



HAL
open science

Conception de protéines artificielles multidomaines

Corentin Léger

► **To cite this version:**

Corentin Léger. Conception de protéines artificielles multidomaines. Biologie structurale [q-bio.BM]. Université Paris Saclay (COMUE), 2018. Français. NNT : 2018SACLS384 . tel-03081249

HAL Id: tel-03081249

<https://theses.hal.science/tel-03081249>

Submitted on 18 Dec 2020

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 de Protéines Artificielles Multidomaines

Thèse de Doctorat
De
l'Université Paris-Saclay
préparée à
l'Université Paris-Sud

École doctorale n°569
Innovation thérapeutique : Du fondamental à l'appliqué (ITFA)
Spécialité de doctorat : Biochimie et Biologie Structurale

Par

M. Corentin Léger

Thèse présentée et soutenue à Orsay, le 12 Novembre 2018 :

Composition du Jury :

Herman Van Tilbeurgh Professeur, Université Paris-Sud (UMR9198)	Président
Nicolas Leulliot Professeur, Faculté de médecine Paris Descartes (UMR8015)	Rapporteur
Alexandre Chenal Chargé de Recherche, Institut Pasteur (UMR3528)	Rapporteur
Marie Erard Professeur, Université Paris-Sud (UMR8000)	Examineur
Jérôme Bonnet Chargé de recherche, INSERM (UMR1054)	Examineur
Philippe Minard Professeur, Université Paris-Sud (UMR9198)	Directeur de thèse

*« En essayant continuellement on finit par réussir.
Donc : plus ça rate, plus on a de chances que ça marche. »*

- Devise Shadok -

Voici donc venu le temps des remerciements ...

Cette page, bien qu'étant synonyme de « délivrance » car écrite en dernier, n'est pas pour autant la plus facile à rédiger. En effet, elle sera très certainement la plus lue de ce manuscrit. Il s'agit ici de n'oublier personne... Ces remerciements seront constitués de trois paragraphes : la famille, le labo et les amis (assez classique me direz-vous).

Tout d'abord, un très grand merci à mes Parents qui ont toujours été derrière moi et ont toujours approuvés mes choix dans les études, comme dans la vie de manière générale. Vous avez su me transmettre des valeurs dont je suis fier. Un grand merci à mes Frères qui malgré quelques conflits enfantins (pas facile d'être le « p'tit dernier ») ont toujours été là quand j'avais besoin d'eux. Merci également à mes deux Belles-sœurs, vous êtes de très bonnes pièces rapp... valeurs ajoutées à la famille Léger. Merci également à Marie-Hélène et Régis pour la correction du résumé. Allant très probablement être la seconde chose la plus lue de ce manuscrit, j'avais besoin d'experts ! Bien évidemment, merci à Fantine qui m'a supporté pendant ces trois ans de thèse (et encore plus durant la dernière année), j'espère que tu arriveras à me supporter encore très longtemps :).

Un grand merci à toute l'équipe MIP qui a fait de ces trois ans de thèse un vrai bonheur. Merci à Philippe (le chef), les discussions avec toi sont toujours un grand moment (scientifique et/ou « historique »). Merci à Agathe et Marielle, vous n'avez jamais perdu patience malgré le grand nombre d'irruptions dans vos bureaux avec mes milliers de questions. Un merci tout particulier à maman Agnès qui en plus de répondre (elle aussi !) à mes questions (pas forcément scientifiques), s'est occupée de moi quand mon pied a dit bonjour à un tesson de verre dans un ruisseau... Merci également à Augustin, tu te fais discret dans le bureau mais ta bonne humeur est toujours un plus. Merci aux anciens, Margarida (dont le rire manque), Thibault (qui m'a initié à la cuisine fine (Bud's Deli), Sébastien (le Post-doc qui aime les licornes et l'alcool (deux qualités importantes)) et Hajar (enlève ton casque !). Merci au p'tit nouveau, Adrien, bon courage à toi pour ta thèse :). Merci aux stagiaires, Stagelotte, Alice et Robin, vous n'êtes pas restés longtemps, mais c'était un plaisir de vous avoir parmi nous ! Merci à Mumtaz qui a mis un peu de piments dans ma vie (surtout dans mes repas). Merci à Pierrette qui trouve toujours des bonnes choses à nous ramener ! Merci aux PIM's (remarquez l'originalité dans le choix des lettres ;)). Merci à Magali, autant d'énergie dans un si petit corps, c'est impressionnant ! Merci à Michel pour ton expertise sur les interactions prot/prot (et pour le vidéo projecteur). Merci également à Christophe, j'aime bien te faire dépenser de l'argent ;) . Merci aussi aux chimistes ! Kalani, qui excelle dans la cuisine comme dans la chimie (presque le même domaine :p) et le dessin. Merci à Rémy qui me fait découvrir des films à presque chacune de ses intrusions dans notre bureau ! Un grand Merci à l'équipe FAAM avec laquelle nous collaborons régulièrement (et parfois plus), Herman, Inès, Sophie, Sylvie, Stéphanie, Sophia, Louisa, Dom' et Stéphane (merci à vous deux de m'avoir porté quand mon pied ne me le permettait plus ...), Bruno, Noureddine et Jordan ! C'est toujours un plaisir de passer dans vos (nombreux) bureaux ! Merci à l'équipe NBP et en particulier à Niko, Akram, Carlos et Shashi qui m'ont guidé lors de mes expériences de FRET. Enfin, merci à l'équipe enseignante en Bio et celle des TPs qui ont fait de ma mission d'enseignement un vrai plaisir !

Nous voilà aux amis, un peu une seconde famille qu'on peut choisir :) Je vous préviens ça vient comme ça vient pas d'ordre particulier allez pas me faire de crise de jalousie ;) Merci à ceux qui ne sont pas là depuis longtemps mais qui occupent une place importante aujourd'hui dans ma vie ! Merci à Romain (mon amour), MaGnuS (IronSkull), Agathe et mon Foie 🍷 ! Merci à notre MJ du tonnerre Léo ! Merci à Matthieu et Majo qui sont là depuis un bout de temps maintenant ! Merci à Mathieu (Gourgandine), Pauline (Dame PauPau), Julien (P(o)uterelle), Eugénie (Blanche Neige), Florian (Heureux), Antoine (Timide), Lyam (Atchoum), bon courage à ceux d'entre vous qui vont soutenir après moi ! Merci aussi à ceux de Licence pas encore cités (Why so serious ?) Marie, Manon, Clément (le cousin), Nico et Patrick ! Merci aux très anciens, Mylène, Thibault, Florian, Chloé et Ludivine, même si on ne se voit/parle pas assez souvent, c'est toujours un grand plaisir d'avoir de vos nouvelles :) Merci aussi à la bande de Nouilles : Charlotte, Jade, Héloïse, Sandrine, Béno, Gauthier, Arnaud, Emilien, Arthur sans oublier Boomy et Rocky ! Merci à ceux d'Alt+3, on ne se voit pas souvent mais par contre, qu'est-ce qu'on s'entend bien : Claire, Sarah, Émilie, Guigui, Thierry, Garyk, Axel, Joël, Laurent, Hugo, Laurent, et tous ceux que j'oublie :|. Un grand merci également à Adélia, Juliette, Marine, Caro, Sarah, Damien, Matthieu (KTA), Mojo, Nico et j'en oublie ... Merci également à trois enseignants qui ont su transmettre leur passion avec une pédagogie et une humanité parfois rare, Benoît Momboisse (Histoire-Géo), Frederic Moraldo (EPS) et Benoît Merlant (SVT). Enfin, merci à Koreus, sans lui, ces remerciements auraient été bien plus courts !

Un grand merci à vous tous !

Vous avez tous joué un rôle important dans la réalisation de ce manuscrit !

Je vous souhaite à présent une bonne lecture pour les plus courageux d'entre vous !

Tables des matières générales

ABRÉVIATIONS	10
PARTIE A - INTRODUCTION	13
I. LE MONDE DES PROTÉINES	16
II. INGÉNIERIE PROTÉIQUE	24
III. OPTIMISATION ET CRÉATION DE RECONNAISSANCE.....	30
IV. CRÉATION D'OUTILS MOLÉCULAIRES COMPLEXES.....	42
V. CONTEXTE SCIENTIFIQUE	68
VI. BIBLIOGRAPHIE	79
PARTIE B : RÉSULTATS ET DISCUSSIONS	89
I. ÉTUDE DU COMPORTEMENT DE DEUX BIDOMAINES AREPS	92
II. OPTIMISATION DES BIOSENSEURS GÉNÉRIQUES.....	127
III. VERS DES BIOSENSEURS GÉNÉTIQUEMENT CODABLES.....	154
IV. BIBLIOGRAPHIE	176
PARTIE C : MATÉRIELS & MÉTHODES	183
I. BIOLOGIE MOLÉCULAIRE	186
II. MICROBIOLOGIE	192
III. PURIFICATION DES PROTÉINES	194
IV. BIOPHYSIQUE	196
V. EXPÉRIENCE DE FRET.....	202
VI. SÉLECTION PAR PHAGE DISPLAY.....	205
VII. BIBLIOGRAPHIE	209
PARTIE D : CONCLUSION ET PERSPECTIVES	213
I. COMPORTEMENT DES PROTÉINES BIDOMAINES	216
II. BIOSENSEURS	223
III. BIBLIOGRAPHIE	230

Abréviations

Ab	Anticorps
BFP	<i>Blue Fluorescent Protein</i>
BiTE	<i>Bispecific T-cell engager</i>
BLAST	<i>Basic Local Alignment Search Tool</i>
BRET	<i>Bioluminescence resonance Energy Transfer</i>
CaM	<i>Calcium-Modulated protein</i> = Calmoduline
CAR T-cell	<i>Chimeric Antigen Receptor T-cell</i>
CD	<i>Cluster of Differentiation</i>
CDR	<i>Complementarity determining regions</i>
CPEC	<i>Circular Polymerase Extension Cloning</i>
CV	Volume de colonne
DARPin	<i>Designed Ankyrin Repeat Protein</i>
DART	<i>Dual-affinity Re-targeting Antibody</i>
DBD	<i>DNA Binding Domain</i>
DHFR	DiHydroxy Folate Réductase
DSC	<i>Differential scanning calorimetry</i>
ERK	<i>Extracellular signal-regulated kinases</i>
Fab	<i>Fragment Antigen Binding</i>
FP	Protéine Fluorescente
FRET	<i>Förster Resonance Energy Transfer</i>
GFP	<i>Green Fluorescent Protein</i>
GST	Glutathion S-transférase
HcAb	<i>Heavy-chain antibody</i>
IL	Interleukine
ITAM	<i>Immunoreceptor Tyrosine-Based Activation</i>
ITC	<i>Isothermal Titration Calorimetry</i>
JNK	c-Jun N-terminal kinases
KD	Constante de dissociation
LBD	<i>Ligand Binding Domain</i>
LBT	<i>Lanthanide Binding Tag</i>
LOV	<i>Light-oxygen-Voltage</i>
LovTAP	<i>LOV- and Tryptophan-Activated Protein</i>

Abréviations

LRR	<i>Leucine Rich Repeat</i>
LUCID	<i>Luciferase-based indicators of drugs</i>
MAP	<i>Mitogen-Activated Protein</i>
MBP	<i>Maltose Binding Protein</i>
Nb	<i>Nanobodies</i>
NFAT	<i>Nuclear Factor of the Activated T-cell</i>
pbdb	<i>Protein Data Bank</i>
PCA	<i>Protein Complementation Assay</i>
PCR	<i>Polymerase Chain Reaction</i>
QD	<i>Quantum Dot</i>
RCA	<i>Rolling Cycle Amplification</i>
RCPG	Récepteurs couplés aux protéines G
RFP	<i>Red Fluorescent protein</i>
RPM	Rotation par minute
SAXS	<i>Small Angles X-ray Scattering</i>
scFv	<i>Single-chain Variable Fragment</i>
SDS-PAGE	<i>Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis</i>
SEC-MALS	<i>Size Exclusion Chromatography coupled with Multi-Angle Light Scattering</i>
SH3	Src Homology 3
SPR	<i>Surface Plasmon Resonance</i>
TAD	<i>Transcription Activated Domain</i>
Tb	Terbium
TPSA	<i>Total Prostate Specific Antigen</i>
Trp	Tryptophane
TrpR	<i>Tryptophan Repressor</i>
TRUCK	<i>T-cell Redirected for Universal Cytokine-mediated Killing</i>
TRX	Thiorédoxine
UCA	Ultra-Centrifugation Analytique
VLR	<i>Variable Lymphocyte Receptor</i>
YFP	<i>Yellow Florescent Protein</i>

Partie A - Introduction

Tables des matières

I.	LE MONDE DES PROTÉINES	16
A.	<i>L'évolution du répertoire protéique</i>	16
B.	<i>Une histoire d'interaction</i>	21
C.	<i>Les linkers : Des entremetteurs protéiques</i>	22
D.	<i>Conclusion</i>	23
II.	INGÉNIERIE PROTÉIQUE	24
A.	<i>Évolution dirigée</i>	24
B.	<i>Conception rationnelle</i>	27
C.	<i>Design semi-rationnel</i>	28
III.	OPTIMISATION ET CRÉATION DE RECONNAISSANCE	30
A.	<i>Les anticorps et leurs dérivés</i>	30
1)	Les anticorps "classiques"	30
2)	Intégration des ScFv dans des constructions plus complexes	31
3)	Les anticorps de camélidés	33
4)	Conclusion	33
B.	<i>Un système immunitaire ancestral</i>	34
C.	<i>Les architectures artificielles comme alternative</i>	36
1)	Une variété d'architectures	36
2)	Les protéines à motifs répétés	38
a)	Quelles ossatures ?	38
b)	L'exemple des Darpins	39
3)	Création de reconnaissance <i>de novo</i>	40
IV.	CRÉATION D'OUTILS MOLÉCULAIRES COMPLEXES	42
A.	<i>Chimeric Antigen Receptor (CAR)</i>	42
B.	<i>De la reconnaissance à la détection : Les biosenseurs</i>	46
1)	Biosenseurs basés sur la variation d'un signal lumineux	47
a)	Le FRET : Förster Resonance Energy Transfer	47
b)	L'apport des protéines fluorescentes et bioluminescente	49
c)	Biosenseurs basés sur du FRET	50
d)	Biosenseurs basés sur le BRET	52
e)	Un nouveau couple Donneur / Accepteur pour voir plus loin :	54
2)	Biosenseurs basés sur une résistance aux antibiotiques	55
3)	Basés sur la permutation circulaire de protéine fluorescente	57
4)	Biosenseurs basés sur la complémentation fonctionnelle	59
5)	Biosenseurs basés sur la stabilité des biosenseurs <i>in vivo</i>	60
C.	<i>Contrôle des fonctions par la lumière</i>	62
D.	<i>Des multidomains pour aider à l'expression et à cristallisation</i>	65
1)	Expression	65
2)	Cristallisation	65
E.	<i>Création d'avidité par clamp</i>	67

V.	CONTEXTE SCIENTIFIQUE	68
A.	<i>Les αReps</i>	68
1)	À l'origine des α Reps : les protéines HEAT	68
2)	Design de la séquence consensus	69
3)	Construction de la bibliothèque AReps	72
a)	Bibliothèque 1.0	72
b)	Bibliothèques 2.0 et 2.1	74
4)	Reconnaissance <i>in vivo</i>	75
5)	Sélection d' α Reps contre des cibles définies	76
B.	<i>Les αReps comme briques élémentaires</i>	77
VI.	BIBLIOGRAPHIE	79

I. Le monde des protéines

A. L'évolution du répertoire protéique

Les protéines sont les composants moléculaires à l'origine de nombreux processus (catalyse, transports, moteurs moléculaires, gestion de l'information génétique, etc.) très caractéristiques des organismes vivants. Ces fonctions, aussi diverses que sophistiquées résultent de l'émergence de chaînes polypeptidiques ayant la capacité de se replier dans l'espace. C'est en effet l'existence d'une structure tertiaire qui confère aux protéines leurs capacités de reconnaissances moléculaires, invariablement nécessaires à leurs fonctions.

Les processus qui ont conduit à l'émergence des structures tertiaires des protéines primordiales restent mal compris. Il semble en effet que les séquences polypeptidiques aléatoires adoptant spontanément une structure tertiaire d'une stabilité comparable aux protéines naturelles soient rares. Il a cependant été très tôt suggéré que toute nouvelle fonction biologique ne nécessitait pas nécessairement l'émergence d'une protéine entièrement nouvelle mais pouvait provenir de la réutilisation d'innovations structurales préexistantes. Ce processus de « bricolage » évolutif (Jacob, F., 1977) repose sur l'idée qu'il est sans doute évolutivement plus efficace, ou plus rapide, de combiner et de modifier des domaines structuraux d'ores et déjà existants pour obtenir des propriétés nouvelles plutôt que d'en recréer de nouveaux. Il apparaît aujourd'hui très clairement de l'étude des données de séquences et de structures que les processus de réutilisation de domaines existants et l'apparition de nouvelles combinaisons de domaines ont effectivement joué un rôle déterminant dans l'évolution biologique.

Un des moteurs les plus importants dans l'évolution des génomes et donc des protéines est la duplication génétique. Cette duplication peut toucher une séquence nucléotidique, un gène voire même le génome complet (paléoploïdie) et cela à plusieurs reprises. (Kellis, M., *et al.*, 2004; Magadum, S., *et al.*, 2013).

Or lorsqu'un gène est dupliqué, il peut par exemple être inséré en fusion à un autre gène. Chaque gène codant pour une protéine au repliement indépendant, la fusion de deux gènes pourra donner une protéine à plusieurs domaines. Les domaines constituent alors les unités structurales et fonctionnelles de bases des protéines. Ce phénomène peut expliquer pourquoi une grande proportion de protéines multidomaines existe chez tous les règnes du vivant. En effet, au moins un tiers des protéines bactériennes et plus des deux tiers des protéines eucaryotes comportent deux domaines ou plus. (Ekman, D., *et al.*, 2005)

De plus, lorsqu'il existe plusieurs copies indépendantes d'un même gène, la pression de sélection évolutive ne s'applique alors que sur l'un des gènes présent dans le génome. Le ou les gènes supplémentaires ont alors plusieurs desseins possibles.

Il peut rester tel quel et permettre une surexpression de la protéine pour laquelle il code. Par exemple, chez l'humain, le gène CCL3L1 (MIP-1 α P), codant pour la chimiokine du même nom, se situe dans une zone fortement dupliquée de notre génome et le nombre de copies de ce gène varie d'un individu à l'autre. Or un nombre de copies du gène CCL3L1 inférieur à la moyenne humaine augmente la susceptibilité au développement d'un SIDA en cas d'infection par le VIH. (Gonzalez, E., *et al.*, 2005)

Dans le cas de la duplication d'un gène codant originellement pour une protéine à deux fonctions, il peut y avoir une subfonctionnalisation. Dans ce cas il n'y a pour ainsi dire pas de création de fonction mais une optimisation d'une seule fonction de la protéine codée par le gène. Par exemple dans le cas des enzymes, des études récentes montrent que la majorité pouvait catalyser originellement plusieurs réactions. Avec le temps, les enzymes se sont spécialisées pour devenir de plus en plus spécifiques et de mieux en mieux contrôlables. Il est intéressant de noter que cette hypothèse est supportée par le fait que de nombreuses enzymes possèdent encore aujourd'hui des activités résiduelles. C'est ce que l'on appelle la plasticité enzymatique (*Enzyme promiscuity*). (Khersonsky, O. and Tawfik, D.S., 2010; Copley, S.D., 2017; Newton, M.S., *et al.*, 2018)

Le gène supplémentaire peut également acquérir des mutations au fil du temps et cela sans impacter nécessairement le fitness de l'organisme puisque le gène ancestral est toujours présent. Ces mutations peuvent alors s'accumuler et parfois entraîner l'apparition de nouvelles fonctions, c'est ce que l'on appelle « l'évolution divergente » (néofonctionnalisation).

Ainsi, à partir d'un répertoire protéique de base il est possible d'obtenir une quantité bien plus importante de protéines aux fonctions variées et complexes. (Figure 1)

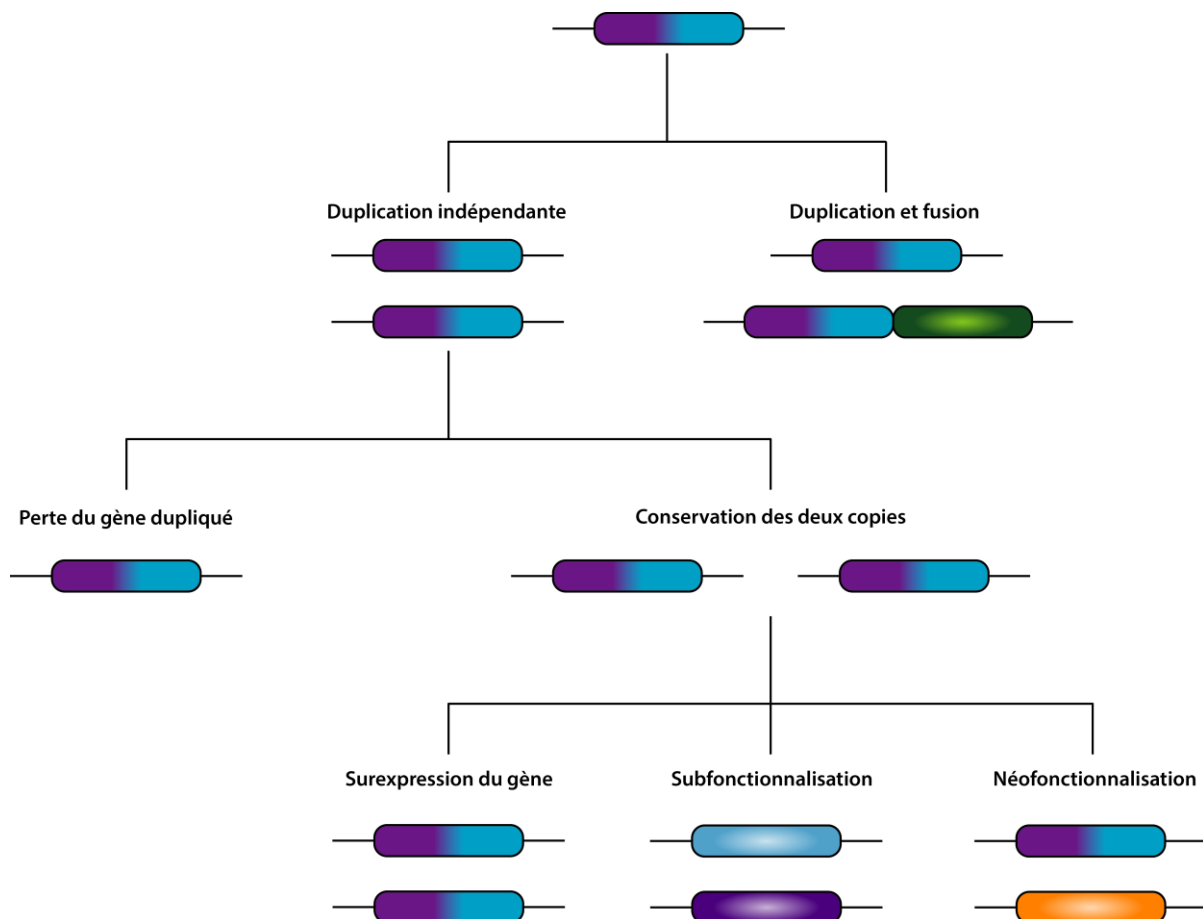


Figure 1 : Dessin d'un gène dupliqué. Les différentes fonctions portées par le gène sont schématisées par les couleurs. Dans le cas d'une subfonctionnalisation, les gènes dupliqués portent chacun une des fonctions initiales du gène originel. Dans le cas de la néofonctionnalisation, le gène acquiert suffisamment de mutations pour changer de fonction.

Ces différents processus d'évolution auraient ainsi permis la création de fonctions complexes; régulation allostérique, régulation intramoléculaire *via* un domaine régulateur, modification de site catalytique et cela à partir d'un répertoire protéique simple. (Bhattacharyya, R.P., *et al.*, 2006)

La famille des récepteurs couplés aux protéines G (RCPG) est une preuve de cette évolution basée sur la réutilisation de domaines. En effet, cette famille compte plus de 800 protéines différentes chez l'homme et est notamment caractérisée par ses sept passages transmembranaires. Les RCPGs peuvent répondre à des stimuli très variés, photons, ions, purines, amines, neurotransmetteurs, chimiokines, hormones, lipides, ... Malgré cette variété de stimuli, la structure est globalement toujours identique. De plus, dans tous les cas, le stimulus entraîne un changement de conformation du RCPG qui transmet alors le signal. (Zhang, D., *et al.*, 2015) Cette réutilisation d'un repliement sans conserver la spécificité n'est pas uniquement observée chez les RCPGs. En effet, pour un même repliement la fonction d'une enzyme peut varier au cours de l'évolution. Par exemple, le *Rossmann Fold* et le *TIM Barrel* sont trouvés dans des enzymes des six différentes classes de la nomenclature enzymatique. (Baier, F., *et al.*, 2016) De plus, l'étude des arbres phylogénétique de 379 superfamilles de domaines a révélé qu'au cours de l'évolution bien que rare, le changement de classe enzymatique est possible. (Furnham, N., *et al.*, 2016)

Concernant la régulation allostérique, on peut par exemple citer la famille des MAP kinase (*Mitogen-activated protein kinases*) au sein de laquelle on trouve les groupes ERK, JNK et p38. Ces trois groupes sont constitués de protéines possédant les mêmes domaines. Pourtant, dans le cas des protéines p38 et ERK, la fixation d'un peptide stimule l'activité catalytique ou l'autophosphorylation. Dans le cas des protéines JNK, la fixation du peptide sur le même site inhibe l'activité catalytique. (Chang, C.-I., *et al.*, 2002; Laughlin, J.D., *et al.*, 2012)

L'un des types de domaines le plus impliqué dans la régulation intramoléculaire est le domaine SH3 (*Src Homology 3*) qui se trouve dans plusieurs centaines de protéines au sein des différents règnes du vivant. Bien qu'ayant des surfaces d'interaction variables, la grande majorité des domaines SH3 interagissent avec des séquences du type PxxP. Cependant afin d'assurer une spécificité de régulation sur les protéines auxquelles ils sont reliés, certains domaines SH3 ont évolué pour reconnaître des séquences non canoniques. Les domaines SH3 des protéines Eps8 et Gads par exemple, reconnaissent respectivement des séquences PxxDY et RxxK. (Zarrinpar, A., *et al.*, 2003)

Parmi les protéines provenant de fusion de gènes, on peut citer la topoisomérase II de levure qui est présente sous la forme de deux protéines, la Gyrase A et la Gyrase B, chez *E. Coli*. Chez l'homme, la Pyrroline-5-carboxylate synthase possédant deux domaines et réalisant les deux premières étapes de la synthèse des prolines est remplacée par deux enzymes indépendantes chez *E. Coli*. (Marcotte, E.M., *et al.*, 1999)

L'ensemble de ces études montre bien qu'au cours de l'évolution, la réutilisation des domaines pour créer de la diversité est un mécanisme général. Cette augmentation de diversité passe aussi par la fusion de domaines dans un but de régulation plus efficace ou d'optimisation sans créer à proprement parler de nouvelles fonctions. L'ensemble de ces processus permet ainsi de mieux comprendre la présence importante des protéines multidomaines et la réutilisation de ces domaines dans tous les règnes du vivant.

La compréhension des mécanismes d'interaction des protéines possédant plusieurs domaines ou non présente un intérêt majeur pour comprendre la complexité du monde vivant. Cette compréhension passe notamment par l'étude des interactions entre protéines et / ou domaines dans leur ensemble et leur prédiction. Dans le cas des interactions domaine-domaine, l'étude des liens ou *linker* les reliant, permet également de mieux appréhender l'organisation générale de telles architectures.

B. Une histoire d'interaction

La majorité des processus biologiques font intervenir des interactions protéine-protéine ou domaine-domaine. Ces interactions entraînent la formation de complexes protéiques. Ces complexes vont jouer différents rôles en fonction de leur composition, de leur affinité mais également de la durée pendant laquelle le complexe reste formé. Les interactions protéine-protéine sont impliquées dans le transport, la régulation de voies métaboliques ou enzymatiques et la signalisation... L'étude et l'identification de ces interactions permettent donc une meilleure compréhension du monde du vivant. (Nooren, I.M. and Thornton, J.M., 2003) Pourquoi un domaine X va interagir avec un domaine Y plutôt qu'avec un domaine Z ? Comment prédire les interactions que peut effectuer un domaine avec une protéine ?

Plusieurs études ont visé à répondre à ces questions et notamment à la prédiction des interactions protéine-protéine et domaine-domaine. Ces études reposent sur l'analyse de bases de données (InterPro, PFAM, SCOP) regroupant les interactions déjà caractérisées entre protéines et domaines. Le principe étant d'identifier des séquences d'acides aminés spécifiques interagissant ensemble au sein de ces bases de données, par la suite divers modèles sont mis en place pour extrapoler à des interactions non caractérisées. (Nye, T.M., et al., 2005; Riley, R., et al., 2005; Doerks, T., et al., 2002; Sprinzak, E. and Margalit, H., 2001; Stelzl, U., et al., 2005) L'équipe de J. M. Carazo a développé ce même genre d'approche en prenant en compte la dimension structurale des interactions, en se basant sur les données de la *protein data bank* (*pdb*). Ils ont ainsi pu améliorer la précision des prédictions tout en augmentant leur nombre. (Segura, J., et al., 2015)

En cas d'une protéine non caractérisée, l'utilisation de ces prédictions peut également aider à comprendre son fonctionnement. Par exemple, si une protéine est prédite pour interagir avec un domaine de liaison à l'ADN, il peut être intéressant d'étudier expérimentalement les effets de cette protéine sur l'ADN.

C. Les *linkers* : Des entremetteurs protéiques

L'étude des *linkers* permet également de mieux comprendre ces interactions et les processus qui en découlent. En effet, les *linkers* ne présentent pas tous la même taille, la même structure ou la même composition. Or, ces paramètres peuvent avoir des conséquences fondamentales dans les protéines multidomaines. En effet, un *linker*, s'il est flexible, pourra permettre des configurations qui ne seront pas accessibles si le *linker* adopte une structure plus rigide et donc par exemple empêcher des interactions intramoléculaires. De la même manière un lien trop court entre deux domaines pourra empêcher leur interaction, la taille est donc également importante.

L'équipe d'Heringa J. a par exemple réalisé une classification détaillée de la composition de 1280 *linkers* (12 776 résidus) en fonction de leurs structures. Cette étude montre qu'au sein des *linkers*, il existe pour chaque type de structure une fréquence d'utilisation des acides aminés spécifique et différente de celle des domaines protéiques. Ainsi, on trouve de manière générale plus de Proline, Arginine, Phénylalanine, Thréonine, Glutamate et Glutamine (par ordre décroissant en termes de préférence) dans les *linkers* que dans les protéines repliées. La présence de proline en tant que résidu le plus présent est probablement due au fait qu'elle ne peut créer de liaison hydrogène et permet alors une isolation structurale entre *linker* et domaines. La composition du *linker* en fonction de la structure varie également, il y a ainsi plus de proline dans les *linkers* non structurés que dans les *linkers* hélicoïdaux. Cela peut s'expliquer par le fait que les prolines vont donner une certaine rigidité au linker, celle-ci peut être nécessaire au repliement des protéines. Fait intéressant, les auteurs ont remarqué que 10% des *linkers* hélicoïdaux étudiés possédaient une proline centrale (contre seulement 2% dans les hélices en général). La présence de cette proline perturbe le réseau de liaisons hydrogènes de l'hélice et peut permettre en fonction des conditions une torsion de l'hélice à la manière d'une charnière. (George, R.A. and Heringa, J., 2003) Les auteurs n'ont cependant pas regardé si les protéines possédant ce genre de *linker* présentaient des mécanismes impliquant de tels changements structuraux. Cela aurait pu expliquer la conservation de cette proline centrale au cours de l'évolution.

D. Conclusion

L'ensemble des études sur l'évolution a permis d'obtenir une meilleure compréhension du monde des protéines. Compréhension sur le fonctionnement de l'évolution et les moyens par lesquelles, de nouvelles fonctions ont pu être obtenues. La réutilisation de domaines aurait ainsi permis une évolution rapide et expliquerait pourquoi certains sont retrouvés dans plusieurs protéines aux fonctions diverses. De plus, les prédictions d'interfaces et l'étude des *linkers* permettent notamment de mieux anticiper et comprendre le comportement des protéines à l'aide de leur séquence.

L'ensemble de ces informations permet d'une part, de comprendre la complexité du monde vivant et d'autre part représente une bonne base de départ pour créer et concevoir des outils moléculaires plus ou moins complexes par ingénierie protéique.

II. Ingénierie protéique

La création de nouvelles fonctions basées sur l'utilisation des protéines représente un enjeu majeur, que cela soit dans le domaine pharmaceutique ou dans le domaine industriel. En effet, d'une part, l'obtention de nouveaux médicaments innovants est toujours nécessaire pour de nombreuses maladies. D'autre part, à l'heure de la chimie verte, les protéines semblent être une excellente alternative pour catalyser des réactions nécessitant jusqu'alors l'utilisation de produits chimiques polluants.

Deux approches bien distinctes sont envisageables pour la conception d'outils moléculaires : L'évolution dirigée et la conception rationnelle.

A. Évolution dirigée

L'évolution dirigée est une approche visant à mimer l'évolution darwinienne à l'échelle de la paillasse. C'est-à-dire qu'une population de départ bien définie va être amenée à évoluer dans un milieu spécifique pour obtenir les propriétés désirées. Cette méthode se base sur la création de diversité génétique *via* des bibliothèques de variants puis dans la majorité des cas, sur une réduction de celle-ci par l'application d'une pression de sélection, et enfin d'un criblage phénotypique des variants présentant les caractéristiques recherchées. Cette pression de sélection n'est cependant pas toujours nécessaire, par exemple, lorsque l'on souhaite améliorer une fonction enzymatique, seule l'étape de criblage est nécessaire après création de diversité. L'évolution dirigée est une méthode permettant l'obtention rapide d'un petit nombre de protéines aux propriétés choisies. (Figure 2-A)

La création de diversité génétique peut se faire par différentes méthodes : mutagénèse aléatoire, recombinaisons d'ADN, utilisations de souches mutagènes... La taille des bibliothèques de variants à créer dépend de l'objectif désiré. Il est généralement préférable expérimentalement d'avoir le plus grand nombre de séquences possibles. Il est cependant impossible de couvrir l'ensemble de l'espace des séquences d'une protéine typique. En effet, pour un décapeptide, le nombre de séquences théoriques possibles est de 10^{13} , en considérant que chacun des vingt acides aminés naturels peut occuper chacune des positions. Or il s'agit de la limite actuelle des méthodes de génération de diversité et de criblage. (Arnold, F.H., 2018; Cobb, R.E., *et al.*, 2013; Packer, M.S. and Liu, D.R., 2015)

Cela pose la question de savoir si la diversité au sein de l'espace des séquences composant une bibliothèque est suffisante pour sélectionner un variant aux propriétés désirées ?! Avec une banque suffisamment exhaustive, l'expérience montre qu'il est possible de sélectionner une séquence, si ce n'est plus, codant pour un ou des variants ayant les propriétés recherchées. Ainsi, bien que n'étant pas totalement exhaustive une bibliothèque de taille suffisante peut amplement suffire à l'obtention d'une, voire de plusieurs protéines, possédant les propriétés recherchées. Il ne s'agira peut-être pas du variant le plus optimal, ce dernier pouvant être obtenu *a priori* avec une banque totalement exhaustive, mais il s'en rapprochera.

Une fois la diversité génétique obtenue, il est nécessaire d'effectuer une étape où une pression de sélection est appliquée. La pression de sélection va permettre de réduire la diversité génétique. Celle-ci peut être basée sur le fitness de bactérie en exprimant les variants dans des conditions variables, sur la reconnaissance de cible, ... Dans le cas où l'on cherche à sélectionner des protéines interagissant avec une cible donnée, la pression de sélection va se faire en se basant sur l'utilisation de méthodes d'expositions («*display*»), les principales étant le Phage-*display*, le Ribosome-*display* ou le Yeast-*display*. (Li, M., 2000) Ces méthodes permettent l'exposition des protéines à une cible d'intérêt fixée sur un support. Pour cela, elles sont liées à des modules d'exposition (phages, ribosomes, levures...) contenant l'information génétique spécifique de la protéine exposée. Si celle-ci présente une affinité pour la cible, alors, le complexe cible-protéine-module d'exposition ne se détachera pas lors de lavages successifs. Il sera ainsi possible en augmentant la stringence des conditions de sélection (lavages, compétiteurs, temps d'incubation...) de sélectionner des protéines de plus en plus affines.

La plasticité enzymatique peut également être exploitée en évolution dirigée. Par exemple, l'équipe de F.H. Arnold a remarqué que les protéines à hème et notamment le cytochrome P450 possède une activité résiduelle de cyclopropanation des alcènes en présence de diazo carbène et cela dans l'eau. Par évolution dirigée, les auteurs ont réussi à augmenter cette activité résiduelle pour produire du cis-cyclopropane, un précurseur du levomilnacipran, un antidépresseur. (Wang, Z.J., et al., 2014)

B. Conception rationnelle

La conception rationnelle présente un avantage par rapport à l'évolution dirigée : il n'est pas nécessaire de cribler expérimentalement les bibliothèques de variants. Par conséquent une partie très lourde de l'évolution dirigée peut être évitée.

Cette approche se base sur les connaissances acquises et sur le postulat que la séquence en acides aminés va définir la structure et les propriétés des protéines. Ainsi grâce à une approche de *design* computationnel, il est possible de créer des protéines aux fonctions désirées en prédisant le repliement de la séquence d'acides aminés. Le design de protéines nécessite alors l'utilisation de base de données structurale (pdb) et d'interaction (BioGRID). (Figure 2-B) (Eriksen, D.T., *et al.*, 2014; Chen, T.S. and Keating, A.E., 2012; Huang, P.S., *et al.*, 2016)

Les stratégies employées sont de plusieurs types. Il peut s'agir de la "simple" ingénierie rationnelle d'un site actif pour augmenter la spécificité ou la sélectivité. Par exemple, Zha *et al.* ont modifié rationnellement l'acide gras synthase, une enzyme bifonctionnelle, afin d'inhiber sa fonction céto réductase et pour uniquement produire de l'acide tri acétique lactone. (Zha, W., *et al.*, 2004)

La conception rationnelle peut aller jusqu'à la création *de novo* de molécule. C'est ce qu'a fait l'équipe de D. Baker en créant la première molécule possédant un repliement n'étant pas observé dans la nature et présentant une haute stabilité avec un Tm supérieur à 100°C. (Kuhlman, B., *et al.*, 2003) Cette même équipe a également réalisé une enzyme capable de catalyser la réaction de Diels-Alder. Le modèle fut conçu *in silico* en utilisant le programme de modélisation Rosetta. Le site actif a été le premier élément créé puis les auteurs ont conçu une structure stable permettant d'obtenir ce site actif. La caractérisation fonctionnelle et structurale de cette enzyme a montré que le modèle se superposait à la structure expérimentale et que l'enzyme catalysait bien la réaction souhaitée. (Siegel, J.B., *et al.*, 2010)

C. Design semi-rationnel

Avec la modernisation des outils informatiques et l'augmentation de leur puissance, la combinaison des deux approches est devenue une stratégie très courante. Désormais, le design semi-rationnel qui consiste en la combinaison de l'évolution dirigée et du design rationnel est souvent l'alternative choisie pour le développement de nouvelles protéines par ingénierie. Cette stratégie se base sur l'utilisation de données structurales, fonctionnelles, et sur les informations extraites des séquences mais aussi sur des algorithmes de design rationnel. Il devient alors possible d'identifier préalablement à la création de diversité génétique, les zones présentant un intérêt à être mutées. Dans le cas d'un site actif par exemple, il est alors possible de cibler les résidus du site actif pour changer la spécificité d'une enzyme. Cela permet de diminuer la diversité génétique initialement nécessaire en ciblant des positions et donc de réduire la taille des bibliothèques. (Figure 2-C) (Lutz, S., 2010)

L'équipe de D. Baker a ainsi identifié et modifié grâce au design semi-rationnel le site actif de huit enzymes pour les rendre capable d'effectuer une élimination de Kemp. Ils ont dans un premier temps créé les sites actifs *de novo*. Par la suite l'application des méthodes d'évolution dirigée a permis d'augmenter la processivité des enzymes d'un facteur allant jusqu'à 200. Cette étape d'évolution dirigée a montré que les positions clés définies par design rationnel n'étaient pas touchées. Seules les positions adjacentes variaient permettant une optimisation très fine de la séquence. (Rothlisberger, D., et al., 2008)

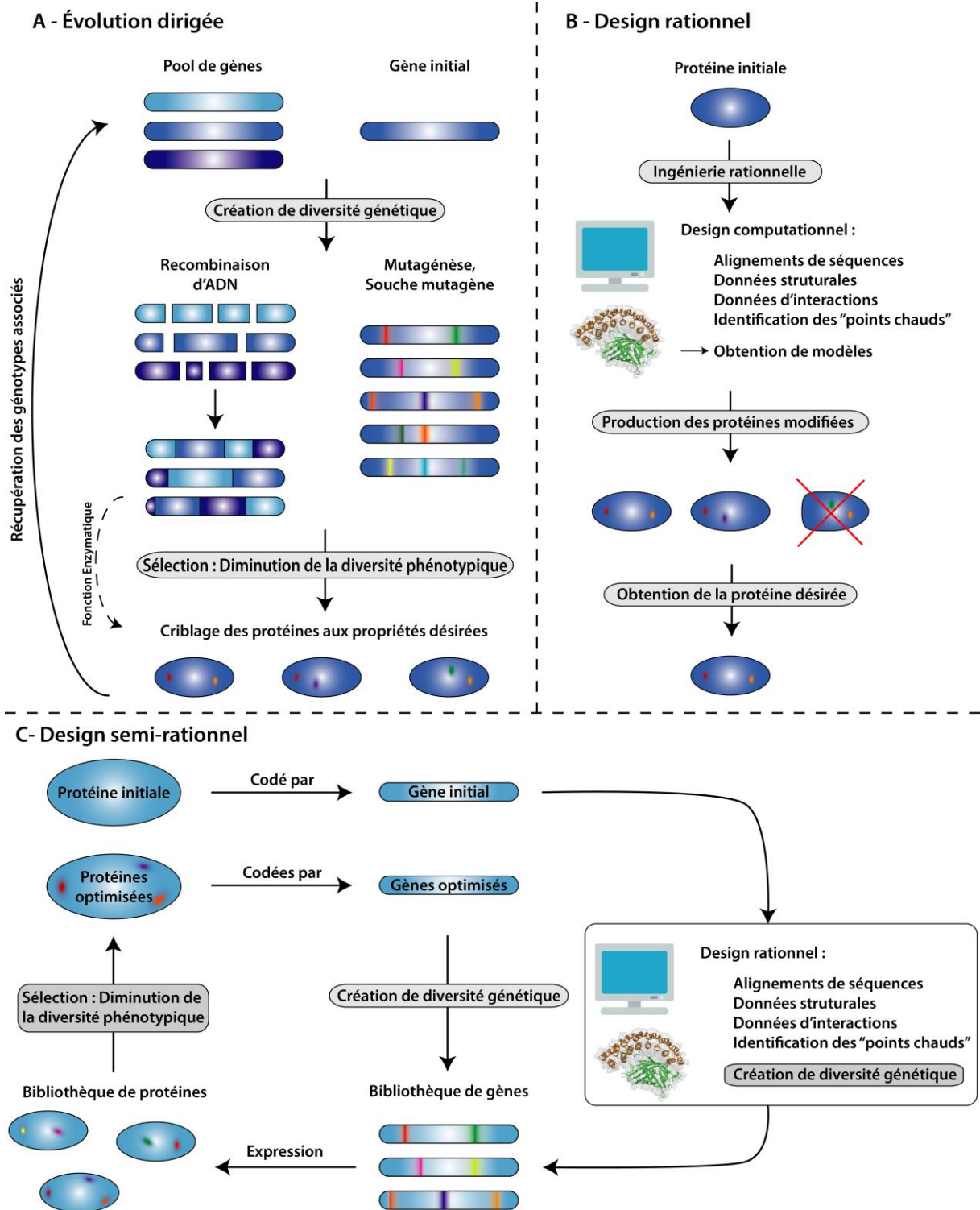


Figure 2 : Différentes approches de l'ingénierie des protéines.

A) Évolution dirigée. À partir d'un gène de départ ou d'un pool de gènes homologues, de la diversité génétique est créée aléatoirement. Une pression de sélection est appliquée puis un criblage des propriétés est effectué. Les génotypes peuvent être récupérés et plusieurs cycles création de diversité, sélection, criblage peuvent être effectués.

B) Design rationnel. Utilisation de l'informatique et de l'ensemble des données (séquence, structure, interaction, point chaud) pour générer des protéines optimisées. S'en suit une caractérisation. Il est possible de partir directement de l'étape de design computationnel : design de novo

C) Design semi-rationnel. À partir d'une protéine initiale, utilisation du design rationnel pour identifier les zones d'intérêt et créer de la diversité génétique (bibliothèque de gènes). S'en suit l'application d'une pression de sélection et un criblage fonctionnel pour obtenir des protéines optimisées qui donneront une bibliothèque de gènes réduites.

III. Optimisation et création de reconnaissance

Les interactions protéine-protéine sont au centre d'un grand nombre de processus du vivant, voies de signalisation, système immunitaire, échafaudage protéique, etc. L'ingénierie des protéines vise depuis longtemps l'amélioration d'affinité d'interaction et la création de reconnaissance spécifique de cibles diverses. Grâce aux méthodes d'ingénierie précédemment abordées, il est possible d'améliorer efficacement des fonctions déjà présente telles que la reconnaissance ou bien encore la stabilité. Il est par ailleurs également possible de créer cette reconnaissance pour des protéines qui n'en étaient pas capable initialement.

A. Les anticorps et leurs dérivés

Les anticorps sont, sans aucun doute, l'exemple type des molécules de reconnaissance. Ces molécules du système immunitaire ont en effet un potentiel quasi illimité de reconnaissance, et cela, à partir d'une même ossature. Les anticorps présentent l'avantage d'être très spécifiques vis-à-vis de leur cible et cela avec des affinités élevées et des constantes de dissociation (K_D) de l'ordre du nanomolaire. Il est donc normal que l'ingénierie des anticorps soit de loin la plus développée.

1) Les anticorps "classiques"

Les approches d'ingénierie ont notamment visé à humaniser leur structure, à créer des sites de reconnaissance et à augmenter leur stabilité. Les anticorps sont donc utilisés aujourd'hui pour de nombreuses applications. Par exemple, ils font de très bons outils biotechnologiques en biologie cellulaire et biochimie, purification de protéines, détection efficace de cible ou bien encore capture de cible pour effectuer du pull-down. Ils sont également d'ores et déjà utilisés dans le cas de traitement de nombreuses maladies. Les études les plus récentes visent à utiliser leur multi-modularité pour créer des anticorps multi-spécifiques, dans le but de développer de nouvelles approches thérapeutiques. (Nunez-Prado, N., *et al.*, 2015)

Les anticorps conventionnels étaient initialement composés de douze domaines. Leur production était peu aisée et coûteuse et permettait l'obtention de sérums polyclonaux. L'optimisation et l'ingénierie des domaines ont finalement permis de réduire la taille des anticorps. Ainsi, en 1988, trois différents types de dérivés d'anticorps furent produits dans *E.Coli* ; les Fab (*Fragment Antigen Binding*), constitués de deux domaines variables et deux domaines constants (Better, M., *et al.*, 1988) ; Les fragments variables indépendamment l'un de l'autre (Skerra, A. and Pluckthun, A., 1988) ; et finalement ces mêmes fragments mais reliés entre eux par un linker flexible "*Single-chain Variable Fragment*", à l'époque appelé sFv, désormais connu sous le nom de scFv. (Huston, J.S., *et al.*, 1988)

Ces ScFv présentant des avantages d'un point de vue de la stabilité, leur exposition sur phage fut démontrée possible en 1990 ouvrant alors la voie à la création de bibliothèques de ScFv. (McCafferty, J., *et al.*, 1990) En 1991, la première étude montrant qu'il était possible de passer au-delà de l'immunisation en combinant bibliothèque de ScFv et exposition sur phage parut et ouvrit alors la porte à de nombreuses possibilités. (Marks, J.D., *et al.*, 1991)

2) Intégration des ScFv dans des constructions plus complexes

Les scFv ont ainsi été utilisés maintes fois en tant que brique élémentaire pour la conception de nouveaux objets de reconnaissance (Figure 3). Des anticorps bivalents et bispécifiques (BiTEs, DARTs, minibody, ...) furent ainsi générés. D'autres constructions furent développées pour mener à des dérivés d'anticorps allant jusqu'à l'hexavalence. Cette augmentation du nombre de sites et la création de bi-spécificité peuvent notamment permettre la double reconnaissance de cible et l'augmentation d'affinité par création d'avidité. L'avidité se définissant comme une augmentation d'affinité apparente lorsqu'une molécule multivalente se lie à une surface *via* plusieurs sites en même temps. (Nunez-Prado, N., *et al.*, 2015)

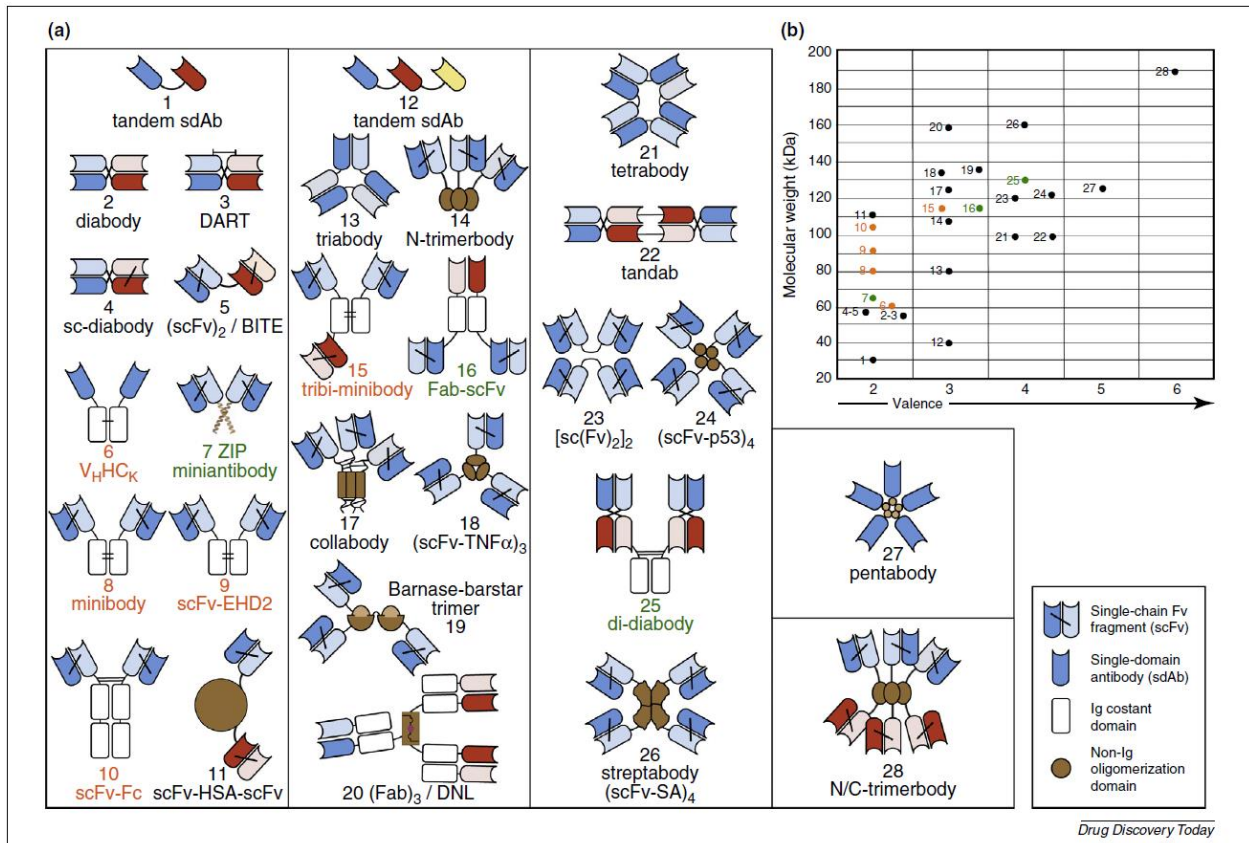


Figure 3 : (a) Schéma des différents anticorps (Ab) multivalents et ayant une masse moléculaire inférieure à 200 kDa. Distribués selon leur valence. (b) Masse moléculaire des constructions basées sur les Ab et leur valence.

Les Ab appartenant au IgG-like sont présentés avec une police orange et des points oranges, ceux n'appartenant pas au IgG-like en noir et ceux appartenant aux deux catégories dit « combinés » en vert. (Nunez-Prado, N., et al., 2015)

3) Les anticorps de camélidés

Outre les anticorps "classiques", les anticorps simple chaîne lourde de camélidés (HcAb) ont également vu leur ossature initiale modifiée. À la différence des anticorps classiques qui possèdent deux chaînes lourdes et deux chaînes légères, les HcAb possèdent seulement deux chaînes lourdes constituées de trois domaines chacune (Hamers-Casterman, C., *et al.*, 1993). Leur structure fut réduite par ingénierie au simple domaine VHH quelques années après leur découverte. (Arbabi Ghahroudi, M., *et al.*, 1997)

Le développement des HcAb a suivi la même voie que les anticorps conventionnels pour donner les nanobodies (Nbs). On trouve ainsi des Nbs bivalents, bispécifiques, pentavalents, ... Les Nbs sont également utilisés à des fins thérapeutiques sous des formes plus ou moins complexes. Ces Nbs permettent de contourner certaines contraintes inhérentes aux anticorps classiques et leurs dérivées. En effet, ils sont plus stables, solubles et l'affinité reste la même entre HcAb et Nbs, ce qui n'est pas forcément le cas entre ScFv et Abs. (Steeland, S., *et al.*, 2016)

4) Conclusion

Bien que l'ingénierie des anticorps ait permis l'avènement d'approches thérapeutiques inédites, des contraintes en rapport à leur architecture existent toujours. Les anticorps et leurs dérivés restent instables en conditions réductrices (ponts disulfures) ce qui empêche par exemple leur utilisation dans le cytoplasme des cellules eucaryotes. Ils présentent également une tendance prononcée à l'agrégation et leur production se révèle généralement coûteuse et très peu efficace en système procaryote. D'autres stratégies basées sur l'utilisation de protéines artificielles ont donc été mises en place.

B. Un système immunitaire ancestral

Une des voies possibles pour s'affranchir des contraintes liées à l'utilisation des anticorps fut découverte en 2004 chez des vertébrés primitifs, les Agnathes. Chez ces derniers, il est possible d'induire une réponse immunitaire adaptative alors qu'il n'y a ni lymphocyte B, ni d'autre signe de présence d'anticorps. Des cellules équivalentes aux lymphocytes B ont cependant été observées chez ces Agnathes. Au sein de ces cellules, des protéines à motifs structuraux répétés « *Leucine Rich Repeat* » (LRR) ont été trouvées et jouent, chez les agnathes, un rôle identique à celui des anticorps chez les mammifères. La recombinaison des motifs LRR entre eux lors du développement des cellules de l'immunité permet d'obtenir un répertoire immunitaire spécifique composé de protéines dont les sites de fixation sont hypervariables. La reconnaissance spécifique de ces protéines appelées « *Variable Lymphocyte Receptor* » (VLR) provient d'une diversification des résidus de la surface concave. La stabilité de ces VLRs est maintenue puisque les acides aminés impliqués dans la structure sont conservés. (Figure 4) (Pancer, Z., et al., 2004)

En 2008, la première structure d'un VLR en complexe avec sa cible a permis de comprendre comment des VLRs interagissent avec leur ligand. (Han, B.W., et al., 2008) Cette structure ouvrit la voie à la réalisation de bibliothèque basée sur les VLRs et la sélection de VLRs contre des cibles choisies. (Velikovsky, C.A., et al., 2009)

Il semblerait donc que l'émergence d'une potentialité de reconnaissance moléculaire illimitée ait initialement eu lieu non à partir de domaines immunoglobulines mais à partir d'architectures protéiques à motifs répétés auto-compatibles. Bien qu'ayant été découvert récemment, les VLRs se révèlent déjà comme étant une alternative efficace aux Abs. (Waters, E.A. and Shusta, E.V., 2018)

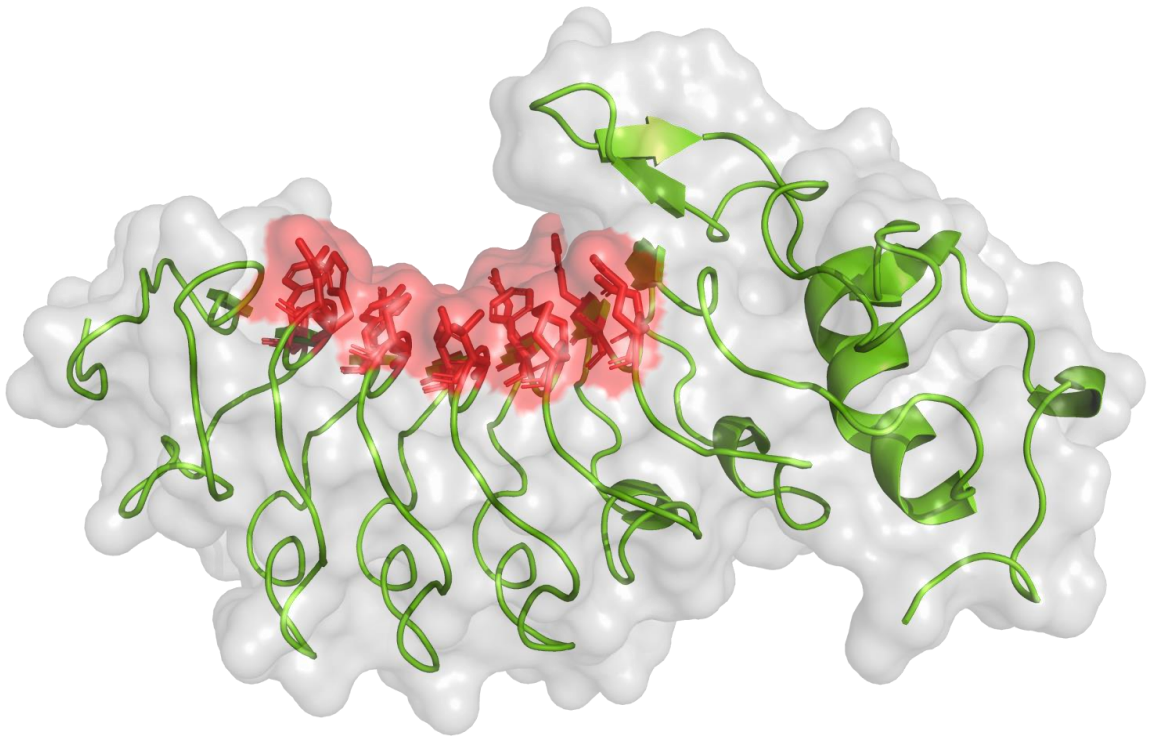


Figure 4 : Structure en cartoon d'un VLR (Variable Lymphocyte Receptor). La surface variable est mise en évidence en rouge. (Han, B.W., et al., 2008)

C. Les architectures artificielles comme alternative

1) Une variété d'architectures

D'autres types d'architectures qui présenteraient des propriétés similaires aux anticorps en termes de reconnaissance (spécificité et affinité) ont été développés en s'inspirant de ce que la nature a à offrir. La grande majorité de ces ossatures possède des avantages considérables en termes de production mais aussi de stabilité. Les protéines artificielles produites n'ont généralement pas de cystéine et donc pas de pont disulfure d'où un repliement possible en milieu réducteur et notamment dans le cytoplasme de cellules procaryotes. De plus, elles présentent une stabilité accrue tout en conservant un grand nombre de positions pouvant être mutées pour créer des surfaces de reconnaissance. Le nombre de ces ossatures ne cesse de croître et on en compte aujourd'hui plus d'une quarantaine. (Skrlec, K., *et al.*, 2015)

Les capacités de reconnaissance moléculaire des protéines artificielles dérivées de ces ossatures viennent, comme pour les VLR, de la possibilité de muter un grand nombre de résidus qui constitueront des sites de liaisons sans impacter le repliement de la protéine. Il est ainsi possible de créer des bibliothèques de protéines artificielles présentant des surfaces d'interaction variables de haute spécificité. La conception de ces protéines peut être basée sur des approches d'ingénierie combinatoire (évolution dirigée et conception rationnelle) ou uniquement rationnelle. (figure 5) (Binz, H.K., *et al.*, 2005)

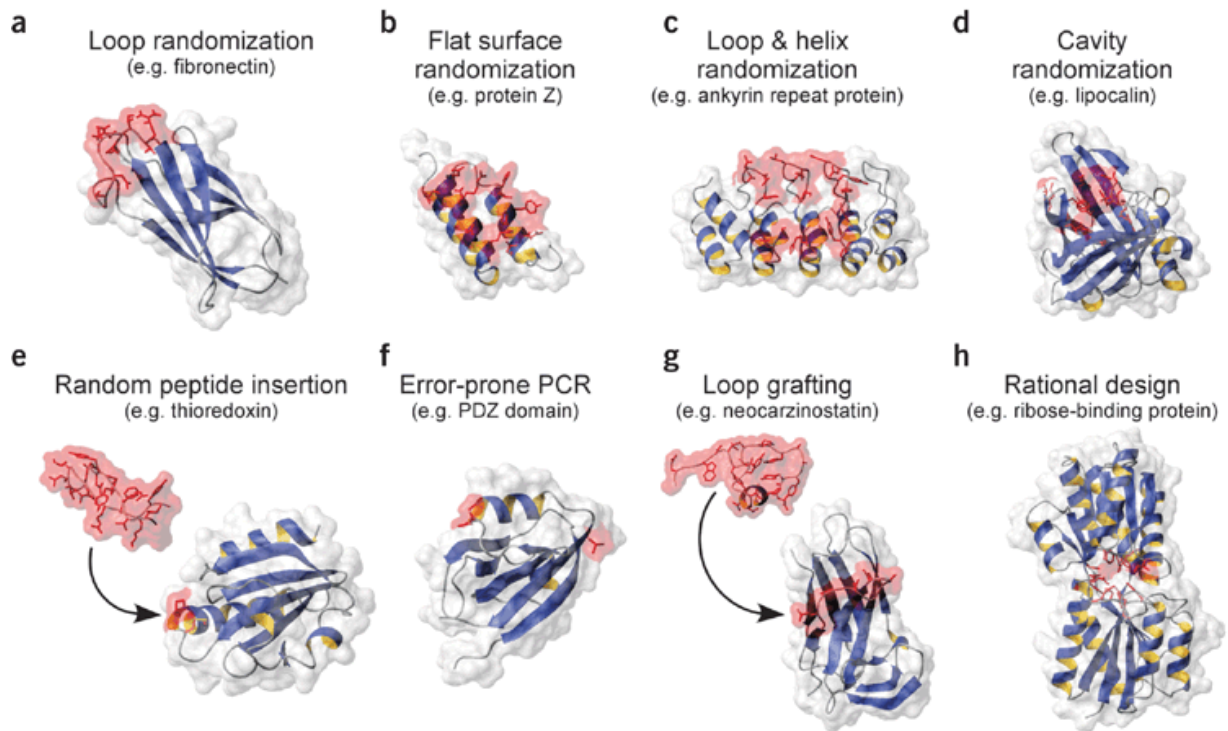


Figure 5 : Mise en évidence des surfaces des liaisons (rouge) et de l'ingénierie réalisée sur les différentes structures alternatives aux anticorps (a-h). Dans le cas d'approches par ingénierie combinatoire, la diversification de positions spécifiques se fait par diversification de codons définis. Cette randomisation peut avoir lieu dans des boucles (a), sur des surfaces (b), dans les boucles et hélices (c), dans des cavités (d), par insertion de peptides aléatoires (e), généralement dans des boucles, ou enfin à des positions aléatoires par « Error-prone PCR ». Dans le cas d'approches rationnelles, des séquences préexistantes connues pour interagir avec des cibles peuvent être insérées (g) ou des sites de liaisons peuvent être insérés dans des structures adaptées grâce au design de novo (h). (Binz, H.K., et al., 2005)

2) Les protéines à motifs répétés

a) Quelles ossatures ?

Parmi les ossatures protéiques choisies pour le développement d'outils de reconnaissance spécifique, les protéines à motifs structuraux répétés possèdent plusieurs caractéristiques intéressantes. En premier lieu, il est aisé par ingénierie de faire varier le nombre de motifs répétés et donc la taille des surfaces d'interaction sur de telles protéines. De plus, la séquence de chaque motif répété possède des positions variables et conservées qui sont identiques d'un motif à l'autre. Il est donc possible de ne raisonner que sur un motif seul et d'extrapoler sur toute une molécule plutôt que sur une molécule complète. Parmi les protéines à motifs structuraux répétés, on peut citer au moins cinq grandes familles de motifs : les *Leucine Rich Repeats* (dont les VLRs) ; *Ankyrine* (DARPin), *Tetratricopeptide Repeat* (TPR), HEAT (α Reps) et Armadillo. (Boersma, Y.L. and Pluckthun, A., 2011) (Figure 6)

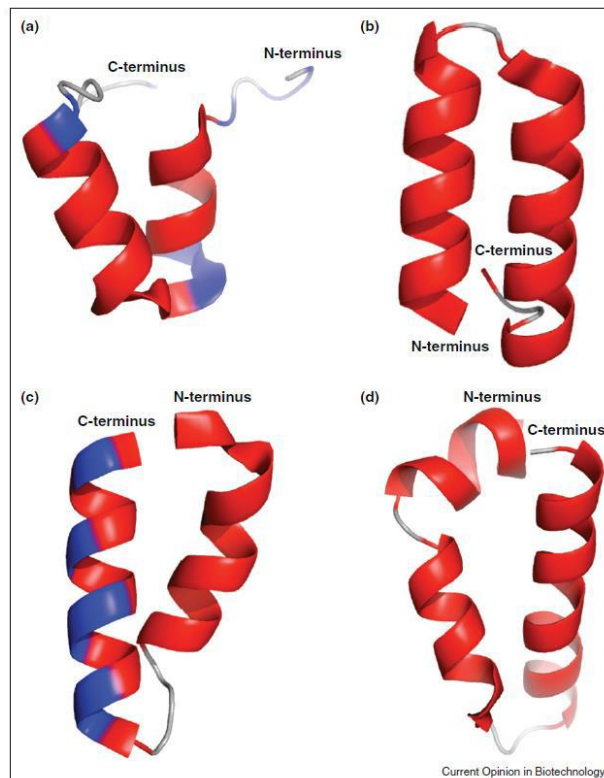


Figure 6 : Représentation cartoon des quatre principaux motifs structuraux répétés. Darpin basée sur le motif Ankyrine (code pdb 2XEE) (a). Motif TPR (code pdb 1NAO) (b). Motif HEAT des protéines α Rep (code pdb 3LTJ) (c). Motif ARM basé sur le mutant Y3MA (d). Les résidus « dégénérés » sont présentés en bleu en (a) et (c). (Boersma, Y.L. and Pluckthun, A., 2011)

Les protéines naturelles constituées de ce genre de motif sont souvent impliquées dans des processus de reconnaissance spécifique de cibles protéiques. À partir de ces protéines à motifs répétés, une séquence consensus peut être définie par alignement de séquences de motifs individuels. Cette analyse permet de repérer les résidus conservés et donc potentiellement engagés dans la stabilité de la structure et les résidus dits hypervariables, vraisemblablement impliqués dans la reconnaissance spécifique. Une approche de «*consensus design*» peut alors être envisagée.

b) L'exemple des Darpins

Ce type d'approche a été mis au point et validée sur les protéines à motif Ankyrine par l'équipe d'A. Plückthun. (Binz, H.K., *et al.*, 2003) La structure d'une DARPins à trois motifs internes est présentée figure 7. Ce même groupe a réalisé des bibliothèques avec ces protéines désormais appelées DARPins pour *Designed Ankyrin Repeat Protein*. Des sélections sur ces bibliothèques ont par la suite eu lieu. (Binz, H.K., *et al.*, 2004) Il est par ailleurs intéressant d'observer que la découverte des VLRs a eu lieu la même année. Depuis, de nombreuses applications ont été développées avec les DARPins que cela soit dans le domaine de la recherche, du diagnostic, ou des applications thérapeutiques. (Pluckthun, A., 2015)

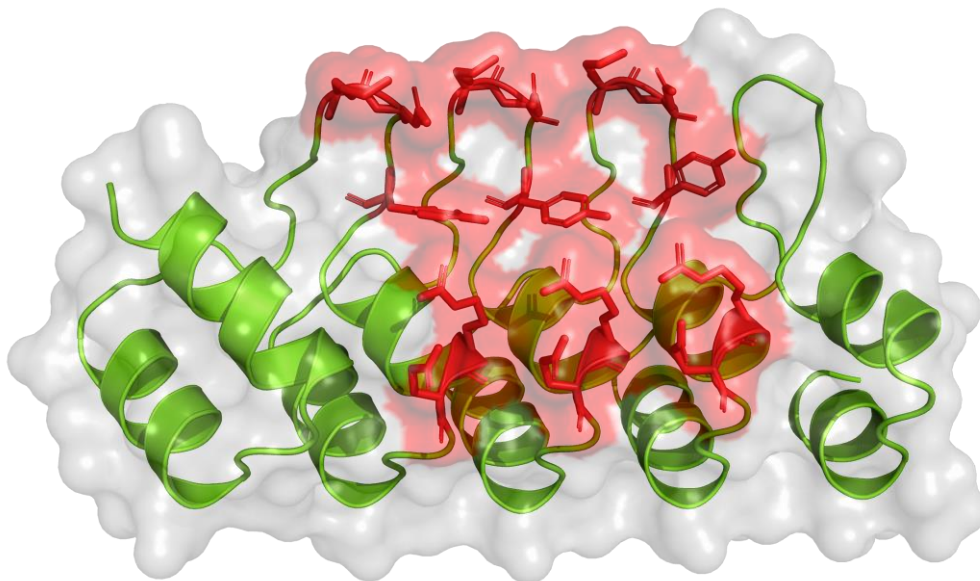


Figure 7 : Structure d'une DARPins à trois motifs internes, les résidus constants sont présentés en cartoon vert, les résidus variables sont visibles en stick rouge.

Une DARPin, l'abicipar, est actuellement en phase clinique III contre la dégénérescence maculaire rétinienne lié à l'âge. Deux DARPins bispécifiques, MP0250 et MP0274, sont en phase clinique II et I respectivement contre différents cancers. Ces deux molécules sont constituées de deux DARPins reliées par un *linker*. La première est capable d'interagir avec l'HGF (*Hepatocyte growth factor*) et le VEGF (*Vascular endothelial growth factor*), tous deux impliqués dans la formation de tumeurs solides. La seconde interagit avec HER2 (*human epidermal growth factor receptor 2*), impliqué dans les cancers du sein, sur deux épitopes strictement différents et rend ce récepteur inactif. (www.molecularpartners.com)

3) Création de reconnaissance *de novo*

L'utilisation de méthodes de conception rationnelle *de novo* permet également d'obtenir des propriétés de reconnaissance en se passant des méthodes d'expositions précédemment citées. Comme expliqué dans la partie présentant le design rationnel, cette approche se base sur des données préexistantes naturellement : identification de séquences d'acides aminés donnant des structures stables, énergies d'interface, variabilité des interfaces, structures. Avec l'accumulation d'informations dans les bases de données, la qualité des algorithmes ne cessent d'augmenter. (Kortemme, T. and Baker, D., 2004)

En 2011, l'équipe de D. Baker a pu créer des interacteurs de l'hémagglutinine, une glycoprotéine de surface du virus de la grippe. Pour obtenir cet interacteur, les auteurs ont analysé l'interaction entre un fragment d'anticorps (Fab) appelé CR6261 et l'hémagglutinine (H1 de 1918). Cette étape permet l'identification d'un « *hotspot* » de fixation, ou point d'ancrage. Ils ont ensuite choisi d'une part une ossature compatible pour remplacer ce Fab et d'autre part analysé quels résidus peuvent se fixer au niveau du site d'interaction. S'en suivent alors des étapes de minimisation énergétique de la structure couplées aux contraintes du « *hotspot* ». Les résidus hors du « *hotspot* » sont alors ajoutés puis les résultats sont validés par expérimentation. Les résultats de cette étude montrent que le design d'interacteur spécifique et très afin avec des K_D de l'ordre du nanomolaire est possible grâce au design *de novo*. (Figure 8) (Fleishman, S.J., et al., 2011)

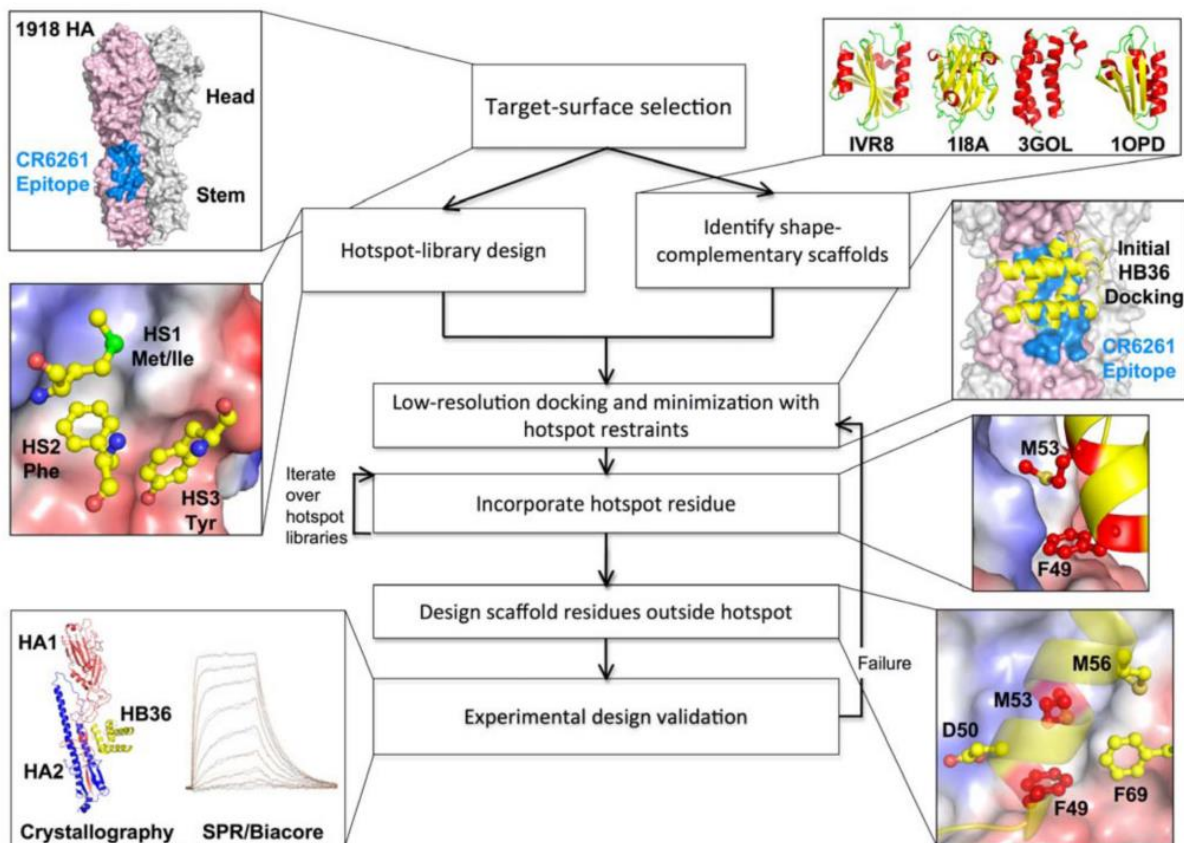


Figure 8 : Principales étapes du design de novo pour la création de nouvelles protéines de liaison. Ici, cette figure illustre le cas de la création de binder pour l'hémagglutinine de la souche de 1918. (Fleishman, S.J., et al., 2011)

IV. Création d'outils moléculaires complexes

Pour créer efficacement des outils complexes, il est possible de s'inspirer de l'évolution. En effet, en regardant dans la nature, les processus plus complexes que la "simple" reconnaissance tels que les activités enzymatiques, la transduction de signaux, la régulation des voies métaboliques font intervenir des protéines multidomaines.

À l'image de l'évolution, les chercheurs ont donc réutilisé des domaines pour les relier à d'autres et modifié des domaines pour changer leur fonction dans le but de créer des outils plus complexes. L'utilisation des méthodes d'ingénierie protéiques appliquées sur des multidomaines a ouvert de nouvelles possibilités : couplage de fonctions, augmentation de stabilité, création d'avidité, interacteurs plurispécifiques, détection de ligand, ...

A. Chimeric Antigen Receptor (CAR)

Les CARs sont, comme leur nom l'indique, des récepteurs antigéniques chimériques. Ils sont couplés aux lymphocytes T (*CAR T-Cell*) ce qui permet de cibler par exemple des cellules cancéreuses afin d'entraîner leur élimination. Ces récepteurs sont dits chimériques car modifiés par ingénierie protéique pour posséder un domaine de reconnaissance en plus des fonctions effectrices et d'ancrage membranaire. Ces molécules développées en 1989 étaient initialement constituées de trois parties, un ectodomaine, un domaine transmembranaire et un endodomaine. (Becker, M.L., *et al.*, 1989)

L'ectodomaine est constitué d'un peptide signal et d'un scFv qui sera spécifique d'une cible choisie grâce à des méthodes d'évolution dirigée ou d'immunisation sur animaux. Le peptide signal a pour fonction d'orienter les protéines chimériques vers le réticulum endoplasmique afin de permettre l'intégration à la surface des cellules T. Le scFv va accomplir la fonction de reconnaissance. L'endodomaine va être la partie effective du CAR et est le plus souvent constitué de la chaîne CD3 ζ contenant trois motifs ITAM (*Immunoreceptor Tyrosine-Based Activation*). (Figure 9)

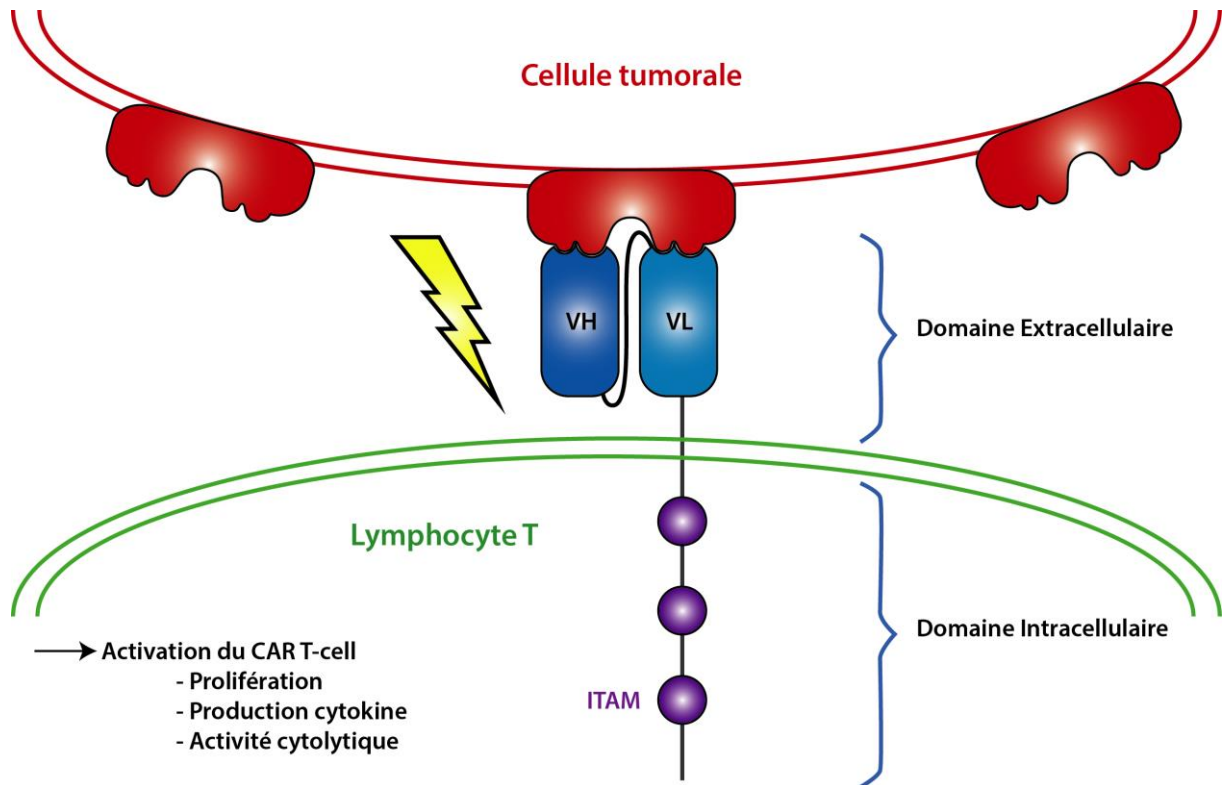


Figure 9 : Schéma du fonctionnement d'un CAR T-Cell. Un ScFv sélectionné contre un récepteur tumoral est relié à un domaine transmembranaire et à plusieurs ITAMs. L'interaction du ScFv avec le récepteur tumoral entraîne l'activation du CAR T-cell. De cette activation découlera une cascade de réaction entraînant la dégradation de la cellule tumorale.

Ces protéines chimériques ont vu leur structure évoluer grâce à de nombreuses modifications par ingénierie au cours des dernières années.

Au niveau de l'endodomaine, l'ingénierie a visé à augmenter la prolifération, la cytotoxicité ainsi qu'une prolongation de la durée de vie *in vivo*. Ainsi, un premier domaine de costimulation a été ajouté (CD28, CD137 ou CD134) avant les ITAMs pour donner les CARs de seconde génération. (Finney, H.M., *et al.*, 1998) La troisième génération de CARs a été obtenue par combinaison d'un second domaine de costimulation (CD28 + CD137 ou CD28 + CD134) afin d'augmenter la capacité d'élimination des lymphocytes T. (Marin, V., *et al.*, 2010) Enfin, une quatrième génération de CARs a vu le jour en 2015. Cette dernière génération ne possède qu'un seul domaine de costimulation comme la seconde génération mais une séquence NFAT (*Nuclear Factor of the Activated T-cell*) a été ajoutée. Cette séquence permet, une fois le récepteur activé, la sécrétion de cytokine et notamment d'interleukine-12 (IL-12). Ces IL-12 vont alors permettre le recrutement d'autres cellules immunitaires capables de détruire les cellules tumorales non reconnues par les CARs. Cette quatrième génération porte le nom de *T-cell Redirected for Universal Cytokine-mediated Killing* (TRUCK). (Chmielewski, M. and Abken, H., 2015; Zhang, C., *et al.*, 2017)

En parallèle de l'ingénierie de l'endodomaine, l'ectodomaine a également vu sa structure changée grâce à l'arrivée de nouvelles molécules de reconnaissance autres que les scFv. En effet, les scFv peuvent parfois être instables ou poser des problèmes d'immunogénicité lorsqu'issus de souris. Ainsi, en 2011, les domaines scFv furent remplacés par un nanobody sur des CARs de deuxième génération. (Iri-Sofla, F.J., *et al.*, 2011; Steeland, S., *et al.*, 2016) Finalement, en 2016, ce sont des VLRs qui viennent remplir la fonction de reconnaissance, toujours sur des CARs de deuxième génération (Moot, R., *et al.*, 2016).

L'ensemble des évolutions des CARs est présentée dans la [figure 10](#).

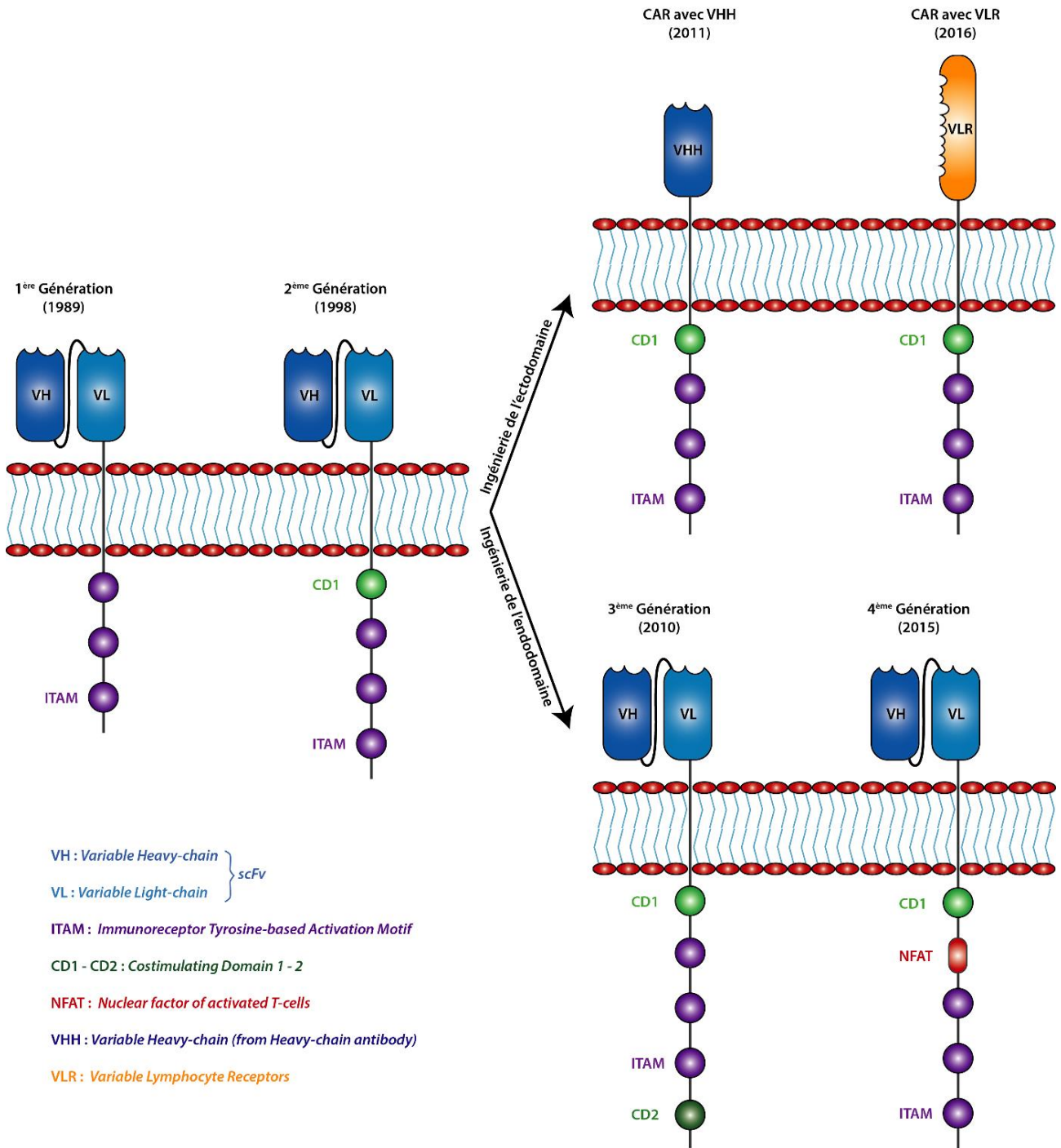


Figure 10 : Évolution des récepteurs antigéniques chimériques (CAR) au cours des dernières années. La 1^{ère} génération de CAR était constituée d'un ectodomaine (scFv + peptide signal), d'un domaine transmembranaire et d'un endodomaine (CD3 ζ avec trois ITAMs). La seconde génération de CAR s'est vue ajouter un domaine de costimulation (CD1), généralement CD28. La 3^{ème} génération possède un second domaine de costimulation (CD2), généralement CD134 ou CD137. Enfin la 4^{ème} génération n'avait qu'un domaine de costimulation mais une séquence NFAT permettant la production d'interleukine 12 après activation du récepteur. Du côté de l'ectodomaine, le scFv a été remplacé soit par un VHH soit par un VLR.

B. De la reconnaissance à la détection : Les biosenseurs

Les biosenseurs peuvent se définir au sens très large comme des outils biologiques ou semi-biologiques permettant de détecter ce qui est invisible pour nos sens. Un biosenseur va coupler deux processus, d'une part il va reconnaître spécifiquement une cible (ions, molécules) et d'autre part il effectuera la transduction de cette reconnaissance sous forme d'un signal détectable (lumineux, résistance à un antibiotique, activité enzymatique, ...).

Un des biosenseurs les plus utilisés au monde permet une mesure électrochimique de la glycémie et est constitué d'une électrode enzymatique. Une enzyme, la glucose déshydrogénase, est immobilisée sur une électrode. Elle réagit spécifiquement avec le glucose et produit des électrons, générant alors un courant mesurable par l'électrode. Ce courant sera directement proportionnel à la glycémie. La spécificité des biosenseurs et donc dans ce cas de l'enzyme, est primordiale pour une détection et / ou quantification correcte. En effet, dans l'exemple présenté ici, l'enzyme initialement utilisée pour mesurer la glycémie était la glucose oxydase. Or, cette enzyme n'était pas totalement spécifique du glucose, des résultats faussés ont donc été obtenus et ont conduit aux décès de plusieurs patients diabétiques par surdosage d'insuline.

Nous nous intéresserons dans cette partie uniquement aux biosenseurs basés sur des constructions moléculaires et utilisables en solution ou *in vivo*.

1) Biosenseurs basés sur la variation d'un signal lumineux

a) Le FRET : *Förster Resonance Energy Transfer*

Une méthode efficace de détection de molécule pour la réalisation de biosenseur est l'utilisation du FRET. Initialement décrit en 1948 par Theodor Förster, cette méthode permet d'estimer la distance entre deux points (fluorophores) distants de 1 à 10 nm. Elle se base sur le transfert d'énergie qui peut exister entre deux molécules fluorescentes (appelées Donneur et Accepteur). Le donneur sera excité par un rayonnement lumineux et passera d'un état électronique fondamental à un état excité métastable. Lors de son retour à l'état fondamental, le donneur peut soit émettre de la lumière (fluorescence), soit retourner à l'état fondamental par relaxation non radiative, soit effectuer un transfert d'énergie avec l'accepteur. Dans ce dernier cas, l'accepteur sera alors excité et pourra revenir à son état fondamental en émettant de la lumière si les conditions le permettent. (Figure 11-A)

Plusieurs conditions doivent en effet être réunies pour observer du FRET. Le Donneur et l'Accepteur doivent avoir des transitions énergétiques de hautes énergies dans l'UV, le visible et l'infra-rouge. Il doit exister une superposition partielle du spectre d'émission du donneur avec le spectre d'absorption de l'accepteur (Figure 11-B). Les molécules doivent être proches, mais pas trop. Enfin, l'orientation des molécules doit être adaptée à l'obtention d'un signal de FRET. (Medintz, I. and Hildebrandt, N., 2014)

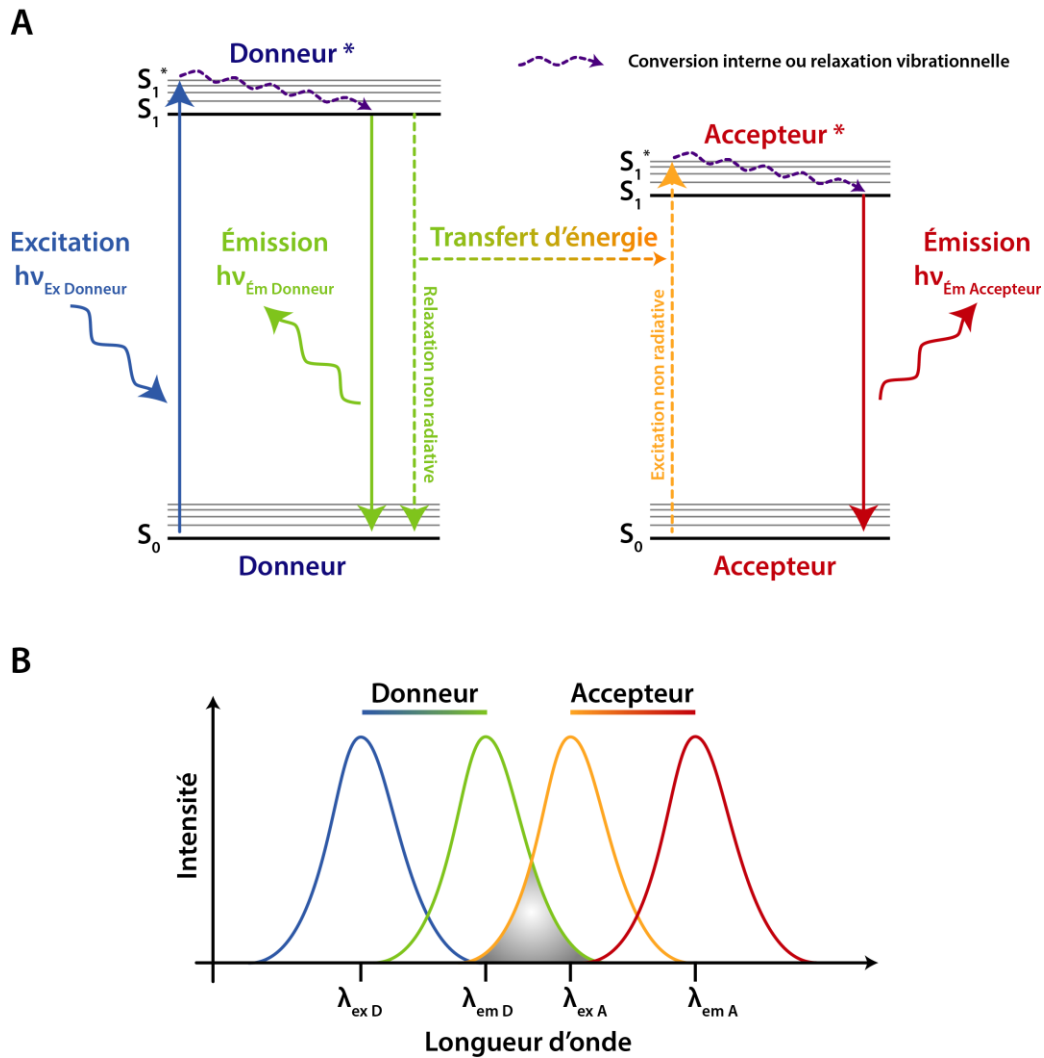


Figure 11 : A) Diagramme d'état d'énergie. Un donneur D est excité et passe d'un état électronique fondamental S_0 à un état excité métastable S_1^* . Par conversion interne ou relaxation vibrationnelle, le donneur passe à un état excité singulet relaxé S_1 . Le donneur peut alors retourner à l'état S_0 en émettant de la fluorescence ou non (via une relaxation non radiative). Lors de la relaxation non radiative, un transfert d'énergie peut avoir lieu vers un accepteur A qui passera d'un état S_0 à S_1^* puis S_1 . Il pourra alors retourner à son état S_0 par émission de lumière. B) Spectre d'absorption et d'émission d'un couple Donneur / Accepteur en fonction de la longueur d'onde. La superposition des spectres d'émission du donneur et d'excitation de l'accepteur est nécessaire pour avoir du FRET.

b) L'apport des protéines fluorescentes et bioluminescente

Bien que découvert en 1948, ce n'est qu'avec la découverte de biomolécules fluorescentes et leur ingénierie que l'utilisation du FRET en biologie et biochimie a réellement explosé. (Zhang, J., *et al.*, 2002; Miyawaki, A. and Niino, Y., 2015)

La découverte et purification de la première protéine fluorescente s'est faite parallèlement à celle d'une protéine bioluminescente. En effet, initialement le Dr O. Shimomura s'intéressait à la bioluminescence de la méduse *Aequorea victoria*. Cette bioluminescence était due à l'aequorine, qui en présence de coelenterazine, de Ca^{2+} et de O_2 , génère une lumière bleue ($\lambda_{\text{max}} = 460 \text{ nm}$) grâce à l'oxydation du substrat. Les auteurs, en plus de purifier l'aequorine, ont réussi à purifier une protéine émettant une lumière verte quand excitée dans l'UV, mais ne produisait pas de bioluminescence en présence de Ca^{2+} et de Coelenterazine. Or, la lumière observée *in vivo* chez *Aequorea victoria* est plutôt verte ($\lambda_{\text{max}} = 510 \text{ nm}$). Il devait donc exister un mécanisme inconnu permettant d'exciter cette protéine verte en se passant d'UV. (Shimomura, O., *et al.*, 1962) Une première hypothèse venant expliquer ce phénomène arriva en 1963, elle suggérait que la lumière émise par l'aequorine venait exciter la protéine verte qui émettait alors à 510 nm. (Johnson, F.H., *et al.*, 1962) Finalement, c'est en 1971 qu'un parallèle fut établi avec le FRET, c'était la découverte du BRET (*Bioluminescence resonance Energy Transfer*). (Morin, J.G. and Hastings, J.W., 1971) Cette protéine verte, n'était autre que la GFP (*Green Fluorescent Protein*) qui est à l'heure actuelle l'un des outils biotechnologiques les plus utilisés. Quelques années plus tard, cette GFP fut produite dans *E. Coli*. (Chalfie, M., *et al.*, 1994) Enfin, de nombreux variants permettant d'obtenir une fluorescence améliorée (eGFP), bleu (BFP), cyan (CFP) et jaune (YFP) ont été obtenus par l'équipe de R. Y. Tsien dans les années qui suivirent. (Shaner, N.C., *et al.*, 2005)

Pour l'ensemble de leurs recherches sur la GFP, O. Shimomura, M. Chalfie et R. Y. Tsien reçurent le prix Nobel de Chimie en 2008.

c) Biosenseurs basés sur du FRET

L'un des premiers biosenseurs basé sur l'utilisation du FRET permettait la détection de l'AMPc. Ce biosenseur créé au sein de l'équipe de R.Y. Tsien en 1991 est basé sur la protéine kinase A (PKA) et son mécanisme d'activation dépendant de l'AMPc. En effet, la régulation de la PKA implique l'interaction entre sous-unités catalytiques et sous-unités régulatrices. La présence d'AMPc entraîne la séparation des sous-unités entre elles. Ainsi, le couplage de deux fluorophores, la fluoescéine et la rhodamine, aux extrémités respectives des sous-unités catalytique et régulatrice permet d'obtenir un signal de FRET dépendant de la concentration d'AMPc. (Figure 12-a) (Adams, S.R., *et al.*, 1991)

Quelques années plus tard, en 1997, cette même équipe a mis au point un biosenseur de Ca^{2+} , fonctionnant sans cofacteur et pouvant être ciblé vers différents compartiments cellulaires de manière spécifique. Cette nouvelle construction, est « génétiquement codable » et se base sur la GFP, ses variants (BFP, CFP et YFP), la calmoduline (ou CaM) et le peptide M13. La construction est constituée de deux protéines fluorescentes (FPs) entre lesquelles se trouvent la CaM et le peptide M13. La CaM s'enroule autour du peptide M13 en présence de calcium et va changer de conformation. Ce changement de conformation va rapprocher les FPs entre lesquelles il pourra y avoir un transfert d'énergie. Ce FRET sera donc dépendant de la concentration en Ca^{2+} (Figure 12-b). Deux couples de FPs sont utilisés, BFP et GFP ou CFP et YFP. Les auteurs ont également montré que l'ajout d'une séquence de localisation permet de guider la construction chimérique vers les différents compartiments cellulaires. Les constructions obtenues sont appelées des "Cameleons". Ce nom provient de la CaM (**Calcium-Modulated protein**) et du fait que la fixation du Ca^{2+} entraîne un changement de longueur d'onde d'émission tel le changement de couleur du caméléon. (Miyawaki, A., *et al.*, 1997)

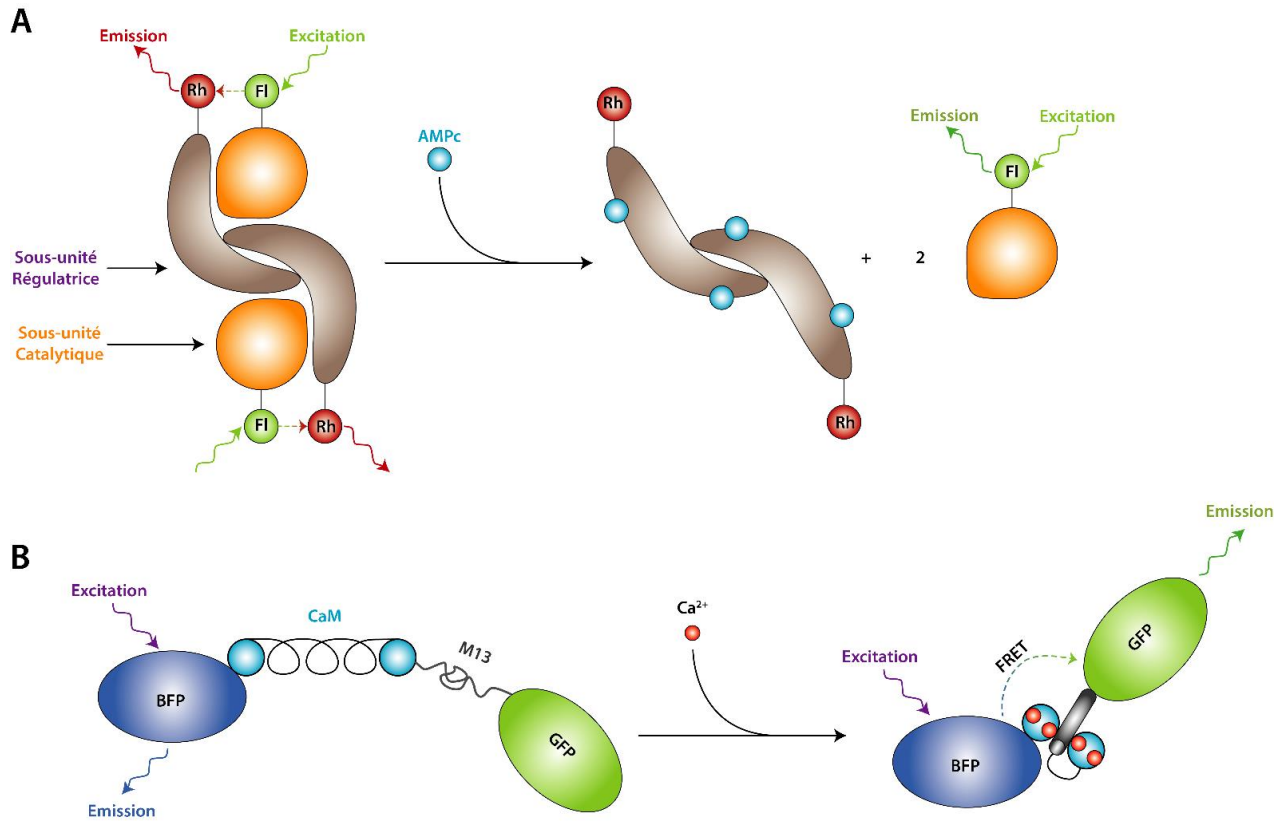


Figure 12 : Deux biosenseurs basés sur le FRET mais dont le principe varie. (A) Bio senseur d'AMPc basé sur la régulation de la Protéine Kinase A. Lorsque l'AMPc se fixe sur les sous-unités régulatrices, il y a une dissociation des sous-unités catalytique. Le couplage de la fluorescéine et de la rhodamine aux extrémités des sous-unités permet de détecter l'AMPc avec la disparition du FRET. (Switch Off) (Adams, S.R., et al., 1991) (B) Biosenseur de type Cameleon. Basé sur la Calmoduline (CaM) et le peptide minimal M13 autour duquel se replie la CaM en présence de Ca^{2+} . Dans ce cas, la fixation du Ca^{2+} entraîne l'apparition d'un signal de FRET. (Switch On) (Miyawaki, A., et al., 1997)

d) Biosenseurs basés sur le BRET

Afin de créer des biosenseurs, il est également possible d'utiliser la bioluminescence comme source de lumière (BRET). La bioluminescence, à la différence de la fluorescence ne nécessite pas d'excitation lumineuse pour passer à un état excité. C'est une réaction biochimique qui est responsable de l'état excité.

L'équipe de K. Johnsson a ainsi mis au point plusieurs senseurs basés sur la nanoKaz, un domaine catalytique optimisé de la luciférase. Cette dernière s'apparente à l'aequorine mais est capable d'émettre une lumière bleue en présence de coelenterazine et d'O₂ uniquement (sans Ca²⁺). Nous reviendront plus en détails dessus dans la partie B de ce manuscrit. Cinq systèmes basés sur un même principe ont été mis en place par l'équipe de K. Johnsson et se sont révélés efficaces sur six molécules thérapeutiques. Les molécules choisies par les auteurs nécessitent un dosage précis et régulier chez les patients : Méthotrexate (anti tumoral), Tacrolimus et Sirolimus (immunosuppresseurs), Cyclosporine A (immunosuppresseurs), Topyramate (antiépileptique) et Digoxine (maladie cardiaque).

Le principe utilisé repose sur la construction d'une protéine chimérique. Celle-ci est constituée d'un domaine de fixation du ligand cible, d'une nanoLuc (nano Luciférase), d'un linker rigide, d'un SNAP-TAG sur lequel est greffé un long *linker* possédant un marqueur fluorescent et un ligand dit "inhibiteur". Le SNAP-TAG est un domaine protéique de 20 kDa réagissant spécifiquement avec les dérivés de benzylguanine (BG) de manière covalente. Il est possible de greffer des étiquettes fluorescentes sur la BG et donc sur n'importe quel SNAP-TAG. Le ligand inhibiteur va pouvoir interagir avec le domaine de fixation du ligand avec une moins bonne affinité que ce dernier. Sans le ligand, l'interaction intramoléculaire inhibiteur/domaine de liaison a lieu, rapprochant ainsi la nanoLuc du marqueur fluorescent et permettant l'obtention d'un signal de BRET. Avec le ligand, l'inhibiteur est décroché et le signal de BRET diminue (figure 13). Ce système appelé "LUCID" (*Luciferase-based indicators of drugs*) permet une mesure simple des concentrations à l'aide d'un simple appareil photo. (Griss, R., *et al.*, 2014)

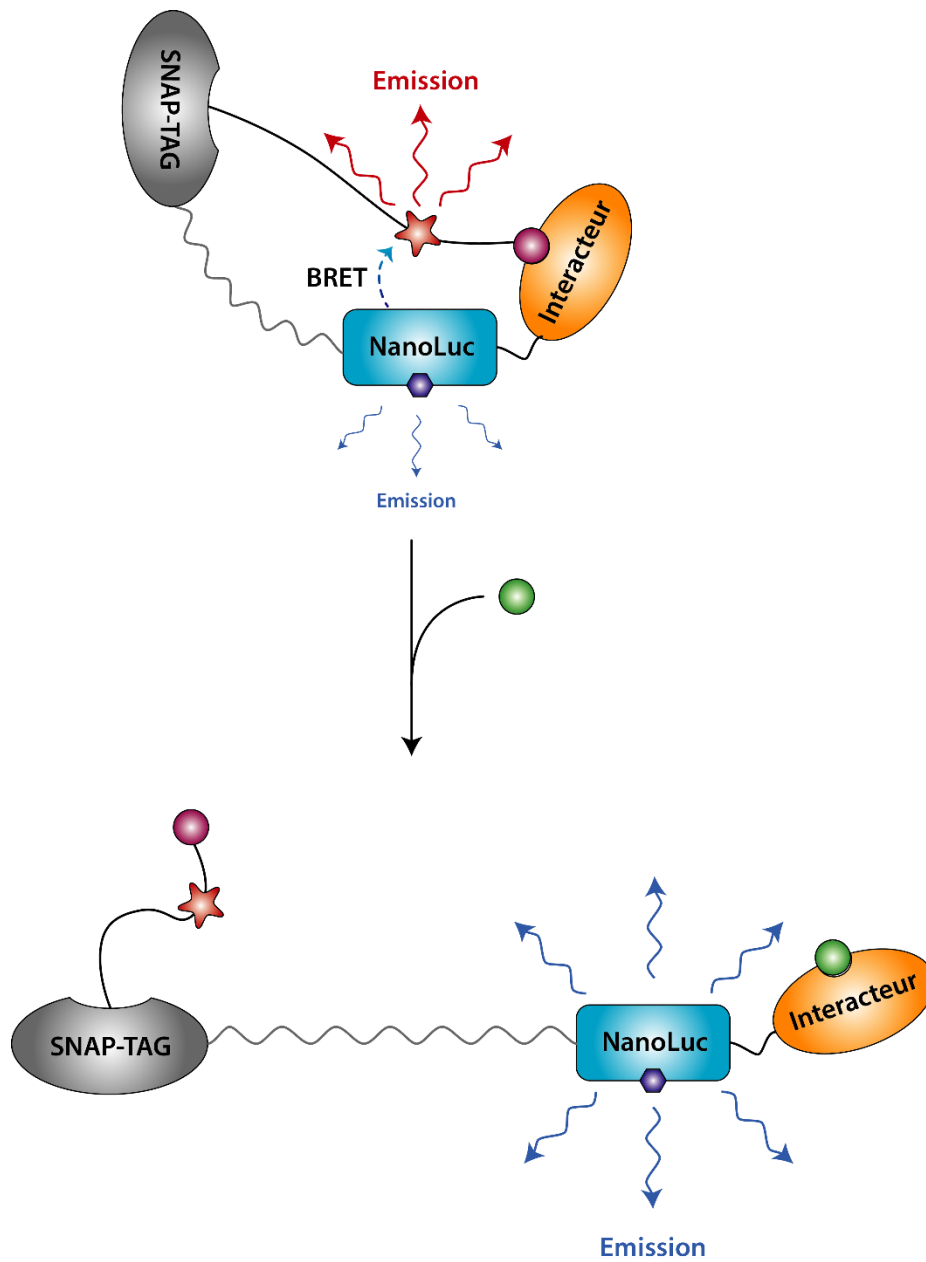


Figure 13 : Principe de fonctionnement des LUCIDs. Sans ligand, un inhibiteur (en violet) se fixe sur le domaine interacteur, rapprochant ainsi le fluorophore (étoile rouge) de la NanoLuc et permettant d'obtenir un signal de BRET. En présence du substrat, ce dernier se fixe sur le domaine interacteur. Le linker rigide reliant le SNAP-TAG de la nanoLUC va alors contraindre la chimère dans une conformation ouverte éloignant le fluorophore de la NanoLuc. Le ratio émission NanoLuc / émission fluorophore est directement proportionnel à la concentration de ligand fixé sur le LUCID.

e) Un nouveau couple Donneur / Accepteur pour voir plus loin :

Plus récemment, deux fluorophores inorganiques ont permis de « repousser » les limites du FRET. Ces deux fluorophores sont les cryptates ou chélates de Lanthanides et les Quantum Dot (QDs). Pris indépendamment, ces éléments présentent des propriétés photophysiques très intéressantes. En effet, les lanthanides ont une durée de vie de fluorescence allant jusqu'à quelques centaines de microsecondes (quelques dizaines de nanosecondes pour les fluorophores « classiques »). Les QDs quant à eux ont une brillance importante et émettent à des longueurs d'ondes dépendantes de leur taille. La combinaison de ces deux fluorophores permet d'augmenter la distance maximale du FRET à 20 nm (contre 10 nm avec des fluorophores classiques). Ces diverses propriétés en font un très bon couple pour la réalisation de biosenseurs basés sur le FRET. (Geissler, D., *et al.*, 2014)

Parmi les études récentes utilisant ce couple, Mattera *et al.* ont réalisé un biosenseur du TPSA (*Total Prostate Specific Antigen*). Les auteurs ont couplé un anticorps anti-TPSA à un cryptate de lanthanide (Terbium) et un QD à un Fab anti-TPSA (épitope strictement différent). Ce biosenseur présente une limite de détection de 0.8 ng/μL soit 0.02 nM, ce qui est en dessous du seuil de limite clinique (4 ng/μL) (Figure 14). (Mattera, L., *et al.*, 2016)

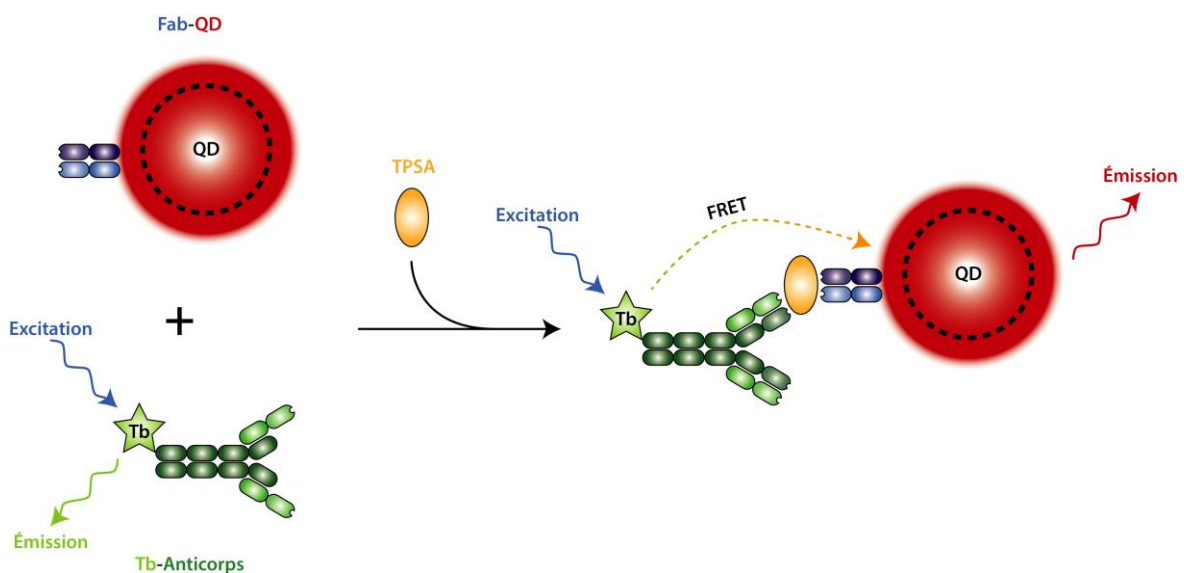


Figure 14 : Principe du biosenseur basé sur l'utilisation des QDs et des lanthanides. La présence de TPSA rapproche le complexe Tb-Anticorps et le Fab-QD permettant d'observer une apparition d'un signal de FRET.

2) Biosenseurs basés sur une résistance aux antibiotiques

Parmi les systèmes de biosenseurs, une tentative de mise au point de détecteurs génériques a été réalisée par l'équipe de M. Ostermeier. Le but de cette étude était de trouver un moyen général de transformer un système en biosenseur *via* la capacité à résister à l'ampicilline. Bien que les résultats ne soient pas généralisables à toutes molécules, l'approche a le mérite d'être originale et prometteuse.

Les auteurs ont dans un premier temps créé une bibliothèque de chimères possédant un domaine *input* et un domaine *output*. Le domaine *input* interagira avec la cible. Il s'agira soit des monobodies, soit des DARPins et en premier lieu ils seront spécifiques de la *Maltose Binding Protein* (MBP). Le domaine *output*, constitué d'une β -lactamase, a pour but de permettre un criblage et une détection aisés. La bibliothèque sera constituée de différentes constructions dans lesquelles les extrémités N et C-Terminale seront à différents endroits dans la séquence du domaine *output* *via* un procédé de permutation circulaire. L'étape de sélection a consisté à trouver pour chaque type de domaine *input* (monobodies ou DARPins) une configuration dans laquelle l'ajout de MBP octroyait une résistance accrue à l'ampicilline. En d'autres termes, avec quelle construction la présence de la MBP change suffisamment la conformation de la β -lactamase pour augmenter son efficacité. Une fois la construction trouvée, le domaine *Input* a été changé pour des domaines interagissant avec l'eGFP, APH(3')IIIA et γ SUMO (Figure 15). Les résistances à l'ampicilline en présence et en absence des ligands susmentionnés ont alors été testées. Les résultats montrent que la résistance à l'ampicilline augmente pour la construction DARPins visant la APH(3')IIIA. Aucune construction n'a cependant fonctionné avec l'eGFP ou avec γ SUMO.

Ces résultats sont cependant encourageants puisque l'on peut supposer que le problème est potentiellement lié uniquement aux interacteurs qui ne subissent pas un changement de conformation suffisant pour activer la β -lactamase. (Nicholes, N., *et al.*, 2016)

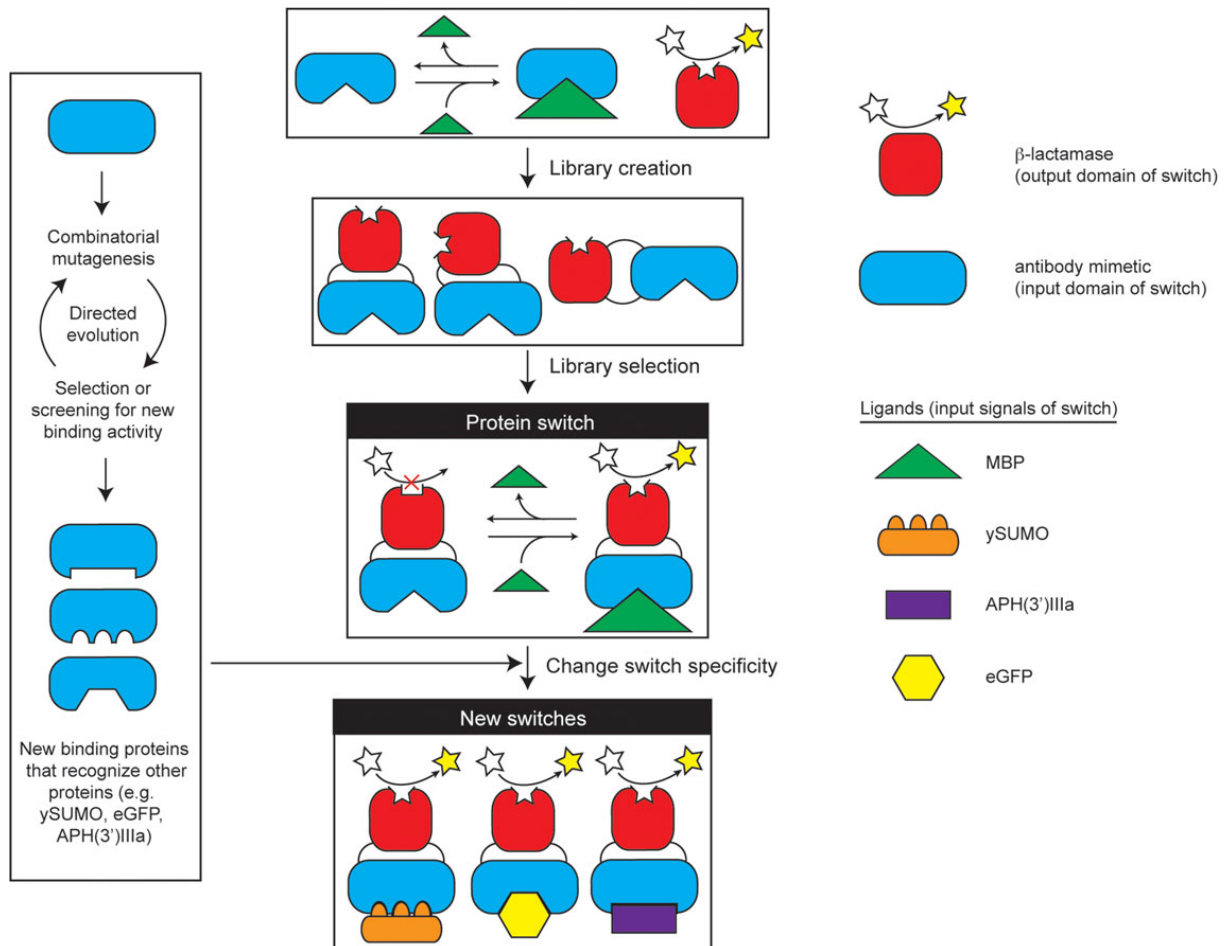


Figure 15 : Méthode mise en place par l'équipe de M. Ostermeier. La première étape consiste à créer une bibliothèque de protéine chimérique N-Ter β -lactamase + domaine d'interaction à la "maltose binding protein" + C-Ter β -lactamase. La 2^{ème} étape consiste à cribler la bibliothèque pour voir quelle construction voit sa résistance à l'ampicilline augmenter en présence de MBP (cible du domaine d'interaction). Enfin, le domaine d'interaction est changé par un domaine de spécificité différente et la résistance à l'ampicilline est testée avec les nouvelles cibles. (Nicholes, N., et al., 2016)

3) Biosenseurs basés sur la permutation circulaire de protéine fluorescente

Une autre méthode simple pour visualiser un changement de conformation et potentiellement détecter une cible est de réaliser ce que l'on appelle une permutation circulaire sur des molécules fluorescentes ou bioluminescentes. Une permutation circulaire consiste à changer les positions N-Terminal et C-Terminal au sein de la protéine. Il est alors possible de les positionner d'une manière favorable aux changements spectraux dans des constructions chimériques. Les premières expériences montrant la tolérance de la GFP et de ses variants vis-à-vis des permutations circulaires ont été effectuées par l'équipe de R. Tsien. Cette équipe a mis au point les premiers biosenseurs de Ca^{2+} basés sur des permutations circulaires. Appelés Camgaroo1, ces biosenseurs étaient basés sur la CaM insérée au sein de la GFP ou de l'YFP (*Yellow Fluorescent Protein*) (Figure 16-A). La fixation du Ca^{2+} entraîne un changement de conformation de la GFP au niveau de son chromophore, entraînant une augmentation de la fluorescence. Ces senseurs présentent des K_D de l'ordre de 7 μM . Enfin, les auteurs ont également inséré un motif en doigt de zinc à la place de la calmoduline et ont pu observer une augmentation de la fluorescence en présence de Zinc avec un K_D estimé à 400 μM . (Baird, G.S., et al., 1999)

Grâce à ces résultats, l'équipe de K. Imoto a réalisé avec la construction M13_(C-Ter)eGFP_(N-TER)eGFP_CaM (où _ représente les *linkers*) un biosenseur du Ca^{2+} plus sensible avec un K_D de l'ordre de 200 nM. Les résultats montrent également que la détection du Ca^{2+} est possible *in vivo* dans les myotubes. (Figure 16-B) (Nakai, J., et al., 2001)

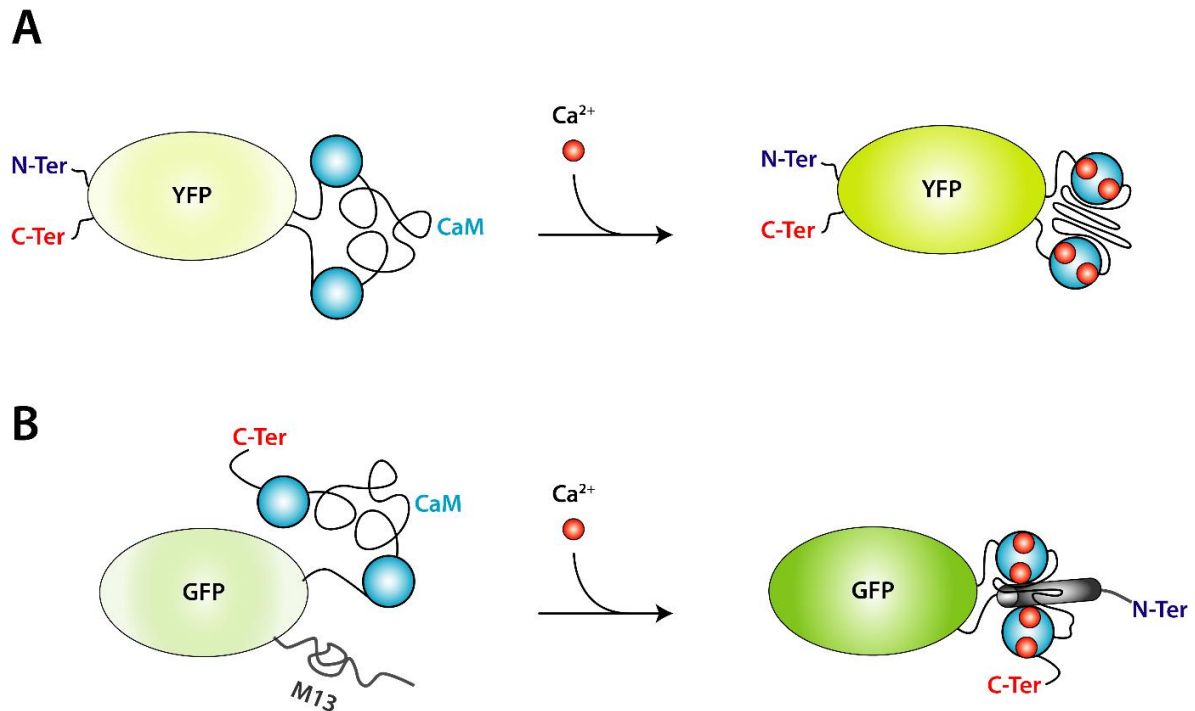


Figure 16 : Modèle présentant les permutations circulaire effectuée par Baird et al et Nakai et al sur la YFP et la GFP respectivement.

L'équipe de A. Miyawaki a travaillé sur des constructions chimériques du type M13_(C-Ter)eYFP_(N-TER)eYFP_CaM appelé "*pericam*". Différentes mutations et plusieurs *linkers* ont été testés afin d'optimiser la réponse au Ca^{2+} du *pericam*. (Nagai, T., et al., 2001) Les auteurs ont ainsi pu obtenir trois *pericams* différents. Les **Flash-pericams** qui présentent une augmentation de l'absorbance à 490 nm en présence de Ca^{2+} et une augmentation de la fluorescence en présence de Ca^{2+} par rapport au *pericam* classique. Les **ratimetric-pericams**, qui ont une augmentation de l'absorbance à 490 nm en présence de Ca^{2+} et présente deux pics d'excitation contrairement à la *flash-pericam*. De plus, il y a un décalage du pic d'émission de 511 à 517 nm et une augmentation de la fluorescence en présence de Ca^{2+} par rapport au *pericam* classique. Enfin, les **Inverse-pericams**, qui n'ont qu'un seul pic d'absorbance à 490 nm et présente une diminution de la fluorescence en présence de Ca^{2+} et non augmentation comme c'est le cas pour les autres *pericams*.

4) Biosenseurs basés sur la complémentation fonctionnelle

La complémentation fonctionnelle (*Protein Complementation Assay* : PCA) est une méthode qui consiste à vérifier l'interaction de deux protéines grâce à la reconstitution d'une activité dite rapporteuse. Pour cela une protéine rapporteuse (DiHydroxy Folate Réductase (DHFR), protéase TEV, GFP et ses variants, Luciférase, ...) est séparée en deux fragments distincts reliés aux extrémités des protéines supposées interagir ensemble. Si l'interaction a lieu, l'activité de la protéine rapporteuse est restituée (Figure 17). Cette approche a en premier lieu été effectuée avec l'ubiquitine clivée en deux et couplée à une DHFR en 1994. La reformation de l'ubiquitine entraîne un clivage libérant la DHFR qui apporte alors une résistance au triméthopime (Johnsson, N. and Varshavsky, A., 1994). L'avantage de la PCA est que cette technique peut être appliquée dans n'importe quel compartiment cellulaire. Par la suite de nombreuses études ont visé à augmenter le nombre de protéines rapporteuses et à augmenter leur stabilité. (Michnick, S.W., et al., 2000; Morell, M., et al., 2009)

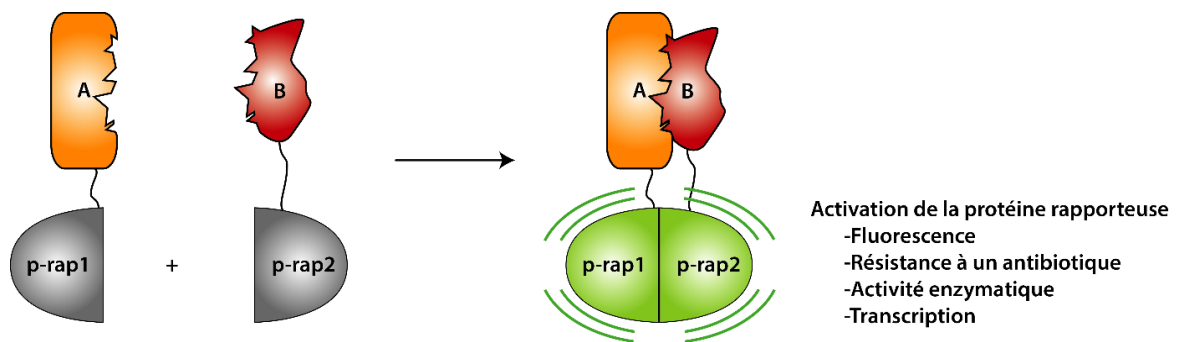


Figure 17 : Principe de la "Protein Complementation Assay". Une protéine A et B dont l'interaction est supposée sont reliées à deux fragments d'une protéine rapporteuse (p-rap1 et p-rap2). Si les protéines A et B interagissent ensemble, l'activité de la protéine rapporteuse est restituée.

5) Biosenseurs basés sur la stabilité des biosenseurs *in vivo*

L'équipe de D. Baker a mis au point une stratégie innovante afin de détecter les petites molécules *in vivo*. Cette méthode se base sur l'instabilité des biosenseurs en absence de ligand. Cela permet d'avoir une accumulation du signal uniquement en présence du ligand. Les auteurs ont ainsi développé deux types de biosenseurs capables de détecter la progestérone et la digoxine (traitement des maladies cardiaques).

Dans un premier temps, les auteurs ont travaillé uniquement dans des levures.

Le premier biosenseur construit est constitué d'un domaine appelé LBD (*Ligand Binding Domain*) capable de reconnaître les molécules désirées et d'une protéine rapporteuse, la GFP ([Figure 18-A](#)). Le LBD est un domaine conçu *de novo*, deux LBD différents ont été testés ici, DIG₀ (Digoxine) et PRO₀ (Progestérone). ([Tinberg, C.E., et al., 2013](#)) Les auteurs ont ensuite muté la structure multidomaine pour la déstabiliser de telle sorte que le biosenseur ne s'accumule que dans les levures contenant le ligand. Pour chaque mutation, les auteurs ont mesuré les signaux de fluorescence de la GFP après excitation. La présence de ligand permet d'augmenter le signal fluorescent d'un facteur maximal de 5.4 fois le signal basal.

Afin d'améliorer la sensibilité de leurs biosenseurs, les auteurs ont modifié la construction initiale. La GFP a été enlevée pour être remplacée par un domaine de liaison à l'ADN appelé DBD (DNA Binding Domain) en N-terminal et un domaine d'activation de la transcription (TAD) en C-Terminal ([Figure 18-B](#)). Le DBD permet une liaison aux sites promoteur de transcription avant le gène rapporteur codant pour la *yEGFP* (*Yeast Enhanced GFP*). Les auteurs ont alors testé différents DBDs (se liant à différent endroit sur l'ADN) et TADs pour déterminer quelle combinaison permettait d'obtenir la plus haute sensibilité. Les résultats montrent que ce sont les promoteurs et activateurs faibles qui permettent d'obtenir le meilleur ratio signal/bruit avec une augmentation de signal égal à 40 fois le signal basal ([Feng, J., et al., 2015](#)).

Les auteurs ont également testé la seconde construction *in vivo* dans des plantes. Ils ont pu mettre en évidence la possibilité de détecter *in vivo* la digoxine grâce à leur système. L'ensemble de cette étude laisse supposer qu'en concevant d'autres LBDs spécifiques de différentes molécules, une intégration dans ce genre de construction permettra une détection aisée.

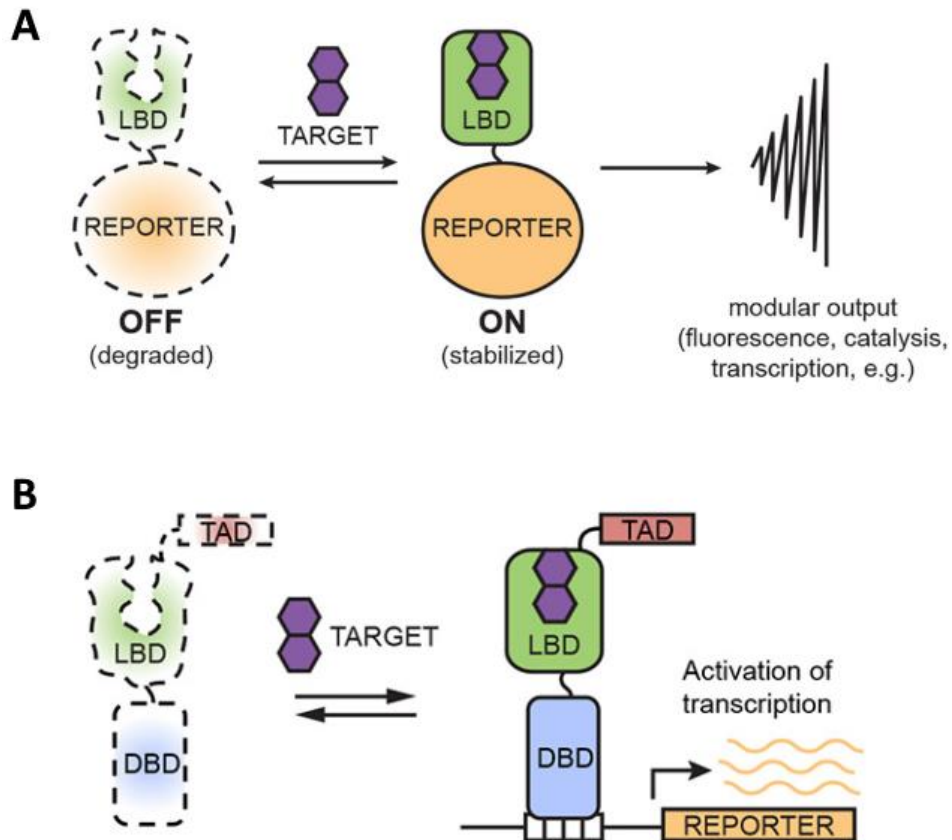


Figure 18 : Deux stratégies de biosenseurs basés sur la stabilité. A) Une protéine rapporteuse (Reporter) est reliée à un domaine de liaison à un ligand (LBD). La fixation du ligand stabilise la structure et donne un signal mesurable. B) Le LBD est situé entre deux domaines. Un Domaine de liaison à l'ADN (DBD) en N-terminal et un domaine d'activation de la transcription (TAD) en C-terminal. La fixation du ligand stabilise le biosenseur et entraîne la transcription du gène rapporteur (protéine fluorescente, enzyme, ...). (Feng, J., et al., 2015)

C. Contrôle des fonctions par la lumière

Il existe des domaines sensibles à la lumière et changeant de conformation en présence de celle-ci. Le plus utilisé, le domaine LOV (*Light-oxygen-Voltage*) a été utilisé de nombreuses fois pour créer ce que l'on appelle des interrupteurs photoactivables. Ce domaine est capable d'interagir dans le noir avec une hélice appelé $J\alpha$. En présence de lumière bleue (450 à 500 nm) l'hélice $J\alpha$ se déplie et ne peut plus interagir avec le domaine LOV. Ce domaine couplé à son hélice $J\alpha$ offre de nombreuses possibilités d'ingénierie multidomaines.

En 2008, J. Lee *et al.* ont inséré ce domaine LOV et son hélice $J\alpha$ au sein d'une DHFR. Les auteurs ont ainsi testé plusieurs positions d'insertion, l'objectif étant d'activer l'enzyme en présence de lumière (Figure 19-A). Les résultats montrent qu'il est effectivement possible, avec l'une des constructions, d'avoir une activité enzymatique deux fois supérieure en présence de lumière. (Lee, J., *et al.*, 2008)

En 2009, Y.I. Wu *et al.* ont fusionné un domaine Rac1, impliqué dans la mobilité cellulaire, au domaine LOV et son hélice $J\alpha$ (Figure 19-B). La proximité de Rac1 et du domaine LOV empêche Rac1 d'interagir avec ses effecteurs et réduit presque à néant la motilité des cellules. La présence de lumière permet d'entraîner un déplacement des cellules vers la zone excitée. Ces résultats prouvent que le contrôle de la motilité cellulaire par la lumière est possible. (Wu, Y.I., *et al.*, 2009)

Une approche intéressante a été mise en place par D. Strickland *et al.* En 2008, les auteurs ont fusionné le domaine LOV et son hélice $J\alpha$ à un facteur de transcription TrpR (*Tryptophane Repressor*). TrpR ne se lie à l'ADN qu'en présence de Tryptophane (Trp). L'originalité de l'approche réside dans le fait que l'hélice $J\alpha$ a été modifiée pour être dans la continuité d'une hélice du domaine TrpR. Ainsi, lorsque l'hélice $J\alpha$ interagit avec le domaine LOV elle ne peut interagir avec le domaine TrpR. Seule la dissociation de l'hélice en présence de lumière permet son interaction avec le domaine TrpR qui peut alors se lier à l'ADN (Figure 19-C). Ils ont appelé cette construction LovTAP pour "*LOV- and Tryptophan-Activated Protein*". L'affinité de ce LovTAP pour l'ADN augmente d'un facteur 5 après une excitation à la lumière bleue. (Strickland, D., *et al.*, 2008) Les auteurs ont par la suite tenté d'augmenter ce facteur. Pour cela, ils ont stabilisé l'interaction entre l'hélice $J\alpha$ et le domaine LOV en introduisant des mutations dans l'hélice. Le choix de ces mutations a été effectué par *design computationnel* en favorisant les résidus impliqués dans les hélices et les interactions électrostatiques. L'introduction de deux mutations au sein de l'hélice a permis de diminuer l'affinité du LovTAP pour l'ADN d'un facteur 9 en absence de lumière sans pour autant diminuer l'affinité de la chimère après une excitation lumineuse. L'augmentation globale d'affinité entre obscurité et lumière avec ces deux mutations est de 70 permettant ainsi d'augmenter par la même occasion la sensibilité de la chimère. (Strickland, D., *et al.*, 2010)

Enfin, en 2014 J.J. Yi *et al.* ont tenté diverses fusions d'un peptide inhibiteur de la PKA (PKI) à l'hélice $J\alpha$ afin de contrôler l'inhibition grâce à la lumière (Figure 19-D). Une seule des constructions s'est révélée efficace pour inhiber la PKA après une illumination. Cette protéine est appelée PA-PKI pour "*Photo-activable PKI*". Les auteurs ont alors essayé de remplacer le PKI par un MKI (*Myosin light chain kinase (MLCK) inhibitor*). La construction ainsi obtenue, PA-MKI, s'est également révélée efficace pour contrôler l'activité de la MLCK. (Yi, J.J., *et al.*, 2014)

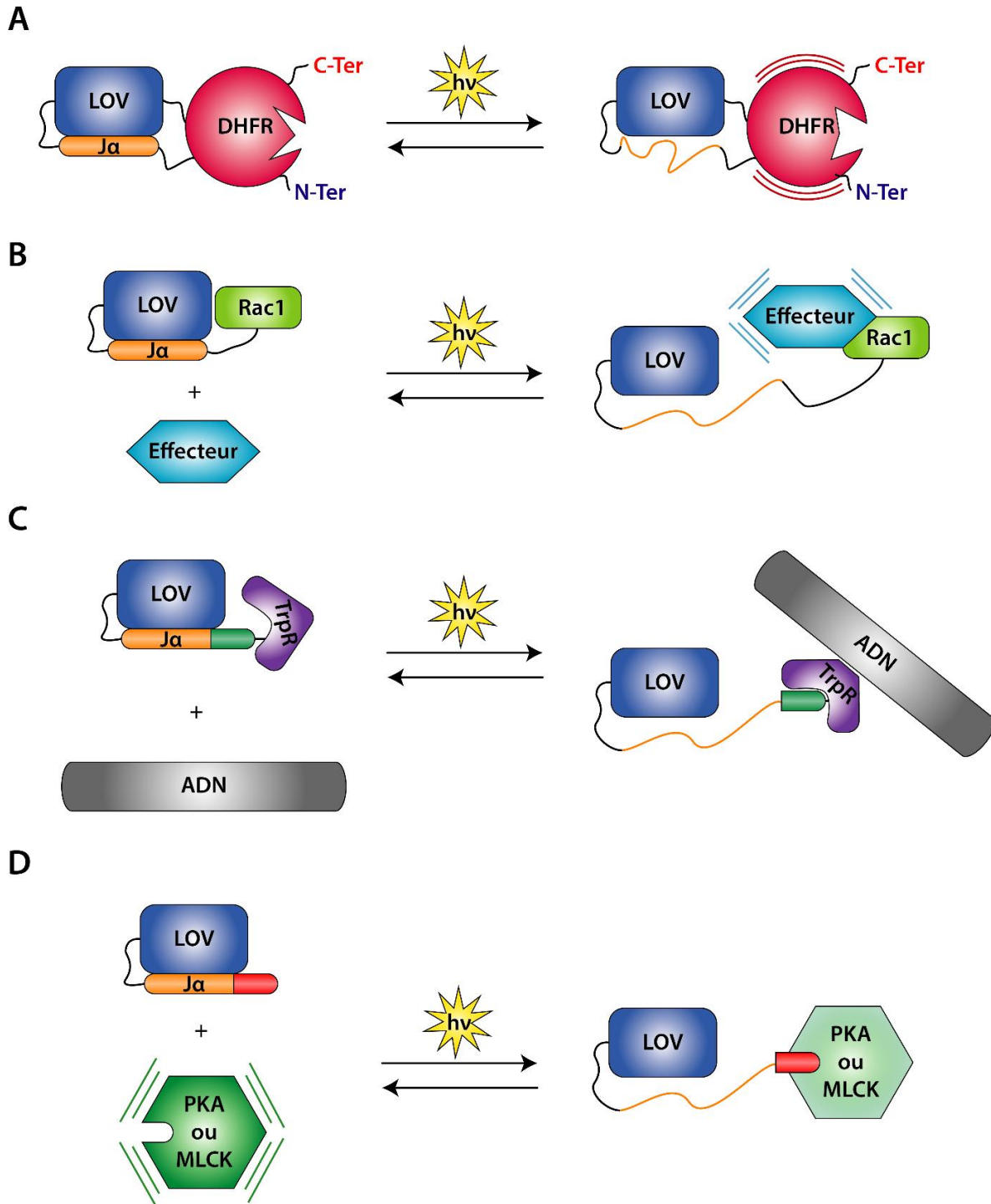


Figure 19 : Représentation schématique des possibilités du domaine LOV.

A) Couplage de la DHFR avec permutation circulaire. L'excitation lumineuse permet d'activer la DHFR. (Lee, J., et al., 2008)

B) Contrôle de la motilité cellulaire via le domaine Rac1 qui ne peut interagir avec ses effecteurs qu'après excitation lumineuse. (Wu, Y.I., et al., 2009)

C) Contrôle de la transcription par fusion de l'hélice Ja à une hélice du domaine TrpR. L'excitation lumineuse rend l'hélice du domaine accessible et permet "d'activer" le domaine TrpR. (Strickland, D., et al., 2010)

D) Inactivation des protéines PKA et MLCK par couplage de l'hélice Ja à des peptides inhibiteurs de ces deux protéines. L'excitation lumineuse libère les peptides inhibiteurs qui peuvent alors se fixer sur la PKA ou la MLCK. (Yi, J.J., et al., 2014)

D. Des multidomains pour aider à l'expression et à cristallisation

1) Expression

La production de protéines solubles et actives a longtemps été problématique. Il a fallu attendre l'utilisation de partenaires de fusion pour produire efficacement certaines protéines. Parmi ces partenaires, trois sont particulièrement connus, thiorédoxine (TRX), Glutathion S-transférase (GST) et la *maltose binding protein* (MBP). En 1999, une étude portant sur les effets de ces trois protéines en tant que partenaires de fusion a montré que la MBP permettait d'une part d'augmenter la part de protéines solubles mais aussi agissait sur le bon repliement des protéines fusionnées (Kapust, R.B. and Waugh, D.S., 1999).

2) Cristallisation

Les protéines capables de reconnaître des cibles diverses (Partie A - III) sont déjà connus comme aidant à la cristallisation de protéine au comportement difficile (i.e. Flexibilité trop importante, trop peu de contacts cristallins). Parmi ces protéines on peut citer les dérivés d'anticorps (Koide, S., 2009), les DARPins (Sennhauser, G. and Grutter, M.G., 2008) ou bien encore les α Reps (Valerio-Lepiniec, M., et al., 2015). Pour les protéines récalcitrantes, le fait de rajouter un linker la reliant à son interacteur s'est révélé être une solution efficace.

A. Batyuk *et al.* ont ainsi fusionné des DARPins à une β -lactamase *via* une hélice α . Cette hélice relie la DARPins en N-ter et la β -lactamase en C-ter de manière rigide (Figure 20). L'ajout de résidus au sein de l'hélice partagé entre les deux protéines permet de faire varier l'angle de liaison de 100 degrés par résidus tout en translatant la β -lactamase d'1.5Å le long de l'hélice. L'ajout de la β -lactamase qui est une protéine rigide et le fait de pouvoir choisir l'angle entre la DARPins et la β -lactamase permet en plus d'augmenter la surface de cristallisation, d'ajouter un paramètre de criblage pour la formation des cristaux en fonction de leur maille. (Batyuk, A., et al., 2016)

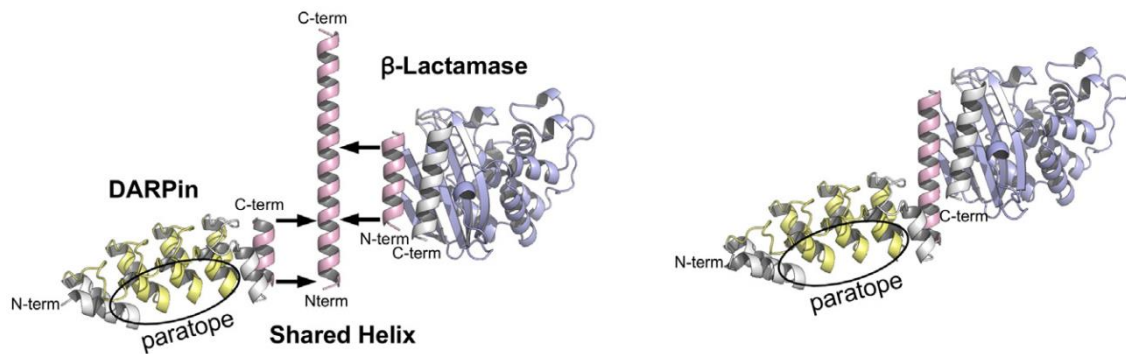


Figure 20 : Fusion des DARPins et de la β -lactamase via une hélice alpha rigide dans le but de faciliter la cristallisation de protéine reconnue par le paratope de la DARPins. (Batyuk, A., et al., 2016)

L'utilisation de protéines interactrices en tant « chaperonne » pour aider la cristallisation laisse supposer que des sélections pour obtenir les interacteurs sont possibles. Ces sélections nécessitent une quantité non négligeable de protéine d'intérêt, ce qui n'est pas toujours possible. Pour cette raison, une méthode innovante a été développée par A. Chevrel *et al* au laboratoire. Cette approche consiste à effectuer des sélections *in vivo* via l'utilisation de la PCA sur la DHFR. Elle permet de sélectionner parmi une bibliothèque de protéines artificielles, appelées α Reps, celles interagissant avec la cible. Pour cela, les protéines α Reps ont été insérées en aval de la partie N-ter de la DHFR. La cible elle a été fusionnée en amont ou en aval de la partie C-ter de la DHFR. Le vecteur codant pour la protéine d'intérêt est inséré par transformation dans les bactéries. Les vecteurs codant pour l'ensemble des α Reps sont insérés grâce à des phages. Seules les bactéries dans lesquelles la DHFR est reconstituée (c.a.d. Au sein desquelles l' α Reps exprimée interagit avec la cible) pourront croître en présence de triméthopime. Par la suite les interacteurs obtenus sont fusionnés par un linker aux protéines d'intérêt, exprimés, purifiés et cristallisés. Cette méthode a permis la sélection d'interacteurs et la résolution des structures des protéines ComD et HP1473, deux protéines difficilement produites ET non cristallisables seules. (Chevrel, A., et al., 2018)

Cette approche présente l'avantage d'être généralisable à d'autres bibliothèques de protéines artificielles. De plus, elle peut être modifiée afin de pouvoir réaliser des sélections sur des cibles non exprimables en bactéries dans des levures par exemple.

E. Création d'avidité par clamp

Une dernière approche intéressante créée par ingénierie multidomaines est la création d'avidité. S. Hansen *et al.* ont créé des interacteurs avec des K_D pico molaire pour la GFP à partir de deux Darpins ayant un K_D pour la GFP de l'ordre de la centaine de nanomolaires. Pour se faire, les auteurs sont partis de deux DARPins interagissant avec la GFP. Ces interacteurs se superposant en partie, un des deux binder a été tronqué. Le design de cette forme tronquée a été effectué *in silico* à l'aide du logiciel Rosetta. La DARPins non tronquée et la DARPins tronquée ont été fusionnées entre elles. La caractérisation de la structure et les mesures des affinités montrent bien que les deux domaines interagissent avec la cible et cela avec un K_D d'environ 20 pM. (Figure 21) (Hansen, S., *et al.*, 2017)

Cette méthode pourrait être généralisable à la création d'interacteurs de haute affinité pouvant inhiber des enzymes ou pour faire de l'immobilisation de cible.

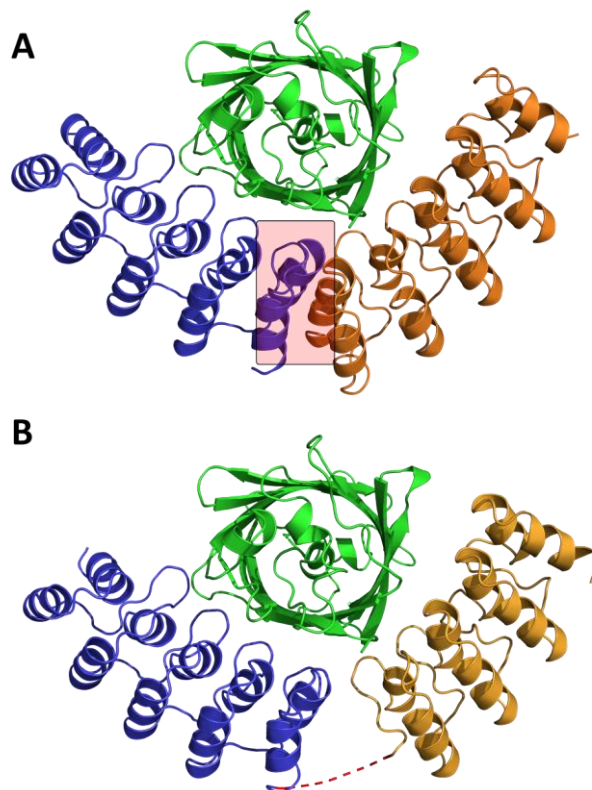


Figure 21 : Structure cartoon, la GFP est visible en vert. A) Superposition des deux DARPins (orange et bleue). Le clash est mis en évidence par le cadre rouge. B) Clamp-GFP des deux DARPins fusionnées, le linker qui est non structuré est visible en pointillé rouge

V. Contexte scientifique

A. Les α Reps

Au sein de l'équipe Modélisation et Ingénierie des Protéines (MIP), nous savons créer des protéines artificielles possédant des propriétés de reconnaissance semblables à celles des anticorps. Dans le cadre de ma thèse, j'ai principalement travaillé avec une famille de protéines artificielles appelées les α Reps.

1) À l'origine des α Reps : les protéines HEAT

Les α Reps sont des protéines artificielles qui ont été conçues en se basant sur des protéines « HEAT *repeat* ». Les protéines HEAT constituent une grande famille basée sur des répétitions du motif structural HEAT. Elles ont dans un premier temps été identifiées chez les eucaryotes puis par la suite chez les procaryotes. Le nom de ce motif est un acronyme pour quatre protéines contenant ce dernier : Huntingtine, Facteur d'élongation 3 (EF3), protéine phosphatase 2A (PP2A), et la kinase de levure TOR1. Ce motif est constitué de deux hélices α reliées par une courte boucle. L'assemblage en tandem de ces motifs HEAT donne généralement des protéines formant un solénoïde qui peut se prolonger sous la forme d'une super-hélice droite. La succession des motifs permet d'obtenir une surface concave *via* laquelle ont généralement lieu des interactions protéine / protéine. (Andrade, M.A. and Bork, P., 1995)

2) Design de la séquence consensus

L'une des protéines de ce groupe se nomme Mth187 et provient de *Methanobacterium thermoautotrophicum* (Archaea). Sa structure fut résolue en 2006 et révéla la présence d'un domaine N-terminal non structuré suivi d'un domaine C-terminal constitué de trois paires d'hélice α formant des motifs HEAT de 30-31 résidus (Figure 22). (Julien, O., et al., 2006)

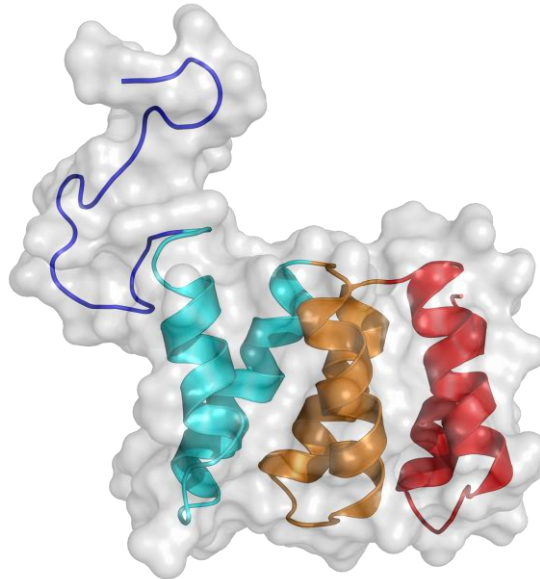


Figure 22 : Structure de Mth187. En bleu, domaine N-terminal non replié. En cyan, motif HEAT N-terminal. En orange, motif HEAT interne. En rouge, motif HEAT C-terminal.

Cette protéine fut utilisée en tant que base pour le design de la séquence consensus des α Reps. Afin de définir la séquence consensus, un BLAST (*Basic Local Alignment Search Tool*) a été effectué contre les deux premiers motifs de Mth187. Le dernier motif présente de grosses différences qui peuvent s'expliquer par son exposition au solvant et il fut donc exclu du BLAST. Ce premier BLAST permis d'obtenir une première séquence consensus à partir des protéines les plus proches de Mth187. Un second BLAST fut alors réalisé en utilisant comme séquence d'entrée la première séquence consensus répétée cinq fois. Une seconde séquence consensus fut alors obtenue à partir des 100 protéines les plus proches et comptabilisant 500 motifs différents.

N-Ter - GDERAVEPLIKALKDEDXXVRXXAAXALGXI - C-Ter

Cette séquence consensus permet de mettre en évidence des positions conservées, probablement impliquées dans la structure, et des positions hypervariables. Cette hétérogénéité de conservation est visible grâce à la représentation en logo séquence (Figure 23-A). Lorsque les positions conservées sont placées sur la structure de Mth187, trois groupes peuvent être observés. Un groupe de résidus apolaires enfouis entre les deux hélices et contribuant aux interactions intra-hélice et au *stacking* des motifs : **A5, L9, L13, V20, A24, A25, A27** et **L28** (Figure 23-B). Deux résidus, **R21** et **D15**, entre lesquels s'établit une interaction électrostatique stabilisant l'interaction entre deux motifs (Figure 23-C). Deux glycine **G1** et **G29** qui permettent la réalisation de coude entre motif. Les positions variables peuvent quant à elle se diviser en deux groupes : d'une part, les positions non enfouies de la première hélice (3, 4, 7 et 11) constituées de résidus aux chaînes polaires ou stabilisant les hélices (K, R, E, Q et A), d'autre part, les positions non enfouies de la seconde hélice, qui sont les plus variables et localisées en **18, 19, 22, 23, 26** et **30** (Figure 23-D). Ce sont ces dernières positions qui ont été choisies pour créer la surface d'interaction des α Reps.

Afin d'assurer une bonne stabilité des ossatures à motifs répétés, le design du premier et du dernier motif, appelé N-Cap et C-Cap, est très important. Bien que le repliement de ces deux motifs soit similaire, les positions exposées au solvant sont différentes et leur conception ne se base pas entièrement sur la séquence consensus. Ainsi, le C-Cap de Mth187 étant bien défini, il a été utilisé tel quel pour concevoir la bibliothèque. Le domaine N-terminal de Mth187, en revanche, n'est pas replié et n'a donc pas été utilisé pour concevoir le N-Cap. Le design de ce dernier s'est donc fait en recherchant des protéines homologues à Mth187 et possédant un N-Cap bien défini. La protéine retenue (identifiant UniProtKB : A0B7C6) possède une forte homologie avec Mth187 et un N-Cap de 34 résidus. De par ces ressemblances, les positions variables identifiées sur la séquence consensus ont pu être transposées à ce N-Cap. (Urvoas, A., *et al.*, 2010)

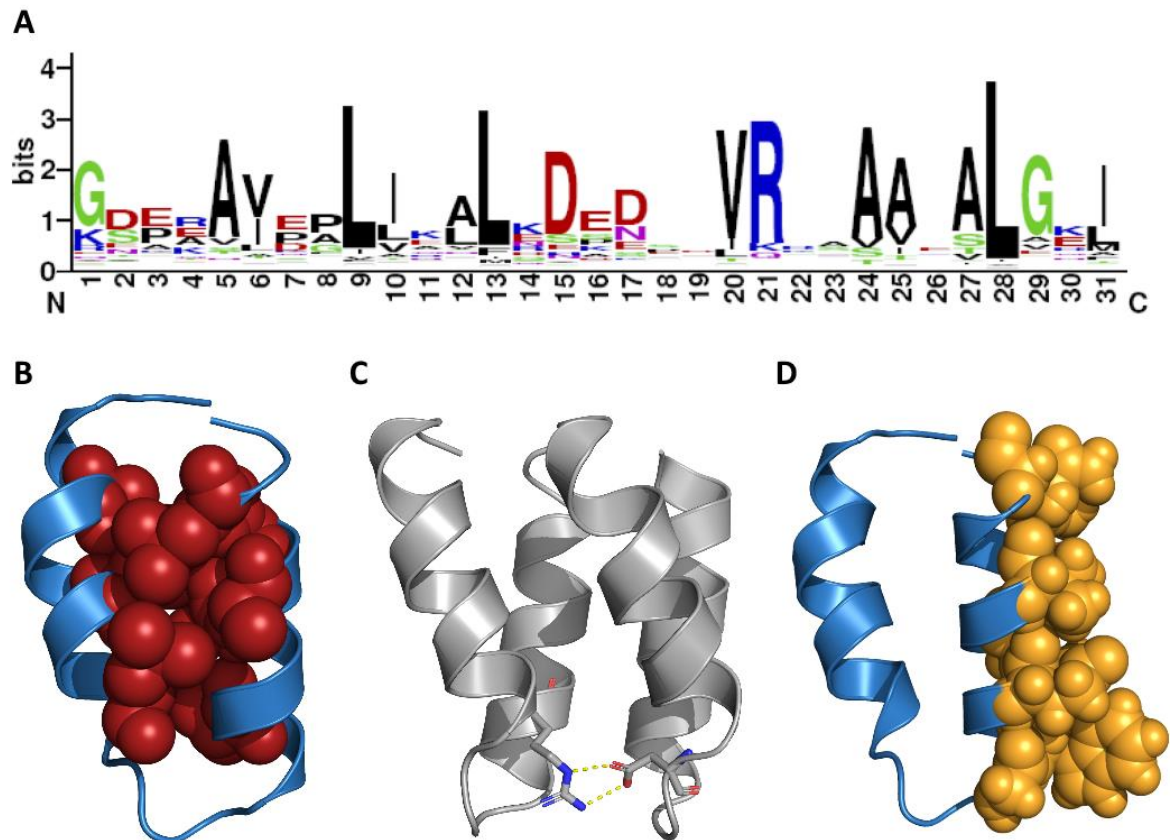


Figure 23 : A) Représentation en logo-séquence de la séquence consensus obtenue par deux BLASTs successifs.

B) Motif HEAT en représentation cartoon avec mise en évidence des résidus constants (sphère rouge) stabilisant les interactions entre les hélices d'un même motif.

C) Liaisons hydrogènes conservées entre l'arginine 21 du motif N et le groupement carboxyle de l'aspartate 15 du motif N+1.

D) Motif HEAT en représentation cartoon avec mise en évidence des résidus variables de la seconde hélice en sphère jaune.

3) Construction de la bibliothèque AReps

a) Bibliothèque 1.0

La séquence consensus, le N-Cap et le C-Cap définis, la construction d'une bibliothèque dite, de première génération, a pu être réalisée. Celle-ci avait pour objectifs de vérifier la viabilité de l'ossature conçue, la tolérance à la variabilité des positions variables et les propriétés d'interactions des α Reps. Pour construire cette bibliothèque, une nouvelle approche basée sur la polymérisation de microgènes a été mise au point. Chaque microgène correspond à un oligonucléotide codant pour un seul motif. Chaque oligonucléotide a les positions variables codées par des codons dégénérés. Les proportions des codons dégénérés est cependant biaisée et permet d'obtenir notamment plus de résidus impliqués dans les interactions (Tyrosine et Tryptophane) et moins de résidus pouvant perturber les structures (Proline et Cystéine). Le processus de polymérisation passe en premier lieu par une circularisation des microgènes qui sont ensuite amplifiés par *Rolling Cycle Amplification* (RCA), puis digérés par l'enzyme de restriction BsmBI. La polymérisation des gènes est effectuée en présence d'un plasmide digéré contenant le C-Cap et d'un excès d'oligonucléotide codant pour les N-Cap. Les différents oligonucléotides (N-Cap et motif) peuvent se greffer sur le plasmide à partir du C-Cap. L'incorporation d'un N-Cap entraîne la fin de la polymérisation. Les plasmides obtenus sont alors différents en nombre de motifs mais également en séquence pour chaque motif et N-Cap. (Figure 24)

La bibliothèque ainsi obtenue (Lib1.0) contient 3×10^8 clones indépendants. Le nombre de motif interne (entre le N-Cap et le C-Cap) varie de 0 à 10 avec une propension supérieure contenant 2 à 5 motifs internes. Cependant au sein de Lib1.0, seulement 33% des clones sont solubles. Cette faible solubilité est due d'une part à un problème de décalage de phase et d'autre part au fait que de nombreux clones n'ont pas de motifs internes et sont insolubles. Cependant, les clones isolés et solubles présentent une stabilité thermique importante avec des T_m supérieurs à 70°C et des rendements allant de 10 à 90 mg par litre de culture. La structure d'un clone a été résolue par cristallographie et confirme le repliement des motifs HEAT en double hélices antiparallèles. (Figure 25) (Urvoas, A., *et al.*, 2010)

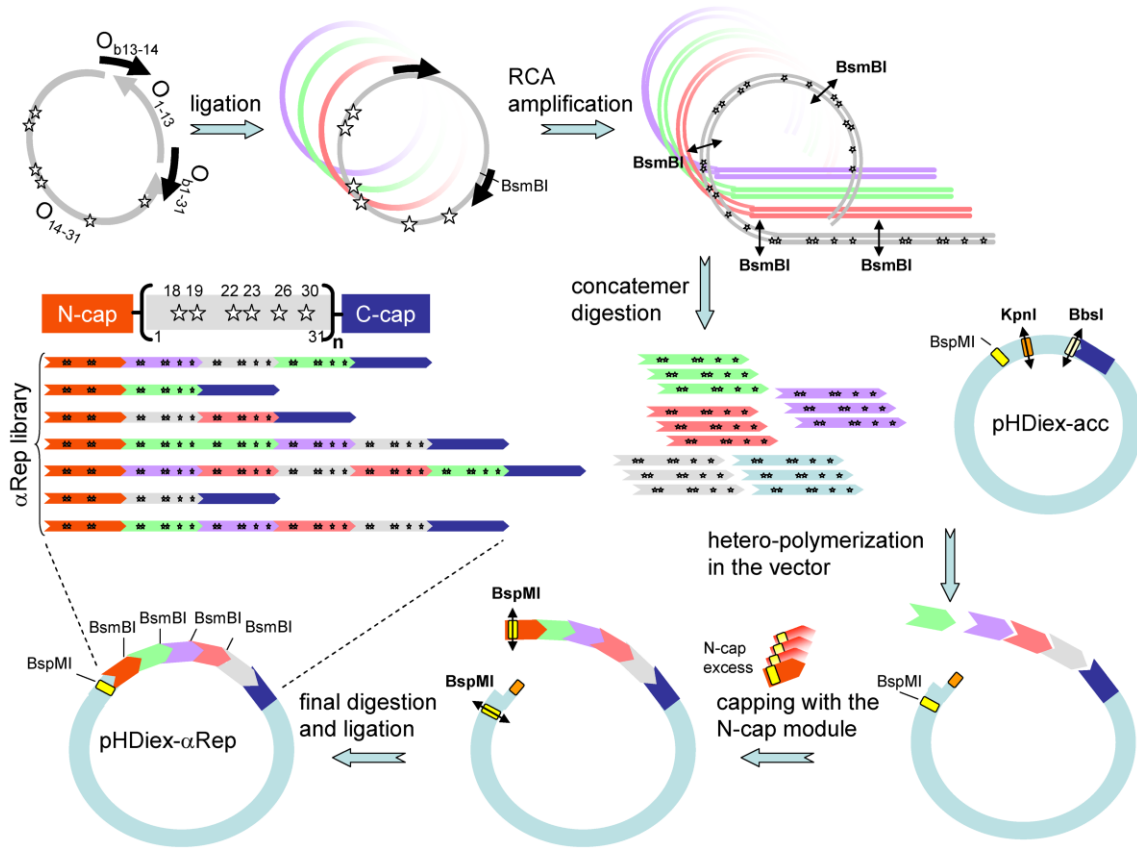


Figure 24 : Conception de la bibliothèque alphaRep 1.0. Des microgènes dégénérés (aux positions variables) et codants chacun pour un motif sont amplifiés par RCA. La polymérisation des vecteurs est ensuite effectuée par digestion enzymatique et ligation en présence un excès de motif N-Cap qui lorsqu'il est inséré arrête la polymérisation. Les vecteurs finaux obtenus sont appelés pHDiex- α Rep.

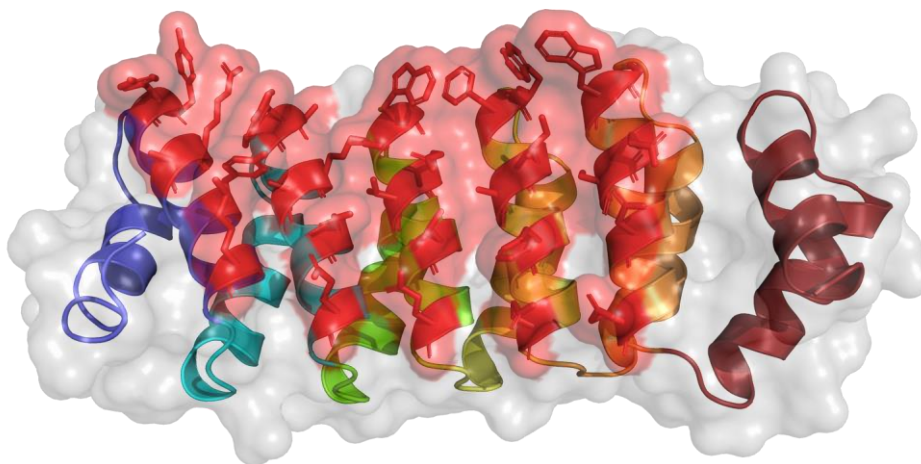


Figure 25 : Structure d'une α Rep à quatre motifs internes représentée en cartoon prélevée dans Lib1.0. N-Cap en bleu. Motifs internes en cyan, vert, jaune et orange. C-Cap en rouge foncé. Les positions variables sont représentées en stick rouge. La surface de la protéine est visible en transparence.

b) Bibliothèques 2.0 et 2.1

La première bibliothèque Lib1.0 possédait plusieurs défauts. Tout d'abord, au niveau des séquences avec une variabilité des positions variables biaisées entraînant une diminution de la diversité possible de la bibliothèque. D'autre part, un taux non négligeable de séquences présente des décalages de phases provenant de l'assemblage des oligonucléotides ou de leur synthèse. Cela contribue à diminuer le nombre total de clones exprimables et par la même occasion l'espace des séquences utilisables. L'optimisation de la bibliothèque était donc nécessaire.

Pour résoudre le problème de variabilité, la proportion des différents résidus aux positions variables a été corrigée pour refléter la répartition naturelle des résidus à ces positions dans les protéines PBS HEAT-like. Une seconde bibliothèque appelée Lib2.0 a été créée en corrigeant ce défaut. Lib2.0, possède 2.7×10^8 clones indépendants mais présente cependant 20% de séquence non codante et présente un nombre moyen de motifs internes d'environ 0.8. Or la variation de taille des surfaces d'interaction et leur augmentation est d'un intérêt majeur afin d'augmenter les cibles potentielles de la bibliothèque d'interacteurs α Reps.

Pour corriger le second problème une nouvelle bibliothèque a été construite en partant d'un échantillon codant des motifs corrects provenant des clones de Lib2.0. Ces motifs ont ensuite été recombinaisonnés pour donner une troisième bibliothèque Lib2.1. Cette troisième bibliothèque possède 1.7×10^9 clones indépendants, le taux de séquence en phase est de plus de 90% et le nombre moyen de motifs passe à environ 2.3. Le nombre de motif interne varie de 0 à 14 augmentant la variabilité de taille de cette bibliothèque. Les températures de dénaturation (T_m) de ces protéines est supérieur à 75°C. (Guellouz, A., *et al.*, 2013)

4) Reconnaissance *in vivo*

Des sélections contre la eGFP ont été effectuées par phage display et ont permis d'obtenir trois interacteurs α Reps (appelées bGFPA, bGFPC et bGFPD) interagissant spécifiquement avec cette dernière. Afin de vérifier si les α Reps sont fonctionnelles dans des cellules eucaryotes vivantes, une eGFP a été fusionnée à une séquence d'adressage à différents compartiments cellulaires, les interacteurs α Reps ont quant à eux été fusionnés à une RFP (*Red Fluorescent Protein*). Les plasmides permettant l'expression de ces deux protéines de fusion ont été transfectés dans des cellules HEK (Human Embryonic Kidney). L'observation au microscope confocal a montré qu'il y avait une co-localisation des deux protéines de fusion et cela peu importe le compartiment cellulaire. Les α Reps peuvent donc être exprimées et sont fonctionnelles *in vivo*, dans des cellules eucaryotes vivantes contrairement à des anticorps classiques. (Figure 26) (Chevrel, A., *et al.*, 2015)

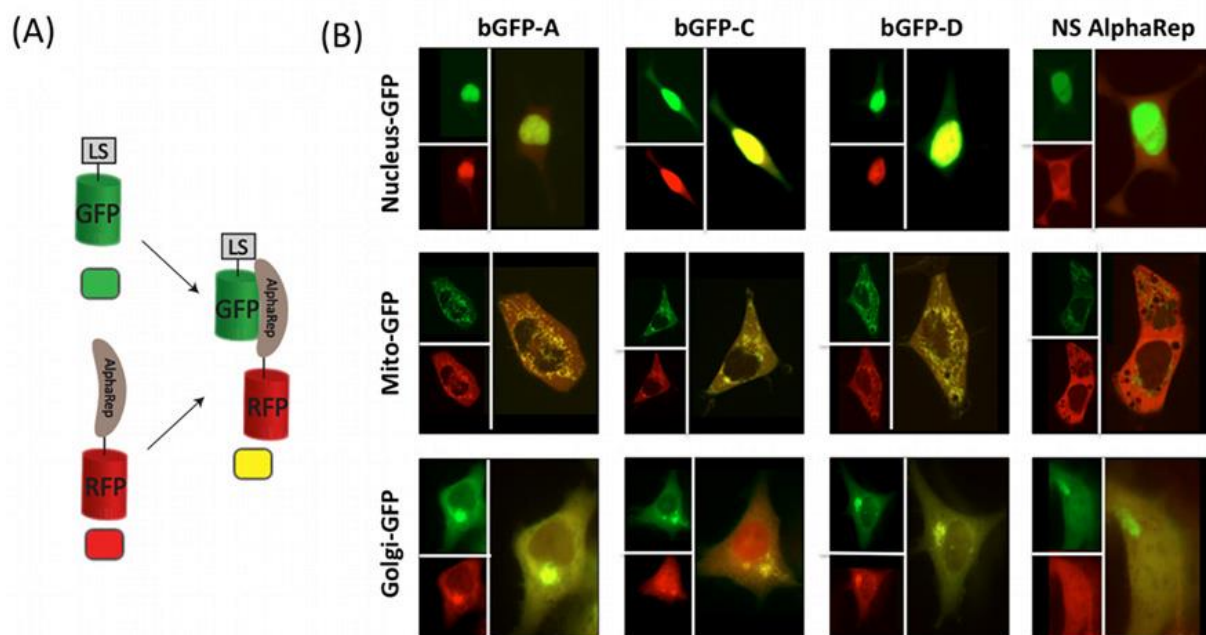


Figure 26 : A) Schéma des protéines de fusion produites. LS : Séquence de localisation. La couleur attendue au microscope confocal est représentée par un rectangle sous chaque construction. B) Images confocales de cellules co-exprimant les constructions avec les trois interacteurs de la eGFP et une α Rep non spécifique (NS). L'émission de lumière jaune témoigne de la colocalisation des α Reps-RFP et de la eGFP.

5) Sélection d' α Reps contre des cibles définies

Des sélections d' α Reps contre diverses cibles définies sont possibles par phage display et par PCA. Chacune des bibliothèques a pu être testée contre différentes cibles et les structures d'une dizaine de complexes ont pu être résolues (Figure 27). Pour toutes les structures, les résidus hypervariables sont ceux interagissant avec les cibles, confirmant le mode d'interaction des α Reps. Les constantes de dissociation des complexes α Reps / cibles vont du micro au nanomolaire. Ces sélections témoignent de la viabilité des bibliothèques α Reps pour l'obtention d'interacteurs spécifiques.

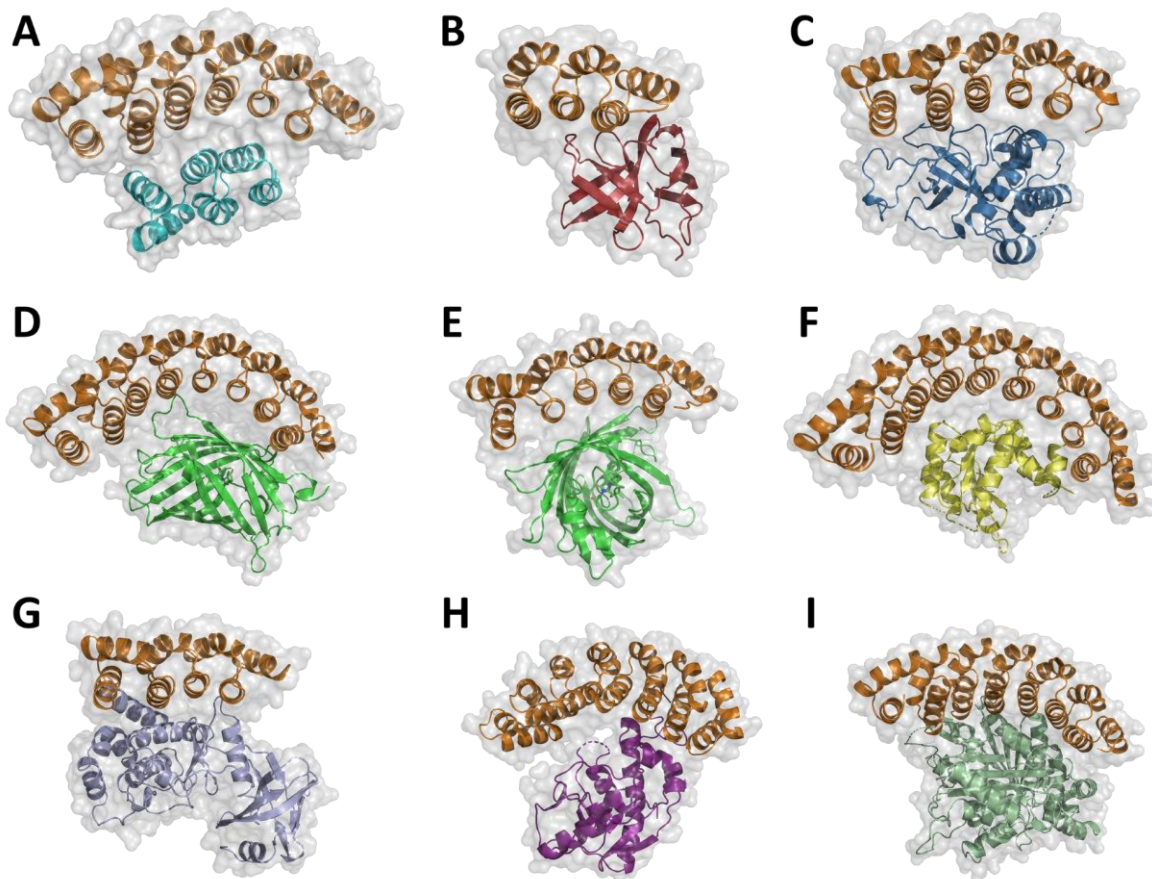


Figure 27 : Structure des α Reps (orange) et de leurs différentes cibles en cartoon. A) bA3-2. B) NCS. (Guellouz, A., et al., 2013) C) FNE. (Tiouajni, M., et al., 2014) D et E) GFP. (Chevrel, A., et al., 2015) F) Octarelline. (Figuroa, M., et al., 2016) G) YabT. (Chevrel, A., et al., 2018) H) HP173. (Sous presse) I) Tubuline. (Sous presse)

B. Les α Reps comme briques élémentaires

L'objectif principal de ma thèse était de voir s'il était possible de créer de nouvelles fonctions basées sur la reconnaissance moléculaire en utilisant des α Reps comme briques élémentaires. Contrairement à la plupart des anticorps et dérivés d'anticorps, les α Reps présentent des propriétés remarquables en termes de production et de stabilité et sont exprimables dans le cytoplasme de cellules eucaryotes. Les α Reps se prêtent donc à être utilisées comme des briques élémentaires en vue d'une ingénierie modulaire à l'image des processus ayant émergé au cours de l'évolution pour augmenter la taille du répertoire protéique.

Deux α Reps élémentaires ont donc été choisies pour concevoir des constructions bidomaines. La première, appelée **A3**, forme un dimère en solution pouvant se dissocier grâce à une deuxième α Rep, **bA3-2**. La troisième α Rep est appelé **bGFPD** et a été sélectionné contre la **eGFP** par phage display. La caractérisation biophysique de ces complexes a été effectuée. La stabilité de ces α Reps est très élevée et elles présentent des T_m supérieurs à 80°C. Les constantes de dissociation de ces complexes sont de l'ordre du nanomolaire. (Guellouz, A., et al., 2013; Chevrel, A., et al., 2015)

La première partie de ce projet de thèse a consisté à construire puis étudier les propriétés biophysiques des deux protéines bidomaines basées sur les α Reps (**A3_A3** et **A3_bGFPD**) afin de mieux comprendre les comportements adoptés par de telles constructions. Cette caractérisation a été effectuée par calorimétrie différentielle à balayage (DSC), chromatographie d'exclusion stérique couplée à de la diffusion de lumière à angles multiples (SEC-MALS), titration calorimétrique isotherme (ITC), résonance plasmonique de surface (SPR), ultracentrifugation analytique (UCA), diffraction des rayons X à petits angles (SAXS) et par cristallographie aux rayons X. Outre l'aspect fondamental de cette question, cette étude a donné « les règles » pour moduler de façon contrôlée les interactions entre ces protéines. Elle a également permis de prouver qu'il était possible de détecter la fixation de ligand par fluorescence.

La seconde partie de ce projet a consisté à concevoir, à partir des protéines bidomaines précédemment étudiées, de nouveaux biosenseurs basés sur le FRET (*Förster Resonance Energy Transfer*) pouvant être utilisés *in vivo* et *in vitro*. Cette deuxième partie présente deux biosenseurs avec des limites de détection en milieu complexe de l'ordre du nanomolaire. Les α Reps utilisées dans ces constructions pouvant être changées en fonction de la cible souhaitée, il s'agit ici d'une preuve de concept pouvant être généralisée à n'importe quelle cible.

Enfin une dernière partie s'est portée sur la conception et l'étude de nouveaux biosenseurs « génétiquement » codables. Dans un premier temps, l'optimisation du LBT (*Lanthanide Binding Tag*), une séquence peptidique fixant le Terbium, a été effectuée pour rendre sa fluorescence dépendante de l'ouverture des bidomaines. Dans un second temps, un biosenseur basé sur une inhibition de la nanoKaz dépendant ici aussi de l'ouverture des bidomaines a été conçu. Les résultats obtenus montrent qu'il est possible de créer de tels biosenseurs, mais une optimisation reste encore nécessaire pour améliorer leur spécificité et leur capacité de détection.

VI. Bibliographie

- Adams, S. R., A. T. Harootunian, Y. J. Buechler, S. S. Taylor, and R. Y. Tsien. 1991. 'Fluorescence ratio imaging of cyclic AMP in single cells', *Nature*, 349: 694-7.
- Andrade, M. A., and P. Bork. 1995. 'HEAT repeats in the Huntington's disease protein', *Nat Genet*, 11: 115-6.
- Arbabi Ghahroudi, M., A. Desmyter, L. Wyns, R. Hamers, and S. Muyldermans. 1997. 'Selection and identification of single domain antibody fragments from camel heavy-chain antibodies', *FEBS Letters*, 414: 521-26.
- Arnold, F. H. 2018. 'Directed Evolution: Bringing New Chemistry to Life', *Angew Chem Int Ed Engl*, 57: 4143-48.
- Baier, F., J. N. Copp, and N. Tokuriki. 2016. 'Evolution of Enzyme Superfamilies: Comprehensive Exploration of Sequence-Function Relationships', *Biochemistry*, 55: 6375-88.
- Baird, G. S., D. A. Zacharias, and R. Y. Tsien. 1999. 'Circular permutation and receptor insertion within green fluorescent proteins', *Proceedings of the National Academy of Sciences*, 96: 11241-46.
- Batyuk, A., Y. Wu, A. Honegger, M. M. Heberling, and A. Pluckthun. 2016. 'DARPin-Based Crystallization Chaperones Exploit Molecular Geometry as a Screening Dimension in Protein Crystallography', *J Mol Biol*, 428: 1574-88.
- Becker, M.L., R. Near, M. Mudgett-Hunter, M.N. Margolies, R.T. Kubo, J. Kaye, and S.M. Hedrick. 1989. 'Expression of a hybrid immunoglobulin-T cell receptor protein in transgenic mice.', *Cell*, 58: 911-21.
- Better, M., C. Chang, R. Robinson, and A. Horwitz. 1988. 'Escherichia coli secretion of an active chimeric antibody fragment', *Science*, 240: 1041-43.
- Bhattacharyya, R. P., A. Remenyi, B. J. Yeh, and W. A. Lim. 2006. 'Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits', *Annu Rev Biochem*, 75: 655-80.
- Binz, H. K., P. Amstutz, A. Kohl, M. T. Stumpp, C. Briand, P. Forrer, M. G. Grutter, and A. Pluckthun. 2004. 'High-affinity binders selected from designed ankyrin repeat protein libraries', *Nat Biotechnol*, 22: 575-82.
- Binz, H. K., P. Amstutz, and A. Pluckthun. 2005. 'Engineering novel binding proteins from nonimmunoglobulin domains', *Nat Biotechnol*, 23: 1257-68.
- Binz, H. Kaspar, Michael T. Stumpp, Patrik Forrer, Patrick Amstutz, and Andreas Plückthun. 2003. 'Designing Repeat Proteins: Well-expressed, Soluble and Stable Proteins from

- Combinatorial Libraries of Consensus Ankyrin Repeat Proteins', *Journal of Molecular Biology*, 332: 489-503.
- Boersma, Y. L., and A. Pluckthun. 2011. 'DARPin and other repeat protein scaffolds: advances in engineering and applications', *Curr Opin Biotechnol*, 22: 849-57.
- Chalfie, M., Y. Tu, G. Euskirchen, W. Ward, and D. Prasher. 1994. 'Green fluorescent protein as a marker for gene expression', *Science*, 263: 802-05.
- Chang, Chung- I., Bing-e Xu, Radha Akella, Melanie H. Cobb, and Elizabeth J. Goldsmith. 2002. 'Crystal Structures of MAP Kinase p38 Complexed to the Docking Sites on Its Nuclear Substrate MEF2A and Activator MKK3b', *Molecular Cell*, 9: 1241-49.
- Chen, T. S., and A. E. Keating. 2012. 'Designing specific protein-protein interactions using computation, experimental library screening, or integrated methods', *Protein Sci*, 21: 949-63.
- Chevrel, A., A. Mesneau, D. Sanchez, L. Celma, S. Quevillon-Cheruel, A. Cavagnino, S. Nessler, I. Li de la Sierra-Gallay, H. van Tilbeurgh, P. Minard, M. Valerio-Lepiniec, and A. Urvoas. 2018. 'A repeat proteins (α Rep) as expression and crystallization helpers', *J Struct Biol*, 201: 88-99.
- Chevrel, A., A. Urvoas, I. Li de la Sierra-Gallay, M. Aumont-Nicaise, S. Moutel, M. Desmadril, F. Perez, A. Gautreau, H. van Tilbeurgh, P. Minard, and M. Valerio-Lepiniec. 2015. 'Specific GFP-binding artificial proteins (α Rep): a new tool for in vitro to live cell applications', *Biosci Rep*, 35.
- Chmielewski, M., and H. Abken. 2015. 'TRUCKs: the fourth generation of CARs', *Expert Opin Biol Ther*, 15: 1145-54.
- Cobb, R. E., N. Sun, and H. Zhao. 2013. 'Directed evolution as a powerful synthetic biology tool', *Methods*, 60: 81-90.
- Copley, S. D. 2017. 'Shining a light on enzyme promiscuity', *Curr Opin Struct Biol*, 47: 167-75.
- Doerks, T., R. R. Copley, J. Schultz, C. P. Ponting, and P. Bork. 2002. 'Systematic identification of novel protein domain families associated with nuclear functions', *Genome Res*, 12: 47-56.
- Ekman, D., A. K. Bjorklund, J. Frey-Skott, and A. Elofsson. 2005. 'Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions', *J Mol Biol*, 348: 231-43.
- Eriksen, D. T., J. Lian, and H. Zhao. 2014. 'Protein design for pathway engineering', *J Struct Biol*, 185: 234-42.

- Feng, J., B. W. Jester, C. E. Tinberg, D. J. Mandell, M. S. Antunes, R. Chari, K. J. Morey, X. Rios, J. I. Medford, G. M. Church, S. Fields, and D. Baker. 2015. 'A general strategy to construct small molecule biosensors in eukaryotes', *Elife*, 4.
- Figuroa, M., M. Sleutel, M. Vandevenne, G. Parvizi, S. Attout, O. Jacquin, J. Vandenameele, A. W. Fischer, C. Damblon, E. Goormaghtigh, M. Valerio-Lepiniec, A. Urvoas, D. Durand, E. Pardon, J. Steyaert, P. Minard, D. Maes, J. Meiler, A. Matagne, J. A. Martial, and C. Van de Weerd. 2016. 'The unexpected structure of the designed protein Octarellin V.1 forms a challenge for protein structure prediction tools', *J Struct Biol*, 195: 19-30.
- Finney, H.M., A.D. Lawson, C.R. Bebbington, and A.N. Weir. 1998. 'Chimeric Receptors Providing Both Primary and Costimulatory Signaling in T Cells from a Single Gene Product', *J Immunol*, 161: 2791-97.
- Fleishman, S. J., T. A. Whitehead, D. C. Ekiert, C. Dreyfus, J. E. Corn, E. M. Strauch, I. A. Wilson, and D. Baker. 2011. 'Computational design of proteins targeting the conserved stem region of influenza hemagglutinin', *Science*, 332: 816-21.
- Furnham, N., N. L. Dawson, S. A. Rahman, J. M. Thornton, and C. A. Orengo. 2016. 'Large-Scale Analysis Exploring Evolution of Catalytic Machineries and Mechanisms in Enzyme Superfamilies', *J Mol Biol*, 428: 253-67.
- Geissler, D., S. Linden, K. Liermann, K. D. Wegner, L. J. Charbonniere, and N. Hildebrandt. 2014. 'Lanthanides and quantum dots as Forster resonance energy transfer agents for diagnostics and cellular imaging', *Inorg Chem*, 53: 1824-38.
- George, R.A., and J. Heringa. 2003. 'An analysis of protein domain linkers their classification and role', *Protein Engineering*, 15: 871-79.
- Gonzalez, E., H. Kulkarni, H. Bolivar, A. Mangano, R. Sanchez, G. Catano, R. J. Nibbs, B. I. Freedman, M. P. Quinones, M. J. Bamshad, K. K. Murthy, B. H. Rovin, W. Bradley, R. A. Clark, S. A. Anderson, J. O'Connell R, B. K. Agan, S. S. Ahuja, R. Bologna, L. Sen, M. J. Dolan, and S. K. Ahuja. 2005. 'The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility', *Science*, 307: 1434-40.
- Griss, R., A. Schena, L. Reymond, L. Patiny, D. Werner, C. E. Tinberg, D. Baker, and K. Johnsson. 2014. 'Bioluminescent sensor proteins for point-of-care therapeutic drug monitoring', *Nat Chem Biol*, 10: 598-603.
- Guellouz, A., M. Valerio-Lepiniec, A. Urvoas, A. Chevrel, M. Graille, Z. Fourati-Kammoun, M. Desmadril, H. van Tilbeurgh, and P. Minard. 2013. 'Selection of specific protein binders for pre-defined targets from an optimized library of artificial helicoidal repeat proteins (α Rep)', *PLoS One*, 8: e71512.
- Hamers-Casterman, C., T. Atarhouch, S. Muyldermans, G. Robinson, C. Hamers, E. B. Songa, N. Bendahman, and R. Hamers. 1993. 'Naturally occurring antibodies devoid of light chains', *Nature*, 363: 446-8.

- Han, B. W., B. R. Herrin, M. D. Cooper, and I. A. Wilson. 2008. 'Antigen recognition by variable lymphocyte receptors', *Science*, 321: 1834-7.
- Hansen, S., J. C. Stuber, P. Ernst, A. Koch, D. Bojar, A. Batyuk, and A. Pluckthun. 2017. 'Design and applications of a clamp for Green Fluorescent Protein with picomolar affinity', *Sci Rep*, 7: 16292.
- Huang, P. S., S. E. Boyken, and D. Baker. 2016. 'The coming of age of de novo protein design', *Nature*, 537: 320-7.
- Huston, J. S., D. Levinson, M. Mudgett-Hunter, M. S. Tai, J. Novotny, M. N. Margolies, R. J. Ridge, R. E. Brucoleri, E. Haber, R. Crea, and et al. 1988. 'Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*', *Proc Natl Acad Sci U S A*, 85: 5879-83.
- Iri-Sofla, F. J., F. Rahbarizadeh, D. Ahmadvand, and M. J. Rasaee. 2011. 'Nanobody-based chimeric receptor gene integration in Jurkat cells mediated by phiC31 integrase', *Exp Cell Res*, 317: 2630-41.
- Jacob, F. 1977. 'Evolution and tinkering', *Science*, 196: 1161-6.
- Johnson, Frank H., Osamu Shimomura, Yo Saiga, Lewis C. Gershman, George T. Reynolds, and John R. Waters. 1962. 'Quantum efficiency of *Cypridina* luminescence, with a note on that of *Aequorea*', *Journal of Cellular and Comparative Physiology*, 60: 85-103.
- Johnsson, N., and A. Varshavsky. 1994. 'Split ubiquitin as a sensor of protein interactions in vivo', *Proceedings of the National Academy of Sciences*, 91: 10340-44.
- Julien, O., I. Gignac, A. Hutton, A. Yee, C. H. Arrowsmith, and S. M. Gagne. 2006. 'MTH187 from *Methanobacterium thermoautotrophicum* has three HEAT-like repeats', *J Biomol NMR*, 35: 149-54.
- Kapust, R. B., and D. S. Waugh. 1999. '*Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused', *Protein Sci*, 8: 1668-74.
- Kellis, M., B. W. Birren, and E. S. Lander. 2004. 'Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*', *Nature*, 428: 617-24.
- Khersonsky, O., and D. S. Tawfik. 2010. 'Enzyme promiscuity: a mechanistic and evolutionary perspective', *Annu Rev Biochem*, 79: 471-505.
- Koide, S. 2009. 'Engineering of recombinant crystallization chaperones', *Curr Opin Struct Biol*, 19: 449-57.
- Kortemme, T., and D. Baker. 2004. 'Computational design of protein-protein interactions', *Curr Opin Chem Biol*, 8: 91-7.

- Kuhlman, B., G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, and D. Baker. 2003. 'Design of a novel globular protein fold with atomic-level accuracy', *Science*, 302: 1364-8.
- Laughlin, J. D., J. C. Nwachukwu, M. Figuera-Losada, L. Cherry, K. W. Nettles, and P. V. LoGrasso. 2012. 'Structural mechanisms of allostery and autoinhibition in JNK family kinases', *Structure*, 20: 2174-84.
- Lee, J., M. Natarajan, V. C. Nashine, M. Socolich, T. Vo, W. P. Russ, S. J. Benkovic, and R. Ranganathan. 2008. 'Surface sites for engineering allosteric control in proteins', *Science*, 322: 438-42.
- Li, M. 2000. 'Applications of display technology in protein analysis', *Nat Biotechnol*, 18: 1251-6.
- Lutz, S. 2010. 'Beyond directed evolution--semi-rational protein engineering and design', *Curr Opin Biotechnol*, 21: 734-43.
- Magadum, Santoshkumar, Urbi Banerjee, Priyadharshini Murugan, Doddabhimappa Gangapur, and Rajasekar Ravikesavan. 2013. 'Gene duplication as a major force in evolution', *Journal of Genetics*, 92: 155-61.
- Marcotte, E. M., M. Pellegrini, H. Ng, D. W. Rice, T. O. Yeates, and D. Eisenberg. 1999. 'Detecting Protein Function and Protein-Protein Interactions from Genome Sequences', *Science*, 285: 751-53.
- Marin, V., I. Pizzitola, V. Agostoni, G. M. Attianese, H. Finney, A. Lawson, M. Pule, R. Rousseau, A. Biondi, and E. Biagi. 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*, 95: 2144-52.
- Marks, James D., Hennie R. Hoogenboom, Timothy P. Bonnert, John McCafferty, Andrew D. Griffiths, and Greg Winter. 1991. 'By-passing immunization: Human antibodies from V-gene libraries displayed on phage', *Journal of Molecular Biology*, 222: 581-97.
- Mattera, L., S. Bhuckory, K. D. Wegner, X. Qiu, F. Agnese, C. Lincheneau, T. Senden, D. Djurado, L. J. Charbonniere, N. Hildebrandt, and P. Reiss. 2016. 'Compact quantum dot-antibody conjugates for FRET immunoassays with subnanomolar detection limits', *Nanoscale*, 8: 11275-83.
- McCafferty, J., A. D. Griffiths, G. Winter, and D. J. Chiswell. 1990. 'Phage antibodies: filamentous phage displaying antibody variable domains', *Nature*, 348: 552-4.
- Medintz, I., and N. Hildebrandt. 2014. *FRET - Förster resonance energy transfer : from theory to applications*.

- Michnick, Stephen W., Ingrid Remy, François- X. Campbell-Valois, Alexis Vallée-Bélisle, and Joelle N. Pelletier. 2000. 'Detection of protein-protein interactions by protein fragment complementation strategies', *Methods in Enzymology*, 328: 208-30.
- Miyawaki, A., J. Llopis, R. Heim, J. M. McCaffery, J. A. Adams, M. Ikura, and R. Y. Tsien. 1997. 'Fluorescent indicators for Ca²⁺ based on green fluorescent proteins and calmodulin', *Nature*, 388: 882-7.
- Miyawaki, A., and Y. Niino. 2015. 'Molecular spies for bioimaging--fluorescent protein-based probes', *Mol Cell*, 58: 632-43.
- Moot, R., S. S. Raikar, L. Fleischer, M. Querrey, D. E. Tylawsky, H. Nakahara, C. B. Doering, and H. T. Spencer. 2016. 'Genetic engineering of chimeric antigen receptors using lamprey derived variable lymphocyte receptors', *Mol Ther Oncolytics*, 3: 16026.
- Morell, M., S. Ventura, and F. X. Aviles. 2009. 'Protein complementation assays: approaches for the in vivo analysis of protein interactions', *FEBS Lett*, 583: 1684-91.
- Morin, J. G., and J. W. Hastings. 1971. 'Energy transfer in a bioluminescent system', *J Cell Physiol*, 77: 313-8.
- Nagai, T., A. Sawano, E. S. Park, and A. Miyawaki. 2001. 'Circularly permuted green fluorescent proteins engineered to sense Ca²⁺', *Proc Natl Acad Sci U S A*, 98: 3197-202.
- Nakai, J., M. Ohkura, and K. Imoto. 2001. 'A high signal-to-noise Ca(2+) probe composed of a single green fluorescent protein', *Nat Biotechnol*, 19: 137-41.
- Newton, M. S., V. L. Arcus, M. L. Gerth, and W. M. Patrick. 2018. 'Enzyme evolution: innovation is easy, optimization is complicated', *Curr Opin Struct Biol*, 48: 110-16.
- Nicholes, N., A. Date, P. Beaujean, P. Hauk, M. Kanwar, and M. Ostermeier. 2016. 'Modular protein switches derived from antibody mimetic proteins', *Protein Eng Des Sel*, 29: 77-85.
- Nooren, I. M., and J. M. Thornton. 2003. 'Diversity of protein-protein interactions', *EMBO J*, 22: 3486-92.
- Nunez-Prado, N., M. Compte, S. Harwood, A. Alvarez-Mendez, S. Lykkemark, L. Sanz, and L. Alvarez-Vallina. 2015. 'The coming of age of engineered multivalent antibodies', *Drug Discov Today*, 20: 588-94.
- Nye, T. M., C. Berzuini, W. R. Gilks, M. M. Babu, and S. A. Teichmann. 2005. 'Statistical analysis of domains in interacting protein pairs', *Bioinformatics*, 21: 993-1001.
- Packer, M. S., and D. R. Liu. 2015. 'Methods for the directed evolution of proteins', *Nat Rev Genet*, 16: 379-94.

- Pancer, Z., C. T. Amemiya, G. R. Ehrhardt, J. Ceitlin, G. L. Gartland, and M. D. Cooper. 2004. 'Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey', *Nature*, 430: 174-80.
- Pluckthun, A. 2015. 'Designed ankyrin repeat proteins (DARPs): binding proteins for research, diagnostics, and therapy', *Annu Rev Pharmacol Toxicol*, 55: 489-511.
- Riley, R., C. Lee, C. Sabatti, and D. Eisenberg. 2005. 'Inferring protein domain interactions from databases of interacting proteins', *Genome Biol*, 6: R89.
- Rothlisberger, D., O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik, and D. Baker. 2008. 'Kemp elimination catalysts by computational enzyme design', *Nature*, 453: 190-5.
- Segura, J., C. O. Sorzano, J. Cuenca-Alba, P. Aloy, and J. M. Carazo. 2015. 'Using neighborhood cohesiveness to infer interactions between protein domains', *Bioinformatics*, 31: 2545-52.
- Sennhauser, G., and M. G. Grutter. 2008. 'Chaperone-assisted crystallography with DARPins', *Structure*, 16: 1443-53.
- Shaner, N. C., P. A. Steinbach, and R. Y. Tsien. 2005. 'A guide to choosing fluorescent proteins', *Nat Methods*, 2: 905-9.
- Shimomura, Osamu, Frank H. Johnson, and Yo Saiga. 1962. 'Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea', *Journal of Cellular and Comparative Physiology*, 59: 223-39.
- Siegel, J. B., A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. St Clair, J. L. Gallaher, D. Hilvert, M. H. Gelb, B. L. Stoddard, K. N. Houk, F. E. Michael, and D. Baker. 2010. 'Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction', *Science*, 329: 309-13.
- Skerra, A., and A. Pluckthun. 1988. 'Assembly of a functional immunoglobulin Fv fragment in Escherichia coli', *Science*, 240: 1038-41.
- Skrlec, K., B. Strukelj, and A. Berlec. 2015. 'Non-immunoglobulin scaffolds: a focus on their targets', *Trends Biotechnol*, 33: 408-18.
- Sprinzak, E., and H. Margalit. 2001. 'Correlated sequence-signatures as markers of protein-protein interaction', *J Mol Biol*, 311: 681-92.
- Steeland, S., R. E. Vandenbroucke, and C. Libert. 2016. 'Nanobodies as therapeutics: big opportunities for small antibodies', *Drug Discov Today*, 21: 1076-113.

- Stelzl, U., U. Worm, M. Lalowski, C. Haenig, F. H. Brembeck, H. Goehler, M. Stroedicke, M. Zenkner, A. Schoenherr, S. Koeppen, J. Timm, S. Mintzlaff, C. Abraham, N. Bock, S. Kietzmann, A. Goedde, E. Toksoz, A. Droege, S. Krobitsch, B. Korn, W. Birchmeier, H. Lehrach, and E. E. Wanker. 2005. 'A human protein-protein interaction network: a resource for annotating the proteome', *Cell*, 122: 957-68.
- Strickland, D., K. Moffat, and T. R. Sosnick. 2008. 'Light-activated DNA binding in a designed allosteric protein', *Proc Natl Acad Sci U S A*, 105: 10709-14.
- Strickland, D., X. Yao, G. Gawlak, M. K. Rosen, K. H. Gardner, and T. R. Sosnick. 2010. 'Rationally improving LOV domain-based photoswitches', *Nat Methods*, 7: 623-6.
- Tinberg, C. E., S. D. Khare, J. Dou, L. Doyle, J. W. Nelson, A. Schena, W. Jankowski, C. G. Kalodimos, K. Johnsson, B. L. Stoddard, and D. Baker. 2013. 'Computational design of ligand-binding proteins with high affinity and selectivity', *Nature*, 501: 212-16.
- Tiouajni, M., D. Durand, K. Blondeau, M. Graille, A. Urvoas, M. Valerio-Lepiniec, A. Guellouz, M. Aumont-Nicaise, P. Minard, and H. van Tilbeurgh. 2014. 'Structural and functional analysis of the fibronectin-binding protein FNE from *Streptococcus equi* spp. *equi*', *FEBS J*, 281: 5513-31.
- Urvoas, A., A. Guellouz, M. Valerio-Lepiniec, M. Graille, D. Durand, D. C. Desravines, H. van Tilbeurgh, M. Desmadril, and P. Minard. 2010. 'Design, production and molecular structure of a new family of artificial α -helical repeat proteins (α Rep) based on thermostable HEAT-like repeats', *J Mol Biol*, 404: 307-27.
- Valerio-Lepiniec, M., A. Urvoas, A. Chevrel, A. Guellouz, Y. Ferrandez, A. Mesneau, I. L. de la Sierra-Gallay, M. Aumont-Nicaise, M. Desmadril, H. van Tilbeurgh, and P. Minard. 2015. 'The α Rep artificial repeat protein scaffold: a new tool for crystallization and live cell applications', *Biochem Soc Trans*, 43: 819-24.
- Velikovskiy, C. A., L. Deng, S. Tasumi, L. M. Iyer, M. C. Kerzic, L. Aravind, Z. Pancer, and R. A. Mariuzza. 2009. 'Structure of a lamprey variable lymphocyte receptor in complex with a protein antigen', *Nat Struct Mol Biol*, 16: 725-30.
- Wang, Z. J., H. Renata, N. E. Peck, C. C. Farwell, P. S. Coelho, and F. H. Arnold. 2014. 'Improved cyclopropanation activity of histidine-ligated cytochrome P450 enables the enantioselective formal synthesis of levomilnacipran', *Angew Chem Int Ed Engl*, 53: 6810-3.
- Waters, E. A., and E. V. Shusta. 2018. 'The variable lymphocyte receptor as an antibody alternative', *Curr Opin Biotechnol*, 52: 74-79.
- Wu, Y. I., D. Frey, O. I. Lungu, A. Jaehrig, I. Schlichting, B. Kuhlman, and K. M. Hahn. 2009. 'A genetically encoded photoactivatable Rac controls the motility of living cells', *Nature*, 461: 104-8.

- Yi, J. J., H. Wang, M. Vilela, G. Danuser, and K. M. Hahn. 2014. 'Manipulation of endogenous kinase activity in living cells using photoswitchable inhibitory peptides', *ACS Synth Biol*, 3: 788-95.
- Zarrinpar, A., R. P. Bhattacharyya, and W. A. Lim. 2003. 'The structure and function of proline recognition domains', *Sci STKE*, 2003: RE8.
- Zha, W., Z. Shao, J. W. Frost, and H. Zhao. 2004. 'Rational pathway engineering of type I fatty acid synthase allows the biosynthesis of triacetic acid lactone from D-glucose in vivo', *J Am Chem Soc*, 126: 4534-5.
- Zhang, C., J. Liu, J. F. Zhong, and X. Zhang. 2017. 'Engineering CAR-T cells', *Biomark Res*, 5: 22.
- Zhang, D., Q. Zhao, and B. Wu. 2015. 'Structural Studies of G Protein-Coupled Receptors', *Mol Cells*, 38: 836-42.
- Zhang, J., R. E. Campbell, A. Y. Ting, and R. Y. Tsien. 2002. 'Creating new fluorescent probes for cell biology', *Nat Rev Mol Cell Biol*, 3: 906-18.

Partie B : Résultats et discussions

Tables des matières

I.	ÉTUDE DU COMPORTEMENT DE DEUX BIDOMAINES AREPS	92
A.	<i>Introduction</i>	92
B.	<i>Considérations et Expériences préliminaires</i>	93
1)	<i>Linker</i> et concentration effective	93
2)	Caractérisation des bidomaines par <i>Differential Scanning Calorimetry</i> (DSC).....	95
C.	<i>Ligand-induced conformational switch in an artificial bidomain protein scaffold</i>	96
1)	Abstract	96
2)	Introduction	97
3)	Results.....	100
a)	Conception of the bidomain proteins	100
b)	Conformational equilibrium for the bidomains in solution.....	102
c)	A3 and bGFPD interact weakly as isolated domains	106
d)	Each domain of each bidomain is still capable of interacting with its target	106
e)	New functional properties emerge from bidomain coupling.....	107
f)	For A3_A3, apparent K_D values depend on the bidomain's conformation.....	109
g)	For A3_bGFPD, binding of the ligands induces apparent positive cooperativity	111
h)	Kinetics of ligand interactions reveal ligand-induced avidity effects	113
i)	Structural analyses of the bidomains are in agreement with the interactions experiments .	115
j)	Conformational transition consecutive to Ligand binding events can be detected by FRET .	119
4)	Discussion.....	122
a)	Chimeric bidomain α Reps adopt a closed form	122
b)	Ligand induced structural transitions in multi-domains α Reps result in functional effects...	122
c)	Ligand induced dimerization	124
d)	Potential applications of complex bidomains such as sensors and switches	124
D.	<i>Conclusion</i>	126

II.	OPTIMISATION DES BIOSENSEURS GÉNÉRIQUES.....	127
A.	<i>Introduction</i>	127
B.	<i>Mise au point</i>	128
	1) Choix des fluorophores	128
	2) Détermination du rayon de Förster du couples L4-Tb / QD625.....	131
	3) Choix des ligands	133
	4) Choix des longueurs d'onde suivies	134
C.	<i>Couplage des bidomaines aux fluorophores</i>	135
	1) Couplage du Lumi4 et du Terbium	135
	2) Couplage des QD625s	137
D.	<i>Détection des ligands</i>	141
	1) Détection en tampon	141
	2) Détection en milieu complexe	144
	3) Détections successives des ligands	145
E.	<i>Mesure des distances grâce au FRET</i>	147
	1) La théorie	147
	2) Résultats.....	148
	3) Le <i>fitting</i>	149
	4) Quelles distances ?.....	150
F.	<i>Conclusion et discussion</i>	153
III.	VERS DES BIOSENSEURS GÉNÉTIQUEMENT CODABLES.....	154
A.	<i>Introduction</i>	154
B.	<i>Lanthanide Binding Tag et dérivés</i>	155
	1) Origine du LBT	155
	2) Modification du LBT	156
	3) Comparaison avec le LBT.....	157
	4) Détection de bA3-2	158
	5) La dockérine	160
	6) Conclusion	162
C.	<i>La nanoKaz</i>	163
	1) Origine de la nanoKaz.....	163
	2) Conception du biosenseur	164
	3) Sélection des interacteurs de la NanoKaz	166
	a) Source des interacteurs de NanoKaz.....	166
	b) Production et purification de la NanoKaz.....	166
	c) Sélection par phage display	167
	d) Phage Elisa Clonal	169
	4) Caractérisation de l'interaction protéines MG / NanoKaz	169
	5) Production et purification des chimères	171
	6) Caractérisation des Chimères.....	171
	a) Spécificité de détection	171
	b) Mesure des affinités par ITC.....	173
	c) Comparaison des chimères et détection de la BFP	174
	7) Conclusion	175
IV.	BIBLIOGRAPHIE	176

I. Étude du comportement de deux bidomains α Reps

A. Introduction

Comme expliqué dans l'introduction de ce manuscrit, l'association de plusieurs domaines entre eux semble être une des voies adoptées au cours de l'évolution afin de créer de nouvelles fonctions.

Afin de vérifier cette hypothèse et de mieux comprendre le comportement des protéines multidomaines, nous avons utilisé les α Reps comme briques élémentaires dans des constructions multidomaines. Les α Reps ont été créées par l'équipe Modélisation et Ingénierie des Protéines. Ces protéines artificielles présentent des propriétés de reconnaissance et une bibliothèque constituée de $1.7 * 10^9$ protéines différentes a été conçue au sein du laboratoire. (Guellouz, A., et al., 2013; Urvoas, A., et al., 2010) Il est possible de sélectionner parmi cette bibliothèque des α Reps reconnaissant des cibles bien définies par Phage Display. Ces protéines ont fait l'objet d'une caractérisation complète et leurs propriétés d'interaction et de stabilité ne sont plus à démontrer. (Guellouz, A., et al., 2013; Chevrel, A., et al., 2015) Ces protéines artificielles sont donc parfaitement adaptées à une utilisation dans des constructions bidomaines.

Le premier objectif de cette thèse était de caractériser en détails le comportement de deux protéines bidomaines. Pour chacune de ces protéines, les domaines α Reps sont liés covalamment par un *linker* flexible de séquence (SGGGG)₂. La première protéine, **A3_A3**, est constituée de deux domaines α Reps identiques alors que la seconde, **A3_bGFPD**, contient deux domaines α Reps différents. L' α Rep **A3** a été prise au hasard dans la bibliothèque. Elle forme un dimère en solution qui peut être dissocié par l' α Rep **bA3-2**, protéine de la banque obtenue par sélection contre la cible **A3**. L' α Rep **bGFPD** a été sélectionnée contre l'**eGFP**, mais elle est également capable d'interagir avec ses variants (BFP, CFP, YFP). L'étude du comportement des deux protéines chimériques **A3_A3** et **A3_bGFPD** a fait l'objet d'une publication en cours de soumission dans le journal « *Scientific Reports* ».

B. Considérations et Expériences préliminaires

1) *Linker* et concentration effective

Une des considérations importantes pour la création de protéines multidomaines est l'effet du lien covalent entre les domaines. Il peut être quantifié par la concentration effective. Ce paramètre est directement lié à la taille des *linkers*. Ce concept initialement proposé par A.J. Kirby permet de mieux quantifier les effets d'une liaison dans une interaction intramoléculaire. Elle correspond à la concentration à laquelle une protéine (un domaine) « perçoit » une autre protéine (ou un autre domaine) à laquelle elle est reliée de façon permanente. Elle se définit comme étant égale au ratio de la constante d'association d'une réaction intramoléculaire sur celle de cette même réaction mais, intermoléculaire, soit K_{intra}/K_{inter} . (Kirby, A.J., 1980)

Lorsque l'objectif souhaité est de créer un *linker* devant favoriser l'interaction entre deux protéines. La taille de celui-ci doit être au minimum égale à la géodésique reliant les deux extrémités devant être jointes par le *linker* lorsque les protéines interagissent. Si ce dernier est plus court, il y aura probablement la formation d'un dimère dans lequel chaque domaine interagira avec son ligand *via* une interaction intermoléculaire qui sera probablement favorisée. Si celui-ci est plus long, les interactions intramoléculaires devraient être favorisées mais il sera plus sensible aux protéases.

L'équipe de G.M. Whitesides a étudié l'effet de la taille des *linkers* sur cette concentration effective. Si celui-ci est trop court, la concentration effective diminue brutalement, s'il est trop long, la diminution est plus amortie. Pour l'anhydrase carbonique et son substrat, la taille idéale du *linker* reliant les deux molécules est de 9 Å avec une concentration effective de 25 mM. Une distance inférieure à 5 Å réduit quasiment à néant la concentration effective (i.e. comme si les protéines n'étaient pas reliées) alors qu'elle est de 5 mM pour une taille de 22 Å (figure 1). Bien que les valeurs déterminées par les auteurs ne soient applicables qu'au système de l'anhydrase carbonique, le comportement global est valable pour tout système à deux domaines reliés par un *linker*. (Krishnamurthy, V.M., et al., 2007)

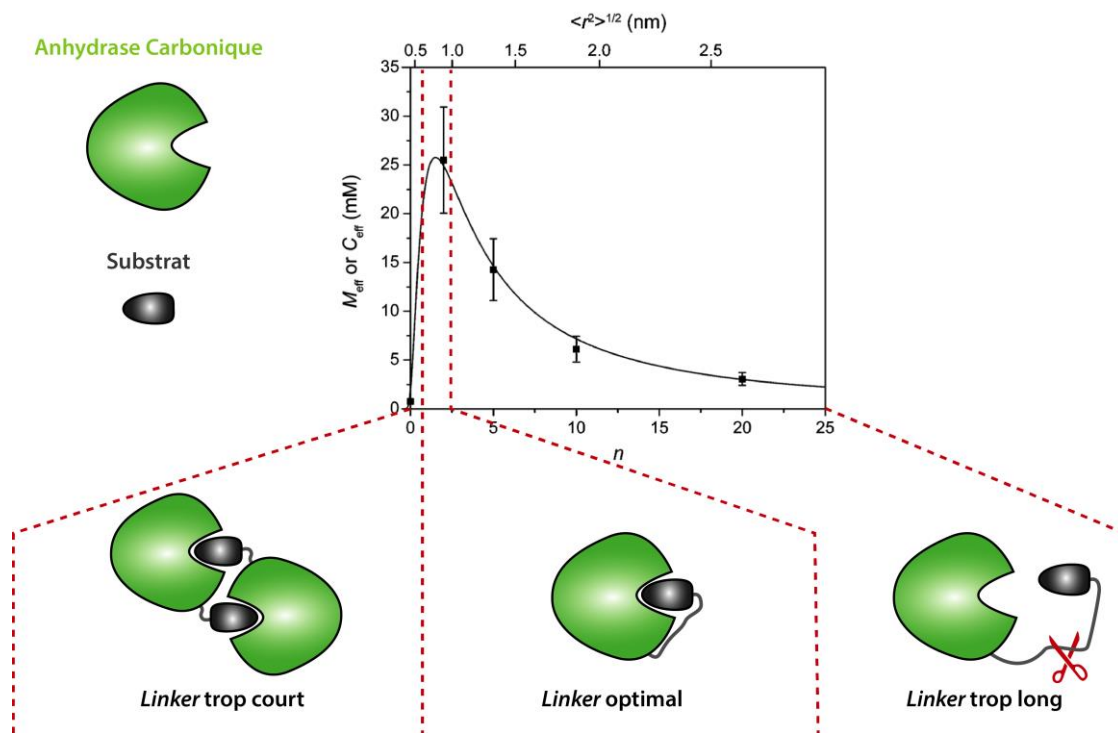


Figure 1 : Dépendance de la concentration effective (M_{eff} ou C_{eff}) en fonction de la longueur du linker. $\langle r^2 \rangle^{1/2}$ correspond à la distance moyenne adoptée par le linker dans l'espace. n correspond au nombre de répétition de l'unité de base (Ethylène Glycol) constituant le linker utilisé dans l'étude de Krishnamurthy et al. Les points noirs représentent les valeurs mesurées, la courbe noire représente les valeurs théoriques obtenues selon les équations des auteurs. (Krishnamurthy, V.M., et al., 2007)

Connaissant la structure du dimère **A3** (pdb id : 3LTJ), le nombre de résidus constituant le linker des bidomains a été choisi pour garantir une interaction intramoléculaire. Le bidomaine **A3_A3**, adoptera donc *a priori* une conformation fermée. En ce qui concerne le bidomaine **A3_bGFPD**, **A3** et **bGFPD** n'ont pas été sélectionnés l'un contre l'autre et n'ont pas de raison d'interagir ensemble. Cependant, il semblerait que le N-cap et le C-Cap de l' α Rep **A3** soient impliqués dans sa dimérisation. Or le C-Cap est le même sur toutes les α Reps, on peut donc se demander si l'interaction N-Cap **A3** / C-cap **bGFPD** sera suffisante pour former une forme fermée en présence du linker pour le bidomaine **A3_bGFPD**.

2) Caractérisation des bidomaines par *Differential Scanning Calorimetry* (DSC)

Le DSC a été l'une des premières expériences effectuées sur les bidomaines. Nous souhaitons en effet en premier lieu observer l'effet du *linker* sur la stabilité des bidomaines. Les résultats présentés [figure 2](#) indiquent que la température de fusion (T_m) du domaine **A3** passe de 83.6 ± 0.1 °C à 86.1 ± 0.1 °C dans le bidomaine **A3_A3**. Cette augmentation laisse penser que le bidomaine adopte bel et bien une conformation fermée plus stable que le simple dimère formé par **A3** en solution. L'interaction intramoléculaire des domaines **A3** serait ainsi renforcée par le *linker* ce qui expliquerait alors l'augmentation de stabilité. Pour le bidomaine **A3_bGFPD**, le T_m passe à 85.9 ± 0.1 °C pour le domaine **A3**. Celui de **bGFPD** passe de 96.2 ± 0.1 °C à 96.8 ± 0.1 °C. L'augmentation du T_m du domaine **A3** dans ce bidomaine laisse ici aussi penser qu'il adopte également une conformation fermée. Il y aurait donc bien une interaction entre **A3** et **bGFPD**, interaction qui serait renforcée par le *linker*. Des expériences complémentaires sont cependant nécessaires afin de confirmer ces hypothèses. Celles-ci sont détaillées dans l'article ci-dessous.

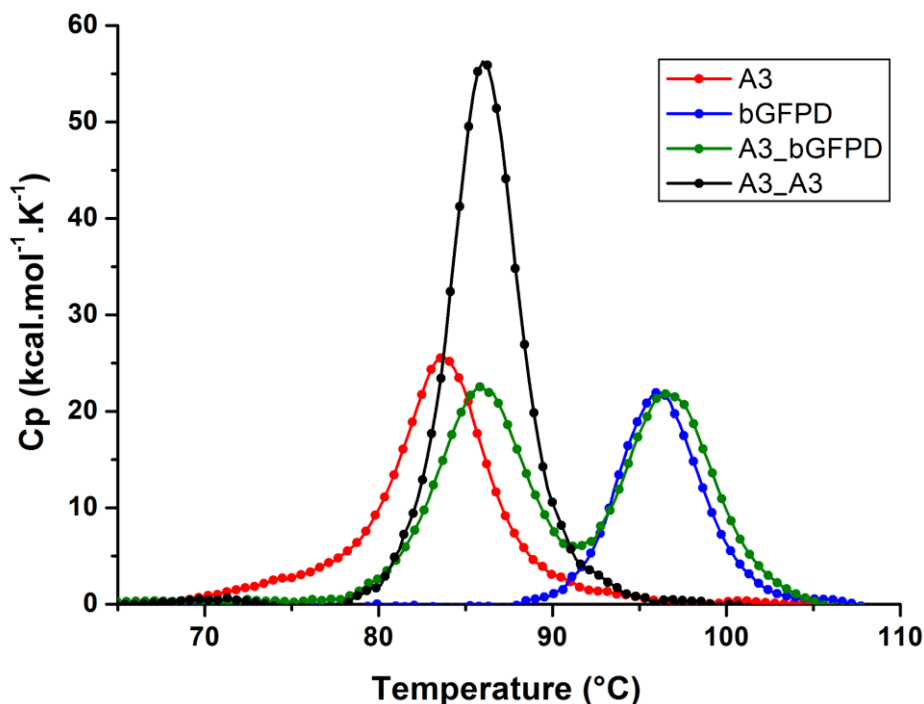


Figure 2 : Stabilités thermiques des bidomaines et des domaines seuls mesurées par DSC. La température de fusion (T_m) du domaine **A3** passe de 83.6 ± 0.1 °C (courbe rouge) à 86.1 ± 0.1 °C dans **A3_A3** (courbe noire) et 85.9 ± 0.1 °C dans **A3_bGFPD** (1er pic courbe verte). Celle de **bGFPD** passe de 96.2 ± 0.1 °C (courbe bleue) à 96.8 ± 0.1 °C dans la chimère **A3_bGFPD** (2nd pic courbe verte). La liaison des domaines semble donc augmenter leur stabilité.

N.B. : Afin de faciliter la lecture de l'article, la mise en page de celui-ci a été révisée. La version originale sera disponible en annexe de ce manuscrit lorsque l'article sera publié. Par ailleurs, la partie « Materials and Methods » a été traduite et insérée dans la partie C de ce manuscrit.

C. Ligand-induced conformational switch in an artificial bidomain protein scaffold

Corentin Léger, Thibault Di Meo, Magali Aumont-Nicaise, Christophe Velours, Dominique Durand, Ines Li de la Sierra-Gallay, Herman van Tilbeurgh, Niko Hildebrandt, Michel Desmadril, Agathe Urvoas*, Marie Valerio-Lepiniec* and Philippe Minard*[§]

[§] Corresponding author: philippe.minard@u-psud.fr

* equal contribution

Institute for Integrative Biology of the Cell (I2BC) CEA, CNRS, Univ. Paris-Sud, Université Paris-Saclay 91198, Gif-sur-Yvette cedex, France

1) Abstract

Artificial proteins binding any predefined "target" protein can now be efficiently generated using combinatorial libraries based on robust protein scaffolds. α Rep is such a family of artificial proteins, based on an α -solenoid protein repeat scaffold. The low aggregation propensity of the specific "binders" generated from this library opens new protein engineering opportunities such as the creation of biosensors within multidomain constructions. Here, we have explored the properties of two new types of artificial bidomain proteins based on α Rep structures. Their structural and functional properties are characterized in detail using biophysical methods.

The results clearly show that both bidomain proteins adopt a closed bivalve shell-like conformation, in the ligand free form. However, the presence of ligands induces a conformational transition, and the proteins adopt an open form in which each domain can bind its cognate protein partner. The open/closed equilibria alter the affinities of each domain and induce new cooperative effects. The binding-induced relative domain motion was monitored by FRET. Crystal structures of the chimeric proteins indicate that the conformation of each constituting domain is conserved but that their mutual interactions explain the emergent properties of these artificial bidomain proteins. The ligand-induced structural transition observed in these bidomain proteins should be transferable to other α Rep proteins with different specificity and could provide the basis of a new generic biosensor design.

2) Introduction

Efficient methods have been recently developed to create artificial proteins binding specifically to almost any desired protein target. The most established strategy is to design and assemble a very large library of protein variants, in which all variants share a same architecture usually based on a natural protein "scaffold" (Banta, S., *et al.*, 2013; Boersma, Y.L. and Pluckthun, A., 2011). Each library variant is characterized by the presence of random amino acid substitutions on its outside surface. New binding properties can emerge from this hypervariable surface, and although very rare in the initial library, specific binders can be selected by genetic sorting methods such as phage, ribosome or yeast display. It is now clear that this strategy, which bypasses immunization, is generally applicable, as illustrated by the fact that a large set of new proteins binding specifically to almost any predefined protein targets were obtained from several published libraries (Banta, S., *et al.*, 2013; Moutel, S., *et al.*, 2016; Tiede, C., *et al.*, 2017). Those specific binding proteins based on alternative scaffolds are usually more efficiently produced and engineered than antibody-based chimeric proteins that suffer from high aggregation propensity. Furthermore, specific binders from disulfide-free protein scaffolds can be efficiently used in reducing environments such as the cytoplasm of living cells.

Engineering of specific proteins based on robust scaffolds opens a range of new opportunities. Sophisticated multidomain proteins such as artificial receptors, sensors or switches could be designed by integrating tailor-made binding proteins into multi-modular constructions (Ha, J.-H. and Loh, S.N., 2012; Stein, V. and Alexandrov, K., 2015; Vallée-Bélisle, A. and Plaxco, K.W., 2010).

In order to convert a specific binder into a specific bio-sensor, the target-binding domain must be integrated with a transducing component in such a way that the target-binding event will give rise to a measurable signal. The signal can result from environment-sensitive fluorescent probes chemically grafted near the binding site of the recognition unit (de Picciotto, S., *et al.*, 2016; Kummer, L., *et al.*, 2013; Miranda, F.F., *et al.*, 2011). Alternatively, the ligand-binding event can induce a structural transition detected in a second "reporter" domain. A simple head to tail fusion of recognition and reporter domains is often not adapted as it usually produces functionally-independent domains. Only few general solutions have been suggested to create such inter-domain coupling. For example, insertion of a peptide within the surface loop of an enzyme (Legendre, D., *et al.*, 1999) or inserting a whole protein into another unrelated enzyme were shown to generate new bi-active proteins with unexpected functional properties (Collinet, B., *et al.*, 2000). The coupling of the inserted and host domains can be further engineered or evolved to create new ligand responsive enzymes (Guntas, G., *et al.*, 2005; Nicholes, N., *et al.*, 2016). Structural transitions were also engineered by designing a bidomain protein in such a way that the folding of the two domains are mutually exclusive (Radley, T.L., *et al.*, 2003). Specific "affinity clamps" have been further developed into protein or peptide specific biosensors by fusing each of the two peptide-binding domains to a fluorescent protein giving rise to reconstituted GFP fluorescence (Carlin, K.B., *et al.*, 2016) or to a ligand-dependent FRET signal (Huang, J. and Koide, S., 2010). Recently, new biosensors using computationally designed protein-binding domain and ligand-induced protein stabilization have been described (Bick, M.J., *et al.*, 2017; Feng, J., *et al.*, 2015; Shelat, N.Y., *et al.*, 2017; Zheng, H., *et al.*, 2014).

In this paper, we propose a new design of a multi-domain protein switch based on an open/closed bivalve-shell-like protein and we describe the experimental behavior of the resulting protein chimera based on this design.

We used a recognition unit based on our previous work with α Rep proteins. α Reps are a family of artificial repeat proteins (Boersma, Y.L. and Pluckthun, A., 2011; Parmeggiani, F. and Huang, P.S., 2017) composed of a HEAT-like motif (Urvoas, A., *et al.*, 2010). Each protein member of the α Rep library has the same general architecture but is endowed with a unique binding surface made by the juxtaposition of hypervariable residues (Guellouz, A., *et al.*, 2013). Specific α Rep binders for a wide range of arbitrarily predefined target proteins with unrelated sequences and structures have been selected by phage display or protein complementation assay (Chevrel, A., *et al.*, 2017). This suggests that the α Rep library is a general source of specific reagents. The crystal structures of five α Rep / target protein complexes (PDB ID: 4JW2, 4JW3, 4XL5, 4XVP, 4ZV6) clearly showed that the targets are bound, as expected, on the hypervariable concave face of the α Rep fold. Given the versatile binding surface grafted on a precisely defined and very stable fold, a designed molecular switch based on α Rep proteins may later be adaptable to other proteins of this family binding different unrelated targets. This prompted us to explore new molecular devices using α Reps as building blocks.

The general design described here is based on the quaternary structure of one the previously described α Rep proteins, named **A3**. This protein spontaneously forms a homodimer in solution (Urvoas, A., *et al.*, 2010). In the dimeric structure, the binding concave surfaces made from the variable residues from each monomer face each other and the dimer resembles a closed bivalve shell. Although relatively tightly associated, we showed that the two monomers can dissociate in presence of another α Rep that binds to the concave surface of **A3**. These results suggest that a synthetic bi-domain protein made by two **A3** monomers linked together should fold as a closed bivalve, but could then be opened by a second distinct protein. This structural transition will *a priori* give rise to a large relative motion of the two linked domains. The opening and closing of a bidomain bivalve shell protein, consecutive to the binding of its ligand appear as a general way to create a conformational switch in this protein family and could later be used to create a range of α Rep-based biosensors.

Despite this opening and closing possibility, it is difficult to predict precisely how such bidomain proteins will behave. For example, it is expected that covalently linking the two subunits of the **A3** homodimer should reinforce the interactions between the two monomers and therefore should stabilize a closed form. But, would this closed bivalve still be able to open in order to bind an external ligand? Would it be possible to observe such a structural transition in α Rep with other binding specificities? For example, how a chimeric protein made by an **A3** domain and a different unrelated α Rep domain would behave in solution? Would it give rise to a closed-bivalve? What will be the behavior of such a bidomain in the presence of ligands specific for each domain? The properties of such proteins are currently unpredictable and a better appreciation of these effects is clearly essential for the conception of multi-modular devices.

Our goal was to address experimentally these questions and to establish clear observations on the possible behavior of such proteins.

3) Results

a) Conception of the bidomain proteins

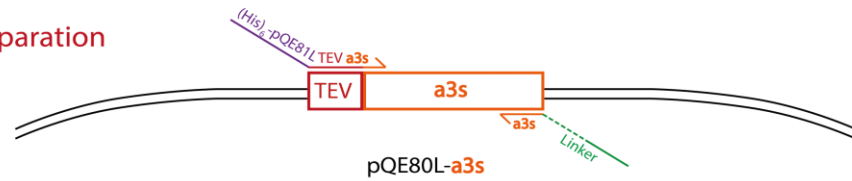
Previously described α Reps **A3** (Urvoas, A., *et al.*, 2010) and **bGFPD** (Chevrel, A., *et al.*, 2015) were connected by a flexible linker (SGGGG)₂, creating two bidomains named **A3_A3** and **A3_bGFPD** (Figure 3).

α Rep **A3** contains four internal repeats between an N-cap and a C-cap motif and its biophysical properties and crystal structure were previously described. The **A3** homodimer dissociates in solution upon interaction with two **bA3-2** and **bA3-17** α Reps that were selected from the library as specific binders for **A3** (Guellouz, A., *et al.*, 2013). The interaction surface was determined using the crystal structure of the **A3** / **bA3-2** complex (PDB ID 4JW2). **bGFPD** was selected from the α Rep library as a specific binder for **eGFP**. It also contains four internal repeats and two caps.

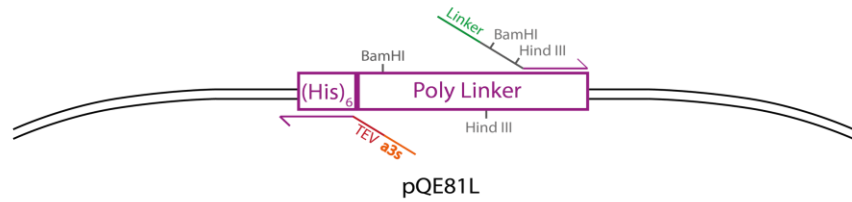
I - CPEC

Step 1: PCR Amplification

a) Insert preparation



b) Preparation of linear vector



Step 2: Final assembly CPEC



II - *a3* or *bgfpD* cloning

a)



or

b)



Figure 3 : The sequences coding for both bidomains, A3_A3 and A3_bgFPD, are cloned in the vector PQE81L. The construction coding for A3_A3 is composed of two distinct DNA sequences coding for the same α Rep A3 protein. In the final chimera both A3_A3 and A3_bgFPD proteins are separated by a linker (SGGGG)₂. The different steps of the construction are detailed in the Materials and Methods section.

b) Conformational equilibrium for the bidomains in solution

SEC-MALS experiments were performed with **A3**, **bGFPD**, **A3_A3** and **A3_bGFPD** alone, and in the presence of their partners (Figures 4 and 5). The experimental molecular weights indicate that **A3** is dimeric and **bGFPD** is monomeric in solution as previously reported (Urvoas, A., *et al.*, 2010; Chevrel, A., *et al.*, 2015). As expected, the bidomain single chain, **A3_A3** is monomeric in solution, as the two linked domains interact together. The hydrodynamic radius of the bidomain **A3_A3** (3.0 nm) is similar to the dimeric **A3** (2.9 nm) indicating that the chimeric bidomain protein also adopts a closed bivalve conformation. SEC-MALS analysis of the **A3_bGFPD** construct yielded a broad and dissymmetric peak. Analysis of the higher elution volume fraction of the peak indicates a 44.8 kDa mass, which corresponds to **A3_bGFPD** monomer mass (Figure 5-B). The mass corresponding to the top of the peak (51.4 kDa) suggests that the bidomain protein **A3_bGFPD** is in equilibrium between a monomeric and a dimeric form. Dimerization could be induced by the interaction of **A3** domains from two distinct molecules.

SEC-SAXS, a useful strategy to separate aggregates and/or oligomers from isolated proteins, has confirmed this conformational equilibrium (Figure 6). In the first SEC-SAXS experiment, SAXS measurements were continuously performed while the proteins eluted from the size exclusion column. The sample concentration, monitored by a UV spectrophotometer installed before the SAXS cell, reached a value of up to about $0.5 \text{ mg}\cdot\text{mL}^{-1}$ ($12 \text{ }\mu\text{M}$). The radius of gyration R_g , the forward scattering intensity $I(0)$ - which is proportional to the protein concentration - and the molecular mass M were extracted from each $I(q)$ curve (Figure 6-A). M decreases strongly at the beginning of the elution, then slower, finally to reach the value of the molecular mass expected from the sequence, when the concentration is smaller than about $0.35 \text{ mg}\cdot\text{mL}^{-1}$ ($8 \text{ }\mu\text{M}$). A simultaneous decrease of R_g is observed. A second SAXS experiment was performed by direct injection of concentrated solutions ($0.5 \text{ mg}\cdot\text{mL}^{-1}$ ($12 \text{ }\mu\text{M}$) $\leq c \leq 8.0 \text{ mg}\cdot\text{mL}^{-1}$ ($189 \text{ }\mu\text{M}$)) into the SAXS cell. The scattering intensities $I(q)/c$ normalized to the concentration change drastically when the concentration increases (Figure 6-B). A strong increase is observed in the small- q range indicating a monomer/dimer distribution shift in the $0.4 - 4 \text{ mg mL}^{-1}$ ($\approx 10 - 100 \text{ }\mu\text{M}$) concentration range. The molecular masses extracted from each $I(q)$ curve are shown in the insert. For the highest concentration ($8 \text{ mg}\cdot\text{mL}^{-1}$, $\approx 200 \text{ }\mu\text{M}$) the value of M is even higher than the mass of the dimer.












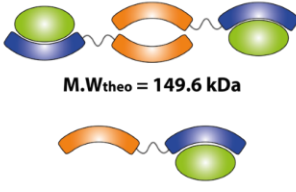


	 bA3-2 M.W _{theo} = 11.9 kDa M.W _{exp} = 11.6 ± 0.9 kDa Rh = N.D	 eGFP M.W _{theo} = 30.6 kDa M.W _{exp} = 32.3 ± 0.7 kDa Rh = 2.6 ± 0.3 nm	 bA3-2 + eGFP
A3  M.W _{theo} = 44.2 kDa M.W _{exp} = 41.8 kDa Rh = 2.90 nm	 M.W _{theo} = 34.1 kDa M.W _{exp} = 33.0 kDa Rh = 2.70 nm	N.D	N.D
bGFPD  M.W _{theo} = 22 kDa M.W _{exp} = 23.1 ± 0.2 kDa Rh = 1.8 ± 0.3 nm	N.D	 M.W _{theo} = 52.6 kDa M.W _{exp} = 56.3 ± 0.3 kDa Rh = 3.2 ± 0.2 nm	N.D
A3-A3  M.W _{theo} = 44.5 kDa M.W _{exp} = 44.8 ± 0.4 kDa Rh = 3.0 ± 0.2 nm	 M.W _{theo} = 68.3 kDa M.W _{exp} = 65.5 ± 0.4 kDa Rh = 4.0 ± 0.2 nm	N.D	N.D
A3-bGFPD  M.W _{theo} = 44.4 kDa M.W _{exp} = 51.4 ± 0.8 kDa Rh = 3.1 ± 0.3 nm	 M.W _{theo} = 56.3 kDa M.W _{exp} = 53.3 ± 0.8 kDa Rh = 3.3 ± 0.2 nm	 M.W _{theo} = 149.6 kDa  M.W _{theo} = 74.8 kDa M.W _{exp} = 140 to 90 kDa Rh = 4.8 ± 0.3 nm	 M.W _{theo} = 86.9 kDa M.W _{exp} = 81.7 ± 0.6 kDa Rh = 4.2 ± 0.3 nm

Figure 4: SEC-MALS characterization of the isolated α Rep domains (A3 and bGFPD), chimeras (A3_A3 and A3_bGFPD) and their respective complexes. A3 is shown in orange, bGFPD in blue, the ligand bA3-2 in cyan and eGFP in green. For each complex, theoretical molecular weight ($M.W_{theo}$) was calculated with the amino acid sequences using ExPASy ProtParam Tool. Schematic representation of the complexes is deduced from experimental values for $M.W_{exp}$ and hydrodynamic radius (R_h) were determined using the SEC-MALS.

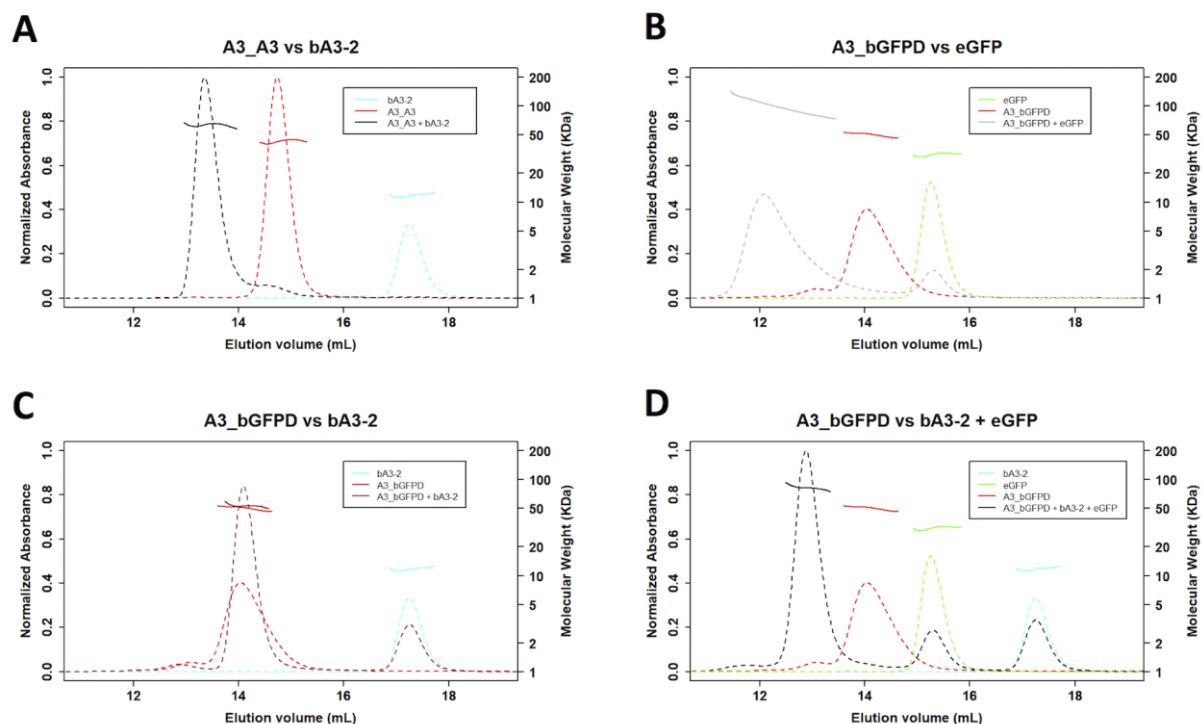


Figure 5 : Molecular Weight (hard lines) and Normalized absorbance (dash lines) are shown for eGFP, bA3-2, A3_A3 and A3_bGFPD. Purified proteins or mixtures (100 μ L, 45 μ M for all proteins except bA3-2 90 μ M) were loaded on a Superdex 200 10/300GL column equilibrated in 50 mM NaP pH 7.5, 150 mM NaCl at a 0.5 mL/min flow rate and detected using a multi angle light scattering detector (Wyatts). For each eluted peak, the experimental molecular weight (M_{Wexp}), the molecular extinction coefficient (ϵ) and the hydrodynamic radius (R_h) were determined using the SEC-MALS Wyatt software. Profiles of the chimera alone are represented in red, profiles of the binder bA3-2 in cyan, of the eGFP in green. The mix of the chimera A3_A3 with the binder bA3-2 is represented in black, the mix of the chimera A3_bGFPD with the eGFP in light grey, with the bA3-2 in brown and with both in black. For panels A, B, D, the shifts of the peaks for the chimera to higher elution volumes in presence of their binders show that the chimeras are still able to interact with their ligand. For panel C, the isolated chimera is in equilibrium between monomeric and dimeric forms (through interactions between A3 domains). bA3-2 binding induces both a shift of this monomer/dimer equilibrium and an increase of the A3_bGFPD apparent mass due to bA3-2 binding. No apparent shift of the main peak is therefore observed. The interactions of bA3-2 and A3_bGFPD can be quantified more precisely using the data obtained by refractometry allowing to calculate the number of moles of proteins in each peak. These data indicate that a fraction of bA3-2 (37 μ g, 3.1 nmol) interacts with A3_bGFPD (137 μ g, 3.1 nmol) to generate a new peak corresponding to the complexes (174.4 μ g). For all proteins, the theoretical and experimental ϵ are comparable.

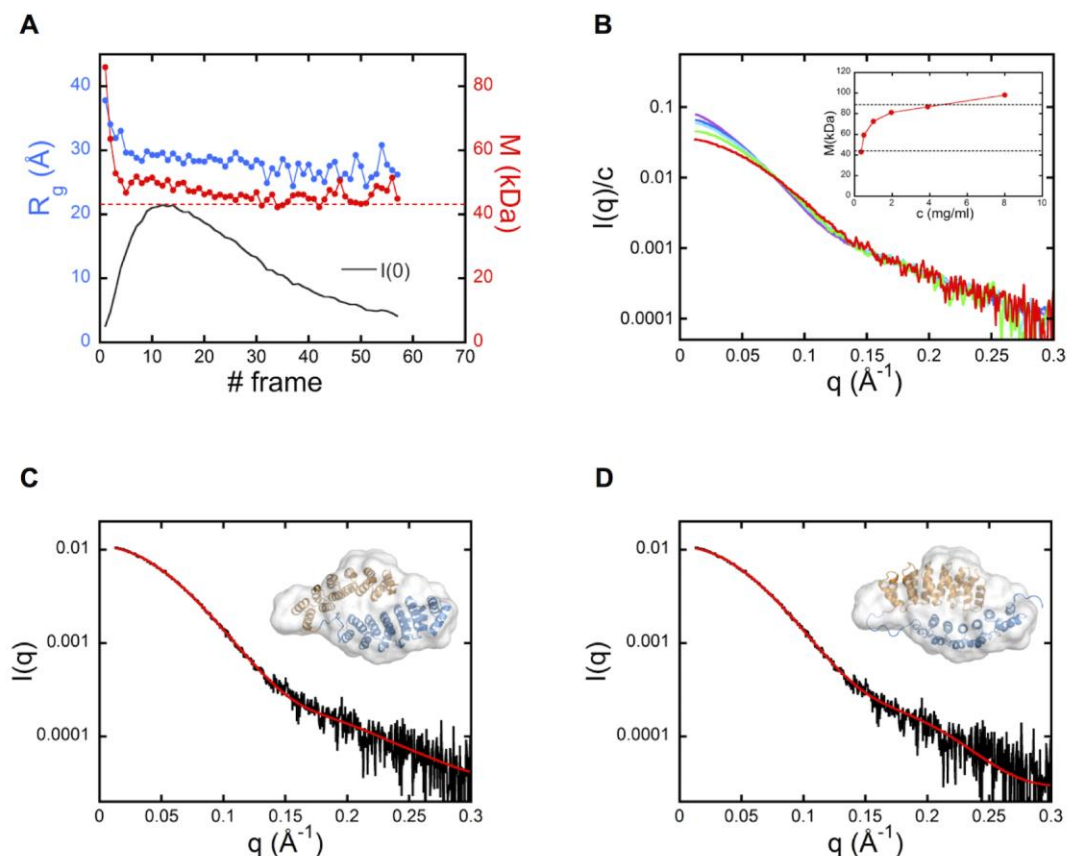


Figure 6 : SAXS analysis of A3_bGFPD

A: SEC-SAXS study of A3_bGFPD. Evolution with time of the radius of gyration R_g , of the forward scattering intensity $I(0)$ and the molecular mass M . For each frame the acquisition time is 2s. The dashed horizontal line corresponds to the molecular mass of the monomer calculated from the sequence.

B: Scattering intensities $I(q)/c$ measured at different concentrations ranging from about 0.35 mg/ml to 8 mg/ml. The curve corresponding to the lowest concentration (red line) is the averaged scattering curve obtained on the SEC-SAXS beamline. The others curves were obtained by direct injection using the in-house SAXS instrument. The molecular masses shown in the insert were deduced from the SAXS curves using the program ScÅtter1 available at the URL <https://bl1231.als.lbl.gov/scatter/>. The two horizontal lines correspond to the calculated molecular masses for the monomer and the dimer, respectively.

C and D: Analysis of the averaged SEC-SAXS curve measured at the end of the elution (black line). The red lines result of fitting the experimental curve using the program CRY SOL2 from the atomic models shown in the inserts. In the model C the two domains A3 and b-GFPD are arranged as in the dimeric A3. The model D corresponds to the crystal structure of A3_bGFPD. Both fits are equally good ($\chi^2=0.64$ and 0.66, respectively). The light-grey colored envelope is deduced from the SEC-SAXS curve using the program ab initio GASBOR3.

c) A3 and bGFPD interact weakly as isolated domains

Concerning the **A3_bGFPD** bidomain, a possible interaction between the **A3** and **bGFPD** linked domains producing a closed monomeric form, was suggested from the SEC-MALS data. We therefore tried to quantify a possible interaction between free (non-linked) **A3** and **bGFPD** domains using ITC: **A3** was titrated with **bGFPD** and the isotherm clearly showed a weak interaction with a K_D of $4.1 \pm 1.3 \mu\text{M}$ and a 1:1 stoichiometry. Analytical ultracentrifugation analyses confirmed the interaction of **A3** with **bGFPD** with a measured K_D value of $3.3 \pm 0.9 \mu\text{M}$ and a 1:1 stoichiometry. This suggests that a relatively weak affinity between the two isolated **A3** and **bGFPD** monomers does exist, and is sufficient to stabilize a closed form when the two domains are linked into a single polypeptide chain.

d) Each domain of each bidomain is still capable of interacting with its target

Upon the addition of one molar equivalent of each ligand (**bA3-2** or **eGFP**) to the bidomain, changes in the SEC elution volumes and/or in the peak area were observed, indicating the formation of complexes of higher molecular weight. The stoichiometry of these complexes was determined by using the measured molar masses and molecular extinction coefficients. In both bidomains, each domain binds one molar equivalent of ligand. Taken together, the SEC-MALS analyses reveal that both bidomains still bind their ligands (**bA3-2** and **eGFP**) and that the two domains of one chimera can bind their ligand simultaneously (**Figure 4 and 5**).

e) New functional properties emerge from bidomain coupling

If these two bi-domains, α Rep adopt a closed form but are still able to bind their respective partners, then, a structural transition has to take place upon ligand binding and should be detected through its effects on the binding properties of the chimeric proteins. The interactions between the different partners were then quantified (Table 1). We first studied the isolated domains **A3** and **bGFPD** with their respective targets, **bA3-2** and **eGFP** using Isothermal Titration Calorimetry (ITC). As previously reported (Chevrel, A., *et al.*, 2015; Guellouz, A., *et al.*, 2013), **bA3-2** interacts with **A3** in a 1:1 stoichiometric *ratio* with a K_D value of 4.5 ± 7 nM and **eGFP** interacts with **bGFPD** with a K_D value of 2.7 ± 2.0 nM. The interaction between two **A3** monomers resulting in the **A3** dimer has been measured using Analytical Ultracentrifugation (AUC) and a K_D value of 37 ± 6 nM ($K_{D_{A3/A3}}$) was determined (Table 1 and Figure 7).

Partner 1	Partner 2	K_D (nM)				
		SPR ^[1]	ITC ^[2]	UCA ^[3]	FRET ^[4]	Constant definition ^[5]
A3	bA3-2	1.1 ± 0.1	4.5 ± 7.0			K_{D1}
bGFPD	eGFP	2.3 ± 0.1	2.7 ± 2.0			
	BFP		5 ± 2.0			
A3	A3		n.a.	37 ± 6		$K_{D_{A3A3}}$
A3	bGFPD		$4.1 \pm 1.3 \cdot 10^3$	$3.3 \pm 0.9 \cdot 10^3$		$K_{D_{A3bGFPD}}$
A3_A3	bA3-2	46.0 ± 2.5	37 ± 7		623 ± 86	K_{app1}
A3_A3 + excess bA3-17	bA3-2	-	9.7 ± 1.5			
A3_A3 + excess bA3-2	bA3-17	-	n.a.			
A3_bGFPD	bA3-2	-	77 ± 13		57 ± 10	K_{app2}
A3_bGFPD	eGFP	12.5 ± 1.2	40 ± 4			
	BFP				134 ± 20	
A3_bGFPD + excess eGFP	bA3-2	-	21 ± 10			
A3_bGFPD + excess bA3-2	eGFP	-	8.9 ± 2.1			

Table 1 : ^[1] SPR data were obtained as described in Figure 3, with partner 2 (analyte) in solution and partner 1 (ligand) immobilized on the sensor chip surface. K_D values were determined using the Langmuir model. Errors on the K_D values were obtained according to: $\Delta K_D = K_D \cdot (\Delta k_{on}/k_{on} + \Delta k_{off}/k_{off})$.

^[2] ITC experiments were obtained as described in Figure 2. Partner 2 injected from the syringe in the cell, containing partner 1. ITC data were treated using a model assuming equivalent and independent sites.

^[3] UCA experiments depicted in Figure 7.

^[4] FRET experiments depicted in Figure 13. FRET intensities plotted as a function of total Ligand were fitted as described in Material and methods section.

^[5] Constants describing the intrinsic properties of the chimeras ($K_{intra_{A3_A3}}$ and $K_{intra_{A3_bGFPD}}$) defined in Figure 9.

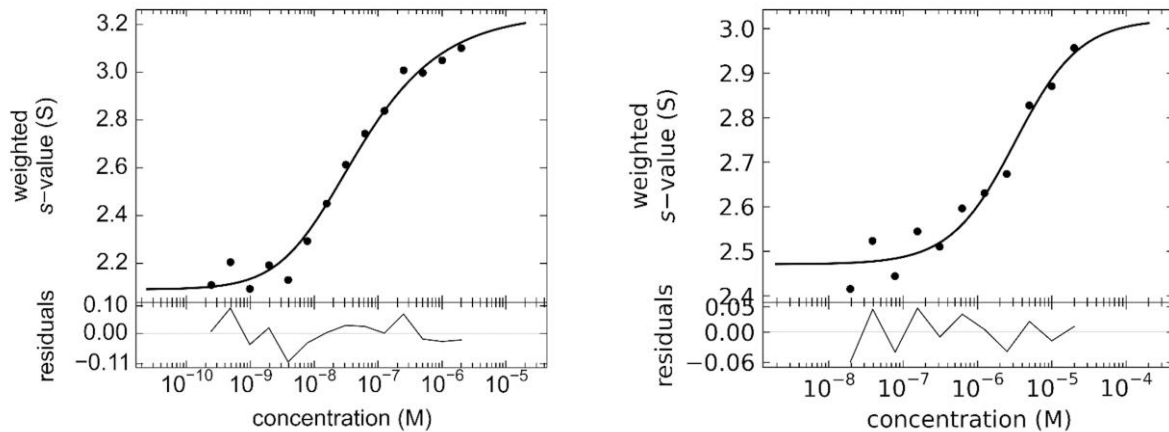


Figure 7 : Analytical Ultracentrifugation analysis of the proteins dimerization.

(left) Determination of the dimerization constant of A3. Ultracentrifugation runs were performed using A3 concentrations ranging from 244 pM to 2 μ M.

(right) A3 was incubated in the presence **bGFPD**: A3 concentration was fixed at 10 nM, **bGFPD** concentrations were from 19 nM to 20 μ M.

In both cases, A3 was labeled with a fluorophore (Monolith NT-115 Nanotemper). Centrifugation was done at 128 297g at 20°C. Sedimentation velocity boundaries were analyzed using SEDFIT software (<https://sedfitsedphat.nibib.nih.gov/software/default.aspx>).

The resulting $c(s)$ distributions were analyzed with GUSI software (<http://biophysics.swmed.edu/MBR/software.html>) using sedimentation velocity isotherm mode. For each concentration, the signal weighted-average sedimentation coefficient (sw) was determined by integration and plotted (Analysis of high affinity self-association...). The sw isotherm was loaded into SEDPHAT (2) for fitting with the homo-dimerization model. The buffer density and viscosity at 20°C were determined by SEDNTERP software (T. Laue, Royal Society of Chemistry, Cambridge), partial specific volume for A3 was calculated based on the amino acid composition using this program.

f) For A3_A3, apparent K_D values depend on the bidomain's conformation

A3_A3 was first titrated with bA3-2 by ITC. The stoichiometry (two bA3-2 for one A3_A3) (Table 1 and Figure 8) indicates that each subunit is able to bind one ligand molecule, which confirms the SEC-MALS stoichiometry data. However, the K_D value of bA3-2 binding on A3_A3 (37 ± 6 nM) is higher than with A3 alone (without linker: 4.5 ± 7 nM), suggesting that the intramolecular link between the two binding sites stabilizes the closed form and induces a lower apparent affinity for an external ligand. The ITC data were reasonably well fitted with a model assuming equivalent and independent sites. It seems however clear that, if the two binding sites of the chimeric protein A3_A3 are nearly identical, the binding of a ligand on the first site will stabilize an open conformation of the bidomain structure and will facilitate the binding on the second site. The K_D determined by this experiment should therefore be considered as an apparent value (named K_{app1} in Figure 9), which averages the binding of two non-independent sites.

In a second experiment, the effects of the open/close equilibrium on ligand binding were probed. The idea was to "pre-open" the chimeric protein with an alternative partner before monitoring ligand binding. The α Rep protein bA3-17 was previously characterized as a ligand for A3 with a lower affinity than bA3-2 (Guellouz, A., *et al.*, 2013). The A3_A3 chimera was pre-incubated with one of its ligands and the second ligand was then added as a competitor. In this experiment, bA3-17 was not able to replace pre-bound bA3-2 as illustrated by the lack of signal in ITC. This result is in accordance with the known lower affinity of bA3-17 for A3, relatively to bA3-2 for A3. In the reverse experiments, bA3-2 was able to replace bA3-17 with a K_D value 9.7 ± 1.5 nM (Table 1, Figure 8-A). Thus, the apparent affinity associated to bA3-2 binding on A3_A3 actually increased when the A3_A3 protein was pre-incubated with another ligand. This suggests that A3_A3, when bound to a first protein ligand (with bA3-17 here), is in an open form while it is in a closed conformation before pre-incubation. In other words, the stability of the closed conformation contributes to lower its apparent affinity for bA3-2.

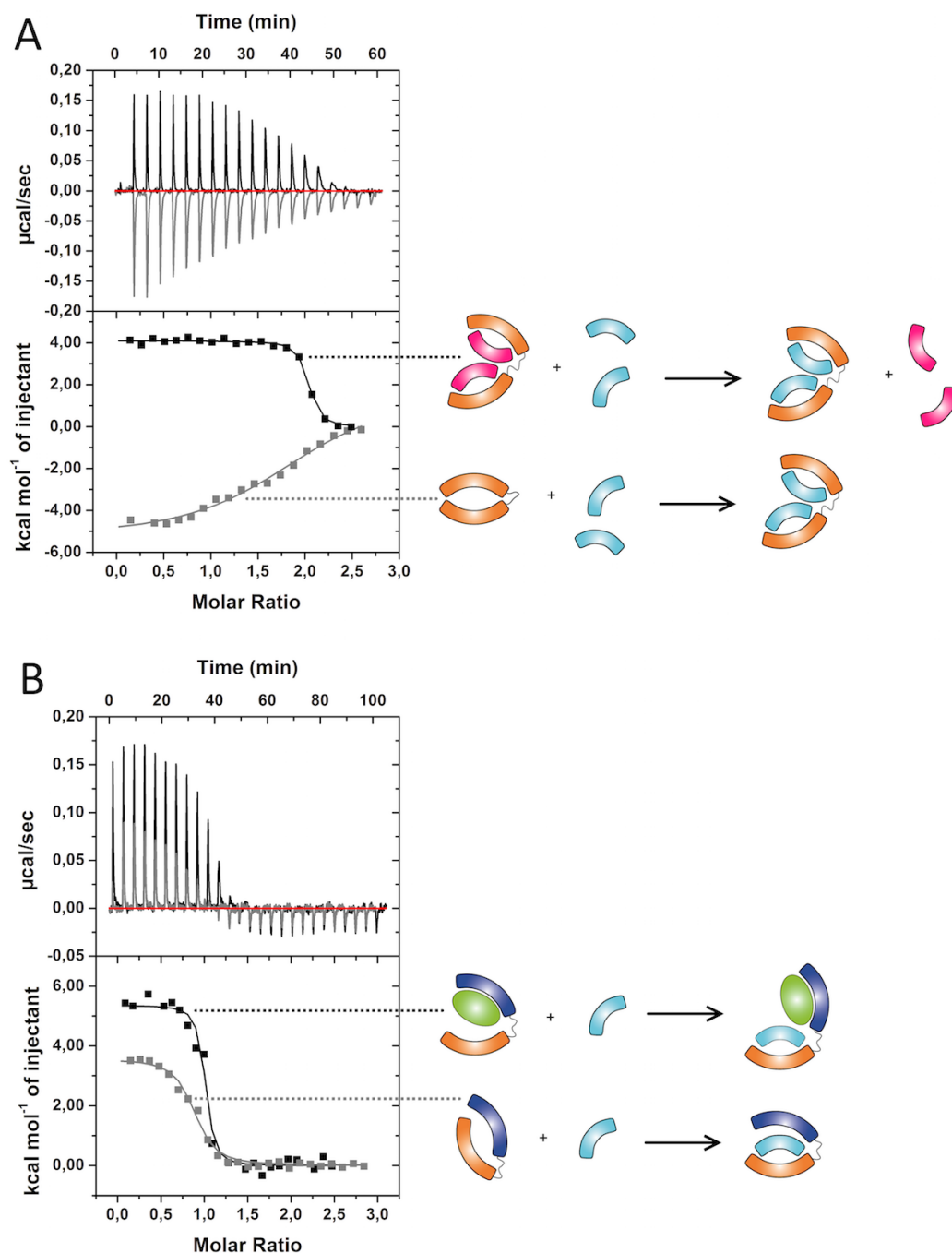


Figure 8 : (A) ITC titration of a solution of A3_A3 (25 μM) either alone (in grey), or pre-incubated with an excess of the protein ligand bA3-17 (100 μM) (black) with bA3-2 (300 μM). These data indicate that bA3-2 can bind the bidomain with a stoichiometry of two bA3-2 ligands per bidomain (one ligand per A3 subunit); bA3-2 binds the bidomain with an increased apparent affinity in the presence of bA3-17 compared with the data obtained without bA3-17 pre-incubation.

(B) ITC titration of monomeric A3_bGFPD (8 μM) alone (grey) or pre-incubated with an excess of the protein ligand eGFP (40 μM) (black) with bA3-2 (100 μM). bA3-2 binds the bidomain with a stoichiometry of one bA3-2 per bidomain (one ligand per A3 subunit) and an increased apparent affinity in the presence of eGFP.

g) For A3_bGFPD, binding of the ligands induces apparent positive cooperativity

According to the SEC-MALS data, the **A3_bGFPD** chimera is in equilibrium between a monomeric and a dimeric form. SAXS experiments revealed that, below 8 μ M, the **A3-bGFPD** chimera is mainly monomeric. To measure the interaction of ligands on the monomeric form of **A3_bGFPD**, we therefore performed the following ITC binding experiment with a 8 μ M solution of chimera.

The binding of **bA3-2** or **eGFP** to **A3_bGFPD** was measured first individually and then sequentially. Stoichiometric interactions with **A3_bGFPD** were observed with K_D values of 77 ± 13 nM for **bA3-2** and 40 ± 4 nM for **eGFP** (Table 1, Figure 8-B). Once again, each ligand is able to interact with its partner domain, but the apparent affinities are lower than those observed for the individual domains. This suggests that the covalent link strengthens interactions between the two domains and stabilizes a closed form for the chimera.

Following pre-incubation of **A3_bGFPD** with **eGFP**, **bA3-2** could bind with a K_D value of 21 ± 10 nM (Table 1). When pre-incubated with **bA3-2**, **eGFP** binds with K_D value of 8.9 ± 2.1 nM. Both pre-incubations could be a way to convert the closed conformation of the chimera into an open state thereby facilitating the binding of the second ligand. Yet, the measured affinity of **bA3-2** and **eGFP** for the **A3_bGFPD** following pre-incubation is still lower than the one for **A3** or **bGFPD** alone, suggesting a steric hindrance of both ligands for the interaction with their respective domain's surface.

Overall, these ligand effects can be considered as positive heterotropic cooperativity between the two sites of the chimera: the binding of the first ligand facilitates the binding of the second one. These binding properties directly support the existence of ligand induced conformational transitions in both chimeric proteins.

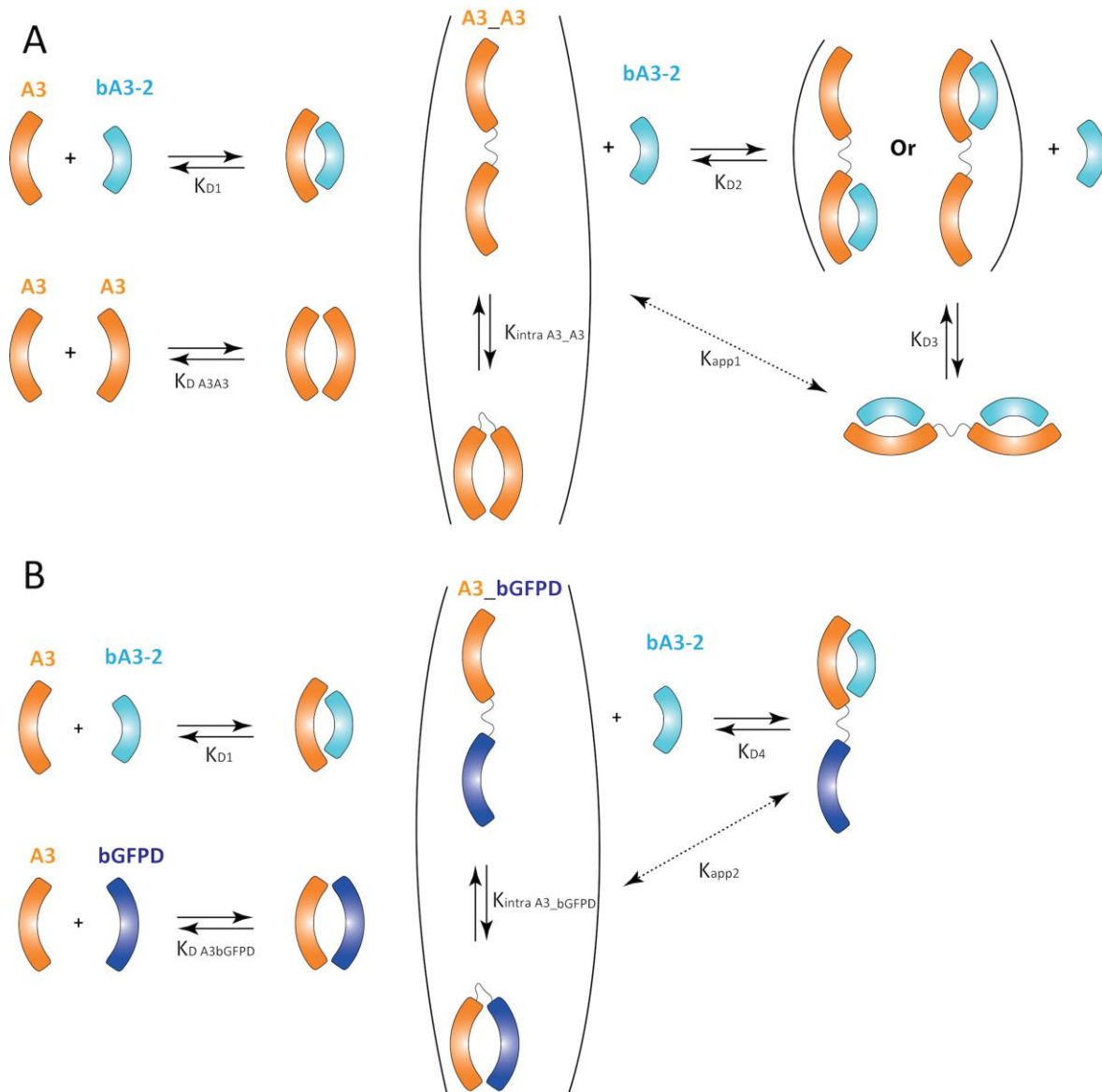


Figure 9 : Models of the interactions in solution with the effects of the equilibrium between the open and closed conformations of the bidomains.

(A) Model of bA3-2 (in cyan) binding on A3 and on A3_A3 bidomain (in orange). The dissociation constant of the A3A3 ($K_{D_{A3A3}}$) homodimer was measured by UCA (Table 1). The dissociation constant K_{D1} for A3 in complex with bA3-2 was measured by SPR with a solution of monomeric A3 (below 2 nM). The apparent dissociation constant K_{app1} obtained by ITC, reflects the presence of the equilibrium between the open and the closed bidomain conformations driven by the constant $K_{intra_{A3_{A3}}}$. $K_{intra_{A3_{A3}}}$ is defined as the ratio of the A3_A3 concentration in the open form to the A3_A3 concentration in the closed form.

(B) Model of bA3-2 (in cyan) binding on A3 and on A3_bGFPD bidomain (in orange-blue). The dissociation constant K_{D1} for A3 in complex with bA3-2 is the same as in (A). Although A3 and bGFPD are non-related α Reps, they interact with low affinity and a micromolar dissociation constant of the A3bGFPD heterodimer ($K_{D_{A3bGFPD}}$) was measured by UCA and ITC (Table 1). The apparent dissociation constant K_{app2} for the binding of bA3-2 obtained by ITC, reflects the presence of the equilibrium between the open and the closed chimera conformations driven by the constant $K_{intra_{A3_{bGFPD}}}$. $K_{intra_{A3_{bGFPD}}}$ is defined as the ratio of the A3_bGFPD concentration in the open form to the A3_bGFPD concentration in the closed form.

h) Kinetics of ligand interactions reveal ligand-induced avidity effects

The kinetics of ligand association and dissociation with the two bidomains were analyzed by Surface Plasmon Resonance (Table 1).

The K_D values calculated from the association and dissociation fitting curves using the Langmuir Model were 2.3 ± 0.1 nM for circulating **bGFPD** with immobilized **eGFP** and 1.5 ± 0.2 nM for circulating **A3** with immobilized **bA3-2**. The maximum concentration of **A3** in this experiment was 4 nM and at this concentration, **A3** is more than 90 % monomeric according to the dimerization constant K_D **A3/A3** determined by AUC (37 ± 6 nM). We thus measured the affinity of **bA3-2** for the monomeric form of **A3**, which is, as expected, higher than for the **A3** dimer.

When the bidomains are immobilized on the chip, the profile of the kinetics of ligand association and dissociation obtained with either **bA3-2** (Figure 10-E) or **eGFP** (Figure 10-F) as circulating analytes are comparable to those observed when each isolated domain is immobilized (Figure 10-A and 10-B).

Bidomains **A3_A3** and **A3_bGFPD** were then used as circulating analytes on **bA3-2** and **eGFP** immobilized ligands. The curves could not be fitted using simple models, but, clearly, the creation of chimeric proteins affects both association and dissociation phases. For both complexes **bA3-2/A3_A3** and **eGFP/A3_bGFPD** (Figure 10-C, 10-D), a slower association was observed for the chimeric bidomain protein relatively to their non-chimeric equivalents (**bA3-2/A3** and **eGFP/bGFPD** complexes) (Figure 10-A, 10-B). This slower association is expected as an equilibrium exists between the open and closed states of the bidomain. The slower dissociation was more unexpected, but can be attributed to a ligand-inducible avidity effect for both chimeras. Indeed, the **A3_A3** bidomain, once opened, can bind simultaneously two immobilized **bA3-2** (Figure 10-C, bottom). Similarly, when the **A3_bGFPD** bidomain binds one **eGFP** ligand immobilized on the chip, its second domain (**A3**) is released and its dimerization is enhanced with an **A3** domain from another bisomain present in the analyte solution, or bound on the surface. Thus, it induces an avidity effect for the second **eGFP** binding site with a very slow dissociation from the chip (Figure 10-D, bottom).

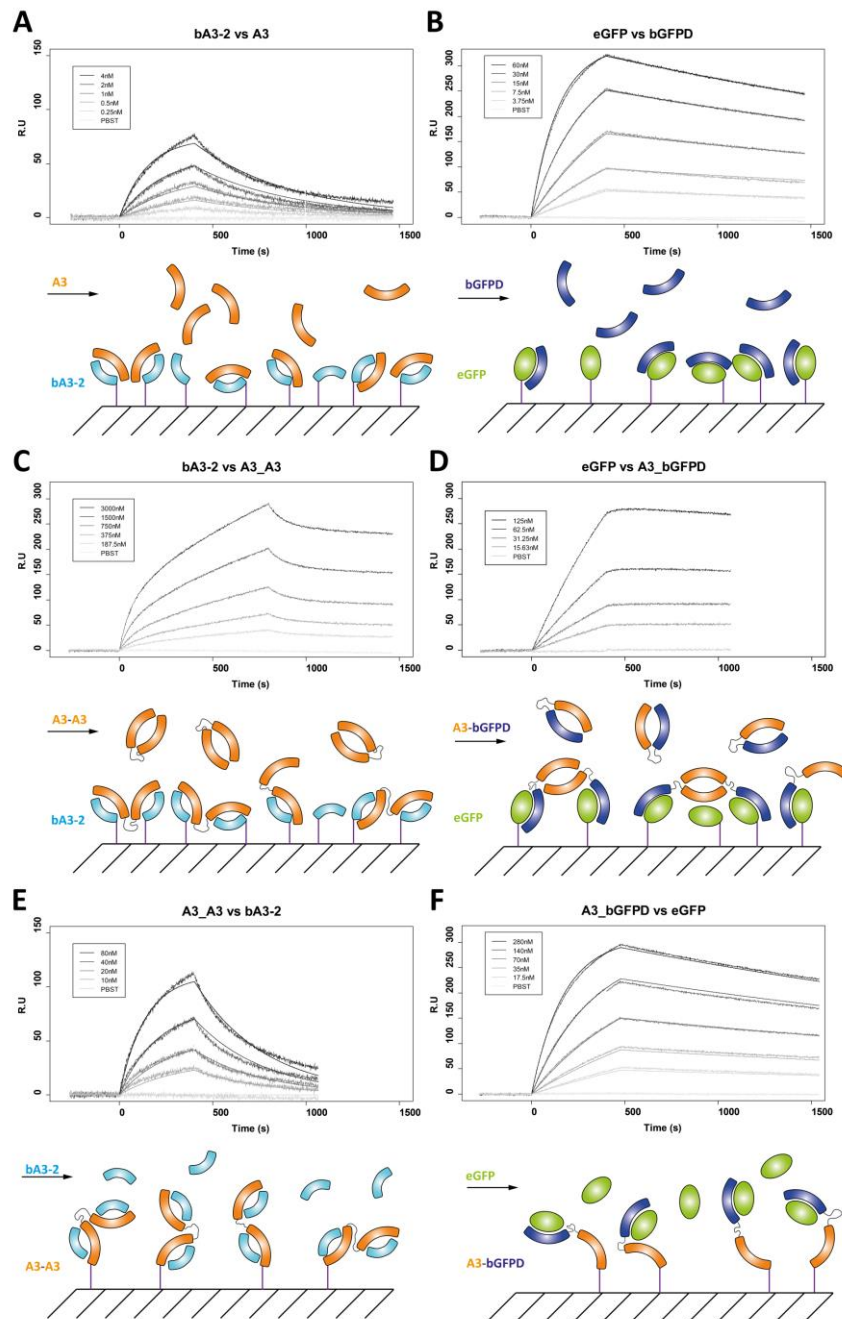


Figure 10 : Surface Plasmon Resonance analysis of interactions of α Reps (A3 and bGFPD) or chimeras (A3_A3 and A3_bGFPD) with their respective partners. His-tagged proteins are immobilized as ligands on Nickel-loaded sensor chips. Proteins used as analytes have no His-tag. A3 has a C-terminal Strep-Tag, bA3-2 a C-terminal Twin Strep-Tag, and the His-tag of the eGFP, A3_A3 and A3_bGFPD were cleaved by the TEV protease.

(A) Five concentrations of monomeric A3 protein (4 nM, 2 nM, 1 nM, 0.5 nM, 0.25 nM) were applied to 5 channels containing immobilized bA3-2. The data were fitted using the Langmuir model with $k_{on} = 1.3 \pm 0.1 \cdot 10^6 \text{ M}^{-1} \cdot \text{s}^{-1}$, $k_{off} = 1.9 \pm 0.1 \cdot 10^{-3} \text{ s}^{-1}$ and $K_D = 1.5 \pm 0.2 \text{ nM}$.

(B) Five concentrations of bGFPD (3.75; 7.5; 15; 30; 60 nM) were applied to immobilized eGFP. The data were fitted using the Langmuir model with $k_{on} = 1.1 \pm 0.1 \cdot 10^5 \text{ M}^{-1} \cdot \text{s}^{-1}$, $k_{off} = 2.5 \pm 0.1 \cdot 10^{-4} \text{ s}^{-1}$ and $K_D = 2.3 \pm 0.2 \text{ nM}$.

(C) Five concentrations of A3_A3 (188; 375; 750; 1500; 3000 nM) were applied to flow cell with immobilized bA3-2.

(D) Five concentrations of A3_bGFPD (15; 31; 62; 125; 250 nM) were applied to immobilized eGFP. The dissociation is slower for chimeras (C and D) than for the isolated domains, which is characteristic of an avidity effect. The curves could not be fitted using the langmuir model.

(E) Four concentrations of bA3-2 (10; 20; 40; 80 nM) were applied to immobilized A3_A3. The data were fitted using the Langmuir model with $k_{on} = 5.8 \pm 0.1 \cdot 10^4 \text{ M}^{-1} \cdot \text{s}^{-1}$ and $k_{off} = 2.7 \pm 0.1 \cdot 10^{-3} \text{ s}^{-1}$ and $K_D = 46.0 \pm 2.5 \text{ nM}$.

(F) Five concentrations of eGFP (17.5; 35; 70; 140; 280 nM) were applied to immobilized A3_bGFPD. The data were fitted using the Langmuir model with $k_{on} = 2.0 \pm 0.1 \cdot 10^4 \text{ M}^{-1} \cdot \text{s}^{-1}$ and $k_{off} = 2.5 \pm 0.1 \cdot 10^{-4} \text{ s}^{-1}$ and $K_D = 12.5 \pm 1.2 \text{ nM}$.

i) Structural analyses of the bidomains are in agreement with the interactions experiments

Structural studies of the two bidomains were undertaken in order to have a better understanding of their behavior. The crystal structures of the two chimeras were determined with a resolution of 1.94 and 2.79 Å for **A3_A3** and **A3_bGFPD** respectively. In neither structure electronic density was observed for the linker, which was expected as it was designed to be totally flexible. **A3_A3** crystallizes with one **A3** subunit in the asymmetric unit. SDS-PAGE of dissolved crystals showed that they contained the intact fusion protein (Figure 11).

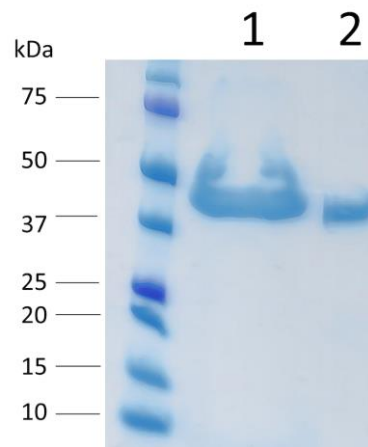


Figure 11 : SDS-PAGE analysis of the crystallized **A3_A3**. Line 1: One crystal was washed three times in water and dissolved in the denaturation buffer. Line 2: Purified protein before crystallisation.

The crystal structure of **A3_A3** (PDB ID 6FT5) is comparable to the previously described dimeric **A3** (PDB ID 3LTM, PDB ID 3LTJ) revealing that the linker does not affect the interface between the two **A3** domains (Figure 12-A). The structure analysis of the **A3** dimer previously solved showed that the two N-cap/C-cap are involved in the interface between the two subunits of the dimer (Urvoas, A., *et al.*, 2010). This suggests that this interface between caps of each monomer contributes to the dimer association. As the conformation of the **A3_A3** bidomain is fully superimposable to the **A3** dimer, the interactions involving the N-cap and C-cap are also present in the chimera and favor the close conformation.

The **A3_bGFPD** (PBD ID 6FSQ) bidomain crystals have a single **A3_bGFPD** in the asymmetric unit. The structure shows an open conformation but the two domains are in contact: the loop linking the consecutive repeats of the **A3** domain interact with the GFP binding surface of the second **bGFPD** domain. Importantly, the crystal symmetry generates a dimer with a neighbor molecule *via* its **A3** subunit. This dimer is identical to the **A3** dimer and to the **A3_A3** chimeric protein. SEC-MALS and SAXS results indicate the presence of an equilibrium between a monomeric and a dimeric form of the chimera and at the concentration used for crystallization, the dimeric form is favored. For the dimer observed in the crystal structure (Figure 12-B), two distinct interfaces are observed. The first one is the dimerization interface established between the two **A3** domains within a dimer, and the second involves the **A3** domain and the **bGFPD** domain of the same molecule. The dimerization interface between two **A3** domains (1170 Å²) is higher than the interface (638 Å²) between the **A3** and **bGFPD** domain of the same molecule. Furthermore, these two interfaces involve different types of residues. The dimeric interface between **A3** domains is established by contact between their variable side chains. In the inter-domain interface, the variable residues of the **bGFPD** domain are in contact with a non-variable part of the surface of **A3** (Figure 12-C). Since the non-variable surfaces of the **A3** and **bGFPD** domains are identical, these **A3/bGFPD** interactions, if stable enough, should also take place between two **bGFPD** molecules. This would lead to an oligomerization of the **bGFPD** protein, which is not observed in solution. Thus, the interaction between the **bGFPD** and **A3** domain observed in the crystal structure may not reflect the inter-domain interactions taking place in solution, for **A3_bGFPD**, but is rather a secondary consequence of crystallization of the dimeric form.

A SEC-SAXS experiment has been used to determine the conformation of **A3_bGFPD** in solution. The successive scattering patterns recorded at the end of the elution peak (Figure 6-A), corresponding to the monomer, are averaged. The resulting curve (Figure 6-C and 6-D) shows that **A3_bGFPD** is rather compact in solution ($R_g = 26.8$ Å, $D_{max} = 90$ Å), and could adopt different possible geometries of the closed forms.

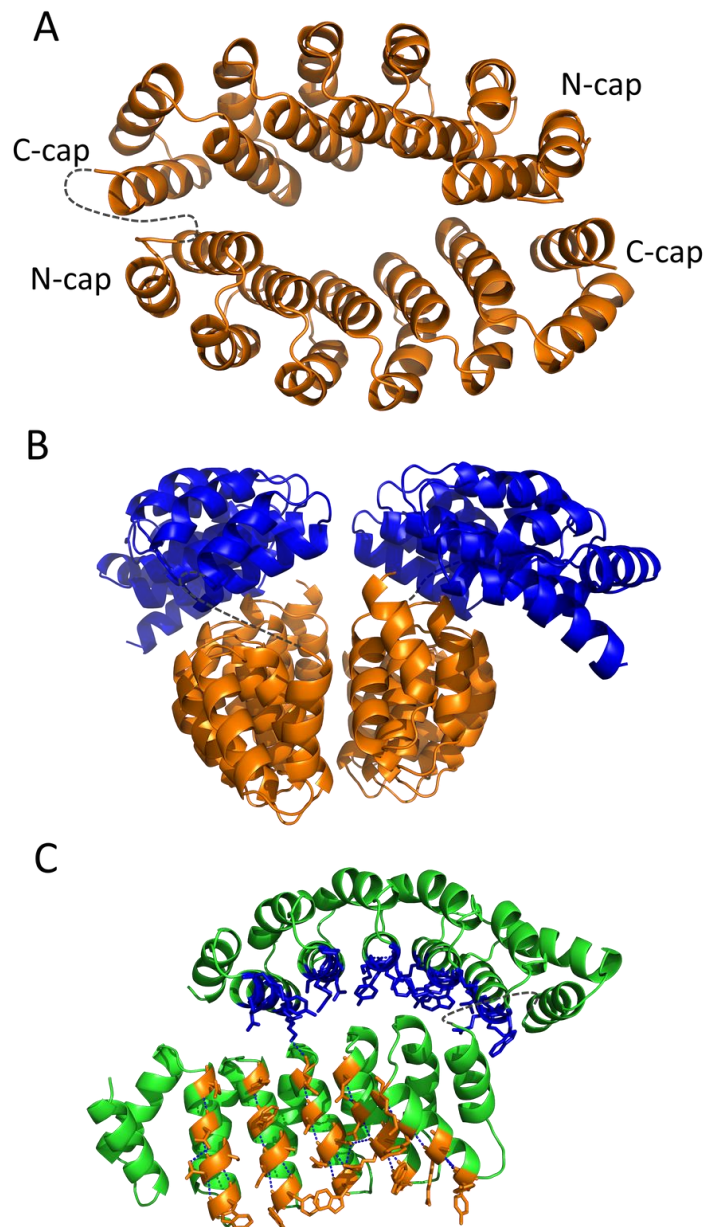


Figure 12 : Crystal structures of the bidomain proteins A3_A3 and A3_bGFPD.

(A) X-Ray structure of the protein A3_A3. The linker is not visible in the bidomain due to its flexibility.

(B) X-Ray structure of A3_bGFPD. A3 domains are in orange and bGFPD domains are in blue. Two chimeric molecules A3_bGFPD dimerize through their A3 domain (orange).

(C) Representation of the variable and constant residues in the structure of A3_bGFPD. Constant residues are in green and variable residues are in orange for the A3 domain and in blue for the bGFPD domain. In the crystal, variable residues (in blue) of the bGFPD domains are in contact with constants residues located on the A3 domain to form a T-shape object.

In all these structures, the linker is schematized with grey dashes.

Table 2: Crystallographic data collection and refinement statistics.*Crystallographic data collection*

	A3_A3	A3_bGFPD
X-ray source	ID29	ID29
Wavelength (Å)	0.97625	0.97
Temperature (K)	100	100
Unit-cell parameters (Å, °)	$a = 89.65, b = 46.55, c = 63.21, \alpha = 90.0, \beta = 128.93, \gamma = 90.0$	$a = 66.90, b = 88.90, c = 143.50, \alpha = \beta = \gamma = 90.0$
Space group	C2	C222
Resolution limits [†] (Å)	44.69 - 1.94 (2.06 - 1.94)	47.87-2.79(2.96-2.79)
Number of observations [†]	68413(10144)	139553(21035)
Number of unique reflections	15038(2340)	10938(1680)
R-meas [†] (%)	9.9(79.6)	21.4(128.3)
Completeness [†] (%)	97.6(94.7)	99.6(98.0)
I/σ^{\dagger} (I)	10.71(1.94)	12.95(2.08)
CC (1/2)	99.8(87.5)	99.7(89.7)

Refinement

	A3_A3	A3_bGFPD
Number of non-hydrogen atoms (protein/other)	1432/97	2829/8
R/R _{free} (%)	17.48/22.21	20.56/26.98
R.M.S.D. Bonds (Å)/angles (°)	0.018/1.855	0.010/1.339
Average temperature factors (protein/other)	44.47/61.80	62.68/61.51
Number of residues (asymmetric unit / chain)	183/409	384/409

[†] Values in parentheses refer to the highest resolution shell.

Accession number

Coordinates and structure factors have been deposited in the Protein Data Bank with accession number 6FT5 (for **A3_A3**) and 6FSQ (for **A3_bGFPD**).

j) Conformational transition consecutive to Ligand binding events can be detected by FRET

In order to monitor if binding events could give rise to a measurable fluorescence output, both bidomains were coupled with a FRET pair of fluorescent dyes. A unique C-terminal cysteine was covalently grafted with an “ATTO 488 maleimide” dye used as FRET donor and the N-terminal His-tag was coupled with a “NTA-ATTO 647N” dye used as a FRET acceptor. For both bidomains, (Figures 13-A and 14-C) a FRET signal (quenching of the donor fluorescence intensity and sensitization of the acceptor fluorescence intensity) between the two dyes is observed, as expected since the closed conformation of the chimera brings the two dyes into close proximity. The labeled chimeras were titrated with their respective ligands (**bA3-2** for both chimeras and **BFP** for **A3_bGFPD**). The blue fluorescent variant **BFP** was chosen as an alternative ligand of **bGFPD**, rather than **eGFP** used in the other experiments, in order to avoid the overlap in the absorbance spectra with **eGFP** (Wachter, R.M., *et al.*, 1997) (Figure 14-A). Control experiments indicated that the affinities of **BFP** and **eGFP** for **bGFPD** are comparable, as assessed by ITC (Figure 14-B). In all cases, the FRET sensitized acceptor emission signal at 664 nm decreased with increasing ligand concentrations (Figure 13-B, 13-C and 13-D), because ligand binding pushes the chimeric proteins into an open conformational transition consecutive to ligand binding and thereby increases the donor-acceptor distance, which leads to a lower FRET efficiency and a concomitant decrease of acceptor emission intensity. For **A3_bGFPD**, the K_D values obtained for both ligands are in the same range as the values obtained by ITC, while for **A3_A3**, the K_D value is significantly higher than the one determined by ITC (Table 1). However, these two experiments do not monitor the same binding events: in the ITC experiment the binding of each ligands on the protein is monitored, whereas in the FRET experiment the signal monitors the binding of the first ligand molecule that opens the chimera and consequently disrupts the FRET between the donor and acceptor. As observed with **A3_bGFPD**, the FRET signal is abolished upon binding of the first Ligand molecule. This suggests that for, the **A3_A3** chimera, the effect of the second bound ligand on the FRET signal is very limited and do not contribute to the estimation of the apparent K_D .

Overall, these fluorescence data clearly show that the conformational changes suggested by the ITC and SPR results can be directly monitored by FRET variation with donor / acceptor pairs located at each extremity of the chimeric proteins.

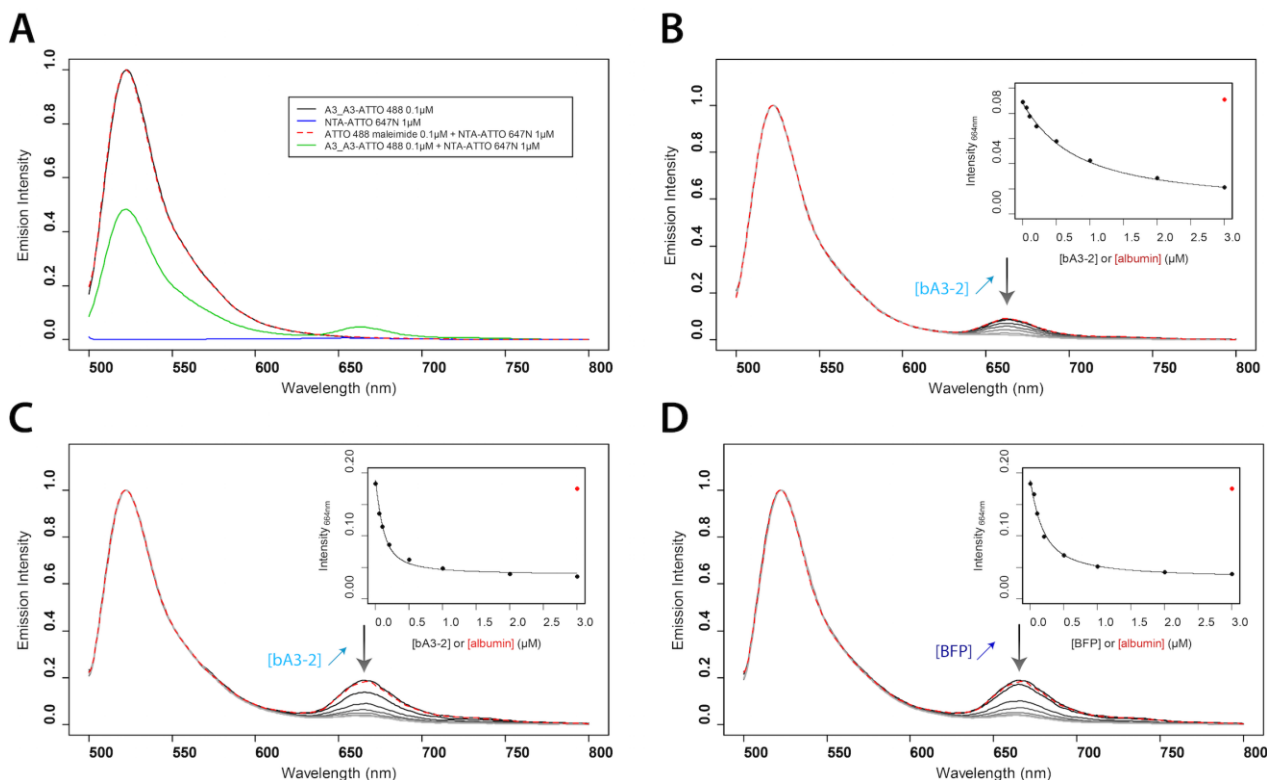


Figure 13 : FRET measurements assessing conformational changes of the bidomains A3_A3 and A3_bGFPD.

A3_A3 or A3_bGFPD bidomains are conjugated with two dyes, NTA-ATTO 647N at the N-Terminal His-tag (Acceptor) and ATTO 488 maleimide covalently attached on a C-terminal cysteine (Donor). 1 μ M ATTO 488-bidomain were mixed with 1 μ M NTA-ATTO 647N. Ligands bA3-2 and BFP were used at concentrations of 0.05, 0.1, 0.2, 0.5, 1, 2 and 3 μ M. without His-Tag. Albumin was used as a control at a concentration of 3 μ M. All emission spectra were recorded with an excitation at 488 nm.

A) Fluorescence emission spectra of A3_A3-ATTO 488 (0.1 μ M) (black line), NTA-ATTO 647N (1 μ M) (blue line), ATTO 488 maleimide (0.1 μ M) mixed with NTA-ATTO 647N (1 μ M) (red dashed line) and a mixture of A3_A3-ATTO 488 (0.1 μ M) and NTA-ATTO 647N (1 μ M) (green line). The emission signal at 662 nm, only for the dually labelled protein suggests fluorescence transfer between the two dyes in the bidomain.

B-C-D) Normalized emission spectra of the dually labelled A3_A3 (B) and A3_bGFPD (C-D) with different concentrations of bA3-2 (B-C) and BFP (D). The emission peak of ATTO 647N (664 nm) decreases for both bidomains in the presence of increasing concentrations of bA3-2 or BFP. Albumin (red dashed line) has no effect on the fluorescence. (Inset) Emission signal at 664 nm, in the presence of increasing ligand concentrations (bA3-2 or BFP) was fitted as a function of total ligand concentration (K_D values reported in Table 1).

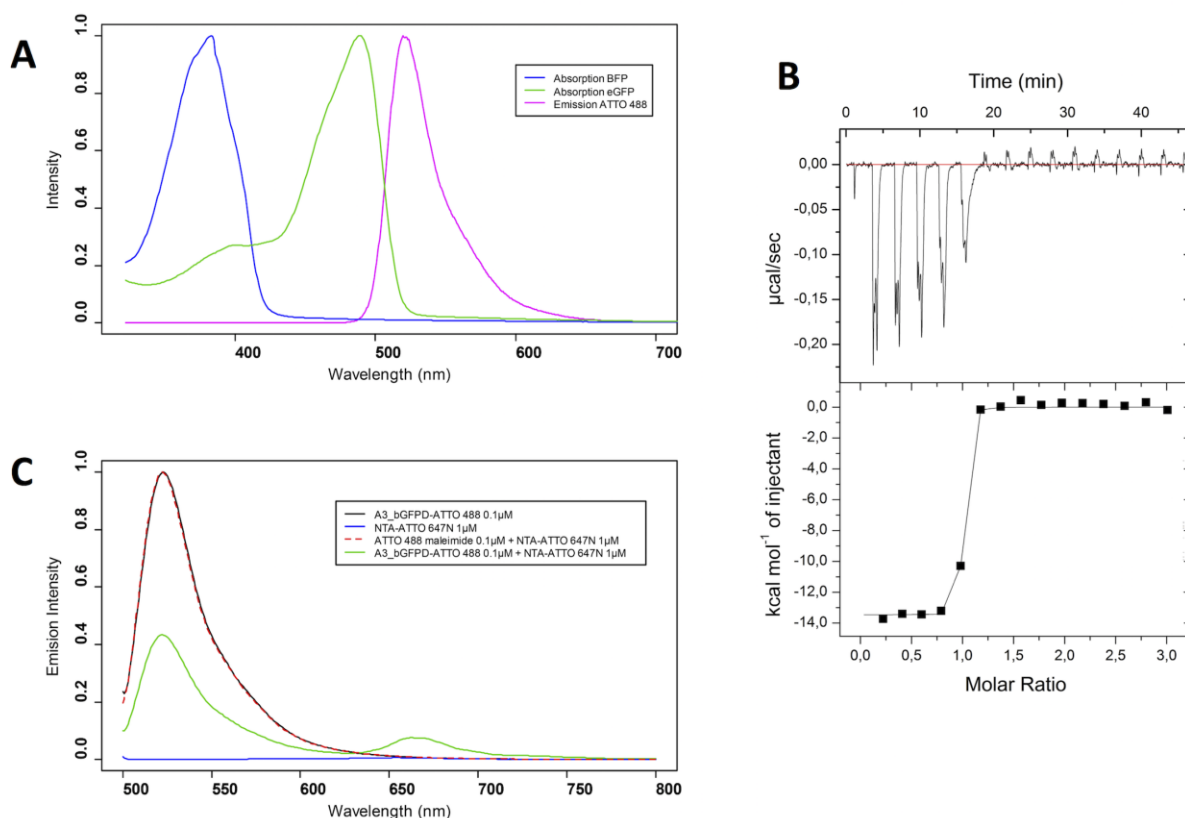


Figure 14 : Control experiments for the FRET assays using A3_bGFPD

A) Normalized absorption spectrum of BFP (blue lines), eGFP (green line) and emission spectrum of ATTO 488 (pink line). Absorption spectra were obtained using a Perkin Elmer Lambda 35 spectrometer. Emission spectrum of the atto488 was obtained via the manufacturer website. There is an overlap around 500 nm between eGFP and ATTO 488. Therefore, the Blue fluorescent Protein (BFP), a variant of eGFP without spectral overlap was selected as a ligand in order to monitor FRET variations, without interferences due to the ligand.

B) ITC titration of a solution of BFP (15.7 μM) with bGFPD (300 μM). These data indicate that the BFP bind to bGFPD with the same affinity as eGFP. The stoichiometry n is equal to 0.94 ± 0.01 , the K_D is 5 ± 2 nM.

C) Fluorescence emission spectra (excitation at 488 nm) of A3_bGFPD-ATTO 488 (0.1 μM) (black line), NTA-ATTO 647N (1 μM) (blue line), ATTO 488 maleimide (0.1 μM) mixed with NTA-ATTO 647N (1 μM) (red dashed line) and a mixture of A3_bGFPD-ATTO 488 (0.1 μM) and NTA-ATTO 647N (1 μM) (green line). The emission signal at 662 nm results from fluorescence transfer between the two dyes in the bidomain.

4) Discussion

a) Chimeric bidomain α Reps adopt a closed form

The SEC-MALS and SAXS results clearly show that, at low micromolar concentration both bidomains adopt a closed bivalve conformation. This was expected for **A3_A3**: when not linked together the two **A3** monomers dimerize with relatively high affinity ($K_{D \text{ A3/A3}} = 37 \text{ nM}$, ultra-centrifugation data, [Table 1](#)), and in the **A3_A3** bidomain, their mutual interactions are further stabilized by their proximity. The closed form of **A3_bGFPD** was not anticipated. Indeed, since the **bGFPD** domain was initially selected to interact specifically with **eGFP**, it was *a priori* not expected to bind the **A3** domain. Nonetheless, ITC and AUC experiments revealed that these two unrelated domains do interact, although with a low affinity ($K_{D \text{ A3/bGFPD}} \approx 4 \mu\text{M}$). However even this relatively weak affinity between domains is sufficient to take place predominantly within a bidomain protein. This fortuitous and relatively weak interaction between **A3** and **bGFPD** moieties may in fact be due to the interactions between the N-cap of the **A3** domain and the C-cap of the **bGFPD** domain. All α Reps share the same C-cap repeat. The **A3** dimer structure clearly indicates that the specific sequence of the **A3** N-cap can interact with the common C-cap repeat and this could contribute to the observed affinity.

b) Ligand induced structural transitions in multi-domains α Reps result in functional effects

The binding surfaces of the bidomain are covering the interior of the bivalve-like structure. The closed forms are therefore not structurally compatible with the presence of bound partners. However, for both bidomains, each domain can still bind its cognate partners, when bidomain proteins undergo a structural transition upon ligand binding.

The *equilibria* between the open and closed conformations for these bivalve proteins modulate their binding affinities for their respective ligands in a cooperative process. This is particularly clear for the **A3_A3** protein, which still binds its ligand although with an apparent lower affinity than **A3**. This low affinity increases if the bidomain protein is pre-opened with another competitive ligand, which is characteristic of positive homotropic cooperativity (Figure 15-A). Similarly, the results suggest that positive heterotropic cooperativity is observed with the **A3_bGFPD** protein when the two ligands (**bA3-2** and **eGFP**) bind together (Figure 15-B). Indeed, when **eGFP** is bound to the chimera first, the affinity of **bA3-2** is increased by a factor of 3.6. When **bA3-2** is bound first, the affinity of **eGFP** increases by a factor of 4.4. However, in neither case do the pre-opened forms recover the affinity observed for the same ligand on the monomeric domain. This suggests that, if the open/closed equilibrium contributes to change the affinity of concatenated α Rep domains, the open form cannot bind the ligand as efficiently as the individual domain.

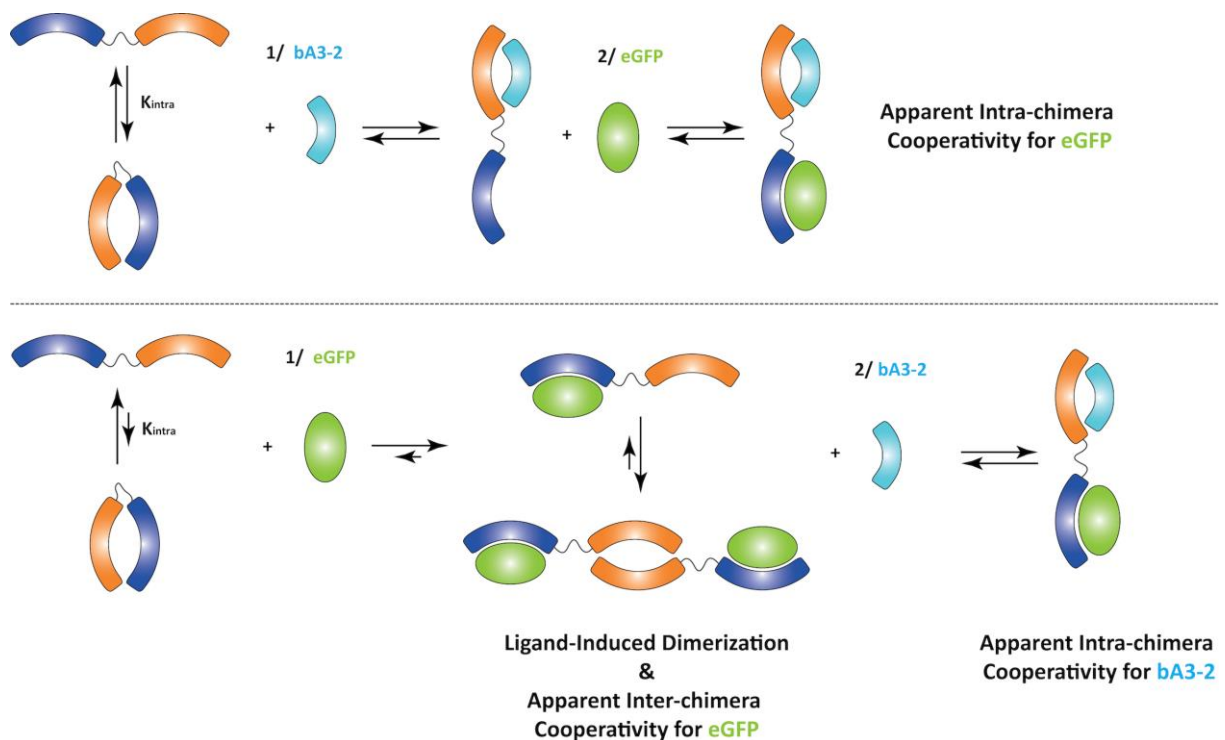


Figure 15 : Model for the binding of the two **A3_bGFPD** ligands whether **bA3-2** or **eGFP** binds first.

The binding of the **eGFP** tends to form a dimeric chimera by releasing **A3** domains in a ligand-induced cooperativity process. When ligands are in excess, the complexes are more prone to form heterotrimers, each domain binding its respective ligand (**A3/bA3-2** and **bGFPD/eGFP**). Such a process could be generalized to develop biosensors with any α Rep binding a specific target of interest.

c) Ligand induced dimerization

SEC-MALS and SPR experiments show that it is possible to induce the dimerization of the **A3_bGFPD** bidomain *via* the **A3** domains by adding **eGFP**. When **eGFP** is immobilized, this dimer of bidomains can therefore bind two anchor points on the chip, thus inducing an avidity effect. This process can considerably slow down the dissociation of the chimeric proteins from the binding surface. This "cooperavidity" effect can be expected for any chimeric protein made from a dimerization domain shielded by a ligand-binding domain. The ligand-binding event releases the dimerization domain, and if the ligand density is sufficient, this in turn induces an avidity effect. The closed form, dominant in absence of the target, should minimize accessibility of the binding surface but dimerization consecutive to binding should strongly favor retention of the bound bidomain on a surface specifically where two target molecules are in close proximity. This effect could be potentially useful for example to enhance contrast in cellular imaging.

d) Potential applications of complex chimeras such as sensors and switches

The experiments described here were conducted with specific α Rep domains (**A3** and **bGFPD**) but the underlying principles can be more generally useful. α Rep binders for various unrelated proteins have already been generated (Ferrandez, Y., *et al.*, 2014; Figueroa, M., *et al.*, 2016; Guellouz, A., *et al.*, 2013; Hadpech, S., *et al.*, 2017; Tiouajni, M., *et al.*, 2014). The binding surface is invariably located, as expected, on the randomized surface. All α Reps share the same general fold and could therefore form a closed bivalve like bidomain structure when fused to an **A3** or **A3**-like closing domain. The "open/closed bivalve effect" described here could therefore possibly be observed on many other α Rep-based recognition unit.

As shown in the FRET assays for these two “model” bidomains, it is clearly possible to monitor the open/closed conformational change into a variation of fluorescence signal correlated to ligand binding. Such a closed/open transition in a bivalve-like protein could later prove useful as a general way to transduce the target-binding event into a measurable quantitative readout. Here, it will be possible to tailor bidomains with domains that bind a target of choice, which is a substantial improvement over existing sensors.

Moreover the cooperative effects observed between the binding partners of a chimeric protein made from two different α Reps with different specificities could also be a basis to design future molecular switches. The structural transition induced by the ligand of one domain can enhance the binding on the second domain of a different, unrelated partner.

D. Conclusion

Cette première étude a permis de montrer qu'il était possible de créer de nouvelles fonctions simplement en ajoutant un *linker* entre deux domaines.

En effet, l'ajout de ce lien covalent a permis de créer un phénomène de coopérativité de fixation des ligands. La fixation du premier ligand va permettre de faciliter la fixation du second. Ce phénomène s'explique par le fait que les deux bidomaines adoptent plusieurs conformations dont l'équilibre est défini par la constante K_{intra} . Cet équilibre tend très largement vers une conformation fermée. La fixation du premier ligand va déplacer l'équilibre vers des conformations ouvertes et ainsi faciliter la fixation du second ligand. Bien que ce phénomène soit surtout visible sur le bidomaine **A3_bGFPD**, ce phénomène est nécessairement présent pour le bidomaine **A3_A3**.

D'autre part, nous avons également pu observer la création d'un phénomène d'avidité en SPR pour les deux bidomaines, ce phénomène pourrait être exploité pour augmenter de manière simple l'affinité de molécules dans le cas de reconnaissance de surface (cellules cancéreuse). Cette propriété pourrait être utilisée afin d'effectuer du tri cellulaire en ajoutant un fluorophore par exemple. Ce phénomène d'avidité est par ailleurs inductible dans le cas du bidomaine **A3_bGFPD**. En effet, lorsque le bidomaine se fixe sur une surface recouverte de la **eGFP**, les domaines **A3** dimérisent ce qui entraîne le phénomène d'avidité.

Le bidomaine **A3_bGFPD** est par ailleurs bispécifique, cette propriété pourrait être utilisée pour rapprocher deux cellules différentes, par exemple une cellule cancéreuse et un lymphocyte T.

Enfin, la modification des bidomaines par ajout de deux fluorophores a prouvé qu'il était possible de détecter l'ouverture par fluorescence et par conséquent la présence de ligand. Sachant désormais la tendance qu'ont des bidomaines à adopter une forme fermée, cette capacité de détection pourrait être généralisée à d'autres cibles *via* des sélections au sein de la bibliothèque α Rep. Cependant les concentrations en bidomaine utilisées sont assez élevées (de l'ordre du μM) et la sensibilité de ce système semble être limitée. Une optimisation semble donc nécessaire, notamment par un changement des fluorophores.

II. Optimisation des biosenseurs génériques

A. Introduction

La détection de cible présente un intérêt majeur que cela soit à des fins thérapeutiques ou de recherche. Actuellement, la détection et la quantification de molécules non immunogènes sont difficiles. Dans le cas de cibles immunogènes, les anticorps et leurs dérivés sont généralement utilisés. Ceux-ci présentent cependant de nombreux désavantages : coût, bien-être animal et séquence parfois aléatoire (anticorps polyclonaux). (Bradbury, A. and Pluckthun, A., 2015) Il est donc crucial de développer des alternatives et si possible des alternatives généralisables à n'importe quelle molécule.

Ayant vu que l'ouverture des bidomaines était détectable par fluorescence à l'aide d'un couple de fluorophores *atto*, j'ai entrepris l'optimisation de ces dernières. D'une part, afin de diminuer les quantités de protéines nécessaires et d'autre part afin d'augmenter la taille des distances mesurables (par le FRET) et la sensibilité de détection. Pour cela, j'ai changé le couple de fluorophores utilisés et évalué les nouvelles propriétés des bidomaines **A3_A3** et **A3_bGFPD**. Le nouveau couple choisi est constitué d'un complexe de Terbium en tant que donneur et d'un Quantum Dot (QD) comme accepteur. Le nombre de bidomaines se fixant sur chaque QD, les limites de détection pour **bA3-2** et la **BFP** ainsi que les distances d'ouverture ont ainsi pu être obtenus.

Si de telles constructions se révèlent efficaces dans la détection de **bA3-2** et de la **BFP** et connaissant les propriétés des α Reps, il est concevable de penser que ce système puisse être adapté à la détection d'autres cibles. En effet, nous disposons d'une bibliothèque d' α Rep parmi laquelle il est possible d'isoler des protéines spécifiques reconnaissant des cibles définies, de plus les bidomaines semblent adopter des conformations fermées. Ainsi le simple ajout des fluorophores serait suffisant pour réaliser des biosenseurs spécifiques.

B. Mise au point

1) Choix des fluorophores

Nous avons précédemment montré que les bidomaines étaient capables de changer de conformation suite à la fixation d'un ou plusieurs ligands. Ce changement de conformation peut être suivi par fluorescence mais des optimisations sont nécessaires afin de rendre ces biosenseurs plus sensibles. En effet, pour un couple de fluorophores donné, la sensibilité est la plus haute entre 0.5 et 2 fois le rayon de Förster, appelé R_0 . Ce R_0 correspond à la distance entre donneur et accepteur pour laquelle l'efficacité du FRET est de 50%. Il dépend de l'orientation des fluorophores l'un par rapport à l'autre, du rendement quantique du donneur, de l'indice de réfraction du milieu et de l'intégrale de recouvrement. Nous verrons plus en détail ces paramètres dans la partie suivante.

L'efficacité de FRET est définie par l'équation suivante :

$$\eta_{FRET} = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6}$$

Équation 1 : Efficacité du FRET en fonction de r et R_0 . η_{FRET} est l'efficacité de FRET, r est la distance en nm entre le donneur et l'accepteur, R_0 le rayon de Förster en nm

Dans le cas des atto-dyes utilisés précédemment, le R_0 est de 4.9 nm, permettant donc d'avoir une haute sensibilité entre 2.45 et 9.8 nm. Or, d'après les structures des deux bidomaines, la distance entre les deux extrémités des bidomaines est d'environ 10 nm dans le cas d'une ouverture totale. Les atto-dyes ne sont donc pas adaptés pour une détection optimale des ligands. J'ai par conséquent changé les deux fluorophores afin d'obtenir un couple avec un R_0 plus grand d'une part et permettant également d'effectuer une mesure sans avoir de bruit de fond de fluorescence résiduelle.

Les fluorophores de ce nouveau couple sont le Terbium (Tb) et le Quantum Dot (QD). Le terbium est un lanthanide aux propriétés de fluorescence intéressantes car il présente des pics d'émission dans le visible avec une longue durée de vie de fluorescence. Cependant, absorbant très mal les photons, il doit être inséré dans une cage qui joue le rôle d'antenne. La cage choisie est la Lumi4, elle va complexer le Terbium pour former un complexe de Terbium (L4-Tb) visible dans la figure ci-dessous.

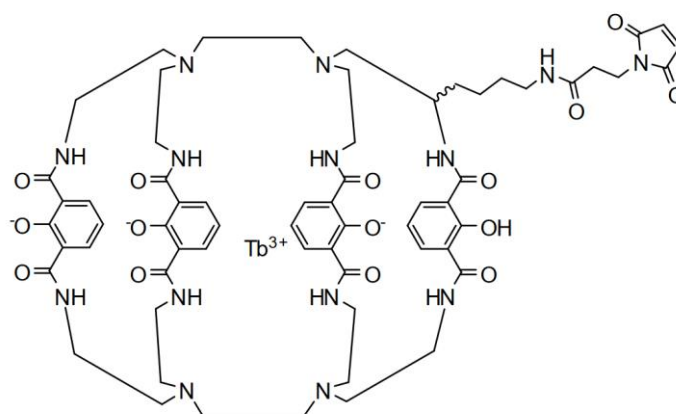


Figure 16 : Structure chimique du Lumi4 - maléimide couplé au Terbium (Tb³⁺).

Les QD choisis possèdent un maximum d'émission à 625 nm (QD625) et leur surface est constituée de Zinc et de Soufre. Le L4-Tb joue le rôle de donneur et le QD625 est l'accepteur. Ces deux fluorophores possèdent des propriétés remarquables pour les mesures de FRET. Tout d'abord, l'émission du L4-Tb a une durée de vie de l'ordre de la milliseconde ce qui est nettement plus long que la fluorescence habituellement mesurée avec d'autre fluorophore qui est de l'ordre de la nanoseconde. Il est donc possible de réduire le bruit de fond de la mesure à néant en ne mesurant le FRET qu'après un délai d'une cinquantaine de microsecondes. Les QD625 quant à eux, ont un spectre d'absorption recouvrant presque l'intégralité du spectre d'émission du L4-Tb et sont quasiment insensibles au photoblanchiment. La durée de vie de la fluorescence des QDs n'étant que de quelques dizaines de nanosecondes, la mesure d'un signal de fluorescence des QDs au-delà de 50 μ s traduira directement du FRET. Pour cette raison, le signal de fluorescence sera mesuré entre 50 et 1000 μ s, c'est ce que l'on appelle une mesure résolue en temps. (Figure 17)

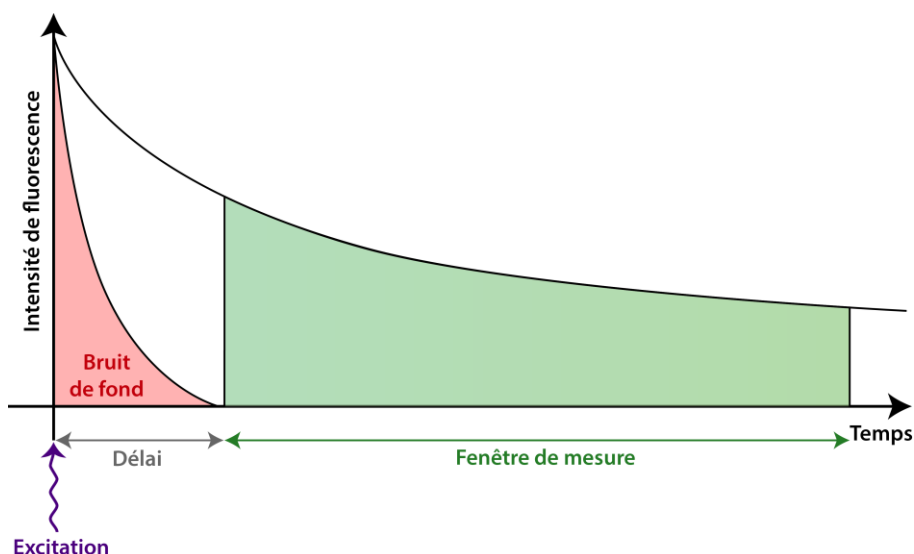


Figure 17 : Principe de d'une mesure de fluorescence résolue en temps. La mesure est effectuée après un délai qui permet de s'affranchir du bruit de fond.

Les couples lanthanide-QD possèdent en général un R_0 d'environ 10 nm permettant donc d'avoir une haute précision dans la mesure entre 5 et 20 nm. Les QD625 ayant un rayon moyen de 5.1 nm, les distances permettant d'obtenir une variation de FRET seront environ entre 0 et 15 nm. Ce couple est donc parfaitement adapté pour être utilisé avec nos bidomaines. (Figure 18)

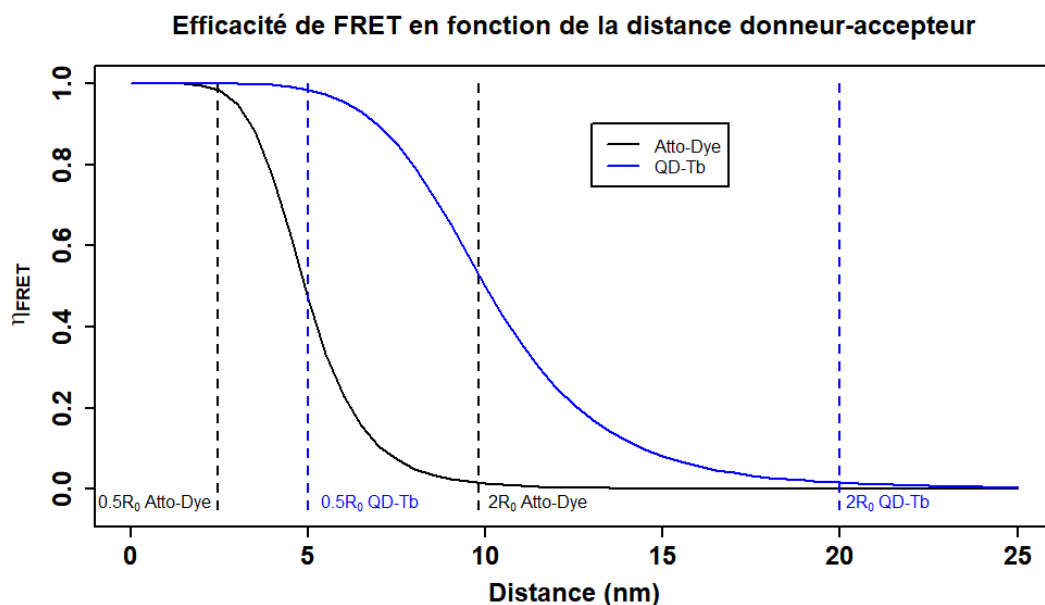


Figure 18 : Efficacité de FRET (η_{FRET}) définie par l'équation $\eta_{FRET} = 1 / (1 + (r/R_0)^6)$ en fonction de la distance entre donneur et accepteur pour deux couple de fluorophore. En pointillé sont visibles pour chacun des couples les valeurs pour $R = 0.5$ et 2 fois le R_0 .

2) Détermination du rayon de Förster du couples L4-Tb / QD625

Il est dans un premier temps nécessaire de rappeler quelques équations permettant d'obtenir le rayon de Förster R_0 du couple L4-Tb / QD625 :

$$R_0 = \left(\frac{9(\ln 10) \kappa^2 \Phi_D}{128 \pi^5 N_A n^4} J \right)^{1/6}$$

Équation 2 : Rayon de Förster R_0 (nm). κ^2 est le facteur d'orientation des fluorophores. Φ_D est le rendement quantique du donneur. N_A le nombre d'Avogadro (mol^{-1}). n l'indice de réfraction du milieu. J l'intégrale de recouvrement ($\text{M}^{-1} \cdot \text{cm}^{-1} \cdot \text{nm}^4$).

L'intégrale de recouvrement J (en $\text{M}^{-1} \cdot \text{cm}^{-1} \cdot \text{nm}^4$) permet de visualiser le recouvrement d'un couple Donneur/Accepteur en prenant en compte l'émission du donneur et l'absorption de l'accepteur et cela à chaque longueur d'onde. Elle est obtenue avec la formule suivante :

$$J = \int f_D(\lambda) \varepsilon_A(\lambda) \lambda^4 d\lambda$$

Équation 3 : Intégrale de recouvrement J . $f_D(\lambda)$ fluorescence du donneur normalisée selon l'intégrale de fluorescence à chaque longueur d'onde. $\varepsilon_A(\lambda)$ Coefficient d'absorption molaire de l'accepteur pour chaque unité de longueur d'onde ($\text{M}^{-1} \cdot \text{cm}^{-1}$). λ Longueur d'onde (nm). Le calcul de cette intégrale est effectué uniquement sur les longueurs d'ondes présentant un intérêt pour le FRET et non de 0 à l'infini.

Avec :

$$\int f_D(\lambda) d\lambda = 1$$

Équation 4 : Fluorescence du donneur normalisée selon l'intégrale de fluorescence.

Il a donc été nécessaire dans un premier temps, d'obtenir les spectres d'émission du L4-Tb (excitation à 337 nm) et le spectre d'absorbance des QD625, cela dans le but d'avoir la fluorescence normalisée de l'accepteur (L4-Tb) et le coefficient d'absorption molaire des QD625 pour chaque longueur d'onde. Il a ensuite été possible d'obtenir dans un premier temps la courbe de recouvrement spectral et donc l'intégrale de ce recouvrement (overlap) (Figure 19).

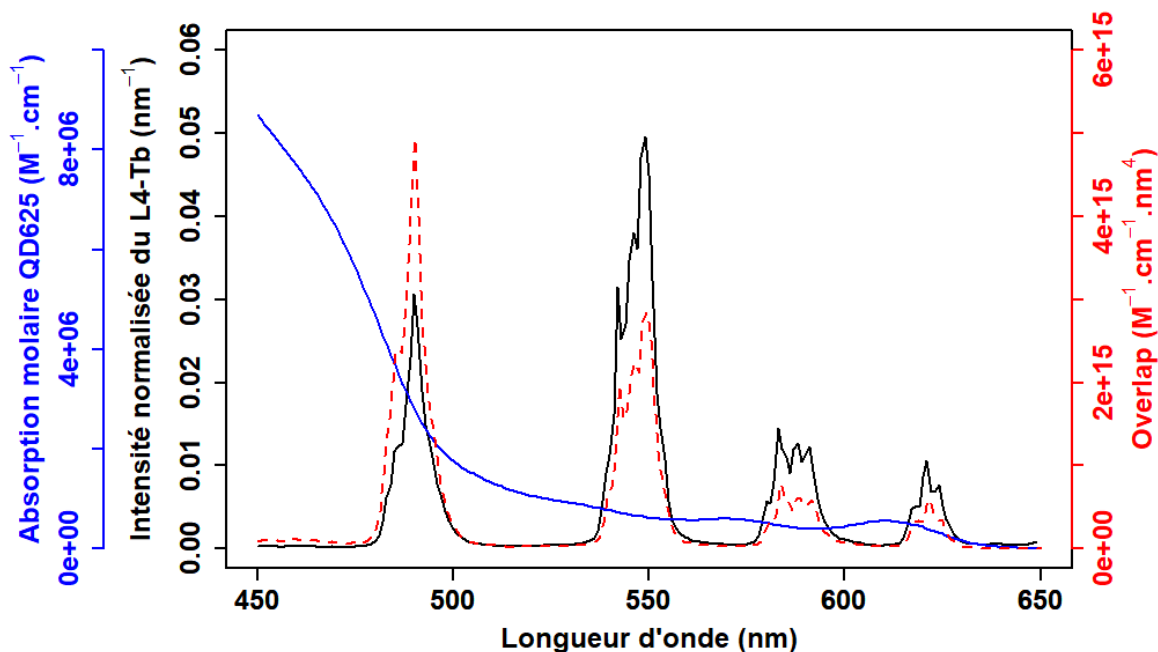


Figure 19 : Intensité de fluorescence normalisée du L4-Tb (nm⁻¹), Coefficient d'absorption molaire des QD625 (M⁻¹.cm⁻¹) et Recouvrement spectral (M⁻¹.cm⁻¹.nm⁴) en fonction de la longueur d'onde.

À partir de la courbe de recouvrement spectral, on peut extraire l'intégrale du recouvrement : $J = 8.414 \cdot 10^{16} \text{ M}^{-1} \cdot \text{cm}^{-1} \cdot \text{nm}^4$. Connaissant les valeurs de κ^2 , Φ_D , N_A et n , respectivement 0.667, 0.81, $6.022 \cdot 10^{23} \text{ mol}^{-1}$ et 1.35, et d'après l'équation 2, le R_0 est de 10.3nm. Ce R_0 et donc le couple QD625 / L4-Tb sont adaptés pour effectuer une détection de l'ouverture des bidomaines.

3) Choix des ligands

L'absorption de l'eGFP (spectre vert foncé) recouvrant en partie le spectre d'émission du terbium (spectre vert clair), elle ne peut être utilisée pour les expériences de FRET. En effet, elle entrerait en compétition avec les QD625 en tant qu'accepteur pour le donneur L4-Tb. Pour cette raison, la BFP (spectre bleu) a une nouvelle fois été utilisée (Figure 20).

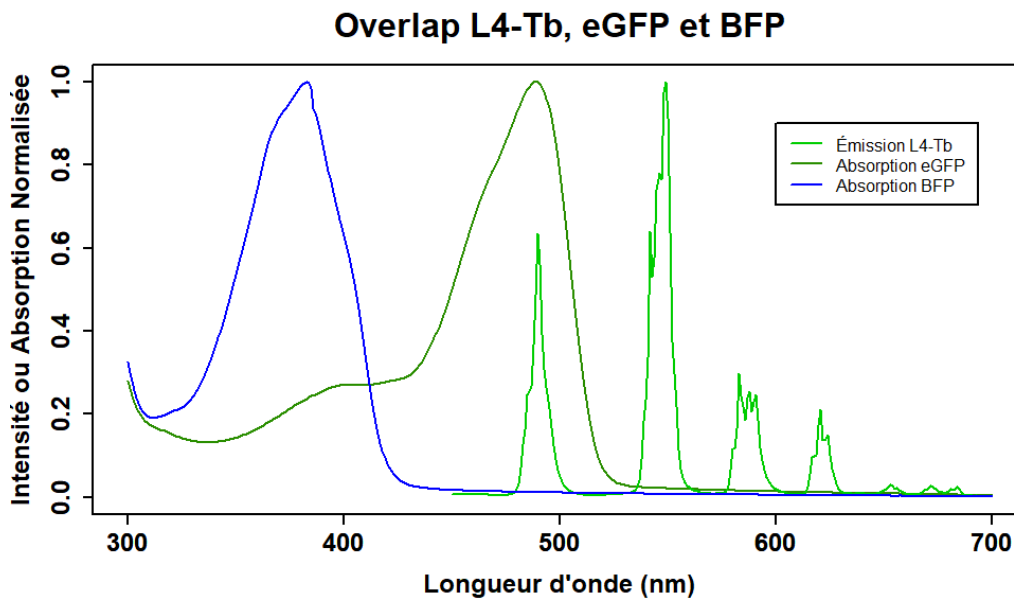


Figure 20 : Spectre d'émission du L4-Tb et spectre d'absorption de la BFP et eGFP. Le spectre de la eGFP se superpose au premier pic d'émission du L4-Tb, par conséquent, elle ne peut être utilisée pour les expériences de FRET.

4) Choix des longueurs d'onde suivies

Le choix des longueurs d'onde pour caractériser le FRET qui s'établit entre les L4-Tb et les QD625 s'est porté sur 494 nm et 640 nm. En effet, bien qu'absorbant à 337 nm, la durée de vie de la fluorescence des QDs n'entraînera pas d'émission due à l'excitation directe du laser à cette longueur d'onde. Bien qu'ayant un maximum d'émission à 625 nm, le suivi de l'émission des QDs s'est fait à 640 nm car le L4-Tb émet également à 625 nm, empêchant un suivi du signal des QDs uniquement. Les déterminations du nombre de protéines par QD et des limites de détections ont été effectuées sur un lecteur microplaque « Kryptor » (*Thermo Fisher*) équipé des filtres 494 ± 12 et 640 ± 10 nm. La détermination des distances a été effectuée avec un lecteur microplaque « *DreamReader* » (*Edinburgh Instruments*) équipé des filtres 494 ± 20 et 640 ± 20 nm. (Figure 21)

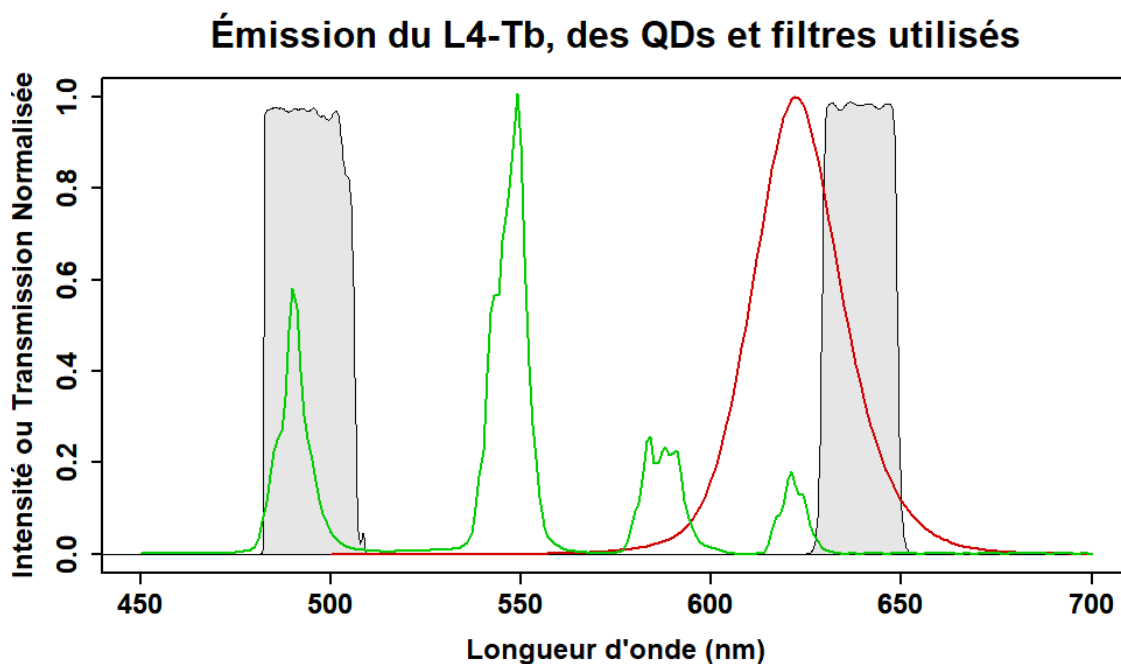


Figure 21 : Émission du L4-Tb (vert), des QDs (rouge) et filtres (gris clair) utilisés sur le Kryptor.

C. Couplage des bidomaines aux fluorophores

1) Couplage du Lumi4 et du Terbium

Le couplage se fait en trois étapes. Tout d'abord, le lumi-4 étant couplé à un maléimide, il réagit (comme ce fut le cas pour l'Atto-488) avec une stœchiométrie 1 : 1 avec la cystéine unique présente en C-terminal des bidomaines. Un excès de Lumi4 est ajouté à raison de quatre Lumi4 par protéine. L'excès de Lumi4 est enlevé grâce à un passage consécutif de l'échantillon sur deux colonnes de dessalage (Zebaspin). Par la suite, du chlorure de terbium (Cl_3Tb) est ajouté en excès, le terbium va se fixer dans le Lumi4 et donner le L4-Tb. L'excès de Tb non fixé n'émettra pas de fluorescence significative et ne perturbera pas les mesures. (Figure 22)

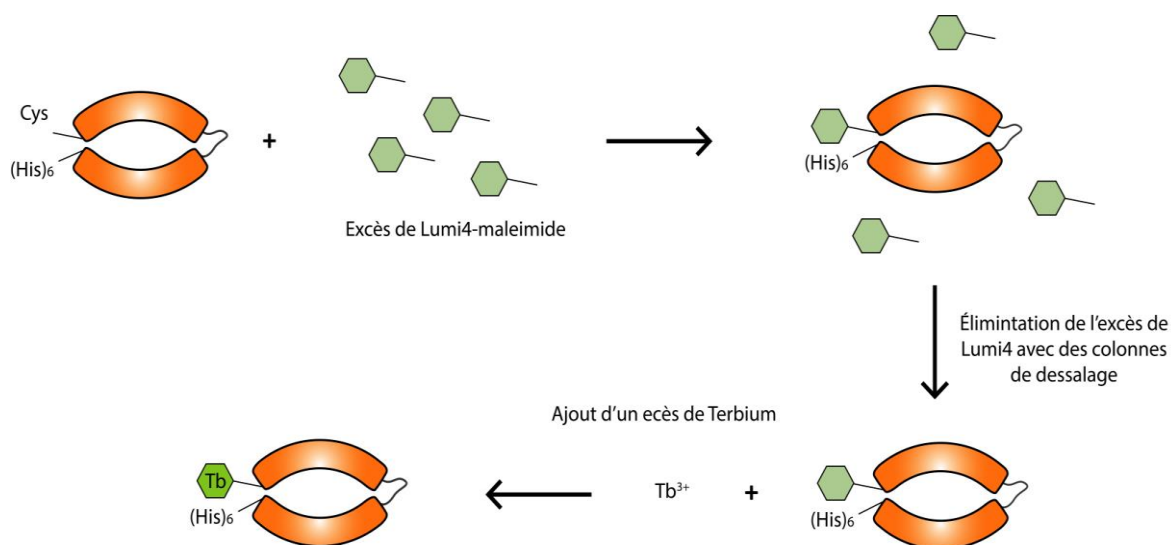


Figure 22 : Processus de couplage des bidomaines avec les Lumi4, le Terbium et les QDs. Les Lumi4 se greffent en C-ter des bidomaines sur la cystéine unique. Un excès de terbium est ensuite ajouté, il sera intégré dans le complexe Lumi4.

Afin de déterminer le taux de couplage du L4-Tb avec les protéines, les spectres d'absorption des protéines (**A3_A3** et **A3_bGFPD**) non couplées, du L4-Tb seul et des protéines couplées au L4-Tb ont été mesurés de 250 à 400 nm. Le maximum d'absorption des protéines est à 280 nm, celui du L4-Tb à 349 nm. Les protéines n'absorbant pas à 349 nm, il est possible de calculer un facteur de correction de l'absorbance à 280 nm (FC_{280}) pour le L4-Tb. Ce facteur permet de soustraire la participation du L4-Tb à l'absorbance à 280 nm et donc de n'obtenir que celle due à la protéine. Ce facteur de correction s'obtient par la formule suivante :

$$FC_{280} = \frac{A_{L4-Tb\ 280}}{A_{L4-Tb\ 349}}$$

Équation 5 : Facteur de correction de l'absorbance à 280 FC_{280} pour le L4-Tb. $A_{L4-Tb\ 280}$ et $A_{L4-Tb\ 349}$ sont les absorbances du L4-Tb à 280 nm et 349 nm respectivement.

Le facteur de correction obtenu ainsi est de 0.20. En appliquant l'équation suivante, il est possible d'obtenir l'absorbance à 280 nm corrigée de la protéine et donc sa concentration.

$$C_{Prot} = \frac{A_{280\ Prot}}{\epsilon_{prot} l} = \frac{A_{280\ complexe} - A_{L4-Tb\ 280}}{\epsilon_{prot} l} = \frac{A_{280\ complexe} - (A_{L4-Tb\ 349} * FC_{280})}{\epsilon_{prot} l}$$

Équation 6 : Concentration en protéine avec correction d'absorbance. $A_{280\ complexe}$ est l'absorbance du complexe à 280nm. ϵ_{prot} est le coefficient d'absorption molaire des bidomaines. l est le chemin optique (1 cm).

Connaissant les coefficients d'absorption molaire à 280 nm de **A3_A3** et **A3_bGFPD** et celui à 349 nm du L4-Tb, respectivement 60390, 49390 et 26000 $M^{-1}.cm^{-1}$, il est possible de déterminer chacune des concentrations et d'en tirer le taux de couplage qui est égal au ratio de la concentration en L4-Tb sur celle en protéine. Les taux de couplages obtenus pour les deux bidomaines sont d'environ 85% et ont permis de poursuivre les expériences.

2) Couplage des QD625s

Les QD625 utilisés ici sont constitués d'un cœur de Cadmium et Sélénium, leur surface elle, est faite de Zinc et de Souffre. Les histidines possédant naturellement une affinité pour le Zinc, l'étiquette histidine des deux bidomains devra se fixer sur les QD625 avec un K_D d'environ 1 nM. (Sapsford, K.E., *et al.*, 2007) (Figure 23)

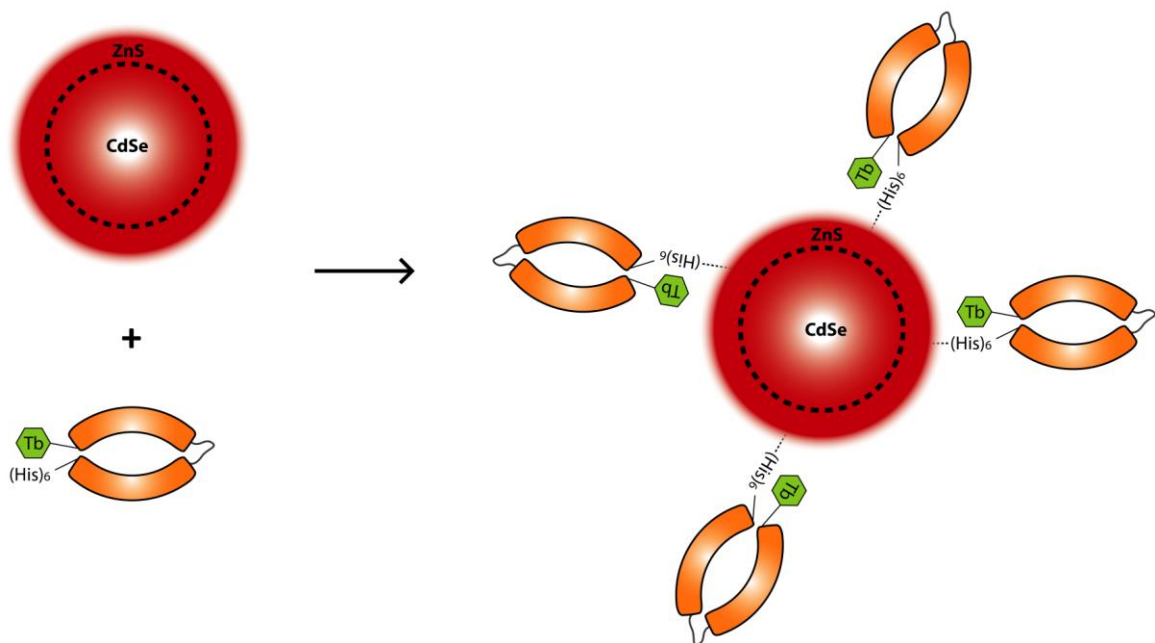


Figure 23 : Couplage des chimères au QD via l'utilisation de l'affinité naturelle de l'étiquette histidine pour le Zinc présent en surface des QD625.

Les QDs utilisés mesurant une dizaine de nanomètres de diamètre, plusieurs bidomains peuvent se fixer dessus. Afin de déterminer la concentration à partir de laquelle tous les bidomains (avec et sans ligand) sont fixés sur les QDs, j'ai effectué une expérience de saturation des QDs. La concentration en bidomaine a été fixée à 4 nM, celles en ligands à 1 μ M et celle en QD625 a varié de 0 à 1.33 nM. Les intensités à 494 et à 640 nm ont été mesurées de 0 à 1 ms sur un appareil Krypton (*Thermo Fisher*) et le ratio de FRET a été calculé. Ce ratio est défini comme la somme des intensités mesurées à 640 nm sur celles mesurées à 494 nm. Pour les raisons citées précédemment, ce ratio a été calculé entre 50 et 900 μ s (« *Time-gated* FRET »). Cet intervalle de temps permet de supprimer toute fluorescence non due au FRET comme expliqué précédemment. L'augmentation de ce ratio traduit une diminution des intensités à 494 nm (L4-Tb) et/ou une augmentation des intensités à 640 nm (QD625). Une augmentation de ce ratio signifie donc une augmentation du signal de FRET. J'ai effectué cette expérience pour les deux bidomains (A3_A3 et A3_bGFPD) avec et sans leur ligand. Dans le cas de l'étude de la fixation des bidomains avec leur ligand, une pré-incubation de 15 min est réalisée pour permettre à l'équilibre d'interaction « bidomaine/ligand » d'être atteint.

Cette expérience, en plus de permettre la détermination de la concentration minimale en QDs nécessaire pour fixer toutes les protéines, permet également d'obtenir le nombre maximal de protéines par QDs. Pour cela, des régressions linéaires ont été effectuées sur les quatre premières et dernières concentrations en QDs pour chaque protéine avec ou sans ligand. Le croisement de ces deux régressions permet d'obtenir le nombre maximal de protéines par QDs. (Figure 24)

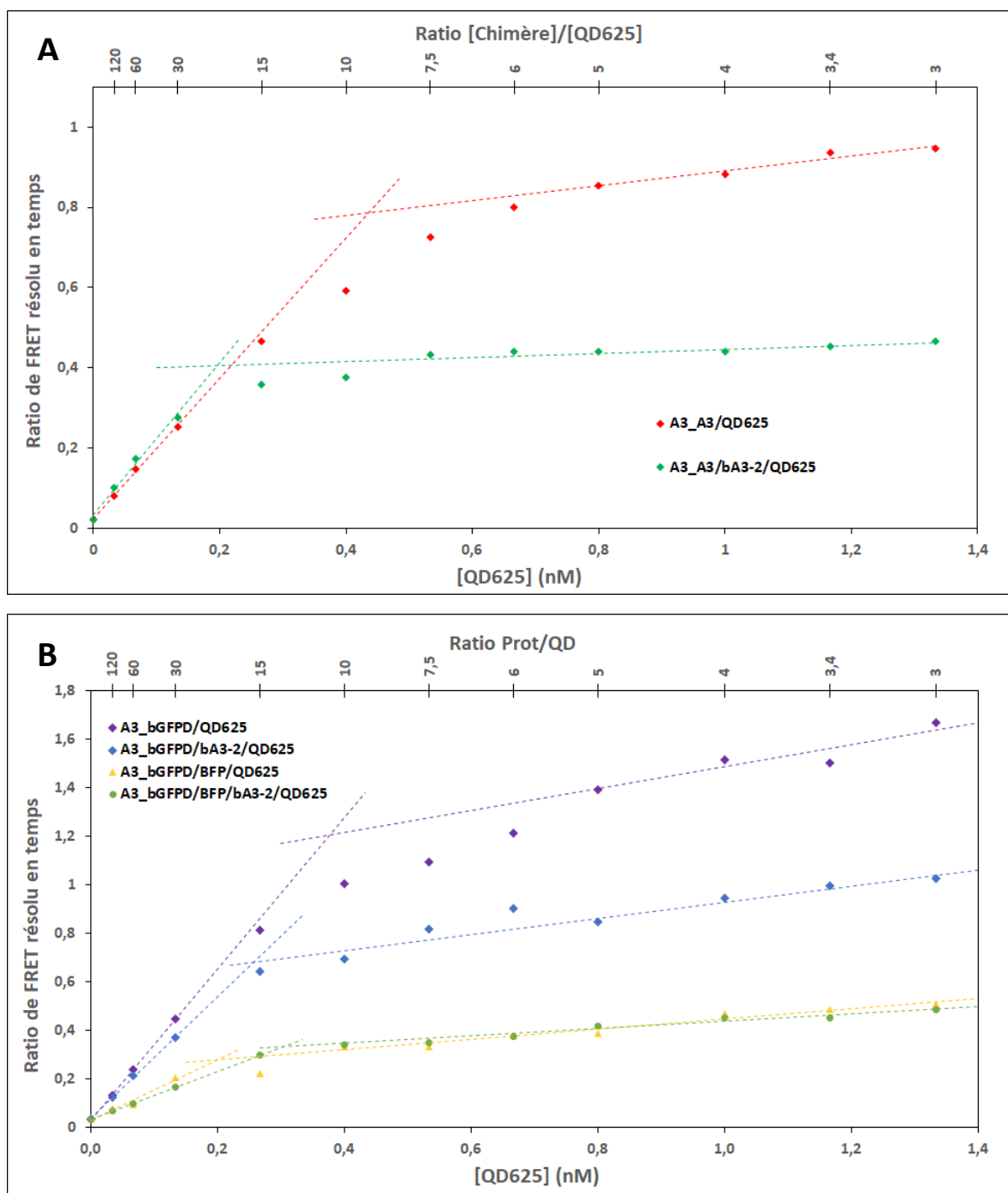


Figure 24 : Courbes de saturation du signal de FRET. Ratio de FRET résolu en temps en fonction de la concentration en QD625.

A) A3_A3 avec et sans bA3-2. B) A3_bGFPD avec et sans bA3-2 et/ou BFP.

Les courbes montrent une saturation du Ratio de FRET aux alentours des 1 nM pour les chimères seules. Cela indique que toutes les chimères sont fixées. En présence de ligand, le plateau est atteint plus tôt entre 0,4 et 0,6 nM.

D'après les résultats obtenus, l'intégralité des bidomains **A3_A3** et **A3_bGFPD** sont fixés aux alentours d'une concentration de 1 nM de QD. En effet, à partir de ces concentrations, le ratio de FRET n'augmente plus, indiquant donc une fixation totale des protéines. L'ajout de QDs ne permet alors plus de fixer des bidomains libres et le ratio de FRET n'augmente plus. Pour les bidomains en présence de ligand, les concentrations nécessaires sont plus faibles entre 0.4 et 0.6 nM. Pour ces raisons, les expériences de détection des ligands ont été effectuées avec des concentrations en bidomaine de 4 nM et des concentrations en QDs de 1.33 nM. Cette concentration en QD garantit la fixation de tous les bidomains et donc un signal de FRET maximal.

De plus, on peut voir que le nombre maximal de protéines par QDs augmente quand les bidomains sont pré-incubés avec leurs ligands. Ce nombre passe de 10 pour les bidomains seuls à 15-20 lorsqu'ils sont incubés avec leurs ligands respectifs. Cela peut s'expliquer par le fait que l'ouverture des bidomains par les ligands réduit la surface occupée sur les QDs permettant alors la liaison de plus de bidomains.

Enfin, en ce qui concerne les amplitudes du ratio de FRET, il est intéressant de voir qu'il n'y a quasiment aucune différence entre **A3_bGFPD/BFP** et **A3_bGFPD/BFP/bA3-2**. Cela laisse supposer que **bA3-2** n'a pas une taille suffisante pour changer significativement le signal de FRET par rapport à la **BFP** seule.

D. Détection des ligands

1) Détection en tampon

Afin d'étudier les capacités de détection des bidomains couplés aux L4-Tb et aux QD625, j'ai mesuré les intensités respectives du L4-Tb et des QDs sur le Kryptor puis calculé le ratio de FRET avec des concentrations croissantes de ligand. Chaque point a été réalisé en triplicata dans ces expériences, le blanc (concentration en ligand nul) a été réalisé dix fois. Une régression linéaire est effectuée sur les douze premières concentrations (0 - 2.33 nM) permettant d'obtenir la pente « a » sur le début de la détection. La limite de détection (LOD) a alors été calculée en utilisant la formule suivante :

$$LOD = \left| \frac{3 * \sigma_{10 \text{ blanc}}}{a} \right|$$

Équation 7 : Limite de détection (LOD) des ligands. $\sigma_{10 \text{ blanc}}$ est l'écart type sur la mesure de dix blancs successifs. a est la pente de la régression linéaire effectuée sur les douze premières concentrations.

Il fut ainsi possible de déterminer les limites de détections pour **bA3-2** et la **BFP** dans un volume de 150 μ L en tampon phosphate. Celles-ci sont respectivement de 170 pM et 440 pM pour **bA3-2** contre **A3_A3** et **A3_bGFPD**. (Figure 25) Il est intéressant d'observer que le signal pour le bidomaine **A3_A3** diminue plus que le signal du bidomaine **A3_bGFPD**. Cela peut être lié au fait que deux **bA3-2** se lient en même temps sur **A3_A3**, au lieu d'un seul sur **A3_bGFPD**. Cependant, cela peut aussi tout simplement être dû aux structures différentes des bidomains qui ne répondraient pas de la même manière à l'ouverture par les ligands.

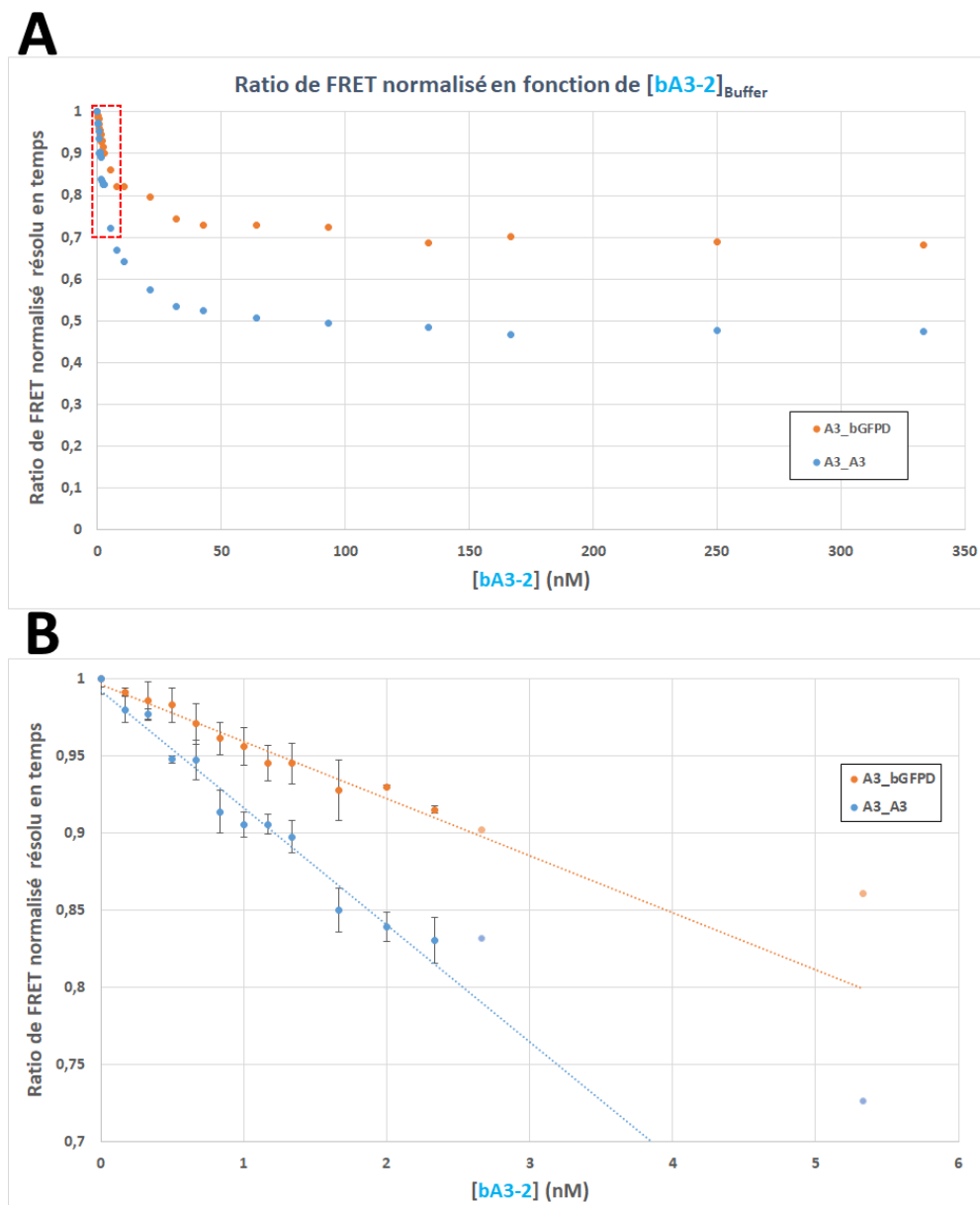


Figure 25 : Détection de $bA3-2$ par les chimères $A3_A3$ et $A3_bGFPD$. Les chimères sont à une concentration constante de 4 nM, les QDs sont à 1.33 nM. Les concentrations en $bA3-2$ vont de 0 à 333 nM. B) Agrandissement des courbes en A (points encadrés par le rectangle en pointillés rouge) afin de présenter les régressions linéaires et les écart-type de chaque mesure pour les deux chimères sur les onze premières concentrations de $bA3-2$ (0 - 2.33 nM)

La LOD de la BFP par le bidomaine A3_bGFPD est de 650 pM (Figure 26). Ces LODs sont du même ordre de grandeur que ce qui existe actuellement avec des systèmes basés sur l'utilisation du FRET. En effet, la majorité des études font part de limites de détection allant de la dizaine de picomolaire à la centaine de nanomolaire (Fu, X., *et al.*, 2017; Jumeaux, C., *et al.*, 2018; Zhang, H., *et al.*, 2017)

Il est intéressant de voir que la BFP entraîne une diminution du ratio de FRET plus importante que bA3-2 sur le bidomaine A3_bGFPD. Cela semble indiquer que les bidomaines n'adoptent pas une conformation totalement ouverte en présence de bA3-2 et que leur ouverture dépend du ligand se fixant. Ce phénomène n'était pas observé avec les atto-dyes où le signal de FRET atteignait un même niveau que cela soit avec bA3-2 ou la BFP. Cela peut s'expliquer par le rayon de Förster plus grand pour le couple L4-Tb / QD qui permet donc de distinguer les deux ouvertures.

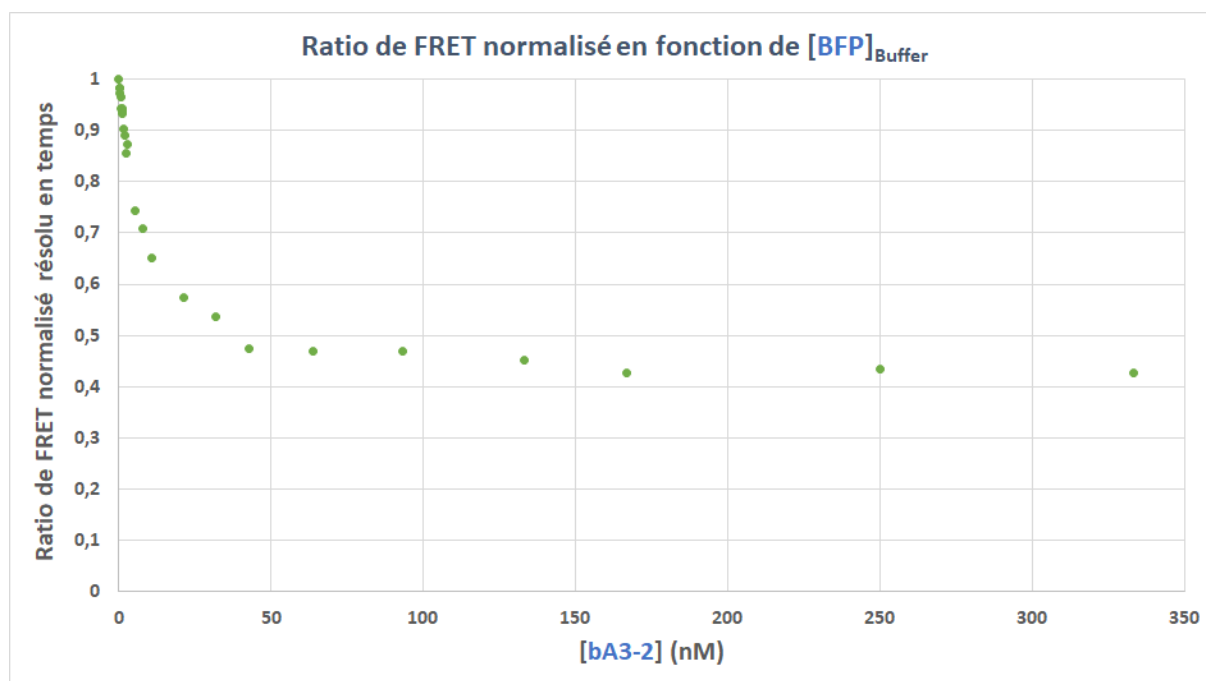


Figure 26 : Détection de la BFP par la chimère A3_bGFPD. A3_bGFPD est à une concentration constante de 4 nM, les QDs sont à 1.33 nM. Les concentrations en BFP vont de 0 à 333 nM.

2) Détection en milieu complexe

La détection de biomarqueur s'effectue généralement en milieu complexe tel que du sérum, du sang, de l'urine, etc. Pour ces raisons, j'ai voulu détecter **bA3-2** dans des milieux plus complexes qu'un simple tampon phosphate, à savoir, du lysat bactérien et du sérum de veaux foetal. (Figure 27)

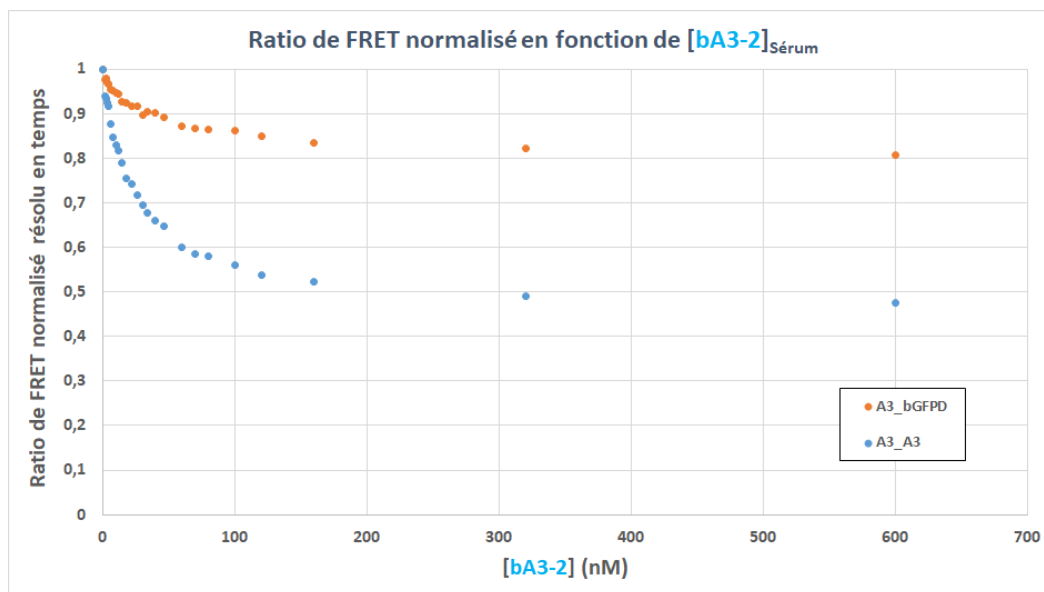


Figure 27 : Détection de bA3-2 en sérum de veaux pour les bidomaines A3_A3 et A3_bGFPD dans du sérum. Les bidomaines sont à une concentration constante de 4 nM, les QDs sont à 1.33 nM. Les concentrations en bA3-2 vont de 0 à 600 nM.

Les limites de détection de **bA3-2** dans le lysat et dans le sérum sont respectivement de 1.2 et 2.2 nM pour **A3_A3** et 3.6 et 7.2 nM pour **A3_bGFPD**. Bien qu'augmentant par rapport aux détections en milieu « simple », ces LODs restent dans les limites typiques observées pour des systèmes basés sur le FRET comme expliqué précédemment.

3) Détections successives des ligands

Afin d'étudier plus en détails le processus d'ouverture des bidomains (et par curiosité), j'ai essayé de détecter successivement les deux ligands du bidomaine **A3_bGFPD**. Dans un premier temps, après un ajout de concentration croissante de **BFP**, des concentrations croissantes de **bA3-2** ont été ajoutées (Figure 28-A). Dans un second temps, l'expérience inverse a été réalisée (Figure 28-B).

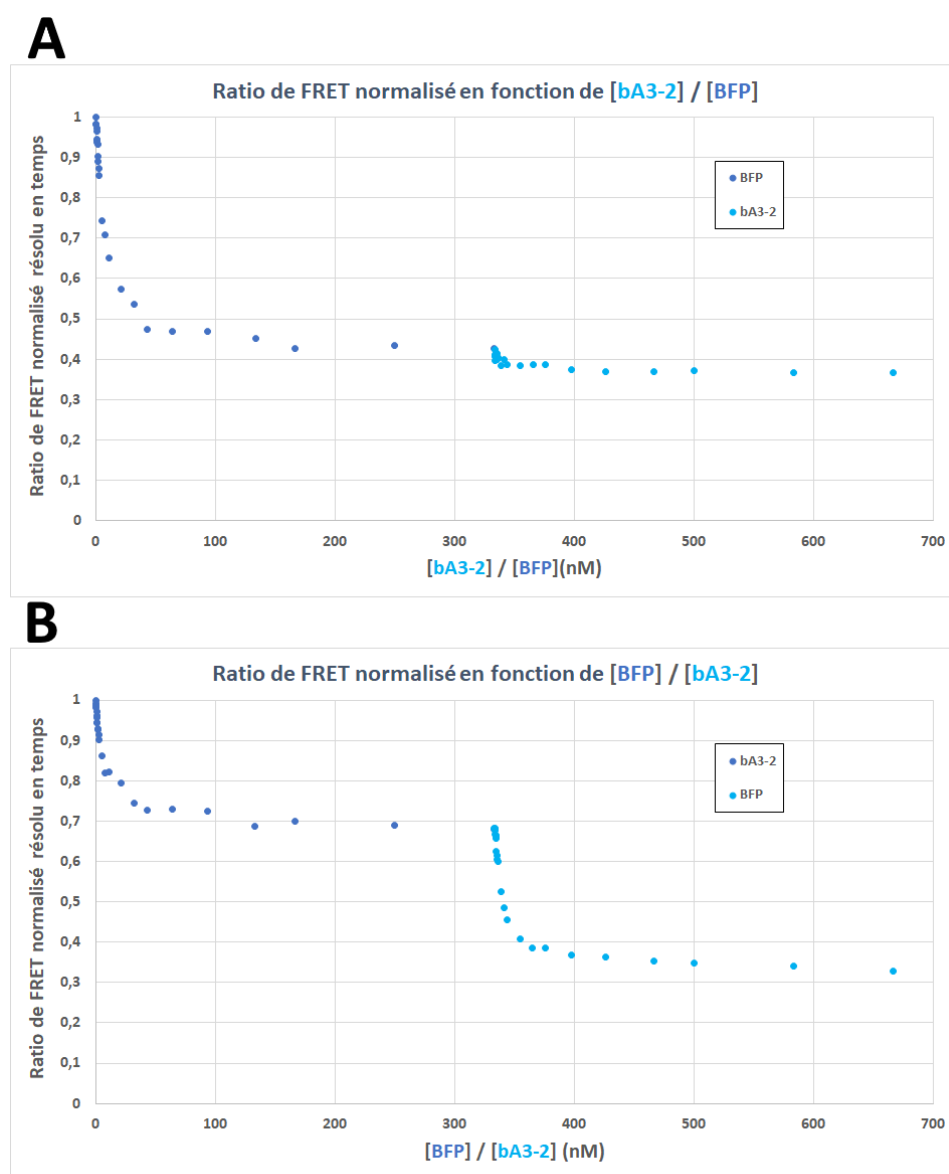


Figure 28 : Détection successive des ligands. A3_bGFPD est à une concentration de 4 nM, les QDs sont à 1.33 nM. Les concentrations en BFP et en bA3-2 vont de 0 à 333 nM pour chaque protéine. L'axe des abscisses représente la somme des concentrations en bA3-2 et en BFP.

Les résultats montrent que la détection successive de la **BFP** puis de **bA3-2** n'est pas possible. En effet, il y a certes une diminution du signal, mais trop faible pour permettre de déterminer une LOD. Cependant la détection en sens inverse (**bA3-2** puis **BFP**) est, elle, possible et donne une LOD pour la **BFP** de 700 pM. Cette LOD est du même ordre de grandeur que celle obtenue lorsque la **BFP** est la première à se fixer (650 pM). L'impossibilité de détecter **bA3-2** en second semble indiquer que lorsque la **BFP** se fixe, la fixation de **bA3-2** n'entraîne pas de changement significatif de conformation et donc ne change pas le signal de FRET. À l'inverse, lorsque **bA3-2** se fixe en premier, la protéine semble pouvoir adopter des conformations plus nombreuses qui sont par la suite inaccessibles lorsque la **BFP** se fixe.

Afin d'étudier le phénomène d'ouverture plus en détail et pour mieux comprendre les implications sur les deux constructions, il est nécessaire de déterminer les distances d'ouvertures des bidomaines.

E. Mesure des distances grâce au FRET

1) La théorie

Afin de mesurer les distances séparant les fluorophores et donc l'ouverture des bidomaines, il est nécessaire d'introduire quelques notions concernant le FRET et les fluorophores utilisés. Tout d'abord, plusieurs moyens sont possibles pour obtenir la distance r séparant les deux fluorophores (Medintz, I. and Hildebrandt, N., 2014). À partir des équations suivantes :

$$\eta_{FRET} = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6} = 1 - \frac{\tau_{DA}}{\tau_D}$$

Équation 8 : Efficacité du FRET en fonction de r et R_0 , τ_D et τ_{DA} . η_{FRET} est l'efficacité de FRET, r est la distance en nm entre le donneur et l'accepteur, R_0 le rayon de Förster en nm. τ_D et τ_{DA} les temps de décroissance de la fluorescence du donneur seul et du donneur en présence d'accepteur respectivement.

On peut déduire que :

$$r = R_0 * \sqrt[6]{\frac{1 - \eta_{FRET}}{\eta_{FRET}}} = R_0 * \sqrt[6]{\frac{\tau_{DA}}{\tau_D - \tau_{DA}}}$$

Équation 9 : Distance r séparant les fluorophores en fonction du rayon de Förster R_0 et de l'efficacité de FRET η_{FRET} et de τ_D et τ_{DA} les temps de décroissance de la fluorescence du donneur seul et du donneur en présence d'accepteur respectivement.

La distance peut donc être calculée à partir des temps de décroissance de la fluorescence.

2) Résultats

Ainsi, afin de déterminer les distances, les bidomaines pré-incubés ou non avec leur ligand ont été ajoutés à des concentrations croissantes de QDs (0 - 1.33 nM). En suivant la fluorescence aux longueurs d'onde d'émission du L4-Tb et des QDs, il est alors possible d'obtenir des courbes de décroissance de fluorescence à l'aide d'un appareil de type DreamReader (*Edinburgh Instrument*). Le suivi de la fluorescence du L4-Tb montre un *quenching* avec l'augmentation de la concentration en QDs. À l'inverse, le suivi du signal des QDs montre une « sensibilisation » progressive par le transfert d'énergie des L4-Tb. Les courbes ainsi obtenues peuvent être *fittées* pour en extraire les τ et les amplitudes de fluorescence du donneur et du donneur en présence des différentes concentrations d'accepteur (Figure 29).

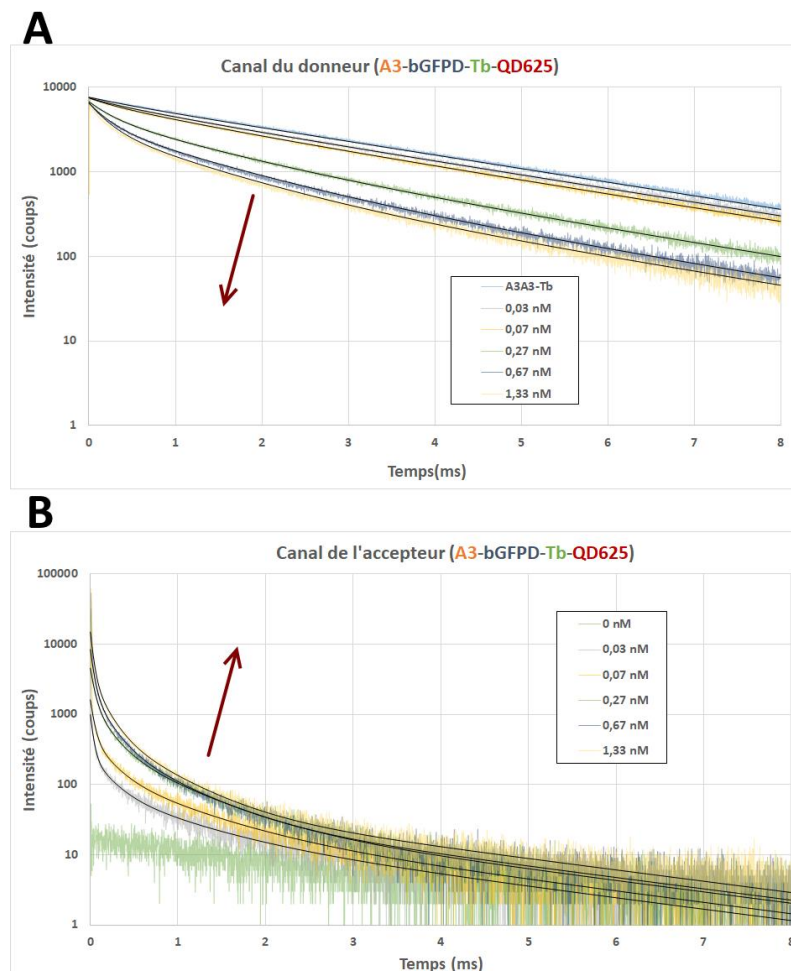


Figure 29 : Courbes de décroissance de la luminescence pour le canal du donneur (L4-Tb) et pour le canal de l'accepteur (QD625) pour le bidomaine A3_bGFPD. La concentration à 0 nM de QD correspond au bidomaine marqué au L4-Tb à 4 nM. Le quenching du L4-Tb est visible en suivant la diminution de l'intensité avec l'augmentation de la concentration en QD. La sensibilisation des QDs est observable par l'augmentation de l'intensité avec l'augmentation de la concentration en QDs ; sans la présence de bidomaine, les QDs ne présentent pas cette augmentation.

3) Le *fitting*

Les *fits* ont été effectués en prenant en compte le fait que les QDs ne sont pas totalement sphériques et tendent plutôt vers une forme ellipsoïde. Par conséquent, en fonction de l'endroit où sera liée la protéine sur le QDs, les distances ne seront pas exactement les mêmes et les temps de décroissances de la fluorescence τ varieront aussi. Il y a donc potentiellement une infinité de valeurs de τ possibles. Afin d'effectuer une approximation correcte pour les *fits*, nous faisons l'hypothèse que seulement deux valeurs de τ coexistent. Le L4-Tb possède quant à lui une décroissance de luminescence qui est quasi monoexponentielle avec une durée de vie d'environ 2,7 ms. En réalité, il existe une très faible contribution d'une durée de vie courte négligeable et une très forte contribution d'une durée de vie longue. Le *quenching* des L4-Tb par les QDs induit une réduction de la durée de vie du L4-Tb. Comme il reste toujours des L4-Tb non-désactivé par le FRET, les courbes monoexponentielles de Tb deviennent multiexponentielles (contribution de L4-Tb et contribution du FRET L4-Tb-QD). Ainsi, de par la forme ellipsoïde des QDs et l'utilisation de deux τ , les *fits* des courbes de décroissance des L4-Tb, seront effectués avec une courbe ajustée sur une équation triple exponentielle (deux τ pour les durées de vie courte FRET L4-Tb-QD et un pour la durée de vie longue L4-Tb).

À cause de l'excitation directe des QDs par le laser à 337 nm, ces derniers présentent une fluorescence très importante sur les premières 10-50 μ s, il est donc nécessaire de prendre en compte ce temps de décroissance en plus des deux dues au FRET *via* les L4-Tb. Enfin, il est également nécessaire de prendre en compte la durée de vie longue des L4-Tb non affectée par le FRET mais visible à la longueur d'onde choisie. Les courbes de décroissance des QDs ont donc été *fittées* à l'aide d'une courbe quadruple exponentielle (un τ très court, deux τ court dus aux FRET, un τ long dû au L4-Tb).

Une amplitude A est couplée à chaque temps de décroissance τ , et correspond à l'intensité de fluorescence due à chacun d'eux. Dans tous les cas, les *fits* ont été ajustés selon l'équation suivante en essayant de minimiser les différences courbe expérimental / fit :

$$I = I_0 + \sum_i A_i * \exp\left(-\frac{t}{\tau_i}\right)$$

Équation 10 : Intensité de luminescence I en fonction de la somme de l'amplitude A_i associée au temps de décroissance τ_i et au temps après excitation t . I_0 représente une intensité constante du bruit de fond.

4) Quelles distances ?

À partir des *fits*, il est possible d'obtenir τ_D qui est pour rappel le temps de décroissance sans l'accepteur (i.e. à concentration de QDs nulle) ainsi que les temps de décroissance du donneur pour chaque concentration en QDs, il est alors possible de faire la moyenne de ces temps pour obtenir une durée de vie FRET moyenne τ_{DA} . Dans notre cas, le *quenching* du donneur était très efficace et les *fits* sur les courbes de ce dernier sont bien meilleurs que ceux sur la décroissance des QDs. Par conséquent, les distances moyennes ont été déterminées avec le canal du donneur uniquement. Elles sont répertoriées dans le tableau ci-dessous.

Échantillon	Distance (nm)	Distance - R_{QD625}
A3_A3	8.1	3
A3_A3/bA3-2	10	4.9
A3_bGFPD	8.7	3.6
A3_bGFPD/ bA3-2	9.4	4.3
A3_bGFPD/BFP	10.1	5
A3_bGFPD/BFP/ bA3-2	10.3	5.2

Tableau 1 : Distances déterminés par FRET pour les bidomaines avec et sans leurs ligands. R_{QD625} correspond au rayon des QDs.

Ces distances nous permettent en premier lieu de mieux comprendre comment **A3** et **bGFPD** interagissent ensemble. En effet, en utilisant le logiciel pymol et en modélisant les QD625 par une sphère de 5.1 nm de rayon, il est possible de vérifier que les distances sont en accord avec le modèle obtenu en SAXS. Pour ce faire, j'ai positionné la structure cristallographique d'**A3_A3** en faisant d'une part en sorte d'avoir une interaction possible entre l'étiquette histidine et la surface du QDs et d'autre part en essayant d'obtenir la distance mesurée par le FRET. J'ai par la suite positionné le modèle d'**A3_bGFPD** obtenu en SAXS pour la forme fermée en superposant le domaine **A3** du bidomaine **A3_bGFPD** au domaine **A3** N-terminal du bidomaine **A3_A3**. La distance ainsi obtenue pour **A3_bGFPD** est la même que celle obtenue en FRET. Cela permet d'ajouter une certaine valeur au modèle de SAXS et montre qu'il est vraisemblable. (Figure 30)

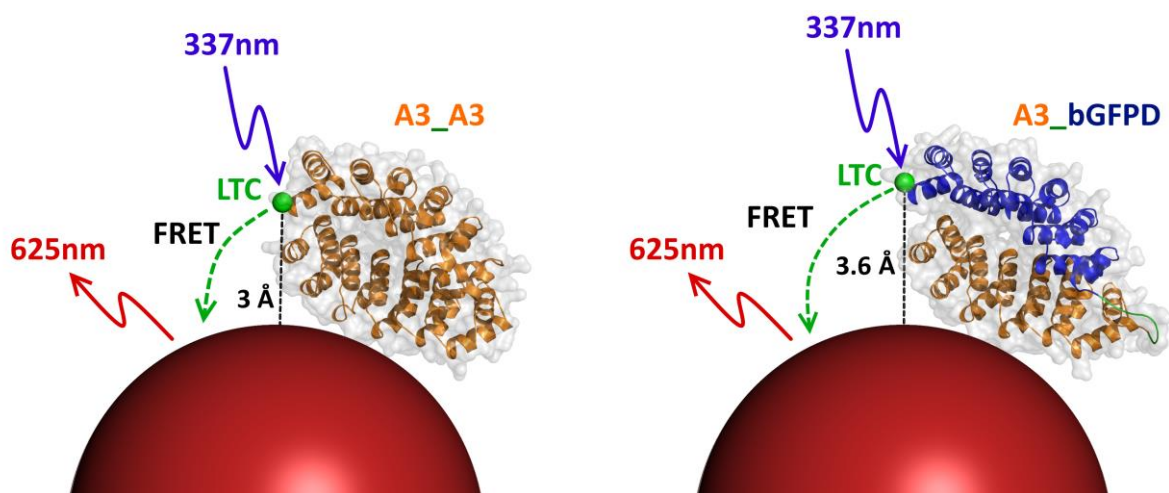


Figure 30 : Modèles représentant l'interaction des bidomaines avec les QDs. Le LTC (Luminescent Terbium Complex) correspond au L4-Tb. Les domaines A3 interagissant avec les QDs sont fixés de la même manière pour les deux bidomaines. Les distances ont été mesurées avec le logiciel pymol et sont en accord avec les distances déterminées par FRET.

D'autre part, ces distances permettent d'obtenir une meilleure compréhension du processus d'ouverture des bidomaines. Tout d'abord, il existe une différence entre les deux protéines pour la fixation de **bA3-2**, en effet, l'ouverture du bidomaine **A3_A3** est de 1.9 nm contre 0.7 nm pour **A3_bGFPD**. L'effet d'ouverture ne semble donc pas additif puisqu'un **bA3-2** ouvre de 0.7 nm pour **A3_bGFPD** alors que dans le cas d'**A3_A3** cette distance serait de 0.95nm (1.9/2). Cela laisse supposer que les mécanismes diffèrent probablement entre les bidomaines. Une des hypothèses envisageables est que la fixation d'un **bA3-2** sur le domaine **A3** C-terminal (sur lequel est fixé le Lumi4-Tb) a plus d'effet sur les conformations adoptables que la fixation d'un **bA3-2** sur le domaine **A3** N-terminal (fixé au QD). (Figure 31)



Figure 31 : Modèle représentant l'ouverture des bidomaines par **bA3-2**. Les distances ont été mesurées avec le logiciel pymol. Les modèles présentés ici ont été construits pour être compatibles avec les distances de FRET. Il faut garder à l'esprit que les distances de FRET représentent une moyenne des distances et donc des conformations que peuvent adopter les systèmes chimères/ligand/QD.

En ce qui concerne la **BFP**, lorsque celle-ci est fixée, l'ajout de **bA3-2** entraîne seulement une ouverture de 0.2 nm. Sachant que le domaine **bGFPD** est en C-terminal et la **BFP** étant plus grosse que **bA3-2**, les conformations possibles en présence de la **BFP** peuvent être plus limitées et tendre vers des formes déjà très ouvertes, réduisant alors l'effet de **bA3-2**. N'ayant pas la structure du complexe **BFP/bGFPD**, il n'est pas possible de mettre en relation les distances obtenues par le FRET avec des modèles structuraux.

F. Conclusion et discussion

Tout d'abord, l'utilisation du couple L4-Tb/QD a permis d'optimiser la détection des ligands par rapport au couple atto-dye. Il a également été possible de déterminer des limites de détection pour chacun des ligands. Ces limites sont par ailleurs dans la gamme de détection des tests actuellement utilisés et permettant la détection de divers biomarqueurs (<https://www.cisbio.com/diagnostics/produits>). Cela permet de confirmer une possible adaptation de ces systèmes pour effectuer de la détection de cibles présentant un intérêt thérapeutique. En effet, il est important de rappeler que cette approche peut être généralisable à toutes les cibles contre lesquelles il est possible de sélectionner des α Reps. Nous avons également vu qu'il était possible de fixer deux ligands simultanément sur un même bidomaine. Cela semble être un moyen d'obtenir une meilleure limite de détection comme le montrent les LODs des bidomaines **A3_A3** et **A3_bGFPD** pour **bA3-2** qui sont 170 et 440 pM respectivement.

Cela nous a également permis d'approfondir la compréhension des systèmes bidomaines/ligands, notamment sur les mécanismes impliqués dans l'ouverture des bidomaines en présence d'un ou plusieurs ligands grâce à la détermination des distances par l'utilisation approfondie des principes du FRET.

Ces biosenseurs génériques pourraient se révéler être une alternative aux tests existant actuellement qui se font très souvent en plaque ELISA ou par *RadiolImmuno Assay*. En effet, ce genre de test se fait généralement en « sandwich », cela signifie que la cible est reconnue à deux endroits différents ce qui permet sa détection. Ces nouveaux biosenseurs présenteraient l'avantage de ne nécessiter une reconnaissance que par un seul épitope.

Bien qu'étant très efficace, ce système présente cependant l'inconvénient de devoir passer par un couplage chimique pour coupler le Lumi4. On peut donc naturellement se demander s'il est possible de s'affranchir de cette étape de couplage pour la conception de biosenseurs.

III. Vers des biosenseurs génétiquement codables

A. Introduction

Les biosenseurs génétiquement codables sont des biosenseurs étant fonctionnels après l'étape de traduction. On peut par exemple citer les *Cameleons* ou les *periCams* présentés dans l'introduction (Partie A - IV - B - 1 et Partie A - IV - B - 4). L'obtention de tels biosenseurs présente plusieurs intérêts. Tout d'abord, d'un point de vue de l'utilisation, en effet, il serait possible avec de tels biosenseurs de s'affranchir de l'étape de couplage ce qui permettrait une utilisation directe après production et purification des protéines. Cela permettrait donc de diminuer nettement les coûts de production des biosenseurs. D'autre part, il n'y aurait plus d'utilisation de la chimie pour les fluorophores et donc moins de polluants associés à leur production. De tel biosenseurs pourraient également être utilisés *in cellulo* sans avoir recours à des injections pour franchir la barrière membranaire des cellules. Il serait même possible de les adresser dans les différents compartiments cellulaires *via* des séquences d'adressage.

Dans un premier temps, j'ai essayé de réaliser un biosenseur en me basant sur l'utilisation d'une séquence peptidique appelée *Lanthanide Binding Tag* (LBT). (Franz, K.J., et al., 2003) Bien qu'utilisant aussi le Terbium, celui-ci serait chélaté non plus par une molécule chimique mais par le LBT pouvant donc être ajouté sur les protéines sans passer par des étapes de couplage chimique. De plus, il a été montré que le Terbium pouvait être absorbé par les bactéries *E. Coli* dans le cytoplasme et le périplasme. (Achyuthan, K.E., et al., 2009) Le LBT a par ailleurs été testé *in vivo* chez *E. Coli*. Les résultats montrent qu'à 50 μM de Terbium dans le milieu de culture, une fluorescence spécifique du Terbium est observée chez les bactéries possédant la protéine couplée au LBT. (Gonzalez, D., et al., 2011)

Dans un second temps, j'ai essayé d'utiliser la *nanoKaz*, une enzyme catalysant une réaction bioluminescente. L'idée étant de réaliser une construction à plusieurs domaines dont un capable d'inhiber la *nanoKaz*, l'ajout d'un ligand entraînerait une levée de l'inhibition et l'apparition d'un signal de bioluminescence.

B. Lanthanide Binding Tag et dérivés

1) Origine du LBT

Le LBT est une séquence peptidique capable de lier le terbium et développée par l'équipe de B. Imperiali en 2003. Sa conception s'est basée sur l'utilisation et l'optimisation d'un motif de type *EF-Hand* liant le Ca^{2+} (présent dans la calmoduline par exemple). Le LBT ainsi obtenu n'est stable que lorsqu'il est relié à une protéine ou quand il est flanqué par des cystéines qui viennent établir un pont disulfure stabilisant la structure. La séquence originale retenue est constituée de douze résidus dont un Tryptophane (Trp) qui fait office d'antenne pour l'ion Tb^{3+} (DYNKDGWYEELE). Il est ainsi possible d'exciter les Trp et par transfert d'énergie une fluorescence est visible avec les quatre pics typiques du Terbium. (Franz, K.J., *et al.*, 2003) La séquence du LBT a par la suite pu varier en fonction des protéines dans lesquelles elle est utilisée sans jamais changer la place du Trp et des résidus permettant de coordonner le Tb^{3+} . (Barthelmes, K., *et al.*, 2011) (Figure 32)

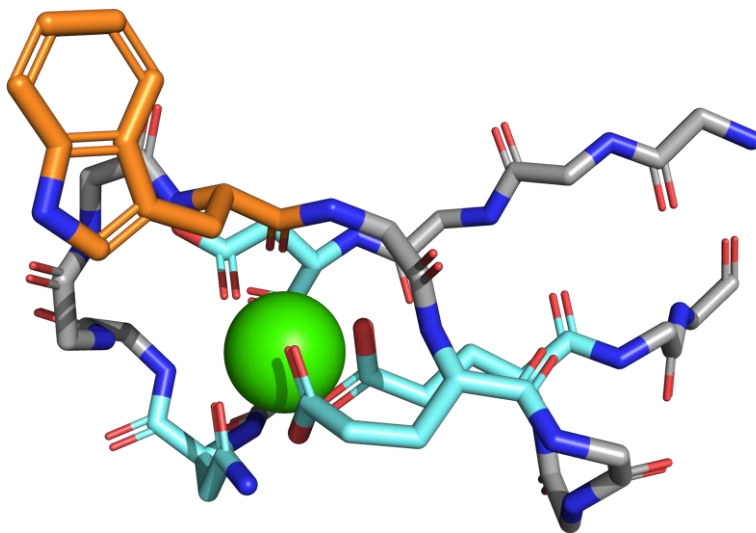


Figure 32 : Structure du Lanthanide Binding Tag (LBT). Le terbium est représenté par la sphère verte. Le tryptophane servant d'antenne est visible en orange. Les résidus coordonnant le Terbium sont visibles en cyan.

2) Modification du LBT

Nous avons décidé d'exploiter ce LBT en l'intégrant au sein des bidomaines. Cependant le LBT tel qu'il est conçu ne permet pas de détecter un changement de conformation des bidomaines en étant seul. Disposant déjà d'une construction LBT_A3_A3, je suis parti de cette protéine pour la transformer en biosenseur. Ainsi, le tryptophane servant d'antenne dans le LBT a été enlevé, remplacé par une sérine et une antenne constituée de deux Trp a été rajoutée à l'extrémité opposée du bidomaine. L'idée étant que lorsque la protéine est fermée, l'antenne Trp capte les photons et par un transfert d'énergie permet une émission de fluorescence du Tb^{3+} . À l'inverse, lorsque le bidomaine s'ouvre en présence des ligands, l'antenne Trp s'éloigne du Tb^{3+} et la fluorescence disparaît. Le bidomaine A3_A3 a ainsi été modifié par un *Gibson Assembly* pour donner LBTs_A3_A3_WW. La production a été effectuée dans des bactéries *E. Coli BL21 Gold* à 37°C dans du milieu 2YT en présence d'ampicilline. La purification a été effectuée en deux étapes, colonne d'affinité His-Trap 5 mL suivie par une étape de filtration sur gel S75-16/600 en tampon HEPES 50 mM, NaCl 150 mM, pH 7.4. Le suivi de la purification a été effectué par SDS-PAGE, la masse moléculaire attendue est de 46.6 kDa (Figure 33).

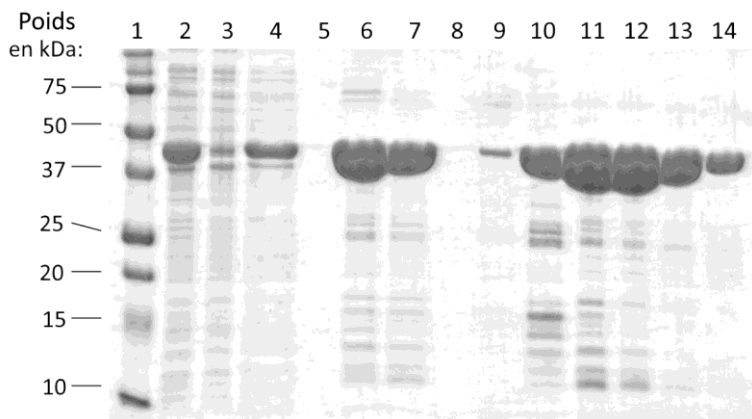


Figure 33 : Gel SDS-PAGE 15% de suivi de purification de la protéine LBTs_A3_A3_WW. 1 : Marqueur de poids moléculaire Protein Dual Color. 2 : Fraction soluble. 3 : Fraction non retenue sur His-Trap. 4 : Lavage avec du tampon 20mM Imidazole. 6 et 7 : Éluat Ni-NTA. 9 à 14 : Éluat de la Gel Filtration.

Le gel SDS-PAGE montre que la protéine est bien produite à la taille attendue, cependant, même après la purification, des contaminants restent présents. Ceux-ci ne devraient cependant pas interférer avec les signaux de fluorescence puisque, ne possédant pas de LBT, les protéines naturelles ne devraient pas entraîner une fluorescence parasite.

3) Comparaison avec le LBT

Afin de mesurer l'efficacité du nouveau LBT couplé à son antenne Trp, j'ai dans un premier temps effectué une comparaison entre le LBT classique et le LBT modifié. Pour cela, j'ai utilisé une protéine déjà réalisée et étudiée au sein de l'équipe, LBT_A3_A3, qui est constituée d'un LBT en position N-Terminale suivi de la protéine A3_A3. L'étude de cette chimère avait révélé qu'à une concentration de 1 μM de chimère et 10 μM de Tb^{3+} (Chlorure de Terbium), un signal typique de la fluorescence du Tb^{3+} était observé après une excitation à 280 nm. Comme pour les expériences impliquant le L4-Tb, la mesure a été résolue en temps. Le signal a ainsi été mesuré après un délai de 50 μs et intégré sur une période de 5 ms. J'ai donc mesuré ce spectre et l'ai comparé à celui de la protéine LBTs_A3_A3_WW (Figure 34).

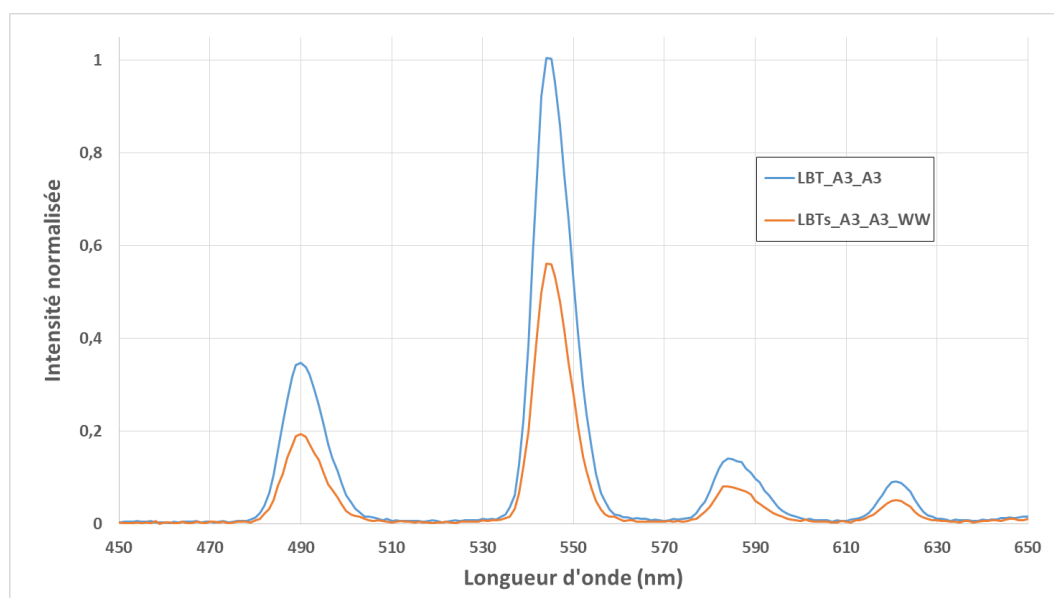


Figure 34 : Spectres normalisés sur la valeur la plus haute. Les protéines LBT_A3_A3 et LBTs_A3_A3_WW à 1 μM en présence de 10 μM de Tb^{3+} dans du tampon HEPES. La mesure a été effectuée 50 μs après excitation avec un temps d'intégration de 5 ms

Ces résultats montrent qu'un transfert d'énergie entre les Trp et le Tb^{3+} semble toujours avoir lieu, même si celui-ci est moins efficace que pour le LBT non modifié. La diminution d'intensité pour le LBT modifié peut être due à une mauvaise orientation des Trp par rapport au Tb^{3+} ou tout simplement à un éloignement plus important de ces derniers, réduisant l'efficacité du transfert d'énergie.

4) Détection de bA3-2

Sachant que le LBT modifié permettait d'obtenir une fluorescence, j'ai par la suite essayé de détecter **bA3-2**. Pour cela j'ai tout d'abord fait varier la concentration de **bA3-2** et mesuré les spectres après un temps d'incubation de deux minutes, en utilisant la bArpin (binder de l'Arpin) comme contrôle négatif. Cette protéine est une α Rep possédant le même nombre de motifs internes que **bA3-2** mais n'interagissant pas avec **A3** d'après des études de SPR. (Figure 35)

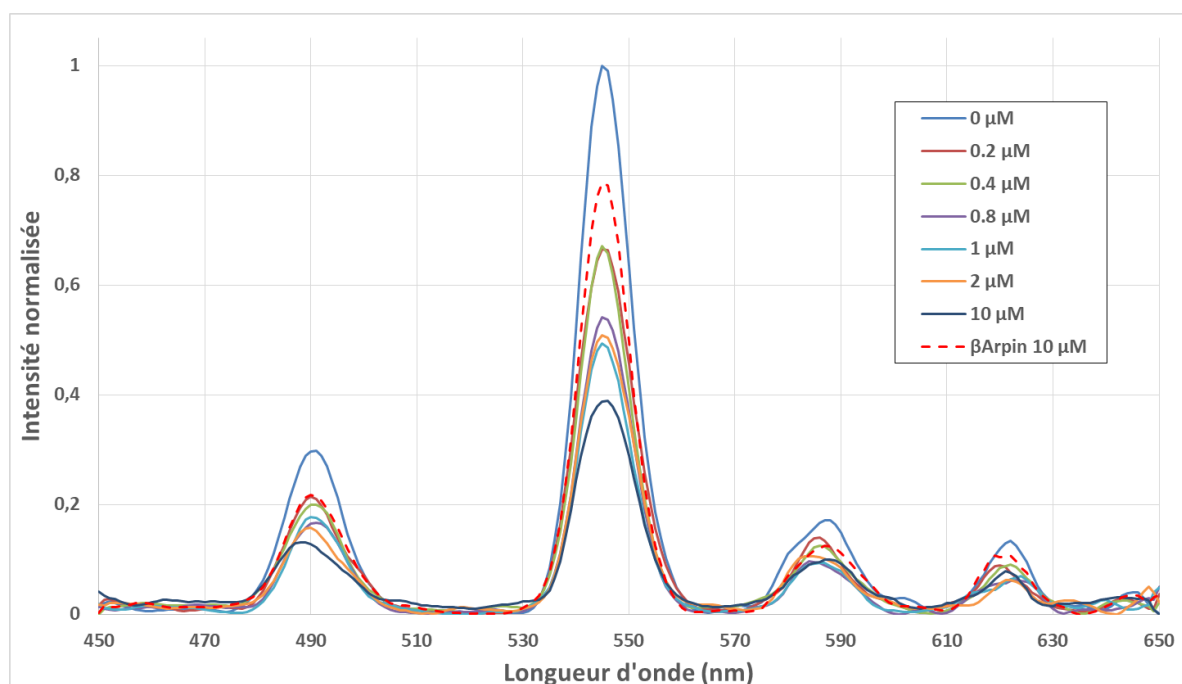


Figure 35 : Spectres normalisés sur la valeur la plus haute. LBTs_A3_A3_WW est à 1 μ M en présence de 10 μ M de Tb^{3+} dans du tampon HEPES. Les concentrations de bA3-2 varient de 0 à 10 μ M. La mesure a été effectuée 50 μ s après excitation avec un temps d'intégration de 1ms.

Ces résultats nous montrent un fait intéressant, même en présence de 10 μ M de **bA3-2** les quatre pics de fluorescence sont toujours présents. Cela pourrait s'expliquer par la présence de quatre Trp le long de la protéine **A3**. Ces derniers pourraient effectuer un homo-FRET (transfert d'énergie de Trp à Trp (Weber, G., 1960)) et transférer leur énergie au Tb^{3+} présent dans le LBT.

Ces résultats montrent qu'il y a bien une diminution du signal en présence de **bA3-2** mais cette diminution n'est pas proportionnelle à la concentration de **bA3-2**. En effet, les concentrations de 0.2 et 0.4 μM se superposent. De plus, le contrôle négatif qui est la bArpin entraîne également une diminution, certes moins importante que **bA3-2** à concentration égale, mais étant tout de même problématique. L'utilisation de la BSA à la place de la bArpin donne également une baisse de signal laissant supposer que le LBT n'est pas stable en présence d'autres protéines ou que le Tb^{3+} se fixe sur les autres protéines non spécifiquement.

Pour vérifier les hypothèses susmentionnées, j'ai effectué une cinétique avec une concentration de bArpin de 8 μM et en mesurant régulièrement le spectre. (Figure 36)

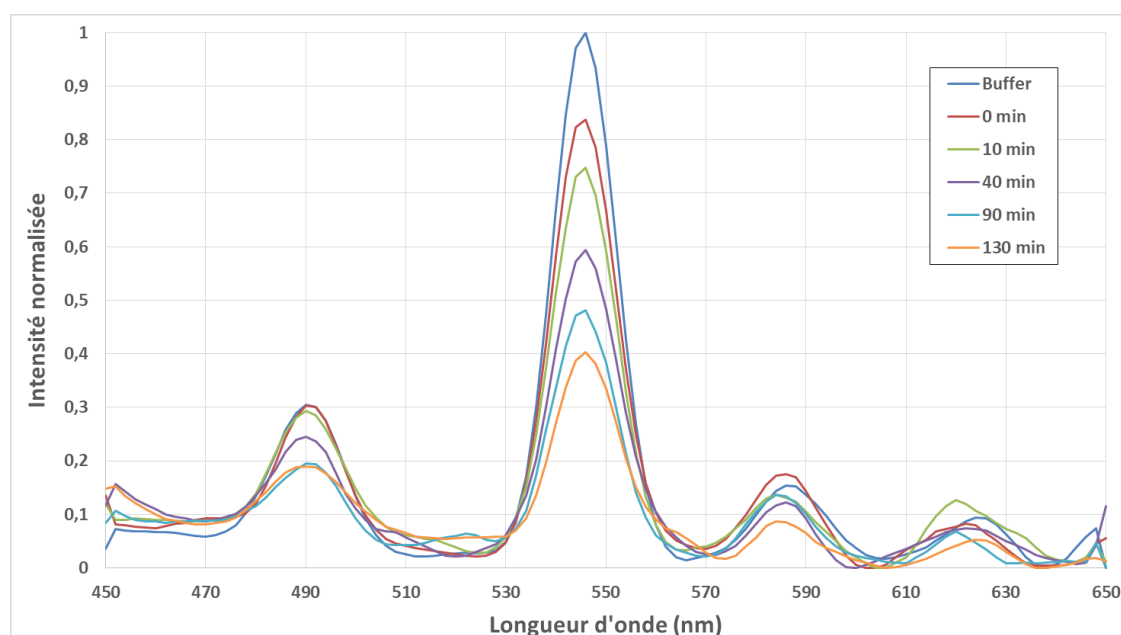


Figure 36 : Spectres normalisés sur la valeur la plus haute. LBTs_A3_A3_WW est à 1 μM en présence de 10 μM de Tb^{3+} dans du tampon HEPES. La concentration en bArpin est de 8 μM . La mesure a été effectuée 50 μs après excitation avec un temps d'intégration de 1ms.

Comme le montrent ces spectres, il y a une cinétique lente de diminution du signal en présence de bArpin. Cette variation du signal avec le temps est également observée pour la BSA mais pas pour **bA3-2**. Cela semble indiquer que les protéines non spécifiques perturbent le LBT et / ou interagissent avec le Terbium, rendant impossible l'exploitation de la construction dans l'état.

Devant ces problèmes, nous avons essayé de créer un LBT plus stable et permettant de donner une intensité de signal d'émission de fluorescence plus élevée. Pour concevoir ce nouveau LBT nous nous sommes inspirés d'une autre séquence capable de fixer le Ca^{2+} .

5) La dockérine

La dockérine est un domaine protéique retrouvé chez les bactéries anaérobie et capable d'interagir avec le domaine cohésine. L'interaction de ces deux domaines est fondamentale pour la formation du cellulosome de ces bactéries. (Ding, S.Y., *et al.*, 2001) Ce domaine est capable de lier deux ions Ca^{2+} *via* deux motifs de type boucle-hélice. (Carvalho, A.L., *et al.*, 2003) La superposition de la structure du LBT sur celle de la dockérine révèle que les deux structures sont similaires. Cependant la structure de la dockérine semble plus robuste puisqu'elle possède des hélices venant stabiliser le domaine. Ainsi, l'ajout d'un Trp sur chacun des motifs de la dockérine à des positions judicieusement choisies pourrait permettre d'obtenir un domaine fixant deux Tb^{3+} et donc permettant d'obtenir une meilleure fluorescence (Figure 37). Par la suite il est possible d'envisager l'insertion de cette dockérine modifiée au sein des bidomaines en supprimant les deux Trp et en ajoutant par exemple une antenne à l'autre extrémité comme pour le LBT.

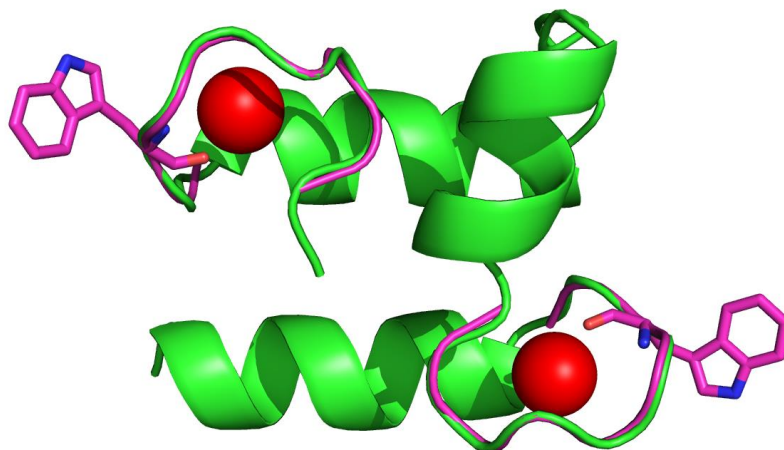


Figure 37 : Représentation cartoon de la dockérine en vert, Calcium (Ca^{2+}) en rouge, LBT en magenta

Dans un premier temps j'ai donc entrepris la modification de la dockérine pour y intégrer deux Trp à des positions similaires à celles du LBT. Sachant que la dockérine se produit difficilement toute seule, j'ai choisi de la fusionner en position C-terminale de la protéine **bGFPD** 3.0 en utilisant un linker PAS puisque le linker (SGGGG)₂ semble présenter une certaine propension à l'hydrolyse. Le linker PAS est constitué de Proline, Alanine et Sérine en proportions variables, il serait *a priori* moins sensible à l'hydrolyse que le linker (SGGGG)₂ et permettrait donc d'obtenir une protéine plus stable. (Schlapschy, M., *et al.*, 2013) Dans notre cas, la séquence choisie est SAPASA. La protéine **bGFPD** 3.0 est un dérivé de **bGFPD** ne possédant pas de C-cap et ayant son 4^{ème} motif interne modifié avec une face externe hydrophile pour permettre une exposition au solvant. La protéine ainsi conçue est appelée **bGFPD_dock**.

La production de cette protéine s'est effectuée dans *E. Coli BL21 (Gold)* à 30°C. La purification qui a suivi consistait ici aussi à une chromatographie d'affinité sur une His-Trap 5 mL suivie d'une colonne de filtration sur gel S75 16/600 dans du tampon Hepes 50 mM, NaCl 150 mM, pH 7.4 contenant 2 mM de CaCl₂ pour permettre à la dockérine d'être repliée autour des Ca²⁺. Le chromatogramme de suivi de la colonne de filtration sur gel présente deux pics (Figure 38).

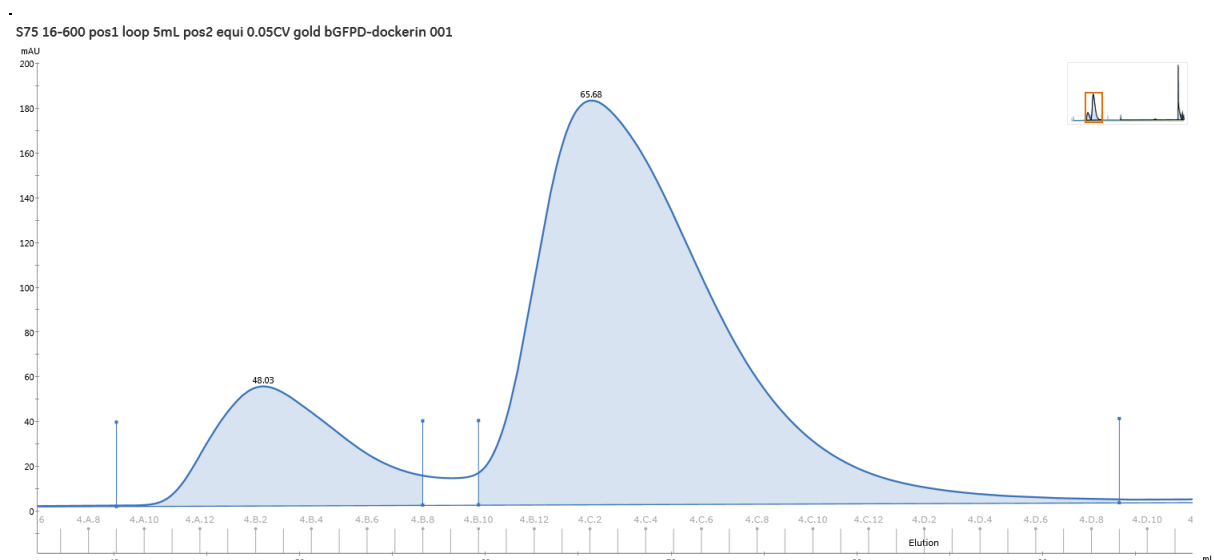


Figure 38 : Chromatogramme à 280 nm lors de la purification sur colonne de gel filtration S75 16-600 de la protéine **bGFPD-Dock**.

La purification a ici été intégralement suivie par SDS-PAGE, la masse moléculaire attendue est de 25kDa (Figure 39).

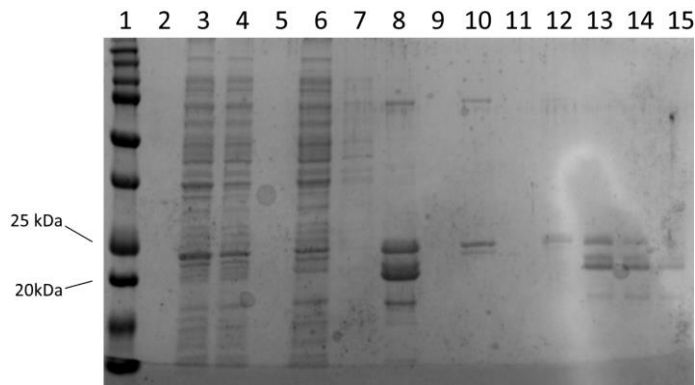


Figure 39 : Gel SDS-PAGE 15% de suivi de purification de la bGFPD_dock.

1: Marqueur de taille Protein Dual Color.
3 : Fraction Totale. 4 : Fraction soluble.
6 : Lavage 20mM imidazole His-Trap.
7 : Fraction non retenue. 8 : Éluion His-Trap.
10 : 1^{er} pic GF S75 16/600. 12-15 : 2nd pic GF S75 16/600.

Le gel SDS-PAGE permet de voir que la protéine est d'une part produite en très faible quantité sous forme soluble et d'autre part présente une forte tendance à d'hydrolyse. Cette hydrolyse semble avoir lieu principalement au niveau de la dockérine puisque les masse des bandes inférieures sont d'environ 5-6 kDa de moins, soit la masse de la dockérine. La fraction déposée dans le puits 10 semblant cependant contenir plus de forme entière, j'ai voulu tester si elle était capable de fixer le Tb³⁺ et d'émettre un signal de fluorescence.

J'ai donc effectué un dessalage dans du tampon Hepes 50 mM, NaCl 150 mM, pH 7.4 sans CaCl₂ puis j'ai ajouté du chlorure de terbium et ai testé cet échantillon pour de la fluorescence en excitant simplement à 280 nm. Je n'ai cependant pas observé de signal d'émission de fluorescence caractéristique d'un lanthanide. Je n'ai hélas pour le moment pas eu le temps de poursuivre ce projet.

6) Conclusion

L'ensemble de ces résultats permet d'entrevoir les possibilités qui existent avec le LBT. Bien que présentant des problèmes de stabilité et de spécificité, l'optimisation de ces constructions pourrait permettre de réaliser des biosenseurs dans un futur plus ou moins proche.

C. La nanoKaz

1) Origine de la nanoKaz

La nanoKaz (ou nanoLuc en fonction de la séquence nucléotidique codante), est une enzyme de 19 kDa capable catalyser une réaction bioluminescente ($\lambda_{\max} = 454 \text{ nm}$) et dérivant d'une luciférase d'*Oplophorus gracilirostris*. Cette luciférase était initialement un hétérotetramère (2 * 35 kDa + 2 * 19 kDa) utilisant la coelenterazine comme substrat. En 2000, l'équipe d'O. Shimomura a montré que seule la sous-unité de 19 kDa était nécessaire pour déclencher de la bioluminescence. (Inouye, S., et al., 2000) Cette bioluminescence se fait suite à l'oxydation d'un substrat appelé Coelenterazine. Par la suite, cette sous-unité a subi des modifications, notamment dans le but d'augmenter sa stabilité et sa spécificité pour un substrat donnant une réaction plus lumineuse. Ainsi, seize mutations du gène codant pour la luciférase ont été trouvées par mutagenèse aléatoire. La protéine ainsi obtenue, appelée NanoLuc ou nanoKaz entraîne une luminescence améliorée d'un facteur $2.5 \cdot 10^6$ et utilise un nouveau substrat dérivant de la coelenterazine, la furimazine. De plus, la durée de vie intracellulaire de la NanoKaz est améliorée et sa stabilité thermique accrue. (Hall, M.P., et al., 2012) (Figure 40)

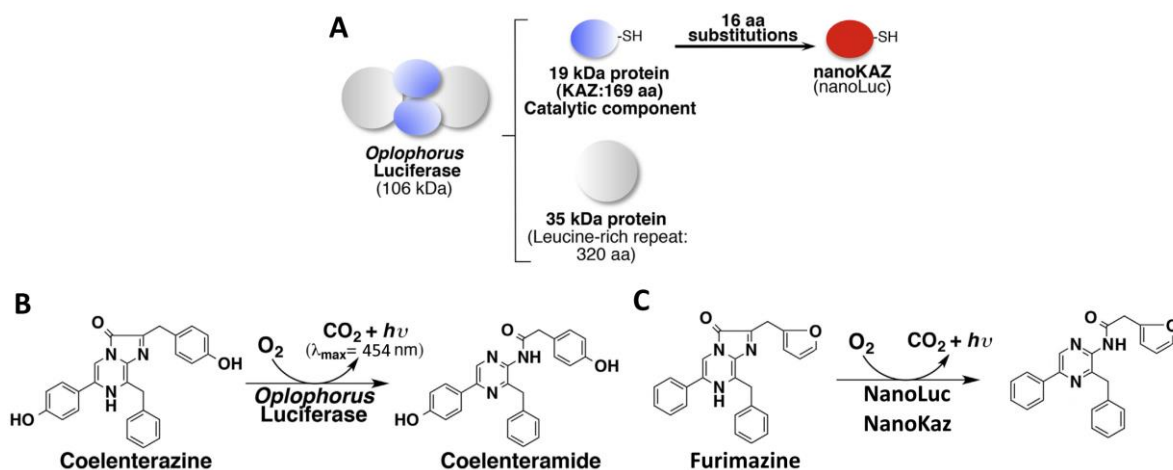


Figure 40 : Origine de la NanoKaz et réaction catalysée. A : Schéma d'ingénierie qui a permis d'obtenir la nanoKaz. B : Réaction initialement catalysée par la luciférase d'*Oplophorus gracilirostris*, utilisant la Coelenterazine. C : Réaction catalysée par la nanoKaz après optimisation, utilisant la furimazine comme substrat. Adapté de (Tomabechi, Y., et al., 2016)

Finalement sa structure a été résolue et a révélé la présence d'une crevasse recouverte par trois hélices. Cette crevasse constituerait le site de fixation du substrat. Les hélices semblent être mobiles pour permettre l'accès à la crevasse. (Tomabechi, Y., *et al.*, 2016)

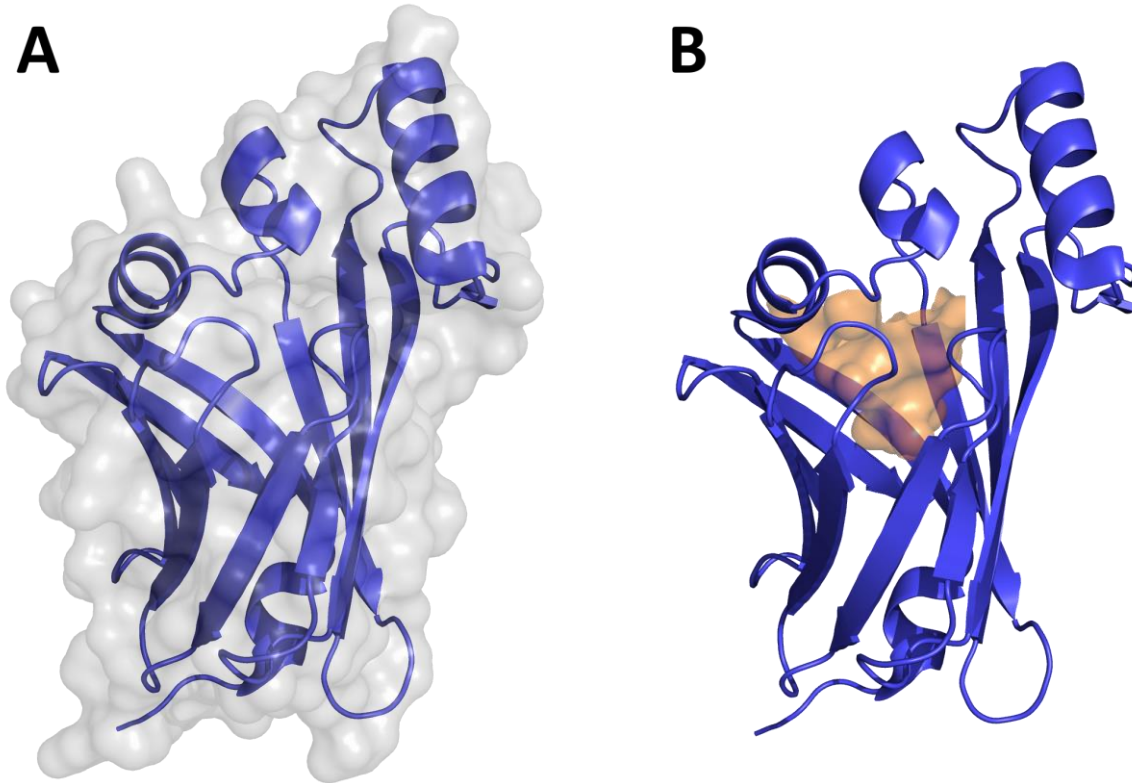


Figure 41 : Structure de la NanoKaz (code pdb : 5b0u). A : Représentation cartoon avec surface en gris. B : Mise en évidence de la crevasse interne accueillant le substrat en orange.

2) Conception du biosenseur

De manière générale, dans le cas des biosenseurs, il est préférable de détecter l'apparition d'un signal plutôt que sa disparition. En effet, la disparition peut être due à une instabilité du biosenseur. Dans le cas de la nanoKaz, il semble possible d'imaginer une protéine inhibitrice interagissant au niveau des hélices ou dans la crevasse et empêchant l'accès du substrat au site catalytique. L'interaction ou la non-interaction de cet inhibiteur pourrait être contrôlée par la présence d'un ligand entraînant un changement de conformation de ce biosenseur.

En partant de ce postulat, nous avons conçu au sein de l'équipe un biosenseur constitué de quatre protéines. Cette chimère est composée de l'extrémité N-terminale à l'extrémité C-terminale, d'une nanoKaz, de deux α Reps **bGFPD** et d'un interacteur de nanoKaz (protéines MG-X). La chimère est entourée d'un TwinStrep-Tag en N-terminal et d'un His-Tag en C-terminal. Afin d'insérer facilement les gènes codant pour les interacteurs de la nanoKaz dans cette construction, les sites BamH I et Hind III flanquent ce dernier. Le gène codant pour cette protéine MG-X est présenté ci-dessous.



Figure 42 : Représentation schématique du gène codant pour la chimère nanoKaz. L'interacteur de la nanoKaz est appelé protéine MG-X.

Nous faisons l'hypothèse qu'une telle protéine à 4 domaines devrait adopter une conformation fermée. En effet, d'une part, **bGFPD** dimérise à partir de 70 μ M, d'autre part l'interacteur de la nanoKaz interagira avec la nanoKaz. Ces deux interactions devraient être donc renforcées au sein de la chimère, garantissant la fermeture de celle-ci. En effet, comme nous l'avons démontré sur les bidomains (Partie B - I), des interactions intramoléculaires sont plus fortes que des interactions équivalentes entre deux molécules non reliées. La présence de la **BFP** ou de ses variants, ligands de **bGFPD**, (une nouvelle fois, la GFP absorbe dans le spectre d'émission de la nanoKaz) entraînerait une ouverture empêchant l'inhibition par l'interacteur et permettant au substrat de la nanoKaz d'interagir. (Figure 43)

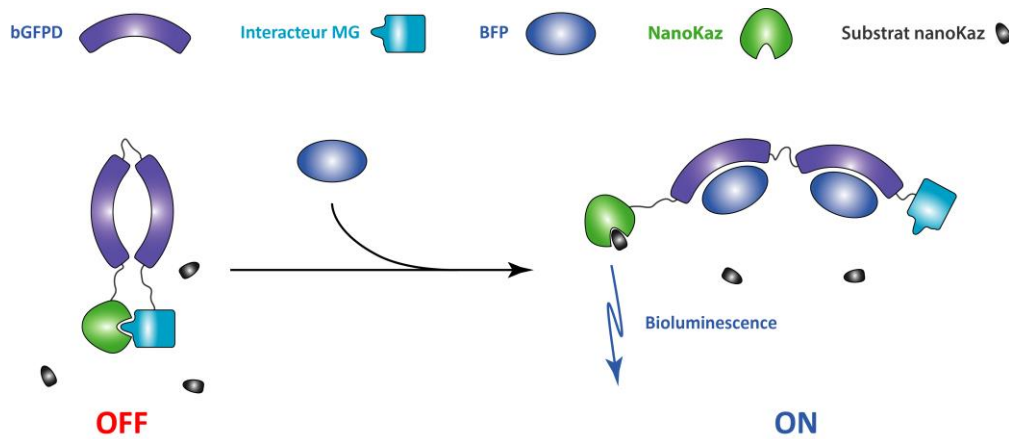


Figure 43 : Principe du biosenseur basé sur l'utilisation de la nanoKaz. En absence de substrat, la chimère adopte une conformation fermée, inhibant la nanoKaz. En présence du substrat, la chimère s'ouvre et transmet un signal de bioluminescence.

3) Sélection des interacteurs de la NanoKaz

a) Source des interacteurs de NanoKaz

Les interacteurs de nanoKaz ont été sélectionnés au sein d'une bibliothèque de protéines artificielles (protéines MG) conçue pour interagir plus spécifiquement avec des crevasses. Cette bibliothèque a été réalisée durant une thèse CIFRE effectuée par la Dr. Margarida Gomez au sein de l'équipe MIP.

b) Production et purification de la NanoKaz

La sélection nécessitant l'immobilisation de la nanoKaz, il a dans un premier temps fallu produire cette dernière. Afin d'augmenter sa stabilité et pour faciliter sa production, elle a été fusionnée à une α Rep constituée de trois motifs internes identiques suivie d'un avi-Tag. L'avi-Tag est un site de biotinylation qui permettra l'immobilisation de la protéine pour la sélection grâce à de la streptavidine. Un linker constitué d'un site TEV et de la séquence (SGG)₂ sépare les deux protéines. Un gène synthétique codant pour cette protéine chimère a ainsi été commandé, la protéine obtenue est appelé NanoKaz_HRN3s. (Figure 44)



Figure 44 : Schéma du gène synthétique codant pour la nanoKaz_HRN3s

Sa production s'est effectuée dans une souche *E. Coli BL21 Gold pBirA* à 37°C. La présence du plasmide pBirA code pour l'enzyme BirA qui permettra la biotinylation de la nanoKaz-HRN3s. Sa purification a suivi un protocole classique His-Trap 5mL et gel filtration S75 16/600 et le suivi a été effectué par SDS-PAGE (Figure 45). La masse moléculaire attendue est de 40 kDa. Ce gel SDS-Page montre que la protéine semble posséder un contaminant de masse moléculaire d'environ 30 kDa. Ce contaminant n'est pas biotinylé d'après un western blot (non montré ici), il ne posera par conséquent pas de problème pour l'immobilisation de la protéine lors des sélections.

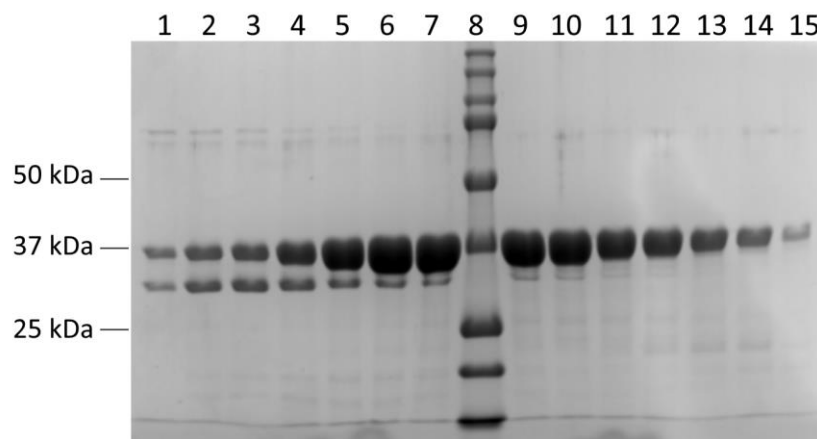


Figure 45 : SDS-PAGE des éluions de la colonne S75 16/600. De 1 à 7 et 9 à 15 : Éluion Gel Filtration. 8 : Marqueur de masse moléculaire Protein dual color

c) Sélection par phage display

Des sélections ont donc été réalisées par phage display contre la nanoKaz-HRN3s à partir des bibliothèques MG. (Smith, G., 1985) Deux éluions ont été effectuées. La première à la TEV protéase va permettre de libérer la nanoKaz de la protéine HRN3s et donc d'éluer spécifiquement les phages fixant la nanoKaz. (Figure 46) La seconde, à la glycine, va décrocher tous les phages restants dans le cas où l'élution par la TEV ne donne aucun clone.

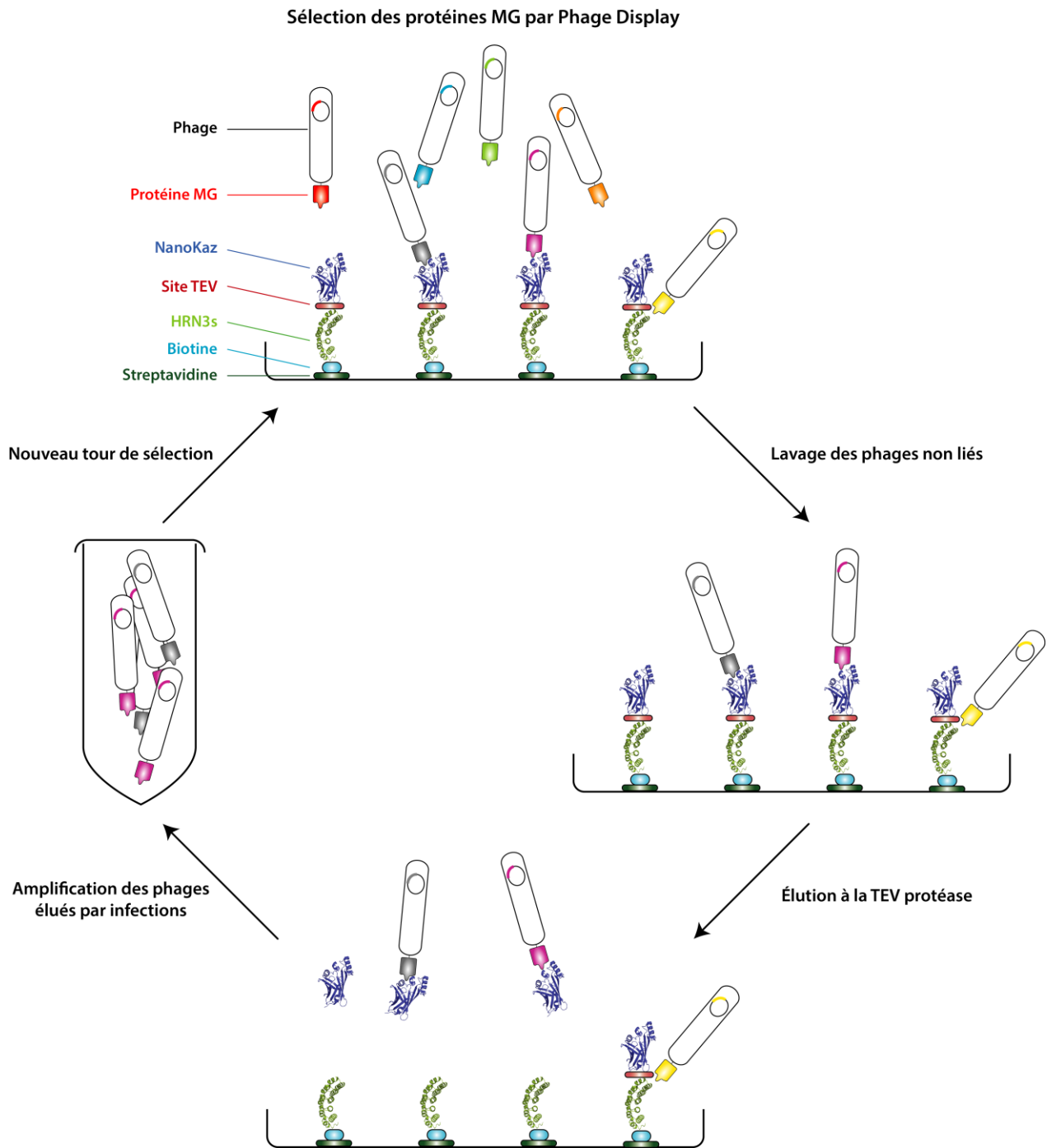


Figure 46 : Sélection par phages display des protéines interagissant avec la protéine NanoKaz_HRN3s. La cible est fixée grâce à sa biotine, reconnaissant la streptavidine recouvrant les plaques. L'étape de lavage permet d'éliminer les phages non spécifiques. L'éluion à la TEV décrochera uniquement les phages exposant une protéine MG fixant la nanoKaz. L'éluion à la glycine n'est pas représenté sur ce schéma par soucis de clarté. L'amplification des phages se fait par infection de bactéries XL1Blue.

d) Phage ELISA Clonal

Au terme de trois tours de sélection, un phage ELISA clonal est effectué selon les protocoles classiques utilisés au laboratoire (Guellouz, A., *et al.*, 2013). Cette étape consiste à tester 96 clones issus de l'input et des trois tours de sélection. Pour ce faire, chaque clone est récupéré, des phages clonaux sont alors produits puis incubés sur deux plaques ELISA préparées où la cible est présente une ligne sur deux afin de tester la spécificité et la fixation des clones. Après plusieurs lavages, des anticorps reconnaissant les phages et couplés à une peroxydase sont ajoutés. Les plaques sont alors à nouveau lavées puis le substrat de la peroxydase (POD) est ajouté. Les plaques ont par la suite été scannées sur un lecteur de type Infinite F200 Pro (TECAN) afin de quantifier le degré de coloration et donc le taux de fixation des phages. La [figure 47](#) montre que les clones spécifiques apparaissent dès le tour 2. Seize clones positifs et spécifiques en phage ELISA ont été séquencés et huit clones différents ont été obtenus. Les séquences correspondantes ont été sous-clonées dans un vecteur d'expression et les protéines correspondantes ont été produites et purifiées.

4) Caractérisation de l'interaction protéines MG / NanoKaz

Afin de vérifier l'interaction des clones MG avec la nanoKaz, la détermination des affinités par ITC a été réalisée et révèle que les clones ont des constantes de dissociation K_D de l'ordre du μM . Bien que peu élevées (valeurs de K_D de l'ordre de la dizaine de micromolaire), ces affinités semblent adaptées pour la construction du biosenseur. En effet puisqu'étant au sein de la même protéine, la concentration effective des protéines MG sera très supérieure au K_D . Par conséquent, les protéines MG devraient quasiment toutes être fixées sur la nanoKaz à l'image de **bGFPD** qui se lie à **A3** dans la chimère **A3_bGFPD**. De plus, un K_D de l'ordre du μM garantit que la chimère puisse toujours être ouverte par les ligands, un K_D plus faible signifiant qu'il est nécessaire d'avoir des ligands avec une affinité plus forte.

Les deux protéines MG (MG-F5 et MG-G10) présentant la plus haute affinité pour la nanoKaz et une non spécifique (MG-WT) ont par la suite été utilisées dans la construction chimérique décrite ci-dessus ([figure 43](#)).

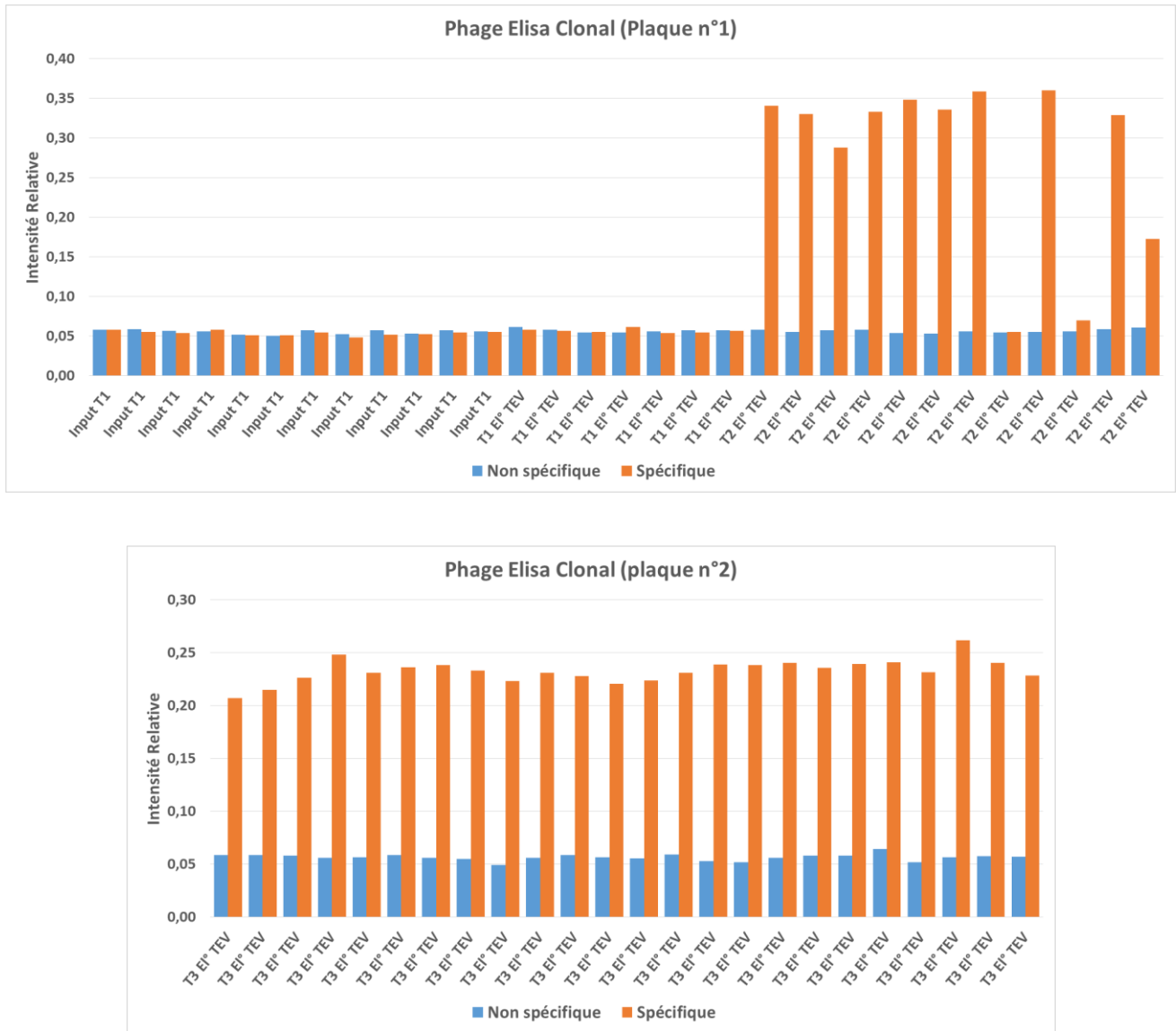


Figure 47 : Résultats du phage ELISA clonal. L'input T1 correspond aux phages provenant de la banque de protéine MG. T1, T2 et T3 correspondent aux tours 1, 2 et 3. E1° TEV correspondent à l'élution à la TEV protéase. Par soucis de clarté, les élutions à la Glycine ne sont pas représenté sur ce graphique. Les temps d'incubation avec le substrat POD soluble n'ayant pas été les mêmes entre les deux plaques réalisées, ces dernières sont représentées séparément.

5) Production et purification des chimères

La production des chimères à quatre domaines a été effectuée dans des bactéries *E. Coli BL21 Gold* à 30°C. Les chimères possédant un His-Tag en C-terminal et un TwinStrep-Tag en N-terminal, leur purification a eu lieu simplement, d'abord sur colonne Ni-NTA puis sur une colonne constituée de résine Streptactine. Un gel SDS-PAGE a été effectué pour suivre la purification des chimères dont la masse attendue est de 72 kDa. Le gel correspondant à chimère contenant la protéine MG-F5 (ChimKaz-F5) est présenté ci-dessous et révèle une purification efficace. Ce fut le cas pour les trois chimères étudiées. Les quantités produites sont de l'ordre de 15 à 30 mg/L de culture.

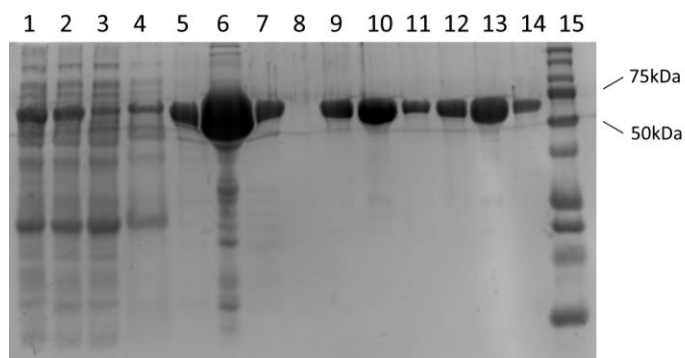


Figure 48 : Gel SDS-PAGE 12% de suivi de purification de la chimKaz-F5. 1 : Fraction Totale 2 : Fraction soluble 3 : Fraction Non retenue 4 : Lavage de la colonne Ni-NTA 5-6-7 : Elution de la colonne NiNTA 9-14 : Elution de la colonne Strep-Trap 15 : Marqueur de masse moléculaire Protein dual color

6) Caractérisation des Chimères

a) Spécificité de détection

Avant de tester l'efficacité des chimères, il était nécessaire de vérifier simplement la spécificité d'ouverture de la chimère. Le contrôle non spécifique utilisé ici est la galactose-mutarotase (GALM), une protéine globulaire de la même masse moléculaire que la BFP. Des expériences d'optimisation des conditions expérimentales ont permis de déterminer les concentrations optimales en nanoKaz-HRN3s (appelé nanoKaz par la suite) de 10 pM et en Furimazine 0.15 % de la solution mère (concentration mère non donnée par Promega®). Ces concentrations seront utilisées pour toutes les expériences réalisées avec les chimères et permettent d'obtenir un signal linéaire sur une période de plusieurs minutes.

Ne disposant au début du projet que de la chimère F5 (ChimKaz-F5) et de la nanoKaz, l'expérience de spécificité a été effectuée uniquement sur ces deux protéines. La comparaison entre nanoKaz et la ChimKaz-F5 montre que les deux ligands GALM et BFP à 1 μM n'ont pas d'effet sur la nanoKaz par rapport au tampon. En revanche, pour la ChimKaz-F5, la bioluminescence augmente de 25 % en présence de la BFP mais ne change pas en présence de GALM. Ces résultats semblent indiquer que le senseur ainsi conçu est spécifique pour la BFP.

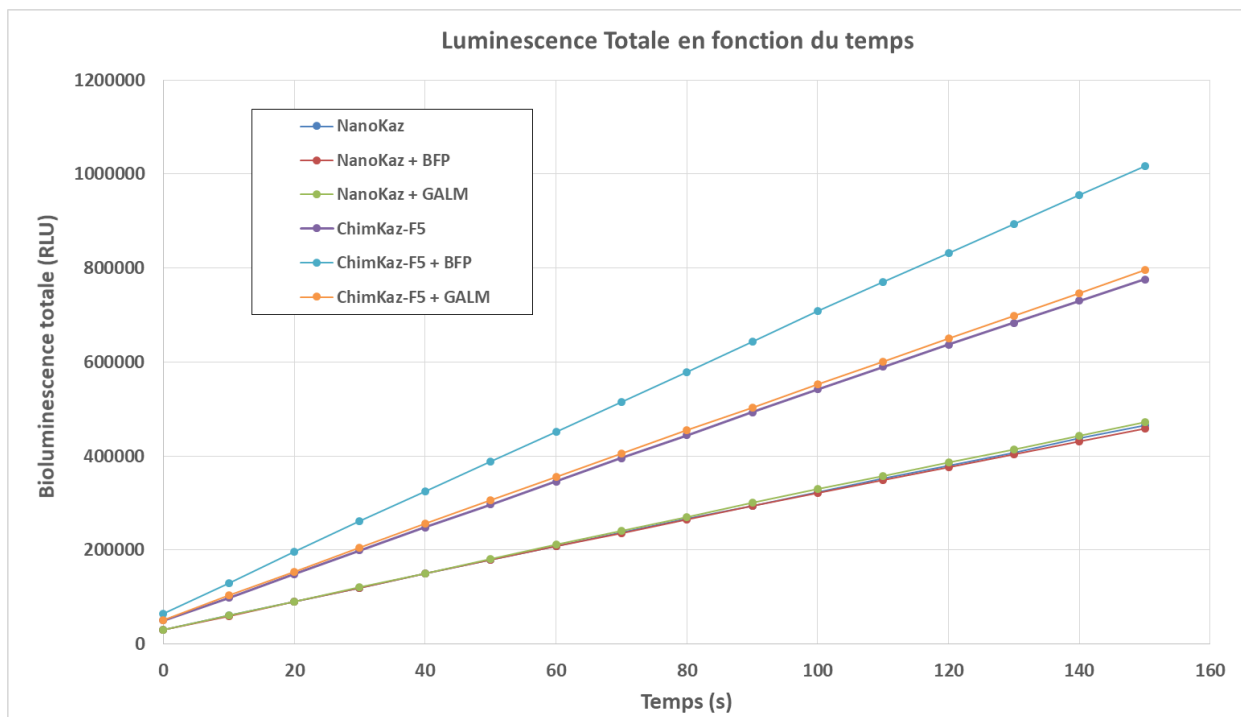


Figure 49 : Cinétique de bioluminescence totale en fonction du temps. La nanoKaz et la ChimKaz-F5 était à 10 μM . La BFP et la GALM sont à 1 μM . Le substrat (nanoGlo) est à une concentration de 0.15%. La bioluminescence totale n'augmente pas pour la nanoKaz que cela soit en présence de BFP ou en présence de GALM (trois courbes inférieures). Pour la ChimKaz-F5, la bioluminescence totale n'augmente qu'en présence de BFP (courbes supérieures).

Il faut noter que la nanoKaz a une bioluminescence totale plus faible que la chimKaz-F5. Cela est très probablement dû au fait qu'il s'agissait d'un lot ayant subi plusieurs congélation et décongélation. Ce résultat n'a pas été reproduit avec des lots frais de nanoKaz.

b) Mesure des affinités par ITC

Disposant par la suite des chimères contenant MG-G10 et MG-WT (ChimKaz-G10 et ChimKaz-WT), les affinités qu'elles ont pour la BFP ont pu être déterminées par ITC. Les constantes de dissociation sont de 87 ± 45 , 417 ± 59 et 532 ± 127 nM pour les chimères WT, F5 et G10 respectivement. (Figure 50) Tout d'abord, le K_D de la ChimKaz-WT pour la BFP est nettement plus élevé que celui de la BFP pour le domaine bGFPD seule. Cela s'explique par le fait que les domaines bGFPD interagissent ensemble au sein de la chimère, rendant plus difficile la fixation de la BFP. D'autre part, bien que possédant une incertitude élevée, ces résultats indiquent clairement que l'ajout de protéine MG spécifique de la nanoKaz vient verrouiller la conformation fermée de la chimère. Cela se traduit par l'augmentation du K_D pour les ChimKaz-F5 et G10 par rapport à celui pour la ChimKaz-WT.

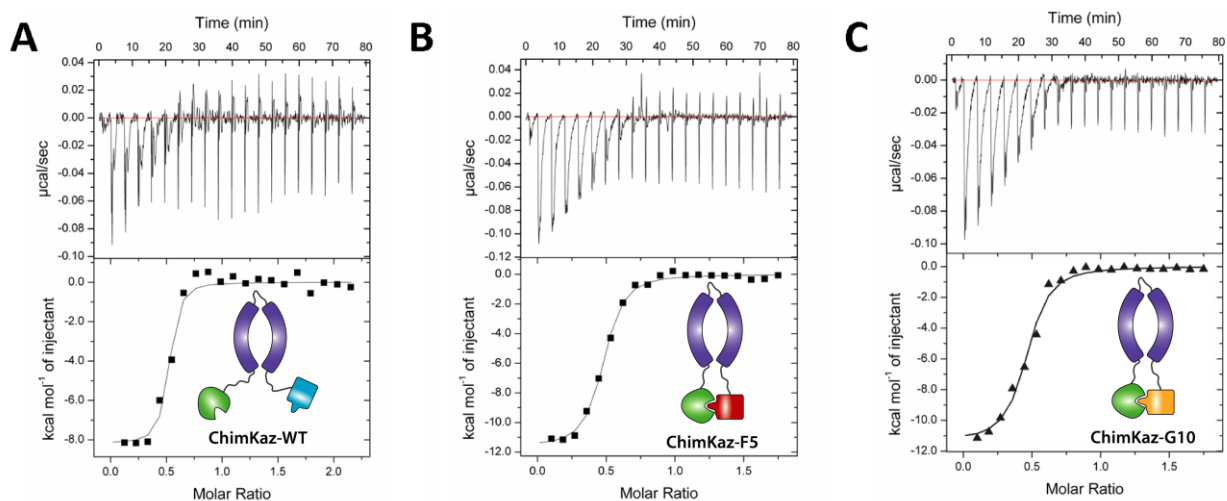


Figure 50 : Détermination des affinités par microcalorimétrie.

A : ChimKaz-WT, adopte une conformation partiellement ouverte sans le verrouillage protéine MG/nanoKaz. $K_D = 87 \pm 45$ nM.

B : ChimKaz-F5 adoptant une conformation complètement fermée. $K_D = 417 \pm 59$ nM.

C : ChimKaz-G10 adoptant aussi une conformation complètement fermée. $K_D = 532 \pm 127$ nM.

c) Comparaison des chimères et détection de la BFP

J'ai par la suite testé si une protéine MG non spécifique, bien que perturbant moins l'ouverture de la chimère que les protéines spécifiques, avait une influence sur la bioluminescence. Ainsi j'ai comparé les luminescences respectives de la nanoKaz, ChimKaz-WT, ChimKaz-F5 et la ChimKaz-G10.

Les mesures d'intensité montrent que la Chimkaz-WT a une bioluminescence égale à 84% de celle de la nanoKaz seule. Il semble donc y avoir soit un effet par la présence d'une molécule proche de la nanoKaz (gène stérique), soit une instabilité de la nanoKaz la rendant moins efficace lorsqu'elle est dans la chimère. Les protéines ChimKaz-F5 et ChimKaz-G10 ont quant à elles une bioluminescence égale à environ 25 % celle de la nanoKaz. Ces valeurs sont en accord avec les mesures d'ITC. Ces résultats prouvent que les protéines MG-F5 et MG-G10 contenues dans les chimères sont capables d'inhiber la nanoKaz. En présence de concentration croissante de BFP, le signal de ChimKaz-WT ne remonte pas et reste au environ de 85 % du signal de la nanoKaz. L'intensité de ChimKaz-F5 remonte lui jusqu'à environ 33 % du signal de la nanoKaz et le signal de ChimKaz-G10 ne varie pas.

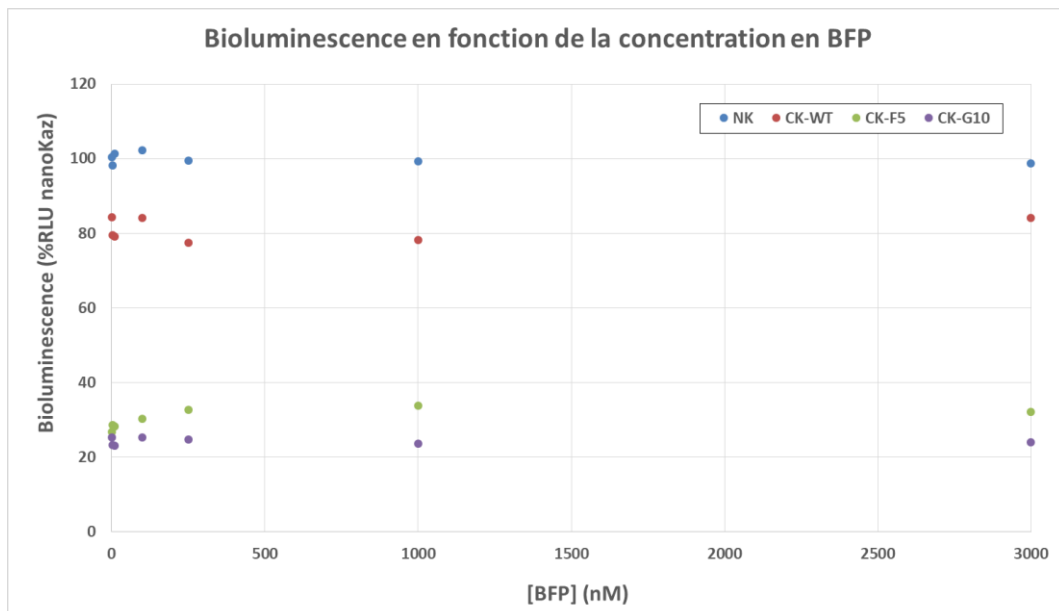


Figure 51 : Bioluminescence de la nanoKaz et des chimKaz-F5 et -G10 (10 pM) en fonction de la concentration de BFP (0-2.5-10-100-250-1000-3000 nM).

7) Conclusion

L'ensemble de ces résultats semble indiquer que les protéines MG sont certes capables d'inhiber la nanoKaz mais pas totalement. Il serait peut-être nécessaire d'obtenir des interacteurs plus affins. Cependant la levée de l'inhibition ne se fait pas totalement en présence de BFP laissant supposer que les interacteurs sont trop affins pour la nanoKaz. Ce paradoxe indique qu'il est nécessaire d'optimiser la construction. Il semble également envisageable de changer le substrat. En effet, la furimazine est le substrat optimisé pour la nanoKaz, or son affinité est plus élevée que celle de la coelenterazine, le substrat « naturel ». On peut donc imaginer que l'utilisation du substrat moins affin pourrait régler le problème d'inhibition partielle des chimères. En effet, ce substrat étant moins affins et en raison de la compétition entre domaine MG et substrat pour l'accès au site actif de la nanoKaz, le substrat aura plus de « difficulté » pour accéder au site actif.

IV. Bibliographie

- Achyuthan, K. E., D. C. Arango, E. L. Carles, C. E. Cutler, L. A. Meyer, and S. M. Brozik. 2009. 'Luminescent investigations of terbium(III) biosorption as a surrogate for heavy metals and radionuclides', *Mol Cell Biochem*, 327: 87-92.
- Banta, Scott, Kevin Dooley, and Oren Shur. 2013. 'Replacing antibodies: engineering new binding proteins', *Annual Review of Biomedical Engineering*, 15: 93-113.
- Barthelmes, K., A. M. Reynolds, E. Peisach, H. R. Jonker, N. J. DeNunzio, K. N. Allen, B. Imperiali, and H. Schwalbe. 2011. 'Engineering encodable lanthanide-binding tags into loop regions of proteins', *J Am Chem Soc*, 133: 808-19.
- Bick, Matthew J., Per J. Greisen, Kevin J. Morey, Mauricio S. Antunes, David La, Banumathi Sankaran, Luc Reymond, Kai Johnsson, June I. Medford, and David Baker. 2017. 'Computational design of environmental sensors for the potent opioid fentanyl', *Elife*, 6.
- Boersma, Y. L., and A. Pluckthun. 2011. 'DARPin and other repeat protein scaffolds: advances in engineering and applications', *Curr Opin Biotechnol*, 22: 849-57.
- Bradbury, A., and A. Pluckthun. 2015. 'Reproducibility: Standardize antibodies used in research', *Nature*, 518: 27-9.
- Carlin, K. B., C. A. Cruz-Teran, J. P. Kumar, C. Gomes, and B. M. Rao. 2016. 'Combinatorial Pairwise Assembly Efficiently Generates High Affinity Binders and Enables a "Mix-and-Read" Detection Scheme', *ACS Synth Biol*, 5: 1348-54.
- Carvalho, A. L., F. M. Dias, J. A. Prates, T. Nagy, H. J. Gilbert, G. J. Davies, L. M. Ferreira, M. J. Romao, and C. M. Fontes. 2003. 'Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex', *Proc Natl Acad Sci U S A*, 100: 13809-14.
- Chevrel, A., A. Mesneau, D. Sanchez, L. Celma, S. Quevillon-Cheruel, A. Cavagnino, S. Nessler, I. Li de la Sierra-Gallay, H. van Tilbeurgh, P. Minard, M. Valerio-Lepiniec, and A. Urvoas. 2017. 'A repeat proteins (α Rep) as expression and crystallization helpers', *J Struct Biol*, 201: 88-99.
- Chevrel, A., A. Urvoas, I. Li de la Sierra-Gallay, M. Aumont-Nicaise, S. Moutel, M. Desmadril, F. Perez, A. Gautreau, H. van Tilbeurgh, P. Minard, and M. Valerio-Lepiniec. 2015. 'Specific GFP-binding artificial proteins (α Rep): a new tool for in vitro to live cell applications', *Biosci Rep*, 35.
- Collinet, B., M. Herve, F. Pecorari, P. Minard, O. Eder, and M. Desmadril. 2000. 'Functionally accepted insertions of proteins within protein domains', *The Journal of Biological Chemistry*, 275: 17428-33.

- de Picciotto, Seymour, Paige M. Dickson, Michael W. Traxlmayr, Bryan S. Marques, Elke Socher, Sixing Zhao, Stephanie Cheung, Jonathan D. Kiefer, A. Joshua Wand, Linda G. Griffith, Barbara Imperiali, and K. Dane Wittrup. 2016. 'Design Principles for SuCESsFul Biosensors: Specific Fluorophore/Analyte Binding and Minimization of Fluorophore/Scaffold Interactions', *Journal of Molecular Biology*, 428: 4228-41.
- Ding, S. Y., M. T. Rincon, R. Lamed, J. C. Martin, S. I. McCrae, V. Aurilia, Y. Shoham, E. A. Bayer, and H. J. Flint. 2001. 'Cellulosomal scaffoldin-like proteins from *Ruminococcus flavefaciens*', *J Bacteriol*, 183: 1945-53.
- Feng, J., B. W. Jester, C. E. Tinberg, D. J. Mandell, M. S. Antunes, R. Chari, K. J. Morey, X. Rios, J. I. Medford, G. M. Church, S. Fields, and D. Baker. 2015. 'A general strategy to construct small molecule biosensors in eukaryotes', *Elife*, 4.
- Ferrandez, Yann, Manuela Dezi, Mickael Bosco, Agathe Urvoas, Marie Valerio-Lepiniec, Christel Le Bon, Fabrice Giusti, Isabelle Broutin, Grégory Durand, Ange Polidori, Jean-Luc Popot, Martin Picard, and Philippe Minard. 2014. 'Amphipol-mediated screening of molecular orthoses specific for membrane protein targets', *The Journal of Membrane Biology*, 247: 925-40.
- Figuroa, M., M. Sleutel, M. Vandevenne, G. Parvizi, S. Attout, O. Jacquin, J. Vandenameele, A. W. Fischer, C. Damblon, E. Goormaghtigh, M. Valerio-Lepiniec, A. Urvoas, D. Durand, E. Pardon, J. Steyaert, P. Minard, D. Maes, J. Meiler, A. Matagne, J. A. Martial, and C. Van de Weerd. 2016. 'The unexpected structure of the designed protein Octarellin V.1 forms a challenge for protein structure prediction tools', *J Struct Biol*, 195: 19-30.
- Franz, K. J., M. Nitz, and B. Imperiali. 2003. 'Lanthanide-binding tags as versatile protein coexpression probes', *Chembiochem*, 4: 265-71.
- Fu, X., X. Fu, Q. Wang, L. Sheng, X. Huang, M. Ma, and Z. Cai. 2017. 'Fluorescence switch biosensor based on quantum dots and gold nanoparticles for discriminative detection of lysozyme', *Int J Biol Macromol*, 103: 1155-61.
- Gonzalez, D., N. Lokhande, S. Vadde, Q. Zhao, A. Cassill, and R. Renthal. 2011. 'Luminescence resonance energy transfer in the cytoplasm of live *Escherichia coli* cells', *Biochemistry*, 50: 6789-96.
- Guellouz, A., M. Valerio-Lepiniec, A. Urvoas, A. Chevrel, M. Graille, Z. Fourati-Kammoun, M. Desmadril, H. van Tilbeurgh, and P. Minard. 2013. 'Selection of specific protein binders for pre-defined targets from an optimized library of artificial helicoidal repeat proteins (α Rep)', *PLoS One*, 8: e71512.
- Guntas, G., T. J. Mansell, J. R. Kim, and M. Ostermeier. 2005. 'Directed evolution of protein switches and their application to the creation of ligand-binding proteins', *Proc Natl Acad Sci U S A*, 102: 11224-9.

- Ha, Jeung-Hoi, and Stewart N. Loh. 2012. 'Protein conformational switches: from nature to design', *Chemistry (Weinheim an Der Bergstrasse, Germany)*, 18: 7984-99.
- Hadpech, Sudarat, Sawitree Nangola, Koollawat Chupradit, Kanda Fanhchaksai, Wilhelm Furnon, Agathe Urvoas, Marie Valerio-Lepiniec, Philippe Minard, Pierre Boulanger, Saw-See Hong, and Chatchai Tayapiwatana. 2017. 'A-helicoidal HEAT-like Repeat Proteins (α Rep) Selected as Interactors of HIV-1 Nucleocapsid Negatively Interfere with Viral Genome Packaging and Virus Maturation', *Scientific Reports*, 7: 16335.
- Hall, M. P., J. Unch, B. F. Binkowski, M. P. Valley, B. L. Butler, M. G. Wood, P. Otto, K. Zimmerman, G. Vidugiris, T. Machleidt, M. B. Robers, H. A. Benink, C. T. Eggers, M. R. Slater, P. L. Meisenheimer, D. H. Klaubert, F. Fan, L. P. Encell, and K. V. Wood. 2012. 'Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate', *ACS Chem Biol*, 7: 1848-57.
- Huang, Jin, and Shohei Koide. 2010. 'Rational conversion of affinity reagents into label-free sensors for Peptide motifs by designed allostery', *ACS chemical biology*, 5: 273-77.
- Inouye, Satoshi, Ken Watanabe, Hideshi Nakamura, and Osamu Shimomura. 2000. 'Secretional luciferase of the luminous shrimp *Oplophorus gracilirostris*: cDNA cloning of a novel imidazopyrazinone luciferase', *FEBS Letters*, 481: 19-25.
- Jumeaux, C., O. Wahlsten, S. Block, E. Kim, R. Chandrawati, P. D. Howes, F. Hook, and M. M. Stevens. 2018. 'MicroRNA Detection by DNA-Mediated Liposome Fusion', *Chembiochem*, 19: 434-38.
- Kirby, Anthony J. 1980. 'Effective Molarities for Intramolecular Reactions', 17: 183-278.
- Krishnamurthy, V. M., V. Semetey, P. J. Bracher, N. Shen, and G. M. Whitesides. 2007. 'Dependence of effective molarity on linker length for an intramolecular protein-ligand system', *Journal of the American Chemical Society*, 129: 1312-20.
- Kummer, Lutz, Chia-Wen Hsu, Onur Dagliyan, Christopher MacNevin, Melanie Kaufholz, Bastian Zimmermann, Nikolay V. Dokholyan, Klaus M. Hahn, and Andreas Plückthun. 2013. 'Knowledge-based design of a biosensor to quantify localized ERK activation in living cells', *Chemistry & Biology*, 20: 847-56.
- Legendre, D., P. Soumillion, and J. Fastrez. 1999. 'Engineering a regulatable enzyme for homogeneous immunoassays', *Nat Biotechnol*, 17: 67-72.
- Medintz, I., and N. Hildebrandt. 2014. *FRET - Förster resonance energy transfer : from theory to applications*.
- Miranda, Frederico F., Elodie Brient-Litzler, Nora Zidane, Frédéric Pecorari, and Hugues Bedouelle. 2011. 'Reagentless fluorescent biosensors from artificial families of antigen binding proteins', *Biosensors & Bioelectronics*, 26: 4184-90.

- Moutel, S., N. Bery, V. Bernard, L. Keller, E. Lemesre, A. de Marco, L. Ligat, J. C. Rain, G. Favre, A. Olichon, and F. Perez. 2016. 'NaLi-H1: A universal synthetic library of humanized nanobodies providing highly functional antibodies and intrabodies', *Elife*, 5.
- Nicholes, N., A. Date, P. Beaujean, P. Hauk, M. Kanwar, and M. Ostermeier. 2016. 'Modular protein switches derived from antibody mimetic proteins', *Protein Eng Des Sel*, 29: 77-85.
- Parmeggiani, F., and P. S. Huang. 2017. 'Designing repeat proteins: a modular approach to protein design', *Curr Opin Struct Biol*, 45: 116-23.
- Radley, Tracy L., Anna I. Markowska, Blaine T. Bettinger, Jeung-Hoi Ha, and Stewart N. Loh. 2003. 'Allosteric Switching by Mutually Exclusive Folding of Protein Domains', *Journal of Molecular Biology*, 332: 529-36.
- Sapsford, Kim E., Thomas Pons, Igor L. Medintz, Seiichiro Higashiya, Florence M. Brunel, Philip E. Dawson, and Hedi Mattoussi. 2007. 'Kinetics of Metal-Affinity Driven Self-Assembly between Proteins or Peptides and CdSe-ZnS Quantum Dots', *The Journal of Physical Chemistry C*, 111: 11528-38.
- Schlapschy, M., U. Binder, C. Borger, I. Theobald, K. Wachinger, S. Kisling, D. Haller, and A. Skerra. 2013. 'PASylation: a biological alternative to PEGylation for extending the plasma half-life of pharmaceutically active proteins', *Protein Eng Des Sel*, 26: 489-501.
- Shelat, Nirav Y., Sidhartha Parhi, and Marc Ostermeier. 2017. 'Development of a cancer-marker activated enzymatic switch from the herpes simplex virus thymidine kinase', *Protein engineering, design & selection: PEDS*, 30: 95-103.
- Smith, G. 1985. 'Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface', *Science*, 228: 1315-17.
- Stein, V., and K. Alexandrov. 2015. 'Synthetic protein switches: design principles and applications', *Trends Biotechnol*, 33: 101-10.
- Tiede, C., R. Bedford, S. J. Heseltine, G. Smith, I. Wijetunga, R. Ross, D. AlQallaf, A. P. Roberts, A. Balls, A. Curd, R. E. Hughes, H. Martin, S. R. Needham, L. C. Zanetti-Domingues, Y. Sadigh, T. P. Peacock, A. A. Tang, N. Gibson, H. Kyle, G. W. Platt, N. Ingram, T. Taylor, L. P. Coletta, I. Manfield, M. Knowles, S. Bell, F. Esteves, A. Maqbool, R. K. Prasad, M. Drinkhill, R. S. Bon, V. Patel, S. A. Goodchild, M. Martin-Fernandez, R. J. Owens, J. E. Nettleship, M. E. Webb, M. Harrison, J. D. Lippiat, S. Ponnambalam, M. Peckham, A. Smith, P. K. Ferrigno, M. Johnson, M. J. McPherson, and D. C. Tomlinson. 2017. 'Affimer proteins are versatile and renewable affinity reagents', *Elife*, 6.
- Tiouajni, M., D. Durand, K. Blondeau, M. Graille, A. Urvoas, M. Valerio-Lepiniec, A. Guellouz, M. Aumont-Nicaise, P. Minard, and H. van Tilbeurgh. 2014. 'Structural and functional analysis of the fibronectin-binding protein FNE from *Streptococcus equi* spp. *equi*', *FEBS J*, 281: 5513-31.

- Tomabechi, Y., T. Hosoya, H. Ehara, S. Sekine, M. Shirouzu, and S. Inouye. 2016. 'Crystal structure of nanoKAZ: The mutated 19 kDa component of *Oplophorus* luciferase catalyzing the bioluminescent reaction with coelenterazine', *Biochem Biophys Res Commun*, 470: 88-93.
- Urvoas, A., A. Guellouz, M. Valerio-Lepiniec, M. Graille, D. Durand, D. C. Desravines, H. van Tilbeurgh, M. Desmadril, and P. Minard. 2010. 'Design, production and molecular structure of a new family of artificial α -helical repeat proteins (α Rep) based on thermostable HEAT-like repeats', *J Mol Biol*, 404: 307-27.
- Vallée-Bélisle, Alexis, and Kevin W. Plaxco. 2010. 'Structure-switching biosensors: inspired by Nature', *Current Opinion in Structural Biology*, 20: 518-26.
- Wachter, R. M., B. A. King, R. Heim, K. Kallio, R. Y. Tsien, S. G. Boxer, and S. J. Remington. 1997. 'Crystal structure and photodynamic behavior of the blue emission variant Y66H/Y145F of green fluorescent protein', *Biochemistry*, 36: 9759-65.
- Weber, G. 1960. 'Fluorescence-polarization spectrum and electronic-energy transfer in proteins', *Biochem J.*, 75(2): 345-52.
- Zhang, H., J. Lv, and Z. Jia. 2017. 'Efficient Fluorescence Resonance Energy Transfer between Quantum Dots and Gold Nanoparticles Based on Porous Silicon Photonic Crystal for DNA Detection', *Sensors (Basel)*, 17.
- Zheng, Huimei, Jing Bi, Mira Krendel, and Stewart N. Loh. 2014. 'Converting a binding protein into a biosensing conformational switch using protein fragment exchange', *Biochemistry*, 53: 5505-14.

Partie C : Matériels & Méthodes

Tables des matières

I.	BIOLOGIE MOLÉCULAIRE	186
1)	Conceptions des protéines	186
a)	Chimères A3_A3 et A3_bGFPD	186
b)	Protéines TwinStrep	188
c)	LBTs_A3_A3_WW	190
d)	NanoKaz et Chimères NanoKaz	191
2)	Séquençage des gènes codant les protéines :	191
II.	MICROBIOLOGIE	192
1)	Milieu de culture	192
2)	Transformation XL1 blue électro-compétentes	192
3)	Transformation bactéries chimio-compétentes.....	193
4)	Expression et production des protéines.....	193
III.	PURIFICATION DES PROTÉINES	194
1)	Purification Ni-NTA des protéines possédant un His-Tag.....	194
2)	Purification Streptactine des protéines possédant un <i>Strep-Tag</i>	194
3)	Colonne de filtration sur gel.....	194
4)	Clivage à la protéase « Tobacco Etch Virus endopeptidase » (TEV).....	195
5)	Dialyse des protéines	195
6)	Gel SDS-PAGE	195
IV.	BIOPHYSIQUE	196
1)	Differential Scanning Calorimetry (DSC)	196
2)	Size Exclusion Chromatography – Multi Angle Light Scattering (SEC-MALS)	197
3)	Isothermal Titration Calorimetry (ITC)	198
4)	Surface Plasmon Resonance (SPR)	199
5)	Ultracentrifugation analytique (UCA).....	200
6)	Small Angle X-rays Scattering (SAXS).....	200
7)	Cristallographie	201
V.	EXPÉRIENCE DE FRET.....	202
1)	Réduction des cystéines	202
2)	Couplage des ATTO-dye	202
3)	Expérience préliminaire de détection des ligands	202
4)	Couplage du cryptate de Terbium.....	203
5)	Couplage des Quantum Dots 625.....	203
6)	Préparation des expériences de FRET	203
7)	Détermination des Limites de détection.....	204
8)	Détermination des distances.....	204

VI. SÉLECTION PAR PHAGE DISPLAY.....	205
1) Production des phages.....	205
2) Préparation des plaques ELISA.....	205
3) Tours de sélection.....	206
a) Sélection.....	206
b) Infection.....	206
c) Production de phages.....	207
4) ELISA Clonal.....	207
VII. BIBLIOGRAPHIE.....	209

I. Biologie Moléculaire

NB : Les analyses de séquences (alignement, design) ont été effectuées à l'aide du logiciel Serial Cloner (SerialBasics).

1) Conceptions des protéines

a) Bidomaines A3_A3 et A3_bGFPD

Les α Reps étant principalement composées de séquence répétées, les étapes de biologie moléculaire, notamment celles basées sur des PCR, peuvent se révéler difficiles à cause des répétitions de codons. La réalisation d' α Reps bidomaines accentue ce phénomène. Afin de simplifier ces étapes, le domaine N-terminal codant pour **A3** a vu sa séquence nucléotidique modifiée. Ainsi, un gène synthétique appelé **a3s** a été conçu pour ne pas entraîner de répétition de séquence avec les gènes codant pour **A3** et **bGFPD**. D'autre part, ce gène possède des codons différents aux positions équivalentes entre motif, garantissant une possibilité d'amplification simple par PCR (Figure 1).

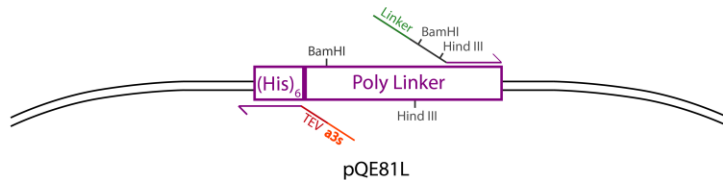
I - CPEC

Étape 1: Amplification par PCR

a) Préparation de l'insert



b) Préparation du vecteur linéaire

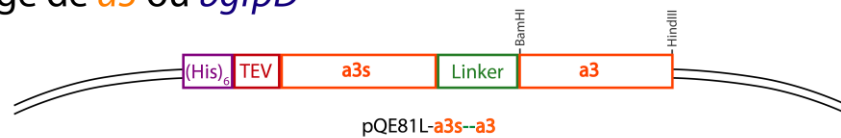


Étape 2: Assemblage final par CPEC



II - Clonage de *a3* ou *bgfpD*

a)



or

b)

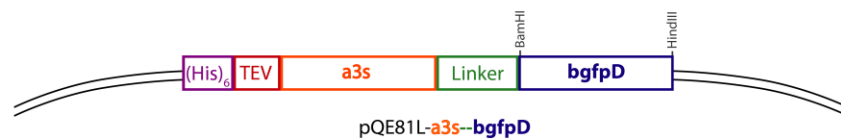


Figure 1 : Conception des gènes codant pour les protéines A3_A3 et A3_bGFPD par CPEC.

I-CPEC : Étape 1-a : Préparation de l'insert : Le gène *a3s* a été au préalable cloné dans un vecteur *pQE80L* avec un site *TEV* en aval. Une 1^{ère} PCR est effectuée avec les amorces « *His-PQE81L-TEVa3s* » et « *a3s---Linker* ». Le produit de PCR correspond à la séquence du gène *a3s* flanqué par la séquence codant l'étiquette Histidine d'un *pQE81L* en aval et par la séquence codant une partie du linker en amont.

Étape 1-b : Préparation du vecteur linéaire : Une 2^{nde} PCR est effectuée avec les amorces « *Linker-BamH I/Hind III* » et « *His-TEVa3s* ». Le produit de PCR correspond à un *pQE81L* linéarisé avec les modifications suivantes. Les sites de restriction du site de clonage multiple supprimés. Une séquence complémentaire du linker suivi des sites *BamH I* et *Hind III* est ajouté en 5'. Une séquence complémentaire au site *TEV* et à une partie du gène *a3s* est ajoutée en 3'.

Étape 2 : Insertion du gène *a3s* dans le vecteur linéarisé : l'assemblage par CPEC (*Circular Polymerase Extension Cloning*) a été effectué en mélangeant les produits de PCR à un ratio molaire de 1 : 1. L'amplification a alors lieu sans ajout d'amorce.

II-Clonage des gènes *a3* et *bGFPD* dans le plasmide *pQE81L-as3-linker*. Les gènes et le vecteur sont doublement digérés par les enzymes de restriction *BamH I* et *Hind III*. Les gènes digérés (*a3* ou *bGFPD*) sont alors mélangés avec le vecteur digéré puis une ligation est effectuée.

b) Protéines TwinStrep

La conception des protéines TwinStrep avait pour objectif de s'affranchir de la séquence His-Tag pouvant mal se cliver lors des étapes de clivage à la TEV. Afin de construire ces protéines, un pQE80L a été modifié pour enlever cette séquence His-tag et rajouter un TwinStrep. Une cassette a ainsi été réalisée à partir de trois couples d'amorces :

K7-FOR-1 5' AA TTC ATT AAA GAG GAG AAA TTA ACT ATG AGA GGA TCT ACG GAT CCG

K7-REV-1 3' G TAA TTT CTC CTC TTT AAT TGA TAC TCT CCT AGA TGC CTA GGC TGC GCA CGC AGC

K7-FOR-2 5' ACG CGT GCG TCG AAA AGC TTA TGG TCA CAT CCG CAG TTT GAA AAA GGT GGT GGT

K7-REV-2 3' TTT TCG AAT ACC AGT GTA GGC GTC AAA CTT TTT CCA CCA CCA TCG C CA CCA CCA AGT

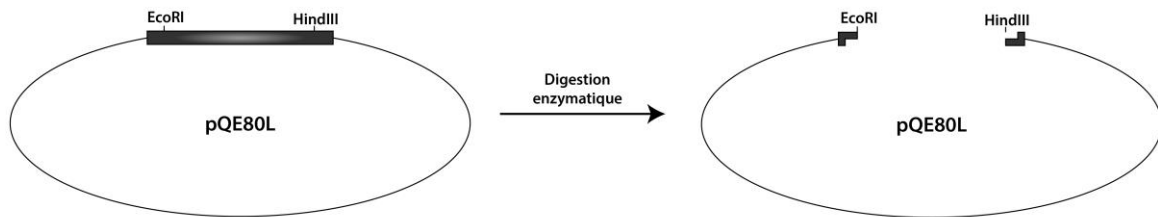
K7-FOR-3 5' AGC GGT GGT GGT TCA GGT GGT AGT GCT TGG AGC CAC CCT CAG TTT GAG AAA TAA C

K7-REV-3 3' CCA CCA TCA CGA ACC TCG GTG GGA GTC AAA CTC TTT ATT GTC GA

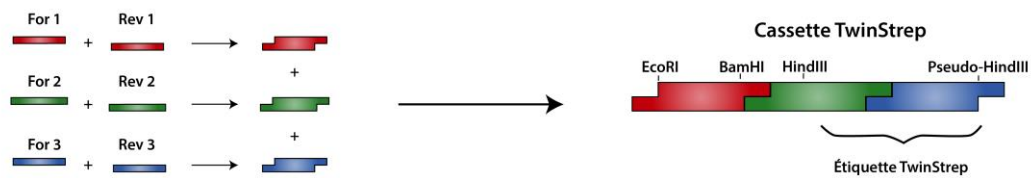
Chacun de ces couples a aux préalable été incubé indépendamment en concentration équimolaire avant d'être mélangé pour former la cassette finale. Celle-ci s'insère dans le vecteur pQE80L digéré par EcoR I et Hind III, cependant, elle possède un pseudo-site Hind III en 5' empêchant par la suite la reformation du site Hind III à cet emplacement. Deux nouveaux sites BamH I et Hind III ont donc été rajouté dans cette cassette (en noir/gras ci-dessus) permettant l'insertion des α Reps. L'étiquette TwinStrep est présente après le site Hind III interne à la cassette. (Figure 2)

Par la suite, les α Reps ont pu être insérées après une simple digestion BamH I/ Hind III suivi d'une ligation. La eGFP et la BFP n'étant pas dans le bon cadre de lecture, des amorces ont été conçues pour amplifier les gènes de ces deux protéines tout en rajoutant un nucléotide pour décaler le site BamH I.

Digestion du plasmide



Assemblage de la cassette



Assemblage du nouveau vecteur

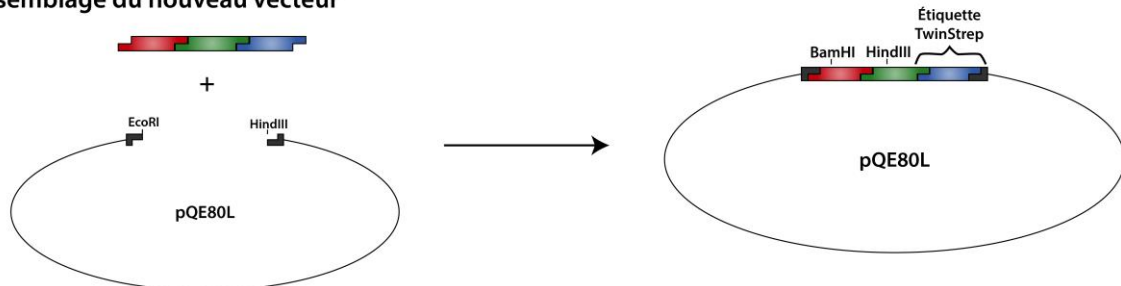


Figure 2 : Conception du vecteur accepteur TwinStrep permettant par la suite l'insertion des alphaReps, de la eGFP et de la BFP.

c) LBTs_A3_A3_WW

Disposant déjà d'un gène codant pour LBT_A3_A3, il était plus simple de modifier cette protéine plutôt que de partir d'une construction A3_A3. L'idée derrière cette construction était pour rappel d'enlever le Trp présent dans le LBT et de rajouter une antenne constituée de deux Trp en C-ter de la protéine. Pour se faire, un *Gibson Assembly* couplé à deux mutagénèses par PCR ont été effectués. Dans un premier il a donc fallut amplifier par PCR le vecteur et l'insert en remplaçant d'une part le Trp par une Ser et d'autre part en ajoutant l'antenne Trp. Dans un second temps, un *Gibson Assembly* a été effectué en mélangeant les deux produits de PCR (Vecteur et Insert).

Amplification et mutation du vecteur et de l'insert

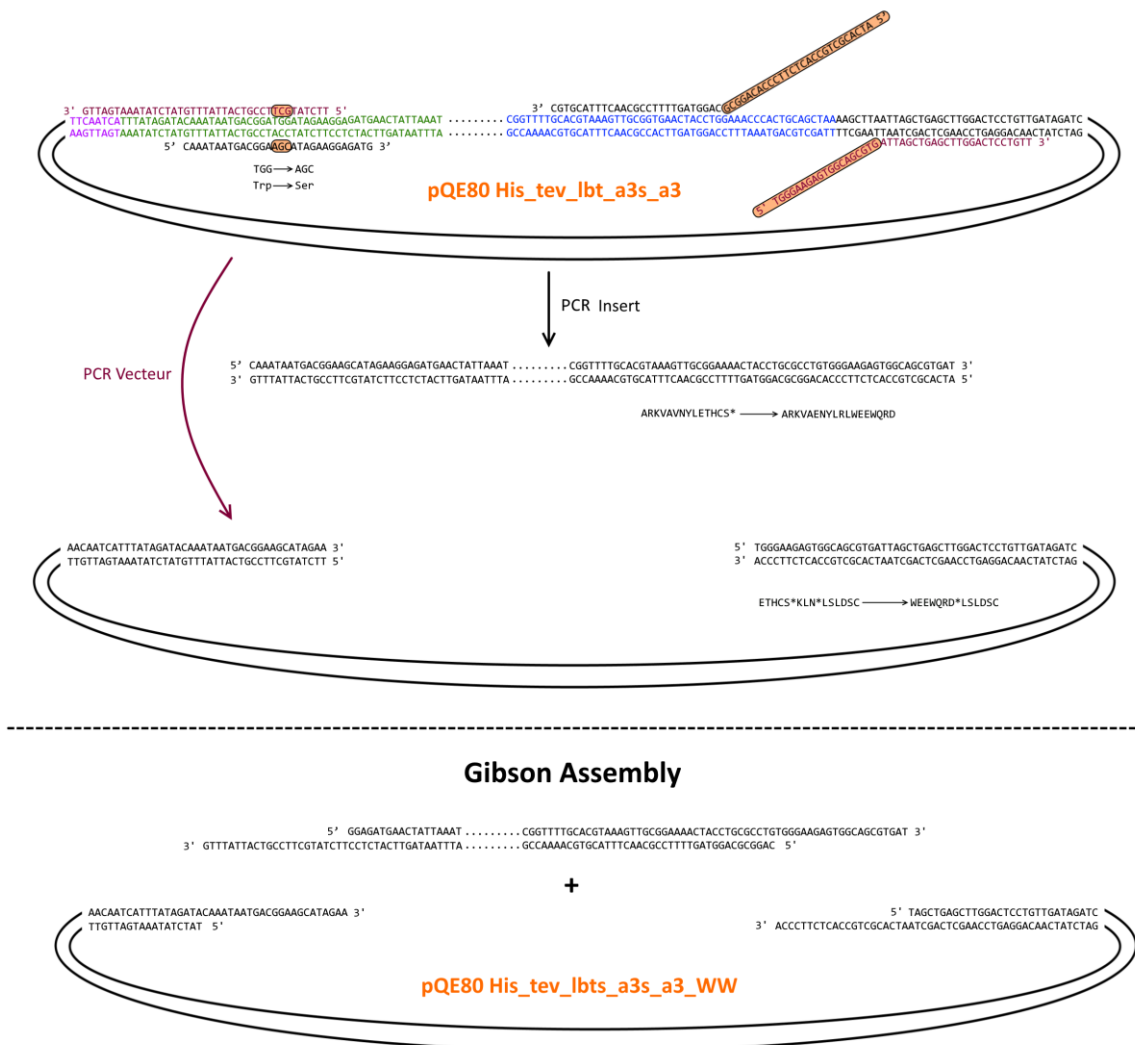


Figure 3 : Conception de la protéine LBTs_A3_A3_WW. Les amorces de la PCR du vecteur sont en rouge foncé, celles de l'insert sont en noire.

d) NanoKaz et Chimères NanoKaz

Dans les deux cas, un gène synthétique pouvant s'insérer dans pQE80L (NanoKaz_HRN3s) ou pQE81 (ChimKaz) a été conçu puis commandé chez *Integrated DNA Technologies*.

Le gène codant pour la nanoKaz_HRN3s peut s'insérer dans pQE80L par digestion BamH I / Hind III suivi d'une ligation. Il est constitué d'un His-Tag, de la nanoKaz, d'un site TEV, d'un linker (SGG)₂, de l'αRep HRN3s et d'un Avi-Tag.

Le gène codant pour la ChimKaz contenait initialement, un TwinStrep Tag, la nanoKaz, deux bGFPD et la protéine MG-F5 comme interacteur de nanoKaz et un His-Tag. Chacune des protéines est séparée par un *linker* constitué de glycine, de sérine et d'alanine. Le gène s'insère par digestion EcoR I / Nhe I dans pQE81. La protéine MG-F5 est flanquée par les sites BamH I et Hind III, il est ainsi possible de la remplacer par les autres protéines MG en digérant le gène par ces enzymes puis en effectuant une ligation.

Les gènes sont représentés schématiquement ci-dessous :

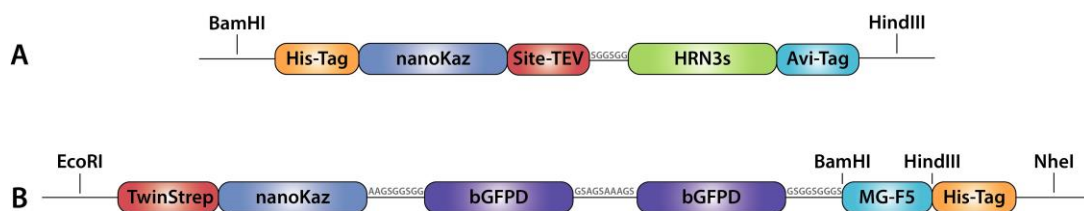


Figure 4 : Schéma des gènes synthétiques commandés. A : NanoKaz-HRN3s B : ChimKaz-F5

2) Séquençage des gènes codant les protéines :

Une préculture de 5 mL de 2xYT amp-tetra-glu a été réalisée à partir d'un clone XL1 Blue MRF' provenant d'une transformation bactérienne avec le plasmide d'intérêt puis incubé à 37°C à 220 RPM O/N. La purification des plasmides a eu ensuite lieu selon le protocole des MiniPreps *MACHEREY-NAGEL*. Les séquençages des gènes codant pour les diverses protéines ont été réalisés par *Eurofins Scientific* puis vérifiés sur *Serial Cloner* en comparant avec les séquences théoriques.

II. Microbiologie

1) Milieu de culture

Toutes les cultures bactériennes ont été réalisées dans du milieu 2xYT (Amresco ou MP). En fonction des conditions nécessaires, les additifs suivants ont été ajoutés :

- Agar : 1,5 %
- Glucose : 1 % (Glu)
- Ampicilline : 100 $\mu\text{g.mL}^{-1}$ (amp)
- Kanamycine : 50 $\mu\text{g.mL}^{-1}$ (kana)
- Tétracycline : 50 $\mu\text{g.mL}^{-1}$ (tetra)
- IPTG : 0.5 μM .
- Biotine :

2) Transformation XL1 blue électro-compétentes

Environ 1 μg de plasmide, soit 1 à 10 μL , ont été ajoutés à 100 μL de bactéries électrocompétentes *E. Coli* XL1blue MRF' décongelées sur glace. Les bactéries sont transférées dans des cuves d'électroporation stériles (VWR) préalablement incubées sur glace. Après un choc électrique (25 μFd / 2,5 volts / 400 Ohms), 900 μL de milieu 2xYT à température ambiante ont été immédiatement ajoutés puis les bactéries ont été mises à incuber 1 h à 37°C à 220 RPM. Elles ont enfin été étalées sur deux boîtes contenant amp, tetra et glu et incubées à 37°C sur la nuit (O/N). 100 μL de culture bactérienne diluée 10 fois ont été étalés sur la première boîte et 100 μL non diluée sur la seconde.

3) Transformation bactéries chimio-compétentes

Les bactéries chimio-compétentes *E. Coli* XL1 Blue MRF' et *E. Coli* BL21 Gold (DE3) ont été transformées selon le même protocole. Environ 100 ng d'ADN plasmidique n'excédant pas 10% du volume de bactéries ont été ajoutés à un aliquote de 200 µL de bactéries compétentes décongelé sur glace. Après une incubation de 20 min, les bactéries ont subi un choc thermique pendant 45 secondes à une température de 42°C puis ont été incubées immédiatement 2 minutes sur glace. Les bactéries ont ensuite été incubées avec 500 µL de 2xYT pendant 1h à 37°C à 220 RPM. Elles ont enfin été étalées sur boîte 2xYT-Glucose 1% avec les antibiotiques nécessaires : tétracycline et ampicilline pour XL1 Blue et uniquement Ampicilline pour Gold.

4) Expression et production des protéines

Une préculture de 50 mL de 2xYT amp-glu a été réalisée à partir d'un clone gold provenant d'une transformation bactérienne puis incubée à 37°C à 220 RPM O/N. La densité optique (DO) à 600 nm de la préculture a été mesurée avec un SmartSpec 3000 (Biorad) puis une culture de 2 x 500 mL de 2xYT amp-kana-glu a étéensemencée avec une DO_{600nm} de 0.15. L'incubation a eu lieu à 37°C à 220 RPM, jusqu'à une DO_{600nm} comprise entre 0.6 et 0.8. A cette valeur, de l'IPTG a été ajouté à la culture qui a, à nouveau, été incubée à 37°C 220 RPM pendant 4 heures. Après une centrifugation à 4000 g pendant 30 minutes, les culots bactérien ont été repris dans 10 mL de tampon NaP 50 mM, NaCl 150 mM pH 7 contenant une pastille « *Protease Inhibitor Cocktail EDTA free* » (Roche). Les culots ont alors subi un cycle de congélation/décongélation dans l'azote liquide puis 1 µL de benzonase a été ajouté et incubé 15 minutes à température ambiante. La benzonase va digérer les acides nucléiques libérés lors de la lyse bactérienne, ces derniers, s'ils ne sont pas digérés rendent la suspension bactérienne visqueuse et difficile à manipuler. Cinq cycles de sonication de 30 secondes ont été réalisés puis les lysats ont été centrifugés 40 minutes à 4°C à 10 000g. Le surnageant également appelé fraction soluble a alors été séparé du culot et conservé à 4°C pour être purifié.

III. Purification des protéines

1) Purification Ni-NTA des protéines possédant un His-Tag

Cette purification a été effectuée sur un Äkta Pure, les débits sont de 5 mL.min⁻¹ durant toute la purification. Le surnageant contenant les protéines (fraction soluble) est appliqué sur une colonne HisTrap FF Crude préalablement équilibrée avec 5 volumes de colonne (CV) de tampon NaP 50 mM, NaCl 300 mM, Imidazole 20 mM, pH 8. La colonne est alors lavée avec ce même tampon jusqu'à obtenir une variation de DO_{280nm} sur 2 min inférieure à 0.1. Les protéines fixées sont alors éluées avec 4 CV de tampon NaP 50 mM, NaCl 300 mM, Imidazole 300 mM, pH 8 et récupérées par fraction de 1 mL.

2) Purification Streptactine des protéines possédant un *Strep-Tag*

Cette purification a été effectuée sur un Äkta™ Pure, les débits sont de 5 mL.min⁻¹ sauf lors de l'application de l'échantillon où celui-ci est de 2 mL.min⁻¹. Le surnageant contenant les protéines (fraction soluble) est appliqué sur une colonne StrepTrap™ HP préalablement équilibrée avec 5 volumes de colonne (CV) de tampon Tris/HCl 100 mM, NaCl 150 mM, 1 mM EDTA, pH 8. La colonne est alors lavée avec ce même tampon jusqu'à obtenir une variation de DO_{280nm} sur 2min inférieure à 0,1. Les protéines fixées sont alors éluées avec 4 CV de tampon Tris/HCl 100 mM, NaCl 150 mM, EDTA 1 mM, d-desthiobiotine 2.5mM, pH 8 et récupérées par fraction de 1 mL.

3) Colonne de filtration sur gel

Les chromatographies d'exclusion préparatoire suivant l'étape de purification sur colonne His-Trap ou Streptrap™ ont été réalisées sur une colonne HighLoad 75 16/600 avec récupération de fraction de 1.5 mL et un débit de 1 mL.min⁻¹. La colonne a été équilibrée avec 1.5 CV de tampon (celui-ci dépend des applications souhaités). 5 mL des fractions les plus concentrées ont été injectés.

4) Clivage à la protéase « Tobacco Etch Virus endopeptidase » (TEV)

La protéine chimérique **A3_bGFPD** possède un His-Tag suivi d'un site de clivage TEV en N-Ter. Afin d'éliminer cette étiquette histidine, le clivage d'**A3_bGFPD** par la protéase TEV a été réalisé. Une concentration de l'ordre de 5-10 μM de TEV pour une concentration en protéine comprise entre 180 et 400 μM . Du tampon 1 mM DTT, 0.5 mM EDTA a été ajouté au milieu réactionnel, celui-ci permet un bon fonctionnement de la TEV protéase. La réaction s'est déroulée à 4°C O/N. La protéine a alors été récupérée par une étape de contre-purification sur colonne Ni-NTA. En effet, la TEV protéase, qui possède un His-Tag, ainsi que les His-Tag clivés ont donc été retenus sur la colonne. Les protéines ne possédant pas de tag hexa-histidine ont théoriquement été récupérées dans la fraction non-retenue.

5) Dialyse des protéines

Les protéines ont été dialysées dans des MIDI ou MINI GeBaFlex-tubes (Gene Bio Application Ltd) avec un cut-off de 6 à 8 kDa. La dialyse a été effectuée en deux étapes à 4°C dans un volume 500 à 1000 fois supérieur à celui des échantillons dialysés en agitation permanente. Une première dialyse d'une durée d'une heure a été réalisée puis une seconde dialyse O/N à 4°C. La dialyse est nécessaire avant les expériences de DSC et d'ITC afin d'avoir l'ensemble des échantillons et des références dans le même tampon.

6) Gel SDS-PAGE

Les gels SDS-PAGE permettant d'analyser les protéines purifiées ont été réalisés à 12 ou 15 % en fonction de la taille des protéines étudiées. Les protéines ont été mélangées à du Laemmli 2X et chauffées 5 min à 95°C. La migration a été faite à ampérage constant de 35-40 A. Les gels ont été colorés avec de l'instant Blue (Expedeon) 30 minutes puis lavés avec de l'eau distillée.

IV. Biophysique

NB : Les expériences de DSC, de SEC-MALS, d'ITC, de SPR et d'UCA ont été effectuées en collaboration avec la Plateforme d'Interaction des Macromolécule (PIM).

NB2 : Le SAXS et la cristallographie ont été effectués en collaboration avec l'équipe Fonction et Architectures des Assemblage Macromoléculaire (FAAM) et la plateforme de cristallisation des protéines qui y est rattachée.

1) Differential Scanning Calorimetry (DSC)

Les mesures ont été effectuées sur un microcalorimètre VP-DSC (MicroCal). La ligne de base a été effectuée avec du tampon NaP 50 mM, NaCl 150 mM, pH 7 puis la protéine a été chargée dans la cuve de mesure à une concentration de 0,5 mg.mL⁻¹. Dans le cas d'étude de complexe, la protéine ayant la masse molaire la plus importante était à 0,5 mg.mL⁻¹ et la seconde protéine était en concentration molaire équivalente. La mesure a été effectuée avec les paramètres suivant : 1 scan de 20 à 120°C, 60 °C/h. La ligne de base a alors été soustraite du signal et les données corrigées ont été traitées en utilisant une spline cubique comme ligne de base pour la transition. Les paramètres thermodynamique ΔH_{cal} et ΔH_{vH} ont été déterminés en ajustant l'équation suivante aux données :

$$\Delta C_p(T) = [K_d(T) \Delta H_{cal} \Delta H_{vH}] / [(1 + K_d(T))^2 RT^2]$$

Où K_d est la constante d'équilibre d'une transition à deux états, ΔH_{vH} l'enthalpie calculée sur la base d'un processus à deux états purs et ΔH_{cal} l'enthalpie mesurée.

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

La chromatographie d'exclusion a été effectuée par High Performance Liquid Chromatography (Shimadzu) sur une colonne Superdex 200 10/300GL increase. Les concentrations des échantillons déposés, en complexe ou non, étaient de 45 μM pour **A3-bGFPD** et **EGFP** et 90 μM pour Rep2. L'HPLC était couplée à un détecteur de diffusion de lumière à 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, toujours selon ce principe, la variation d'intensité de lumière diffusée en fonction de l'angle de mesure est directement reliée au rayon hydrodynamique des molécules. Le couplage de la chromatographie d'exclusion stérique à la diffusion de lumière à multiples angles et à la réfractométrie (SEC-MALS) permet donc d'obtenir des informations précises sur la masse molaire des protéines, la formation de complexes protéique mais également, le cas échéant, la stœchiométrie de ceux-ci. L'obtention du rayon hydrodynamique des objets par ces méthodes pourra permettre de déterminer la forme (ouverte ou fermée) d'une protéine chimérique. Enfin, il est possible grâce à l'ensemble des paramètres mesurés ($\text{DO}_{280\text{nm}}$, MALS et dRI) de déterminer la proportion et la masse moléculaire de chaque molécule constituant un complexe. La concentration protéique totale est tout d'abord obtenue grâce au signal de réfractométrie. Le signal d'absorbance est ensuite déconvolué en utilisant les coefficients d'absorption molaire de chaque protéine afin d'obtenir une concentration protéique identique à celle de réfractométrie. La proportion de chaque protéine peut alors être déterminée. Par déconvolution des signaux dRI et MALS les masses moléculaires des constituants d'un complexe peuvent ensuite être obtenues.

3) Isothermal Titration Calorimetry (ITC)

Les mesures ont été effectuées sur un ITC200 (Microcal) à 20°C avec une agitation de 800 rpm. Les concentrations utilisées étaient (sauf indication contraire dans les résultats) de 25 µM dans la cuve et 300 µM dans la seringue. Les données obtenues ont été intégrées pour générer les courbes où l'aire de chaque pic d'injection est présentée en fonction du ratio d'échantillon injecté par rapport au contenu de la cuve. Les données ont ensuite été analysées avec le logiciel MicroCal Origin® selon un modèle à un site de fixation. Les changements d'énergie et d'entropie lors de la fixation du ligand ont été calculés à partir des paramètres à l'équilibre en utilisant l'équation : $-RT\ln(K_A) = \Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$ où R est la constante universelle des gaz parfaits (1,9872 cal.mol⁻¹.K⁻¹), T la température en degré Kelvin, K_A est la constante d'association, ΔG° le changement d'énergie libre standard de Gibbs, ΔH° le changement d'enthalpie standard et ΔS° le changement d'entropie standard. La constante de dissociation de chaque interaction est exprimée comme $\frac{1}{K_A} = \frac{k_{off}}{k_{on}} = \frac{[A][B]}{[AB]} = K_D$ (en M⁻¹) où k_{off} et k_{on} sont les constantes cinétiques de dissociation et d'association ; [A], [B] les concentrations en protéines A et B ; [AB] la concentration en complexe AB.

NB : Les données de DSC, d'ITC ont été traitées avec le logiciel MicroCal Origin®.

4) Surface Plasmon Resonance (SPR)

Les mesures de SPR ont été effectuées sur un ProteOn (Biorad) à l'aide d'une puce HTG (Tris-Ni-NTA). Cette puce fixe les ligands ayant une étiquette poly-Histidine grâce à des groupements Tris-Ni-NTA qui garantissent une fixation efficace. En effet, les puces Ni-NTA classique ne présentent qu'un ion nickel par chaîne NTA pouvant établir deux liaisons de coordination, soit deux histidines parmi les six de l'étiquette histidine des protéines. Le tris-Ni-NTA permet d'avoir trois ions nickel à proximité de l'étiquette hexa-histidine qui peuvent potentiellement fixer les six histidines (deux par Nickel) du tag avec une affinité bien supérieure à celle du Ni-NTA simple. Les ligands sont **bA3-2**, **eGFP A3_A3** et **A3_bGFPD** (toutes avec His-tag), les analytes **A3**, **bGFPD**, **A3_A3** et **A3_bGFPD** (sans His-Tag). Le principe de cette technique repose sur le fait que les rayons incidents arrivant sur la surface couverte de ligand sont réfractés avec un certain angle et que la fixation d'un partenaire (analyte) sur ce ligand induit un changement d'angle de réfraction des rayons incidents. La mesure des variations d'angle de réfraction pour différentes concentrations d'analytes en solution, permettent d'estimer les constantes cinétiques d'interaction (k_{on} et k_{off}) et la constante d'association à l'équilibre (K_A) 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 :

$$R_L = \frac{R_{max}}{n} * \frac{M_{ligand}}{M_{Analyte}}$$

Où n est la stœchiométrie de la réaction, R_{max} (=200) l'unité de résonance maximale souhaitée, M_{Ligand} et $M_{Analyte}$ les masses molaires du ligand et de l'analyte respectivement.

Trois pistes étaient utilisées sur la puce HTG : la première avec une quantité de ligand permettant l'obtention d'un signal équivalent à 2 RL ; la seconde avec un signal à 1 RL ; la dernière sans ligand fixé servant de témoin. Une gamme de concentration en analyte entourant le K_D obtenu en ITC a été effectuée pour chaque analyte (sauf indication contraire dans les résultats). Pour la détermination des paramètres cinétiques, 6 concentrations différentes de **A3** ou **A3-bGFPD** 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.

5) Ultracentrifugation analytique (UCA)

Toutes les expériences de vitesse de sédimentation ont été effectuées à 20°C sur une ultracentrifugeuse XLI (Beckman Coulter, Palo Alto, USA) avec un rotor An-50Ti équipé d'un système de détection de fluorescence (AVIV Biomedical, Lakewood, NJ). Chaque canal de fluorescence à un chemin de 12mm (Epon charcoal-filled centerpieces). La protéine **A3** a été marquée avec un Monolith NT.115 (*Protein Labeling Kit BLUE NHS*, Nanotemper Technologies). Les échantillons ont été préparés par dilution des concentrations stock dans du PBS, 0.1 mg.mL⁻¹ BSA. Pour les concentrations de 1 à 60 nM, seule la protéine marquée a été utilisée. Pour les concentrations en protéine supérieures à 60 nM, 60 nM de protéine marquée ont été mélangés avec de la protéine non marquée pour obtenir la concentration désirée. 400 µL de chaque échantillon de la protéine **A3** ont été centrifugés à 42 000 rpm (128297g). Les profils de sédimentations ont été collectés toutes les cinq minutes et analysés comme décrit. (Chaturvedi, S.K., *et al.*, 2017; Schuck, P., 2003; Zhao, H., *et al.*, 2013)

6) Small Angle X-rays Scattering (SAXS)

Le principe du SAXS repose sur l'interaction élastique des photons avec les nuages électroniques. La diffusion des photons à travers un échantillon apportant des informations sur les variations des densités électroniques, il sera ainsi possible de remonter jusqu'à la forme d'une protéine, voire sa structure. Le SAXS a été couplé à une HPLC sur colonne Agilent Bio SEC-3. La protéine **A3-bGFPD** sans His-Tag (digéré à la TEV) est passée sur la ligne BM29 dédiée au BioSAXS de l'European Synchrotron Radiation Facility (ESRF). L'acquisition a eu lieu tout au long de l'élution à une température de 293 K dans du tampon Tris/HCl 20 mM, NaCl 200 mM, pH 7,5. La distance échantillon - récepteur est de 2.867m et couvre une gamme de vecteur de diffusion de 0.005 Å⁻¹ à 0.5 Å⁻¹. $q = 4\pi \sin\theta / \lambda$ où 2θ est l'angle de diffusion et λ (0.992 Å) est la longueur d'onde des rayons X. Le détecteur utilisé est un Pilatus. Les intensités de diffusion ont été normalisées en soustrayant l'intensité du faisceau arrivant et le bruit de fond en utilisant le logiciel IspyB (Delagenière, S., *et al.*, 2011). Le module US-SOMO HPLC-SAXS a été utilisé pour déterminer l'évolution du rayon de giration R_g et la masse moléculaire au cours de l'élution (Brookes, E., *et al.*, 2016). L'ensemble des représentations 3D des molécules a été obtenu avec le logiciel PyMol (Schrödinger).

7) Cristallographie

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 Classic, MB Class I, PEG I, JCSG+). Les deux bidomaines ont cristallisé dans diverses conditions. Pour **A3_A3**, la condition de cristallisation optimale est 0.1 M Sodium-citrate pH 5.6, 2.4 M AmSO₄. Pour **A3_bGFPD**, 1.6 M Tri-Sodium citrate pH6.5. Les cristaux ont été congelés dans l'azote liquide après trempage dans la solution de cristallisation complétement avec 30 % de glycérol pour **A3_A3** et 3.8M malonate pour **A3_bGFPD**. Les données de diffraction ont été collectées à 100K sur la ligne de lumière ID29 de l'ESRF en utilisant un détecteur PILATUS 6M. Les images ont par la suite été analysées avec le programme XDS (Kabsch, W., 2010). Les structures ont été résolues par remplacement moléculaire en utilisant PHASER 7 (McCoy, A.J., 2007) implémenté dans la suite de logiciel ccp4 (Winn, M.D., et al., 2011). L' α Rep **A3**, code pdb : 3LTJ, a été utilisée comme modèle. La carte de densité électronique expérimentale a été traitée en utilisant le programme DM (Cowtan, K., 1999). Les modèles initiaux ont ensuite été complétés manuellement avec le logiciel COOT (Emsley, P., et al., 2010). L'affinement des structures a été effectué à l'aide de REFMAC (Murshudov, G.N., et al., 2011).

La structure cristallographique de **A3_A3** a été obtenue avec une résolution de 1.94 Å avec un R et un R_{Free} de 17.48 et 22.21 %. La structure cristallographique de **A3_bGFPD** a été obtenue avec une résolution de 2.79 Å avec un R et un R_{Free} de 20.56 et 26.98 %.

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

V. Expérience de FRET

1) Réduction des cystéines

Une cystéine unique a été ajoutée par mutagénèse dirigée en C-Terminal des protéines **A3_A3** et **A3_bGFPD**. La réduction de ces cystéines uniques a été effectuée par ajout d'un excès de DTT (10 mM) pendant 45 min à température ambiante dans du tampon de marquage (Tris-HCl 100 mM, 150 mM NaCl, pH 7.4). L'excès de DTT a été éliminé grâce à deux colonnes de dessalage Zebaspin® (Thermo Fisher Scientific) 7 KDa successives.

2) Couplage des ATTO-dye

Une solution d'ATTO 488 maleimide (Sigma) à 10 mM a été préparée dans du DMSO anhydre et mélangée aux protéines **A3_A3** et **A3_bGFPD** fraîchement réduites dans du PBS (10 équivalents molaires d'ATTO488 par protéine). Les mélanges ont alors été incubés à 4°C dans le noir pendant toute la nuit. L'excès d'ATTO 488 a été éliminé par à deux colonnes de dessalage Zebaspin® 7KDa successives. La concentration d'ATTO 488 a été déterminée en utilisant l'absorbance mesurée à 501 nm en utilisant le coefficient d'absorption molaire fourni par le fabricant, 90 000 M⁻¹.cm⁻¹. (<https://www.atto-tec.com>)

Les protéines couplées à l'ATTO 488 ont par la suite été couplées au NTA-ATTO 647N *via* l'étiquette histidine. Dix équivalents molaires ont été ajoutés à des solutions de 0.1µM de bidomaine_Atto-488 et incubés 5 minutes dans le noir avant mesure.

3) Expérience préliminaire de détection des ligands

L'expérience de détection de liaison par FRET avec les ATTO-Dye a été effectuée sur un spectrophotomètre SAFAS (Monaco) flx-Xenius V2. Les spectres d'émission ont été mesurés de 500 à 800 nm en excitant les échantillons à 488nm. Les spectres ont été enregistrés avec un pas de 1 nm, une fente de 10 nm pour chaque canal, un temps d'intégration de 0.1s et un voltage du photomultiplicateur à 500 V. Des aliquotes de **bA3-2**, de BFP et d'albumine (sans His-Tag) ont été ajoutés successivement à une solution de 0.1 µM de protéine marquée dans des cuves en quartz. Les solutions ont alors été incubées deux minutes dans le noir avant la mesure du spectre.

4) Couplage du cryptate de Terbium

Le Lumi4-maleimide (*Lumiphore Inc*) a été resuspendu à 8mM dans du DMF anhydre puis mélangé en excès avec les protéines fraîchement réduites dans le tampon de marquage. Les mélanges ont alors été incubés à température ambiante sur un *intellimixer*[®] à 25 rpm pendant 3 h puis à 4°C sur la nuit. L'excès de Lumi4-maléimide n'ayant pas réagi a été éliminé grâce à deux colonnes de dessalage Zebaspin[®] 7 KDa successives. Le complexe protéine-Lumi4 a ensuite été chargé en Tb³⁺ avec une concentration en TbCl₃ en excès pendant 1h à 4°C. La concentration de Tb a été déterminée par mesure de l'absorbance à 340nm en utilisant le coefficient d'absorption molaire du fabricant à savoir 26 000 M⁻¹.cm⁻¹. La concentration en protéine a été déterminée par mesure de l'absorbance à 280 nm.

5) Couplage des Quantum Dots 625

Les protéines marquées avec le L4-Tb ont été couplées aux QDs *via* leur étiquette poly-histidine en mélangeant simplement au ratio désiré dans du tampon de marquage. Aucune purification n'est nécessaire puisque le complexe histidine / QD est de haute affinité et permet d'obtenir un assemblage quasi-total.

6) Préparation des expériences de FRET

Les protéines marquées avec le L4-Tb et les QD ont été diluées dans 50µL de tampon de marquage avec des concentrations finales de 12 nM et de 0 à 4 nM respectivement. Ils ont ainsi été mélangés et incubés 4 h à température ambiante. 50 µL de ligands ont par la suite été ajoutés aux concentrations désirées dans les 100µL de mélange protéine-L4-Tb + QD et incubés 15min à température ambiante avant mesure. L'ensemble des manipulations a été effectué en tampon Nap 50 mM, NaCl 150 mM, BSA 0.1%, pH 7.4. Les plaques utilisées sont des plaques opaques 96 puits 150 µL.

7) Détermination des Limites de détection

La mesure de la luminescence a été effectuée entre 0.1 et 0.9 ms sur un *Kryptor compact plus fluorescence plate reader* (Cézanne / Thermo Fisher Scientific) en effectuant 500 mesures successives avec un temps d'intégration de 2 μ s. L'excitation a été effectuée par un laser à azote (337.1 nm, 20 Hz, 100 pulses). Les deux canaux, donneur et accepteur ont pu être suivis en utilisant un filtre à 494 \pm 12 nm (Semrock) pour le Terbium et un à 640 \pm 10 nm (Semrock) pour les QD625. Chaque expérience de FRET (Terbium/QD625) a été effectuée dans une plaque opaque noire de 96 puits d'un volume optimal de 150 μ L. Chaque échantillon contenait différentes concentrations de ligand (**bA3-2** ou BFP) et a été préparé trois fois. Les blancs ont été préparés dix fois. La fluorescence de chaque échantillon a été mesurée en triplicata. Les barres d'erreur non visibles sur les graphiques de limite de détection sont plus petites que les points du graphique.

8) Détermination des distances

Les spectres de décroissance ont été obtenus sur un spectrofluorimètre *DreamReader* (*Edinburgh Instrument*) en effectuant 1000 mesures successives par puits avec une intégration de 2 μ s. L'excitation (337.1 nm, 20 Hz) a été effectuée par un laser à azote (MNL 300, LTB Lasertechnik Berlin). Les deux canaux, donneur et accepteur ont pu être suivis en utilisant un filtre à 494 \pm 20 nm (Semrock) pour le Terbium et un à 640 \pm 20 nm (Semrock) pour les QD625. Chaque expérience de FRET (Terbium/QD625) a été effectuée dans une plaque opaque noire de 96 puits d'un volume optimal de 150 μ L. Les données ont été fittées avec le logiciel FAST (*Edinburgh Instrument*) afin d'en extraire les différents temps de décroissance ainsi que leurs amplitudes.

VI. Sélection par Phage Display

1) Production des phages

Les phages ont été produits par Margarida Gomes en infectant des bactéries XL1 Blue MRF¹ contenant la bibliothèque de protéines MG à l'aide d'un phage *helper* M13KO1 (Soltes, G., *et al.*, 2003). Afin d'avoir une représentation totale de la bibliothèque, la quantité de bactéries infectées est au moins dix fois supérieure à la diversité de la banque. 10^{10} bactéries ont donc étéensemencées à un DO_{600} de 0.1 dans 100 mL de milieu AT à 37°C à 225 RPM. L'infection a été effectuée en phase exponentielle de croissance bactérienne ($DO_{600} = 0.6-0.7$) avec un volume de phage *helper* M13KO1 permettant d'avoir au moins 20 fois plus de phages. Le milieu est alors incubé pendant 30 min à 37°C sans agitation puis 30 min à 37°C à 100 RPM. Après infection, les bactéries sont récupérées par centrifugation à 5000 g 20min à 4°C (élimination des phages *helper* qui restent dans le surnageant) puis mises en culture dans 100 mL de milieu 2xYT AK pendant 4h à 30°C à 200 RPM. Seules les bactéries ayant d'une part un plasmide de la banque MG (résistance à l'ampicilline) ET le vecteur du phage *helper* (résistance à la kanamycine) pourront croître et produire des phages. Afin de récupérer les phages produits, la culture est centrifugée à 10 000 g 1h à 4°C, le surnageant contenant les phages est alors récupéré.

2) Préparation des plaques ELISA

La NanoKaz contre laquelle a été effectuée la sélection correspond à la protéine NanoKaz-HRN3s (voir partie C-I-d). Elle a été préalablement *biotinylée via* l'avitag (GLNDIFAQKIEWHE) lors de la production afin d'être fixée de manière spécifique sur la streptavidine. La sélection a été effectuée en plaque ELISA 96 puits Maxisorp (Nunc). 100 μ L de streptavidine à 20 μ g/mL sont directement incubés sur 8 puits de deux plaques différentes appelées **P1 (Spécifique)** et **P2 (non spécifique)** pendant 4 h à 20°C à 300 RPM. Les deux plaques sont alors lavées trois fois avec 300 μ L de TBS Tween 0.1% puis bloquées avec 300 μ L de TBS Tween 0.1 % BSA 4 % sur la nuit à 4°C à 300 RPM. Les deux plaques sont à nouveau lavées trois fois avec 300 μ L de TBS Tween 0.1 %.

3) Tours de sélection

a) Sélection

100 μ L de nanoKaz-HRN3s à 2 μ M dans du TBS sont incubés sur **P1** pendant 1 h à 20°C à 300 RPM. En parallèle, 130 μ L de phages (à environ 10^{12} phages/mL pour le tour 1) contenant la banque MG sont déposés dans la plaque **P2**, également 1 h à 20°C à 300 RPM. Cette étape va permettre d'éliminer les phages non spécifiques et se fixant sur la plaque, la streptavidine ou la BSA. La plaque **P1** est alors lavée trois fois avec 300 μ L de TBS Tween 0.1%. 100 μ L de phages provenant de la plaque **P2** sont ensuite transférés dans la plaque **P1** et incubés 2 h à 20°C 300 RPM. La plaque **P1** est lavée dix fois avec 300 μ L de TBS Tween 0.1 % et dix fois avec du TBS. Après les lavages, l'élution des phages spécifiques est effectuée par ajout de 100 μ L de TEV à 2 μ M dans du tampon DTT 20 mM, EDTA 10 mM pendant la nuit à 4°C 300 RPM. Les éluions sont récupérées et une seconde élution TEV a été effectuée pendant 4 h à 20°C 300 RPM. La seconde élution n'ayant pas permis d'éluer plus de clones, elle n'a pas été effectuée aux tours 2 et 3.

b) Infection

À l'issue de chaque tour de sélection, 1.5 mL de bactéries XL1 Blue MRF' à une $DO_{600} = 0.6-0.8$ ont été infectés par les phages élués et incubés 30 min à 37°C à 100 RPM puis étalés sur milieu 2xYT ATG agar (boîte 25 cm x 25 cm) et incubés à 37°C sur la nuit. Les bactéries ayant poussé sur boîte sont alors récupérées par ajout de 5-10 mL de milieu de culture 2xYT Glu 1% + Glycérol 15 %. Une partie est stockée à -80°C, l'autre est utilisée pour produire les phages du tour suivant.

c) Production de phages

25mL de culture XL1 Blue MRF' en milieu ATG à $DO_{600} = 0.15$ sont incubés à 37°C 200 RPM. Quand la DO_{600} atteint environ 0.8, les phages M13KO1 sont ajoutés à raison de 10 phages / bactérie puis incubés 30 min à 37°C sans agitation puis 30 min à 37°C à 100 RPM. Après infection, les bactéries sont récupérées par centrifugation à 5000 g 20min à 4°C puis mises en culture dans 100mL de milieu 2xYT AK pendant 4h à 30°C à 200 RPM. Afin de récupérer les phages produits, la culture est centrifugée à 10 000 g 1h à 4°C, le surnageant contenant les phages est alors récupéré et est utilisé pour le tour suivant.

4) ELISA Clonal

Afin de vérifier la spécificité des phages élués au cours des sélections, un crible par phage ELISA clonal a été effectué. Pour cela, 96 clones provenant d'infection effectuée à différents niveaux des tours de sélection ont été incubés dans une plaque 96 puits stérile dans 150µL de milieu 2xYT ATG 37°C sur la nuit à 300 RPM. 15 µL de chaque puits sert alors àensemencer une nouvelle plaque 96 puits contenant 150 µL de milieu 2xYT ATG qui est incubé 2h à 37°C 300 RPM. Chaque culture est alors infectée avec 12 µL de phage *helper* M13KO1 (Stock à 10^{12} - 10^{13} phages/mL généralement) et incubée 30 min à 37°C sans agitation puis 30 min à 37°C 100 RPM. 100 µL de bactéries infectées sont alors transférées dans une plaque 96 puits deepwell stérile dans 1.5mL de milieu 2xYT AK et incubées sur la nuit à 30°C 300 RPM. La plaque est alors centrifugée 1 h à 2000 g. 100 µL de surnageant contenant les phages sont alors transférés dans une plaque ELISA préparée comme pour les sélections (sur l'ensemble de la plaque) et incubés en présence (spécifique) ou en absence (non spécifique) de la nanoKaz 1h30 à 20°C 300 RPM. La plaque est alors lavée trois fois avec du TBS Tween 0.1 % puis incubée 1 h à 4°C 300 RPM avec 100 µL d'anticorps anti-M13 couplé à la HRP (*horseradish peroxidase*) dilués aux 1/5000^{ème} (*GE Life Sciences*). Trois lavages avec du TBS Tween 0.1 % sont alors effectués et la révélation est réalisée à l'aide de 100 µL de substrat POD soluble (*Roche*). La réaction est alors arrêtée avec 100 µL de HCL 1 M quand la coloration est suffisante. L'absorbance est alors mesurée à 405 nm sur un TECAN.

VII. Expériences impliquant le LBT

Les expériences de détection de liaison par utilisation d'un LBT modifié ont été effectuées sur un spectrophotomètre SAFAS (Monaco) flx-Xenius V2. Les spectres d'émission ont été mesurés de 450 à 650 nm en excitant les échantillons à 280nm. Les spectres ont été enregistrés avec un pas de 1 nm, une fente de 10 nm pour chaque canal, un temps d'intégration de 0.1s et un voltage du photomultiplicateur à 500 V. Des aliquotes de **bA3-2** et des protéines ayant servi au contrôle négatif (BSA ou bArpin) ont été ajoutés successivement pour obtenir les concentrations souhaitées à une solution de 250 µL à 1 µM de LBTs **A3_A3_WW** ou **LBT_A3_A3** pré-incubée avec 10 µM de Tb³⁺ dans des cuves en quartz. Les solutions ont alors été incubées deux minutes avant la mesure du spectre. L'ensemble des manipulations a été effectué en tampon Hepes 20 mM, NaCl 150 mM, pH 7.4.

VIII. Test des chimères nanoKaz

Les mesures de bioluminescence des chimKaz ont été effectuées en tampon Tris HCl 20 mM, NaCl 200 mM, MgCl₂ 5mM, BSA 0,1%, pH 8.0 Les plaques utilisées sont des plaques opaques 96 puits 150 µL. Des solutions de NanoGlo à 0.45%, de NanoKaz ou des ChimKaz à 30pM et de BFP ou GALM à des concentrations allant de 15 à 9000nM ont été préparées. 50 µL de NanoKaz ou de ChimKaz sont d'abord préparé et incubé au minimum 15 min. 50 µL de NanoGlo est ensuite ajouté juste avant mesure à l'aide d'une pipette multicanaux. Ainsi, huit mesures sont effectuées en parallèle. L'intégralité de la lumière émise a été mesurée par un Infinite F200 Pro (TECAN) avec une intégration de 100 ms par mesure.

IX. Bibliographie

- Brookes, E., P. Vachette, M. Rocco, and J. Perez. 2016. 'US-SOMO HPLC-SAXS module: dealing with capillary fouling and extraction of pure component patterns from poorly resolved SEC-SAXS data', *J Appl Crystallogr*, 49: 1827-41.
- Chaturvedi, Sumit K., Jia Ma, Huaying Zhao, and Peter Schuck. 2017. 'Use of fluorescence-detected sedimentation velocity to study high-affinity protein interactions', *Nature Protocols*, 12: 1777-91.
- Cowtan, K. 1999. 'Error estimation and bias correction in phase-improvement calculations', *Acta Crystallographica. Section D, Biological Crystallography*, 55: 1555-67.
- Delagenière, Solange, Patrice Brechereau, Ludovic Launer, Alun W. Ashton, Ricardo Leal, Stéphanie Veyrier, José Gbadinho, Elspeth J. Gordon, Samuel D. Jones, Karl Erik Levik, Seán M. McSweeney, Stéphanie Monaco, Max Nanao, Darren Spruce, Olof Svensson, Martin A. Walsh, and Gordon A. Leonard. 2011. 'ISPyB: an information management system for synchrotron macromolecular crystallography', *Bioinformatics (Oxford, England)*, 27: 3186-92.
- Emsley, P., B. Lohkamp, W. G. Scott, and K. Cowtan. 2010. 'Features and development of Coot', *Acta Crystallographica. Section D, Biological Crystallography*, 66: 486-501.
- Kabsch, Wolfgang. 2010. 'XDS', *Acta Crystallographica. Section D, Biological Crystallography*, 66: 125-32.
- McCoy, Airlie J. 2007. 'Solving structures of protein complexes by molecular replacement with Phaser', *Acta Crystallographica. Section D, Biological Crystallography*, 63: 32-41.
- Murshudov, Garib N., Pavol Skubák, Andrey A. Lebedev, Navraj S. Pannu, Roberto A. Steiner, Robert A. Nicholls, Martyn D. Winn, Fei Long, and Alexei A. Vagin. 2011. 'REFMAC5 for the refinement of macromolecular crystal structures', *Acta Crystallographica. Section D, Biological Crystallography*, 67: 355-67.
- Schuck, Peter. 2003. 'On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation', *Analytical Biochemistry*, 320: 104-24.
- Soltes, Glenn, Heather Barker, Kristine Marmai, Elaine Pun, Amy Yuen, and Erik J. Wiersma. 2003. 'A new helper phage and phagemid vector system improves viral display of antibody Fab fragments and avoids propagation of insert-less virions', *Journal of Immunological Methods*, 274: 233-44.

- Winn, Martyn D., Charles C. Ballard, Kevin D. Cowtan, Eleanor J. Dodson, Paul Emsley, Phil R. Evans, Ronan M. Keegan, Eugene B. Krissinel, Andrew G. W. Leslie, Airlie McCoy, Stuart J. McNicholas, Garib N. Murshudov, Navraj S. Pannu, Elizabeth A. Potterton, Harold R. Powell, Randy J. Read, Alexei Vagin, and Keith S. Wilson. 2011. 'Overview of the CCP4 suite and current developments', *Acta Crystallographica. Section D, Biological Crystallography*, 67: 235-42.
- Zhao, Huaying, Suwendu Lomash, Carla Glasser, Mark L. Mayer, and Peter Schuck. 2013. 'Analysis of high affinity self-association by fluorescence optical sedimentation velocity analytical ultracentrifugation of labeled proteins: opportunities and limitations', *PLoS One*, 8: e83439.

PARTIE D : Conclusion et Perspectives

Tables des matières

I.	COMPORTEMENT DES PROTÉINES BIDOMAINES.....	216
A.	<i>Bilan</i>	216
B.	<i>Perspectives</i>	219
1)	Enzyme activable.....	219
2)	Commutateurs modulables.....	221
II.	BIOSENSEURS	223
A.	<i>Basés sur l'utilisation du FRET</i>	223
1)	Conclusion	223
2)	Optimisation des bidomaines L4Tb/QD625	224
B.	<i>Biosenseurs basés sur l'utilisation du LBT</i>	225
1)	Conclusion	225
2)	Optimisation de la dockérine	225
C.	<i>Biosenseurs basés sur la nanoKaz</i>	226
1)	Bilan.....	226
2)	Optimisation du biosenseur NanoKaz	226
D.	<i>Conclusion générale sur les biosenseurs</i>	229
III.	BIBLIOGRAPHIE	230

I. Comportement des protéines bidomaines

A. Bilan

L'étude sur le comportement des bidomaines **A3_A3** et **A3_bGFPD** a permis d'apporter une meilleure compréhension sur les propriétés de telles protéines.

Nous avons ainsi pu montrer que l'ajout d'un *linker* entre deux α Reps permettait de former des bidomaines fermés. Cet effet lié aux interactions intramoléculaires a lieu même lorsque les deux α Reps reliées ont des affinités l'une pour l'autre modérées, avec des constantes de dissociation de l'ordre du μ M.

L'ouverture de ces bidomaines reste possible par les ligands respectifs des domaines et les affinités alors mesurées sont plus faibles d'un facteur dépendant de l'affinité initiale des domaines non liés. Ainsi, plus les domaines ont une affinité élevée, plus l'affinité mesurée pour les ligands diminuera. Cette conformation fermée des bidomaines donne par ailleurs lieu à un phénomène de coopérativité de fixation des ligands. Pour **A3_bGFPD**, la fixation de l'**eGFP** va faciliter la fixation de **bA3-2**. Les ligands étant différents, il s'agit d'hétéro-coopérativité. Pour **A3_A3**, ce phénomène se voit en comparant les constantes de dissociation de **bA3-2** déterminées en fluorescence (avec un $K_D \approx 600$ nM) avec celles obtenues en ITC ($K_D \approx 36$ nM). En effet, le K_D mesuré en fluorescence correspond à la fixation du premier **bA3-2** qui bloque la fermeture du bidomaine. Un second **bA3-2** peut se fixer mais sa contribution au signal de fluorescence est nulle comme en témoigne le fait que pour **A3_bGFPD**, un seul **bA3-2** entraîne un retour au signal basal. Ainsi, la variation de fluorescence et donc la détermination du K_D de **bA3-2** pour **A3_A3** traduisent uniquement de la première fixation. L'ITC mesure les échanges de chaleur, ainsi la détermination du K_D par ITC traduira de la fixation de deux **bA3-2**. Le K_D mesuré représente donc une résultante des échanges de chaleur (en kcal/mol) correspondant à la fixation des deux ligands, donnant alors un K_D représentant la moyenne des chaleurs échangées lors des deux fixations. Les ligands étant identiques pour **A3_A3**, il s'agit ici d'homo-coopérativité.

Sachant que l'interaction est fortuite et de faible affinité entre **A3** et **bGFPD**, elle est probablement due en grande partie à l'interaction N-Cap (**A3**), / C-Cap (**bGFPD**). Cette capacité à obtenir des bidomains fermés pourrait ainsi être généralisable à tout bidomaine possédant un domaine **A3**. Il semble également, d'après l'étude de plusieurs protéines de la bibliothèque d' α Reps, qu'une grande proportion d' α Reps puisse dimériser, facilitant alors la création de bidomains fermés.

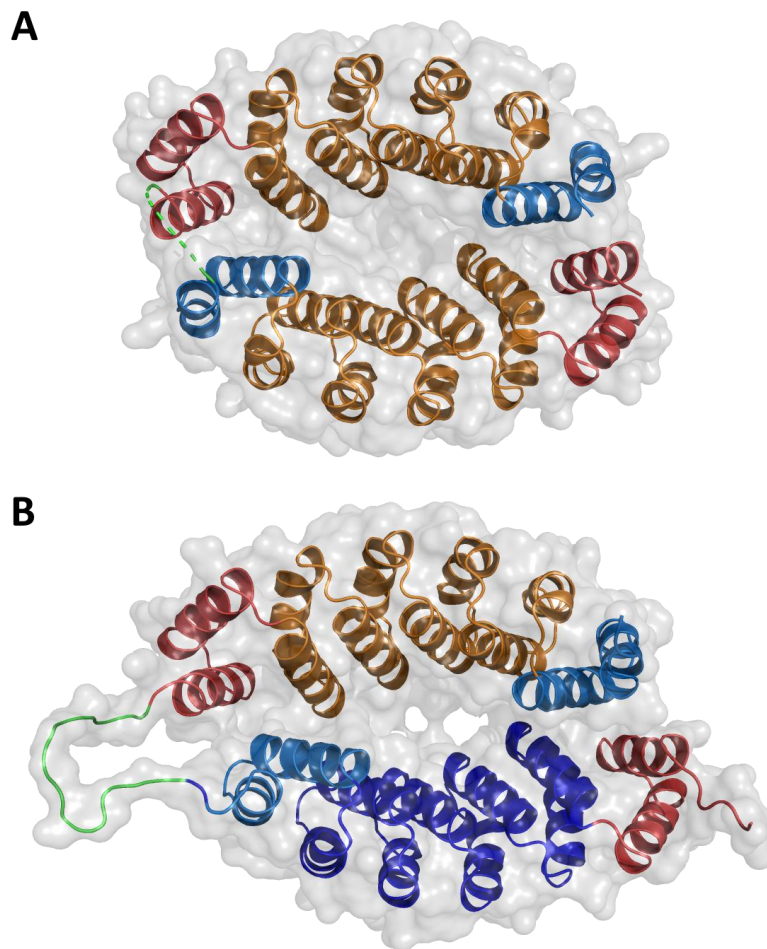


Figure 1 : Représentation des deux bidomains en cartoon. A : A3_A3 et B : A3_bGFPD. Les N-Caps de chaque domaines sont visibles en bleu marine, les C-Caps sont en rouge.

D'autre part nous avons pu voir que les bidomaines étaient capables de fixer simultanément deux ligands (identiques ou non). Cette propriété peut également être utilisée pour effectuer le rapprochement de deux cellules (par exemple, une cellule tumorale et un lymphocyte T) ou deux protéines d'une même voie enzymatique. Un financement de l'ANR a par ailleurs été attribué en partie à l'équipe MIP afin de concevoir une plateforme de d'assemblage pour améliorer *in vivo* les voies de biosynthèses impliquant les *cyclopeptides synthases* et les *peptides synthases* non ribosomales. L'idée de ce projet est de fusionner les enzymes à des α Reps qui vont d'une part servir de partenaire d'expression et d'autre part vont se fixer sur une « *landing platform* » constituée des domaines ciblés par les α Reps fusionnés. (Figure 2)

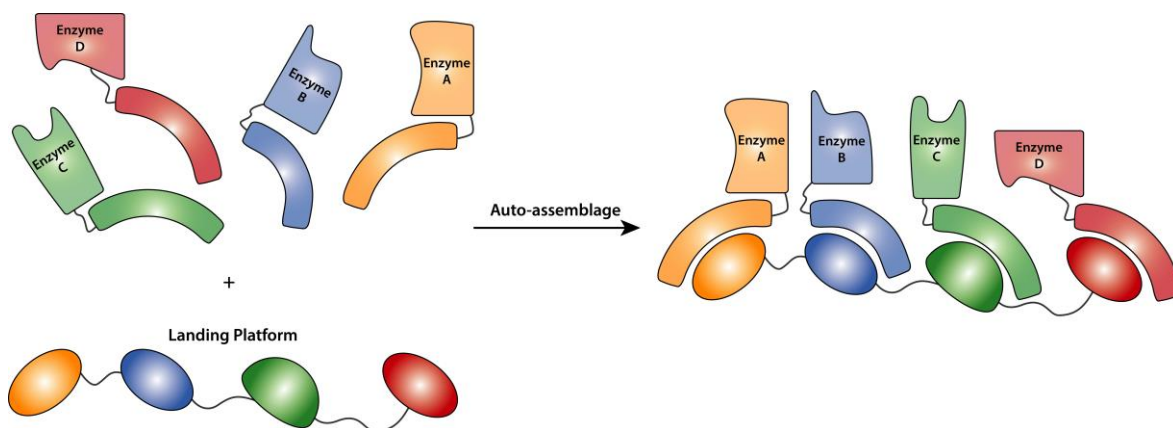


Figure 2 : Projet de « *landing platform* » capable d'entraîner l'auto-assemblage d'enzymes fusionnées à des α Reps.

Enfin, un phénomène d'avidité a pu être observé quand les ligands d'un bidomaine sont fixés sur une même surface. Ce phénomène peut être exploité pour effectuer du tri cellulaire, en couplant un fluorophore à un bidomaine reconnaissant un récepteur d'une cellule par exemple, le phénomène d'avidité garantirait une fixation quasi covalente sur la cellule.

L'ensemble de ces comportements serait généralisable à d'autres α Reps. Or nous disposons au sein du laboratoire d'une bibliothèque d' α Reps pouvant reconnaître potentiellement n'importe quelle cible. Il semble donc raisonnable d'affirmer que les nouvelles propriétés créées par ajout du *linker* soient utilisables pour n'importe quelle cible.

B. Perspectives

Connaissant les nouveaux comportements créés par ajout du *linker*, il est possible d'imaginer des constructions plus complexes non étudiées dans ce manuscrit.

1) Enzyme activable

L'ouverture des bidomains **A3_A3** et **A3_bGFPD** par la fixation des ligands a été démontrée dans ce manuscrit. Cette ouverture inductible pourrait par exemple être utilisée pour activer ou inhiber une activité enzymatique à la demande. En effet, il a été démontré que le dimère de la protéine **A3** pouvait être transformé en une métallo-enzyme capable de catalyser la réaction de Diels Alder. Pour effectuer cette modification, un résidu situé dans la crevasse du dimère **A3** a été muté en cystéine et une phénanthroline de cuivre a été greffée sur chaque sous-unité de l'homodimère. (Di Meo, T., *et al.*, 2017)

On peut ainsi imaginer que dans un bidomaine **A3_A3**, il serait possible de greffer un seul ligand au sein de la crevasse et générer ainsi un autre site actif que celui créé sur le dimère **A3**. L'ajout de **bA3-2** pourrait alors ouvrir le bidomaine en interagissant avec le domaine **A3** non modifié et ainsi activer l'enzyme. (Figure 3-A)

De plus, nous avons également vu que des interactions faibles avec des constantes de dissociation de l'ordre du μM peuvent tout de même induire une conformation fermée pour une protéine chimérique quand les domaines sont reliés par un linker. Ainsi, il serait possible à partir des αReps de créer différents « métallo-bidomaine » formant des sites catalytique différents pouvant alors permettre de changer la spécificité de substrat d'une réaction, voire même de catalyser différentes réactions. (Figure 3-B)

Des travaux d'optimisation de la protéine **A3_A3** sur son utilisation en tant qu'enzyme font actuellement partie du projet de thèse de Kalani Kariyawasam en collaboration avec l'Institut de Chimie Moléculaire et des Matériaux d'Orsay (ICMMO).

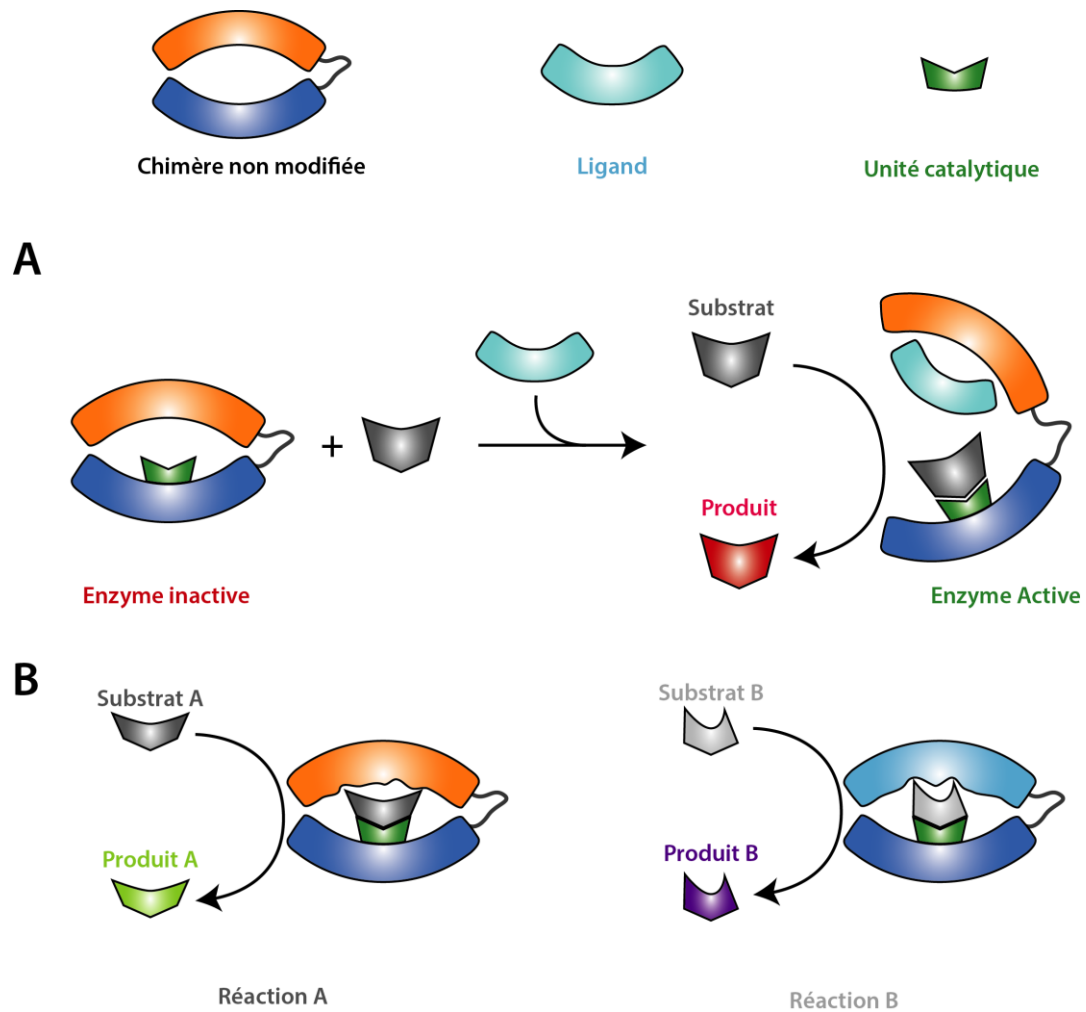


Figure 3 : Utilisation des bidomaines dans le cadre de réaction chimique.

A : Ouverture et activation du bidomaine enzyme suite à la fixation d'un ligand.

B : Changement de spécificité de substrat en fonction du domaine faisant face au domaine catalytique.

NB : Les proportions réelles des unités catalytiques et des substrats par rapport à l'enzyme ne sont pas respectées par soucis de clarté. Ceux-ci sont nettement plus petits que les bidomaines

2) Commutateurs modulables

Il semble également imaginable d'adapter des bidomains pour créer des commutateurs modulables. C'est-à-dire des interacteurs dont il serait possible de déclencher l'interaction par ajout d'un ligand. Cela permettrait de faire de l'interférence protéique contrôlable par exemple. Il a en effet été démontré au sein de l'équipe MIP qu'il était possible de sélectionner des « anti- α Reps » reconnaissant la face convexe (dos) des α Reps. Pour cela, les α Reps ont été modifiées à deux niveaux. Tout d'abord, leur dos a été dé-répété, cela signifie que les acides aminés normalement identiques entre chaque motif ont été changés pour différer d'un motif à l'autre. De plus, deux tyrosines ont été ajoutées sur ce dos. Il a été démontré que les tyrosines sont davantage représentées dans les CDRs des anticorps. Cela serait *a priori* dû à leurs multiples propriétés (hydrophile, hydrophobe et aromatique) qui permettraient de réaliser une grande variété d'interactions. (Collis, A.V.J., *et al.*, 2003) Les deux tyrosines ainsi ajoutées sur le dos des α Reps ont fait office de point chaud ou « Hot-spot » d'interaction. Il a alors été possible d'effectuer des sélections à partir de la bibliothèque d' α Reps « classiques ». Ces sélections ont permis d'isoler trois α Reps appelées anti- α Reps capables de reconnaître spécifiquement le dos modifié et tyrosiné. Ces anti- α Reps ne reconnaissent pas les α Reps ne possédant pas ces modifications. La structure cristallographique d'une anti- α Rep avec une α Rep au dos modifié a été résolue. Cette structure montre bien que l'interaction de l'anti- α Rep est établie en partie avec les deux tyrosines sur le dos modifié.

On peut alors imaginer un bidomaine possédant des α Reps différentes capables de reconnaître deux épitopes strictement différents d'une même cible. Le dos de chacun des domaines pourra être modifié afin d'être reconnu par des anti- α Reps. La conception d'un anti-bidomaine constitué de deux anti- α Reps peut être envisagée. Sa structure pourrait être rigidifiée de telle sorte que lorsqu'il interagit avec le bidomaine reconnaissant la cible, l'interaction bidomaine / cible soit rendue impossible. (Figure 4)

Des optimisations seront nécessaires notamment vis-à-vis des affinités des interacteurs qui devront être assez élevées pour que ceux-ci puissent interagir avec la cible mais pas trop fortement pour pouvoir être décrochés par « l'anti-bidomaine ». Une thèse effectuée par Adrien Fleck au sein de l'équipe MIP a pour objectif de réaliser de telles constructions.

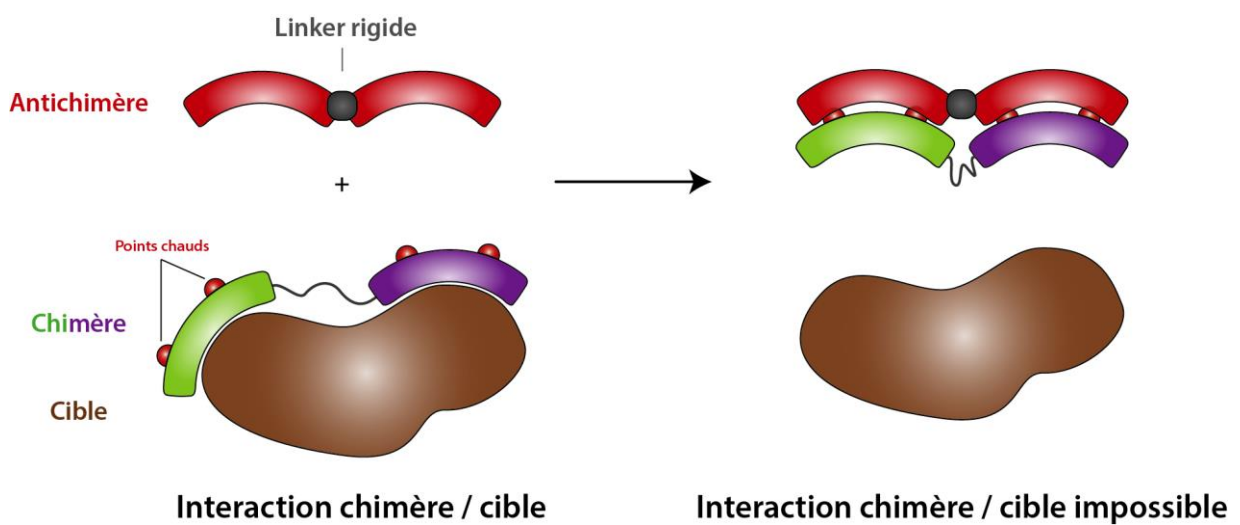


Figure 4 : Schéma de l'interacteur modulable exploitant le comportement flexible et bispécifique d'un bidomaine.

II. Biosenseurs

A. Basés sur l'utilisation du FRET

1) Conclusion

Afin de concevoir une nouvelle famille de biosenseurs potentiels, les bidomains **A3_A3** et **A3_bGFPD** ont été fonctionnalisés en les couplant à des sondes fluorescentes de telle façon que l'événement de liaison à la cible produise un signal fluorescent mesurable basé sur le FRET. Deux couples de fluorophores ont été testés.

Le premier constitué de deux atto-dyes a permis de voir qu'il était possible de détecter la fixation des ligands en suivant le signal de FRET. Cependant ce couple présentait l'inconvénient de ne pas permettre de différencier le ligand se fixant sur **A3_bGFPD**. La diminution maximale de signal étant la même pour **bA3-2** ou la **BFP**. De plus, la sensibilité de ce couple ne permettait pas une diminution des concentrations en bidomains. En effet, bien que permettant d'obtenir un pic d'émission de l'accepteur bien défini par rapport au pic du donneur, il était impossible de différencier le signal de l'accepteur du bruit de fond à des concentrations inférieures à 100 nM. La sensibilité de ce type de biosenseur pouvait donc être améliorée en changeant le couple de fluorophores associés.

Un second couple basé sur le Terbium et les QDs s'est révélé plus efficace. Possédant un rayon de Förster plus grand, ce couple permettait de différencier le ligand fixé sur le bidomaine **A3_bGFPD**. De plus, la possibilité d'effectuer des mesures résolues en temps réglait le problème de bruit de fond. Enfin, ce couple a permis de déterminer les distances et donc de mieux comprendre les mécanismes d'ouverture des bidomains. La réalisation de biosenseurs basés sur ces systèmes serait possible en test de routine. En effet, le Kryptor utilisé lors des détections de ligand est un appareil utilisé pour les diagnostics dans de nombreux laboratoires. Cependant, il semble encore nécessaire d'optimiser cette construction puisque les limites de détections, bien que correctes en tampon avec environ 0.5 nM, augmentent pour passer à environ 5 nM en sérum, qui est le milieu le plus utilisé pour les détections de cible. Cette optimisation pourrait notamment passer par une optimisation du *linker* pour le rendre plus rigide.

2) Optimisation des bidomaines L4Tb/QD625

Afin d'optimiser ces bidomaines pour diminuer les limites de détection. Une des pistes envisageables est de modifier le *linker* pour le rendre plus rigide sans empêcher la fermeture des bidomaines. On peut imaginer l'utilisation d'un *linker* hélicoïdal possédant des résidus déstabilisants en son centre. Cela permettrait d'adopter la forme fermée tout en garantissant une ouverture totale en cas de présence des ligands. L'ouverture étant plus grande, la variation du signal de FRET le sera également et permettra donc de diminuer la limite de détection. Ce type d'hélice capable d'adopter à la fois une forme « hélice » et une forme « hélice partielle et non structurée » a par exemple été observé dans la protéine AcrB d'*E. Coli*. (Zwama, M., et al., 2017)

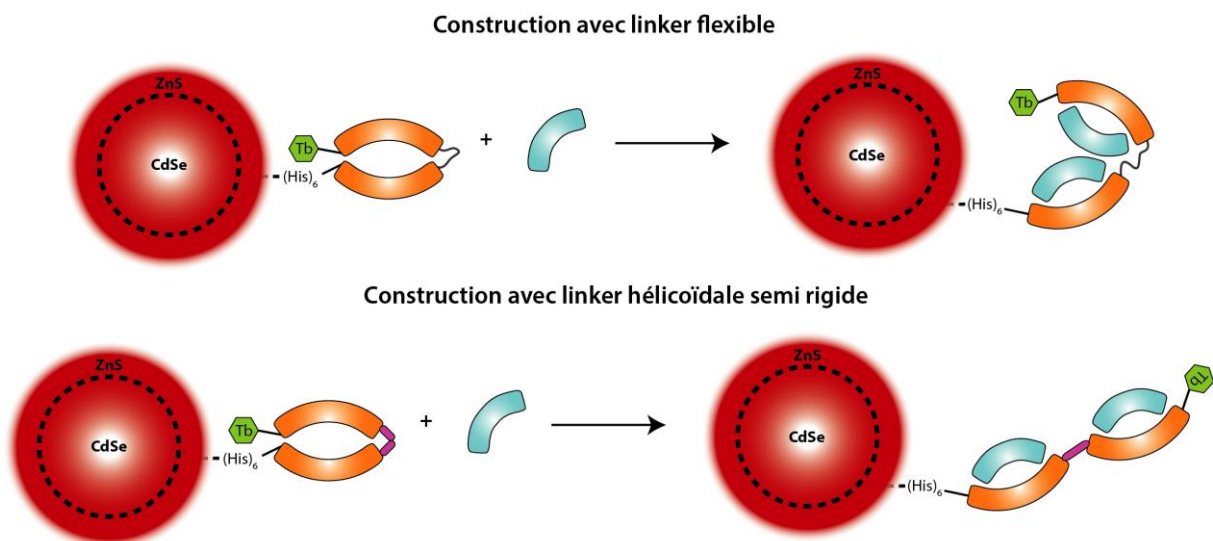


Figure 5 : Optimisation du bidomaine A3_A3 avec un linker hélicoïdale semi rigide permettant une ouverture totale.

B. Biosenseurs basés sur l'utilisation du LBT

1) Conclusion

Concernant les biosenseurs basés sur le LBT, les résultats semblent prometteurs bien que le concept ne puisse être utilisé pour le moment en raison d'une mauvaise spécificité et d'une instabilité. La dockérine semble *a priori* être une bonne piste mais la construction dans laquelle elle est insérée semble instable, notamment au niveau du *linker*. L'optimisation de ce dernier semble donc être une idée envisageable.

2) Optimisation de la dockérine

En ce qui concerne la dockérine, il a été observé qu'il y avait une hydrolyse, très probablement au niveau de la dockérine et du *linker* plus exposés au solvant et donc aux protéases. Dans l'hypothèse où des conditions de production sans hydrolyse ne soient pas trouvées, la fusion des deux protéines sans utiliser de *linker* est envisageable. En effet, la dockérine possède trois hélices dont une C-terminale, il est donc possible de fusionner cette hélice à la première hélice du motif N-ter d'une α Rep (Figure 6). Cela permettrait d'obtenir une forme stable de la dockérine, possédant alors moins de flexibilité et limitant l'exposition aux protéases. On peut donc par la suite imaginer un déplacement des tryptophanes pour obtenir le même effet antenne pour le transfert d'énergie au Tb, qu'avec le LBT.

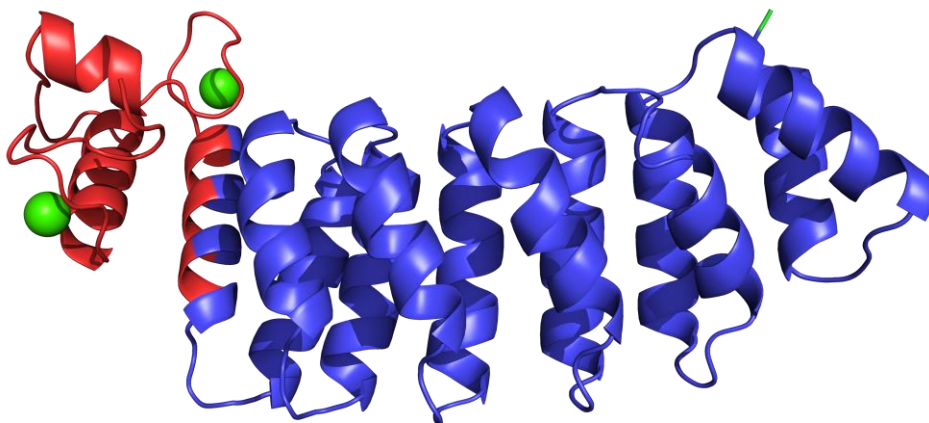


Figure 6 : Représentation en cartoon d'une dockérine (rouge) fusionnée via son hélice C-ter à l'hélice N-Ter d'une α Rep (bleu). Le Tb^{3+} pouvant se fixer dans cette protéine est représenté par des sphères vertes.

C. Biosenseurs basés sur la nanoKaz

1) Bilan

Les biosenseurs basés sur la nanoKaz semblent également être prometteurs, cependant ici aussi, une optimisation est nécessaire. En effet, l'inhibition de la nanoKaz par les protéines MG n'est pas suffisante pour entraîner un signal basal nul des biosenseurs. De plus, la présence des ligands ne semble pas suffisante pour entraîner un retour à une activité de la bioluminescence équivalente à celle de l'enzyme seule en solution. Cela serait très probablement dû au fait que les ligands n'ouvrent pas totalement la chimère.

Bien évidemment pour l'ensemble de ces biosenseurs, une fois totalement optimisés, des sélections contre des cibles présentant un intérêt thérapeutique ou expérimental devront être effectuées.

2) Optimisation du biosenseur NanoKaz

Afin d'optimiser la chimère, il semble ici aussi possible de rigidifier les *linkers* afin d'entraîner une ouverture totale de la chimère et donc une désinhibition plus importante. Cette rigidification peut ici aussi être effectuée à l'aide d'hélice semi-rigide par exemple.

Un autre moyen d'optimiser cette construction serait d'utiliser la PCA sur la nanoKaz. Il est en effet possible de séparer le dernier brin beta C-terminal, le peptide obtenu à une masse de 1.3 kDa et possède un K_D pour la nanoKaz de 190 μ M. Lorsque ce fragment est absent la bioluminescence de la nanoKaz diminue d'un facteur 300. (Dixon, A.S., *et al.*, 2016) Cette perte de luminescence peut être en partie due à l'absence de l'arginine 166 qui semble être impliquée dans la bioluminescence de la nanoKaz. (Inouye, S., *et al.*, 2014) Ce résidu fait par ailleurs partie des seize résidus mutés dans la nanoKaz par rapport à la Kaz sauvage. L'analyse de structure révèle que ce résidu pourrait potentiellement être impliqué dans le site actif (Figure 7).

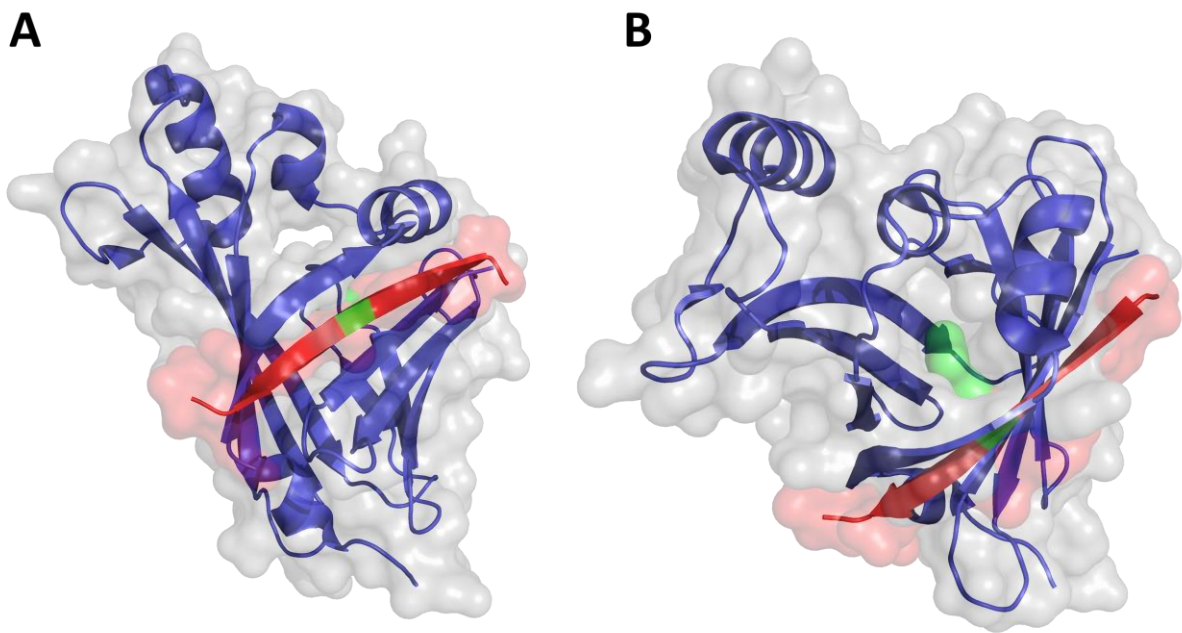


Figure 7 : Structure de la nanoKaz avec mise en évidence du peptide utilisé dans la PCA (en rouge) et de l'arginine 166 (en vert) impliquée dans la structure de la poche contenant le site actif.

Sachant que le peptide est proche du site actif, on peut imaginer une construction où en plus d'inhiber le site actif, la protéine MG se fixerait également sur l'emplacement du peptide. Cela permettrait d'empêcher l'accès du substrat au site actif et l'interaction du peptide avec la nanoKaz entraînant une diminution plus importante de la luminescence. L'affinité du peptide pour la nanoKaz étant faible par rapport à l'affinité des interacteurs qu'il est possible de sélectionner ($K_{D \text{ peptide}} \approx 190 \mu\text{M}$ et $K_{D \text{ MG}} \approx 1\mu\text{M}$), il serait possible de relier le peptide à la nanoKaz dans la chimère. La fermeture du bidomaine αRep (ici **bGFPPD**) viendrait rapprocher la protéine MG de la nanoKaz. Celle-ci pourra alors interagir avec la nanoKaz et empêcher la fixation du peptide. Cela réduirait le niveau basal d'activité des chimères.

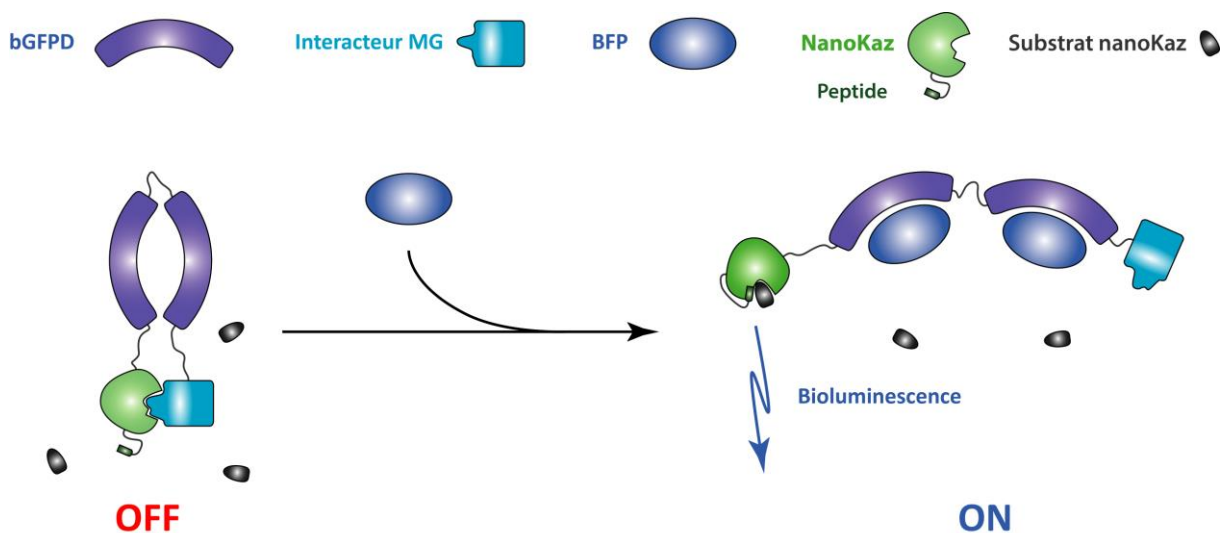


Figure 8 : Schéma de la construction optimisée du biosenseur nanoKaz. Le peptide (vert foncé) vient se loger à côté du site actif uniquement quand la protéine MG (cyan) est décrochée.

De tels biosenseurs impliquant la nanoKaz seraient eux aussi généralisables à n'importe quelle cible contre laquelle il est possible de sélectionner des αReps . Ils permettraient également de pouvoir détecter les cibles à l'aide d'un simple appareil photo comme les biosenseurs LUCID abordés dans l'introduction de ce manuscrit. (Griss, R., et al., 2014)

D. Conclusion générale sur les biosenseurs

L'ensemble des biosenseurs étudiés dans ce manuscrit, présente deux avantages considérables.

Tout d'abord, les propriétés des α Reps, et notamment le fait de former des bidomains fermés par simple ajout d'un linker, semblent être des propriétés quasi-générales. Disposant d'une bibliothèque d' α Rep, les constructions présentées ici pourraient être adaptées à n'importe quelles α Reps spécifiques d'une cible donnée. Les biosenseurs présentés ici ont donc l'avantage d'être généralisables pour la détection de n'importe quelle cible contre laquelle il est possible de sélectionner des α Reps (Phage display et PCA).

D'autre part, ces biosenseurs fonctionnent en ne reconnaissant qu'un seul épitope et non pas en « sandwich » comme la majorité des tests diagnostiques utilisant les anticorps et leurs dérivés disponibles actuellement. L'ingénierie des anticorps est par ailleurs plus contraignante que celle impliquant les α Reps. En effet, leur ossature complexe impliquant des ponts disulfures et leur grande taille rend leur fonctionnalisation plus difficile. De plus, la production d'anticorps monoclonaux s'effectue très largement en cellule de souris par la technologie de l'hybridome. (Köhler, G. and Milstein, C., 1975) Cela entraîne d'une part une augmentation des coûts de production. D'autre part de nombreuses protéines humaines présentent une forte homologie avec les modèles murins utilisés, ainsi de nombreux antigènes humains n'entraînent pas de réponse immunitaire chez ces modèles. Les α Reps ne présentent *a priori* pas d'inconvénients de fonctionnalisation et, étant produites dans *E. Coli*, elles ne nécessitent pas de réponse immunitaire pour être sélectionnées. De plus, elles sont exprimables et fonctionnelles en cellules eucaryotes, les biosenseurs présentés ici pourraient donc être utilisés *in vivo*.

III. Bibliographie

- Collis, Abigail V. J., Adam P. Brouwer, and Andrew C. R. Martin. 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*, 325: 337-54.
- Di Meo, T., W. Ghattas, C. Herrero, C. Velours, P. Minard, J. P. Mahy, R. Ricoux, and A. Urvoas. 2017. ' α Rep A3: A Versatile Artificial Scaffold for Metalloenzyme Design', *Chemistry*, 23: 10156-66.
- Dixon, A. S., M. K. Schwinn, M. P. Hall, K. Zimmerman, P. Otto, T. H. Lubben, B. L. Butler, B. F. Binkowski, T. Machleidt, T. A. Kirkland, M. G. Wood, C. T. Eggers, L. P. Encell, and K. V. Wood. 2016. 'NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells', *ACS Chem Biol*, 11: 400-8.
- Griss, R., A. Schena, L. Reymond, L. Patiny, D. Werner, C. E. Tinberg, D. Baker, and K. Johnsson. 2014. 'Bioluminescent sensor proteins for point-of-care therapeutic drug monitoring', *Nat Chem Biol*, 10: 598-603.
- Inouye, S., J. Sato, Y. Sahara-Miura, S. Yoshida, and T. Hosoya. 2014. 'Luminescence enhancement of the catalytic 19 kDa protein (KAZ) of Oplophorus luciferase by three amino acid substitutions', *Biochem Biophys Res Commun*, 445: 157-62.
- Köhler, G., and C. Milstein. 1975. 'Continuous cultures of fused cells secreting antibody of predefined specificity', *Nature*, 256: 495-97.
- Zwama, M., K. Hayashi, K. Sakurai, R. Nakashima, K. Kitagawa, K. Nishino, and A. Yamaguchi. 2017. 'Hoisting-Loop in Bacterial Multidrug Exporter AcrB Is a Highly Flexible Hinge That Enables the Large Motion of the Subdomains', *Front Microbiol*, 8: 2095.

“Goodbye? Oh no, please. Can’t we just go back to page one and start all over again?”
- Winnie the Pooh -



Titre : Conception de Protéines Artificielles Multidomaines

Mots clés : Protéines artificielles, Protéines multidomaines, Biochimies des protéines, Biologie moléculaire, Biophysique.

Résumé : La création de nouvelles fonctions basées sur la reconnaissance protéique et sur l'assemblage de domaines est un enjeu majeur en biotechnologie et est un moyen de comprendre les relations structures/fonctions des protéines engagées dans des processus d'interactions. Aujourd'hui, des bibliothèques de protéines artificielles obtenues par ingénierie peuvent être sources de protéines aux propriétés de reconnaissance analogues à celles des dérivés d'anticorps.

L'équipe Modélisation et Ingénierie des Protéines a ainsi construit une banque de protéines à motifs structuraux répétés appelées « α Reps ». Les α Reps présentent des propriétés remarquables en termes de production et de stabilité. Contrairement à la plupart des anticorps et dérivés d'anticorps, elles peuvent même s'exprimer sous forme fonctionnelle dans le cytoplasme de cellules eucaryotes. De tels objets peuvent donc maintenant être utilisés comme des briques élémentaires en vue d'une ingénierie modulaire. Ainsi la construction de nouvelles fonctions de reconnaissance optimisées tant au niveau de la spécificité que de l'affinité sera possible en réarrangeant et/ou dupliquant ces briques élémentaires.

Un premier volet de ce projet de thèse a consisté à construire puis étudier les propriétés biophysiques de protéines bidomaines basées sur les α Reps afin de mieux comprendre les comportements adoptés par de telles constructions. Outre l'aspect fondamental de cette question, cette étude donnera « les règles » pour moduler de façon contrôlée les interactions entre ces protéines. Les résultats montrent qu'il est possible de créer de nouvelles fonctions par simple ajout d'un linker entre deux α Reps : avidité, coopérativité, changement de conformation.

Dans un second temps, l'objectif a été de développer, à partir des protéines bidomaines précédemment étudiées, de nouveaux biosenseurs basés sur le FRET (Förster Resonance Energy Transfer) pouvant être utilisés *in vivo* et *in vitro*. Cette deuxième partie présente deux biosenseurs avec des limites de détection de l'ordre du nanomolaire. Les α Reps utilisées dans ces constructions pouvant être changées en fonction de la cible souhaitée, il s'agit ici d'une preuve de concept pouvant être généralisée à n'importe quelle cible.

Enfin la dernière partie de cette thèse s'est portée sur la conception et l'étude de nouveaux biosenseurs génétiquement codables. Ces biosenseurs présentent notamment l'avantage d'être utilisables immédiatement après production et ne nécessitent donc plus d'étape de couplage chimique. Les résultats obtenus montrent que la création de tels biosenseurs est possible mais qu'une optimisation reste encore nécessaire pour améliorer leur spécificité, leur stabilité et leur capacité de détection.

Title: Conception of multidomain artificial proteins

Keywords: Artificial Proteins, Multidomain proteins, Biochemistry, Molecular Biology, Biophysics

Abstract: The creation of new protein functions based on recognition and molecular assembly is not only a major goal in biotechnology but is also a means to understand the relation structure/function of proteins involved in interaction processes. Today, libraries of artificial proteins obtained by engineering can be a source of proteins with recognition properties similar to the properties of antibodies.

The team Protein Engineering and Modeling has thus created a library of proteins with structural repeats called the " α Reps". The α Reps present remarkable properties in terms of production and stability. Unlike most of the antibodies and their derivatives, they can even be expressed and functional in the cytoplasm of eukaryotic cells. Such objects can therefore be used as building bricks in modular engineering. The construction of new optimized recognition functions both in specificity and in affinity can then be possible by rearranging or duplicating these elementary bricks.

The first part of this thesis project consisted in the construction and study of the biophysical properties of bidomain proteins based on α Rep in order to have a better understanding of the behaviour of such constructions. Beside the fundamental aspect of this question, this study will give the "rules" to modulate the interactions between these proteins in a controlled way. The results show that it is possible to create new functions such as avidity, cooperativity, conformational change, simply by adding a linker between two α Reps. In a second step, the goal was to develop, with the bidomain proteins previously studied, new biosensors based on the FRET (Förster Resonance Energy Transfer) which can be used *in vivo* and *in vitro*. This second part presents two biosensors with limits of detection in the nanomolar range. Since the α Reps used in these constructions can be changed depending on the chosen target, it is a proof of concept which can be adapted to any desired target.

Finally, the third part of this thesis focused on the development of genetically codable biosensors. These biosensors have the particular advantage of being usable directly after production and therefore no longer require a chemical coupling step. The results show that the development of such biosensors is worth considering but an optimization is still required in order to improve their specificity, their stability and their detection capacity.

