cible (verte), l'ajout du modulateur rigide spécifique de l'interacteur provoque le décrochage de la cible. En effet l'interaction entre le modulateur et l'interacteur va contraindre la géométrie relative des 2 domaines de l'interacteur, induisant probablement le décrochage d'au moins un des 2 domaines (non représenté ici) ce qui va considérablement diminuer son affinité pour la cible et faciliter ainsi le décrochage du deuxième domaine.

Un outil permettant l'isolement et la remise en culture des CTCs présente donc un réel enjeu.

De même, la purification de protéines non étiquetées intéresse particulièrement les cristallographes et l'industrie pharmaceutique lorsqu'il s'agit de purifier des composés à visées thérapeutiques. Un outil qui puisse permettre de réaliser de telles purifications s'inscrirait donc pleinement dans le contexte scientifique actuel.

Dans ce cadre, mon projet de thèse propose un outil qui permet de purifier des cellules entières ou des protéines non étiquetées, ce qui s'inscrit dans une logique d'innovation et d'ingénierie protéique.

Contexte scientifique

Mes travaux de thèse ont consisté en la conception d'une bibliothèque inédite de protéines artificielles bi-domaine, ayant pour objectif de fixer des cibles protéiques et de les relâcher en condition natives par ajout dans le milieu d'une molécule appelée modulateur.

Tout d'abord, il a fallu identifier une nouvelle ossature protéique qui servirait de base à la conception de ma bibliothèque de protéines. Cette ossature devait répondre aux critères suivants :

- Un comportement monomérique en solution
- Une petite taille (< 30 kDa)
- Une haute stabilité
- Une forte tolérance à l'ingénierie protéique
- L'absence de ponts disulfures
- Une structure secondaire riche en hélices α

Pour identifier une telle ossature protéique, un examen approfondi de la PDB (*Protein Data Bank*) a été mené et a permis d'identifier une protéine de fonction inconnue, chez *Thermus thermophilus*, répondant à ces critères (code PDB : 2CWY). Un BLAST (*Basic Local Alignment Search Tool*) a permis d'identifier une seconde protéine candidate pour la conception de la bibliothèque bi-domaine (une ossature par domaine), homologue à 2CWY, et identifiée chez *Thermus oshimai*. *Thermus thermophilus* et *Thermus oshimai* sont des organismes thermophiles dont la température optimale de croissance se situe entre 65°C et 72°C. Les protéines issues des organismes thermophiles sont généralement très stables ce qui en fait de bons candidats à l'ingénierie protéique. Les 2 protéines ainsi retenues n'ont pas de fonction connue. Elles sont constituées de 94 (*Thermus thermophilus*) ou 93 (*Thermus oshimai*) acides aminés repliés en 5 hélices α antiparallèles et organisées en 2 couches ([Figure 12](#)), sont monomériques et ne possèdent pas de ponts disulfures. Nous avons décidé de les appeler pentahélines. Des tests ont été réalisés au laboratoire afin de vérifier la surexpression des 2 pentahélines, leur solubilité et leur stabilité. Leur tolérance aux mutations a également été vérifiée par génération, pour chaque pentahéline, d'un mutant qui diffère de la protéine sauvage par 18 acides aminés.

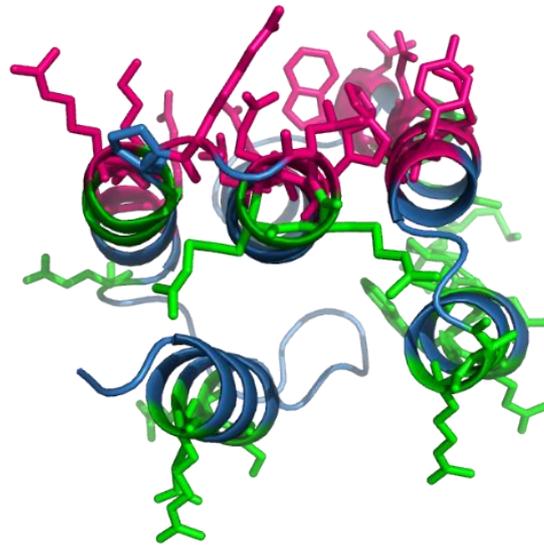


Figure 12 : Structure cristallographique de 2CWY

La structure a été représentée en cartoon. Les chaînes latérales qui seront mutées sur la face tri-hélicoïdale ont été colorées en rose et les positions accessibles de la face bi-hélicoïdale en vert. Le reste de la protéine est représenté en bleu.

Les travaux décrits précédemment ont été réalisés avant mon arrivée au laboratoire.

Mon projet de thèse a donc consisté à créer d'une part, la bibliothèque de protéines bi-domaines mentionnée précédemment, dont la conception sera décrite dans la première partie des résultats. La seconde partie des résultats correspondra à la conception du modulateur, cette molécule qui doit dissocier la cible des interacteurs issus de la bibliothèque. Enfin, une troisième partie sera consacrée aux applications du système conçu dans les deux premières parties.

Bibliographie

- Adams, A. A., Okagbare, P. I., Feng, J., Hupert, M. L., Patterson, D., Göttert, J., McCarley, R. L., Nikitopoulos, D., ... Soper, S. A., (2008), Highly Efficient Circulating Tumor Cell Isolation from Whole Blood and Label-Free Enumeration Using Polymer-Based Microfluidics with an Integrated Conductivity Sensor, *Journal of the American Chemical Society*, vol. 130, n°27, p. 8633-8641.
- Almeida, M., García-Montero, A. C. & Orfao, A., (2015), Cell Purification: A New Challenge for Biobanks, *Pathobiology*, vol. 81, n°5-6, p. 261-275.
- Bacac, M., Umaña, P., Herter, S., Colombetti, S., Sam, J., Le Clech, M., Freimoser-Grundschober, A., Richard, M., ... Klein, C., (2016), CD20 Tcb (RG6026), a Novel « 2:1 » T Cell Bispecific Antibody for the Treatment of B Cell Malignancies, *Blood*, vol. 128, n°22, p. 1836-1836.
- Bargou, R., Leo, E., Zugmaier, G., Klinger, M., Goebeler, M., Knop, S., Noppeney, R., Viardot, A., ... Kufer, P., (2008), Tumor Regression in Cancer Patients by Very Low Doses of a T Cell-Engaging Antibody, *Science*, vol. 321, n°5891, p. 974-977.
- Bates, A. & Power, C. A., (2019), David vs. Goliath: The Structure, Function, and Clinical Prospects of Antibody Fragments, *Antibodies*, vol. 8, n°2, p. 28.
- Bathula, N. V., Bommadevara, H. & Hayes, J. M., (2021), Nanobodies: The Future of Antibody-Based Immune Therapeutics, *Cancer Biotherapy and Radiopharmaceuticals*, vol. 36, n°2, p. 109-122.
- Batyuk, A., Wu, Y., Honegger, A., Heberling, M. M. & Plückthun, A., (2016), DARPIn-Based Crystallization Chaperones Exploit Molecular Geometry as a Screening Dimension in Protein Crystallography, *Journal of Molecular Biology*, vol. 428, n°8, p. 1574-1588.
- Becker, M. L. B., Near, R., Mudgett-Hunter, M., Margolies, M. N., Kubo, R. T., Kaye, J. & Hedrick, S. M., (1989), Expression of a hybrid immunoglobulin-T cell receptor protein in transgenic mice, *Cell*, vol. 58, n°5, p. 911-921.
- Beste, G., Schmidt, F. S., Stibora, T. & Skerra, A., (1999), Small antibody-like proteins with prescribed ligand specificities derived from the lipocalin fold, *Proceedings of the National Academy of Sciences*, vol. 96, n°5, p. 1898-1903.
- Bezverbnaya, K., Mathews, A., Sidhu, J., Helsen, C. W. & Bramson, J. L., (2017), Tumor-targeting domains for chimeric antigen receptor T cells, *Immunotherapy*, vol. 9, n°1, p. 33-46.
- Binz, H. K., Amstutz, P., Kohl, A., Stumpp, M. T., Briand, C., Forrer, P., Grütter, M. G. & Plückthun, A., (2004), High-affinity binders selected from designed ankyrin repeat protein libraries, *Nature Biotechnology*, vol. 22, n°5, p. 575-582.

Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Plückthun, A., (2003), Designing Repeat Proteins: Well-expressed, Soluble and Stable Proteins from Combinatorial Libraries of Consensus Ankyrin Repeat Proteins, *Journal of Molecular Biology*, vol. 332, n°2, p. 489-503.

Boder, E. T. & Wittrup, K. D., (1997), Yeast surface display for screening combinatorial polypeptide libraries, vol. 15, p. 5.

Boersma, Y. L., (2018), Advances in the Application of Designed Ankyrin Repeat Proteins (DARPs) as Research Tools and Protein Therapeutics, In : A. K. UDIT (éd.), *Protein Scaffolds*, Springer New York, New York, NY, vol. 1798, , p. 307-327.

Boersma, Y. L., Chao, G., Steiner, D., Wittrup, K. D. & Plückthun, A., (2011), Bispecific Designed Ankyrin Repeat Proteins (DARPs) Targeting Epidermal Growth Factor Receptor Inhibit A431 Cell Proliferation and Receptor Recycling, *Journal of Biological Chemistry*, vol. 286, n°48, p. 41273-41285.

Bogan, A. A. & Thorn, K. S., (1998), Anatomy of hot spots in protein interfaces, *Journal of Molecular Biology*, vol. 280, n°1, p. 1-9.

Brocchieri, L. & Karlin, S., (2005), Protein length in eukaryotic and prokaryotic proteomes, *Nucleic Acids Research*, vol. 33, n°10, p. 3390-3400.

Campanacci, V., Urvoas, A., Cantos-Fernandes, S., Aumont-Nicaise, M., Arteni, A.-A., Velours, C., Valerio-Lepiniec, M., Dreier, B., ... Gigant, B., (2019), Insight into microtubule nucleation from tubulin-capping proteins, *Proceedings of the National Academy of Sciences*, vol. 116, n°20, p. 9859-9864.

Cayrefourcq, L., Mazard, T., Joosse, S., Solassol, J., Ramos, J., Assenat, E., Schumacher, U., Costes, V., ... Alix-Panabières, C., (2015), Establishment and Characterization of a Cell Line from Human Circulating Colon Cancer Cells, *Cancer Research*, vol. 75, n°5, p. 892-901.

Chames, P. & Wurch, T., (2019), Les anticorps et **scaffold** bispécifiques, des médicaments innovants en oncologie impliquant le ciblage des cellules immunitaires, *médecine/sciences*, vol. 35, n°12, p. 1072-1082.

Chatenoud, L., Ferran, C., Legendre, C., Thouard, I., Merite, S., Reuter, A., Gevaert, Y., Kreis, H., ... Bach, J.-F., (1990), IN VIVO CELL ACTIVATION FOLLOWING OKT3 ADMINISTRATION: SYSTEMIC CYTOKINE RELEASE AND MODULATION BY CORTICOSTEROIDS, *Transplantation*, vol. 49, n°4, p. 697-702.

Chen, J., Sawyer, N. & Regan, L., (2013), Protein-protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area: Protein-protein Interactions, *Protein Science*, vol. 22, n°4, p. 510-515.

Chevrel, A., Mesneau, A., Sanchez, D., Celma, L., Quevillon-Cheruel, S., Cavagnino, A., Nessler, S., Li de la Sierra-Gallay, I., ... Urvoas, A., (2018), Alpha repeat proteins (α Rep) as expression and crystallization helpers, *Journal of Structural Biology*, vol. 201, n°2, p. 88-99.

Chevrel, A., Urvoas, A., de la Sierra-Gallay, I. L., Aumont-Nicaise, M., Moutel, S., Desmadril, M., Perez, F., Gautreau, A., ... Valerio-Lepiniec, M., (2015), Specific GFP-binding artificial proteins (α Rep): a new tool for in vitro to live cell applications, *Bioscience Reports*, vol. 35, n°4, p. e00223.

Chmielewski, M., Kopecky, C., Hombach, A. A. & Abken, H., (2011), IL-12 Release by Engineered T Cells Expressing Chimeric Antigen Receptors Can Effectively Muster an Antigen-Independent Macrophage Response on Tumor Cells That Have Shut Down Tumor Antigen Expression, *Cancer Research*, vol. 71, n°17, p. 5697-5706.

Courtenay-Luck, N. S., Epenetos, A. A., Moore, R., Lã, M., Dhokia, B. & Ritter, M. A., (1986), Development of Primary and Secondary Immune Responses to Mouse Monoclonal Antibodies Used in the Diagnosis and Therapy of Malignant Neoplasms, p. 6.

De Gast, G. C., Van Houten, A. A., Haagen, I.-A., Klein, S., De Weger, R. A., Van Dijk, A., Phillips, J., Clark, M. & Bast, B. J. E. G., (1995), Clinical Experience with CD3 X CD19 Bispecific Antibodies in Patients with B Cell Malignancies, *Journal of Hematotherapy*, vol. 4, n°5, p. 433-437.

Deisenhofer, J., (s. d.), Crystallographic Refinement and Atomic Models of a Human Fc Fragment and Its Complex with Fragment B of Protein A from *Staphylococcus aureus* at 2.9- and 2.8-Å Resolution+, p. 10.

Diem, M. D., Hyun, L., Yi, F., Hippensteel, R., Kuhar, E., Lowenstein, C., Swift, E. J., O'Neil, K. T. & Jacobs, S. A., (2014), Selection of high-affinity Centyrin FN3 domains from a simple library diversified at a combination of strand and loop positions, *Protein Engineering, Design and Selection*, vol. 27, n°10, p. 419-429.

Dreier, B., Honegger, A., Hess, C., Nagy-Davidescu, G., Mittl, P. R. E., Grutter, M. G., Belousova, N., Mikheeva, G., ... Pluckthun, A., (2013), Development of a generic adenovirus delivery system based on structure-guided design of bispecific trimeric DARPin adapters, *Proceedings of the National Academy of Sciences*, vol. 110, n°10, p. E869-E877.

Dunlevy, F. K., Martin, S. L., de Courcey, F., Elborn, J. S. & Ennis, M., (2012), Anti-inflammatory effects of DX-890, a human neutrophil elastase inhibitor, *Journal of Cystic Fibrosis*, vol. 11, n°4, p. 300-304.

Edelman, G. M., (1959), DISSOCIATION OF γ -GLOBULIN, *Journal of the American Chemical Society*, vol. 81, n°12, p. 3155-3156.

Eggel, A., Baravalle, G., Hobi, G., Kim, B., Buschor, P., Forrer, P., Shin, J.-S., Vogel, M., ... Jardetzky, T. S., (2014), Accelerated dissociation of IgE-Fc ϵ RI complexes by disruptive inhibitors actively desensitizes allergic effector cells, *Journal of Allergy and Clinical Immunology*, vol. 133, n°6, p. 1709-1719.e8.

Eggel, A., Baumann, M. J., Amstutz, P., Stadler, B. M. & Vogel, M., (2009), DARPins as Bispecific Receptor Antagonists Analyzed for Immunoglobulin E Receptor Blockage, *Journal of Molecular Biology*, vol. 393, n°3, p. 598-607.

Ekman, D., Björklund, Å. K., Frey-Skött, J. & Elofsson, A., (2005), Multi-domain Proteins in the Three Kingdoms of Life: Orphan Domains and Other Unassigned Regions, *Journal of Molecular Biology*, vol. 348, n°1, p. 231-243.

Feldhaus, M. J., Siegel, R. W., Opresko, L. K., Coleman, J. R., Feldhaus, J. M. W., Yeung, Y. A., Cochran, J. R., Heinzelman, P., ... Wittrup, K. D., (2003), Flow-cytometric isolation of human antibodies from a nonimmune *Saccharomyces cerevisiae* surface display library, *Nature Biotechnology*, vol. 21, n°2, p. 163-170.

Figueroa, M., Sleutel, M., Vandevenne, M., Parvizi, G., Attout, S., Jacquin, O., Vandenameele, J., Fischer, A. W., ... Van de Weerd, C., (2016), The unexpected structure of the designed protein Octarellin V.1 forms a challenge for protein structure prediction tools, *Journal of Structural Biology*, vol. 195, n°1, p. 19-30.

Finney, H. M., Lawson, A. D. G., Bebbington, C. R. & Weir, A. N. C., (1998), Chimeric Receptors Providing Both Primary and Costimulatory Signaling in T Cells from a Single Gene Product, *The Journal of Immunology*, vol. 161, n°6, p. 2791.

Forrer, P., Binz, H. K., Stumpp, M. T. & Plückthun, A., (2004), Consensus Design of Repeat Proteins, *ChemBioChem*, vol. 5, n°2, p. 183-189.

Fudenberg, H. H., Drews, G. & Nisonoff, A., (1964), SEROLOGIC DEMONSTRATION OF DUAL SPECIFICITY OF RABBIT BIVALENT HYBRID ANTIBODY, *Journal of Experimental Medicine*, vol. 119, n°1, p. 151-166.

Getmanova, E. V., Chen, Y., Bloom, L., Gokemeijer, J., Shamah, S., Warikoo, V., Wang, J., Ling, V. & Sun, L., (2006), Antagonists to Human and Mouse Vascular Endothelial Growth Factor Receptor 2 Generated by Directed Protein Evolution In Vitro, *Chemistry & Biology*, vol. 13, n°5, p. 549-556.

Geyer, C. R., McCafferty, J., Dübel, S., Bradbury, A. R. M. & Sidhu, S. S., (2012), Recombinant Antibodies and In Vitro Selection Technologies, In : G. PROETZEL et H. EBERSBACH (éd.), *Antibody Methods and Protocols*, Humana Press, Totowa, NJ vol. 901, , p. 11-32.

Gilbreth, R. N., Esaki, K., Koide, A., Sidhu, S. S. & Koide, S., (2008), A Dominant Conformational Role for Amino Acid Diversity in Minimalist Protein-Protein Interfaces, *Journal of Molecular Biology*, vol. 381, n°2, p. 407-418.

Godar, M., de Haard, H., Blanchetot, C. & Rasser, J., (2018), Therapeutic bispecific antibody formats: a patent applications review (1994-2017), *Expert Opinion on Therapeutic Patents*, vol. 28, n°3, p. 251-276.

Goldberg, S. D., Cardoso, R. M. F., Lin, T., Spinka-Doms, T., Klein, D., Jacobs, S. A., Dudkin, V., Gilliland, G. & O'Neil, K. T., (2016), Engineering a targeted delivery platform using Centyrins, *Protein Engineering Design and Selection*, p. proeng;gzw054v1.

Grant, S. D. (2021). Therapeutic Antibodies. *Drug Discovery and Development E-Book: Technology in Transition*, 108.

Gross, G. G., Junge, J. A., Mora, R. J., Kwon, H.-B., Olson, C. A., Takahashi, T. T., Liman, E. R., Ellis-Davies, G. C. R., ... Arnold, D. B., (2013), Recombinant Probes for Visualizing Endogenous Synaptic Proteins in Living Neurons, *Neuron*, vol. 78, n°6, p. 971-985.

Gu, G. J., Friedman, M., Jost, C., Johnsson, K., Kamali-Moghaddam, M., Plückthun, A., Landegren, U. & Söderberg, O., (2013), Protein tag-mediated conjugation of oligonucleotides to recombinant affinity binders for proximity ligation, *New Biotechnology*, vol. 30, n°2, p. 144-152.

Guellouz, A., Valerio-Lepiniec, M., Urvoas, A., Chevrel, A., Graille, M., Fourati-Kammoun, Z., Desmadril, M., van Tilbeurgh, H. & Minard, P., (2013), Selection of Specific Protein Binders for Pre-Defined Targets from an Optimized Library of Artificial Helicoidal Repeat Proteins (alphaRep), *PLoS ONE*, vol. 8, n°8, p. e71512.

Gunneriusson, E., Nord, K., Uhlén, M. & Nygren, P.-Å., (1999), Affinity maturation of a Taq DNA polymerase specific affibody by helix shuffling, *Protein Engineering, Design and Selection*, vol. 12, n°10, p. 873-878.

Guo, K.-T., SchÄfer, R., Paul, A., Gerber, A., Ziemer, G. & Wendel, H. P., (2006), A New Technique for the Isolation and Surface Immobilization of Mesenchymal Stem Cells from Whole Bone Marrow Using High-Specific DNA Aptamers, *Stem Cells*, vol. 24, n°10, p. 2220-2231.

Hackel, B. J., Ackerman, M. E., Howland, S. W. & Wittrup, K. D., (2010), Stability and CDR Composition Biases Enrich Binder Functionality Landscapes, *Journal of Molecular Biology*, vol. 401, n°1, p. 84-96.

Hackel, B. J., Kapila, A. & Dane Wittrup, K., (2008), Picomolar Affinity Fibronectin Domains Engineered Utilizing Loop Length Diversity, Recursive Mutagenesis, and Loop Shuffling, *Journal of Molecular Biology*, vol. 381, n°5, p. 1238-1252.

Hackel, B. J. & Wittrup, K. D., (2010), The full amino acid repertoire is superior to serine/tyrosine for selection of high affinity immunoglobulin G binders from the fibronectin scaffold, *Protein Engineering Design and Selection*, vol. 23, n°4, p. 211-219.

Heyd, B., Pecorari, F., Collinet, B., Adjadj, E., Desmadril, M. & Minard, P., (2003), In Vitro Evolution of the Binding Specificity of Neocarzinostatin, an Eneidyne-Binding Chromoprotein, *Biochemistry*, vol. 42, n°19, p. 5674-5683.

Hohlbaum, A. M. & Skerra, A., (2007), Anticalins[®] : the lipocalin family as a novel protein scaffold for the development of next-generation immunotherapies, *Expert Review of Clinical Immunology*, vol. 3, n°4, p. 491-501.

Hou, H. W., Warkiani, M. E., Khoo, B. L., Li, Z. R., Soo, R. A., Tan, D. S.-W., Lim, W.-T., Han, J., ... Lim, C. T., (2013), Isolation and retrieval of circulating tumor cells using centrifugal forces, *Scientific Reports*, vol. 3, n°1, p. 1259.

Hulme, J. T., D'Souza, W. N., McBride, H. J., Yoon, B.-R. P., Willee, A. M., Duguay, A., Thomas, M., Fan, B., ... Li, P., (2017), Novel protein therapeutic joint retention strategy based on collagen-binding Avimers: CARTILAGE TETHERING COLLAGEN-BINDING AVIMERS, *Journal of Orthopaedic Research*.

Iri-Sofla, F. J., Rahbarizadeh, F., Ahmadvand, D. & Rasaei, M. J., (2011), Nanobody-based chimeric receptor gene integration in Jurkat cells mediated by PhiC31 integrase, *Experimental Cell Research*, vol. 317, n°18, p. 2630-2641.

Jamieson, C. H. M., Ailles, L. E., Dylla, S. J., Muijtjens, M., Jones, C., Zehnder, J. L., Gotlib, J., Li, K., ... Weissman, I. L., (2004), Granulocyte–Macrophage Progenitors as Candidate Leukemic Stem Cells in Blast-Crisis CML, *New England Journal of Medicine*, vol. 351, n°7, p. 657-667.

Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S. & Winter, G., (1986), Replacing the complementarity-determining regions in a human antibody with those from a mouse, *Nature*, vol. 321, n°6069, p. 522-525.

Jost, C., Schilling, J., Tamaskovic, R., Schwill, M., Honegger, A. & Plückthun, A., (2013), Structural Basis for Eliciting a Cytotoxic Effect in HER2-Overexpressing Cancer Cells via Binding to the Extracellular Domain of HER2, *Structure*, vol. 21, n°11, p. 1979-1991.

Julien, O., Gignac, I., Hutton, A., Yee, A., Arrowsmith, C. H. & Gagné, S. M., (2006), MTH187 from *Methanobacterium thermoautotrophicum* has three HEAT-like Repeats, *Journal of Biomolecular NMR*, vol. 35, n°2, p. 149-154.

Karatan, E., Merguerian, M., Han, Z., Scholle, M. D., Koide, S. & Kay, B. K., (2004), Molecular Recognition Properties of FN3 Monobodies that Bind the Src SH3 Domain, *Chemistry & Biology*, vol. 11, n°6, p. 835-844.

Khan, S. N., Sok, D., Tran, K., Movsesyan, A., Dubrovskaya, V., Burton, D. R. & Wyatt, R. T., (2018), Targeting the HIV-1 Spike and Coreceptor with Bi- and Trispecific Antibodies for Single-Component Broad Inhibition of Entry, *Journal of Virology*, vol. 92, n°18, .

Kim, B., Eggel, A., Tarchevskaya, S. S., Vogel, M., Prinz, H. & Jardetzky, T. S., (2012), Accelerated disassembly of IgE–receptor complexes by a disruptive macromolecular inhibitor, *Nature*, vol. 491, n°7425, p. 613-617.

Klein, C., Schaefer, W. & Regula, J. T., (2016), The use of CrossMAb technology for the generation of bi- and multispecific antibodies, *MAbs*, vol. 8, n°6, p. 1010-1020.

Koduri, V. & Blacklow, S. C., (2001), Folding Determinants of LDL Receptor Type A Modules, *Biochemistry*, vol. 40, n°43, p. 12801-12807.

Köhler, G. & Milstein, C., (1975), Continuous cultures of fused cells secreting antibody of predefined specificity, *Nature*, vol. 256, n°5517, p. 495-497.

Koide, A., Abbatiello, S., Rothgery, L. & Koide, S., (2002), Probing protein conformational changes in living cells by using designer binding proteins: Application to the estrogen receptor, *Proceedings of the National Academy of Sciences*, vol. 99, n°3, p. 1253-1258.

Koide, A., Bailey, C. W., Huang, X. & Koide, S., (1998), The fibronectin type III domain as a scaffold for novel binding proteins, *Journal of Molecular Biology*, vol. 284, n°4, p. 1141-1151.

Koide, A., Gilbreth, R. N., Esaki, K., Tereshko, V. & Koide, S., (2007), High-affinity single-domain binding proteins with a binary-code interface, *Proceedings of the National Academy of Sciences*, vol. 104, n°16, p. 6632-6637.

Kummer, L., Hsu, C.-W., Dagliyan, O., MacNevin, C., Kaufholz, M., Zimmermann, B., Dokholyan, N. V., Hahn, K. M. & Plückthun, A., (2013), Knowledge-Based Design of a Biosensor to Quantify Localized ERK Activation in Living Cells, *Chemistry & Biology*, vol. 20, n°6, p. 847-856.

Kyle, S., (2018), Affimer Proteins: Theranostics of the Future?, *Trends in Biochemical Sciences*, vol. 43, n°4, p. 230-232.

Labrijn, A. F., Janmaat, M. L., Reichert, J. M. & Parren, P. W. H. I., (2019), Bispecific antibodies: a mechanistic review of the pipeline, *Nature Reviews Drug Discovery*, vol. 18, n°8, p. 585-608.

Lanzoni, G., Alviano, F., Marchionni, C., Bonsi, L., Costa, R., Foroni, L., Roda, G., Belluzzi, A., ... Paolo Bagnara, G., (2009), Isolation of stem cell populations with trophic and immunoregulatory functions from human intestinal tissues: potential for cell therapy in inflammatory bowel disease, *Cytotherapy*, vol. 11, n°8, p. 1020-1031.

Lewis, S. M., Wu, X., Pustilnik, A., Sereno, A., Huang, F., Rick, H. L., Guntas, G., Leaver-Fay, A., ... Demarest, S. J., (2014), Generation of bispecific IgG antibodies by structure-based design of an orthogonal Fab interface, *Nature Biotechnology*, vol. 32, n°2, p. 191-198.

Liao, H.-I., Olson, C. A., Hwang, S., Deng, H., Wong, E., Baric, R. S., Roberts, R. W. & Sun, R., (2009), mRNA Display Design of Fibronectin-based Intrabodies That Detect and Inhibit Severe Acute Respiratory Syndrome Coronavirus Nucleocapsid Protein, *Journal of Biological Chemistry*, vol. 284, n°26, p. 17512-17520.

Lipovsek, D., (2011), Adnectins: engineered target-binding protein therapeutics, *Protein Engineering Design and Selection*, vol. 24, n°1-2, p. 3-9.

Lipovšek, D., Lippow, S. M., Hackel, B. J., Gregson, M. W., Cheng, P., Kapila, A. & Wittrup, K. D., (2007), Evolution of an Interloop Disulfide Bond in High-Affinity Antibody Mimics Based on Fibronectin Type III Domain and Selected by Yeast Surface Display: Molecular Convergence with Single-Domain Camelid and Shark Antibodies, *Journal of Molecular Biology*, vol. 368, n°4, p. 1024-1041.

Macey, M. G., (s. d.), Principles of Flow Cytometry, *Introduction to Flow Cytometry*, p. 15.

Magadum, S., Banerjee, U., Murugan, P., Gangapur, D. & Ravikesavan, R., (2013), Gene duplication as a major force in evolution, *Journal of Genetics*, vol. 92, n°1, p. 155-161.

Mahara, A. & Yamaoka, T., (2010), Continuous separation of cells of high osteoblastic differentiation potential from mesenchymal stem cells on an antibody-immobilized column, *Biomaterials*, vol. 31, n°14, p. 4231-4237.

Majeti, R., Becker, M. W., Tian, Q., Lee, T.-L. M., Yan, X., Liu, R., Chiang, J.-H., Hood, L., ... Weissman, I. L., (2009), Dysregulated gene expression networks in human acute myelogenous leukemia stem cells, *Proceedings of the National Academy of Sciences*, vol. 106, n°9, p. 3396-3401.

Marin, V., Pizzitola, I., Agostoni, V., Attianese, G. M. P. G., Finney, H., Lawson, A., Pule, M., Rousseau, R., ... Biagi, E., (2010), Cytokine-induced killer cells for cell therapy of acute myeloid leukemia: improvement of their immune activity by expression of CD33-specific chimeric receptors, *Haematologica*, vol. 95, n°12, p. 2144-2152.

McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J., (1990), Phage antibodies: filamentous phage displaying antibody variable domains, *Nature*, vol. 348, n°6301, p. 552-554.

Moot, R., Raikar, S. S., Fleischer, L., Querrey, M., Tylawsky, D. E., Nakahara, H., Doering, C. B. & Spencer, H. T., (2016), Genetic engineering of chimeric antigen receptors using lamprey derived variable lymphocyte receptors, *Molecular Therapy - Oncolytics*, vol. 3, p. 16026.

Morrison, C., (2020), Fresh from the biotech pipeline—2019, *Nature Biotechnology*, vol. 38, n°2, p. 126-131.

Morrison, C. & Lähteenmäki, R., (2017), Public biotech in 2016—the numbers, *Nature Biotechnology*, vol. 35, n°7, p. 623-629.

Münch, R. C., Muth, A., Muik, A., Friedel, T., Schmatz, J., Dreier, B., Trkola, A., Plückthun, A., ... Buchholz, C. J., (2015), Off-target-free gene delivery by affinity-purified receptor-targeted viral vectors, *Nature Communications*, vol. 6, n°1, p. 6246.

Neuberger, M. S., Williams, G. T., Mitchell, E. B., Jouhal, S. S., Flanagan, J. G. & Rabbitts, T. H., (1985), A hapten-specific chimaeric IgE antibody with human physiological effector function, *Nature*, vol. 314, n°6008, p. 268-270.

Nichols, A. L., Noridomi, K., Hughes, C. R., Jalali-Yazdi, F., Eaton, J. B., Lai, L. H., Advani, G., Lukas, R. J., ... Roberts, R. W., (2018), α 1-FANGs: Protein Ligands Selective for the α -Bungarotoxin Site of the α 1-Nicotinic Acetylcholine Receptor, *ACS Chemical Biology*, vol. 13, n°9, p. 2568-2576.

Nilsson, B., Moks, T., Jansson, B., Abrahmsén, L., Elmlblad, A., Holmgren, E., Henrichson, C., Jones, T. A. & Uhlén, M., (1987), A synthetic IgG-binding domain based on staphylococcal protein A, *Protein Engineering, Design and Selection*, vol. 1, n°2, p. 107-113.

Nisonoff, A., Wissler, F. C. & Lipman, L. N., (1960), Properties of the Major Component of a Peptic Digest of Rabbit Antibody, *Science*, vol. 132, n°3441, p. 1770-1771.

Nooren, I. M. A., (2003), NEW EMBO MEMBER'S REVIEW: Diversity of protein-protein interactions, *The EMBO Journal*, vol. 22, n°14, p. 3486-3492.

Nord, K., Nilsson, J., Nilsson, B., Uhlén, M. & Nygren, P.-Å., (1995), A combinatorial library of an α -helical bacterial receptor domain, *Protein Engineering, Design and Selection*, vol. 8, n°6, p. 601-608.

North, C. L. & Blacklow, S. C., (1999), Structural Independence of Ligand-Binding Modules Five and Six of the LDL Receptor, *Biochemistry*, vol. 38, n°13, p. 3926-3935.

Olson, C. A., Liao, H.-I., Sun, R. & Roberts, R. W., (2008), mRNA Display Selection of a High-Affinity, Modification-Specific Phospho-I κ B α -Binding Fibronectin, *ACS Chemical Biology*, vol. 3, n°8, p. 480-485.

Orfao, A. & Ruiz-Argüelles, A., (1996), General concepts about cell sorting techniques, *Clinical Biochemistry*, vol. 29, n°1, p. 5-9.

Orlova, A., Magnusson, M., Eriksson, T. L. J., Nilsson, M., Larsson, B., Höidén-Guthenberg, I., Widström, C., Carlsson, J., ... Nilsson, F. Y., (2006), Tumor Imaging Using a Picomolar Affinity HER2 Binding Affibody Molecule, *Cancer Research*, vol. 66, n°8, p. 4339-4348.

Pancer, Z., Amemiya, C. T., Ehrhardt, G. R. A., Ceitlin, J., Larry Gartland, G. & Cooper, M. D., (2004), Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey, *Nature*, vol. 430, n°6996, p. 174-180.

Parker, M. H., Chen, Y., Danehy, F., Dufu, K., Ekstrom, J., Getmanova, E., Gokemeijer, J., Xu, L. & Lipovsek, D., (2005), Antibody mimics based on human fibronectin type three domain engineered for thermostability and high-affinity binding to vascular endothelial growth factor receptor two, *Protein Engineering, Design and Selection*, vol. 18, n°9, p. 435-444.

Perez, P., Hoffman, R. W., Shaw, S., Bluestone, J. A. & Segal, D. M., (1985), Specific targeting of cytotoxic T cells by anti-T3 linked to anti-target cell antibody, *Nature*, vol. 316, n°6026, p. 354-356.

Philibert, P., Stoessel, A., Wang, W., Sibler, A.-P., Bec, N., Larroque, C., Saven, J. G., Courtête, J., ... Martineau, P., (2007), A focused antibody library for selecting scFvs expressed at high levels in the cytoplasm, *BMC Biotechnology*, vol. 7, n°1, p. 81.

Plückthun, A., (2015), Designed Ankyrin Repeat Proteins (DARPs): Binding Proteins for Research, Diagnostics, and Therapy, *Annual Review of Pharmacology and Toxicology*, vol. 55, n°1, p. 489-511.

Porter, R. R., (s. d.), The Hydrolysis of Rabbit γ -Globulin and Antibodies with Crystalline Papain, p. 9.

Richards, J., Miller, M., Abend, J., Koide, A., Koide, S. & Dewhurst, S., (2003), Engineered Fibronectin Type III Domain with a RGDWXE Sequence Binds with Enhanced Affinity and Specificity to Human α v β 3 Integrin, *Journal of Molecular Biology*, vol. 326, n°5, p. 1475-1488.

Riethmüller, G., (2012), Symmetry breaking: bispecific antibodies, beginnings, and 50 years on, *Cancer Immunity*, vol. 12, p. 7.

Roberts, B. L., Markland, W., Ley, A. C., Kent, R. B., White, D. W., Guterman, S. K. & Ladner, R. C., (1992), Directed evolution of a protein: selection of potent neutrophil elastase inhibitors displayed on M13 fusion phage., *Proceedings of the National Academy of Sciences*, vol. 89, n°6, p. 2429-2433.

Röthlisberger, D., Honegger, A. & Plückthun, A., (2005), Domain Interactions in the Fab Fragment: A Comparative Evaluation of the Single-chain Fv and Fab Format Engineered with Variable Domains of Different Stability, *Journal of Molecular Biology*, vol. 347, n°4, p. 773-789.

Schilling, J., Schöppe, J. & Plückthun, A., (2014), From DARPins to LoopDARPins: Novel LoopDARPin Design Allows the Selection of Low Picomolar Binders in a Single Round of Ribosome Display, *Journal of Molecular Biology*, vol. 426, n°3, p. 691-721.

Schlehuber, S. & Skerra, A., (2005), Lipocalins in drug discovery: from natural ligand-binding proteins to 'anticalins', vol. 10, n°1, p. 11.

Schneck, H., Gierke, B., Uppenkamp, F., Behrens, B., Niederacher, D., Stoecklein, N. H., Templin, M. F., Pawlak, M., ... Disseminated Cancer Cell Network (DCC Net) Duesseldorf, (2015), EpCAM-Independent Enrichment of Circulating Tumor Cells in Metastatic Breast Cancer, *PLOS ONE*, vol. 10, n°12, p. e0144535.

Schonfeld, D., Matschiner, G., Chatwell, L., Trentmann, S., Gille, H., Hulsmeyer, M., Brown, N., Kaye, P. M., ... Skerra, A., (2009), An engineered lipocalin specific for CTLA-4 reveals a combining site with structural and conformational features similar to antibodies, *Proceedings of the National Academy of Sciences*, vol. 106, n°20, p. 8198-8203.

Schütz, M., Batyuk, A., Klenk, C., Kummer, L., de Picciotto, S., Gülbakan, B., Wu, Y., Newby, G. A., ... Plückthun, A., (2016), Generation of Fluorogen-Activating Designed Ankyrin Repeat Proteins (FADAs) as Versatile Sensor Tools, *Journal of Molecular Biology*, vol. 428, n°6, p. 1272-1289.

Seeger, M. A., Zbinden, R., Flütsch, A., Gutte, P. G. M., Engeler, S., Roschitzki-Voser, H. & Grütter, M. G., (2013), Design, construction, and characterization of a second-generation DARPIn library with reduced hydrophobicity: Second-Generation DARPIn Library, *Protein Science*, vol. 22, n°9, p. 1239-1257.

Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A., (2021), Cancer Statistics, 2021, *CA: A Cancer Journal for Clinicians*, vol. 71, n°1, p. 7-33.

Silverman, J., Lu, Q., Bakker, A., To, W., Duguay, A., Alba, B. M., Smith, R., Rivas, A., ... Stemmer, W. P. C., (2005), Multivalent avimer proteins evolved by exon shuffling of a family of human receptor domains, *Nature Biotechnology*, vol. 23, n°12, p. 1556-1561.

Simon, M., Stefan, N., Plückthun, A. & Zangemeister-Wittke, U., (2013), Epithelial cell adhesion molecule-targeted drug delivery for cancer therapy, *Expert Opinion on Drug Delivery*, vol. 10, n°4, p. 451-468.

Simon, M., Zangemeister-Wittke, U. & Plückthun, A., (2012), Facile Double-Functionalization of Designed Ankyrin Repeat Proteins using Click and Thiol Chemistries, *Bioconjugate Chemistry*, vol. 23, n°2, p. 279-286.

Staerz, U. D., Kanagawa, O. & Bevan, M. J., (1985), Hybrid antibodies can target sites for attack by T cells, *Nature*, vol. 314, n°6012, p. 628-631.

Stefan, N., Zimmermann, M., Simon, M., Zangemeister-Wittke, U. & Plückthun, A., (2014), Novel Prodrug-Like Fusion Toxin with Protease-Sensitive Bioorthogonal PEGylation for Tumor Targeting, *Bioconjugate Chemistry*, vol. 25, n°12, p. 2144-2156.

Swers, J. S., Grinberg, L., Wang, L., Feng, H., Lekstrom, K., Carrasco, R., Xiao, Z., Inigo, I., ... Baca, M., (2013), Multivalent Scaffold Proteins as Superagonists of TRAIL Receptor 2-Induced Apoptosis, *Molecular Cancer Therapeutics*, vol. 12, n°7, p. 1235-1244.

Tamaskovic, R., Schwill, M., Nagy-Davidescu, G., Jost, C., Schaefer, D. C., Verdurmen, W. P. R., Schaefer, J. V., Honegger, A. & Plückthun, A., (2016), Intermolecular biparatopic trapping of ErbB2 prevents compensatory activation of PI3K/AKT via RAS-p110 crosstalk, *Nature Communications*, vol. 7, n°1, p. 11672.

Terry, V. H., Johnston, I. C. D. & Spina, C. A., (2009), CD44 MicroBeads accelerate HIV-1 infection in T cells, *Virology*, vol. 388, n°2, p. 294-304.

Theurillat, J.-P., Dreier, B., Nagy-Davidescu, G., Seifert, B., Behnke, S., Zürcher-Härdi, U., Ingold, F., Plückthun, A. & Moch, H., (2010), Designed ankyrin repeat proteins: a novel tool for testing epidermal growth factor receptor 2 expression in breast cancer, *Modern Pathology*, vol. 23, n°9, p. 1289-1297.

Tibben, J. G., Boerman, O. C., Massuger, L. F. A. G., Schijf, C. P. T., Claessens, R. A. M. J. & Corstens, F. H. M., (1996), Pharmacokinetics, biodistribution and biological effects of intravenously administered bispecific monoclonal antibody OC/TR F(ab')₂ in ovarian carcinoma patients, *International Journal of Cancer*, vol. 66, n°4, p. 477-483.

Tiede, C., Tang, A. A. S., Deacon, S. E., Mandal, U., Nettleship, J. E., Owen, R. L., George, S. E., Harrison, D. J., ... McPherson, M. J., (2014), Adhiron: a stable and versatile peptide display scaffold for molecular recognition applications, *Protein Engineering, Design and Selection*, vol. 27, n°5, p. 145-155.

Tiouajni, M., Durand, D., Blondeau, K., Graille, M., Urvoas, A., Valerio-Lepiniec, M., Guellouz, A., Aumont-Nicaise, M., ... van Tilbeurgh, H., (2014), Structural and functional analysis of the fibronectin-binding protein FNE from *Streptococcus equi* spp. *equi*, *FEBS Journal*, vol. 281, n°24, p. 5513-5531.

Tomlinson, M. J., Tomlinson, S., Yang, X. B. & Kirkham, J., (2013), Cell separation: Terminology and practical considerations, *Journal of Tissue Engineering*, vol. 4, p. 204173141247269.

Urvoas, A., Guellouz, A., Valerio-Lepiniec, M., Graille, M., Durand, D., Desravines, D. C., van Tilbeurgh, H., Desmadril, M. & Minard, P., (2010), Design, Production and Molecular Structure of a New Family of Artificial Alpha-helical Repeat Proteins (α Rep) Based on Thermostable HEAT-like Repeats, *Journal of Molecular Biology*, vol. 404, n°2, p. 307-327.

Vila, A., Abal, M., Muinelo-Romay, L., Rodriguez-Abreu, C., Rivas, J., López-López, R. & Costa, C., (2016), EGFR-Based Immunoisolation as a Recovery Target for Low-EpCAM CTC Subpopulation, *PLOS ONE*, vol. 11, n°10, p. e0163705.

Vogt, M. & Skerra, A., (2004), Construction of an Artificial Receptor Protein (“Anticalin”) Based on the Human Apolipoprotein D, *ChemBioChem*, vol. 5, n°2, p. 191-199.

Weisser, N., Wickman, G., Davies, R. & Rowse, G., (2017), Abstract 31: Preclinical development of a novel biparatopic HER2 antibody with activity in low to high HER2 expressing cancers, In : *Experimental and Molecular Therapeutics*, American Association for Cancer Research, p. 31-31.

Will, B. & Steidl, U., (2010), Multi-parameter fluorescence-activated cell sorting and analysis of stem and progenitor cells in myeloid malignancies, *Best Practice & Research Clinical Haematology*, vol. 23, n°3, p. 391-401.

Winter, G. & Milstein, C., (1991), Man-made antibodies, *Nature*, vol. 349, n°6307, p. 293-299.

Wojcik, J., Hantschel, O., Grebien, F., Kaupe, I., Bennett, K. L., Barking, J., Jones, R. B., Koide, A., ... Koide, S., (2010), A potent and highly specific FN3 monobody inhibitor of the Abl SH2 domain, *Nature Structural & Molecular Biology*, vol. 17, n°4, p. 519-527.

Wolff, A. C., Hammond, M. E. H., Hicks, D. G., Dowsett, M., McShane, L. M., Allison, K. H., Allred, D. C., Bartlett, J. M. S., ... Hayes, D. F., (2013), Recommendations for Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Update, *Journal of Clinical Oncology*, vol. 31, n°31, p. 3997-4013.

Wu, X., Yuan, R., Bacica, M. & Demarest, S. J., (2018), Generation of orthogonal Fab-based trispecific antibody formats, *Protein Engineering, Design and Selection*, vol. 31, n°7-8, p. 249-256.

Xu, D., Alegre, M.-L., Varga, S. S., Rothermel, A. L., Collins, A. M., Pulito, V. L., Hanna, L. S., Dolan, K. P., ... Zivin, R. A., (2000), In Vitro Characterization of Five Humanized OKT3 Effector Function Variant Antibodies, *Cellular Immunology*, vol. 200, n°1, p. 16-26.

Xu, L., Aha, P., Gu, K., Kuimelis, R. G., Kurz, M., Lam, T., Lim, A. C., Liu, H., ... Lipovsek, D., (2002), Directed Evolution of High-Affinity Antibody Mimics Using mRNA Display, *Chemistry & Biology*, vol. 9, n°8, p. 933-942.

Yang, W.-P., Green, K., Pinz-Sweeney, S., Briones, A. T., Burton, D. R. & Barbas III, C. F., (1995), CDR Walking Mutagenesis for the Affinity Maturation of a Potent Human Anti-HIV-1 Antibody into the Picomolar Range, *Journal of Molecular Biology*, vol. 254, n°3, p. 392-403.

Zielinski, R., Lyakhov, I., Jacobs, A., Chertov, O., Kramer-Marek, G., Francella, N., Stephen, A., Fisher, R., ... Capala, J., (2009), Affitoxin—A Novel Recombinant, HER2-specific, Anticancer Agent for Targeted Therapy of HER2-positive Tumors, *Journal of Immunotherapy*, vol. 32, n°8, p. 817-825.

Résultats et discussion

Table des matières

Résultats et discussion	63
Partie 1 : Conception et construction de la bibliothèque de biP5H.....	65
I. Identification du <i>scaffold</i> de départ	66
II. Production des P5H et caractérisation de leurs propriétés de stabilité	74
II. Ingénierie des bi-domaines P5H.....	78
IV. Stratégie d'assemblage de la bibliothèque	89
Partie 2 : Conception du modulateur allostérique.....	104
I. Sélections par phage display d' α Reps reconnaissant les P5H.....	105
II. Caractérisation des α Reps reconnaissant les P5H.....	115
III. Conception du modulateur	118
IV. Caractérisation des modulateurs	125
Partie 3 : Applications de la bibliothèque de bi-P5H.....	136
I. Evaluation de la bibliothèque de bi-P5H sur l'eGFP.....	138
II. Evaluation de la bibliothèque de bi-P5H sur des protéines membranaires.....	145
V. Bilan et perspectives	158
Bibliographie.....	165

Partie 1 : Conception et construction de la bibliothèque de biP5H

I. Identification du *scaffold* de départ

Dans un premier temps, il a fallu identifier un repliement robuste qui conviendrait pour construire la bibliothèque. Un examen de la PDB a donc été réalisé, en focalisant les recherches sur des petites protéines, riches en hélices α , sans ponts disulfures, et issues d'organismes thermophiles. La recherche s'oriente autour de protéines issues d'organismes thermophiles car ces protéines sont généralement très stables. Or, il est préférable de partir d'une ossature protéique stable pour réaliser une bibliothèque de protéines puisque les mutations introduites afin de créer la surface d'interaction vont souvent avoir tendance à déstabiliser la structure. Il est préférable de choisir une ossature protéique ne contenant pas de pont disulfure afin de pouvoir l'exprimer facilement dans le cytoplasme des bactéries. Les ossatures protéiques riches en hélices α sont plus appropriées pour la construction de bibliothèques car l'expérience acquise au laboratoire a montré que ces structures étaient plus tolérantes aux mutations que les structures riches en feuillets β . Cela peut être dû au fait que les liaisons hydrogène dans une hélice α se font entre les résidus d'une même hélice alors que les liaisons hydrogène au sein d'un feuillet β se font entre des brins voisins, rendant les mutations plus propices à déstabiliser la structure secondaire. Enfin, on a également recherché une protéine de petite taille, car d'une part elles sont plus facilement manipulables au niveau de leur gène, et d'autre part car l'ossature recherchée devait constituer un seul des deux domaines de la future protéine de la bibliothèque. Il était donc préférable de choisir une ossature de départ relativement petite, notamment pour faciliter les expériences de biologie moléculaire.

Une protéine a été retenue, car remplissant les critères évoqués précédemment. En effet, cette protéine (code PDB : 2CWY), dont la structure est présentée [Figure 13](#), a été identifiée chez *Thermus thermophilus*, un organisme thermophile. Elle est monomérique, composée de 5 hélices α antiparallèles disposées en deux couches, et ne possède pas de fonction connue. Le repliement a été baptisée pentahéline (P5H) en raison de sa structure. Cette protéine peut donc servir de base à la construction d'un des deux domaines de la bibliothèque de protéines bi-domaine.

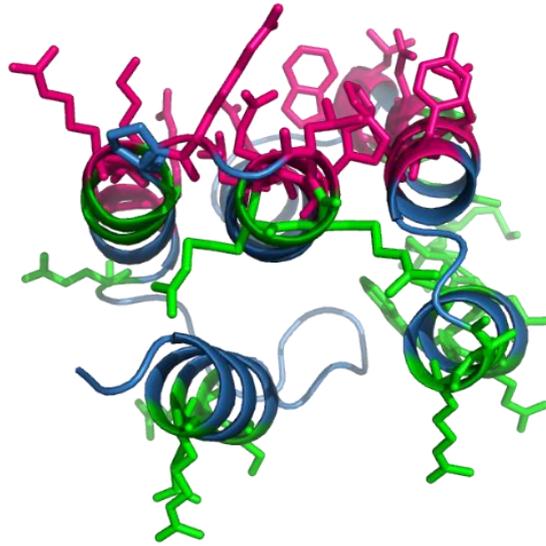


Figure 13 : Structure cristallographique de 2CWY à 1.85 Å

La structure a été représentée en *cartoon*. Les chaînes latérales qui seront mutées sur la face tri-hélicoïdale ont été colorées en rose et les positions accessibles de la face bi-hélicoïdale en vert. Le reste de la protéine est représenté en bleu.

Afin de constituer le 2^{ème} domaine de la protéine de la bibliothèque, il était nécessaire d'identifier une seconde ossature protéique homologue à la première (2CWY), possédant une séquence différente mais une structure similaire. Dans cette optique, un BLAST a été réalisé contre 2CWY et a révélé l'existence de plusieurs protéines homologues chez d'autres organismes thermophiles apparentés à *Thermus thermophilus* (Figure 14).

```

>gi|55771450|dbj|BAD69891.1| conserved hypothetical protein [Thermus
thermophilus HB8]
MVPDWEEVLGLWRAGRYEYEVHEVLEPYWLKATGEERRLLQGVILLAAALHQRRLLGRPGLRNLKAEARLEGLPCP
LMGLD
WRSLLQEARRRLGA
>gi|551066405|ref|WP_022797721.1| hypothetical protein [Thermus islandicus]
MEVLGEALRLWREGRYFEVHEVLEEAWREARAGEERRFLOGLILLAAALHQOGLGKSGLRNLKAEARLRGLPSY
LGLDW
RPLLEARRRLGA
>gi|517272448|ref|WP_018461266.1| hypothetical protein [Thermus oshimai]
MEVLKEALRLWEEGRYFEVHEVLEEAWREARAGEERRFLOGLILLAAALHQKALGKTGLGNLRKAEARLKGLPSY
LGVDWEALLGEARRRLGA
>gi|516808267|ref|WP_018111047.1| hypothetical protein [Thermus igniterrae]
MEAFKEAWRLFQEGKYFEAHEVLEGPWREARAGEERRFLOGLILLAAALHQAKMGRRGLRNLKKAEGKLGFPSPY
LGLDW
QPLLQEARRRLGA
>gi|218243290|gb|EED09820.1| protein of unknown function DUF309 [Thermus
aquaticus Y51MC23]
MEALEAFVRHFQEGRYFEAHEVLEEAWLKAEGEERRFLOGLILLAAALHRGKGGVRNLKAEAKLSGLPSPYLGL
DWELL
WQEARRRLGA
>gi|518381188|ref|WP_019551395.1| hypothetical protein [Thermus
scotoductus]
MEALGEAWRLFQEGQYFEAHEALEEVWREAQGEKRRFLOGLILLAAALHQAKSGRGGRLNLKKAEGKLLGFPSPY
LSLDW
RPLFEEARRRLGA
>gi|328450263|gb|AEB11164.1| protein of unknown function DUF309
[Marinithermus hydrothermalis DSM 14884]
MAFTDTPAWREAQALWRAGRYWEVHEALEPLWHEAQGAERAFLOGVILLAAALHKARTSPTGGRRNYAKALRHLLA
PLPDVDFHGVQVRALELEVHRALLDPGYTPCLPTVDFPS
>gi|481062567|gb|AGK05730.1| hypothetical protein K649_12210 [Meiothermus
ruber DSM 1279]
MQGIDLLNRPEYTEALELWHQGRFWEVHEALEPLWQQLSGPDRELTHGVILLAAALHKARDNPRGGWRNFHKALK
HLEGLPATYRGVRVVDLIEATRQALQKGPQIPAFPRL
>gi|517278330|ref|WP_018467148.1| hypothetical protein [Meiothermus
timidus]
MDWLEDDPLWDGASPIHLANTPEFARARELWQRGAFWEVHEVLEPLWVRLEGEQREVVHGLILLAAALHKARSSP
TGGWRNFAKALRHLEGLPGVYAGLDLSQTIAEVRRALEQPGYHPRLCFASS
>gi|517841859|ref|WP_019012067.1| hypothetical protein [Deinococcus
aquatilis]
MNATEVGGWQEGITLNFAGHWWEAHEAWEEVWLQATGDERACLSALILLAAALHKRWHHGS LAHRNFIKAEKYLD
ALPEQYAGLDLTELRAVVAALQASDIDQARPQLQRGVN

```

Figure 14 : Résultat du BLAST contre 2CWY.

Le BLAST a été réalisé sans a priori contre une base de données de séquences protéiques non redondantes. Les résultats ont été retenus sur la base de leur pourcentage d'identité de séquence et en éliminant les redondances.

Un alignement multiple a été réalisé entre 2CWY et ses plus proches homologues issus du BLAST (Figure 15.A). Puis, un nouvel alignement a été réalisé avec les 4 protéines les plus proches de 2CWY d'un point de vue évolutif (Figure 15.B/C). À ce stade, il a fallu choisir deux protéines qui serviraient de base aux deux domaines des futures protéines de la bibliothèque d'interacteurs bi-domaines. Les deux

pentahélines choisies ont été les pentahélines issues des organismes *Thermus thermophilus* et *Thermus oshimai*. En effet, la structure de la première est connue (code pdb : 2CWY), et la deuxième possède le plus d'identité de séquence avec 2CWY, soit 71% d'identité de séquence. De plus, 2CWY se replie en un seul domaine, et étant donné leur identité de séquence ainsi que leur proximité phylogénétique, notre hypothèse est que la pentahéline issue de *Thermus oshimai* se replierait également en un seul domaine similaire à celui de *Thermus thermophilus*. Ces critères ne sont cependant pas suffisants pour valider ces architectures comme point de départ d'une nouvelle bibliothèque d'interacteurs.

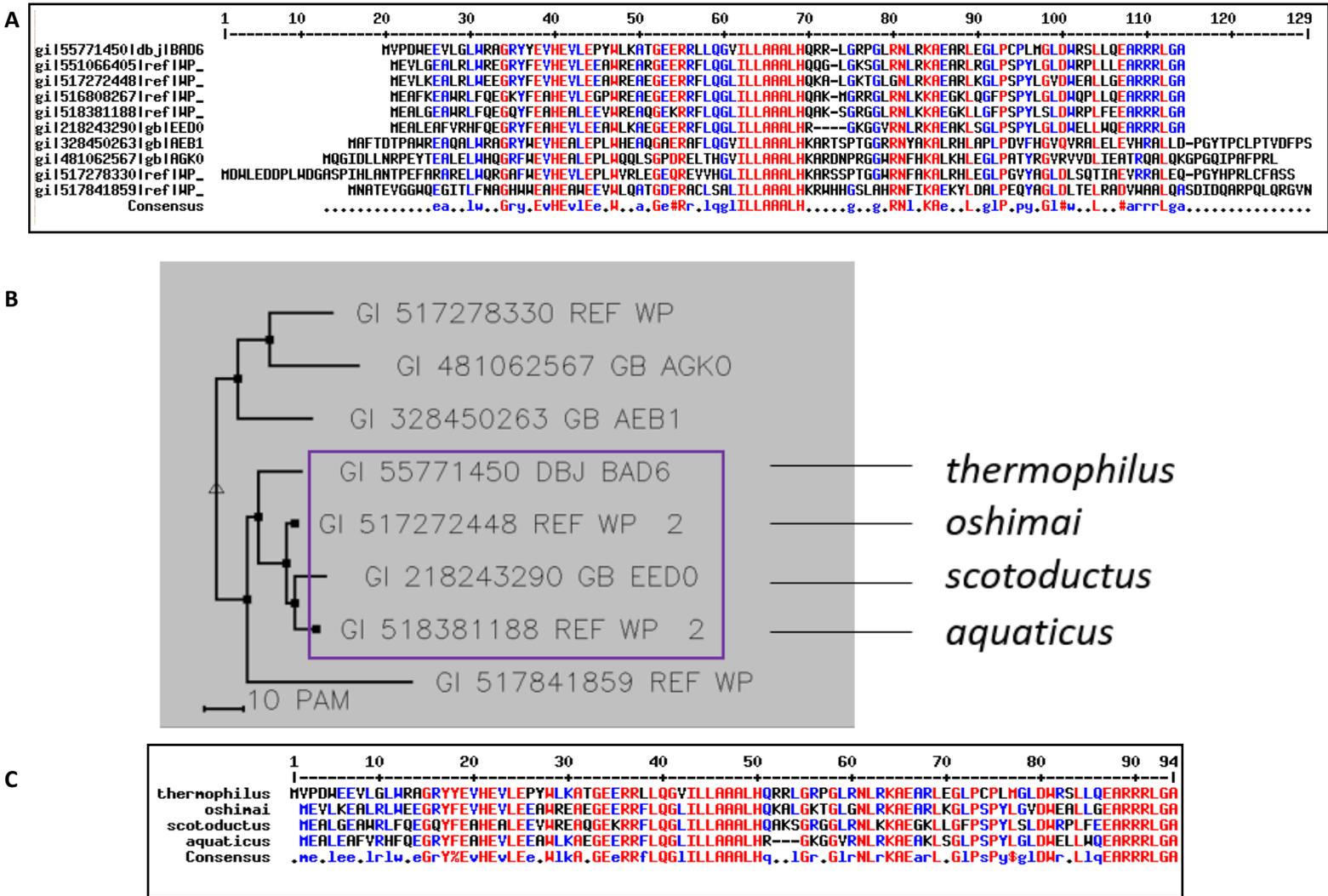


Figure 15 : Stratégie d'identification d'un repliement robuste pour la bibliothèque de protéine.

(A) Alignement des séquences issues du BLAST contre 2CWY (B). Arbre phylogénétique des protéines issues du BLAST contre 2CWY (C). Alignement des 4 protéines les plus proches phylogénétiquement de 2CWY (encadrées en violet Figure 15.B). Les protéines sont nommées d'après l'organisme dont elles sont originaires

Dans un second temps, nous avons donc étudié l'impact de mutations multiples sur l'expression, la solubilité, et la stabilité de ces architectures. En effet, la construction d'une bibliothèque aléatoire

repose sur la création d'un grand répertoire de protéines toutes issues d'une même protéine matrice sur laquelle des positions ont été choisies pour être diversifiées. Le choix de ces positions est déterminant puisqu'il faut un nombre suffisant de mutations pour générer une surface diversifiée conséquente, tout en maintenant la stabilité de l'ossature protéique. Dans cette optique, les séquences des deux pentahélines, P5H T (issue de *Thermus thermophilus*) et P5H O (issue de *Thermus oshimai*) ont été annotées sur la base de leur structure, afin d'identifier les positions accessibles sur leur deux « faces » (Figure 16).

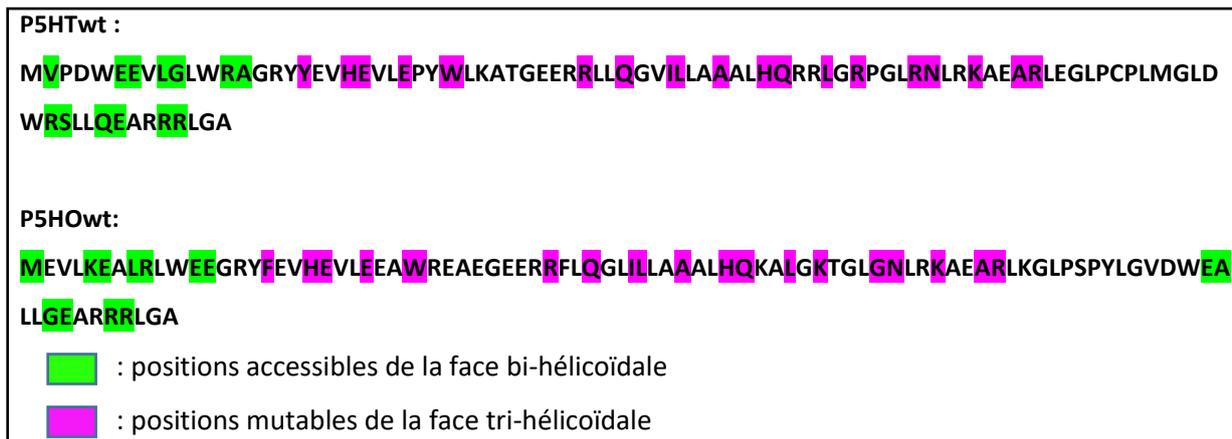


Figure 16 : Annotation des séquences *wild-type* des pentahélines issues de *Thermus thermophilus* et *Thermus oshimai*

Les séquences des pentahélines issues de *Thermus thermophilus* et *Thermus oshimai* ont été annotées à l'aide de leur structure cristallographique. Les positions surlignées en vert ont des chaînes latérales exposées sur la face bi-hélicoïdale, et les positions surlignées en magenta ont des chaînes latérales exposées sur la face tri-hélicoïdale des P5H.

Pour rappel, l'idée était de pouvoir imaginer sur chaque ossature deux surfaces distinctes constituant d'une part le site de fixation pour les cibles (cette surface sera diversifiée dans la bibliothèque finale) et d'autre part un site de fixation pour le modulateur (Figure 17). Les positions surlignées en magenta sur la Figure 16 forment une surface continue sur la face tri-hélicoïdale choisie pour être diversifiée dans la future bibliothèque afin de créer un vaste répertoire de surfaces d'interaction pouvant reconnaître toute cible d'intérêt. Les positions surlignées en vert correspondent aux chaînes latérales accessibles sur la face bi-hélicoïdale qui pourraient constituer un site de fixation pour le modulateur. Deux gènes synthétiques ont été conçus pour chacune des pentahélines. Un gène dit « *wild type* » : P5H Twt1 et P5H Owt, et un gène dit « mutant » : P5H Tmut1 et P5H Omut. En réalité les gènes *wild type* diffèrent des véritables séquences *wild type* par quelques résidus, dans le but de rapprocher les séquences du consensus défini à partir du BLAST, et de rajouter une tyrosine accessible sur la face bi-hélicoïdale afin d'induire un *hot spot* de reconnaissance (ce concept sera expliqué ultérieurement) pour le modulateur. Il s'agissait enfin de supprimer une cystéine unique chez P5H T pour éviter la

formation de ponts disulfures illégitimes entre protéines ou avec des composants de milieu de production. Les gènes mutants correspondent aux gènes *wild type* avec les 18 positions indiquées en magenta sur la [Figure 16](#) mutées par un même jeu de résidus pour les deux pentahélines. Quatre gènes synthétiques ont donc été commandés : P5H Twt1, P5H Tmut1, P5H Owt, et P5H Omut ([Figure 18](#)), avec des sites de restrictions BamH I/Hind III pour les insérer dans le vecteur pQE80L. Les P5H T possèdent la même face bi-hélicoïdale, mais une face tri-hélicoïdale différente ([Figure 17](#)). Il en va de même pour les P5H O, et on peut noter que les faces tri-hélicoïdales ont subi des mutations conduisant à la même surface pour les deux P5H. En revanche, il faut noter que la face bi-hélicoïdale de P5H T est différente de la face bi-hélicoïdale de P5H O, ce qui est une condition nécessaire pour sélectionner des α Reps reconnaissant spécifiquement chaque domaine afin de construire le modulateur allostérique.

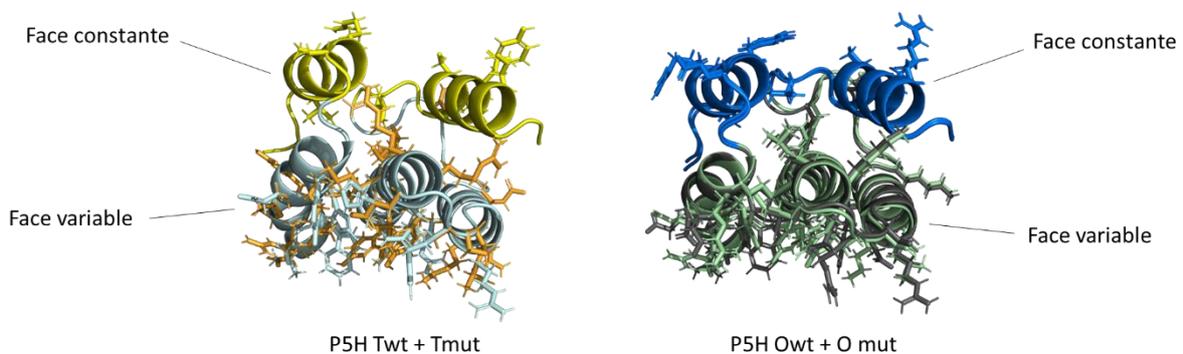


Figure 17 : Superposition des P5H T et des P5H O

Les P5H sauvages ont été superposées à leur mutant. Les positions accessibles de la face bi-hélicoïdale ont été colorées de la même couleur, et les positions accessibles de la face tri-hélicoïdale ont été colorées différemment pour visualiser les 18 mutations simultanées sur les mutants.

Des tests d'expression ont été menés pour ces 4 protéines (P5H Twt1, P5H Tmut1, P5H Owt, et P5H Omut), dans des souches d'*E.coli* M15 à 30°C et 37°C (résultats non montrés). Ces tests ont révélé une surexpression dans la fraction soluble pour P5H Owt et P5H Omut. En revanche, P5H Twt1 s'exprimait dans la fraction insoluble, de même que P5H Tmut1. La séquence codant pour les P5H T a donc été modifiée pour améliorer la stabilité du repliement. Ces modifications ont été conçues en s'aidant de la version *Standalone* du logiciel Foldit, une version très intuitive d'un logiciel de manipulation des structures. Foldit est un logiciel de visualisation et de manipulation des structures initialement développé comme un outil de science participative sous forme de jeu vidéo ([Koepnick, B. et al., 2019](#)). Ce logiciel est dérivé de la suite Rosetta développée dans le groupe de David Baker pour réaliser de la prédiction de structure ou du *design* de protéines ([Rocklin, G. J. et al., 2017](#)). Ces modifications ont globalement consisté à remplacer des résidus hydrophobes exposés au solvant par des résidus hydrophiles. Les protéines ainsi modifiées sont présentées [Figure 18](#) sous les noms « P5H Twt 2 » et

« P5H Tmut 2 ». Ces protéines se sont bien surexprimées dans les fractions solubles des lysats d'*E.coli* M15 à 30°C et à 37°C.

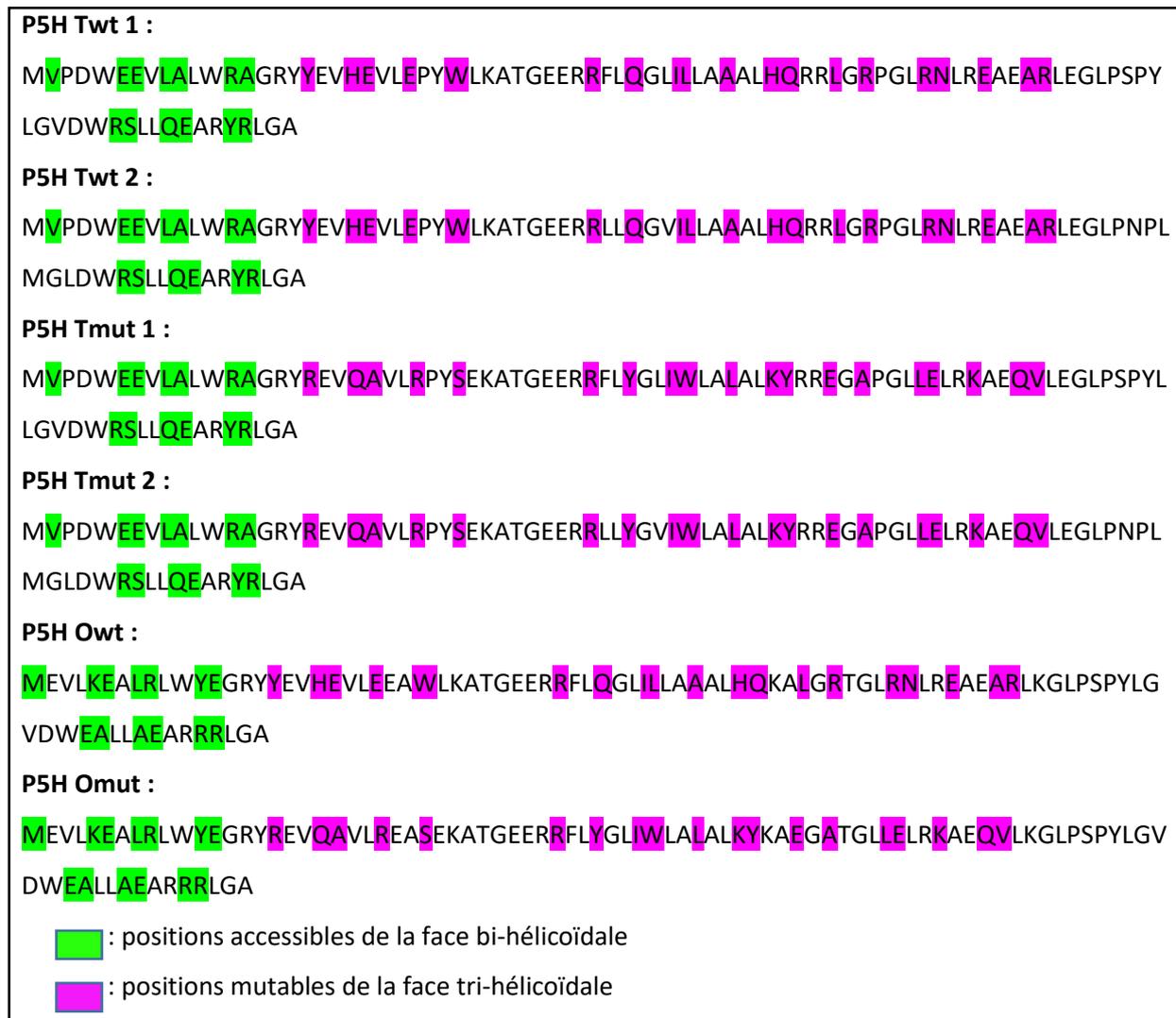


Figure 18 : Séquences peptidiques des pentahélines matrices

Les séquences ci-dessus correspondent aux séquences peptidiques des pentahélines de départ. Les deux P5H T (P5H Twt1 et P5H Tmut1) ont été modifiées car elles s'exprimaient principalement en corps d'inclusion (voir ci-dessous) et leur production a donc dû être optimisée pour augmenter leur solubilité. Les séquences ainsi modifiées ont été renommées P5H Twt 2 et P5H Tmut 2. Les gènes de ces séquences ont été insérés dans un vecteur pQE80L grâce à des sites de restriction BamH I/Hind III.

Ces expériences préliminaires ont apporté des éléments de réponse importants pour la suite du projet. En effet, les P5H mutées simultanément sur 18 positions (sur un total de 92-93 résidus) sont correctement surexprimées dans les fractions solubles. Cela montre la tolérance aux mutations de ce repliement, ce qui est indispensable puisque ces positions exposées sur la face tri-hélicoïdale seront mutées aléatoirement dans la future bibliothèque. De plus, les 4 P5H se sont produites efficacement (20 à 50 mg/litre de culture bactérienne).

Des expériences supplémentaires ont été menées afin d'étudier la stabilité thermique des 4 P5H. Ces données de dénaturation thermique obtenues en thermofluor ont montré un T_m , température à laquelle la moitié des protéines sont dénaturées, autour de 92°C pour les P5H T et non déterminé pour les P5H O. Ces dernières ne présentant pas de transition correspondant à une dénaturation jusqu'à 95°C, seraient peut-être encore plus stables que les P5H T. Des données de SEC-MALS ont indiqué des masses moléculaires observées très proches des masses moléculaires attendues (autour de 12 kDa) pour les 4 pentahélines.

Ces données préliminaires confirment la stabilité des P5H et valident le choix de cette architecture comme base d'une future bibliothèque de protéines artificielle bi-domaines.

II. Production des P5H et caractérisation de leurs propriétés de stabilité

A. Production et purification des P5H

Les 4 pentahélines ont été produites dans des souches d'*E. coli* Gold et purifiées sur résine Ni-NTA puis par chromatographie d'exclusion stérique (matériel et méthode). Les protéines ont été correctement surexprimées (plusieurs milligrammes par litre de culture), et leur pureté a été vérifiée sur gel de polyacrylamide (Figure 19). Ces protéines ont ensuite été utilisées pour des expériences visant à étudier leurs propriétés biophysiques, notamment leur stabilité.

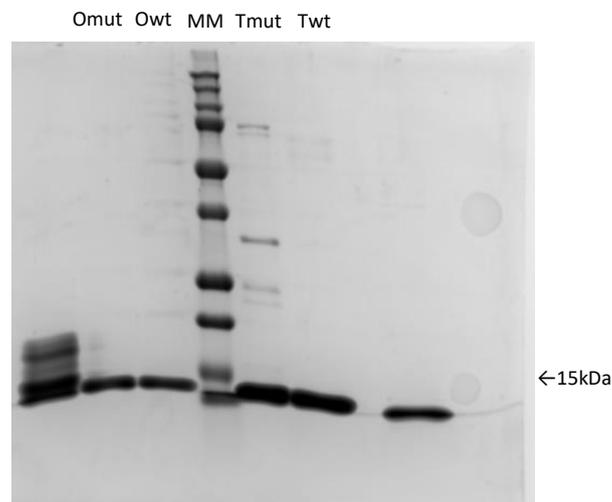


Figure 19 : profil de migration des 4 pentahélines sur gel d'acrylamide

Électrophorèse sur gel de polyacrylamide des 4 pentahélines et de deux protéines non pertinentes pour le projet (puits 1 et 7).

B. Propriétés biophysiques des P5H

Afin de s'assurer du bon repliement et de l'état d'oligomérisation des pentahélines produites, des expériences de SEC-MALS et de dichroïsme circulaire ont été réalisées.

Pour le SEC-MALS, il est nécessaire de concentrer les pentahélines à 2 mg/mL. Ces concentrations ont été atteintes pour toutes les pentahélines, hormis la P5H Twt qui n'a pas pu être concentrée à plus de 0.6 mg/mL. 200 µL de chaque P5H ont ensuite été analysés par SEC-MALS. Les résultats obtenus (Tableau 1) ont montré que les quatre P5H sont monomériques, et ont la masse molaire attendue. Pour les deux P5H O, un rayon hydrodynamique a pu être calculé (2.1 nm ± 0.2 nm pour P5H Owt et 1.8 nm ± 0.2 nm pour P5H Omut). Ce rayon est en accord avec la distance calculée dans pymol entre deux résidus opposés sur la structure (2.19 nm ; résultats non montrés). Tous ces résultats tendent à montrer que les P5H sont des monomères solubles et correctement repliés.

Protéine	Masse molaire théorique (kDa)	Masse molaire observée (kDa)
P5H Twt	14.9	14.9 ± 5.7
P5H Tmut	14.9	15.5 ± 4.3
P5H Owt	14.7	14.6 ± 2.6
P5H Omut	14.6	15.2 ± 0.6

Tableau 1 : Comparaison des masses molaires obtenues par SEC-MALS avec les masses molaires théoriques

Des expériences de dichroïsme circulaire ont alors été réalisées pour étudier la stabilité des quatre pentahélines. En effet, il est crucial que l'architecture servant de base à une bibliothèque de protéines soit la plus stable possible, car les mutations aléatoires introduites pour créer la surface hypervariable d'interaction ont une probabilité non nulle d'être déstabilisantes pour l'ossature protéique.

Pour les 4 protéines, les spectres dichroïques obtenus sont caractéristiques de structures secondaires en hélices α , et leur superposition montre que les 4 P5H ont des structures secondaires similaires (Figure 20.A).

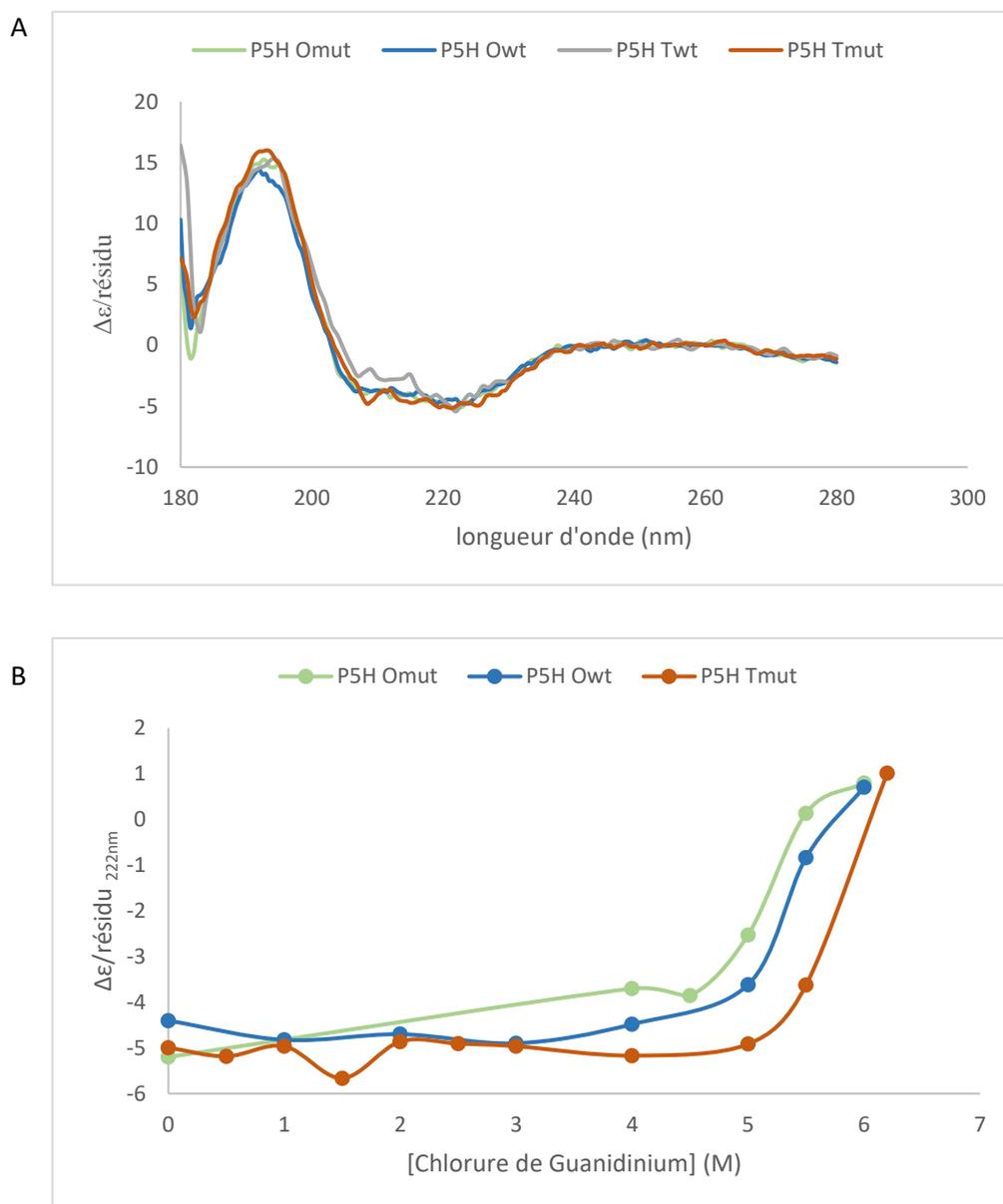


Figure 20 : Dichroïsme circulaire réalisé sur les 4 pentahélines.

(A) : Superposition des spectres dichroïques des différentes pentahélines. (B) : Courbe de dénaturation chimique suivie par dichroïsme circulaire à 222 nm. Une gamme de chlorure de guanidinium a été réalisée sur 3 des 4 pentahélines concentrées à 10 μM . Cette gamme n'a pas pu être réalisée sur P5H Twt en raison de sa trop faible concentration de départ.

Une gamme de chlorure de guanidinium (agent dénaturant des protéines) a ensuite été réalisée sur 3 des 4 pentahélines concentrées à 10 μM finale, afin d'étudier leur stabilité. La P5H Twt n'était pas disponible dans des quantités suffisantes pour réaliser une gamme en présence de chlorure de guanidinium (GdmCl). Le $\Delta\epsilon/\text{résidu}$ à 222 nm (Figure 20.B), caractéristique des hélices α , permet de suivre la dénaturation des P5H en fonction de la concentration en GdmCl. Bien que le plateau de dénaturation ne soit pas tout à fait atteint, on peut constater que la concentration en agent dénaturant

nécessaire pour dénaturer 50% de la protéine (Cm) est supérieure à 5 M de GdmCl, ce qui rend compte d'une très grande stabilité des 3 pentahélines testées.

L'ensemble des données collectées montre que les 4 pentahélines sont solubles, correctement repliées en hélices α , et monomériques. Les données de dichroïsme circulaire montrent également une grande stabilité pour 3 des pentahélines. Pour atteindre le « plateau de dénaturation » des P5H par dichroïsme circulaire, il faudrait réaliser une dénaturation thermique en présence de GdmCl, car la gamme de GdmCl seule n'a pas permis d'atteindre ce plateau.

Ces expériences ont montrées que les P5H étaient une bonne architecture de départ. En effet, bien que 18 positions sur 95 aient été mutées, cela n'a pas destabilisé l'architecture qui devrait donc être compatible avec la diversification requise pour créer une bibliothèque de protéines.

II. Ingénierie des bi-domaines P5H

A. Conception des *linker*

Ces premiers résultats ont permis de valider l'architecture protéique choisie pour réaliser la bibliothèque, puisqu'elle est stable et se produit correctement dans des souches d'*E. coli*.

Pour mener à bien le projet, il est nécessaire d'avoir des protéines bi-domaines. En effet, le fait d'avoir une bibliothèque d'interacteurs bi-domaines présente plusieurs avantages.

D'une part, la stratégie que l'on souhaite développer pour décrocher les interacteurs de leur cible nécessite d'avoir des interacteurs bi-domaines, dont chacun des deux domaines contribue à l'énergie de liaison avec la cible. Ainsi, leur interaction avec le modulateur devrait, en décrochant un des deux domaines, diminuer l'énergie de liaison suffisamment pour que la constante de dissociation K_d de l'interaction augmente et conduise l'interacteur à se décrocher de sa cible.

D'autre part, un interacteur bi-domaines a une affinité apparente plus forte pour sa cible grâce au phénomène d'avidité. En effet, il faut que les deux domaines d'un interacteur bi-domaines se décrochent simultanément pour qu'il y ait dissociation entre l'interacteur et sa cible. Or, la fixation d'un des deux domaines entraîne une augmentation de la concentration effective de l'autre domaine localement et favorise son interaction avec la cible. Il est donc plus difficile de décrocher un interacteur possédant plusieurs sites de fixations sur une cible qu'un interacteur ne possédant qu'un seul site de fixation. De plus, en fonction de la nature du lien peptidique entre les deux domaines, il est possible de donner plus ou moins de flexibilité à l'architecture protéique afin de lui permettre « d'explorer » un plus grand espace des conformations. Si les surfaces d'interaction sur chacun des domaines parviennent à interagir simultanément avec une même cible, une telle bibliothèque de protéines pourrait donner des interacteurs possédant une meilleure affinité qu'une bibliothèque mono-domaine.

L'objectif est donc de générer les P5H bi-domaines (bi-P5H) en décidant d'un protocole d'assemblage de la bibliothèque qu'il faudra tester. Comme discuté précédemment, le choix du lien entre les deux P5H aura une incidence directe sur les propriétés physiques des protéines de la bibliothèque. Dans notre cas, nous avons choisi un lien peptidique semi-flexible, car nous voulions de la flexibilité pour permettre aux deux domaines d'explorer un espace des conformations diversifié. Cependant, nous voulions également garder une certaine distance entre les deux domaines pour éviter que les épitopes reconnus soient trop proches spatialement, et que la fixation du modulateur n'entraîne pas de décrochage de l'interacteur.

C'est dans cette optique que j'ai conçu un premier connecteur peptidique de 14 acides aminés (Figure 21), en me basant sur les travaux de George et Heringa (George, R. A. & Heringa, J., 2002). Ces travaux regroupent des données statistiques sur les propensions de chaque acide aminé à être dans un connecteur de type boucle, hélice, ou brin bêta. La séquence du premier connecteur contient deux prolines, avec pour objectif de rigidifier le connecteur peptidique, et de défavoriser la formation de structures secondaires. Le résidu en N-ter de chaque proline a une grande importance car il influe sur la conformation de la proline. En effet, la proline peut prendre une conformation *cis* ou *trans*, et selon sa conformation, son rôle dans le connecteur sera très différent. Une *cis*-proline provoquerait un virage à 180° dans la chaîne polypeptidique et rapprocherait les deux domaines, ce qui va à l'encontre de notre objectif. L'idée est donc de choisir des résidus en N-ter de chaque proline qui tendent à favoriser au maximum leur conformation *trans*. Pour cela, je me suis servi des données présentées dans l'article de George et Heringa, et j'ai conçu deux connecteurs peptidiques : un de 12 acides aminés et l'autre de 28 acides aminés (Figure 21).

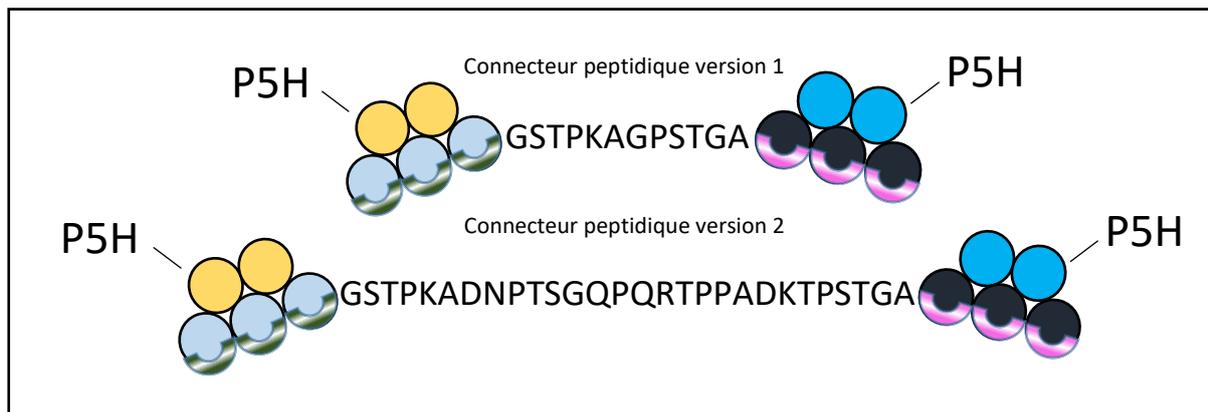


Figure 21 : Séquences des connecteurs peptidiques utilisés dans la bibliothèque de bi-P5H.

Les bi-P5H ainsi conçues ont été testées pour leur expression et leur production dans des souches adéquates, afin de vérifier que la fusion des deux P5H n'avait pas d'effets délétères sur leur production et leur stabilité.

B. Test expression / Production bi-P5H

L'objectif est de vérifier la production soluble des bi-P5H par les bactéries. Pour cela, des tests d'expression ont été réalisés, sur Twt-Owt dans un premier temps. Un test d'expression consiste à produire la protéine d'intérêt dans différentes conditions pour cribler la condition dans laquelle la protéine est la plus surexprimée de façon soluble. Dans notre cas, nous avons décidé de tester différentes souches d'*E.coli* : Gold, M15, et Rosetta, et différentes températures. Les photos des gels sont présentées ci-dessous (Figure 22).

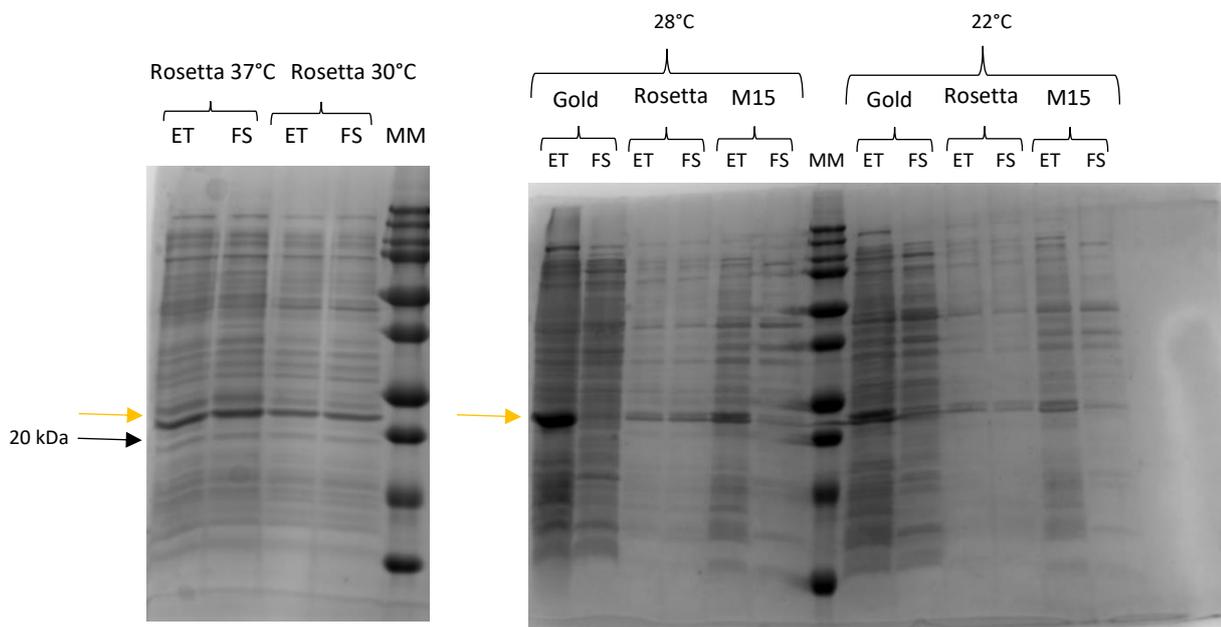


Figure 22 : Tests d'expression de la bi-P5H Twt-Owt dans Gold, Rosetta, ou M15 à différentes températures.

Des tests d'expression ont été réalisés sur la bi-P5H Twt-Owt afin de déterminer la condition permettant d'obtenir le plus de protéine soluble. Pour cela, différentes souches d'*E.coli* ont été utilisées pour produire la protéine Twt-Owt. Les cultures ont été lysées, et une partie de ces lysats correspondant aux extraits totaux (ET) ont été déposés sur gel. Le reste des lysats a été centrifugé pour séparer les fragments de membranes et les agrégats de la fraction soluble (FS). L'objectif d'un test d'expression est donc de trouver la condition dans laquelle la fraction soluble contient le plus de protéine d'intérêt. Ici, la souche Rosetta à 37°C est celle qui permet d'obtenir le plus de Twt-Owt soluble.

La souche Rosetta a été retenue car l'expression de la bi-P5H (indiquée par la flèche jaune) y est soluble (taille attendue : 23.6 kDa). À 37°C, le niveau de production est satisfaisant. La souche Gold est celle qui surexprime le plus la bi-P5H mais son expression se fait majoritairement en corps d'inclusion.

Ces expériences permettent de montrer que Twt-Owt se produit bien de façon soluble dans les bactéries. La prochaine étape consiste donc à la produire et à la purifier.

La protéine Twt-Owt a donc été produite et purifiée sur colonne Ni-NTA puis par chromatographie d'exclusion de taille (Figure 23).

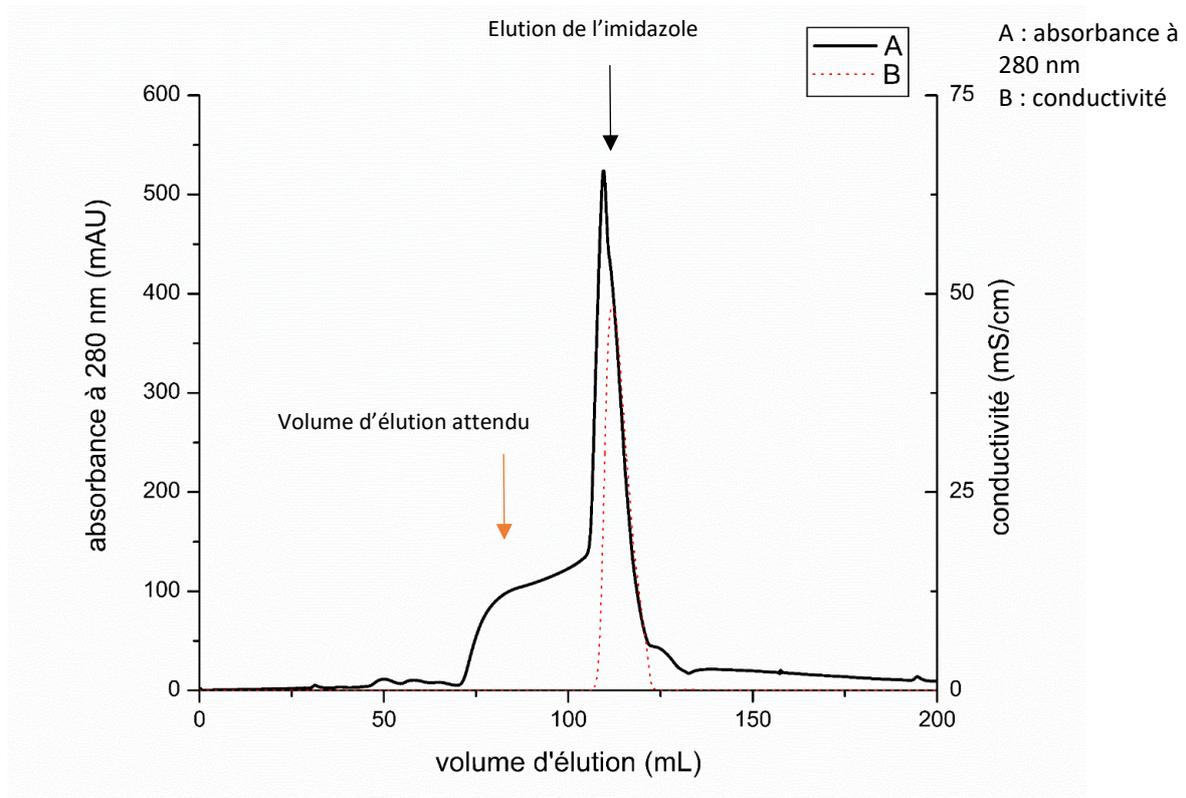


Figure 23 : Profil d'élution de la bi-P5H Twt-Owt par chromatographie d'exclusion de taille.

La protéine Twt-Owt a été produite et purifiée par chromatographie d'affinité sur une colonne Ni-NTA. Les fractions les plus concentrées de cette première étape de purification ont été purifiées par chromatographie d'exclusion stérique dans un tampon Tris 20 mM NaCl 150 mM à un pH de 7.5.

La Figure 23 montre que le profil d'élution atypique de Twt-Owt, très étiré, avec un pic principal sortant en même temps que l'imidazole (dont l'élution est indiquée par la courbe orange qui représente la conductivité de l'éluât). Différentes fractions d'élution ont été déposées sur gel afin de vérifier leur composition (Figure 24).

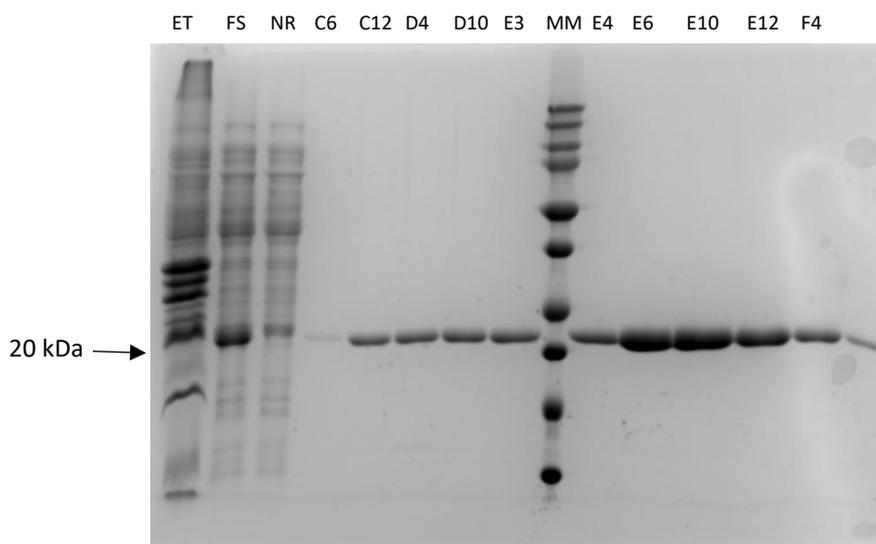


Figure 24 : Contrôle de la chromatographie d'exclusion de taille de Twt-Owt

Différentes fractions d'élutions de la chromatographie d'exclusion de taille présentée Figure 23 ont été déposés sur gel afin d'en contrôler le contenu. On remarque que le profil très étendu correspond bien au bi-domaine qui semble assez pur dans toutes les fractions déposées.

La [Figure 24](#) montre que la protéine est pure dans toutes les fractions d'élution. Le profil d'élution représenté [Figure 23](#) correspond donc bien à celui de la bi-P5H Twt-Owt. Cependant, l'allure du profil d'élution est anormale et l'hypothèse est que cette protéine est retenue par la résine au cours de l'élution. Plusieurs hypothèses peuvent être émises pour l'expliquer, comme la rétention de la protéine sur la résine composant la colonne. Pour tester ces différentes hypothèses, la même purification a été réalisée dans plusieurs tampons afin d'obtenir le pic le plus étroit et symétrique possible à l'issue de la chromatographie d'exclusion stérique.

Dans cette optique, différentes compositions de tampon ont été testées pour optimiser la purification, à la fois au niveau de la séparation sur la colonne Ni-NTA et au niveau de la séparation par chromatographie d'exclusion stérique. Les différents tampons testés contenaient des concentrations croissantes de sels, dans le but d'inhiber d'éventuelles interactions de type électrostatique avec la matrice des colonnes, ainsi que du glycérol. L'ajout de glycérol dans le tampon doit permettre de masquer d'éventuelles régions hydrophobes exposées de la protéine, grâce aux propriétés amphiphiles du glycérol qui ont été discutées avant. La résine superdex est composée de dextran, un polysaccharide dérivé du glucose, sur lequel est fixé covalamment de l'agarose, un autre polysaccharide.

Les profils d'élution de différentes chromatographies d'exclusion stérique sont représentés [Figure 25](#).

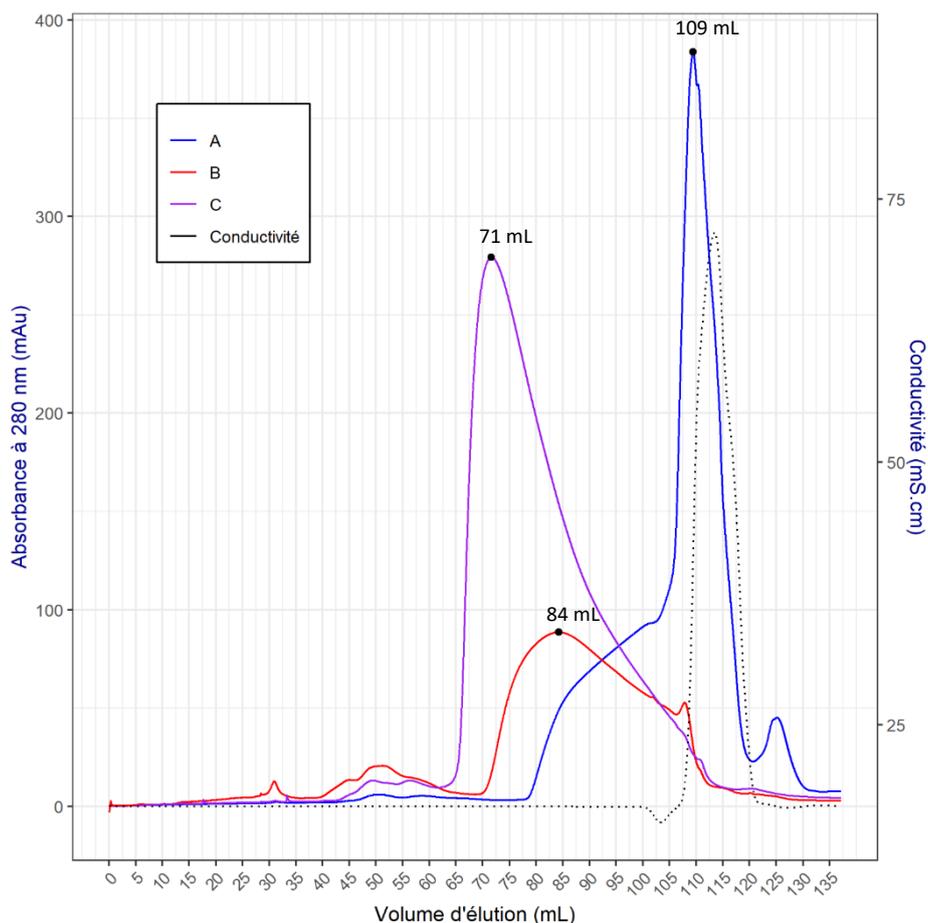


Figure 25 : Optimisation de la purification de la bi-P5H Twt-Owt par chromatographie d'exclusion de taille.

La bi-P5H Twt-Owt issue de la purification sur colonne Ni-NTA dans différents tampons a été injectée sur une colonne s75 16/600. Ces tampons contenaient 300 mM d'imidazole, 50 mM de phosphate, mais variaient au niveau de la concentration en sels : (A) 1 M NaCl (B) 2 M NaCl (C) 1 M NaCl + 5 % glycérol. Les purifications par chromatographie d'exclusion stérique ont été réalisées dans un tampon **20 mM Tris 150 mM NaCl** (A et B) ou dans un tampon **20 mM Tris 300 mM NaCl 5% glycérol** (C).

On peut voir que selon le tampon dans lequel était stockée la protéine avant la chromatographie d'exclusion stérique, le volume d'élution change. Plus la concentration en sels est élevée avant la SEC, plus le volume d'élution est faible (courbe rouge Figure 25). Cette expérience suggère que la protéine interagit avec la résine, probablement par des interactions de type électrostatiques et que ces interactions peuvent être inhibées par l'augmentation de la force ionique du tampon. Afin d'obtenir un profil plus fin, une nouvelle purification a été réalisée, en augmentant la concentration de NaCl dans le tampon d'élution de la SEC et en y ajoutant du glycérol (courbe violette Figure 25).

Les changements apportés au tampon d'élution de la SEC ont dissocié l'élution de la protéine de celle de l'imidazole, et le profil d'élution obtenu est bien plus resserré autour de son maximum d'absorbance à 280 nm. Cela semble donc confirmer l'hypothèse selon laquelle la protéine interagit avec la résine superdex par des interactions de type électrostatique.

Les deux bi-P5H test (Twt-Owt et Owt-Twt) ont pu être produites et purifiées de façon efficace grâce à ce protocole optimisé, qui consiste à réaliser la purification par Ni-NTA dans un tampon contenant 1 M NaCl et 5% glycérol, et la SEC dans un tampon contenant 300 mM NaCl et 5% glycérol. Une autre particularité des P5H observée lors de ces expériences avec les mono-domaines comme les bi-domaines, est leur tendance à précipiter au froid. Cette précipitation est cependant réversible en quelques minutes à température ambiante.

C. Test d'exposition sur phages des bi-P5H

Lors des sélections par phage *display*, les variants de la bibliothèque sont exposés à la surface de phages filamenteux M13. Pour cela, les gènes codant pour les protéines de la bibliothèque sont fusionnés par leur extrémité C-terminale à une protéine de la capside du phage, la protéine pIII (Figure 26). De cette façon, chaque variant de la bibliothèque est exposé à la surface des phages, ce qui permet par exemple de les présenter à des cibles d'intérêt en vue de réaliser les sélections d'interacteurs.

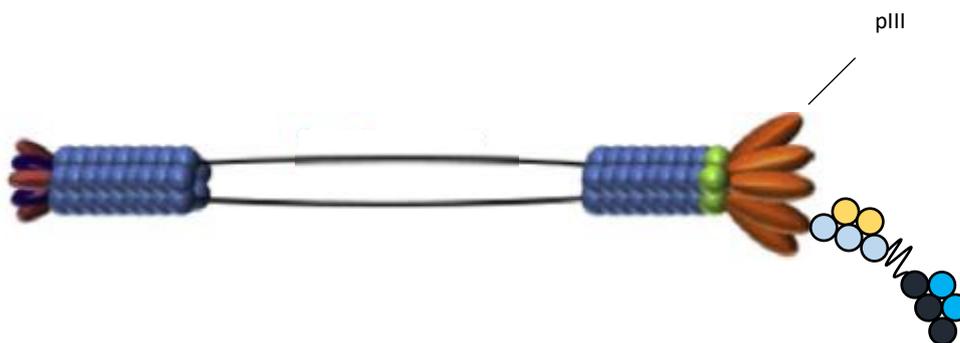


Figure 26 : représentation d'un phage exposant une protéine *via* la protéine pIII

La première étape est donc de vérifier si les bi-P5H sont efficacement exposées à la surface des phages M13. C'est dans cette optique que des tests d'exposition sur phage ont été réalisés. Un test d'exposition sur phage consiste à évaluer la proportion des phages de la bibliothèque exposant la protéine d'intérêt. Le taux d'exposition sur phages est en effet un paramètre très important pour le processus de sélection par phage *display*. Si les phages exposent trop peu la protéine d'intérêt, la sélection risque d'échouer car pas assez de variants de la bibliothèque auront été mis en contact de la cible. À l'inverse, si les phages exposent plus qu'une protéine, le phénomène d'avidité généré par cette exposition multiple risque d'entraîner la sélection d'interacteurs peu affins pour la cible. Pour éviter ce problème, on souhaite généralement obtenir un taux d'exposition compris entre 1% et 0.001%, soit un variant exposé pour 10^2 à 10^5 phages. Ces tests ont été réalisés de deux façons : par ELISA sur plaque, ou par comptage de bactéries sur boîtes.

Le test d'exposition par ELISA sur plaque repose sur le même principe qu'un test ELISA (Figure 27). Les puits sont recouverts par un anticorps capable de capturer les phages exposants les protéines d'intérêt étiquetées par une séquence reconnue par l'anticorps (ici *via* un tag d'affinité de type FLAG ou HA). La capture est alors vérifiée par la détection des phages M13 dans les puits grâce à un anticorps anti-phages M13 couplée à une peroxydase (Figure 27).

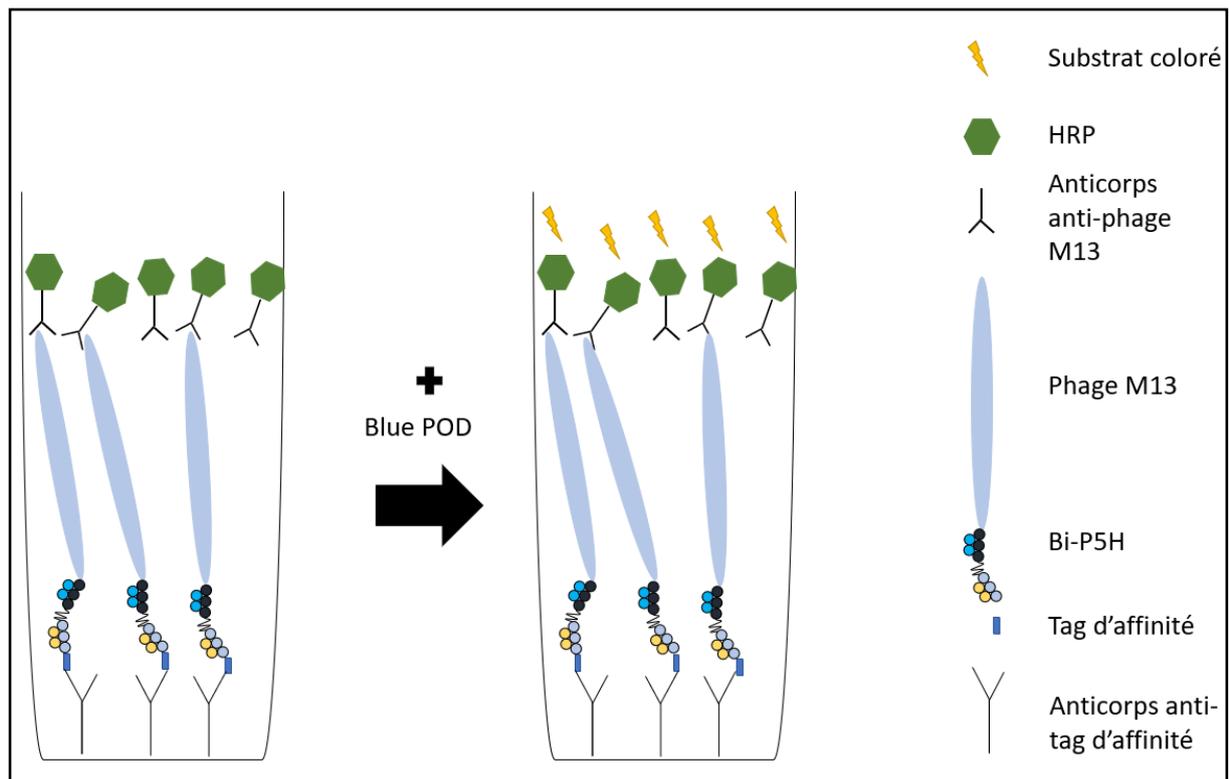


Figure 27 : Principe d'un phage ELISA clonal

Dans cette expérience, les phages exposent un variant de la bibliothèque qui possède un *tag* d'affinité. Des puits sont recouverts avec un anticorps spécifique de ce tag d'affinité, qui sont alors capables de capturer les phages si ces derniers exposent un variant de la bibliothèque. La présence de phages au fond des puits est vérifiée par ajout d'un anticorps spécifique de l'autre extrémité des phages. Ce deuxième anticorps est couplé à une peroxydase, ce qui permet de révéler par une mesure d'absorbance la présence de l'anticorps (et donc du phage) par ajout de Blue POD, le substrat coloré de la peroxydase.

Dans notre cas, nous avons voulu comparer l'efficacité de capture par les anticorps anti-FLAG-tag par rapport à la capture par les anticorps anti-HA-tag. En effet, les bibliothèques déjà développées par le laboratoire présentent un tag d'affinité de type FLAG. Nous avons cherché à savoir si l'affinité entre l'anticorps anti-FLAG-tag et le FLAG-tag était suffisante pour que les tests d'expression rendent compte du taux d'exposition réel des variants de la bibliothèque à la surface des phages. En effet, si l'affinité entre l'anticorps anti-tag et le tag est trop faible, des phages exposant la protéine d'intérêt pourraient ne pas être capturés, et donc ne pas être pris en compte dans le calcul du taux d'exposition. Dans la littérature ([Ranawakage, D. C. et al., 2019](#)), nous avons vu qu'un anticorps commercial qui reconnaît le HA-tag possède une meilleure affinité pour son tag que plusieurs anticorps anti-FLAG-tag. Afin de comparer l'efficacité de capture avec les deux tags, des phages exposants une bi-P5H (Twt-Owt) précédée soit d'un FLAG-tag, soit d'un HA-tag ont été produits. Des phages exposant l' α Rep A8 (une des α Reps sélectionnées contre P5H T) précédée d'un FLAG-tag ont également été produits pour servir

de contrôle, puisque le taux d'exposition des α Reps est connu. Les résultats de ces tests d'exposition sont présentés [Figure 28](#).

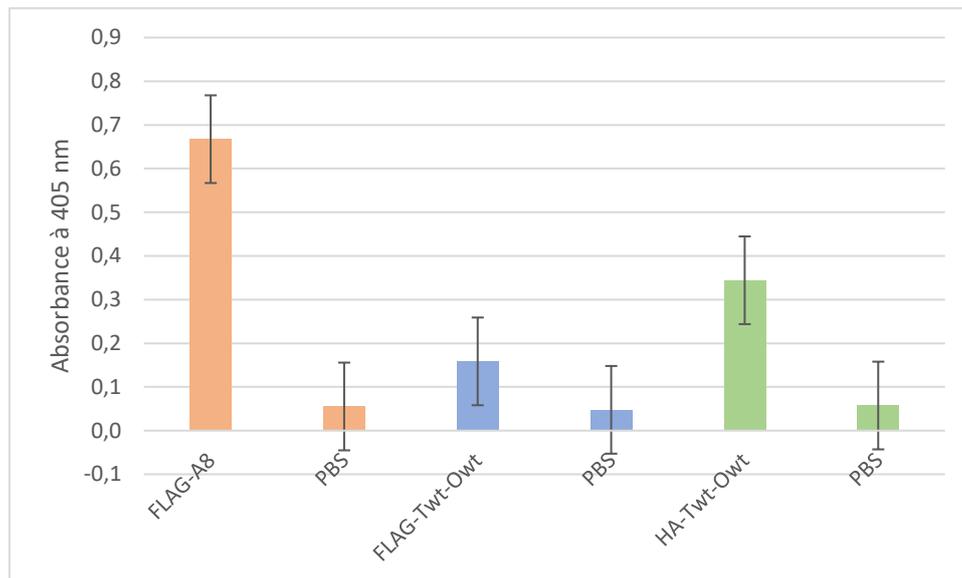


Figure 28 : Test d'exposition sur phages de la bi-P5H Twt-Owt par ELISA

On peut voir [Figure 28](#) que deux fois plus de phages exposant la bi-P5H Twt-Owt sont capturés lorsque celle-ci présente un **HA-tag** à la place d'un **FLAG-tag**. Cette expérience montre donc que le HA-tag est préférable au FLAG-tag pour évaluer la quantité de phages exposants les variants de la bibliothèque, puisqu'il rendra mieux compte du nombre de phages qui présentent un variant à leur surface.

Parallèlement, des tests d'exposition sur phages par comptage de bactéries sur boîtes ont été réalisés. Ce deuxième type de test d'exposition permet d'avoir accès au nombre de phages déposés dans chaque puit et au nombre de phages capturés, ce qui permet de calculer un taux d'exposition. En effet, au lieu de révéler la présence des phages avec un anticorps anti-phages M13, les phages capturés ont été élués et différentes dilutions de ces éluats ont servi à infecter des bactéries. Les bactéries infectées ont été mises en culture sur milieu solide en présence d'un antibiotique dont la résistance est portée par le phagemide. En comptant le nombre de colonies sur chaque boîte, il est donc possible de connaître le nombre de phages présents dans chaque éluat, puisqu'en théorie une bactérie ne peut être infectée que par un phage. Les résultats de ce test sont présentés ci-dessous ([Tableau 2](#)).

	10^{-1}	10^{-2}	10^{-3}	10^{-8}	10^{-9}
Input FLAG-A8				218	22
Input FLAG-Twt-Owt				140	13
Input HA-Twt-Owt				161	16
Output FLAG-A8	$> 10^3$	$> 10^3$	224		
Output FLAG-Twt-Owt	796	49	3		
Output HA-Twt-Owt	tapis	tapis	$> 10^3$		

Tableau 2 : Test d'exposition sur phages de la bi-P5H Twt-Owt par infection de bactéries

Ce deuxième test d'exposition a permis de confirmer que le HA-tag était plus efficace pour la capture des phages que le FLAG-tag. En effet, les bactéries ne sont pas dénombrables ($>10^3$) car trop nombreuses, lorsqu'elles ont été infectées par les phages exposant la bi-P5H Twt-Owt précédée d'un HA-tag. En revanche, on peut compter les bactéries infectées par des phages exposant la même bi-P5H précédée d'un FLAG-tag, soit environ 8000 bactéries infectées par des phages pour 1.4×10^{10} phages mis en contact de la culture bactérienne.

Le taux d'exposition a été calculé à partir des comptages sur boîtes selon la formule $Output/Input \times 100$.

FLAG-A8 : $Output/Input \times 100 = 1.02 \times 10^{-5}$ soit un phage exposant pour 100000 phages

FLAG-TO : 5.7×10^{-5} soit un phage exposant pour 570000 phages

HA-TO : Plus que FLAG-A8 mais impossibilité de compter les bactéries.

Ce test d'exposition a permis de montrer que peu importe le tag d'affinité, la bi-P5H est bien exposée à la surface des phages. On parvient donc à exposer correctement une protéine bi-domaine qui contient un *linker* flexible, malgré la complexité que cela apporte par rapport à une protéine mono-domaine. De plus, cette expérience permet également de vérifier que la protéine exposée est entière (du N-ter au C-ter). En effet, la protéine est exposée à la surface des phages par son côté C-ter, qui est fusionné à la protéine P3. Comme elle est capturée par son tag qui est localisé en N-ter de sa séquence, les phages capturés exposent nécessairement des protéines entières. Nous avons choisi d'utiliser le HA-tag pour la bibliothèque de bi-P5H puisqu'il rend mieux compte du nombre de phages exposants les P5H.

IV. Stratégie d'assemblage de la bibliothèque

L'architecture choisie au départ s'est avérée compatible avec l'ensemble des critères préliminaires essentiels à la création d'une bibliothèque. En effet, nous avons vérifié la surexpression des bi-P5H, leur solubilité, et leur exposition sur phages. La suite du projet a donc consisté à créer une bibliothèque de bi-P5H compatible avec le phage *display*. Pour créer une bibliothèque de protéines, la première étape consiste à choisir une ossature protéique, et des positions à diversifier pour créer la surface hypervariable d'interaction. Nous avons montré que les pentahélines possédaient toutes les qualités requises pour créer une telle bibliothèque, étant donné leur stabilité ainsi que leur tolérance aux mutations. L'étape suivante est alors de choisir la fréquence souhaitée pour chaque acide aminé à chaque position diversifiée. La séquence nucléotidique codant pour les variants est découpée en plusieurs oligonucléotides, dont certains sont dégénérés afin de coder pour les positions hypervariables. Ces oligonucléotides doivent alors être assemblés pour reconstituer une séquence codant pour une protéine entière. Une fois la séquence complète reconstituée, une étape d'amplification est souhaitable pour s'assurer qu'on dispose d'assez de matériel pour les étapes suivantes. Il est ensuite nécessaire d'insérer les gènes codants pour les variants de la bibliothèque dans le phagemide (pour les bibliothèques compatibles avec le phage *display*), le vecteur qui permet d'exprimer les protéines de la bibliothèque en fusion avec la protéine pIII du phage M13. La dernière étape consiste alors à transformer le plus grand nombre de bactéries possibles avec les phagemides codants pour les variants de la bibliothèque, dans le but d'obtenir la bibliothèque sous forme de bactéries infectées par des phages exposants chaque variant de la bibliothèque. Ces différentes étapes sont synthétisées [Figure 29](#).

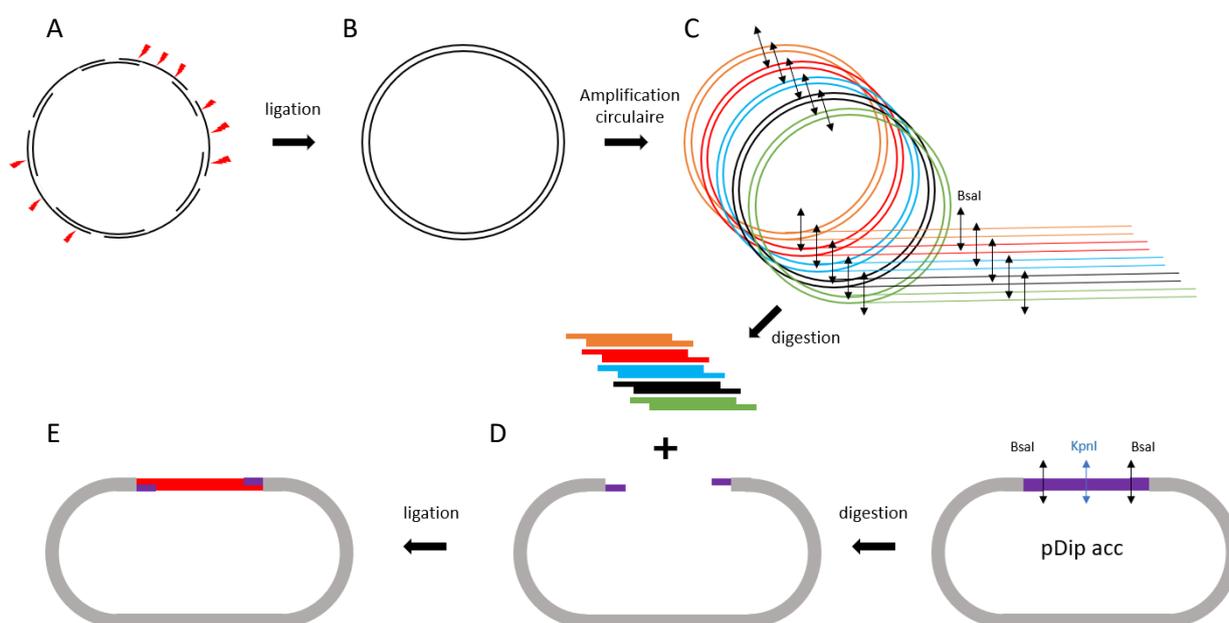


Figure 29 : schéma de la stratégie d'assemblage des oligonucléotides codants pour les variants de la bibliothèque.

Les oligonucléotides codants pour une bi-P5H entière sont assemblés grâce à la complémentarité des bases entre les brins sens et anti-sens (A). Une étape de ligation permet d'obtenir des cercles d'ADN double brins qui contiennent la séquence des variants bi-P5H de la bibliothèque (B). Ces cercles sont ensuite amplifiés par RCA (C), et digérés par l'enzyme Bsa I. Simultanément, le vecteur accepteur de la bibliothèque est également digéré par Bsa I. Les inserts et le vecteur digéré possèdent des extrémités complémentaires, ce qui permet l'insertion des séquences codants pour les bi-P5H dans le vecteur accepteur (phagemide)(D). Une étape de ligation permet d'obtenir les inserts correctement insérés dans le phagemide en vue de transformer des bactéries pour constituer la bibliothèque.

A. Diversification des bi-P5H

Les positions sélectionnées pour être diversifiées et ainsi former une surface hypervariable d'interaction sont presque toutes les positions présentées initialement [Figure 13](#) pour chacun des domaines. Deux positions ont été retirées des positions diversifiées pour des raisons qui seront expliquées dans la sous-partie suivante. Il reste 16 positions diversifiées par domaine, soit 32 positions diversifiées dans la bibliothèque de bi-P5H et pour lesquelles il a fallu choisir la distribution en acides aminés. Notre choix s'est porté sur une répartition équivalente aux résidus qu'on retrouve sur les faces accessibles des hélices α ([Guellouz, A. et al., 2013](#)), en excluant prolines et cystéines et en augmentant la représentation des tyrosines. Les prolines ont été exclues pour éviter de casser la structure des hélices α , et les cystéines ont été exclues pour éviter la formation de ponts disulfures. Les tyrosines

ont été surreprésentées car elles sont régulièrement impliquées dans les interactions et font figure de *hot spots* d'interaction (cf. Introduction). On obtient donc la répartition présentée [Tableau 3](#).

A	Ala	12.0%
C	Cys	
D	Asp	4.0%
E	Glu	12.0%
F	Phe	3.0%
G	Gly	4.0%
H	His	2.0%
I	Ile	2.0%
K	Lys	8.0%
L	Leu	5.0%
M	Met	3.0%
N	Asn	4.0%
P	Pro	
Q	Gln	7.0%
R	Arg	9.0%
S	Ser	7.0%
T	Thr	2.0%
V	Val	2.0%
W	Trp	4.0%
Y	Tyr	10%
S U M		100%

Tableau 3 : Fréquences des acides aminés à chaque position diversifiée de la bibliothèque de bi-P5H

Ce tableau représente la proportion souhaitée pour chaque acide aminé à chaque position diversifiée dans la bibliothèque de bi-P5H.

La stratégie employée a été d'assembler des oligonucléotides dégénérés porteurs des mutations, afin d'obtenir directement les gènes codants pour chaque variant, en misant sur l'appariement aléatoire de ces oligonucléotides.

B. Assemblage des oligonucléotides dégénérés

Les oligonucléotides porteurs des mutations ont été générés grâce à la technologie des TRIMers (Yagodkin, A. *et al.*, 2007). La synthèse des oligonucléotides est arrêtée à chaque position qu'on souhaite diversifier. Le mélange réactionnel est alors mis en présence d'un ensemble de codons codants pour les 18 acides aminés désirés et dans les proportions souhaitées (Tableau 3). Cette méthode permet donc d'insérer de façon aléatoire les résidus qu'on souhaite, dans des proportions contrôlées. Les oligonucléotides ont été conçus selon le schéma présenté Figure 30.

Oligonucléotides sens	FOR 1													FOR 2																						
Séquence codante	TTA	TCT	GGT	CTC	A	GCA	AGT	TAT	CCA	TAT	GAT	GTT	CCA	GAT	TAT	GCT	GAA	GAC	TCC	GTT	CCC	GAT	TGG	GAA	GAG	GTA	CTA	GCT	TTA	TGG	CGT	GCG	GGT	CGC	TAT	
Séquence en acides aminés					A	S	Y	P	Y	D	V	P	D	Y	A	E	D	S	V	P	D	W	E	E	V	L	A	L	W	R	A	G	R	Y		
Séquence non codante	AAT	AGA	CCA	GAG	T	CGC	TCA	ATA	GGT	ATA	CTA	CAA	GGT	CTA	ATA	CGA	CTT	CTG	AGG	CAA	GGG	CTA	ACC	CTT	CTC	CAT	GAT	CGA	AAT	ACC	GCA	CGC	CCA	GCG	ATA	
Oligonucléotides anti-sens	REV 13				REV 1													REV 2																		
	FOR 3													FOR 4																						
	XXX	GAA	GTT	XXX	XXX	GTG	CTG	XXX	CCG	TAT	XXX	XXX	AAG	GCC	ACG	GGT	GAA	GAG	CGC	CGG	CTT	CTC	XXX	GGT	GTA	ATT	XXX	CTG	GCG	XXX	GCC	CTG	XXX	XXX	CGT	
	X	E	V	X	X	V	L	X	P	Y	X	X	K	A	T	G	E	E	R	R	L	L	X	G	V	I	X	L	A	X	A	L	X	X	R	Y
													TTC	CGG	TGC	CCA	CTT	CTC	GCG	GCC	GAA	GAG														GCA
	FOR 5													FOR 6																						
	CGC	TTG	GGG	CGC	CCA	GGC	CTC	XXX	XXX	TTA	CGC	XXX	GCG	GAA	XXX	XXX	CTG	GAA	GGC	TTA	CCA	AAT	CCG	CTT	ATG	GGT	CTG	GAT	TGG	CGC	AGC	CTG	CTT	CAG	GAA	GCC
	R	L	G	R	P	G	L	X	X	L	R	X	A	E	X	X	L	E	G	L	P	N	P	L	M	G	L	D	W	R	S	L	L	Q	E	A
	GCG	AAC	CCC	GCA	GGT	CCC	GAG										GAC	CTT	CCG	AAT	GGT	TTA	GGC	GAA	TAC	CCA	GAC	CTA	ACC	GCG	TCG	GAC	GAA	GTC	CTT	CGG
	REV 4				REV 3													REV 5																		
	FOR 7													FOR 8																						
	CGT	TAT	CGC	CTG	GGC	GCA	GCG	TCT	TCC	GGA	TCT	ACA	CCT	AAA	GCG	GAT	AAC	CCA	ACT	AGC	GCC	CAG	CCG	CAA	CGT	ACG	CCA	CCG	GCG	GAT	AAA	ACC	CCG	TCC	ACG	GGG
	R	Y	R	L	G	A	A	S	S	G	S	T	P	K	A	D	N	P	T	S	G	Q	P	Q	R	T	P	P	A	D	K	T	P	S	T	G
	GCA	ATA	GCG	GAC	CCG	CGT	CGC	AGA	AGG	CCT	AGA	TGT	GGA	TTT	CGC	CTA	TTG	GGT	TGA	TCG	CCG	GTC	GGC	GTT	GCA	TGC	GGT	GCC	CGC	CTA	TTT	TGG	GGC	AGG	TGC	CCC
	REV 6						REV 7													REV 8																
	GCT	TCG	GAG	GTA	CTA	AAG	GAA	GCT	CTA	AGG	CTT	TGG	TAC	GAA	GGC	CGT	TAC	XXX	GAA	GTT	XXX	XXX	GTG	CTG	XXX	GAG	GCG	XXX	XXX	AAA	GCA	ACT	GGC	GAG	GAA	CGT
	A	S	E	V	L	K	E	A	L	R	L	W	Y	E	G	R	Y	X	E	V	X	X	V	L	X	E	A	X	X	K	A	T	G	E	E	R
	CGA	AGC	CTC	CAT	GAT	TTT	CGT	GAT	TCC	GAA	ACC	ATG	CTT	CCG	GCA	ATG												TTT	CGT	TGA	CCG	CTC	CTT	GCA		
	REV 9													REV 10																						
	CGC	TTT	TTA	XXX	GGT	CTG	ATT	XXX	CTG	GCA	XXX	GCG	CTG	XXX	XXX	AAG	GCG	CTG	GGC	CGT	ACG	GGT	CTG	XXX	XXX	TTA	CGC	XXX	GCG	GAA	XXX	XXX	CTC	AAG	GGG	CTG
	R	F	L	X	G	L	I	X	L	A	X	A	L	X	X	K	A	L	G	R	T	G	L	X	X	L	R	X	A	E	X	X	L	K	G	L
	GCG	AAG	AAT													TTC	CGC	GAC	CCG	GCA	TGC	CCA	GAC							GAG	TTC	CCC	GAC			
	FOR 11													FOR 12																						
	CCG	TCT	CCA	TAT	TTA	GGC	GTT	GAC	TGG	GAA	GCG	CTT	TTA	GCA	GAG	GCA	CGC	CGT	CCG	CTT	GGT	GCC	GGC	TGG	AGA	CCC	TGT									
	P	S	P	Y	L	G	V	D	W	E	A	L	L	A	E	A	R	R	R	L	G	A	G	W	R	P	C									
	GCG	AGA	GGT	ATA	AAT	CCG	CAA	CTG	ACC	CTT	CGC	GAA	AAT	CGT	CTC	CGT	GCG	GCA	GCC	GAA	CCA	CGG	CCG	ACC	TCT	GGG	ACA									
	REV 12													REV 13																						

Figure 30 : Conception des oligonucléotides à assembler pour créer la bibliothèque de bi-P5H

Schéma des oligonucléotides à assembler pour créer la bibliothèque de bi-P5H et leur séquence correspondante. Les sites de coupure aux enzymes de restriction Bsa I (5'-GGTCTC-3') et Bbs I (5'-GAAGAC-3') sont représentés en **gras**, et les extrémités générées par la coupure par l'une ou l'autre enzyme sont représentées en **bleu** (Bsa I) ou en **vert** (Bbs I). Les positions diversifiées sont marquées par un codon XXX ou un acide aminé X, et les positions non diversifiées par rapport aux mutants de départ (P5H Tmut et P5H Omut) sont représentées en **rouge**.

La stratégie employée a été d'assembler les oligonucléotides du brin sens avec les oligonucléotides complémentaires du brin anti-sens. Trois oligonucléotides portent les mutations pour chaque domaine : les oligonucléotides notés FOR3, FOR4, et FOR5 pour le domaine T et les oligonucléotides notés FOR10, FOR11 et FOR12 pour le domaine O. Ces oligonucléotides sont assemblés grâce à leurs extrémités qui sont constantes, en utilisant un oligonucléotide anti-sens chevauchant les deux extrémités des oligonucléotides sens juxtaposés. Deux positions initialement candidates à la

diversification ont été rendues constantes (acides aminés en rouge Figure 30) pour permettre l'appariement des fragments FOR4 et FOR5 d'une part, et des fragments FOR11 et FOR12 d'autre part. En effet, le nombre de bases constantes entre ces fragments était trop faible pour concevoir les oligonucléotides anti-sens suffisamment chevauchant.

Pour optimiser un appariement correct, l'assemblage des oligonucléotides s'est fait par étapes. Tout d'abord, les oligonucléotides ont été assemblés en blocs de trois oligonucléotides consécutifs, puis ces blocs ont été assemblés entre eux (Figure 31) et ensuite ligués pour former des cercles d'ADN double brin correspondant à la séquence de chaque variant de la bibliothèque (cf. matériel et méthode).

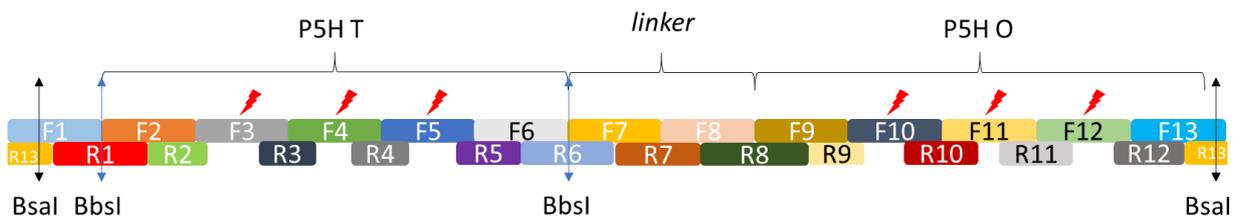


Figure 31 : schéma de l'assemblage des oligonucléotides codants pour les variants de la bibliothèque de bi-P5H

Afin d'avoir une quantité de matériel suffisante pour la construction de la bibliothèque, les cercles d'ADN ont été amplifiés par *rolling circle amplification* (RCA), une technique qui permet l'amplification d'une matrice d'ADN circulaire. Le principe de l'amplification par RCA repose sur une polymérase capable de déplacer le brin d'ADN qu'elle utilise comme matrice (Figure 32), ce qui permet de la répliquer en tandem. L'utilisation de la polymérase Phi27, qui possède une activité de déplacement d'un brin dans un ADN double brin, associée à des amorces de polymérisation constituées d'hexamères aléatoires permet l'amplification isotherme exponentielle de n'importe quelle séquence d'ADN préférentiellement circulaire ou de grande longueur. La molécule qui en résulte possède une très haute masse moléculaire puisqu'il s'agit d'une seule molécule d'ADN comportant à la suite les répétitions de la matrice de base.

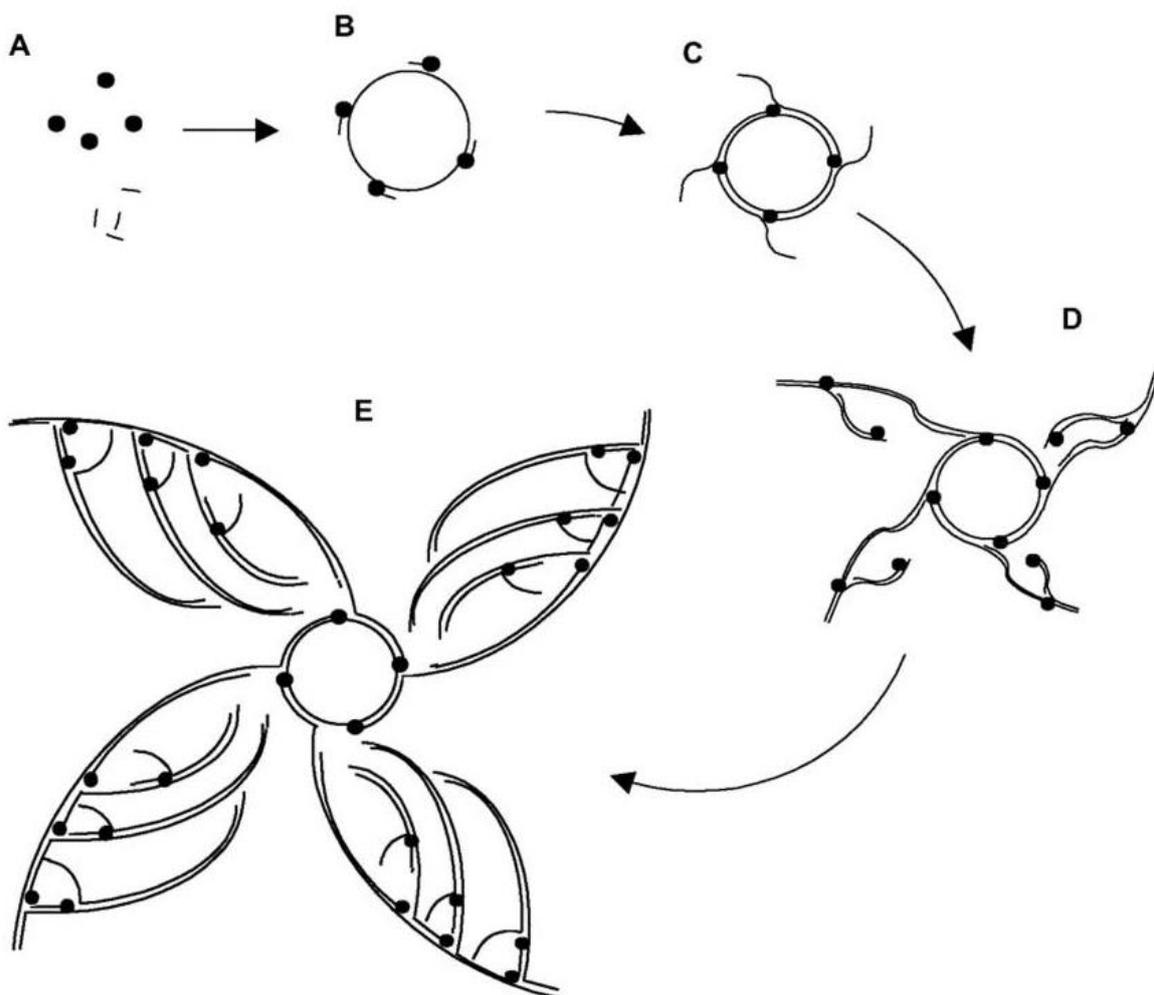


Figure 32 : schéma du principe de l'amplification d'ADN circulaire par RCA (*Rolling Circle Amplification*)

Amplification TempliPhi d'une matrice d'ADN circulaire. L'hexamère aléatoire (~) et l'ADN polymérase Phi29 (•) (A) s'hybrident à une matrice d'ADN circulaire simple ou double brin dénaturée (B). L'ADN polymérase Phi29 synthétise l'ADN à partir de la matrice en déplaçant les brins produits par les événements de réplication en aval (C). L'amplification continue,

produisant des répétitions en tandem complémentaires de la matrice d'origine. L'hexamère aléatoire se lie aux brins concatémères, amorçant une amplification additionnelle (D). Le déplacement continu des brins entraîne la libération exponentielle d'ADN double brin (E). (Reagin, M. J. *et al.*, 2003)

Les produits de RCA ont été déposés sur gel afin de s'assurer de l'amplification des cercles d'ADN (Figure 33).

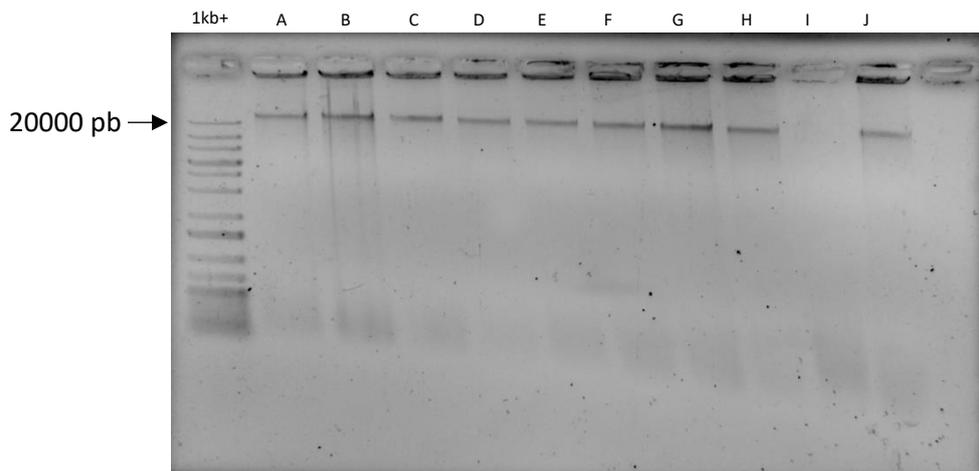


Figure 33 : Amplification des assemblages d'oligonucléotides de la bibliothèque de bi-P5H par RCA

3 μ L de chacune des 10 RCA ont été déposés sur gel d'agarose pour contrôler l'amplification des cercles d'ADN ligués.

Les bandes caractéristiques d'une amplification par RCA ne migrent pas dans le gel du fait de leur haute masse moléculaire (sauf celle déposée dans le puit I dont l'amplification n'a pas marché) (Figure 33). Cela confirme donc que les amplifications par RCA ont fonctionné, et on peut donc confirmer *a posteriori* que la ligation des cercles d'ADN codant pour les bi-P5H a été efficace. En effet, la matrice d'ADN doit être circulaire (ou de grande taille) pour donner lieu à une amplification par RCA.

C. Stratégie de clonage des inserts codant les variants de bi-P5H

La stratégie générale du clonage des inserts codant les variants de la bibliothèque de bi-P5H a été présentée Figure 29. Des sites de coupure à l'enzyme BsaI ont été placés de part et d'autre de la séquence codant pour les bi-P5H pour permettre leur clonage dans un vecteur accepteur compatible avec le phage *display*, dont la carte est présentée Figure 34.

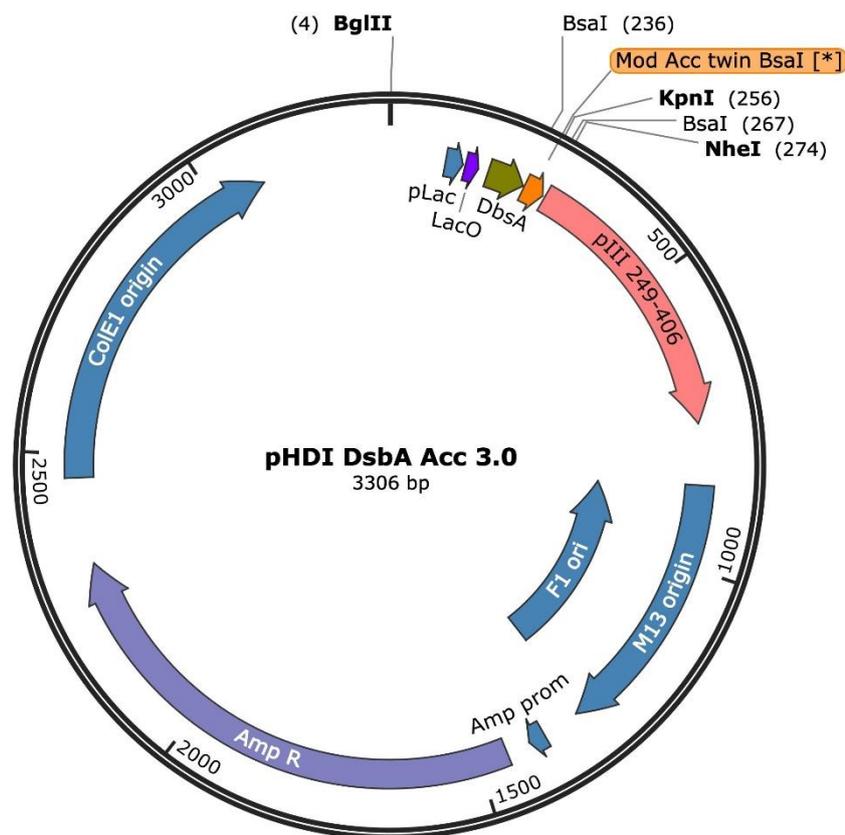


Figure 34 : carte du vecteur accepteur de la bibliothèque (pDip 3.0)

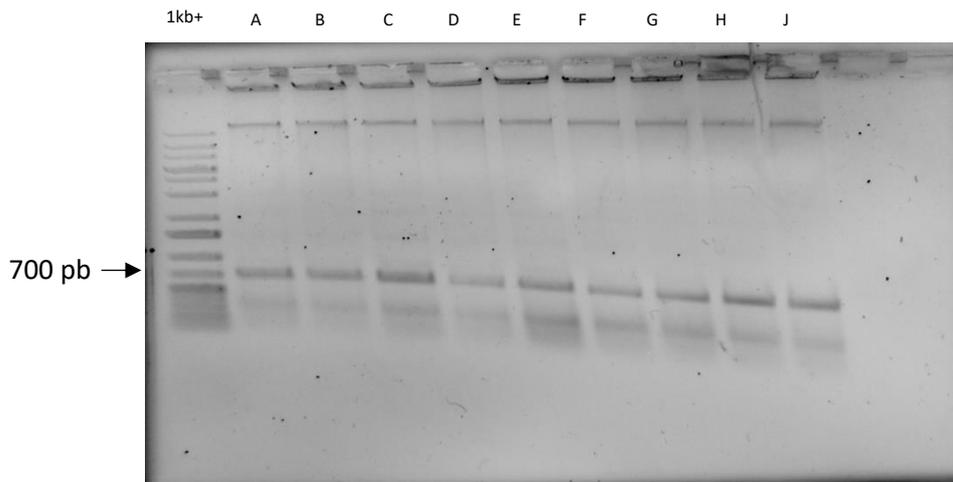


Figure 35 : Digestion des produits d'amplification de la RCA par Bsal

Les cercles d'ADN codants pour les variants de la bibliothèque et amplifiés par RCA ont été digérés par l'enzyme Bsal. On peut voir que la digestion d'une molécule de plusieurs dizaines à plusieurs centaines de kilobases donne l'insert de 721 paires de bases digéré et prêt à être inséré dans un vecteur accepteur. On peut également noter la présence d'une bande contaminante vers 300 paires de bases qui correspond aux mono-P5H.

La digestion simultanée du vecteur accepteur et des cercles d'ADN amplifiés par l'enzyme de restriction Bsal (Figure 35) a alors permis d'obtenir des extrémités complémentaires entre les gènes codants pour chaque variant de la bibliothèque d'une part, et le vecteur de destination de la bibliothèque qu'on appelle pDip 3.0. Ce vecteur est compatible avec le phage *display* car il permet d'exprimer les protéines de la bibliothèque en fusion avec la protéine pIII du phage, pour permettre leur exposition à sa surface. Les produits de la digestion par Bsal ont été purifiés sur colonne (kit *gel and PCR clean up*, Macherey Nagel), et une digestion par l'enzyme KpnI a été réalisée sur le vecteur pDip 3.0 digéré par Bsal, afin de linéariser le vecteur non digéré. En effet, un site de coupure à l'enzyme KpnI est présent dans la région encadrée par les deux sites Bsal (Figure 34), et ce site n'est donc pas présent sur les molécules de pDip3.0 doublement coupées par Bsal. La quantité d'ADN nécessaire pour construire la bibliothèque est très importante, et 70 µg de pDip 3.0 ont dû être digérés simultanément dans 10 réactions.

À l'issue d'une étape de ligation (8 réactions dans un ratio molaire insert : vecteur = 1.3 : 1), on a donc inséré les gènes codants pour une multitude de variants bi-P5H dans un vecteur permettant leur exposition à la surface de phages filamenteux M13, ce qui permet la sélection d'interacteurs par phage *display*. Les produits de ligation ont été rassemblés puis purifiés sur colonne en vue de réaliser la transformation des bactéries.

Lors de la conception de la bibliothèque de bi-P5H, nous avons réfléchi à introduire un outil qui permettrait de générer de la diversité, même une fois la bibliothèque terminée. C'est pour cette raison

que le domaine P5H T (en N-ter des bi-P5H) est entouré de 2 sites de coupure à l'enzyme BbsI (Figure 31). L'idée est de pouvoir digérer l'ADN codant pour les protéines de la bibliothèque par l'enzyme BbsI, pour couper l'ADN codant les domaines T de toute la bibliothèque. Les domaines T ainsi digérés vont se ré-apparier au hasard parmi tous les plasmides disponibles, et on aura alors un brassage entre les domaines T et O de toute la bibliothèque. On appelle ce processus le *shuffling* de la bibliothèque, qui permet de générer de la diversité à partir de la diversité déjà existante. En effet, par ce processus, il n'y aura pas création de nouveaux domaines T ou O, mais création de nouvelles combinaisons entre les domaines T et O présents dans la bibliothèque.

D. Transformation et caractérisation de la diversité de la bibliothèque bi-P5H

L'ultime étape de la construction de la bibliothèque consiste alors à transformer le plus grand nombre de bactéries possible par les plasmides codants pour chaque variant de bi-P5H. Dans le cas d'une bibliothèque construite pour le phage *display*, ces plasmides sont appelés phagemides. Cette étape est peut-être la plus importante du processus de construction d'une bibliothèque compatible avec le phage *display*, puisque c'est l'étape limitante du point de vue de la diversité atteignable (de Haard, H. J. et al., 1999). En effet, c'est l'efficacité de la transformation des bactéries par les phagemides qui va déterminer la diversité de la bibliothèque, et c'est pour cette raison qu'il est intéressant de réaliser le plus grand nombre de transformations possibles. Pour la bibliothèque de bi-P5H, des *E. coli* supercompétentes ont été transformées par électroporation par 4 à 5 µg de plasmides pour donner une bibliothèque d'une diversité théorique de 1.4×10^9 clones indépendants (cf. matériel et méthode). L'ADN codant pour les différents variants de la bibliothèque a été extrait et digéré pour contrôler la taille des inserts et ainsi avoir une idée de la quantité de protéines bi-domaines par rapport aux protéines mono-domaine au sein de la bibliothèque (Figure 36).

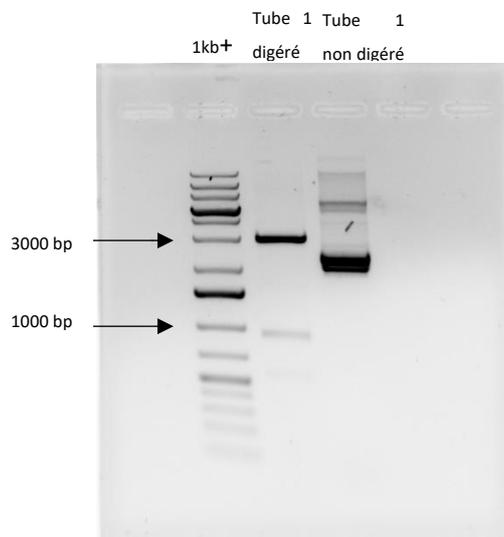


Figure 36 : profil de migration de l'ADN de la bibliothèque de bi-P5H doublement digéré par Bgl II et Nhe I.

400 ng d'une miniprep de la bibliothèque digérée par les enzymes de restriction BglII et NheI ont été déposés sur gel dans le deuxième puit. 1 µg de la même miniprep non digérée a été déposé dans le dernier puit.

La double digestion d'une bi-P5H insérée dans le vecteur accepteur de la bibliothèque par les enzymes de restriction Bgl II et Nhe I doit donner une bande à 3030 paires de bases et une bande à 928 paires

de bases. Si on remarque une légère bande à ~550 paires de bases, qui correspond à la taille d'un insert codant pour un mono-domaine, la grande majorité de l'ADN digéré donne les bandes à la taille attendue. Cela signifie que la majorité des protéines de la bibliothèque de bi-P5H sont bien des bi-domaines.

Les mono-P5H sont formées par l'hybridation aspécifique de l'oligonucléotide FOR12 sur le fragment REV4 à la place de FOR5 (Figure 31). En effet, les deux P5H de départ T et O possèdent un grand nombre de résidus en commun (71% d'identité de séquence) et l'amorce FOR12 code pour les mêmes résidus à son extrémité que le fragment FOR5. La seule différence entre les sites d'hybridation est la troisième base de chaque codon, ce qui suffit à FOR12 pour s'hybrider au fragment REV4. Selon le même principe, le fragment FOR11 peut s'hybrider au fragment REV3 à la place de FOR4. Les P5H qui résultent de ces mésappariements ne possèdent qu'un seul domaine, chimère entre un domaine T et un domaine O.

E. Séquençage de 48 clones pris au hasard dans la bibliothèque

Pour évaluer la qualité de la bibliothèque créée, 48 clones pris au hasard dans la bibliothèque ont été séquencés. Le résultat de ces séquençages sont présentés [Tableau 4](#).

Nombre de clones séquencés	48
Séquences exploitables	39
bi-P5H en phase	15 (38,5%)
mono-P5H en phase	7 (17,9%)
Séquences hors phase	10
Anomalies de séquences	7

Tableau 4 : Séquençage des 48 clones de la bibliothèque de bi-P5H

Sur 48 clones envoyés au séquençage, 39 ont donné des séquences exploitables, c'est-à-dire avec un score de qualité supérieur au seuil fixé par la compagnie qui a réalisé le séquençage (Eurofins Genomics). Parmi ces séquences, on retrouve 15 bi-P5H (38.5% des séquences exploitables), et 7 mono-P5H (17.9% des séquences exploitables), soit 56.4% de séquences codants pour des variants correctement insérées dans le vecteur de la bibliothèque. On retrouve également 10 séquences entières de P5H hors phase et 7 anomalies de séquences qui semblent correspondre à des cercles mal assemblés.

Nous avons également analysé les positions diversifiées des séquences des 15 bi-P5H correctement insérées dans le phagemide pour voir si les [fréquences](#) d'acides aminés [observées](#) à ces positions sont conformes aux [fréquences attendues](#) ([Figure 37](#)). La [Figure 37](#) montre que la diversité observée aux position diversifiée est conforme à la diversité souhaitée dans la conception de la bibliothèque. On peut noter que l'alanine, l'arginine et la sérine semblent moins représentées que prévu, tandis que la glutamine, le tryptophane, et la tyrosine semblent plus représentés qu'initialement prévu. Ces variations ne sont pas forcément représentatives de l'ensemble de la bibliothèque car seules 15 séquences ont été analysées.

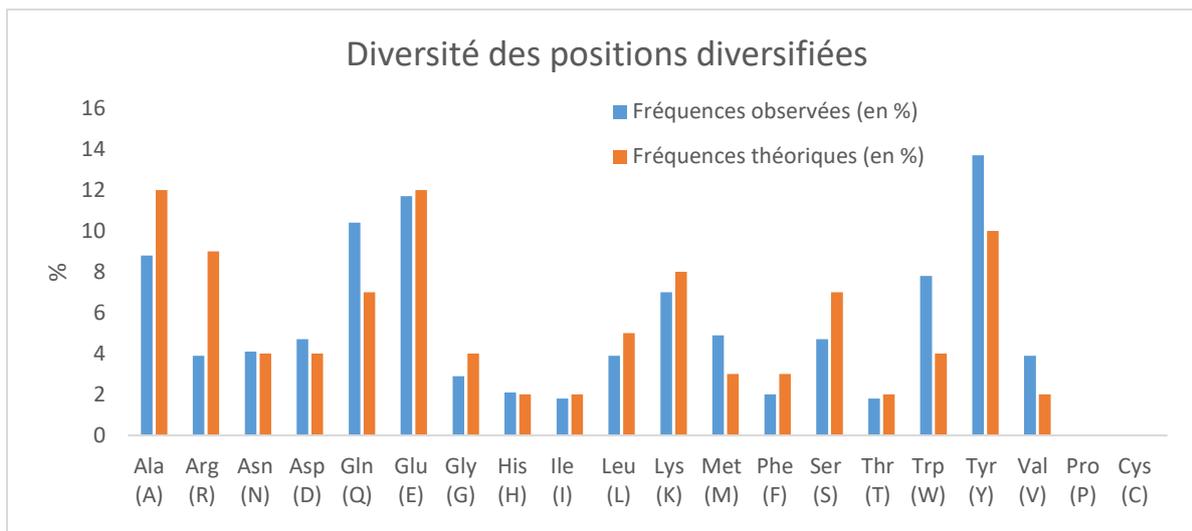


Figure 37 : Diversité observée contre diversité théorique au niveau des positions diversifiées de la bibliothèque de bi-P5H

La fréquence de chaque acide aminés au niveau des 32 positions diversifiées au sein de la bibliothèque a été comparée aux fréquences cibles présentées Tableau 3.

Globalement, le séquençage de la bibliothèque valide la stratégie d'assemblage de la bibliothèque, et l'utilisation de la technologie des TRIMers pour générer une diversité ciblée à chaque position diversifiée dans la bibliothèque.

Cette expérience conclut la création de la bibliothèque de bi-P5H, et donc la première partie des résultats. La seconde partie des résultats est consacrée à la conception du modulateur allostérique.

Partie 2 : Conception du modulateur allostérique

I. Sélections par phage display d' α Reps reconnaissant les P5H

Simultanément à la construction de la bibliothèque, l'autre axe du projet a consisté à créer le modulateur allostérique, cette molécule qui doit libérer la cible des interacteurs issus de la bibliothèque. Pour rappel, ce modulateur doit être constitué de deux α Reps dirigées chacune contre la face bi-hélicoïdale (constante) d'un domaine des bi-P5H, puisque la face tri-hélicoïdale de chaque domaine est différente pour chaque variant de la bibliothèque. Pour obtenir des α Reps dirigées contre la face constante des P5H, nous avons imaginé une stratégie originale de sélection par phage *display* à partir de la bibliothèque d' α Reps déjà développée par le laboratoire (Urvoas, A. *et al.*, 2010).

Le phage *display* est la méthode utilisée au laboratoire afin d'isoler des interacteurs spécifiques d'une cible protéique à partir d'une bibliothèque de protéines (Scott, J. K. & Smith, G. P., 1990). Cette technique consiste à exposer chaque protéine de la bibliothèque à la surface de bactériophages filamenteux M13, puis de les mettre en contact avec une surface sur laquelle a été immobilisée la cible protéique pour laquelle on souhaite obtenir des interacteurs (Figure 38).

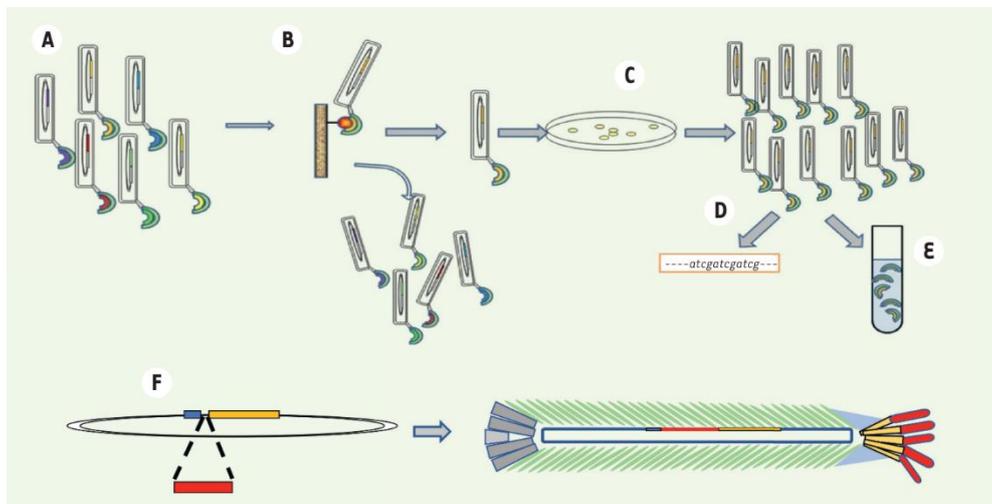


Figure 38 : Principe du phage *display*

L'exposition sur phage ou phage *display* est une puissante méthode de sélection moléculaire qui permet d'obtenir des interacteurs d'une cible spécifique à partir d'une bibliothèque de protéines. L'idée est d'avoir une très vaste population (typiquement 1 milliard ou plus) de protéines ou de peptides différents exposés à la surface de bactériophages (A). Ces bactériophages sont mis en contact avec une molécule cible (la sphère rouge) fixée sur un support (B). Les très rares protéines dont la séquence leur permet d'interagir avec la cible sont retenues tandis que les autres sont éliminées. Les phages portant ces protéines peuvent être alors amplifiés (C, D), puis utilisés pour déterminer les séquences des protéines capables de reconnaître la cible (D), et produire ces protéines sous forme recombinante (E). La méthode originellement décrites pour sélectionner des séquences courtes (des peptides) a pu être ensuite transposée aux protéines. Le coeur de l'idée est de fusionner (F) le gène (en rouge) de la protéine à exposer avec le gène d'une des protéines du phage, la protéine PIII. Les

phages alors produits présentent la protéine (rouge) à leur surface, ce qui permet de les trier ; la particule virale intègre le gène correspondant, ce qui permet de les amplifier. (Minard, P., 2019)

Les phages exposant des variants de la bibliothèque qui interagissent avec la cible seront retenus sur cette surface, tandis que les autres seront éliminés lors des différents lavages. Une étape d'élution est alors réalisée afin de récupérer les phages retenus sur la surface. Des bactéries sont ensuite infectées par les phages élués afin de les multiplier, puis un nouveau tour de sélection identique au premier est réalisé avec les phages issus du premier tour. Au bout de 3 à 5 tours de sélection, on considère que la population de phages élués est assez enrichie en phages exposants des variants d'intérêt et on réalisera un ELISA afin d'identifier parmi un échantillon de clones choisis aléatoirement lesquels interagissent le mieux avec la cible. Pour réaliser la détection des clones positifs par ELISA, il est nécessaire d'immobiliser la cible de la même façon que lors du phage *display*. Les phages provenant de clones uniques, c'est à dire qui exposent la même protéine, sont ensuite mis au contact de la cible de façon individuelle. La présence de phages dans les puits, qui reflète une interaction entre la protéine exposée à la surface des phages et la cible, est révélée à l'aide d'un anticorps anti-phageM13 couplé à la HRP, une peroxydase dont le produit de réaction est coloré. Une mesure d'absorbance de la plaque suffit alors à comparer l'efficacité de fixation des protéines exposées sur phage.

Pour orienter la sélection par phage *display* vers des α Reps qui reconnaissent la face bi-hélicoïdale des P5H, nous avons utilisé les propriétés des formes sauvages et mutantes des P5H. En effet, les P5H ont été mutées sur leur face tri-hélicoïdale pour tester la tolérance aux mutations de l'architecture. Pour un même domaine (P5H T ou P5H O), les formes sauvages et mutantes partagent donc la face bi-hélicoïdale (dite constante) mais possèdent des faces tri-hélicoïdales (dites variables) différentes. La stratégie mise en place consiste à alterner à chaque tour de sélection entre les formes sauvages et mutantes pour chaque domaine, afin de sélectionner des α Reps qui interagissent au niveau de la partie commune aux deux formes, c'est-à-dire la face bi-hélicoïdale. Le principe de cette sélection est illustré à travers l'exemple de la sélection contre P5H T (Figure 39), mais le même protocole a été utilisé pour la sélection contre le domaine P5H O.

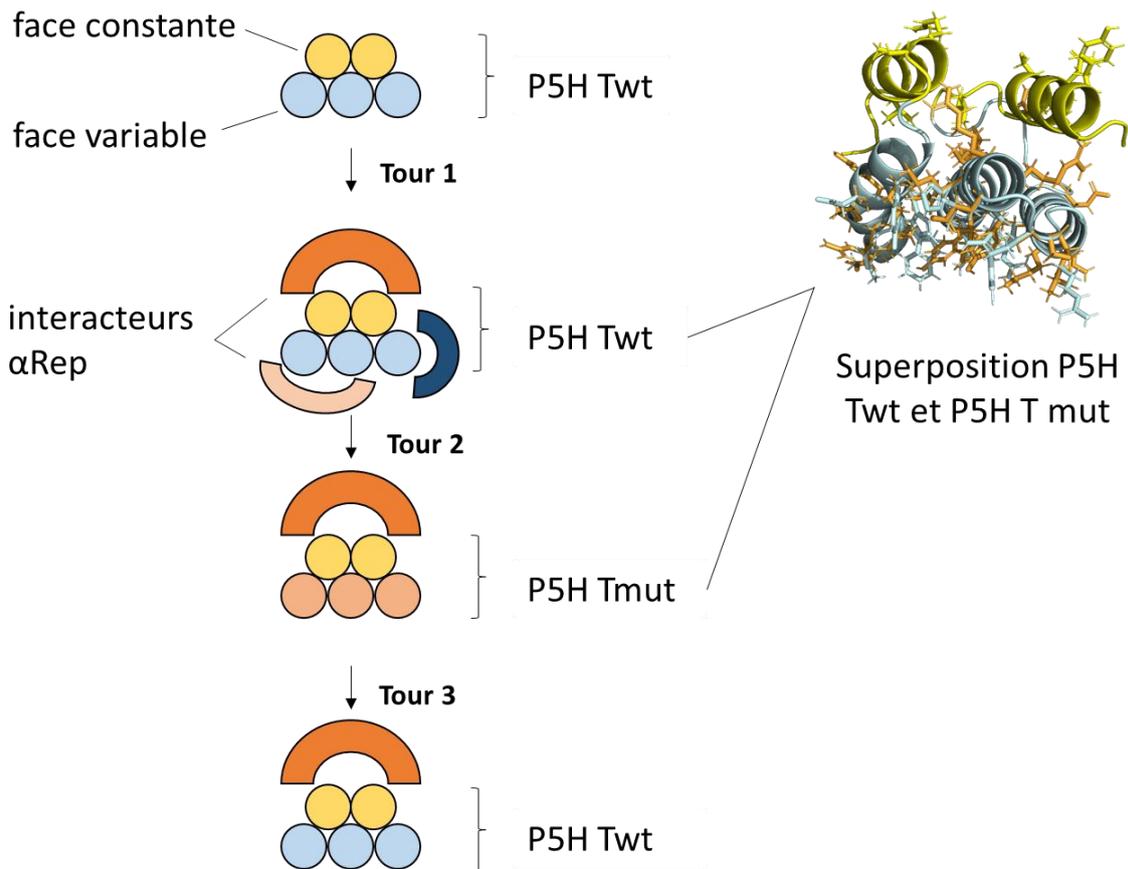


Figure 39 : Schéma de la stratégie utilisée pour la sélection par *phage display* des interacteurs de la face bi-hélicoïdale

Pour une même pentahéline (O ou T) deux formes existent : la forme sauvage (wt) ou la forme mutée (mut). Les formes mutées ont été générées par mutation de 18 positions sur la face tri-hélicoïdale de la protéine. Ainsi, les formes mutées et sauvages partagent la même face bi-hélicoïdale. La stratégie employée pour obtenir des interacteurs de la face bi-hélicoïdale consiste à immobiliser alternativement les formes wt et mut sur la plaque ELISA à chaque tour de sélection. Or, le point commun entre ces deux formes étant leur face bi-hélicoïdale, cette alternance devrait avoir pour effet de faire converger la sélection vers la reconnaissance de cette face des pentahélines.

Afin de favoriser l'interaction entre le modulateur et la face bi-hélicoïdale de chacune des P5H, une tyrosine a été rajoutée sur celle-ci. La tyrosine doit jouer un rôle de *hotspot* d'interaction, puisque sa nature amphiphile lui permet de réaliser des interactions polaires par son groupement $-OH$, et des interactions par son cycle aromatique. Cette particularité lui permet d'être impliquée dans de nombreuses interactions, ce qui pourrait expliquer qu'elle soit fréquemment retrouvée au niveau des boucles CDR des anticorps (Collis, A. V. J. et al., 2003).

Pour réaliser cette sélection originale par *phage display*, il a tout d'abord fallu établir une stratégie d'immobilisation des cibles pour les présenter de façon homogène lors de la sélection. La stratégie retenue a été de produire la cible avec un His-tag en N-ter et un avitag « biotinylable » en C-ter, séparé

de la protéine par un site de coupure à la TEV protéase. De cette façon, les cibles ont été produites biotinylées, ce qui a permis de les immobiliser avec de la streptavidine et d'éluer le complexe cible-phage par coupure à la TEV protéase (Figure 40). Cette stratégie nécessite dans un premier temps de sous-cloner le gène codant pour les différentes pentahélines (P5H Twt/mut et P5H Owt/mut) dans un vecteur répondant aux critères décrits précédemment.

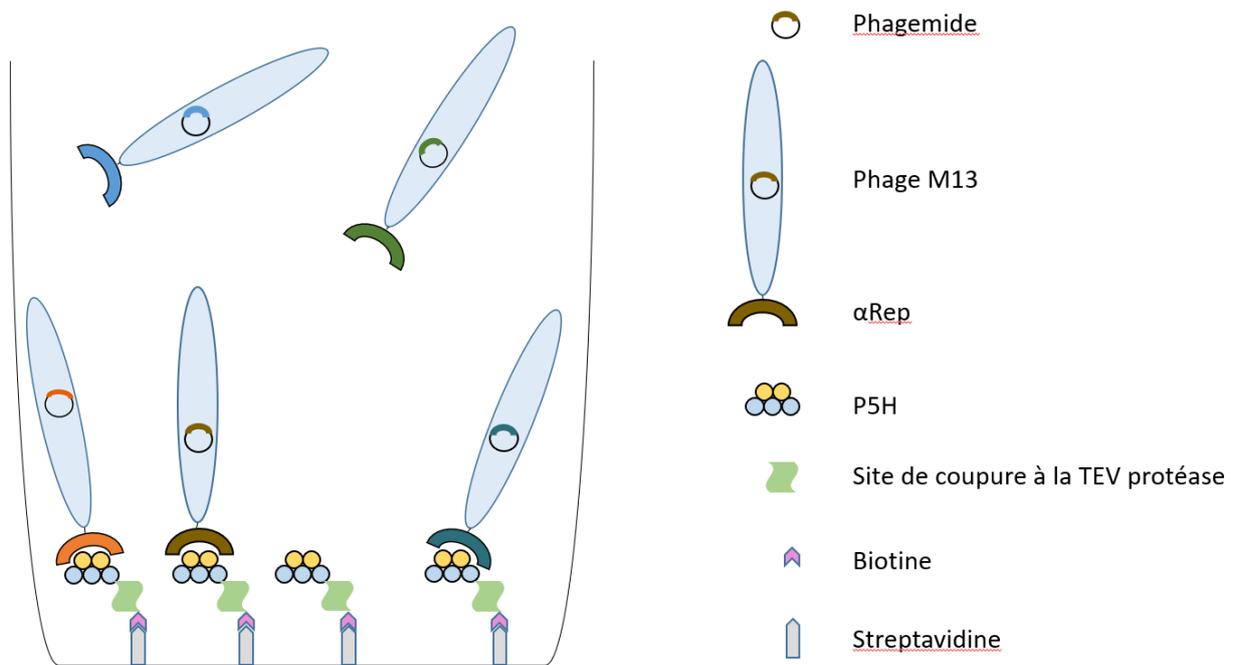


Figure 40 : Schéma de la stratégie d'immobilisation utilisée pour le projet

Les cibles protéiques, produites avec un His-tag en N-ter et un Avitag biotinylable précédé d'un site de coupure à la TEV protéase en C-ter sont immobilisées sur une surface préalablement recouverte de streptavidine. Les protéines de la bibliothèque (ici des α Reps) sont exposées à la surface de phages filamenteux. Les phages qui exposent des interacteurs de la cible vont être capturés par celle-ci. Les complexes interacteur-cible pourront être élués en ajoutant de la TEV protéase dans le milieu.

Une approche basée sur le Gibson Assembly (Gibson, D. G. *et al.*, 2009) (GA) a donc été réalisée afin d'insérer les gènes codants pour les différentes pentahélines dans un vecteur pTeTar2, qui permet d'exprimer une protéine avec un His-tag à son extrémité N-ter et un Avitag précédé d'un site de coupure à la TEV protéase à son extrémité C-ter.

La pureté des produits PCR ainsi formés a été vérifiée sur gel (Figure 41).

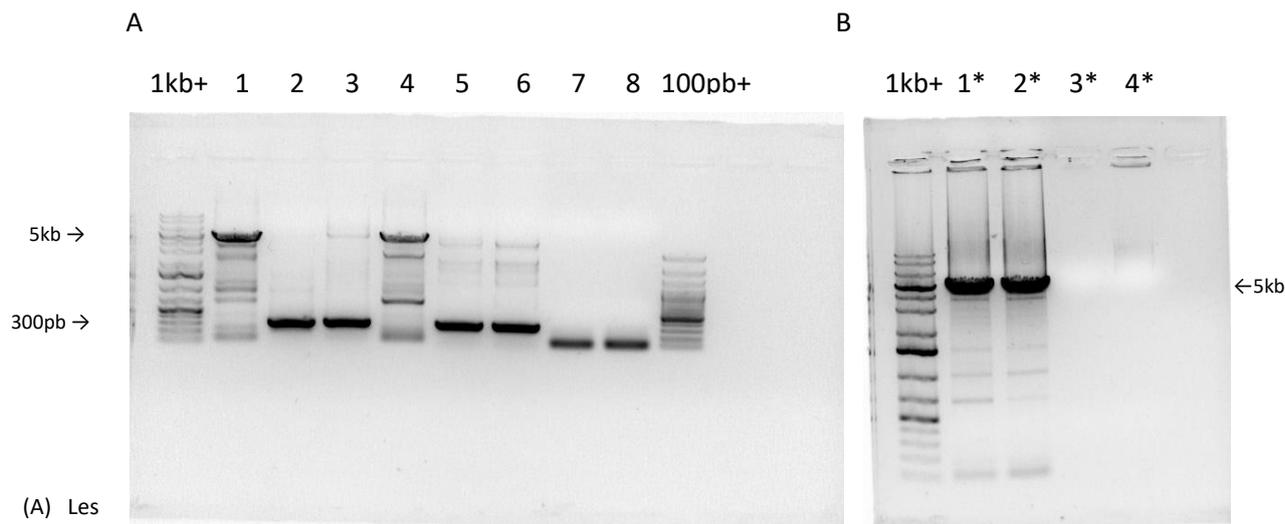


Figure 41 : Contrôle sur gel de la pureté des produits PCR pour le Gibson Assembly (GA).

produits PCR utilisés pour le GA ont été déposés sur gel d'agarose. 1kb+ : marqueur DNA ladder ; 1 : pTeTar2 pour P5H T ; 2 : P5H Twt ; 3 : P5H Tmut ; 4 : pTeTar2 pour P5H O ; 5 : P5H Owt ; 6 : P5H Omut ; 7 : contrôle sans vecteur avec les mêmes amorces que 1 ; 8 : contrôle sans vecteur avec les mêmes amorces que 4 ; 100pb+ : marqueur DNA ladder. Le programme utilisé est décrit dans le matériel et méthode. (B) Les PCR 1 et 4 ont été refaites du fait de la présence de trop nombreux contaminants. La température d'hybridation a été augmentée afin de défavoriser l'hybridation aspécifique. 1* : idem 1 ; 2* : idem 4 ; 3* : contrôle sans vecteur ; 4* : contrôle sans amorces.

Le Gibson Assembly a été réalisé selon un protocole détaillé dans le matériel et méthode, puis l'insertion des gènes codants pour les 4 pentahélines dans le vecteur d'expression des cibles biotinyllées (pTeTar2) a été vérifiée sur gel, après une double digestion par les enzymes de restriction BamH I/Hind III dont les sites de coupure entourent le gène d'intérêt (Figure 42).

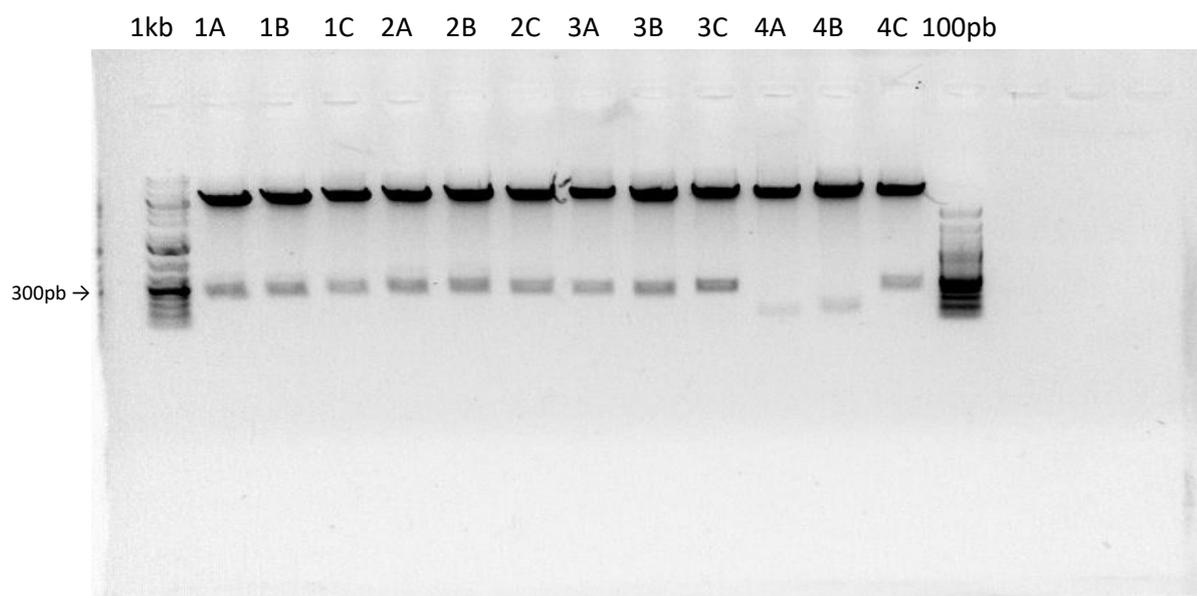


Figure 42 : Contrôle sur gel de l'insertion des gènes d'intérêt par digestion BamH1/Hind III des produits du GA

Les plasmides, synthétisés par les bactéries transformées par les produits du GA, ont été extraits par miniprep et digérés doublement par BamH1/Hind III puis déposés sur gel d'agarose. 1kb+ : marqueur DNA ladder ; 1A/B/C : P5H Twt dans pTeTar2 ; 2A/B/C : P5H Tmut dans pTeTar2 ; 3A/B/C : P5H Owt dans pTeTar2 ; 4A/B/C : P5H Omut dans pTeTar2.

Les fragments générés de cette façon font bien la taille attendue d'environ 320 pb à l'exception des puits 4A et 4B. Les échantillons ont été séquencés et l'exactitude des séquences sous-clonées a ainsi été vérifiée.

A. Selection d' α Reps dirigées contre la face bi-hélicoïdale des P5H

Une fois les gènes codant pour les P5H insérés dans le vecteur d'expression pTetar2, les protéines ont été produites et purifiées, et leur biotinylation a été vérifiée par western blot (Figure 43).

Environ 2 μ g de protéines ont été déposés dans chaque puits, sur 2 gels de polyacrylamide 15%. L'un des gels a été coloré au

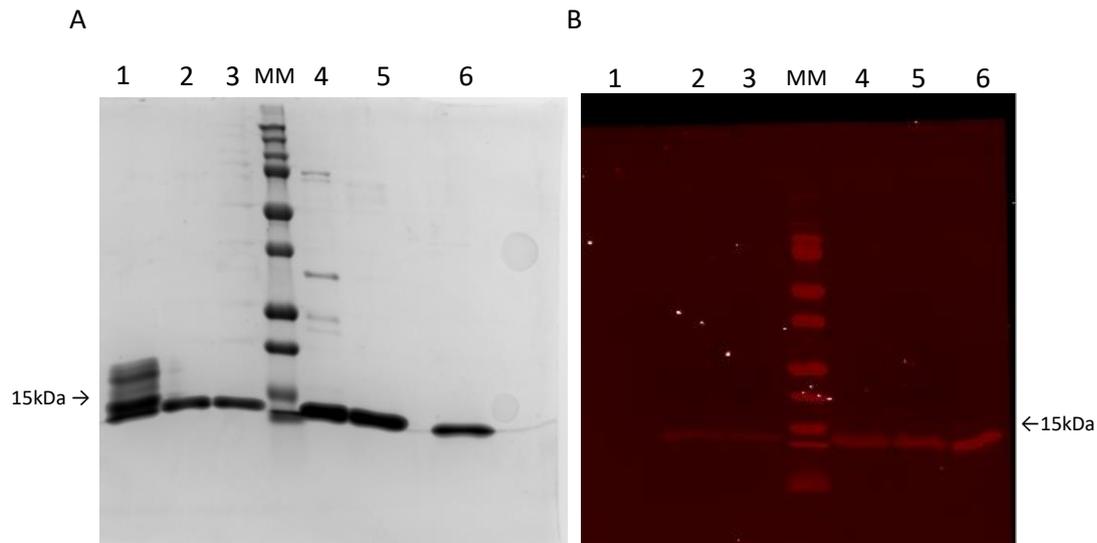


Figure 43 : Contrôle de la biotinylation des 4 P5H par Western blot

bleu de Coomassie (A), et l'autre gel a été traité selon le protocole décrit dans le matériel et méthode pour le Western blot (B). Les protéines ont été déposées dans l'ordre suivant : 1 : α Rep « B8 » non biotylée ; 2 : P5H Omut ; 3 : P5H Owt ; MM : marqueur de taille Dual Color ; 4 : P5H Tmut ; 5 : P5H Twt ; 6 : α Rep « B8 » biotylée.

La bande correspondant aux P5H dans le western blot (Figure 43.B) est moins intense que la bande obtenue pour les mêmes protéines sur le gel coloré au bleu de coomassie, alors que la même quantité de protéines a été déposée sur les deux gels. Il est donc probable que pour chaque P5H, toutes les protéines de l'échantillon ne soient pas biotylées. Cependant, pour les applications que nous souhaitons réaliser (sélection par phage *display* et ELISA clonal), nous n'avons pas besoin que 100% des protéines purifiées soient biotylées. Il est seulement nécessaire d'avoir assez de protéine biotylée à immobiliser sur la surface, et nous avons donc continué les expériences avec ces échantillons.

Les sélections par phage *display* d' α Reps dirigées contre la face constante de chaque domaine P5H a été réalisée selon un protocole détaillé dans la partie matériel et méthode conçu à partir de la stratégie discutée précédemment. À l'issue des 3 tours de sélection, un ensemble de phages exposant des variants potentiellement capables de reconnaître les cibles ont été sélectionnés. Il était alors nécessaire d'identifier ceux qui possèdent l'affinité et la spécificité les plus adaptées à l'objectif fixé.

Dans cette optique, une centaine de clones ont été testés par phage ELISA clonal (Figure 44). Le phage ELISA clonal permet de tester les phages issus de colonies isolées, c'est-à-dire des populations de phages qui n'expriment qu'un variant de la bibliothèque, et le protocole utilisé est présenté dans le matériel et méthode. Le principe du phage ELISA clonal est présenté Figure 44.

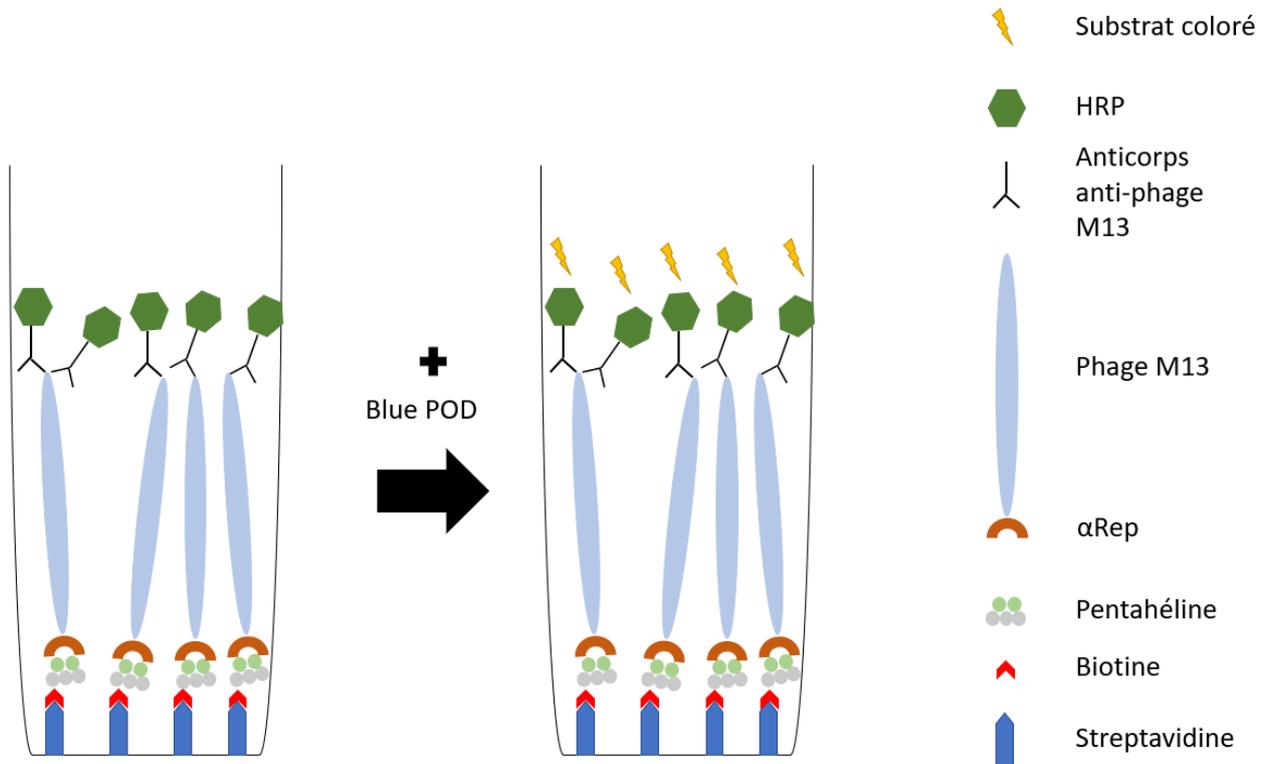


Figure 44 : Schéma d'un phage ELISA clonal

Les cibles biotinylées ont été immobilisées au fond des puits d'une plaque 96 puits par de la streptavidine, puis les phages exposants des interacteurs de ces cibles ont été capturés. Après lavages, les anticorps anti-M13 couplés à la peroxydase HRP ont été rajoutés dans chaque puit, et enfin le substrat de cette peroxydase, le Blue POD, a été ajouté pour révéler la présence de l'anticorps. Le produit de la réaction est bleu mais lorsque la réaction est arrêtée par ajout de HCl dans le milieu, la coloration devient jaune. Dans chaque puit, il n'y a qu'une population de phages, exprimant un variant sélectionné lors du phage *display*.

Pour réaliser cet ELISA clonal, la cible sous sa forme sauvage (P5H Twt ou P5H Owt) a été immobilisée sur une plaque ELISA de la même façon que lors de la sélection. Les 24 meilleurs interacteurs des formes sauvages de chaque P5H ont été testés selon le même protocole dans un nouvel ELISA clonal, cette fois sur les 4 P5H (sauvages et mutantes) afin de tester leur spécificité pour l'une ou l'autre forme des P5H (Figure 45).

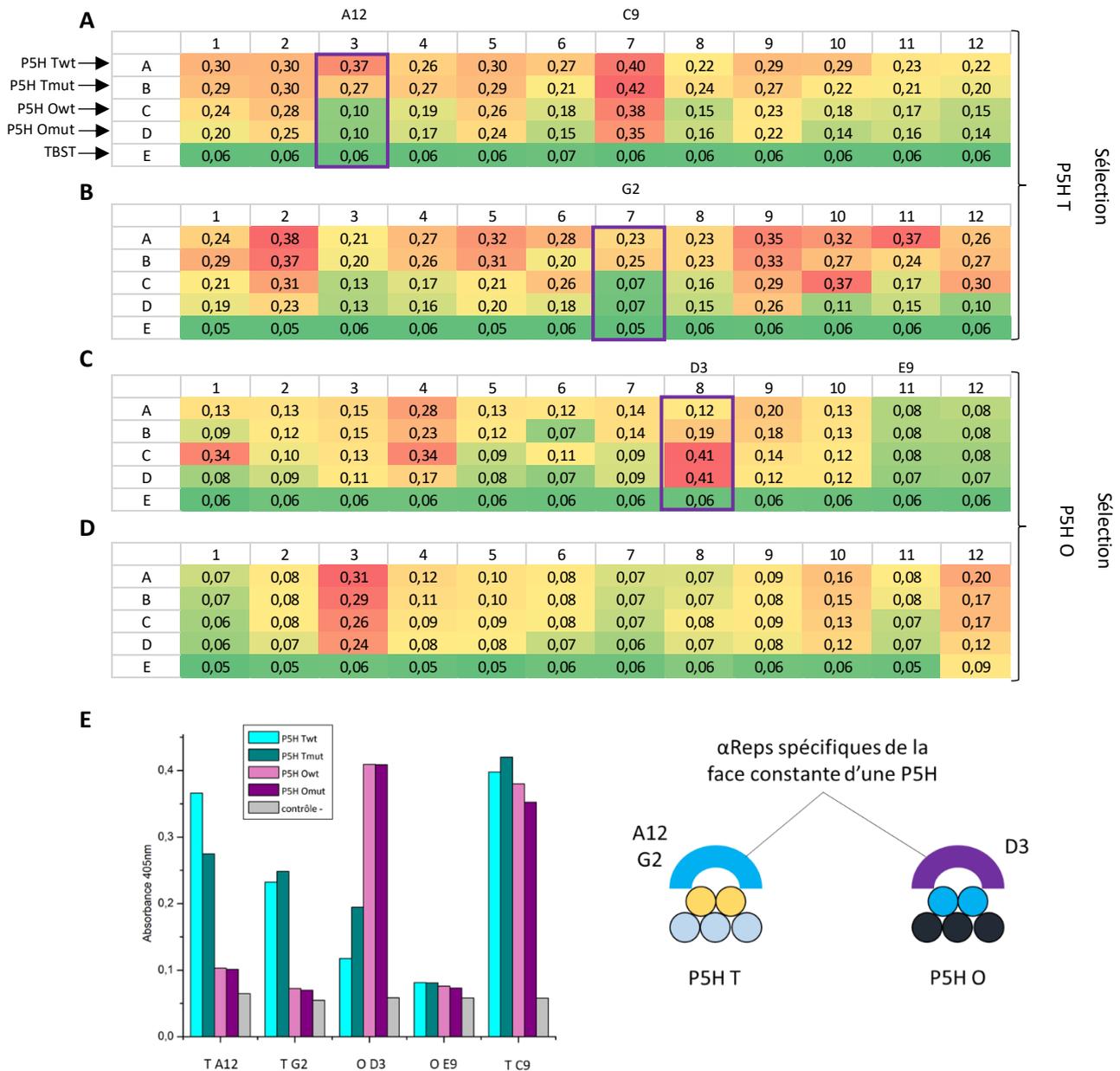


Figure 45 : Sélections d'αReps spécifiques des P5H T ou des P5H O

Les cibles ont été déposées dans l'ordre suivant pour toutes les plaques : ligne A : P5H Twt ; ligne B : P5H Tmut ; ligne C : P5H Owt ; ligne D : P5H Omut ; ligne E : TBST (*Tris buffer saline* tween 0.1%). Plus la valeur d'absorbance est haute, plus la coloration vert-jaune-rouge va tirer sur le rouge. Les 24 meilleurs interacteurs de P5H Twt (A et B) et P5H Owt (C et D) ont été testés par phage ELISA clonal sur les 4 P5H. Pour cela, des plaques ELISA 96 puits ont été recouvertes avec de la streptavidine, puis par les 4 pentahélines après avoir été bloquées par de la BSA (*Bovine serum albumine*). Les plaques ELISA ont ensuite été mises à incuber avec les phages exposant les 24 meilleurs interacteurs des P5H sauvages, puis avec un anticorps anti-phage_M13 couplé à la HRP (*horseradish peroxidase*) (voir matériel et méthode). Enfin la révélation a été faite par ajout de BluePOD soluble et la réaction a été stoppée à l'apparition de la couleur par ajout de HCl 1 M. Les valeurs d'absorbance à 405 nm des interacteurs encadrés en violet, qui montrent une nette différence d'affinité entre les P5H T et les P5H O, ont été représentées sous forme d'histogramme (E), avec deux autres clones représentatifs des clones criblés : un qui reconnaît toutes les P5H (C9) et un autre qui n'en reconnaît aucune (E9).

Ce phage ELISA clonal sur les 4 P5H a permis d'identifier 3 interacteurs montrant une spécificité vis-à-vis d'une seule des deux P5H (P5H T ou P5H O). Les valeurs d'absorbance à 405 nm de ces 3 clones ont été représentées sous forme d'histogramme [Figure 45.E](#), pour faire ressortir la préférence de ces α Reps pour une des deux P5H. Ces résultats semblent donc valider la stratégie mise en place pour sélectionner des interacteurs de la face constante des P5H, et sont conformes à notre objectif initial de sélectionner des α Reps dirigées contre chacun des deux domaines des bi-P5H. On peut cependant noter que la plupart des interacteurs identifiés (~34) reconnaissent les 4 pentahélines avec plus ou moins d'affinité, à l'image du clone C9. Tous les interacteurs ont été séquencés à l'exception du clone de la 12^{ème} colonne de la plaque D car ce clone donne un signal positif même dans le puit qui ne contient pas de P5H. Les séquençages ont montré que la majorité des séquences n'étaient pas redondantes, et que les sélections contre les P5H ont donné des interacteurs α Reps de taille et de séquence variables. Nous avons réalisé la suite des expériences avec une sélection d' α Reps composée des α Reps spécifiques d'un seul domaine P5H ainsi que des meilleurs interacteurs spécifiques des 4 P5H.

II. Caractérisation des α Reps reconnaissant les

P5H

Des α Reps qui interagissent potentiellement avec la face constante des P5H ayant été identifiées, nous avons cherché à caractériser l'interaction entre ces α Reps et les P5H pour choisir les meilleurs candidats à la création du modulateur. Les gènes codants pour ces α Reps ont d'abord été sous-clonés dans un vecteur d'expression, puis les protéines ont été produites et purifiées (voir matériel et méthode). Les protéines purifiées ont été utilisées dans des tests d'interaction par SPR.

A. Production des interacteurs α Reps anti-bi-P5H

Les interacteurs sélectionnés contre les P5H ont d'abord été sous-clonés dans le vecteur d'expression pQE 81, qui présente un Histag en N-ter de la protéine produite, afin de les purifier en vue d'étudier leur interaction avec les P5H. Une coupure aux enzymes de restriction BamH I et Hind III suivie d'une étape de ligation a permis d'insérer correctement le gène d'intérêt dans le pQE 81. L'insertion a été vérifiée par une coupure avec les mêmes enzymes de restriction (BamH I et Hind III) des gènes clonés pour vérifier la taille des inserts.

Les gènes ainsi clonés ont été séquencés pour vérifier l'exactitude des séquences, puis les interacteurs ont été produits dans des souches *d'E.coli* M15, et purifiés sur colonne Ni-NTA et SEC (voir matériel et méthode). À ce stade, nous disposons de plusieurs α Reps de taille différente et possédant une surface d'interaction différente. Nous avons alors cherché à déterminer lesquelles répondaient le mieux aux attentes que nous avons, à savoir interagir spécifiquement avec un des deux domaines P5H (T ou O). Pour répondre à cette question, l'interaction entre les α Reps et les P5H a été étudiée par SPR.

B. SPR P5H vs α Reps anti-bi-P5H

Les expériences de SPR (*Surface Plasmon Resonance*) ont été mises au point afin d'étudier l'interaction entre les 4 P5H et les différentes α Reps sélectionnées contre les P5H T ou les P5H O. Les P5H, qui ont été produites biotinylées, ont été fixées sur une puce neutravidine SA (Biacore) et les α Reps sélectionnées contre les P5H ont été passées comme analytes. Le tableau ci-dessous (Tableau 5) récapitule les résultats obtenus lors des expériences de SPR.

	Twt	Tmut	Owt	Omut
	KD (en μ M)			
α Rep A8 (sélection vs P5H T)	1,2 \pm 0,5	0,3 \pm 0,1	ND	0,3 \pm 0,2
α Rep A12 (sélection vs P5H T)	1,1 \pm 0,3	1,2 \pm 0,3	ND	ND
α Rep C9 (sélection vs P5H T)	1,4 \pm 0,4	0,8 \pm 0,2	ND	0,7 \pm 0,0
α Rep E6 (sélection vs P5H T)	ND	0,5 \pm 0,2	ND	ND
α Rep G2 (sélection vs P5H T)	0,4 \pm 0,0	0,3 \pm 0,0	ND	ND
α Rep G10 (sélection vs P5H T)	1,8 \pm 0,7	1,6 \pm 0,6	ND	ND
α Rep D3 (sélection vs P5H O)	0,8 \pm 0,4	0,5 \pm 0,1	0,6 \pm 0,0	0,5 \pm 0,1
α Rep F1 (sélection vs P5H O)	1,4 \pm 0,4	0,7 \pm 0,2	0,2 \pm 0,1	0,4 \pm 0,2

Tableau 5 : Étude des interactions entre les α Reps sélectionnées contre les P5H et leur cible par SPR

Cette expérience a notamment permis de montrer la différence d'affinité des différentes α Reps vis-à-vis de chaque P5H, et de vérifier si la méthode mise en place pour la sélection a permis d'obtenir des interacteurs spécifiques de P5H T ou de P5H O. On peut voir que certaines α Reps issues de la sélection contre P5H T montrent une différence d'affinité pour les P5H. En effet, les α Reps A12, G2 et G10 interagissent avec les P5H T avec un K_D de l'ordre du micromolaire, mais leur interaction avec les P5H O était trop faible pour permettre de calculer un K_D . Ce résultat tend à valider la stratégie que nous avons employée pour favoriser l'obtention d'interacteurs de la face constante des P5H, à savoir de rajouter une tyrosine sur la face constante des P5H et d'alterner entre les formes sauvages et mutantes à chaque tour de sélection lors du phage *display*. Cependant, on ne retrouve pas ce résultat pour les α Reps issues de la sélection contre P5H O. En effet l' α Rep D3, qui donnait un signal ELISA plus important pour les P5H O que pour les P5H T, semble interagir de façon équivalente avec les 4 P5H. En réalité, il serait possible de réaliser un modulateur composé de deux α Reps capables d'interagir avec

les deux domaines P5H de façon équivalente, puisqu'un tel modulateur serait capable d'interagir avec les bi-P5H et de moduler leur interaction avec une cible. L'intérêt de choisir des α Reps spécifiques d'un seul domaine P5H était d'avoir un moyen de contrôler facilement que l'interaction se faisait au niveau de la face constante des P5H. Si les α Reps capables d'interagir avec les deux domaines P5H de façon équivalente se fixent au niveau de la face constante des P5H, il n'y aurait aucun problème à concevoir le modulateur avec ces α Reps.

Pour la suite du projet, nous avons décidé d'utiliser les α Reps A12 et D3 pour réaliser le modulateur allostérique. L' α Rep A12 a en effet montré une interaction spécifique avec les P5H T à la fois en ELISA clonal et en SPR. L' α Rep D3 n'a montré une interaction spécifique avec les P5H O qu'en ELISA clonal et pas en SPR, mais l'autre α Rep testée (F1) n'était spécifique pour P5H O ni en ELISA clonal ni en SPR.

III. Conception du modulateur

A. Cristallisation des complexes A12-Tmut et D3-Omut

Les α Reps qui constitueront le modulateur ayant été choisies (A12 et D3), il était nécessaire de réfléchir à une stratégie de construction du modulateur. Bien que nous ayons décidé de fusionner les deux α Reps par une hélice, plusieurs constructions étaient envisageables, et afin de discriminer lesquelles étaient les plus pertinentes, nous avons décidé de cristalliser les complexes α Rep-P5H pour orienter la conception du modulateur. De plus, cela permettait d'identifier les surfaces d'interaction entre les α Reps et les P5H, et de pouvoir ainsi vérifier que la stratégie de sélection ciblée sur la face constante des P5H avait bien permis de sélectionner des interacteurs spécifiques de cette zone (au moins pour l' α Rep A12).

Des expériences de cristallisation ont été lancées avec les complexes A12-Tmut et D3-Omut en ratio molaire 1:1 (voir matériel et méthode).

Des cristaux ont été obtenus pour les deux complexes mais seuls les cristaux du complexe A12-Tmut (Figure 46) ont diffracté à des résolutions permettant l'exploitation des données (3.02 Å), tandis que les cristaux du complexe D3-Omut n'ont diffracté qu'à 5 Å.

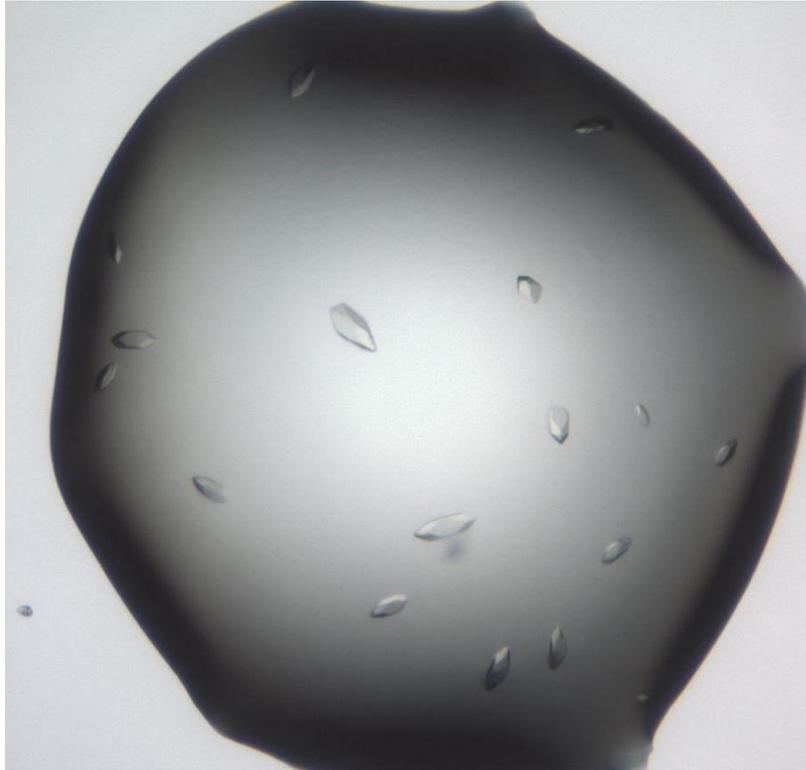


Figure 46 : photographie d'une goutte de cristallisation du complexe A12-Tmut

L'orientation relative des deux protéines l'une par rapport à l'autre a donc pu être déterminée, ce qui était le principal objectif de cette expérience. En effet, on voulait confirmer que la stratégie utilisée pour la sélection avait bien favorisé l'obtention d'interacteurs de la face bi-hélicoïdale des P5H. C'est ce que montre la structure du complexe (Figure 47) dans laquelle la pentahéline Tmut est représentée en rose et l' α Rep A12 en vert. Seules les chaînes latérales supposées être en interaction sont représentées. Afin de les identifier, la structure a été analysée sur le serveur en ligne PDBsum. Cette figure nous montre bien que la zone d'interaction sur la P5H est constituée d'acides aminés de la face bi-hélicoïdale et d'acides aminés accessibles sur le « dos » des hélices tri-hélicoïdales. On remarque également l'implication de la tyrosine 101, qui avait été ajoutée pour jouer le rôle de *hotspot* d'interaction. Globalement, cette interaction est caractérisée par deux ponts salins entre des arginines et des aspartates et des liaisons hydrogène. Cette structure vient confirmer que la stratégie que nous avons mis en place pour sélectionner les α Reps a bien permis d'obtenir un interacteur de la face bi-hélicoïdale de P5H T.

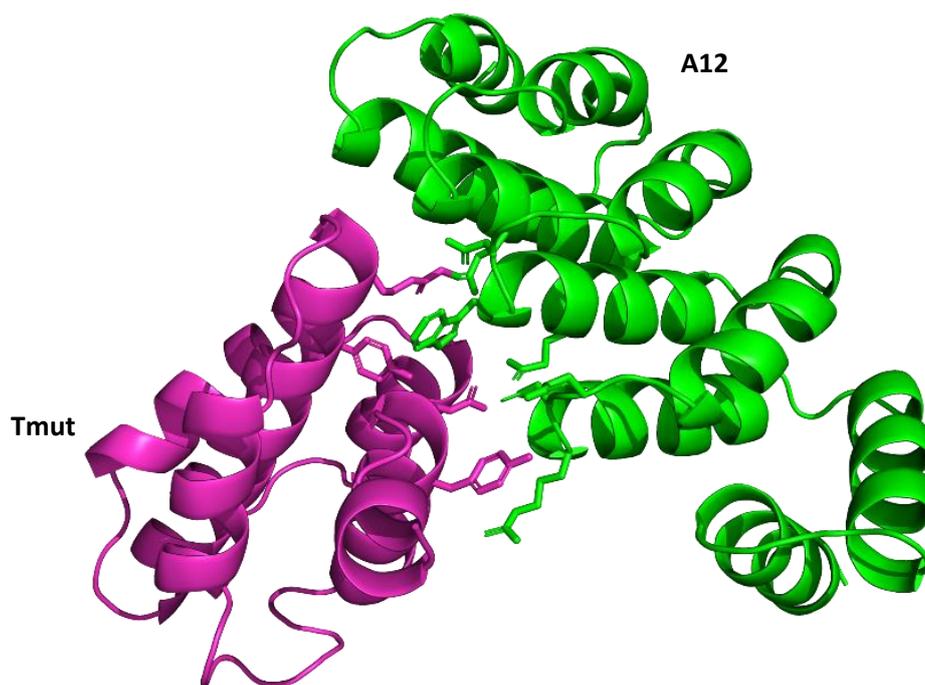


Figure 47 : Structure cristallographique du complexe A12 (vert) - Tmut (rose) à 3.02 Å

B. Conception du modèle in silico

La stratégie a consisté à utiliser la structure cristallographique d'une α Rep possédant le même nombre de motifs que les α Reps A12 et D3, qui ont été sélectionnées contre les P5H, afin de générer un modèle structural d'une fusion de ces deux α Reps (voir matériel et méthode). Comme discuté précédemment, l'objectif de cette fusion était de générer un modulateur « rigide ». En effet, afin de contraindre la conformation relative des interacteurs de la bibliothèque pour les décrocher, il est nécessaire que la structure du modulateur ne s'adapte pas à celle de la bi-P5H fixée sur sa cible. Par-là, on entend que la structure du modulateur, composée de deux domaines α Reps reconnaissant chacun un des deux domaines des bi-P5H, ne doit pas se déformer au niveau du point de jonction entre les deux α Reps pour permettre l'interaction, mais qu'au contraire c'est la bi-P5H qui doit se déformer au niveau du point de fusion entre les deux P5H lors de sa fixation sur le modulateur. La nature du lien peptidique entre les α Reps constituant le modulateur doit donc nécessairement être rigide, et nous avons décidé de connecter ces deux α Reps par un connecteur peptidique qui se replie en hélice α pour s'assurer de sa rigidité.

Comme la structure des α Reps est une succession d'hélices α , la stratégie employée a consisté à prolonger l'hélice C-ter (rose Figure 48) d'un des interacteurs α Rep (A12 ou D3) par la 2^{ème} hélice N-ter (verte Figure 48) de l'autre α Rep (D3 ou A12) afin d'obtenir un modulateur rigide. Dans les α Reps, les extrémités N-ter et C-ter sont du même côté, ainsi pour obtenir la structure décrite Figure 48, la fusion a été réalisée avec la 2^{ème} hélice du N-cap de l' α Rep localisée en C-ter de la fusion. La première hélice du N-cap de cette α Rep a donc été supprimée.

La structure des complexes A12-Tmut et D3-Omut n'étant pas connue au moment de la conception du modulateur, nous avons décidé de modéliser les deux types de modulateurs possibles : A12D3 et D3A12. La différence entre ces deux modulateurs est l'ordre dans lequel les deux α Reps les composants ont été fusionnées. Ainsi, A12 est du côté N-ter dans A12D3, et du côté C-ter dans D3A12 et inversement. De cette façon, on devrait pouvoir tester les différentes possibilités d'interaction entre le modulateur et les bi-P5H.

La conception des protéines chimères a été faite en deux étapes. Dans un premier temps une superposition géométrique de l'hélice commune aux deux structures initiales a été réalisée. Cette superposition a permis ensuite de construire un fichier PDB unique où les deux domaines sont reliés dans un même repère, avec une seule hélice de jonction dont la position résulte de la superposition. Les chaînes latérales de l'hélices de jonction de même que celles en contact avec cette hélice dans le modèle ont ensuite été optimisées, principalement pour éviter les conflits stériques, et lorsque c'est possible pour créer des interactions favorables. Cette optimisation ne permet pas de changement de

conformation du bi-domaine, mais explore des combinaisons de chaînes latérales favorables à cette conformation. La partie conception de séquence (*design*) a été réalisée en utilisant la version Standalone de Foldit (Kleffner, R. et al., 2017) qui permet de localiser de façon très rapide les zones à optimiser et de tester manuellement des combinaisons de chaînes latérales plus appropriées. Il n'est pas possible d'explorer exhaustivement et automatiquement toutes les combinaisons de chaînes latérales possibles mais il est possible de tester rapidement de nombreuses combinaisons et de retenir celles qui paraissent favorables, en se servant de la fonction de score de Foldit comme indicateur de stabilité.

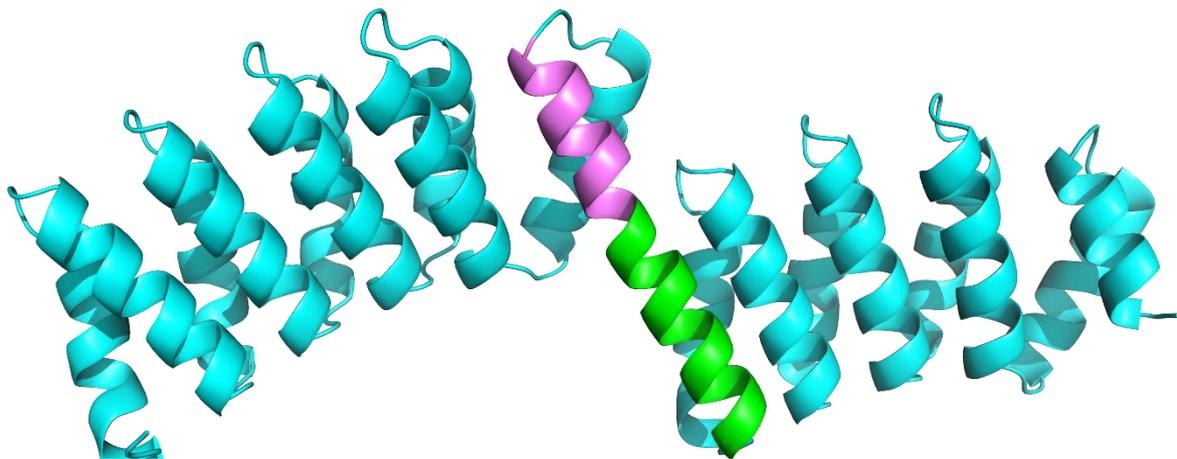


Figure 48 : Modèle d'anti-interacteur A12D3

Modèle d'une protéine résultant de la fusion entre les α Reps A12 et D3. La structure de la bFNE (une α Rep à 3 motifs HEAT internes sélectionnée contre la fibronectine) a été dupliquée, puis mutée dans Pymol pour faire correspondre sa séquence avec celle des α Reps A12 ou D3. La première hélice du N-cap de D3 a ensuite été supprimée. Les 10 derniers résidus de la dernière hélice de A12 (rose) ont été alignés dans l'axe de la deuxième hélice de l' α Rep D3 (verte) comme si ces deux hélices faisaient partie de la même hélice. Les deux protéines ont été fusionnées pour ne former qu'une seule chaîne polypeptidique. L'hélice qui sert de point de fusion a subi une minimisation d'énergie dans Foldit.

Les gènes synthétiques correspondants aux deux séquences des modulateurs bi-domaines modélisés ont été commandés et clonés dans PUR 23, un vecteur d'expression qui permet de cloner la séquence de la protéine d'intérêt par *Golden Gate* (Engler, C. et al., 2009) avec un His-tag en N-ter de sa séquence en acides aminés, afin de pouvoir tester leur expression dans différentes souches d'*E.coli*. Les gels correspondants à ces tests d'expression sont représentés ci-dessous (Figure 49).

A12D3 :

D3A12 :

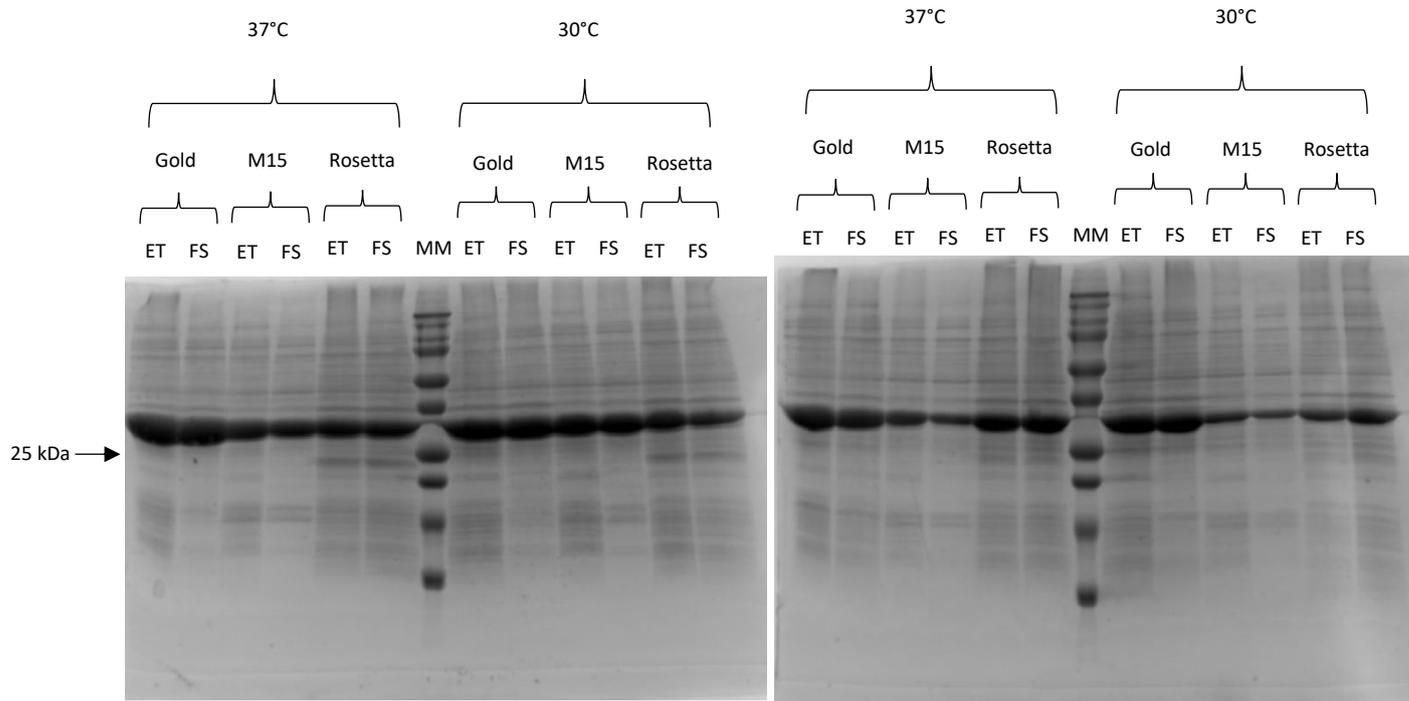


Figure 49 : Profils de migration des tests d'expression de A12D3 et D3A12 dans Gold, Rosetta, ou M15 à différentes températures.

Les anti-interacteurs A12D3 et D3A12 ont été produits dans les différentes souches d'E.coli indiquées et des fractions d'extraits totaux de ces cultures (ET) ou de leur fraction soluble (FS) ont été déposés sur gel de polyacrylamide.

Ces tests montrent que la production la plus efficace est celle réalisée dans Gold à 37°C, où les protéines sont fortement surexprimées de façon soluble. C'est donc dans ces conditions que les deux modulateurs ont été produits. Ils ont ensuite été purifiés d'abord sur colonne Ni-NTA puis par chromatographie d'exclusion de taille (Figure 50). Le profil d'élution des deux modulateurs semble indiquer que A12D3 et D3A12 se produisent et se purifient correctement (plusieurs dizaines de mg par litre de culture).

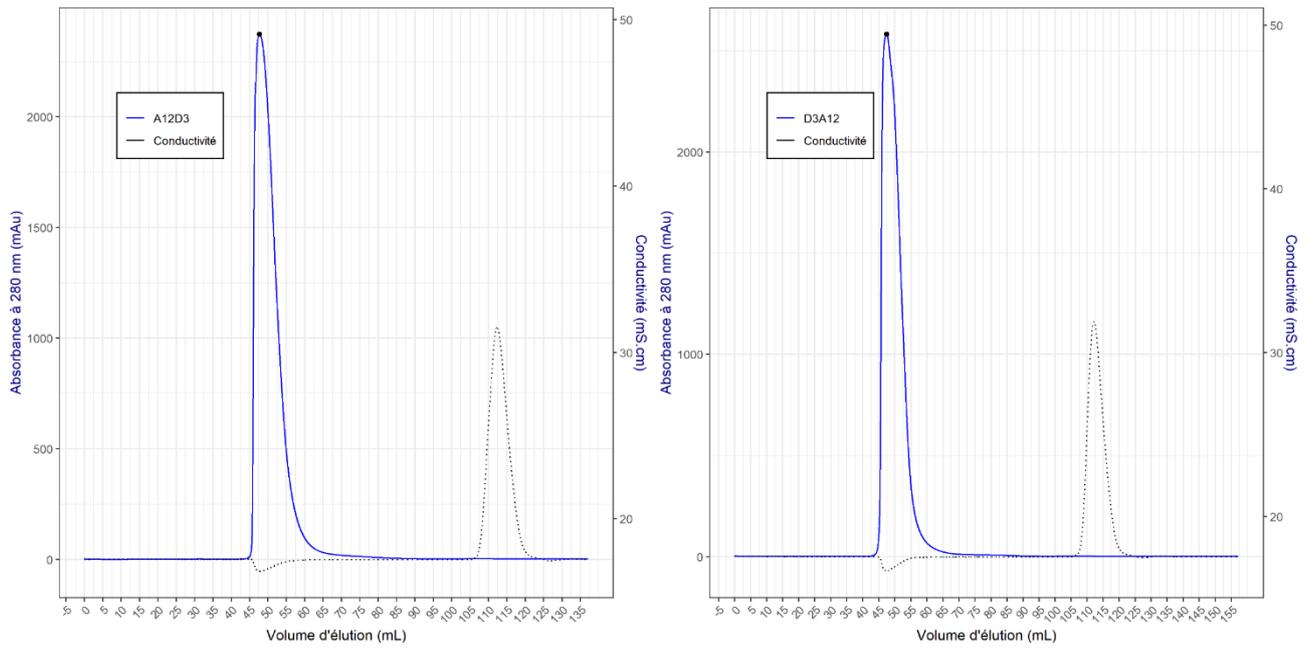


Figure 50 : Purification des modulateurs A12D3 et D3A12 par chromatographie d'exclusion stérique

IV. Caractérisation des modulateurs

A. Cristallisation des modulateurs

Des expériences de cristallographie ont été menées afin de vérifier si la structure du modulateur était proche du modèle *in silico*. Plus particulièrement, il était important de contrôler la structure secondaire au niveau du site de fusion entre les deux α Reps, pour s'assurer de la rigidité du modulateur.

Si A12D3 a donné des cristaux et une structure exploitable, ce n'est pas le cas de D3A12 qui n'a donné de cristaux ni dans les plaques commerciales de type PEGs ni dans les plaques de criblage. La structure obtenue pour A12D3 à 3.54 Å est présentée ci-dessous (Figure 52), superposée au modèle présenté Figure 48.

On peut tout de même discuter de l'intérêt d'obtenir la structure pour D3A12. En effet, d'après les résultats expérimentaux obtenus en DSC (Figure 51), il y a une différence de comportement entre A12D3 et D3A12. On observe deux pics de dénaturation pour A12D3, un premier assez faible vers 66°C et un deuxième plus grand vers 77°C. En revanche, pour D3A12, on n'observe qu'un seul pic à 77°C. Ces résultats peuvent être associés à une structure tridimensionnelle différente, et il serait donc intéressant d'obtenir la structure de D3A12 afin de vérifier si les différences observées sont liées à sa structure.

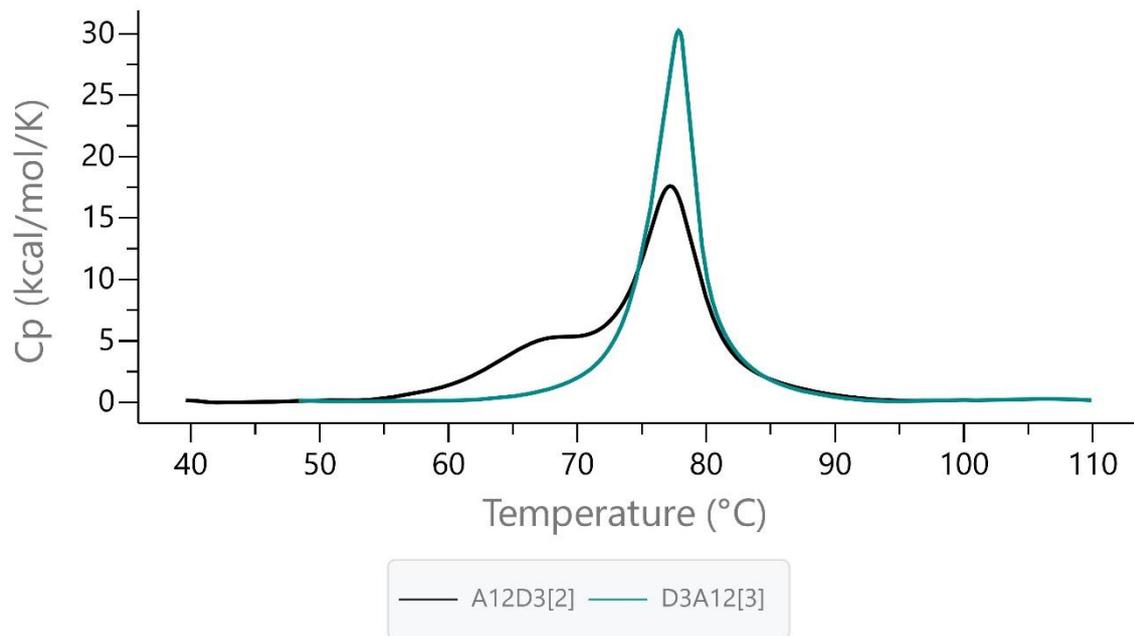


Figure 51 : Profils de dénaturation des anti-interacteurs A12D3 (courbe noire) et D3A12 (courbe verte) en DSC

Stabilité thermique des modulateurs A12D3 (noir) et D3A12 (vert) mesurée par des expériences de DSC (Differential Scanning Calorimetry).

Analyse comparée de la structure expérimentale de l' α Rep bi-domaine
relativement à la structure conçue

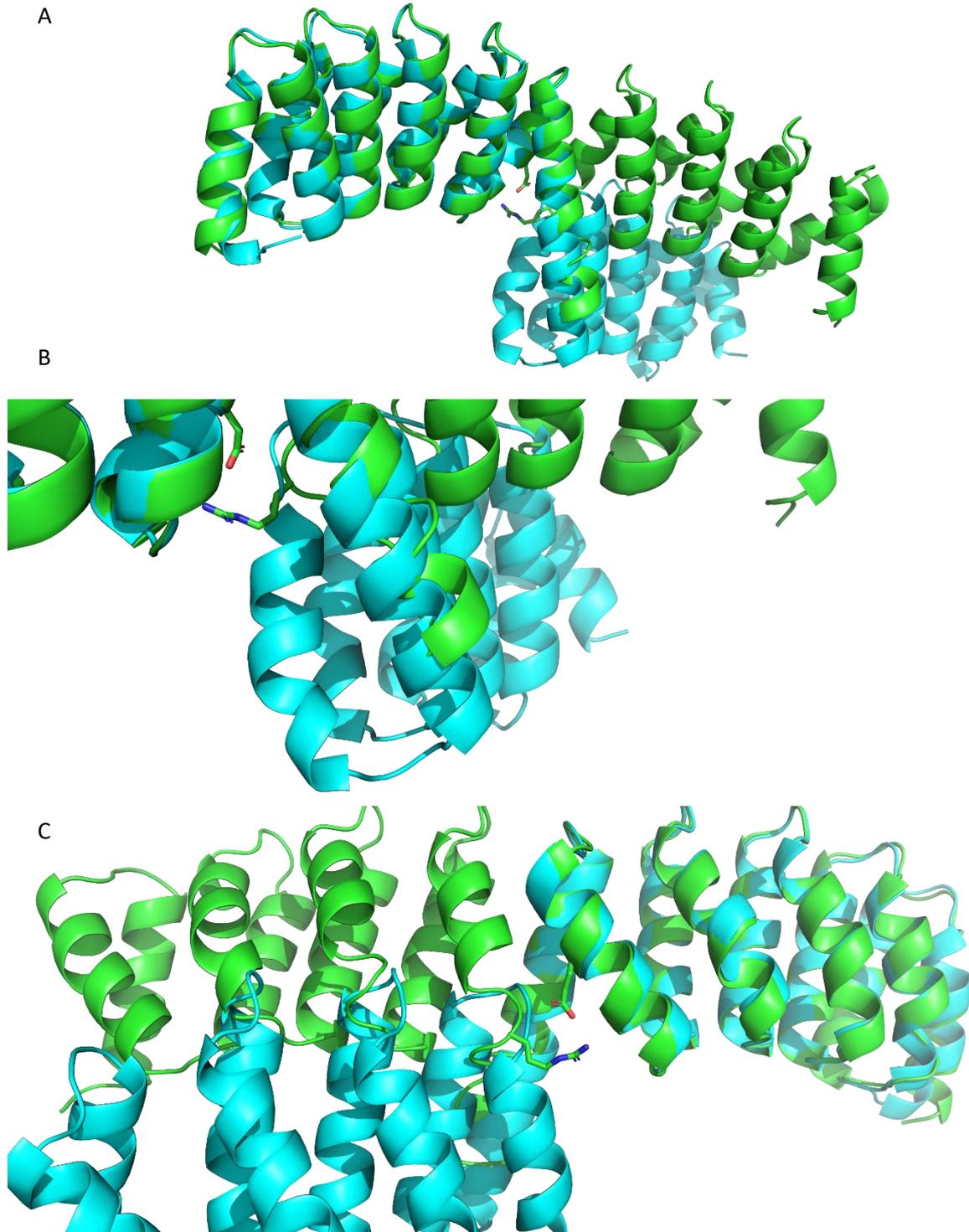


Figure 52 : superposition de la structure de A12D3 à 3.54 Å (vert) avec le modèle *in silico* (cyan)

A : vue d'ensemble ; B : zoom sur la fin de l'hélice partagée ; C : zoom sur le pont salin de la structure d'A12D3 (vert)

On peut voir sur cette superposition qu'il est possible de superposer l'un ou l'autre domaine du modèle sur la structure mais que superposer les deux simultanément n'est pas possible. Cela résulte d'une différence d'orientation relative des deux domaines de A12D3 entre le modèle et la structure expérimentale. La différence principale se situe au niveau de l'hélice qui a servi à fusionner les deux α Reps. La structure expérimentale indique qu'il « manque » un tour d'hélice par rapport au modèle (Figure 52.B), ce qui donne lieu à une différence dans l'orientation de la boucle qui suit. Le second domaine adopte une orientation qui lui permet d'interagir avec l'hélice de jonction de façon plus compacte que ce qui était conçu. Il est intéressant de voir l'existence dans la structure d'un pont salin qui n'était pas prévu, entre le glutamate 144 et l'arginine 179 (Figure 52.C). Ce pont salin pourrait en partie expliquer la différence de « *packing* » entre la structure et le modèle, la structure possédant un *packing* plus compact.

Dans l'ensemble, la structure du modulateur A12D3 comporte des caractéristiques recherchées dans le modèle. Les deux domaines α Reps adoptent une conformation fixe, avec leur surface d'interaction accessible. Bien que l'hélice partagée entre les deux α Reps soit plus courte que dans le modèle, le point de jonction est bien l'hélice α , comme initialement conçu. Si cette architecture est rigide, elle doit permettre l'interaction simultanée de la bi-P5H avec les deux domaines du modulateur sans que celui-ci subisse de changement conformationnel majeur.

L'étape suivante consiste à confirmer l'interaction entre les bi-P5H et l'anti-interacteur par des expériences d'interactions.

B. Etudes des interactions entre les modulateurs et les bi-P5H

Des expériences de SPR ont été menées entre les bi-P5H et les deux modulateurs, pour confirmer l'interaction d'une part, mais également afin d'évaluer l'affinité relative du modulateur avec l'une et l'autre des mono-P5H, ainsi qu'avec la P5H bi-domaine. La mesure de la constante de dissociation K_D pour ces partenaires nous permettrait de vérifier l'augmentation attendue de l'affinité des modulateurs pour les bi-P5H comparé aux mono-P5H. En effet, le passage de un à deux sites de fixation fait intervenir le phénomène d'avidité discuté précédemment et qui devrait augmenter l'affinité apparente des modulateurs pour les bi-P5H, d'une part en diminuant le k_{off} (les deux domaines doivent se décrocher simultanément pour qu'il y ait dissociation), et d'autre part en facilitant la fixation d'un domaine de la bi-P5H si l'autre domaine est fixé sur le modulateur (augmentation de la concentration effective localement au niveau du site de fixation). Les modulateurs (A12D3 et D3A12) ont été produits biotinylés, et immobilisés à la surface d'une puce recouverte de streptavidine (voir matériel et méthode). Les bi-P5H ont été utilisées comme analytes circulants.

Les résultats de ces expériences sont synthétisés [Figure 53](#).

Model	Buffer	K_D (nM)
TOLong - AD		
1:1	TBS-P20	26,9 ± 0,7
1:1	PBS-P20	15 ± 15
TOLong - DA		
1:1	TBS-P20	6,3 ± 0,3
1:1	PBS-P20	21 ± 12
OTLong - AD		
1:1	TBS-P20	13,0 ± 0,5
1:1	PBS-P20	5,7 ± 0,2
OTLong - DA		
1:1	TBS-P20	2,9 ± 0,1
1:1	PBS-P20	5,7 ± 0,2

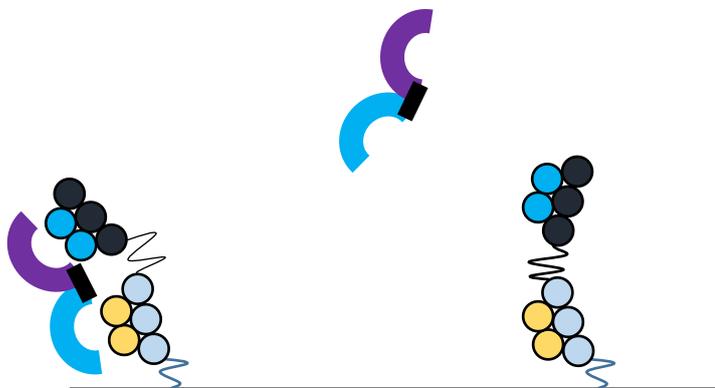


Figure 53 : Etude des interactions entre les modulateurs et les bi-P5H par SPR

L'affinité des deux modulateurs A12D3 (noté AD) et D3A12 (noté DA) pour les deux modèles de bi-P5H reliés par le linker long, Twt-Owt (noté TOLong) et Owt-Twt (noté OTLong) a été mesurée par SPR dans un tampon 20 mM Tris 300 mM NaCl 0.05% Tween (noté TBS-P20) ou dans un tampon Dulbecco's Phosphate Buffered Saline (Sigma) contenant 150 mM NaCl et complété avec 0.05% Tween (noté PBS-P20). Les données présentées ici ont été obtenues avec un modèle 1:1 qui rend compte de l'affinité globale d'un partenaire pour l'autre. Les bi-P5H biotinylées ont été immobilisées sur des puces recouvertes de streptavidine, et les modulateurs ont été passés comme analytes.

Ces résultats montrent que globalement, les modulateurs fixent les deux bi-P5H de façon équivalente (avec une légère préférence pour Owt-Twt long). On peut également noter que les interactions semblent meilleures dans le tampon PBS 150 mM NaCl pour A12D3, alors qu'elles semblent meilleures dans le tampon TBS 300 mM NaCl pour D3A12. Cette information pourrait nous renseigner sur la nature des interactions puisque les interactions électrostatiques ont tendance à être inhibées par l'augmentation de la concentration en NaCl.

Ces expériences valident donc l'existence d'une interaction entre les modulateurs et les bi-P5H avec une constante de dissociation de l'ordre de quelques nM, c'est-à-dire 3 ordres de grandeur de moins que les K_D obtenus avec les mono-P5H.

L'interaction simultanée des deux domaines du modulateur avec les deux domaines d'une bi-P5H est donc possible et conduit effectivement à une interaction de forte affinité entre les bi-domaines. Cette interaction des bi-P5H avec les modulateurs ayant été confirmée, l'objectif était désormais d'obtenir la structure des complexes pour avoir accès aux surfaces d'interaction impliquées, et à l'orientation du modulateur vis à vis de la bi-P5H.

C. Cristallogénèse des complexes bi-P5H/modulateur

Des plaques 96 puits correspondant à des kits commerciaux ont été utilisées pour cristalliser les complexes modulateur/bi-P5H. Cependant, ces cribles n'ont pas donné de cristaux à ce jour. Cela est peut-être dû à la faible concentration des bi-P5H dans la goutte (1.3 mg/mL environ). En effet, les bi-P5H (et les mono-P5H) sont difficiles à concentrer, et leur concentration s'arrête généralement autour de 1 à 2 mg/mL, ce qui est un peu faible pour obtenir des cristaux. Une des pistes envisagées serait de concentrer les bi-P5H avec les modulateurs pour voir si la présence d'un ligand permet de mieux concentrer les bi-P5H et d'atteindre environ 10 mg/mL, concentration plus adaptée à l'obtention de cristaux.

D. Modélisation du modulateur avec AlphaFold

AlphaFold est le logiciel développé par l'entreprise DeepMind pour réaliser de la prédiction de structure des protéines (Jumper, J. *et al.*, 2021), et qui a conduit à une spectaculaire amélioration des capacités de prédiction de structures, lors des tests de prédictions à l'aveugle CASP. Nous avons profité de la mise à disposition publique très récente du logiciel AlphaFold pour comparer *a posteriori* le modèle que nous avons conçu pour A12D3 au meilleur modèle généré par AlphaFold (appelé *rank_1*) ainsi qu'à la structure cristallographique. Les 3 structures ont été superposées par rapport à leur domaine N-ter qui correspond à l' α Rep A12 (Figure 54).

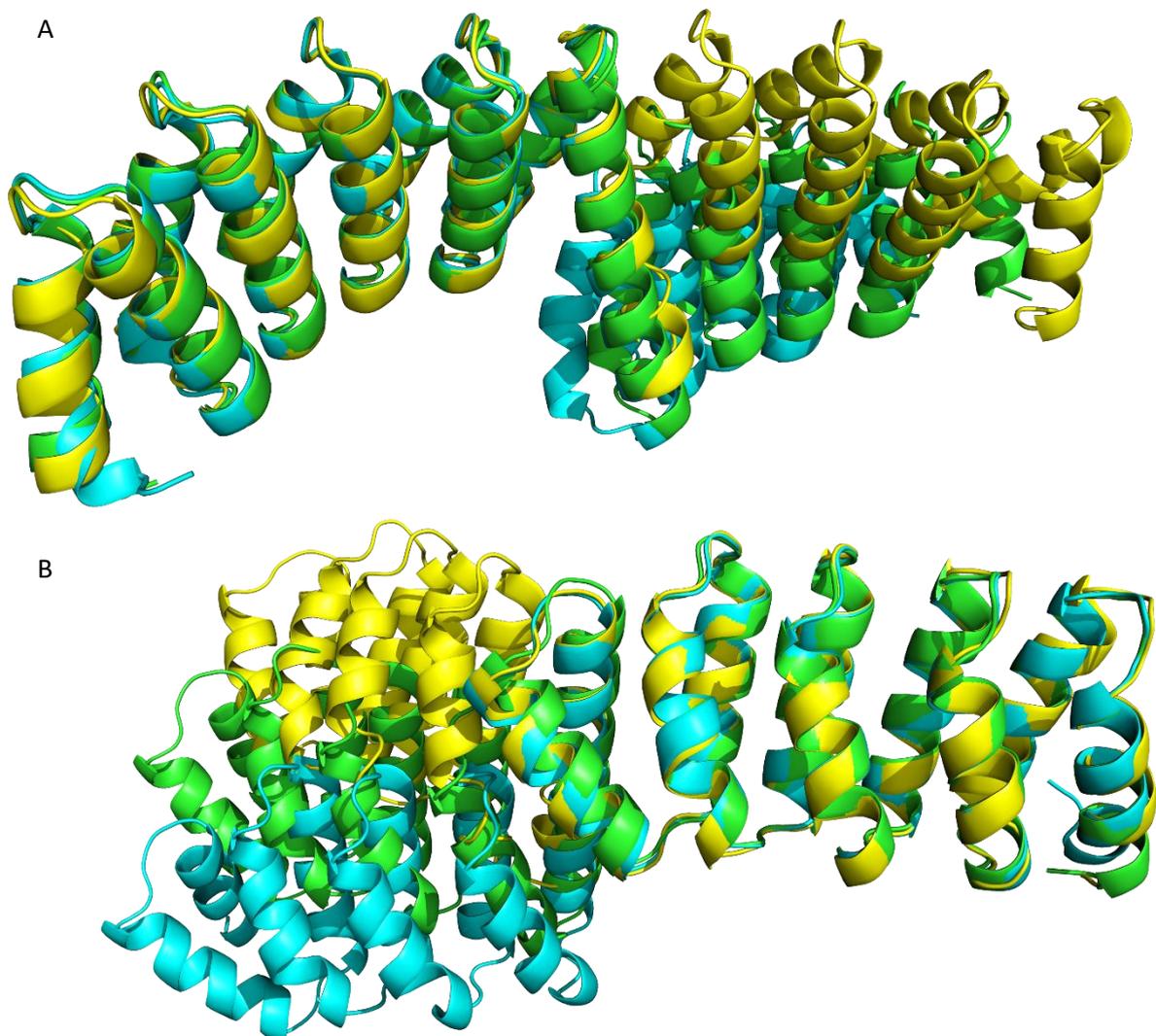


Figure 54 : comparaison de la structure d'A12D3 avec le modèle conçu par l'équipe et le modèle généré par AlphaFold

Il est intéressant de voir ici que [modèle prédit par AlphaFold](#) est intermédiaire entre le [modèle que nous avons conçu](#) et la [structure cristallographique](#). Si on regarde le profil LDDT (*Local Distance Difference Test*) ([Figure 55](#)), qui correspond à un indice de confiance position-dépendant donné par AlphaFold, on remarque que l'hélice fusion entre les deux α Reps est la zone la moins bien prédite.

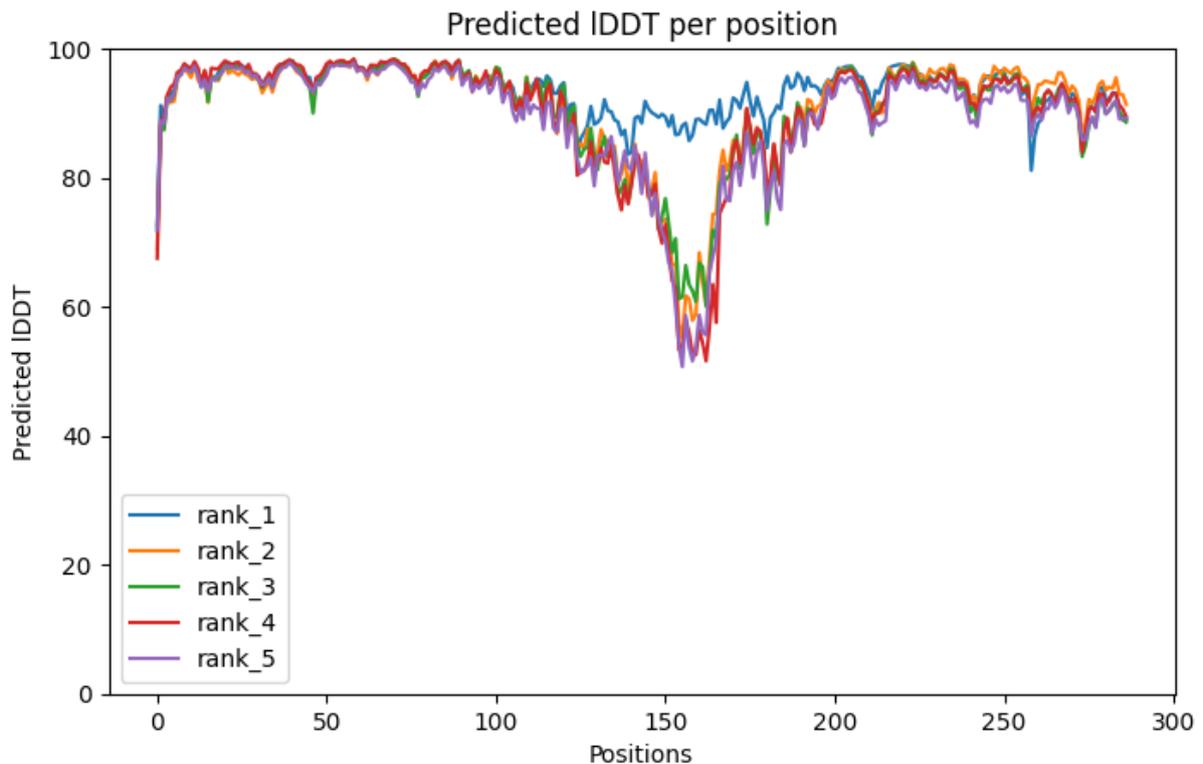


Figure 55 : profil IDDT des 5 meilleurs modèles générés par AlphaFold pour le modulateur A12D3

À partir de ce modèle, nous avons exploré d'autres combinaisons de 5 résidus jugés sous-optimaux et effectué une nouvelle prédiction avec AlphaFold à partir de la séquence mutée, appelée A12D3v2 (A12D3 version 2). Nous avons également utilisé la fonction de calcul d'énergie du logiciel FoldIt pour comparer le niveau d'énergie des différents modèles. Avant l'avènement d'AlphaFold, Rosetta, dont l'outil de calcul d'énergie est utilisé dans FoldIt, était le logiciel permettant de prédire les structures avec le plus de précision. Il est intéressant de noter qu'AlphaFold ne calcule pas l'énergie de la protéine, ce qui peut causer des erreurs de prédiction lorsqu'on décide d'évaluer le repliement d'un mutant par exemple. Nous avons donc décidé d'évaluer de façon complémentaire les modèles proposés avec la fonction de calcul d'énergie de Rosetta implémentée dans FoldIt. Le modèle prédit par AlphaFold pour A12D3v2 donne un score d'énergie plus faible (donc meilleur) que le modèle que nous avons généré. De même, le profil IDDT des 5 meilleurs modèles générés par AlphaFold pour A12D3v2 ([Figure 56](#)) montre un indice de confiance bien plus élevé, même au niveau de l'hélice-fusion. Ces analyses suggèrent que si AlphaFold n'est pas un outil adapté au *design* de protéines comme l'est Rosetta, cela peut constituer un outil utile de vérification indépendant et complémentaire (*back*

checking) des séquences conçues *in silico*. Il est remarquable d'observer que nous n'avons pas identifié la structure exacte que AlphaFold propose, alors même que dans l'outil de *design* utilisé son score de stabilité est plus favorable que la meilleure solution que nous avons identifiée. Il est ensuite possible de poursuivre l'amélioration du *design* (comme pour les substitutions que nous proposons pour la version 2) par une approche itérative. Nous avons commandé un gène synthétique codant pour le modulateur A12D3v2, dans le but d'obtenir sa structure cristallographique et de voir si les mutations ont permis à la protéine de se replier selon le modèle prédit par AlphaFold.

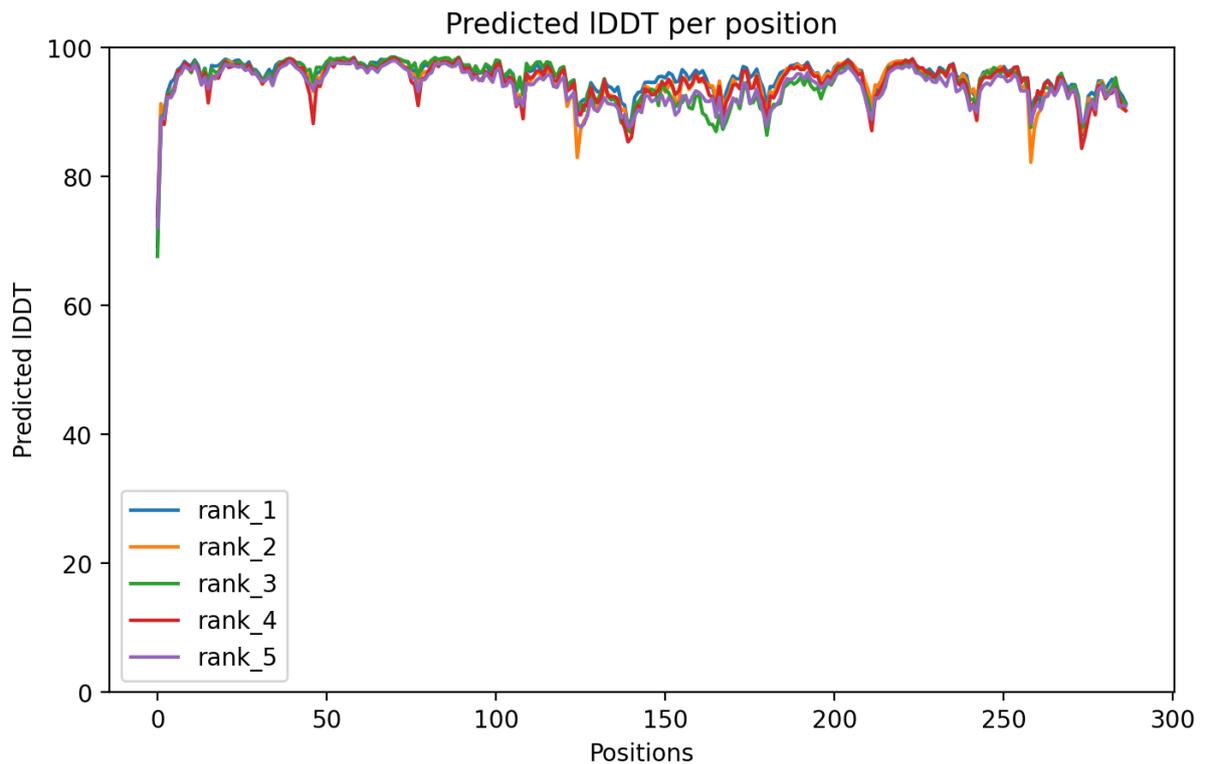


Figure 56 : Profil IDDT des 5 meilleurs modèles générés par AlphaFold pour A12D3v2

Partie 3 : Applications de la bibliothèque de bi- P5H

La bibliothèque de bi-P5H et le modulateur ayant été produits, la suite de mon projet de thèse a consisté à vérifier d'une part que la bibliothèque de bi-P5H pouvait donner des interacteurs de plusieurs cibles protéiques non apparentées, et d'autre part, que le modulateur pouvait permettre de moduler cette interaction. Dans ce but, des sélections par phage *display* ont été réalisées avec la bibliothèque de bi-P5H contre l'eGFP, choisie comme protéine modèle puis contre les cibles EpCAM et TROP 2, des marqueurs de cellules tumorales circulantes.

Nous avons choisi de travailler sur l'eGFP pour établir la preuve de concept car c'est une protéine stable, bien repliée, et qui a déjà été utilisée comme cible par le laboratoire pour faire des sélections d' α Reps. Il sera donc possible de comparer les résultats obtenus avec la bibliothèque d' α Reps développée par le laboratoire et ceux obtenus avec la bibliothèque de bi-P5H développée pendant mon projet de thèse.

Par ailleurs, nous avons cherché à savoir si les interacteurs issus de la bibliothèque étaient capables de reconnaître des protéines membranaires. En effet, les sélections par phage *display* de protéines membranaires sont réalisées sur des protéines solubilisées en détergent ou sur les domaines extracellulaires de ces protéines. Les interacteurs issus de ces sélections ne sont pas toujours capables de reconnaître leur cible dans leur environnement naturel, c'est-à-dire à la surface des cellules. Cette capacité à fixer des protéines de surface cellulaire a été évaluée pour la bibliothèque de bi-P5H comme pour la bibliothèque d' α Rep, puisqu'aucune de ces deux bibliothèques n'avait été utilisée pour ce type de cible. Les cibles choisies pour répondre à cette question sont les protéines EpCAM et TROP 2, les marqueurs de cellules tumorales circulantes présentés dans l'introduction.

I. Evaluation de la bibliothèque de bi-P5H sur l'eGFP

A. Sélection par phage *display* d'interacteurs bi-P5H contre l'eGFP

Les sélections d'interacteurs par phage *display* ont d'abord été réalisées sur l'eGFP (Tsien, R. Y., 1998), pour vérifier la capacité de la bibliothèque de bi-P5H à fournir des interacteurs d'une cible protéique. L'eGFP est une protéine fluorescente rendue plus brillante que la GFP *wild type* par la substitution S65T, et qui adopte également un repliement en tonneau β (Figure 57).

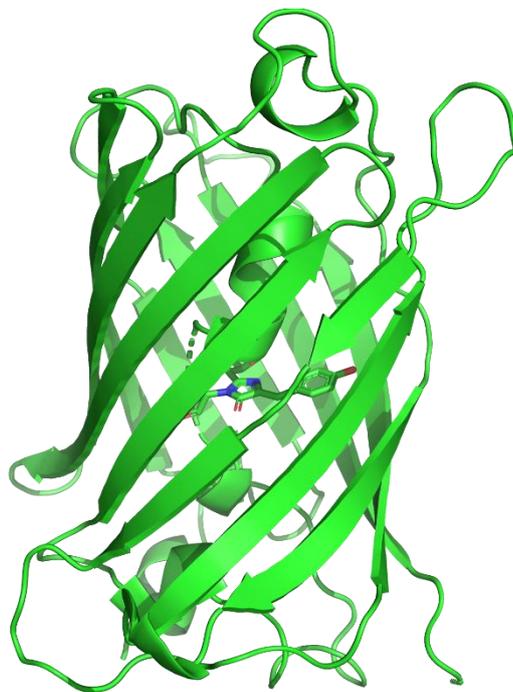


Figure 57 : structure cristallographique de l'eGFP (code PDB : 6YLQ) à 1.65 Å

La protéine est représentée en *cartoon*, et le chromophore est représenté en *stick*. Le chromophore correspond aux résidus 65 à 67 et est responsable de la fluorescence de la protéine.

De l'eGFP possédant un His-tag en N-ter a été produite et purifiée (résultats non montrés) pour réaliser les sélections.

L'eGFP ainsi produite a été immobilisée sur plaque selon un protocole décrit en détail dans le matériel et méthode. Dans les grandes lignes, l'eGFP a été immobilisée par son His-tag grâce à une molécule appelée Tris-NTA (Tris- acide nitrilotriacétique) (BiotechRabbit), qui permet de capturer les protéines

possédant un His-tag avec une forte affinité (K_D de l'ordre du nM) (Reichel, A. *et al.*, 2007). Cette molécule consiste en trois groupements NTA et peut être biotinylée, ce qui permet de l'immobiliser par de la streptavidine. Huit puits d'une plaque ELISA ont donc été recouverts par de la streptavidine, puis par du Tris-NTA biotinylé, ce qui a permis d'immobiliser l'eGFP possédant un His-tag. L'eGFP ainsi immobilisée a été mise au contact des phages de la bibliothèque de bi-P5H, et après plusieurs étapes de lavages, les phages retenus sur l'eGFP ont été élués par l'application d'un pH acide (glycine pH 2.5). Les phages ainsi élués ont été utilisés pour infecter des bactéries, dans le but de reconstituer un ensemble de phages enrichi en phages interagissant avec l'eGFP, et c'est ce nouvel ensemble qui a été utilisé lors d'un deuxième tour de sélection identique au premier. Après 3 tours de sélection, l'ensemble de phages récupérés peut être assez enrichi en phages interagissant avec la cible. Des colonies isolées issues des tours de sélection 1, 2, et 3 ont alors été repiquées sur une plaque de culture 96 puits, et un phage ELISA clonal a été réalisé selon le protocole décrit dans le matériel et méthode. Les résultats de ce phage ELISA clonal sont présentés Figure 58.

A

	1	2	3	4	5	6	7	8	9	10	11	12
A	0,089	0,086	0,129	0,098	0,081	0,177	0,169	0,083	0,078	0,08	0,082	0,11
B	0,197	0,1	0,09	0,091	0,084	0,102	0,083	0,106	0,091	0,093	0,094	0,094
C	0,109	0,132	0,084	0,093	0,087	0,079	0,081	0,08	0,092	0,08	0,082	0,087
D	0,508	0,109	0,085	0,093	0,148	0,08	0,107	0,102	0,127	0,097	0,094	0,093
E	0,193	0,137	0,111	0,096	0,101	0,13	0,184	0,111	0,142	0,154	0,143	0,078
F	0,092	0,097	0,085	0,119	0,097	0,081	0,093	0,112	0,088	0,145	0,09	0,086
G	0,129	0,122	0,08	0,096	0,606	0,108	0,171	0,074	0,1	0,093	0,331	0,079
H	0,181	0,094	0,08	0,094	0,08	0,08	0,149	0,077	0,085	0,079	0,09	0,088

B

	1	2	3	4	5	6	7	8	9	10	11	12
A	0,109	0,097	0,117	0,102	0,122	0,093	0,204	0,136	0,154	0,134	0,09	0,133
B	0,095	0,104	0,107	0,091	0,096	0,097	0,089	0,099	0,087	0,083	0,082	0,097
C	0,133	0,098	0,1	0,167	0,128	0,085	0,08	0,102	0,081	0,109	0,098	0,918
D	0,11	0,096	0,102	0,105	0,095	0,093	0,081	0,128	0,079	0,085	0,078	0,095
E	0,148	0,128	0,116	0,087	0,093	0,089	0,15	0,126	0,082	0,082	0,076	0,106
F	0,112	0,093	0,1	0,1	0,108	0,079	0,104	0,098	0,126	0,09	0,084	0,086
G	0,157	0,113	0,13	0,136	0,14	0,086	0,133	0,083	0,08	0,087	0,128	0,112
H	0,105	0,091	0,105	0,09	0,126	0,113	0,154	0,12	0,101	0,088	0,086	0,091

Figure 58 : Crible des sélections bi-P5H contre l'eGFP par phage ELISA clonal

Dans cette expérience l'eGFP est immobilisée sur deux plaques ELISA 96 puits dans les puits A, C, E, G. Les puits B, D, F, G ne présentent donc pas la cible et servent à déceler les clones qui interagissent de façon non spécifique. Ici, on a considéré que l'interaction était spécifique à partir d'un ratio signal spécifique/signal aspécifique supérieur à 2.

96 clones de bi-P5H obtenus à l'issue de la sélection sont exposés sur phages et incubés dans chaque puit (1 clone/puit). La présence des phages dans les puits est révélée à l'aide de l'anticorps anti-phage M13 couplé à l'HRP. Le produit coloré formé dans la réaction catalysée par la peroxydase est visible et quantifiable à 410 nm, et proportionnel au nombre de phages retenus dans le puit qui par extrapolation peut être associé à l'existence d'une affinité de la protéine exposée pour la cible. Sur les 96 clones testés, 8 semblent donner un signal spécifique (encadrés violet Figure 58) et ont été mis en

culture pour extraire les phagemides correspondants. Leur ADN a été digéré par les enzymes de restriction BglIII et NheI afin de vérifier si ces interacteurs correspondaient à des P5H mono-domaine ou bi-domaines. Les résultats de cette double digestion sont présentés [Figure 59](#).

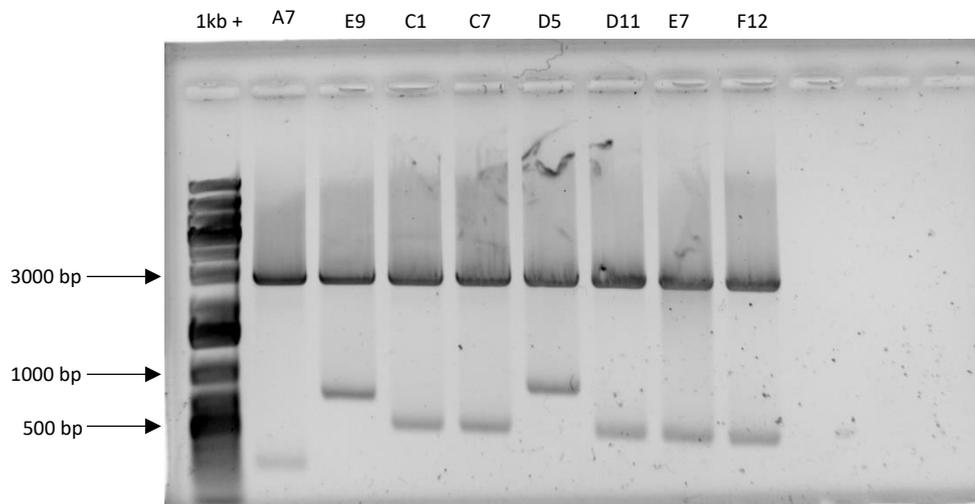


Figure 59 : profil de migration par électrophorèse des produits de digestion par BglIII et NheI des interacteurs bi-P5H issus des sélections contre l'eGFP

Environ 700 ng d'ADN doublement digéré par les enzymes de restriction BglIII et NheI ont été déposés sur gel d'agarose 1%. Pour la double digestion par les enzymes de restriction BglIII/NheI d'une bi-P5H dans le vecteur pDip 3.0, on attend une bande à 3030 pb et une bande à 928 pb. La même digestion d'une mono-P5H dans ce vecteur donne une bande à 3030 pb et une bande à 559 pb.

On peut voir [Figure 59](#) que les clones E9 et D5 semblent correspondre à des bi-P5H (bande à 928 pb) tandis que les autres clones semblent correspondre à des mono-P5H (bande à 559 pb). Le clone A7 donne un fragment trop petit pour être une mono-P5H et pourrait correspondre à un artefact de la sélection, ou de l'ELISA.

Tous ces clones ont été séquencés. Les clones correspondant à des mono-P5H (C1, C7, D11, E7, et F12) sont redondants et codent pour la même protéine. Le clone E9 correspond bien à une bi-P5H mais possède un décalage du cadre de lecture assez tôt dans sa séquence en acides aminés. Seul le clone D5 correspond bien à une bi-P5H entière.

On peut retenir de cette sélection qu'elle a permis de montrer que la bibliothèque de bi-P5H semblait en mesure de fournir des interacteurs de cibles protéiques. En effet, à l'issue de la sélection, on a bien obtenu la séquence d'une bi-P5H sélectionnée contre l'eGFP. La prochaine étape consiste donc à confirmer l'interaction entre cette bi-P5H (D5) et l'eGFP pour valider la bibliothèque de bi-P5H.

B. Étude des interactions bi-P5H sélectionnées/cibles par tamis analytique

L'interaction entre la bi-P5H sélectionnée contre l'eGFP (D5) et sa cible a tout d'abord été étudiée par tamis moléculaire analytique. Il est en effet possible d'observer la formation d'un complexe par tamis analytique car le complexe protéique, plus volumineux, aura un volume d'éluion plus faible que celui des protéines seules. Pour étudier cette interaction, la bi-P5H D5 (Figure 60, courbe bleue), l'eGFP (Figure 60, courbe verte), ou un mélange équimolaire des deux (Figure 60, courbe rouge) ont été analysés par chromatographie d'exclusion analytique qui permet de séparer les molécules selon leur taille. Les échantillons ont été injectés sur une colonne de chromatographie d'exclusion de taille superdex s75 10/300 GL, et les profils d'éluion obtenus sont représentés sur la Figure 60.

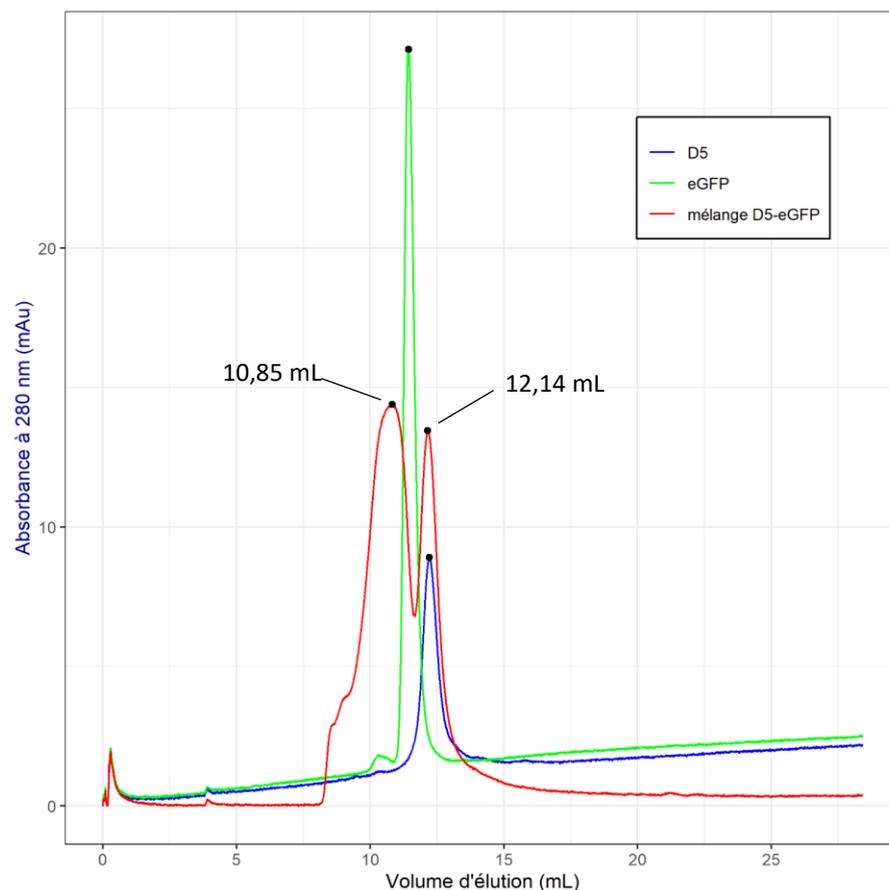


Figure 60 : Étude de l'interaction eGFP-D5 par tamis analytique.

Profil d'éluion d'une centaine de nanomoles de la bi-P5H D5 (courbe bleue), de l'eGFP (courbe verte), ou d'un mélange équimolaire des deux (courbe rouge) sur une colonne superdex 75 10/300 GL. L'absorbance à 280 nm est représentée sur l'axe des ordonnées et le volume d'éluion de l'échantillon est représenté en abscisse.

Le mélange équimolaire de D5 et eGFP donne deux pics, un dont le maximum est à 10.85 mL et un autre dont le maximum se situe à 12.14 mL. Le premier pic a été élué à un volume plus faible (c'est-à-

dire avec un rayon de Stokes plus élevé) que ceux obtenus avec les protéines seules. Ceci pourrait donc à la formation d'un complexe composé d'une copie du monomère D5 et d'une copie du monomère d'eGFP. Pour vérifier cette hypothèse, les fractions d'élution correspondant à chaque pic d'élution des 3 expériences ont été déposés sur gel d'acrylamide, et une électrophorèse a été réalisée. Le profil de

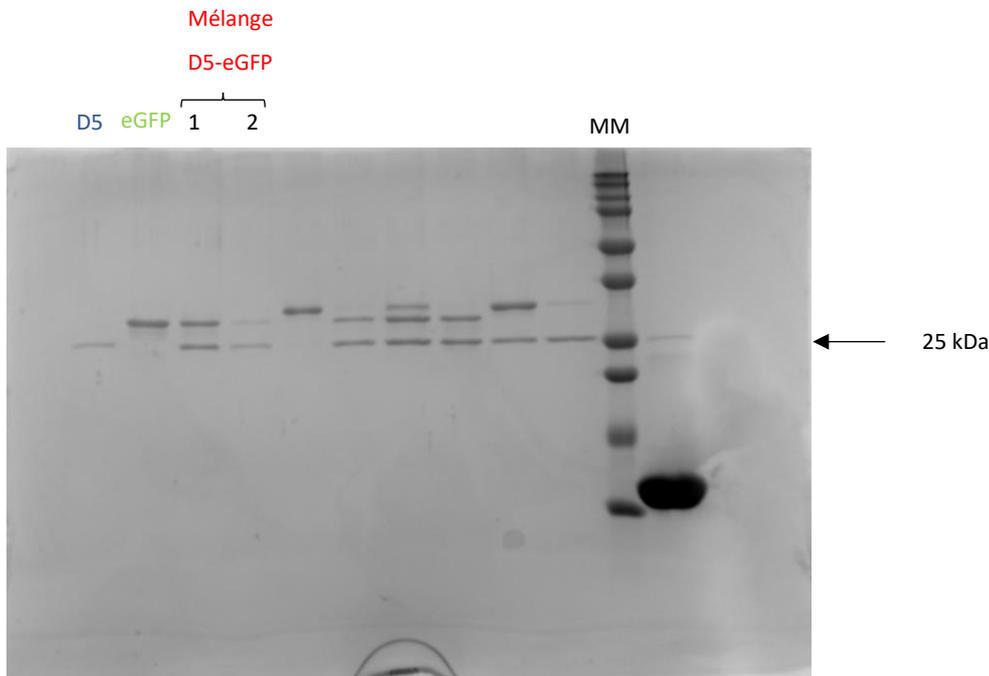


Figure 61 : Étude de l'interaction D5-eGFP par tamis analytique.

migration est présenté [Figure 61](#).

Les fractions collectées dans les différentes expériences de tamis ([Figure 60](#)) ont été déposées sur gel d'acrylamide pour vérifier les protéines présentes dans chaque pic. On peut voir que le pic d'élution sortant à 10.85 mL (noté 1) lors du tamis analytique du mélange D5-eGFP contient les deux protéines, confirmant ainsi la formation du complexe. Le pic sortant à 12.14 mL est noté 2.

La [Figure 61](#) nous renseigne sur les protéines présentes dans chaque pic d'élution lors du tamis analytique. Les deux protéines, D5 et eGFP, sont présentes dans la fraction qui correspond au premier pic d'élution (10.85 mL) [Figure 60](#). Cette expérience nous permet donc de dire que la bi-P5H D5 reconnaît l'eGFP, contre laquelle elle a été sélectionnée, avec une affinité suffisamment forte pour que le complexe soit visible en tamis analytique. Il serait intéressant de refaire cette expérience en diminuant la quantité de D5 dans le mélange, pour qu'elle soit majoritairement dans le complexe et ainsi voir disparaître le pic d'élution de D5 seule.

L'interaction entre la bi-P5H D5 et l'eGFP ayant été confirmée, nous avons cherché à connaître plus précisément la constante d'affinité du complexe, en réalisant des expériences de *Bi-layer interferometry* (BLI).

C. Étude des interactions bi-P5H sélectionnées/cibles par BLI

Une fois cette interaction validée notre objectif a été d'une part de déterminer quelle était l'affinité de ce binder pour l'eGFP et d'autre part si l'on pouvait moduler cette interaction avec le modulateur allostérique A12D3.

L'interaction entre la bi-P5H D5 et l'eGFP a tout d'abord été testée par BLI. Cette technique permet de caractériser les constantes cinétiques d'une interaction, le k_{on} et le k_{off} de l'interaction, ce qui donne accès à la constante de dissociation du complexe K_D , égale au rapport k_{off}/k_{on} (cf. Introduction). Pour cela, l'eGFP a été produite biotinylée et fixée sur des sondes recouvertes de streptavidine. Ces sondes ont alors été mises en contact d'une solution contenant différentes concentrations de la bi-P5H D5 (résultats non montrés).

On obtient un K_D de 16.4 ± 0.1 nM entre la bi-P5H D5 et l'eGFP, confirmant ainsi la capacité de la bibliothèque à fournir des interacteurs de forte affinité pour des cibles protéiques.

Nous avons ensuite testé l'effet du modulateur allostérique sur cette interaction, en pré-incubant la bi-P5H D5 avec le modulateur A12D3 avant de mettre ce mélange en présence de l'eGFP immobilisée sur la sonde streptavidine. Les résultats de ces expériences sont présentés [Figure 62](#).

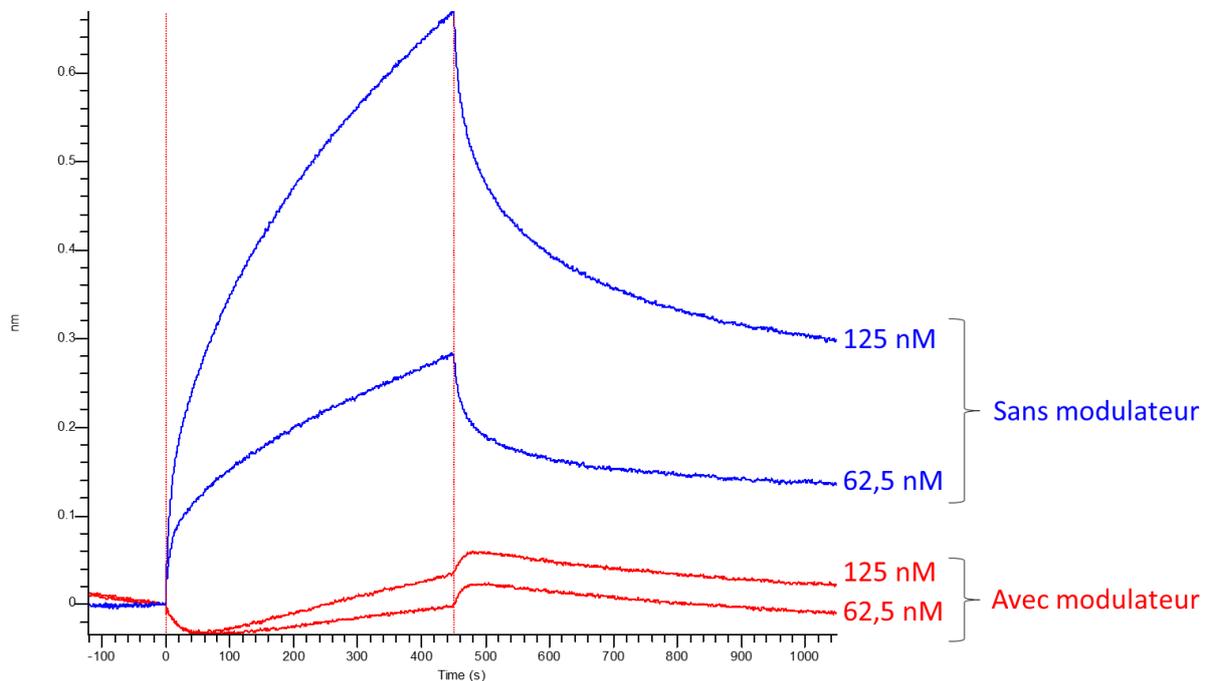


Figure 62 : Mesure de l'effet du modulateur sur l'interaction eGFP-D5

Les résultats présentés [Figure 62](#) montrent une forte diminution dans la fixation de la bi-P5H D5 sur l'eGFP lorsque D5 est pré-incubée avec le modulateur A12D3. Cela témoigne donc de la capacité du modulateur à inhiber l'interaction entre la bi-P5H et sa cible.

À ce stade, nous avons pu montrer la capacité du modulateur à inhiber l'interaction entre une bi-P5H et sa cible. Nous n'avons pas montré ici que le modulateur est capable de décrocher la bi-P5H lorsque cette dernière est déjà engagée dans un complexe avec sa cible. Des expériences complémentaires sont donc nécessaires, et des expériences de « pull-down » et de purification de GFP non étiquetée sont planifiées.

II. Evaluation de la bibliothèque de bi-P5H sur des protéines membranaires

La capacité de la bibliothèque de bi-P5H à fournir des interacteurs de cibles protéiques ayant été validée avec la cible eGFP, nous avons cherché à obtenir des interacteurs de protéines membranaires. La purification de cellules entières fait en effet partie des potentielles applications de la bibliothèque et nous avons déjà discuté de l'intérêt de purifier des CTCs dans l'introduction.

Les marqueurs choisis sont les protéines EpCAM et TROP 2. Ces protéines sont considérées comme des marqueurs de cellules épithéliales, et ne devraient donc pas être présents sur les cellules de la lignée hématopoïétique qui se trouvent dans le sang, ce qui permet de discriminer des CTCs d'origine épithéliales des cellules saines. Ces deux marqueurs sont donc régulièrement utilisés comme marqueurs des CTCs (Goldenberg, D. M. *et al.*, 2018; Schnell, U. *et al.*, 2013). EpCAM (ou TROP 1) et TROP 2 sont deux protéines homologues (49% d'identité de séquence et 67% de similarité de séquence) qui font partie de la famille des CAMs (*cell adhesion molecules*). Elles comportent toutes deux un domaine transmembranaire. Leur domaine extracellulaire est riche en cystéines et composé de trois domaines : un domaine thyroglobuline-*like* entouré d'un domaine EGF-*like* (*epidermal growth factor-like*), et un domaine non apparenté à une famille connue de domaines. Cependant, pour EpCAM, il a été montré que la façon dont sont connectées les cystéines dans le domaine EGF-*like* n'est pas retrouvé dans ce type de domaine et on peut questionner la classification de ce domaine comme un domaine EGF-*like*. EpCAM possède 3 sites de N-glycosylation pour une masse moléculaire totale de 38kDa tandis que TROP 2 possède 4 sites de N-glycosylation (dont deux partagés avec EpCAM) pour une masse moléculaire totale de 50 kDa.

A. Sélection par phage display d'interacteurs bi-P5H contre EpCAM et TROP 2

Ces deux protéines sont ancrées à la membrane par une seule hélice transmembranaire. Les sélections ont donc été réalisées sur leur domaine extracellulaire, qui peuvent être produits de façon soluble indépendamment de la partie transmembranaire. Ces domaines extracellulaires ont été commandés chez ACROBiosystems, qui les a produits glycosylés dans des cellules HEK293, avec un His-tag en C-ter de la protéine (du côté où se trouve l'hélice transmembranaire dans la protéine entière).

Les sélections ont été réalisées selon le même protocole que pour la sélection contre l'eGFP.

Pour la sélection de bi-P5H contre EpCAM, nous n'avons pas observé une amplification du nombre de clones sur boîte à l'issue du premier tour de sélection. Cela peut signifier que la sélection d'interacteurs bi-P5H contre EpCAM par phage *display* n'a pas fonctionné. Un phage ELISA en pool a été réalisé pour vérifier que les quelques clones présents sur les boîtes ne correspondaient pas à des interacteurs d'EpCAM, mais ces tests se sont révélés être négatifs.

Plusieurs hypothèses permettraient d'expliquer que le phage *display* n'a pas permis d'obtenir d'interacteur d'EpCAM. Tout d'abord, il est possible que la diversité de la bibliothèque ne soit pas suffisante pour donner des interacteurs d'EpCAM. Une des solutions possibles pour y remédier serait de réaliser le *shuffling* évoqué plus tôt dans le manuscrit (partie 1.IV), et qui consiste à mélanger entre eux les domaines T et O de la bibliothèque pour générer de nouvelles combinaisons et recréer de la diversité. Enfin, il est également possible que le problème vienne des glycosylations (3 sites de N-glycosylation sur EpCAM). On peut imaginer que les glycosylations viennent masquer des « épitopes » qui autrement, seraient reconnus par les bi-P5H, réduisant ainsi la surface d'interaction potentielle et rendant la sélection d'interacteurs plus compliquée. Enfin, il est possible que la protéine commandée ne soit pas correctement repliée. ACROBiosystems fournit en effet les protéines sous forme lyophilisée, et ce processus est peut-être déstabilisateur pour la structure d'EpCAM.

Nous nous sommes concentrés sur l'autre cible d'intérêt : la protéine TROP 2. La sélection de bi-P5H contre TROP 2 a été effectuée toujours selon le même protocole que pour l'eGFP, et a donné des résultats positifs. Les résultats du phage ELISA clonal réalisé à l'issue des 3 tours de sélection sont représentés [Figure 63](#).

A												
	1	2	3	4	5	6	7	8	9	10	11	12
A	0,17	0,089	0,074	0,158	0,307	0,08	0,376	0,322	0,317	0,306	0,242	0,415
B	0,08	0,095	0,105	0,172	0,092	0,083	0,099	0,081	0,073	0,07	0,085	0,076
C	0,18	0,096	0,128	0,247	0,218	0,244	0,203	0,119	0,175	0,121	0,096	0,123
D	0,09	0,112	0,083	0,097	0,086	0,079	0,071	0,083	0,071	0,074	0,074	0,072
E	0,16	0,206	0,075	0,156	0,143	0,229	0,272	0,344	0,075	0,274	0,166	0,251
F	0,09	0,096	0,085	0,084	0,071	0,084	0,093	0,101	0,069	0,081	0,073	0,204
G	0,16	0,233	0,128	0,242	0,14	0,167	0,255	0,128	0,249	0,274	0,159	0,128
H	0,08	0,091	0,082	0,077	0,092	0,078	0,072	0,075	0,083	0,068	0,08	0,076

B												
	1	2	3	4	5	6	7	8	9	10	11	12
A	0,101	0,101	0,136	0,15	0,119	0,184	0,137	0,113	0,064	0,27	0,268	0,143
B	0,07	0,081	0,073	0,084	0,073	0,068	0,07	0,07	0,066	0,066	0,062	0,172
C	0,159	0,205	0,206	0,204	0,175	0,136	0,194	0,175	0,121	0,082	0,158	0,495
D	0,082	0,081	0,094	0,073	0,073	0,071	0,069	0,074	0,07	0,073	0,072	0,142
E	0,065	0,149	0,201	0,329	0,352	0,134	0,114	0,373	0,096	0,098	0,157	0,296
F	0,072	0,078	0,078	0,058	0,084	0,077	0,073	0,081	0,072	0,069	0,077	0,076
G	0,19	0,163	0,581	0,27	0,264	0,101	0,245	0,364	0,34	0,194	0,145	0,235
H	0,07	0,07	0,094	0,079	0,081	0,075	0,076	0,081	0,074	0,081	0,079	0,061

Figure 63 : Crible des sélections bi-P5H contre TROP 2 par phage ELISA clonal

Dans cette expérience le domaine extracellulaire de TROP 2 est immobilisé sur deux plaques ELISA 96 puits dans les puits A, C, E, G. Les puits B, D, F, G ne présentent donc pas la cible et servent à déceler les clones qui interagissent de façon non spécifique. 96 clones de bi-P5H obtenus à l'issue de la sélection sont exposés sur phages et incubés dans chaque puit (1 clone/puit). La présence des phages dans les puits est révélée à l'aide de l'anticorps anti-phage M13 couplé à l'HRP. Le produit coloré formé dans la réaction catalysée par la peroxydase est visible et quantifiable à 410 nm. On observe la reconnaissance spécifique d'un grand nombre de clones bi-P5H pour TROP 2.

Pour beaucoup de clones, le signal spécifique (lignes A, C, E, G) est plus important que le signal aspécifique (lignes B, D, E, H). On peut voir que beaucoup de clones donnent un ratio signal/bruit supérieur à 2, ce qui est le seuil que nous avons fixé pour identifier les interacteurs potentiels. Tous les clones avec un ratio signal/bruit > 2 ont été récoltés et séquencés.

Ces séquençages ont montré que tous les clones séquencés ont la même séquence codant pour une mono-P5H. Contrairement à la sélection de bi-P5H contre l'eGFP, aucune bi-P5H n'a été identifiée parmi les interacteurs. Dans un tel cas de figure, des stratégies alternatives sont proposées pour générer des interacteurs bi-domaines. En effet, la libération de la cible implique d'avoir un interacteur bi-domaine dont l'énergie de liaison soit partagée entre ses deux domaines. Il est cependant possible, à partir de deux mono-P5H sélectionnées par phage *display*, de muter la face bi-hélicoïdale (constante) de l'une avec les résidus présents sur la face bi-hélicoïdale du domaine P5H T (en jaune Figure 64). La face bi-hélicoïdale de l'autre mono-P5H peut alors être mutée avec les résidus présents sur la face bi-

hélicoïdale du domaine P5H O (en bleu Figure 64). Les deux mono-P5H ainsi mutés peuvent être fusionnés, ce qui permet de générer une bi-P5H reconnue par le modulateur allostérique, tout en gardant les faces tri-hélicoïdales des mono-P5H qui la compose et qui constituent la surface d'interaction avec la cible.

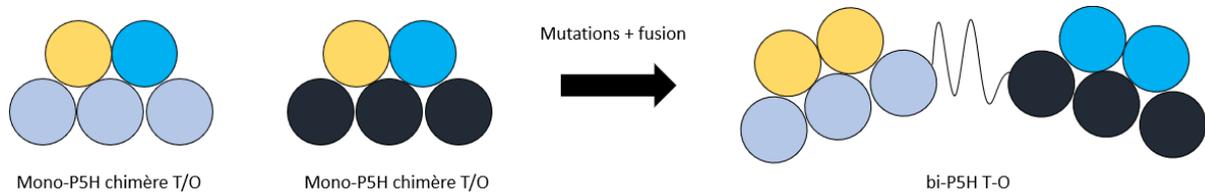


Figure 64 : Ingénierie sur deux mono-P5H pour générer une bi-P5H

Si une seule mono-P5H est sélectionnée par phage *display*, le même principe pourra être mis en place. La bi-P5H qui résultera de cette ingénierie pourra alors se fixer de façon bivalente par le même épitope sur deux molécules cible distinctes, comme un anticorps.

B. Sélection par phage display d'interacteurs α Rep contre TROP 2

Parallèlement aux sélections de bi-P5H contre EpCAM et TROP 2, une sélection contre TROP 2 avec la bibliothèque d' α Reps développée par le laboratoire a été réalisée. La bibliothèque d' α Reps a donné des interacteurs pour plus de 40 cibles, mais la fixation sur cellules d' α Reps sélectionnées contre une protéine de surface cellulaire purifiée n'avait jamais été testée.

La sélection a été réalisée dans les mêmes conditions que les sélections réalisées avec la bibliothèque de bi-P5H, et les résultats du phage ELISA clonal effectué à l'issue des trois tours de sélection sont présentés [Figure 65](#).

	1	2	3	4	5	6	7	8	9	10	11	12
A	17,119	10,885	1,448	2,105	1,712	17,850	14,698	1,865	13,164	1,593	21,742	20,439
B	1,510	1,388	1,286	6,716	1,553	1,261	1,511	1,865	4,405	16,525	1,322	20,906
C	1,269	9,500	9,403	12,554	14,797	8,865	10,148	13,212	11,327	6,771	11,037	1,115
D	5,667	1,610	9,032	9,950	4,400	8,604	6,480	14,628	8,900	14,340	1,417	1,148
E	4,695	0,864	7,950	5,276	10,907	5,845	3,082	7,467	6,222	6,630	9,629	16,852
F	2,544	2,017	1,373	12,030	1,061	1,519	9,880	1,154	12,250	15,961	9,281	18,017
G	1,214	8,655	1,096	10,015	1,113	14,019	16,554	7,245	14,135	14,364	8,140	10,063
H	7,556	7,583	8,118	8,091	8,292	10,633	11,204	11,192	12,192	13,115	15,211	14,000

Figure 65 : Crible des sélections α Reps contre TROP 2 par phage ELISA clonal

Ce tableau représente le ratio du signal ELISA spécifique sur le signal ELISA aspécifique pour les 96 clones α Reps criblés à l'issue de la sélection. La présence des phages dans les puits a été révélée à l'aide de l'anticorps anti-phage M13 couplé à l'HRP. Le produit coloré formé dans la réaction catalysée par la peroxydase est visible et quantifiable à 410 nm. Chaque case correspond au ratio signal spécifique / signal aspécifique d'un clone.

Le domaine extracellulaire de TROP 2 a été immobilisé sur une première plaque, tandis qu'une autre plaque a servi à mesurer l'interaction non spécifique de chaque clone testé. Le ratio signal spécifique / signal aspécifique pour chaque clone a été représenté [Figure 65](#). On peut voir que beaucoup de clones donnent un ratio signal/bruit très élevé, ce qui rend compte d'une interaction spécifique des clones testés. Ceux qui possédaient un ratio signal/bruit > 8 ont été envoyés au séquençage.

Comme pour la sélection des bi-P5H contre TROP 2, les résultats du séquençage ont montré que tous les clones séquencés codaient pour la même α Rep. Ce résultat est inhabituel pour une sélection faite avec la bibliothèque d' α Rep. En effet, lors des sélections, plusieurs α Rep sont généralement sélectionnés contre les cibles (de 5 à 15 selon les sélections). Ce résultat suggère que la protéine TROP 2 est une cible particulièrement difficile pour la sélection d'interacteurs, que ce soit des α Reps ou des bi-PH5. Notre hypothèse pour expliquer cette difficulté de sélection repose sur la nature de la cible.

En effet, TROP 2 est glycosylée contrairement aux cibles testées jusqu'ici avec la bibliothèque d' α Rep. Ces glycosylations pourraient rendre la surface de TROP 2 moins favorable à la reconnaissance par les variants de la bibliothèque, en masquant des sites reconnus par les α Reps ou les bi-P5H.

Il est également possible que ce soit la méthode de criblage des interacteurs qui ne permette pas d'identifier toutes les bi-P5H ou toutes les α Reps dirigées contre les cibles. Lorsqu'on réalise le phage ELISA clonal, on crible un nombre restreint de clones (96 clones) par rapport au nombre de clones totaux sélectionnés par phage *display*. Il est possible que dans notre cas, étant donné la nature difficile de la cible étudiée, ce nombre ne soit pas suffisant pour identifier l'ensemble des interacteurs sélectionnés.

Une des solutions envisagées pour caractériser les profils de sélection serait de réaliser du séquençage nouvelle génération (NGS : *New Generation Sequencing*) sur la population de clones issus du tour 3 de la sélection par phage *display* pour avoir accès à un répertoire exhaustif des séquences codant pour des variants retenus sur la cible. Cette méthode permettrait d'avoir accès à l'ensemble des séquences codant pour des variants sélectionnés par phage *display*. Il serait ainsi possible de comparer ces séquences entre elles, et d'identifier des regroupements d'interacteurs qui varient très peu les uns des autres, et qu'on pourrait regrouper en « familles » d'interacteurs. Cette méthode pourrait renseigner sur les positions les plus importantes pour l'interaction, et les positions mutables pour améliorer l'interaction.

Pour résumer, les sélections contre la protéine TROP 2 ont donné deux interacteurs : une mono-P5H appelée B1, et une α Rep appelée A1. Nous avons par la suite cherché à savoir si ces interacteurs sont capables de reconnaître leur cible à la surface des cellules.

C. Test de l'interaction sur cellules eucaryotes

Nous nous sommes intéressés à montrer que les interacteurs sélectionnés contre TROP 2 étaient capables de reconnaître la protéine native à la surface de cellules eucaryotes. En effet, les sélections par phage *display* sont réalisées sur des cibles protéiques purifiées qui ne sont pas présentées dans leur contexte cellulaire. Les interacteurs issus de telles sélections ne sont donc pas forcément aussi affins pour la protéine telle que présentée à la surface des cellules.

Mise en place d'un modèle cellulaire exprimant TROP2-GFP

La première étape a donc consisté à créer un modèle d'étude présentant la protéine TROP 2 à la surface des cellules eucaryotes. Comme TROP 2 est glycosylée, notre choix s'est porté sur un modèle d'étude eucaryote, les cellules HEK293. Les cellules HEK (*Human Embryonic Kidney*) sont des cellules embryonnaires humaines immortalisées, et sont un bon modèle d'étude notamment par la facilité avec laquelle elles peuvent être transfectées (Pulix, M. *et al.*, 2021). Cela permet de facilement exprimer des protéines recombinantes, dans notre cas la protéine TROP 2. De plus, les cellules HEK293 non transfectées n'expriment pas TROP 2, ce qui permet de s'affranchir d'une interaction basale. Nous avons choisi pour notre système d'étude d'exprimer une protéine issue de la fusion entre TROP 2 et la GFP. La fusion avec la GFP apporte en effet plusieurs avantages. Cela permet par une simple analyse visuelle des plaques de culture au microscope de contrôler si la transfection a fonctionné. De plus, cela permettrait de visualiser une interaction entre TROP 2 et un interacteur couplé à un autre fluorophore que la GFP, par co-localisation des signaux lumineux. Un plasmide codant pour la protéine TROP 2 dont le domaine cytosolique a été fusionné à la GFP a donc été exprimé à partir d'un gène synthétique.

Des cellules HEK293 ont alors été transfectées selon un protocole décrit en détail dans le matériel et méthode. Les cellules HEK293 ont été cultivées jusqu'à une confluence d'environ 80%, puis mises en contact du plasmide codant pour TROP 2-GFP et de lipofectamine. La lipofectamine est un composé de nature lipidique qui va entourer les molécules d'acides nucléiques (ici notre plasmide d'intérêt) de façon à les présenter dans un liposome chargé positivement. Cela facilite la fusion avec la membrane plasmique des cellules, chargée négativement, et permet aux molécules d'acides nucléiques, également chargées négativement, de passer la membrane plasmique malgré la répulsion électrostatique. Le matériel génétique inséré dans les cellules peut alors être intégré au génome, et on parlera alors de transfectant stable. Si le plasmide n'est pas intégré au génome des cellules, il est alors progressivement dégradé et les cellules transformées, qui peuvent exprimer une résistance à l'antibiotique utilisé pour le criblage pendant quelques jours, sont appelées transfectants transitoires puisqu'elles n'expriment que transitoirement le gène d'intérêt. Dans notre cas, nous souhaitons obtenir des transfectants stables, pour avoir des lignées cellulaires sur lesquelles nous pouvons

travailler sans recommencer la transfection. Les transfectants stables ont été sélectionnés par propagation sur un milieu contenant de la généticine, un antibiotique qui permet d'éliminer les cellules n'ayant pas intégré le plasmide d'intérêt à leur génome.

Les cellules HEK293 ainsi transfectées ont été observées au microscope pour déterminer si la transfection avait été efficace (Figure 66).

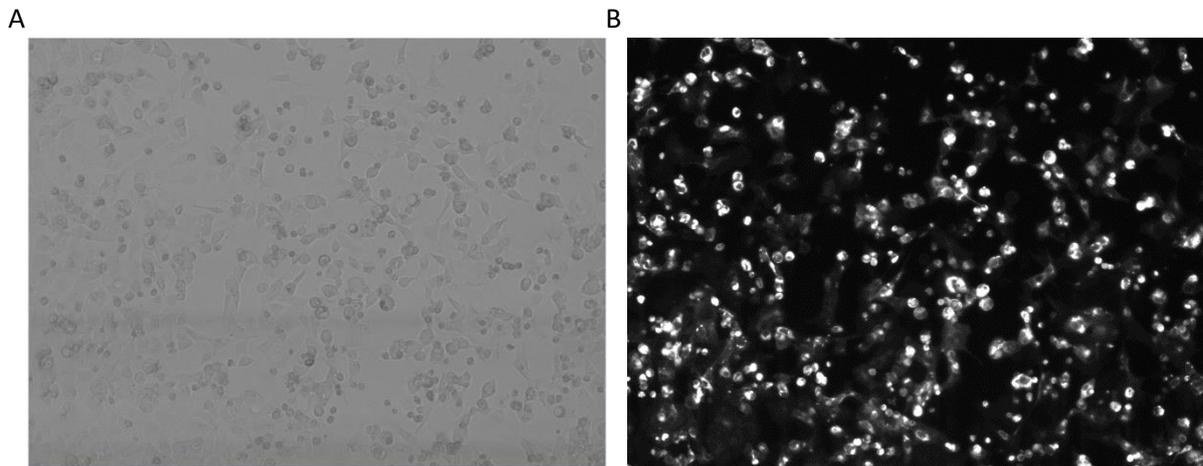


Figure 66 : Culture des HEK293 transfectées par un plasmide codant pour TROP 2-GFP vue au microscope

Les cultures de cellules HEK293 transfectées par le plasmide codant pour la fusion TROP 2-GFP ont été observées au microscope. Un cliché a été pris en transmission de lumière blanche (A) ou en présence d'un filtre qui ne laisse passer que les longueurs d'ondes correspondant à l'émission de la GFP (B).

On peut voir Figure 66 que la culture présente un très grand nombre de cellules fluorescentes, ce qui traduit une transfection très efficace. L'étape suivante a alors consisté à cultiver les cellules à un état polyclonal, c'est-à-dire avec quelques clones (2 à 5) par puit. Ces lignées polyclonales ont été observées au microscope pour déterminer lesquelles seraient les plus adaptées aux expériences de FACS. Cette évaluation s'est faite selon deux critères : l'intensité de fluorescence d'une part, et le rapport cellules fluorescente / cellules non fluorescentes d'autre part. Les meilleures lignées ont alors été sélectionnées pour réaliser les expériences de FACS.

Nous avons donc réussi à obtenir des lignées stables qui surexpriment la protéine fusion TROP 2-GFP, et nous pouvons utiliser ce système pour tester l'interaction des deux protéines sélectionnées contre TROP 2, la mono-P5H B1 et l' α Rep A1.

Études des interactions bi-P5H / Trop2 par FACS

Des expériences de FACS ont alors été réalisées avec la lignée polyclonale qui donnait le signal de fluorescence le plus fort, en mettant les cellules exprimant TROP 2-GFP au contact de la mono-P5H (B1) et de l' α Rep (A1) issues des sélections contre TROP 2. Les interacteurs A1 et B1 ont été produits biotinylés et peuvent être ainsi associés à de la streptavidine couplée à un fluorophore : l'AlexaFluor680 (AF680) qui émet à 680 nm.

Dans un premier temps, nous avons étudié l'interaction des cellules HEK293 surexprimant TROP 2-GFP avec l' α Rep A1, afin de mettre au point le protocole pour les expériences de cytométrie en flux qui sont nouvelles pour le laboratoire. L'interaction des cellules transfectées avec l'interacteur B1 sera présentée ultérieurement.

Des cellules HEK293 surexprimant TROP 2-GFP ont été cultivées, puis incubées avec différentes concentrations de l' α Rep A1. Après une étape de lavage, la présence de A1 à la surface des cellules a été révélée par ajout de streptavidine-AF680 dans le tube contenant les cellules. Les échantillons ont ensuite été analysés par cytométrie en flux, et les résultats de cette expérience sont synthétisés [Figure 67](#).

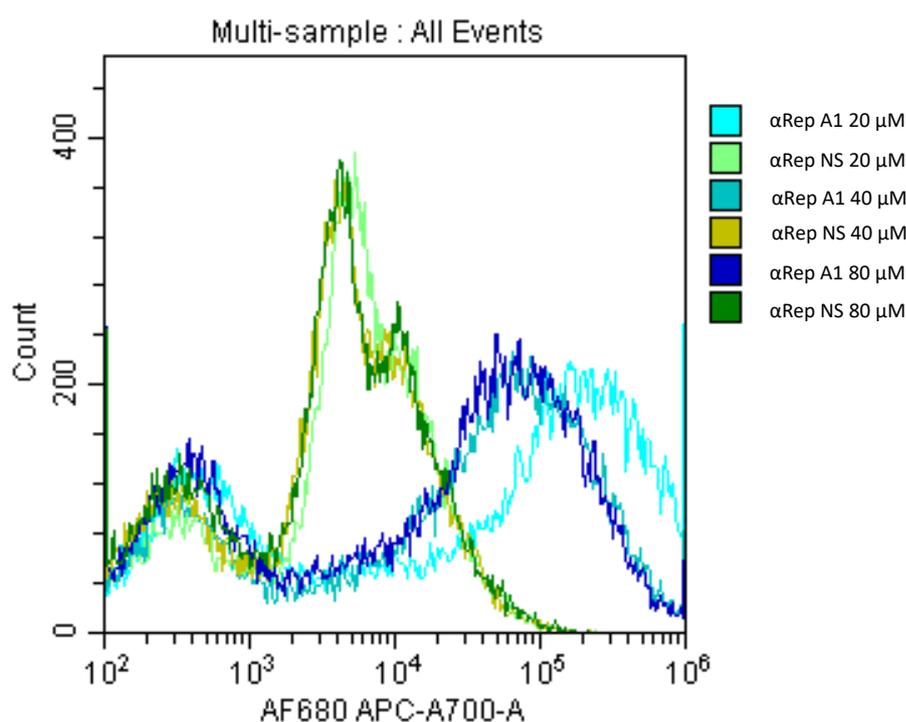


Figure 67 : Étude de l'interaction entre le mono-domaine A1 et TROP 2 par FACS.

Les graphes présentés dans cette figure représentent l'intensité de la fluorescence à 680 nm en abscisse (AF680 APC-A700-A) et le nombre de cellules comptées en ordonnée (*count*). Ces graphiques rendent compte de la présence de la streptavidine-AF680, et donc par extrapolation, du ligand biotinylée à la surface des cellules HEK293 qui surexpriment TROP 2-GFP.

Le monomère de A1 à 20 μM donne la meilleure intensité de fluorescence (signal décalé vers la droite), tandis que les concentrations plus élevées (40 et 80 μM) donnent un signal légèrement plus faible. Dans l'ensemble, les trois concentrations testées pour le monomère de A1 donnent un signal presque 20 à 100 fois supérieur que l' $\alpha\text{Rep NS}$ aux mêmes concentrations. Un signal de fluorescence plus fort rend compte d'un plus grand nombre de molécules de streptavidine à la surface des cellules, et donc de ligand biotinylé. Cette première expérience semble donc montrer une interaction spécifique de l' $\alpha\text{Rep A1}$ avec la protéine TROP 2 présente à la surface des cellules. Nous avons alors cherché à savoir s'il était possible d'augmenter la sensibilité de la détection par un effet d'avidité.

La streptavidine possédant 4 sites de fixation pour la biotine, nous avons incubé la streptavidine-AF680 avec 4 équivalents molaire d' $\alpha\text{Rep A1}$ avec pour objectif de créer des multimères fluorescents d'interacteurs. La formation de ces multimères fluorescents d'interacteurs devrait augmenter l'avidité de ces derniers pour la cible, et ainsi minimiser le décrochage des interacteurs lors des étapes de lavage. Les résultats obtenus avec les multimères sont présentés Figure 68.

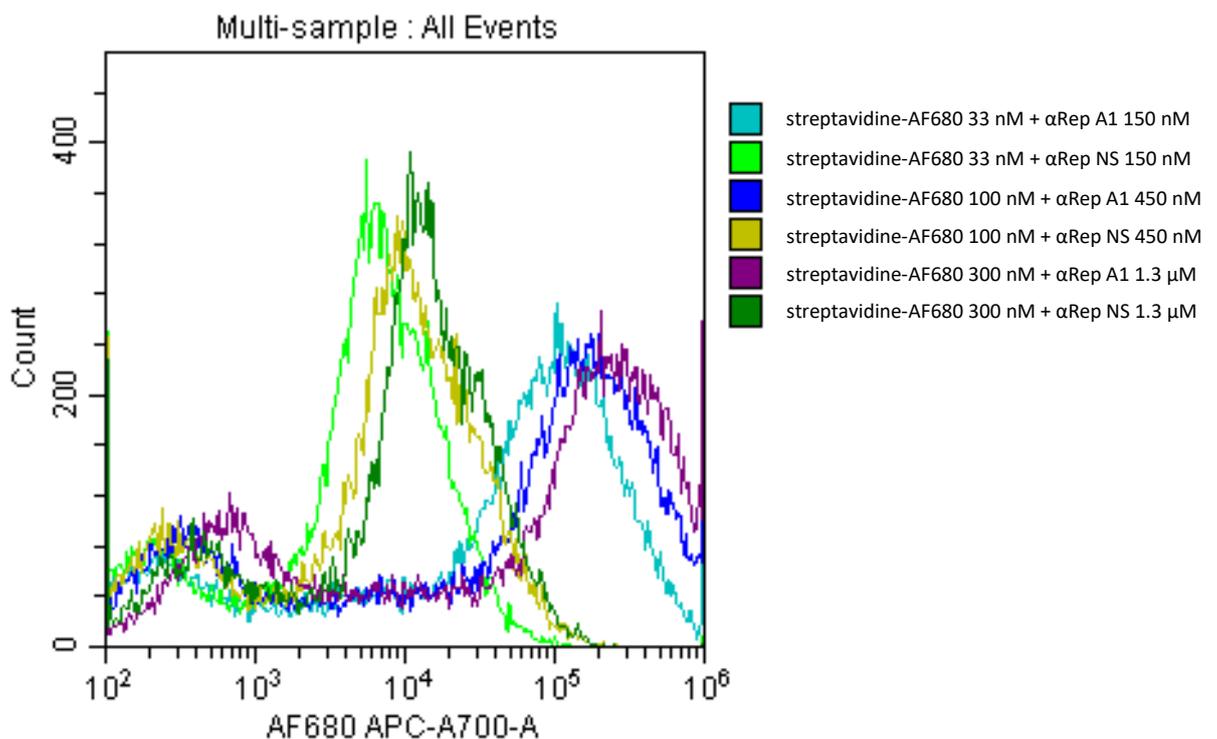


Figure 68 : Étude de la spécificité de l'interaction entre les multimères de A1 et TROP 2 par FACS.

On observe une fixation des multimères fluorescents de A1 même à la concentration la plus faible (33 nM). L'interaction semble à nouveau spécifique, puisque le multimère d' $\alpha\text{Rep non spécifique}$ donne un signal au moins 10 fois plus faible qu'avec le multimère de A1. L'intensité du signal de fluorescence semble corrélée à la concentration du multimère fluorescent, même pour les $\alpha\text{Reps non spécifiques}$.

Nous avons montré que les multimères d'interacteurs étaient capables d'interagir avec les cellules HEK293 surexprimant TROP 2-GFP, même à de faibles concentrations. Pour s'assurer de la spécificité de l'interaction, nous avons réalisé une compétition avec un complexe non fluorescent entre de la streptavidine non marquée et l'interacteur A1. Les résultats de cette compétition sont présentés Figure 69.

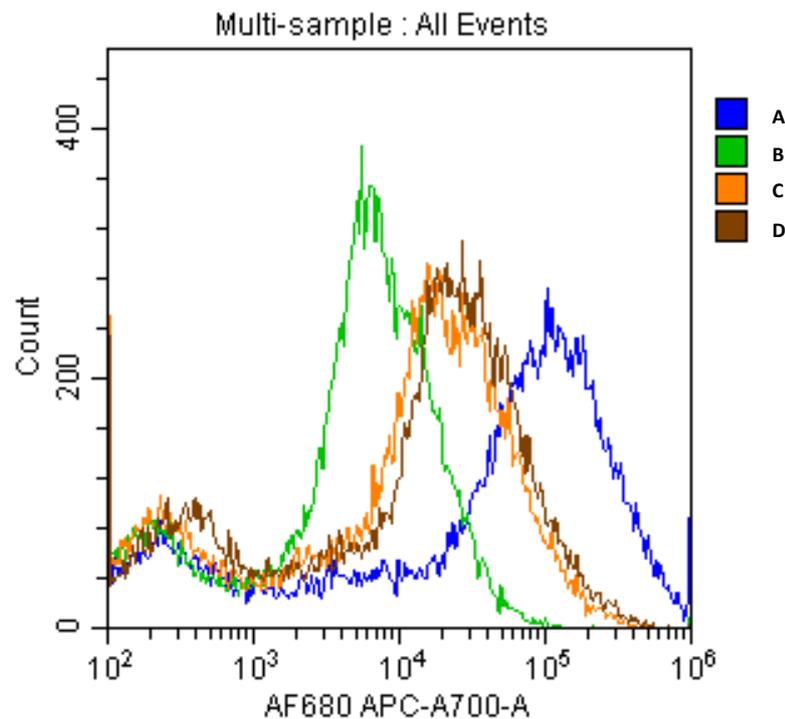


Figure 69 : Compétition entre un multimère fluorescent et un multimère non fluorescent de A1

Des cellules HEK293 surexprimant TROP 2-GFP ont été incubées avec le multimère fluorescent de A1 à 33 nM (courbe A) ou avec un multimère fluorescent d'une α Rep non spécifique à 33 nM (courbe B). Une pré-incubation des cellules HEK293 transfectées avec un multimère non fluorescent de A1 avant de les incuber avec le multimère fluorescent de A1 a été effectuée pour obtenir les courbes C et D. Le multimère non fluorescent est 15 fois plus concentré que le multimère fluorescent, soit 495 μ M contre 33 nM (courbe C) et 1500 nM contre 100 nM (courbe D).

Dans cette expérience, les cellules HEK293 surexprimant TROP 2-GFP ont été mises en contact avec le multimère fluorescent de A1 (A), ou pré-incubées avec un multimère non fluorescent de A1 (C et D) avant la mise en contact avec le multimère fluorescent de A1. On peut voir que la pré-incubation des cellules avec un multimère non fluorescent de A1 avant la mise en contact avec le multimère fluorescent de A1 (C et D) provoque un décalage de l'intensité de fluorescence vers la gauche par rapport à la même condition sans la pré-incubation (A), synonyme d'une fluorescence moins intense. Cela témoigne donc de la capacité du multimère non fluorescent de A1 à inhiber de façon compétitive la fixation du multimère fluorescent de A1 à la surface des cellules HEK293 qui surexpriment TROP 2-GFP, et montre la spécificité de l'interaction entre A1 et la protéine TROP 2 à la surface des cellules.

Globalement, ces expériences montrent la spécificité de l'interacteur A1 pour la protéine TROP 2 dans son état natif, à la surface des cellules. Ces résultats sont encourageants, et nous cherchons maintenant à développer une interaction modulable à partir de l' α Rep A1. Une des pistes envisagées consiste à placer un site de coupure spécifique de la protéase 3C dans l'une des boucles situées entre deux motifs interne de l' α Rep A1. L'interaction pourrait alors être modulée par ajout dans le milieu de la protéase 3C, puisque celle-ci séparerait la surface d'interaction de l' α Rep en deux moitiés, ce qui favoriserait sa dissociation de la cible. Une autre possibilité consisterait à utiliser un interacteur de la face convexe des α Reps (non diversifiée) pour constituer un modulateur de la même façon que le modulateur de la bibliothèque de bi-P5H. Cela supposerait toutefois de disposer d' α Rep dont la surface non fixatrice ait été modifiée et qu'un interacteur spécifique de cette surface modifiée soit généré. Par la suite, nous avons étudié l'interaction des cellules HEK293 surexprimant TROP 2-GFP avec la mono-P5H B1. Les résultats préliminaires réalisés à 4°C n'ont pas donné de résultats. Une expérience à température ambiante a alors été réalisée en présence d'azide de sodium pour inhiber les mouvements membranaires et éviter l'internalisation de l'interacteur. Les résultats de cette expérience sont présentés [Figure 70](#).

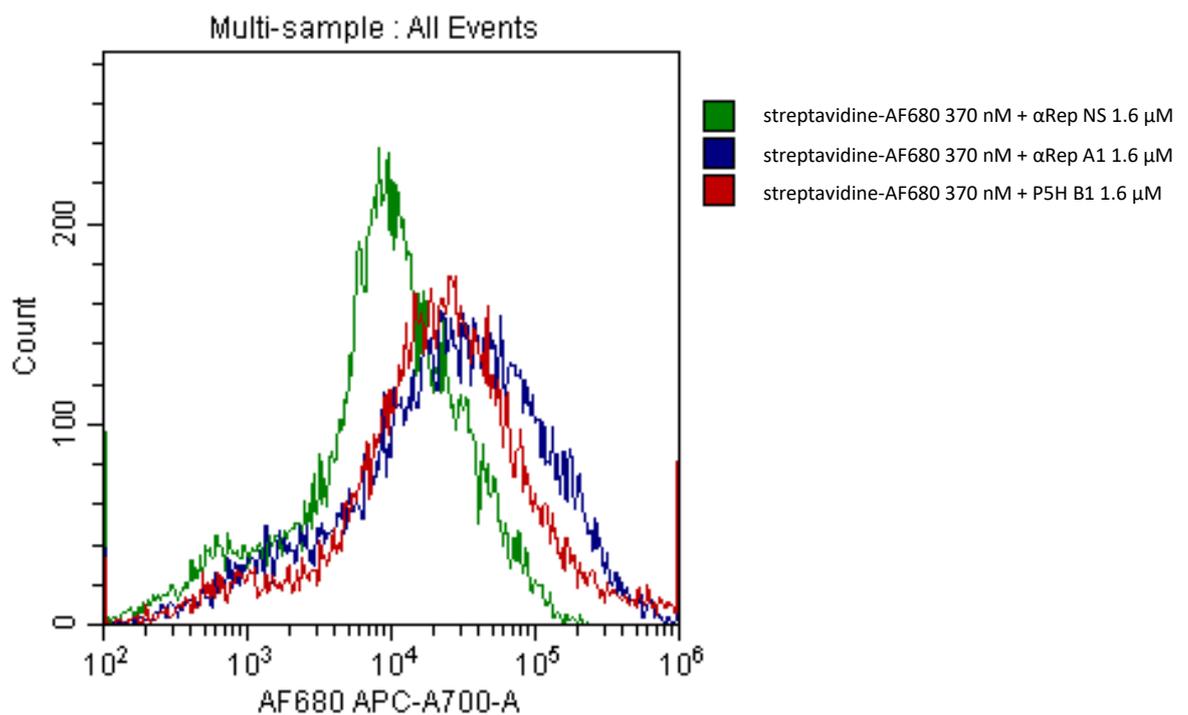


Figure 70 : Étude de l'interaction entre les cellules HEK293 surexprimant TROP 2-GFP et les interacteurs A1 et B1 à 20°C

Les signaux obtenus pour les interacteurs A1 et B1 sont comparables, et décalés par rapport au signal obtenu avec l' α Rep NS. Des expériences supplémentaires sont en cours pour l'interacteur B1 qui n'a pas été aussi bien caractérisé que A1. En effet, l'interaction observée à température ambiante dans

l'azide de sodium n'a pas été reproduite à 4°C, et une mise au point du protocole est donc nécessaire pour étudier cet interacteur. La différence observée en fonction de la température pourrait s'expliquer par la tendance des P5H à précipiter à froid. En effet, les P5H naturelles viennent d'un organisme thermophile, et cette ossature a donc été optimisée pour des températures élevées, ce qui pourrait expliquer son instabilité dans un milieu froid.

Les résultats obtenus pour B1 sont encourageants puisque dans les mêmes conditions l' α Rep A1, qui par ailleurs interagit de manière spécifique avec TROP 2 même à faible concentration, donne le même signal. La suite du projet consiste donc à optimiser le protocole de cytométrie en flux pour étudier l'interaction de la protéine B1 avec la protéine TROP 2 ancrée à la membrane plasmique, notamment en réalisant l'expérience à 37°C.

V. Bilan et perspectives

A. Conception et construction de la bibliothèque de bi-P5H

Les travaux réalisés durant ce projet de thèse ont permis de valider le protocole mis au point pour construire une bibliothèque de protéines bi-domaine, nouvelle au laboratoire. La taille obtenue pour la bibliothèque ($1.4 \cdot 10^9$ variants) est satisfaisante et en accord avec la taille des bibliothèques compatibles avec le phage *display* (Carmen, S., 2002). Peu d'interacteurs ont été obtenus pour les sélections réalisées jusqu'ici, ce qui pourrait être attribué à une diversité trop faible au sein de la bibliothèque. Si ce problème venait à se confirmer dans les sélections à venir, nous avons une solution qui permet de générer de la diversité à partir de la bibliothèque actuelle. En effet nous avons placé des sites BbsI de part et d'autre du domaine P5H T dans la bibliothèque. Ceci qui offre la possibilité de réaliser un *shuffling* qui permettra de générer de la diversité en associant de nouvelles paires de domaines P5H T et P5H O entre eux. Ce processus peut permettre d'augmenter le nombre de cibles reconnues par les variants issus de la bibliothèque si la diversité de la bibliothèque initiale venait à être limitante pour l'obtention d'interacteurs.

Dans la bibliothèque, 56.4% de clones codants pour des pentahélines sont correctement insérées dans le phagemide, dont plus de deux-tiers de bi-domaines. Ce chiffre est comparable à la première bibliothèque d' α Reps développée par le laboratoire, qui avait donné 63% de clones correctement insérés pour une bibliothèque de $3 \cdot 10^8$ clones (Urvoas, A. et al., 2010). Par la suite, cette bibliothèque a été optimisée par une filtration des phages exposant une protéine entière *via* leur FLAG-tag, puis par une étape de *shuffling* pour recréer de la diversité. Cette librairie optimisée de $1.7 \cdot 10^9$ clones contient 93% de clones correctement insérés dans le phagemide (Guellouz, A. et al., 2013), et ce type d'optimisation pourrait tout à fait être réalisé avec la bibliothèque de bi-P5H. En effet, il est possible de filtrer les phages exposant une protéine entière grâce au HA-tag présent en N-ter des bi-P5H, et nous avons déjà discuté de la possibilité d'effectuer une étape de *shuffling* pour recréer de la diversité. Dans la bibliothèque de bi-P5H, un tiers des séquences codantes insérées dans le phagemide sont des mono-domaines. Cependant, il est possible de générer des bi-P5H à partir de deux mono-P5H reconnaissant des épitopes différents. Pour rappel, les mono-P5H de la bibliothèque sont des chimères issues de l'appariement aspécifique des oligonucléotides codants pour la partie C-ter du domaine P5H O sur les oligonucléotides codants pour la partie N-ter du domaine P5H T. Nous avons discuté de la façon dont il est possible de générer une bi-P5H à partir d'une ou deux mono-P5H, ce qui permet de valoriser les mono-P5H de la bibliothèque. Une autre application intéressante de la bibliothèque de bi-P5H serait de pouvoir facilement obtenir des interacteurs de deux épitopes non chevauchants sur une

cible. En effet lors d'une sélection par phage *display*, les deux domaines d'une bi-P5H ne peuvent pas interagir simultanément sur le même épitope, et nous espérons donc que la sélection favorisera l'obtention de bi-P5H qui interagissent par leurs deux domaines sur des épitopes non chevauchants. Enfin, l'aboutissement de cette construction vient confirmer l'expertise du laboratoire dans la conception de bibliothèques de protéines artificielles complexes, depuis le choix d'une ossature protéique robuste jusqu'à la stratégie d'assemblage des oligonucléotides dégénérés. La réussite de cet assemblage repose sur l'assemblage de 13 oligonucléotides sens, dont 6 sont dégénérés, et qui doivent s'hybrider dans l'ordre grâce à 13 oligonucléotides anti-sens pour générer une bi-P5H entière et fonctionnelle. Le pourcentage de bi-P5H dans la bibliothèque finale est donc la preuve que ce protocole pourtant compliqué est opérationnel.

B. Conception du modulateur allostérique

La conception du modulateur allostérique a constitué la partie la plus ambitieuse de mon projet de thèse d'un point de vue technique. Le concept qui a conduit à construire la bibliothèque de bi-P5H repose en effet sur la capacité des interacteurs de la bibliothèque à pouvoir libérer leur cible en interagissant avec un modulateur allostérique. Dans cette optique, nous avons décidé de concevoir une protéine capable de fixer simultanément les deux domaines des interacteurs de la bibliothèque. Cependant, pour que le modulateur soit générique, c'est-à-dire, que la même molécule puisse interagir avec tous les variants de la bibliothèque, il a fallu imaginer une stratégie pour que le modulateur reconnaisse spécifiquement la face bi-hélicoïdale, c'est-à-dire la face constante des P5H. De ce point de vue, la stratégie employée pour orienter la sélection vers des interacteurs de la face constante des P5H est un succès, puisque la structure obtenue pour le complexe A12-Tmut montre une interaction majoritairement au niveau de la face bi-hélicoïdale de la P5H, avec une implication de la tyrosine ajoutée pour jouer le rôle de *hot spot* d'interaction (Chen, J. *et al.*, 2013). Un jeu de données à 2.6 Å a également été collecté très récemment pour le complexe D3-Owt, mais la structure n'a pas encore été résolue à ce jour. De plus, des tests d'exposition sur phages ont été réalisés sur la bibliothèque finie, en comparant la capture des phages de la bibliothèque par l'anticorps anti-HA-tag ou par les deux modulateurs A12D3 et D3A12. Ces tests ont montré que les modulateurs capturaient 10^5 phages sur 10^9 présentés, soit un taux de capture de 1/10000. En comparaison, l'anticorps anti-HA-tag capturait environ 10^6 phages sur 10^9 présentés, soit un taux de capture de 1/1000. Cette expérience confirme la capacité des modulateurs à interagir avec des variants de la bibliothèque qui ne présentent pas la même face tri-hélicoïdale, ce qui implique que la surface d'interaction soit au niveau de la face bi-hélicoïdale. La différence de taux de capture observée entre les modulateurs et l'anticorps anti-HA-tag peut s'expliquer par le fait que l'anticorps reconnaît seulement le HA-tag, qui ne présente pas de structure secondaire. Une protéine dénaturée devrait donc être reconnue de la même façon qu'une protéine correctement repliée. À l'inverse, les modulateurs reconnaissent les protéines correctement repliées, mais pas les protéines dénaturées ou mal repliées. Une autre hypothèse pour expliquer cette différence de taux de capture pourrait être que la diversité au niveau de la face tri-hélicoïdale puisse induire chez certains variants de la bibliothèque un léger changement conformationnel à distance au niveau de la face bi-hélicoïdale, diminuant ainsi l'affinité du modulateur pour les P5H concernées. Ces résultats encourageants nous ont également permis de progresser vers la conception d'une fusion rigide entre deux α Reps, en utilisant comme élément de jonction une superposition d'hélice suivie d'une adaptation de la séquence. Il est difficile de concevoir une protéine-fusion dont le site de jonction adopte une structure définie, et lorsque la conception d'une telle protéine fusion est une

réussite, cela implique souvent un calcul intensif de minimisation de l'énergie de la protéine à l'aide (Wu, Y. *et al.*, 2017). Si le modèle que nous avons conçu ne se superpose pas parfaitement à la structure, l'objectif initial de créer une fusion rigide entre deux α Reps *via* une hélice α a été rempli. L'équipe MIP ne dispose pas d'une expertise en conception *ab initio* de structures de protéines. Cette conception *in silico* a donc été réalisée par assemblage modulaire à partir de données structurales sur les α Reps. Il est donc très encourageant de constater que le modèle que nous avons conçu rationnellement est proche de la structure cristallographique, et proche du modèle généré par AlphaFold. La modélisation réalisée avec AlphaFold nous a également donné une piste d'amélioration du *design*, et les séquences ainsi conçues permettront d'évaluer si la version 2 du modulateur A12D3 correspond à la structure conçue.

Concernant la libération de la cible des bi-P5H par le modulateur, les résultats préliminaires obtenus en BLI ont permis de montrer que le modulateur inhibait la fixation de la bi-P5H sur sa cible. Pour évaluer la capacité du modulateur à décrocher une cible d'intérêt en complexe avec son interacteur bi-P5H, nous avons programmé une expérience dont le principe est présenté Figure 71.

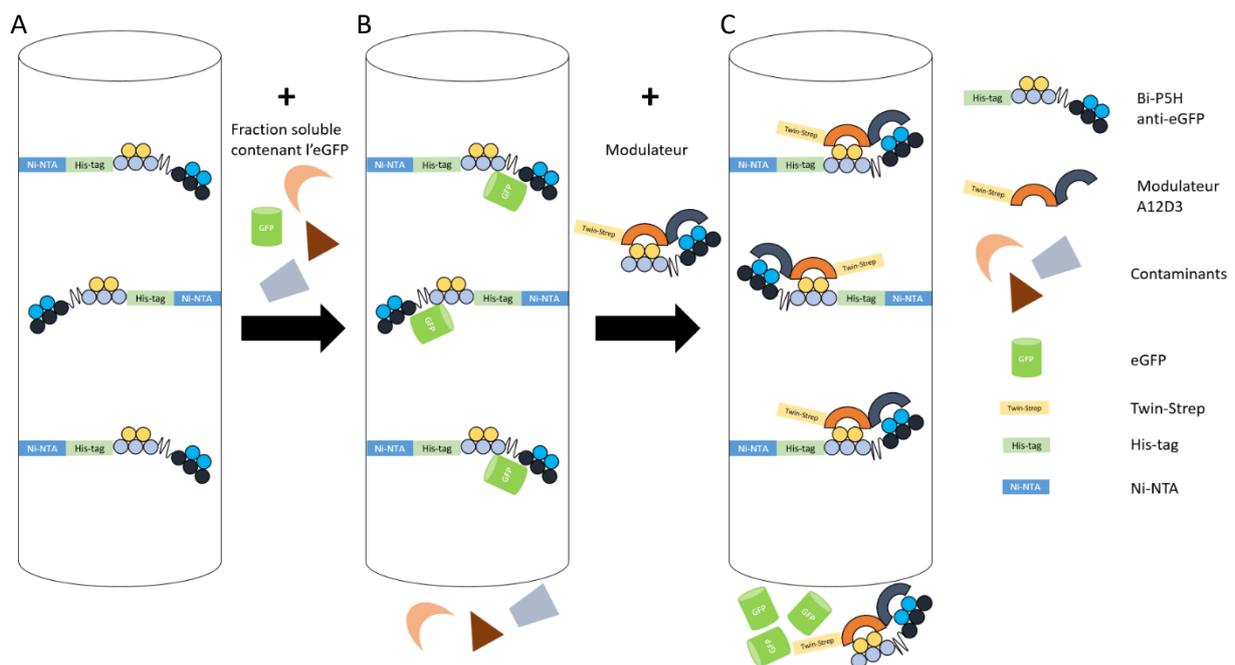


Figure 71 : Principe de la purification d'une protéine sans étiquette avec la bibliothèque de bi-P5H

Dans cette expérience, la bi-P5H D5 sélectionnée contre l'eGFP est immobilisée sur une résine Ni-NTA par son His-tag (Figure 71.A). La fraction soluble d'une culture bactérienne contenant de l'eGFP est alors déposée sur la colonne, et l'eGFP est capturée par la bi-P5H tandis que les contaminants sont lavés (Figure 71.B). L'ajout du modulateur sur la colonne doit alors permettre de libérer l'eGFP, qui est collectée avec un excès de modulateur (Figure 71.C). Pour obtenir la protéine pure, le modulateur qui possède une étiquette *Twin-Strep* peut être capturé par des billes magnétique recouvertes de

streptactine, ce qui permet de récupérer un surnageant contenant l'eGFP native et sans étiquette. Le principe de cette expérience peut être généralisé à toute protéine soluble pour laquelle il existerait un interacteur bi-P5H. La purification de protéines non étiquetées représente un enjeu important pour les structuralistes, ainsi que dans l'industrie pharmaceutique. En effet, la présence d'une étiquette sur les protéines est parfois suspectée de jouer un rôle dans l'interaction de certaines protéines, ou d'être source d'immunogénicité. Si une étiquette est utile à la purification des protéines, elle est souvent clivée à l'aide d'une protéase après l'étape de purification pour obtenir la protéine dans son état natif, et la bibliothèque de bi-P5H pourrait donc permettre de s'affranchir de cette étape.

C. Applications de la bibliothèque de bi-P5H

Enfin, les sélections réalisées avec la bibliothèque de bi-P5H ont permis d'obtenir un interacteur bi-domaine contre l'eGFP (D5) et un interacteur mono-domaine contre la protéine TROP 2 (B1). L'interaction de ces P5H avec leur cible a été vérifiée par BLI, tamis analytique, ou par FACS, ce qui a permis de valider la bibliothèque créée comme répertoire capable de fournir des interacteurs de cibles protéiques non apparentées. Ces résultats confirment la capacité du laboratoire à créer des bibliothèque complexes, puisque la bibliothèque créée est composée majoritairement de protéines bi-domaines qui sont reconnues par une même protéine, le modulateur.

Les expériences de FACS ont nécessité la mise en place d'un modèle cellulaire, les HEK293. Ces cellules ont été transfectées avec le plasmide codant pour la fusion entre la protéine TROP 2 et la GFP, et des transfectants stables ont pu être cultivés sous forme de lignée stable. La mise en place de ces lignées cellulaires constitue également un résultat important pour le laboratoire, puisque cela a permis de développer la compétence de l'équipe sur les cellules eucaryotes et que les lignées mises en place peuvent servir à de futures expériences pour les interacteurs de TROP 2.

Si l'interaction de D5 avec l'eGFP a été étudiée par l'intermédiaire de plusieurs techniques, l'interaction de B1 avec TROP 2 doit cependant encore être caractérisée. En effet, nos efforts se sont d'abord concentrés sur l' α Rep A1, également sélectionnée contre TROP 2, car c'est l'interacteur qui a donné les meilleurs résultats préliminaires, et que sa caractérisation nous permet de mettre au point le protocole d'étude d'interaction par FACS. Cette technique est nouvelle pour le laboratoire, et les résultats obtenus avec A1 nous permettent d'acquérir l'expertise nécessaire pour mener à bien l'étude de l'interaction entre B1 et TROP 2. Les études d'interaction se faisant généralement sur des protéines purifiées et non sur des protéines ancrées à la membrane cellulaire, les premiers résultats obtenus qui semblent montrer une interaction spécifique avec la protéine ancrée dans la membrane constituent une avancée importante.

Ces interacteurs pourraient alors servir à la purification de divers composants, allant des protéines aux cellules. Au niveau protéique, la bibliothèque de bi-P5H pourrait permettre la purification de protéines non étiquetées comme discuté précédemment. Au niveau cellulaire, il serait possible de réaliser la purification de cellules à haute valeur ajoutée comme les cellules tumorales circulantes, les lymphocytes T, ou les cellules souches. La purification de cellules entières pourrait être réalisée en immobilisant les bi-P5H soit sur des billes magnétiques, ce qui s'apparenterait au MACS (cf. Introduction), ou sur une surface de polysaccharide de type dextran pour mimer une purification d'affinité sur colonne.

Les travaux réalisés au cours de ce travail de thèse a permis d'explorer des conceptions innovantes en ingénierie des protéines, et l'équipe dans laquelle a été réalisé ce travail va maintenant s'attacher à explorer les applications qu'il est possible de réaliser avec ces outils nouveaux.

Bibliographie

- CARMEN, S., (2002), Concepts in antibody phage display, *Briefings in Functional Genomics and Proteomics*, vol. 1, n°2, p. 189-203.
- CHEN, J., SAWYER, N. & REGAN, L., (2013), Protein-protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area: Protein-protein Interactions, *Protein Science*, vol. 22, n°4, p. 510-515.
- COLLIS, A. V. J., BROUWER, A. P. & MARTIN, A. C. R., (2003), Analysis of the Antigen Combining Site: Correlations Between Length and Sequence Composition of the Hypervariable Loops and the Nature of the Antigen, *Journal of Molecular Biology*, vol. 325, n°2, p. 337-354.
- DE HAARD, H. J., VAN NEER, N., REURS, A., HUFTON, S. E., ROOVERS, R. C., HENDERIKX, P., DE BRÛINE, A. P., ARENDS, J.-W. & HOOGENBOOM, H. R., (1999), A Large Non-immunized Human Fab Fragment Phage Library That Permits Rapid Isolation and Kinetic Analysis of High Affinity Antibodies, *Journal of Biological Chemistry*, vol. 274, n°26, p. 18218-18230.
- ENGLER, C., GRUETZNER, R., KANDZIA, R. & MARILLONNET, S., (2009), Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes, *PLoS ONE*, vol. 4, n°5, p. e5553.
- GEORGE, R. A. & HERINGA, J., (2002), An analysis of protein domain linkers: their classification and role in protein folding, *Protein Engineering, Design and Selection*, vol. 15, n°11, p. 871-879.
- GIBSON, D. G., YOUNG, L., CHUANG, R.-Y., VENTER, J. C., HUTCHISON, C. A. & SMITH, H. O., (2009), Enzymatic assembly of DNA molecules up to several hundred kilobases, *Nature Methods*, vol. 6, n°5, p. 343-345.
- GOLDENBERG, D. M., STEIN, R. & SHARKEY, R. M., (2018), The emergence of trophoblast cell-surface antigen 2 (TROP-2) as a novel cancer target, *Oncotarget*, vol. 9, n°48, p. 28989-29006.
- GUELLOUZ, A., VALERIO-LEPINIEC, M., URVOAS, A., CHEVREL, A., GRAILLE, M., FOURATI-KAMMOUN, Z., DESMADRIL, M., VAN TILBEURGH, H. & MINARD, P., (2013), Selection of Specific Protein Binders for Pre-Defined Targets from an Optimized Library of Artificial Helicoidal Repeat Proteins (alphaRep), *PLoS ONE*, vol. 8, n°8, p. e71512.
- JUMPER, J., EVANS, R., PRITZEL, A., GREEN, T., FIGURNOV, M., RONNEBERGER, O., TUNYASUVUNAKOOL, K., BATES, R., ... HASSABIS, D., (2021), Highly accurate protein structure prediction with AlphaFold, *Nature*, vol. 596, n°7873, p. 583-589.
- KLEFFNER, R., FLATTEN, J., LEAVER-FAY, A., BAKER, D., SIEGEL, J. B., KHATIB, F. & COOPER, S., (2017), Foldit Standalone: a video game-derived protein structure manipulation interface using Rosetta, *Bioinformatics*, vol. 33, n°17, p. 2765-2767.
- KOEPNICK, B., FLATTEN, J., HUSAIN, T., FORD, A., SILVA, D.-A., BICK, M. J., BAUER, A., LIU, G., ... BAKER, D., (2019), De novo protein design by citizen scientists, *Nature*, vol. 570, n°7761, p. 390-394.

MINARD, P., (2019), L'évolution dirigée des protéines, *médecine/sciences*, vol. 35, n°2, p. 169-175.

PULIX, M., LUKASHCHUK, V., SMITH, D. C. & DICKSON, A. J., (2021), Molecular characterization of HEK293 cells as emerging versatile cell factories, *Current Opinion in Biotechnology*, vol. 71, p. 18-24.

RANAWAKAGE, D. C., TAKADA, T. & KAMACHI, Y., (2019), HiBiT-qIP, HiBiT-based quantitative immunoprecipitation, facilitates the determination of antibody affinity under immunoprecipitation conditions, *Scientific Reports*, vol. 9, n°1, p. 6895.

REAGIN, M. J., GIESLER, T. L., MERLA, A. L., RESETAR-GERKE, J. M., KAPOLKA, K. M. & MAMONE, J. A., (2003), TempliPhi: A Sequencing Template Preparation Procedure That Eliminates Overnight Cultures and DNA Purification, p. 6.

REICHEL, A., SCHAIBLE, D., AL FUROUKH, N., COHEN, M., SCHREIBER, G. & PIEHLER, J., (2007), Noncovalent, Site-Specific Biotinylation of Histidine-Tagged Proteins, *Analytical Chemistry*, vol. 79, n°22, p. 8590-8600.

ROCKLIN, G. J., CHIDYUSIKU, T. M., GORESHNIK, I., FORD, A., HOULISTON, S., LEMAK, A., CARTER, L., RAVICHANDRAN, R., ... BAKER, D., (2017), Global analysis of protein folding using massively parallel design, synthesis, and testing, *Science*, vol. 357, n°6347, p. 168-175.

SCHNELL, U., CIRULLI, V. & GIEPMANS, B. N. G., (2013), EpCAM: Structure and function in health and disease, *Biochimica et Biophysica Acta (BBA) - Biomembranes*, vol. 1828, n°8, p. 1989-2001.

SCOTT, J. K. & SMITH, G. P., (1990), Searching for Peptide Ligands with an Epitope Library, *Science*, vol. 249, n°4967, p. 386-390.

TSIEN, R. Y., (1998), THE GREEN FLUORESCENT PROTEIN, *Annual Review of Biochemistry*, vol. 67, n°1, p. 509-544.

URVOAS, A., GUELLOUZ, A., VALERIO-LEPINIEC, M., GRAILLE, M., DURAND, D., DESRAVINES, D. C., VAN TILBURGH, H., DESMADRIL, M. & MINARD, P., (2010), Design, Production and Molecular Structure of a New Family of Artificial Alpha-helical Repeat Proteins (α Rep) Based on Thermostable HEAT-like Repeats, *Journal of Molecular Biology*, vol. 404, n°2, p. 307-327.

WU, Y., BATYUK, A., HONEGGER, A., BRANDL, F., MITTL, P. R. E. & PLÜCKTHUN, A., (2017), Rigidly connected multispecific artificial binders with adjustable geometries, *Scientific Reports*, vol. 7, n°1, p. 11217.

YAGODKIN, A., AZHAYEV, A., ROIVAINEN, J., ANTOPOLSKY, M., KAYUSHIN, A., KOROSTELEVA, M., MIROSHNIKOV, A., RANDOLPH, J. & MACKIE, H., (2007), Improved Synthesis of Trinucleotide Phosphoramidites and Generation of Randomized Oligonucleotide Libraries, *Nucleosides, Nucleotides and Nucleic Acids*, vol. 26, n°5, p. 473-497.

Conclusion

Les bibliothèques de protéines associées aux méthodes d'évolution dirigée fournissent un moyen efficace d'explorer de façon expérimentale un vaste espace des séquences, mais centré autour d'une architecture protéique stable, pour trouver une protéine qui possède des propriétés particulières, comme la fixation d'un partenaire spécifique. Les progrès réalisés dans les domaines des biotechnologies et de la biologie moléculaire ont permis de créer des bibliothèques de grande taille, et le répertoire de bibliothèques disponible aujourd'hui permet d'obtenir rapidement des interacteurs pour une cible d'intérêt.

C'est dans ce contexte que nous avons développé une nouvelle bibliothèque de protéines artificielles, comportant non pas un mais deux domaines de reconnaissance, dans le but de créer un outil de reconnaissance moléculaire qui puisse présenter une plasticité conformationnelle, lui permettant ainsi d'être modulable. Nous avons également conçu, produit et caractérisé une molécule effectrice susceptible d'affecter de façon générique l'affinité des interacteurs sélectionnés pour leur cible. Les résultats obtenus sont encore préliminaires, en ce sens que la capacité de la bibliothèque à générer des interacteurs contre de nombreuses cibles cellulaires n'a pas été encore démontrée. Elle l'a seulement été sur une cible modèle : l'eGFP. Cette bibliothèque a été conçue pour être améliorée et diversifiée par recombinaison de façon efficace. Il sera donc essentiel d'exploiter cette possibilité et d'évaluer le potentiel de la bibliothèque ainsi diversifiée. Il restera également à démontrer l'efficacité du modulateur sur les interacteurs spécifiques de cibles variées, et à tester les applications de ces nouveaux outils, par exemple en biologie cellulaire. Le concept exploré dans ce travail de thèse peut être également abordé sous un angle différent : puisque la bibliothèque d' α Reps, ou d'autres comme les bibliothèques de DARPins sont des sources génériques d'interacteurs, il serait utile de développer une solution générique pour les rendre modulables.

Une autre façon d'explorer l'espace des séquences est d'en réaliser une exploration computationnelle, en calculant l'énergie associée à un arrangement tridimensionnel d'une séquence d'acides aminés. L'hypothèse de Christian B. Anfinsen ([Anfinsen, C. B., 1973](#)), qui suppose que la structure d'une protéine correspond à sa conformation de plus basse énergie, suggère qu'il devrait être possible de prédire la structure d'une protéine en calculant l'arrangement spatial de sa séquence d'acides aminés qui possède l'énergie la plus basse. Cependant cette possibilité est restée longtemps non réalisée tant le problème est complexe. En effet, la réussite de cette prédiction dépend d'une part de l'efficacité de l'exploration conformationnelle, et d'autre part de la justesse du modèle mathématique qui attribue un score d'énergie à une conformation donnée. Le logiciel Rosetta développé par le groupe de David Baker a été développé initialement avec cet objectif de prédiction de structure ([Park, H. et al., 2019](#)). Cependant, Rosetta permet d'aller plus loin et de faire le cheminement inverse, c'est-à-dire de trouver la séquence d'acides aminés qui permet de donner une structure prédéfinie ([Baker, D., 2019](#)). Cette

application paraît extrêmement importante, puisqu'elle permet de concevoir des protéines qui n'ont jamais existé dans la nature, et donc de dépasser les limites du répertoire des protéines naturelles. En effet, la nature n'a exploré qu'une petite fraction de l'espace des séquences (Huang, P.-S. *et al.*, 2016), les protéines naturelles étant soumises à une pression de sélection basée sur leur fonction biologique. Cette dernière notion explique d'ailleurs pourquoi les protéines conçues *de novo* sont généralement plus stables que les protéines naturelles qui n'ont pas été soumises à une pression de sélection sur leur stabilité thermique (Baker, D., 2019). De nouvelles ossatures artificielles ont été conçues par le groupe de David Baker (Koepnick, B. *et al.*, 2019; Rocklin, G. J. *et al.*, 2017) et pourraient servir de base à la création de nouvelles bibliothèques de protéines artificielles, puisqu'elles possèdent une grande stabilité thermique.

Le logiciel AlphaFold développé par l'entreprise DeepMind et récemment publié, a dépassé Rosetta dans les compétitions CASP (Jumper, J. *et al.*, 2021), où il s'agit de prédire la structure tridimensionnelle à partir de la seule séquence d'acides aminés. AlphaFold utilise le *machine learning* et des alignements multiples de séquences pour prédire avec une efficacité jusque-là jamais atteinte la structure tridimensionnelle des protéines à partir de leur seule séquence. AlphaFold n'est pas, ou pas encore, développé pour la conception de protéines artificielles ou pour la création d'interactions nouvelles. Lors de notre travail nous avons pu illustrer, à travers un exemple, la complémentarité des approches de type intelligence artificielle qui explore d'une façon nouvelle l'espace des conformations, et les approches de modélisations type Rosetta, qui permettent d'évaluer la qualité d'un *design* sur des critères d'énergie. En effet, s'il intègre implicitement de très nombreuses règles observées dans les structures de protéines, AlphaFold n'est pas guidé explicitement par une grandeur reliée à l'énergie de la protéine. Des tests simples montrent par exemple que ce n'est pas un outil adapté pour évaluer les effets d'une mutation sur la stabilité d'une protéine. Il est donc vraisemblable que les années qui viennent conduiront à une convergence des outils de prédiction utilisant les concepts de l'intelligence artificielle et des outils de conception (*design*) de séquence utilisant des procédures d'exploration et des fonctions de score qui ont préalablement démontré leur efficacité. Rosetta a par ailleurs commencé à inclure le *machine learning* dans un outil baptisé trRosetta qui utilise un réseau neuronal convolutionnel pour prédire les distances et les contacts inter-résidus à partir de d'alignements multiples de séquences. L'utilisation de trRosetta associée à Rosetta a donné de meilleurs résultats par rapport à leur utilisation isolée (Baek, M. *et al.*, 2021; Yang, J. *et al.*, 2020).

Le succès de ces méthodes computationnelles peut poser la question de l'utilité future des bibliothèques de protéines, qui explorent l'espace des séquences de façon expérimentale en créant physiquement un répertoire de protéines diversifiées. L'espace des séquences exploré par les calculs *in silico* est en effet beaucoup plus vaste, et ces méthodes ont connu une progression fulgurante au cours des dernières années. Actuellement la conception uniquement par le calcul de protéines fixant

spécifiquement et à haute affinité une cible donnée n'est pas opérationnelle. En revanche plusieurs travaux, dont ceux du groupe de David Baker, montrent que les approches de *design* associées à des approches de tri d'un grand nombre de solutions sont très efficaces. L'épidémie de COVID-19 (*Coronavirus Disease 19*) est un bon exemple pour comparer sur un cas réel l'efficacité des différentes options possibles pour générer des interacteurs spécifiques. En effet, à partir du début de l'année 2020, un très grand nombre de laboratoires se sont intéressés à trouver des interacteurs des protéines de surface du SARS-Cov-2 (*Severe Acute Respiratory Syndrome Coronavirus 2*) afin d'inhiber l'entrée du virus dans les cellules des voies respiratoires humaines. L'entreprise Molecular Partners exploitant les DARPins développées par le groupe d'Andreas Plückthun, a par exemple utilisé la bibliothèque des DARPins pour isoler un ensemble d'interacteurs, puis les a exploités pour construire un interacteur multi-domaine constitué de la fusion de 3 DARPins, chacune dirigée contre un domaine différent de la protéine *spike* du SARS-CoV-2 (Walser, M. *et al.*, 2020). La protéine *spike* est responsable de la fusion du virion avec les cellules respiratoires de l'hôte en interagissant avec le récepteur humain ACE2 (*Angiotensin-Converting Enzyme 2*) (Cai, Y. *et al.*, 2020), et les interacteurs de *spike* peuvent donc constituer des inhibiteurs de l'infection virale. L'équipe MIP a également réalisé des sélections par phage *display* sur la bibliothèque d' α Reps pour obtenir en quelques semaines des interacteurs possédant une haute affinité pour la protéine *spike* (travaux non publiés). Les α Reps ont été testés avec succès à la fois sur des modèles cellulaires et sur des modèles animaux. Enfin, le groupe de David Baker a conçu par des approches de calcul un grand nombre de candidats interacteurs potentiels du domaine RBD (Receptor Binding Domain) de la protéine *spike*. Mais pour identifier les solutions optimales parmi l'ensemble des séquences conçues théoriquement il a été nécessaire de synthétiser des bibliothèques de gènes correspondant à l'ensemble de ces *designs*, puis de les cribler par exposition sur levure (*yeast display*). Cette stratégie a permis d'identifier des interacteurs de ce domaine avec des affinités allant de 100 nM à 2 μ M. Des étapes de maturation d'affinité ont permis d'obtenir 10 interacteurs possédant des affinités très élevées, allant du picomolaire au nanomolaire (Cao, L. *et al.*, 2020).

Les deux approches - conception computationnelle ou criblage de bibliothèques – sont parvenues à générer des interacteurs de la protéine *spike* en quelques mois. Le groupe de David Baker a cependant adopté une méthode hybride qui a consisté à utiliser la modélisation pour guider un *design* de bibliothèque restreinte, criblée par *yeast display*. De plus, les interacteurs issus de ce criblage ont nécessité des étapes de maturation d'affinité par mutagenèse aléatoire pour parvenir aux dix interacteurs de haute affinité. Cela témoigne de l'importance des bibliothèques de protéines et de l'exploration expérimentale de l'espace des séquences pour concevoir rapidement des interacteurs d'une cible d'intérêt. Il reste que la conception d'interacteurs par calcul computationnel suppose que la structure de la cible soit connue. Ce n'était pas le cas d'un grand nombre de cibles jusqu'à très

récemment, mais la mise à disposition de l'ensemble des structures de protéines humaines prédite par AlphaFold constitue une évolution importante. Il reste à établir si ces structures prédites seront suffisamment précises pour permettre la conception, par le calcul seul, d'interacteurs spécifiques. L'évolution des méthodes et l'augmentation de la puissance de calcul permettra peut-être un jour de s'affranchir des bibliothèques de protéines pour générer des interacteurs de haute affinité, mais dans un futur proche l'association des deux types de méthode - conception computationnelle ou criblage de bibliothèques - semble être la manière la plus fiable d'obtenir rapidement des interacteurs affins et spécifiques pour leur cible.

Bibliographie

- ANFINSEN, C. B., (1973), Principles that Govern the Folding of Protein Chains, *Science, New Series*, vol. 181, n°4096, p. 223-230.
- BAEK, M., DIMAIO, F., ANISHCHENKO, I., DAUPARAS, J., OVCHINNIKOV, S., LEE, G. R., WANG, J., CONG, Q., ... BAKER, D., (2021), Accurate prediction of protein structures and interactions using a three-track neural network, *Science*, vol. 373, n°6557, p. 871-876.
- BAKER, D., (2019), What has de novo protein design taught us about protein folding and biophysics?, *Protein Science*, vol. 28, n°4, p. 678-683.
- CAI, Y., ZHANG, J., XIAO, T., PENG, H., STERLING, S. M., JR, R. M. W., RAWSON, S., RITS-VOLLOCH, S. & CHEN, B., (2020), Distinct conformational states of SARS-CoV-2 spike protein, p. 8.
- CAO, L., GORESHNIK, I., COVENTRY, B., CASE, J. B., MILLER, L., KOZODOY, L., CHEN, R. E., CARTER, L., ... BAKER, D., (2020), De novo design of picomolar SARS-CoV-2 miniprotein inhibitors, p. 7.
- HUANG, P.-S., BOYKEN, S. E. & BAKER, D., (2016), The coming of age of de novo protein design, *Nature*, vol. 537, n°7620, p. 320-327.
- JUMPER, J., EVANS, R., PRITZEL, A., GREEN, T., FIGURNOV, M., RONNEBERGER, O., TUNYASUVUNAKOOL, K., BATES, R., ... HASSABIS, D., (2021), Highly accurate protein structure prediction with AlphaFold, *Nature*, vol. 596, n°7873, p. 583-589.
- KOEPNICK, B., FLATTEN, J., HUSAIN, T., FORD, A., SILVA, D.-A., BICK, M. J., BAUER, A., LIU, G., ... BAKER, D., (2019), De novo protein design by citizen scientists, *Nature*, vol. 570, n°7761, p. 390-394.
- PARK, H., LEE, G. R., KIM, D. E., ANISHCHENKO, I., CONG, Q. & BAKER, D., (2019), High-accuracy refinement using Rosetta in CASP13, *Proteins: Structure, Function, and Bioinformatics*, vol. 87, n°12, p. 1276-1282.
- ROCKLIN, G. J., CHIDYUSIKU, T. M., GORESHNIK, I., FORD, A., HOULISTON, S., LEMAK, A., CARTER, L., RAVICHANDRAN, R., ... BAKER, D., (2017), Global analysis of protein folding using massively parallel design, synthesis, and testing, *Science*, vol. 357, n°6347, p. 168-175.
- WALSER, M., ROTHENBERGER, S., HURDISS, D. L., SCHLEGEL, A., CALABRO, V., FONTAINE, S., VILLEMAGNE, D., PALADINO, M., ... AMSTUTZ, P., (2020), *Highly potent anti-SARS-CoV-2 multivalent DARPIn therapeutic candidates* (preprint), preprint, Immunology.
- YANG, J., ANISHCHENKO, I., PARK, H., PENG, Z., OVCHINNIKOV, S. & BAKER, D., (2020), Improved protein structure prediction using predicted interresidue orientations, *Proceedings of the National Academy of Sciences*, vol. 117, n°3, p. 1496-1503.

Matériels et méthodes

Table des matières

Matériels et méthodes	173
PCR pour le Gibson Assembly :	176
Extraction d'ADN sur gel :	176
Gibson Assembly :	176
Clonages par <i>Golden Gate</i> :	177
Extraction de plasmides :	177
Séquençage des gènes :	178
Transformation bactéries chimiocompétentes :	178
Transformation bactéries électrocompétentes :	178
Milieux de cultures :	179
Production des protéines :	179
Production des protéines biotinylées :	179
Lyse des bactéries :	180
Purification sur colonne Ni-NTA :	180
Filtration sur gel (ou SEC) :	181
Western blot :	181
Size Exclusion Chromatography – Multi Angle Light Scattering (SEC-MALS) :	181
Dichroïsme circulaire :	182
Surface plasmon Resonance (SPR) :	182
Bi-layer interferometry (BLI) :	183
Cristallogénèse :	183
Sélection par phage display d' α Reps contre les P5H (T ou O) :	184
Sélection par phage display de bi-P5H contre l'eGFP :	185
Sélection par phage display de bi-P5H contre EpCAM et TROP 2 :	186
Phage ELISA clonal :	186
Dialyses :	187
	174

Tests d'expression des bi-P5H	187
Construction de la bibliothèque de bi-P5H :	188
Assemblage des oligonucléotides :	188
Amplification des assemblages par RCA :.....	189
Insertion des gènes dans le phagemide :	189
Transformation des bactéries par les plasmides de la bibliothèque :.....	190
Conception <i>in silico</i> du modulateur allostérique :	192
Mise en place du modèle HEK293 surexprimant TROP 2-GFP :.....	192
Fluorescent Activated Cell Sorting (FACS):	195

PCR pour le Gibson Assembly :

Afin de réaliser le Gibson Assembly, il a été nécessaire de rajouter les extrémités compatibles avec cette technique de clonage pour les différents inserts et vecteurs par PCR. Le protocole qui suit a été utilisé pour chaque PCR : dans un tube PCR ont été ajoutés de l'eau distillée (QSP 50 µL), du tampon phusion 5X (10 µL), un vecteur adéquat (25 ng), les 2 amorces correspondantes (2 µM), les dNTPs (2.5 nM), du DMSO pur (1 µL), de la polymérase phusion ThermoFisher F530S (1 µL d'une solution à 2 U/µL). Les conditions suivantes ont été utilisées pour les PCRs : 2 mins à 98°C, 10 cycles de 30 secs à 98°C puis 30 secs à 50°C et 5 mins à 72°C, 20 cycles de 30 secs à 98°C puis 30 secs à 65°C et 5 mins à 72°C, 10 mins à 72°C, puis un maintien à 4°C. Pour les 2 PCRs servant à produire le vecteur pTetar2, les températures d'hybridations utilisées lors des cycles ont été optimisées (60°C pour les 10 cycles et 68°C pour les 20 cycles) ainsi que 3 mins d'élongation au lieu de 5. Pour la PCR servant à produire l'insert T2wt les températures d'hybridation des cycles ont également été optimisées (40°C pour les 10 cycles avec 1 min d'élongation et 65°C pour les 20 cycles avec 5 mins d'élongation). 3 µL des produits PCR ont ensuite été déposés sur gel d'agarose 1% pour réaliser une électrophorèse 20 mins de migration à 100V, et le reste a été conservé à -20°C.

Extraction d'ADN sur gel :

Les bandes d'intérêt ont été découpées sur gel après une migration permettant de bien séparer la bande d'intérêt des contaminants, puis l'ADN a été extrait en utilisant le kit « Nucleospin Gel and PCR Clean-Up » de Macherey Nagel. La bande découpée a été mise à incuber 10 mins à 50°C avec 200 µL pour 100 mg de gel d'un tampon de fixation, afin de dissoudre l'agarose et d'optimiser la fixation de l'ADN. L'ADN a ensuite été fixé sur la membrane de silice par centrifugation en présence de sels chaotropiques afin de dénaturer d'éventuels ligands protéiques, puis cette membrane a été lavée 2 fois par une solution éthanolée afin de laver les impuretés sans éluer l'ADN. Enfin, l'ADN a été élué par une solution faiblement alcaline (5mM Tris/HCl, pH 8.5), après avoir placé la colonne dans un tube eppendorf neuf.

Gibson Assembly :

Les amorces nécessaires à la réalisation du Gibson Assembly ont été conçues à l'aide du logiciel Serial Cloner et commandées chez Eurofins Genomics. Les inserts ont été produits par PCR et purifiés sur gel d'agarose 1% après 20 mins de migration à 100V ou 30 mins à 100V pour les 2 vecteurs. Le Gibson Assembly a ensuite été réalisé dans un appareil à PCR en suivant le protocole qui suit (pour chacun des assemblages) : dans un tube PCR ont été mélangés 100 ng de vecteur, 300 ng d'inserts, 10 µL de master mix Gibson Assembly 2X, de l'eau (QSP 20 µL). Les tubes ont été mis à incuber 15 mins à 50°C puis dans

la glace. 100 µL de bactéries XL1Blue électrocompétentes ont été transformées avec 4 µL de mix Gibson Assembly pour chacun des mix, puis mises en culture dans 900 µL de 2x YT 1h, à 37°C, 100 rpm. 2 x 450 µL ont été étalés pour chacun des mix Gibson Assembly sur boîte 2x YT AG puis mis à incuber ON, à 37°C. Ces boîtes ont ensuite été utilisées pour réaliser des cultures à partir d'un clone isolé afin de faire des minipreps et de récupérer les plasmides d'intérêt.

Clonages par *Golden Gate* :

Les clonages ont été réalisés par *Golden Gate* à partir de gènes synthétiques ou de produits PCR possédant des extrémités adaptées. Le *Golden Gate* est une technique de clonage en une étape qui repose sur l'utilisation d'une enzyme de restriction de type IIs, c'est-à-dire dont le site de coupure est distinct de la séquence reconnue par l'enzyme. Cette particularité a l'avantage de ne pas permettre la reformation du site de coupure dans le vecteur accepteur une fois que le gène est inséré. En d'autres termes, une fois que le gène d'intérêt est inséré dans le vecteur accepteur, celui-ci ne présente plus le site de coupure et ne peut plus être digéré, ce qui tire l'équilibre de la réaction vers l'insertion du gène d'intérêt. Un autre avantage de cette méthode est que la digestion, l'hybridation et la ligation se font en une seule étape. Pour réaliser le *Golden Gate* (kit ThermoFischer), 75 ng de vecteur sont mélangés à l'insert dans un ratio molaire insert : vecteur = 2 : 1, puis du tampon de ligase 10 X est ajouté ainsi que le mix enzymatique du kit qui contient l'enzyme de restriction Bsa I et une ligase. De l'eau miliQ est ajoutée pour compléter au volume final et la réaction est réalisée 1h à 37°C. Les enzymes sont inactivées par 5 mins d'incubation à 60°C. Le produit du Golden Gate peut être utilisé directement pour transformer des bactéries, purifié sur colonne, ou digéré par une autre enzyme de restriction qui permet de linéariser le plasmide de départ.

Extraction de plasmides :

Les extractions de plasmides (ou minipreps) ont été réalisées avec le kit « Nucleospin plasmid » de Macherey Nagel en utilisant le protocole du fournisseur. Dans un premier temps, les bactéries ont été culottées dans un tube eppendorf puis resuspendues dans une solution tampon optimale pour la lyse alcaline. Dans un second temps une lyse alcaline a été réalisée par ajout d'une deuxième solution permettant la libération du contenu cellulaire. Le lysat a ensuite été neutralisé par une troisième solution, qui optimise aussi la fixation de l'ADN plasmidique à la membrane de silice utilisée plus tard dans l'extraction. Les protéines agrégées, l'ADN génomique, et les débris cellulaires ont été culottés par centrifugation, et le surnageant a été chargé sur une colonne du fournisseur qui possède une membrane de silice. L'ADN a été fixé sur la membrane par centrifugation puis les contaminants tels que les sels, les métabolites, et les macromolécules solubles ont été lavées par une solution contenant

de l'éthanol qui va faire précipiter l'ADN et ne va donc pas l'éluer. La membrane a été séchée, puis l'ADN plasmidique a été élué en deux fois par une solution légèrement alcaline (5 mM Tris/HCl pH 8.5) chauffée à 70°C après avoir changé le tube collecteur.

Séquençage des gènes :

Le séquençage a été réalisé par Eurofins Genomics ou par IDT. Pour cela, les échantillons d'ADN à séquencer ont été préparés (15 µL à 50-100 ng/µL) et conservés au froid jusqu'au passage du transporteur.

Transformation bactéries chimiocompétentes :

Les bactéries XL1Blue MRF' chimiocompétentes, Gold BirA chimiocompétentes, Gold chimiocompétentes, ou Rosetta chimiocompétentes (50-100 µL) décongelées sur glace ont été transformées par 100 à 500 ng de plasmide. Après 20 minutes de repos sur glace, un choc thermique de 45 secondes à 42°C a été réalisé (création de pores dans la membrane), puis les bactéries ont été remises 2 minutes sur glace afin de stabiliser les membranes. 300 µL de milieu 2xYT ont ensuite été ajoutés, puis les bactéries ont été mises à incuber 1h, 37°C, 200 rpm. 80 µL de cette culture ont ensuite été étalés sur boîte de Pétri contenant du glucose 1% et les antibiotiques adéquats (ampicilline uniquement quand l'application en aval n'implique pas une infection par des phages et que le maintien des pili par la tétracycline n'est pas nécessaire) puis la boîte a été mise à l'étuve à 37°C, ON.

Transformation bactéries électrocompétentes :

30 à 200 ng de plasmides ont été ajoutés à 50 à 100 µL de bactéries XL1Blue MRF' électrocompétentes décongelées sur glace. Le mélange a été déposé doucement dans une cuvette stérile pré-refroidie dans la glace, puis un choc électrique a été réalisé à 2.5V / 25µFd / 400 Omhs dans un micro-pulser BioRad. 500 µL de milieu 2xYT à température ambiante ont été ajoutés immédiatement après, et les bactéries ont été mises à incuber 1h à 37°C et 100 rpm. 50 µL de cette culture ont ensuite été étalés sur boîte, puis incubés à 37°C ON.

Milieux de cultures :

Les cultures bactériennes ont été réalisées dans du milieu 2xYT (MP biomedicals) avec addition au besoin d'ampicilline (100 µg/mL) pour maintenir le plasmide de la bibliothèque, de tétracycline (15.5 µg/mL) pour maintenir l'expression du pili, de chloramphénicol (35 µg/mL) pour maintenir le plasmide exprimant BirA, de kanamycine (50 µg/mL) pour maintenir le plasmide du phage helper, d'IPTG (119 µg/mL) pour induire l'expression des protéines, de biotine (12 µg/mL) pour permettre la biotinylation *in vivo* des protéines, de glucose 1% pour maintenir l'inhibition de l'expression des protéines, ou d'agar 1.5% (VWR chemicals) pour réaliser un milieu solide.

Production des protéines :

La production des protéines a été réalisée dans des souches d'*E.coli* Gold (αReps et eGFP) ou Rosetta (P5H). A partir des boîtes obtenues à l'issue de la transformation, un clone isolé a été repiqué dans 50 mL de milieu 2xYT AG et mis à incuber à 37°C, 200 rpm, ON. Le lendemain, 2 fois 500 mL de milieu 2xYT Amp ont été inoculés à $DO_{600} = 0.15$ par la préculture dans des erlenmeyers de 2.5L puis mis à incuber à 37°C, 200 rpm. A $DO_{600} = 0.6-0.8$, la production des protéines a été induite par ajout de 500 µL d'IPTG dans chaque erlenmeyer, puis les cultures ont été mises à incuber 4h, à 37°C, 200 rpm. Les cultures ont ensuite été centrifugées 15 mins, 5000g, 4°C. Chaque culot a été repris dans 40 mL de tampon de lavage 20 mM imidazole 50 mM NaP 300 mM NaCl (αReps et eGFP) ou 1M NaCl + 5% glycérol (P5H) à un pH = 8. Une pastille d'antiprotéase a été dissoute dans chaque culot, qui ont été congelés et stockés à -80°C.

Production des protéines biotinylées :

Les productions de protéines biotinylées ont été réalisées dans une souche d'*E. coli* Gold BirA chimiocompétente qui surexprime le gène BirA, codant pour une biotine ligase d'*E. coli*, laquelle biotinye la lysine d'un Avitag placé en C-ter des protéines produites. Le protocole qui suit a été utilisé pour produire les 4 pentahélines à partir des 4 boîtes obtenues après transformation de la souche Gold BirA. Un clone isolé a été repiqué dans 50 mL de milieu 2xYT ACG et mis à incuber à 37°C, 200 rpm, ON. Le lendemain, 2 fois 500 mL de milieu 2xYT AC ont été inoculés à $DO_{600} = 0.15$ par la préculture dans des erlenmeyers de 2.5L puis mis à incuber à 37°C, 200 rpm. A $DO_{600} = 0.6-0.8$, la production des protéines biotinylées a été induite par ajout de 500 µL de biotine + 500 µL d'IPTG dans chaque erlenmeyer, puis les cultures ont été mises à incuber 4h, à 30°C, 200 rpm. Les cultures ont ensuite été centrifugées 15 mins, 5000g, 25°C. Chaque culot a été repris dans 20 mL de TBS + antiprotéase, puis congelé et stocké à -80°C.

La production des bi-P5H a été réalisée à 37°C, 4H, dans des *E.coli* Rosetta selon le même protocole que décrit précédemment.

Lyse des bactéries :

Les bactéries ont été lysées par 6 cycles de sonication de 30 secondes, puis 2 µL de benzonase ont été ajouté au lysat afin de dégrader les acides nucléiques libérés lors de la lyse, rendant le lysat visqueux. Les lysats ont ensuite été centrifugés 45 mins à 11000g, 4°C (αReps et eGFP) ou à 20°C (P5H) afin de culotter les fragments de membrane et les agrégats. Le surnageant (ou fraction soluble) a été transféré dans un nouveau tube en vue de purifier les protéines d'intérêt.

Purification sur colonne Ni-NTA :

Deux types de purification sur colonne Ni-NTA ont été réalisées : les purifications sur paillasse et les purifications sur AKTA pure. Selon la protéine purifiée et l'expérience, différents tampons ont été utilisés.

Tampons de lavage : 20 mM imidazole 150/300/1000/2000 mM NaCl 50mM NaH₂PO₄ (+ 5% de glycérol pour les P5H)

Tampons d'élution : 300 mM d'imidazole 150/300/1000/2000 mM NaCl 50mM NaH₂PO₄ (+ 5 % de glycérol pour les P5H)

Purification sur paillasse : le surnageant a été déposé sur une colonne contenant 4 ou 5 mL de résine Ni NTA préalablement équilibrée avec 10 CV de tampon de lavage. La colonne a été lavée par 20 CV de ce même tampon de lavage, puis la dernière goutte a été collectée dans un tube eppendorf afin de vérifier au Nanodrop-1000 (ThermoScientific) l'absence de protéines en mesurant la DO₂₈₀. La protéine d'intérêt a ensuite été éluée par fractions de 1 CV avec un tampon d'élution, et les fractions d'élution ont été collectées dans des tubes à hémolyse de 5 mL jusqu'à ce que les DO₂₈₀ mesurées soient négligeables (<0.08 mg/mL).

Purification sur AKTA pure : le même principe que la purification sur paillasse a été appliqué mais de façon automatisée. Une colonne HisTrap FF crude de 5 mL a été utilisée (capacité de fixation de 200 mg de protéine marqué (His)₆ selon le fournisseur). L'équilibration a été réalisée avec 3 CV d'eau distillée puis 5 CV de tampon de lavage identique à celui utilisé pour la purification du paillasse. Le lavage a été réalisé avec 20 CV de ce même tampon, et l'élution a été réalisée avec 5 CV de tampon d'élution identique à celui utilisé pour la purification du paillasse, par fractions de 1mL sur une plaque 96 puits de type deep-well Thermo scientific.

Filtration sur gel (ou SEC) :

Les chromatographies d'exclusion de taille (SEC) ont été réalisées sur AKTA pure ou sur AKTA purifier en utilisant une colonne HiLoad 16/600 Superdex 75 prep grade, permettant de séparer avec une très bonne résolution les biomolécules de poids moléculaire compris entre 3000 Da et 70000Da. La collecte a été réalisée par fractions de 2 mL dans des plaques 96 puits deep-well Thermo scientific, à un débit de 1 mL.min⁻¹. Les colonnes ont été équilibrées avec 2 CV de tampon phosphate (NaP 50 mM, NaCl 150 mM pH7) et les échantillons ont été injectés par fraction de 5 mL.

Western blot :

Le Western Blot est une technique qui permet de séparer et d'identifier spécifiquement une protéine d'intérêt à partir d'un mélange de protéines. Pour cela, le mélange contenant la protéine que l'on souhaite identifier a été déposé sur un gel d'acrylamide et une électrophorèse a été réalisée. Les protéines du gel sont ensuite transférées sur une membrane de nitrocellulose pendant 1h à 80V. La membrane a ensuite été bloquée par 20 mL de TBST 0.5% BSA 4% pendant 1h à température ambiante ou sur la nuit à 4°C. La membrane a alors été lavée deux fois 2 mins par 20 mL de TBST 0.1%, puis marquée pendant 30 mins à température ambiante par 10 mL de streptavidine-AlexaFluor680 dilué au 1/20000^{ème} dans du TBST 0.1%. Quatre lavages de deux minutes chacun ont été réalisés, deux par 20 mL de TBST 0.1 % puis deux par 20 mL de TBS. La membrane a été lue sur un scanner Odyssey à 700 nm.

Size Exclusion Chromatography – Multi Angle Light Scattering (SEC-MALS) :

La chromatographie d'exclusion de taille a été effectuée par High Performance Liquid Chromatography (HPLC) sur une colonne Superdex 200 10/300 GL increase. Les échantillons ont été déposés aux concentrations suivantes : P5H Twt = 0.26 mg/mL ; P5H Tmut = 1.75 mg/mL ; P5H Owt = 3.78 mg/mL ; P5H Omut = 3.08 mg/mL. L'HPLC a été couplée à un détecteur de diffusion de lumière aux petits angles miniDAWN TREOS à 3 (Wyatt) et à un réfractomètre T-REX (Wyatt). Le MALS se base sur le principe de diffusion de Rayleigh-Gans-Debye selon lequel la quantité de lumière diffusée est directement proportionnelle au produit de la masse molaire et de sa concentration. De plus, selon ce même principe, la variation d'intensité de lumière diffusée en fonction de l'angle de mesure est directement reliée au rayon hydrodynamique des molécules. Le SEC-MALS permet donc d'obtenir des informations précises sur la masse molaire des protéines, leur état d'oligomérisation, et leur rayon hydrodynamique. Dans un premier temps, la concentration protéique totale est mesurée grâce au signal de réfractométrie. Le signal d'absorbance est alors déconvolué en utilisant les coefficients d'extinction molaire de chaque protéine afin d'obtenir une concentration protéique identique à celle

obtenue par réfractométrie. La proportion de chaque protéine peut alors être déterminé. Ces expériences ont été réalisées en collaboration avec Christophe Velours sur la plateforme SEC-MALS du CNRS.

Dichroïsme circulaire :

Les expériences de dichroïsme circulaire ont été réalisées sur un appareil Jasco, dans le laboratoire de Human Rezaei à l'INRA de Jouy-en-Josas. Les spectres ont été enregistrés 6 fois chacun, entre 180 nm et 280 nm, avec un pas de 0.5 nm. Une gamme de chlorure de guanidinium allant de 0 M à 6 M a été réalisée pour les 4 pentahélines (10 µM finale) à l'exception de P5H Twt qui n'était pas assez concentrée pour réaliser la gamme. 50 µL des échantillons à 10 µM avec ou sans chlorure de guanidinium ont été déposés dans une cuve en quartz de 100 µm de trajet optique. Les données générées ont été traitées sur Excel 2016. Le signal dichroïque a été converti en $\Delta\epsilon$ moyen par résidu.

Surface plasmon Resonance (SPR) :

Les mesures de SPR ont été réalisées sur un ProteOn (Biorad) avec une puce NLC recouverte de neutravidine ou sur un T100 (Biacore) avec une puce neutravidine SA. Ces puces sont recouvertes de neutravidine et fixent donc les ligands biotinylés. Les ligands immobilisés sur les puces ont été soit les 4 P5H de départ (Twt, Tmut, Owt, Omut), soit les modulateurs A12D3 et D3A12. Les analytes ont été respectivement les α Reps sélectionnées contre les P5H, ou les bi-P5H modèles (Twt-Owt ou Owt-Twt). Le principe de cette technique est que les rayons incidents arrivant sur la surface couverte de ligand vont être réfractés avec un certain angle, et que l'interaction d'un partenaire (analyte) avec ce ligand va modifier l'angle de réfraction des rayons incidents. La mesure des variations d'angle de réfraction, pour différentes concentrations d'analytes en solution, permet d'estimer les constantes cinétiques d'interaction (k_{on} et k_{off}) et la constante de dissociation à l'équilibre (K_D) des deux partenaires. Les changements d'angle de réfraction sont mesurés et quantifiés en RU (Resonance Unit). Lors d'une expérience SPR, les concentrations en ligand à fixer doivent permettre d'obtenir le nombre de RL (Resonance Unit Ligand) donné par la formule suivante :

$$RL = R_{max} n * M_{ligand} M_{Analyte}$$

Où n est la stoechiométrie de la réaction, R_{max} l'unité de résonance maximale souhaitée, M_{Ligand} et $M_{Analyte}$ les masses molaires du ligand et de l'analyte respectivement.

Des tests à différentes concentrations ont été réalisés pour chaque analyte dans le but d'estimer le K_D et de pouvoir tester une gamme de concentration de l'analyte entourant le K_D . Pour la détermination des paramètres cinétiques, 6 concentrations différentes d'analytes ont été injectées en parallèle pendant 400 secondes (s) ; le signal de dissociation a été ensuite mesuré pendant 600 s. Les données

ont été analysées avec le logiciel ProteOn Manager, en utilisant un ajustement basé sur le modèle de Langmuir, ou sur le logiciel de traitement de données de Biacore. Les expériences ont été réalisées avec Magali Aumont-Nicaise sur la plateforme d'interaction des molécules (PIM) à l'I2BC, ou avec Magali Noiray sur la plateforme INTERMOL à l'IPSIT.

Bi-layer interferometry (BLI) :

Le BLI est une technique de mesure d'interaction dont le principe physique repose sur la mesure des variations d'épaisseur optique à la surface d'un biocapteur. Cette mesure est évaluée grâce aux interférences créées lors de la réflexion d'un faisceau lumineux à la surface du biocapteur, sur lequel une première molécule (le ligand) est capturé, et la fixation du deuxième partenaire (l'analyte) est mesurée à la surface du biocapteur. La réponse est mesurée en temps réel (sensorgramme), et donne accès aux constantes cinétiques de l'interaction, le k_{on} et le k_{off} . Cette approche permet donc de mesurer les interactions moléculaires entre deux partenaires.

Pour nos expériences, nous avons fixé l'eGFP biotinylée sur des sondes recouvertes de streptavidine. Ces sondes ont été plongées dans des solutions contenant différentes concentrations de la bi-P5H D5 (de 31 nM à 500 nM), puis la dissociation a été réalisée dans le tampon des protéines. Pour les expériences d'inhibition par A12D3, les mêmes gammes de D5 ont été réalisées en présence de 300 nM d'A12D3. Les expériences ont été réalisées sur un OctetRED96e (ForteBio), et les résultats ont été analysés avec le logiciel du fournisseur.

Cristallogénèse :

Toutes les expériences de cristallographie ont été effectuées à 293K en utilisant la méthode de diffusion en phase vapeur. Un criblage des conditions initiales a été effectué à partir des kits commerciaux (Qiagen Classics, MB Class I, PEG I, PEG II, JCSG+). Le modulateur A12D3 a donné des cristaux dans plusieurs conditions, mais la structure a été résolue à partir des cristaux obtenus dans une solution contenant 7% de PEG8000 et 0.2 M d'acétate de magnésium dans un rapport protéine : agent cristallisant 2 : 1. Les cristaux ont été congelés dans l'azote liquide après trempage dans la solution de cristallisation complétée avec 30 % de glycérol. Les données de diffraction ont été collectées à 100K sur la ligne de lumière Proxima 1 du synchrotron SOLEIL. Les images ont par la suite été analysées avec le programme XDS. Les structures ont été résolues par remplacement moléculaire en utilisant PHASER 7 implémenté dans la suite de logiciel ccp4. Le modèle que nous avons conçu *in silico* a été utilisé comme modèle. La carte de densité électronique expérimentale a été traitée en utilisant le programme DM. Les modèles initiaux ont ensuite été complétés manuellement avec le logiciel COOT. L'affinement des structures a été effectué à l'aide de REFMAC.

La structure cristallographique de A12D3 a été obtenue avec une résolution de 3.54 Å avec un R et un RFree de 29.8 et 31.8 %.

L'ensemble des représentations 3D des molécules a été obtenu avec le logiciel PyMol (Schrödinger).

Sélection par phage display d' α Reps contre les P5H (T ou O) :

Le principe du phage display a été décrit dans la partie résultats, mais consiste globalement à cribler un large répertoire de protéines dans le but d'identifier des interacteurs potentiels d'une cible d'intérêt. Pour cela, la cible est immobilisée sur une surface, puis mise en contact de la bibliothèque de protéines exposée à la surface de phages filamenteux M13. L'hypothèse est que les phages exposants des interacteurs de la cible seront retenus sur la surface, et ces phages sont alors récupérés, amplifiés, et utilisés pour réaliser un nouveau tour de sélection. À la fin de la sélection, on obtient donc un ensemble de phages exposants potentiellement des interacteurs de la cible d'intérêt.

Tour 1 : 2 plaques ELISA (P1) et (P2) ont été recouvertes (colonne 1) par 100 μ L/puits de streptavidine à 20 μ g/mL dans du TBS et incubée 2h à 37°C et 300 rpm. La cible a été immobilisée sur P1 et la plaque P2 a été utilisée pour éliminer les phages fixés de façon non spécifique. Après 3 lavages par 300 μ L/puits de TBST 0.1%, les 2 plaques ont ensuite été incubées avec 300 μ L/puits de TBST 0.1% BSA 4% ON à 4°C, 300 rpm, afin de bloquer les zones de la plaque n'ayant pas fixé de streptavidine.

Les 2 plaques ont été lavées 3 fois par 300 μ L/puits de TBST 0.1% puis la cible (P5H Twt ou P5H Owt) a été déposée sur P1 à 8 μ M (100 μ L/puits) tandis que les phages de la bibliothèque d' α -Rep ont été déposés sur P2 (130 μ L/puits). P1 et P2 ont été incubées 1h30 à 25°C, puis P1 a été lavée 3 fois par 300 μ L/puits de TBST 0.1%. 100 μ L/puits des phages de P2 ont ensuite été transférés dans les puits correspondant sur P1 et la plaque a été incubée 1h30 à 25°C. 20 lavages par 300 μ L/puits de TBST 0.1% puis 20 lavages par 300 μ L/puits de TBS ont été réalisés sur la plaque, puis l'élution à la TEV protéase a été menée par ajout de 100 μ L/puits de TEV protéase à DO=0.7 diluée 14 fois dans du TBS 1X, DTT 1 mM, EDTA 0.5 mM, après avoir transféré le 40^{ème} lavage dans un falcon de 15 mL. Après 1h d'incubation à 37°C, 300 rpm, l'élution à la TEV protéase a été récoltée dans un tube de 1.5 mL, puis une élution par ajout de 100 μ L/puits de glycine pH 2.5 a été réalisée pendant 10 mins à température ambiante sous agitation. Les éluats ont été neutralisés par ajout de 5 μ L/puits de Tris 1 M, puis collectés dans un tube de 1.5 mL. Une série de dilutions en cascades a ensuite été réalisée afin de réaliser des infections de bactéries pour le comptage de colonies sur boîte le lendemain. Pour cela, les dilutions de phages suivantes ont été réalisées :

input (phages de la bibliothèque d' α -rep au tour 1) : 10^{-2} 10^{-4} 10^{-6} 10^{-7} 10^{-8} 10^{-9}

output (élution à la TEV protéase) : 10^{-2} 10^{-3} 10^{-4}

40^{ème} lavage : 10^{-2} 10^{-3} 10^{-4}

Parallèlement à la sélection sur plaque, un clone isolé de XL1 Blue MRF' a été repiqué dans 10 mL de 2xYT puis mis à incuber ON, 37°C, 200 rpm. Le lendemain, une culture de XL1 Blue MRF' a été inoculée à DO600=0.15 à partir de la préculture de nuit, puis mise à incuber à 37°C, 200 rpm, jusqu'à DO600= 0.6-0.8. 100 µL de chaque dilution de phage ont été ajoutés à 200 µL de la culture de XL1 Blue MRF' puis mis à incuber 30 min à 37°C, 200 rpm. 1.5 mL de la culture de XL1 Blue MRF' ont été infectés par le reste de l'éluion à la TEV protéase et mis à incuber 30 mins à 37°C, 200 rpm. Le reste des dilutions et l'éluion à la glycine pH 2.5 ont été conservé au frigidaire à 4°C. Les 1.5 mL de bactéries ont été étalés sur grande boîte 15x15 cm 2xYT ATG et 200 µL des infections par les dilutions ont été étalés sur des boîtes de Pétri « classiques » 2xYT ATG. Toutes les boîtes ont été mises à l'étuve ON à 37°C. Le lendemain, le comptage a été réalisé puis les bactéries obtenues sur la grande boîte 15x15 ont été récupérées dans un volume final de 13 mL de 2xYT glucose 1% glycérol 15%. La DO600 a été mesurée et une culture de 50 mL de 2xYT AT a été ensemencée à DO600= 0.15 et mise à incuber à 37°C, 200 rpm. A DO600=0.5-0.8, la culture a été infectée par 20x plus de phages helper que de bactéries, puis mise à incuber 30 min à 37°C, 0 rpm, et 30 min à 37°C, 100 rpm. Le contenu de l'erlenmeyer a été centrifugé 15 mins à 4000g, 25°C sur rotor FX6100, puis les culots ont été repris dans 50 mL de 2xYT AK et mis à incuber ON à 30°C, 200 rpm. Le lendemain, le contenu de l'erlenmeyer a été centrifugé 1h, 4°C, 10000g. Les surnageants (phages non dialysés du tour 1) ont été dialysés ON dans un boudin de dialyse de MWCO 300 kDa dans 1 L de TBS renouvelé le soir. Le reste des phages non dialysés a été aliquoté par fractions de 13 mL et stockés à -80°C. Le lendemain, le contenu du boudin a été récupéré (phages dialysés du T1) puis aliquoté par fractions de 1.4 mL puis stocké à -80°C. Tour 2 : Le même protocole qu'au tour 1 a été appliqué avec pour seuls changements la cible déposée sur P1 (P5H Tmut et P5H Omut pour le tour 2) et les phages déposés sur P2 (dépôt des phages dialysés du tour 1). Pour les comptages une dilution 10⁻⁵ a été ajoutée pour l'output et une dilution 10⁻¹ a été ajoutée pour le lavage.

Tour 3 : Le même protocole qu'au tour 1 a été appliqué avec ajout des mêmes dilutions qu'au tour 2 pour le comptage.

Sélection par phage display de bi-P5H contre l'eGFP :

Pour les sélections contre l'eGFP, le même protocole que décrit pour la sélection d'αReps contre les P5H a été suivi à l'exception de quelques paramètres. Pour la fixation de la cible dans les puits, l'eGFP n'était pas biotinylée. À la place, nous l'avons immobilisée grâce à son His-tag par l'intermédiaire de Tris-NTA-biotin (BiotechRabbit) lui-même immobilisé par de la streptavidine au fond des puits. Le Tris-NTA-biotin est une molécule biotinylée capable d'interagir avec les étiquettes poly-histidines avec une

très forte affinité. Les plaques P1 et P2 ont été recouvertes (colonne 1) de 100 µL/puits de streptavidine à 20 µg/mL, puis incubées 2h à 37°C et 300 rpm. Les lavages et le blocage ont été réalisés comme décrit précédemment. Le lendemain, un mélange a été réalisé entre du Tris-NTA et une solution de NiSO₄ à une concentration finale de 1 µg/mL de Tris-NTA et 5 mM de NiSO₄. Ce mélange a été incubé 30 mins à température ambiante. Les plaques P1 et P2 ont été lavées comme décrit précédemment, puis 100 µL/puits du mélange Tris-NTA/NiSO₄ ont été ajoutés. Après 1h d'incubation à 20°C et 300 rpm, les plaques ont été lavées et 100 µL/puits d'eGFP ont été déposés à 1 µM sur la plaque P1 tandis que 130 µL/puits de phages de la bibliothèque de bi-P5H ont été déposés sur la plaque P2. Après une incubation d'1h à 20°C et 300 rpm, la plaque P1 a été lavée, et 100 µL/puits des phages de la plaque P2 ont été transférés sur la plaque P1. La plaque P1 a été incubée 1h à 20°C et 300 rpm, puis 20 lavages par 300 µL/puits de TBST 0.1% ont été réalisés. Le dernier lavage a été récupéré pour évaluer le nombre de phages retenus de façon aspécifique. Les phages encore fixés sur la plaque ont été élués par ajout de 100 µL/puits de glycine à pH = 2.5 pendant 10 mins sous agitation. La glycine a été neutralisée par ajout de 5 µL/puits de Tris 1 M, et les éluats ont été utilisés pour infecter des bactéries comme décrit précédemment.

Les tours 2 et 3 ont été réalisés de la même façon que le tour 1, en utilisant les phages amplifiés du tour précédent.

Sélection par phage display de bi-P5H contre EpCAM et TROP 2 :

Le même protocole que la sélection contre l'eGFP a été utilisé à quelques paramètres près. Seulement 6 lavages ont été réalisés aux tours 1, puis 20 lavages aux tours 2 et 3. La cible a également été déposée à 1 µM, et a été immobilisée par son étiquette poly-histidine.

Phage ELISA clonal :

Le protocole qui suit a été utilisé pour chaque sélection. Une plaque matrice de culture de cellules 96 puits a été préparée par ajout de 150 µL/puits de 2xYT ATG. 1 clone isolé/puits a été repiqué à partir des boîtes de comptage de l'éluat à la TEV protéase (2 lignes avec des clones du tour 2 et 6 lignes avec des clones du tour 3). La plaque a été mise à incuber ON à 37°C, 600 rpm. Le lendemain, une plaque infection a été préparée par ajout de 150 µL/puits de 2YT ATG puis inoculée par 15 µL/puits de la plaque matrice. La plaque infection a été incubée 4h à 37°C, 600 rpm tandis que 50 µL/puits de glycérol 70% dans du 2xYT ont été ajoutés à la plaque matrice, stockée par la suite à -80°C. La plaque infection a été infectée par 5 µL/puits de phage helper K1 à 6.10¹³ phages/mL puis incubée 30 min à 37°C, 0 rpm, et 30 min à 37°C, 100 rpm. Pour chaque puits, 100 µL de la plaque d'infection ont été transférés dans une plaque deep-well (Thermo scientific) contenant 1.5 mL/puits de 2xYT AK. La plaque

deep-well a été incubée ON à 37°C, 200 rpm. Parallèlement, deux plaques ELISA ont été recouvertes par 100 µL/puits de streptavidine à 20 µg/mL puis incubées 2h à 25°C, 300 rpm. Après 3 lavages par 300 µL/puits de TBST 0.1%, les plaques ont été bloquées par 300 µL/puits de TBST 0.1% BSA 4% ON à 4°C, 300 rpm. Le lendemain, les plaques ELISA ont été lavées 3 fois par 300 µL/puits de TBST 0.1% puis la cible (P5H Twt ou P5H Owt) a été déposée (100 µL/puits à 8µM) 1 ligne sur 2, tandis que les autres lignes ont été remplies par 100 µL/puits de TBST 0.1%. Les plaques ELISA ont ensuite été incubées 1h30 à 25°C, 300 rpm. La plaque deep-well a été centrifugée 1h à 4°C, 2000g sur rotor SX4750A, puis le surnageant (environ 1.2-1.3mL de phages) a été transféré sur une nouvelle plaque deep-well qui sera utilisée par la suite. Les plaques ELISA ont été lavées 3 fois par 300 µL/puits de TBST 0.1% puis 100 µL/puits des phages de la plaque deep-well ont été déposés sur les plaques ELISA en répliquant chaque ligne sur celle du dessous (1 ligne spécifique recouverte avec la cible et 1 ligne non spécifique recouverte avec du TBST). Les plaques ELISA ont ensuite été incubées 1h30 à 25°C, 300 rpm, puis lavées 3 fois par 300 µL/puits de TBST 0.1%. Les plaques ELISA ont été remplies par 100 µL/puits d'anticorps anti-phages M13 dilués 4000 fois dans du TBST, puis incubées 1h30 à 25°C, 300 rpm. Après 3 lavages par 300 µL/puits de TBST 0.1% et 1 lavage par 300 µL/puits de TBS, 100 µL/puits de substrat BM Blue POD soluble (Roche Diagnostics) ont été ajoutés aux plaques ELISA. La réaction a été arrêtée à l'apparition de la couleur bleutée par ajout de 100 µL/puits de HCl 1 M. L'absorbance des plaques a ensuite été lue à 405 nm au lecteur de plaque TECAN Infinite F200 Pro.

Dialyses :

Pour réaliser ces expériences (CD et SEC MALS), les 4 pentahélines, obtenues à partir des fractions d'élution de la purification sur colonne Ni-NTA décrite précédemment, ont été dialysées dans un tampon phosphate 50 mM NaCl 150 mM pH = 8.0 afin d'éliminer les 300 mM d'imidazole contenus dans le tampon d'élution. Les 4 protéines ont ensuite été purifiées par SEC, et les fractions pures ont été regroupées. Pour les expériences de dichroïsme circulaire, les 4 pentahélines ont été dialysées dans un tampon phosphate 150 mM pH = 8.0 sans NaCl. En effet, le NaCl absorbant aux faibles longueurs d'onde (≤ 205 nm), il est préférable de l'éliminer afin de ne pas interférer avec l'absorbance des échantillons à ces longueurs d'onde.

Tests d'expression des bi-P5H

Les gènes codants pour les bi-P5H modèle (Twt-Owt et Owt-Twt) ont été insérés dans un phagemide, qui permet d'exposer les protéines à la surface des phages filamenteux M13. Des plaques 96 puits ont été recouvertes soit de 100 µL/puits d'anticorps anti-HA-tag ou anti-FLAG-tag à 10 µg/mL, soit de 100 µL/puits de streptavidine à 20 µg/mL. Après 3 lavages par 300 µL/puits de TBST 0.1%, les puits recouverts de streptavidine ont été incubés 1h à 20°C avec 200 µL/puits d'une solution purifiée des

modulateurs A12D3 ou D3A12 à plusieurs mg/mL. Les puits ont été blosués selon le même protocole que pour les ELISA clonaux, puis lavés. Tous les puits ont été mis en contact de 200 µL/puits de phages exposants les différentes constructions à 10¹¹ pfu/mL, puis la révélation des phages dans les puits a été effectuée comme pour le phage ELISA clonal. L'absorbance à 405 nm a été lue sur un TECAN Infinite F200 Pro.

Construction de la bibliothèque de bi-P5H :

Assemblage des oligonucléotides :

Les oligonucléotides nécessaires à la construction de la bibliothèque ont été commandés chez EllaBiotech. Les oligonucléotides constants ont été commandés phosphorylés, contrairement aux oligonucléotides dégénérés. Les oligonucléotides dégénérés (FOR 3, FOR 4, FOR 5, FOR 10, FOR 11, FOR 12) ont été repris à 119 µM dans de l'eau miliQ, tandis que les oligonucléotides constants ont été repris à 100 µM dans de l'eau miliQ. Les oligonucléotides dégénérés ont été phosphorylés selon le protocole suivant :

Dans un tube ont été introduit

- 42 µL d'oligonucléotides dégénérés (à 119 µM)
- 5 µL de tampon ligase 10X (pour avoir de l'ATP dans le tampon)
- 2 µL d'eau miliQ
- 1 µL de T4 Polynucleotide Kinase (PNK)

Le mélange a été incubé 1h à 37°C, puis 10 mins à 75°C pour inactiver l'enzyme. On obtient alors 50 µL d'une solution contenant les oligonucléotides dégénérés phosphorylés à 100 µM.

Afin d'optimiser l'hybridation correcte des oligonucléotides entre eux, ils ont été préassemblés pour former des « blocs » selon le schéma suivant :

BLOC 1 : FOR 1 + REV 1 + REV 2 + FOR 2 (1 µL de chaque oligonucléotide à 100 µM) + 1 µL de tampon NEB 3 (New England Biolab) + 5 µL eau miliQ

BLOC 2 : FOR 3 + REV 3 + FOR 4 (1 µL de chaque oligonucléotide à 100 µM) + 1 µL de tampon NEB 3 + 6 µL eau miliQ

BLOC 3 : REV 4 + FOR 5 (1 µL de chaque oligonucléotide à 100 µM) + 1 µL de tampon NEB 3 + 7 µL H₂O eau miliQ

BLOC 4 : REV 5 + FOR 6 + REV 6 (1 µL de chaque oligonucléotide à 100 µM) + 1 µL de tampon NEB 3 + 6 µL eau miliQ

BLOC 5 : FOR 7 + REV 7 + FOR 8 (1 µL de chaque oligonucléotide à 100 µM) + 1 µL de tampon NEB 3 + 6 µL eau miliQ

BLOC 6 : REV 8 + FOR 9 + REV 9 (1 μL de chaque oligonucléotide à 100 μM) + 1 μL de tampon NEB 3 + 6 μL eau miliQ

BLOC 7 : FOR 10 + REV 10 + FOR 11 (1 μL de chaque oligonucléotide à 100 μM) + 1 μL de tampon NEB 3 + 6 μL eau miliQ

BLOC 8 : REV 11 + FOR 12 (1 μL de chaque oligonucléotide à 100 μM) + 1 μL de tampon NEB 3 + 7 μL eau miliQ

BLOC 9 : REV 12 + FOR 13 + REV 13 (1 μL de chaque oligonucléotide à 100 μM) + 1 μL de tampon NEB 3 + 6 μL eau miliQ

Les tubes contenant les blocs ont été incubés 15 mins à 70 °C pour casser d'éventuelles structures secondaires puis 15 mins à 44°C pour hybrider les oligonucléotides entre eux de façon spécifique.

Le mélange suivant a alors été réalisé à 44°C en attendant 5 mins entre l'ajout de deux blocs pour laisser l'hybridation se faire entre les blocs précédents :

- 9 μL de chaque bloc
- 10 μL de tampon ligase 10 X
- 9 μL d'eau miliQ

Le mélange a reposé 5 mins à 44°C puis 1 μL de T4 DNA ligase (ThermoFischer) à 10U/ μL ont été ajoutés, et la ligation a été réalisée pendant 1h à 37°C pour donner un cercle d'ADN double brin codant pour les variants de la bibliothèque.

Amplification des assemblages par RCA :

Les cercles d'ADN ainsi formés ont été amplifiés par 10 réactions de RCA (kit TempliPhi de Cytiva) pour générer la quantité de matériel nécessaire à la construction de la bibliothèque. 10 tubes ont été préparés comme indiqué ci-après. 1 μL du produit d'assemblage a été mélangé à 10 μL du tampon « *sample buffer* » du kit qui contient les hexamères aléatoires permettant d'amorcer la réaction d'amplification. Ce mélange a été incubé 3 mins à 95 °C pour séparer les brins d'ADN, puis refroidi sur pailleuse pour permettre l'hybridation progressive des amorces. 10 μL du tampon « *reaction buffer* » puis 0.4 μL du mix enzymatique du kit ont alors été ajoutés au mélange qui a été incubé environ 18h à 30°C. La polymérase a été activée en incubant les mélanges réactionnels 15 mins à 65°C, puis 70 μL d'eau miliQ et 10 μL de tampon NEB 3 ont y été ajoutés pour fluidifier le mélange réactionnel rendu très visqueux par l'amplification de l'ADN. L'amplification a été vérifiée sur gel (voir résultats).

Insertion des gènes dans le phagemide :

Le produit des 9 RCA qui ont fonctionné et le phagemide (le vecteur accepteur qui permet de réaliser le phage *display*) ont été digérés par l'enzyme de restriction BsaI pour générer des extrémités complémentaires et permettre l'insertion des gènes codants pour les variants de la bibliothèque dans le phagemide. Pour cela, une solution mère 10 fois concentrée a été préparée contenant 870 μL d'eau

miliQ, 100 µL de tampon NEB 3, et 30 µL de l'enzyme de restriction Bsa I. Pour chacune des 9 RCA ayant fonctionné, 100 µL du produit d'amplification ont été ajoutés à 100 µL de la solution mère et 800 µL d'eau miliQ, puis une incubation à 37°C a été réalisée sur la nuit (environ 18h). La digestion des produits d'amplification a été vérifiée sur gel (voir résultats), et la quantité d'insert correctement digéré a été estimée à 1.5 µg d'ADN par tube, soit 13.5 µg totaux. Comme la digestion n'était pas totale, 2 µL de Bsa I ont été rajoutés dans chaque tube qui ont alors été incubés 2h à 37°C.

Parallèlement, 70 µg du phagemide ont été digérés par Bsa I en 10 réactions contenant 7 µg de plasmide chacune. Une solution mère pour 11 réactions a été préparée, contenant 55 µL de tampon NEB 3, 22 µL de l'enzyme Bsa I, et 198 µL d'eau miliQ. 25 µL de phagemide (7µg) ont été mélangés à 25 µL de la solution mère (20U d'enzyme) et ce mélange réactionnel a été incubé 2h à 37°C. La digestion totale a été vérifiée sur gel.

Les produits de digestion des produits d'amplification d'une part et du phagemide d'autre part ont été purifiés sur colonne avec le kit « *gel and PCR clean up* » (Macherey Nagel) selon le protocole du fournisseur optimisé, c'est-à-dire en réalisant une élution en deux fois avec un tampon chauffé à 70°C. Cette étape a permis de récupérer les inserts et le vecteur accepteur digérés et donc avec des extrémités compatibles. 8 réactions de ligation ont alors été réalisées pour insérer les inserts codants pour les variants de la bibliothèque dans le vecteur accepteur compatible avec le phage *display*. Une solution mère pour 11 réactions a été préparée contenant 110 µL de tampon T4 DNA ligase, 33 µL de T4 DNA ligase, et 77 µL d'eau miliQ. 20 µL de la solution mère ont été mélangés à 40 µL de phagemide digéré (4.6 µg) et 40 µL d'inserts digérés (1.3 µg) dans un ratio molaire insert : vecteur = 1.3 : 1. Les produits de ligation ont été purifiés sur colonne comme décrit précédemment mais en réalisant l'élution avec un tampon d'élution dilué 10 fois pour diminuer la force ionique. En effet, l'étape suivante consiste à transformer des bactéries par électroporation avec les produits de ligation purifiés, et une trop forte salinité du mélange peut causer la formation d'arcs électriques défavorable à la survie des bactéries transformées.

Transformation des bactéries par les plasmides de la bibliothèque :

La dernière étape a consisté à transformer des bactéries avec les plasmides codants pour tous les variants de la bibliothèque pour générer la bibliothèque. Dans cette optique, 36 tubes contenant 50 µL d'*E.coli* XL1Blue supercompétentes (Agilent) ont été transformés par électroporation par 4 µL des produits de ligation (soit environ 120 ng de la bande d'intérêt) dans des cuves Eurogentec 1mm CE-0001-50 refroidies sur glace. 500 µL de milieu 2xYT ont été ajoutés immédiatement après le choc électrique, et les produits de transformation ont été incubés 1h à 37°C sous une agitation de 200 rpm. Les 36 tubes ont été réunis en 1 seul tube, à partir duquel des dilutions allant de 10^{-2} à 10^{-6} ont été réalisées. 200 µL des dilutions ont été étalées sur des boîtes ATG, et 450 µL de la culture non diluée

ont été étalés sur 36 grandes boîtes ATG (20 x 20 cm). Les boîtes ont été incubées dans une étuve à 37°C sur la nuit.

Le lendemain, une solution de 2xYT contenant 1% de glucose et 15% de glycérol a été préparée. Les grandes boîtes ont été grattées en deux fois avec 7 mL puis 3 mL de la solution préparée, dans le but de récupérer les bactéries ayant poussées sur la nuit. Les bactéries collectées ont été réunies et une densité optique à 600 nm de 35.6 a été mesurée pour l'ensemble des bactéries. Les bactéries ont alors été distribuées par fraction de 1.6 mL dans des tubes de 2 mL (197 tubes), puis stockées à – 80°C. Les boîtes sur lesquelles les dilutions ont été étalées ont été comptées et ont servi à déterminer la taille de la bibliothèque. 17 colonies ont été obtenues pour 200 µL de la dilution 10^{-6} étalés, soit 1.377×10^9 clones pour les 16.2 mL étalés sur les grandes boîtes.

Conception *in silico* du modulateur allostérique :

Pour concevoir le modulateur allostérique, nous avons décidé de fusionner les α Reps A12 et D3 par une jonction en hélice α (voir résultats). Pour cela, nous avons établi un modèle de structure pour ces deux α Reps, en utilisant la structure cristallographique d'une α Rep à 3 motifs HEAT interne. En effet, A12 et D3 possèdent toutes deux 3 motifs internes, et comme la structure des α Reps est superposable d'une structure à l'autre, nous avons émis l'hypothèse que le modèle créé serait proche de la vraie structure de ces deux α Rep. La séquence de l' α Rep choisie a été mutée pour correspondre à celle de la protéine A12, ou D3. La description qui suit concerne le modulateur A12D3, mais le même protocole a été réalisé pour le modulateur D3A12 en inversant les deux domaines. La première hélice α (en N-ter) du modèle de D3 a été supprimée, car son orientation n'était pas compatible avec la fusion que nous souhaitions réaliser. Dans le logiciel Foldit, 10 alanines ont été rajoutées à l'extrémité N-terminale de D3, en imposant une conformation en hélice α de façon à prolonger la première hélice de D3. Ces 10 alanines ont vocation à guider l'alignement des hélices α entre A12 et D3 et ne sont pas présentes dans le modèle final. Dans le visualisateur de protéines pymol, les 10 derniers résidus de A12 ont été superposés aux 10 alanines formant une extension de la première hélice α de D3. De cette façon, la dernière hélice α de A12 et la première hélice α de D3 étaient alignées comme s'il ne s'agissait que d'une seule hélice. Les 10 alanines servant de guide ont été supprimées, et les deux chaînes polypeptidiques ont été fusionnées en une seule, afin de générer un fichier correspondant aux deux domaines A12 et D3 joints par une hélice étendue. Ce fichier a été importé dans Foldit, et l'hélice servant de jonction a subi une minimisation d'énergie. Nous avons pris le parti de ne pas utiliser cette fonction de minimisation d'énergie sur l'intégralité de la protéine pour éviter que des biais ne soient introduits. Trois positions jugées déstabilisantes (résidus hydrophobes) ont été mutés en résidus hydrophiles.

Mise en place du modèle HEK293 surexprimant TROP 2-GFP :

Des cellules HEK293 adhérentes, cultivées dans un flacon de culture, ont été traitées avec 2.5 mL de Trypsine-EDTA après élimination des 5mL de milieu de culture DMEM contenant 10% FCS + NEAA + Pyruvate+ PS. Le flacon a été incubé à 37°C pendant environ 10 minutes. Les cellules n'étaient plus adhérentes et après 3 ou 4 rinçages, les cellules étaient bien individualisées. Les 2,5 mL de cellules non adhérentes ont été ajoutés à 7,5 mL de milieu complet DMEM et centrifugées à 1200 rpm pendant 10 minutes à 6°C. Après centrifugation, le surnageant a été jeté et le culot cellulaire a été remis en suspension avec précaution. 5 mL de milieu complet DMEM frais ont été ajoutés et les cellules ont été comptées dans une lame KOVA. Deux plaques de six puits (qu'on nommera A et B) ont été préparées.

Dans chaque puits, 3 mL de milieu DMEM complet préchauffé à 37°C ont été ajoutés. Les cellules HEK293 ont été ajoutées selon le plan suivant pour chacune des plaques :

Puit n°1 : 117000 cellules /puits

Puit n°2 : 58800 cellules/puits

Puit n°3 : 29400 cellules/puits

Puit n°4 : 14700 cellules/puits

Puit n°5 : 7350 cellules/puits

Puit n°6 : 3675 cellules/puits

Les deux plaques de 6 puits ont été incubées à 37°C pendant le week-end. Nous avons ensuite examiné attentivement les plaques de six puits afin de déterminer quels puits étaient les plus appropriés pour réaliser la transfection. Nous avons trouvé que le puits n°1 avait une confluence adéquate de cellules HEK293 (environ 80%) pour la transfection.

Les surnageants des puits ont été retirés dans les puits n°1 des deux plaques A et B. 2,5 mL de milieu DMEM complet préchauffé à 37°C a été ajouté à chaque puits et les plaques ont été incubées à 37°C pendant environ 1 heure.

Deux tubes ont été préparés comme indiqué ci-après.

Tube A : 0,1 mL d'Opti-MEM + 24 µL de lipofectamine LTX. Homogénéisation et incubation à température ambiante pendant 5 minutes.

Tube B : 0,3 mL d'Opti-MEM + 5 µL de réactif Plus + 50 µL d'ADNc plasmidique (5 µg). Ce mélange a été doucement homogénéisé et incubé à température ambiante pendant 5 minutes.

Les contenus des tubes A et B ont été mélangés ensemble, puis incubé à température ambiante pendant 15 minutes. Après 5 minutes, 280 µL de ce mélange ont été ajoutés goutte à goutte dans les puits n°1 des plaques A et B. Les plaques ont été doucement agitées et incubées à 37°C.

La plaque A a été incubée pendant 3h30 à 37°C. Puis le surnageant du puits n°1 a été soigneusement retiré et 2,5 mL de DMEM (milieu complet) préchauffé ont été ajoutés. La plaque a été mise à incuber dans une étuve à 37°C.

La plaque B a été incubée pendant une nuit à 37°C. Puis, le surnageant du puits n°1 a été retiré et 2,5 mL de DMEM préchauffé (milieu complet) ont été ajoutés. La plaque a été mise à incuber dans une étuve à 37°C.

Les deux plaques de 6 puits ont été examinées avec un microscope possédant un filtre GFP pour déterminer si la transfection a été efficace.

Le puits n°1 de la plaque A (3h30) et de la plaque B (ON) contenaient de nombreuses cellules fortement fluorescentes. Les deux puits n°1 contenaient des cellules mortes mais il semble qu'il y avait plus de cellules mortes dans la plaque B (ON) que dans la plaque A (3H30). Après une agitation douce, les surnageants des puits n°1 ont été retirés et 1,2 mL de Trypsine EDTA ont été ajoutés à chaque plaque,

qui ont alors été incubées à 37°C pendant 10 minutes. Les cellules ont été remises en suspension et rincées plusieurs fois pour dissocier les agrégats. Ensuite, les surnageants ont été mélangés à 3,5 mL de milieu DMEM complet frais pour éteindre l'activité de la trypsine. Puis, les deux tubes ont été centrifugés à 1200 rpm pendant 10 minutes à 6°C. Après élimination des surnageants, les culots cellulaires ont été remis en suspension et 2,5 mL de DMEM complet ont été ajoutés. La suspension cellulaire a été mélangée soigneusement. Les suspensions cellulaires ont été comptées comme suit : 50 µL de chaque suspension cellulaire ont été dilués dans 1 mL de milieu DMEM. Les deux dilutions ont été comptées dans une lame KOVA. La suspension HEK 293 (3h30) contenait $6,07 \times 10^5$ cellules/mL, soit un nombre total des cellules récupérées de 1.517×10^6 . La suspension HEK 293 (ON) contenait 2.1×10^5 cellules/mL, soit un nombre total de cellules récupérées de $5,25 \times 10^5$.

Deux plaques de 24 puits ont été utilisées pour chaque suspension cellulaire. Chaque puits a été rempli avec 500 µL de milieu complet DMEM contenant 1 mg/mL de G-418 (également appelée généticine). Puis 100 cellules/puits ont été ajoutées à une plaque et de 500 cellules/puits pour la seconde plaque. Les plaques ainsiensemencées ont été incubées à 37°C. Deux flacons de culture ont été préparés pour cultiver les cellules restantes après les avoir déposées dans les plaques 24 puits. Le flacon N°1 contient des cellules HEK293 transfectées avec le plasmide encodant TROP2-GFP (3h30) et le flacon N°2 contient des cellules HEK293 transfectées avec le plasmide encodant TROP2-GFP (ON). Les cellules ont été cultivées dans du DMEM complet contenant 1mg/ml de G-418. Les cellules des flacons et des plaques de 24 puits ont été examinées tous les deux jours pour vérifier la viabilité des cellules et l'état de confluence de la culture. Lorsque les cellules étaient confluentes, pour les deux types de cultures (plaques 24 puits et flacons), le surnageant a été retiré et les cellules adhérentes ont été traitées avec de la trypsine EDTA pendant environ 10 minutes à température ambiante. Les cellules soigneusement rincées pour dissocier les agrégats ont été mélangées avec 5 mL de leur propre surnageant pour chaque suspension cellulaire. Les cellules ainsi remises en suspension ont été centrifugées à 1200 rpm pendant 10 minutes. Les surnageants ont été éliminés et les cellules remises en suspension. Du DMEM contenant du G-418 à 1 mg/mL a été ajouté à chaque suspension, et une goutte de ce mélange a servi à relancer les cultures en flacon, ou 50 µL pour les cultures sur plaques 24 puits.

Les cellules HEK293 provenant des flacons ont été examinées au microscope pour déterminer combien de cellules expriment la GFP. De nombreuses cellules étaient fluorescentes mais les niveaux de fluorescence étaient très hétérogènes. Ainsi, les deux flacons de

(1) HEK293 TROP2-GFP 3h30 (cultivé dans du DMEM contenant du G-418 à 1 mg/mL)

(2) HEK293 TROP2-GFP OVN (cultivé dans du DMEM contenant du G-418 à 1 mg/mL)

ont été traitées avec 2,5 ml de Trypsine-EDTA pendant environ 10 minutes. Les cellules ont ensuite été recueillies dans des tubes de 15 mL contenant du milieu DMEM complet. Les cellules ont été centrifugées à 1200 rpm pendant 10 minutes. Ensuite, les surnageants ont été éliminés et les cellules

ont été remises en suspension dans le milieu de congélation (4 mL). Ainsi, les deux types de cellules HEK293 TROP2-GFP (transfectées 3h30 ou ON) ont été réparties dans des tubes de congélation (4 tubes pour chaque suspension cellulaire). Les tubes ont été immédiatement congelés à -80°C.

Deux cultures en flacons (un pour chaque type de transfection) ont été relancées avec une goutte de suspension cellulaire trypsine-EDTA. Dans chaque flacon, 5 mL de milieu DMEM complet contenant du G-418 à 1mg/ml ont été ajoutés et les flacons ont été incubés à 37°C :

(1) HEK293 TROP2-GFP 3h30 (cultivé dans du DMEM contenant du G-418 à 1 mg/mL)

(2) HEK293 TROP2-GFP OVN (cultivé dans du DMEM contenant du G-418 à 1 mg/mL)

Des plaques de 24 puits contenant 100 cellules HEK293 TROP2GFP 3h30 par puits (cultivées dans du DMEM contenant du G-418 à 1 mg/mL) ont été examinées au microscope pour identifier le nombre de clones qui se développaient dans chaque puits. Il est intéressant de noter que de nombreux puits contenaient plusieurs clones. Cependant, quelques puits ne contenaient que des cellules mortes. Tous les puits contenant des cellules vivantes ont été marqués et certains de ces puits ont été examinés au microscope pour détecter la GFP. De toute évidence, plusieurs groupes de cellules étaient clairement fluorescents mais la fluorescence était hétérogène.

A ce stade, l'observation des puits au microscope a permis d'identifier des clones plus ou moins bons. Les critères d'évaluation étaient le nombre de cellules non fluorescentes au sein d'un même clone, et l'intensité de la fluorescence. Les cultures contenant les clones les plus intéressants (minimum de cellules non fluorescentes et forte intensité de fluorescence) ont été maintenues, puis des échantillons de ces cultures ont été congelés comme décrit précédemment. Les expériences de FACS ont été réalisées avec tous ces clones dans un premier temps, ce qui a permis d'identifier un clone « D1 » qui possédait la meilleure intensité de fluorescence. Les expériences de FACS ont ensuite été réalisées uniquement sur ce clone.

Fluorescent Activated Cell Sorting (FACS):

Les cellules HEK293 non transfectées et les cellules HEK293 transfectées avec TROP2-GFP (lignée oligoclonale D1) ont été traitées environ 10-15 minutes avec 2 mL de trypsine-EDTA à 37°C, puis mélangées à 5 mL de milieu DMEM (surnageant de chaque flacon). Les suspensions ont été centrifugées à 1200 rpm à 10°C pendant 10 minutes. Après centrifugation, les surnageants ont été jetés et les cellules culottées ont été remises en suspension dans 5 mL de milieu DMEM pour les cellules HEK293 non transfectées ou 15 mL de volume final pour la lignée D1. Les cellules ont été comptées et un volume de chaque suspension a été ajouté pour obtenir 5×10^5 cellules par tube. Tous les tubes ont

été centrifugés à 1200 rpm pendant 10 minutes. Les surnageants ont été jetés et le culot cellulaire remis en suspension après élimination de la dernière goutte.

Les cellules ont alors été mises au contact de différentes concentrations des réactifs suivants : un anticorps anti-TROP 2, la mono-P5H B1 biotinylée, l' α Rep A1 biotinylée, un anticorps non spécifique, une α Rep non spécifique, des tétramères fluorescents de B1 ou de A1, des tétramères fluorescent d' α Rep non spécifique. Après un temps d'incubation allant de 30 mins à 1 h selon les expériences, les cellules ont été lavées par ajout dans chaque tube de 700 μ L de DMEM complet, puis centrifugées 10 mins à 1200 rpm. Les surnageants ont été éliminés jusqu'à la dernière goutte, et la présence des interacteurs a été révélée par ajout dans les tubes d'un réactif fluorescent secondaire : la streptavidine-AlexaFluor680 pour révéler les interacteurs biotinylés ou un anticorps anti-IgG de souris couplé à un AlexaFluor580 pour révéler les anticorps. Après un temps d'incubation de 45 mins à 1h selon les expériences, les cellules ont été lavées comme décrit précédemment puis remises en suspension 400 μ L de PBS + 2% de sérum de veau fœtal avant d'être analysées sur le CytoFlex de la plateforme d'imagerie de l'I2BC à Gif-sur-Yvette (ImageGif).