Ingénierie de l'architecture protéique artificielle αRep : élaboration de catalyseurs biohybrides par couplage covalent de complexes métalliques

Thibault Di Méo

▶ **To cite this version:**

Thibault Di Méo. Ingénierie de l'architecture protéique artificielle αRep : élaboration de catalyseurs biohybrides par couplage covalent de complexes métalliques. Biochimie [q-bio.BM]. Université Paris Saclay (COmUE), 2017. Français. NNT : 2017SACLS012 . tel-01618227

HAL Id: tel-01618227

https://theses.hal.science/tel-01618227

Submitted on 17 Oct 2017

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.

NNT : 2017SACLS012

THESE DE DOCTORAT
DE
L'UNIVERSITE PARIS-SACLAY
PREPAREE A
L'UNIVERSITE PARIS-SUD

ECOLE DOCTORALE N°569
Innovation thérapeutique : du fondamental à l'appliqué

Spécialité de doctorat : Biochimie et Biologie Structurale

Par

M. Thibault DI MÉO

Ingénierie de l'architecture protéique artificielle αRep : élaboration de catalyseurs biohybrides par couplage covalent de complexes métalliques

Thèse présentée et soutenue à Orsay, le 19 janvier 2017 :

Composition du Jury :

Mme. SIMAAN, Jalila	Directeur de recherche – Université Aix-Marseille	Rapporteur
M. SOUMILLION, Patrice	Professeur – Université Catholique de Louvain	Rapporteur
M. MAHY, Jean-Pierre	Professeur – Université Paris-Sud	Examinateur
M. NIZAK, Clément	Chargé de recherche – ESPCI ParisTech	Examinateur
Mme. URVOAS, Agathe	Maitre de conférences – Université Paris-Sud	Co-directrice de thèse
M. RICOUX, Rémy	Ingénieur de recherche – Université Paris-Sud	Co-directeur de thèse

Avant toutes choses…

Il est ici venu l'exercice remerciements. Travail complexe s'il en est, car indubitablement cette page sera, sans nul doute, la plus scrutée de tout ce manuscrit.

Il s'agit de s'y prendre de façon méthodique, ces remerciements s'articuleront donc en trois parties : La famille, le labo, les amis (la patrie, la science et la gloire).

Merci donc tout d'abord à mes parents pour m'avoir autorisé, si ce n'est de naître, de poursuivre mes études dans une carrière faite de biologie, recherche (d'emploi ?) et avant tout de passion. Merci également à mes frères pour me rappeler qu'on ne choisit pas sa famille, mais que des fois le hasard fait bien les choses (indice : un compliment est caché dans cette phrase). Merci enfin à ma sœur, présidente et unique membre de mon fan club, pour tes encouragements au moins aussi fervents que ceux que tu réserves à l'ASM. Merci enfin à la famille au sens large : aux Mousavisadeh, aux grand-mères, à Tonton Philippe, aux amis du rugby du paternel, tous sans exception avez participé à faire de moi la personne que je suis aujourd'hui.

Merci ensuite aux gens qui m'ont accueilli pendant folles années : tout d'abord Agathe, pour ta patience maternelle et Rémy pour ton positivisme communicatif. Merci à vous deux de m'avoir guidé, soutenu (supporté ?) et surtout permis de mener à terme ce projet pas évident : vous êtes au top ! Merci à Philippe (« Pr. Minard ») pour m'avoir donné l'envie de travailler sur les protéines, votre enthousiasme concernant les extraordinaires possibilités de l'évolution dirigée est contagieux.

Merci à tous les membres du MIP (et j'inclus les PIM's) ! Ceux qui m'avaient accueilli comme le petit nouveau et qui sont parti : Anne (l'Exemple !), Yann, Danielle … Ceux que j'ai accueilli comme les petits nouveaux et qui me survivront : Corentin (ne fait jamais confiance aux nains de l'espace), Sébastien (Post d'oc !), Margarida (many things to say, je résumerais donc par un 💩🔴 , I'll see you on the dark side of the thesis !).. Et ceux qui nous survivront tous : les permanents. Marielle, Magali et Agnès, rien ne ternit votre bonne humeur (si ce n'est les étudiants pas doués ou les problèmes de H&S). Michel et Christophe, vous m'avez prouvé que les geeks étaient partout ! Merci enfin à Mumtaz pour sa gentillesse et à Pierrette pour tout ce qui est bon dans la vie : les confitures et les palets bretons !!!

Merci à tous les membres du ECBB ! Jean-Pierre, Wadih, Frédéric, Yann, Alexandre, Khaled, Kalani… Que ce soit pour boire des bières ou parler de science, entre gueule de bois et nouvelles connaissances sur la chimie, être à votre contact m'a toujours apporté quelque chose ! Sans oublier tous ceux avec qui j'ai échangé, lors des séminaires, du

barbecue annuel, ou tout simplement dans les couloirs… La liste est malheureusement bien longue pour tenir sur une page : à grosse équipe petits remerciements !

Merci enfin aux amis ! Les anciens, YAN et Fabze, toujours présents même dans le plus profond des black-out. Sans oublier l'gros Geo si loin en Bretagne… Bulie, Daïe (on ne crie pas dans le hameau), Léa, un peu de féminité dans ce monde de brutes (ahem..). Les copinous de la Fac, partenaires de Cité U, les gens du terter du 336, tout ce petit monde maintenant éparpillé aux quatre du globe : Attendez-moi, j'arrive ! Et pour finir, la Faisandière, encore une nouvelle famille de joyeux drilles qui ont égayé ces deux dernières années dans les plus sombres heures de la thèse. Prolongation du rêve éveillé de la cité U et occasion de nouvelles rencontres formidables, grâce à vous ces années resteront inoubliables.

À tous ceux qui auraient maintenant l'envie de poursuivre la lecture de ce manuscrit, je leur souhaite une bonne lecture.

Thibault

Table des matières

Partie A - Introduction 13

Préambule Chimie verte et catalyse sélective, des enjeux d'actualité 15

Chapitre I - La réaffectation enzymatique 20

I.1 L'évolution dirigée d'enzymes : une approche darwinienne 20

I.2 « Voyage » dans l'espace des séquences 21

I.3 La promiscuité enzymatique, un point de départ pour l'ingénierie 25

I.4 La génération de diversité : vers des banques de plus en plus rationnelles 26

I.5 Réduction de la diversité : sélection *versus* criblage 37

I.6 Conclusion 47

Chapitre II - Les métaux dans les enzymes: métalloenzymes artificielles et biohybrides 49

II.1 Le design de centre métallique datif 52

II.2 Insertion d'un cofacteur métallique au sein d'une architecture protéique: les catalyseurs biohybrides. 57

Chapitre III - Contexte scientifique de la thèse 103

III.1 Une architecture protéique modulaire 103

III.2 Construction de la banque de protéines αRep. 105

III.3 La structure αRep, une architecture de choix pour la conception de catalyseurs biohybrides ? 110

III.4 La protéine αRep-A3, un représentant particulier de la banque αRep 110

III.5 Choix des complexes métalliques 111

Bibliographie 113

Partie B - Résultats 131

Chapitre I - Utilisation du dimère αRep-A3 133

I.1 Etude structurale du complexe αRep-A3 et choix des positions 133

I.2 Obtention des biohybrides 139

I.3 Analyse biophysique des variants fonctionnalisés 142

I.4 Mime d'une activité RNAse par les biohybrides dérivés d'αRep A3 150

I.5 Etude d'une activité catalytique énantiosélective : la réaction de Diels-Alder. 157

I.6 Analyse structurale des différents variants 178

I.7 Résumé de l'étude du dimère A3s 183

Chapitre II - Utilisation de la protéine αRep-A3sA3 185

II.1 Construction génétique codant la protéine A3sA3 185

II.2 Caractérisation de la protéine A3sA3 186

II.3 Couplage de petits ligands phénanthroline et terpyridine 188

II.4 Couplage de la protéine A3sA3 avec une porphyrine. 202

II.5 Résumé de l'utilisation des variants de A3sA3 pour le greffage d'une porphyrine 217

Bibliographie 218

Partie C - Matériel et méthodes 223

Chapitre I - Chimie 225

I.1 Synthèse des ligands 225

I.2 Synthèse des substrats et authentiques 229

I.3 Cristallographie du complexe Cu(II)[(phen)2(NO_3)]$^+$ 231

Chapitre II - Biologie 232

II.1 Constructions génétiques 232

II.2 Obtention des protéines 241

Chapitre III - Biohybride 246
III.1 Couplage des cofacteurs 246
III.2 Caractérisation 247
III.3 Catalyse 251

Chapitre IV - Bio-informatique 255
IV.1 Représentations moléculaires 255
IV.2 Modélisation de docking moléculaire 255
IV.3 PDB-Mining des protéines binucléaires à zinc 255
Bibliographie 256

Partie D - Conclusion 259

Chapitre I - La protéine αRep A3 : une architecture prometteuse pour la génération de nouveaux biocatalyseurs hybrides 261

Chapitre II - Plasticité du dimère et dynamique entre les sous-unités 263

Chapitre III - Poursuite du projet 269
III.1 Vers une stratégie d'évolution dirigée basée sur A3sA3 269
III.2 Utilisation des αRep pour une autre stratégie : la liaison directe du métal 275
III.3 Utilisation d'un catalyseur commutable 279
Bibliographie 280

Partie E - Annexes 283
Tables cristallographiques 285
Article 289

Abréviations

⊂	Insertion supramoléculaire d'un complexe dans une protéine
Å	Ångström
AA	Acide Aminé
AcOEt	Acétate d'éthyle
ADN	Acide désoxyribonucléique
Aza	Azachalcone
CAS	Enzyme Commission number
CAST	Mutagénèse combinatoire du site actif
CATH	Méthode de classification des protéines par leur structure
CCM	Chromatographie sur couche mince
CG	Chromatographie Gazeuse
CLHP	Chromatographie Liquide Haute Performance
Cp	Cyclopentadiène
Cu	Cuivre
Cu(NO3)2	Nitrate de zinc
Da	Dalton
DCM/CH2Cl2	Dichlorométhane
DMF	Diméthylformamide
DMSO	Diméthylsulfoxyde
dNTP	Désoxyribonucléotide triphospate
DO	Densité optique (absorbance)
E	Facteur d'énantiosélectivité
EDA	diazoacétate d'éthyle
EDC	1-Éthyl-3-(3-diméthylaminopropyl)carbodiimide

EDTA	Éthylène diamine tétra-acétique
ee	Excès énantiomérique
ESI-ToF	Spectrométrie de masse en temps de vol avec ionisation par electrospray
GC	Chromatographie Gazeuse
h	Heure
H_2	Dihydrogène
H2O2	Peroxyde d'hydrogène
HCl	Acide Clorhydrique
HEPES	Acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique
HPLC	Chromatographie Liquide Haute Performance
ICP-MS	Spectroscopie de masse à plasma à couplage inductif
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ISM	Mutagénèse à saturation itérative
ITC	Titration Calorimétrique Isotherme
kcat	Nombre de cycles catalytiques par catalyseur par unité de temps kcat ≡ TOF
KD	Constante de dissociation
kDa	Kilo Dalton
KHSO5	Hydrogénopersulfate de potassium
KM	Constante de Michaëlis
KOH	Hydroxyde de Potassium
L	Ligand
L (m,µ,n,p,f)	litre (milli, micro, nano, pico, femto)
M	molaire (mol/L)
m (milli, micro, nano, pico, femto)	mètre (milli, micro, nano, pico, femto)

mal	maléimide
MALDI-ToF	Désorption et ionisation par LASER assisté d'une matrice, couplé à une analyse par temps de vol.
MeOH	Méthanol
MES	Acide 2-(N-morpholino)ethanesulfonique
MgSO4	Sulfate de magnésium
MHz	Mégaherz
min	Minute
Mn	Manganèse
MnTPP-mal	Mn(III) 5,10,15-triphényl-20-(4-maléimidopropylamido)phényl)-porphyrine
MOPS	Acide 3-(N-morpholino)propanesulfonique
Myo	Myoglobine
Na	Sodium
NAD(P)H	Nicotinamide adénine dinucléotide (phosphate)
NaOH	Hydroxyde de Sodium
O/N	Sur la nuit
O_2	Dioxygène
Oac	Acétate
Otf	Triflate
Oxone	Hydrogénopersulfate de potassium
PCR	Réaction en chaine par polymérase
pH	Potentiel hydrogène
phen	Phénanthroline
ppm	Partie par million
Rf	Rapport frontal
RMN	Résonance magnétique nucléaire

RPE	Résonance paramagnétique électronique
s	Seconde
SCOP	Méthode de classification des protéines par leur structure
SDS-PAGE	Électrophorèse en gel de polyacrylamide en condition dénaturante
SPAAC	Cycloaddition alcyne-azide favorisée par contraintes
TEV	Virus de la mosaïque du tabac
TFA	Acide Trifluoroacétique
TMPYP	*5,10,15,20*-(4-N-méthylpyridiniumyl)-porphyrine
TMPYP-SCN	5,10,15-tri(4-N-méthylpyridiniumyl)-20-(4-iothiocyanatophényl)-porphyrine
TOF	Nombre de cycles catalytiques par catalyseur par unité de temps TOF ≡ kcat
TON	Nombre de cycles catalytiques par catalyseur
Tp	Tampon
TPP	Tetraphenylporphyrine
TPPMn	Mn(III) 5,10,15-triphényl-20-(4-maléimidopropylamido)phényl)-porphyrine couplée à A3sA3
Tpy/terpy	Terpyridine
TTN	Nombre total de cycles catalytiques
UCA	Ultracentrifugation analytique
UV	Ultraviolet
Zn(NO3)2	Nitrate de zinc
ΔG	Variation d'enthalpie libre

Partie A - Introduction

Préambule

Chimie verte et catalyse sélective, des enjeux d'actualité

Au cours du siècle dernier l'évolution de la chimie nous a permis d'obtenir pléthore de matières et de molécules (polymères, phythosanitaires, pharmaceutiques) : présente dans tous les domaines, la chimie a ainsi profondément modifié notre quotidien. Il est cependant aujourd'hui nécessaire de répondre à de nouvelles problématiques. L'avenir se dessinant avec des ressources pétrolières limitées, nous devons penser la chimie d'une manière plus sobre. D'un point de vue environnemental, il est également nécessaire de parvenir à limiter les déchets produits. La chimie doit, aujourd'hui plus que jamais, devenir plus précise et plus efficace. De ce constat est né le concept de la chimie verte, introduit dans les années 90 par Paul Anastas et John C. Warner [1], qui prône l'utilisation de réactions de synthèse moins consommatrices d'énergie et plus sélectives. En effet, un type et un nombre donné d'atomes peut constituer différentes molécules : des isomères. La chiralité est un type particulier d'isomérie où, à l'instar de deux mains, les deux molécules (dites énantiomères) sont l'image l'une de l'autre dans un miroir (figure 1).

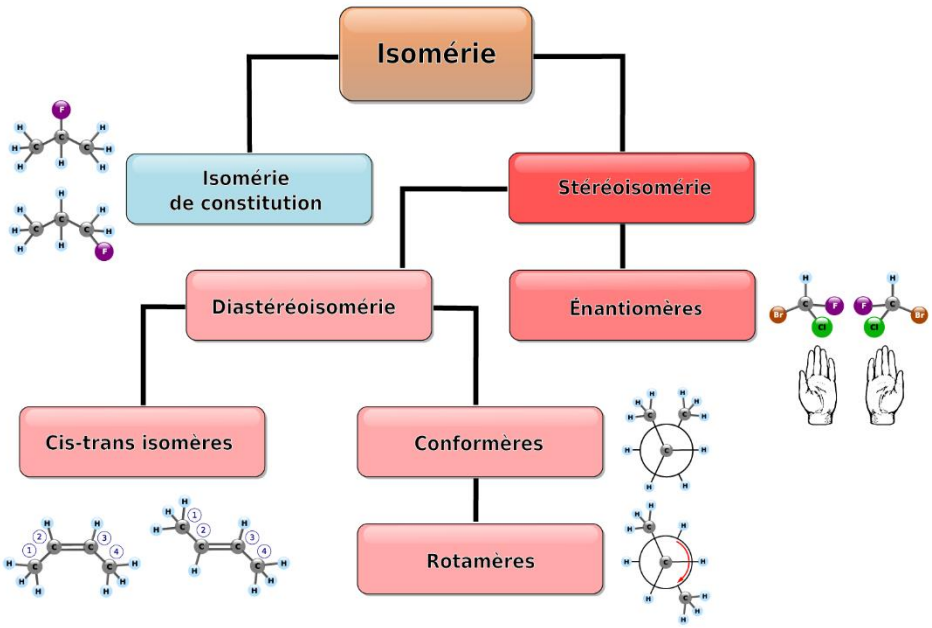


Figure 1 : La sélectivité d'une réaction chimique peut concerner tous les types d'isomérie présentés sur cette figure. Nous nous concentrerons principalement sur l'énantiosélectivité.

La sélectivité, et plus particulièrement l'énantiosélectivité, est d'un intérêt majeur dans la synthèse de molécules biologiquement actives. En effet, ces molécules agissent pour la plupart en se fixant à des biomolécules, que ce soit des récepteurs ou des enzymes, ou des acides nucléiques afin de moduler leur activité. La spécificité de cette fixation dépend de l'adéquation entre les deux partenaires, suivant un modèle de clé-serrure ou d'ajustement induit.[2,3] En effet, les biomolécules étant des molécules chirales énantiopures, deux énantiomères d'un composé de même formule chimique mais de configurations absolues différentes (voir figure 2) pourront avoir des capacités d'interactions très différentes. Que ce soit dans le cas de médicaments, de phytosanitaires ou de parfums, la structure spatiale des molécules conditionne donc leur fonction. Beaucoup d'exemples sont disponibles pour illustrer ce phénomène, le plus éloquent étant le scandale de la thalidomide. Ce médicament était utilisé durant les années 50 et 60 comme sédatif et anti vomitif administré aux femmes enceintes. Il s'est avéré qu'alors que la cette molécule dans la configuration absolue S possédait effectivement l'effet désiré, la molécule dans la configuration R avait un effet tératogène. La racémisation de cette molécule au sein du corps humain rendit impossible l'utilisation de cette molécule énantiopure et elle fut retirée du marché. Depuis, une étude complète sur l'activité biologique de chacun des énantiomères d'une molécule thérapeutique est demandée par les instances de certification avant toute mise sur le marché si le produit est distribué sous forme racémique. D'autres exemples moins dramatiques mais tout aussi importants en terme de marché mondial peuvent être cités: l'aspartame par exemple n'a un pouvoir sucrant que dans l'un de ses 4 diastéréoisomères, les autres étant amers ou sans goût. Il est alors évident que l'obtention de l'énantiomère désiré seul, sans avoir recours à une séparation ultérieure des différents isomères doit être privilégiée en termes économiques en portant les efforts sur l'efficacité de production et, à plus longue échéance, en termes écologiques en limitant l'impact environnemental.

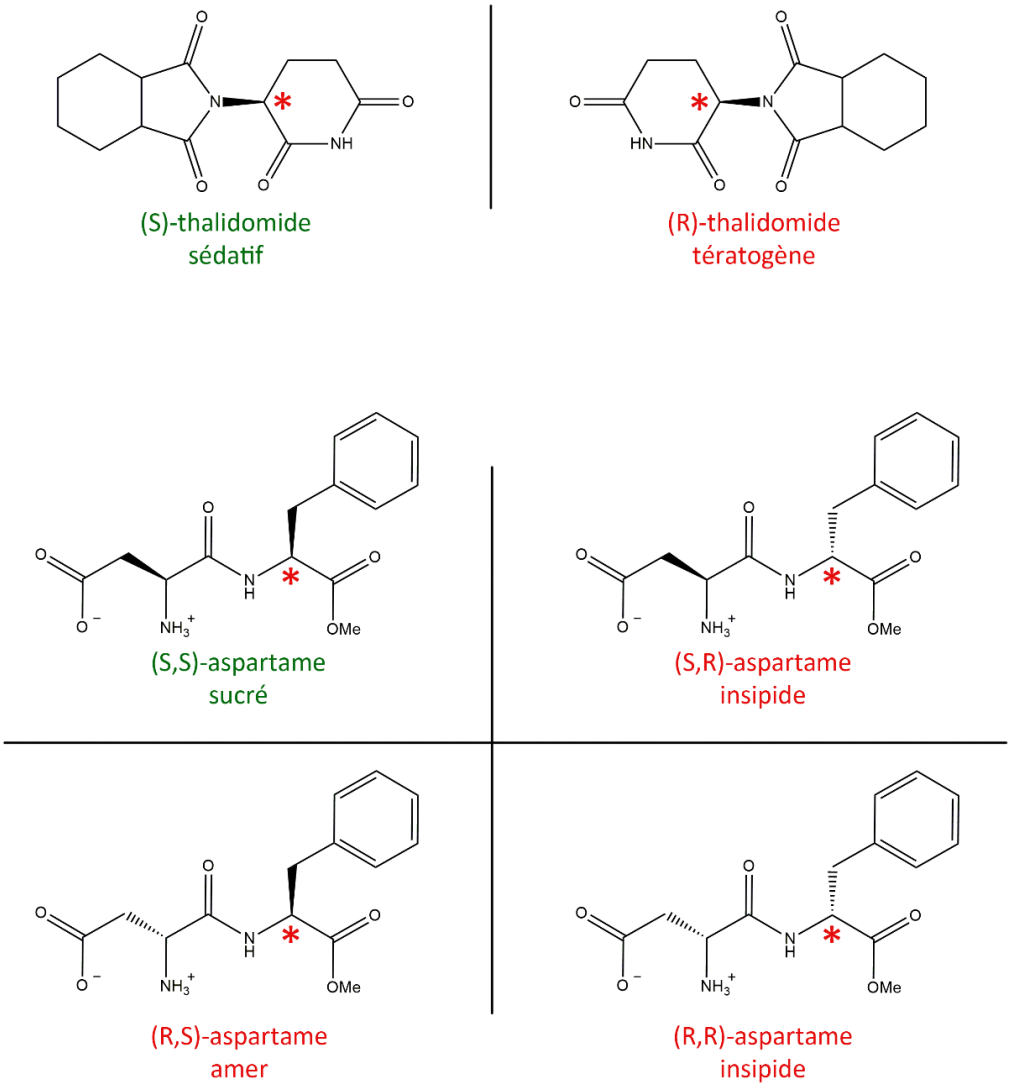


Figure 2 : Exemples de molécules chirales dont la réponse biologique dépend de l'isomère

Afin d'éviter une coûteuse et laborieuse séparation des différents énantiomères, la meilleure manière d'obtenir efficacement des produits énantiopurs est d'utiliser des procédés catalytiques énantiosélectifs.[4]

Parmi les options disponibles, la catalyse asymétrique à base de catalyseurs organométallique chiraux (figure 3) a révolutionné la chimie de la fin du 20^{e} siècle grâce aux travaux pionniers de Sharpless, Noyouri, Knowles et Kagan. Ces travaux ont alors valu aux trois premiers le prix Nobel de chimie en 2002, confirmant l'intérêt porté par la communauté scientifique à la catalyse asymétrique. Cependant ces catalyseurs de petite taille, bien que très efficaces, ne permettent obtenir une régio et stéréo sélectivité comparable à celle d'enzymes naturelles. En effet, ces dernières possèdent, grâce à leur grande structure tridimensionnelle et à la chiralité des acides aminés qui les composent,

une capacité de reconnaissance particulièrement précise, permettant le placement optimal des substrats en faveur de la réaction catalysée.

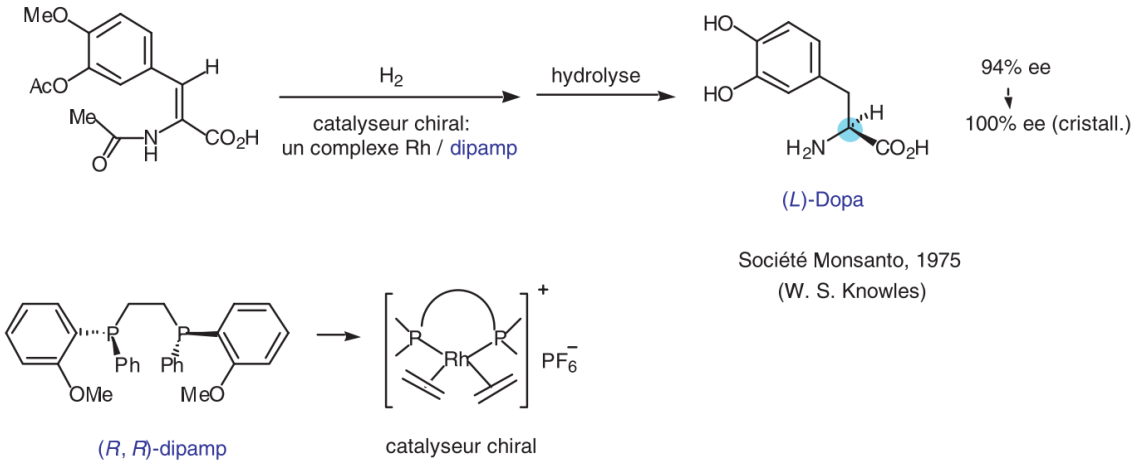


Figure 3 : Synthèse asymétrique industrielle de la (L)-DOPA (selon [5])

Les enzymes sont donc des catalyseurs d'une extrême efficacité utilisés dans le monde de la synthèse organique pour leur avantage en terme de sélectivité [6], tout en fonctionnant dans des conditions douces de température et de solvant. Cependant, les enzymes naturelles sont issues d'un processus d'évolution pour une fonction donnée dans un contexte cellulaire particulier, qui n'est pas forcément propice à une utilisation industrielle. Il existe également des réactions qui ne sont pas catalysées par les enzymes naturelles. Le champ de l'ingénierie enzymatique consiste donc en leur modification afin de catalyser de nouvelles réactions dans des conditions requises par l'industrie. Il est également intéressant de coupler la chimie des métaux de transition à la puissante capacité d'évolution des protéines.

Nous verrons donc dans cette introduction des méthodes utilisées en ingénierie des protéines pour la réaffectation d'une enzyme vers une nouvelle réaction ou un nouveau substrat. Nous aborderons ensuite plus précisément le sujet de la conception de métalloenzymes artificielles. Pour finir, je présenterai les architectures protéiques que nous avons exploitées en vue du développement d'une nouvelle génération de métalloenzymes artificielles.

Excès énantiomérique, pureté optique et sélectivité

La manière de quantifier l'efficacité des catalyseurs pour l'obtention de produits énantiopurs est diverse et parfois déconcertante, il est donc nécessaire de définir les termes employés.[194]

Pureté optique, excès énantiomérique et composition énantiomérique :

L'indice historique de la composition d'une solution contenant deux énantiomères est la « pureté optique » (noté op). Cet indice provient du moyen traditionnel de quantification des substances optiquement actives, qui est le rapport entre la rotation de la lumière polarisée observée ($[\alpha]_{obs}$) et la rotation maximale possible du produit énantiopur ($[\alpha]_{max}$). L'apparition de méthodes analytiques comme la chromatographie chirale ou la RMN, permettant la quantification précise des deux énantiomères, a progressivement rendu obsolète la mesure de l'activité optique. Cependant l'indice appelé « **excès énantiomérique** » (noté ***ee***) et son cousin l' « excès diastéréomérique » (noté *de*) ont été défini en adéquation avec le terme de pureté optique. Les équations sont donc, pour deux énantiomères *R* et *S* (avec *R* + *S* = 1, mais souvent exprimé en pourcentage) :

$$ee = op = \frac{[\alpha]_{obs}}{[\alpha]_{max}} = \frac{R - S}{R + S}$$

Cet indice d'excès énantiomérique est contestable, car il représente de manière peu intuitive la composition d'une solution. En effet cet indice implique que le « contaminant » n'est pas un des énantiomères mais le mélange racémique. Il est donc possible d'utiliser le rapport entre les deux produits sous la forme d'un ratio simple $q = \frac{R}{S}$ $q \in]0; +\infty[$, ou en un ratio ayant 1 comme dénominateur (ex. 100/1, 1000/1), ou encore normalisé par rapport à 100 (ex. 98:2, 60:40), ce dernier que Kagan nomme « **composition énantiomérique** » (notée ***ec***).[195] Malgré ses désavantages, l'excès énantiomérique est encore largement répandu et sera l'indice principalement utilisé dans cette thèse dans un soucis de comparaison.

Sélectivité :

Lors d'un dédoublement cinétique d'énantiomères, les excès énantiomérique du substrat et du produit sont dépendants du taux de conversion. Il a alors été introduit Le facteur de **sélectivité S**, également noté **ratio énantiomérique E,** qui est indépendant du taux de conversion.[196] Ce facteur peut être calculé à un instant *t* grâce au taux de conversion (*C*, compris entre 0 et 1) et l'excès énantiomérique ee_s des substrats n'ayant pas réagi ou des produits ee_P, grâce à la relation (pour une réaction irréversible) :

$$E = S = \frac{\ln[1 - C(1 + ee_p)]}{\ln[1 - C(1 - ee_p)]} = \frac{\ln[(1 - C)(1 - ee_s)]}{\ln[(1 - C)(1 + ee_s)]}$$

En pratique, un ratio inférieur à 15 est considéré comme mauvais, il devient bon entre 15 et 30 et il devient excellent au-delà.[197]

Chapitre I - La réaffectation enzymatique

I.1 L'évolution dirigée d'enzymes : une approche darwinienne

La sélection naturelle, théorie développée par Darwin dans son livre « On the Origin of Species »[7], expliquait l'évolution des espèces animales par la survie, au sein d'une population donnée, du représentant le plus adapté à un contexte extérieur donné, et la transmission de ses traits sélectifs à sa descendance. Le moteur de cette évolution, point essentiel avancé par Darwin, est l'apparition d'une diversité suivie d'un processus de sélection. Cette théorie fut plus tard enrichie de la découverte de la génétique puis du lien entre ADN et protéines, qui permirent d'expliquer la relation génotype-phénotype. Le point central de ce mécanisme d'évolution, est le lien créé entre le phénotype (l'observable sélectif, qui permet une sélection), et le génotype (l'encodage de cet observable, qui permet la conservation de l'information sélectionnée et son amplification).

De cette compréhension des mécanismes moléculaires de l'évolution est né le concept d'évolution dirigée des protéines, sujet qui sera abordé lors de cette première partie. L'évolution dirigée résume l'évolution naturelle en gardant, d'une manière ou d'une autre, les principes de génotype et phénotype, de sélection et amplification. Mimer l'évolution nécessite donc trois prérequis (qui s'inscrivent dans un schéma général décrit dans la figure 4) : (1) Conserver un lien entre génotype et phénotype. Cette problématique est généralement abordée avec des méthodes d'encapsulation et d'exposition (« display »). L'information génétique est compartimentée dans une structure, qu'il s'agisse d'une cellule, d'un virus ou une compartimentation artificielle (microgouttelette, puits de microplaque). La protéine codée par cette information génétique est, soit exposée à la surface du compartiment, soit contenue à l'intérieur. Deux méthodes de réduction extrême du lien phénotype/génotype sont le *ribosome display* et le *mRNA display*, où la protéine est directement liée covalemment au gène qui la code, par l'intermédiaire du ribosome l'ayant synthétisée. (2) Il est ensuite nécessaire de pouvoir générer de la diversité, pour l'obtention d'une bibliothèque (ou banque) de gènes codant chacun pour un variant différent de la protéine de départ. (3) Enfin, il faut disposer d'une méthode permettant de discriminer et favoriser les individus présentant des caractéristiques désirées. Cette discrimination peut se faire par une méthode de sélection ou une méthode de criblage, termes qui seront définis ci-dessous. Ces trois composantes s'inscrivent dans un schéma cyclique, fait de plusieurs itérations où la

rigueur des étapes de diversification et de discrimination peut être variée (ou même nulle).

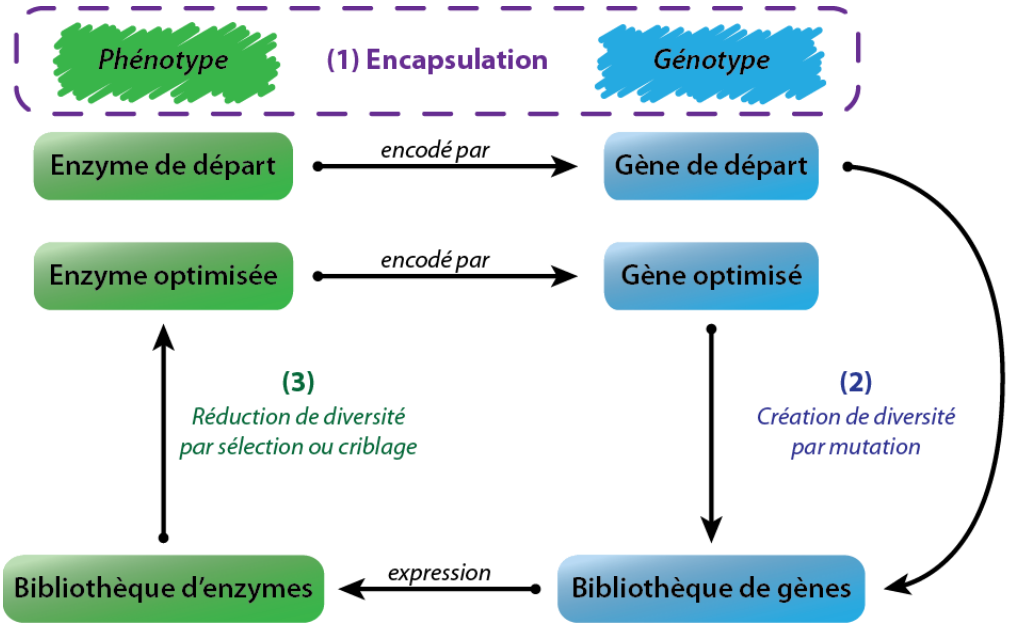


Figure 4 : Schéma canonique d'une stratégie d'évolution dirigée d'une protéine qui nécessite (1) un lien génotype-phénotype, (2) une méthode de création de diversité génétique et (3) une méthode de discrimination des variants intéressants.

Le point préliminaire de cette stratégie est le choix d'une protéine – ou enzyme – de départ, pour ensuite « naviguer » intelligemment dans l'espace des séquences génétiques à la recherche du phénotype désiré.

I.2 « Voyage » dans l'espace des séquences

Afin d'exploiter les principes d'évolution en vue de la création de nouvelles enzymes, il est nécessaire de comprendre comment une protéine peut acquérir une fonction catalytique. En effet, pour une protéine de 100 acides aminés par exemple, étant donné l'alphabet de 20 d'acides aminés disponibles, l'espace des séquences possibles est de 20^{100}, soit 10^{130} séquences différentes soit 10^{50} fois plus de possibilités que d'atomes dans l'univers. Il est évidemment impossible que toutes ces possibilités soient parcourues et sélectionnées en un temps raisonnable. Autrement dit, l'émergence de nouvelles fonctions pour une protéine ne peut passer par l'exploration exhaustive de l'espace des séquences. Ce problème du nombre est central dans le design d'enzymes et nous verrons dans cette partie les différentes stratégies pour naviguer intelligemment dans l'espace des séquences (« *beat the number problem* » selon l'expression de Manfred T. Reetz [8]).

L'étude structurale des protéines nous donne un premier indice percutant sur les voies de l'évolution naturelle. Les différents types de repliements de protéines selon la nomenclature SCOP, les « folds », sont définis par des mêmes structures secondaires majeures, dans le même ordre et avec les mêmes arrangements topologiques. On peut observer sur le graphique (figure 5) que, malgré une croissance quasi-exponentielle du nombre de nouvelles structures protéiques résolues depuis les années 70, la découverte de nouveaux « folds » a cessé depuis une dizaine d'année, se stabilisant autour de seulement 1400 sur 120 000 structures résolues à ce jour. Une autre classification, CATH, montre un résultat similaire.

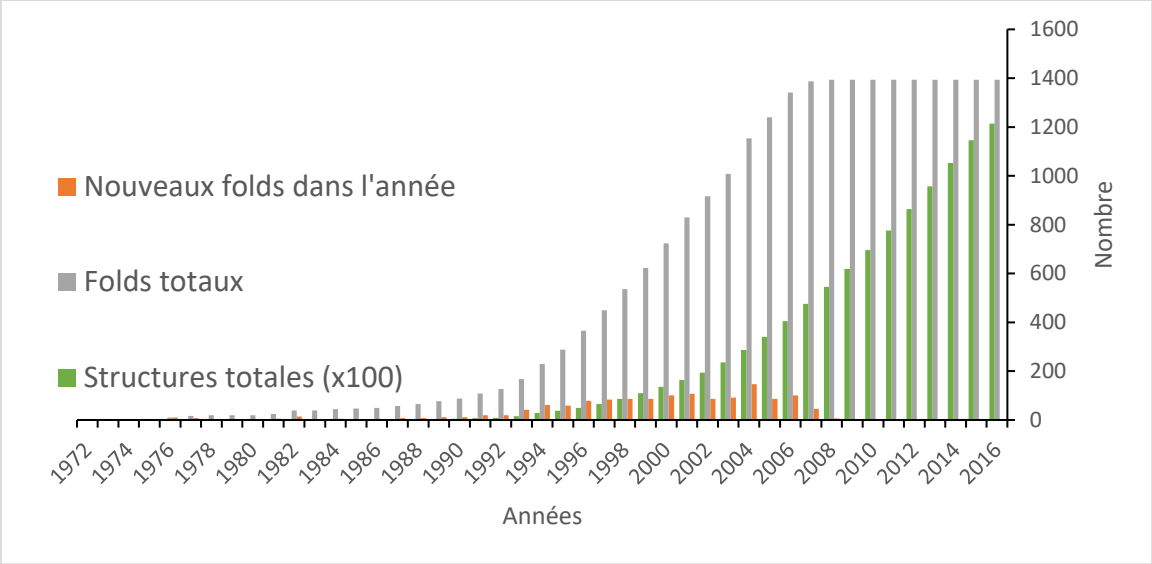


Figure 5 : Découverte de nouveaux « folds » dans la « Protein Data Bank » selon la classification SCOP. Le nombre total de folds différents stagne autour de 1400 depuis 2008 malgré une croissance exponentielle de nouvelles structures résolues chaque année.

Ce nombre de « folds » incroyablement bas au regard du nombre de structures connues de protéines aux fonctions extrêmement variées (1%) nous indique que de nombreuses fonctions peuvent être adaptées à partir d'une architecture donnée. Ceci réduit drastiquement l'espace des séquences nécessaire à l'obtention d'une protéine fonctionnelle. Cela est vrai pour des propriétés d'interaction protéine-protéine, protéine-ligand, mais également pour des fonctions de catalyseur. Par ailleurs, une même fonction peut être effectuée par deux protéines différentes: ainsi un même type d'activité catalytique (représenté par un numéro *EC*) est présent en moyenne dans deux familles structurales indépendantes.[9] Les chloroperoxidases (EC:1.11.1.10) en sont un exemple, représentées par trois familles de protéines évolutivement distinctes et au mécanisme cataytique différent. Il est donc évident que l'exploration de l'espace entier des séquences n'est pas nécessaire pour obtenir un catalyseur satisfaisant. Pourtant, si des propriétés équivalentes peuvent être obtenues en deux points de l'espace des séquences, leur

émergence peut se faire au sacrifice de la découverte du meilleur catalyseur théorique. La question se résume donc à savoir à quoi ressemble le « paysage adaptatif » des protéines: si l'on représente les différentes séquences possibles par une surface dont deux points contigus sont séparés d'une seule mutation (ce qui est bien sûr impossible car l'espace des séquences d'une protéine à N acide aminés est à N dimensions), et que l'altitude de chaque point de cette surface représente la performance dans la fonction recherchée, à quel type de paysage peut-on s'attendre ?

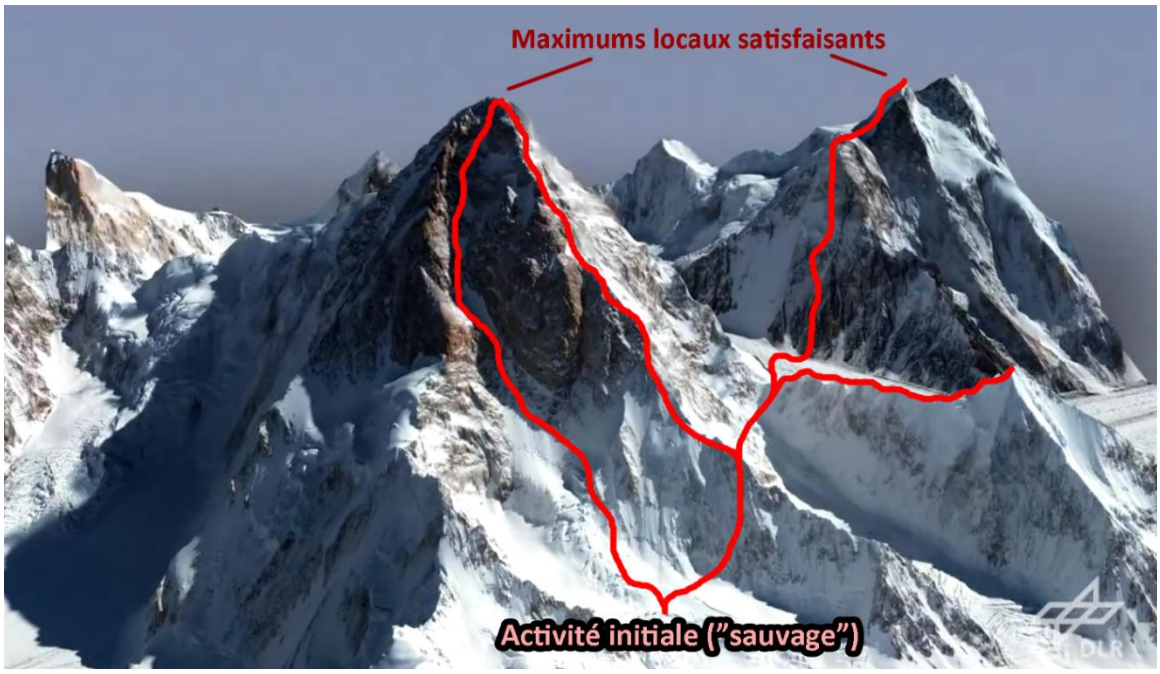


Figure 6 : Exploration du paysage adaptatif des séquences. Si l'altitude représente la fonction désirée (activité catalytique, sélectivité ...) et les coordonnées latitude/longitude représentent les séquences protéiques, plusieurs chemins évolutifs peuvent amener au même maximum local, ou un autre tout aussi satisfaisant.

D'après les observations phylogénétiques et structurales exposées ci-dessus, on obtient un paysage varié basé sur deux postulats: 1/ Plusieurs structures peuvent avoir la même activité : il y a plusieurs « sommets » ; 2/ De nombreuses séquences relativement proches donnent une structure et une activité comparable : les « pentes » ne semblent pas trop abruptes, voir sont parfois plates.

Deux paramètres clés entrent en jeu dans la forme de ce paysage : l'additivité et l'épistasie (figure 7) et agit plus précisément dans chaque « pas » effectué dans ce paysage (figure 8). La pure **additivité** des mutations sur le phénotype implique une augmentation graduelle du phénotype recherché au cours des mutations (figure 7a). Dans ce scénario, toutes les mutations bénéfiques sont indépendantes et se côtoient dans l'espace des séquences pour converger en un point d'adaptation maximale. Ce comportement a été observé pour des petites portions du paysage adaptatif,[10] et est largement exploité dans

les stratégies d'évolution dirigée. L'**épistasie** est un terme qui désigne en génétique l'interdépendance de plusieurs gènes. Ici, dans l'espace des séquences, l'épistasie vient de l'interdépendance des mutations entre elles: un acide aminé à une position donnée ne pourra être optimal qu'en présence d'un autre acide aminé ailleurs dans la séquence. L'épistasie explique la « rugosité » du paysage évolutif des protéines (figure 7b). Les relations épistatiques, c'est-à-dire l'interdépendance des mutations sélectionnées, sont la source de difficulté principale dans la conception de banques de protéines pour l'évolution dirigée, car elles peuvent agir à longue distance, être cryptiques et invasives. [11] C'est-à-dire qu'une mutation spécifique pourra être indispensable dans une portion du paysage évolutif ne conduisant pas à un maximum local satisfaisant, conduisant à un cul-de-sac évolutif.

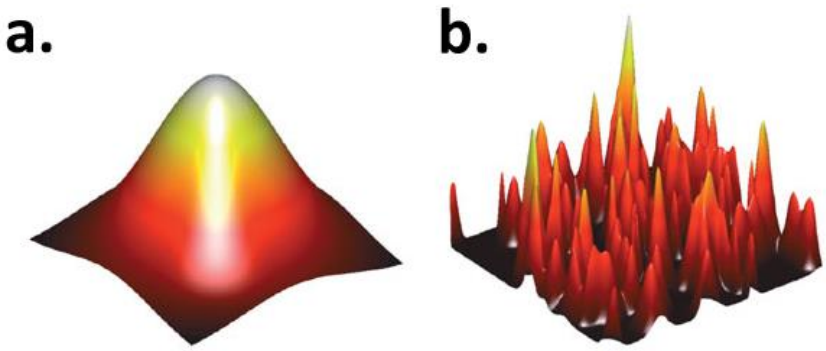


Figure 7 : Deux exemples extrêmes de ce à quoi peut ressembler un paysage adaptatif (a) Paysage adaptatif sans épistasie : chaque mutation a une faible influence, un seul maximum global. (b) Paysage adaptatif avec une forte épistasie : chaque changement a une forte influence, présence de maximums locaux. Selon [12]

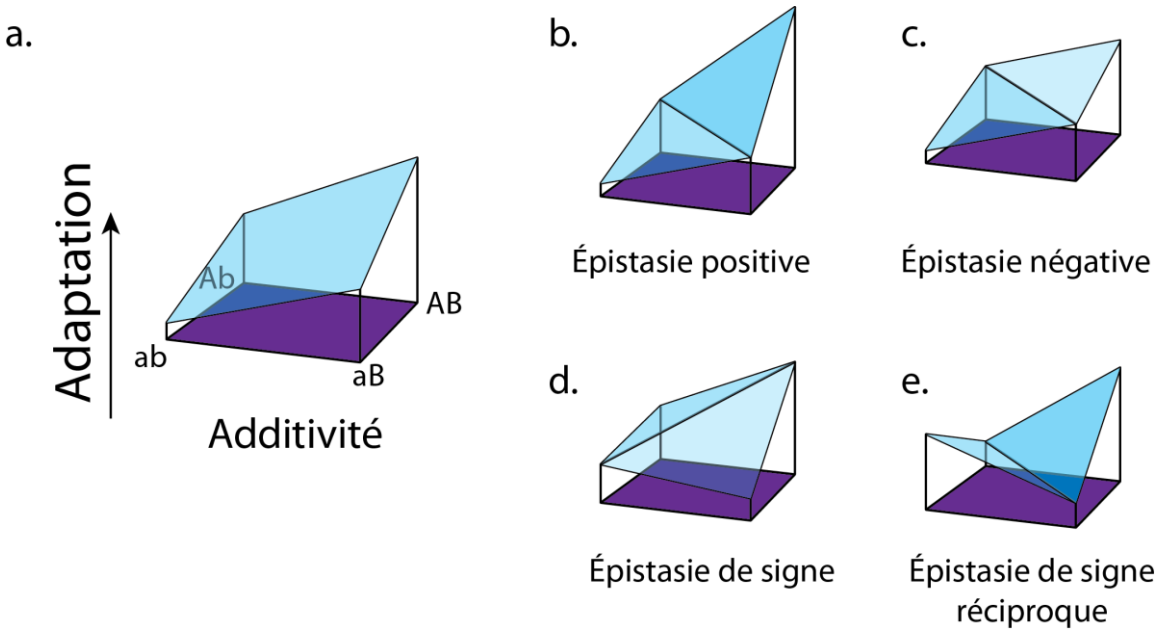


Figure 8 : Différents types d'épistasie formant chaque « pas » dans le paysage adaptatif chaque coin du carré violet représente l'état de deux mutations A et B, la hauteur de la surface bleue représente l'avantage adaptatif résultant. ***a.*** *Additivité simple, pas d'épistasie entre les deux positions.* ***b.*** *Épistasie de magnitude positive, l'activité du double mutant est plus importante que l'addition des deux.* ***c.*** *Épistasie de magnitude négative, l'activité du double mutant est plus faible l'addition des deux.* ***d.*** *Épistasie de signe, la mutation B est délétère sans la mutation A, mais la mutation A ne dépend pas de la mutation B.* ***e.*** *Épistasie de signe réciproque, les mutations A et B sont délétères si elles ne sont pas concomitantes, mais bénéfiques si elles le sont.*

Afin de naviguer efficacement dans l'espace des séquences, le plus simple est de trouver une enzyme présentant en partie la nouvelle fonction catalytique recherchée pour ensuite l'améliorer : c'est la promiscuité enzymatique.

I.3 La promiscuité enzymatique, un point de départ pour l'ingénierie

Bien que les enzymes soient des catalyseurs d'une remarquable spécificité, il est commun qu'elles présentent d'autres propriétés catalytiques, en conditions non physiologiques. On parle alors de promiscuité enzymatique. Ce terme ambigu regroupe deux définitions spécifiques : il peut s'agir d'une promiscuité de substrat (une enzyme catalysant le même type de réaction sur deux substrats différents) ou d'une promiscuité de réaction (deux réactions différentes à partir du même substrat). Cette promiscuité peut provenir soit du fait que l'activité secondaire n'a pas été sélectionnée négativement au cours de l'évolution, soit parce que la protéine n'est pas suffisamment spécifique pour discriminer les deux substrats. La ribulose Bis-phosphate Carboxylase Oxygenase (RuBisCO), enzyme responsable de la fixation du CO_2 par les organismes photosynthétiques, est un parfait exemple de promiscuité de réaction et de substrat. Son activité principale est la carboxylation du ribulose bis-phosphate par le CO_2 et son activité secondaire, l'oxydation de ce même substrat par une molécule d'O_2. Malgré l'effet très

délétère de cette réaction secondaire sur le rendement de fixation du CO_2 par les végétaux, il n'est pas possible d'améliorer la spécificité sans baisser la vitesse de la réaction, et la nature a dû mettre en place des mécanismes très couteux en énergie pour répondre à ce problème.[13]

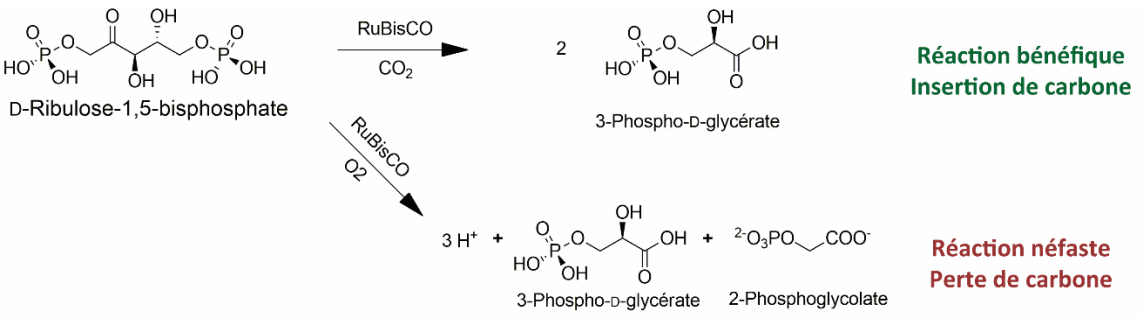


Figure 9 : Promiscuité enzymatique de la RuBisCO a) réaction bénéfique conduisant à l'incorporation de carbone dans la plante b) réaction néfaste, conduisant à un cycle appelé photorespiration nécessaire à la régénération des composés, provoquant une perte nette en carbone incorporé par la photosynthèse.

La promiscuité enzymatique est le point de départ de la réaffectation d'une enzyme vers une autre activité, tant dans la nature qu'au laboratoire. Une fois qu'une activité catalytique, aussi faible soit-elle, a été observée dans une protéine, il est *a priori* possible de l'améliorer. La question se pose alors de la méthode à employer pour créer suffisamment de diversité pour explorer le paysage adaptatif de la molécule pour la réaction désirée. Nous verrons ici les différentes méthodes de création de diversité génétique, classées en termes d'efficacité croissante. Le terme efficacité se réfère ici à la réduction de la taille de la banque de gènes nécessaire pour obtenir un catalyseur désiré, éliminant au maximum les protéines peu intéressantes. Nous aborderons ensuite différents moyens utilisés pour discriminer les variants d'une banque, et les difficultés que pose l'amélioration de l'énantiosélectivité. Nous verrons qu'une banque plus efficace nécessite également plus de connaissances structurales et fonctionnelles sur l'objet à faire évoluer.

I.4 La génération de diversité : vers des banques de plus en plus rationnelles

Différentes méthodes de génération de diversité peuvent être utilisées pour la génération d'enzymes stéréosélectives. Celles-ci seront présentées en s'appuyant sur l'exemple de l'évolution dirigée d'une lipase, qui a bénéficié d'une étude comparative entre les différentes méthodes par l'équipe du Pr. Manfred T. Reetz pour l'optimisation du dédoublement cinétique d'un mélange racémique de p-nitrophenyl2-methyldecanoate (figure 10).

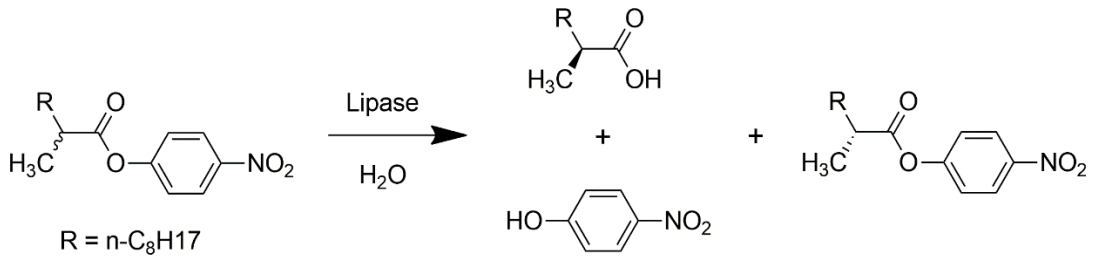
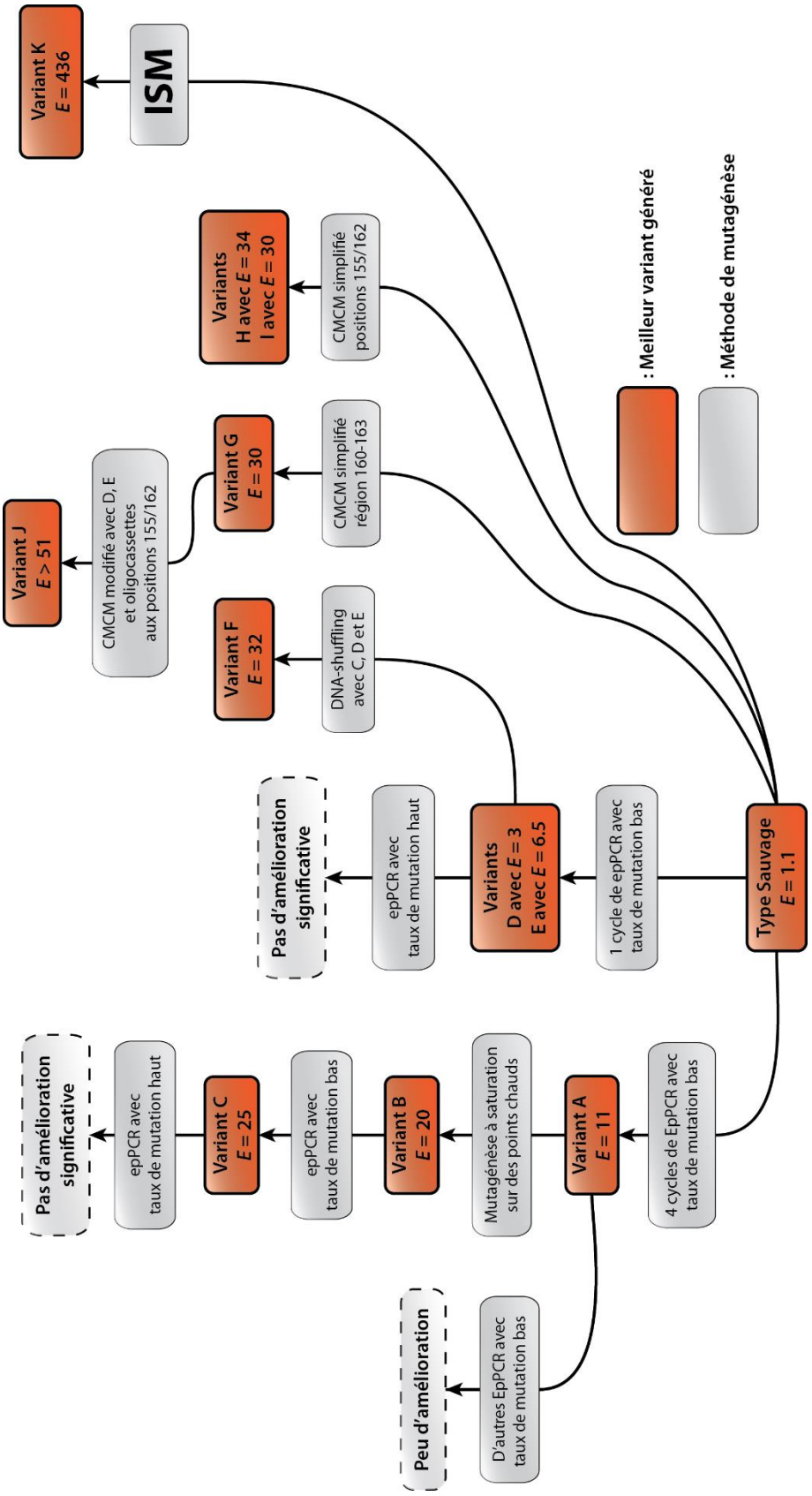


Figure 10 : Réaction standard énantiosélective (ici en faveur de l'hydrolyse de l'énantiomère (R)) étudiée par l'équipe de Reetz, la sélectivité peut être étudiée par la réalisation de deux cinétiques séparées d'hydrolyse d'un mélange énantiopur du substrat. La réaction est suivie grâce à l'absorbance à 405 nm du p-Nitrophénol.

Ce substrat libère lors de son hydrolyse du *p*-Nitrophénol dont la forme phénolate absorbe à 405 nm. L'excès énantiomérique et le facteur E sont calculés grâce une cinétique d'apparition du produit lors de deux expériences jumelées. Dans un puits, la réaction est effectuée sur l'un des deux énantiomères du substrat, l'autre énantiomère étant utilisé dans le puits adjacent. Ce mode opératoire permet le criblage de 48 mutants différents par plaque, il est ainsi possible de comparer les stratégies de mutagénèse. Ces différentes stratégies sont résumées dans la figure 11, où sont représentées par des étiquettes oranges les différents variants obtenus et leur sélectivité, et par des étiquettes grises les différentes stratégies de mutagénèse ayant permis d'obtenir lesdits variants.

Figure 11 : Evolution de la lipase selon différentes méthodes d'évolution dirigée. Les étiquettes grises représentent les méthodes de mutagénèse utilisées pour constituer les différentes banques de protéines. Les étiquettes oranges représentent les meilleurs catalyseurs obtenus pour la réaction d'hydrolyse énantiosélective présentée en figure 10. Selon[14]



I.4.1 Les méthodes non ciblées : mutations sur la totalité du gène

Les méthodes non ciblées nécessitent le moins de connaissances préalables pour la génération de banques géniques, et permettent d'introduire des mutations au hasard sur la totalité du gène. Un autre avantage est de pouvoir trouver des mutations que l'on n'aurait pas pu prédire du fait d'effets épistatiques à longue distance. Cependant cette approche possède deux inconvénients : la fraction utile des mutations est très faible, certaines positions fixées au cours de l'évolution étant très délétères pour l'intégrité structurale de la protéine si mutées, ou d'autres positions étant tout à fait neutres. De plus, comme vu plus haut, la combinatoire des séquences mutées aléatoirement sur toute la longueur d'un gène étant gigantesque, l'exploration de l'espace des séquences ne peut être que très partielle.[15]

I.4.1.1 La mutagénèse aléatoire

La mutagénèse aléatoire est la méthode la moins ciblée de toutes, mais également la plus facile à mettre en place. La technique *in vitro* de base utilise d'une version modifiée de la réaction en chaine par polymérase (PCR), elle est désignée par son acronyme anglais EpPCR pour Error prone PCR, le but étant de réduire la fidélité de la polymérase afin d'introduire des erreurs avec un taux contrôlé (généralement 1 à 10 mutations par 1000 pb). La polymérase est généralement rendue moins fidèle par l'ajout de Mn^{2+} au lieux de Mg^{2+} dans le milieu réactionnel [16]. D'autres techniques existent également comme l'utilisation de souches de *E. coli* possédant une polymérase infidèle [17], la Sequence Saturation Mutagenesis (SeSaM) [18], l'insertion/deletion aleatoire (RID) [19] ou d'autres encore [20,21], mais nous nous focaliseront sur l'EpPCR. Cette dernière est celle encore actuellement la plus utilisée dans la construction de banques géniques naïves pour sa simplicité bien qu'elle possède deux biais majeurs : i/ La fréquence de permutation est dépendante du type de base, les bases A-T étant généralement plus sujettes aux mutations que les G-C. Des kits d'enzymes spécialement optimisés promettent une diminution de ce biais. ii/ La dégénérescence du code génétique favorise l'apparition des acides aminés les plus représentés.

La première utilisation de l'EpPCR pour l'évolution dirigée d'une enzyme date du début des années 90 avec les travaux pionniers de Frances Arnold sur la subtilisine E [22–24]. Ces travaux visaient à augmenter l'activité de l'enzyme en présence de diméthylformamide (DMF). La méthode de criblage employée était visuelle : les protéines étaient sécrétées par *B. subtilis* sur des boites de pétri contenant des concentrations variables DMF et de la caséine, la taille du halo de dégradation de la caséine autour de la colonie reflétant l'activité de la protéine. Les gènes présents dans les colonies présentant un plus grand halo que le contrôle étaient utilisés pour un criblage d'activité en tube contre le substrat

chromogénique de protéase s-AAPF-pNA. Le meilleur candidat était utilisé pour nouveau cycle de mutagénèse. Après plusieurs tours de mutagénèse et sélection successifs, des enzymes portant 3, 10, puis 13 acides aminés changés ont été trouvées et caractérisées. Ces variant possédaient une efficacité catalytique supérieure à celle de la protéine native dans un solvant purement aqueux, et la conservaient (et même l'augmentaient !) en présence de moins de 40% de DMF. Il a été décrit que ces améliorations étaient dues à la baisse de l'affinité apparente pour le substrat modèle, ainsi qu'à une plus grande stabilité de l'enzyme modifiée en présence de DMF. Il est intéressant de noter que toutes ces mutations se situent sur la surface de la protéine, la plupart étant dans la poche de fixation du substrat ou autour du site catalytique et qu'un grand nombre sont également présentes dans d'autres subtilisines. *A contrario*, dans une étude cherchant à augmenter la thermostabilité de l'enzyme, les mutations stabilisantes se trouvaient à longue distance du site catalytique. Cependant F. Arnold ne s'est pas intéressée à la stéréosélectivité dans ces études.

C'est en 1997 que Reetz publia la première optimisation de l'énantiosélectivité de la lipase de *P. Aeruginosa* envers le p-nitrophenyl2-methyldecanoate par mutagénèse aléatoire. Les paramètres de l'EpPCR étaient empiriquement mis au point de manière à n'avoir qu'une seule mutation en moyenne par gène. Le premier postulat de cette approche était que les mutations étaient essentiellement additives. Le second était que la présence d'un trop grand nombre de mutations sur le même gène pouvait cacher une mutation bénéfique à cause d'une deuxième mutation très délétère. La taille de la banque pour une substitution d'un acide aminé par un des 19 autres disponibles dans une protéine est donnée par l'équation :

$$N = \frac{19^M X!}{[(X-M)!\,M!]}$$

X étant la taille de la protéine (ici 285 aa) et M le nombre de mutations par protéine (ici 1). Le nombre de protéines différentes est donc de 5415. On peut noter qu'avec un taux de mutation de 2 on passe à 15.10^6 protéines différentes, et 52.10^9 pour seulement 3 mutations, agrandissant considérablement le criblage nécessaire. Quatre tours de sélection de seulement 1000 à 2400 clones de cette banque (8000 au total) suffirent à trouver des candidats passant de 2% d'excès énantiomérique pour le type sauvage à 81% pour le type évolué (Variant A, figure 11), correspondant à des ratios énantiomériques de 1,1 et 11,3 respectivement. Bien que remarquable pour seulement 4 tours de sélection avec si peu de clones criblés, un ratio énantiomérique de 11,3 est cependant insuffisant pour une utilisation pratique. Une tentative d'amélioration par d'autres cycles à un taux de mutation bas ne s'avéra pas concluante. Une approche par EpPCR à haut taux de mutation donna après criblage de 15 000 variants, deux clones à l'énantiosélectivité faible

(Variant D (*E* = 3) et E (*E* = 6,5), figure 11), confirmant le premier postulat selon lequel un nombre plus élevé de mutations augmente drastiquement la fraction non utile (possiblement instable) des variants de la banque.

Bien que relativement peu efficace, l'approche de mutagénèse aléatoire peut permettre de déterminer les positions intéressantes à faire varier (les « hot spots » de mutation). C'est également un moyen pour créer de la diversité génétique pour un autre type d'approche non ciblé : les méthodes par recombinaison de gènes.

I.4.1.2 Les méthodes par recombinaison de gènes

Les méthodes de recombinaison se basent sur le fait que les acides aminés importants pour l'intégrité de la structure d'une protéine sont conservés au cours de l'évolution (naturelle ou artificielle). L'alignement de différentes séquences homologues codant pour des enzymes équivalentes chez des espèces différentes, fait apparaitre des points chauds et des points froids de mutation dans la séquence d'un gène. Une analyse de la séquence « consensus » issue de ces alignements permet de recombiner ces séquences, en s'assurant de la pertinence des mutations, contrairement aux méthodes de mutagénèse aléatoire décrites précédemment. La première méthode appelée « gene shuffling » ou « PCR sexuelle », introduite par Willem P. C. Stemmer dans les années 1990, utilise la PCR sur des gènes fragmentés, se recombinant par homologie durant la phase d'hybridation. [25] Cette méthode induit cependant un biais en faveur de la recombinaison des fragments les plus homologues, réduisant la diversité créée. Nombre de variantes ont été imaginées pour palier à ce problème, avec ou sans PCR.[26–35] Dans le cas de la lipase de *P. Aeruginosa* de l'étude de Reetz, la recombinaison entre les variants sélectionnés au sein d'une banque générée par des méthodes de PCR mutagène à bas taux d'erreur s'est soldée par un échec. Une recombinaison entre ces mêmes enzymes et les autres variants peu performants provenant de la banque à haut taux de mutation citée plus haut a cette fois permis de passer d'un facteur *E* de 25 avant recombinaison à 32 après recombinaison puis sélection (variants C et F respectivement, figure 11). Bien que peu performants, les variants provenant de la banque à haut taux d'erreur possédaient donc tout de même des mutations bénéfiques pour la sélectivité.

I.4.2 Les méthodes à alea ciblé

Les méthodes ciblées, qui consistent à choisir précisément les positions qui seront mutées, nécessitent une connaissance plus profonde de l'objet à modifier. Les points chauds de mutations peuvent être déterminés par alignement de séquences natives comme des séquences des variants provenant de banques aléatoires. Les positions

peuvent également être sélectionnés de manière rationnelle par inspection de la crevasse d'un site actif (CAST) par exemple ou par des méthodes de modélisation moléculaire.

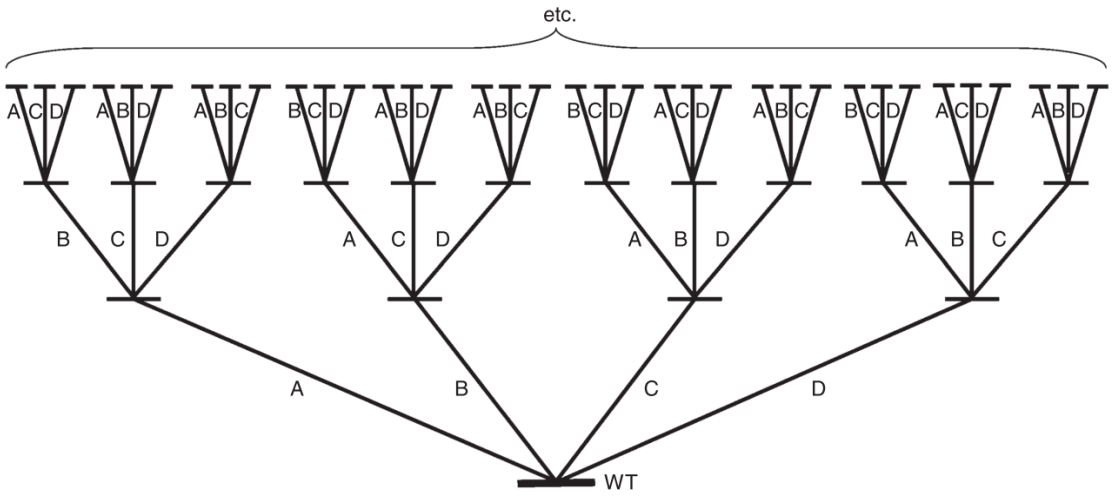
I.4.2.1 La mutagénèse combinatoire à cassettes

La méthode de mutagénèse combinatoire à multiples cassettes (MCMC) se rapproche des méthodes de recombinaison évoquées précédemment, cependant elle prend en compte la présence de points chauds de mutations, elle est donc un point intermédiaire entre les méthodes de recombinaison et les méthodes de mutagénèse ciblées. Son utilisation sur seulement deux positions (155 et 162) de la Lipase de *P.aeruginosa* a permis de découvrir des variants possédant un facteur E jusqu'à 34 (variants I et J, figure 11). Par l'utilisation de la combinaison de ce type de mutagénèse et des méthodes de recombinaison classiques, l'équipe de Reetz a obtenu des facteurs E encore meilleurs (*E* de >51: Variant J, figure 11).

I.4.2.2 La mutagénèse à saturation

La mutagénèse à saturation est employée pour muter un acide aminé ciblé de la protéine. La méthode couramment utilisée est semblable à la mutagenèse dirigée, utilisant des amorces dégénérées pour une amplification PCR circulaire. Les schémas de mutagénèse peuvent cependant être volontairement biaisés pour réduire la diversité, éviter certains acides aminés ou des codons non-sens, ou contrer la dégénérescence naturelle du code génétique. Ainsi le triplet NNK est souvent utilisé car il encode les 20 acides aminés en divisant le nombre de codons par deux et éliminant deux codons stop sur trois. Cependant différentes expériences d'évolution dirigés utilisant des alphabets très réduits d'acides aminés ont donné des résultats très satisfaisants tout en réduisant drastiquement le nombre de combinaisons possibles, prouvant qu'une très grande diversité n'était pas forcément nécessaire. [8,36–38]

Postulant que l'énantiosélectivité d'une enzyme est principalement gouvernée par les acides aminés proches du substrat dans le site actif, le principe de mutagénèse à saturation de site actif apparut (CAST), réduit considérablement l'espace des possibles. [39] L'utilisation de la méthode CAST de manière itérative sur chacune des positions choisies l'une après l'autre est nommée mutagénèse à saturation itérative (ISM, figure 12). [40] Afin d'illustrer l'efficacité de cette méthode, le groupe de Reetz a optimisé l'énantiosélectivité de la lipase de *P. aeruginosa* jusqu'à un facteur *E* de 436 avec seulement 10 000 variants testés (variant K, figure 11).[41] Elle a depuis été reprise avec succès par nombre de groupes de recherche dans l'évolution de la spécificité de substrat de différentes enzymes. [42–51] Cette méthode semble donc être une des plus efficaces dans l'exploration de l'espace des séquences pour l'amélioration de l'énantiosélectivité.


etc.
A C D
A B D
A B C
B C D
A B D
A B C
B C D
A C D
A B C
B C D
A C D
A B D
B C D
A C D
A B D
A B C
A C D
A B C
B C D
A B D
A B C
A C D
A B D
A B C
A C D
A B C
A B C
A B C

Ces analyses peuvent s'appuyer sur des **homologies de séquences ou de structures** : L'application web 3DM, conçu par l'entreprise Bio-Prodict, utilise les comparaisons de protéines d'une même famille ainsi que la littérature existante pour prédire les résidus concernés dans la sélectivité d'une enzyme, et a été utilisé avec succès dans des stratégies d'évolution dirigée.[46] Dans cette étude de Bornscheuer, les comparaisons de séquence de la famille d'une estérase ont servi à « conseiller » la mutation de quatre positions avec un alphabet spécifique pour chacune. Une comparaison d'un peu plus de 500 clones provenant de banques CAST conçues selon trois schémas de dégénérescences a été effectuée : codons NNK (tous les acides aminés possibles, graphe vert figure 14), codons « conseillés » (graphe bleu, figure 14) et « déconseillés » (graphe rouge, figure 14). Cette étude montre de façon très convaincante que le schéma déduit de la comparaison de séquences permet la découverte de plus d'enzymes, plus actives et plus énantiosélectives. En d'autres termes la banque conçue avec les acides aminés conseillés par l'alignement de séquence est plus efficace.

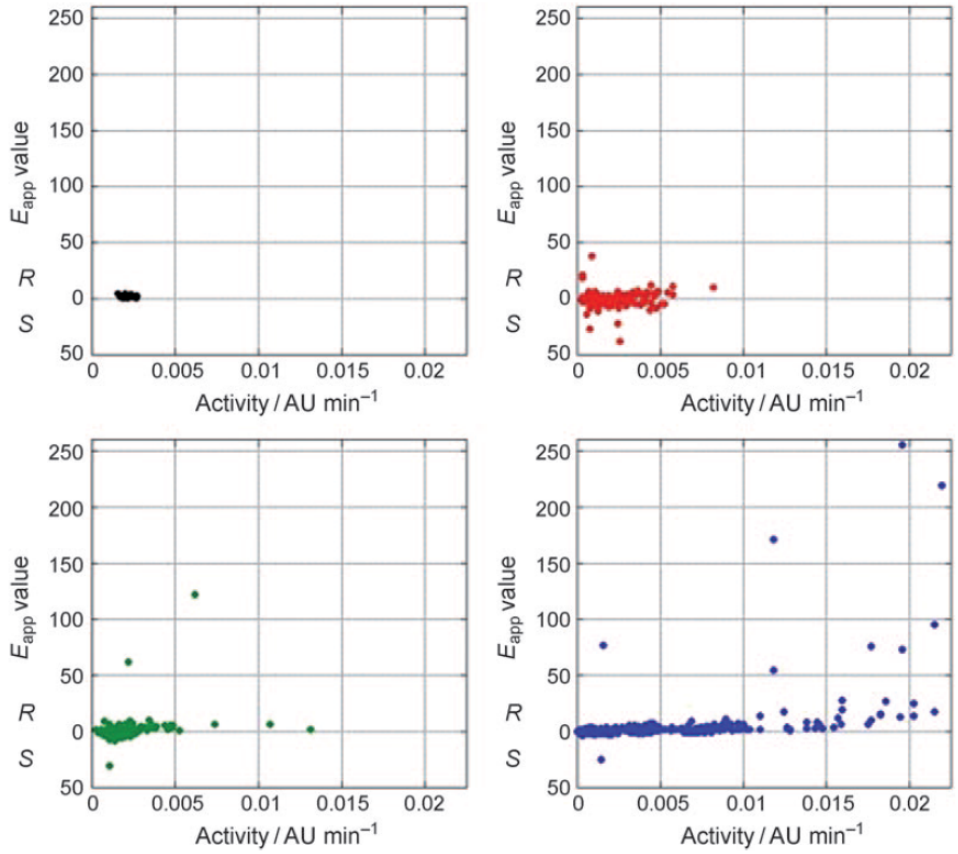


Figure 14 : Comparaison d'environ 500 variants provenant de banques CAST sur quatre acides aminés, conçues selon 3 schémas de dégénérescence en comparaison d'une collection d'enzymes sauvages (graphe noir), pour l'évolution dirigée d'une estérase de P. fluorescens pour la conversion énantiosélective de l'acide 3-phenylbutyrique. Rouge : acides aminés déconseillés, Vert : tous les acides aminés, bleu : acides aminés conseillés. On remarque que le graphe violet contient plus de protéines plus actives et également plus spécifiques.

Deux types d'approches concluantes par **design *de novo*** ont également été décrites par l'équipe de David Baker. La première s'appuie sur le logiciel de modélisation moléculaire Rosetta, et particulièrement son module RosettaMatch : Son équipe utilise une approche par « théozyme ». Un théozyme consiste en la modélisation de l'état de transition de la molécule attendue dans la conformation voulue ainsi que des chaines latérales d'acides aminés nécessaires pour activer la réaction. La base de données structurale de protéines (PDB) est ensuite scannée afin de trouver les potentielles architectures pouvant accommoder ce ligand virtuel. Ainsi concernant la cycloaddition de Diels-Alder, une architecture de « β-propeller » a été trouvée puis modifiée afin de placer judicieusement des donneurs et accepteurs de liaison hydrogène. Une activité modeste a donc été créée *de novo* avec une bonne énantiosélectivité. Elle fut ensuite améliorée par évolution dirigée, jusqu'à l'obtention de la Diels Alderase la plus active décrite à ce jour.[52] Une combinaison d'évolution dirigée et de design rationnel s'avère donc un mariage très

fructueux. Ces approches de design *de novo* se multiplient mais restent cependant pour le moment cantonnées à l'équipe de Baker, leader dans la modélisation de protéines.[53–56]

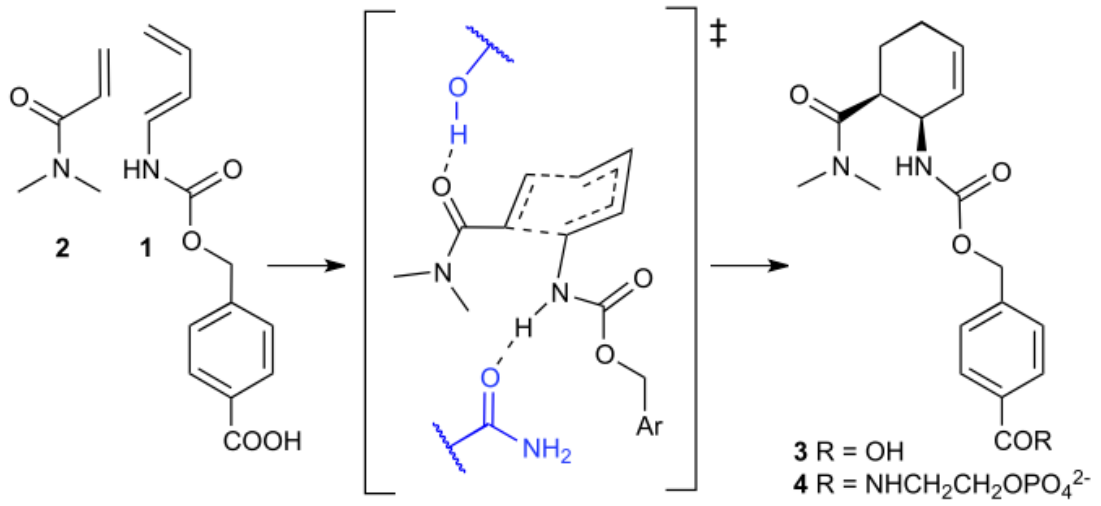


Fig. 1. Diels–Alder reaction between 4-carboxybenzyl-*trans*-1,3-butadiene-1-carbamate (**1**) and *N*,*N*-dimethylacrylamide (**2**). The theozyme for promoting formation of the 3R,4S *endo* cyclohexene product (**3**) is shown in brackets. The phosphorylated product analog (**4**) was used for inhibition and crystallization experiments.

Figure 15 : Approche de l'équipe de Baker : l'état de transition couplé à des acides aminés judicieusement placés forment un théozyme.

Une deuxième stratégie de design *in silico* est la modélisation d'un site actif d'enzyme préexistant afin d'orienter la sélectivité vers un énantiomère donné d'un substrat d'intérêt. Par l'utilisation d'une approche combinant design puis criblage *in silico*, appelée CASCO (« catalytic selectivity by computational design »), l'équipe de Baker en collaboration avec celle de Dick B. Janssen est parvenu à convertir une hydrolase de l'époxyde de limonène (LEH) pour l'hydrolyse énantiosélective de différents meso époxydes (figure 16). Le résultat de cette hydrolyse dépend entièrement de la face d'attaque de la molécule d'eau.

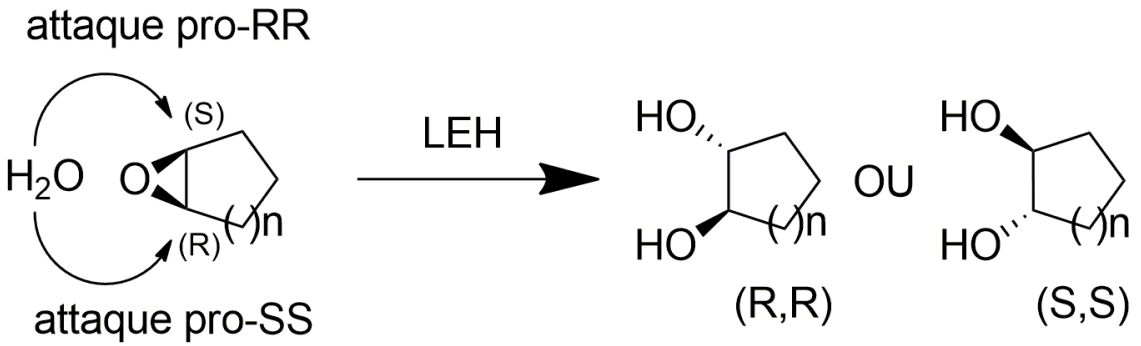


Figure 16 : Réaction catalysée par l'hydrolase d'epoxyde de limonène modifiée par l'équipe de Baker et Janssen. La configuration absolue du produit dépend de la face d'attaque de la molécule d'eau.

Dans cette étude, c'est une combinaison de rosettaDesign (mutagénèse *in silico*) se basant sur un pré-état de transition (appelé NAC pour « near attack conformation ») et d'une méthode de crible virtuel utilisant la modélisation moléculaire qui a permis l'obtention des différentes mutations nécessaires à augmenter la stéréosélectivité jusqu'à 90% et 85% envers la forme (*S,S*) et (*R,R*) du produit respectivement.

I.5 Réduction de la diversité : sélection *versus* criblage

Différents approches ont été développées pour l'ingénierie des protéines en général mais il faut distinguer les méthodes permettant de chercher des protéines pour leurs capacités d'interagir avec une cible donnée (dérivés d'anticorps et d'architecture protéiques alternatives) et celles qui sont utilisées en ingénierie enzymatique auxquelles nous nous intéresserons ici. La mise en évidence de l'énantiosélectivité d'une enzyme est un problème complexe dans l'évolution dirigée des enzymes, car il n'existe aucune méthode qui soit applicable de façon générique. Chaque cas doit donc être abordé d'une manière particulière par différents procédés, ou une combinaison de ces derniers. On peut classer les méthodes utilisées couramment en deux groupes: le criblage et la sélection. La stricte définition des deux termes n'étant pas universellement acceptée, ils sont souvent intervertis l'un et l'autre[57,58].

I.5.1 Les méthodes de réduction de la diversité par sélection

La sélection consiste à réduire la diversité de la population sous une pression de sélection en assurant toujours un lien physique entre le phénotype sélectionné et son génotype.

Parmi les approches développées en biotechnologie, on distingue i/ des approches de « display » ou exposition des protéines à la surface de cellules ou de bactériophages, et ii/ des approches de compartimentation dans des cellules vivantes. Il est ensuite possible de trier les cellules en fonction des propriétés recherchées ou bien d'exploiter la capacité de survie des cellules dans les conditions de sélection.

I.5.1.1 Sélection par affinité : « bio-panning »

Les méthodes de « bio-panning » utilisent l'affinité des protéines pour un ligand comme pression de sélection. Les protéines de la banque sont liées physiquement à leur génome et affichées à l'extérieur, à la surface de phages par exemple. Dans ce type de sélection, le ligand cible est fixé sur un support physique tel que le fond d'un puits de microplaque ou des billes magnétiques par exemple. Les protéines de la banque interagissant avec le ligand resteront alors sur le support solide, résistant aux étapes de lavage. Les candidats retenus sont enfin élués et amplifiés (grâce à l'infection de bactéries par les phages dans le cas de la figure 17).

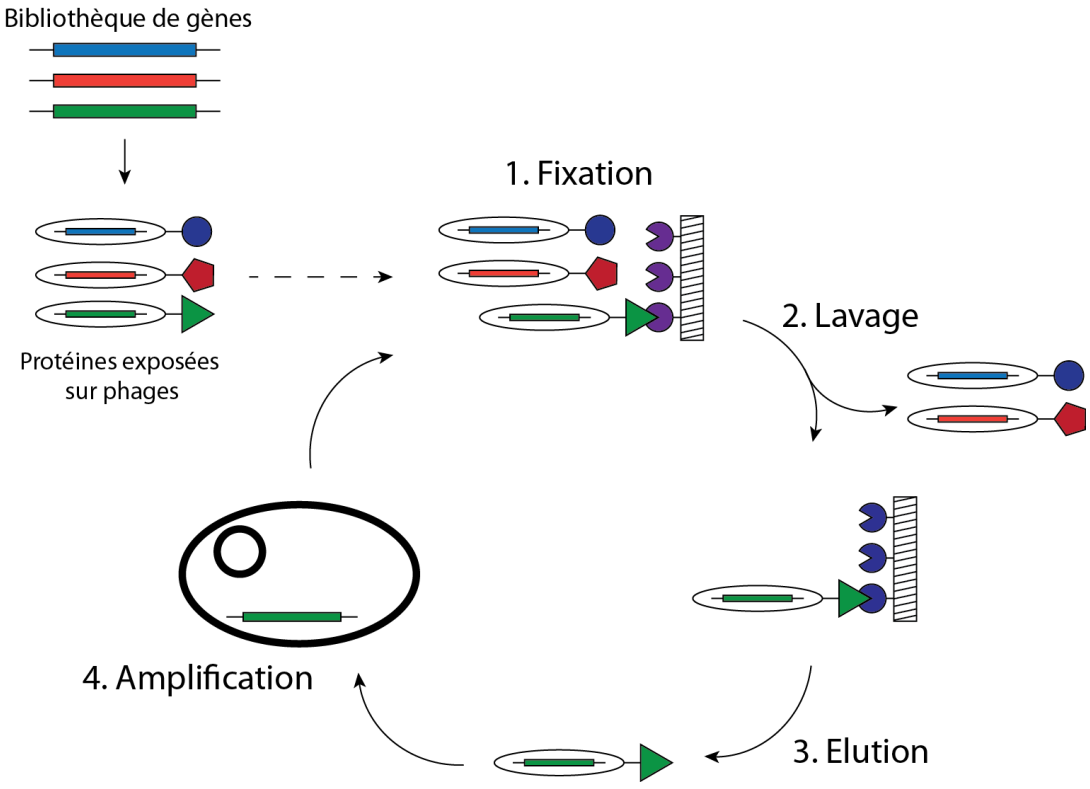


Figure 17 : Schéma général d'une expérience de bio-panning (ici le phage display). Les protéines codées par la banque de gènes sont exposées à la surface de bactériophages, chacun encapsulant le gène codant pour la protéine exposée à sa surface. 1. Les phages sont incubés en présence de la cible immobilisée au fond du tube de sélection. 2. Les protéines ayant la capacité de se fixer à la cible résistent à l'étape de lavage 3. Les protéines affines sont ensuite éluées d'une manière sélective ou non. 4. Les variants sélectionnés sont ensuite amplifiés grâce au gène qu'ils portent (ici, par infection de bactérie) et un nouveau cycle peut être effectué.

Dans une étude de 2001, le groupe de Jacques Fastrez a mis à profit cette méthode pour la sélection de métalloenzymes dépendantes du zinc.[59] Le principe reposait sur trois étapes : i/ L'inactivation des enzymes potentielles dans la banque de phages par séquestration du zinc, ii/ sélection des protéines affichées selon leur affinité pour le substrat cible, puis iii/ élution par ajout de zinc, seules les protéines ayant une activité catalytique induite par le zinc pouvant cliver le substrat et se libérer. La méthode a été testée avec la métallo β-lactamase de *B. cereus* affichée sur un phage filamenteux. Il a été montré que cette méthode de sélection enrichissait la fraction de protéine sauvage par rapport à un mutant inactif. Le groupe a également tenté d'améliorer l'activité de cette enzyme en utilisant une banque de protéines conçue par PCR mutagène, les résultats étant une discrimination des mutants peu actifs ou mal conformés et une sélection de variants d'activité similaire ou supérieure.

I.5.1.2 Sélection par survie

La survie d'individus parmi une population sous une pression de sélection commune est le mode de sélection le plus proche de la sélection naturelle (la survie du plus adapté) et a l'avantage de pouvoir sélectionner des variants parmi une très grande banque de départ, car seuls les clones positifs se détachent du bruit de fond. Dans l'étude pionnière conduite en 1994 par Stemmer expliquant la méthode de « gene shuffling » abordée plus haut, l'évolution dirigée par sélection a été utilisée pour l'amélioration de la β-lactamase TEM-1 d'*E. coli*.[25] Dans cette étude, les cellules étaient sélectionnées sur boite de pétri selon leur résistance à un antibiotique de type β-lactame, la cefotaxime : seules les bactéries résistant à la concentration d'antibiotique ajouté pouvaient se multiplier, permettant leur sélection. Trois cycles de *shuffling* et deux cycles de rétro-croisement avec le gène sauvage (pour éliminer les mutations non essentielles) ont été effectués avec une sélection sur un milieu contenant une quantité croissante d'antibiotique. Les variants sélectionnés conféraient aux bactéries une résistance à une concentration 32000 fois plus élevée d'antibiotique. Cet exemple illustre parfaitement la simplicité et l'efficacité des approches de sélection par survie.

Si on l'applique à la recherche d'activités enzymatiques énantiosélectives, cette approche possède néanmoins certaines limitations: Tout d'abord, il faut que les substrats de la réaction ne soient pas trop toxiques, ni que les conditions de réaction recherchées soient trop délétères pour la cellule, sans cela aucune sélection ne sera possible (ou la sélection d'une enzyme dégradant le substrat ou le produit risque d'être effectuée). Ensuite, les réactifs et produits doivent être raisonnablement solubles afin de pouvoir accéder à l'enzyme et suffisamment stables pour rester actifs le temps de la sélection. Enfin, il est

nécessaire de disposer d'une molécule conférant un avantage sélectif sous la forme de l'énantiomère ciblé.

Cette dernière condition est évidemment centrale dans le principe de sélection. Concernant la problématique d'énantiosélectivité, la plupart des solutions ont été trouvées dans le dédoublement cinétique d'énantiomères par des moyens ingénieux de «pseudo- racémiques»[57].

Dans un mélange pseudo-racémique, deux molécules portent l'atome chiral relié soit à une molécule source de carbone (pour l'énantiomère souhaité) soit à un poison (pour l'énantiomère non souhaité). La réaction désirée doit consister à cliver ces molécules, provoquant la libération de la source de carbone ou du poison. Ce faisant, seule la cellule ayant une enzyme présentant l'activité souhaitée provoquera un atout sélectif qui sera sélectionné sur boite de pétri. Peu d'exemples sont actuellement disponibles concernant cette stratégie.

La première utilisation de la sélection génétique pour l'amélioration de l'énantiosélectivité a été publiée en 2008 par l'équipe de W. J.Quax.[60] Elle consiste en la réversion de l'énantiopréférence de la Lipase A (LipA) de *B. subtilis* dans l'hydrolyse de l'ester butanoate de 1,2-O-isopropylidene-sn-glycerol (IPG) en faveur de *(S)*-(+)-IPG, l'enzyme sauvage ayant une faible préférence pour l'énantiomère inverse. L'IPG dans sa forme *(S)* est un précurseur dans la synthèse d'un antagoniste des β-adrenorécepteurs. Les auteurs ont pour cela utilisé des mimes de substrats : un ester issu de la condensation de l'aspartate sur le S-(+)-IPG et un phosphoester issu de la condensation d'un phosphonate sur le R-(–)-IPG. L'hydrolyse du premier libère une source d'aspartate, tandis que l'hydrolyse du second libère le phosphonate, un puissant inhibiteur de la lipase (figure 18). Les chercheurs ont d'abord montré que les taux de conversion des substrats synthétiques et du substrat naturel étaient similaires et que l'enzyme avait conservé sa faible énantiosélectivité. Une banque de $3{,}5 \times 10^4$ transformants de la lipase a été transformée dans une souche de bactéries *E.coli* auxotrophe pour l'aspartate. Après plusieurs tours de sélection sur un milieu minimum supplémenté des deux analogues du précurseur d'IPG dans des conditions inhibant les bactéries porteuses du gène sauvage, plusieurs clones ont été sélectionnés. Alors que la Lipase A a une préférence envers la forme (R) (e.e = 23%, E = 1.8), après le troisième tour de sélection le meilleur variant sélectionné avait une préférence pour les esters de la forme *(S)* (*ee* = 73.1%, E = 8.5). En montrant qu'il était possible d'inverser l'énantiosélectivité d'une enzyme, cette étude pionnière a permis de valider, l'approche des mimes de substrat pour l'évolution dirigée de la sélectivité enzymatique.

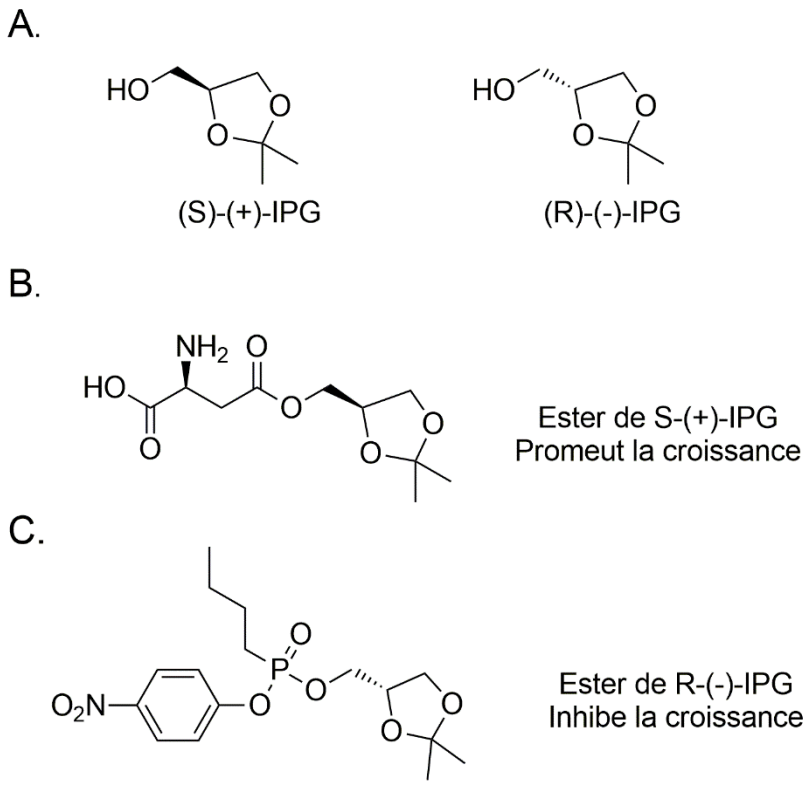


Figure 18 : A) Deux formes énantiomériques de l'IPG. La lipase naturelle hydrolyse préférentiellement l'ester butanoate de (R)-IPG. B) Mime du butanoate d'IPG où le butanoate est remplacé par un aspartate. Précurseur de (S)-IPG, il libère de l'aspartate après hydrolyse et promeut la croissance. C) Mime du butanoate d'IPG où le butanoate est remplacé par un phosphonate, précurseur de (R)-IPG, il libère un inhibiteur de la lipase après hydrolyse.

La même année, le groupe de Reetz a publié une approche similaire pour le clivage hydrolytique d'un ester d'acétate ou de fluoroacétate et l'IPG par la Lipase de *C. antartica*.[61] Cette lipase présente une préférence pour l'acétate d'IPG formant après hydrolyse du *(S)*-IPG (attention, la configuration absolue des esters d'IPG passent de *(R)* à *(S)* après hydrolyse selon la convention de nommage CIP) et le but de l'étude était l'inversion de cette sélectivité (figure 19). Dans cette approche, la réaction désirée génère l'énantiomère *(R)* de l'IPG et de l'acétate (source de carbone), tandis que l'autre réaction génère l'énantiomère *(S)* de l'IPG et du fluoroacétate (un poison cellulaire). Après ajustement des paramètres (pH et concentrations de la pro-source de carbone et du pro-poison), une petite banque de 240 variants sécrétés par *P. pastoris* a été étalée sur des boites de sélection. Sur les 80 colonies ayant poussé, 10 des plus grandes ont été récupérées pour des études d'énantiosélectivité. Il est apparu que 8 des 10 clones récupérés contenaient une lipase dont la sélectivité avait été inversée envers l'énantiomère *(S)*, le neuvième ayant une légère sélectivité pour *(R)* et le dernier étant inactif. Les ratios énantiomériques (*E*) trouvés allaient de 3 à 8.

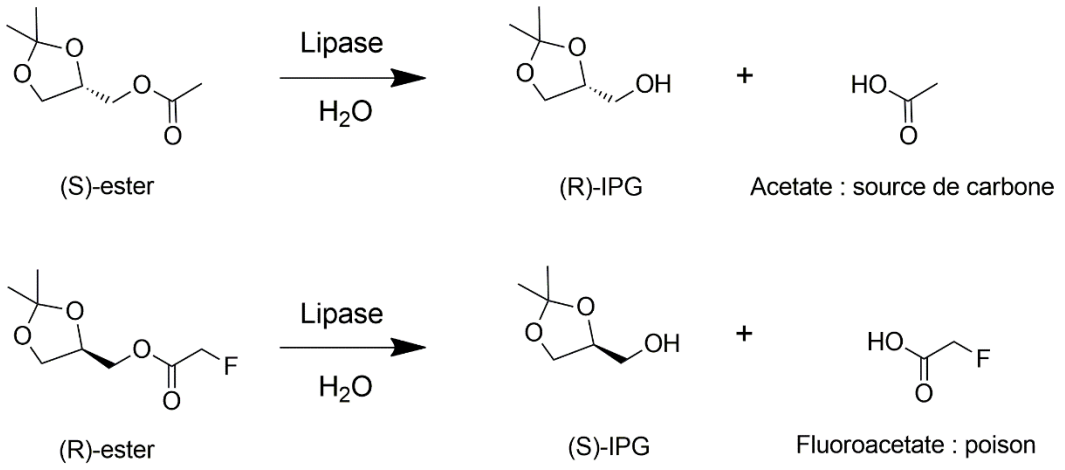


Figure 19 : Mimes du substrat IPG utilisés par l'équipe de Reetz pour la sélection de variants catalysant l'hydrolyse de l'ester précurseur de (S)-IPG. La réaction désirée permet l'obtention par la bactérie d'une source de carbone (de l'acétate) tandis que la réaction non désirée provoque la libération d'un poison cellulaire (du fluoroacétate).

Ces deux études sont, à notre connaissance, les seuls exemples de l'application de la sélection génétique pour l'amélioration de l'énantiosélectivité. Bien que très puissante pour cribler de grandes banques, elles ne sont applicables que sur des substrats particuliers et aucun résultat n'a été publié sur des banques conséquentes, alors que c'est l'avantage même de cette approche. L'utilisation de mimes de substrats, malgré leur utilisation avec succès dans ces deux études, est également un problème majeur.

Une autre faiblesse des systèmes de sélection dans des cellules est la complexité du milieu cellulaire vivant, qui permet de trouver des possibilités d'échappatoire à la pression de sélection. Une étude du groupe de M. Hecht est une parfaite illustration de ce phénomène.[62] Une banque de protéines synthétiques (basées sur un minimaliste « 4-helix bundle ») a été transformée dans une souche *E. coli* déficiente pour un gène codant la phosphosérine phosphatase, essentielle à la synthèse de sérine, en espérant trouver par ce moyen une protéine complémentant cette fonction catalytique manquante. Il est apparu qu'un clone sélectionné ne contenait de gène codant pour un variant catalytique, mais pour un facteur de transcription induisant l'expression du gène codant pour l'histidinol phosphate phosphatase, qui par promiscuité catalytique complémentait la fonction disparue. Ceci illustre la première règle de l'évolution dirigée : « You get what you select for ».[24]

I.5.2 Les méthodes par criblage

A partir d'une population d'objets divers, éventuellement, après des étapes de sélection permettant d'enrichir la population en variants présentant les propriétés recherchées, il est nécessaire de trier les clones individuellement pour isoler les meilleurs candidats recherchés : c'est ce que l'on appelle le criblage.

Les méthodes de criblage nécessitent d'analyser, un à un, différents variants. Ceci permet une plus grande flexibilité dans les conditions de réaction même si celles-ci doivent être compatibles avec la méthode employée. Dans le cas d'un crible catalytique, il est nécessaire non seulement de créer un lien entre le gène et la protéine, mais également de compartimenter le produit issu de la réaction. La méthode d'analyse est également différente selon l'observable utilisé : principalement par mesure d'absorbance ou de fluorescence, mais également par spectrométrie de masse ou spectroscopie infrarouge. Ici nous aborderons les deux méthodes de compartimentation communément employées, ainsi que les méthodes de détection compatibles avec ces dernières.

I.5.2.1 Le criblage en microplaque

La microplaque est encore la plus utilisée des méthodes de compartimentation. Malgré son débit plutôt faible (dépendant des moyens mis en œuvre), elle présente l'avantage d'être rapide à mettre en place et accessible dans la plupart des laboratoires. Dans chaque puits, une colonie unique d'un transformant de la banque est mise en culture et exprime le variant codé par le plasmide qu'elle contient. La protéine est généralement sécrétée par la cellule dans le milieu de culture, qui est alors utilisé comme solution de catalyseur.[63] La procédure peut être manuelle mais l'automatisation grâce à des robots piqueurs de colonies, incubateurs, pipeteurs etc., permet un débit de production plusieurs dizaines de milliers de transformant par jour. L'avantage principal de la compartimentation en plaque est la connaissance de la composition de chaque puits, permettant de dupliquer une réaction en changeant un ou plusieurs paramètres. Ce type de compartimentation peut être couplé à plusieurs types d'analyse.

Les techniques de chromatographie telles que la chromatographie liquide haute performance, la chromatographie gazeuse ou encore l'électrophorèse capillaire (en phase chirale) permettent d'avoir la plus grande précision sur les rendements et excès énantiomériques. Leur principal inconvénient est leur caractère sériel par nature et leur temps d'exécution allant de quelques minutes à plusieurs dizaines de minutes, ce qui conduit à un très bas débit de criblage (une centaine d'échantillons par jour). Des essais de multiplexage se sont malgré tout révélés concluants mais restent marginaux. Comme

preuve technique, une étude de Welch *et al.* décrit un montage de huit colonnes HPLC en parallèle pour la quantification des énantiomères de la benzoïne, les échantillons étant directement prélevés dans une plaque multipuits. [64] Malgré cette parallélisation, même une analyse HPLC de seulement dix minutes prendrait deux heures pour l'analyse d'une seule plaque de 96 puits.

De très élégantes approches par spectrométrie de masse ont été introduites pour quantifier la spécificité de biocatalyseurs dans des cas de résolution cinétique dynamique, ou de désymmétrisation de molécules *meso* ou prochirales.[65] Les substrats utilisés étaient en fait des pseudo-racémiques, pseudo-*meso* ou pseudo-prochiraux, se distinguant par leur composition isotopique et donnant aux produits de leur réaction une masse différente selon le groupement modifié par la réaction (figure 20). La spécificité des réactions pouvait donc être suivie en analysant les intensités des masses des produits formés. L'avantage de la spectrométrie de masse est, malgré un caractère sériel, une grande vitesse d'analyse. L'utilisation de pseudo-énantiomères, pseudo-*meso* ou pseudo-chirales isotopiques permet l'utilisation de parfaits mimes du substrat d'intérêt.

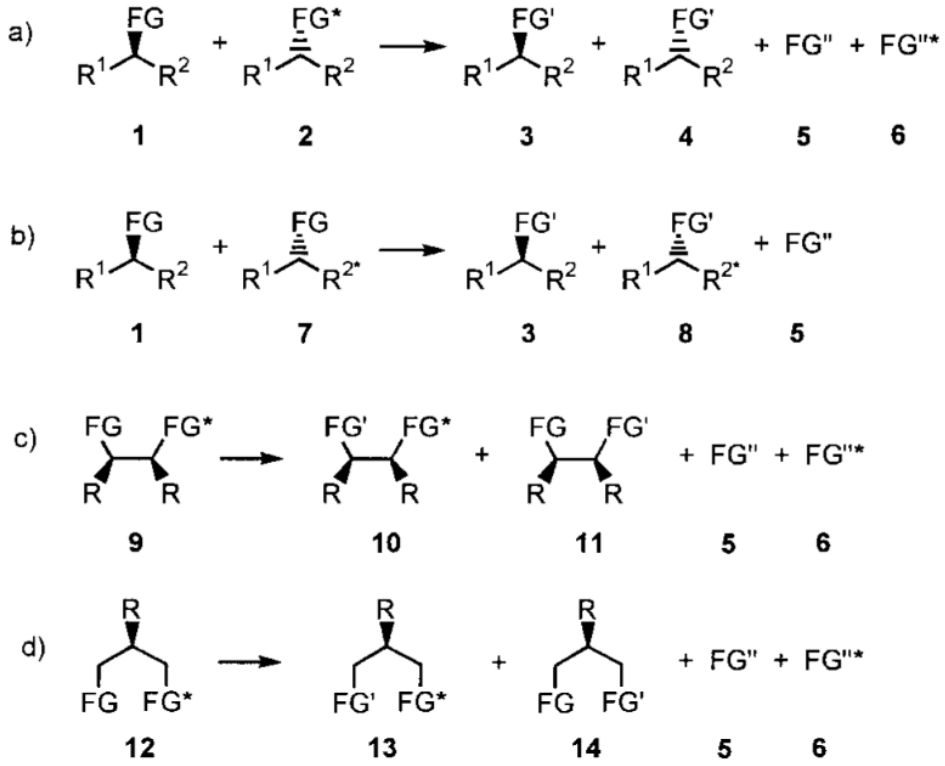


Figure 20 : a,c,d) Transformation asymétrique d'une mixture de pseudo-énantiomères, pseudo-meso ou pseudo-prochiraux impliquant le clivage du groupe fonctionnel FG ou du groupe fonctionnel marqué isotopiquement FG. Les intensités des produits 5 et 6* permettent de calculer l'énantiospécificité. b) Transformation asymétrique d'une mixture de pseudo-énantiomères impliquant soit le clivage, soit la formation d'une liaison avec le groupe fonctionnel FG. Le marquage isotopique est effectué sur le groupe R^2. Les intensités des produits 3 et 8* permettent de calculer l'énantiospécificité. Selon [65]*

Des méthodes optiques telles la fluorescence, l'absorbance, la luminescence, le dichroïsme circulaire et l'infra-rouge ont également été utilisées pour cribler des variants sur leurs propriétés de catalyse énantiospécifique. Les méthodes optiques sont idéales car rapides à analyser et hautement parallélisables mais peuvent très rarement être utilisées dans la détection « énantiospécifique » comme « énantiosélective ». En effet, peu de substrats ou produits d'intérêt sont directement observables, il est donc nécessaire de passer par des mimes de substrat chromogènes ou par des méthodes de détection secondaires présentant des inconvénients comme des étapes de dérivatisation ou pré-traitement avec des sélecteurs chiraux, une gamme de substrats étroite, un équipement spécialisé et une mise en place complexe. [66]

I.5.2.2 La compartimentation microfluidique

Durant la dernière décennie les méthodes de compartimentation microfluidique se sont considérablement développées. Ces méthodes utilisent des canaux de taille micrométrique convoyant des gouttes d'eau de l'ordre du picolitre au sein d'un fluide hydrophobe. La réduction de la taille du volume réactionnel couplée à la vitesse de création et d'analyse de ces gouttes ont ouvert de nouvelles perspectives dans la taille des banques de variants criblées.

Dans une étude du groupe d'Andrew Griffith, l'évolution dirigée de la peroxidase de raifort a été effectuée grâce à un système microfluidique avec compartimentation en microgouttes et suivi des réactions par la fluorescence (figure 21).[67] Un tableau comparatif des coûts et temps entre une méthode par plaque et une méthode microfluidique montre avec éloquence l'avantage de cette dernière approche (tableau 22).

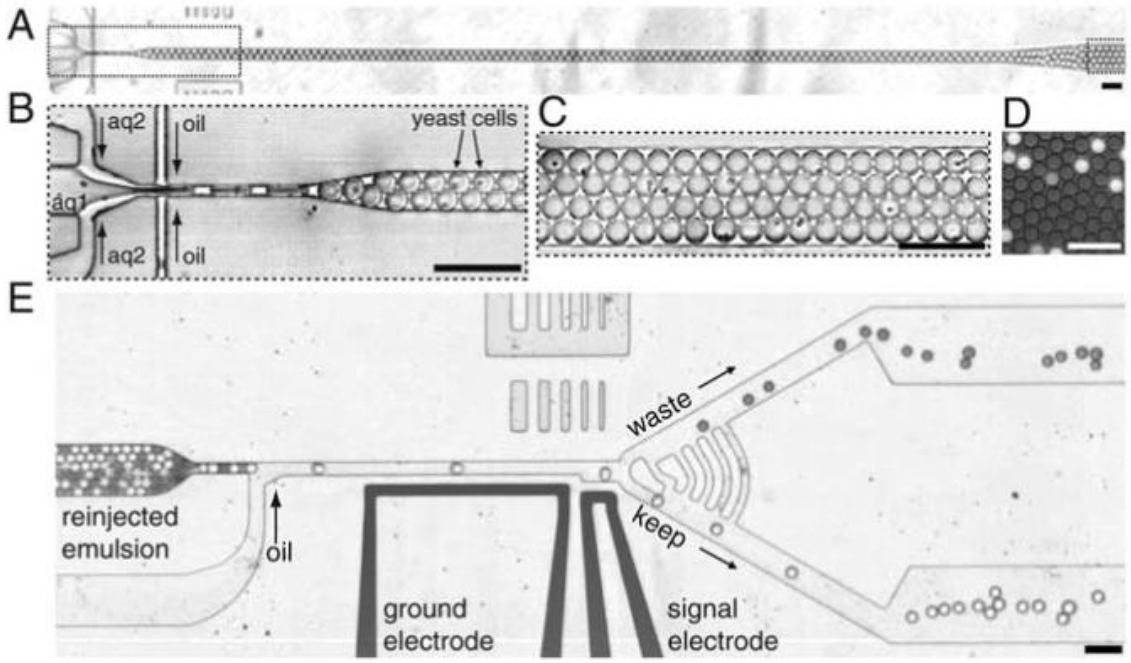


Figure 21 : Modules de la plateforme microflidique utilisée pour l'étude d'évolution dirigée de la HRP. A. Vue d'ensemble du système de création de gouttes. B. Vue rapprochée de la synthèse des gouttes, une solution contenant des levures affichant les variants HRP à leur surface est injectée dans l'entrée aq1 tandis que le substrat fluorogénique est ajouté en aq2. La vitesse de synthèse des gouttes est de 2 kHz. Le contenu des gouttes étant régit par une distribution de Poisson, ~22% des gouttes contiennent une cellule unique. C. Vue rapprochée de la ligne d'incubation, les gouttes stationnent dans cette zone pendant environ 5 minutes. D. Un tapis de goutte, les gouttes fluorescentes sont visibles en blanc. E. Séparation des gouttes après la période d'incubation analyse et ségrégation des gouttes fluorescentes grâce à un champ électrique.

	Robot	Gouttes microfluidiques
Réactions totales	5×10^7	5×10^7
Volumes de réaction	100 µL	6 pL
Volume total	5,000 L	150 µL
Reactions/jour	73	1×10^8
Temps total	~2 ans	~7 heures
Nombre de plaques/instruments	260	2
Coût des plaques/instruments	$520,000	$1.00
Coût des cônes de pipetage	$10 million	$0.30
Coût amorti des instruments	$280,000	$1.70
Coût des substrats	$4.75 million	$0.25
Coût total	$15.81 million	$2.50

Tableau 22 : Étude comparative d'une méthode de criblage grande échelle pour l'évolution dirigée d'une protéine au substrat fluorogénique, tel qu'effectué pour une peroxydase. Les couts d'utilisation d'un robot pipeteur/analyseur en microplaque et un système microfluidique sont comparés en terme de temps, volume et coût selon [67]

Malgré son avantage évident en matière de débit, les études utilisant cette approche se basent en grande partie sur des méthodes de détection optique, qui nécessitent encore une fois l'utilisation de mimes fluorogéniques (pour les cas de résolution cinétique) ou une dérivatisation du substrat (dans les réactions génératrices de chiralité). [68] La détection des produits de réaction reste actuellement le point limitant de la microfluidique.[66] De nouvelles études intégrant des circuits microfluidiques à de l'électrophorèse capillaire chirale et une détection en spectrométrie de masse ou fluorescence à UV profonds ont été récemment menées avec des résultats prometteurs, mais ne sont pas encore appliquées à des expériences d'évolution dirigée (figure 23). [69–71]

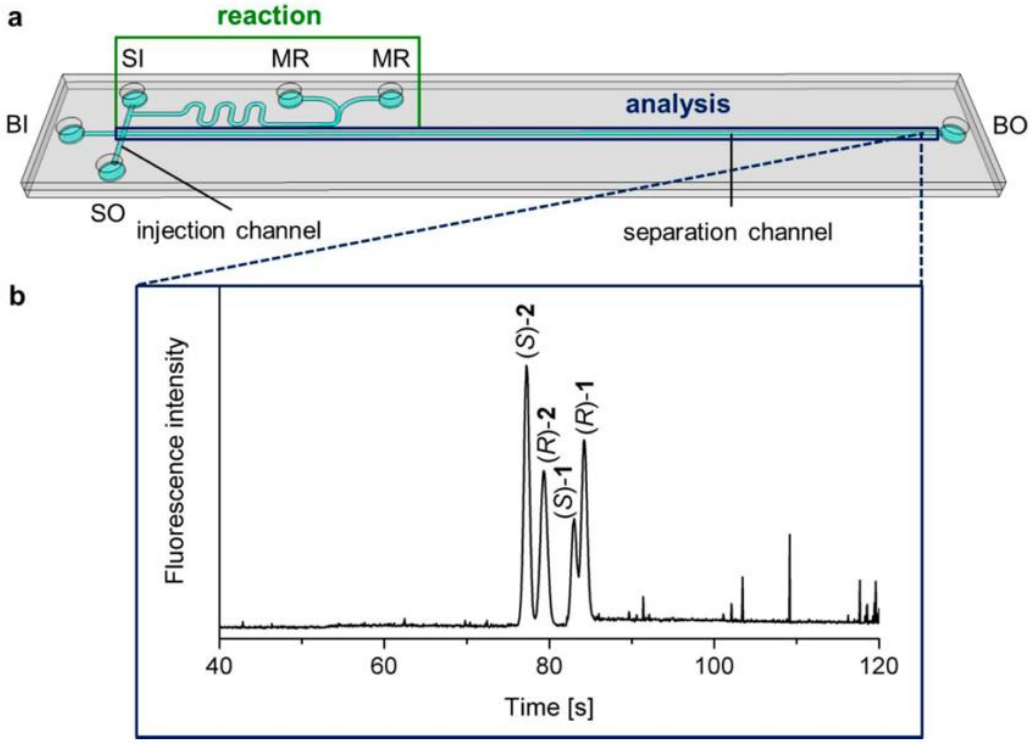


Figure 23 : Exemple de chromatographie intégrée en microfluidique. BI : entrée de tampon de migration, BO : sortie du tampon de migration, MR : microréacteur, SI : entrée de l'échantillon, SO : sortie de l'échantillon. L'analyse se fait en fluorescence à UV profonds pour détecter des molécules sans marquage.

I.6 Conclusion

Ces approches décrites d'ingénierie d'enzyme par évolution dirigée utilisent, à l'exception notable de l'approche par théozyme, une promiscuité de substrat ou de mécanisme réactionnel proche de l'activité naturelle des enzymes et se résument plutôt à une réaffectation de substrat ou une amélioration de sélectivité envers un substrat d'intérêt. Cependant, il existe des réactions n'étant catalysée par aucune enzyme

(hydrogenation, substitution allylique, hydroformylation).[72] Notamment, les réactions catalysées par des métaux de transition sont largement sous-représentées. Ainsi une voie intéressante est actuellement en train d'émerger : celle des catalyseurs biohybrides ou métalloenzymes artificielles.

Chapitre II - Les métaux dans les enzymes: métalloenzymes artificielles et biohybrides

Les métalloenzymes sont des enzymes contenant des ions de métaux de transition fixés dans leur site actif, soit directement à la protéine par liaisons dites datives, soit par le biais d'un groupe prosthétique (typiquement une porphyrine). Environ un tiers des enzymes existantes connues sont des métalloenzymes.[73] Les métalloenzymes catalysent les réactions chimiques les plus difficiles, grâce aux propriétés remarquables des métaux de transition.

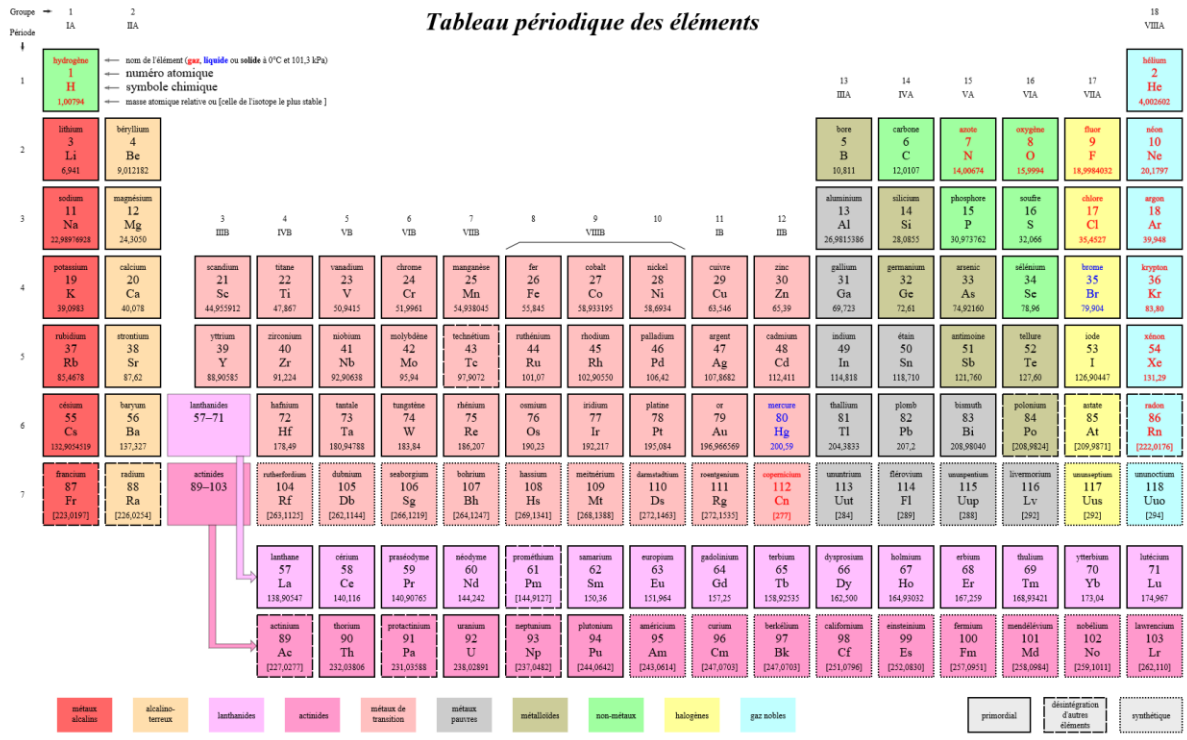


Figure 24 : Tableau périodique des éléments. En saumon sont représentés les métaux de transition.

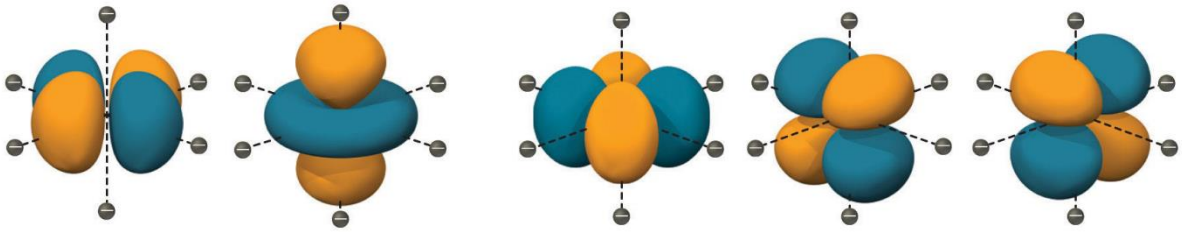


Figure 25 : L'orbitale d incomplète est caractéristique des métaux de transition

Les règles de remplissage des couches électroniques des atomes suivent les règles de la chimie orbitalaire établies au milieu du 20ème siècle (règle de Klechkowski, principe d'exclusion de Pauli, règle de Hund…). Cependant, les métaux de transition présentent souvent des exceptions à ces principes. La définition courante d'un métal de transition, utilisée par l'Union internationale de chimie pure et appliquée (IUPAC) est : « un élément ayant une orbitale *d* incomplète, ou pouvant former un ion stable avec une orbitale *d* incomplète ».[74]

Le cuivre dans son état atomique par exemple, a pour configuration électronique la plus stable $1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 3d^{10}\ 4s^1$. Avec sa couche d remplie et un seul électron sur la couche 4s, il fait ainsi une exception à la règle de Klechkowski. L'ion Cu^{2+} avec la configuration électronique $1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 3d^9$ présente, lui, une couche *d* incomplète et un électron non apparié qui lui confère ses propriétés paramagnétiques et fait rentrer le cuivre dans la définition des métaux de transition. L'ion Cu^+ ($1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 3d^{10}$) a une couche *d* complète (et est de fait diamagnétique). Les différents états redox du cuivre peuvent jouer un rôle très important, en particulier dans les enzymes à cuivre.

Le zinc ($1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 4s^2\ 3d^{10}$) fait office d'exception à la définition générale d'un métal de transition. En effet, l'ion Zn^{2+} stable en solution a une couche *d* complète (configuration électronique $1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^6\ 3d^{10}$). Cependant, le zinc est un métal très présent dans les enzymes et les protéines en général, pouvant avoir un rôle structural ou catalytique. Il sera donc inclus dans le propos développé ci-dessous.

Le remplissage incomplet des couches électroniques supérieures (s ou d) confère aux métaux de transition leurs propriétés particulières d'acides de Lewis (accepteur de doublets d'électrons) en permettant ainsi leur complexation avec des ligands non métalliques qui se comportent alors comme des bases de Lewis, en partageant une paire d'électrons.

La liaison établie entre un acide et une base de Lewis dépend du caractère « dur » ou « mou » de ceux-ci, selon la théorie de Pearson (principe HSAB). Un acide ou une base dur(e) est faiblement polarisable et possède une charge totale ou partielle très localisée, contrairement aux acides mous et bases molles qui sont fortement polarisables et possèdent une charge partielle et peu localisée. En chimie sont appelés « ligands » les atomes participant à la sphère de coordination d'un métal, et plus généralement les molécules impliquées dans la complexation d'un métal. Les (atomes) ligands présents dans les protéines peuvent être classés en fonction de leur électronégativité croissante (O > N > S), ce qui leur confère un caractère mou ou dur et donc une comptabilité avec les métaux qu'ils sont susceptibles de coordiner. Un acide dur forme une liaison plus stable

avec une base dure, et inversement.[75] Ces différences de propriétés électroniques expliquent en partie les diverses réactivités des métaux.

Le ligand (en tant que molécule) est caractérisé par sa denticité : Un ligand tel que la phénanthroline par exemple possède deux atomes d'azote permettant la liaison avec un métal, c'est un ligand bidenté. La denticité détermine de façon prépondérante l'affinité de la liaison du ligand avec le métal qu'il coordine et joue également un rôle dans la géométrie du complexe. Le nombre de liaisons de coordination qu'un ion métallique peut former dépend de sa configuration électronique, elle peut aller de 2 à 9.[76] L'arrangement des ligands autour de l'ion est appelé la géométrie de coordination. Cette géométrie dépend du nombre d'atomes ligands, de leur nature (dur ou mou) mais également des contraintes géométriques imposées par leur taille et leur structure. Ainsi une géométrie de complexe peut être couramment distordue.[76] Le potentiel redox est un autre paramètre influant sur la réactivité de chaque métal, l'état d'oxydation changeant la composition électronique des couches de valence du métal, les types de liaisons et réactivités sont modifiées. [76]

Tout comme le ligand modifie la réactivité d'un métal, le métal module en retour la réactivité du ligand. Ces ligands peuvent être également substrats d'une réaction, la complexation permettant l'acte catalytique en induisant le rapprochement des différents partenaires ainsi que leur activation par contraintes physiques, polarisation ou transfert d'électron.

C'est par la combinaison de tous ces paramètres (configuration électronique, ligands, potentiel redox) que la réactivité des métaux de transition est si diverse. Dans la nature, les métaux de transition disponibles en grande quantité sous forme soluble sont utilisés dans de nombreuses réactions biologiques, ce qui explique l'abondance de métaux non précieux dans les enzymes naturelles : fer, zinc, cuivre, manganèse…

La diversité des réactions chimiques catalysées par les enzymes naturelles avec un répertoire si restreint de métaux est déjà impressionnante, on peut donc envisager d'exploiter les propriétés des métaux et de leurs ligands pour étendre le spectre catalytique des catalyseurs bio-inorganiques.

II.1 Le design de centre métallique datif

La première approche que l'on peut envisager pour étendre le spectre catalytique d'une enzyme est de remplacer son métal naturel par un autre.

Dans les protéines à liaison dite « dative » ou « directe » de métal, les chaines latérales des acides aminés ou le squelette peptidique sont directement impliqués dans la coordination du métal. Un grand nombre de protéines utilisent la liaison dative du zinc pour maintenir leur structure tridimensionnelle, mais le zinc est également présent dans les sites catalytiques. Lorsque l'ion Zn^{2+} joue un rôle structural, sa sphère de coordination est complètement remplie par des atomes de la protéine, alors que lorsque le zinc est catalytique, sa sphère de coordination présente au moins un site vacant, pouvant être occupé par une molécule d'eau ou de substrat. Une molécule d'eau présente comme ligand des ions Zn^{2+} catalytiques peut être activée par ionisation ou polarisation, ou encore être remplacée par un substrat. C'est le cas de l'anhydrase carbonique, qui possède un ion zinc coordiné par trois histidines dans son site actif (figure 26).

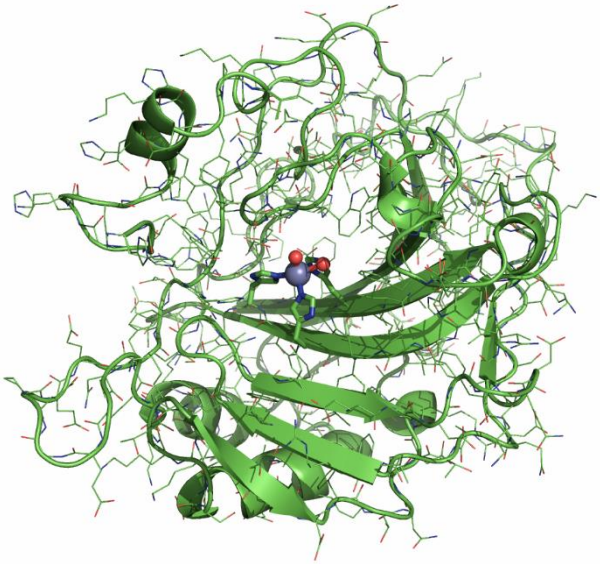


Figure 26 : Structure de l'anhydrase carbonique humaine (PDB : 3KS3) avec le site de chélation du zinc en son centre (bâtons) : le zinc est coordiné par trois histidines et deux molécules de solvant.

II.1.1 Changement de la réactivité de l'anhydrase carbonique par échange du cation catalytique

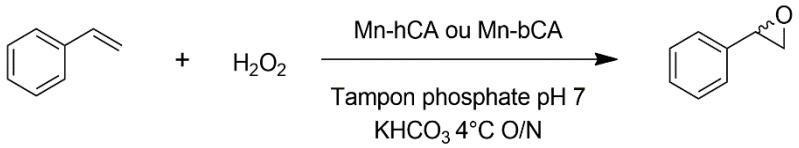
L'anhydrase carbonique est une enzyme fondamentale pour le fonctionnement de la vie aérobie. Elle catalyse la conversion, en présence d'une molécule d'eau, du dioxyde de carbone en acide carbonique et bicarbonate dans les deux sens de réaction. Chez les animaux, elle permet la dissolution du CO_2 produit par la respiration dans le sang et son

relargage au niveau des poumons (ou branchies).[77] Chez les plantes également, une anhydrase carbonique permet le transport du CO_2.[78] Il existe cependant trois classes évolutivement distinctes : alpha (monomériques), beta (dimériques) et gamma (trimériques).[79] La première existe chez les mammifères, la deuxième chez les plantes et la troisième chez les bactéries méthanogènes des sources chaudes. Malgré leur très faible identité de séquence, les trois ont besoin d'un ion zinc dans leur site catalytique.

Le zinc étant un métal très versatile (intermédiaire entre un atome dur et un atome mou), il peut former des liaisons datives avec des ligands durs tels que les oxygènes des carboxylates, intermédiaires tels que les azotes des histidines, ainsi que mous tels que les thiols des cystéines. La mutation en alanine d'une des histidines coordinant le zinc diminue l'affinité pour le zinc d'un facteur 10^5. Ceci est vrai pour chacune des trois histidines, ce qui indique une forte coopérativité entre les trois sites.[80] De même, le remplacement en aspartate provoque une chute d'affinité de 10^4. En revanche, un remplacement par une cystéine s'accompagne d'une baisse d'un facteur 2 uniquement.[80] Le changement de ligand diminue également son efficacité catalytique tant dans l'hydratation du CO_2 que dans l'hydrolyse d'esters, qui est une autre réaction catalysée par cette protéine, probablement en modifiant le pKa de la molécule d'eau présente dans la sphère de coordination du zinc. Ces résultats démontrent que la nature des résidus et leurs positions sont cruciales pour l'activité catalytique du métal. La conception d'un nouveau site catalytique par liaison directe du métal au sein d'une protéine s'avère donc être un travail d'orfèvre.

En 2006, les équipes de Romas J Kazlauskas et Patrice Soumillion ont publié le changement de portée de réaction et de substrat de l'anhydrase carbonique bovine et humaine par un simple changement de métal.[81,82] La protéine était déplétée de son zinc par une dialyse sur la nuit en présence d'un chélatant (2,6-pyridinedicarboxylate), puis reconstituée par l'ajout de chlorure de manganèse. Ces protéines modifiées catalysent l'époxydation du styrène avec le peroxyde d'hydrogène comme oxydant. Il avait été postulé que la catalyse était dépendante de la formation d'un peroxymonocarbonate HCO_4^-, la réaction a alors été effectuée en présence de bicarbonate. Une poche hydrophobe connue pour accepter des inhibiteurs de l'anhydrase carbonique a été prise en compte dans le choix du substrat. Différents types d'alcènes aromatiques ont ainsi pu être époxydés avec des excès énantiomériques allant jusqu'à 40% pour la protéine humaine et 50% pour la protéine bovine, mais avec des mauvais rendements pour cette dernière. L'augmentation de la concentration en oxydant augmentaient le rendement mais diminuaient la sélectivité. Okrasa et Kazlauskas ont testé l'époxydation du *p*-chlorostyrène et sont parvenus à une énantiosélectivité de 67% pour l'énantiomère (*R*) en utilisant du tampon MES, mais avec des rendements très faibles (~4%). Différents

tampons conféraient à la réaction des énantiosélectivités et des rendements variables. La conclusion de ces études était que la protéine ralentissait la réaction en empêchant l'accès au métal. De plus, la protéine semblait se dégrader au cours de la réaction, probablement du fait de la génération d'intermédiaires réactionnels radicalaires hautement réactifs.



Enzyme	Epoxidation yield [%]	*ee* [%]	Ratio H_2O_2/styrene
Mn-bCA	9	52	10
Mn-hCA	29	40	10
No apoCA	0	0	10
Mn-bCA	23	33	36
Mn-hCA	57	18	36
$Mn\text{-}hCA_{DEPC}$[a]	9	37	36
No apoCA	0.9	0	36
Zn-bCA	5	0	36
Zn-hCA	2	0	36

Figure 27 : Epoxydation du styrène par l'anhydrase carbonique humaine (hCA) ou bovine (bCA) dont le zinc a été remplacé par un sel de manganèse (selon [81]). La modification par pyrocarbonate d'éthyle (DEPC) est utilisée pour modifier les histidines impliquées dans la coordination du métal.

Plus tard la même équipe a gardé cette approche en utilisant le rhodium sous forme Rh(I) pour catalyser l'hydrogénation de stilbène avec une stéréosélectivité allant jusqu'à 20:1 en faveur du cis-stilbène.[83]

II.1.2 Création d'un site catalytique via la fixation d'un métal dans la protéine tHisF

Le groupe de Reetz s'est intéressé en 2010 à la création *de novo* d'un site de fixation de métal catalytique au sein d'une protéine. Le choix de l'architecture protéique s'est porté sur l'imidazole glycérol phosphate synthase (tHisF) de *Thermotoga maritima*. [84] Cette protéine avec une structure de type « TIM-barrel » a été choisie pour sa robustesse (venant d'un organisme thermophile), sa production efficace chez *E. coli,* et sa purification aisée par traitement thermique. Ces différents atouts en faisaient un candidat de choix pour une ingénierie génétique subséquente. Dans l'étude de Podtenieff *et al*, un site de chélation constitué de deux histidines et un carboxylate a été conçu dans l'idée d'accueillir un ion Cu^{2+} (figure 28).[84] Les autres acides aminés risquant de coordiner le cuivre ont été également remplacés par des alanines : 4 histidines de surface ainsi qu'une cystéine proche de la triade catalytique.

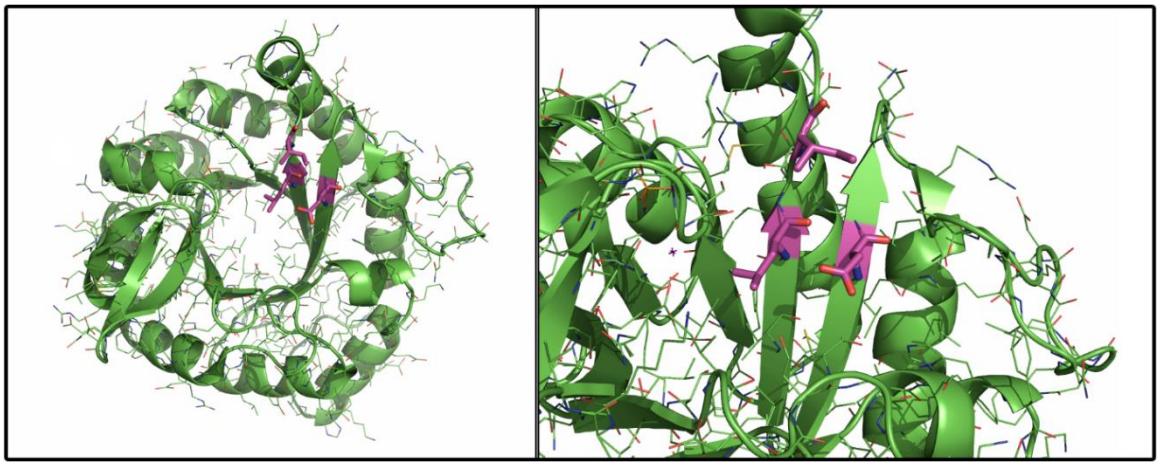
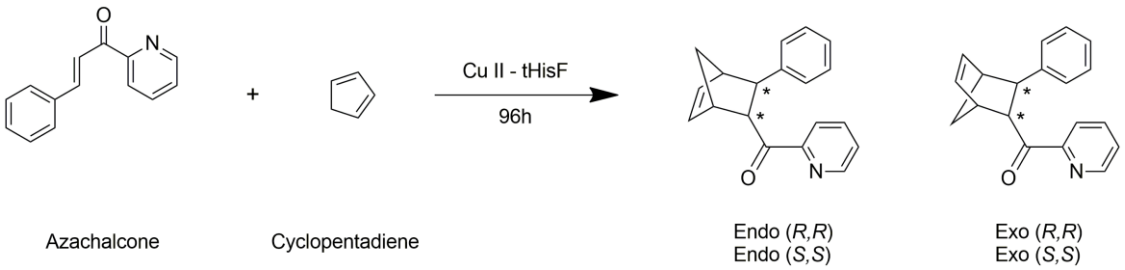


Figure 28 : Structure cristallographique de la protéine tHisF utilisée par l'équipe de Reetz pour la conception d'un site catalytique par liaison directe d'un cation métallique au sein de la protéine (selon [84]). Les acides aminés utilisés pour créer le variant L50H/I52H (l'aspartate D11 n'est pas modifié) sont représentés en bâtons violets.

Un test de catalyse de la réaction de cycloaddition de Diels-Alder entre l'azachalcone et le cyclopentadiène a été effectué avec ces différents variants (figure 29). Il est apparu, en présence de cuivre, que seule la protéine possédant la triade catalytique, (HHD-4xala dans le tableau, figure 29) catalysait la réaction avec un excès énantiomérique conséquent (jusqu'à 46%), ce qui semblait conforter l'hypothèse de la création d'un site de fixation du cuivre. [84]



Catalyst	Conversion [%]	*ee* [%]	*endo/exo*
Cu^{II}/HHD-4xala	73 ± 1.6	46 ± 0.5	13:1
Cu^{II}/NC-4xala	61 ± 0.5	5 ± 0.1	9:1
Cu^{II}HHA-4xala	56 ± 1.2	4 ± 0.9	8:1
free Cu^{II}	44 ± 2.9	1 ± 0.3	7:1
no catalyst (buffer)	26 ± 3.1	1.5 ± 0.5	8:1

Figure 29 : Réaction énantiosélective de Diels-Alder catalysée par les mutants de tHisF. Les excès énantiomériques sont calculés sur la forme endo, qui est thermodynamiquement favorisée. Les réaction ont été effectués avec 153 µM de protéine, 115 µM de $CuSO_4$,1.05 mM azachalcone et 5.25 mM cyclopentadiène pendant 96h (tableau reproduit selon [84]).

Des expériences de résonance paramagnétique électronique (RPE) ont permis de montrer que le cuivre se trouve dans des environnements atomiques différents, s'il est en présence de la protéine possédant la triade ou bien en présence de la protéine native. L'interprétation de spectre obtenu avec la triade montrant un facteur de Landé *g* différent de celui mesuré pour des électrons d'un cuivre en solution, et un couplage hyperfin des signaux a permis de suggérer la fixation du cuivre par une liaison à des histidines. [84] Il a été relevé en revanche que le signal était complexe, dénotant une flexibilité dans l'environnement du cuivre. De plus la protéine sans triade présentait également un signal, l'auteur argumentant qu'il était courant d'observer des sites de coordination fortuits de métaux au sein de protéines. Des études de RPE plus poussées de corrélation de couplage hyperfin ont permis de confirmer la nature des ligands du cuivre dans les deux cas : une complexation par des imidazoles dans la protéine possédant la triade et une complexation par un atome d'azote du squelette protéique dans la protéine sans la triade. Malheureusement, malgré les résultats encourageants et une architecture adaptée à une stratégie d'évolution dirigée, aucun autre résultat n'a été publié par ce groupe sur ce projet.

Ces deux exemples nous enseignent d'une part qu'il est possible de créer une métalloenzyme ayant des propriétés catalytiques inédites avec une coordination directe d'un métal, mais d'autre part que ce type d'ingénierie est particulièrement délicat. En effet, le type de mutations et la position des résidus modifiés sont des facteurs cruciaux pour la fixation et la réactivité du métal, le plus léger changement pouvant être extrêmement délétère. De plus, la fixation du cation métallique peut être en compétition avec d'autres sites fortuits de faible affinité couramment présents à la surface des protéines. Si le site modélisé n'est pas suffisamment affin et spécifique, il est fortement probable que les modifications introduites soient responsables d'une perte de sélectivité conséquente.

L'ingénierie de sites catalytiques datifs est un domaine de recherche à part entière, impliquant une part croissante de modélisation théorique à la fois sur les modèles dynamiques des protéines mais également sur les champs de force pour modéliser les propriétés des métaux. Ne pouvant aborder ces domaines exhaustivement ici, nous nous arrêterons à ces deux exemples récents pour nous intéresser à une autre approche pour l'insertion de métaux au sein d'architectures protéiques : l'utilisation de cofacteurs métalliques.

II.2 Insertion d'un cofacteur métallique au sein d'une architecture protéique: les catalyseurs biohybrides.

Le principe d'un catalyseur biohybride est de séparer l'acte catalytique de l'acte de positionnement du substrat, ces actions étant effectuées respectivement par le cofacteur métallique et l'architecture protéique (figure 30). Ceci présente de multiples avantages : L'activité catalytique peut être ainsi modifiée suivant la nature du cofacteur et du métal utilisé. Nous verrons qu'il existe cependant de nombreux cas où des résidus de l'architecture complètent également la première sphère de coordination du métal et modulent ainsi sa réactivité. Le cofacteur artificiel portant le métal (appelé « complexe », ou « ligand » si il ne porte pas de métal, ou encore « groupement prosthétique » s'il s'agit d'un hème) permet également de s'assurer de fortes affinités pour le métal, ce qui facilite la conception de métalloenzymes. Enfin, la sélectivité envers un ligand ou un produit donné peut être modulée en modifiant l'architecture protéique par ingénierie génétique.

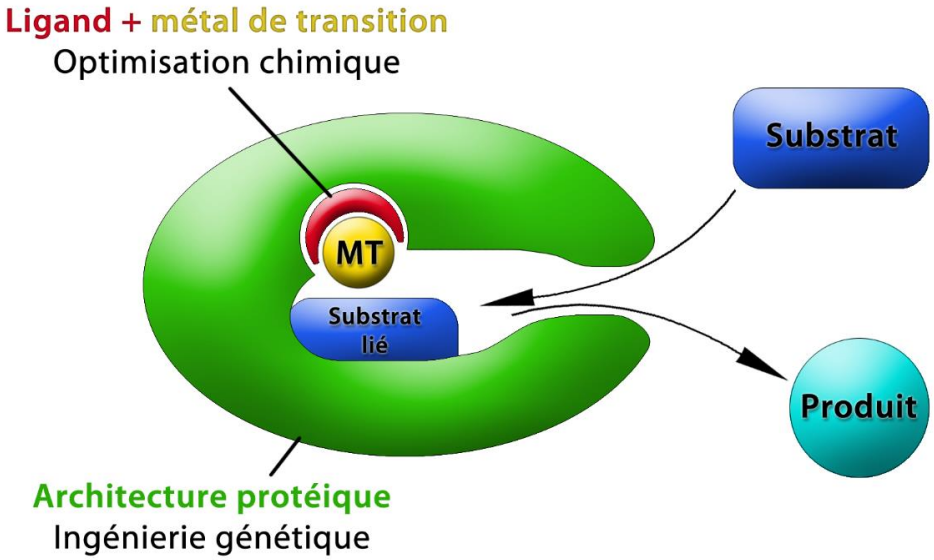


Figure 30 : Principe d'un biohybride. Un complexe métallique (rouge) est lié au sein d'une architecture protéique (vert). La protéine est en charge du positionnement du substrat et induit une sélectivité tandis que l'ion métallique effectue l'acte catalytique.

Trois stratégies différentes existent pour la création de catalyseurs biohybrides. La première consiste en la fixation directe et non covalente du ligand au sein de la protéine, il s'agit de la stratégie « hôte-invité ».[85] La deuxième exploite également une fixation non covalente mais indirecte, le cofacteur se fixant à la protéine par l'intermédiaire d'une autre molécule, il s'agit de la stratégie dite du cheval de Troie.[85] La troisième stratégie implique enfin un lien covalent entre protéine hôte et cofacteur métallique. Dans ce

chapitre, les différents exemples illustreront ces stratégies et un intérêt particulier sera porté sur les capacités d'évolution de chaque système en vue d'améliorer l'activité ou l'énantiosélectivité.

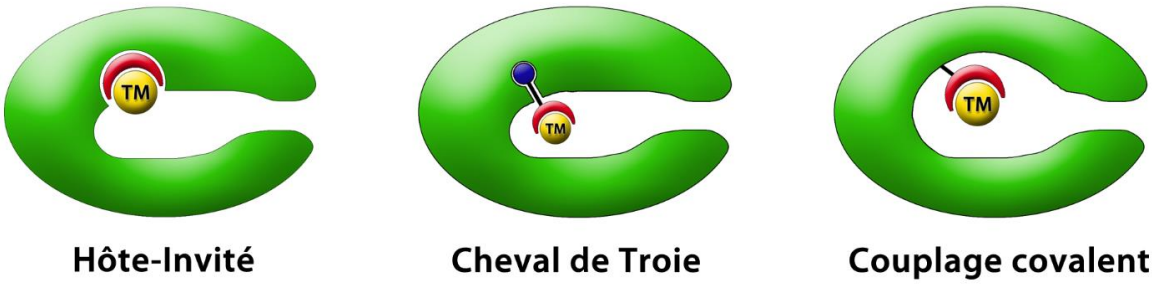


Figure 31 : Présentation schématique des trois stratégies de création d'un catalyseur biohybride. Hôte-Invité : le cofacteur métallique est lié à la protéine de façon non covalente. Cheval de Troie : Le cofacteur métallique est lié covalemment à un ligand de la protéine. Couplage covalent : le cofacteur métallique est directement lié de façon covalente avec la protéine.

II.2.1 La stratégie Hôte-Invité

Il est courant d'associer la stratégie hôte-invité à l'incorporation non naturelle de complexes métalliques dans des protéines. Cependant cette définition n'inclurait pas les approches consistant en la sélection d'un anticorps contre un hème, qui sont tous deux des molécules présentes dans le vivant. Il est aussi impossible de ne pas aborder dans cette partie le domaine de l'ingénierie des protéines héminiques telles que les cytochromes P450 et les myoglobines. Nous utiliserons donc une définition large qui est : l'ingénierie de toute protéine, naturelle ou artificielle, contenant un complexe métallique, naturel ou artificiel, lié de manière directe mais non covalente, dans le but d'obtenir une activité catalytique nouvelle.

II.2.1.1 Ingénierie des cytochromes P450

Les cytochromes P450 sont des enzymes héminiques ubiquitaires catalysant des réactions d'oxydation d'alcènes, d'hétéroatomes ou de liaisons C-H inertes utilisant l'oxygène moléculaire comme oxydant. Leur rôle principal au sein de l'organisme est l'oxydation des xénobiotiques dans la première phase de leur métabolisme. Ces composés souvent hydrophobes sont ainsi rendus plus hydrophiles, permettant leur élimination par les voies d'excrétion (métabolite de phase I). Leur rôle ne s'arrête cependant pas là, certains cytochromes P450 (CyP450) étant présents dans de nombreuses voies de biosynthèse de produits naturels comme les terpènes (dont les stéroïdes), alcaloïdes et polycétides. En plus de cette grande palette de substrats, les P450 (naturelles et artificielles) sont capables de catalyser une grande variété de réactions : hydroxylations,

époxydations, sulfoxydations, couplages aryl-aryl, nitrations (en présence d'oxyde nitrique [86]), dehalogénations oxydatives et réductrices, ainsi que des réactions non naturelles telles que les transferts d'alcènes.[87–89]

De nombreuses études ont permis d'élucider le mécanisme réactionnel commun aux cytochromes P450. La protéine dans son état oxydé possède un hème ferrique Fe (III) lié à un résidu cystéinate en position proximale et à une molécule d'eau en position distale. En présence de ces ligands, le fer est à l'état bas spin, ce qui empêche sa réduction en Fer (II) (figure 32 (A)). La fixation du substrat hydrophobe chasse la molécule d'eau et le fer se retrouve alors en état activé (haut spin), qui peut alors être réduit en Fer (II) par le domaine réductase de la protéine. Le dioxygène peut alors se fixer pour former l'intermédiaire superoxyde ferrique (figure 32 (B)). Une nouvelle étape de réduction donne l'intermédiaire fer-peroxo (figure 32 (C1)) qui après protonation médiée par l'hydroxyle d'une thréonine donne l'intermédiaire fer-hydroperoxo (figure 32 (C2)). Après libération d'une molécule d'eau, l'intermédiaire de haute valence est formé (radical cathion, nommé par convention composé I (figure 32 (D)). [89] C'est généralement cet intermédiaire qui est responsable des réactions d'oxydation par un mécanisme d'abstraction d'hydrogène (la plupart du temps) ou de transfert séquentiel à un électron à la manière des peroxydases. Cependant, les intermédiaires superoxydes ainsi que peroxo et hydroperoxo peuvent également être responsables de réactions de nitration, époxydations et coupure de liaison C-C.

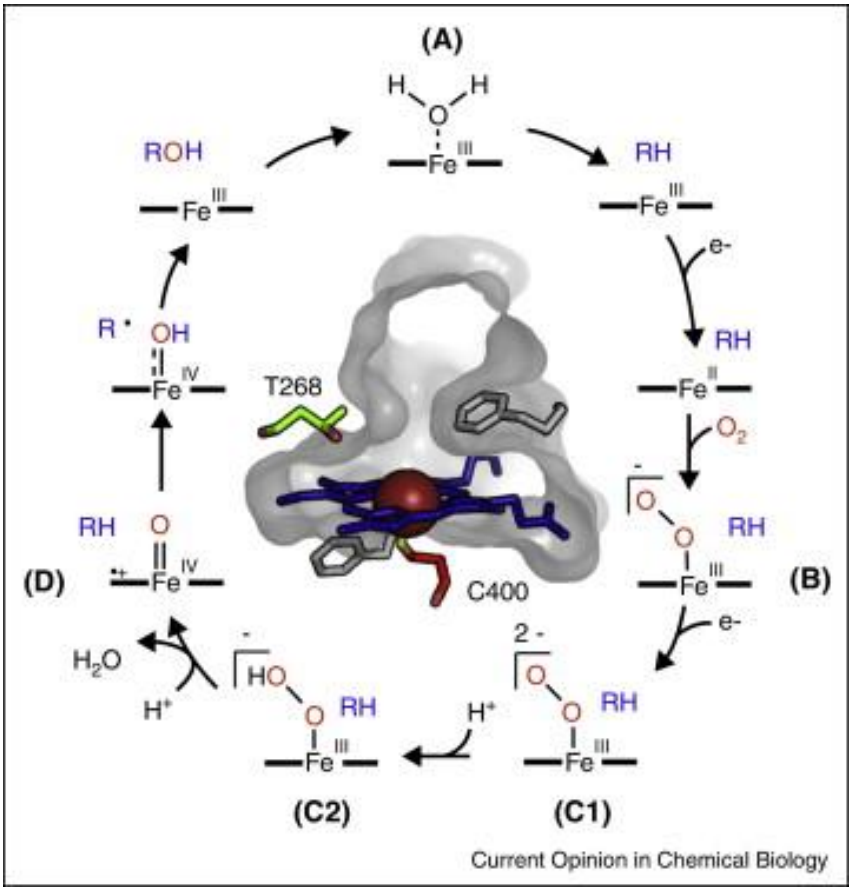
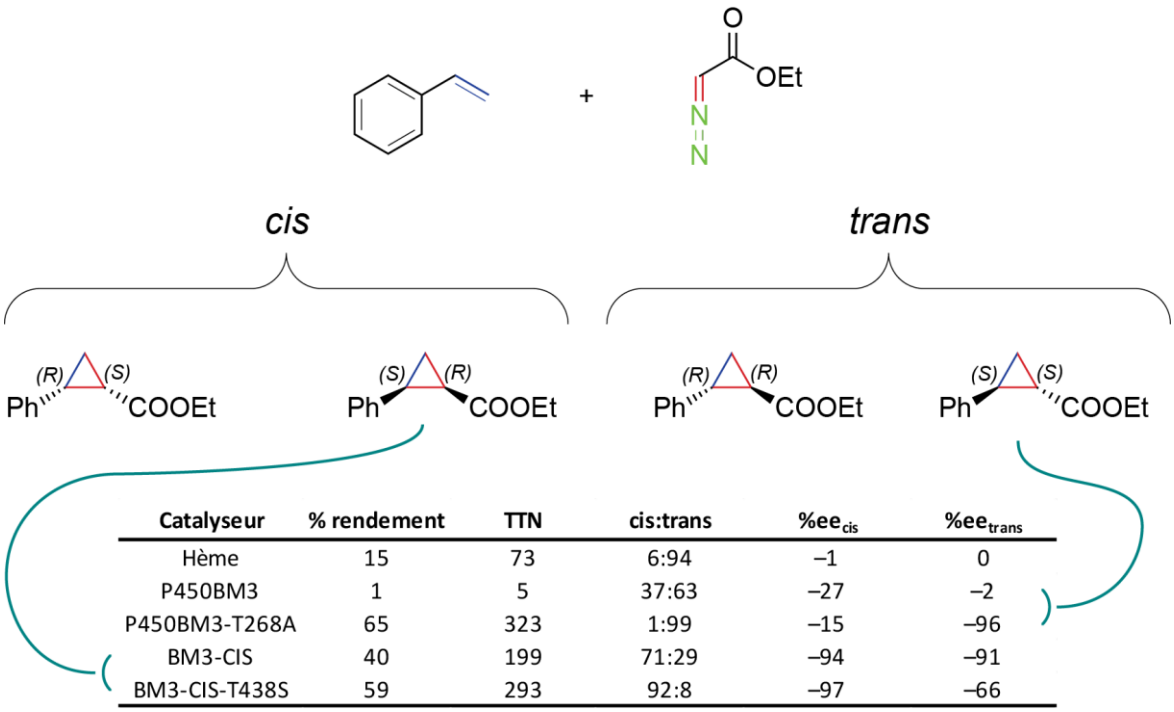


Figure 32 : Cycle catalytique des cytochromes de type P450 (selon [89]). Dans l'état oxydé, au repos, le fer est en état d'oxydation III et est complexé en face distale par une molécule d'eau (A), lui conférant une configuration électronique de spin bas, non réactive. La complexation du substrat chasse cette molécule d'eau, passant la configuration du fer à spin haut, dans un état activé permettant sa réduction en Fe (II). La complexation d'une molécule d'oxygène forme le composé superoxyde (B). Une nouvelle réduction donne l'intermédiaire peroxo (C1) puis hydroperoxy (C2). La coupure hétérolytique de la liaison O-O forme le composé I (D), radical cation à haute valence. Ce composé est en général responsable de la catalyse par abstraction d'hydrogène.

Le groupe de Frances Arnold s'est intéressé à l'exploitation de cette classe de protéines pour catalyser le transfert de carbènes. Des travaux pionniers à ce sujet ont montré que le cytochrome $P450_{BM3}$ de *B. megaterium* pouvait catalyser la synthèse de dérivés du cyclopropane à partir de styrènes en conditions anaérobies en utilisant du diazoacétate d'éthyle (EDA) comme source de carbène (figure 33). L'efficacité de ce catalyseur était inférieure à l'hème en solution, cependant la mutation de la thréonine 268 (inutile dans ce type de réaction) en une valine moins volumineuse permis la cyclopropanation *trans*-sélective avec de bons rendements (323 TTN, 96% *ee* forme (*S,S*), 1 :99 *cis*:*trans*). Le criblage d'une petite bibliothèque de variants permis d'isoler le variant BM3-CIS, porteur de 14 mutations et présentant une diastéréosélectivité envers la forme *cis*. A l'issue d'une expérience de mutagenèse systématique (CAST) sur 5 acides aminés du site actif, un variant optimisé de seconde génération, BM3-CIS-T438S, a été obtenu. Il présente une

très bonne préférence pour la forme *cis* ainsi qu'une excellente énantiosélectivité (293 TTN, 97% *ee* forme (*S,R*), 92:8 *cis*:*trans*). [90]



Catalyseur	% rendement	TTN	cis:trans	%ee$_{cis}$	%ee$_{trans}$
Hème	15	73	6:94	–1	0
P450BM3	1	5	37:63	–27	–2
P450BM3-T268A	65	323	1:99	–15	–96
BM3-CIS	40	199	71:29	–94	–91
BM3-CIS-T438S	59	293	92:8	–97	–66

Figure 33 : Réaction de cyclopropanation entre le styrène et un carbène activé par un groupement diazo (EDA) (selon [90]). La réaction engendre 4 isomères (deux énantiomères dans la forme cis, deux énantiomères dans la forme trans). Les isomères favorisés par les différents variants sont représentés par les flèches bleues.

Un autre variant catalysant l'insertion de carbène dans une liaison N-H a également été isolé (figure 34). [91]

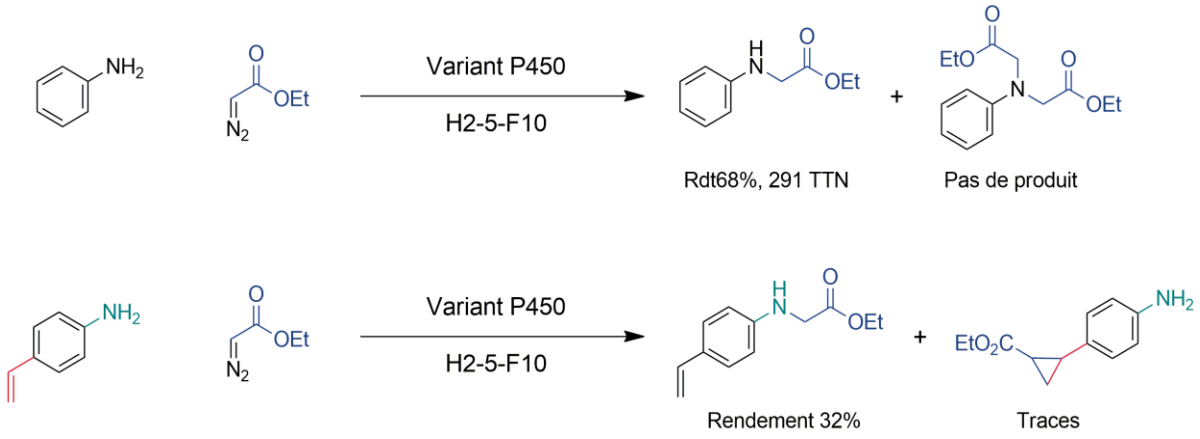


Figure 34 : Réactions sélectives d'insertion de carbène au sein d'une liaison N-H catalysée par des variants de P450 (selon [91])

Comme dit précédemment, la protéine sauvage au repos possède un hème dans l'état Fe (III). Étant donné que l'espèce active est sous forme Fe (II), il était intéressant d'essayer de faciliter la réduction initiale de l'hème. L'équipe de Arnold a donc étudié l'impact du ligand proximal sur le potentiel redox du fer. Il s'est avéré qu'une mutation de la cystéine 400, strictement conservée dans la famille des CyP450, en une sérine provoquait un changement du potentiel redox passant de -430 mV à -293 mV, le rendant réductible par le NAD(P)H sans fixation préalable du substrat. Ce nouveau variant fut renommé $P411_{BM3}$ du fait du décalage à 411 nm de la bande de Soret de l'hème lié au CO. Cette nouvelle espèce permet donc d'éviter l'utilisation d'un réducteur fort comme le dithionite et de pouvoir effectuer les catalyses de transfert de carbène en conditions douces, dans un milieu intracellulaire. De plus, l'affinité pour le dioxygène est réduite, ce qui permet d'effectuer la catalyse en milieu aérobie sans empoisonner le catalyseur. Ainsi les réactions de transfert de carbènes asymétriques ont pu être effectuées avec de bonnes diastéréo- et énantiosélectivités avec un rendement accru, même en milieu intracellulaire.[92]

Cette équipe s'est également intéressée à l'insertion de nitrènes au sein de liaisons C-H et S-H (sulfimidation) par l'utilisation de précurseurs azides (figure 35).[93,94] L'enzyme P411 BM3-CIS-T438S était également le plus actif des variants testés pour cette réaction. Cette étude a permis de conclure que la mutation de deux acides aminés essentiels pour les réactions d'oxydation (C400 et T268) permet la conversion des enzymes pour la catalyse de réactions jamais réalisées dans les cellules vivantes.

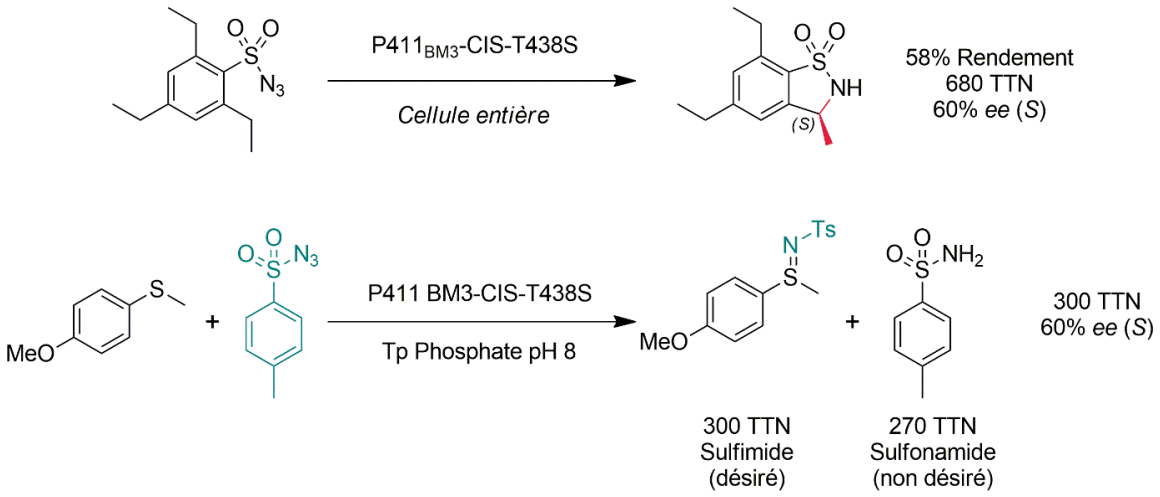
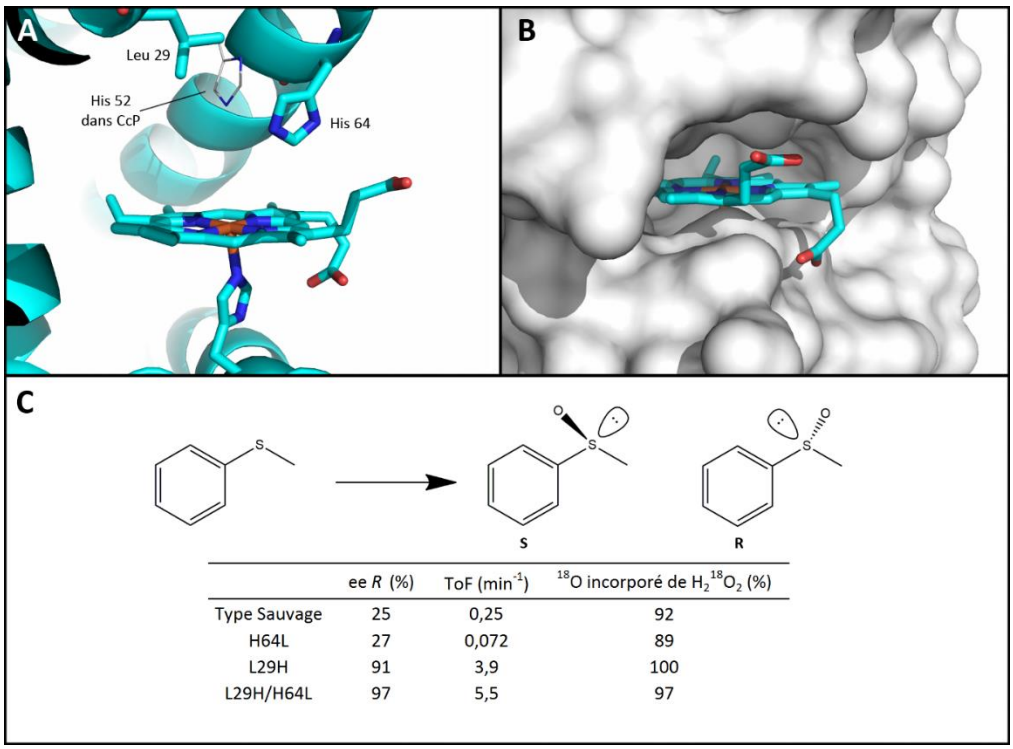


Figure 35 : Réaction d'insertion intramoléculaire de nitrène au sein d'une liaison C-H (haut) et S-H (bas), catalysée par les variants P411 de cytochrome P450 (selon l'équipe de F. Arnold [93,94]).

II.2.1.2 La myoglobine : une architecture protéique propice à l'ingénierie de catalyseurs biohybrides

La myoglobine, à l'instar de l'hémoglobine, est une protéine qui transporte le dioxygène grâce à une porphyrine de fer. Elle est en revanche un piètre catalyseur des réactions d'oxydation car son site actif ne participe pas à l'activation de la coupure hétérolytique du H_2O_2 pour la formation de l'espèce active Fe (IV) oxo (composé I) comme dans le cas des peroxydases héminiques de type P450. L'étude de la crevasse contenant la porphyrine indique également qu'elle est encombrée, empêchant l'entrée d'éventuels substrats. Les équipes de Tasamasa Hayashi, Yoshihito Watanabe et Takafumi Ueno ont étudié l'effet des résidus du site actif sur l'époxydation énantiosélective du thioanisole (methyl phenyl sulfide) en methyl phenyl sulfoxyde. Il y a 20 ans déjà, une comparaison de l'environnement de la porphyrine entre la myoglobine et la Cytochrome C peroxydase a suggéré à l'équipe de Watanabe la mutation L29H pour placer un résidu histidine à une position équivalente à l'histidine 52 de la Cytochrome C peroxydase (figure 36A). En effet, cette mutation a permis de conduire à une augmentation sensible de l'activité et une bonne énantiosélectivité, malgré une vitesse de réaction toujours très faible (figure 36C).[95]



	ee *R* (%)	ToF (min^{-1})	^{18}O incorporé de $H_2{}^{18}O_2$ (%)
Type Sauvage	25	0,25	92
H64L	27	0,072	89
L29H	91	3,9	100
L29H/H64L	97	5,5	97

Figure 36 : A. Acides aminés proximaux et distaux dans la myoglobine. La leucine 29 et l'histidine 64 ne participent pas à l'acte catalytique. La position relative de l'histidine de la Cytochrome C peroxydase est représentée en lignes fines. B. Hème de la myoglobine vu de l'extérieur, on remarque que la poche est très encombrée et ne permet pas l'entrée de substrat. C. Réaction de sulfoxydation énantiosélective du thioanisole catalysée par les différents mutants de la myoglobine utilisés par Watanabe, Hayashi et Ueno .[95]

Des études ultérieures ont montré qu'une histidine à une position différente (F43H) était un catalyseur général acidobasique pouvant conduire à la formation putative de la forme ferryl radical cation (O=Fe(IV) $Por^{+\bullet}$). La structure cristallographique a confirmé que la distance entre le N^{ϵ} et le Fe (III) de l'hème était effectivement similaire à celles mesurées dans les peroxydases (L29H : 6.6 Å, F43H : 5.3 Å, H64 : 4.5 Å comparé aux 5.5-6 Å des peroxydases).[96] Dans une étude concomitante, la mutation H64D permit également la formation H_2O_2 dépendante du composé 1, qui a alors pu être observé pour la première fois spectroscopiquement. La sulfoxydation du thioanisole fut étudiée pour ce mutant mais malgré une forte accélération de la réaction (145 min^{-1}) un mauvais excès énantiomérique fut constaté (6% *ee*).[97]

En 1999, dans une approche pionnière mêlant ingénierie génétique et chimique, le même groupe remplaça l'hème en ajoutant des carboxylates sur les groupements propionates (figure 37). Ainsi modifiée, la myoglobine sauvage présentait déjà une affinité et réactivité supérieure dans l'oxydation du guaiacol.[98] Le mutant H64D portant l'hème modifié avait une activité démultipliée (k_{cat} passant de 10,8 à 336 min^{-1}) se rapprochant de celle d'une peroxydase naturelle comme la peroxydase de raifort (HRP) (1 920 min^{-1}).[99] L'efficacité catalytique (k_{cat}/K_M) était seulement 3 fois inférieure à celle de la HRP (23 000 $M^{-1}.s^{-1}$ et 72 000 $M^{-1}.s^{-1}$ respectivement).

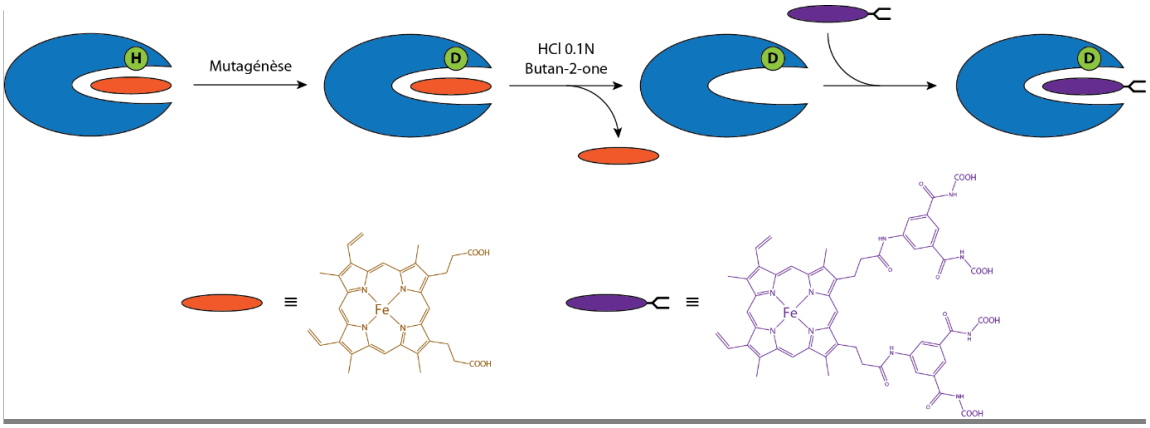


Figure 37 : Stratégie de modification de la myoglobine par l'équipe de Watanabe et Ueno. Le gène codant la myoglobine est mutée par génie génétique, le variant H64D étant alors produit avec l'hème endogène. Le variant est ensuite traité au Butan-2-one afin de retirer l'hème pour ensuite le remplacer par un hème synthétique (selon [98]).

Ces travaux illustrent l'efficacité de la modification du groupement prosthétique d'une protéine pour modifier son activité, même si la porphyrine modifiée ne changeait pas les propriétés d'oxydation du métal. Il est intéressant de noter que la modification de l'hème, en modifiant le K_D de la protéine pour ses substrats hydrophobes, joue le rôle d'une poche de fixation, rôle normalement assuré par la protéine. Cependant, la méthode de

remplacement de l'hème de la myoglobine est trop lourde et peu efficace pour envisager une méthode d'évolution dirigée sur cette architecture.

Le groupement prosthétique peut être également radicalement modifié. En effet, en insérant différents salophènes de chrome en lieu et place de l'hème, ce même groupe parvint à oxyder le thioanisole, avec des sélectivités toutefois modestes (figure 38). Il fut prouvé que le salen, de par sa conformation planaire, s'insérait bien à la place de l'hème et que le chrome était coordiné par la même histidine.[100]

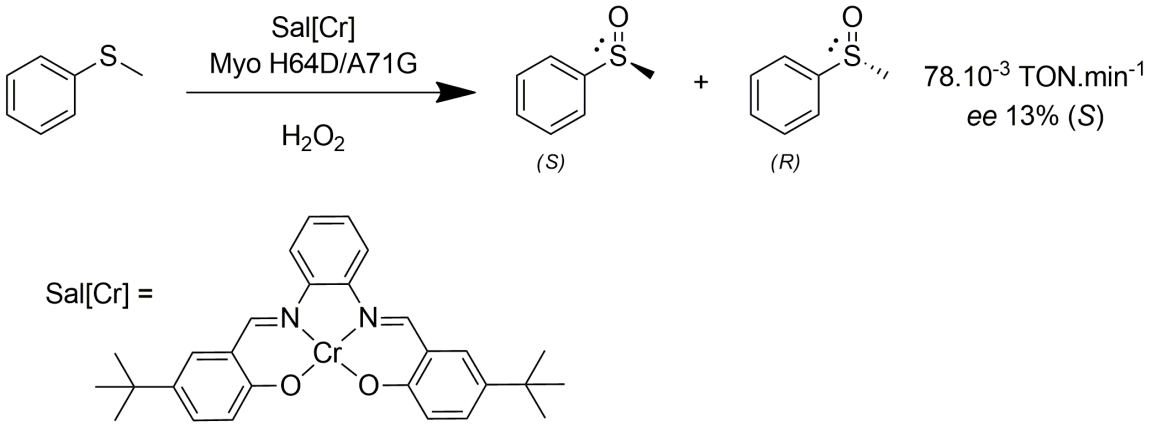


Figure 38 : Complexe salophène de chrome inséré dans la l'apo-myoglobine (selon [100]).

Plus récemment, l'équipe de Rudi Fasan s'est intéressée à des réactions de transfert de carbène catalysées par la myoglobine. En effet, plusieurs études ont démontré que les porphyrines de fer sont capables de promouvoir la cyclopropanation d'oléfines ainsi que la fonctionnalisation de liaisons X-H (X = C, Si, N, S) en présence de diazoacétates.[101,102] Dans cette réaction, le carbène est stabilisé sur le métal après libération d'une molécule de diazote, avant de se lier à la liaison alcène ou l'atome à fonctionnaliser. Ce type de réaction serait particulièrement intéressant pour la conception de produits de haute valeur ajoutée, avec l'avantage d'utiliser des composés simples, et de libérer uniquement du diazote comme sous-produit.

Dans une approche de mutations sélectives limitées à 4 acides aminés tapissant la crevasse (L29, F43, H64, V68), le mutant H64V/V68A s'est montré efficace pour catalyser dans la cyclopropanation du styrène en présence de diazoacétate d'éthyle (EDA) (figure 39, milieu). Le produit de cette réaction contenant deux centres de chiralité, quatres stéréoisomères sont possibles (deux paires d'énantiomères). Ce mutant présentait une grande énantiosélectivité et diastéréosélectivité pour une large gamme de substrats dérivés du styrène.[103] Dans une approche similaire, le même mutant s'est avéré capable d'effectuer l'insertion intermoléculaire de carbène dans une liaison N-H de différentes aryl amines substituées en *ortho*, *meta* et *para* avec de bons rendements (mais pas de centre stéréogène créé) (figure 39, haut).[104] La nécessité de conditions fortement

réductrices (dithionite de sodium) et anaérobies montrait que le fer ferreux était responsable de la fixation de l'EDA, en compétition avec la fixation du dioxygène. Cette même équipe est enfin parvenue à effectuer la synthèse de thioéthers par transfert de carbène sur un grand nombre de substrats volumineux et hydrophobes avec de bons nombres de cycles catalytiques (TON jusqu'à 5400 avec le mutant H64V/L29A), mais avec une énantiosélectivité peu spectaculaire (49% *ee* maximum avec le mutant F43V). Il s'agit du premier exemple d'insertion d'un carbène au sein d'une liaison S-H catalysé par une protéine après ingénierie rationelle (figure 39, bas).[105]

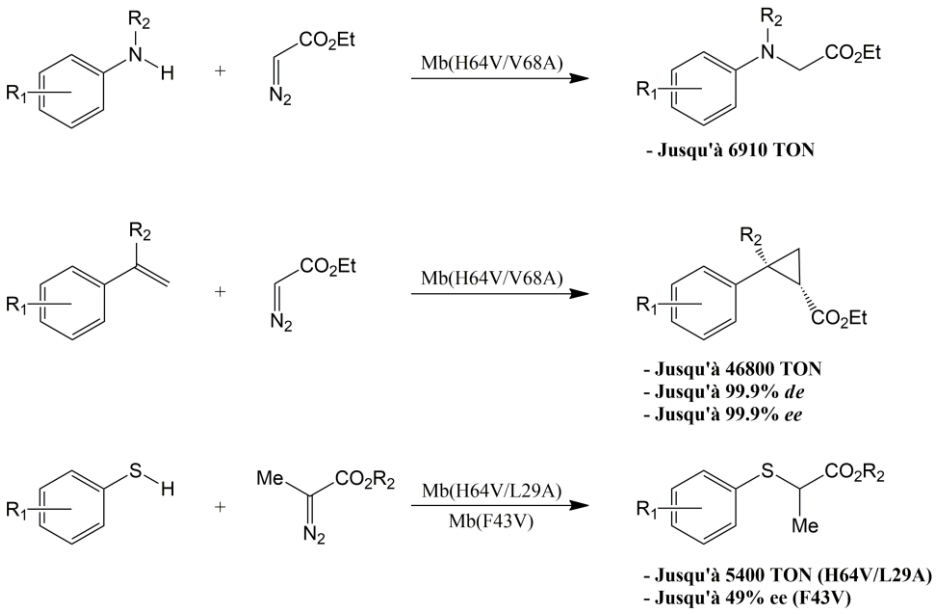


Figure 39 : Différents types de transfert d'alcènes par le biais de composés diazo, catalysées par des mutants de myoglobine (Equipe de Fasan) Haut : Insertion dans une liaison N-H, milieu : cyclopropanation, bas : insertion dans une liaison S-H.

L'équipe de Clark et Hartwig a très récemment utilisé une stratégie de double évolution (chimique & génétique) pour modifier les propriétés catalytiques de la myoglobine.[106] Considérant que les possibilités de catalyse étaient limitées par le métal porté par la porphyrine, leurs travaux ont d'abord consisté en un remplacement du fer par d'autres métaux non présents habituellement dans les molécules biologiques. Afin d'obtenir facilement la forme *apo* de la myoglobine, leur stratégie a consisté, dans un premier temps, à optimiser une production de la protéine sans son groupement prosthétique. L'utilisation d'*E.coli* comme hôte dans un milieu minimum déplété en fer et à basse température a permis la production à haut rendement de l'apoprotéine (Myo) repliée (entre 5 et 40 mg/L de culture). Huit variants pour l'histidine proximale (H93) ont été ainsi produits avec succès. Après reconstitution avec des cofacteurs protoporphyrine IX (PIX) portant différents métaux (Fe(Cl), Co(Cl), Cu, Mn(Cl), Rh, Ir(Cl),Ir(Me), Ru(CO), Ag), les nouveaux biohybrides ont été le sujets de tests de catalyse de transfert de carbène

intramoléculaire et d'addition de carbènes dans une fonction alcène interne (figure 40 a et b respectivement).

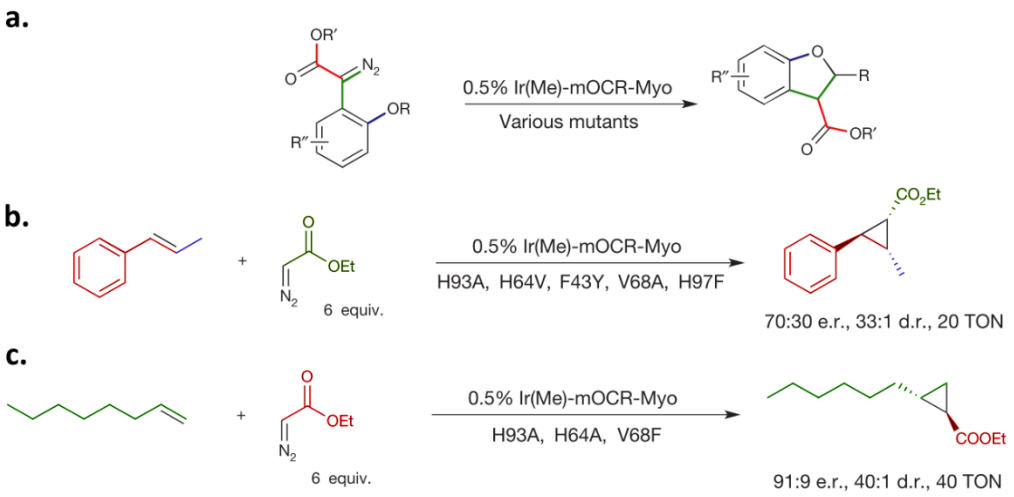


Figure 40 : Réactions catalysées par les biohybrides myoglobine reconstituées avec une porphyrine d'iridium. a. Réaction de transfert de carbène intramoléculaire. b. Réaction de cyclopropanation par addition de carbène dans une fonction alcène interne. c. Réaction d'addition de carbène dans une fonction alcène non activée

Le biohybride Ir(Me)-PIX ⊂ Myo (⊂ signifie l'inclusion d'un complexe, au sein de la protéine Myo ici) est capable de catalyser les deux types de réaction avec le plus haut rendement dans tous les biohybrides testés. Une stratégie d'évolution dirigée s'approchant de l'ISM a ensuite été effectuée utilisant l'Ir(Me)-PIX, en modifiant de manière incrémentielle les acides aminés de la crevasse de plus en plus éloignés, utilisant un alphabet d'acides aminés restreint (hydrophobes). Une série de 450 variants au total ont été criblés pour la réaction de transfert de carbène intramoléculaire (figure 40a) et l'addition de carbène dans un alcène aliphatique (figure 40c). Des variants ayant une énantiosélectivité en faveur de chacun des énantiomères ont été trouvés pour chaque substrat concernant l'insertion de C-H, avec un TON allant jusqu'à 7260. Pour l'addition de carbène dans des alcènes non activés, les activités obtenues sont moindres (40 TON) mais présentent de bonnes énantio- et diastéréosélectivités.

Ces exemples basés sur la stratégie hôte-invité illustrent la puissance de l'ingénierie du cofacteur métallique et de l'architecture protéique pour accéder à réactions non naturelles avec de bonnes énantiosélectivités. Cependant les expériences d'évolution dirigée se limitent encore à des protéines purifiées individuellement, ne permettant que l'analyse de quelques centaines de variants (ce qui représente déjà une masse de travail conséquente). Ainsi, les résultats *in cellulo* des enzymes de type P411 sont très

prometteurs pour la création de cribles catalytiques qui pourraient être utilisables dans des approches d'évolution dirigée avec des librairies de grande taille.

II.2.1.3 La xylanase A

Des travaux précurseurs de l'équipe de Satoshi Nakamura présentaient la xylanase B, une enzyme dégradant les polysaccharides linéaire β-1,4-xylane en monomères xylose, de *T. maritima* comme pouvant accepter une porphyrine de fer synthétique au sein de sa structure. Cette association formant une hémoprotéine artificielle capable de fixer l'oxygène de la même manière que la myoglobine.[107] L'équipe de Jean-Pierre Mahy a porté son choix sur la xylanase A de *S. lividans* (figure 41). Cette protéine est en effet simple à produire en grande quantité dans son organisme d'origine, elle est thermostable et sa structure cristallographique est connue. Cette protéine d'architecture « TIM barrell » présente une charge globale positive à pH neutre, et possède une large poche. Cette équipe a montré que la xylanase A est capable d'incorporer un cofacteur tel qu'une porphyrine anionique de type Fe(III)(TpCPP) avec une stœchiométrie 1:1 et un K_D de l'ordre du micromolaire.[108]

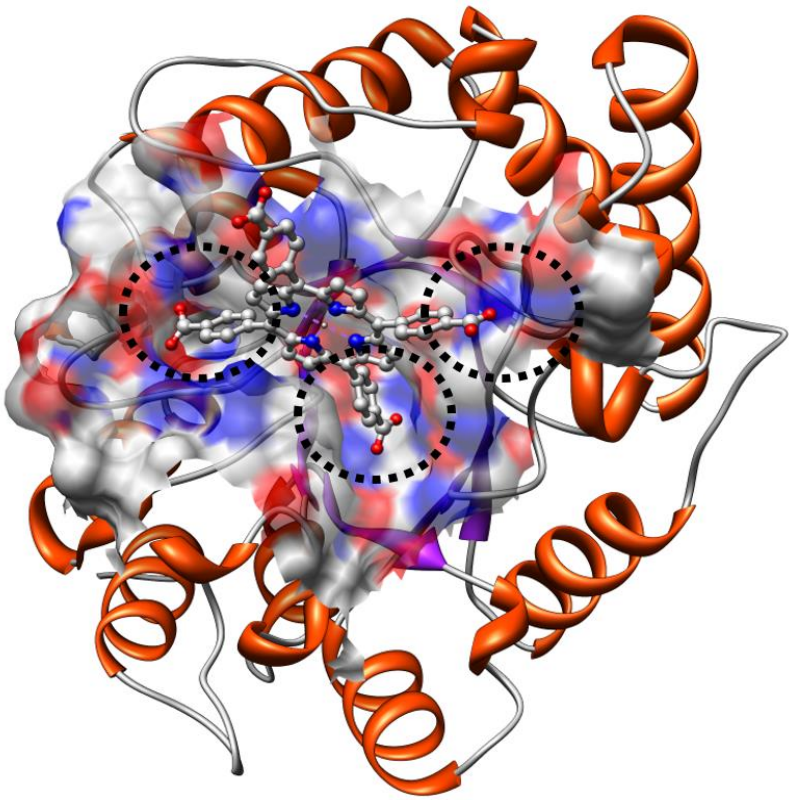
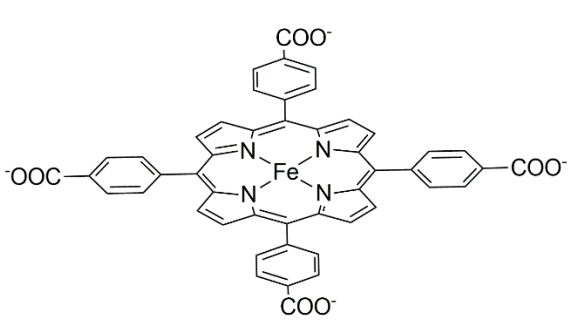
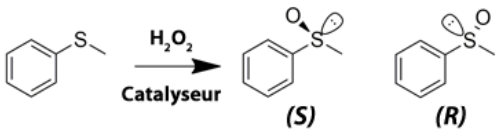


Figure 41 : Surface de potentiel électrostatique de la Xylanase 10A de S. lividans à pH 7 avec une porphyrine anionique modélisée à sa surface (modèle boules et bâtons), les carboxylates de la porphyrine étant censés former des liaisons ioniques avec les patchs positifs de la crevasse.

Ricoux *et al.* ont donc utilisé des porphyrines anioniques Fe(III)(TpCPP) et Fe(III)(TpSPP)) (figure 42) pour l'insertion au sein de la protéine via la stratégie hôte-invité. Il a été

montré que le biohybride Fe(III)(TpCPP) ⊂ Xln10A est capable de catalyser la sulfoxydation du thioanisole avec un excès énantiomérique pouvant atteindre 40% en faveur de l'énantiomère (*S*). Il a été montré que la présence d'imidazole augmente la vitesse de réaction par un facteur de plus de 3, soit en se positionnant comme ligand axial du fer, soit en catalysant la coupure hétérolytique de la liaison O-O par un mécanisme acide-base. [109] De même, une porphyrine de manganèse insérée au sein de la crevasse a pu catalyser l'oxydation du métoxystyrène par le $KHSO_5$ avec un excès énantiomérique de 80% envers la forme (*R*).[110]





Catalyseur	TOF (h^{-1})	*ee* (%)
Fe(III)(TpSPP)	37 ± 1	–
Fe(III)(TpSPP) ⊂ XylA	20 ± 1	10 ± 1
Fe(III)(TpCPP)	25 ± 1	–
Fe(III)(TpCPP) ⊂ XylA	18 ± 1	36 ± 3
Fe(III)(TpCPP) ⊂ XylA + ImH	63 ± 2	40 ± 3

Figure 42 : Porphyrine Fe(III)(TpCPP) insérée au sein de la xylanase A et réaction catalysée (selon [109])

II.2.1.4 Les albumines sériques

Les albumines sont des protéines sanguines permettant le transport d'un nombre varié de petites molécules, préférentiellement hydrophobes, telles que les acides gras, hormones, hémines ou molécules exogènes. Leur capacité de fixation de cofacteurs a été mise à profit à de nombreuses reprises pour la création de métalloenzymes. Elles sont de plus aisément disponibles en grande quantité par extraction à partir des sérums de mammifères. Ce sont des protéines de grande taille (71 kDa pour l'albumine humaine (HSA)) composées d'hélices alpha et de forme complexe. Cette protéine présente plusieurs poches hydrophobes permettant la fixation des différents ligands. La figure 43 présente la HSA contenant sept molécules d'acide palmitique, chacune étant dans une poche différente.

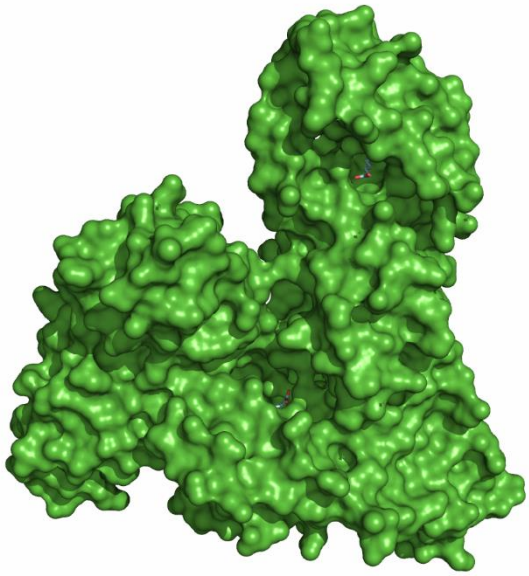


Figure 43 : Les albumines sériques (ici la HSA) sont des protéines de grande taille et de forme complexe. Elles possèdent plusieurs poches hydrophobes où se fixent des molécules principalement hydrophobes. Deux molécules d'acide palmitique sont visibles en bâtons bleus et rouges (pour les oxygènes) sont visibles sur cette structure, sur un total de sept.

Atif Mahammed et Zeev Gross ont publié l'utilisation des albumines de sérum Humain (HSA), de porc (PSA), de mouton (SSA), de bovin (BSA) et de lapin (RSA) pour l'insertion de corroles (figure 44) amphiphiles de fer et manganèse. La fixation de ces deux partenaires présentait une stœchiométrie 1:1 et une affinité nanomolaire, indiquant un ancrage spécifique permettant d'accéder potentiellement à des sélectivités de réaction. Ces différents biohybrides ont permis la sulfoxydation du thioanisole et d'autres sulfides par le H_2O_2 avec des excès énantiomériques jusqu'à 98%.[111]

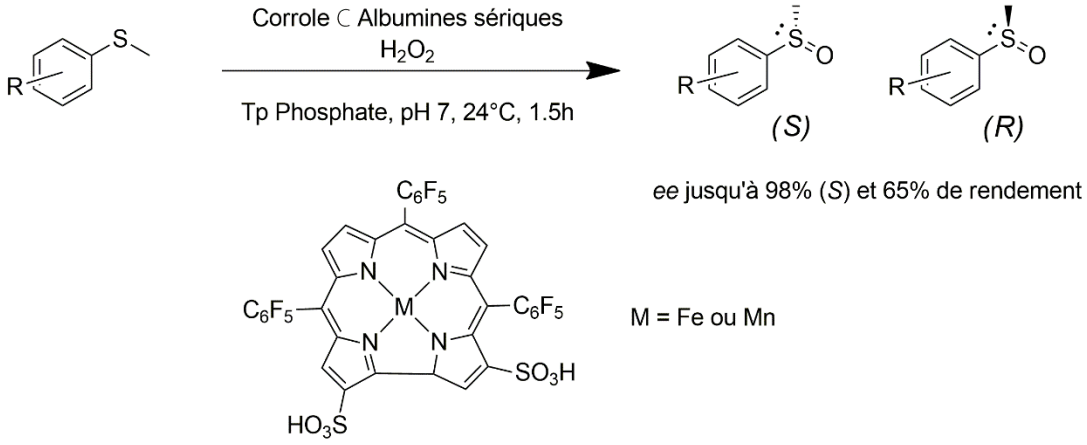


Figure 44 : Sulfoxydation du thioanisole catalysée par des coroles insérées dans différentes albumines sériques (d'après [111])

Partant de ce principe, l'équipe de Reetz a utilisé en 2006 une phtalocyanine de cuivre insérée dans différentes albumines sériques pour catalyser la cycloaddition de

Diels-Alder entre le cyclopentadiène et différentes chalcones (figure 45). L'excès énantiomérique de la réaction catalysée par ces biohybrides a atteint 98% dans la forme *endo*, avec une sélectivité endo/exo de 91:9 (la forme *endo* est privilégiée thermodynamiquement dans cette réaction).[112]

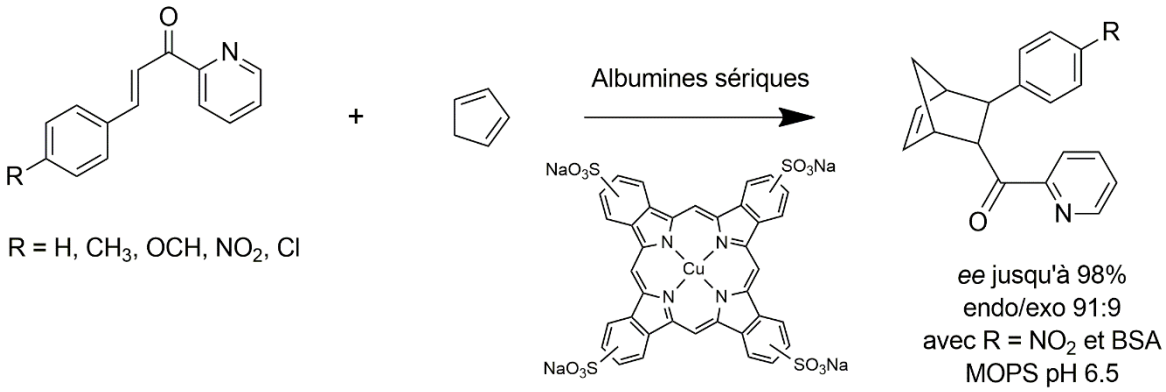


Figure 45 : Cycloaddition de Diels-Alder par des biohybrides d'albumines et phtalocyanines de cuivre (d'après [112])

Enfin, l'équipe de Stéphane Ménage a utilisé l'albumine Humaine (HSA) pour l'insertion de salènes de manganèse. Ce biohybride a pu catalyser la réaction de sulfoxydation du thioanisole par l'hypochlorite de sodium (NaOCl) avec une excellente chimiosélectivité (absence de sur-oxydation) envers le produit sulfoxide par rapport au sulfone (figure 46). Cette sélectivité a été imputée au caractère hydrophobe de la poche catalytique, empêchant la fixation du substrat mono-oxydé. Il a été montré que la sélectivité décroit quand la stœchiométrie salen:HSA dépasse 1:1 et donc qu'un site de fixation préférentiel est présent. Cependant, ces biohybrides ne sont pas énantiosélectifs.[113]

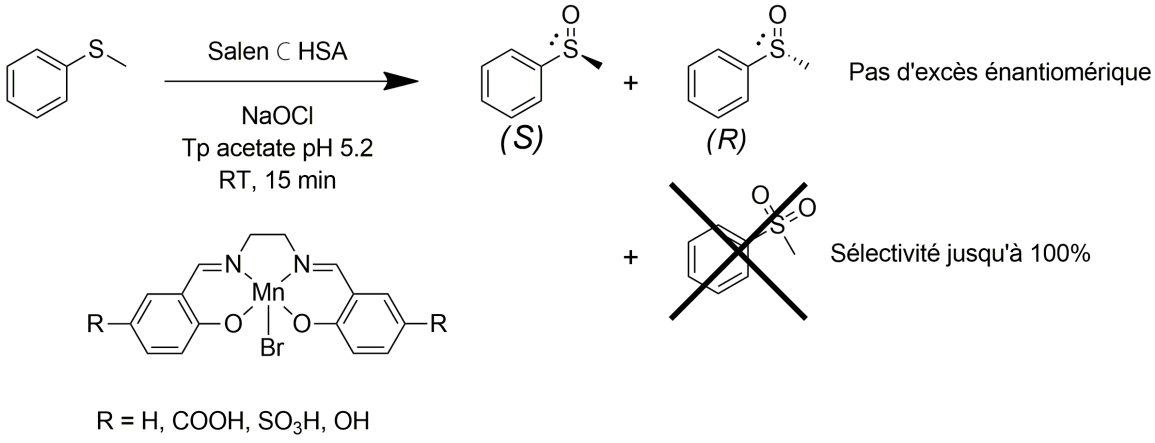


Figure 46 : Sulfoxydation chimiosélective du thioanisole par des complexes salen de manganèse insérés dans la HSA (selon [113])

Dernièrement, l'équipe de Liang Hong a inséré dans l'albumine sérique bovine (BSA) un complexe de type salen de cobalt pour la sulfoxydation du thioanisole (et d'autres sulfides

prochiraux) par le peroxyde d'hydrogène (H_2O_2). Après optimisation des conditions de catalyse concernant la température, la concentration des substrats et les cosolvants, des taux de conversion de 98%, 100% d'énantiosélectivité pour un excès énantiomérique de 50% pour la forme (*R*) ont été obtenus pour le thioanisole. Des énantiosélectivités plus importantes ont été obtenues pour d'autres dérivés, comme le *meta*-chlorothioanisole (71% *ee*) ou *para*-chlorothioanisole (87% *ee*) mais au détriment de la chimiosélectivité (80 et 68% respectivement).[114]

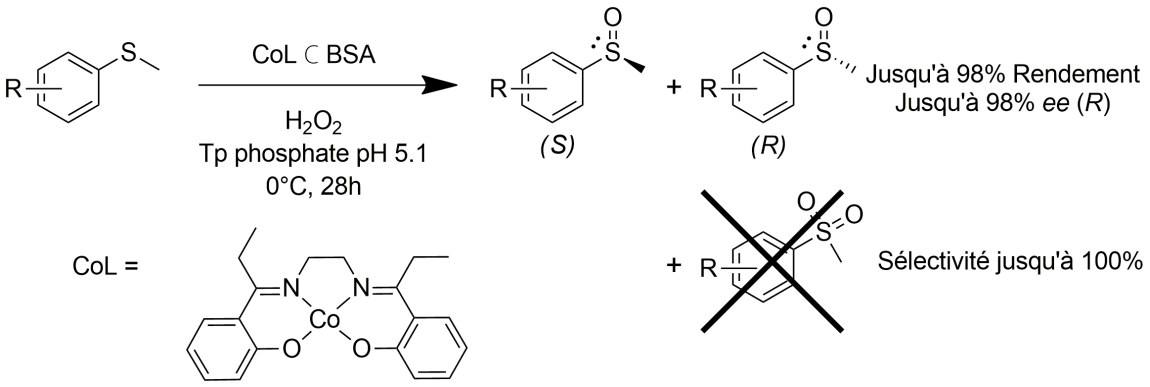


Figure 47 : Sulfoxidation énantio et chimiosélective de dérivés du thioanisole par des complexes de cobalt inclus dans la BSA. (selon [114])

Les albumines sériques possèdent l'avantage d'être disponibles en grande quantité commercialement, mais elles sont peu utilisées pour de l'ingénierie génétique. Une telle approche nécessite un système d'expression efficace et compatible avec les méthodes de sélection.[115] La HSA peut cependant être produite dans l'hôte *P. pastoris*, et un système d'évolution dirigée avec cet hôte a été publié en 2010, ce qui ouvre la voie à l'optimisation de ces nouvelles métalloenzymes.[116]

II.2.1.5 La protéine LmrR

Très récemment, l'équipe de G. Roelfes a utilisé la crevasse hydrophobe résultant de la dimérisation du facteur de transcription procaryote Lactococcal multidrug resistance Regulator (LmrR) pour l'insertion non covalente d'un cofacteur métallique de type phénanthroline.[117] Cette protéine est un homodimère dans lequel deux acides aminés tryptophanes équivalents sur chacun des monomères et situés l'un en face de l'autre dans la structure semblaient être idéalement positionnés pour l'insertion du ligand et sa stabilisation par des « π stacking » (figure 48). Le biohybride résultant de cet assemblage supramoléculaire est capable de catalyser une réaction de Friedel-Kraft avec des excès énantiomériques jusqu'à 94%. Le remplacement de ces tryptophanes par des alanines conduit à une diminution de l'affinité de la protéine pour la phénanthroline (K_D passant de 2.6 µM à 45 µM) et à une perte de l'énantiosélectivité. Ceci a permis de confirmer l'importance de ces résidus tryptophane pour l'insertion du cofacteur au sein du dimère.

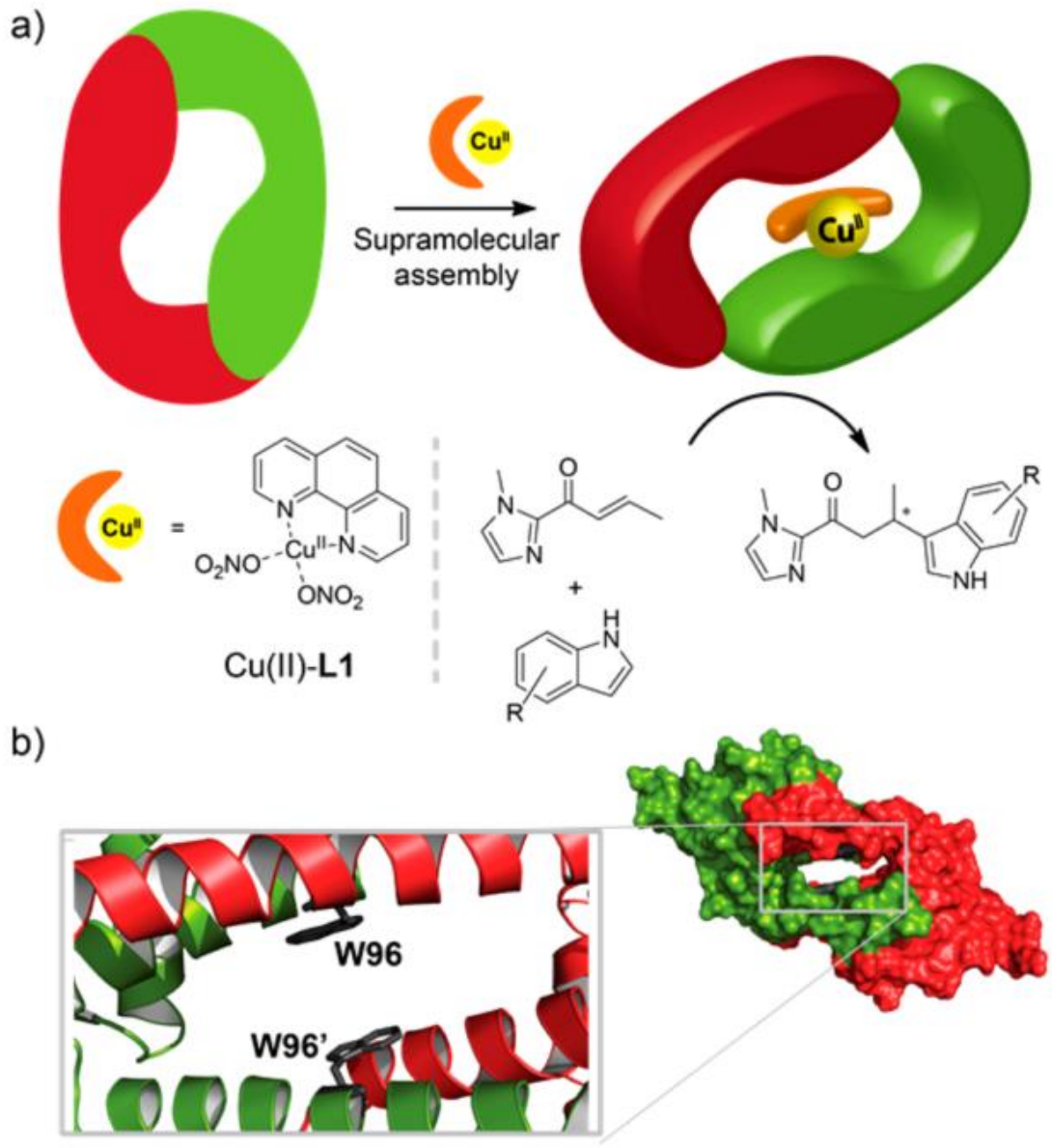


Figure 48 : a) Représentation schématique de l'insertion d'une phénanthroline au sein du dimère LmrR (dessus). Réaction de Friedel-Kraft catalysée par un complexe phénanthroline de cuivre (dessous) b) En bâtons sont représentés les deux tryptophanes responsables du « π stacking » (selon [117]).

II.2.1.6 Ingénierie d'anticorps : conversion de reconnaisseurs en catalyseurs

L'idée d'utiliser des anticorps en vue de construire des catalyseurs biohybrides a été exploitée durant les années 90. Le principe était d'utiliser la propriété intrinsèque de reconnaissance spécifique des anticorps en effectuant des sélections contre un cofacteur métallique afin d'effectuer un ancrage supramoléculaire spécifique. Le cofacteur de choix étant la porphyrine, un grand nombre d'études se sont portées sur la sélection d'anticorps anti-porphyrines diversement substituées, contenant un métal ou non, pour effectuer des

réactions d'oxydation de type P450 ou peroxydases.[118–127] Ces différents complexes, nommés hémoabzyme pour « hemo-antibody-enzyme », se sont malheureusement révélés de modestes catalyseurs, notamment en raison du manque d'un acide aminé proximal comme la cystéine du cytochrome P450 (voir *ci-dessus*).

Afin d'obtenir une sélectivité de substrat, Shai Nimri et Ehud Kenan ont utilisé une porphyrine hydrosoluble d'étain Sn (IV) portant un ligand axial α-naphtoxy. Cette construction avait été choisie pour mimer l'état de transition d'une réaction de sulfoxydation de sulfides aromatiques ainsi que l'approche de l'oxydant iodosylbenzène sur le centre métallique. Un anticorps sélectionné fixant spécifiquement ce complexe avec une affinité nanomolaire a ensuite été testé pour la catalyse en présence d'une porphyrine de ruthénium portant les mêmes groupements fonctionnels. Cet objet a permis la sulfoxydation stéréosélective du thioanisole par l'iodosylbenzene avec un excès énantiomérique de 43% pour la forme (*S*) (figure 49).[128] Cette approche est un exemple canonique d'un biohybride, avec la protéine responsable de l'affinité pour le substrat et le cofacteur métallique responsable de l'acte catalytique. Cependant, une ingénierie génétique de la protéine en vue d'améliorer ses propriétés n'a pas été tentée.

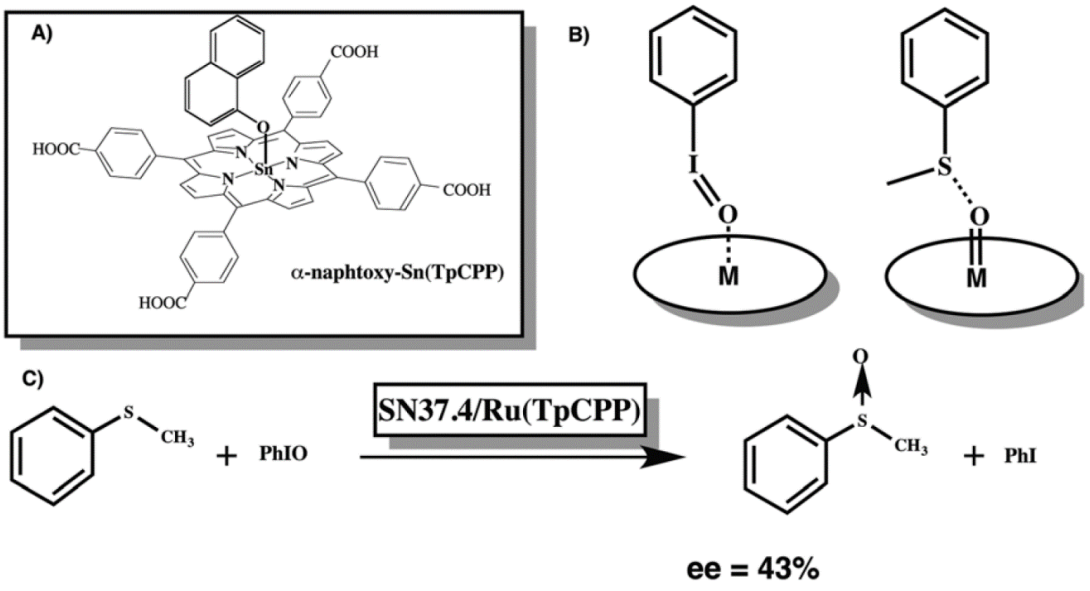


Figure 49 : A) Porphyrine utilisée par Nimri et Keinan pour la sélection d'un anticorps B) État de transition de la réaction d'oxydation du métal (gauche) ou de la sulfoxydation de thioanisole (droite). C) Réaction de sulfoxydation énantiosélective du thioanisole par un anticorps complexé avec une porphyrine de ruthénium Ru(TpCPP) (selon [85]).

L'équipe de Mahy a exploré l'idée d'utiliser un ligand possédant déjà un acide aminé coordinant le métal d'une porphyrine en position proximale : la microperoxydase 8 (MP-8). En effet, la microperoxydase est issue de la digestion trypsique du cytochrome C de cheval et est constituée d'un peptide de 8 acides aminés (CAQCHTVE) lié à l'hème par

deux cystéines. L'histidine 4 du peptide constitue le ligand axial de l'ion métallique. Plusieurs carboxylates pouvaient également servir de point d'ancrage à l'anticorps (figure 50). De plus, comme son nom l'indique, la MP-8 possède une activité de type peroxydase et monooxygénase, et catalyse un grand nombre de réactions telles que l'oxydation d'o-dianisidine, acide 2,2'-azino-bis(3-éthylbenzothiazoline-6-sulphonique) (ABTS) et guaiacol [129,130], la nitration de phénols [131] ou la sulfoxydation énantiosélective de sulfures.[132] L'immunisation de souris contre la MP-8 a permis d'isoler un anticorps monoclonal (3A3). Il a été montré que la protéine permet de protéger le cofacteur vis-à-vis de sa dégradation l'empêcher sa dimérisation qui est observée en solution aqueuse.[133] La sulfoxydation du thioanisole par le peroxyde d'hydrogène catalysée par le couple 3A3-MP8 a donné un excès énantiomérique de 45% envers la forme (*R*).

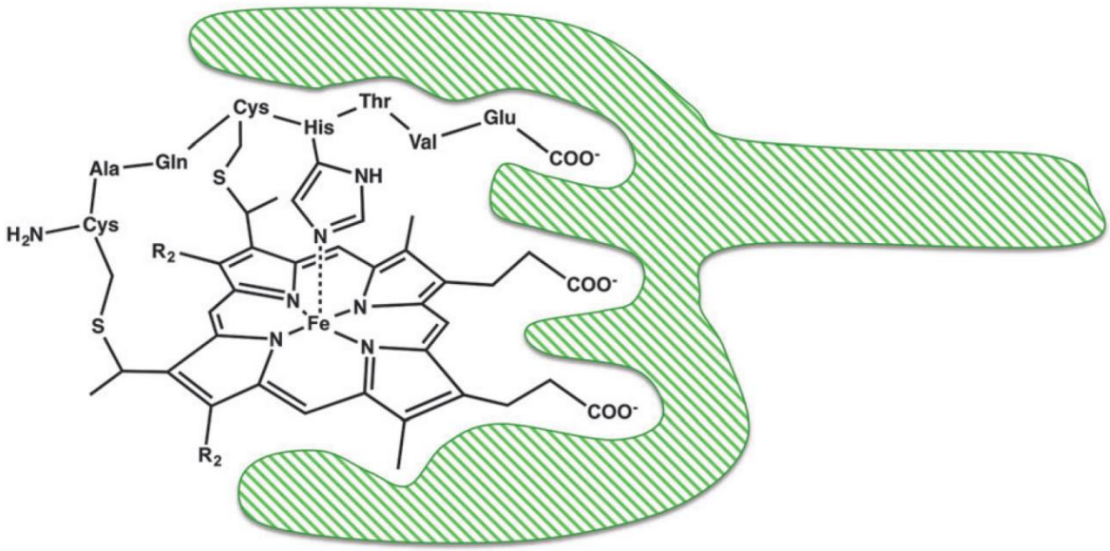


Figure 50 : Vue schématique de l'anticorps 3A3 dirigé contre la microperoxydase 8 (selon [85])

L'incorporation d'une porphyrine dans un anticorps permet donc la protection de la porphyrine en empêchant sa dimérisation et son auto-dégradation oxydative, et en provoquant aussi une diffusion contrôlée de l'oxydant vers le métal. Cette incorporation précise du complexe est également source de sélectivité. Cependant les anticorps sont des protéines complexes difficiles à produire sous forme recombinante dans des hôtes tels qu'*E.coli*, ce qui limite leur optimisation par ingénierie génétique et limite la production en grosse quantité. Enfin, la géométrie d'insertion des porphyrines dans la crevasse des anticorps est généralement homologue à celle de peroxydases, favorisant une approche du substrat par le bord de la porphyrine, alors que les cytochromes de type P450 privilégient une approche faciale au sein d'une large crevasse hydrophobe. Ces raisons ont conduit les différentes équipes de par le monde à porter leurs efforts vers l'utilisation de protéines hôtes différentes, stables et plus faciles à produire.

II.2.1.7 Conclusion de la stratégie hôte-invité

Les différentes stratégies hôte-invité que nous avons abordées dans cette partie montrent que le positionnement d'un complexe métallique au sein d'une crevasse protéine peut être une approche efficace pour générer de la catalyse énantiosélective. Cependant, ces exemples présentent tous un même inconvénient : l'affinité de la protéine pour le ligand dépend du type de ligand utilisé d'une part et des acides aminés en contact avec le cofacteur d'autre part. Ceci limite donc les possibilités d'ingénierie chimique et génétique, car une modification trop importante de chacun des deux partenaires risque de déstabiliser l'interaction. Pour tenter de répondre à cette limitation, une solution est de séparer la partie du ligand responsable de l'interaction de la partie responsable de l'acte catalytique. C'est le principe de la stratégie du cheval de Troie.

II.2.2 La stratégie du cheval de Troie

II.2.2.1 Le couple (strept)avidine/biotine : cheval de bataille de cette stratégie

La stratégie du « cheval de Troie », également appelé ancrage supramoléculaire, consiste à coupler le cofacteur métallique à une molécule possédant une affinité pré-existante pour la protéine. Le couple avidine-biotine, qui possède une affinité des plus fortes connues (avec un K_D de l'ordre du femtomolaire) dans la nature entre une protéine et une petite molécule, a naturellement été exploité par plusieurs équipes pour ancrer à la protéine un complexe métallique lié à une biotine et explorer plusieurs réactions catalytiques.

Les travaux de Georges Whitesides ont ouvert la voie à ce type d'approche.[134] Ces travaux consistaient en un couplage d'un complexe phosphine de rhodium avec la biotine pour effectuer la réduction d'acide α-acétamidoacrylique en N-acétylalanine avec un excès énantiomérique de 40% en faveur de l'énantiomère S.

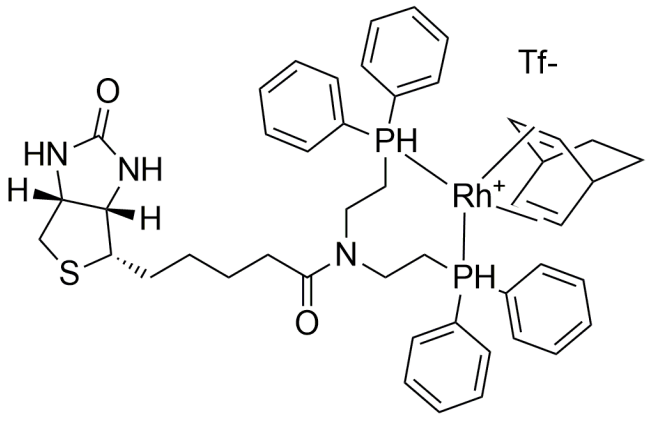
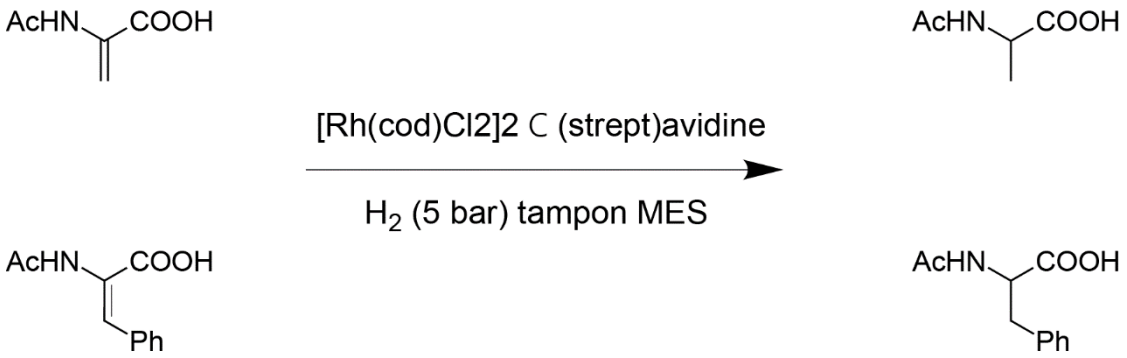


Figure 51 : Complexe de rhodium couplé à une biotine utilisé par Whitesides pour insertion dans l'avidine

L'équipe de Thomas R. Ward s'est spécialisée dans cette approche, utilisant la streptavidine au lieu de l'avidine. La streptavidine, plus facile à obtenir de manière recombinante chez *E.coli,* est également une protéine homotétramérique, liant la biotine avec une affinité sub-picomolaire, qui diffère de l'avidine par son point isoélectrique (6.2 pour la streptavidine, 10.4 pour l'avidine). En répétant l'expérience de Whitesides avec la streptavidine, Ward atteint des excès énantiomériques conséquents (92%) en faveur de l'énantiomère opposé, ainsi que pour la réduction de l'acide acétamidocinnaminique en (R)-acétamidophénylalanine (figure 52). Cette inversion dans l'énantiosélectivité de la réaction a pu être expliquée par des changements dans le type de liaison entre la biotine et le complexe. Des expériences de mutations de l'acide aminé S112 qui, par étude cristallographique avait été identifié comme en face du complexe métallique une fois logé dans la poche de la streptavidine, ont permis de confirmer le positionnement du complexe au sein de la crevasse de la protéine.[135–137]



Ligand	Protein	% *ee* (a)	% conv (a)	% ee (b)	% conv (b)
Biot-L	WT avi	37 (*S*)	90	-	-
Biot-L	WT sav	92 (*R*)	Quant.	93 (*R*)	84
Biot-L	Sav S112G	96 (*S*)	Quant.	94 (*R*)	77
Biot-4^{meta}-L	Sav S112K	63 (*S*)	Quant.	88 (*S*)	89
Biot-(R)-Pro-L	Sav S112W	95 (*S*)	Quant.	95 (*S*)	Quant

Figure 52 : Hydrogénations énantiosélectives par des complexes de rhodium couplés à la biotine, insérés dans l'avidine ou la streptavidine (d'après [63])

L'équipe de Reetz a ensuite tenté, pour la première fois en 2006, une approche d'évolution dirigée de la streptavidine pour améliorer l'énantiosélectivité de l'hydrogénation d'un ester de l'acide acétamidoacrylique. Malheureusement, les complications dans l'expression des gènes mutants ont empêché une production parallélisée.[138] Une expérience de mutagenèse itérative du site actif (ISM) a été employée pour générer une bibliothèque de variants et le criblage a permis d'identifier un variant

pour lequel un excès énantiomérique de 65% fut atteint. Il est à noter que la plupart des mutations, notamment les plus distantes du complexe, rendaient les protéines insolubles.

L'équipe de Ward a entrepris, également en 2006, une approche d'ingénierie par mutagenèse à saturation sur une unique position de la streptavidine pour catalyser le transfert d'hydrogénation de cétones avec des ligands de type « tabouret de piano » (figure 53). En faisant varier à la fois le ligand arène coiffant le métal et la position S112 sur la protéine, Letondor *et al.* ont obtenu des variants catalysant la réduction 4-méthylacétophénone en présence de formiate, avec de bonnes énantiosélectivités (ex. 91% (*R*) pour [(η_6-cymene)Ru(Biot-p-L)Cl] ⊂ S112A). [139]

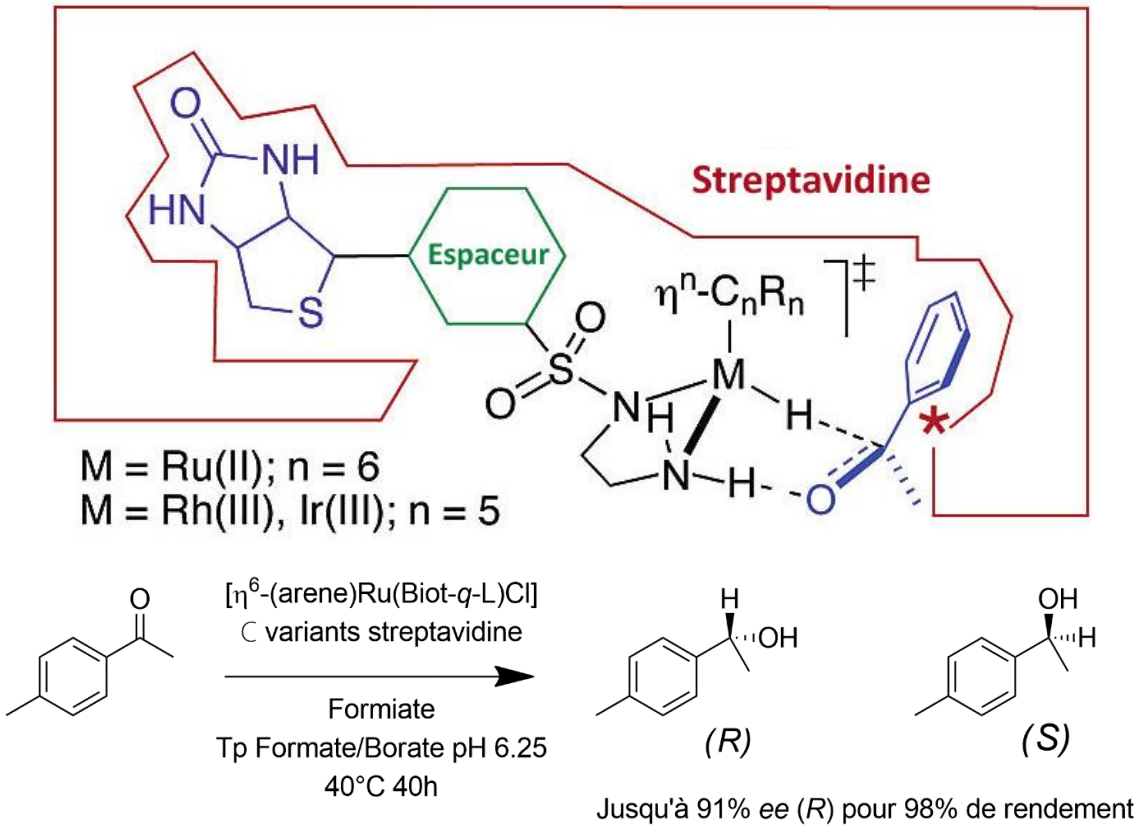


Figure 53 : (haut) Schéma des complexes de type tabouret de piano couplés à la biotine, au sein de la streptavidine. (bas) Transfert d'hydrogénation du 4-méthylacétophénone catalysé par des ligands ruthénium (selon [139])

Une approche d'évolution dirigée de la streptavidine a ensuite permis de réduire des substrats dialcyl cétone par du formiate de manière énantiosélective. Dans cette approche originale, des billes de biotine-sépharose ont été utilisées pour sélectionner les variants d'intérêt. La streptavidine étant tétramérique, l'ajout d'un excès de streptavidine sur les billes provoque la fixation d'une biotine par tétramère, laissant théoriquement 3 sites accessibles pour une catalyse hétérogène. Cette méthode de purification rapide a permis d'accélérer grandement le processus de criblage. Des séries de variants de la

streptavidine (ou « bibliothèques ») ont été construits par mutagénèse à saturation d'acides aminés déterminés sur la base d'une structure cristallographique de la protéine en présence du complexe. La lysine 121 et la leucine 124 ont été mutées sur la base des variants S112A et S112K (et sauvage WT). Après plusieurs tours de sélection, des protéines portant le complexe de ruthénium coiffé de benzène ou *p*-cymène capables de réaliser la réduction de différentes diacyl cétones avec une excellente énantiosélectivité et des rendements quantitatifs ont été obtenus.[140] Une représentation de l'évolution des meilleurs variants pour la réduction du 4-phényl-2-butanone est visible en figure 54.

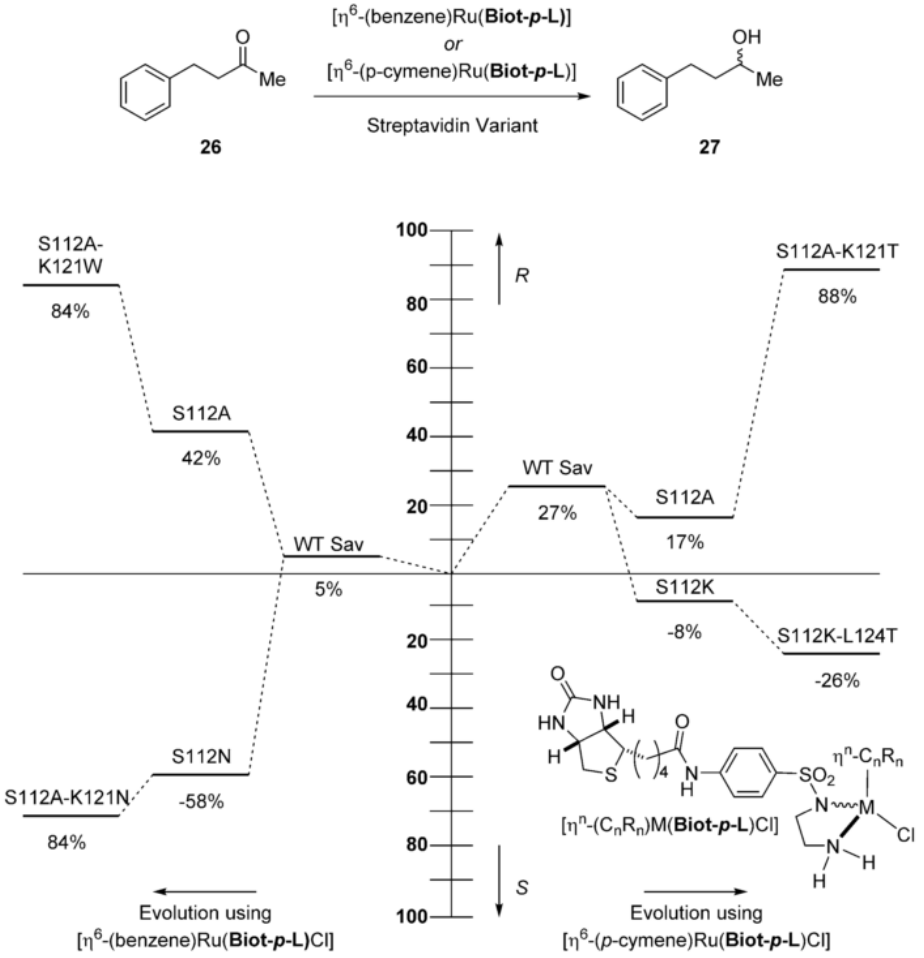


Figure 54 : Évolution de la streptavidine avec deux ligands différents pour la réduction du 4-phenyl-2-butanone obtenue par une banque de mutagénèse à saturation (CAST) sur la base des protéines sauvage (WT), S112A et S112K. En abscisse sont les variants présentant de plus en plus de mutations en présence de catalyseurs au ruthénium coiffés de benzene (gauche) ou cymène (droite). En ordonnées sont les excès énantiomériques obtenus pour les différents variants, favorisant l'énantiomère R (haut) ou S (bas) (selon [140])

En 2011, Ward et son équipe ont utilisé avec succès des complexes d'iridium pour la réduction d'imines cycliques par des variants de streptavidine optimisés. Les énantiomères (*R*) et (*S*) ont été obtenus avec des excès de 91% et 78% respectivement, pour des variants S112A et S112K. Dans ce cas particulier, la vitesse de réaction et l'affinité

ont pu être améliorées par la mutation d'un acide aminé proche dans la structure (K121T).[141]

Dans le cas de l'énantiomère (*R*), un excès énantiomérique de 99% a pu être atteint en couplant l'activité du biohybride à une monoamine oxydase, effectuant la réaction inverse spécifiquement pour la forme (*S*). Cette étude montre un élégant cas de résolution cinétique d'énantiomères couplée à une synthèse asymétrique.[142]

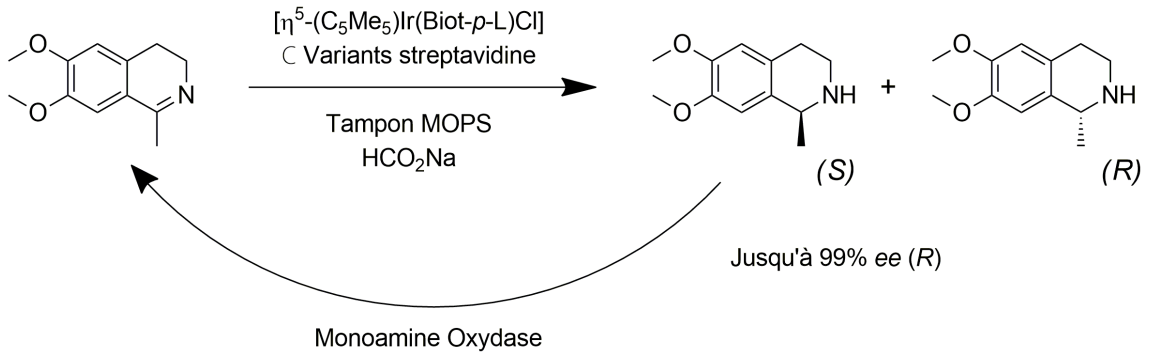


Figure 55 : Réduction d'amines cycliques par le formiate grâce à des catalyseurs insérés dans des variants de streptavidine (couplé à la monoamine oxydase) (selon [141])

Dernièrement, l'insertion de complexes de palladium dans des variants de streptavidine a permis de générer des biohybrides capables de catalyser des réactions de Suzuki énantiosélectives. Un criblage de différents ligands, puis une approche de mutagénèse à saturation sur les positions S112, N118 et K121 ont permis d'obtenir la première « Suzukiase » artificielle énantiosélective à ce jour.[143]

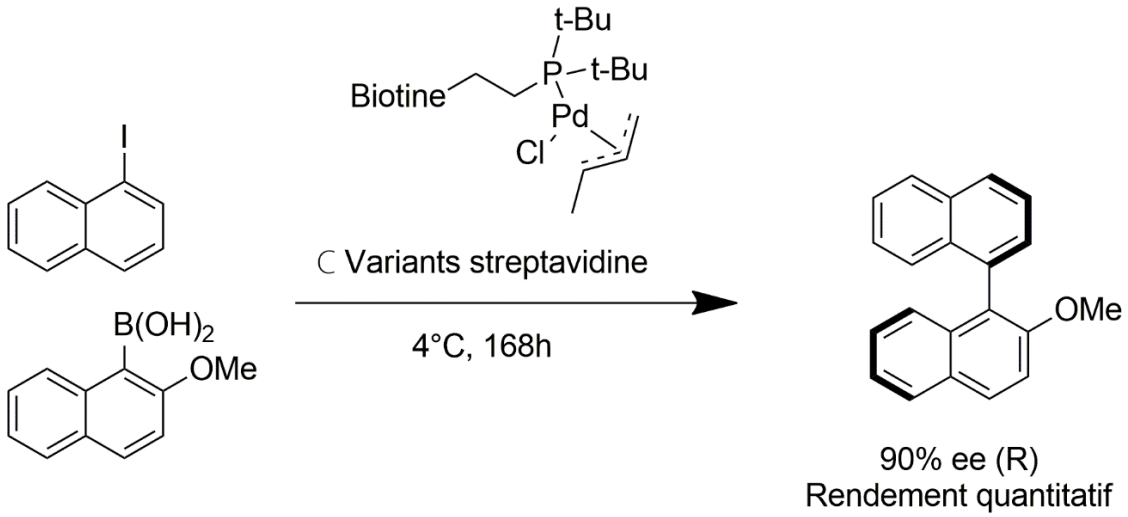


Figure 56 : L'insertion de complexes de palladium au sein de la streptavidine ont permis l'obtention de la première « Suzukiase » énantiosélective à ce jour (selon [143])

Ces travaux sur la streptavidine combinent l'ingénierie du ligand métallique, tant dans sa nature que dans sa liaison avec la biotine, avec l'ingénierie génétique. Ils illustrent parfaitement la possibilité, et souvent la nécessité, d'une ingénierie des deux partenaires pour atteindre de bonnes énantiosélectivités. La streptavidine comme architecture protéique comporte de nombreux avantages : une poche catalytique, une bonne stabilité thermique, une production aisée et une haute affinité pour la molécule utilisée comme ancre pour accrocher le complexe métallique (la biotine). Cependant l'ingénierie génétique est limitée, se concluant par une baisse de la solubilité et parfois une production inexistante de nombreux variants. Ainsi, les cribles d'évolution sont souvent restés limités à un nombre très restreint d'acides aminés.

Très récemment, Ward a développé une nouvelle approche prometteuse permettant un criblage à plus haut débit et donc une meilleure exploration de l'espace des séquences.[144] Cette nouvelle approche dans laquelle la protéine est sécrétée dans le périplasme de la bactérie, a permis à l'équipe de Ward d'effectuer des expériences d'évolution dirigée pour la création d'un biohybride catalysant la métathèse d'oléfines. Une stratégie d'ISM a alors été mise en place avec un crible de fluorescence du produit en plaques. La réaction testée n'était cependant pas source de produits chiraux. [145]

Cette approche est très intéressante pour l'évolution de métalloenzymes artificielles en particulier, car il a été démontré que la forte concentration intracellulaire en glutathion réduit pouvait inhiber la plupart des cofacteurs métalliques artificiels.[146] Il est fort probable que cette étude ouvre la porte à de nouvelles générations de biohybrides issus de l'évolution dirigée, l'enjeu restant de pouvoir cribler à haut débit des réactions énantiosélectives.

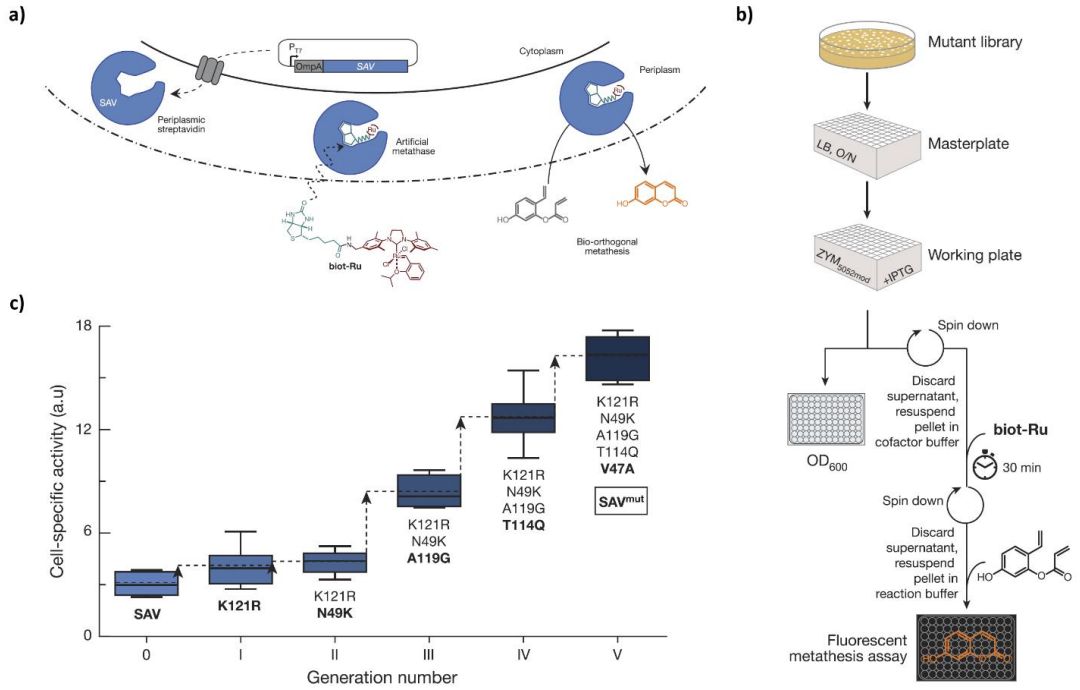


Figure 57 : Evolution dirigée d'une métathase artificielle. a) Principe de la méthode : expression périplasmique, insertion du catalyseur, catalyse in situ b) Principe du crible en plaque c) augmentation de l'activité au cours de l'expérience d'ISM (selon [145])

II.2.2.2 L'approche du cheval de Troie étendue à d'autres protéines

Il existe peu d'autres exemples de la stratégie du cheval de Troie car peu de protéines résistantes à l'ingénierie sont connues pour interagir avec des petites molécules avec des affinités suffisamment fortes.

Les anticorps peuvent présenter une bonne affinité pour les petites molécules contre lesquels ils ont été sélectionnés. L'équipe de Mahy a ainsi repris le concept d'hémoabzyme dans la stratégie du cheval de Troie, utilisant des porphyrines de fer conjuguées à de l'œstradiol Fe(III)(TMPyP) et également la Fe(III)(TpSPP) (figure 58). Un anticorps très affin pour l'œstradiol a été utilisé comme architecture protéique. Le complexe formé est capable de catalyser la sulfoxydation énantiosélective du thioanisole avec du peroxyde d'hydrogène comme oxydant. L'énantiosélectivité est cependant modeste, avec un *ee* d'environ 10% en faveur de l'énantiomère (*S*) pour les deux porphyrines. Il a été remarqué que l'architecture protéique conférait une protection de la porphyrine contre les dégradations dues à l'oxydant.[147,148]

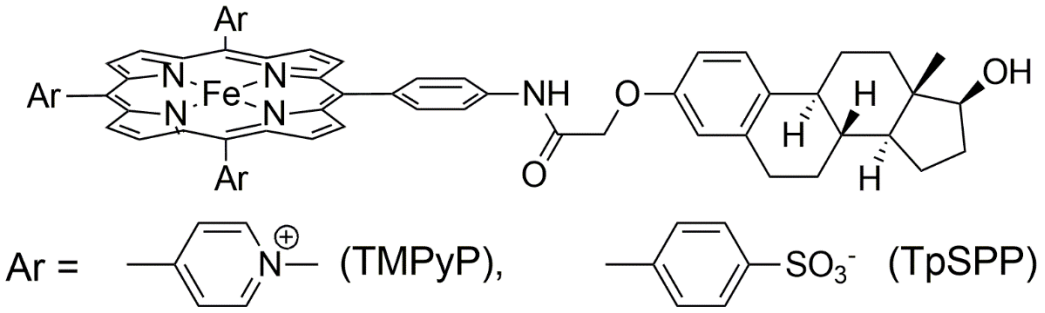


Figure 58 : Cofacteurs porphyrine de fer II développé par Raffy et al pour la création d'un biohybride par la stratégie du cheval de Troie. L'architecture protéique étant l'anticorps 7A3 ayant une constante de dissociation de l'ordre de 10^{-10} M avec l'oestradiol. Les biohybrides Fe(III)(TMPyP) ⊂ 7A3 et Fe(III)(TpSPP) ⊂ 7A3 catalysent la sulfoxydation du thioanisole par le H_2O_2 avec un ee de 8% et 10% respectivement (selon [147,148])

Cependant les anticorps sont des protéines peu stables et coûteuses à produire. La même équipe a donc tenté d'explorer la même stratégie avec une protéine alternative aux anticorps.

Dans une étude initiale d'évolution dirigée, une bibliothèque de variants de la néocarzinostatine (NCS), une protéine bactérienne, avait été conçue dans l'équipe de Philippe Minard et des variants fixant la testostérone avaient été sélectionnés par la méthode du « phage display ».[149] Le variant résultant de ces sélections (NCS-3.24) liait la testostérone dans sa crevasse (comme l'a indiquée la structure cristallographique du complexe), avec une affinité de l'ordre de 10 µM.[150]

Lors d'une première collaboration entre les équipes Mahy et Minard, cette protéine a donc été utilisée comme hôte du « cheval de Troie », en couplant la testostérone à différents complexes métalliques et en étudiant les activités catalytiques associées. Une porphyrine anionique de Fe(III) a ainsi permis de catalyser la réaction de sulfoxidation du thioanisole par H_2O_2 avec un très faible excès énantiomérique (13%) [151]; un complexe Zn(II)-bipyridyl a permis de catalyser la transestérification de phosphodiesters et ainsi de mimer une activité de type RNAse [152] ; et enfin, le couplage de la testostérone avec une phénantholine de Cu (II) a conduit à la catalyse de la réaction de cycloaddition de Diels-Alder entre l'azachalcone et le cyclopentadiène avec un rapport endo:exo différent du cuivre en solution mais sans excès énantiomérique.[153]

Des études de modélisation moléculaire suggèrent que les complexes associés à la testostérone seraient peu accessibles ou trop encombrants pour la crevasse de la NCS-3.24 et seraient donc plus ou moins mobiles à l'extérieur de la protéine. Cette interprétation expliquerait l'absence d'énantiosélectivité des différents biohybrides générés d'après cette approche. Par ailleurs, du fait de l'affinité relativement faible de la

protéine pour son ligand, dans les conditions d'utilisation des biohybrides pour la catalyse, une partie du ligand se trouve sous forme libre en solution en équilibre avec sa forme liée.

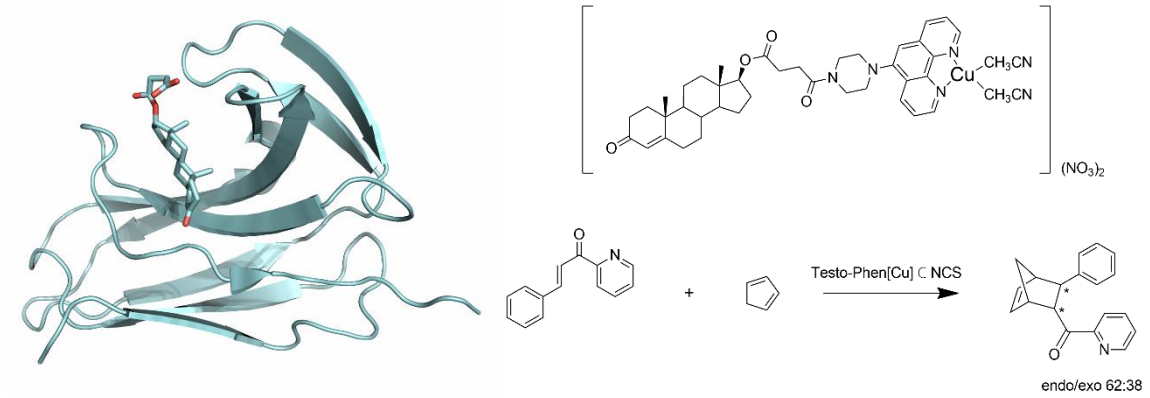


Figure 59 : Structure cristallographique de la néocarzinostatine (NCS) liée à la testostérone (en bâtons). Le complexe de cuivre fonctionnalisé par la testostérone puis inséré dans la NCS catalyse la réaction de Diels-Alder avec un rapport endo:exo de 62 :38 mais sans excès énantiomérique.

L'équipe de Michèle Salmain a utilisé la beta lactoglobuline, une protéine extraite du lait de vache et connue pour fixer différents acides gras dont les acides laurique et palmitique. Dans leur étude, des composés dipyridylamine de rhodium ont été liés covalemment à des chaînes lauroyl ou palmitoyl. Les différent biohybrides sont capables d'effectuer des transferts d'hydrogénation de trifluoroacétophénone (TFAP) avec des excès énantiomériques jusqu'à 25% (figure 60).[154]

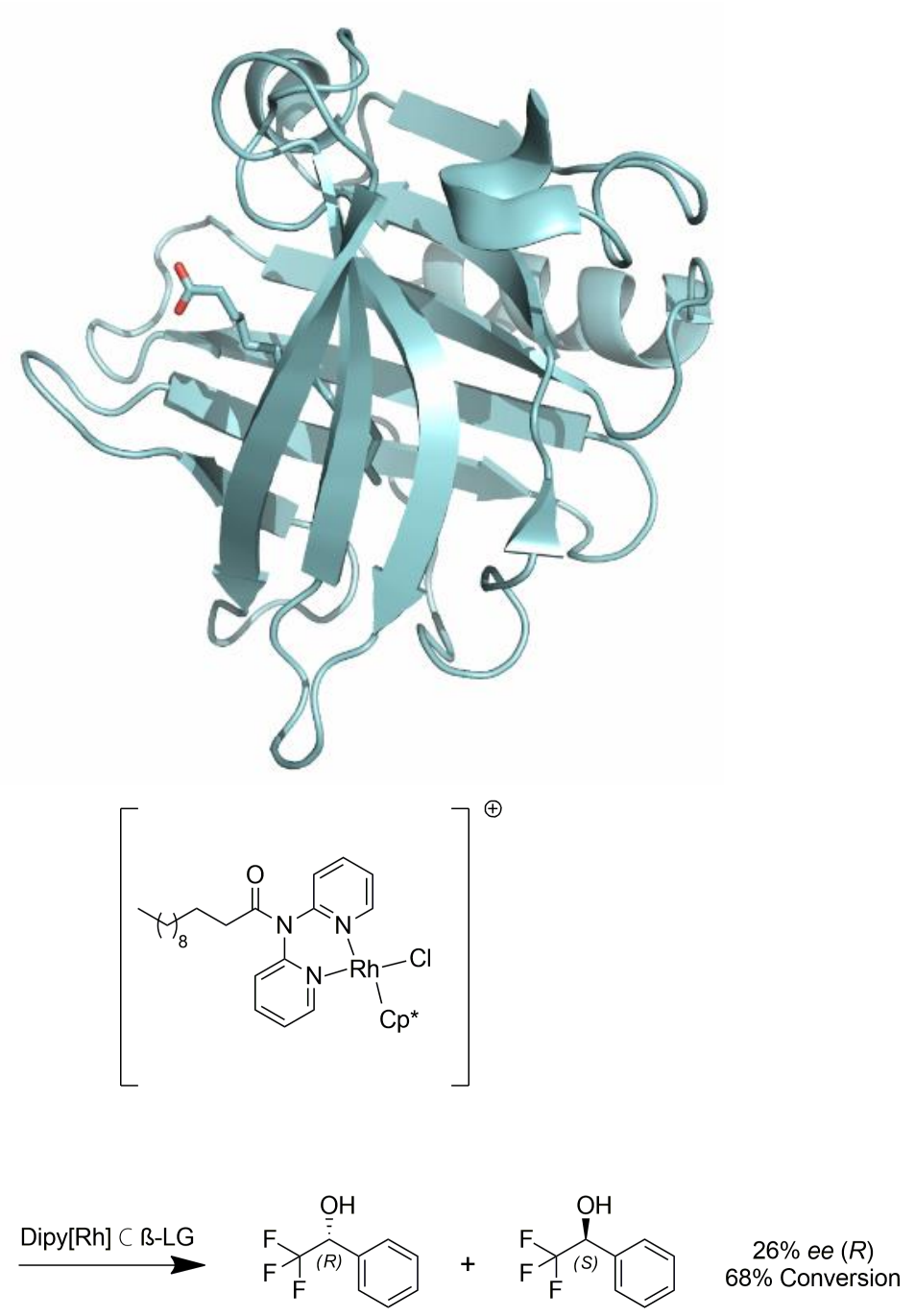


Figure 60 : Structure de labeta lactoglobuline liée à un acide gras (en bâtons). Le complexe de rhodium fonctionnalisé par une chaîne alkyle (au dessus) se positionne à la même place que l'acide gras, le biohybride catalysant le transfert d'hydrogénation du trifluoroacétophénone (TFAP)

La stratégie du cheval de Troie, comme la stratégie hôte-invité, présente l'avantage de pouvoir effectuer avec une relative simplicité la combinatoire entre variants protéiques et catalyseurs chimiques de différentes natures. De bonnes activités et sélectivités ont été observées, à la condition que le ligand soit raisonnablement enfoui.

Cependant ces stratégies possèdent un désavantage inhérent au caractère non covalent des interactions entre protéine et cofacteur : en effet, il est nécessaire de travailler à des concentrations suffisantes en protéine (> 10 X K_D) et supérieure à celle du ligand pour limiter la présence de ligand libre lors des expériences de catalyse.

II.2.3 Stratégie du couplage covalent

La troisième approche pour générer des catalyseurs biohybrides consiste à coupler covalemment le cofacteur métallique dans la protéine. Ceci a pour principal avantage de passer outre le problème lié à l'affinité entre les deux molécules. Ainsi, les acides aminés entourant le site catalytique nouvellement formé peuvent donc être modifiés sans craindre une baisse d'affinité comme dans le cas des stratégies de couplage non covalent. La stabilité de l'interaction permet également de ne pas avoir de fraction libre du cofacteur dans la solution pour les cas d'interactions peu affines. Elle peut enfin permettre d'augmenter la stabilisation conformationnelle du cofacteur lui-même, comme nous le verrons ci-dessous. Afin d'effectuer une liaison covalente stable et spécifique entre une protéine et un complexe métallique, différentes stratégies sont possibles et seront précisées pour chacune des constructions présentées dans ce chapitre.

Les propriétés physico-chimiques des acides aminés peuvent être mises à profit pour la liaison covalente spécifique avec un ligand. La cystéine est de loin l'acide aminé le plus couramment employé pour plusieurs raisons : elle est peu présente dans les protéines naturelles, permettant une insertion spécifique par mutagénèse dirigée. Sa chaine latérale est déjà utilisée naturellement pour former des liens covalents par le biais de ponts disulfures. La forme thiolate étant un nucléophile puissant, elle réagit avec les réactifs de Michael ou les haloacétamides plus rapidement que les amines primaires (lysine, résidu N-terminal) à pH neutre. L'utilisation de thiols réactifs pour le couplage de ligands n'est pas l'approche idéale pour toutes les architectures protéiques. En effet, les protéines stabilisées par des ponts disulfures nécessitent des conditions oxydantes pour leur repliement, ce qui peut conduire à la formation de ponts disulfures impliquant les cystéines libres réactives.

II.2.3.1 La papaïne, une protéine modèle pour la fonctionnalisation d'un thiol

Les travaux précurseurs dans la fonctionnalisation d'une protéine au sein de son site actif concernaient la papaïne. La papaïne est une protéase à cystéine, cet acide aminé est donc idéalement placé au sein d'une crevasse. De plus, une seule cystéine libre (Cys25) est présente dans la protéine, ce qui permet de cibler le point d'ancrage (Figure 61).

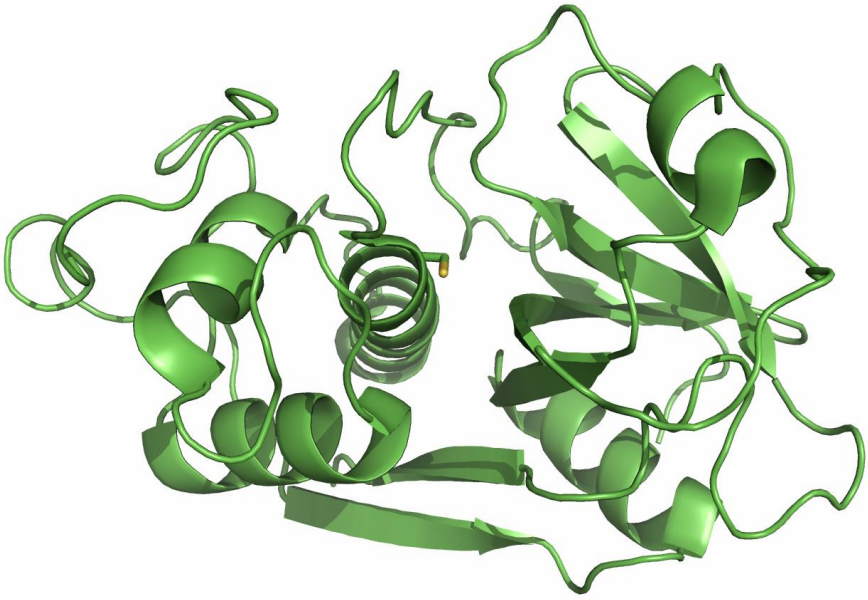


Figure 61 : Structure cristallographique de la papaïne. La cystéine 25 utilisée pour le greffage covalent des cofacteurs métalliques est représentée en bâtons.

Les premiers travaux concernaient la modification de la papaïne avec des dérivés de flavine pour l'oxydation de nicotinamides. [155] L'ancrage utilisait alors des dérivés bromoacétamides, qui par substitution nucléophile formaient une liaison thioéther. L'équipe de Reetz réutilisa cette protéine pour insérer des complexes salen de manganèse, pour l'époxydation du styrène par l'hydroperoxyde de tertiobutyle. Ces complexes étaient fonctionnalisés avec un composé maléimide, qui, par réaction de Michael avec la cystéine, formaient un lien covalent avec la protéine. L'énantiosélectivité de la réaction était cependant faible (10% *ee*). [156]

Dans une étude de Panella *et al*, un complexe de rhodium a été couplé avec succès à la papaïne. Ce biohybride catalysait spécifiquement l'hydrogénation de 2-acétamidoacrylate de méthyle, mais sans excès énantiomérique (figure 62).[157] Le couplage était effectué grâce à un bras bromoacétamide.

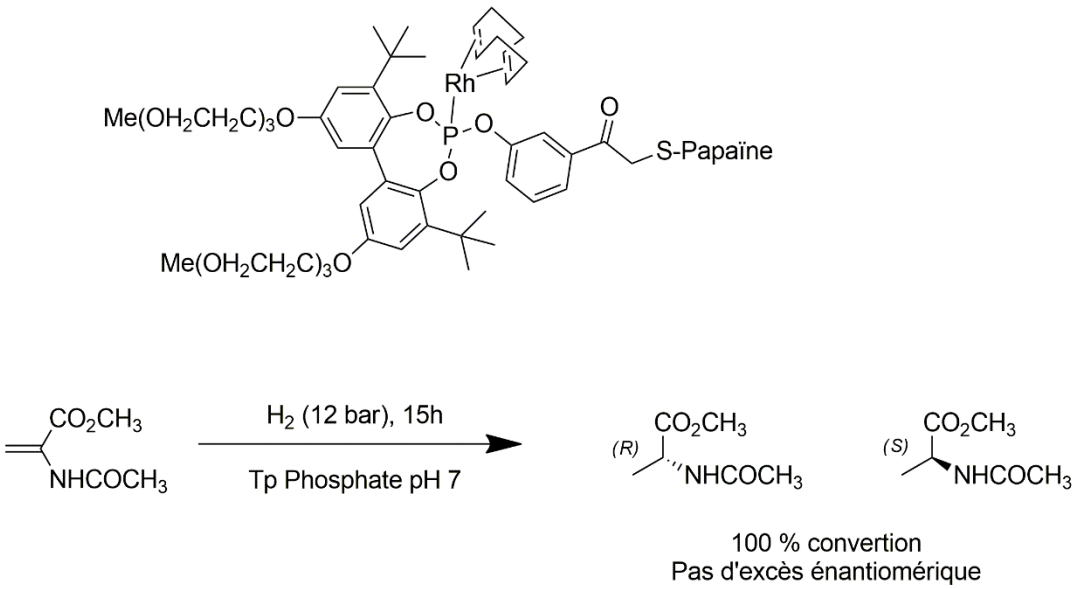


Figure 62 : Complexe rhodium covalemment ancré à la papaïne, catalysant l'hydrogénation non énantiosélective de 2-acétamidoacrylate de méthyle (selon [157])

L'équipe de Salmain s'est également intéressée à cette protéine. Des composés dipyridylamine et phénanthroline de ruthénium, coiffés de différents ligands η^6-arènes, ont été fonctionnalisés par un composé maléimide. En faisant varier le type d'arène (benzène ou *p*-cymène) ainsi que le nombre de carbones de la chaine alkyle reliant le complexe au maléimide, cette équipe est parvenue à effectuer la réduction du $NAD(P)^+$ en NAD(P)H avec le formate comme donneur d'hydrure, ainsi que la réaction de Diels-Alder entre l'acroléine et le cyclopentadiène (figure 63). Malheureusement ces catalyseurs n'ont conduit à aucun excès énantiomérique.[158]

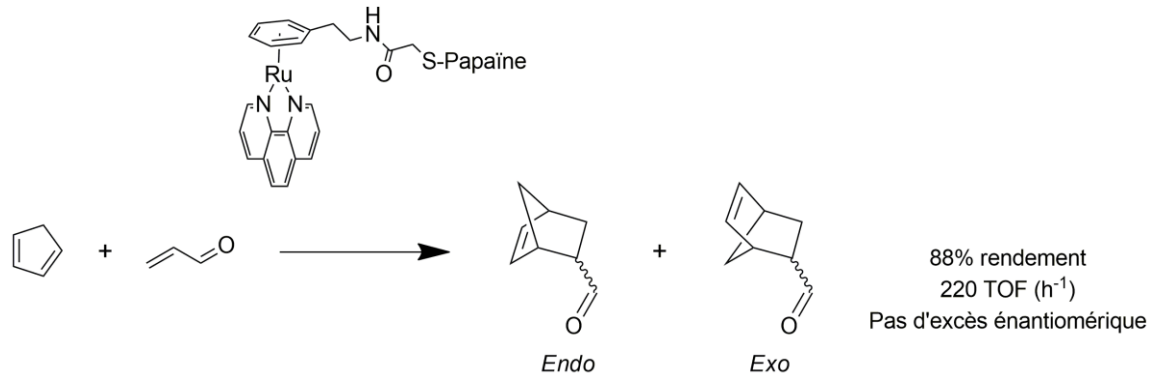


Figure 63 : Réaction de Diels-Alder catalysée par le biohybride issu de la liaison covalente entre un complexe de ruthénium et la papaïne (selon [158]).

La mauvaise énantiosélectivité de chacun de ces biohybrides a été expliquée soit par une cavité trop grande, un ancrage trop flexible ou trop long, faisant ressortir le cofacteur métallique en dehors de la cavité de la protéine. En effet plus récemment une structure cristallographique de la papaïne avec un ligand de ruthénium a été déterminée par

l'équipe de Salmain (PDB : 4KP9) (figure 64). On peut clairement constater que le complexe de rhuténium est à l'extérieur de la crevasse.

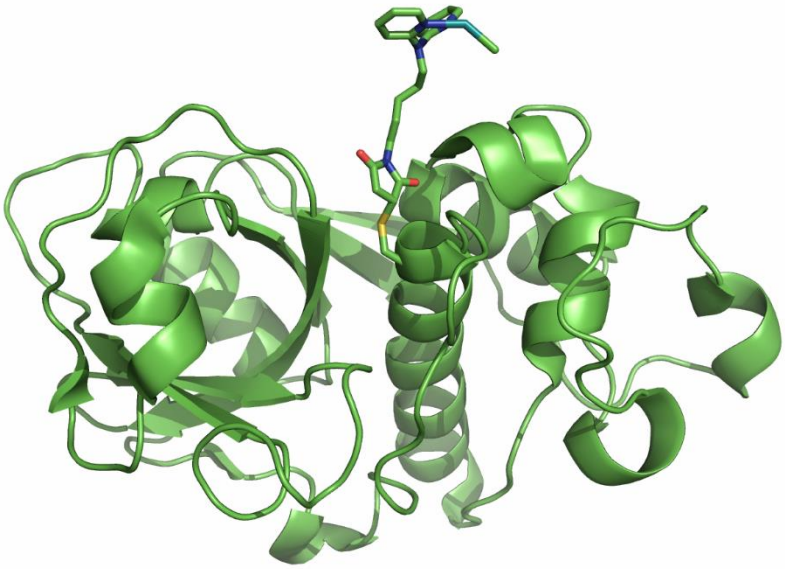


Figure 64 : Structure de la papaïne fonctionnalisée par un complexe de ruthénium (bâtons) déterminée par l'équipe de Salmain. On peut constater que le ligand est à l'extérieur de la crevasse.

II.2.3.2 La myoglobine

Il a été précédemment mentionné l'insertion par stratégie hôte-invité d'un complexe salophène de chrome dans l'apo-myoglobine. L'équipe de Lu a émis l'hypothèse que la mauvaise énantiosélectivité du biohybride était due à un manque de rigidité du complexe au sein de la crevasse. Partant de ce postulat, des complexes de type salen de manganèse ont été insérés dans la crevasse en lieu et place de l'hème. Afin de rigidifier l'ancrage, deux mutations en cystéines ont été effectuées sur la base de la structure cristallographique de la protéine naturelle (L72C/Y103C). Le ligand salen, fonctionnalisé par deux groupements méthane thiosulfonate, était alors ancré à la protéine par deux ponts disulfures. Ainsi fonctionnalisé, le biohybride a été capable d'effectuer la sulfoxydation du thioanisole avec une énantiosélectivité de 51% pour la forme (*S*) (Figure 65). La nécessité de ce double ancrage a été vérifiée par l'utilisation d'un mutant unique (Y103C) qui possédait après fonctionnalisation une énantiosélectivité de seulement 12% pour la même forme.[159] Un autre point d'ancrage fut trouvé (T39C/L72C) portant l'énantiosélectivité à 60%, prouvant l'importance des points d'ancrage pour la sélectivité de la réaction.[160]

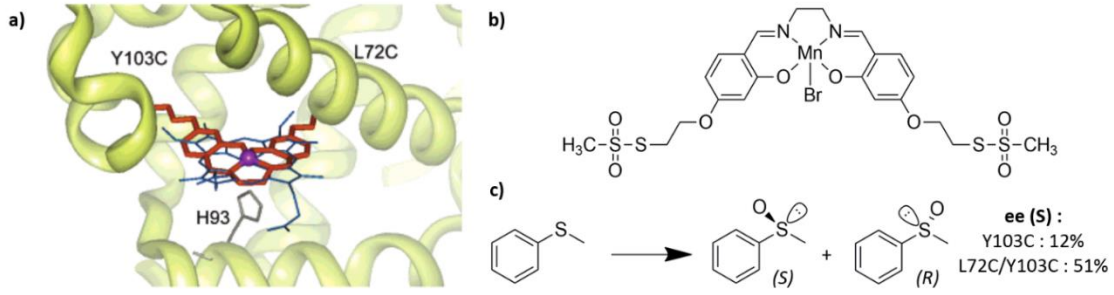


Figure 65 : Catalyse de sulfoxydation du thioanisole avec un complexe salophène de manganèse doublement ancré dans la myoglobine (selon [159]).

II.2.3.3 La Photoactive Yellow Protein

La Photoactive Yellow Protein (PYP) est une protéine fluorescente, originaire de *Halorhodospira halophila*. Le fluorophore (acide *p*-hydroxycinnamique) est covalemment lié au sein d'une crevasse hydrophobe à un résidu cystéine unique. C'est cette cystéine qu'a utilisée l'équipe de Paul C. J. Kamer pour le couplage covalent de différents complexes de rhodium, ruthénium et palladium. La modification de la cystéine était effectuée grâce à des ligands imidazolines, formant des liaisons thioester avec la cystéine (figure 66b). Ce type de réaction a permis le couplage de ligands phosphines, impossibles à coupler par les méthodes classiques (maléimide ou haloacétamide) du fait du caractère nucléophile des phosphines. Un biohybride fonctionnalisé par un complexe allyle de palladium a été utilisé pour l'amination allylique de la benzylamine sur l'acétate de phenylprop-2-enyl (figure 66c). Malheureusement, aucun excès énantiomérique n'a été obtenu. [161]

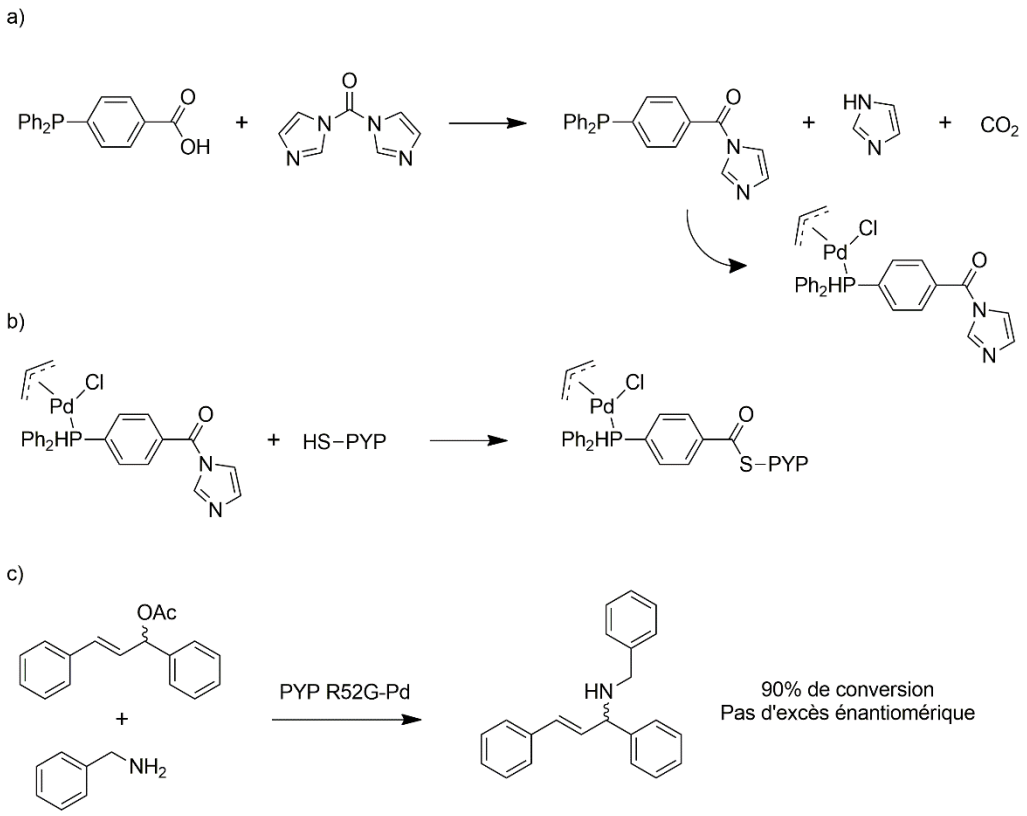


Figure 66 : a) Conception du ligand phosphine de palladium avec un bras imidazoline. b) Couplage du ligand sur une cystéine libre de la Photoactive Yellow Protein (PYP) c) Substitution allylique de l'acétate de 1,3-diphényl-2-propenyl avec la benzylamine catalysée par la PYP modifiée (selon [161])

II.2.3.4 La protéine LmrR

Nous avons discuté plus haut de l'assemblage supramoléculaire de cofacteurs au sein de la crevasse hydrophobe formée par la dimérisation de la protéine LmrR. Les travaux originels de l'équipe de Roelfes concernaient le couplage covalent de cofacteurs avec une cystéine insérée dans sa crevasse. La protéine LmrR ne comportant ni pont disulfure, ni cystéine, deux variants portant un acide aminé cystéine unique à différents points de la crevasse ont été modifiés par un ligand phénanthroline ou bipyridine grâce à l'utilisation d'un bras bromoacétamide. La capacité de dimérisation de la protéine provoquait donc l'apparition de deux centres catalytiques équivalents. Les différents biohybrides ont été utilisés avec succès pour la réaction de Diels-Alder énantiosélective entre le cyclopentadiene et différents dérivés de chalcone. En effet, le biohybride LmrR_M89C_Phen_Cu catalysait cette réaction avec jusqu'à 97% d'excès énantiomérique et un taux de conversion de 97% (figure 67). [162]

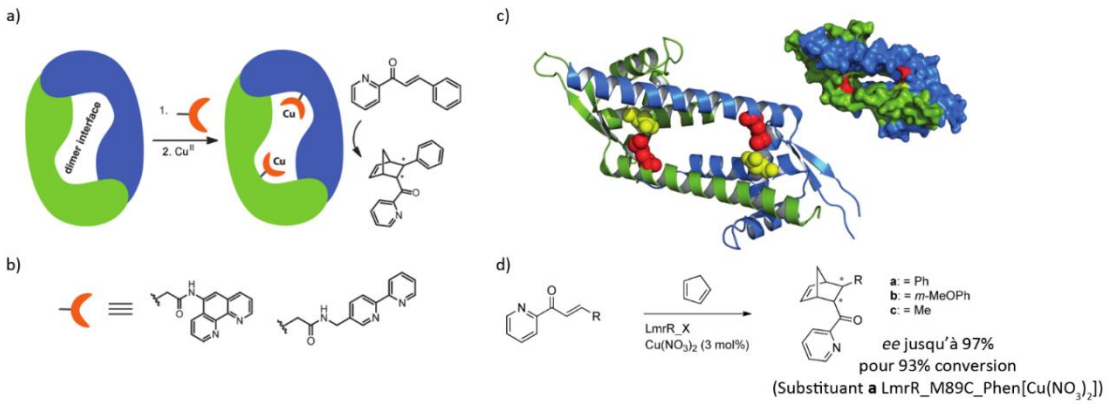


Figure 67 : Approche de l'équipe de Roelfes avec la protéine LmrR catalysant la réaction de Diels-Alder énantiosélective (selon [162])

Dans une étude ultérieure, il a été montré que le même biohybride était également capable de catalyser la réaction d'hydratation de dérivés de chalcone avec jusqu'à 75% d'excès énantiomérique (figure 68). Différentes mutations ponctuelles autour du site actif nouvellement créé ont eu un effet sur le rendement ou l'énantiosélectivité de la réaction, indiquant que la réaction était bien guidée par les acides aminés environnants. Malheureusement aucune amélioration dans la sélectivité n'a été obtenue. [163] Aucune expérience d'évolution dirigée n'a été effectuée sur cette protéine à ce jour.

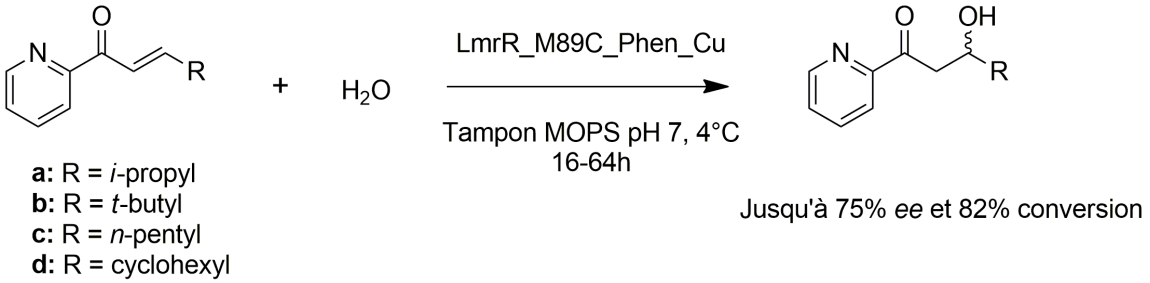


Figure 68 : Réaction d'hydratation catalysée par le biohybride LmrR couplé à une phénanthroline de cuivre (selon [163])

II.2.3.5 Modification de la protéine SCP-2L en une Diels-Alderase artificielle

Le groupe de Kamer s'est également intéressé au couplage covalent de complexes de cuivre pour la création d'une Diels-Alderase artificielle. Son groupe a alors utilisé une protéine du type transporteur de stérol, possédant une cavité hydrophobe cylindrique propice à l'insertion covalente d'un ligand métallique. Deux différents variants présentant une cystéine unique furent utilisés : V83C et A100C. Des complexes de type phénanthroline ou bis(pyridin-2-ylmethyl)amine fonctionnalisés avec un bras maléimide ont été couplés avec succès à ces deux variants (figure 69). Le métal (du cuivre) ne peut pas être ajouté avant le couplage, car il est montré dans leur étude que ces complexes de cuivre catalysent la création de ponts disulfures intermoléculaires avant que la cystéine ne puisse réagir avec le maléimide. Les biohybrides formés ont été capables de catalyser

la réaction de Diels-Alder avec des énantiosélectivités pour la forme *endo* allant jusqu'à 25% (pour le biohybride SCP-2L-V83C-Phen[$Cu(NO_3)_2$]). [72]

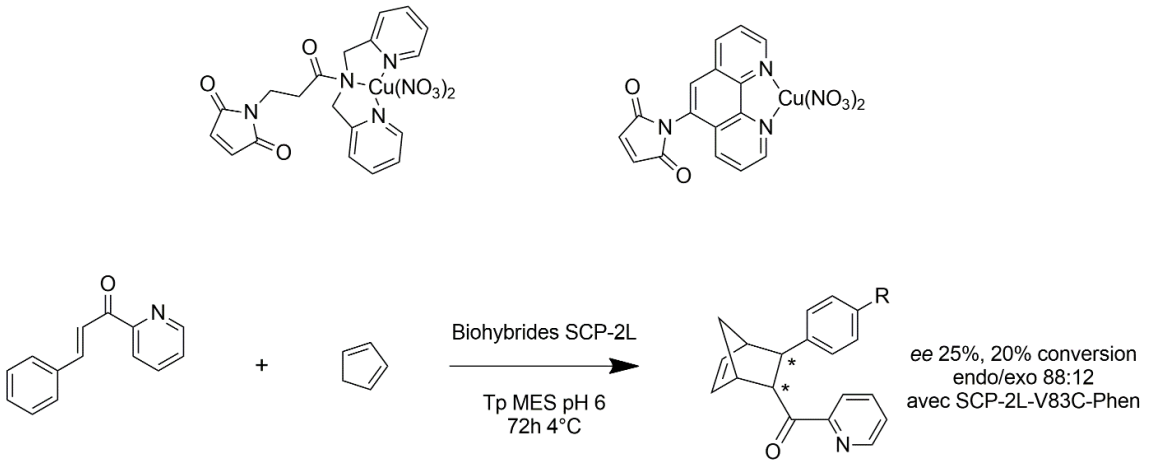


Figure 69 : Ligands couplés covalemment aux variants de la protéine SCP-2L par l'équipe de Kamer. bis(pyridin-2-ylméthyl)amine à gauche, phénanthroline à droite. Réaction catalysée par les différents catalyseurs biohybrides (le meilleur résultat en termes d'énantiosélectivité est noté). Selon [72]

II.2.3.6 Modification d'une lipase en une Diels-Alderase artificielle

Dans le même esprit, l'équipe de José M. Palomo s'est récemment intéressée à la modification de la lipase de *G. thermocatenulatus* (GTL) en vue de la création d'une Diels-Alderase artificielle. Une phénanthroline de cuivre fixée covalemment (à l'aide d'un bras iodoacétamide, figure 70a) à une cystéine unique permis d'effectuer la cycloaddition entre azachalcone et cyclopentadiène avec 92% d'excès énantiomérique dans la forme *endo* (figure 70b). La catalyse était effectuée en phase solide, la protéine étant attachée covalemment à des billes de sépharose par les différentes lysines à la surface de la protéine. L'originalité de cette approche, outre l'utilisation de la catalyse hétérogène, réside dans le lieu d'insertion du complexe phénanthroline, différent de la poche catalytique naturelle de l'enzyme. Ainsi l'activité native n'était pas affectée par la modification, qui était encore capable d'effectuer l'hydrolyse d'esters tels qu'un butyrate de nitrophényl (figure 70c, première flèche). Ce dernier pouvait ensuite être réduit en cascade par le complexe phénanthroline de cuivre situé dans l'autre poche catalytique (figure 70c, deuxième flèche).

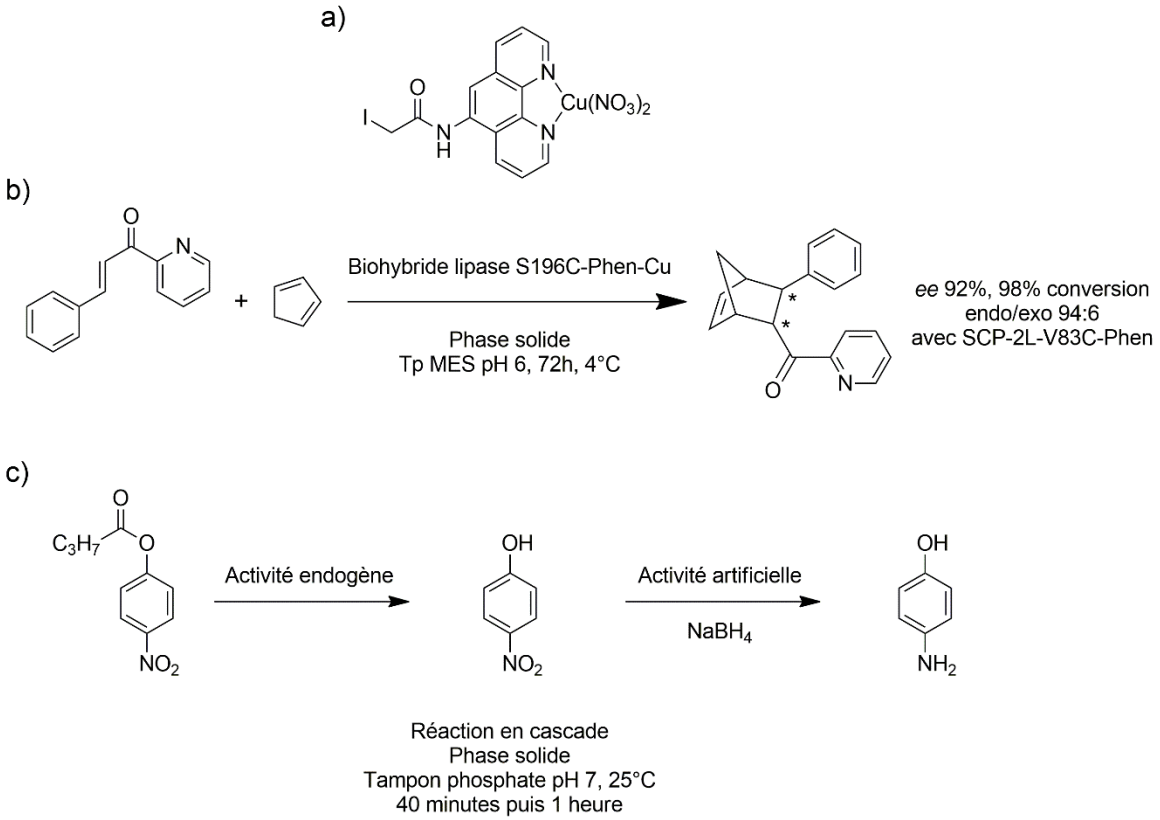


Figure 70 : a) Complexe phénanthroline lié covalemment à la la cystéine de la protéine grâce à un bras iodoacétamide (le cuivre est en fait ajouté après). b) réaction de Diels-Alder énantiosélective catalysée par le biohybride. c) Réactions d'hydrolyse d'ester puis de réduction catalysées en cascade dans la poche catalytique naturelle de la protéine pour la première et dans la poche catalytique artificielle pour la seconde.

Le point commun de ces différentes approches est l'utilisation d'acides aminés naturels puis leur modification post-traductionnelle. Le couplage covalent entre le cofacteur métallique et la protéine confère une stabilité au couple protéine/cofacteur difficilement atteignable par les stratégies non covalentes où, à l'exception du couple streptavidine/biotine, les affinités sont trop faibles pour durer dans le temps. Cependant la modification post-traductionnelle, notamment sur les cystéines, présente plusieurs inconvénients : la réaction ne peut se faire dans un milieu complexe, d'autres protéines ou molécules nucléophiles pouvant interagir en provoquant des couplages non désirés et diminuant l'efficacité du couplage désiré. De plus, bien que la cystéine soit le nucléophile majeur d'une protéine à pH neutre, d'autres groupements ont également un caractère nucléophile (par exemple les amines primaires : lysine et N-terminal). Il en résulte des couplages supplémentaires non contrôlés, nécessitant l'optimisation de la réaction de fonctionnalisation (contrôle de la stœchiométrie, temps de réaction, pH), ce qui rend difficile l'utilisation dans des milieux complexes tels que des lysats cellulaires ou même des cellules entières.

Afin de répondre à ce problème, une solution émergente consiste à introduire dans la séquence peptidique des acides aminés non-naturels portant des fonctions chimiques n'existant pas dans les biomolécules (azoture, alcyne….). Il est courant et facile d'insérer des acides aminés artificiels lors d'une synthèse peptidique chimique, certains efforts dans ce sens ont d'ailleurs conduit à la conception de métallopeptides artificiels catalysant des réactions énantiosélectives.[164] Cependant cette approche ne permet pas de bénéficier de la capacité d'évolution, c'est pourquoi nous nous contenterons d'aborder le domaine récent de la modification du code génétique, introduit par Schultz.

II.2.4 L'extension du code génétique

Le code génétique, qui permet d'associer chaque triplet de nucléotides (« codons ») à un acide aminé correspondant, dépend d'un **quartet** : Le **codon** est reconnu par un ARN de transfert (**ARNt**) spécifique, cet ARNt est lui-même reconnu par une aminoacyl-ARNt synthétase (**AA-RS**) spécifique qui charge cet ARNt avec **l'acide aminé** correspondant. Il est possible de créer par ingénierie génétique un système permettant l'insertion d'un acide aminé non naturel en créant un quartet « codon/ARNt/AA-RS/acide aminé » orthogonal. Le terme orthogonal désigne deux systèmes qui n'interfèrent pas entre eux. Il est donc nécessaire que l'ARNt ne soit reconnu par aucune AA-RS endogène, que l'AA-RS artificielle ne reconnaisse aucun ARNt naturel et qu'elle charge l'ARNt uniquement avec l'acide aminé désiré. Il faut enfin que l'ARNt reconnaisse uniquement le codon désiré et que celui-ci ne soit reconnu par aucun autre ARNt endogène (figure 71). [165] La méthode la plus simple pour faire évoluer une AA-RS spécifique d'un acide aminé artificiel (aa*) est d'utiliser un acide aminé de structure proche de l'acide aminé naturellement reconnu. C'est pourquoi la plupart des acides aminés artificiels disponibles actuellement sont des dérivés de la tyrosine et de la pyrrolysine (évolution dirigée d'une Tyr-RS de *M. janaschii* [166] et d'une pyrLys-RS de *M. barkeri* [167]). Ces systèmes aa*/aa*-RS/ARNt sont dérivés d'archées, car leurs ARNt ne sont pas reconnus par les systèmes bactériens. Le codon orthogonal communément utilisé est le codon stop de type ambre (UAG), les deux autres codons stops n'étant pas reconnus par un ARNt spécifique.[168]

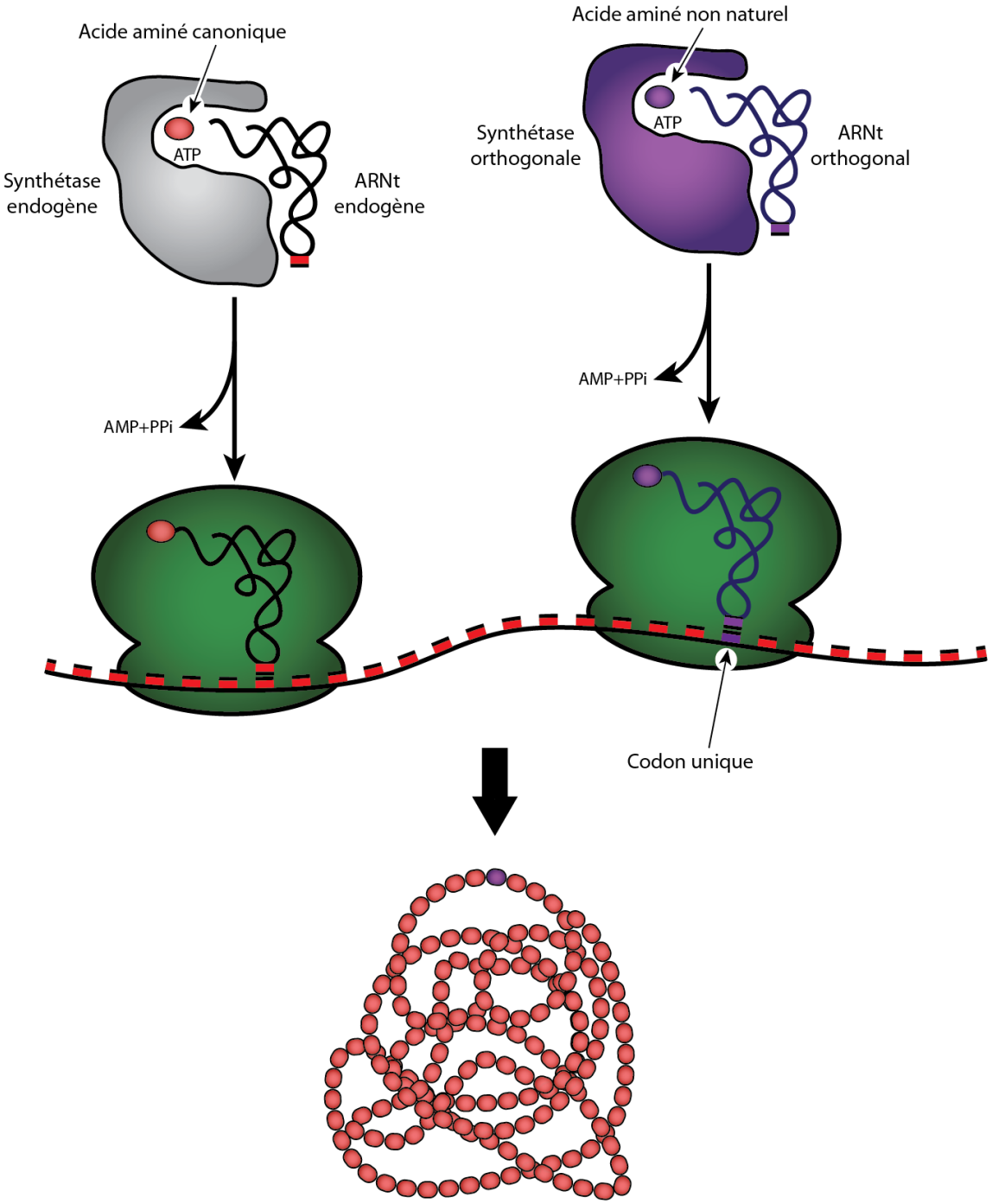


Figure 71 : Mécanisme nécessaire à la modification du code génétique. Une ARNt synthétase (AA-RS) charge spécifiquement un acide aminé non naturel sur un ARNt orthogonal. L'ARNt reconnait un codon unique afin d'intégrer l'acide aminé non naturel à une position voulue (selon [165]).

Il existe maintenant un répertoire d'acides aminés non naturels insérables dans les protéines [169], la plupart étant des dérivés de la tyrosine. Deux types d'acides aminés artificiels sont utilisés dans la conception de métalloenzymes. Ces acides aminés peuvent être de bons chélateurs de métaux de transition qu'ils peuvent donc directement complexer (figure 72, flèche haut), ou ils peuvent porter une fonction chimique permettant un couplage bioorthogonal d'un ligand chimique (figure 72, flèche bas).

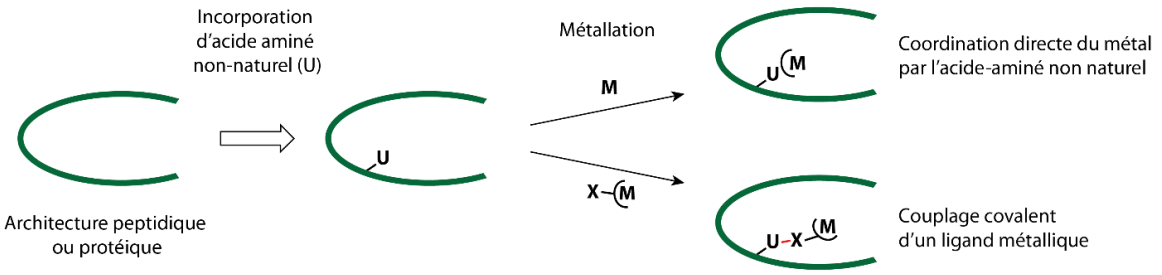


Figure 72 : Deux utilisations des acides aminés non naturels (U) pour la création de métalloenzymes artificielles : insertion d'un acide aminé possédant une forte affinité pour les métaux, ou possédant un groupement permettant un couplage par chimie « click ».

II.2.4.1 Les acides aminés artificiels pour la chimie bioorthogonale

La chimie de couplage bioorthogonale (aussi appelée chimie « click ») a été introduite par l'équipe de Barry Sharpless au début des années 2000 et est caractérisée par une chimie de couplage dans des conditions biocompatibles (milieu aqueux complexe et aérobie), avec absence de réactions secondaires comme un couplage indésirable aux biomolécules, ainsi qu'une réaction rapide et quantitative. La chimie « click » la plus utilisée utilise la cycloaddition azide-alcyne catalysée par le cuivre Cu (I) (CuAAC). Cependant l'utilisation de cuivre réduit Cu (I) est contraignante et cytotoxique. C'est pourquoi il lui est préféré sa variante utilisant des alcynes contraintes (cyclooctyne) permettant de se passer du catalyseur (réaction de type SPAAC).[170,171] Il existe cependant d'autres types de réaction. La figure 73 présente différents acides aminés artificiels incorporés au code génétique étendu permettant d'effectuer ces différents types de chimie.

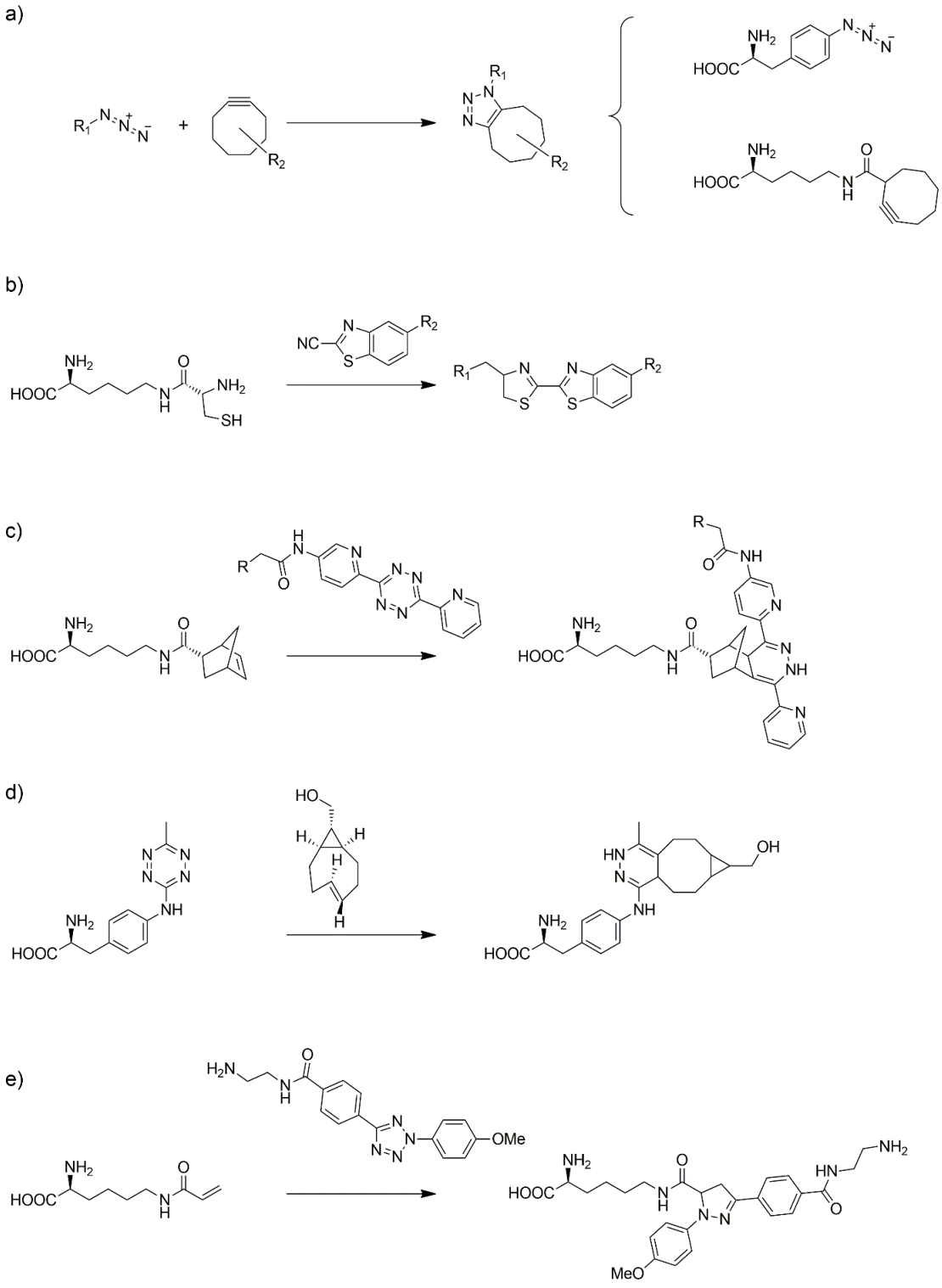


Figure 73 : Différents types d'acides aminés artificiels présents dans le répertoire du code génétique étendu actuel capable d'effectuer des réactions de couplage bioorthogonales (selon [172,173]) a) Cycloaddition alcyne/azide de type SPAAC, des acides aminés portant la fonction azide ou cyclooctyne existent b) Condensation de cyanobenzothiazole sur un acide aminé cysteinyl-lysine. c) Ligation norbornene/tetrazine avec un acide aminé portant une fonction norbornene. d) Ligation tetrazine/trans-cyclooctene avec un acide aminé portant la fonction tétrazine e) Réaction photoclick (activée par illumination à 365 nm) avec un acide aminé portant un groupement alcène déficient en électrons et un ligand tetrazole.

La chimie orthogonale présente un grand intérêt dans le domaine de la chimie de couplage des protéines en milieux complexes, cependant il existe encore peu d'exemples utilisant cette technologie pour insérer des ligands de métaux de transition afin de catalyser des réactions. L'équipe de Jared C. Lewis a très récemment exploré cette approche, en utilisant la protéine tHisF comme architecture protéique.[170] Par génie génétique, un acide aminé de l'intérieur du pore de la protéine (position 50) a été muté en un codon ambre pour permettre l'insertion de l'acide aminé *p*-azido-L-phénylalanine. Des complexes de type tetracarboxylate de di-rhodium ainsi que la terpyridine de manganèse et de cuivre, portant une fonction octyle, ont pu alors être liés par chimie « click » (stratégie **a** de la figure 73). Les efficacités de couplage allaient de 50 à 90% suivant les points de mutation utilisés et le type de complexe. Le biohybride portant le ligand de di-rhodium en position 176 est capable de catalyser des réactions de cyclopropanation et d'insertion de carbène dans une liaison Si-H mais malheureusement sans excès énantiomérique et avec des rendements plus faibles que le catalyseur seul (figure 74).

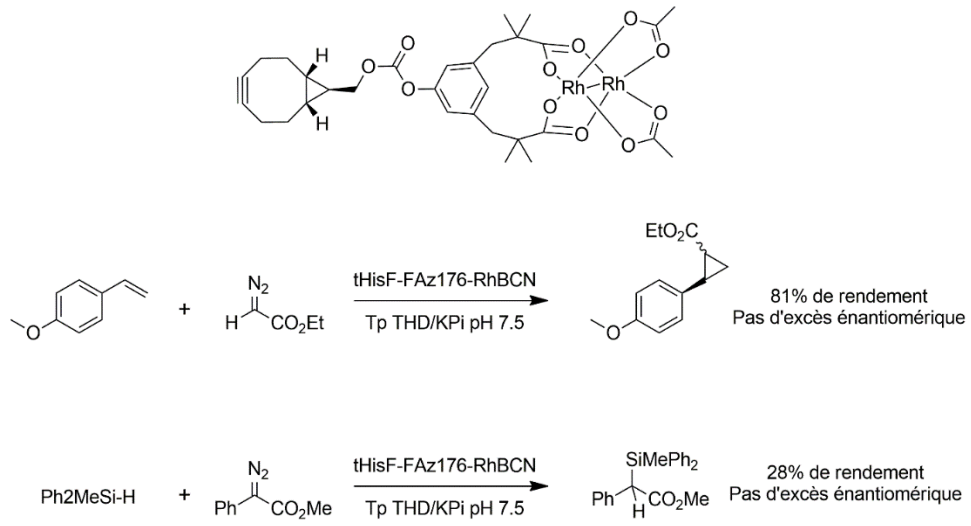


Figure 74 : Complexe de rhodium portant une fonction cyclooctyne pour liaison à l'acide aminé p-azido-L-phenylalanine en position 176 de la protéine tHisF (dessus). Réactions de cyclopropanation et d'insertion de carbène dans une liaison Si-H catalysées par le biohybride, sans excès énantiomérique.

Ce même groupe a alors changé d'architecture protéique, pour utiliser la prolyl organopeptidase (POP) de *P. furiosus*.[174] Cette protéine est constituée de deux domaines : un domaine α/β hydrolase (en forme de tonneau). Un autre domaine de type tonneau β vient recouvrir le premier, de manière à former une grande cavité interne avec un pore étroit traversant le second domaine (figure 75). L'utilisation d'une protéine issue d'un organisme hyperthermophile permet de partir d'une architecture protéique très stable, pour laquelle l'effet déstabilisateur induit par les mutations n'est généralement pas délétère pour le repliement de la protéine. La possibilité d'effectuer des réactions à des

températures plus hautes est également intéressante. La même stratégie que pour la protéine tHisF a été utilisée. Cependant, la position mutée en azido-phénylalanine à l'intérieur de la cavité (mutation S477Z, Z signifiant le résidu azido-phénylalanine) étant inaccessible au cofacteur, il a été entrepris d'agrandir le pore de la protéine (figure 75, droite) par une quadruple mutation en alanine, ce qui permit le couplage du cofacteur en conditions douces (4°C, tampon pH 7,4). La protéine fut testée pour la catalyse de réactions de cyclopropanation (réaction figure 75). Après optimisation des conditions de réaction (PIPES pH 7,4, 1,75M NaBr), la protéine ne portant que la mutation conférait un excès énantiomérique encourageant de 38%, mais un mauvais ratio cyclopropane/alcool de 0,6.

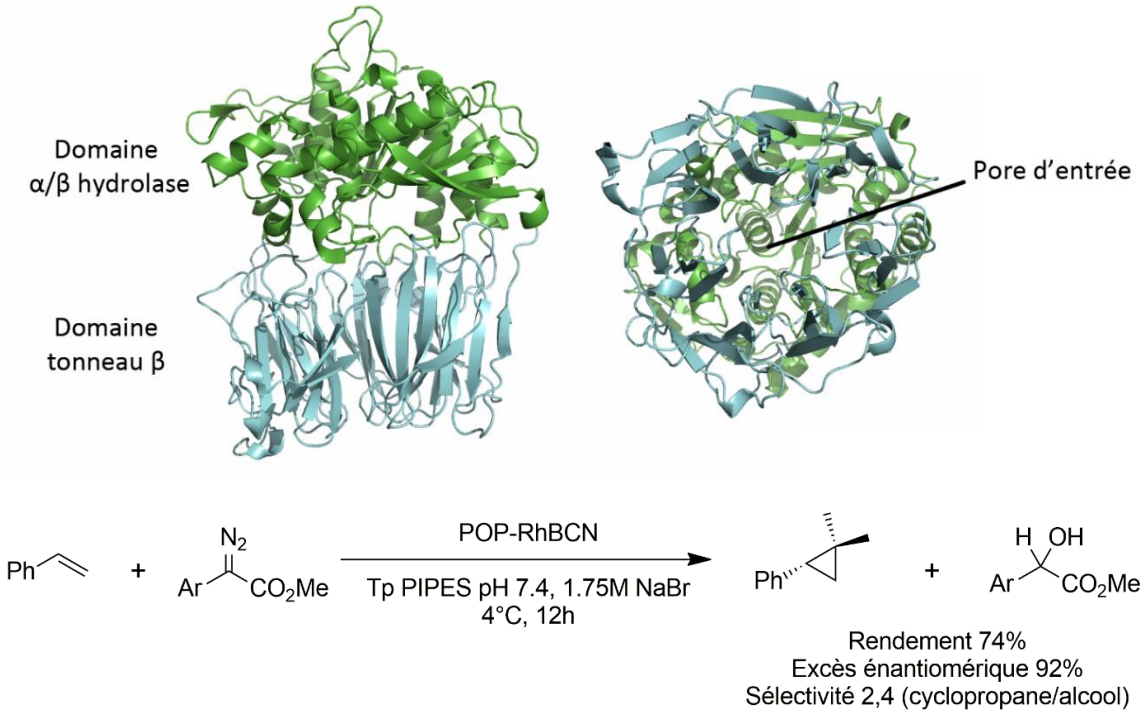


Figure 75 : Prolyl organopeptidase porcine utilisée par l'équipe de J.C. Lewis (homologue à celle de P. furiosis). Réaction catalysée par le biohybride créé. Les résultats de rendements concernent la protéine portant 4 mutations alanine au niveau du pore d'entrée, ainsi que les mutations S477Z (p-azido-L-phenylalanine), L328H, G99F, G594F.

Les auteurs de cette étude ont alors cherché à positionner une histidine à proximité du métal afin de rigidifier le complexe en coordinant un ion de rhodium et améliorer la sélectivité. Ainsi, la mutation L328H parvint à augmenter l'énantiosélectivité jusqu'à 85% et le ratio cyclopropane/alcool à 1,6. Des mutations subséquentes en acides aminés encombrants (phénylalanines) autour du site catalytique putatif ont permis de multiplier le rendement général par 3 par rapport au biohybride sans mutations, ainsi que de passer à un excès énantiomérique de 92% et un ratio cyclopropane/alcool à 2,4. Cette approche par mutations rationnelles s'est avérée fructueuse et a permis de cartographier les acides aminés bordant le site catalytique.

II.2.4.2 Les acides aminés artificiels pour la liaison dative

Comme nous l'avons vu plus haut, la principale difficulté lors de l'ingénierie d'une liaison dative entre la protéine et un métal concerne le positionnement des résidus de manière à créer un ligand multidenté. Cette limitation peut être contournée par l'insertion d'acides aminés ayant directement une forte affinité pour le métal. Ceci revient à l'idée d'insérer directement un complexe métallique lors de la synthèse protéique. Plusieurs acides aminés artificiels de ce type existent, mais un seul a pour le moment été utilisé pour la création de métalloenzyme artificielle : la bipyridine-alanine (BpyAla, nommé X ci-dessous).

L'équipe de Roelfes s'est en effet intéressée à la modification du facteur de transcription LmrR avec cet acide aminé artificiel. Cette protéine avait déjà été modifiée covalemment avec une phénanthroline pour donner une Diels-Alderase artificielle présentant une très bonne énantiosélectivité (voir chapitre couplage covalent), ainsi qu'une métallo-hydratase et un catalyseur de Friedel-Craft avec une série de dérivés indoles. Cette nouvelle approche utilise le codon ambre pour l'insertion directe de l'acide aminé non naturel pour à nouveau la réaction de Friedel-Craft.

Trois positions ont été testées pour son incorporation : N19X, M89X et F93X. De manière intéressante, la position donnant le meilleur catalyseur (M89) est la même que celle utilisée précédemment lors du couplage covalent de cofacteurs à une cystéine. La mutation systémique en alanine des acides aminés bordant la position modifiée (« alanine-scanning ») a été effectuée (figure 76, acides aminés en bâton sur l'image) et a permis de montrer que toutes les positions sauf K22 jouent un rôle significatif soit sur le taux de conversion, soit sur l'énantiosélectivité. Ces observations ont prouvé que la catalyse était bien dépendante de la sphère de coordination due aux acides aminés environnants. Les positions amenant aux plus grands changements dans l'énantiosélectivité ont été choisies pour un second tour de mutagénèse (H86 et F96). Ces deux positions ont été changées indépendamment en 4 autres acides aminés couvrant diverses propriétés (I/S/W/D pour H86 et I/H/W/D pour F93) et l'activité catalytique des différents variants testée pour la réaction de Friedel-Craft (figure 76, substituants **a**). Il est apparu que la position 86 s'accordait bien avec des acides aminés apolaires, quelle que soit leur taille, tandis que la position 93 nécessitait des acides aminés aromatiques, laissant croire que cette position participait à des interactions de type « π stacking ». Indépendamment de cette étude de mutagénèse, un double mutant H86A/E107A a également été test après avoir observé que les variants indépendants présentaient une sélectivité accrue. Il s'est avéré que les mutations n'étaient pas additives.

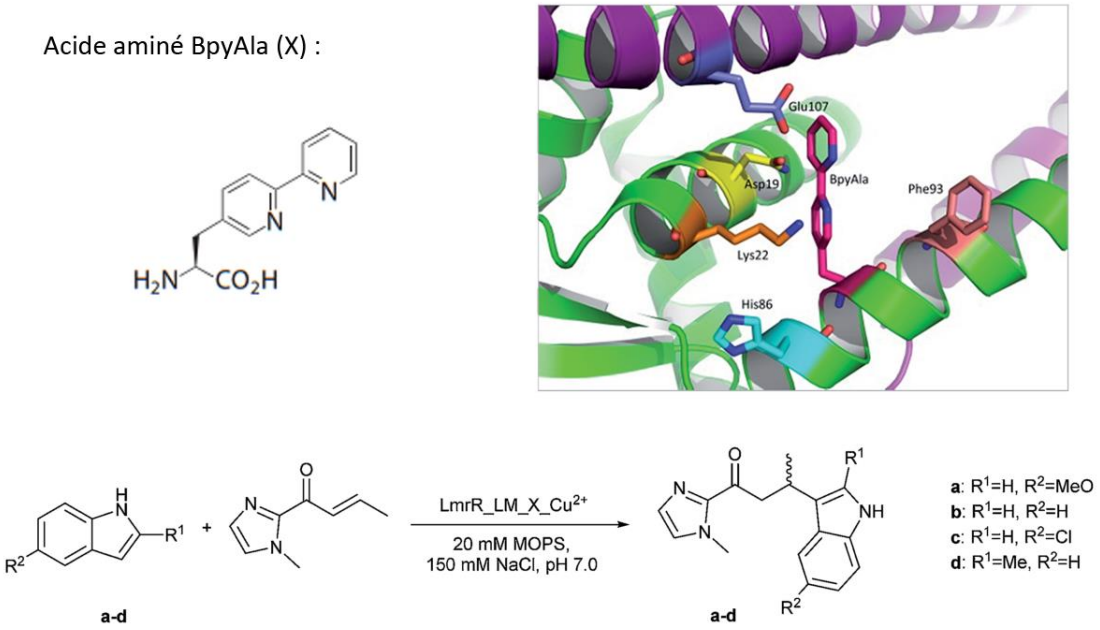


Figure 76 : Protéine LmrR possédant un acide aminé non naturel BpyAla (construction PyMOL en encart). Réaction de Friedel-Craft catalysée en milieu aqueux par le biohybride de Cu (II) (selon [175]).

Cette étude illustre encore une fois l'intérêt mais aussi les inconvénients d'une approche par tâtonnement : il a été apparemment possible de cartographier la sphère de coordination du substrat par une expérience de « alanine-scanning », cependant bien que les mutagénèses à saturation à des points précis aient donné des informations quant à la nature des interactions mises en jeu, la plupart des mutations se sont révélées non additives. Il est donc clair qu'une approche par évolution dirigée serait une solution intéressante pour pallier à ce problème.

Chapitre III - Contexte scientifique de la thèse

Nous avons pu constater au cours de cette introduction que la question principale lors de l'ingénierie de protéines et plus spécifiquement de la création d'un catalyseur biohybride est le choix de l'architecture protéique. Celle-ci doit pouvoir résister aux conditions de couplage chimique ou celles imposées lors de la catalyse, c'est-à-dire tolérer une certaine proportion de solvant organique et être stable dans la plus grande gamme de température possible. Elle doit également être disponible facilement et en grande quantité, les vitesses de catalyse étant généralement loin d'égaler celles de protéines naturelles, une grande quantité est nécessaire pour leur étude. Enfin, comme décrit lors de la première partie, la capacité d'évolution par ingénierie est primordiale, car malgré la promiscuité enzymatique, il n'est pas possible d'avoir une enzyme spécialiste de plusieurs réactions. Il est donc nécessaire de pouvoir adapter le biohybride, le tailler sur mesure pour la réaction désirée.

Le laboratoire de Modélisation et Ingénierie des Protéines de l'I2BC travaille sur la conception de protéines artificielles nommées αRep. Nous verrons ici en quoi consistent les protéines à motifs répétés, dont les αRep sont des représentants, et pourquoi ce sont des candidats de choix pour la conception de biohybrides entièrement artificiels.

III.1 Une architecture protéique modulaire

Le premier exemple venant à l'esprit concernant les protéines modulaires est bien entendu les anticorps. De structure globulaire, ces protéines modulables par la recombinaison aléatoire des boucles CDR forment le pilier du système immunitaire acquis. Elles sont majoritairement utilisées en laboratoire pour la détection de protéines comme de petites molécules. Cependant ces protéines n'ont pas le privilège de la modularité ni de la capacité d'interaction. Par exemple, le système immunitaire des agnates (vertébrés démunis de mâchoire tels que les lamproies) utilise une architecture protéique radicalement différente nommée VLR (Variable Lymphocyte Receptors). [176] Ces protéines sont dites à motifs structuraux répétés car elles sont constituées d'une succession d'un motif nommé LRR pour « Leucine Rich Repeats ». Ce dernier est composé de 24 acides aminés et seulement deux éléments de structures secondaires : une hélice α et un brin β. C'est la répétition d'un nombre variable de ce motif simple qui donne des protéines en forme de solénoïde de différentes tailles (figure 77). La composition en acides aminés du motif peut également varier tout en conservant sa structure générale et son interaction

avec les motifs adjacents. La surface concave créée par le feuillet β varie alors en taille ainsi qu'en composition des acides aminés pour donner une surface d'interaction avec les « antigènes » reconnus.

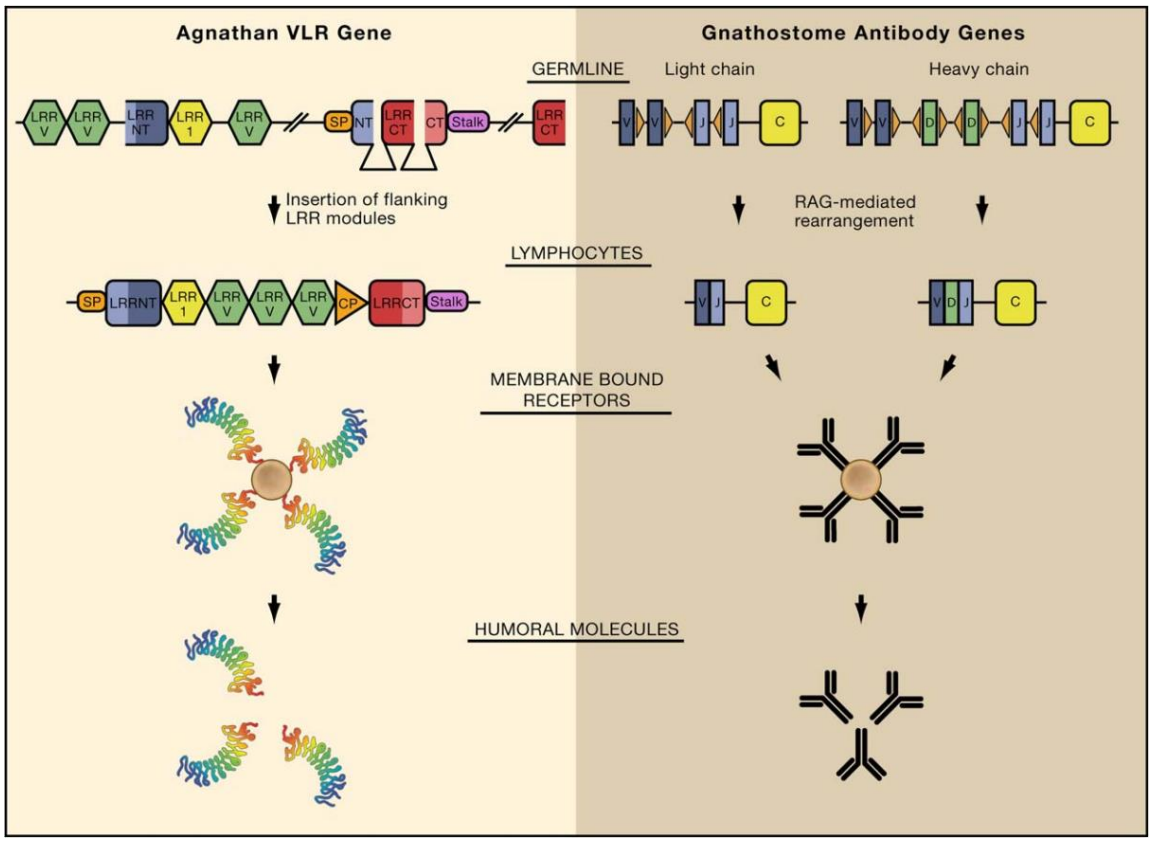


Figure 77 : Comparaison des protéines du système immunitaire acquis des agnates et des gnatostomes. Tandis que les anticorps des gnatostomes sont variables dans la séquence des boucles CDR, les protéines de type VLR des agnates sont variables en taille grâce au nombre de motifs répétés, tout comme dans la séquence de chacun des motifs (selon [176]).

Les protéines répétées en forme de solénoïde allongé sont dites de classe III selon la classification proposée par Kajava.[177] Outre l'exemple probant du système immunitaire des agnates, ces protéines de classe III sont la base d'un vaste répertoire.[58] Leur point commun est leur composition en petits motifs structuraux simples (de 20 à 50 acides aminés) placés en sandwich entre deux motifs particuliers nommés « caps ». Les motifs interagissent entre eux par des interactions hydrophobes et hydrogènes, tandis que les « caps » sont composés d'acides aminés plus hydrophiles dans les parties exposées (appelées exofaces) qui forment des embouts contribuant à la solubilité des protéines. L'étude des séquences des motifs d'une même famille de protéine permet de déterminer les acides aminés conservés responsables de la cohérence de la structure, et les acides aminés variables responsables de la diversité de la surface exposée. Ces protéines ont

depuis une quinzaine d'année été l'objet d'une attention particulière dans la conception de nouveaux interacteurs en alternative aux anticorps.

En effet, une fois le motif consensus « idéal » obtenu et les positions hypervariables repérées, il est virtuellement possible de concevoir un très grand nombre de protéines solubles par assemblage de motifs différents les uns des autres uniquement aux positions hypervariables. Mieux encore, il a été montré que la stabilité de ce genre de protéines est proportionnelle au nombre de motifs qu'elles contiennent.[178] Chaque motif stabilise les motifs voisins et, par effet cumulatif, la protéine entière.

La méthode de *design* de protéines artificielles par cette approche du consensus a été appliquée avec succès pour différentes familles de protéines à motifs répétés [179] : les ankyrines (ANK ou DARPins) [180], les tetratricopeptides (TPR) [181], les LRR [182], les armadillo repeats (ARM) [183] et les HEAT repeats [184]. C'est cette dernière famille de motifs qui a été utilisée au laboratoire pour la création de la banque αRep.

III.2 Construction de la banque de protéines αRep.

Les protéines à motifs répétés de type HEAT forment une famille de protéines présentes aussi bien chez les eucaryotes que les procaryotes. La plupart de ces protéines sont impliquées dans des interactions protéines-protéines bien que des enzymes soient également présentes (une enzyme particulière de cette famille sera abordée en conclusion de cette thèse). Le nom « HEAT » provient de quatre protéines de cette famille : la **H**untingtine, le facteur d'élongation « **E**longated factor 3 », la « protein phosphatase 2**A »**, et la kinase **T**OR (« Target Of Rapamycin »). Le motif est constitué de deux hélices antiparallèles d'une taille allant de 37 à 47 résidus. La divergence entre les protéines de cette famille est telle qu'il est encore possible de les classer dans trois sous-groupes. Cette divergence est problématique pour la création d'un motif consensus : il faut être prudent dans le choix des protéines utilisées dans l'alignement de séquences car l'utilisation de protéines trop éloignées risque fausser l'alignement et proposer des acides aminés incompatibles et délétères pour la stabilité de la structure. Le sous-groupe des protéines de type « PBS HEAT repeats » a donc été utilisé, nommé d'après la **P**hyco**b**ilin **S**ynthase accessory protein, première protéine identifiée appartenant à ce sous-groupe. Ces protéines sont constituées de motifs de longueur et séquence homogène, facilitant la création d'un consensus. Elles sont également très présentes chez les organismes thermophiles, signe de leur stabilité. La protéine Mth187, dont la structure a été publiée en 2006 [185] (code PDB : 1TE4) a été le point de départ de la création du consensus utilisé pour construire la banque αRep (figure 78).

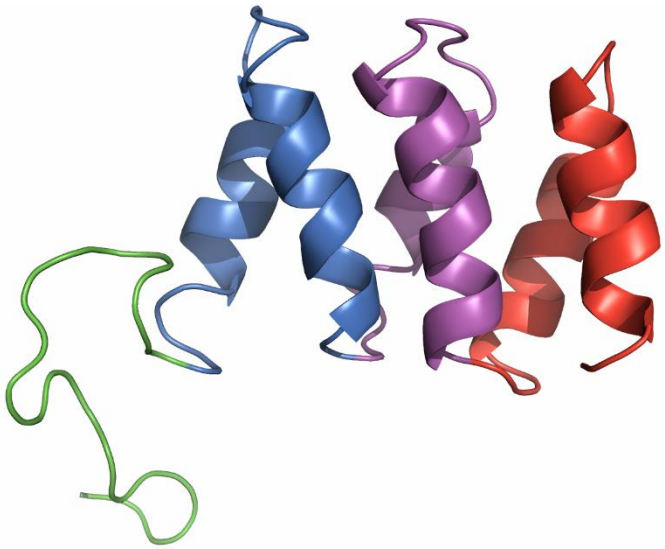


Figure 78 : Protéine Mth187 de M. thermoautotrophicum, de type PDB HEAT repeat, a été le point de départ de la création du motif consensus des αReps. Les motifs bleu et mauve sont des motifs internes tandis que le motif rouge est un « cap » C-terminal.

Cette protéine de *M. thermoautotrophicum* est composée de trois motifs de type HEAT repeat, dont deux motifs dit « internes » et un motif « cap » C-terminal de séquence clairement divergente. La portion N-terminale est constituée d'une boucle désordonnée. Une recherche BLAST effectuée sur la base des deux premiers motifs a permis de trouver les homologues les plus proches et de créer un premier consensus. Une seconde recherche sur la base de ce consensus a permis d'obtenir 500 séquences issues de 100 protéines différentes. [184] La représentation en logo-séquence permet de visualiser le consensus de 31 résidus qui s'en dégage (figure 79A).

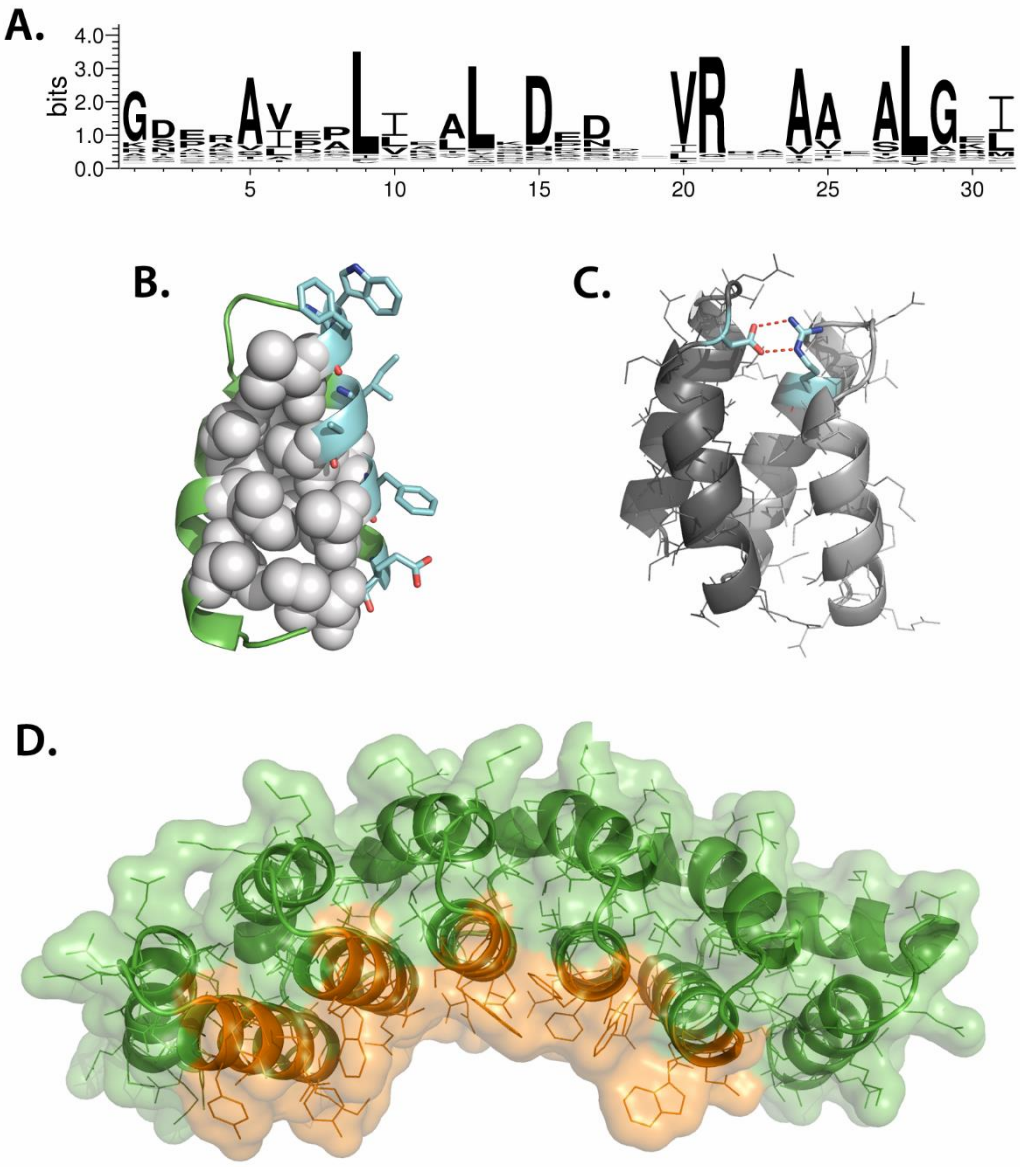


Figure 79 : A. Représentation en logo séquence de l'alignement des homologues du motif de Mth187. La taille des lettres représente la conservation de ces acides aminés à ces positions du motif. B. Structure d'un motif de la banque αRep créée sur la base du motif consensus. En sphères grises sont les acides aminés hydrophobes responsables du maintien de la structure et en bâtons bleus sont les acides aminés variables, tous sur la même face du motif. C. Structure de deux motifs internes (l'un en gis clair, l'autre en gris foncé). En bleu sont représentés les deux acides aminés arginine et aspartate établissant deux liaisons hydrogene entre deux motifs successifs. Ce réseau de liaisons hydrogènes et la continuité du cœur hydrophobe contribuent à la stabilité de l'ensemble de la protéine. D. Structure cristallographique d'une protéine αRep à quatre motifs internes, la surface variable est colorisée en orange. On remarque que le motif C-cap (à droite) est constant.

Le consensus obtenu est donc GDERAVEPLIKALKDED**XX**VR**XX**AA**X**ALG**X**I, où les X représentent les positions variables qui sont donc modifiées aléatoirement au sein de la bibliothèque αRep. La variabilité des acides aminés à ces positions n'a pas été faite de manière totalement aléatoire (par l'utilisation de codons NNK par exemple). Il a fallu en effet respecter les biais en acides aminés propres à chaque position (figure 80), comme la nécessité d'acides aminés de petite taille (sérine et alanine) en position 23 ou encore l'absence de proline à certaines positions. D'autres biais viennent de considérations pratiques telles que le choix d'éviter la présence de cystéines ou la nécessité d'avoir des codons en position 30 compatibles avec un site de restriction BsmBI utilisé lors de la création de la banque. Concernant les modules N- et C-terminaux, le motif en partie N-terminale est basé sur une protéine possédant un consensus très proche de Mth187 et possédant un motif N-terminal de 34 résidus exposé au solvant (protéine A0B7C6_METTP). Les positions de ce N-cap correspondant aux positions variables ont été également randomisées. Le motif en partie C-terminale correspond exactement à celui de la protéine Mth187, sans résidus randomisés.

Position	Codons	Encoded Amino Acids
18	snd (4)	A, D, E, G, H, L, P, R, V
	tsg (1)	S, W
	tmy (1)	S, Y
19	khy (4)	A, D, F, S, V, Y
	van (1)	D, E, H, K, N, Q
	tgg (1)	W
22	vdr (5)	E, G, K, L ,M ,Q ,R ,V
	tsg (1)	S, W
23	kck	A, S
26	khy (5)	A, D, F, S, V, Y
	tgg (1)	W
30	vaa	E, K, Q

Figure 80 : Schéma de randomisation propre à chaque position variable de la banque αRep. La première colonne indique la position, la deuxième indique les codons utilisés (avec leur proportion relative entre parenthèses) et la troisième indique les acides aminés codés par chaque codon. La distribution des acides aminés provenant de clones choisis aléatoirement dans la banque correspond à celle encodée dans ce tableau (selon [184]).

La création des banques αRep tire plein avantage du caractère répété de ces protéines. Elles sont créées par polymérisation de micro-gènes dégénérés. La méthode d'amplification circulaire de l'ADN utilisée (RCA pour « rolling circle amplification »), couplée à la polymérisation orientée (grâce à l'utilisation de l'enzyme de restriction BsmBI) puis à la ligation intramoléculaire des plasmides créés permet d'obtenir de très grandes efficacités de clonage et donc des banques de très grande taille (figure 81).[184]

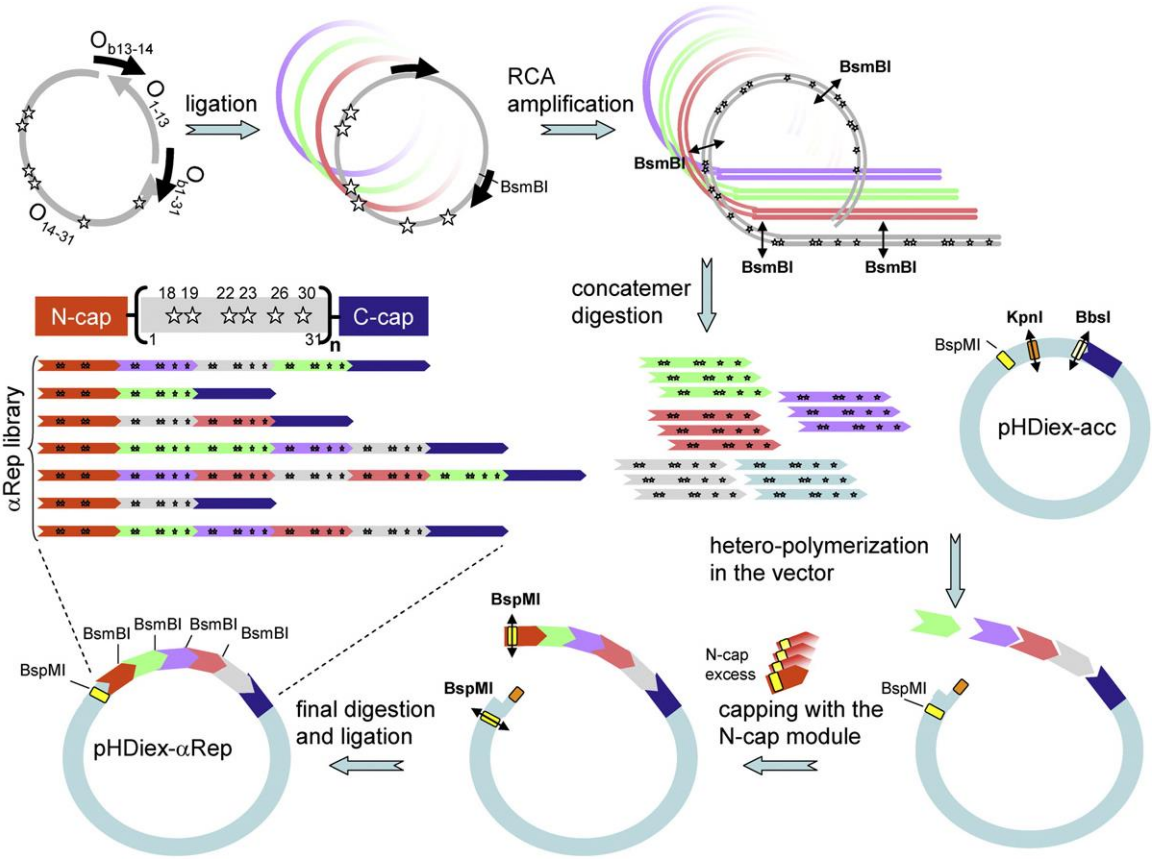


Figure 81 : Méthode de création des banques αRep. Les microgènes dégénérés (codant chacun pour un motif différent) sont amplifiés par la méthode RCA. Ils sont ensuite digérés selon un site de restriction asymétrique grâce à l'enzyme BsmBI. Les microgènes sont ensuite polymérisés dans le vecteur accepteur, la polymérisation étant arrêtée par un ajout d'une grande quantité du microgène N-cap. Les plasmides sont enfin refermés par le site unique BspMI, cette ligation intramoléculaire est bien plus efficace que la ligation d'un insert dans un vecteur accepteur (selon [184])

Une première banque, appelée αRep-1.0 contenant 3.10^8 clones indépendants a été construite sur la base de ces considérations, selon un schéma de dégénérescence aux positions hypervariables conduisant à un enrichissement en acides aminés aromatiques.[184] Une banque optimisée de seconde génération contenant 2.10^9 clones indépendants a ensuite été construite[186]. Dans cette banque appelée αRep-2.1, le nombre de séquences non codantes induites par des mésappariements survenus lors de la construction de la banque a été minimisé et la variabilité des acides aminés aux positions variables mime la diversité naturelle des protéines de cette sous-famille de motifs HEAT.

L'étude de plusieurs protéines issues de ces deux banques contenant un nombre variable de motifs internes, a permis de montrer que les protéines obtenues grâce à cette analyse de séquences peuvent être produites chez *E. coli* en grande quantité (plusieurs dizaine de milligrammes pas litre de culture), sous forme soluble et repliée. Des expériences de

dénaturation ont permis de constater que leur résistance à la chaleur croissait effectivement avec le nombre de motifs internes, une protéine contenant quatre motifs internes possédant une température de demi-dénaturation de 80°C contre 71°C pour une protéine contenant un seul motif interne, et 95°C pour une protéine avec 6 motifs internes. [184]

III.3 La structure αRep, une architecture de choix pour la conception de catalyseurs biohybrides ?

La banque αRep-2.1 est régulièrement utilisée dans les différents projets du laboratoire MIP, pour sélectionner des variants αRep interagissant spécifiquement avec des cibles protéiques choisies afin d'obtenir des outils moléculaires tels que des chaperons de cristallisation [187], des outils de ciblage intracellulaire [186] ou des briques de nano-assemblages [188]. Afin d'élargir les champs d'application de cette charpente protéique entièrement artificielle et hautement modulable, il nous paraissait intéressant d'explorer quelques pistes pour développer une nouvelle génération de catalyseurs biohybrides.

Mon projet de thèse s'inscrit dans ce contexte avec pour objectif la conception de nouveaux biocatalyseurs hybrides selon l'approche par couplage covalent décrite dans le chapitre 2 : des complexes métalliques responsables de l'acte catalytique seront donc greffés sur l'architecture protéique modulable αRep pouvant intervenir dans le contrôle et le positionnement du substrat.

III.4 La protéine αRep-A3, un représentant particulier de la banque αRep

Pour démarrer notre projet, nous avons choisi une protéine particulière issue de la banque αRep-1.0, le variant αRep-A3. La protéine analysée contenant quatre motifs internes, présente la particularité de former un dimère stable en solution (non dissocié en tamis moléculaire). La structure cristallographique de cette protéine révèle que le dimère est formé par l'interaction des faces variables de la protéine, selon une symétrie cristallographique de type C2 (180° autour d'un axe).[184] Cette dimérisation provoque la formation d'une crevasse (figure 82).

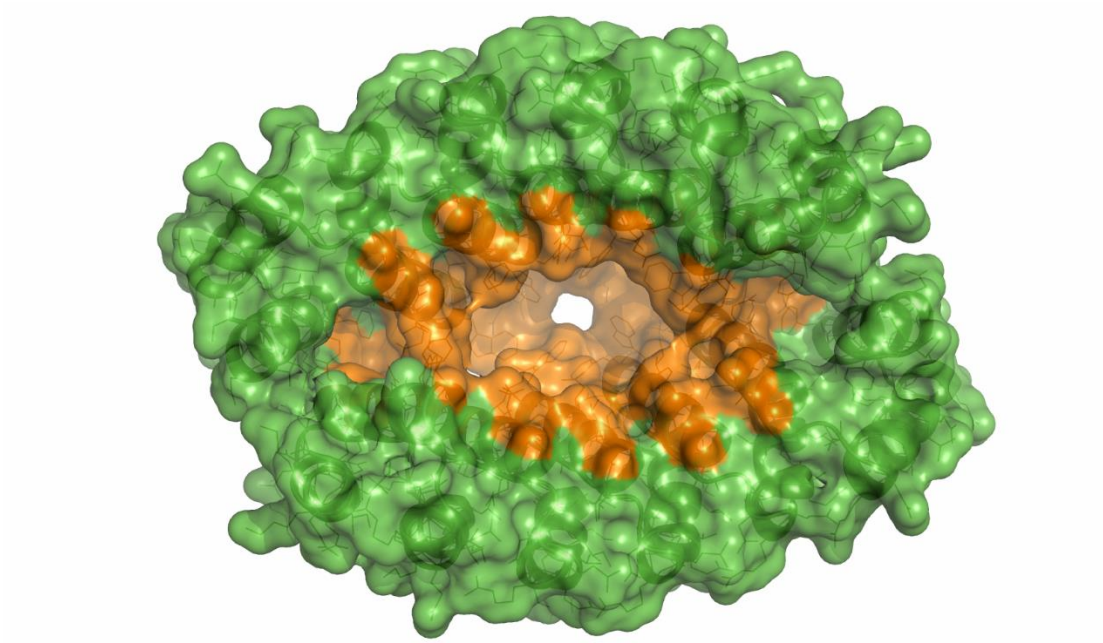


Figure 82 : Structure du dimère de la protéine αRep-A3 (code PDB : 3LTJ), en orange sont représentées les surfaces variables. Le dimère possède une symétrie cristallographique parfaite de type C2 dans ce cristal.

Cette protéine semble posséder tous les prérequis nécessaires à la formation d'un catalyseur biohybride : Elle est facile à produire en grande quantité dans un hôte d'expression simple (*E. coli*). Elle ne possède pas de cystéines libres ni de ponts disulfures, ce qui permet d'utiliser l'approche par couplage covalent. Elle est soluble et très stable avec une température de fusion de 80°C. Elle possède une crevasse, conférant un milieu enfoui et chiral, propice à une catalyse énantiosélective. Enfin, la surface de la crevasse étant composée des acides aminés variables de la banque αRep, nous savons par l'expérience qu'il est possible de les modifier sans affecter la structure tridimensionnelle de la protéine, ce qui est un atout majeur pour la mise en place d'une stratégie d'évolution dirigée d'un biohybride.

III.5 Choix des complexes métalliques

Dans un premier temps notre choix s'est porté sur de complexes de petite taille comme la phénanthroline et la terpyridine. En effet celles-ci sont connues non seulement pour fixer un grand nombre de métaux mais aussi pour posséder des réactivités aussi variées que des réactions de Diels alder[72,162], Friedel-Craft [189], hydroxylation[190] ou époxydation d'oléfines[191]. La taille réduite de ces deux ligands leur permet de facilement s'insérer dans une crevasse de protéine.

J'ai également utilisé des ligands de taille plus importante tel que les porphyrines. Comme vu au cours de cette introduction, les porphyrines sont capables de catalyser de nombreuses réactions d'oxydation. De plus, le potentiel redox peut être modulé par l'ajout de groupements électrodonneurs ou électroattracteurs sur le cycle porphyrinique.[192] Dans ce manuscrit je décrirai en particulier l'utilisation d'une TetraPhenylPorphyrine de manganèse (MnTPP).

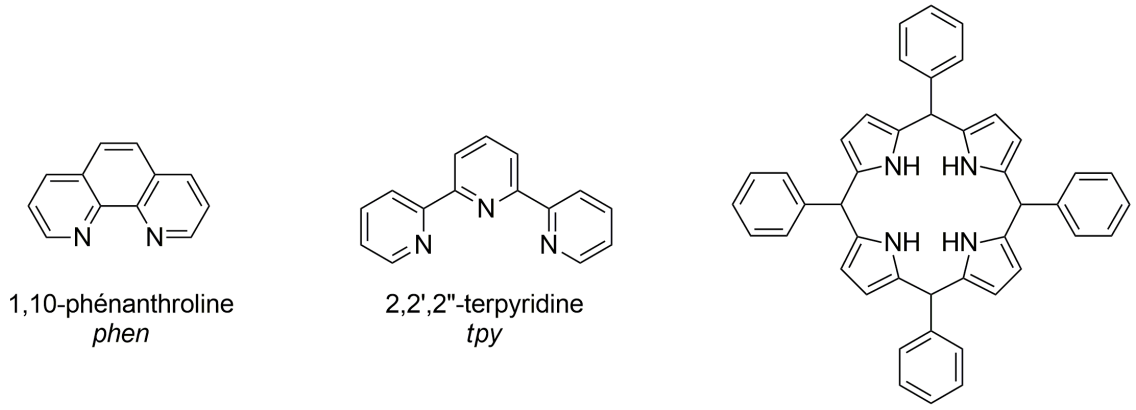


Figure 83 : Formule topologique des trois ligands utilisés dans ces travaux. En romain est notée la formule IUPAC, en italique est le nom raccourci.

Le but de ma thèse a donc été de modifier cette protéine αRep-A3, par la stratégie du couplage covalent, avec différents types de cofacteurs métalliques, afin de créer le premier catalyseur biohybride issu d'une protéine artificielle à motifs répétés connu à ce jour. Ce travail s'articule en trois axes : le premier consiste en la création d'un catalyseur grâce au couplage covalent de deux ligands de type phénantroline par dimère. Ce biohybride a été testé pour des réactions de Diels-Alder, une réaction dont il n'existe pas d'enzyme naturelle dans l'industrie.[193] Par la suite, nous avons cherché à améliorer notre métalloenzyme artificielle par l'ajout d'un lien entre les deux sous-unités du dimère permettant ainsi le couplage covalent d'un unique ligand de type phénantroline ou terpyridine dans la crevasse. Cette nouvelle métalloenzyme a été comparée pour la même réaction. Enfin, cette stratégie permettant le greffage d'un unique complexe dans la crevasse, nous avons pu explorer l'utilisation d'une porphyrine. Chacun de ces types de biohybrides a été caractérisé de manière biophysique et biochimique avant d'étudier ses propriétés catalytiques dans des réactions appropriées.

Bibliographie

(1) (2000) Green Chemistry By Paul T. Anastas and John C. Warner. Oxford University Press: Oxford. 2000. Paperback. 135 pp. £14.99. ISBN 0-19-850698-9. *Org. Process Res. Dev. 4*, 437-438.

(2) Sun, H., Chen, P., Li, D., Li, Y., et Hou, T. (2016) Directly Binding Rather than Induced-Fit Dominated Binding Affinity Difference in (*S*)- and (*R*)-Crizotinib Bound MTH1. *J. Chem. Theory Comput. 12*, 851-860.

(3) Wilson, C., Agafonov, R. V, Hoemberger, M., Kutter, S., Zorba, A., Halpin, J., Buosi, V., Otten, R., Waterman, D., Theobald, D. L., et Kern, D. (2015) Using ancient protein kinases to unravel a modern cancer drug's mechanism. *Science (80-.). 347*, 882-886.

(4) Knowles, W. S. (2002) Asymmetric hydrogenations (Nobel lecture). *Angew. Chem. Int. Ed. Engl. 41*, 1999-2007.

(5) Kagan, H. (2003) La synthèse asymétrique de composés biologiquement actifs. *Actual. Chim.* 10-14.

(6) Gröger, H., et Asano, Y. (2012) Introduction – Principles and Historical Landmarks of Enzyme Catalysis in Organic Synthesis, in *Enzyme Catalysis in Organic Synthesis*, p 1-42. CHAP, Wiley-VCH Verlag GmbH & Co. KGaA.

(7) Darwin, C. (1859) On the Origin of the Species. *Darwin*.

(8) Sun, Z., Wikmark, Y., Bäckvall, J. E., et Reetz, M. T. (2016) New Concepts for Increasing the Efficiency in Directed Evolution of Stereoselective Enzymes. *Chem. - A Eur. J. 22*, 5046-5054.

(9) Holliday, G. L., Andreini, C., Fischer, J. D., Rahman, S. A., Almonacid, D. E., Williams, S. T., et Pearson, W. R. (2012) MACiE: Exploring the diversity of biochemical reactions. *Nucleic Acids Res. 40*.

(10) Lunzer, M., Miller, S. P., Felsheim, R., et Dean, A. M. (2005) The biochemical architecture of an ancient adaptive landscape. *Science 310*, 499-501.

(11) Lunzer, M., Brian Golding, G., et Dean, A. M. (2010) Pervasive cryptic epistasis in molecular evolution. *PLoS Genet. 6*, 1-10.

(12) Romero, P. A., et Arnold, F. H. (2009) Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol. 10*, 866-76.

(13) Galmés, J., Conesa, M. À., Díaz-Espejo, A., Mir, A., Perdomo, J. A., Niinemets, Ü., et Flexas, J. (2014) Rubisco catalytic properties optimized for present and future climatic conditions. *Plant Sci. 226*, 61-70.

(14) Reetz, M. T. (2004) Controlling the enantioselectivity of enzymes by directed evolution: practical and theoretical ramifications. *Proc. Natl. Acad. Sci. U. S. A. 101*, 5716-5722.

(15) Urvoas, A., Valerio-Lepiniec, M., et Minard, P. (2012) Artificial proteins from combinatorial approaches. *Trends Biotechnol. 30*, 512-520.

(16) Cadwell, R. C., et Joyce, G. F. (1992) Randomization of genes by PCR mutagenesis. *Genome Res.*, p 28-33.

(17) Camps, M., Naukkarinen, J., Johnson, B. P., et Loeb, L. A. (2003) Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I. *Proc. Natl. Acad. Sci. U. S. A. 100*, 9727-32.

(18) Wong, T. S., Tee, K. L., Hauer, B., et Schwaneberg, U. (2004) Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. *Nucleic Acids Res. 32*, e26.

(19) Murakami, H., Hohsaka, T., et Sisido, M. (2002) Random insertion and deletion of arbitrary number of bases for codon-based random mutation of DNAs. *Nat. Biotechnol. 20*, 76-81.

(20) Jones, D. D. (2005) Triplet nucleotide removal at random positions in a target gene: The tolerance of TEM-1 ??-lactamase to an amino acid deletion. *Nucleic Acids Res. 33*, 1-8.

(21) Liu, J., et Cropp, T. A. (2012) A method for multi-codon scanning mutagenesis of proteins based on asymmetric transposons. *Protein Eng. Des. Sel. 25*, 67-72.

(22) Chen, K. Q., et Arnold, F. H. (1991) Enzyme engineering for nonaqueous solvents: random mutagenesis to enhance activity of subtilisin E in polar organic media. *Biotechnol. Nat. Publ. Co. 9*, 1073-1077.

(23) Chen, K., et Arnold, F. H. (1993) Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl. Acad. Sci. USA 90*, 5618-5622.

(24) You, L., et Arnold, F. H. (1994) Directed evolution of subtilisin E in Bacillus subtilis to enhance total activity in aqueous dimethylformamide. *Protein Eng. 9*, 77-83.

(25) Stemmer, W. P. (1994) Rapid evolution of a protein in vitro by DNA shuffling. *Nature*.

(26) Bergquist, P. L., Reeves, R. A., et Gibbs, M. D. (2005) Degenerate oligonucleotide gene shuffling (DOGS) and random drift mutagenesis (RNDM): Two complementary techniques for enzyme evolution. *Biomol. Eng. 22*, 63-72.

(27) Ness, J. E., Kim, S., Gottman, A., Pak, R., Krebber, A., Borchert, T. V, Govindarajan, S., Mundorff, E. C., et Minshull, J. (2002) Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. *Nat. Biotechnol. 20*, 1251-1255.

(28) Lee, S. H., Ryu, E. J., Kang, M. J., Wang, E. S., Piao, Z., Choi, Y. J., Jung, K. H., Jeon, J. Y. J., et Shin, Y. C. (2003) A new approach to directed gene evolution by recombined extension on truncated templates (RETT). *J. Mol. Catal. B Enzym. 26*, 119-129.

(29) Patrick, W. M., et Gerth, M. L. (2014) ITCHY: Incremental Truncation for the Creation of Hybrid enzYmes. *Methods Mol. Biol. 1179*, 225-244.

(30) Kawarasaki, Y., Griswold, K. E., Stevenson, J. D., Selzer, T., Benkovic, S. J., Iverson, B. L., et Georgiou, G. (2003) Enhanced crossover SCRATCHY: construction and high-throughput screening of a combinatorial library containing multiple non-homologous crossovers. *Nucleic Acids Res. 31*, e126.

(31) Villiers, B. R. M., Stein, V., et Hollfelder, F. (2010) USER friendly DNA recombination (USERec): A simple and flexible near homology-independent method for gene library construction. *Protein Eng. Des. Sel. 23*, 1-8.

(32) O'Maille, P. E., Bakhtina, M., et Tsai, M. D. (2002) Structure-based combinatorial protein engineering (SCOPE). *J. Mol. Biol. 321*, 677-691.

(33) Herman, A., et Tawfik, D. S. (2007) Incorporating Synthetic Oligonucleotides via Gene Reassembly (ISOR): A versatile tool for generating targeted libraries. *Protein Eng. Des. Sel. 20*, 219-226.

(34) Li, M. Z., et Elledge, S. J. (2007) Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat. Methods 4*, 251-256.

(35) Engler, C., Gruetzner, R., Kandzia, R., et Marillonnet, S. (2009) Golden gate shuffling: A one-pot DNA shuffling method based on type ils restriction enzymes. *PLoS One 4*.

(36) Walter, K. U., Vamvaca, K., et Hilvert, D. (2005) An active enzyme constructed from a 9-amino acid alphabet. *J. Biol. Chem. 280*, 37742-37746.

(37) Yang, G., et Withers, S. G. (2009, novembre 23) Ultrahigh-throughput FACS-based screening for directed enzyme evolution. *ChemBioChem*. WILEY-VCH Verlag.

(38) Parra, L. P., Agudo, R., et Reetz, M. T. (2013) Directed evolution by using iterative saturation mutagenesis based on multiresidue sites. *ChemBioChem 14*, 2301-2309.

(39) Reetz, M. T., Bocola, M., Carballeira, J. D., Zha, D., et Vogel, A. (2005) Expanding the range of substrate acceptance of enzymes: Combinatorial active-site saturation test. *Angew. Chemie - Int. Ed. 44*, 4192-4196.

(40) Reetz, M. T., et Carballeira, J. D. (2007) Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc. 2*, 891-903.

(41) Prasad, S., Bocola, M., et Reetz, M. T. (2011) Revisiting the lipase from pseudomonas aeruginosa: Directed evolution of substrate acceptance and enantioselectivity using iterative saturation mutagenesis. *ChemPhysChem 12*, 1550-1557.

(42) Agudo, R., Roiban, G. D., et Reetz, M. T. (2013) Induced axial chirality in biocatalytic asymmetric ketone reduction. *J. Am. Chem. Soc. 135*, 1665-1668.

(43) Savile, C. K., Janey, J. M., Mundorff, E. C., Moore, J. C., Tam, S., Jarvis, W. R., Colbeck, J. C., Krebber, A., Fleitz, F. J., Brands, J., Devine, P. N., Huisman, G. W., et Hughes, G. J. (2010) Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science (80-.). 329*, 305-309.

(44) Abrahamson, M. J., Vázquez-Figueroa, E., Woodall, N. B., Moore, J. C., et Bommarius, A. S. (2012) Development of an amine dehydrogenase for synthesis of chiral amines. *Angew. Chemie - Int. Ed. 51*, 3969-3972.

(45) Ranoux, A., Karmee, S. K., Jin, J., Bhaduri, A., Caiazzo, A., Arends, I. W. C. E., et Hanefeld, U. (2012) Enhancement of the Substrate Scope of Transketolase. *ChemBioChem 13*, 1921-1931.

(46) Jochens, H., et Bornscheuer, U. T. (2010) Natural diversity to guide focused directed evolution. *ChemBioChem 11*, 1861-1866.

(47) Zhang, Z. G., Roiban, G. D., Acevedo, J. P., Polyak, I., et Reetz, M. T. (2013) A new type of stereoselectivity in Baeyer-Villiger reactions: Access to E- and Z-olefins. *Adv. Synth. Catal. 355*, 99-106.

(48) Pham, S. Q., Pompidor, G., Liu, J., Li, X.-D., et Li, Z. (2012) Evolving P450pyr hydroxylase for highly enantioselective hydroxylation at non-activated carbon atom. *Chem. Commun. 48*, 4618.

(49) Dennig, A., Marienhagen, J., Ruff, A. J., Guddat, L., et Schwaneberg, U. (2012) Directed Evolution of P450 BM3 into a p-Xylene Hydroxylase. *ChemCatChem 4*, 771-773.

(50) Agudo, R., Roiban, G. D., et Reetz, M. T. (2012) Achieving Regio- and Enantioselectivity of P450-Catalyzed Oxidative CH Activation of Small Functionalized Molecules by Structure-Guided Directed Evolution. *ChemBioChem 13*, 1465-1473.

(51) van Leeuwen, J. G. E., Wijma, H. J., Floor, R. J., Van der Laan, J. M., et Janssen, D. B. (2012) Directed Evolution Strategies for Enantiocomplementary Haloalkane Dehalogenases: From Chemical Waste to Enantiopure Building Blocks. *ChemBioChem 13*, 137-148.

(52) Preiswerk, N., Beck, T., Schulz, J. D., Milovník, P., Mayer, C., Siegel, J. B., Baker, D., et Hilvert, D. (2014) Impact of scaffold rigidity on the design and evolution of an artificial Diels-Alderase. *Proc. Natl. Acad. Sci. U. S. A. 111*, 8013-8.

(53) Wijma, H. J., Floor, R. J., Bjelic, S., Marrink, S. J., Baker, D., et Janssen, D. B. (2015) Enantioselective enzymes by computational design and in silico screening. *Angew. Chemie - Int. Ed. 54*, 3726-3730.

(54) Khare, S. D., Kipnis, Y., Greisen, P. J., Takeuchi, R., Ashani, Y., Goldsmith, M., Song, Y., Gallaher, J. L., Silman, I., Leader, H., Sussman, J. L., Stoddard, B. L., Tawfik, D. S., et Baker, D. (2012) Computational redesign of a mononuclear zinc metalloenzyme for organophosphate hydrolysis. *Nat. Chem. Biol. 8*, 294-300.

(55) Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Rothlisberger, D., Zanghellini, A., Gallaher, J. L., Betker, J. L., Tanaka, F., Barbas, C. F., Hilvert, D., Houk, K. N., Stoddard, B. L., et Baker, D. (2008) De Novo Computational Design of Retro-Aldol Enzymes. *Science (80-.). 319*, 1387-1391.

(56) Röthlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J. L., Althoff, E. A., Zanghellini, A., Dym, O., Albeck, S., Houk, K. N., Tawfik, D. S., et Baker, D. (2008) Kemp elimination catalysts by computational enzyme design. *Nature 453*, 190-5.

(57) Acevedo-Rocha, C. G., Agudo, R., et Reetz, M. T. (2014) Directed evolution of stereoselective enzymes based on genetic selection as opposed to screening systems. *J. Biotechnol. 191*, 3-10.

(58) Chevrel, A. (2014) Ingénierie d'une ossature à motifs structuraux répétés par évolution dirigée : développements et applications d'un nouvel outil de reconnaissance moléculaire. Université Paris Sud.

(59) Ponsard, I., Galleni, M., Soumillion, P., et Fastrez, J. (2001) Selection of metalloenzymes by catalytic activity using phage display and catalytic elution. *Chembiochem 2*, 253-259.

(60) Boersma, Y. L., Dröge, M. J., van der Sloot, A. M., Pijning, T., Cool, R. H., Dijkstra, B. W., et Quax, W. J. (2008) A novel genetic selection system for improved enantioselectivity of Bacillus subtilis lipase A. *ChemBioChem 9*, 1110-1115.

(61) Reetz, M. T., Höbenreich, H., Soni, P., et Fernández, L. (2008) A genetic selection system for evolving enantioselectivity of enzymes. *Chem. Commun. (Camb).* 5502-5504.

(62) Digianantonio, K. M., et Hecht, M. H. (2016) A protein constructed de novo enables cell growth by altering gene regulation. *Proc. Natl. Acad. Sci.* 201600566.

(63) Hyster, T. K., et Ward, T. R. (2016) Genetic Optimization of Metalloenzymes: Enhancing Enzymes for Non-Natural Reactions. *Angew. Chemie - Int. Ed.* 7344-7357.

(64) Sajonz, P., Schafer, W., Gong, X., Shultz, S., Rosner, T., et Welch, C. J. (2007) Multiparallel microfluidic high-performance liquid chromatography for high-throughput normal-phase chiral analysis. *J. Chromatogr. A 1145*, 149-154.

(65) Reetz, M. T., Kühling, K. M., Wilensek, S., Husmann, H., Häusig, U. W., et Hermes, M. (2001) A GC-based method for high-throughput screening of enantioselective catalysts. *Catal. Today 67*, 389-396.

(66) Leung, D., Kang, S. O., et Anslyn, E. V. (2012) Rapid determination of enantiomeric excess: a focus on optical approaches. *Chem. Soc. Rev. 41*, 448.

(67) Agresti, J. J., Antipov, E., Abate, A. R., Ahn, K., Rowat, A. C., Baret, J.-C. J.-C., Marquez, M., Klibanov, A. M., Griffiths, A. D., et Weitz, D. A. (2010) Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc. Natl. Acad. Sci. 107*, 4004-4009.

(68) Belder, D., Ludwig, M., Wang, L. W., et Reetz, M. T. (2006) Enantioselective catalysis and analysis on a chip. *Angew. Chemie - Int. Ed. 45*, 2463-2466.

(69) Krone, K. M., Warias, R., Ritter, C., Li, A., Acevedo-Rocha, C. G., Reetz, M. T., et Belder, D. (2016) Analysis of Enantioselective Biotransformations Using a Few Hundred Cells on an Integrated Microfluidic Chip. *J. Am. Chem. Soc. 138*, 2102-2105.

(70) Ohla, S., Beyreiss, R., Fritzsche, S., Glaser, P., Nagl, S., Stockhausen, K., Schneider, C., et Belder, D. (2012) Monitoring on-chip pictet-spengler reactions by integrated analytical separation and label-free time-resolved fluorescence. *Chem. - A Eur. J. 18*, 1240-1246.

(71) Fritzsche, S., Ohla, S., Glaser, P., Giera, D. S., Sickert, M., Schneider, C., et Belder, D. (2011) Asymmetric organocatalysis and analysis on a single microfluidic nanospray chip. *Angew. Chemie - Int. Ed. 50*, 9467-9470.

(72) Deuss, P. J., Popa, G., Slawin, A. M. Z., Laan, W., et Kamer, P. C. J. (2013) Artificial Copper Enzymes for Asymmetric Diels-Alder Reactions. *ChemCatChem 5*, 1184-1191.

(73) Holm, R. H., Kennepohl, P., Solomon, E. I., et Solomon, E. I. (1996) Structural and Functional Aspects of Metal Sites in Biology. *Chem. Rev. 96*, 2239-2314.

(74) Nič, M., Jirát, J., Košata, B., Jenkins, A., et McNaught, A. (Éd.). (2009) IUPAC Compendium of Chemical Terminology. IUPAC, Research Triagle Park, NC.

(75) Minkin, V. I. (1999) Glossary of terms used in theoretical organic chemistry. *Pure Appl. Chem. 71*, 1919-1981.

(76) Housecroft, C., et Sharpe, A. (2005) Inorganic Chemistry. *Inorg. Chem. 2nd Ed*.

(77) Breton, S. (2001) The Cellular Physiology of Carbonic Anhydrases. *JOP. J. Pancreas JOP. J. Pancreas – http//www.joplink.net – 22*, 159-164.

(78) Badger, M. R., et Price, G. D. (1994) The Role of Carbonic Anhydrase in Photosynthesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol. 45*, 369-392.

(79) Supuran, C. T. (2016) Structure and function of carbonic anhydrases. *Biochem. J 473*, 2023-2032.

(80) Kiefer, L. L., et Fierke, C. a. (1994) Functional characterization of human carbonic anhydrase II variants with altered zinc binding sites. *Biochemistry 33*, 15233-15240.

(81) Fernández-Gacio, A., Codina, A., Fastrez, J., Riant, O., et Soumillion, P. (2006) Transforming carbonic anhydrase into epoxide synthase by metal exchange. *ChemBioChem 7*, 1013-1016.

(82) Okrasa, K., et Kazlauskas, R. J. (2006) Manganese-substituted carbonic anhydrase as a new peroxidase. *Chemistry 12*, 1587-96.

(83) Jing, Q., Okrasa, K., et Kazlauskas, R. J. (2009) Stereoselective hydrogenation of olefins using rhodium-substituted carbonic anhydrase - A new reductase. *Chem. - A Eur. J. 15*, 1370-1376.

(84) Podtetenieff, J., Taglieber, A., Bill, E., Reijerse, E. J., et Reetz, M. T. (2010) An artificial metalloenzyme: Creation of a designed copper binding site in a thermostable protein. *Angew. Chemie - Int. Ed. 49*, 5151-5155.

(85) Mahy, J.-P., Maréchal, J.-D., et Ricoux, R. (2015) From « hemoabzymes » to « hemozymes »: towards new biocatalysts for selective oxidations. *Chem. Commun. 51*, 2476-2494.

(86) Barry, S. M., Kers, J. A., Johnson, E. G., Song, L., Aston, P. R., Patel, B., Krasnoff, S. B., Crane, B. R., Gibson, D. M., Loria, R., et Challis, G. L. (2012) Cytochrome P450–catalyzed L-tryptophan nitration in thaxtomin phytotoxin biosynthesis. *Nat. Chem. Biol. 8*, 814-816.

(87) Guengerich, F. P., et Munro, A. W. (2013) Unusual cytochrome P450 enzymes and reactions. *J. Biol. Chem. 288*, 17065-17073.

(88) Podust, L. M., et Sherman, D. H. (2012) Diversity of P450 enzymes in the biosynthesis of natural products. *Nat. Prod. Rep. 29*, 1251-66.

(89) McIntosh, J. A., Farwell, C. C., et Arnold, F. H. (2014) Expanding P450 catalytic reaction space through evolution and engineering. *Curr. Opin. Chem. Biol.*

(90) Coelho, P. S., Brustad, E. M., Kannan, A., et Arnold, F. H. (2013) Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science 339*, 307-10.

(91) Wang, Z. J., Peck, N. E., Renata, H., et Arnold, F. H. (2014) Cytochrome P450-catalyzed insertion of carbenoids into N-H bonds. *Chem. Sci. 5*, 598-601.

(92) Coelho, P. S., Wang, Z. J., Ener, M. E., Baril, S. a, Kannan, A., Arnold, F. H., et Brustad, E. M. (2013) A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins in vivo. *Nat. Chem. Biol. 9*, 485-7.

(93) Farwell, C. C., McIntosh, J. A., Hyster, T. K., Wang, Z. J., et Arnold, F. H. (2014) Enantioselective imidation of sulfides via enzyme-catalyzed intermolecular nitrogen-atom transfer. *J. Am. Chem. Soc. 136*, 8766-8771.

(94) McIntosh, J. A., Coelho, P. S., Farwell, C. C., Wang, Z. J., Lewis, J. C., Brown, T. R., et Arnold, F. H. (2013) Enantioselective intramolecular C-H amination catalyzed by engineered cytochrome P450 enzymes in vitro and in vivo. *Angew. Chemie - Int. Ed. 52*, 9309-9312.

(95) Ozaki, S. I., Matsui, T., et Watanabe, Y. (1996) Conversion of myoglobin into a highly stereospecific peroxygenase by the L29H/H64L mutation. *J. Am. Chem. Soc. 118*, 9784-9785.

(96) Matsui, T., Ozaki, S.-I., Liong, E., Phillips, G. N., et Watanabe, Y. (1999) Effects of the Location of Distal Histidine in the Reaction of Myoglobin with Hydrogen Peroxide. *J. Biol. Chem. 274*, 2838-2844.

(97) Matsui, T., Ozaki, S. I., et Watanabe, Y. (1999) Formation and catalytic roles of compound I in the hydrogen peroxide- dependent oxidations by His64 myoglobin mutants. *J. Am. Chem. Soc. 121*, 9952-9957.

(98) Hayashi, T., Hitomi, Y., Ando, T., Mizutani, T., Hisaeda, Y., Kitagawa, S., et Ogoshi, H. (1999) Peroxidase activity of myoglobin is enhanced by chemical mutation of heme-propionates. *J. Am. Chem. Soc. 121*, 7747-7750.

(99) Sato, H., Hayashi, T., Ando, T., Hisaeda, Y., Ueno, T., et Watanabe, Y. (2004) Hybridization of Modified-Heme Reconstitution and Distal Histidine Mutation to Functionalize Sperm Whale Myoglobin. *J. Am. Chem. Soc. 126*, 436-437.

(100) Ohashi, M., Koshiyama, T., Ueno, T., Yanase, M., Fujii, H., et Watanabe, Y. (2003) Preparation of artificial metalloenzymes by insertion of chromium(III) Schiff base complexes into apomyoglobin mutants. *Angew. Chemie - Int. Ed. 42*, 1005-1008.

(101) Lai, T.-S., Chan, F.-Y., So, P.-K., Ma, D.-L., Wong, K.-Y., et Che, C.-M. (2006) Alkene cyclopropanation catalyzed by Halterman iron porphyrin: participation of organic bases as axial ligands. *Dalton Trans.* 4845-51.

(102) Intrieri, D., Carminati, D., et Gallo, E. (2016) The Ligand Influence in Stereoselective Carbene Transfer Reactions Promoted by Chiral Metal Porphyrin Catalysts. *Dalt. Trans.*

(103) Bordeaux, M., Tyagi, V., et Fasan, R. (2015) Highly diastereoselective and enantioselective olefin cyclopropanation using engineered myoglobin-based catalysts. *Angew. Chemie - Int. Ed. 54*, 1744-1748.

(104) Sreenilayam, G., et Fasan, R. (2015) Myoglobin-catalyzed intermolecular carbene N-H insertion with arylamine substrates. *Chem. Commun. 51*, 1532-4.

(105) Tyagi, V., Bonn, R. B., et Fasan, R. (2015) Intermolecular carbene S–H insertion catalysed by engineered myoglobin-based catalysts. *Chem. Sci. 6*, 2488-2494.

(106) Key, H. M., Dydio, P., Clark, D. S., et Hartwig, J. F. (2016) Abiological catalysis by artificial haem proteins containing noble metals in place of iron. *Nature 534*.

(107) Komatsu, T., Ishihara, S., Tsuchida, E., Nishide, H., Morokuma, C., et Nakamura, S. (2005) Heat-resistant oxygen-carrying hemoproteins consist of recombinant xylanases and synthetic iron(II) porphyrin. *Biomacromolecules 6*, 1489-1494.

(108) Ricoux, R., Dubuc, R., Dupont, C., Marechal, J., Martin, A., Sellier, M., et Mahy, J.-P. (2008) Hemozymes Peroxidase Activity Of Artificial Hemoproteins Constructed From the Streptomyces lividans Xylanase A and Iron(III)-Carboxy-Substituted Porphyrins. *Bioconjug. Chem. 19*, 899-910.

(109) Ricoux, R., Allard, M., Dubuc, R., Dupont, C., Maréchal, J.-D., et Mahy, J.-P. (2009) Selective oxidation of aromatic sulfide catalyzed by an artificial metalloenzyme: new activity of hemozymes. *Org. Biomol. Chem. 7*, 3208-11.

(110) Allard, M., Dupont, C., Muñoz Robles, V., Doucet, N., Lledós, A., Maréchal, J.-D., Urvoas, A., Mahy, J.-P., et Ricoux, R. (2012) Incorporation of manganese complexes into

xylanase: new artificial metalloenzymes for enantioselective epoxidation. *Chembiochem 13*, 240-51.

(111) Mahammed, A., et Gross, Z. (2005) Albumin-conjugated corrole metal complexes: Extremely simple yet very efficient biomimetic oxidation systems. *J. Am. Chem. Soc. 127*, 2883-2887.

(112) Reetz, M. T., et Jiao, N. (2006) Copper-phthalocyanine conjugates of serum albumins as enantioselective catalysts in Diels-Alder reactions. *Angew. Chemie - Int. Ed. 45*, 2416-2419.

(113) Rousselot-Pailley, P., Bochot, C., Marchi-Delapierre, C., Jorge-Robin, A., Martin, L., Fontecilla-Camps, J. C., Cavazza, C., et Ménage, S. (2009) The protein environment drives selectivity for sulfide oxidation by an artificial metalloenzyme. *ChemBioChem 10*, 545-552.

(114) Tang, J., Huang, F., Wei, Y., Bian, H., Zhang, W., et Liang, H. (2016) Bovine serum albumin–cobalt(II) Schiff base complex hybrid: an efficient artificial metalloenzyme for enantioselective sulfoxidation using hydrogen peroxide. *Dalt. Trans. 45*.

(115) Reetz, M. T. (2012) Artificial metalloenzymes as catalysts in stereoselective Diels-Alder reactions. *Chem. Rec. 12*, 391-406.

(116) Fernández, L., Jiao, N., Soni, P., Gumulya, Y., de Oliveira, L. G., et Reetz, M. T. (2010) An efficient method for mutant library creation in Pichia pastoris useful in directed evolution. *Biocatal. Biotransformation 28*, 122-129.

(117) Bos, J., Browne, W. R., Driessen, A. J. M., et Roelfes, G. (2015) Supramolecular Assembly of Artificial Metalloenzymes Based on the Dimeric Protein LmrR as Promiscuous Scaffold. *J. Am. Chem. Soc. 21*, 150804112052007.

(118) de Lauzon, S., Desfosses, B., Mansuy, D., et Mahy, J. (1999) Studies of the reactivity of artificial peroxidase-like hemoproteins based on antibodies elicited against a specifically designed ortho -carboxy substituted tetraarylporphyrin. *FEBS Lett. 443*, 229-234.

(119) Kawamura-Konishi, Y., Asano, A., Yamazaki, M., Tashiro, H., et Suzuki, H. (1998) Peroxidase activity of an antibody-ferric porphyrin complex. *J. Mol. Catal. - B Enzym. 4*, 181-190.

(120) De Lauzon, S., Quilez, R., Lion, L., Desfosses, B., Desfosses, B., Lee, I., Sari, M. A., Benkovic, S. J., Mansuy, D., et Mahy, J. P. (1998) Active site topology of artificial peroxidase-like hemoproteins based on antibodies constructed from a specifically designed ortho-carboxy-substituted tetraarylporphyrin. *Eur. J. Biochem. 257*, 121-130.

(121) Harada, A., Fukushima, H., Shiotsuki, K., Yamaguchi, H., Oka, F., et Kamachi, M. (1997) Peroxidation of Pyrogallol by Antibody-Metalloporphyrin Complexes. *Inorg. Chem. 36*, 6099-6102.

(122) Quilez, R., De Lauzon, S., Desfosses, B., Mansuy, D., et Mahy, J. P. (1996) Artificial peroxidase-like hemoproteins based on antibodies constructed from a specifically designed ortho-carboxy substituted tetraarylporphyrin hapten and exhibiting a high affinity for iron-porphyrins. *FEBS Lett. 395*, 73-76.

(123) FENG, Y., LIU, Z., GAO, G., GAO, S. J., LIU, X. Y., et YANG, T. S. (1995) Study of the Abzyme with Peroxidase Catalytic Activity. *Ann. N. Y. Acad. Sci. 750*, 271-276.

(124) Takagi, M., Kohda, K., Hamuro, T., Harada, A., Yamaguchi, H., Kamachi, M., et Imanaka, T. (1995) Thermostable peroxidase activity with a recombinant antibody L chain-porphyrin Fe(III) complex. *FEBS Lett. 375*, 273-276.

(125) Savitsky, A. P., Nelen, M. I., Yatsmirsky, A. K., Demcheva, M. V., Ponomarev, G. V., et Sinikov, I. V. (1994) Kinetics of oxidation of o-dianisidine by hydrogen peroxide in the presence of antibody complexes of iron(III) coproporphyrin. *Appl. Biochem. Biotechnol. 47*, 317-327.

(126) Cochran, A. G., et Schultz, P. G. (1990) Peroxidase Activity of an Antibody-Heme Complex A New General Synthesis for Polylithium Organic. *J Am Chem Soc 112*, 9414-9415.

(127) Keinan, E., Sinha, S. C., Sinha-Bagchi, A., Benory, E., Ghozi, M. C., Eshhar, Z., et Green, B. S. (1990) Towards antibody-mediated metallo-porphyrin chemistry. *Pure Appl. Chem. 62*, 2013-2019.

(128) Nimri, S., et Keinan, E. (1999) Antibody-metalloporphyrin catalytic assembly mimics natural oxidation enzymes. *J. Am. Chem. Soc. 121*, 8978-8982.

(129) Baldwin, D. A., Marques, H. M., et Pratt, J. M. (1987) Hemes and Hemoproteins. 5: Kinetics of the Peroxidatic Activity of Microperoxidase-8: Model for the Peroxidase Enzymes. *J. Inorg. Biochem. 30*, 203-217.

(130) Adams, P. (1990) The peroxidasic activity of the haem octapeptide microperoxidase-8 (MP-8): the kinetic mechanism of the catalytic reduction of H2O2 by MP-8 using 2, 2'-azinobis-(3-. *J. Chem. Soc., Perkin Trans. 2 2*, 1407-1414.

(131) Ricoux, R., Boucher, J. L., Mansuy, D., et Mahy, J. P. (2001) Microperoxidase 8 catalyzed nitration of phenol by nitrogen dioxide radicals. *Eur. J. Biochem. 268*, 3783-3788.

(132) Colonna, S., Gaggereo, N., Carrea, G., et Pasta, P. (1994) The microperoxidase-11 catalyzed oxidation of sulfides is enantioselective. *Tetrahedron Lett. 35*, 9103-9104.

(133) Ricoux, R., Sauriat-Dorizon, H., Girgenti, E., Blanchard, D., et Mahy, J. P. (2002) Hemoabzymes: Towards new biocatalysts for selective oxidations. *J. Immunol. Methods 269*, 39-57.

(134) Wilson, M. E., et Whitesides, G. M. (1978) Conversion of a protein to a homogeneous asymmetric hydrogenation catalyst by site-specific modification with a diphosphinerhodium(I) moiety. *J. Am. Chem. Soc. 100*, 306-307.

(135) Klein, G., Humbert, N., Gradinaru, J., Ivanova, A., Gilardoni, F., Rusbandi, U. E., et Ward, T. R. (2005) Tailoring the active site of chemzymes by using a chemogenetic-optimization procedure: Towards substrate-specific artificial hydrogenases based on the biotin-avidin technology. *Angew. Chemie - Int. Ed. 44*, 7764-7767.

(136) Skander, M., Humbert, N., Gradinaru, J., Loosli, A., Zocchi, A., Ward, T. R., et Neucha, C.-. (2004) Artificial Metalloenzymes : (Strept) avidin as Host for Enantioselective Hydrogenation by Achiral Biotinylated Rhodium - Diphosphine Complexes by the 2001 Nobel Prize in Chemistry , which was shared by. *J Am Chem Soc 126*, 14411-14418.

(137) Collot, J., Gradinaru, J., Humbert, N., Skander, M., Zocchi, A., et Ward, T. R. (2003) Artificial metalloenzymes for enantioselective catalysis based on biotin-avidin. *J. Am. Chem. Soc. 125*, 9030-1.

(138) Reetz, M. T., Peyralans, J. J.-P., Maichele, A., Fu, Y., et Maywald, M. (2006) Directed evolution of hybrid enzymes: Evolving enantioselectivity of an achiral Rh-complex anchored to a protein. *Chem. Commun. (Camb).* 4318-4320.

(139) Letondor, C., Pordea, A., Humbert, N., Ivanova, A., Mazurek, S., Novic, M., et Ward, T. R. (2006) Artificial transfer hydrogenases based on the biotin-(strept)avidin technology: Fine tuning the selectivity by saturation mutagenesis of the host protein. *J. Am. Chem. Soc. 128*, 8320-8328.

(140) Creus, M., Pordea, A., Rossel, T., Sardo, A., Letondor, C., Ivanova, A., LeTrong, I., Stenkamp, R. E., et Ward, T. R. (2008) X-ray structure and designed evolution of an artificial transfer hydrogenase. *Angew. Chemie - Int. Ed. 47*, 1400-1404.

(141) Dürrenberger, M., Heinisch, T., Wilson, Y. M., Rossel, T., Nogueira, E., Knörr, L., Mutschler, A., Kersten, K., Zimbron, M. J., Pierron, J., Schirmer, T., et Ward, T. R. (2011) Artificial transfer hydrogenases for the enantioselective reduction of cyclic imines. *Angew. Chemie - Int. Ed. 50*, 3026-3029.

(142) Köhler, V., Wilson, Y. M., Dürrenberger, M., Ghislieri, D., Churakova, E., Quinto, T., Knörr, L., Häussinger, D., Hollmann, F., Turner, N. J., et Ward, T. R. (2013) Synthetic cascades are enabled by combining biocatalysts with artificial metalloenzymes. *Nat. Chem. 5*, 93-99.

(143) Chatterjee, A., Mallin, H., Klehr, J., Vallapurackal, J., Finke, A., Vera, L., Marsh, M., et Ward, T. R. (2016) An Enantioselective Artificial Suzukiase Based on the Biotin-Streptavidin Technology. *Chem. Sci. 7*, 673-677.

(144) Jeschek, M., Panke, S., et Ward, T. R. (2016) Periplasmic Screening for Artificial Metalloenzymes, in *Methods in enzymology*.

(145) Jeschek, M., Reuter, R., Heinisch, T., Trindler, C., Klehr, J., Panke, S., et Ward, T. R. (2016) Directed evolution of artificial metalloenzymes for in vivo metathesis. *Nature advance on*.

(146) Wilson, Y. M., Dürrenberger, M., Nogueira, E. S., et Ward, T. R. (2014) Neutralizing the detrimental effect of glutathione on precious metal catalysts. *J. Am. Chem. Soc. 136*, 8928-8932.

(147) Raffy, Q., Ricoux, R., et Mahy, J. P. (2008) Synthesis of a new estradiol-iron metalloporphyrin conjugate used to build up a new hybrid biocatalyst for selective oxidations by the « Trojan horse » strategy. *Tetrahedron Lett. 49*, 1865-1869.

(148) Sansiaume, E., Ricoux, R., Gori, D., et Mahy, J. P. (2010) Oxidation of organic molecules in homogeneous aqueous solution catalyzed by hybrid biocatalysts (based on the Trojan Horse strategy). *Tetrahedron Asymmetry 21*, 1593-1600.

(149) Heyd, B., Pecorari, F., Collinet, B., Adjadj, E., Desmadril, M., et Minard, P. (2003) In vitro evolution of the binding specificity of neocarzinostatin, an enediyne-binding chromoprotein. *Biochemistry 42*, 5674-5683.

(150) Drevelle, A., Graille, M., Heyd, B., Sorel, I., Ulryck, N., Pecorari, F., Desmadril, M., van Tilbeurgh, H., et Minard, P. (2006) Structures of in Vitro Evolved Binding Sites on Neocarzinostatin Scaffold Reveal Unanticipated Evolutionary Pathways. *J. Mol. Biol. 358*, 455-471.

(151) Sansiaume-Dagousset, E., Urvoas, A., Chelly, K., Ghattas, W., Maréchal, J.-D., Mahy, J.-P., et Ricoux, R. (2014) Neocarzinostatin-based hybrid biocatalysts for oxidation reactions. *Dalton Trans. 43*, 8344-54.

(152) Urvoas, A., Ghattas, W., Maréchal, J. D., Avenier, F., Bellande, F., Mao, W., Ricoux, R., et Mahy, J. P. (2014) Neocarzinostatin-based hybrid biocatalysts with a RNase like activity. *Bioorganic Med. Chem. 22*, 5678-5686.

(153) Ghattas, W., Cotchico-Alonso, L., Maréchal, J.-D., Urvoas, A., Rousseau, M., Mahy, J.-P., et Ricoux, R. (2016) Artificial Metalloenzymes with the Neocarzinostatin Scaffold: Toward a Biocatalyst for the Diels–Alder Reaction. *ChemBioChem 17*, 433-40.

(154) Chevalley, A., et Salmain, M. (2012) Enantioselective transfer hydrogenation of ketone catalysed by artificial metalloenzymes derived from bovine β-lactoglobulin. *Chem. Commun. (Camb). 48*, 11984-6.

(155) Kaiser, E. T., et Lawrence, D. S. (1984) Chemical mutation of enzyme active sites. *Science (80-.). 226*, 505-511.

(156) Reetz, M. T., Rentzsch, M., Pletsch, A., et Maywald, M. (2002) Towards the Directed Evolution of Hybrid Catalysts. *Chim. Int. J. Chem. 56*, 721-723.

(157) Panella, L., Broos, J., Jin, J., Fraaije, M. W., Janssen, D. B., Jeronimus-Stratingh, M., Feringa, B. L., Minnaard, A. J., et de Vries, J. G. (2005) Merging homogeneous catalysis with biocatalysis; papain as hydrogenation catalyst. *Chem Commun 23*, 5656-5658.

(158) Talbi, B., Haquette, P., Martel, A., de Montigny, F., Fosse, C., Cordier, S., Roisnel, T., Jaouen, G., et Salmain, M. (2010) (η6-Arene) ruthenium(ii) complexes and metallo-papain hybrid as Lewis acid catalysts of Diels–Alder reaction in water. *Dalt. Trans. 39*, 5605.

(159) Carey, J. R., Ma, S. K., Pfister, T. D., Garne, D. K., Kim, H. K., Abramite, J. a., Wang, Z., Guo, Z., et Lu, Y. (2004) A site-selective dual anchoring strategy for artificial metalloprotein design. *J. Am. Chem. Soc. 126*, 10812-10813.

(160) Zhang, J.-L., Garner, D. K., Liang, L., Chen, Q., et Lu, Y. (2008) Protein scaffold of a designed metalloenzyme enhances the chemoselectivity in sulfoxidation of thioanisole. *Chem. Commun. (Camb).* 1665-1667.

(161) Laan, W., Mu??oz, B. K., Den Heeten, R., et Kamer, P. C. J. (2010) Artificial metalloenzymes through cysteine-selective conjugation of phosphines to photoactive yellow protein. *ChemBioChem 11*, 1236-1239.

(162) Bos, J., Fusetti, F., Driessen, A. J. M., et Roelfes, G. (2012) Enantioselective artificial metalloenzymes by creation of a novel active site at the protein dimer interface. *Angew. Chem. Int. Ed. Engl. 51*, 7472-5.

(163) Bos, J., García-Herraiz, A., et Roelfes, G. (2013) An enantioselective artificial metallo-hydratase. *Chem. Sci. 4*, 3578.

(164) Agarkov, A., Greenfield, S., Xie, D., Pawlick, R., Starkey, G., et Gilbertson, S. R. (2006) Synthesis of phosphine containing amino acids: Utilization of peptide synthesis in ligand design. *Biopolymers 84*, 48-73.

(165) Wang, Q., Parrish, A. R., et Wang, L. (2009) Expanding the genetic code for biological studies. *Chem. Biol. 16*, 323-36.

(166) Kobayashi, T., Nureki, O., Ishitani, R., Yaremchuk, A., Tukalo, M., Cusack, S., Sakamoto, K., et Yokoyama, S. (2003) Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nat. Struct. Biol. 10*, 425-32.

(167) Polycarpo, C., Ambrogelly, A., Bérubé, A., Winbush, S. M., McCloskey, J. A., Crain, P. F., Wood, J. L., et Söll, D. (2004) An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *Proc. Natl. Acad. Sci. U. S. A. 101*, 12450-4.

(168) Hu, C., et Wang, J. (2016) Method for Enzyme Design with Genetically Encoded Unnatural Amino Acids, in *Methods in Enzymology*, p 109-133.

(169) Liu, C. C., et Schultz, P. G. (2010) Adding new chemistries to the genetic code. *Annu. Rev. Biochem. 79*, 413.

(170) Yang, H., Srivastava, P., Zhang, C., et Lewis, J. C. (2014) A general method for artificial metalloenzyme formation through strain-promoted azide-alkyne cycloaddition. *Chembiochem 15*, 223-7.

(171) Fekner, T., Li, X., Lee, M. M., et Chan, M. K. (2009) A pyrrolysine analogue for protein click chemistry. *Angew. Chemie - Int. Ed. 48*, 1633-1635.

(172) Nguyen, D. P., Elliott, T., Holt, M., Muir, T. W., et Chin, J. W. (2011) Genetically encoded 1,2-aminothiols facilitate rapid and site-specific protein labeling via a bio-orthogonal cyanobenzothiazole condensation. *J. Am. Chem. Soc. 133*, 11418-11421.

(173) Hu, C., Chan, S. I., Sawyer, E. B., Yu, Y., et Wang, J. (2014) Metalloprotein design using genetic code expansion. *Chem. Soc. Rev. 43*.

(174) Srivastava, P., Yang, H., Ellis-Guardiola, K., et Lewis, J. C. (2015) Engineering a dirhodium artificial metalloenzyme for selective olefin cyclopropanation. *Nat. Commun. 6*, 7789.

(175) Drienovská, I., Rioz-Martínez, A., Draksharapu, A., et Roelfes, G. (2015) Novel artificial metalloenzymes by in vivo incorporation of metal-binding unnatural amino acids. *Chem. Sci. 6*, 770-776.

(176) Cooper, M. D., et Alder, M. N. (2006) The evolution of adaptive immune systems. *Cell 124*, 815-822.

(177) Kajava, A. V. (2012) Tandem repeats in proteins: From sequence to structure. *J. Struct. Biol. 179*, 279-288.

(178) Wetzel, S. K., Settanni, G., Kenig, M., Binz, H. K., et Plückthun, A. (2008) Folding and Unfolding Mechanism of Highly Stable Full-Consensus Ankyrin Repeat Proteins. *J. Mol. Biol. 376*, 241-257.

(179) Boersma, Y. L., et Plückthun, A. (2011) DARPins and other repeat protein scaffolds: Advances in engineering and applications. *Curr. Opin. Biotechnol. 22*, 849-857.

(180) Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P., et Plückthun, A. (2003) Designing Repeat Proteins: Well-expressed, Soluble and Stable Proteins from Combinatorial Libraries of Consensus Ankyrin Repeat Proteins. *J. Mol. Biol. 332*, 489-503.

(181) Main, E. R. G., Xiong, Y., Cocco, M. J., D'Andrea, L., et Regan, L. (2003) Design of stable α-helical arrays from an idealized TPR motif. *Structure 11*, 497-508.

(182) Lee, S.-C. S.-G., Park, K., Han, J., Lee, J. -j., Kim, H.-S. J., Hong, S., Heu, W., Kim, Y. J., Ha, J.-S., Lee, S.-C. S.-G., Cheong, H.-K., Jeon, Y. H., Kim, D., et Kim, H.-S. J. (2012) Design of a binding scaffold based on variable lymphocyte receptors of jawless vertebrates by module engineering. *Proc. Natl. Acad. Sci. 109*, 3299-3304.

(183) Varadamsetty, G., Tremmel, D., Hansen, S., Parmeggiani, F., et Plückthun, A. (2012) Designed armadillo repeat proteins: Library generation, characterization and selection of peptide binders with high specificity. *J. Mol. Biol. 424*, 68-87.

(184) Urvoas, A., Guellouz, A., Valerio-Lepiniec, M., Graille, M., Durand, D., Desravines, D. C., van Tilbeurgh, H., Desmadril, M., et Minard, P. (2010) Design, production and molecular structure of a new family of artificial alpha-helicoidal repeat proteins (αRep) based on thermostable HEAT-like repeats. *J. Mol. Biol. 404*, 307-27.

(185) Julien, O., Gignac, I., Hutton, A., Yee, A., Arrowsmith, C. H., et Gagné, S. M. (2006) MTH187 from Methanobacterium thermoautotrophicum has three HEAT-like repeats. *J. Biomol. NMR 35*, 149-154.

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

(187) Figueroa, M., Sleutel, M., Vandevenne, M., Parvizi, G., Attout, S., Jacquin, O., Vandenameele, J., Fischer, A. W., Damblon, C., Goormaghtigh, E., Valerio-Lepiniec, M., Urvoas, A., Durand, D., Pardon, E., Steyaert, J., Minard, P., Maes, D., Meiler, J., Matagne,

A., Martial, J. A., et Van de Weerdt, C. (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.

(188) Gurunatha, K. L., Fournier, A. C., Urvoas, A., Valerio-Lepiniec, M., Marchi, V., Minard, P., et Dujardin, E. (2016) Nanoparticles Self-Assembly Driven by High Affinity Repeat Protein Pairing. *ACS Nano 10*, 3176-3185.

(189) Park, S., Ikehata, K., Watabe, R., Hidaka, Y., Rajendran, A., et Sugiyama, H. (2012) Deciphering DNA-based asymmetric catalysis through intramolecular Friedel–Crafts alkylations. *Chem. Commun. 48*, 10398.

(190) Wang, Y., Zhou, C., et Wang, R. (2015) Copper-catalyzed hydroxylation of aryl halides: efficient synthesis of phenols, alkyl aryl ethers and benzofuran derivatives in neat water. *Green Chem. 17*, 3910-3915.

(191) Zhang, C., Srivastava, P., Ellis-Guardiola, K., et Lewis, J. C. (2014) Manganese terpyridine artificial metalloenzymes for benzylic oxygenation and olefin epoxidation. *Tetrahedron 70*, 4245-4249.

(192) Meunier, B. (1992) Metalloporphyrins as versatile catalysts for oxidation reactions and oxidative DNA cleavage. *Chem. Rev. 92*, 1411-1456.

(193) Funel, J.-A., et Abele, S. (2013) Industrial applications of the Diels-Alder reaction. *Angew. Chem. Int. Ed. Engl. 52*, 3822-63.

(194) Gawley, R. E. (2006) Do the terms « % ee » and « % de » make sense as expressions of stereoisomer composition or stereoselectivity? *J. Org. Chem. 71*, 2411-2416.

(195) Kagan, H. B. (2010) Is there a preferred expression for the composition of a mixture of enantiomers? *Recl. des Trav. Chim. des Pays-Bas 114*, 203-205.

(196) Faber, K. (2011) Biotransformations in Organic Chemistry. Springer Berlin Heidelberg, Berlin, Heidelberg.

(197) Faber, K., et Kroutil, W. (2012) SELECTIVITY : A Computer Program for the Determination of the Enantioselectivity (E-Value) in the Kinetic Resolution of Enantiomers 5-7.

Partie B - Résultats

Chapitre I - Utilisation du dimère αRep-A3

I.1 Etude structurale du complexe αRep-A3 et choix des positions

I.1.1 Analyse structurale

Le premier objectif était la détermination des positions de la crevasse propices au greffage covalent d'un ligand de type phénantroline. L'analyse des deux structures cristallographiques de la protéine αRep-A3 non modifiée, dite « sauvage », nous a permis de proposer des positions appropriées. Il est en effet nécessaire que la fonctionnalisation de la protéine n'altère pas sa capacité à dimériser, sans quoi le métal ne serait plus suffisamment enfoui.

Les deux structures cristallographiques se différentient majoritairement par la présence d'une molécule de polyéthylène glycol (PEG) au sein de la crevasse. La forme générale du complexe (figure 84) est comparable à un coquillage bivalve), présentant une ouverture d'un côté (l'entrée de la crevasse) et un fond hydrophobe constitué majoritairement de tryptophanes, phénylalanines et tyrosines formant une sorte de ceinture hydrophobe (figure88E). Cette molécule de PEG issue de la solution de cristallisation provoque un basculement des deux monomères, conduisant à une légère asymétrie du dimère et une ouverture de la crevasse.

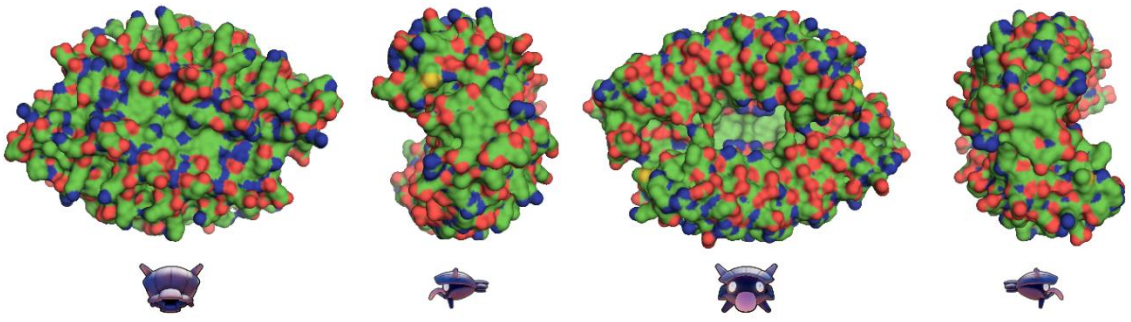


Figure 84: Différentes vues du dimère αRep-A3 dans la forme ouverte (la molécule de PEG a été retirée). Le coquillage illustre l'orientation de la protéine. (PDB : 3LTM) En jaune est visible la méthionine présente dans le module N-terminal.

La forme sans PEG présente, elle, une symétrie axiale d'ordre 2 (C2) où les deux monomères se font face. Elle a également une crevasse plus refermée, ce qui ouvre la ceinture hydrophobe et forme donc un pore (figure 85).

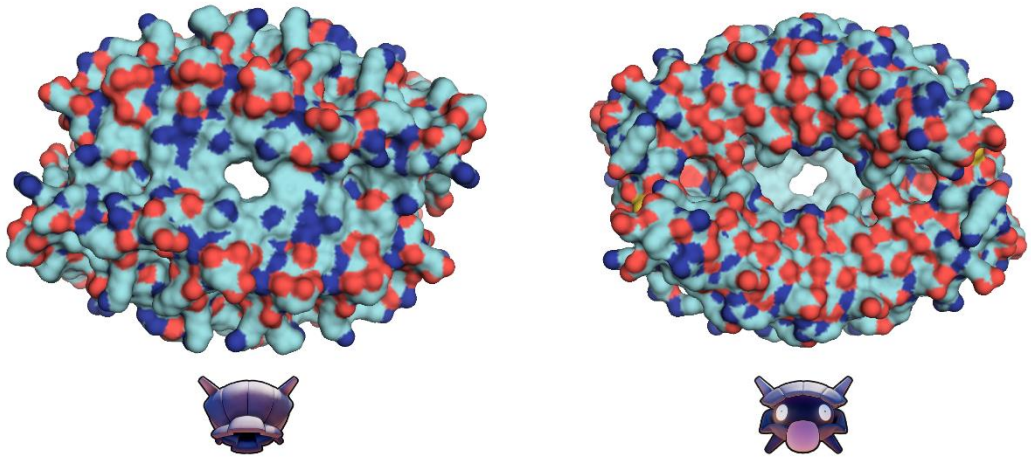


Figure 85 : Structure cristallographique de la protéine αRep-A3 dans sa forme fermée, on peut voir l'ouverture du pore provoquée par l'éloignement des acides aminés composant la ceinture hydrophobe. En contrepartie, la surface d'interaction entre les modules terminaux (caps) est agrandie. (PDB : 3LTJ)

Ces différences dans les formes ouvertes et fermées nous donnent un indice sur la flexibilité du dimère, qui semble s'ouvrir et se fermer par glissement. Il semblerait que ces deux structures représentent deux formes extrêmes de la configuration ouverte et fermée du dimère. J'ai donc tenté de déterminer les acides aminés importants pour la dimérisation dans ces deux formes grâce au serveur Robetta [1].

I.1.2 Analyse par le serveur Robetta

Le serveur Robetta est une mise en service en libre accès de l'algorithme de prédiction structurale Rosetta conçu dans le laboratoire de David Baker.[2] Les points chauds d'interaction sont calculés grâce à un scan d'alanine en corps rigide. La variation d'enthalpie libre (ΔG) associée à l'interaction entre les deux monomères est calculée pour les différents variants où les acides aminés sont changés en alanine les uns après les autres. La différence d'énergie d'interaction entre le dimère muté et le dimère sauvage est calculée (ΔΔG), donnant un indice sur l'importance du résidu en question pour la dimérisation. La modélisation en corps rigide se prête particulièrement bien aux αRep, l'ossature protéique n'étant pas affectée dans toutes les structures connues.

Il apparait que les acides aminés de la ceinture hydrophobe sont effectivement pointés comme responsables de plus grande énergie d'interaction ((ΔΔG) supérieur à 1 kcal/mol). La figure 3 résume les résidus en jeu. Sur cet alignement est représentée la séquence d'un monomère de la protéine αRep-A3, les modules étant alignés selon la séquence canonique de 31 acides aminés pour chacun des motifs répétés. En vert sont représentés les acides aminés variables dans la banque αRep. Les résidus encadrés sont ceux présentant un ΔΔG supérieur à 1 kcal/mol dans les deux formes. Le résidu W111 souligné

est noté comme participant à la dimérisation uniquement dans la forme ouverte. Un dernier résidu semble intéressant, il s'agit de l'arginine en position 22 du N-cap. Cette dernière effectue des liaisons hydrogène avec l'aspartate en position 26 du 4e motif dans la forme fermée, alors qu'elle en est très éloignée dans la forme ouverte.

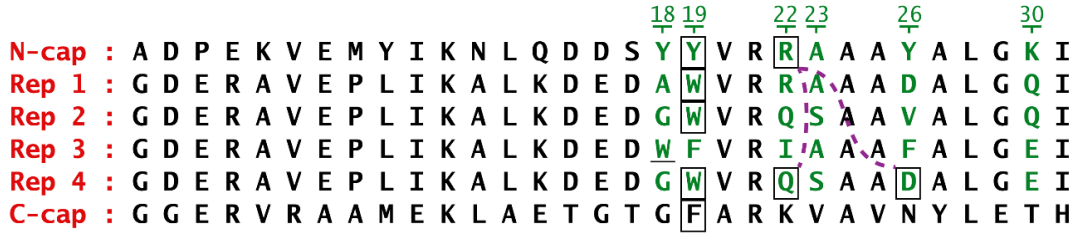


Figure 86 : Représentation de la séquence d'un monomère d'αRep-A3 par alignement des séquences des modules répétés. Les acides aminés hypervariables de la banque αRep sont en vert. Résultats de l'analyse Robetta : Les résidus encadrés semblent participer majoritairement à la formation du dimère dans les deux formes connues du dimère, le tryptophane souligné n'y participe que dans la forme ouverte. Le résidu R22 effectue une liaison hydrogène avec le résidu D150 de l'autre monomère uniquement dans la forme fermée, alors qu'elle est en interaction avec le résidu Q146 dans la forme ouverte.

Une analyse de l'interface grâce au serveur ePISA donne des résultats concordants : dans la forme fermée l'acide aminé R22 est présenté comme formant de fortes liaisons hydrogènes et ponts salins avec le résidu D150 (en dans une moindre mesure avec le résidu Q146), dans la forme ouverte, la liaison entre R22 et D150 disparait. Les surfaces d'interface sont : 1110 Å^2 pour la forme fermée et 1019 Å^2 pour la forme ouverte. Ceci est notamment dû au basculement des deux monomères, éloignant les modules terminaux (caps) se faisant face.

Ces analyses bio-informatiques semblent confirmer l'utilité de la ceinture hydrophobe dans la formation du dimère αRep-A3, et nous pousse donc à exclure les acides aminés en position 18 et 19 des candidats au greffage covalent d'un ligand. Il faut toutefois être conscient que cette approche n'est pas entièrement fiable. Les études menées par l'équipe de D. Baker sur cet algorithme ont donné un taux de réussite de 79% dans la prédiction de points chauds d'interaction.[1] Les positions candidates 22, 26 et 30 ont donc été évaluées principalement sur la base d'une analyse visuelle des structures grâce au logiciel PyMOL. La position 23 a été exclue car, du fait de son encombrement, seuls de petits acides aminés peuvent s'y accommoder.

I.1.3 Choix d'un complexe de petite taille : la phénanthroline.

La phénanthroline est un petit ligand bidenté grâce à ses deux atomes d'azote (figure 87). D'une part, elle chélate la plupart des métaux de transition, incluant le zinc et le cuivre. D'autre part, plusieurs études ont montré que des complexes phénantroline-métal peuvent catalyser des réactions d'intérêt biologique [3–6]. Sa petite taille comparée à celle des ligands de type porphyrine nous permet d'envisager le greffage d'une phénanthroline par monomère d'A3, et donc d'avoir deux phénanthrolines présentes au sein de la crevasse du dimère.

Il a fallu fonctionnaliser cette phénanthroline avec un bras réactif aux cystéines pour la création d'un lien covalent. Notre choix s'est porté sur une fonction bromoacétamide, pour son faible encombrement et sa bonne sélectivité envers les thiols pour des valeurs de pH légèrement basiques.

J'ai donc effectué la fonctionnalisation d'une 1,10-phénantholine-5-amine commerciale (Sigma-Aldrich) avec du bromoacétyl bromide selon la méthode décrite par Bos *et al* [7] ref pour récupérer le composé **2**. Le composé a été purifié par recristallisation, analysé par RMN ^{1}H, spectrométrie de masse ESI-ToF à haute résolution et la structure a été résolue par diffraction aux rayons X.

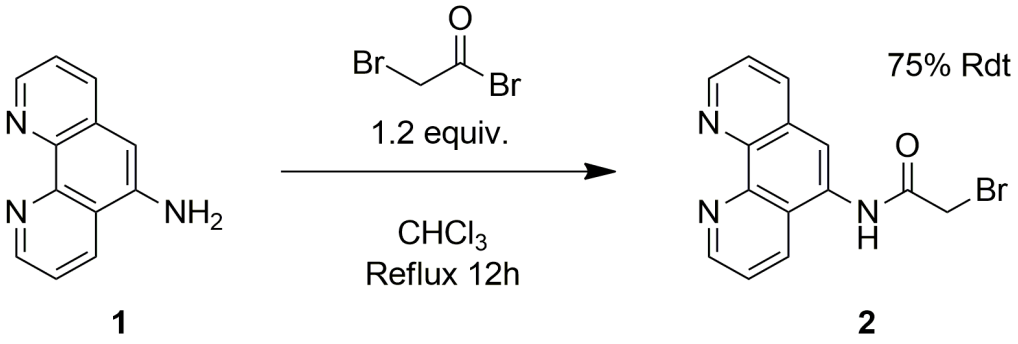


Figure 87 : Synthèse d'une phénanthroline fonctionnalisée avec un bras bromoacétamide réagissant avec les cystéines pour la formation d'un lien covalent.

C'est la structure cristallographique de la molécule **2** que j'ai utilisée pour construire un modèle à l'aide du logiciel PyMOL. En effet l'étude du comportement des complexes métalliques dans les champs de force conventionnels de modélisation moléculaire étant un domaine encore peu prédictif, nous avons fait le choix de nous reposer sur une simple inspection visuelle. Supposant que l'ajout du ligand phénanthroline au sein du dimère avait une forte probabilité de provoquer un basculement du dimère similaire au cas de l'insertion d'une molécule de PEG, c'est la structure de la forme ouverte qui fut utilisée. Les positions candidates ont été mutées *in silico* en cystéine, puis la chimie de couplage

avec le composé **2** a été mimée en tenant compte des distances et angles habituels de ces types de liaison.

J'ai retenu quatre positions intéressantes pour démarrer l'étude de la protéine αRep-A3, qui sont représentées sur la figure 88. Les positions Y26 (figure 88A) et F119 (figure 88C) correspondent toutes les deux à un acide aminé hypervariable en position 26 (figure 88F) et pointent en direction de la crevasse, la différence principale étant l'éloignement entre les deux points de fixation au sein du dimère. En effet, si l'on mesure la distance entre les carbones β des acides aminés se faisant face, on trouve dans la forme ouverte de la protéine une distance de 27,3 Å entre Y26 et Y26', et de 21,3 Å entre F119 et F119'. Cette distance est différente dans la forme fermée de la protéine : 28,9 Å et 18,3 Å respectivement. Ces positions nous permettront d'évaluer l'effet de la distance entre les deux centres métalliques sur la catalyse avec une potentielle création d'un catalyseur binucléaire.

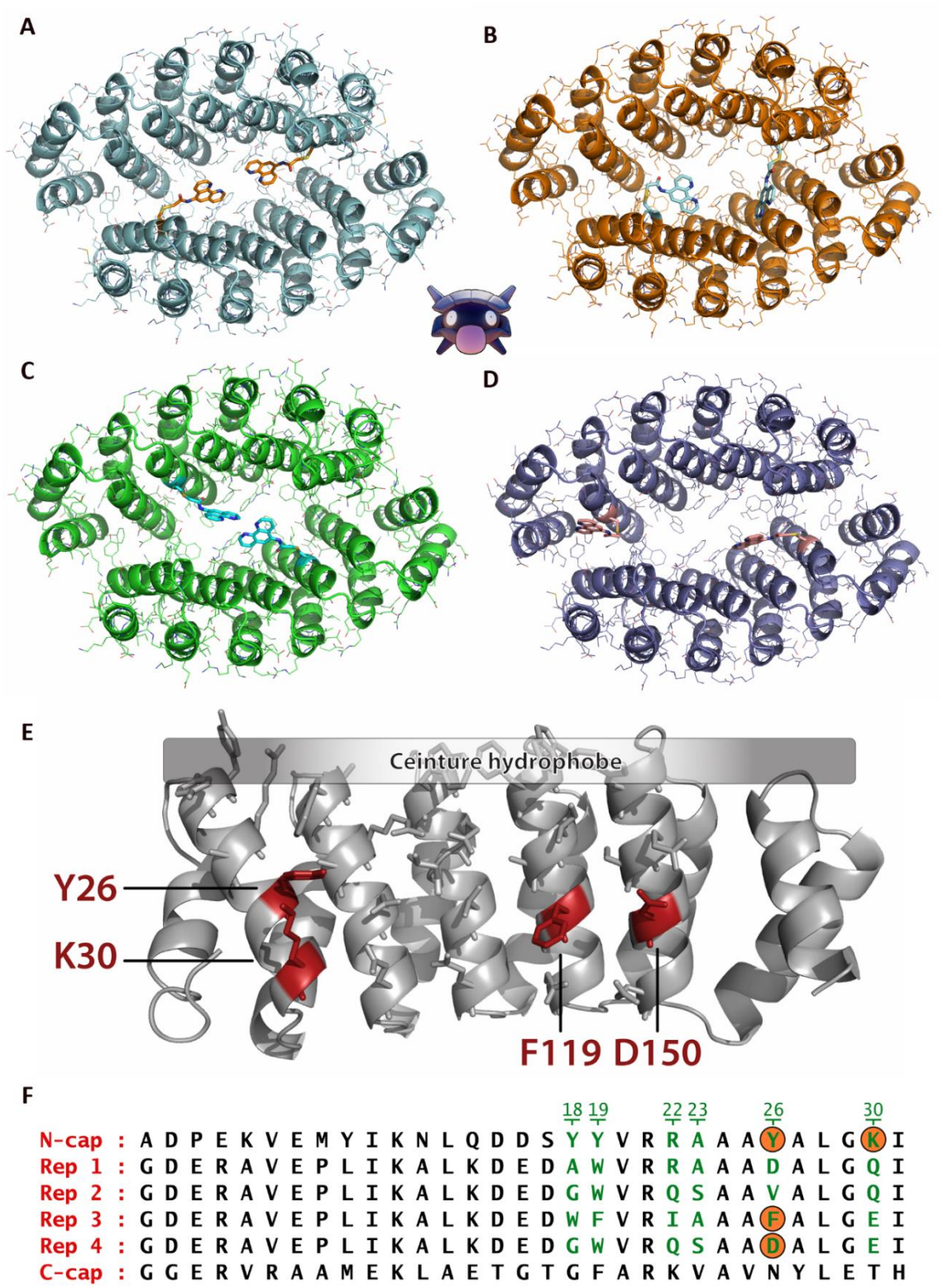


Figure 88 : A–D : Modèles obtenus après greffage manuel d'une phénanthroline aux positions Y26C (A), K30C (B), F119C (C) et D150C (D) à l'aide de PyMOL. E : Représentation des positions choisies sur un monomère de αRep-A3 vu de dessus. On remarque que la position K30 se situe au bord de l'ouverture de la crevasse tandis que les positions Y26, F119 et D150 se situent à des positions plus intermédiaires. F : Représentation des positions choisies sur l'alignement des modules. En vert sont les positions hypervariables de la banque, les points orange représentent les positions candidates au greffage de la phénanthroline après mutation en cystéine.

L'acide aminé D150 (figure 88D) est également à une position hypervariable n°26 et la distance entre les résidus D150 et D150' est relativement proche de celle entre Y26 et Y26'. Cette position a été choisie car la chaine latérale de l'acide aminé D150 est tournée à l'opposé de la crevasse centrale, vers une crevasse secondaire de faible volume. On peut alors se représenter deux orientations possibles de la phénanthroline : soit tournée vers la crevasse principale, soit tournée vers la crevasse secondaire. Cette orientation inédite de la phénanthroline peut se révéler utile si la crevasse principale est trop volumineuse pour apporter une sélectivité à la réaction. Les deux orientations ont été représentées dans la figure 88D.

Enfin, l'acide aminé K30 (figure 88B) est situé à l'entrée de la crevasse. Cette position est très accessible au solvant et peut être utilisée comme témoin d'accessibilité du substrat (distance entre carbones β de K30 et K30' : 22,8 Å dans la forme ouverte et 24,8 Å dans la forme fermée).

I.2 Obtention des biohybrides

I.2.1 Construction génétique et expression des variants

Le gène codant pour les protéines αRep étant constitué de séquences répétées de par sa conception (excepté aux positions hypervariables), cette forte homologie entre les modules pose un problème lors de l'hybridation des oligonucléotides dans des expériences de biologie moléculaire telles que la PCR. L'hybridation partielle d'oligonucléotides sur différents modules peut diminuer l'efficacité de l'amplification ainsi que de provoquer des mutations non désirées lors d'expériences de mutation dirigée (observations faites par les membres du laboratoire). Pour remédier à ce problème, il fut nécessaire de concevoir un gène synthétique non répétitif, codant pour la même protéine, tout en respectant au maximum l'usage codon d'*E.coli*. Pour ce faire, un outil d'optimisation génétique (genscript) a été utilisé : dans ce processus, chaque codon est remplacé aléatoirement par un codon synonyme en respectant les probabilités de l'usage codon d'*E.coli*. Après quelques ajustements, afin d'éviter les enchainements de codons rares ou les zones encore trop identiques (figure 89), l'utilitaire JCat m'a permis de calculer l'adéquation de notre gène synthétique avec l'organisme d'expression grâce à l'indice CAI (codon adaptation index)[8].

Figure 89 : Alignement des séquences de chacun des motifs des gène A3 (haut) et A3s (bas). Les différences par rapport au premier motif sont surlignées en rouge.

Le CAI du gène d'origine étant de 0,52 et celui du synthétique étant de 0,50, nous avons pu le faire synthétiser et de le nommer *A3sWT* : « A3 » pour la protéine codée, « s » pour synthétique et « WT » pour le gène « sauvage ». Pour toute la suite de notre étude, les variants seront nommés *A3sX*, X étant le nom de la position mutée (ex. *A3sF119C*). Suivant la nomenclature généralement utilisée, le gène sera écrit en italique (*A3sF119C*) tandis que la protéine portera le même nom en romain (A3sF119C). Cette protéine étant un dimère, il serait normalement nécessaire de considérer la concentration du dimère en solution. Dans un souci de simplicité cependant, j'indiquerais régulièrement la concentration en termes de monomère (ou concentration de cystéine libre).

Une séquence codant une étiquette hexahistidine (HisTag) a été placée en amont du gène pour une purification très efficace sur colonne d'affinité au nickel. Cependant, la capacité du HisTag à chélater fortement les ions métalliques peut s'avérer problématique pour la conception d'une métalloenzymes artificielle. J'ai donc inclus la séquence de clivage de la TEV protéase entre l'étiquette histidine et le gène codant pour la protéine A3s. Un StrepTag a été placé en aval du gène.

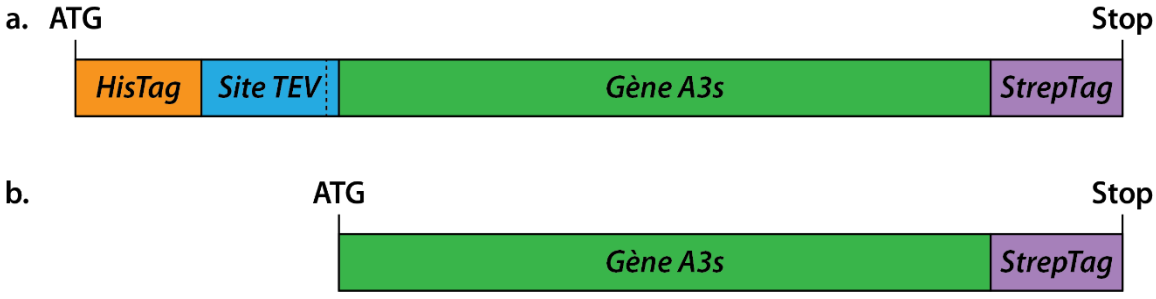


Figure 90 : Constructions génétiques utilisées pour la synthèse de la protéine A3s. a. Une séquence codant une étiquette HisTag suivie d'un site de clivage à la protéase TEV est inclue en amont du gène A3s afin de purifier la protéine en grande quantité. b. Dans cette construction, la séquence codant l'étiquette HisTag et le site TEV a été retirée, permettant une purification plus rapide bien que moins efficace.

Le gène synthétique a ensuite été muté grâce à la méthode QuikChange et transféré dans un plasmide d'expression de type pQE. La production dans la souche d'expression *E.coli* M15 permet d'atteindre une quantité moyenne 100 mg de protéine par litre de culture.

Après l'étape de clivage à la TEV protéase, c'est une quantité d'environ 70 mg qui peut être obtenue avec un litre de culture bactérienne. L'immunodétection en utilisant un anticorps dirigé contre le HisTag nous a permis d'observer l'absence de signal pour la protéine purifiée, indiquant qu'il avait été retiré avec succès. La bonne capacité de production de cette protéine et de ses variants sous forme soluble est encourageante car elle dénote une faible cytotoxicité et une faible propension à l'agrégation.

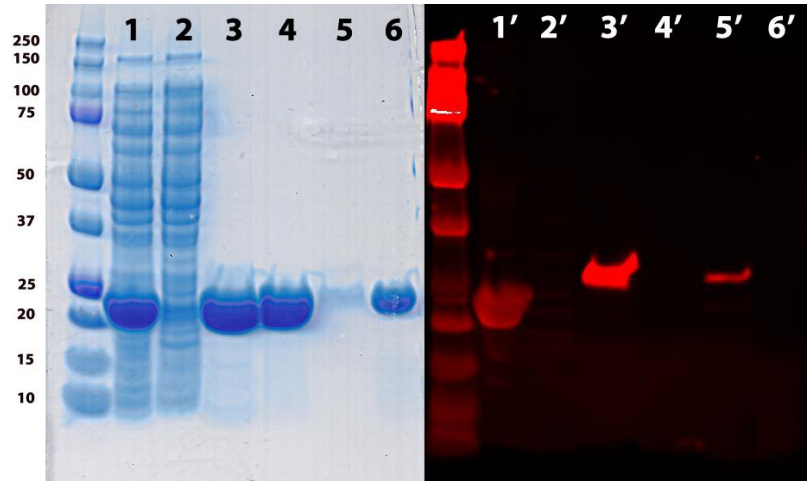


Figure 91 : Gel d'électrophorèse en conditions dénaturantes (gauche) et western blot (droite) révélé par un anticorps anti-HisTag récapitulant les étapes de purification des variants (ici pour le variant A3sF119C). (1, 1') Fraction soluble bactérienne. (2, 2') Fraction non retenue de la colonne d'affinité au nickel. (3, 3') Fraction éluée après la colonne d'affinité au nickel, puis incubée avec la protéase TEV (possédant elle-même un HisTag). (4, 4') Fraction non retenue après contre-purification sur colonne d'affinité au nickel (variant A3sF119C débarassé de son HisTag). (5, 5') Élution de la contre-purification (essentiellement la TEV protéase) (6, 6') fraction obtenue après polissage par chromatographie d'exclusion (Variant A3sF119C). MM du monomère A3sF119C (après clivage du HisTag): 22270 Da, MM de la protéase TEV : 28617 Da.

Cette construction est très efficace pour purifier une grande quantité de protéine mais se révèle particulièrement longue, j'ai donc également effectué une deuxième construction en retirant la partie N-terminale comprenant le HisTag et le site de clivage à la TEV protéase. La purification à l'aide du StrepTag étant plus rapide bien que moins efficace (30 mg/L de culture), cette construction a pu être utilisée pour des analyses biophysiques. L'unique différence entre les deux protéines obtenues est la présence des quatre acides aminés « MRGS » en amont de la protéine issue de la deuxième construction ; cette séquence est présente dans toutes les constructions génétiques d'αRep dont la structure cristallographique a été résolue et ne provoque aucun changement structural.

Après étape de clivage à la TEV protéase, les protéines sont conservées dans des tampons traités au CHELEX 100 afin de retirer les éventuels métaux contaminants qui pourraient interférer avec la phénanthroline et ainsi oxyder les cystéines.

I.3 Analyse biophysique des variants fonctionnalisés

I.3.1 Test d'Ellman

La présence des cystéines sur nos différents variants a pu être estimée grâce au test d'Ellman[9] résolu en temps. Ce test utilise la réaction entre l'acide 5,5-dithio-bis-(2-nitrobenzoïque) (DTNB) et les composés sulfhydryl tels que les cystéines non engagées dans des ponts disulfure. La quantité de cystéines libres peut donc être mesurée grâce à l'absorbance du TNB^{2-} libéré à l'issue de cette réaction (ε = 14150 $M^{-1}.cm^{-1}$ à 412 nm). Il a été observé que la concentration en cystéine libre mesurée correspondait, dans tous les variants, à la concentration de monomère mesuré à 280 nm grâce à l'absorbance des acides aminés aromatiques. La protéine sauvage ne provoquant pas d'apparition de signal d'absorbance à 412 nm, nous pouvons donc conclure que les variants ne possèdent pas de cystéines engagées dans un pont disulfure, même après une conservation prolongée en absence de réducteur (dans la glace ou à -80°C). Une autre observation rendue possible par le test d'Ellman résolu en temps est l'accessibilité des cystéines. L'apparition du signal d'absorbance à 412 nm est en effet corrélée à la concentration des deux partenaires et à l'accessibilité de la cystéine. En fixant les concentrations de DTNB (100 µM) et de protéine (~90 µM monomère), il est donc possible d'avoir accès à l'accessibilité relative de la cystéine entre les différents variants. Il apparait que dans ces conditions, les variants Y26C, K50C et D150C ont tous les trois une grande accessibilité au DTNB : le signal atteint le plateau dans les quelques secondes nécessaires à l'ajout de protéine. En revanche, le résidu F119C a une cinétique bien plus lente : le temps d'apparition de la moitié du signal ($t_{½}$) est de 1 min 40 s. Cette accessibilité plus faible est un indice encourageant concernant l'enfouissement de la phénanthroline au sein de la crevasse.

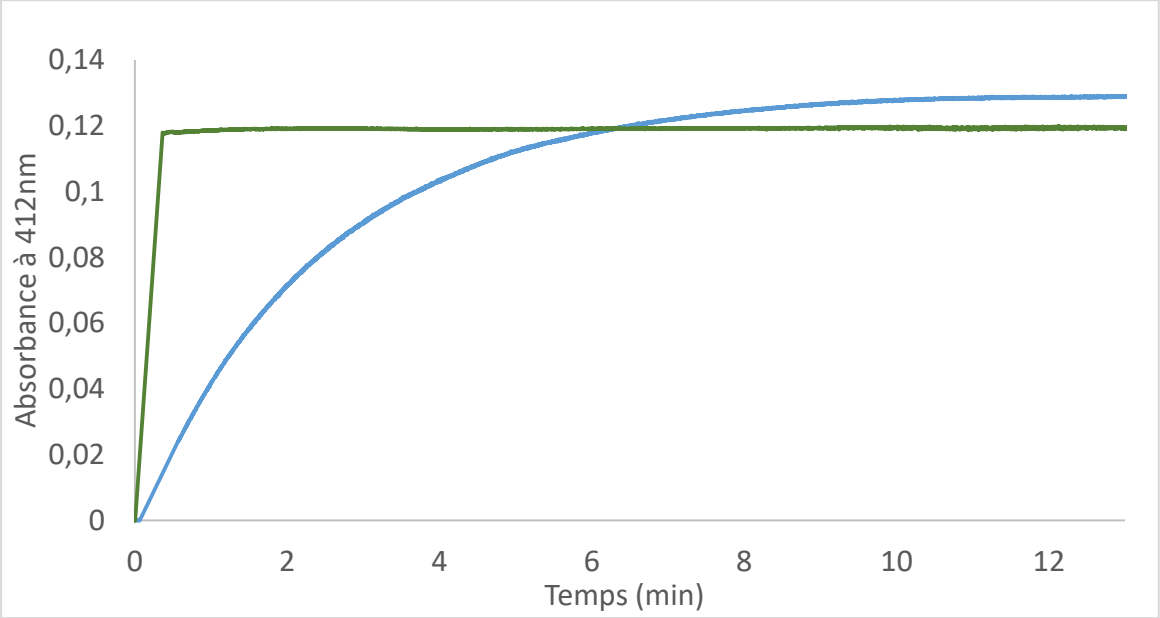


Figure 92 : Comparaison de la cinétique du test d'Ellman pour les protéines A3sD150C (vert) et A3sF119C (bleu). On observe une cinétique plus lente pour le variant A3sF119C, indiquant une moindre accessibilité de la cystéine. Les autres variants se comportent comme A3sD150C : la réaction est déjà terminée au moment de la relance de la mesure indiquant une cystéine très accessible et/ou réactive.

I.3.2 Couplage de la phénantroline sur la protéine

Le couplage s'effectue selon un mécanisme de substitution nucléophile, où le thiol du résidu cystéine de la protéine quand il est sous sa forme cystéinate attaque le carbone en α du brome. Le résultat de la réaction est un acide aminé modifié appelé résidu NPH.

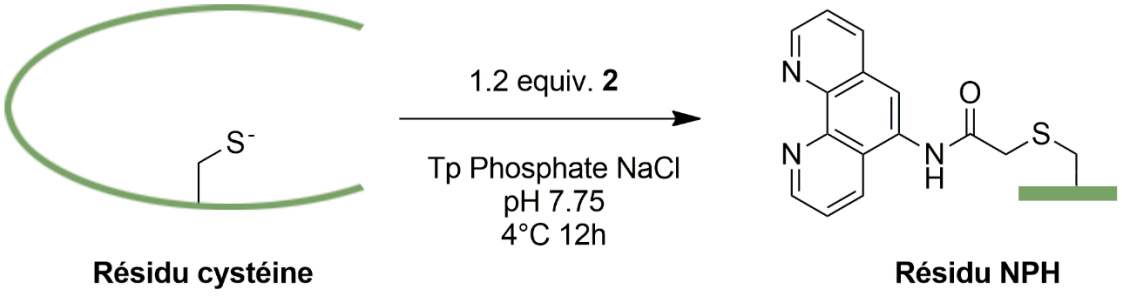


*Figure 93 : Réaction de couplage entre la phénanthroline bromoacétamide (composé **2**) et le résidu cystéine sous forme cystéinate. Le couplage s'effectue selon un mécanisme de substitution nucléophile, où le thiol du résidu cystéine de la protéine quand il est sous sa forme cystéinate attaque le carbone en α du brome. Le résultat de la réaction est un acide aminé modifié appelé résidu NPH.*

Le couplage a été effectué dans du tampon Na-phosphate 50 mM à pH 7,75, 150 mM NaCl, et traité avec de la résine CHELEX 100 afin de retirer les traces de métaux. En effet la phénanthroline de cuivre est connue pour son action d'oxydation des résidus cystéine, ce qui serait tout à fait contre-productif dans notre cas. Le ratio idéal de phénanthroline sur cystéines libres pour un couplage total après une nuit d'incubation à 4°C a été déterminé comme étant 1,2 par caractérisation au test d'Ellman et MALDI, comme expliqué ci-dessous.

I.3.3 Confirmation couplage par le test d'Ellman :

Après couplage avec la phénanthroline, le signal observé avec le réactif d'Ellman diminue drastiquement, démontrant la disparition de cystéines libres. On peut déterminer par ce biais que l'efficacité du couplage est de 95 à 100%. Il peut néanmoins s'agir d'une oxydation des cystéines, avec une molécule d'oxygène ou de la création d'un pont disulfure, des traces de cuivre pouvant être éventuellement présentes.

Il est également possible que la fonction bromoacétamide du composé **2** réagisse avec les amines primaires présentes à l'extrémité des résidus lysine, à l'amine N-terminale, et autres nucléophiles. La qualité du couplage a donc été contrôlée par spectrométrie de masse.

I.3.4 Spectrométrie de masse MALDI-ToF

La spectrométrie de masse est une technique précise mais généralement pas considérée comme quantitative (bien que dans ce cas la différence entre les deux molécules à différentier est minime). Il est cependant possible d'analyser la qualité du couplage en mesurant la différence de masse entre protéine couplée et non couplée. Il a été constaté qu'après optimisation des conditions de couplage (1,2 équivalent molaire de composé **2**, pH 7,75) il était possible d'avoir une disparition presque complète du pic correspondant à la masse de la protéine non couplée, avec apparition d'un pic présentant une différence de masse d'environ 235 Da, correspondant à l'ajout d'une seule phénanthroline par monomère.

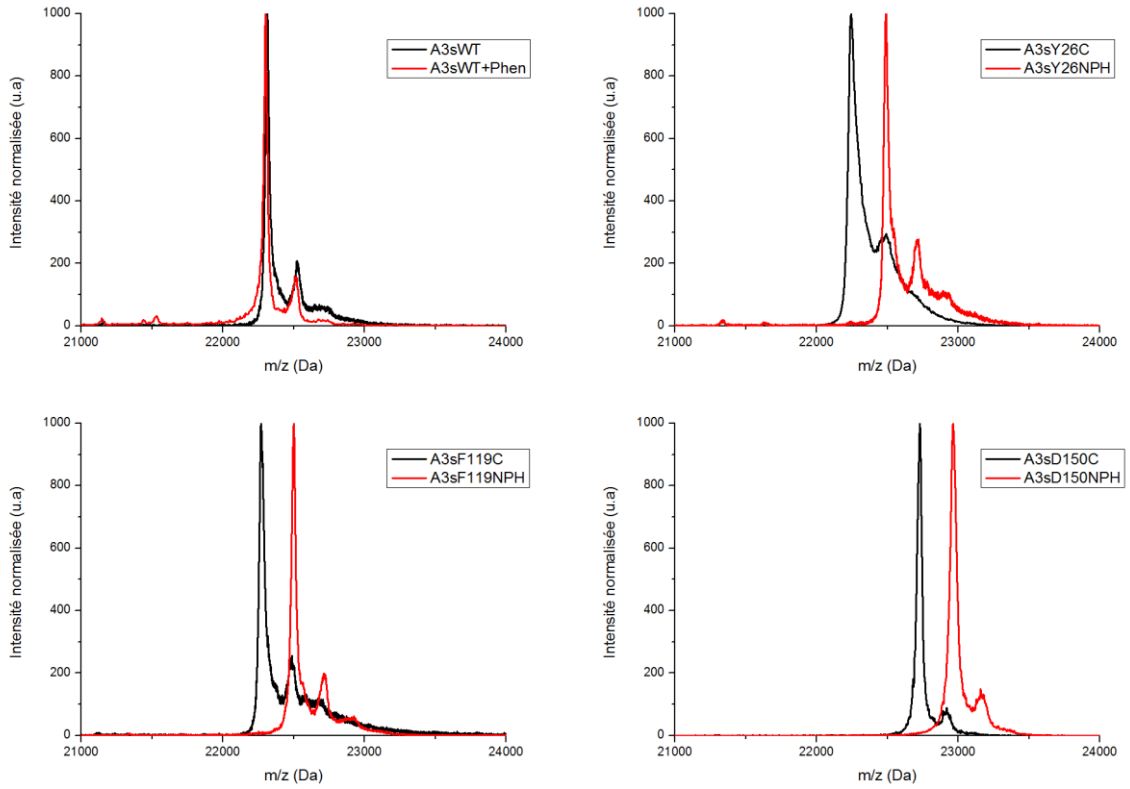


*Figure 94 : Spectre MALDI-ToF des variants avant (courbe noire) ou après (courbe rouge) couplage sur la nuit avec 1,2 équiv. de composé **2**. Les variants fonctionnalisés présentent un décalage de masse d'environ 235 Da comme attendu pour le greffage de la phénanthroline. La protéine sauvage (en haut à gauche) ne présente pas de décalage de masse, le greffage est donc bien spécifique à la cystéine. Le variant A3sK30C (non représenté) présente le même résultat.*

D'un autre côté, le spectre de la protéine A3sWT traitée de la même manière ne présente pas de pic correspondant à une protéine couplée, ce qui confirme la sélectivité du couplage de la phénanthroline envers la cystéine des différents variants.

I.3.5 Tamis moléculaire analytique

L'état d'oligomérisation des protéines couplées avec la phénantroline a été déterminé grâce au tamis moléculaire analytique. Il apparait que chacun des variants couplés à la phénanthroline possède un volume d'élution proche de la protéine sauvage. Le report sur une droite d'étalonnage effectué grâce à des protéines globulaires de taille connue montre que les variants couplés à la phénanthroline sont effectivement sous forme dimérique en solution.

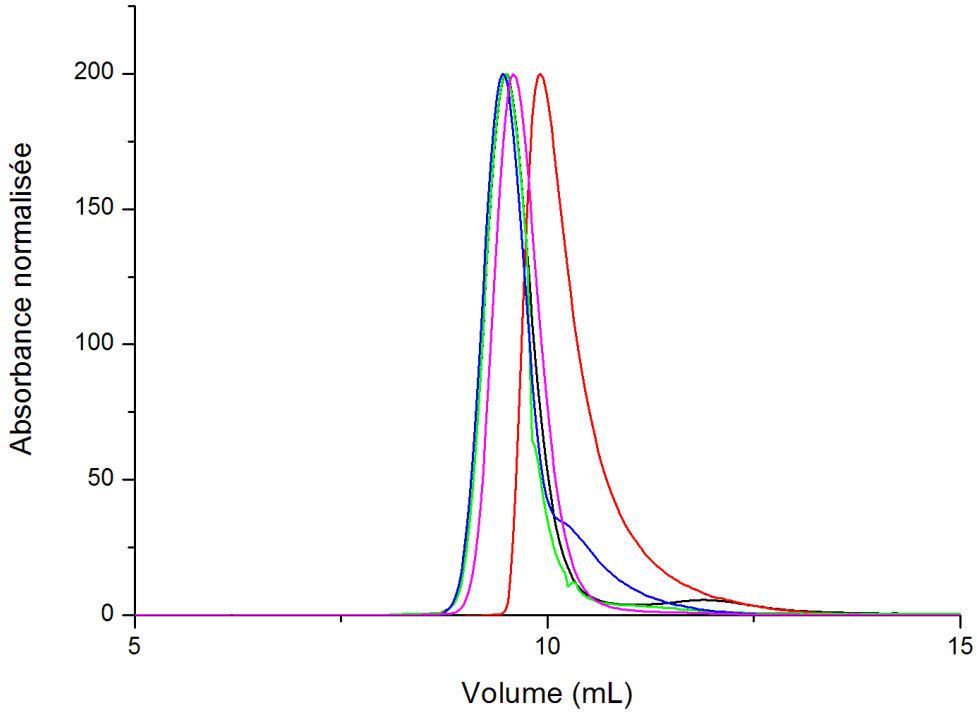


Figure 95 : Gel filtration analytique des variants couplés en présence ou non de $Zn(NO_3)_2$.(rouge) A3sWT (vert) A3sY26NPH_Zn (noir) A3sK30NPH_Zn (bleu) A3sF119NPH_Zn (magenta) A3sF119NPH

On note cependant un léger décalage vers des volumes d'élution plus faibles, ce qui suggère un volume occupé plus grand à la suite du couplage. Cette observation pourrait s'expliquer par une forme plus ouverte du dimère. La masse moléculaire apparente calculée selon la courbe d'étalonnage reste entre 37 et 44 kDa pour tous les biohybrides et la protéine sauvage, pour une masse du dimère de 44 kDa.

I.3.6 Ultracentrifugation analytique

Des expériences d'ultracentrifugation analytique menées par le Dr Christophe Velours de la plateforme PiM du CNRS ont permis de déterminer que la constante de dissociation du complexe A3 sauvage était de 43,8 nM.

La même expérience menée avec le variant A3sF119NPH non métallé n'a pas permis de déterminer précisément la constante de dissociation, la concentration la plus basse utilisée n'étant pas suffisamment diluée pour dissocier totalement le complexe. Nous pouvons donc conclure que l'affinité du complexe de variant A3sF119NPH est encore meilleure que la protéine sauvage. Ce résultat peut s'expliquer par des interactions de type liaison hydrogène ou interaction hydrophobe, soit entre les deux phénanthrolines se

faisant face, soit entre les phénanthrolines et l'architecture protéique du monomère opposé.

La même expérience menée sur le variant A3sY26NPH non métallé semble indiquer une constante de dissociation plus haute que la protéine sauvage, signe d'une déstabilisation du dimère par la phénanthroline. Ce résultat peut s'expliquer par une gêne stérique causée par l'addition de la phénanthroline, les analyses structurales semblant montrer que la position Y26 était plus encombrée. Malgré le manque de mesures à hautes concentrations, nous pouvons donner une constante de dissociation de l'ordre de la centaine de nanomolaire.

I.3.7 Calcul du coefficient d'absorption molaire des biohybrides

La concentration des protéines est communément mesurée à 280 nm, méthode rapide et fiable si le coefficient d'extinction molaire est connu avec précision.

La phénanthroline absorbant fortement à 280 nm, il a fallu corriger son effet afin de pouvoir mesurer avec la plus grande précision possible la concentration en protéine. La première méthode utilisée fut celle du dosage de Bradford, basée sur le changement d'absorbance du bleu de Coomassie après liaison avec les acides aminés basiques et hydrophobes de la protéine. Malheureusement cette méthode s'est avérée étonnamment peu reproductible entre les différents lots de biohybride (malgré des courbes d'étalonnage ayant un coefficient R^2 satisfaisant de 0,99). Il a donc été nécessaire de trouver une méthode systématique et fiable afin d'avoir une cohérence dans le dosage des différentes protéines.

Pour une protéine classique (sans pont disulfure), l'outil ProtParam du portail Expasy calcule le coefficient d'absorption à 280 nm grâce à la séquence selon l'équation :

$$\varepsilon_{Prot} = N_{Tyr} \times \varepsilon_{Tyr} + N_{Trp} \times \varepsilon_{Trp}$$

La principale cause de déviation entre coefficients calculé et mesuré étant la structure tridimensionnelle de la protéine, il est utile de comparer les spectres de la protéine native et de la protéine dénaturée dans 6M de chlorure de guanidinium pour déceler une éventuelle différence. Constatant que le spectre d'absorption de nos biohybrides changeait très peu entre l'état replié ou dénaturé, il nous a semblé raisonnable d'étendre l'équation précédente par :

$$\varepsilon_{BioH} = \varepsilon_{Prot} + N_{NPH} \times \varepsilon_{Phen}$$

Le coefficient d'absorption molaire de la phénanthroline a été mesuré dans une solution de 70% éthanol comme étant égal à 22428 $M^{-1}.cm^{-1}$. C'est cette valeur que nous additionnerons au coefficient calculé sur ProtParam de la protéine non fonctionnalisée.

I.3.8 Dichroïsme circulaire

Le dichroïsme circulaire mesure la différence d'absorbance de la lumière polarisée circulairement à droite et à gauche. Lorsque l'on mesure ce signal dans l'UV lointain, on peut observer l'absorbance de la liaison peptidique, aux alentours de 215 nm. Les structures secondaires telles les hélices α et les feuillets β ont une signature caractéristique dans cette région du spectre dichroïque, ainsi la comparaison des spectres de dichroïsme circulaire avec la protéine sauvage nous informe si la fonctionnalisation de la protéine a un effet sur ses structures secondaires. Les αRep étant des protéines possédant uniquement hélices α et des boucles, nous nous attendons à un spectre caractéristique des hélices α. La figure 96 présente les spectres des différents variants fonctionnalisés (le variant K30NPH est testé en présence de zinc). On peut alors observer une signature spectrale proche, caractéristique des protéines en hélice α. Tous les spectres sont superposables ce qui indique que la mutation et le couplage de phénantoline sur les variants n'ont pas altéré la composition en structures secondaires des protéines. Les légères variations obtenues peuvent être expliquées par les imprécisions de mesure de la concentration de la protéine couplée à la phénanthroline.

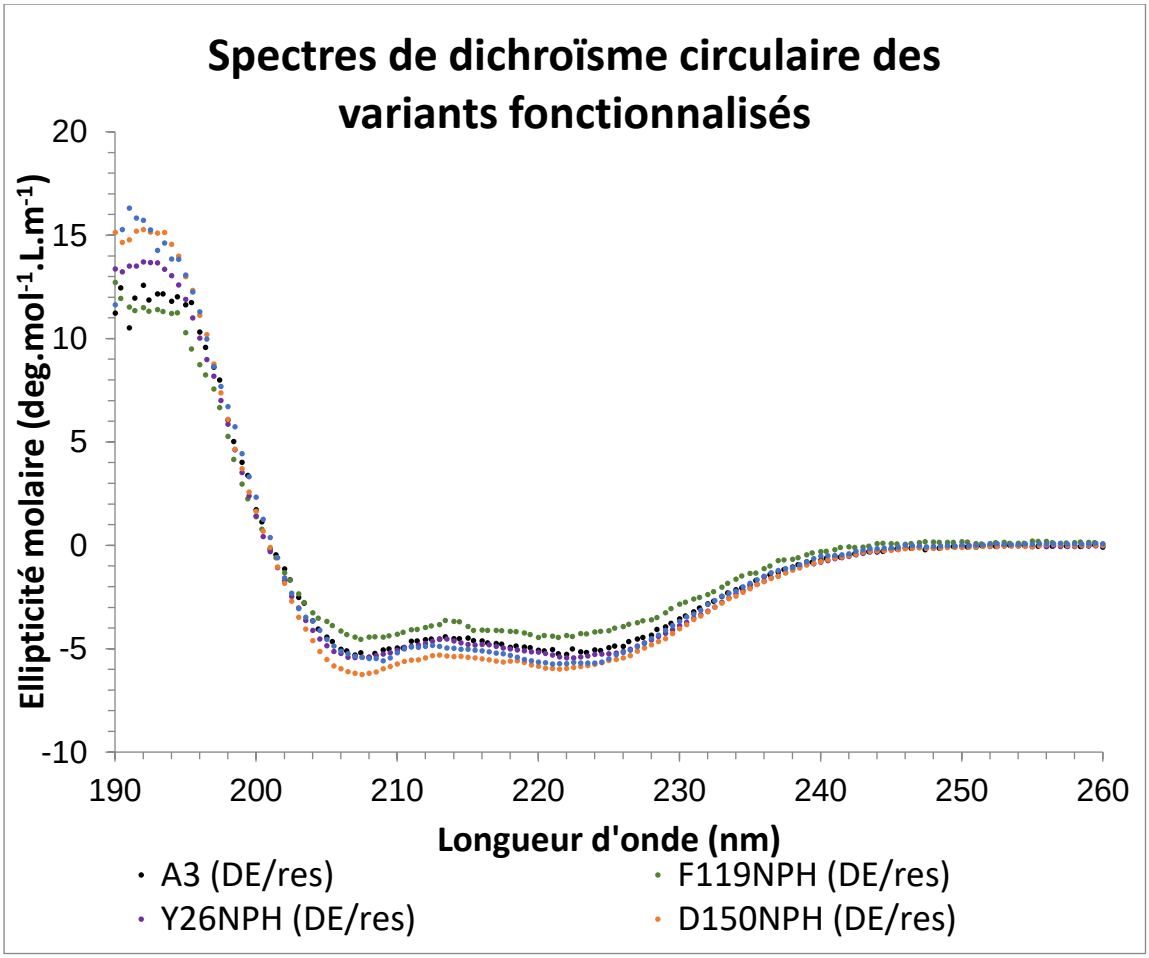


Figure 96 : Spectres de dichroïsme circulaire des variants fonctionnalisés, la protéine sauvage "A3" apparait en noir. On peut constater que les protéines conservent leurs structures secondaires avec une signature spectrale typique des hélices α.

L'ensemble des expériences de caractérisation des biohybrides obtenus après couplage de la phénantroline, nous indique que les protéines ont donc très vraisemblablement conservé leur structure tertiaire et quaternaire. L'incorporation de métal au sein de ces biohybrides était l'étape suivante pour convertir ces nouveaux objets en catalyseurs.

Dans un premier temps, nous avons testé l'activité catalytique de notre biohybride en étudiant la réaction de trans-estérification de phosphodiesters (réaction modèle d'activité RNAse). Dans un second temps, nous avons cherché à déterminer s'il peut induire une énantiosélectivité, en étudiant la réaction de cycloaddition de Diels-Alder avec des subtrats modèles.

I.4 Mime d'une activité RNAse par les biohybrides dérivés d'αRep A3

Nos biohybrides possédant deux sites de chélation du métal au sein de la crevasse nous avons décidé d'étudier l'effet de la distance entre les deux centres métalliques sur l'activité catalytique, et la possibilité d'un mécanisme binucléaire.

I.4.1 Analyse bio-informatique sur les enzymes binucléaires à zinc

Avant de tester si nos biohybrides pouvaient avoir une activité binucléaire, nous avons voulu savoir si la distance entre les deux métaux pouvait être compatible avec les distances présentes dans les enzymes binucléaires naturelles. J'ai pour cela utilisé une méthode bio-informatique en m'appuyant sur les protéines à zinc. En analysant les structures contenant du zinc dans la base de données structurale (PDB), nous pouvons classer les protéines en fonction de la distance minimale séparant deux ions zinc dans leur structure. Le graphique représentant cette distance minimale en fonction de l'ordre des protéines dans le classement (figure 97) adopte une forme particulière : un plateau est visible entre 3 et 4 Å de distance Zn-Zn. Afin de donner un indice supplémentaire, la proportion d'enzymes caractérisées dans une fenêtre glissante de 20 structures a été calculée (en orange sur la figure 97). On peut alors constater que dans la zone plateau la proportion d'enzyme est visiblement plus élevée que dans le reste de la courbe. En effet : 55% des protéines sont des enzymes entre 3 et 4 Å, contre seulement 29% au-delà de 4 Å.

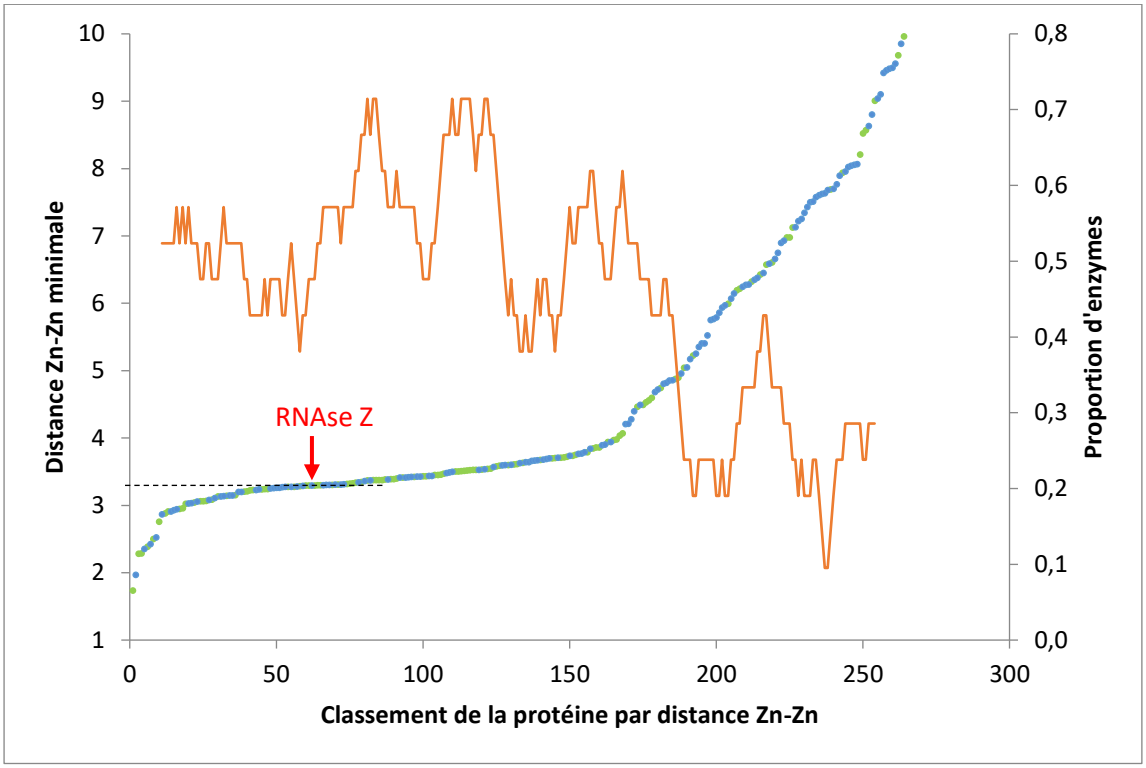
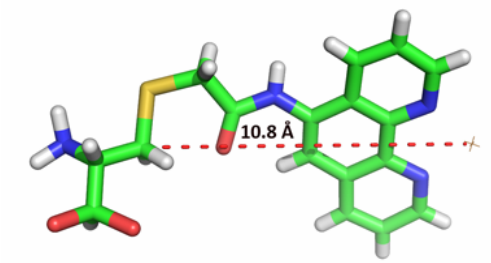


Figure 97 : Analyse bio-informatique des protéines contenant du Zinc. Bleu/vert : Distance minimale entre deux zincs selon le classement de la protéine (vert : enzymes, bleu : autres). On observe un plateau entre 3 et 4 Å. Orange : Proportion relative d'enzymes dans une fenêtre glissante de 20.

Cette analyse simple semble montrer que les enzymes ont généralement une distance Zn-Zn située entre 3 et 4 Å, notamment la RNAse Z qui catalyse la réaction d'hydrolyse de l'ARN grâce à un centre binucléaire de zinc (distance Zn-Zn : 3,30 Å). Nous avons donc fait l'hypothèse qu'il s'agissait d'une distance classique pour une enzyme au mécanisme binucléaire. Ces données sont tout à fait compatibles avec les variants Y26, K30 et F119 : un modèle du résidu NPH a été conçu à partir des données cristallographiques de la phénanthroline de cuivre et minimisé avec le logiciel Chem3D. La distance entre le carbone β et le métal a été ensuite mesurée comme étant 10,8 Å (figure 98, gauche). Par simple différence de distance entre les carbones β des structures cristallographiques de la protéine A3 dans les deux formes, nous pouvons avoir une idée de la distance possible entre les deux métaux (figure 98, droite). Les variants A3sY26NPH, A3sK30NPH et A3sF119NPH ont été considérés comme potentiels catalyseurs binucléaires compte tenu de la flexibilité du dimère et du résidu NPH. Il était attendu que les phénanthrolines se repoussent par gène stérique dans le cas de la position F119.



	Forme ouverte		Forme fermée	
Position	Distance β-β'	Distance M-M	Distance β-β'	Distance M-M
Y26	27,3	5,7	28,9	7,3
K30	22,8	1,2	24,8	3,2
F119	21,3	-0,3	18,3	-3,3
D150	26,7	5,1	30,6	9

Figure 98 : Gauche : Distance carbone ß-métal dans un modèle du résidu NPH métallé. Droite : Comparaison des distances entre carbones ß- ß' et distance calculée entre les deux métaux (M-M) grâce à l'équation (Distance ß-ß')-2x10,8. La distance négative pour la position F119 indique que les deux phénanthrolines peuvent se repousser par gène stérique.

I.4.2 Test enzymatique binucléaire : le transfert hydrolitique de liaison phosphoester

Nous avons donc tenté de déterminer si nos biohybrides fonctionnalisés en position Y26, K30 ou F119 étaient capables d'effectuer des réactions de catalyse binucléaire. Nous avons pour cela utilisé une catalyse binucléaire typique : le clivage du 2-hydroxypropyl-4-nitrophenyl phosphate (HPNP) par transfert de liaison phosphoester. Le HPNP est utilisé comme modèle d'ARN pour la création de nucléases artificielles, ses principaux avantages étant sa solubilité dans l'eau et la possibilité de suivre la cinétique de la réaction à 405 nm grâce au composé nitrophénolate libéré. Dans une étude de Kazuki Saito, un modèle du mécanisme de clivage du HPNP à l'aide d'un complexe binucléaire de zinc a été proposé.[10] Ce mécanisme implique que l'hydroxyde lié à un ion zinc agit comme simple base de Brønsted, déprotonant le groupement alcool du HPNP, ce dernier effectuant l'attaque nucléophile sur le phosphate (ce modèle a été appuyé par une récente étude théorique [11]). L'étude démontrait la nécessité d'un centre binucléaire pour une catalyse efficace de la réaction : alors qu'un catalyseur de zinc mononucléaire accélérait la réaction par un facteur 22, un complexe binucléaire l'accélérait par un facteur 6500. C'est cette forte augmentation de la catalyse que nous tenté d'observer afin de confirmer ou d'infirmer le caractère binucléaire de nos biohybrides.

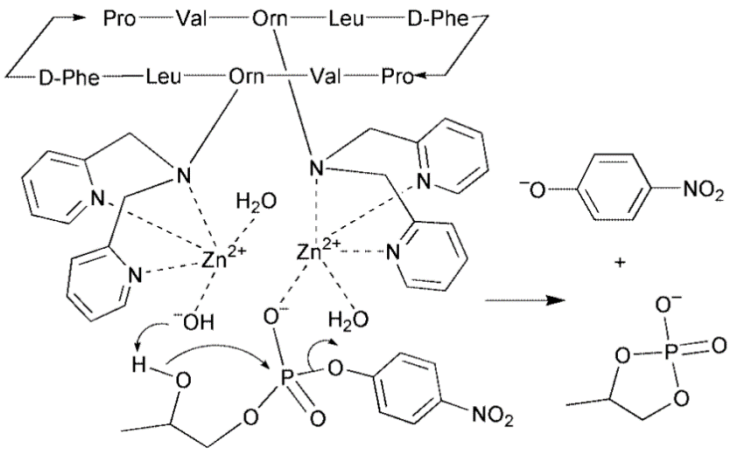


Figure 99 : Mécanisme de phosphotransfert du HPNP catalysé par un complexe binucléaire de zinc (selon [10]).

Les variants A3sY26NPH, A3sK30NPH et A3sF119NPH ont donc été testés pour l'hydrolyse de HPNP en présence d'un léger excès de nitrate de zinc (1,1 équivalent molaire par rapport à la phénanthroline). La catalyse mettant en jeu des bases de Brønsted, cette réaction est sensible au pH (« profil pH »). J'ai donc effectué les réactions à différents pH afin d'observer si l'environnement du biohybride avait un effet par rapport au zinc en solution. Pour s'affranchir des possibles perturbations dues au changement de tampon, j'ai utilisé uniquement un tampon HEPES. Le changement de pH induit par l'ajout du catalyseur dans sa solution tamponnée a été corrigé, ainsi que l'absorbance du 4-nitrophénolate libéré à l'issu de la réaction. Les résultats sont présentés dans la figure 100.

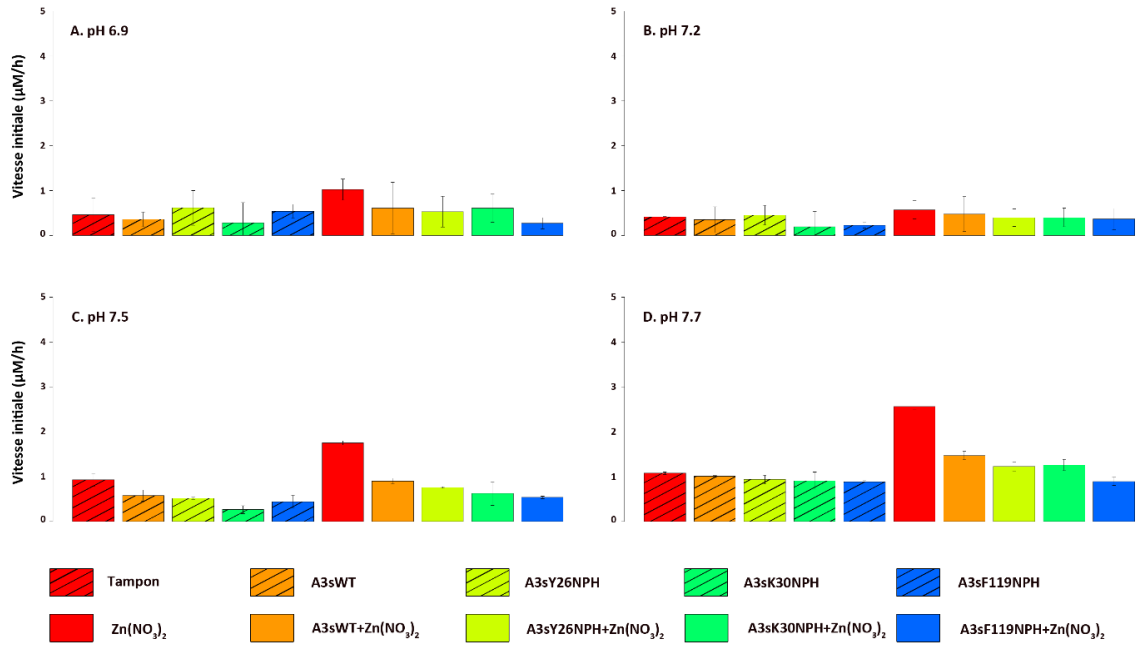


Figure 100 : Hydrolyse du HPNP par les protéines A3sWT, A3sY26NPH, A3sK30NPH et A3sF119NPH en présence ou non de zinc (II), au pH 6,9 (A), 7,2 (B), 7,5 (C) et 7,7 (D). Conditions de réaction : 16,5 µM de monomère A3s, 16,5 µM de $Zn(NO_3)_2$ (1 équivalent), tampon HEPES 20 mM, 150 mM NaCl, 5 mM HPNP. Chaque réaction a été effectuée en duplicate.

Il ressort de cette expérimentation que dans chacun des pH, les biohybrides en présence de zinc ont une activité plus basse que le zinc en solution. Il semble donc que les biohybrides empêchent l'activité catalytique centrée sur le métal. La protéine sauvage ayant également le même comportement, il semblerait que même en absence de phénanthroline la protéine ait une capacité à capturer le zinc et diminuer son accessibilité envers le substrat. Il est également à noter que les vitesses de réactions croissent avec le pH, qu'il s'agisse du zinc en solution ou en présence de protéine. Il n'y a donc pas d'effet de la protéine sur le profil pH du zinc, et *a fortiori* aucune activité binucléaire. Afin de pouvoir tester la réaction à un pH plus élevé encore j'ai utilisé un tampon d'acide N-Cyclohexyl-2-aminoethanesulfonique (CHES) pH 9, les conclusions sont identiques.

J'ai effectué la même expérience avec du cuivre (II), capable également d'effectuer la catalyse du HPNP. Les résultats sont visibles sur la figure 101. Les conclusions sont similaires, il est à noter que le cuivre est un meilleur catalyseur de la réaction d'hydrolyse du HPNP dans ces conditions. La protéine sauvage, cette fois, inhibe bien moins le cuivre que les biohybrides. Il est donc raisonnable d'émettre l'hypothèse que le métal se fixe effectivement à la phénanthroline et que c'est cette complexation qui diminue d'autant l'activité du cuivre (par un manque d'accessibilité du substrat). Étonnamment, même le biohybride fonctionnalisé en position K30, qui est située au bord de la crevasse et donc

très accessible au solvant, présente une vitesse de réaction du même ordre que les autres variants, et empêche donc la réaction.

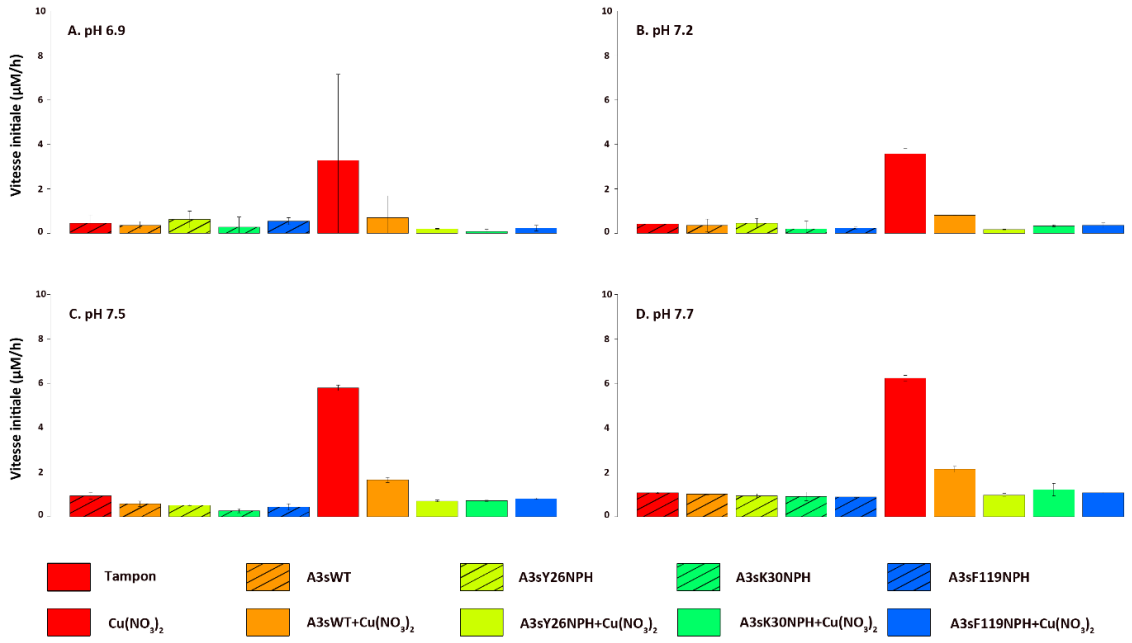


Figure 101 : Hydrolyse du HPNP par les protéines A3sWT, A3sY26NPH, A3sK30NPH et A3sF119NPH en présence ou non de nitrate de cuivre (les conditions sans cuivre (II) sont hachurées) au pH 6,9 (A), 7,2 (B), 7,5 (C) et 7,7 (D). Conditions de réaction : 16,5 µM de monomère, 16,5 µM de $Cu(NO_3)_2$ (1 équivalent), tampon HEPES 20 mM, 150 mM NaCl, 5 mM HPNP. Chaque réaction a été effectuée en duplicate.

I.4.3 Résumé

En résumé, (1) tous les biohybrides empêchent la réactivité du cuivre ou du zinc soit par complétion de la sphère de coordination du métal par des acides aminés de la protéine, soit par un manque d'accessibilité du substrat. (2) La protéine sauvage semblerait également fixer les métaux en provoquant un effet similaire.

Il est courant que les protéines possèdent des sites fortuits de fixation d'ions métalliques. La connaissance de la constante de dissociation entre protéine et métal pourrait être d'un grand avantage. Malheureusement les études de titration calorimétrique isotherme (ITC) se sont révélées peu concluantes, les signaux étant bruités par l'interaction entre les métaux et les molécules de la solution tampon.

La baisse de l'efficacité catalytique des métaux lors de la réaction d'hydrolyse du HPNP peut être interprétée comme un enfouissement du métal au sein de la protéine, limitant l'accès au substrat. Nous avons alors émis l'hypothèse que cet enfouissement pouvait provoquer une énantiosélectivité lors de la catalyse. Nous avons décidé d'utiliser la catalyse de Diels-Alder pour tester cette hypothèse.

I.5 Etude d'une activité catalytique énantiosélective : la réaction de Diels-Alder

La catalyse de Diels-Alder est une réaction de cyclisation entre un diène et un diénophile. Lors de cette réaction, les électrons π des deux molécules se réarrangent selon un mécanisme concerté pour former deux nouvelles liaisons σ. De manière générale, c'est l'orbitale basse vacante (BV) du diénophile qui est en interaction avec l'orbitale haute occupée (HO) du diène. Cette configuration (nommée « Diels-Alder à demande normale d'électrons ») est favorisée par la présence de groupements électroattracteurs sur le diénophile et/ou de groupements électrodonneurs sur le diène. Les deux molécules partenaires de la réaction peuvent être substituées sur leurs différents carbones, provoquant lors de la cyclisation la création de jusqu'à quatre centres stéréogènes. La configuration adoptée par le produit formé dépendra de la géométrie des deux partenaires au moment de la réaction.

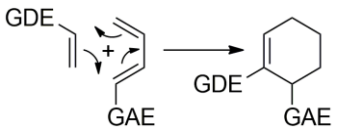


Figure 102 : Schéma général de la réaction de Diels-Alder. GDE : donneur d'électrons, GAE : attracteur d'électrons

Les métaux de transition sont connus pour catalyser efficacement cette réaction par leur propriété d'acides de Lewis en solvant organique. La complexation de l'ion métallique avec le groupement électroattracteur du diénophile (dans le cas d'une réaction à demande normale d'électrons) diminue l'énergie de l'orbitale basse vacante et donc la différence d'énergie HO-BV, facilitant la réaction. Un milieu aqueux favorise également la réaction, ce qui est expliqué par la combinaison de deux effets : (i) Le diène et diénophile étant souvent des molécules hydrophobes, l'eau diminue sa surface de contact avec celles-ci, forçant les deux partenaires à se rapprocher et stabilisant leur état de transition. (ii) L'eau peut également effectuer une liaison hydrogène avec le groupement électroattracteur du diénophile et potentialiser son effet, à l'instar des acides de Lewis.[12]

Il est cependant à remarquer que les métaux de transition sont rarement efficaces dans l'eau, en effet un substrat diénophile monodenté est alors en compétition avec les molécules d'eau complexées autour du métal. Or l'eau est une bonne base de Lewis et en concentration bien plus grande que le substrat.[13] C'est pour cette raison que les substrats communément utilisés lors de la réaction de Diels-Alder catalysée par les métaux de transition dans l'eau sont bidentés. Un couple diène-diénophile particulièrement utilisé est le couple cyclopentadiène et le (*E*)-3-phényl-1-(pyridine-2-yl)prop-2-ène-1-one (azachalcone) (figure 103). Quatre produits sont possibles à l'issue de la réaction : deux produits *endo* et deux produits *exo*. La forme *endo*, bien que stériquement plus encombrée, est largement favorisée lors de cette réaction. La raison communément

acceptée étant un recouvrement favorable des orbitales de l'oxygène de l'azachalcone et du cyclopentadiène (effet orbitalaire secondaire ou « effet endo »).

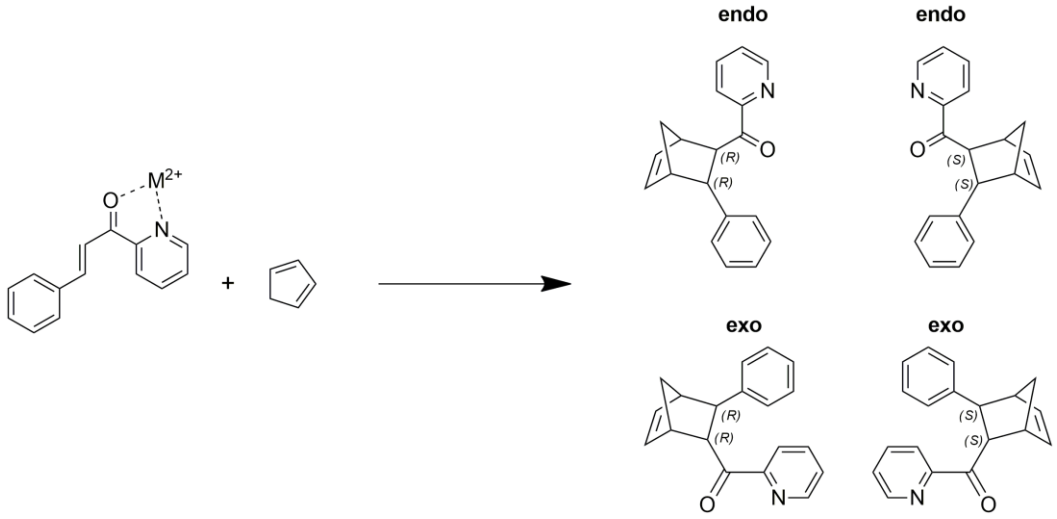


Figure 103 : Azachalcone (sous la forme proposée complexée au métal de transition [13]) et cyclopentadiène, utilisés comme substrat de la réaction de Diels-Alder. Quatre produits sont possibles, la forme endo étant favorisée par intéraction orbitalaire secondaire entre le diène et l'oxygène de l'azachalcone.

Parmi les deux formes *endo*, la réaction accélérée par un catalyseur achiral n'est pas énantiosélective. C'est l'excès énantiomérique de la forme *endo*, ainsi que le rapport *endo*:*exo*, qui sont communément utilisés pour examiner et comparer le caractère sélectif des catalyseurs (et notamment des catalyseurs biohybrides). Le produit de la réaction est typiquement analysé par chromatographie chirale après extraction en solvant organique. C'est ce couple de substrats que nous avons choisi pour l'étude de la sélectivité de nos différents variants.

Les réactifs utilisés pour mettre en place la réaction de cycloaddition de Diels-Alder ainsi que les authentiques des produits de cette réaction ont été fournis par Dr. Wadih Ghattas (ICMMO de l'Université Paris-Sud, CNRS). La méthode classique de préparation du dièneophile (*E*)-3-phényl-1-(pyridine-2-yl)prop-2-ène-1-one (azachalcone) comprend une étape finale de purification par cristallisation or ce composé sous forme cristalline est sensible à la lumière [14]. Cette étape a été alors remplacée par une purification par chromatographie. Le cyclopentadiène a été obtenu par distillation suite au craquage de son dimère commercial. Il s'est avéré stable pour une durée de conservation d'un mois à -20°C. Enfin, La réaction entre azachalcone et cyclopentadiène a été catalysé par du chlorure d'aluminium dans du dichlorométhane pour fournir le mélange authentique des quatre produits isomères.

I.5.1 Méthode d'extraction, rendements et contaminants

Avant d'aborder la description des résultats, il est important d'expliquer la façon dont ils ont été obtenus. Plus que la simple méthode (décrite dans la partie éponyme), il est utile de comprendre les sources de variation ayant pu entacher ou compliquer les différentes expérimentations et analyses. J'introduirai alors les méthodes calculatoires utilisées pour les corriger. Plus qu'un regard *a posteriori* sur les résultats obtenus, cette partie se veut également être un retour d'expérience et une invitation à la prudence lors de ce type d'étude.

Il est tout d'abord nécessaire de prendre en compte le dispositif expérimental menant de la catalyse au résultat. Les différentes catalyses ont été menées dans un volume de 500 µL. Après le temps d'incubation voulu, allant de 24h à 96h, les produits ont été extraits trois fois avec un volume équivalent d'acétate d'éthyle. Lors de cette extraction la protéine se trouvait généralement sous forme précipitée, ce qui nécessitait une centrifugation afin de pouvoir récupérer la phase organique. L'extrait était ensuite injecté sur colonne d'HPLC chirale. Les chromatogrammes à différentes longueurs d'onde pouvaient enfin être traitées afin de déterminer l'excès énantiomérique et le rendement de la réaction (figure 104).

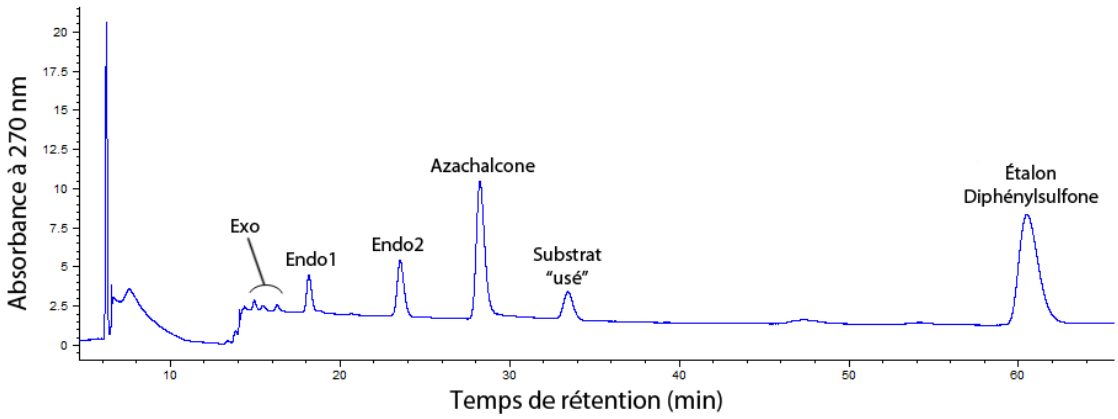


Figure 104 : Exemple d'une courbe obtenue après injection des extraits de réaction de Diels-Alder sur HPLC chirale. Le contaminant n'est pas visible sur ce chromatogramme.

Ces expériences ont été menées de nombreuses fois et j'ai été confronté à des difficultés dans le traitement des données qui ne semblaient pas reproductibles. L'analyse des résultats m'a permis de comprendre partiellement les raisons de ces variabilités que j'expose ci-dessous.

La principale source de contamination vient du matériel utilisé lors du protocole d'extraction à l'acétate d'éthyle. En effet, par soucis de miniaturisation, de maintien de la chaîne du froid et du besoin de centrifuger afin de séparer les phases, les outils classiques

de biologie ont été utilisés (tubes microcentrifugeuse en polypropylène, micropipettes, centrifugeuse réfrigérée). Ce n'est qu'après une analyse attentive des chromatogrammes HPLC à différentes longueurs d'onde (270 et 230 nm en particulier) que j'ai pu constater sur certains chromatogrammes qu'un des pics de produit *endo* de la réaction pouvait être parfaitement superposé avec un autre produit contaminant, faussant le calcul de l'excès énantiomérique. Le ratio d'absorbance du produit *endo* a donc été calculé à partir de l'aire du pic non contaminé de plusieurs chromatogrammes (équation 1).

$$R_{endo} = \frac{A_{endo}^{270}}{A_{endo}^{230}} = 0{,}501 \pm 0{,}03 \quad (1)$$

Connaissant cette valeur, et sachant que le cuivre catalyse la formation des deux produits *endo* de manière racémique, il a été possible d'estimer le ratio 270/230 (R_{conta}) du contaminant grâce aux aires du pic contaminé ($A_{endoConta}$) et du pic non contaminé ($A_{endo'}$) (équation 2).

$$R_{conta} = \frac{A_{endoConta}^{270} - A_{endo'}^{270}}{A_{endoConta}^{230} - A_{endo'}^{230}} \cong 0{,}157 \quad (2)$$

C'est un contaminant possédant ce ratio R_{conta} et un volume d'élution proche de celui du produit *endo* que j'ai donc recherché. Après avoir effectué le protocole d'extraction sur chacun des composés indépendants du milieu réactionnel, j'ai pu à chaque fois observer un pic non négligeable de contaminant possédant un ratio R_{conta} de 0,154 ± 0,003. Ce contaminant était également présent sous forme de traces dans l'acétate d'éthyle pur. C'est en comparant de l'acétate d'éthyle pipeté avec les cônes du laboratoire avec d'autres cônes de la même composition (*Corning® DeckWorks™ low binding tips* et *Corning® universal fit pipet tips* respectivement, tous deux en polypropylène) que j'ai pu constater que nos cônes (et dans une proportion moindre, les tubes de microcentrifugeuse) semblaient être la source de cette contamination.

Cette analyse a été compliquée par la deuxième source de variabilité découverte au cours de l'étude : la mise à température de la colonne HPLC prenait plusieurs heures contrairement à l'indication de la sonde, influant grandement les volumes d'élution du contaminant comme on peut le constater sur la figure 105. La variation du volume d'élution en fonction de la température étant dépendant du type de molécule, nous avions l'explication de la présence ou non du contaminant dans nos pics d'intérêt.

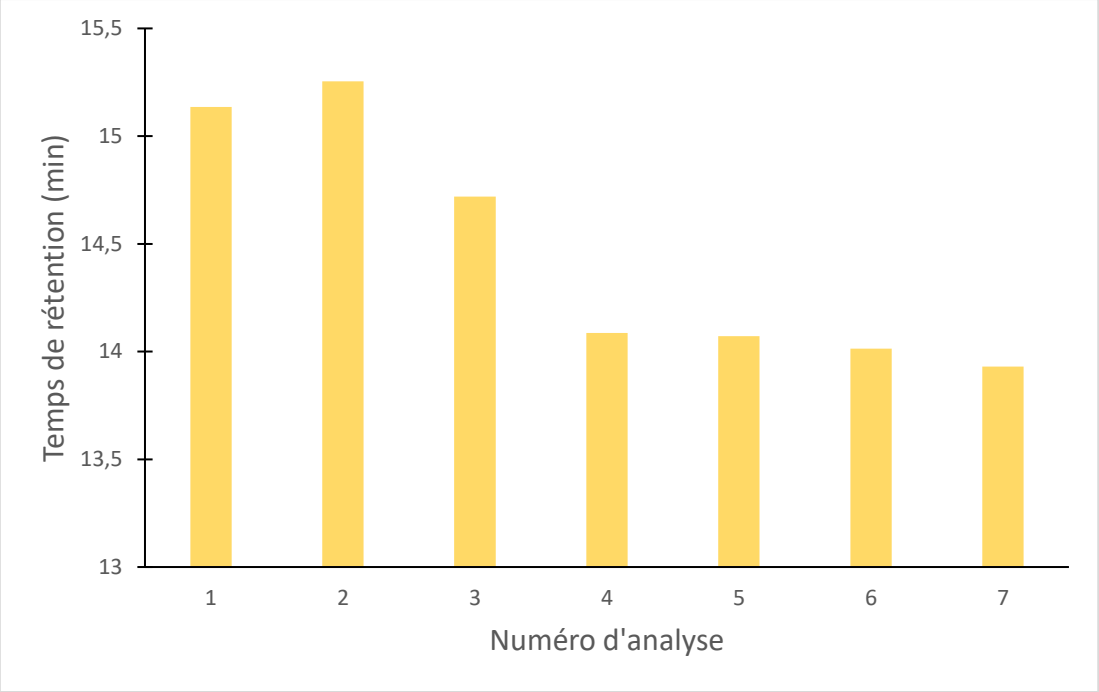


Figure 105 : Temps de rétention du contaminant en fonction du nombre d'analyses (analyses de 50 min). Le temps de rétention est très sensible à l'équilibration de la température de l'appareil.

Les rapports d'absorbance 270/230 étant connus pour le produit *endo* (R_{endo}) et le contaminant (R_{conta}), nous pouvions alors détecter si le contaminant est présent dans un pic analysé et le corriger si nécessaire grâce à l'équation (3).

$$A_{endo}^{270} = \frac{A_{endoConta}^{270} - (R_{conta} \times A_{endoConta}^{230})}{1 - \frac{R_{conta}}{R_{endo}}} \quad (3)$$

Un dernier effet à prendre en compte est celui de la dégradation du substrat. Le substrat a été dilué dans un volume d'eau correspondant au volume réactionnel et extrait selon la méthode usuelle, comparé avec du substrat non extrait, directement dilué dans un volume équivalent d'acétate d'éthyle. Après analyse HPLC, trois pics étaient observables (outre le contaminant précédemment cité) : l'étalon interne, le substrat et un pic correspondant à une forme dégradée du substrat (appelée « substrat usé ») apparaissant quelques minutes après le pic de substrat (figure 104). Or le pic correspondant au substrat usé augmentait d'un facteur 2,5 après le protocole d'extraction (figure 106). Une hypothèse est que le substrat se dimérise par action de la lumière selon une cycloaddition [2+2], formant un cyclobutane (figure 107). Cette dimérisation a déjà été décrite pour des dérivés de l'azachalcone en solvant organique [15] et en milieu solide [14]. Il a cependant été remarqué sur certains chromatogrammes effectués avec une colonne récente que le pic de substrat présentait un pont jusqu'au pic de substrat usé. Ceci étant tout à fait

caractéristique d'une dégradation durant l'analyse, cette observation va à l'encontre de l'hypothèse de dimérisation induite par la lumière.

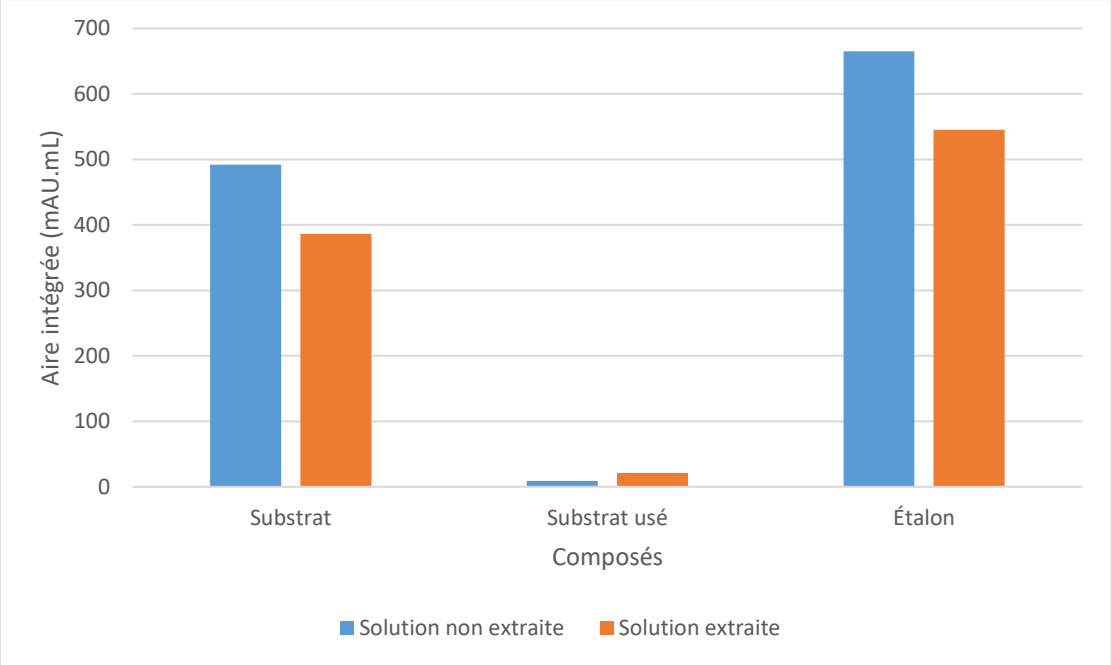


Figure 106 : Comparaison des aires non normalisées des pics obtenus pour des composés de la solution de substrat avant (bleu) et après extraction (orange). On observe que l'extraction provoque la formation de « substrat usé ».

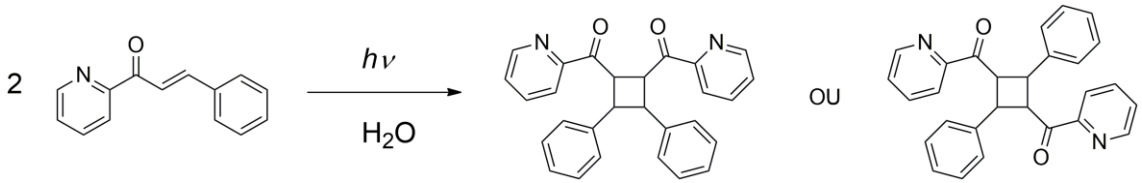


Figure 107 : Produit de la cycloaddition [2+2] de l'azachalcone sous l'effet de la lumière.

Cette dégradation du substrat est source de variabilité lors du calcul du taux de conversion. Il est ainsi plus fiable de comparer les rendements de réaction grâce à l'apparition des produits attendus : le substrat étant en excès et la réaction n'étant pas totale dans les conditions utilisées, la dégradation du substrat a *a priori* peu d'impact sur la catalyse.

L'analyse des produits donne en théorie 4 pics correspondant aux 2 diastéréoisomères sous la forme *endo* et aux 2 diastéréoisomères sous la forme *exo*. En pratique, les pics correspondant aux formes *exo* étaient trop petits pour permettre le calcul de leur aire, représentant moins de 5% de la surface des formes *endo*, et les calculs d'excès enantiomériques ont donc été réalisés en comparant les aires des 2 pics correspondant à la forme *endo* (figure 104).

Pour la suite, nous présenterons donc les résultats de catalyse, en calculant le rendement de la réaction (équation 4), tenant compte des corrections décrites ci-dessus, et les excès énantiomériques (équation 5).

$$Rendement\ (\%) = \frac{Quantité\ de\ produit\ mesuré}{Quantité\ de\ substrat\ initial} \times 100 \quad (4)$$

$$ee\ (\%) = \frac{A_{endo2} - A_{endo1}}{A_{endo2} + A_{endo1}} \times 100 \quad (5)$$

I.5.2 Premier crible d'activité

Les 4 différents variants ont tout d'abord été criblés pour un test d'activité avec l'azachalcone et le cyclopentadiène. Un équivalent de cuivre (par rapport à la phénanthroline) a été ajouté extemporanément au variant et la catalyse a été lancée dans des conditions standards : 30 µM catalyseur (monomère A3s), 34 mM (1133 équivalents) de cyclopentadiène et 1 mM (33 équivalents) d'azachalcone dans un volume final de 500 µL pendant 48h à 4°C. Les résultats d'excès énantiomérique et de rendement (calculé grâce à l'apparition du substrat) sont reportés dans le tableau 108.

Monomère	Ratio Cu(II):Monomère	Tampon	pH	Durée	Cosolvant	*ee endo* (%)	Rendement (%)
-	1:-	HEPES/MES (80/20)	7,5	48 h	-	2 ± 2	15 ± 3
A3sWT	1:1					1,0	30,0
A3sY26NPH						0,2	7,0
A3sK30NPH						5,8	2,9
A3sF119NPH						**27,8**	8,2
A3sD150NPH						1,6	5,0

Tableau 108 : Catalyse de la réaction de Diels Alder catalysée par le cuivre (30 µM) entre azachalcone (1 mM) et cyclopentadiène (34 mM) en présence ou non de 30 µM de monomère de A3s. Tampon HEPES/MES 20 mM pH 7,5, 150 mM NaCl, 48h

Nous voyons tout d'abord qu'à l'exception de la protéine sauvage le rendement de la réaction est, pour tous les variants, inférieur à celui du nitrate de cuivre libre en solution. Cette diminution de l'activité correspond aux observations faites lors des expériences d'hydrolyse du HPNP suggérant que la protéine modifie l'accessibilité du cuivre en solution ou provoque une inhibition par le produit (ou le substrat). Le variant A3sF119NPH semblait être le plus intéressant, présentant une énantiosélectivité de presque 30%. C'est donc celui-ci que nous avons choisi d'analyser plus en détails.

Une cinétique de réaction dans ce tampon a été effectuée avec un point à 48h, 96h et 144h. Par ailleurs, un test catalyse dans différents ratios de MES/HEPES avec un pH final

de 5,5 ; 6,5 et 7,5 a été effectué, avec 64h de temps d'incubation. Ces résultats sont consignés dans le tableau 109.

Monomère	Ratio Cu(II):Monomère	Tampon	pH	Durée	Cosolvant	*ee endo* (%)	Rendement (%)
-	1:-	HEPES/MES (80/20)	7,5	48h	-	2	16
				96h		3	18
				144h		2	21
A3sF119NPH	1:1			48h		**28**	8
				96h		**33**	14
				144h		**34**	19

Tableau 109 : Cinétique de la catalyse par le variant A3sF119NPH. Réactions de cycloaddition de Diels Alder avec le variant A3sF119NPH en présence de cuivre. Ratio molaire MES/HEPES de 20/80 pour un pH de 7,5. 20 mM tampon, 150 mM NaCl, 30 µM de $Cu(NO_3)_2$, 30 µM monomère de A3s, 1 mM azachalcone (33 équiv.) dilué dans CH_3CN (1% final), 34 mM cyclopentadiène (1133 équiv.).

Nous pouvons observer dans l'expérience de cinétique que le rendement augmente pour le biohybride et le cuivre en solution (16 à 21% pour le cuivre et 8 à 19% pour le biohybride, à 48h et 144h respectivement), tandis que l'excès énantiomérique semble peu affecté (toujours autour de 30%). De même, le cuivre semble atteindre un plateau après 96h.

Monomère	Ratio Cu(II):Monomère	Tampon	pH	Durée	Cosolvant	*ee endo* (%)	Rendement (%)
-	1:-	HEPES/MES (0/100)	5,5	64h	-	4	32
		HEPES/MES (40/60)	6,5			1	52
		HEPES/MES (80/20)	7,5			2	29
A3sF119NPH	1:1	HEPES/MES (0/100)	5,5			3	39
		HEPES/MES (40/60)	6,5			10	26
		HEPES/MES (80/20)	7,5			**30**	18

Tableau 110 : Effet du pH sur la catalyse par le variant A3sF119NPH. Réactions de cycloaddition de Diels Alder avec le variant A3sF119NPH en présence de cuivre. Ratio molaire MES/HEPES de 100/0 ; 60/40 ; 20/80 pour un pH de 5,5 ; 6,5 ; 7,5 respectivement. 20 mM tampon, 150 mM NaCl, 30 µM de $Cu(NO_3)_2$, 30 µM monomère de A3s, 1 mM azachalcone (33 équiv.) dilué dans CH_3CN (1% final), 34 mM cyclopentadiène (1133 équiv.).

Concernant les résultats de catalyse à différents pH, nous pouvons voir tout d'abord que l'excès énantiomérique du biohybride diminue proportionnellement avec le pH, allant de 30% à pH 7,5 jusqu'à seulement 3% à pH 5,5. Une dénaturation de la protéine est à exclure, des expériences de dichroïsme circulaire nous ayant indiqué que la protéine conservait sa structure en hélices α même après 64h à pH 3 (données non montrées).

Deux autres explications sont possibles : une protonation de la phénanthroline du fait du pH acide pourrait libérer le cuivre qui catalyserait librement en solution, affectant l'excès énantiomérique. Cette explication pourrait être valable pour le pH le plus bas étant donné

que le pKa de la première protonation de la phénanthroline est de 5.[16] En revanche elle n'explique pas pourquoi la baisse d'énantiosélectivité est déjà visible à pH 6,5. La deuxième explication viendrait du ratio molaire de tampon MES. Étant plus haut dans les pH plus bas, la perte de sélectivité peut être due à une interaction avec ce tampon. Ces considérations nous ont poussés à déterminer si le cuivre était effectivement là où nous l'attendions (c'est-à-dire complexé par la phénantroline) avant de d'approfondir la recherche des paramètres physico-chimiques favorables à une réaction énantiosélective.

Enfin, on constate que les rendements à 64h sont anormalement plus hauts que ceux à 96h dans la même condition mais au cours d'une expérience réalisée à un autre moment. Cette différence ne peut être imputée aux lots de protéine utilisés dans ces 2 expériences car un effet similaire est également observé dans l'expérience contrôle de la catalyse par le cuivre sans protéine. Cette marge d'erreur dans les calculs des rendements entre deux expériences indépendantes pourrait provenir d'une incertitude supplémentaire liée à la pesée et au pipetage des substrats et de la précipitation l'étalon interne dans la solution mère de substrat azachalcone, faussant légèrement les calculs.

I.5.3 Résonnance Paramagnétique Electronique

Le cuivre (II) étant une espèce paramagnétique (possédant un électron de valence non apparié sur son orbitale *d*), il est possible de l'étudier par résonnance paramagnétique électronique (RPE). Cette méthode, analogue à la résonance magnétique nucléaire (RMN) repose sur la mesure de l'absorption par l'échantillon d'une onde électromagnétique haute fréquence (fixe). L'échantillon étant dans un champ magnétique variable, la fréquence absorbée par l'échantillon correspond à la différence d'énergie entre les deux états de spins de l'électron non apparié (induite par le champ magnétique). Cette technique est sensible et quantitative, l'intégration de la courbe d'absorption nous permettant de quantifier l'ion cuivre en solution. Il est également possible, par analyse qualitative du spectre d'absorption, de suggérer un environnement électronique autour du cuivre.

J'ai tout d'abord mis en présence 50 µM de monomère de A3sF119NPH avec un équivalent molaire de nitrate de cuivre dans du tampon HEPES et analysé l'échantillon après incubation sur la nuit. Les spectres RPE ont été acquis grâce au Dr. Christian Herrero, responsable de la plateforme RPE de l'ICMMO, tous à une température de 100K et les mêmes paramètres. Aucun signal n'a été détecté (non montré) tant pour le cuivre en présence de la protéine que pour le cuivre seul dans le tampon. Cette observation nous a permis de conclure que le tampon HEPES masquait le signal du cuivre.

J'ai alors préparé les échantillons dans du tampon MOPS, qui est peu coordinant. Dans ce tampon également, le cuivre en solution est silencieux (donnée non montrée). Cependant, après ajout d'un équivalent molaire de cuivre (par rapport au monomère d'A3s) dans une solution de protéine sauvage à 200 µM de monomère, il est possible d'observer un signal caractéristique (figure 111, courbe bleue). Ce résultat indique que le cuivre (II) est complexé avec la protéine sauvage, ce qui confirmerait les observations faites précédemment lors de l'hydrolyse du HPNP. La fixation du cuivre (II) au sein de la protéine non modifiée est surprenante car la protéine issue de la banque d'αRep n'a pas été spécifiquement conçue de posséder un site de fixation du cuivre. Cependant, la présence de plusieurs de résidus portant des fonctions chargées à la surface de la protéine (Asp, Glu, Arg, Lys) pourraient conduire à une fixation de cations métalliques de manière non spécifique. Un tel effet de fixation fortuite, non spécifique et non désirée a déjà été décrite dans la littérature.[17] La simulation du spectre laisse apparaitre une superposition de deux signaux. L'un contribuant à 83% du signal ($g_{//}$ 2,28 ; $g_{\perp}$ 2,06) et l'autre à 17% ($g_{//}$ 2,24 ; $g_{\perp}$ 2,06). Ces signaux semblent montrer que deux types et/ou deux sites de coordination de Cu(II) existent dans la protéine, ce qui va dans le sens d'une coordination aspécifique.

La même préparation a donc été injectée sur tamis moléculaire analytique afin de décrocher le Cu^{2+} qui serait coordiné avec une faible affinité, et la protéine éluée a été à nouveau analysée par RPE à la même concentration. Le spectre RPE donne un signal plat (figure 111, courbe jaune), prouvant l'absence de cuivre. Le cuivre a donc été éliminé par l'étape de filtration sur gel, ce qui confirme une faible affinité du cuivre pour la protéine sauvage. Nous avons ensuite voulu savoir si le biohybride fixait le cuivre Cu(II) de manière stable, pouvant résister à une étape de tamis moléculaire.

La protéine A3sF119NPH a donc été métallée avec 1 équivalent de cuivre par rapport au monomère. Après une période de métallation sur la nuit en chambre froide, le biohybride a été injecté sur tamis moléculaire. Le spectre RPE acquis pour le biohybride à la même concentration que la protéine sauvage présente un signal observable (figure 111, courbe grise). La protéine contenant le résidu NPH semble donc bien posséder une plus forte affinité pour le cuivre. La simulation de ce spectre semble indiquer un seul signal en solution ($g_{//}$ 2,25 ; $g_{\perp}$ 2,06), bien que le signal soit assez faible. On remarque que le signal est nettement inférieur à celui de la protéine sauvage. La double intégration des courbes de la figure 111 donne une indication sur la concentration en cuivre. En comparant le signal de la protéine sauvage avec le biohybride, on calcule que le biohybride représente 55% de la concentration de la protéine.

Si l'on admet que la protéine sauvage, malgré sa faible affinité, a fixé la plupart des ions cuivre métalliques de la solution et donc que le signal de la protéine sauvage présente les

200 µM de Cu(II) en solution, alors la concentration de cuivre dans la protéine A3sF119NPH représente 110 µM de cuivre pour 200 µM de monomère. Nous avons en conséquence émis l'hypothèse que le ratio véritable entre le cuivre et le variant A3sF119NPH était de 1:2 (Cu(II):Monomère). Afin de vérifier cette hypothèse, nous avons choisi d'analyser la quantité de métal présent dans le biohybride par une deuxième méthode quantitative, l'ICP-MS.

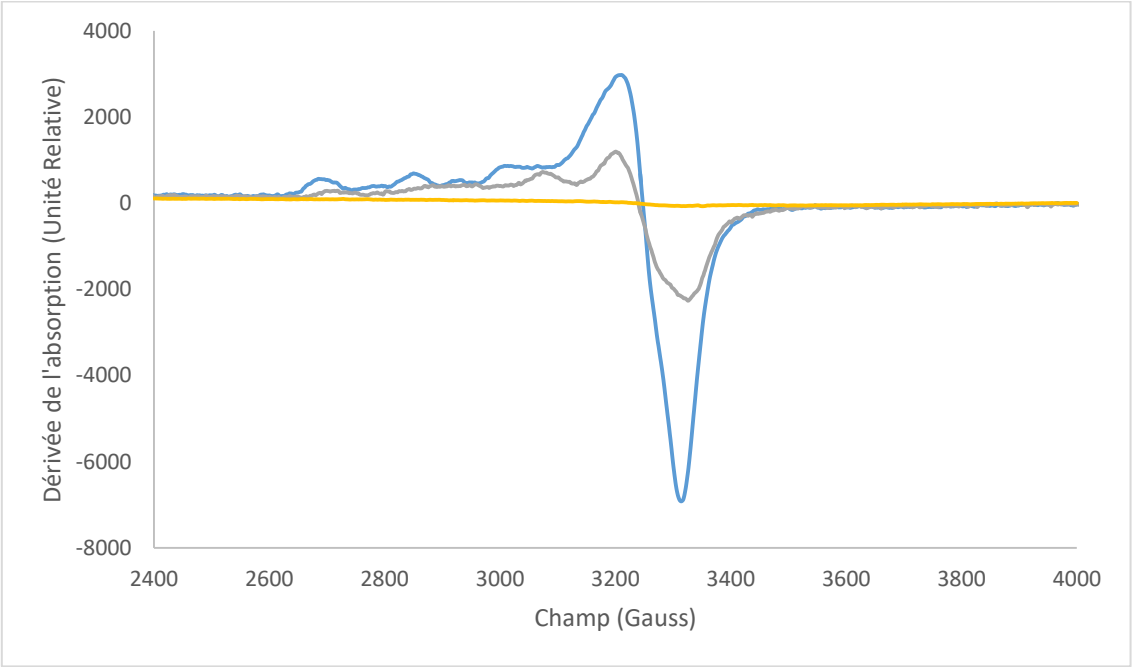


Figure 111 : Spectres RPE des variants A3s ou A3sWT. Bleu : protéine sauvage en présence d'un équivalent molaire de cuivre (100 µM). Jaune : même protéine à 100 µM après tamis moléculaire. On observe la disparition du signal du cuivre. Gris : spectre RPE du variant A3sF119NPH (100 µM) mis en présence d'un équivalent molaire de cuivre et après tamis moléculaire. On observe un signal environ correspondant à une quantité environ deux fois inférieure à un équivalent molaire de cuivre.

I.5.4 Spectrométrie de masse à plasma à couplage inductif (ICP-MS)

L'ICP-MS est une technique où l'échantillon est ionisé grâce à une torche à plasma puis analysé par spectrométrie de masse (ici avec séparation des ions selon leur temps de vol). La destruction de l'échantillon en particules élémentaires permet d'identifier et de quantifier avec une grande précision les métaux. L'avantage de ce type d'ionisation est de permettre la dissociation totale entre le métal et la protéine et de mesurer de manière fiable la quantité de métal dans l'échantillon. La mesure du tampon seul peut également nous éclairer sur la présence de contaminants.

La protéine a été préparée avec un excès de 1,5 équivalent de cuivre par rapport au monomère A3s (ratio 3:2 Cu(II):Monomère) afin de s'assurer que tous les sites soient occupés. Après incubation sur la nuit, la protéine a été traitée par tamis moléculaire

contre un tampon compatible avec la technique (phosphate 10 mM pH 7, 10 mM NaCl). La concentration de la protéine a été mesurée par absorbance à 280 nm afin de la comparer au dosage. Les résultats obtenus indiquent que pour une concentration de monomère A3sF119NPH mesurée de 39 µM, la concentration de cuivre dosée par ICP-MS est de 21 µM. Ce résultat confirme le ratio de deux monomères (et donc deux phénanthrolines) pour un ion Cu (II).

I.5.5 Effet du ratio Cu(II):Protéine sur la catalyse

Il a fallu confirmer par l'expérimentation que l'espèce catalytique permettant d'obtenir un excès énantiomérique correspondait bien au ratio de cuivre observé en RPE et ICP-MS. J'ai donc effectué des tests catalytiques avec différents ratios Cu(II):Monomère dans le tampon MOPS à pH 7, 150 mM NaCl avec 48h d'incubation. Dans cette expérience, la concentration de monomère est laissée constante (égale à précédemment) et le ratio de cuivre varie. L'entité catalytique étant sensé être une bis-phénanthroline de cuivre, j'ai également effectué les réactions avec cette dernière avec la même concentration que le monomère de la protéine A3sF119NPH. Les résultats sont représentés dans le tableau 112.

Monomère	Ratio Cu(II):Monomère	Tampon	pH	Durée	Cosolvant	*ee endo* (%)	Rendement (%)
-	0:-	MOPS	7	48h	-	3 ± 1	3 ± 1
	1:-					0 ± 2	31 ± 2
	2:-					1 ± 1	34 ± 1
phen	1:2					2 ± 1	14 ± 1
	2:2					1 ± 1	27 ± 1
A3sF119NPH	0:1					11 ± 4	4 ± 1
	1:2					**40 ± 5**	15 ± 1
	2:2					3 ± 1	26 ± 11
	3:2 + Tamis					**62 ± 3**	11 ± 1

Tableau 112 : Effet du ratio Cu(II):Monomère sur la catalyse par le variant A3sF119NPH. Catalyse de Diels-Alder entre l'azachalcone (1 mM) et le cyclopentadiène (34 mM) avec différents ratios de métal vs protéine (monomère). Conditions : 30 µM de monomère A3s du variant A3sF119NPH, 0 µM (0 équiv.), 15 µM (0,5 équiv.) ou 30 µM (1 équiv.) de $Cu(NO_3)_2$. 34 mM cyclopentadiène (1133 équiv.), 1 mM azachalcone (33 équiv.) dans CH_3CN (1% final), tampon MOPS 20 mM pH 7, 150 mM NaCl, 48h, 4°C. Toutes les expériences ont été faites en duplicat.

On peut tout d'abord remarquer qu'en l'absence de cuivre, le variant A3sF119NPH présente un léger excès énantiomérique. Cet excès est à relativiser car le rendement étant très bas (proche de celui du tampon sans cuivre), l'incertitude de la mesure est grande. Deux hypothèses sont néanmoins possibles : bien que les tampons soient traités à l'aide

de CHELEX 100, censé chélater les cations métalliques en solution, les dosages d'ICP-MS lors de l'étalonnage du tampon nous ont indiqué que des traces de métaux étaient toujours présentes. La deuxième hypothèse est que la protéine, possédant des acides aminés hydrophobes (dont le résidu artificiel NPH), peut induire une légère catalyse stéréosélective par interaction hydrophobe avec les substrats (effet dit « template »).

Il est également possible de remarquer que nous n'obtenons pas un rendement linéaire avec la quantité de cuivre libre en solution. Ce résultat à priori étrange est en fait en accord avec la cinétique effectuée plus haut. Nous avions alors remarqué que la réaction catalysée par le cuivre atteignait un plateau dans les premières 48h. La présence d'une quantité double de cuivre fait donc atteindre ce plateau plus rapidement.

On observe également qu'avec 1 équivalent molaire de cuivre le biohybride perd toute stéréosélectivité, la condition équivalente dans le mélange HEPES/MES donnant un excès énantiomérique de 30% (figure 108). Il est tout à fait probable que les tampons MES ou HEPES aient une action sur l'activité catalytique de la protéine, tout comme nous avons pu remarquer plus haut que le tampon HEPES impactait le signal du Cu(II) en RPE même au sein de la protéine. Des modifications de sélectivité dues au tampon sont régulièrement observées et il est nécessaire d'optimiser ce paramètre. Ce point sera abordé plus bas. À ce ratio, le complexe phénanthroline de cuivre comme le biohybride atteignent le rendement du cuivre en solution.

Finalement, nous pouvons observer qu'avec un ratio de 1:2 (Cu(II):monomère), l'excès énantiomérique est grandement augmenté. Le rendement est alors deux fois plus bas que le cuivre libre en solution à la même concentration, et proche de celui de la bis-phénanthroline de cuivre. La réaction est donc plus lente, mais énantiosélective.

Enfin, après avoir mis en présence un léger excès de cuivre (3:2 Cu(II):monomère) pendant la nuit, et procédé à une étape de tamis moléculaire, nous observons que l'excès énantiomérique est maximal (62% *ee*) et que le rendement est du même ordre que la protéine en présence d'un ratio 1:2 (Cu(II):monomère).

Ceci nous permet de conclure qu'une fois formé, le complexe stable d'une stœchiométrie Cu(II):Monomère de 1:2 est bien responsable de la catalyse énantiosélective. Nous pouvons alors proposer une interprétation pour les résultats obtenus en présence d'excès de cuivre, en absence d'étape de gel filtration : le cuivre en excès participerait à la catalyse en améliorant le rendement, par contre, il ne contribuerait pas à améliorer l'excès énantiomérique. Ce comportement équivalent à celui du cuivre libre (ou fixé dans un site non spécifique de la protéine) peut alors masquer l'effet du cuivre présent dans l'entité catalytique qui, elle, provoque un excès énantiomérique.

Il est à noter que, la concentration de cuivre effective étant finalement moitié moins que celle attendue, les équivalents de substrat ajoutés sont en fait doublés : 2267 équivalents de cyclopentadiène et 66 équivalents d'azachalcone.

I.5.6 Recherche de conditions optimales pour la catalyse

Maintenant que nous savons qu'il est nécessaire de ne pas dépasser le ratio 1:2 (Cu(II):monomère), j'ai pu chercher à optimiser les conditions de réaction en faisant varier d'autres paramètres physicochimiques tels que la nature du tampon, la température ou les co-solvants.

I.5.6.1 Effet du tampon

Trois différents tampons ont été testés dans l'espoir d'améliorer l'énantiosélectivité de la catalyse. Les résultats sont résumés dans le tableau 113. Nous observons tout d'abord que les rendements du cuivre (II) sont plus bas que d'ordinaire, et du même ordre de grandeur voir au-dessous de ceux du biohybride. Ce résultat, notamment concernant le MOPS qui est un réplicat de l'expérience ci-dessus, est pour l'instant inexpliqué. Les résultats concernant le biohybride sont néanmoins cohérents entre eux et peuvent donc être analysés.

Monomère	Ratio Cu(II):Monomère	Tampon	pH	Durée	Cosolvant	*ee endo* (%)	Rendement (%)
-	0:-	MOPS	7	48h	-	2 ± 3	2 ± 1
	1:-					1 ± 1	7 ± 4
	0:-	NaP				2 ± 3	2 ± 1
	1:-					1 ± 3	6 ± 1
	0:-	HEPES				3 ± 2	2 ± 1
	1:-					2 ± 1	8 ± 2
A3sF119NPH	0:2	MOPS				18 ± 2	2 ± 1
	1:2					**57 ± 2**	8 ± 1
	0:2	NaP				12 ± 4	3 ± 1
	1:2					**53 ± 1**	10 ± 1
	0:2	HEPES				11 ± 6	2 ± 1
	1:2					**53 ± 2**	9 ± 1

Tableau 113 : Effet du tampon sur la catalyse par le variant A3sF119NPH. Catalyse de Diels-Alder entre l'azachalcone (1 mM) et le cyclopentadiène (34 mM) avec différents tampons. Conditions : 30 µM de monomère A3s du variant A3sF119NPH, 15 µM (0,5 équiv.) de $Cu(NO_3)_2$. 34 mM cyclopentadiène (2267 équiv.), 1 mM azachalcone (66 équiv.) dans CH_3CN (1% final), tampon MOPS 20 mM pH 7,HEPES 20 mM pH 7, Na-Phosphate 50 mM pH 7, tous les tampons contenant 150 mM NaCl, 48h, 4°C. Toutes les expériences ont été faites en duplicat. La métallation est faite la veille de l'expérimentation, la protéine étant dans le tampon MOPS.

Concernant le biohybride, nous pouvons observer d'une manière générale que la nature du tampon semble peu affecter le rendement. Concernant la sélectivité, nous pouvons observer une légère baisse dans les tampons phosphate et HEPES. De manière intéressante, le tampon HEPES ne semble donc pas beaucoup affecter la sélectivité somme il était soupçonné précédemment. Tout porte donc à croire que c'est le tampon MES qui agirait d'une manière négative. Il est néanmoins à savoir que, dans un soucis de reproductibilité, la métallation s'est effectuée dans le tampon MOPS et que le mélange réactionnel en contient donc à hauteur de 17%. Il se peut donc que la métallation soit l'étape critique pour la sélectivité du biohybride, et qu'une fois le complexe formé il est alors plus stable dans différentes conditions. N'ayant pas obtenu d'amélioration, nous avons décidé de tester les autres paramètres que sont la température et le cosolvant.

Les prochaines catalyses ont été effectuées avec un léger sous-excès de cuivre (0,6:2 Cu(II):monomère) par rapport à un demi-équivalent molaire, afin d'assurer la formation la plus complète du supposé complexe de bis-phénanthroline de cuivre. La condition standard utilisée jusqu'ici étant une température de 4°C et pas de cosolvant ajouté (à part le 1% d'acétonitrile final correspondant à l'ajout du substrat).

I.5.6.2 Effet de la température

Trois températures été utilisés : 4°C, 15°C et 25°C. Les résultats de catalyses sont donnés dans le tableau 114.

Monomère	Ratio Cu(II):Monomère	Tampon	pH	Durée	Température	Cosolvant	*ee endo* (%)	Rendement (%)
A3sF119NPH	0:2	MOPS	7	48h	4°C	-	22 ± 11	5 ± 1
					15°C		21 ± 2	3 ± 1
					25°C		8 ± 1	2 ± 1
	0,6:2				4°C		**60 ± 2**	12 ± 2
					15°C		**64 ± 2**	8 ± 1
					25°C		45 ± 17	3 ± 1

Tableau 114 : Effet de la température sur la catalyse par le variant A3sF119NPH. Catalyse de la réaction de Diels-Alder entre azachalcone et cyclopentadiène dans différentes conditions de température. Catalyseur A3sF119NPH incubé ou non avec 0,3 équiv. de $Cu(NO_3)_2$ (par rapport au monomère A3s) La condition à 4°C est celle utilisée précédemment. 34 mM cyclopentadiène (3778 équiv.), 1 mM azachalcone (110 équiv.) dans CH_3CN (1% final), tampon MOPS 20 mM pH 7, 150 mM NaCl, 48h. Toutes les expériences ont été faites en duplicat.

Une augmentation de la température sur le rendement de réaction semble néfaste : partant de 7,2% à 4°C, le rendement du biohybride décroit jusqu'à 1,6% à 25°C. Ceci est probablement dû à la dimérisation spontanée du cyclopentadiène, qui diminue la quantité de réactif disponible pour la réaction. Ceci est également valable pour la protéine non métallée dont le rendement, bien que déjà faible, passe de 2,6% à 1,2% entre 4°C et 25°C.

Cet effet de dimérisation du substrat a déjà été décrit et justifie pourquoi la plupart des études de catalyse de la réaction de Diels-Alder sont effectuées à 4°C [18].

Concernant l'effet de la température sur l'excès énantiomérique, on observe tout d'abord que la protéine non métallée possède toujours une sélectivité (soit par contamination par le cuivre, soit par effet « template », comme expliqué précédemment). Partant d'un excès énantiomérique d'environ 22% à 4°C, la sélectivité décroit jusqu'à 8% à 25°C. Concernant le biohybride, la sélectivité reste haute (60-64% à 4-15°C) même si elle semble décroitre à 25°C pour atteindre 45%, ce chiffre est à relativiser, la marge d'erreur étant grande à cette température.

Finalement, une augmentation de la température ne permet pas d'améliorer l'excès énantiomérique et conduit par aussi à une baisse du rendement de la réaction. La température de 4°C précédemment utilisée est donc bien la condition de température optimale.

I.5.6.3 Effet des cosolvants

Les trois cosolvants, pouvant favoriser la solubilité des substrats, ont été choisis pour leur coefficient de partage octanol/eau différents (logP de -0,3 ; 0,05 et 0,4 pour acétonitrile, isopropanol et tert-butanol respectivement). Les résultats sont résumés dans le tableau 115.

Monomère	Ratio Cu(II):Monomère	Tampon	pH	Durée	Température	Cosolvant	*ee endo* (%)	Rendement (%)
A3sF119NPH	0:2	MOPS	7	48h	4°C	-	22 ± 11	5 ± 1
						5% ACN	11 ± 5	6 ± 1
						10% ACN	6 ± 3	7 ± 0
						5% IsoOH	7	4
						10% IsoOH	8 ± 4	5 ± 1
						5% t-ButOH	9 ± 7	5 ± 1
						10% t-ButOH	4 ± 2	6 ± 1
	0,6:2					-	**60 ± 2**	12 ± 2
						5% ACN	51 ± 1	13 ± 2
						10% ACN	36 ± 2	11 ± 2
						5% IsoOH	47 ± 4	10 ± 1
						10% IsoOH	47 ± 1	8 ± 1
						5% t-ButOH	52 ± 1	11 ± 1
						10% t-ButOH	47 ± 2	9 ± 1

Tableau 115 : Effet des cosolvants sur la catalyse par le variant A3sF119NPH Catalyse de la réaction de Diels-Alder entre azachalcone et cyclopentadiène dans différentes conditions de cosolvant. Catalyseur A3sF119NPH incubé ou non avec 0,3 équiv. de $Cu(NO_3)_2$ (par rapport au monomère A3s). Dans toutes les conditions de cosolvant, 1% d'acétonitrile est présent en plus, provenant de la solution de substrat. 34 mM cyclopentadiène (3378 équiv.), 1 mM azachalcone (110 équiv.) dans CH_3CN (1% final), tampon MOPS 20 mM pH 7, 150 mM NaCl, 48h, 4°C. Toutes les expériences ont été faites en duplicat.

Concernant l'effet des cosolvants, les tendances sont comparables à celles pour la température : chez la protéine avec ou sans métal la sélectivité et le rendement semblent affectés négativement, quel que soit le type de cosolvant utilisé. Nous avions vérifié par des expériences de dichroïsme circulaire que la protéine restait repliée en hélices α en présence de 10% d'acétonitrile, ces mesures ayant été faites 24h heures après l'ajout de solvant. Cependant, il est possible que la présence de solvant organique modifie les propriétés dynamiques de la protéine sans pour autant affecter ses structures secondaires, qui perd donc sa capacité à générer un environnement propice à la catalyse énantiosélective.

La condition la plus favorable à une catalyse énantiosélective est donc un taux minimal de cosolvant.

I.5.6.4 Effet des concentrations en substrats

Fort de ces résultats, j'ai tenté de déterminer si la concentration en substrats pouvait avoir un impact positif sur l'excès énantiomérique ou le rendement, l'hypothèse principale étant une inhibition par le produit. J'ai donc fait varier les deux paramètres indépendamment en diminuant de moitié ou du quart la quantité ajoutée de chacun des substrats. Les concentrations finales sont donc de 1 mM ; 0,5 mM ; 0,25 mM d'azachalcone et 33,8 mM ; 16,9 mM ; 8,45 mM de cyclopentadiène. Les résultats sont reportés dans le tableau 116.

Monomère	Ratio Cu(II):Monomère	Tampon	pH	Durée	Cosolvant	[azachalcone] (mM)	[cyclopentadiene] (mM)	*ee endo* (%)	Rendement (%)
A3sF119NPH	3:2 + Tamis	MOPS	7	48h	-	1	34	40 ± 2	28 ± 4
						0,5	34	52 ± 1	52 ± 11
						0,25		**60**	56
						1	17	43 ± 1	8 ± 2
							8,5	43 ± 2	5 ± 1

Tableau 116 : Effet de la concentration en substrats sur la catalyse par le variant A3sF119NPH. Catalyse de la réaction de Diels-Alder entre azachalcone et cyclopentadiène par le biohybride A3sF119NPH métallé puis dessalé par tamis moléculaire. Condition normale : 34 mM cyclopentadiène (2266 équiv.), 1 mM azachalcone (66 équiv.) dans CH_3CN (1% final), tampon MOPS 20 mM pH 7, 150 mM NaCl, 48h, 4°C. Toutes les expériences ont été faites en duplicat.

Nous pouvons tout d'abord remarquer que dans les conditions normales (identiques à la ligne « 1,5 équiv. + tamis » de la figure 112), l'excès énantiomérique est plus faible (40% contre 62%) tandis que le rendement est plus élevé (28% contre 11%). Cette disparité n'est pas encore tout à fait expliquée et peut être due à un lot plus récent du substrat azachalcone ainsi qu'à une nouvelle préparation de la protéine.

Lorsque l'on baisse la concentration en azachalcone, le rendement et l'excès énantiomérique se trouvent augmentés. L'augmentation du rendement peut s'expliquer

par une inhibition par le produit, étant donné qu'à rendement équivalent la concentration de produit dans le milieu réactionnel est plus faible.

La baisse de la concentration de cyclopentadiène ne semble pas impacter l'excès énantiomérique de la réaction catalysée par le biohybride (43% contre 40% *ee*). En revanche, le rendement de la réaction se trouve très fortement abaissé (de 28% à 8% puis 5% pour des concentrations respectives en cyclopentadiène de 33,8 mM, 16,9 mM et 8,45 mM). Cette baisse de rendement dénote une mauvaise affinité entre la protéine et le cyclopentadiène. Nos résultats confortent un mécanisme réactionnel en deux étapes décrit précédemment [13] : tout d'abord la fixation de l'azachalcone sur le cuivre dans une position privilégiée par l'environnement protéique favorisant l'exposition d'une face donnée de la molécule vers le cyclopentadiène.

I.5.7 Modèle proposé pour le biohybride A3sF119NPH_Cu

Compte tenu de l'ensemble de ces résultats, nous soumettons donc l'hypothèse selon laquelle le centre catalytique au sein du biohybride A3sF119NPH serait une bis-phénanthroline de cuivre. Le dimère étant symétrique, les deux monomères devraient se comporter de façon équivalente face à la fixation d'un objet achiral : si les deux phénanthrolines ne chélatent qu'un seul cuivre, il est probable qu'elles le fassent de la même façon. Dans cette configuration, l'entité catalytique formerait une symétrie de 180° autour d'un axe passant par le cuivre. Le site catalytique demeure chiral, mais l'accès est alors possible au substrat dans deux orientations.

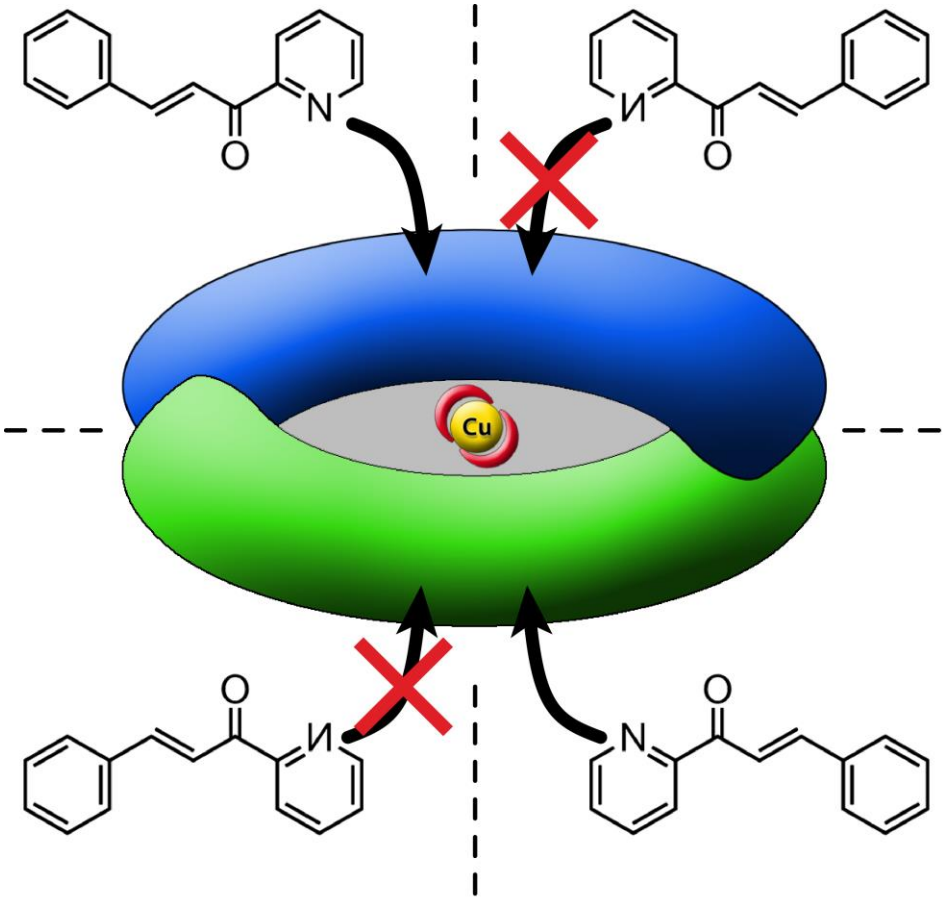


Figure 117 : Modèle proposé pour la structure de l'entité catalytique de A3sF119NPH en présence de cuivre. Ce catalyseur présentant une symétrie cristallographique de type C2, l'environnement resterait chiral avec cependant deux sens d'arrivée du substrat possibles après une rotation de 180°. Le cyclopentadiène viendrait après la fixation de l'azachalcone pour former la forme endo ou exo correspondante.

I.5.8 Structure cristallographique de la phénantroline de cuivre

J'ai par ailleurs obtenu des cristaux du complexe $[Cu(1,10\text{-phénanthroline})_2(NO_3)]^+$ (selon une méthode dérivée de [19]) dont le Dr Régis Guillot a déterminé la structure cristallographique (figure 118). L'unité asymétrique du cristal contient le complexe cationique le contre-ion NO_3^-, ainsi qu'une molécule de solvant méthanol (MeOH). Ce type de complexe est déjà décrit dans la littérature[20–22], avec cependant différents contre-ions ($[AuBr_2(CN)_2]^-$, $C_4N_3^-$ (CCDC : EDUPOK[22]) ou OH^- (CCDC : PEYWEW[20])). La structure avec le contre-ion tricyanomethanide ($C_4N_3^-$) est décrite comme possédant une géométrie tétragonale-bipyramidale asymétrique, les positions équatoriales étant occupées par les deux atomes d'oxygène du ligand nitrato ainsi qu'un atome d'azote de chacune des ligands *phen*. Les positions axiales sont occupées par le deuxième atome d'azote des ligands *phen*. Cette description correspond à ce que nous observons dans notre structure cristallographique. Dans chacune des structures disponibles une différence de distance Cu-O est présente (2.154(2) et 2.586(2) Å dans la structure EDUPOK, 2,014(2) et 2,332(2) dans la structure PEYWEW), dans notre structure les distances sont 2.156(2) Å et 2.760(2) Å, le cuivre est donc seulement partiellement coordiné par le deuxième atome d'oxygène. La structure possédant le contre-ion OH^- (CCDC : PEYWEW) est également intéressante car elle présente le complexe sous la forme hexahydratée, la géométrie est alors décrite comme un octaèdre déformé. En comparant les angles des plans formés par les carbones des phénanthrolines, on obtient les valeurs suivantes : 57° (UYEFOU), 77° (EDUPOK), 48° (notre travail). Ces différentes structures illustrent la flexibilité de ce type de complexe en fonction de son environnement, il n'est donc pas possible de prévoir une géométrie attendue au complexe inséré dans la protéine.

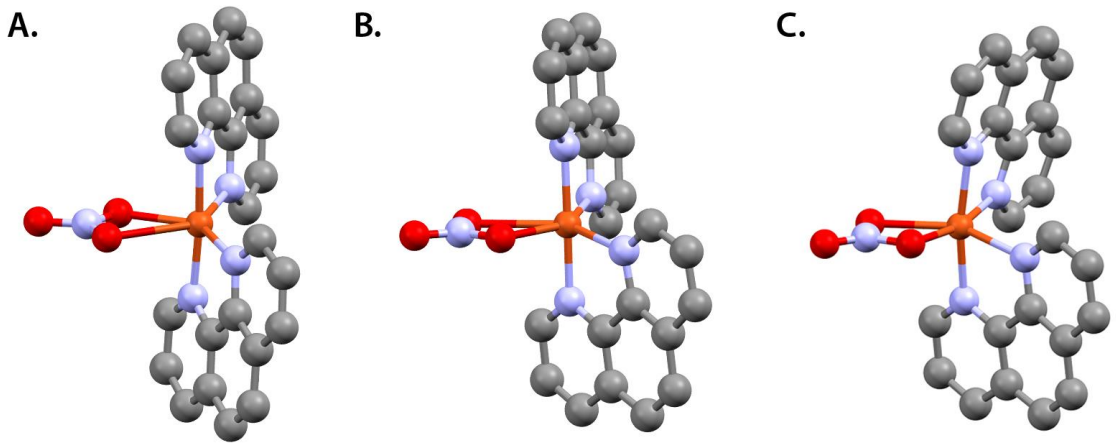


Figure 118 : Structure cristallographique du complexe $[Cu(phen)_2(NO_3)]^+$, les molécules de contre-ion et solvants ont été masquées pour plus de clarté. A. Structure issue de notre travail, contre-ion NO_3^-, solvant MeOH. B. Structure EDUPOK, contre-ion $C_4N_3^-$, pas de molécule de solvant. C. Structure PEYEW, contre ion OH^-, solvant $(H_2O)_6$. Images réalisées avec le programme Mercury.

Il est à noter que le complexe [Cu(1,10-phenanthroline)$_2$(NO_3)]$^+$ étant chiral, deux orientations sont présentes dans le cristal, il est cependant possible qu'une des deux configurations sont sélectionnée au sein de la protéine étant donné que la métallation est effectuée après le greffage du ligand.

Il ne sera possible de confirmer la présence d'une bis-phénanthroline de cuivre qu'avec des expériences de RPE plus poussées ou la structure cristallographique du biohybride.

I.5.9 Comparaison avec la littérature

Biohybride [a]	**Conversion (%)**	***endo/exo* (%:%)**	***ee* (%)**	**ref**
LmrR_M89C-Phen[b]	93	95:5	97	40
SB-Lys-GTL*/196-Phen[b]	98	93:7	93,5	41
HSA-phtalocyanine	89	91:9	85	42
A3sF134C-Phen[b] (A3sF134NPH)	11	86:14	62	ce travail
NCS-Phen[b]	84	84:16	0	43
NB-Tpy[c]	22	90:10	90	44
FhuA-Tpy[c]	15	66:34	ND[e]	45
FhuA-NHC[d]	62	90:10	ND[e]	36

Tableau 119 : Comparaison des résultats du biohybride A3sF134NPH avec la littérature actuelle. [a] Nomenclature utilisée pour les protéines : LmrR, Lactococcal multidrug resistance Regulator ; SBSB-Lys-GTL/196, Sepabeads Lys196 mutant of Geobacillus thermocatenulatus lipase; HSA, human serum albumin; A3, aRep A3 protein; SCP-2L, sterol carrier protein type 2-like domain; NCS, neocazinostatin; NB, nitrobindin; FhuA, ferric hydroxamate uptake protein component A. [b] Phen, phénanthroline. [c] Tpy, terpyridine. [d] NHC, carbène N-heterocyclique. [e] Non déterminé.*

Dans le tableau 119 sont reportés les taux de conversion, compositions endo/exo et excès énantiomériques obtenus par les différents biohybrides dans la littérature pour la même réaction de Diels-Alder. Comparant le rendement de la réaction, notre biohybride donne un rendement plus bas que les autres catalyseurs biohybrides, qui s'échelonnent de 15% pour le biohybride FhuA-Tpy-Cuivre(II) [45] jusqu'à 98% pour le biohybride SB-Lys-GTL*/196-Phen-Cuivre(II) [41]. Ce résultat peut s'expliquer soit par un accès limité des substrats envers le centre catalytique (dûe à une faible hydrophobicité de la crevasse), ou par une inhibition par les substrats ou produits de la réaction. On remarque également que l'excès énantiomérique donné par notre biohybride (62%) se démarque de ceux observés dans la littérature, qui sont soit très haut (de 85% à 97%) soit très bas (0%). Ce résultat peut être expliqué par une plasticité du biohybride ou par différentes conformations du complexe de bis-phénanthroline de cuivre(II) au sein de la crevasse. Ce complexe étant chiral, il est possible qu'une forme de ce complexe au sein de la crevasse catalyse la réaction de façon plus énantiosélective que l'autre forme. Il n'est cependant pas possible de confirmer ou infirmer cette hypothèse à l'aide des données dont nous possédons actuellement.

I.6 Analyse structurale des différents variants

La connaissance de l'environnement protéique autour de la phénanthroline pouvant être d'une grande aide pour trouver les résidus constituant la « seconde sphère de coordination » des substrats, j'ai donc entrepris d'obtenir la structure cristallographique avec les conseils du Dr Inès Gallay. Le dimère αRep-A3 présente également une capacité à former des cristaux dans un large éventail de conditions.

Il est tout d'abord apparu que les biohybrides cristallisaient bien moins aisément que la protéine sauvage, avec une grande propension à former des sphérulites (figure 120). Je suis néanmoins parvenu à obtenir un certain nombre de cristaux, notamment pour les variants A3sY26NPH et A3sD150NPH.

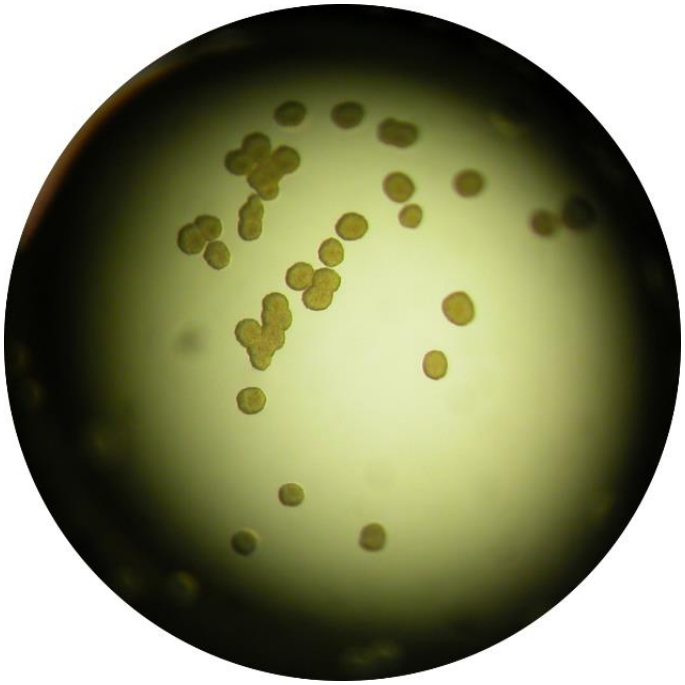
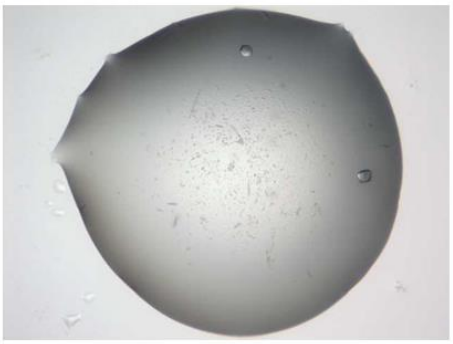


Figure 120 : Formation de sphérulites, phénomène couramment observé lors des essais de cristallisation des différents variants.

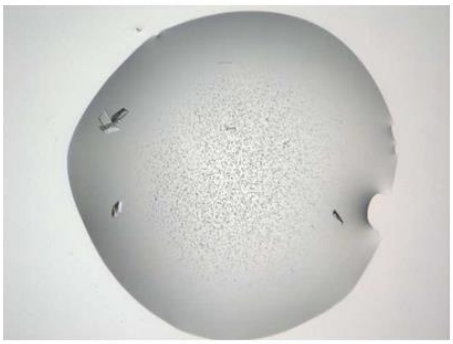
I.6.1 Structures du variant A3sY26NPH

A.



Sel	Tampon	pH	Précipitant
0,1 M NaCl	0,1 M Citrate de sodium	5,6	18 % (w/v) PEG 400

B.



Sel	Tampon	pH	Précipitant
	0,1 M Acétate de sodium	4,6	2 M Sulfate d'ammonium

Figure 121 : Cristaux obtenus en kit de criblage de cristallisation pour le variant A3sY26NPH contenant du zinc. Les conditions sont indiquées en dessous des images.

Plusieurs types de cristaux ont été obtenus pour le variant A3sY26NPH. Un premier cristal provenant d'une préparation contenant du zinc (figure 121A a été analysé sur la ligne PROXIMA1 du synchrotron SOLEIL. Nous avons pu obtenir des clichés à une résolution de 1,9 Å. Après traitement par remplacement moléculaire à l'aide de la structure de la protéine A3 sauvage, nous avons vite constaté que la structure était très proche de celle de la protéine non modifiée. Cependant la densité électronique se limitait uniquement à un résidu cystéine. Il semblerait donc que le peu de protéine non couplée à la phénanthroline présent dans cette préparation ait cristallisé. Nous n'avons donc pas continué l'affinement de la structure.

Un autre cristal (figure 121B) a été observé pour une préparation de protéine également en présence de zinc. Ce cristal a mis plusieurs mois à apparaitre (entre le 35^{e} jour et le 205^{e} jour). Après diffraction à 2 Å sur la ligne PROXIMA1 également, j'ai tenté de résoudre la structure. Un remplacement moléculaire par la protéine A3 sauvage a tout d'abord révélé que la structure du dimère était clairement modifiée. Comme nous pouvons voir sur la figure 122, les deux monomères sont décalés par un glissement des sous-unités encore plus important que celui de la forme « ouverte » de la protéine contenant la molécule de PEG.

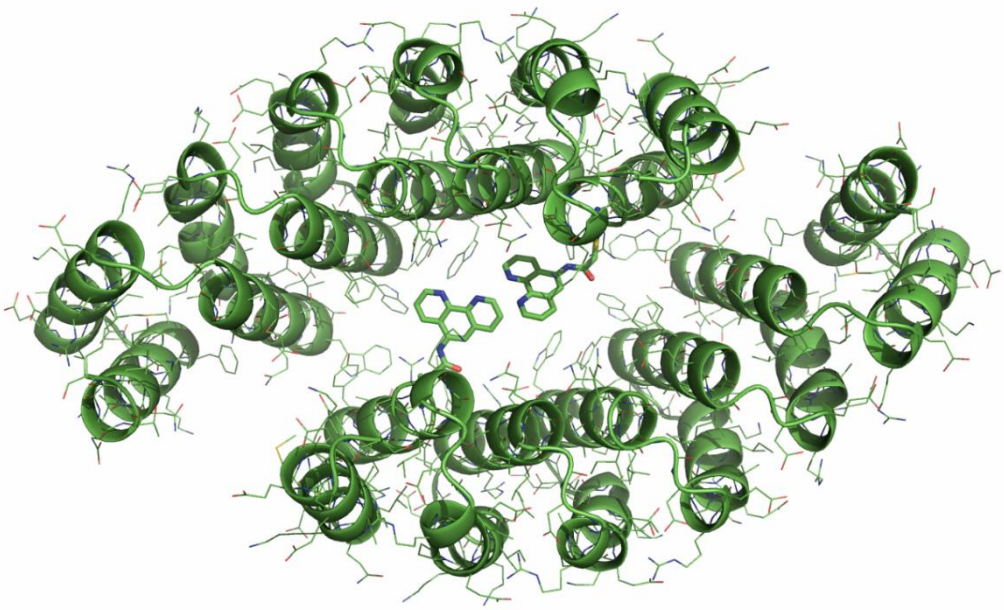
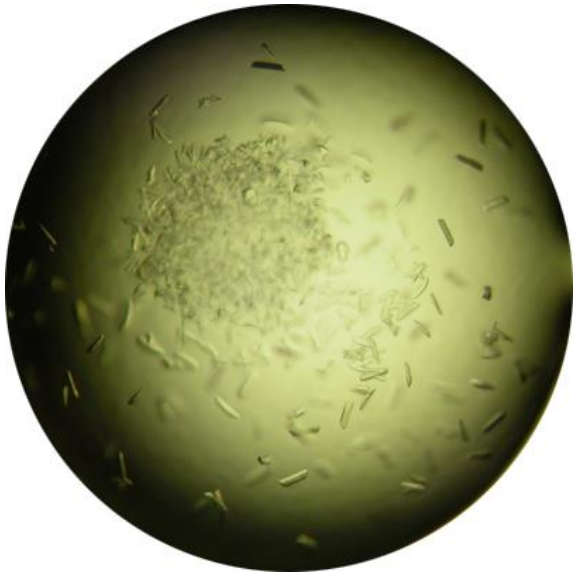


Figure 122 : Structure cristallographique du variant A3sY26NPH en présence de zinc. Les HisTag ont été retirés de la structure pour plus de clarté. La position des phénanthrolines est peu précise, la densité électronique dans cette zone étant particulièrement diffuse.

Nous pouvons conclure de ce résultat que le dimère αRep-A3, bien que très stable, présente une flexibilité inattendue. Il est donc possible que les différents biohybrides puissent présenter un glissement des sous-unités analogue, un ion métallique faisant le pont entre les deux sous-unités. Il faut cependant rester prudent concernant la possibilité de formation d'un complexe bis-phénanthroline de cuivre lors des expériences de catalyse concernant le variant A3sY26NPH. Tout d'abord, la structure cristallographique a été résolue pour la protéine en présence de zinc (des tentatives d'obtention de cristaux pour la forme avec cuivre n'ayant pas été concluantes). Ensuite, après une expérience de RPE menée sur ce variant ayant subi une étape de filtration sur gel, la double intégration du signal mesuré donnait une valeur intermédiaire entre le variant A3sF119NPH et la protéine sauvage en présence de cuivre. Il n'est donc pas à exclure la présence de deux populations différentes, passant progressivement d'un état de mono phénanthroline de cuivre à bis phénanthroline de cuivre après glissement des deux monomères. Ceci est en accord avec le fait que les cristaux aient mis un grand laps de temps avant d'apparaître : la population de forme bis-phénanthroline de zinc a probablement augmenté avec le temps jusqu'à atteindre un seuil de nucléation.

I.6.2 Structure du variant A3sD150NPH



Sel	Tampon	pH	Précipitant
0,45 M NaCl	0,1 M Phosphate de sodium	6,4	15 % (w/v) PEG 2000

Figure 123 : Cristaux obtenus après optimisation pour le variant A3sD150NPH en présence de zinc, les conditions de cristallisation sont notées en dessous.

Des cristaux du variant A3sD150NPH ont été également obtenus dans une condition en présence de zinc (figure 123). Après diffraction à une limite de 2 Å sur la ligne ID23 du synchrotron de l'ESRF de Grenoble, j'ai également effectué l'analyse par remplacement moléculaire en utilisant le monomère αRep-A3. On remarque une structure générale extrêmement proche de la forme « ouverte » de la protéine (PDB : 3LTM) et pour cause : la protéine semble également contenir une molécule de PEG au sein de la grande crevasse (non montré sur la figure).

La position D150 avait été choisie car le résidu était pointé non pas vers la crevasse principale mais vers une crevasse secondaire. La structure cristallographique confirme que la phénanthroline est bien dirigée vers cette crevasse (en rose, figure 124). Cependant aucun métal n'est visible dans la structure : la phénanthroline forme effectivement une liaison hydrogène avec la seule histidine de la protéine située à l'extrémité C-teminale.

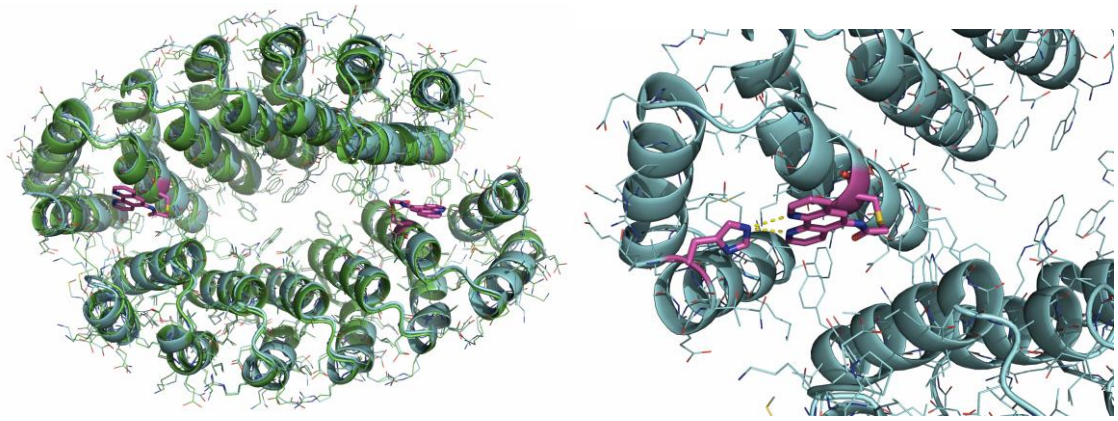


Figure 124 : Gauche : Alignement des structures de la protéine A3 (en vert, PDB : 3LTM) et de la structure déterminée pour le variant A3sD150NPH (en bleu). La protéine semble également contrenir une molécule de PEG provenant de la solution de cristallisation, à l'instar de la structure de la protéine sauvage (PDB : 3LTM) La phénanthroline est représentée en violet. Droite : vue de près de l'environnement de la phénanthroline, montrant les liaisons hydrogènes qu'elle effectue avec une histidine à la place d'un ion zinc.

I.6.3 Cristaux du variant A3sK30NPH

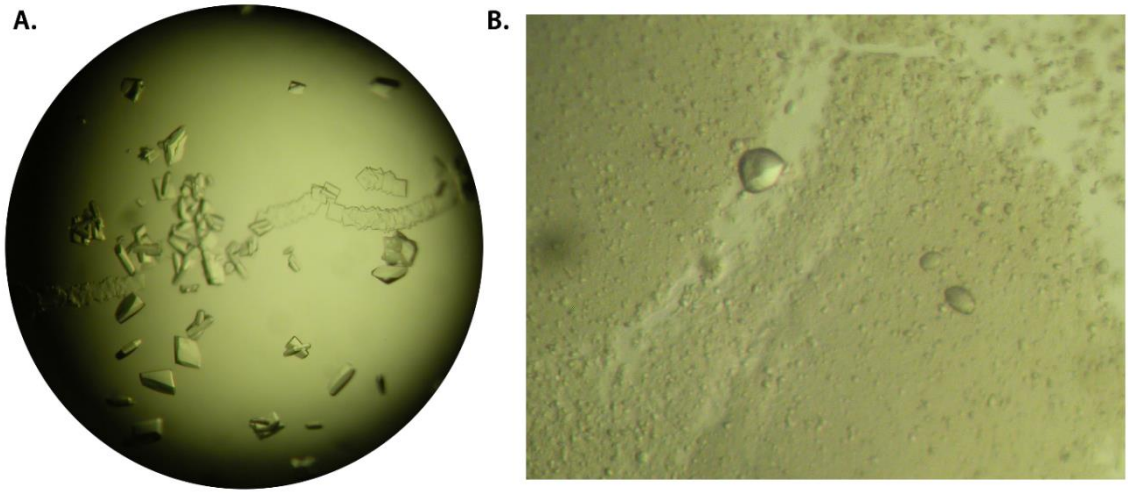


Figure 125 : A. Cristaux du variant A3sK30NPH en présence de zinc, aucun jeu de données en dessous de 7 Å de diffraction n'a pu être obtenu. B. Cristaux du variant A3sK30NPH en présence de cuivre, aucune diffraction n'a été obtenue.

Plusieurs types de cristaux ont été obtenus pour le variant A3sK30NPH (en présence de zinc ou cuivre). Cependant la résolution n'a pas pu être descendue en dessous de 7 Å, ce qui ne permettait pas d'avoir d'informations sur la position de la phénanthroline.

I.7 Résumé de l'étude du dimère A3s

Nous avons vu durant cette étude que la protéine αRep-A3 pouvait être modifiée par greffage covalent d'une phénanthroline sur des résidus cystéines placés dans des positions choisies au sein de la crevasse générée par le dimère. Les protéines ainsi modifiées que nous appelons biohybrides ont conservé leurs structures secondaires et restent dimériques en solution. Les conditions de couplages ont été mises au point pour un rendement de couplage quantitatif. Parmi les différents variants, A3sF119NPH est le meilleur candidat pour la catalyse énantiosélective de la réaction de cycloaddition de Diels-Alder entre l'azachalcone et le cyclopentadiène, permettant d'atteindre jusqu'à 60% d'excès énantiomérique. L'ensemble de données de RPE, d'ICP-MS et de catalyse apportent un faisceau d'indices nous permettant de proposer un modèle de biohybride dans lequel la catalyse énantiosélective serait assurée par un complexe bis-phénanthroline de cuivre au sein du dimère, et non par un complexe binucléaire comme attendu au départ de notre étude. Un tel complexe a été décrit dans la littérature comme catalysant des réactions d'hydrolyse de l'ADN, dont il serait intéressant de tester une éventuelle spécificité induite par la protéine.[23]

Plusieurs explications sont possibles quant à l'énantiosélectivité plus élevée du variant A3sF119NPH. La présence d'un complexe bis-phénanthroline de cuivre provoque un pontage entre les deux sous-unités. Il est possible que ce double ancrage diminue la mobilité du complexe au sein de la crevasse et de fait améliore l'énantiosélectivité, comme il a été observé dans les travaux de l'équipe de Yi Lu.[24] Ce double ancrage n'est *a priori* pas présent dans le cas des autres biohybrides, ceci peut expliquer leur plus basse énantiosélectivité. Cependant un point vient à l'encontre de cette explication : la structure du variant A3sY26NPH en présence de zinc présente un glissement qui semble conduire à la formation d'un complexe bis-phénanthroline de zinc, pourtant le nuage électronique autour des ligands est très diffus, signe d'une mobilité importante. Un complexe bis-phénanthroline ne serait donc pas forcément signe de stabilité.

Plusieurs autres points sont encore à éclaircir. Nous avons remarqué des variations pouvant atteindre jusqu'à 20% concernant les rendements et excès énantiomériques entre les différents lots de substrat ou de protéine, bien que les excès énantiomériques restent cohérents. Ces différences pourraient être dues à une instabilité au cours du temps de nos réactifs, ou à une présence de contaminants entre les différentes préparations de protéines. Il est donc nécessaire d'apporter une attention particulière aux conditions de conservation, notamment de l'azachalcone.

Enfin, bien que les essais de cristallographie n'aient pas permis de résoudre la structure du biohybride catalyseur avec précision, elles nous ont permis de mettre en évidence la

plasticité du dimère A3, tout comme dans le cas du variant A3sY26NPH en présence de zinc, avec d'une part un glissement des sous-unités protéiques l'un par rapport à l'autre et d'autre part une mobilité des groupements phénantrolines au sein de la crevasse. Cette importante plasticité du dimère αRep-A3 est un inconvénient pour la conception d'une banque de biohybrides : la configuration de la crevasse étant mal contrôlée, il serait difficile de concevoir une banque de petite taille présentant des cavités. De plus la présence d'une bis-phénanthroline oblige *a priori* à respecter une symétrie C2 centrée sur le métal, ce qui diminue les possibilités de créer une crevasse asymétrique.

Afin de tenter de contrôler la formation de la cavité et de générer des biohybrides plus performants, nous avons décidé d'utiliser une nouvelle construction génétique permettant la production d'une unique protéine à deux domaines. Cette protéine que nous appellerons chimère A3sA3, est constituée des deux monomères de l'αRep-A3 reliés par une séquence peptidique. L'étude de nouveaux biohybrides issus de cette chimère sera décrite dans le chapitre suivant.

Chapitre II - Utilisation de la protéine αRep-A3sA3

Après avoir constaté que l'insertion de complexes métalliques au sein de la protéine αRep-A3 pouvait conduire de manière fortuite à un glissement des monomères pouvant être induit par exemple par la formation de bis-phénantroline de cuivre, nous avons décidé d'explorer une nouvelle piste en ne greffant qu'un seul ligand au sein de la crevasse. Le problème de la symétrie de l'homodimère se posait alors, car elle impose naturellement que le point de greffage (la cystéine) soit présent de chaque côté de la crevasse. Il était donc nécessaire de ne plus utiliser un dimère αRep-A3 mais une protéine bi-domaine. Les deux monomères de la protéine A3 (dorénavant appelés « domaines ») étant reliés par une liaison peptidique encodée génétiquement (appelée « *linker* »), la protéine était alors asymétrique et il était à présent possible de modifier l'un des deux domaines sélectivement. Un autre atout issu de la liaison des deux domaines est l'augmentation de la concentration apparente des deux domaines l'un pour l'autre [25], diminuant considérablement le risque d'échange de domaines entre deux dimères distincts (figure 126).

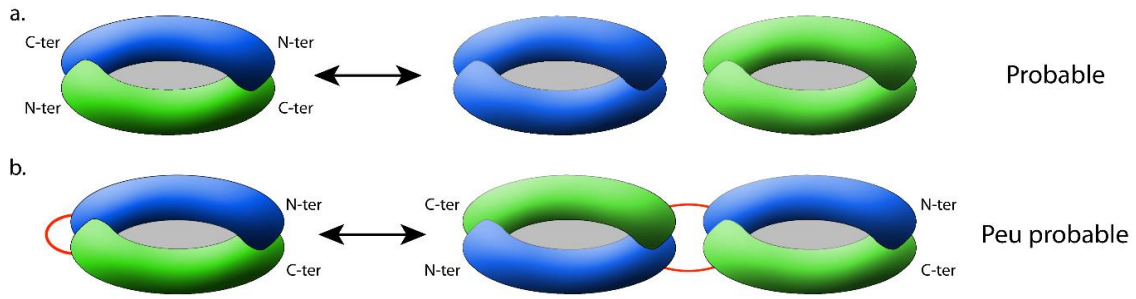


Figure 126 : a. Schéma illustrant l'échange des monomères du dimère A3. Les vitesses d'association et de dissociation du dimère n'étant pas connues, il n'est pas possible de prédire la vitesse de cet échange. b. Schéma illustrant la nouvelle protéine αRep-A3sA3 faite de deux protéines A3 reliées par un linker encodé génétiquement (en rouge). Ainsi liées, l'échange de domaines est très peu probable du fait de la grande concentration apparente entre les deux domaines liés.

II.1 Construction génétique codant la protéine A3sA3

Un autre problème se posait pour l'obtention de gènes mutés sur un seul des deux domaines. En effet, la méthode de mutagénèse « QuikChange » utilise l'appariement des amorces de PCR sur la séquence ciblée, un gène dupliqué risquait donc de provoquer une mutation dans chacun des gènes. Il a donc été décidé de construire le gène codant la protéine bi-domaine en plaçant le gène synthétique dé-répété *A3s* (présenté dans la première partie des résultats) en amont du gène répété *A3,* d'où le nom de la protéine

A3sA3. Les deux gènes sont séparés par une séquence codant un peptide *linker* flexible classique [26] $(SG_4)_2$ (figure 127). Ce *linker* artificiel est constitué d'acides aminés polaires et flexibles, afin de ne provoquer aucune contrainte sur les deux domaines. De même que pour la protéine A3s, la construction génétique implique une étiquette HisTag suivie d'un site de clivage à la protéase TEV à des fins de purification. Cette construction a été effectuée par le stagiaire ERASMUS Conor Nelson sous la direction du Dr Marielle Valerio-Lepiniec.

Figure 127 : Construction génétique utilisée pour la synthèse de la protéine A3sA3. Le gène synthétique A3s est placé en amont du gène répété A3, les deux étant séparés par un linker de séquence peptidique $(SG_4)_2$. Seul le gène synthétique peut être la cible des amorces lors des expériences de mutagénèse dirigée. L'étiquette HisTag est présente pour la purification, elle peut être retirée après purification par action de la TEV protéase.

II.2 Caractérisation de la protéine A3sA3

II.2.1 Tamis moléculaire analytique

Une expérience de tamis moléculaire analytique, nous a montré que le volume d'élution de la protéine A3sA3 était équivalent à celui de la protéine A3s (figure 128). Ceci confirme la forme « fermée » de la protéine en solution qui ressemble à celle du dimère. La plus grande symétrie du pic de la protéine A3sA3 semble même indiquer une plus grande homogénéité de la forme de la protéine qui ne peut plus subir d'échanges entre sous-unités.

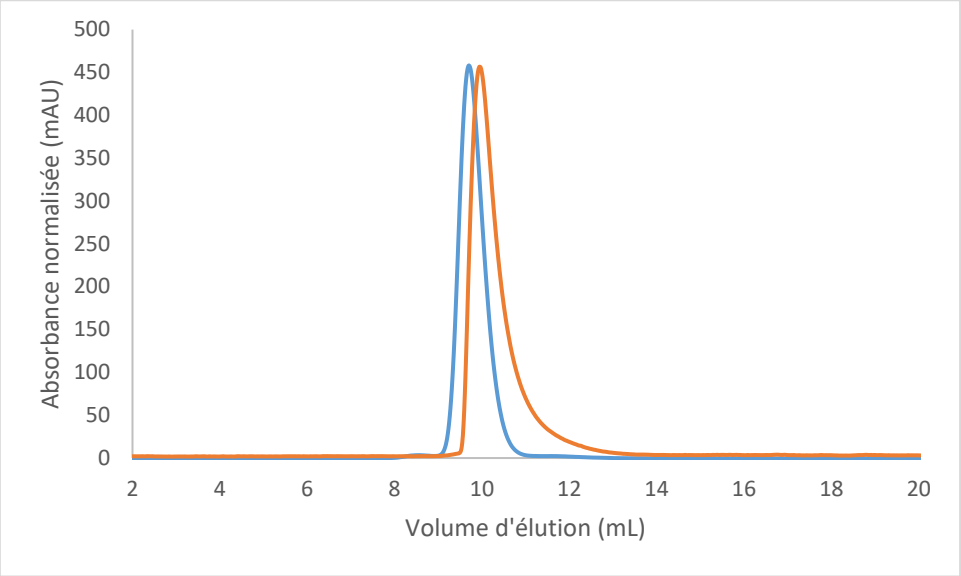


Figure 128 : Analyse par tamis moléculaire analytique. Comparaison de la protéine A3s (orange) et A3sA3 (bleu). On remarque que les volumes d'élution se correspondent tous deux, confirmant que les formes en solution sont équivalentes. Le pic de la protéine A3sA3 est plus symétrique que celui de A3s, semblant indiquer une plus grande stabilité du dimère (rayon de giration plus constant)

II.2.2 Étude cristallographique

J'ai rapidement obtenu des cristaux de la protéine dans différentes conditions de cristallisation. J'ai choisi volontairement une solution ne contenant pas de PEG afin de ne pas avoir cette molécule au sein de la crevasse. Un jeu de données d'un cristal diffractant à 1,9 Å de résolution a été collecté, j'ai donc pu résoudre la structure par remplacement moléculaire à partir de la structure de la protéine A3 (figure 129a). Nous pouvons tout d'abord remarquer que le linker n'est pas visible dans la structure, ce qui est cohérent avec sa grande flexibilité attendue. L'intégrité de la protéine issue des cristaux a par ailleurs été contrôlée sur gel par SDS-PAGE et ne montre pas de protéolyse de la protéine en deux sous-unités A3 distinctes. La comparaison de la structure de la protéine A3sA3 avec les deux autres structures d'A3 disponibles (figure 129b) nous montre que A3sA3 est plus proche de la forme fermée (ce qui est cohérent avec l'absence de PEG dans la crevasse). Cependant elle est légèrement plus ouverte.

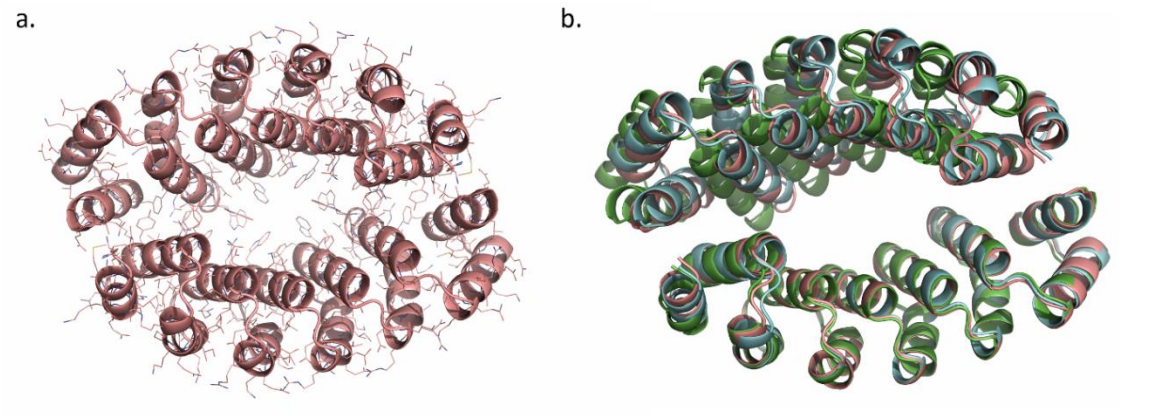


Figure 129 : a. Structure de la protéine A3sA3 résolue à partir d'un jeu de données à 1,9 Å de résolution. Le linker n'est pas visible dans la structure, probablement du fait d'une grande flexibilité. b. Comparaison de la structure de A3sA3 (rose) avec les structures des protéines forme ouverte (vert) et forme fermée (bleu). La structure de la protéine A3sA3 est plus proche de la forme fermée (cohérent avec l'absence de PEG dans la crevasse), cependant elle est légèrement plus ouverte.

Fort de ces résultats nous avons pu nous lancer dans la construction de biohybrides basés sur A3sA3. Deux approches ont été tentées sur cette protéine : dans un premier temps nous avons effectué un couplage covalent de petits ligands de type phénanthroline et terpyridine. Après avoir caractérisé ces couplages, nous avons effectué l'étude de l'énantiosélectivité des biohybrides créés pour la réaction de Diels-Alder afin de comparer ce système avec le système A3s. Dans un deuxième temps, la présence d'un unique site de couplage dans cette cavité volumineuse nous a permis d'effectuer un couplage covalent d'un complexe de plus grande taille, une porphyrine de manganèse, ceci afin d'explorer de nouvelles réactivités que sont les réactions d'oxydation.

II.3 Couplage de petits ligands phénanthroline et terpyridine

II.3.1 Synthèse des ligands

Le ligand de type phénanthroline est le même que celui utilisé dans le premier chapitre des résultats.

La terpyridine a été synthétisée par la doctorante Kalani Kariyawasam selon la procédure décrite dans la littérature [27] : la première étape correspond à une substitution nucléophile du chlorure en position 4' de la 4'-chloro–2,2':6',2''-terpyridine par le groupement hydroxyle de l'éthanolamine. L'intermédiaire 2-([2,2':6',2''-terpyridin)-4'-yloxy)éthylamine a été obtenu sous forme de solide blanc légèrement jaune avec un rendement de 83 %. Cet intermédiaire est ensuite mis en présence de N-méthoxycarbonyl-maléimide. Cette dernière étape de la réaction permet d'ajouter le groupement maléimide. Le composé **3** (appelé par la suite *terpy* ou *tpy*) a été obtenu sous

forme de solide blanc avec un rendement de 80 %, caractérisé par RMN-H^1 et spectrométrie de masse. Le coefficient d'extinction molaire à 280 nm de ce ligand a été déterminé par la même méthode que pour la phénanthroline et une valeur de 8000 $M^{-1}.cm^{-1}$ a été obtenue. La concentration des protéines fonctionnalisées a été mesurée grâce à l'absorbance à 280 nm grâce à la même équation que pour la phénanthroline (addition des coefficients d'extinction molaire).

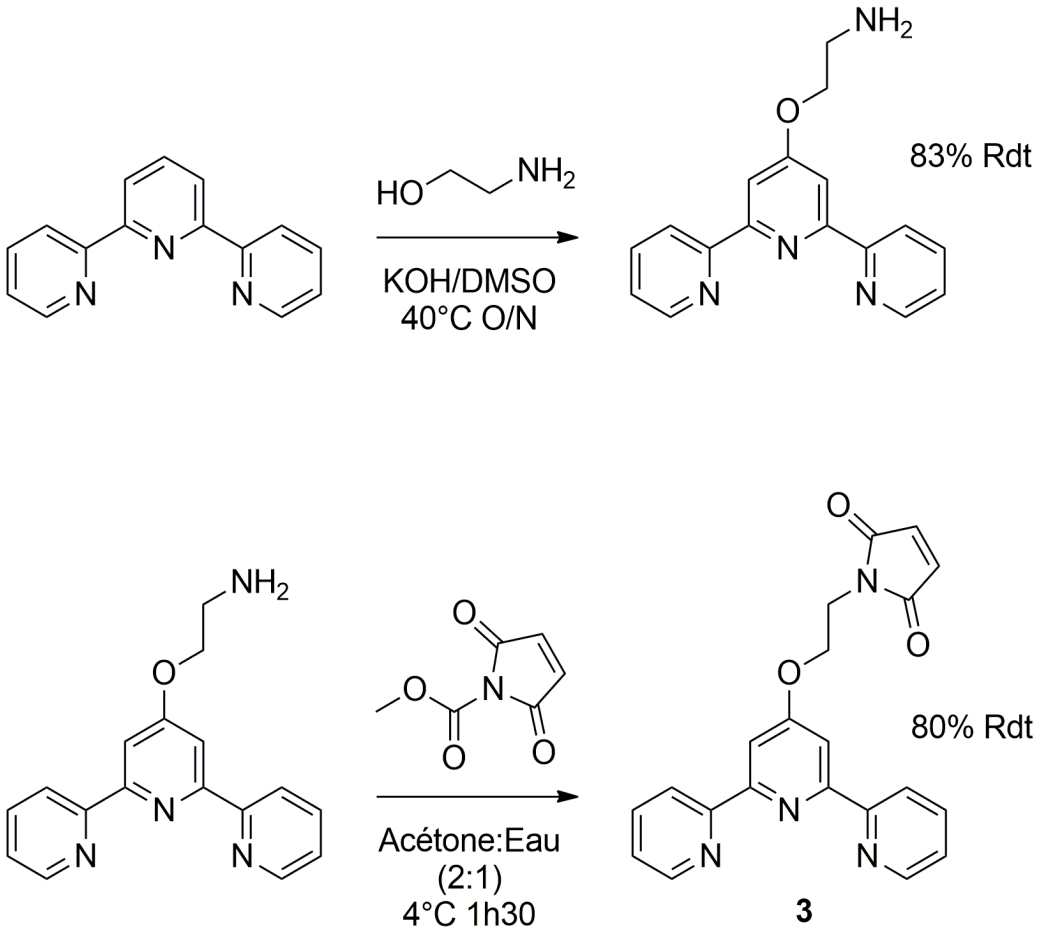


*Figure 130 : Étapes de la synthèse du ligand N-[2-((2,2':6',2''-terpyridin)-4'-yloxy)éthyl]maléimide (composé **3**).*

II.3.2 Mutation des gènes A3sA3

Le gène codant A3sA3 a été muté grâce à la méthode « QuikChange » aux positions Y26 et F119. La production dans la souche d'expression *E.coli* M15 permet également d'atteindre une production d'environ 100 mg de protéine par litre de culture. Après l'étape de clivage à la TEV protéase, c'est une quantité d'environ 70 mg qui peut être obtenue avec un litre de culture bactérienne. L'immunodétection en utilisant un anticorps dirigé contre le HisTag nous a permis d'observer l'absence de signal pour la protéine

purifiée, indiquant qu'il avait été retiré avec succès. L'analyse du gel SDS-PAGE montre l'apparition d'une très légère bande à la moitié de la taille attendue, indiquant que la protéine est parfois très légèrement protéolisée au niveau de son linker. Il est impossible de se débarrasser de ce contaminant étant donné que la protéine clivée a quasiment les mêmes propriétés que la protéine entière. Cependant la faible quantité de cette dégradation est très minoritaire et nous avons travaillé par la suite avec ces échantillons.

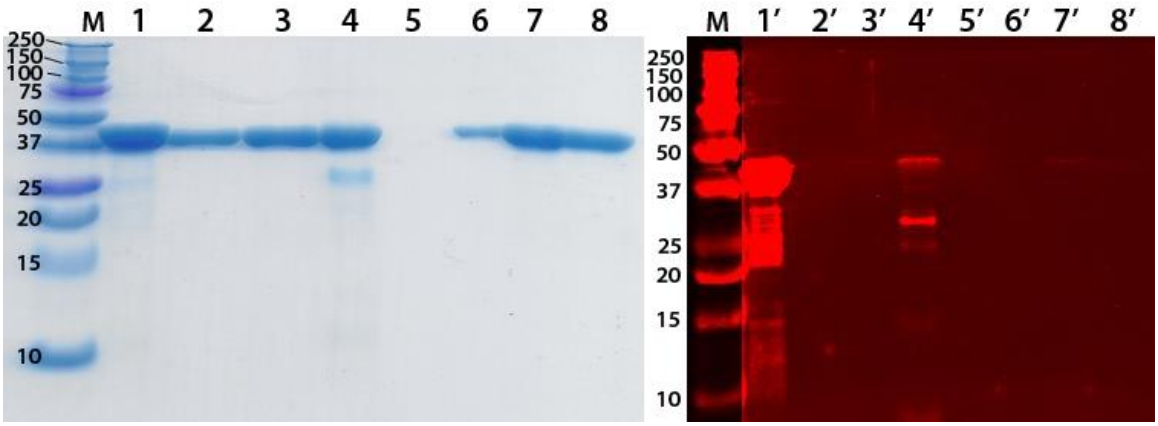


Figure 131 : Gel d'électrophorèse en conditions dénaturantes (gauche) et western blot révélé par un anticorps anti-HisTag récapitulant les étapes de purification des protéines A3sA3 (ici pour le variant A3sA3Y26C). L'image du Western Blot a été volontairement surexposée afin de bien observer les différentes contaminations, puis superposé avec une image moins surexposée pour faire apparaitre le marqueur de taille. (1, 1') Protéine purifiée en sortie de colonne d'affinité au nickel. (2, 2') et (3,3') Première et dernière fractions non retenues de la contre purification après clivage TEV. La disparition du signal sur western blot indique que la protéine a bien été débarrassée de son HisTag. (4, 4') Fraction d'élution de la contre purification : les protéines non clivées et la TEV (à ~30 kDa) sont visibles dans cette fraction.

II.3.3 Test d'Ellman sur les variants non couplés

L'analyse des tests d'Ellman résolu en temps nous donne un temps de temps d'apparition de la moitié du signal ($t_{½}$) de 3 min 30 s pour A3sA3F119C contre 1 min 40 s pour la protéine A3sF119C, indiquant que la cystéine est moins accessible au sein de cette protéine que dans le dimère équivalent. Le même phénomène est visible avec la protéine A3sA3Y26C qui a un $t_{½}$ de 30 s alors que celui de A3sY26C est si faible qu'il n'est pas mesurable. L'analyse des tests d'Ellman nous donne un temps d'apparition de la moitié du signal ($t_{½}$) de 3 min 30 s pour A3sA3F119C contre 1 min 40 s pour la protéine A3sF119C à concentration égale de thiols. Le même phénomène est visible avec la protéine A3sA3Y26C qui a un $t_{½}$ de 30 s alors que celui de A3sY26C est si faible qu'il n'est pas mesurable. Si l'on considère que la protéine est en équilibre entre deux formes, une ouverte ($A3sA3_{ouverte}$) favorisant la réaction de la cystéine avec le DTNB et une fermée ($A3sA3_{fermée}$) empêchant (ou limitant) cette réaction, le $t_{½}$ est alors dépendant des

populations relatives des deux formes (figure 132). Nous pouvons donc supposer que le *linker* de la protéine A3sA3 favorise la forme fermée par rapport à la protéine A3s. Cette observation est expliquée par l'augmentation de la concentration apparente des deux monomères comme décrit dans la littérature, limitant l'accessibilité de la cystéine. [25]

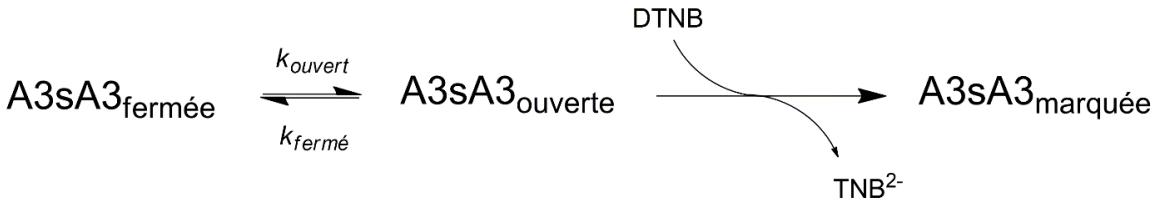


Figure 132 : Schéma expliquant la différence d'accessibilité d'une cystéine au même point de mutation entre la protéine A3s et A3sA3. La présence du linker favorise la forme fermée, où la cystéine est moins accessible au DTNB.

II.3.4 Analyse du couplage des variants par spectrométrie de masse MALDI-ToF

Les conditions de couplage des cofacteurs sont les mêmes que pour la phénanthroline et les variants A3s, avec cependant 2 équivalents molaires entre ligand terpyridine et cystéine libre contre 1,2 pour le ligand phénanthroline. Les couplages ont d'abord été vérifiés par test d'Ellman qui confirmait un taux de coupage supérieur à 95%.

Les protéines ont ensuite été analysées par spectrométrie de masse MALDI-ToF. Il est possible d'observer une augmentation de la masse des variants, avec un delta de masse correspondant à l'ajout des deux ligands (figure 133). Nous pouvons donc conclure que le couplage se fait de manière quantitative, sans sur-couplage.

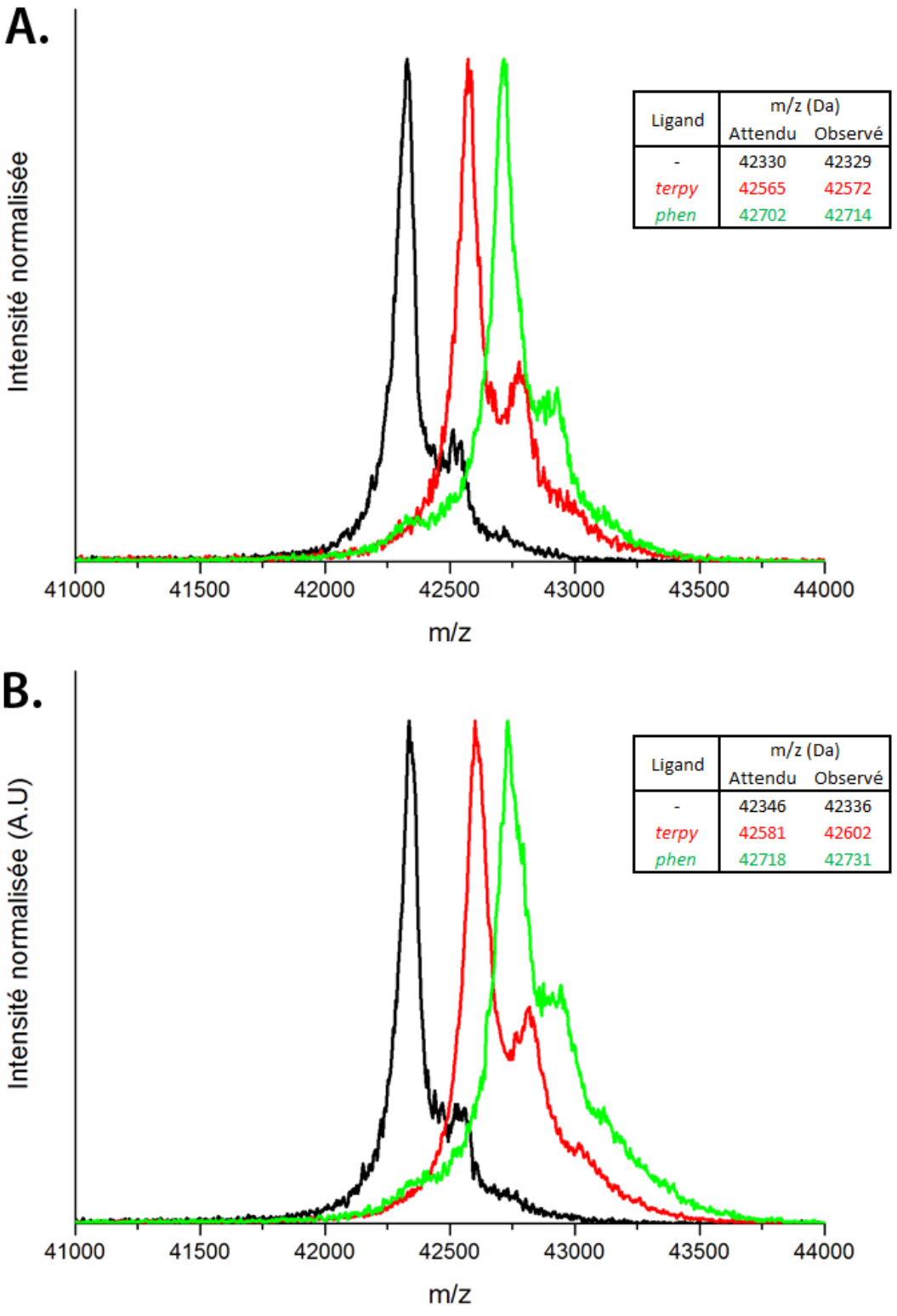
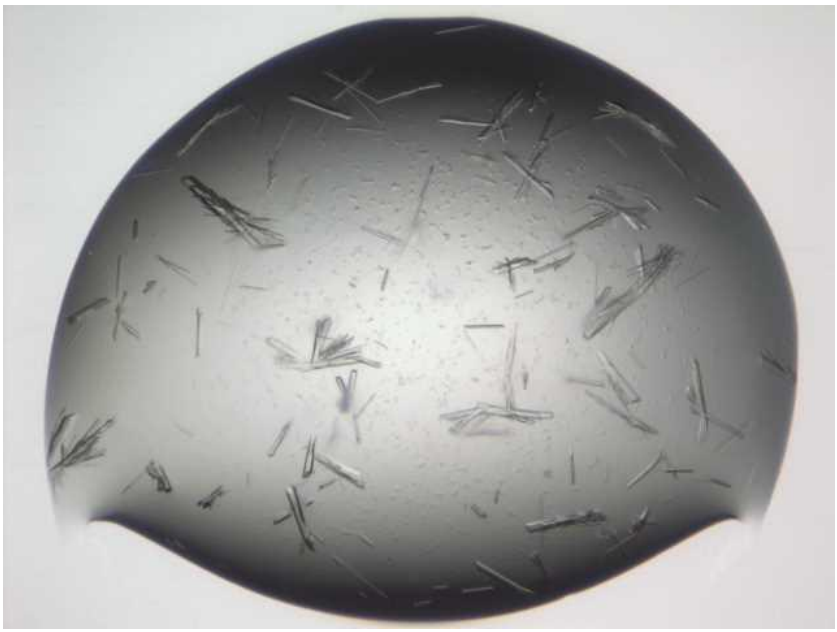


Figure 133 : Spectres d'analyse MALDI-ToF en mode linéaire du couplage de A3sA3Y26C (A) et A3sA3F119C (B) avec les ligands terpy et phen. Les m/z (en Dalton) attendus et observés pour les espèces monochargées sont représentées dans les tableaux. Le pic secondaire de plus haut m/z correspond à un adduit de matrice (acide sinapinique).

II.3.5 Analyse structurale du variant A3sA3F119NPH

Des essais de cristallisation ont été menés sur le biohybride A3sA3F119NPH, sous la forme apo ou après une incubation en présence d'un léger excès de nitrate de cuivre, suivi d'un passage sur tamis moléculaire. Des cristaux sont apparus rapidement dans des conditions proches que celles de la protéine A3sA3 sauvage. Un jeu de données à 2 Å a pu être collecté, et j'ai donc effectué le traitement des données, encore une fois par remplacement moléculaire.



Sel	Tampon	pH	Précipitant
	0,1 M Acide citrique	5	1,6 M sulfate d'ammonium

Figure 134 : Cristaux obtenus pour la protéine A3sA33F119NPH en présence de nitrate de cuivre. La composition de la solution de cristallisation est notée en dessous. On remarque l'absence de PEG dans la composition, ainsi que le pH très bas.

Premièrement il est possible de voir que la structure globale est tout à fait identique à la protéine sauvage (RMSD = 0,26). De manière étonnante, la structure ne semble pas présenter de densité électronique au niveau de la position F119 qui pourrait correspondre à une phénanthroline. Il est à noter que la densité correspondant à une phénylalanine (sur le domaine non modifié) n'est pas claire non plus. Sachant que la protéine est couplée de façon covalente et quantitative à la phénanthroline, une première hypothèse expliquant ce résultant suggère que la protéine contenue dans le cristal soit en fait un contaminant provenant de la coupure du linker et d'une réassociation des domaines de séquence sauvage. Cependant, en déposant les cristaux lavés dans du tampon sur un gel SDS-PAGE on observe une bande à la taille attendue pour la protéine

A3sA3. La protéine étant entière et couplée covalemment au cofacteur de manière quantitative, la phénanthroline est donc nécessairement dans le cristal.

La seconde hypothèse pouvant expliquer l'absence de densité électronique est une grande flexibilité de la phénanthroline au sein de la crevasse. Les différentes orientations dans le cristal étant moyennées, ceci expliquerait une densité électronique diffuse à cette position. Cependant ceci n'explique pas pourquoi la phénylalanine de l'autre domaine est également diffuse.

Une autre hypothèse expliquant ces densités électroniques diffuses au niveau de la position 119 des deux sous-unités, pourrait provenir de l'empilement cristallin lui-même. En effet, les contacts cristallins n'étant affectés ni par le linker ni la phénanthroline à l'intérieur de la crevasse et la protéine étant de symétrie C2, il se pourrait que l'empilement cristallin ne soit en fait pas régulier mais fait de protéines orientées aléatoirement dans un sens ou dans l'autre. Cet empilement désordonné conduirait alors à une moyenne des deux orientations, c'est-à-dire un taux d'occupation des éléments asymétriques (dont la phénanthroline) divisé par deux.

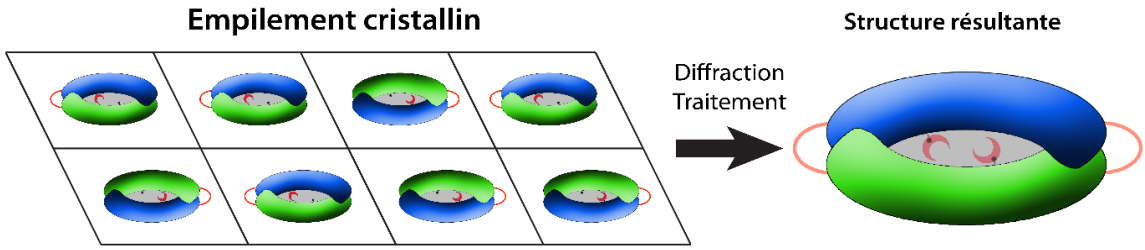


Figure 135 : Figure explicative du problème d'orientation de la protéine au sein de l'empilement cristallin. La protéine étant de symétrie C2 et présentant les mêmes contacts cristallins, le cristal est un empilement des deux orientations dont la moyennation résulte en 50% de présence des éléments asymétriques.

Cette dernière hypothèse ne peut pas être uniquement vraie car si la phénanthroline était figée, bien que moyenné le taux d'occupation serait de 50%. La première hypothèse (flexibilité du complexe) est donc également vraie.

II.3.6 Titration par métallation d'un biohybride terpyridine

Alors que les propriétés spectroscopiques de la phénantroline ne nous permettaient pas de suivre la fixation du métal sur le ligand, la terpyridine, elle, nous permet d'observer la complexation du cuivre grâce à deux bandes apparaissant à 320 et 330 nm (attribuées à une transition $\pi \rightarrow \pi^*$ du complexe [27]) Il a donc été possible de réaliser une titration permettant de vérifier la fixation du métal dans la protéine.

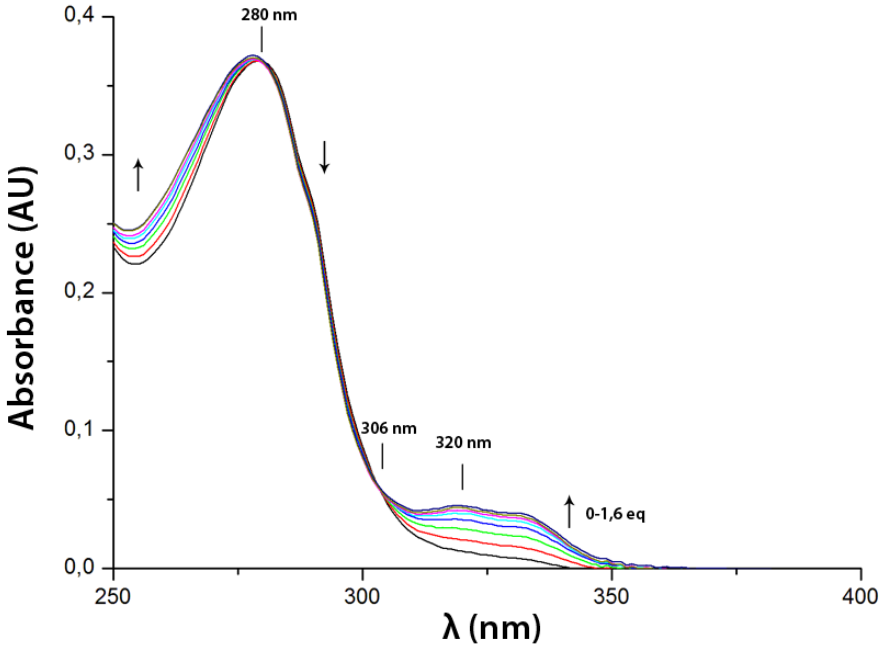


Figure 136 : Spectres de métallation de la protéine A3sA3Y26Tpy avec des ajouts progressifs de 0,2 équiv. de Cu(NO3)2. On remarque l'apparition de deux bandes à 320 et 330 nm, ainsi que deux points isobestiques à 280 et 306 nm.

Avec une protéine A3sA3Y26Tpy concentrée à 5,6 µM, l'ajout de cuivre par ajouts successifs de 0,2 équivalents dans la solution a permis d'observer une apparition proportionnelle et immédiate des bandes à 320 et 330 nm. En traçant la courbe d'absorbance à 320 nm en fonction des concentrations de Cu(II) ajouté, il est possible de voir une courbe de saturation correspondant à une augmentation du signal proportionnelle à la concentration en Cu(II) puis une inflexion pour une stoechiométrie d'environ 1:1 (figure 137). La linéarité de la réponse entre 0 et 1 équivalent est signe d'une formation du complexe $[Cu(tpy)]^{2+}$ avec une constante de dissociation très inférieure aux concentrations de cuivre et protéine utilisées. La légère augmentation de l'absorbance après ajout d'une quantité supérieure à 1 équivalent peut s'expliquer par la fixation aspécifique du métal au site de fixation fortuit découvert lors de l'étude de la protéine A3s.

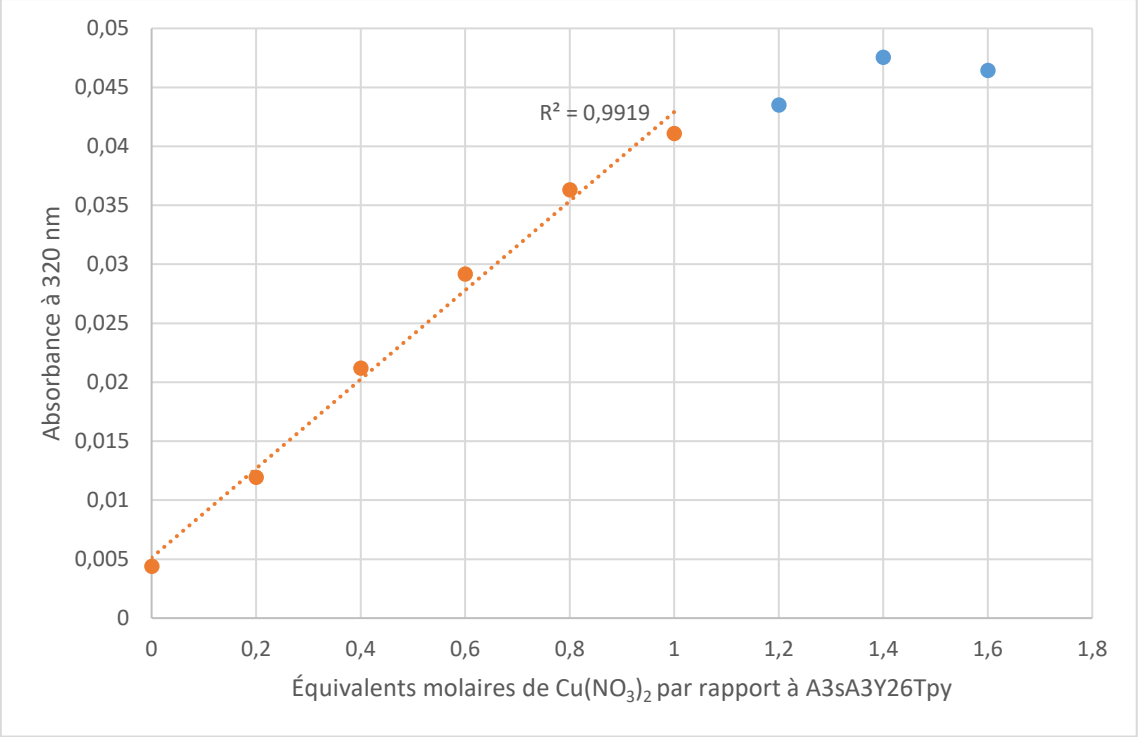


Figure 137 : Expérience de métallation du biohybride A3sA3Y26Tpy. L'absorbance est suivie à 320 nm après des ajouts progressifs de 0,2 équivalents de $Cu(NO_3)_2$ dans une solution à 5,6 µM de protéine. On remarque une saturation du signal à partir d'un équivalent ajouté. La linéarité dose/réponse est due à une concentration des deux partenaires très au-dessus de la constante de dissociation du complexe.

En ajoutant un grand excès d'EDTA à un biohybride A3sA3Y26Tpy métallé, il est également possible de faire disparaitre le signal à 320nm, confirmant que le cuivre est bien lié à la terpyridine et qu'il est possible de l'en retirer. Le décalage des points isobestiques est dû à la formation du complexe cuivre/EDTA qui absorbe également dans ces longueurs d'onde.

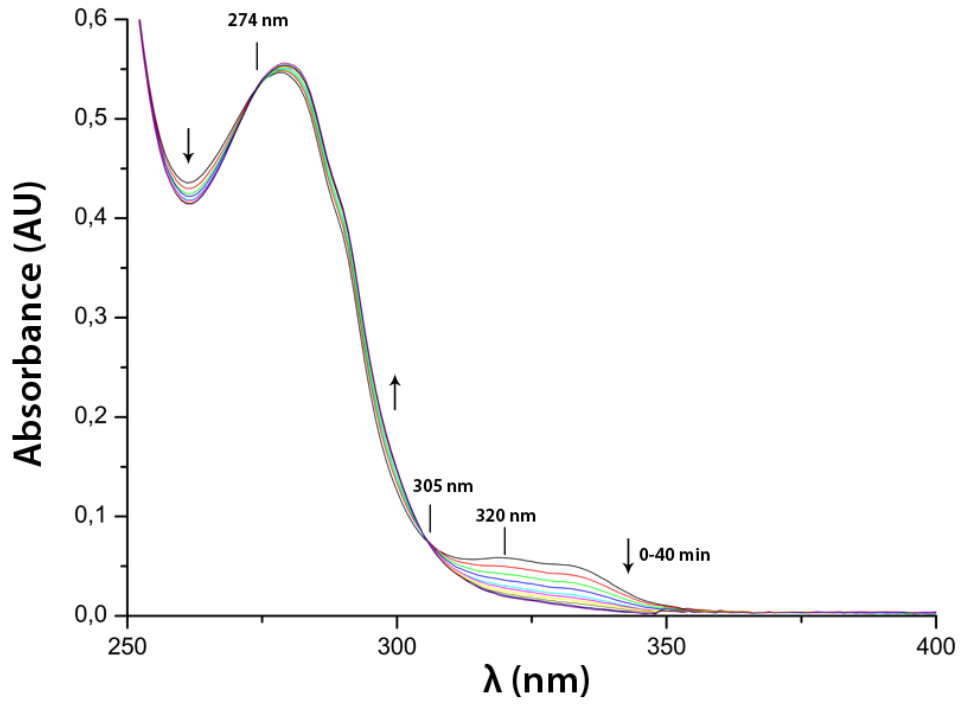


Figure 138 : Spectres de cinétique de démétallation par d'EDTA du biohybride A3sA3Y26Tpy_Cu. On remarque la disparition des bandes à 320 et 330 nm ainsi que la présence de deux points isobestiques. Le décalage des points isobestiques par rapport à l'expérience de métallation est dû à la formation du complexe cuivre/EDTA absorbant également dans l'ultraviolet.

Nous avons ensuite pu vérifier, du moins pour les variants fonctionnalisés avec la terpyridine, si le métal était toujours présent après un passage sur tamis moléculaire. J'ai effectué les prochaines expériences avec Kalani Kariyawasam.

II.3.7 Analyse des biohybrides terpyridine sur tamis moléculaire

D'après les résultats obtenus sur le dimère A3s, les catalyses ne seront effectuées qu'après une étape de tamis moléculaire afin d'enlever tout cuivre fixé de façon aspécifique. Les dessalages sont effectués sur tamis moléculaire analytique, ce qui nous permet également d'obtenir des informations sur le rayon hydrodynamique des biohybrides métallés.

Nous avons tout d'abord remarqué un comportement étrange sur un lot de variant A3sA3Y26Tpy métallé avec moins d'un équivalent de Cu(II). La figure 139 résume ce qui a été observé. La première expérience de tamis moléculaire (courbe bleue) a été effectuée avec cette préparation. On observe qu'il existe une forme au volume d'élution plus faible, c'est-à-dire au rayon hydrodynamique plus important. Supposant qu'il puisse s'agir d'un dimère de protéines A3sA3Y26Tpy reliées par une bis-terpyridine de cuivre, nous avons

décidé de rassembler ces fractions et de rajouter un excès de cuivre (1,5 équivalent). Une nouvelle expérience de tamis moléculaire analytique a été effectuée (courbe orange), on constate alors que la proportion de « dimère » a diminué. La dissociation semble cependant difficile à effectuer. Ce comportement n'avait pas été observé lors de l'expérience de titration, peut être car le complexe de bis-terpyridine de cuivre se forme avec une cinétique lente (sur la nuit). Suite à une métallation d'un lot frais avec un excès de cuivre, le profil d'élution ne montre pas de pic correspondant au dimère, ce qui confirme l'hypothèse d'une métallation rapide et d'une dimérisation lente (coube verte).

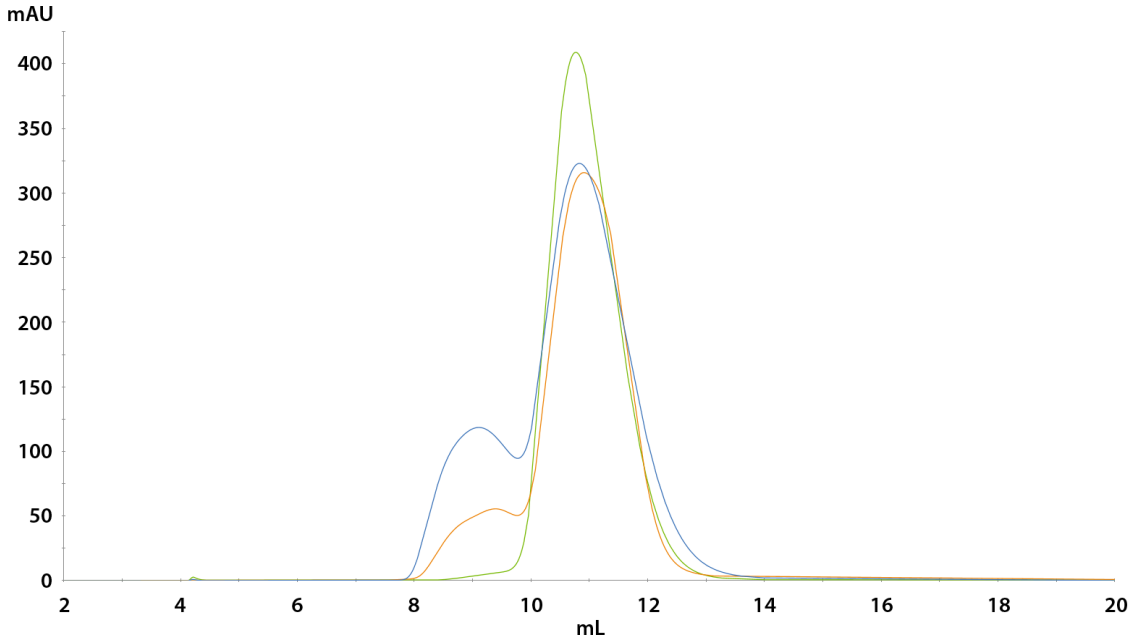


Figure 139 : Chromatogrammes de tamis analytique du variant A3sA3Y26Tpy après ajout de cuivre. Courbe bleue : biohybride A3sA3Y26Tpy_Cu après métallation avec une quantité substœchiométrique de $Cu(NO_3)_2$. Une espèce de plus haut rayon hydrodynamique est présente, pouvant correspondre à une dimérisation des protéines A3sA3. Courbe orange : Même préparation après remétallation avec un excès de $Cu(NO_3)_2$ sur la nuit. Le pic de dimère soupçonné est plus faible mais toujours présent, signe qu'une fois formé, il est difficile de le dissocier. Courbe verte : biohybride A3sA3Y26Tpy_Cu après métallation avec un excès de $Cu(NO_3)_2$ sur la nuit. Le pic de dimère est absent, signe qu'après métallation totale, le dimère ne se forme pas.

Ce résultat intéressant suggère que la protéine A3sA3Y26Tpy pourrait former un dimère ponté par une bis-terpyridine. Il est pour le moment impossible de savoir si cette dimérisation se produit avec une forme ouverte de la protéine (échange de domaine, figure 126) ou avec la forme fermée, la terpyridine dépassant de la crevasse.

Les expériences de tamis analytique sur les autres biohybrides A3sA3F119Tpy_Cu, A3sA3F119NPH_Cu, A3sA3Y26NPH_Cu montrent des volumes d'élution sensiblement identiques, signe d'un rayon hydrodynamique proche. Seul le biohybride A3sA3F119Tpy présente un très petit pic en amont, ressemblant au « dimère » vu précédemment. Le ligand terpyridine semblerait donc favoriser ce comportement.

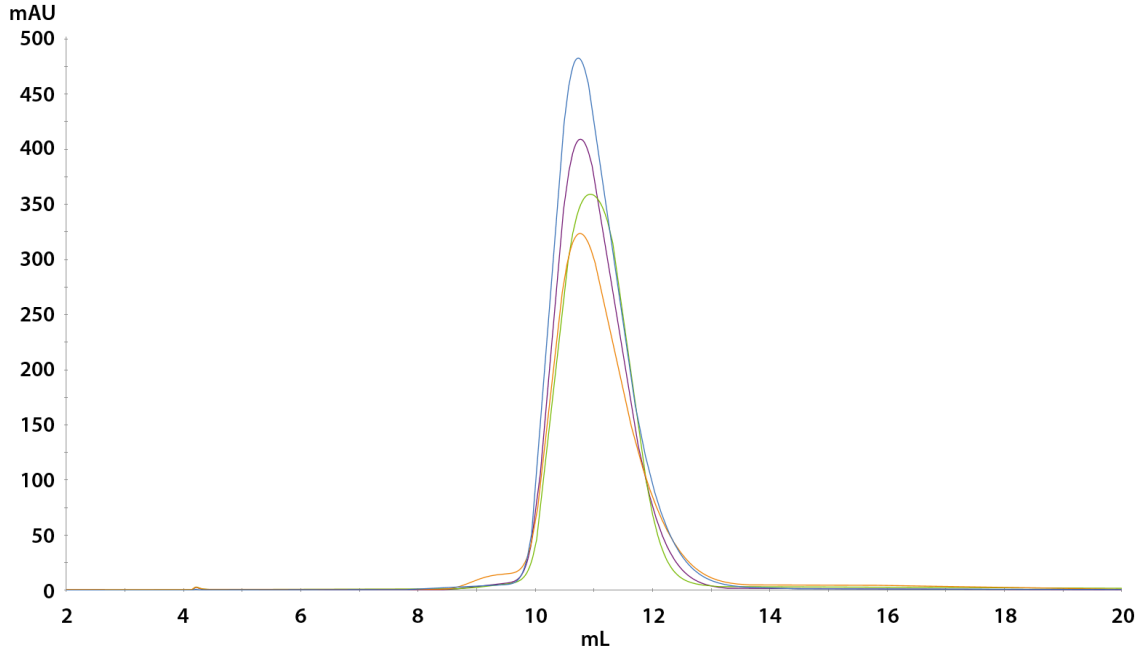


Figure 140 : Chromatogrammes des biohybrides A3sA3 métallés. Courbe verte : A3sA3F119NPH_Cu Bleue : A3sA3Y26NPH_Cu ; Orange : A3sA3F119Tpy_Cu ; Violette : A3sA3Y26Tpy_Cu. Les volumes d'élution sensiblement identiques indiquent un rayon hydrodynamique équivalent. Le biohybride A3sA3F119Tpy_Cu présente un très léger pic correspondant à un rayon hydrodynamique plus grand, semblable au « dimère » du variant Y26 également fonctionnalisé avec la terpyridine.

Les différents biohybrides ayant de volumes d'élution comparables, nous avons décidé de directement effectuer des catalyses de la réaction de Diels-Alder dans les conditions optimales de tampon, temps et température obtenues pour le biohybride A3sF119NPH_Cu dans l'optique de les comparer.

II.3.8 Catalyse de la réaction de Diels-Alder par les biohybrides dérivés de A3sA3

Les catalyses de la réaction de Diels-Alder ont été effectuées avec les biohybrides dans un tampon MOPS pH 7 à 4°C pendant 48h avec 30 µM ou 15 µM de catalyseur en fonction de la quantité de protéine disponible pour la catalyse. Les résultats sont reportés dans le tableau 141. Il est à noter que les excès énantiomériques, quand ils sont significatifs, sont toujours en faveur du même énantiomère, le même également que dans l'étude sur A3sF119NPH.

Catalyseur	[catalyseur] (µM)	*ee* endo (%)	Rendement (%)
$Cu(NO_3)_2$	15	2 ± 1	30 ± 1
	30	1 ± 1	44 ± 2
A3sA3Y26Tpy_Cu	15	6 ± 1	6 ± 1
	30	4 ± 15	18 ± 6
A3sA3Y26NPH_Cu	15	**16** ± 2	10 ± 1
	30	**17** ± 2	17 ± 5
A3sA3F119Tpy_Cu	30	**24** ± 13	11 ± 6
A3sA3F119NPH_Cu	30	1 ± 4	**38** ± 2

Tableau 141 : Résultats d'excès énantiomérique et rendement d'apparition du produit pour la réaction de Diels-Alder entre azachalcone et cyclopentadiène pour des catalyseurs ($Cu(NO_3)_2$ ou biohybride A3sA3) à des concentrations de 15 µM ou 30 µM. Conditions : 34 mM cyclopentadiène, 1 mM azachalcone dans CH_3CN (1% final), tampon MOPS 20 mM pH 7, 150 mM NaCl, 48h, 4°C. Toutes les expériences ont été faites en réplica.

En comparant les rendements de réaction à 48h des différents catalyseurs, on peut observer que les biohybrides présentent tous un rendement largement inférieur celui du cuivre, excepté le biohybride A3sA3F119NPH_Cu(II) obtenant un rendement proche de celui du cuivre (38% vs 44%). Ce fort rendement est accompagné d'une absence d'excès énantiomérique pour ce biohybride A3sA3F119NPH_Cu. Cette absence d'énantiosélectivité peut s'expliquer grâce aux résultats cristallographiques indiquant une flexibilité de la phénanthroline au sein de la crevasse. En effet, des mouvements de la phénantroline au sein de la crevasse induiraient alors un environnement moyen isotrope pour le substrat et donc empêcherait une catalyse énantiosélective.

Les biohybrides présentant le plus haut excès énantiomérique sont A3sA3Y26NPH_Cu et A3sA3F119Tpy_Cu(II), étant l'ordre de 20%. Ces excès énantiomériques sont accompagnés d'un rendement à 48h de 11% et 17% respectivement pour 30 µM de catalyseur contre 44% pour le cuivre dans les mêmes conditions. Ces excès énantiomériques et rendements peuvent être le signe d'une baisse de flexibilité du cofacteur au sein de la crevasse. Cette baisse de flexibilité peut provenir de son enfouissement ou de la complétion de la sphère de coordination du cuivre par des chaines latérales d'acides aminés proches.

Le biohybride A3sA3Y26Tpy_Cu(II) est présente lui un excès énantiomérique moyen entre 4% et 6%, et d'un rendement de 6% à 15 µM ou 18% à 30 µM, ce qui est le rendement le plus bas de tous les biohybrides testés.

II.3.9 Résumé de l'utilisation des variants A3sA3 pour l'insertion de petits ligands

En résumé, grâce à une construction permettant de lier génétiquement par un *linker* deux gènes différents codant pour deux domaines A3 identiques, nous sommes parvenus à construire une protéine αRep-A3 bi-domaine, permettant le greffage covalent de complexes terpyridine ou phénanthroline au sein de la crevasse de manière asymétrique. Cette nouvelle protéine, du fait de la concentration apparente d'un domaine envers l'autre beaucoup plus élevée, semble adopter comme prévu une conformation fermée que l'on a mise en évidence dans la structure cristallographique du biohybride A3sA3F119NPH_Cu. Les variants fonctionnalisés semblent également beaucoup moins prompts à former des complexes bis-phénanthroline ou bis-terpyridine, bien qu'ils soient peut-être observables lors d'une présence prolongée d'une présence sub-stoechiométrique de Cu(II), comme vu lors d'expériences de tamis moléculaire analytique.

Concernant la sélectivité des quatre biohybrides obtenus pour la catalyse de la réaction de Diels-Alder, nous avons constaté que l'insertion de complexes différents à la même position provoque l'apparition ou non d'un excès énantiomérique. De manière intéressante les meilleures sélectivités obtenues pour les ligands phénanthroline ou terpyridine sont aux positions Y26 et F119 respectivement : pour obtenir un biohybride sélectif, il ne semble donc y avoir ni de position privilégiée pour le couplage, ni de ligand privilégié.

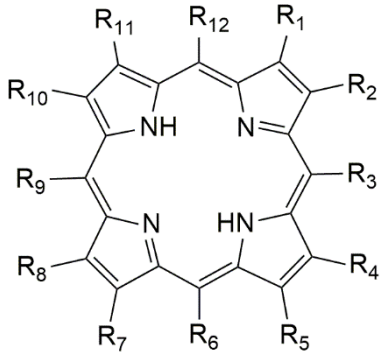
Il serait alors intéressant d'une part de modifier le bras d'ancrage des cofacteurs afin de les rendre plus rigides, d'autre part de faire varier les acides aminés de la crevasse pour tenter d'atteindre des sélectivités plus hautes.

Bien qu'ayant des sélectivités moindres que le biohybride A3sF119NPH_Cu, les biohybrides basés sur A3sA3 semblent donc plus prometteurs pour mettre en place une stratégie d'évolution dirigée.

La présence d'un seul ligand laissant beaucoup de place disponible au sein de la crevasse, nous avons donc décidé d'étudier la création d'un biohybride avec un ligand plus complexe et de plus grande taille : une porphyrine.

II.4 Couplage de la protéine A3sA3 avec une porphyrine.

Les porphyrines sont des cofacteurs ubiquitaires dans la nature, la porphyrine principale du monde animal étant la protoporphyrine XI de fer, ou hème. Les protéines à hème effectuent grâce à leur groupement prosthétique des fonctions variées telles que la fixation de dioxygène (hémoglobine, myoglobine), des réactions d'oxydation (cytochrome P450) ou le transfert d'électrons (Cytochrome C). L'ingénierie des hémoprotéines constitue un véritable challenge dans le domaine de la chimie bioinorganique, comme l'illustrent les travaux sur la myoglobine ou le cytochrome P450 présentés dans l'introduction de cette thèse.



$R_{3,}\ R_{6,}\ R_{9,}\ R_{12}$ = 5, 10, 15, 20 = *meso*

Figure 142 : Structure générale d'une porphyrine. Les porphyrines naturelles sont généralement fonctionnalisées sur les cycles pyrroles tandis que les porphyrines artificielles le sont généralement aux positions meso.

Les porphyrines sont basées sur un cycle tetrapyrrole, appelé macrocycle, formant un système π à 18 électrons très délocalisé. Les quatre atomes d'azote pointant au centre du macrocycle peuvent complexer un métal avec une haute affinité. [28] Le système électronique est influencé par les substituants électrodonneurs ou électroattracteurs du macrocycle, qui à son tour influence les propriétés du métal. Ce dernier influence en retour le système électronique de la porphyrine par son état d'oxydation ou la présence d'un ou deux ligands axiaux. Les propriétés des porphyrines sont donc nombreuses et chaque combinaison porphyrine/métal/ligand possède des propriétés physicochimiques qui lui sont propres (voir *The Porphyrin Handbook*). Plusieurs propriétés peuvent émerger du couplage d'une porphyrine à une architecture protéique : la solubilisation de la porphyrine dans un solvant aqueux, la protection de cette dernière face aux dégradations oxydatives, la modification de son potentiel redox par la coordination entre un acide aminé et le métal et l'augmentation de la sélectivité de la réaction.[29] Obtenir un biohybride basé sur une porphyrine est un projet ambitieux et prometteur, ouvrant un large panel de possibilités, notamment dans les réactions d'oxydation sélectives.[29] Nous

avons choisi d'utiliser pour ces expériences des porphyrines de manganèse qui sont connues comme étant de bons catalyseurs d'oxydation et ne forment pas de ponts μ-oxo en milieu aqueux susceptibles d'induire la formation de dimères de porphyrine.

II.4.1 Études préliminaires

Des tentatives d'association entre la protéine A3 et différentes porphyrines par la stratégie hôte-invité avaient déjà été tentées par le passé, sans succès (master 2 Recherche de Hicham Kardaoui). C'est pourquoi nous avons décidé d'utiliser la stratégie du couplage covalent pour associer différentes porphyrines. J'ai tout d'abord synthétisé une porphyrine Mn(III) 5,10,15-triphényl-20-(*4*-(maléimidopropylamido)phényl)porphyrine (MnTPP-mal, P1 figure 143) qui a un caractère hydrophobe.

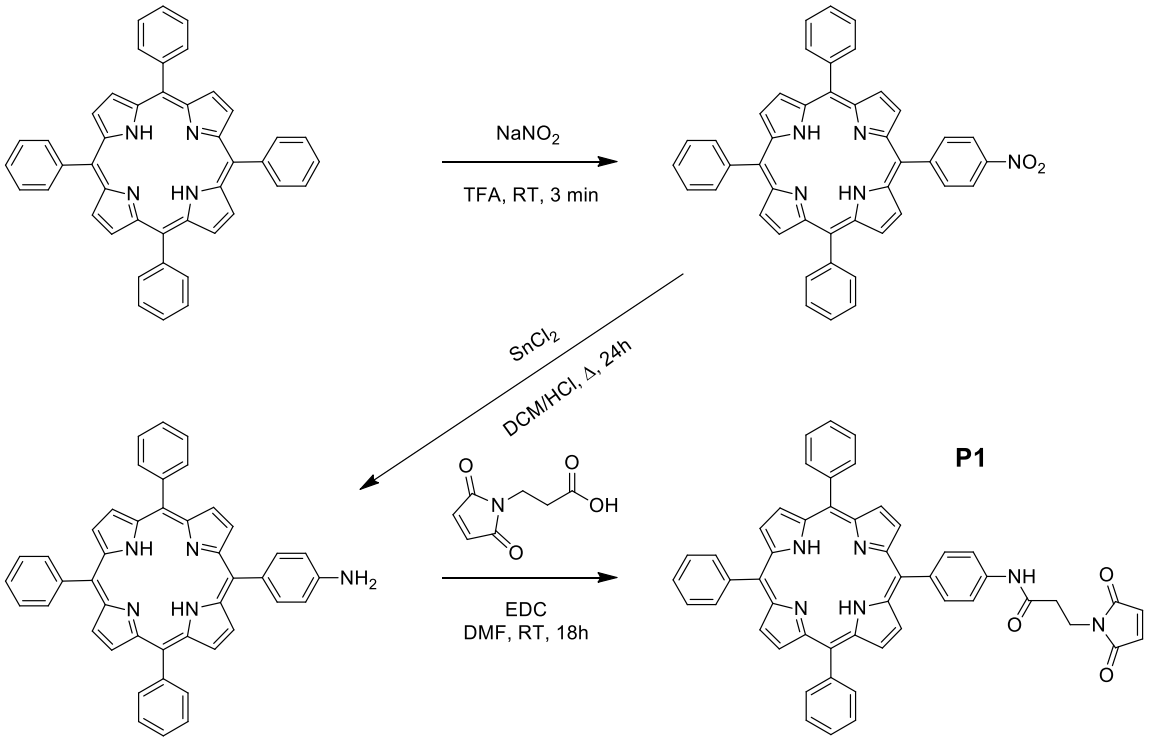


Figure 143 : Synthèse de la porphyrine P1 (MnTPP-mal ou Mn(III) 5,10,15-triphényl-20-(4-(maléimidopropylamido)phényl)porphyrine)

J'ai également utilisé d'autres porphyrines présentes au laboratoire, telle que la 5,10,15-tri(4-N-méthylpyridiniumyl)-20-(4-isothiocyanatophényl)-porphyrine (TMPyP-SCN, P2 figure 146) qui a un caractère hydrophile. J'ai tout d'abord effectué le couplage de ces deux porphyrines au variant A3sK30C, un des variants du **dimère** A3s. Ces expériences préliminaires montraient une oligomérisation de la protéine (figure 144, gauche). Les oligomères résistant à la dénaturation thermique (95°C durant 30 min, figure 144, droite), provenaient probablement de l'empilement des porphyrines dans l'eau, un problème typique de ce genre de cofacteur. [30,31] Le dimère A3s ne semblait donc pas adapté au greffage de deux porphyrines dans sa crevasse, qu'elle soit hydrophile ou hydrophobe.

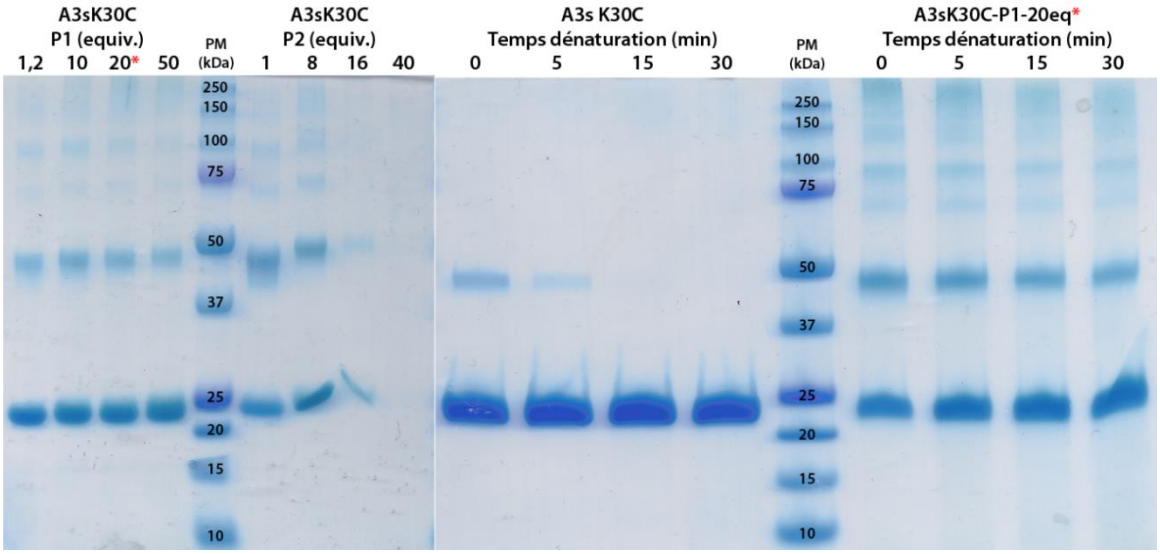


*Figure 144 : Gels SDS-PAGE des expériences préliminaires de couplage covalent de porphyrine hydrophobe (P1 : MnTPP-mal) ou hydrophile (P2 : TMPyP-SCN) au variant A3sK30C. Masse attendue : 23 kDa. Gauche : couplage de 1,2 à 50 équivalents de Porph1 ou de 1 à 40 équivalents de Porph2 par rapport au monomère de A3sK30C. Une multimérisation résistante aux conditions dénaturantes est clairement visible. L'ajout d'un excès de Porph2 provoque la précipitation de la protéine. Droite : Comparaison de la résistance à l'étape de dénaturation du protocole SDS-PAGE du dimère A3sK30C ou du produit de couplage A3sK30C-Porph1-20équiv. Les multimères sont clairement résistants à la dénaturation. *même préparation dans les deux gels.*

Dès lors, il nous a semblé judicieux d'utiliser la protéine bi-domaine A3sA3 pour effectuer le greffage d'une porphyrine. Une modélisation simple par le programme AutoDock entre la protéine A3 sauvage et une porphyrine libre simple (*5,10,15,20*-(4-N-méthylpyridiniumyl)-porphyrine, TMPyP) nous a confirmé que la porphyrine pouvait s'insérer dans la crevasse (figure 145). On peut également voir sur la figure que la simulation montrait un autre site de fixation possible à l'extérieur de la protéine, confirmant le besoin d'un ancrage covalent afin de maintenir la porphyrine au sein de la crevasse.

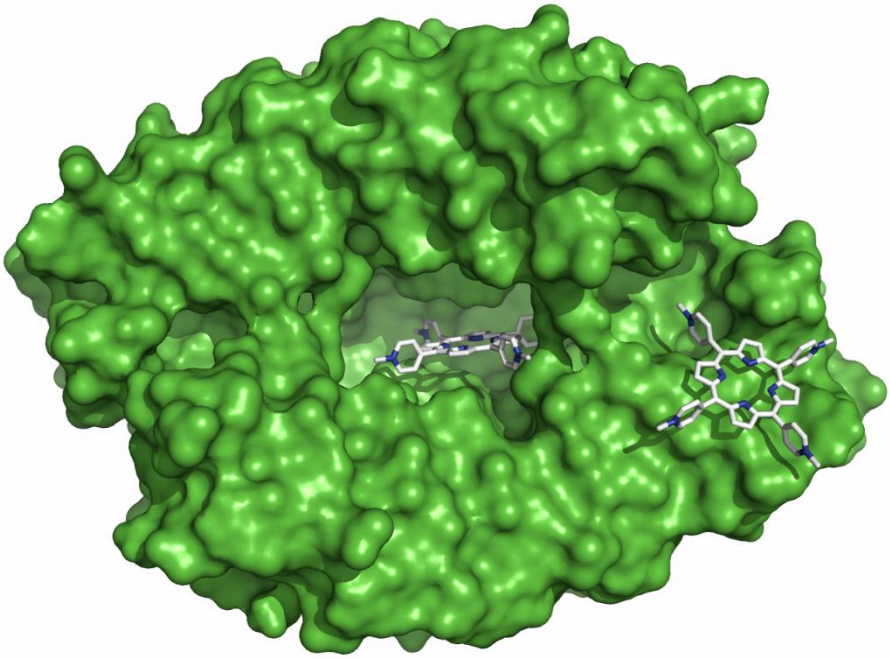


Figure 145 : Simulation AutoDock de fixation d'une porphyrine TMPyP sur la protéine A3sA3. Deux postitions sont privilégiées, une dans la crevasse et une autre à l'extérieur.

II.4.2 Tests de couplage de différentes porphyrines

J'ai alors tenté le couplage de quatre porphyrines disponibles au laboratoire (figure 146) sur les trois variants A3sA3Y26C, A3sA3K30C et A3sA3F119C. Le variant A3sA3D150C n'a pas été retenu car orienté vers une crevasse secondaire trop petite pour accueillir une porphyrine. L'analyse du couplage a été effectuée en spectrométrie de masse MALDI-ToF, les résultats sont présentés dans le tableau 147.

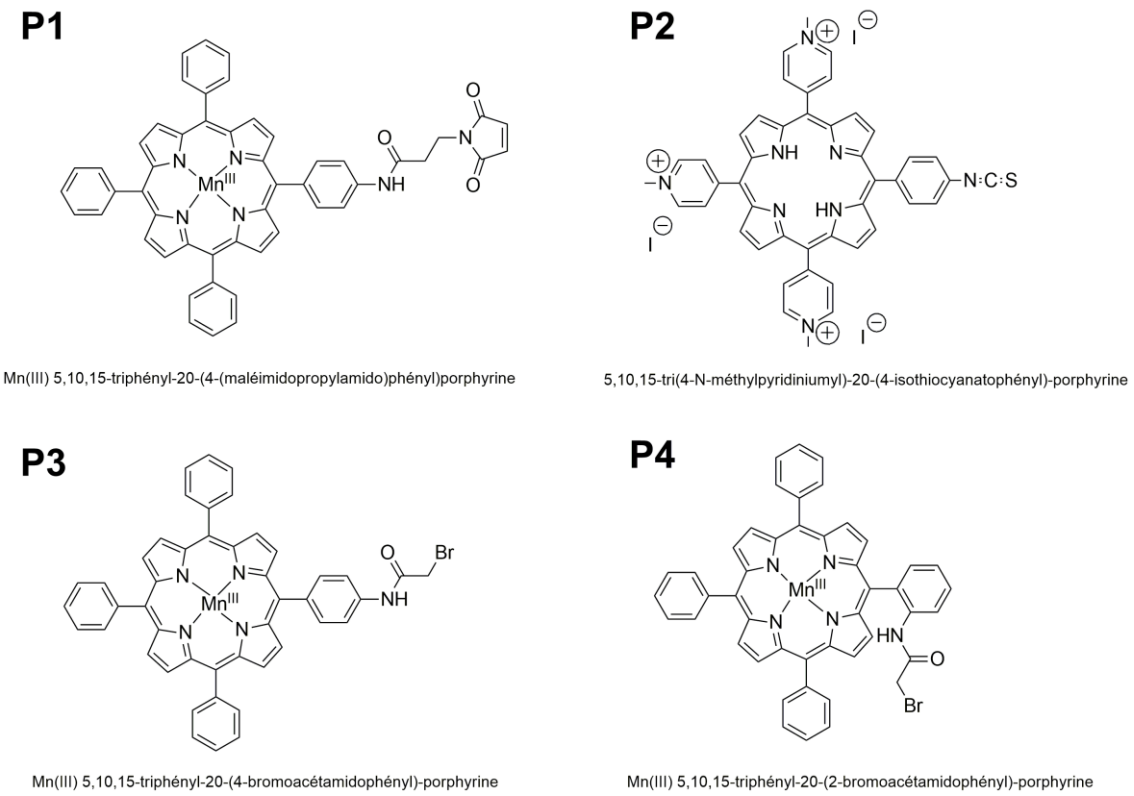


Figure 146 : Différentes porphyrines couplées aux variants A3sA3Y26C, A3sA3K30C, A3sA3F119C. Conditions de couplage : 5 µM variant A3sA3, 25 µM porphyrine (5 équiv.) dans DMSO pour 10% DMSO final, tampon Na-Phosphate 50 mM pH 7,75, 150 mM NaCl.

		Porphyrines			
		P1	P2	P3	P4
Variants A3sA3	Y26C	+++	+	-	-
	K30C	+++	+	-	-
	F119C	+++	++	-	-

Tableau 147 : Résumé des analyses MALDI-ToF des réactions de couplages: (-) Absence pic de protéine fonctionnalisée. (+) Pic de protéine fonctionnalisée inférieur au pic de protéine non fonctionnalisée. (++) Pic de protéine fonctionnalisée

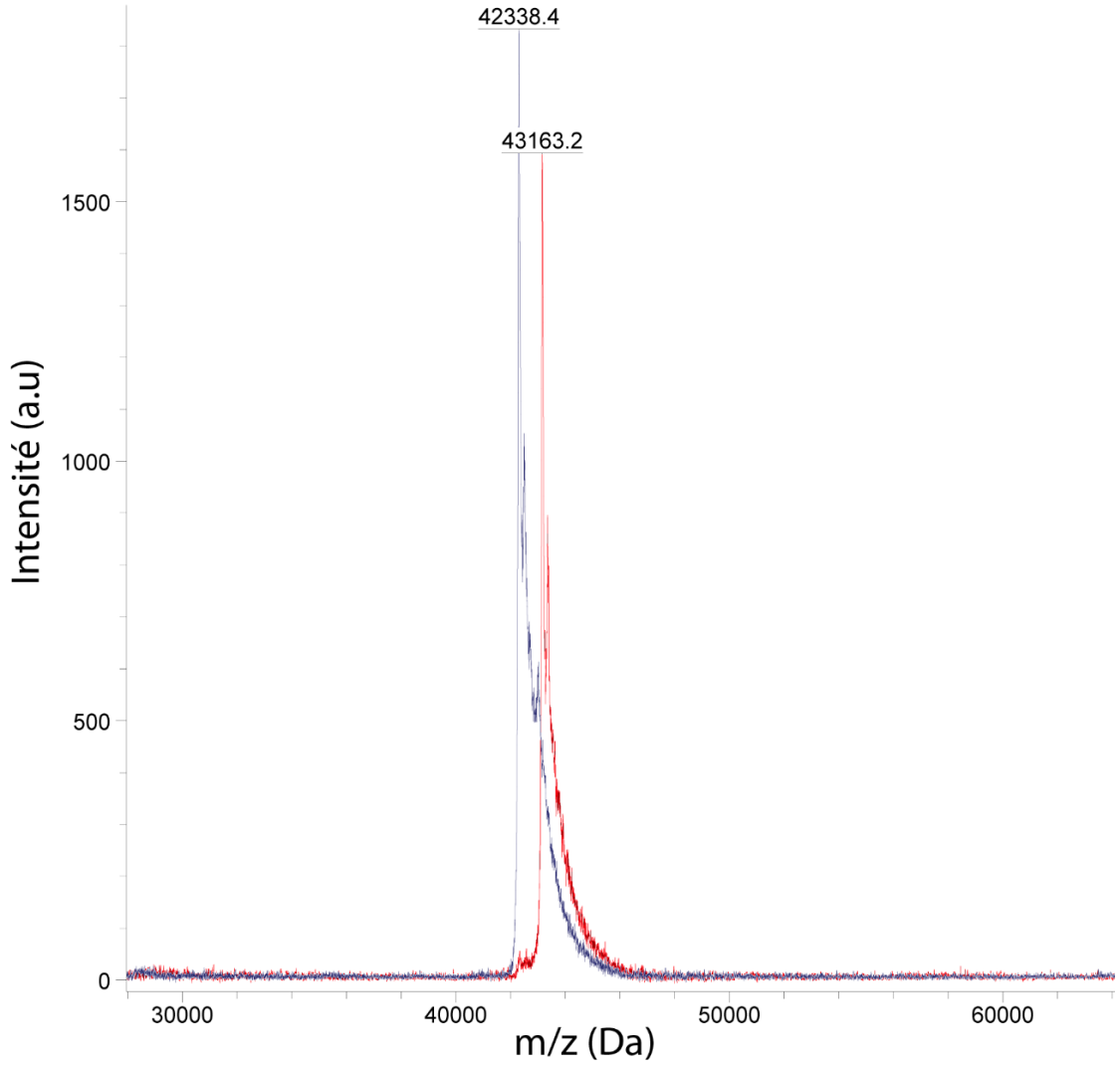


Figure 148 : Spectre de masse MALDI-ToF du couplage entre le variant A3sA3Y26C et la porphyrine P2 (MnTPP-mal). Courbe bleue : A3sA3Y26C (masse calculée : 42330 Da, masse mesurée : 42338 Da). Courbe rouge : A3sA3Y26TPPMn (masse calculée : 42163 Da, masse mesurée : 42163 Da)

L'analyse par spectrométrie de masse MALDI-ToF a montré que les porphyrines P3 et P4, des MnTPP fonctionnalisées par un bras bromoacétamide en position *para* et *ortho* d'un phényl, respectivement, ne provoquaient pas de décalage de masse des différents variants. Nous en avons conclu une absence de couplage covalent malgré l'utilisation du même bras que pour la phénanthroline. La porphyrine P2, fonctionnalisée par un bras isothiocyanate réagissant avec les nucléophiles de la protéine (amines et sulfhydryles), ne présentait qu'un couplage partiel des différents variants. Seule la porphyrine P1, une MnTPP fonctionnalisée par un bras maléimide en position *para* d'un phényl, provoquait la disparition totale du pic correspondant à la protéine non couplée pour former un pic unique avec une variation de masse attendu pour la fixation d'une porphyrine par protéine.

Nous avons donc décidé d'utiliser la porphyrine P1 avec le variant A3sA3Y26C. La position 26, étant excentrée par rapport à la crevasse et relativement enfouie, semblait la plus propice à l'insertion de la porphyrine, comme visible sur la figure 149. Le couplage de cette porphyrine étant également quantitatif pour les deux autres variants, cette étude sera poursuivie ultérieurement sur les deux autres biohybrides générés.

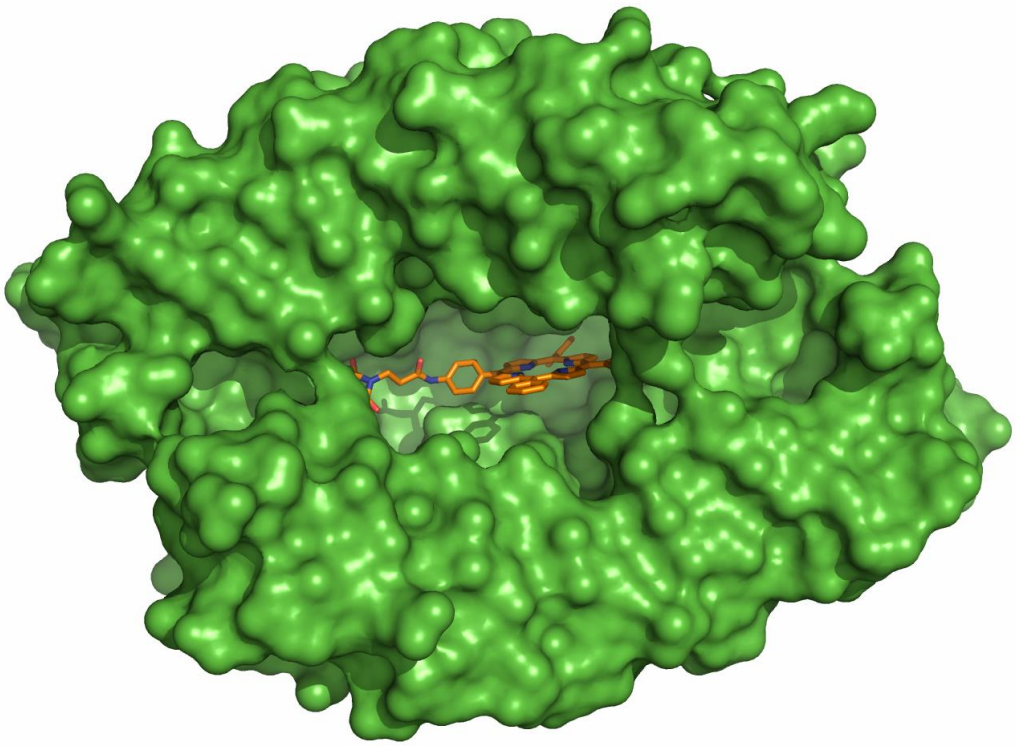


Figure 149 : Construction PyMol du variant A3sA3Y26C fonctionnalisé avec la porphyrine P1.

II.4.3 Analyse sur le biohybride A3sA3Y26TPPMn

II.4.3.1 Méthode d'obtention du biohybride

Après avoir réussi à montrer la possibilité d'un couplage covalent entre les variants A3sA3 et la porphyrine P1, le second but a été de trouver un moyen de séparer la porphyrine en excès de la protéine. La porphyrine TPP étant très hydrophobe, elle présente une forte propension à se fixer de manière forte et aspécifique à un grand nombre d'objets : protéines, colonnes de dessalages, concentrateurs, verrerie… Il était donc exclu d'utiliser tout matériel non jetable pour la séparation de la porphyrine en excès. Une mise au point a alors été effectuée en utilisant la protéine A3sA3WT qui ne forme pas de lien covalent avec la porphyrine P1.

L'utilisation d'une colonne de dessalage usage unique par centrifugation de type Zeba Spin était d'une efficacité limitée : une partie de la porphyrine en excès était effectivement retirée (aisément visible par la couleur verte de la colonne), cependant l'éluat de la protéine A3sA3WT présentait toujours une couleur verte, signe que de la porphyrine était encore fixée non covalemment à la protéine.

J'ai donc utilisé une version modifiée de la méthode de Teale.[32] Cette méthode découverte en 1959 est encore aujourd'hui utilisée pour séparer les groupements prosthétiques des hémoprotéines, elle repose sur une extraction à la butanone-2 en conditions acides. J'ai montré que les conditions acides n'étaient pas nécessaires et qu'une simple extraction sur glace permettait de retirer toute la porphyrine de la protéine A3sA3WT, la phase aqueuse contenant la protéine devenant incolore.

En effectuant la même manipulation avec la protéine A3sA3Y26TPPMn, la phase aqueuse reste vert pâle. La porphyrine covalemment liée à la protéine est à présent complètement solubilisée, elle ne se fixe plus aux matériaux tels que les filtres de concentrateurs, il est donc possible d'utiliser les équipements tels que le tamis moléculaire. Pour ce faire, il est cependant nécessaire de pouvoir concentrer la protéine, or une fraction de butanone se dissout dans la phase aqueuse au moment de l'extraction (à hauteur de 20%). Une telle concentration de butanone dissout les plastiques des concentrateurs, en effet cette cétone est couramment utilisée pour effectuer des soudures chimiques de plastiques.

Deux options ont donc été considérées : (1) il est possible de diluer la solution aqueuse de façon à avoir environ 3% de butanone, ainsi diluée la solution ne dissout plus le plastique des concentrateurs, cependant le volume à concentrer est gigantesque (30 mL pour 1 mg de protéine). Il est également possible d'effectuer une dialyse dans des boudins fait de cellulose regénérée, qui semble résister à 20% de butanone, avant de concentrer. (2) Enfin, la méthode qui semble la plus rapide et efficace est de laisser reposer la phase

aqueuse dans de la glace puis de la centrifuger à froid. Dans ces conditions, un culot de protéines précipitées blanc-vert se forme au fond du tube et il est alors possible de séparer le culot de protéines du surnageant. Un rinçage à l'aide de tampon permet de re-solubiliser la protéine tout en la concentrant et en se débarrassant de la butanone. Cette re-solubilisation presque totale est rendue possible par le caractère réversible de la dénaturation des αRep.[33] A l'issue de l'une ou l'autre des deux méthodes, il est nécessaire de se débarrasser de quelques agrégats en passant à travers un filtre de 0,2 µm.

Il est alors enfin possible d'injecter la protéine sur tamis moléculaire afin de vérifier son agrégation. Afin d'éviter d'endommager la colonne analytique avec la porphyrine, une ancienne colonne préparative a dû être utilisée. Les chromatogrammes des protéines obtenues par les méthodes (1) et (2) sont reportés dans sur figure 150.

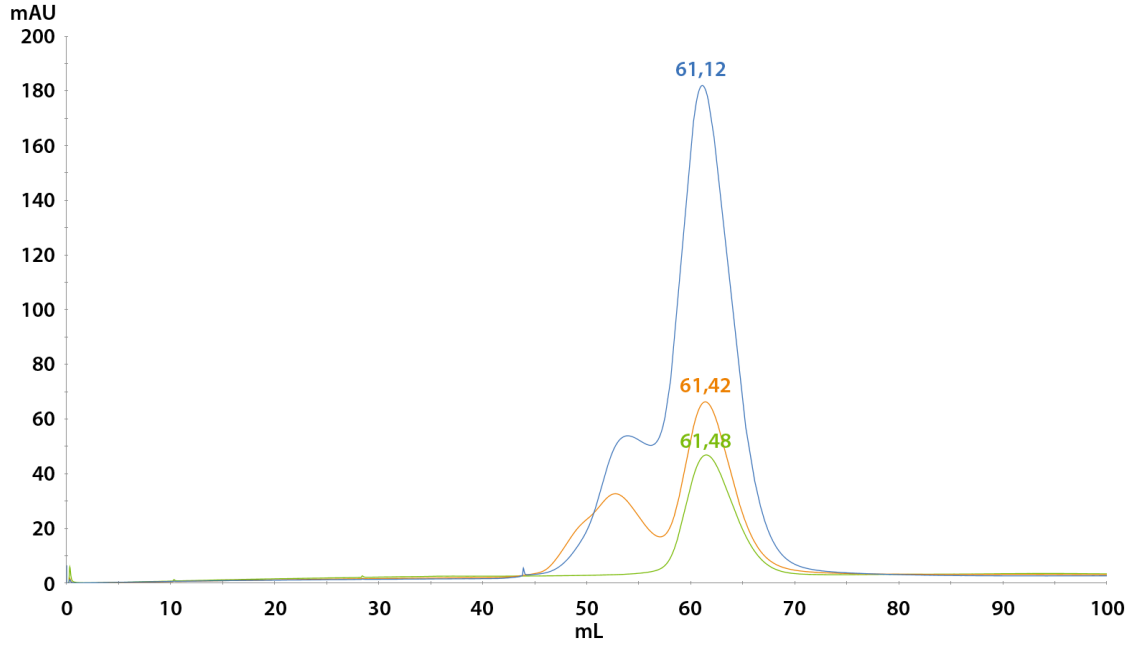


Figure 150 : Chromatogrammes de tamis moléculaires sur colonne préparative des protéines A3sA3. Courbe verte : Protéine A3sA3Y26C ; Courbe bleue : Biohybride A3sA3Y26TPPMn obtenu par la méthode (1) ; Courbe orange : Biohybride A3sA3Y26TPPMn obtenu par la méthode (2) On remarque des agrégats différents pour les différentes méthodes d'extraction.

On constate tout d'abord dans les deux méthodes la présence majoritaire de protéine au volume d'élution correspondant à la protéine non fonctionnalisée. Cependant la présence d'un pic de haut rayon hydrodynamique, peut-être de multimères comme vu dans le cas des biohybrides fonctionnalisés avec la terpyridine est observé. D'une part, il n'est pas étonnant de constater des formes mutimériques ou agrégées au vu des conditions rudes d'extraction et d'autre part, il est intéressant de constater que les deux méthodes ne conduisent pas aux mêmes types de contaminants. De manière intéressante, les volumes

d'élution des pics majoritaires sont légèrement différents, avec 61,12 mL pour le biohybride obtenu par la méthode (1) contre 61,42 mL et 61,48 mL pour le biohybride obtenu par la méthode (3) et la protéine non fonctionnalisée, respectivement. Cette faible différence a d'abord été perçue comme peu significative du fait de l'utilisation d'une colonne préparative, elle a cependant été observée plusieurs fois pour des préparations indépendantes, ce qui invite à penser qu'elle aurait un sens.

Les pics d'intérêt des biohybrides obtenus avec les deux différentes méthodes ont été analysés par SDS-PAGE (figure 151), il en ressort que le biohybride extrait selon la méthode (1) forme une bande plus diffuse à la taille attendue (ligne 1) comparé au biohybride extrait selon la méthode (2), qui forme une bande compacte et unique à la taille attendue (ligne 2). Les lignes 3 et 4 représentent le pic contaminant lors de l'injection du tamis moléculaire du biohybride obtenu par la méthode (2). On constate effectivement la présence de multimères résistants à la dénaturation thermique, comme lors des expériences préliminaires.

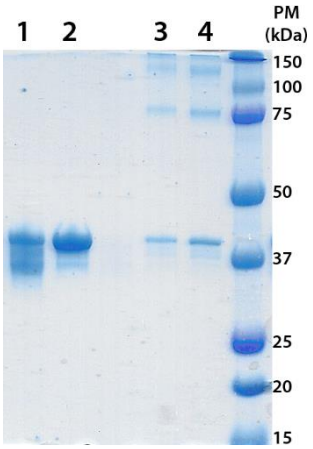


Figure 151 : Gel SDS-PAGE des biohybrides obtenus par les deux méthodes présentées. 1. Biohybride obtenu après tamis moléculaire suivant la méthode (1) 2. Biohybride obtenu après tamis moléculaire suivant la méthode (2) 3-4. Fractions du pic de haut rayon hydrodynamique après SDS-PAGE de la méthode (2). Masse attendue du biohybride : 43 kDa.

La méthode (2) a donc été privilégiée pour plusieurs raisons : sa rapidité et simplicité, la pureté de la protéine observée sur SDS-PAGE, le volume d'élution du biohybride plus proche de la protéine non fonctionnalisée et la plus grande séparation entre le biohybride et les multimères permettant d'obtenir une solution homogène, ce qui n'est pas possible avec la méthode (1). Il n'est pas possible de donner une explication rationnelle à la différence de comportement du biohybride entre les deux méthodes employées. Il pourrait s'agir d'un effet dû à la présence d'une concentration différente de DMSO, en effet ce solvant est éliminé par dialyse ou dilution en même temps que la butanone dans la méthode (1), tandis qu'il est concentré par sa cristallisation dans la méthode (2), or il a été montré que le DMSO pouvait agir spécifiquement sur les hélices α en fonction de sa concentration.[34]

La quantité de protéine perdue peut sembler conséquente (~50%), cependant il existe très peu d'information dans la littérature sur les rendements de la méthode de Teale, le protocole étant lui-même très imprécis. Il est sûrement possible d'améliorer le rendement de la préparation de notre biohybride, nous avons néanmoins continué avec cette préparation.

Des analyses complémentaires de ces biohybrides par dichroïsme circulaire permettront de vérifier si la protéine a bien conservé l'intégralité de ses structures secondaires, et donc son repliement lors de ces traitements en présence de solvant organique. Ne disposant à ce stade que d'une quantité limitée de matériel, nous avons fait le choix de l'utiliser pour les essais préliminaires d'activité catalytique. Etant donnés les profils de tamis analytique et les propriétés biophysiques des αRep, nous avons postulé que la protéine a bien été renaturée dans le tampon une fois le solvant éliminé. Ayant enfin obtenu le biohybride A3sA3Y26TPPMn, nous avons donc pu tenter d'analyser sa réactivité.

II.4.3.2 Test d'oxydation de la tétraméthylbenzidine

Comme réaction test nous avons utilisé la catalyse de l'oxydation de la tétramethylbenzidine (TMB) par le peroxyde d'hydrogène (H_2O_2). Deux degrés d'oxydation sont possibles avec ce substrat, les deux produits absorbant à des longueurs d'onde différentes (figure 152). La première oxydation du TMB donne le produit radical cation $TMB^{\bullet+}$ formant dynamiquement un complexe de transfert de charge avec une autre molécule de TMB. Ce complexe de transfert présente deux maximums d'absorption, à 370 nm et 652 nm (ε = 39 000 $M^{-1}.cm^{-1}$).[35] Après une deuxième étape d'oxydation, le composé diimine est formé, absorbant à 450 nm (ε= 59 000 $M^{-1}.cm^{-1}$). [36,37]

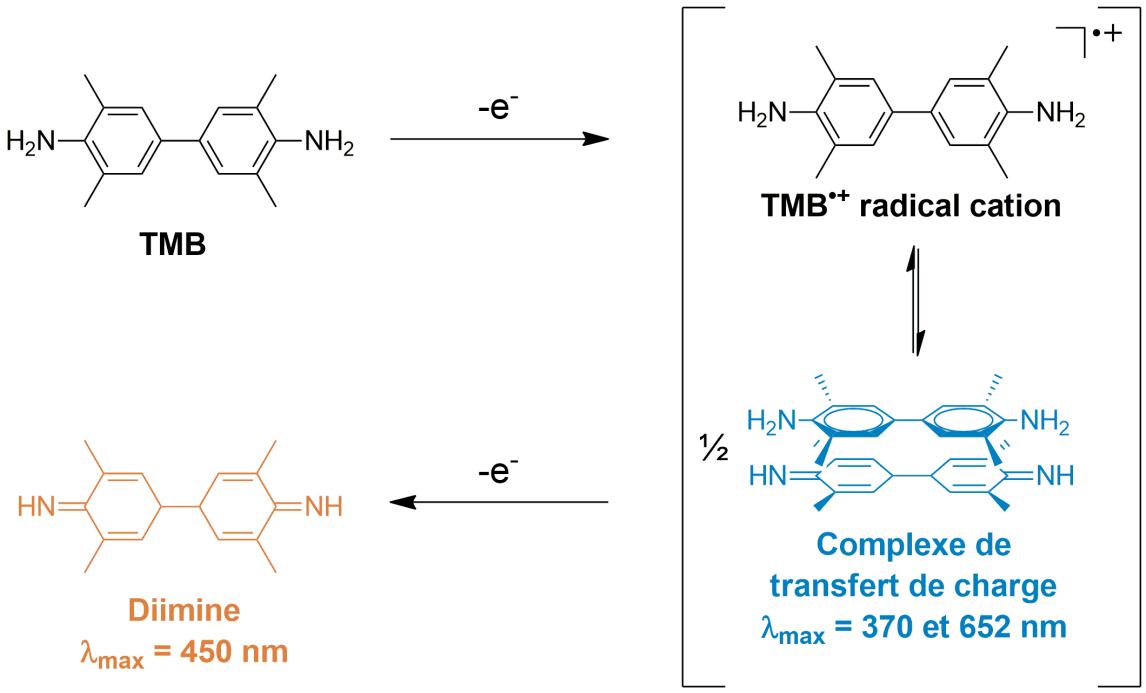


Figure 152 : Formation des produits d'oxydation du TMB. Un premier transfert d'électron conduit à l'espèce radical cation TMB•+, formant un complexe de transfert de charge absorbant à 370 et 652 nm, une deuxième étape d'oxdation donne le composé diimine absorbant à 450 nm.

Nous avons choisi d'effectuer les tests dans les meilleures conditions pour observer l'effet catalytique de la porphyrine libre en solution Mn(TPP). Il a donc fallu ajouter 20% d'acétonitrile pour favoriser la solubilité de cette porphyrine très hydrophobe, et utiliser des tampons sans NaCl afin de ne pas faire précipiter le substrat TMB. Les résultats de cinétique suivie à 652 nm sont reportés sur la figure 153.

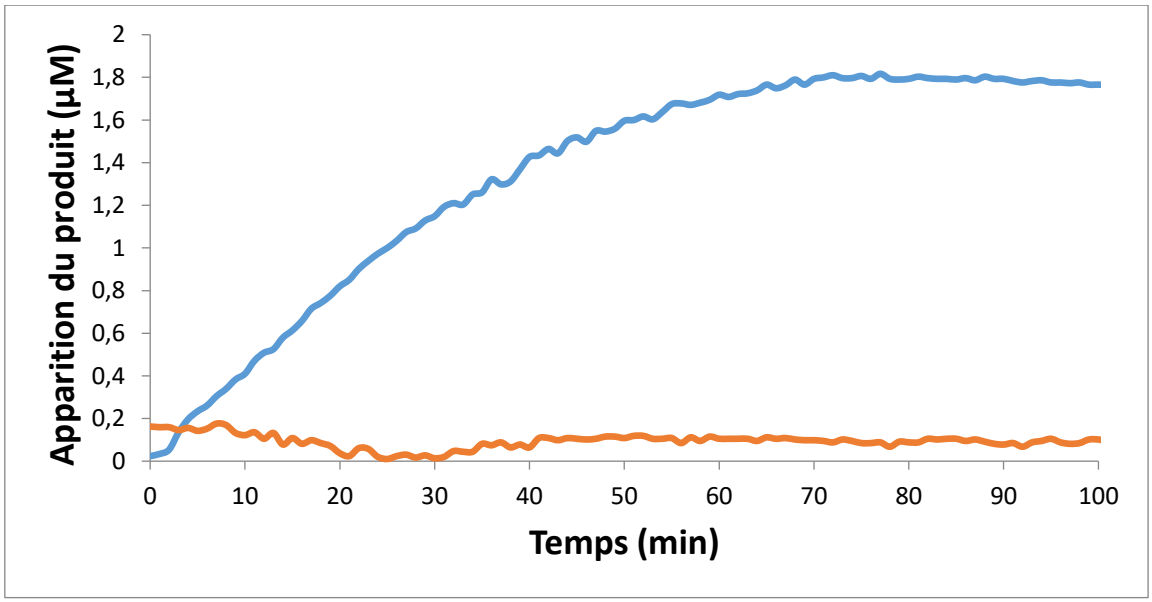


Figure 153 : Cinétique d'oxydation du substrat TMB par le complexe Mn(TPP) (bleu) ou le biohybride A3sA3Y26TPPMn (orange) suivi à 652 nm. 5 µM de catalyseur, 20% ACN, 250 µM H_2O_2, 250 µM TMB, HEPES 10 mM pH 7, 0 mM NaCl.

Voyant que le biohybride ne possédait aucune activité de type peroxydase, il a été supposé que l'oxydant et/ou le substrat ne pouvaient entrer au contact du cofacteur afin d'effectuer la catalyse. Par ailleurs, lors d'autres expériences d'oxydation à l'aide d'un oxydant plus puissant (hydrogénopersulfate ou « oxone ® ») j'ai remarqué que le biohybride conservait sa couleur, contrairement à la porphyrine libre ou liée de façon non covalente à la protéine A3sA3WT (figure 154). Ceci suggère que la porphyrine ne serait pas dégradée par cet oxydant au sein du biohybride contrairement à la porphyrine seule en solution. Une telle observation traduirait alors un effet « bouclier protecteur » [38] de la protéine envers la porphyrine.

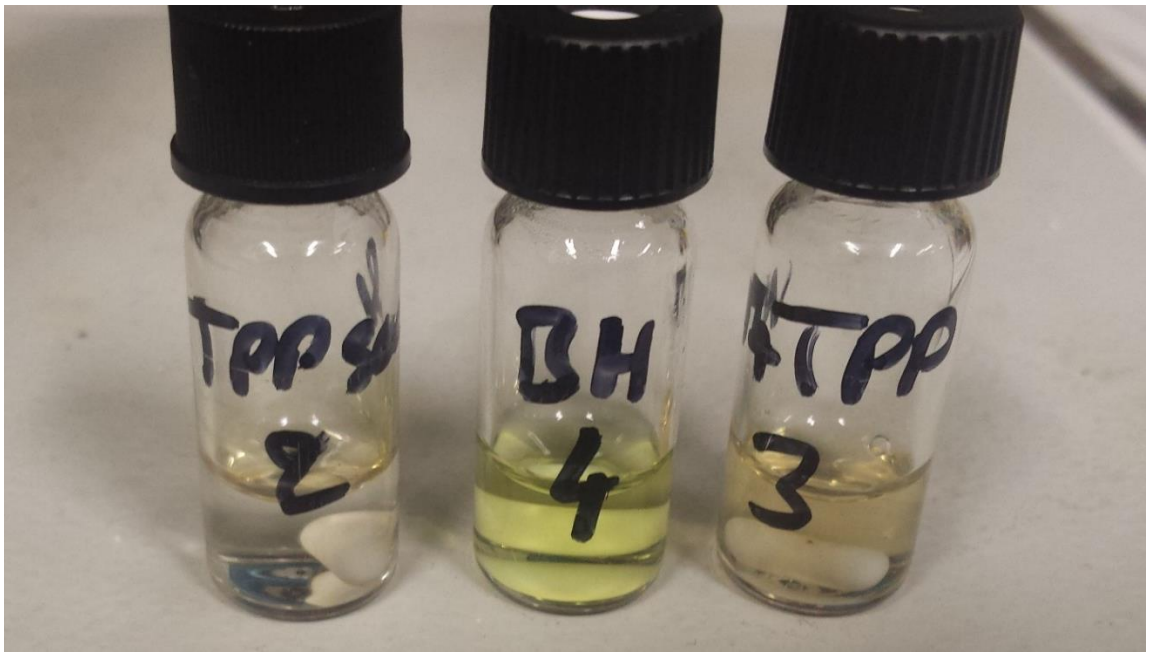


Figure 154 : Catalyseurs mis en présence de 20 mM d'oxone pendant 2h. Gauche : MnTPP seule. Milieu : Biohybride A3sA3Y26TPPMn. Droite : Hôte-invité A3sA3WT+MnTPP. Seul le biohybride conserve la couleur d'origine, signe d'un effet « bouclier protecteur ».

II.4.3.3 Effet bouclier protecteur

Nous avons voulu avoir une preuve plus précise de l'effet bouclier protecteur, j'ai pour cela effectué un ajout de 27 µM (20 équivalents) d'oxone sur la porphyrine seule ou le biohybride et un spectre UV-visible entre 350 et 600 nm a été enregistré. Nous avons pu observer qu'immédiatement après l'ajout de l'oxydant sur la porphyrine libre, sa bande de Soret à 471 nm disparaissait totalement, indiquant la destruction de la porphyrine. En revanche aucun effet sensible n'était observé sur le spectre du biohybride. Et celui-ci reste relativement constant malgré 6 ajout successifs de 27 µM (162 µM, 120 équivalents) d'oxone pendant une heure (figure 155). Ces résultats montrent que la porphyrine est très fortement protégée d'un oxydant aussi puissant que l'oxone.

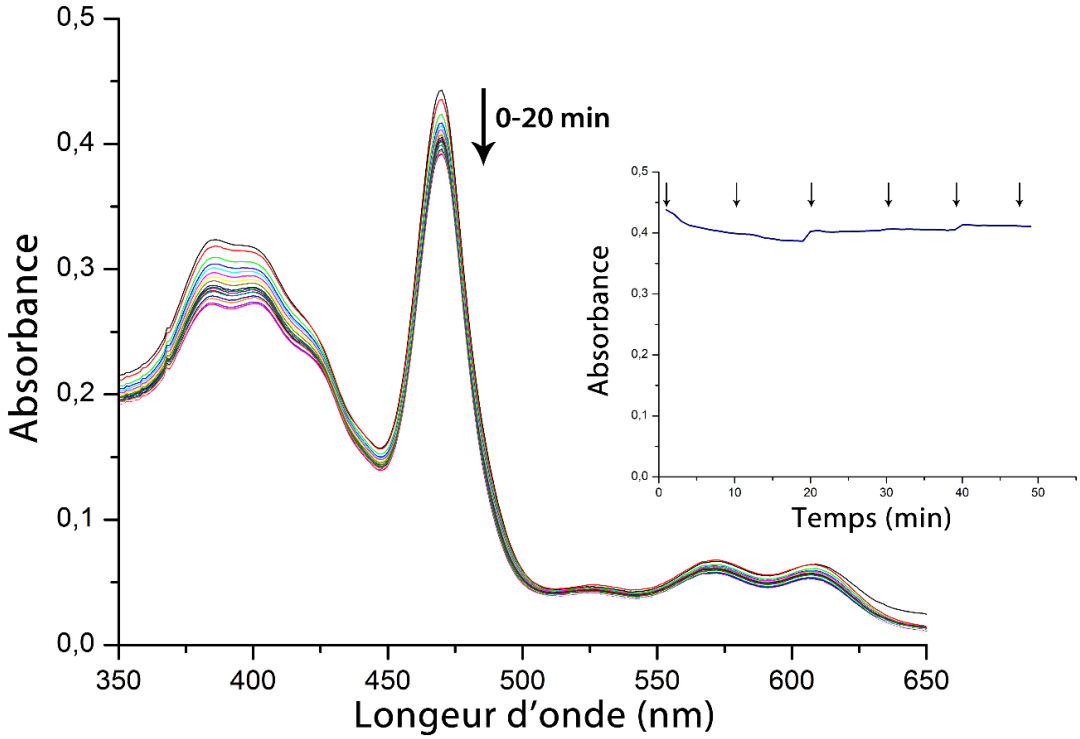


Figure 155 : Évolution du spectre d'absorption du biohybride A3sA3Y26TPPMn après ajout de 27 µM d'oxone (20 équivalents) toutes les 10 minutes. Encart : Suivi de la bande de Soret (à 472 nm) de la porphyrine au cours du temps, les flèches représentent les ajouts d'oxone.

Ces résultats semblent confirmer que la porphyrine enfouie dans la crevasse de la protéine est inaccessible à l'oxydant, et ne peut donc pas effectuer de catalyse. Il est donc nécessaire d'augmenter l'accessibilité de la porphyrine si on veut l'utiliser pour catalyser une réaction.

II.4.3.4 Ouverture du dimère par la protéine α2

Une protéine de la banque αRep, que nous appellerons α2, avait été sélectionnée par la technique de Phage Display contre la protéine αRep-A3.[39] Cette protéine possède un seul module interne, donc au total trois motifs en comptant les modules N-cap et C-cap. Des expériences de titration isotherme (ITC) menées au laboratoire ont conclu à une constante de dissociation apparente de la protéine α2 pour le monomère A3 de 3,7 nM et une stœchiométrie 1:1. La structure cristallographique du complexe A3/α2 a montré que les deux protéines interagissent via leurs surfaces variables respectives, ce qui implique que la protéine α2 dissocie le dimère A3 avant de se fixer (figure 156).

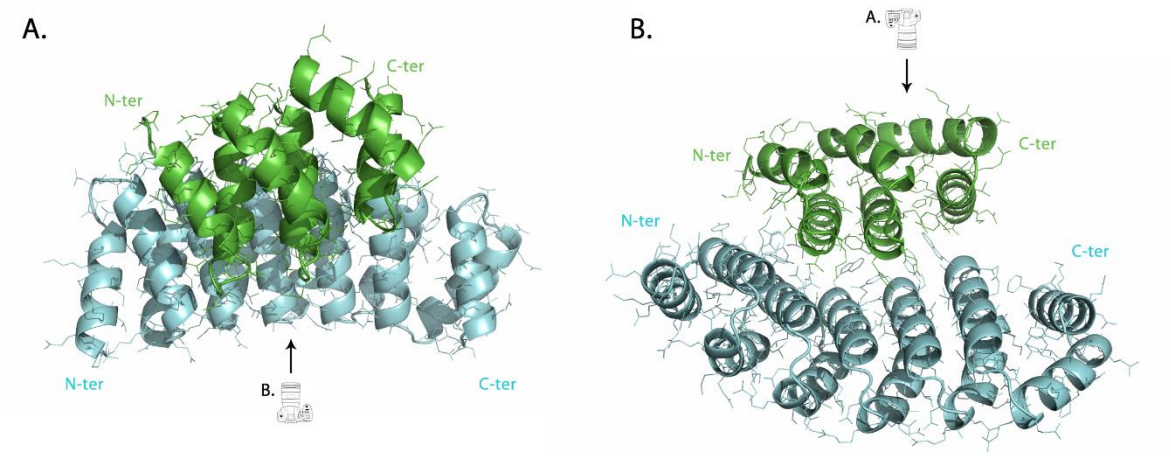


Figure 156 : Modèle d'après la structure cristallographique A3/α2 (PDB : 4JW2) sous différent angles (indiqués par les flèches). La protéine α2 (en vert) lie la protéine A3 (en bleu) au niveau de sa face variable, ouvrant le dimère. Dans la structure cristallographique la protéine α2 présente un clivage puis réarrangement de l'hélice du module C-terminal, ce clivage a été montré comme non nécessaire pour l'interaction.[39]

Il nous est apparu intéressant d'utiliser ce partenaire protéique afin d'ouvrir le biohybride A3sA3Y26TPPMn. Cette dernière possédant un domaine non modifié, il est toujours possible que la protéine α2 s'y fixe. L'autre domaine étant fonctionnalisé par la porphyrine au niveau de la surface d'interaction A3/α2, il est peu probable qu'une fixation soit possible.

Nous avons donc effectué la même expérience d'oxydation du TMB en présence de 1 ou 5 équivalents de protéine α2. Les résultats sont reportés sur la figure 157.

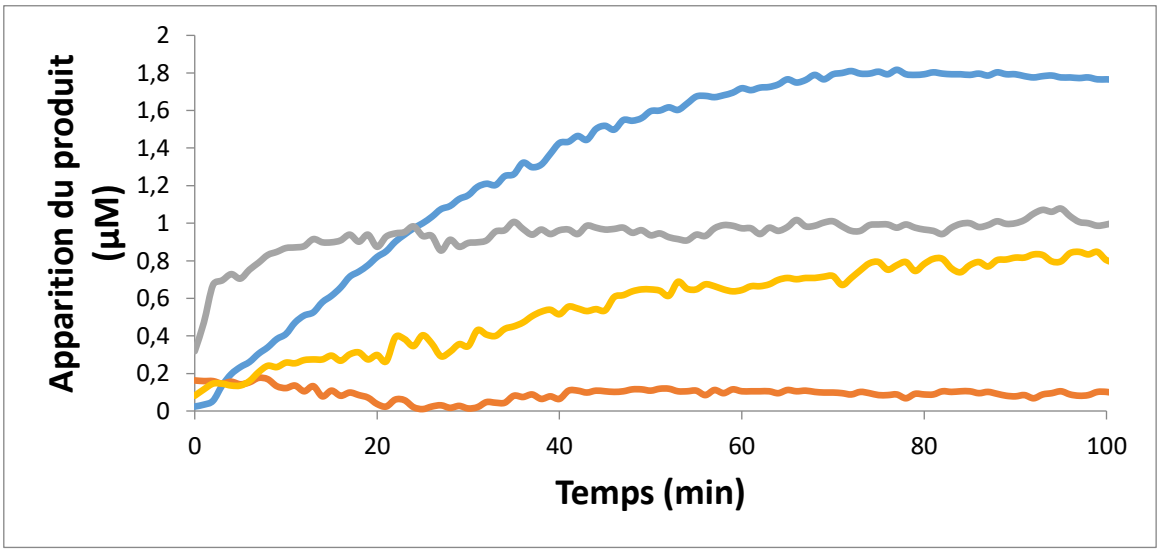


Figure 157 : Cinétique d'oxydation du substrat TMB par le complexe Mn(TPP) (bleu), le biohybride A3sA3Y26TPPMn seul (orange), le biohybride A3sA3Y26TPPMn:α2 en proportion 1:1 (courbe jaune), le biohybride A3sA3Y26TPPMn:α2 en proportion 1:5 (courbe grise) suivi à 652 nm. La protéine α2 seule a la même activité que le substrat dans le tampon (basale). 5 µM de catalyseur, 20% ACN, 250 µM H2O2, 250 µM TMB, HEPES 10 mM pH 7, 0 mM NaCl.

On peut remarquer qu'après l'ajout de la protéine α2, les courbes de cinétiques du biohybride se détachent du bruit de fond, ce qui indique clairement une activité catalytique. Cette observation confirme que la protéine α2 peut augmenter l'accessibilité de la porphyrine certainement en ouvrant la protéine A3sA3. Ceci confirme de manière indirecte que le biohybride A3sA3Y26TPPMn a conservé sa conformation tridimensionnelle, lui permettant d'interagir avec son partenaire protéique. On peut également observer que pour un ratio A3sA3:α2 de 1:1 (courbe jaune) la vitesse de réaction est plus lente que pour un ratio A3sA3:α2 de 1:5 (courbe grise). En effet, le biohybride en présence d'un excès de protéine α2 atteint très vite un plateau (environ 20 min) alors que la courbe jaune semble presque linéaire jusqu'à 100 min. Il est à noter que la cinétique du biohybride avec un excès de protéine α2 atteint plus vite le plateau que la porphyrine en solution, mais que ce plateau est plus bas. La porphyrine semble donc plus active au sein de la protéine mais elle s'inactive également plus vite. Il sera nécessaire de reproduire et d'effectuer de nouveaux essais de catalyse dans différentes conditions afin de confirmer et d'expliquer ces résultats très prometteurs.

II.5 Résumé de l'utilisation des variants de A3sA3 pour le greffage d'une porphyrine

Nous avons vu au cours de cette partie que la porphyrine MnTPP pouvait être couplée covalemment aux différents variants de la protéine A3sA3. Ce lien covalent permet de se débarrasser des porphyrines liées de façon non spécifique à la protéine grâce à l'extraction par la méthode de Teale modifiée. Ceci confère plusieurs avantages à notre système: le biohybride une fois isolé ne forme plus d'oligomères solubles, le catalyseur est donc plus homogène. La porphyrine est également solubilisée, ce qui représentait un défi pour un cofacteur aussi hydrophobe que la TPP. Enfin, il a été observé que la protéine peut protéger la porphyrine à l'encontre de l'oxydation par des oxydants forts tels que l'oxone.

La porphyrine ainsi protégée est également inactive, car inaccessible aux substrats. L'utilisation d'un partenaire protéique de la protéine A3 (α2) semble réussir à ouvrir le complexe et de ce fait favoriser l'accès de la porphyrine aux substrats. Il en résulte que le biohybride A3sA3Y26TPPMn est un catalyseur commutable, protégeant la porphyrine des dégradations en milieu oxydatif et ne libérant l'activité catalytique qu'en présence de son partenaire protéique α2.

Si ces résultats très encourageants sont confirmés, il s'agirait à notre connaissance du premier catalyseur biohybride commutable par ajout d'un partenaire protéique conçu à ce jour.

Bibliographie

(1) Kortemme, T., Kim, D. E., et Baker, D. (2004) Computational Alanine Scanning of Protein-Protein Interfaces. *Sci. Signal. 2004*, pl2-pl2.

(2) Kim, D. E., Chivian, D., et Baker, D. (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res. 32*, 526-531.

(3) Bijloo, G. J., van der Goot, H., Bast, A., et Timmerman, H. (1990) Copper complexes of 1,10-phenanthroline and related compounds as superoxide dismutase mimetics. *J. Inorg. Biochem. 40*, 237-244.

(4) Kellett, A., O'Connor, M., McCann, M., McNamara, M., Lynch, P., Rosair, G., McKee, V., Creaven, B., Walsh, M., McClean, S., Foltyn, A., O'Shea, D., Howe, O., et Devereux, M. (2011) Bis-phenanthroline copper(II) phthalate complexes are potent in vitro antitumour agents with « self-activating » metallo-nuclease and DNA binding properties. *Dalt. Trans. 40*, 1024-7.

(5) Kobashi, K. (1968) Catalytic oxidation of sulfhydryl groups by o-phenanthroline copper complex. *Biochim. Biophys. Acta - Gen. Subj. 158*, 239-245.

(6) Hirohama, T., Kuranuki, Y., Ebina, E., Sugizaki, T., Arii, H., Chikira, M., Selvi, P. T., et Palaniandavar, M. (2005) Copper(II) complexes of 1,10-phenanthroline-derived ligands: Studies on DNA binding properties and nuclease activity. *J. Inorg. Biochem. 99*, 1205-1219.

(7) Bos, J., Fusetti, F., Driessen, A. J. M., et Roelfes, G. (2012) Enantioselective artificial metalloenzymes by creation of a novel active site at the protein dimer interface. *Angew. Chem. Int. Ed. Engl. 51*, 7472-5.

(8) Sharpl, P. M., et Li, W. (1987) The codon Adaptation Index - A measure of directional synonymous codon usage bias, and its possible applications. *Nucleic Acids Res. 15*, 1281-1295.

(9) Ellman, G. L. (1959) Tissue Sulfyhydryl Groups. *Arch. Biochem. Biophys.* 70-77.

(10) Yamada, K., Takahashi, Y., Yamamura, H., Araki, S., Kawai, M., et Saito, K. (2000) Phosphodiester bond cleavage mediated by a cyclic β-sheet peptide-based dinuclear zinc(ii) complex. *Chem. Commun. 2*, 1315-1316.

(11) Daver, H., Das, B., Nordlander, E., et Himo, F. (2016) Theoretical Study of Phosphodiester Hydrolysis and Transesterification Catalyzed by an Unsymmetric Biomimetic Dizinc Complex. *Inorg. Chem. 55*, 1872-1882.

(12) Otto, S., et Engberts, J. B. F. N. (2000) Diels–Alder reactions in water. *Pure appl. chem. 72*, 1365-1372.

(13) Otto, S., Bertoncin, F., et Engberts, J. B. F. N. (1996) Lewis acid catalysis of a Diels-Alder reaction in water. *J. Am. Chem. Soc. 118*, 7702-7707.

(14) Yamada, S., et Tokugawa, Y. (2009) Cation - π Controlled Solid-State Photodimerization of 4-Azachalcones 2098-2099.

(15) Yayli, N., Üçüncü, O., Yaşar, A., Yayli, N., Burnaz, N. A., Karaoğlu, Ş. A., et Küçük, M. (2009) Photochemistry of nitro-substituted (E)-2-azachalcones with theoretical calculations and biological activities. *J. Photochem. Photobiol. A Chem. 203*, 85-91.

(16) De Robertis, A., Foti, C., Gianguzza, A., et Rigano, C. (1996) Protonation Thermodynamics of 1 , 10-Phenanthroline in Aqueous Solution . Salt Effects and Weak Complex Formation. *J. Solut. Chem. 25*, 597-606.

(17) Podtetenieff, J., Taglieber, A., Bill, E., Reijerse, E. J., et Reetz, M. T. (2010) An artificial metalloenzyme: Creation of a designed copper binding site in a thermostable protein. *Angew. Chemie - Int. Ed. 49*, 5151-5155.

(18) Hönicke, D., Födisch, R., Claus, P., et Olson, M. (2000) Cyclopentadiene and Cyclopentene, in *Ullmann's Encyclopedia of Industrial Chemistry*. CHAP, Wiley-VCH Verlag GmbH & Co. KGaA.

(19) Kumar, S., Pal Sharma, R., Venugopalan, P., Jezierska, J., Wojciechowska, A., et Ferretti, V. (2015) Synthesis, Characterization and X-ray Diffraction/Structural Studies of three Hybrid Inorganic-organic Compounds: Silver(I)-, Lead(II)- and Tris(phenanthroline)copper(II)- 2,6-naphthalenedisulfonate (2,6-nds). *Inorganica Chim. Acta 432*, 221-230.

(20) Zhi-min, C., Wei, L., Ying-qun, Y., Dai-zhi, K., Yong-lan, F., Jian-qiu, W., et Fu-xing, Z. (2005) Study on Synthesis, Structure and Quantum Chemistry of the Three-dimentional Network Complex [Cu(Phen)2NO3](OH)·6H2O. *J.Nat.Sci.Hunan Norm. Univ.* 54-58.

(21) Lacková, K., et Potočňák, I. (2012) (Nitrato-κ 2 O , O ')bis(1,10-phenanthroline-κ 2 N , N ')copper(II) tricyanomethanide. *Acta Crystallogr. Sect. E Struct. Reports Online 68*, m1553-m1554.

(22) Ovens, J. S., Geisheimer, A. R., Bokov, A. A., Ye, Z.-G., et Leznoff, D. B. (2010) The Use of Polarizable [AuX 2 (CN) 2] − (X = Br, I) Building Blocks Toward the Formation of Birefringent Coordination Polymers. *Inorg. Chem. 49*, 9609-9616.

(23) Molphy, Z., Slator, C., Chatgilialoglu, C., et Kellett, A. (2015) DNA oxidation profiles of copper phenanthrene chemical nucleases. *Front. Chem. 3*, 28.

(24) Carey, J. R., Ma, S. K., Pfister, T. D., Garne, D. K., Kim, H. K., Abramite, J. a., Wang, Z., Guo, Z., et Lu, Y. (2004) A site-selective dual anchoring strategy for artificial metalloprotein design. *J. Am. Chem. Soc. 126*, 10812-10813.

(25) van Dongen, E. M. W. M., Evers, T. H., Dekkers, L. M., Meijer, E. W., Klomp, L. W. J., et Merkx, M. (2007) Variation linker length in ratiometric fluorescent sensor protein allows rational tuning of Zn(II) affinity in the pM to fM range [CFP Atox1 FRET to YFP]. *J. Am. Chem. Soc. 129*, 3494-3495.

(26) Chen, X., Zaro, J. L., et Shen, W. C. (2013) Fusion protein linkers: Property, design and functionality. *Adv. Drug Deliv. Rev. 65*, 1357-1369.

(27) Himiyama, T., Sauer, D. F., Onoda, A., Spaniol, T. P., Okuda, J., et Hayashi, T. (2016) Construction of a hybrid biocatalyst containing a covalently-linked terpyridine metal complex within a cavity of aponitrobindin. *J. Inorg. Biochem. 158*, 55-61.

(28) Wang, J. H. (1964) Porphyrins and Metalloporphyrins. *Yale J. Biol. Med. 37*, 170.

(29) Mahy, J.-P., Maréchal, J.-D., et Ricoux, R. (2015) From « hemoabzymes » to « hemozymes »: towards new biocatalysts for selective oxidations. *Chem. Commun. 51*, 2476-2494.

(30) Pasternack, R. F., Francesconi, L., Raff, D., et Spiro, E. (1973) Aggregation of Nickel(II), Copper(II), and Zinc(II) Derivatives of Water-Soluble Porphyrins. *Inorg. Chem. 12*, 2606-2611.

(31) Kim, J.-O., Lee, Y.-A., Yun, B. H., Han, S. W., Kwag, S. T., et Kim, S. K. (2004) Binding of meso-tetrakis(N-methylpyridinium-4-yl)porphyrin to AT oligomers: effect of chain length and the location of the porphyrin stacking. *Biophys. J. 86*, 1012-7.

(32) Teale, F. W. (1959) Cleavage of the haem-protein link by acid methylethylketone. *Biochim. Biophys. Acta 35*, 543.

(33) Urvoas, A., Guellouz, A., Valerio-Lepiniec, M., Graille, M., Durand, D., Desravines, D. C., van Tilbeurgh, H., Desmadril, M., et Minard, P. (2010) Design, production and molecular structure of a new family of artificial alpha-helicoidal repeat proteins (αRep) based on thermostable HEAT-like repeats. *J. Mol. Biol. 404*, 307-27.

(34) Batista, A. N. L., Batista, J. M., Bolzani, V. S., Furlan, M., et Blanch, E. W. (2013) Selective DMSO-induced conformational changes in proteins from Raman optical activity. *Phys. Chem. Chem. Phys. 15*, 20147-52.

(35) Josephy, D., Eling, T., et Mason, R. (1982) The Horseradish Peroxidase-catalyzed Oxidation of 3,5,3',5'- Tetramethylbenzidine. *J. Biol. Chem. 257*, 3669-3675.

(36) Josephy, P. D., Eling, T., et Mason, R. P. (1982) The Horseradish Peroxidase-catalyzed Oxidation of 3,5,3',5'- Tetramethylbenzidine. *J. Biol. Chem. 257*, 3669-3675.

(37) Marquez, L. a, et Dunford, H. B. (1997) Mechanism of the oxidation of 3,5,3',5'-tetramethylbenzidine by myeloperoxidase determined by transient- and steady-state kinetics. *Biochemistry 36*, 9349-9355.

(38) Allard, M. (2011) Élaboration de nouvelles métalloenzymes artificielles pour des réactions d'oxydation sélectives selon différentes stratégies. Université Paris Sud XI.

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

(40) Bos, J., Fusetti, F., Driessen, A. J. M., et Roelfes, G. (2012) Enantioselective artificial metalloenzymes by creation of a novel active site at the protein dimer interface. Angew. Chem. Int. Ed. Engl. 51, 7472-5.

(41) Filice, M., Romero, O., Gutiérrez-Fernández, J., de las Rivas, B., Hermoso, J. a., et Palomo, J. M. (2015) Synthesis of a heterogeneous artificial metallolipase with chimeric catalytic activity. Chem. Commun. 51, 9324-9327.

(42) Reetz, M. T., et Jiao, N. (2006) Copper-phthalocyanine conjugates of serum albumins as enantioselective catalysts in Diels-Alder reactions. Angew. Chemie - Int. Ed. 45, 2416-2419.

(43) Ghattas, W., Cotchico-Alonso, L., Maréchal, J.-D., Urvoas, A., Rousseau, M., Mahy, J.-P., et Ricoux, R. (2016) Artificial Metalloenzymes with the Neocarzinostatin Scaffold: Toward a Biocatalyst for the Diels–Alder Reaction. ChemBioChem 17, 433-40.

(44) Himiyama, T., Sauer, D. F., Onoda, A., Spaniol, T. P., Okuda, J., et Hayashi, T. (2016) Construction of a hybrid biocatalyst containing a covalently-linked terpyridine metal complex within a cavity of aponitrobindin. J. Inorg. Biochem. 158, 55-61.

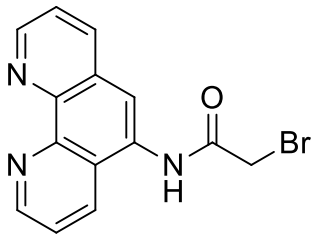
(45) Osseili, H., Sauer, D. F., Beckerle, K., Arlt, M., Himiyama, T., Polen, T., Onoda, A., Schwaneberg, U., Hayashi, T., et Okuda, J. (2016) Artificial Diels-Alderase based on the transmembrane protein FhuA. Beilstein J. Org. Chem. 12, 1314-1321.

Partie C - Matériel et méthodes

Chapitre I - Chimie

I.1 Synthèse des ligands

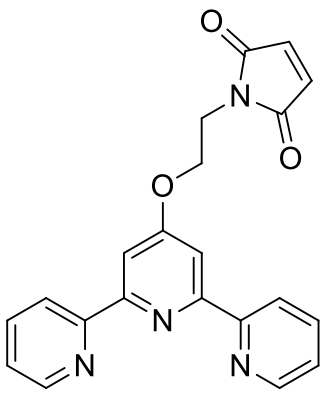
I.1.1 Synthèse du ligand 2-bromo-N-(1,10-phenanthrolin-5-yl)acetamide hydrobromide



2

195 mg (1 mmol, 1 équiv.) de 1,10-phenantroline-5-amine sont suspendus dans 30 mL de Dichlorométhane distillé. La solution est placée dans un ballon bicol sous atmosphère d'argon et sous agitation à reflux. 243 mg de Bromoacetyl Bromide (1,2 mmol, 1,2 équiv.) sont ajoutés au goutte à goutte et le mélange est mis à incuber sur la nuit. Le produit de réaction est enfin cristallisé au méthanol chaud afin de récupérer le produit sous forme d'une poudre jaunâtre. Le produit est analysé en ESI-ToF à l'aide d'un instrument micrOToF-Q II en mode posifif. Masses : deux doublets (316 Da;318 Da) et (317 Da;319 Da) caractéristiques d'un massif isotopique de Brome (Masse moyenne attendue : 316,007 Da, masse moyenne mesurée : 316,008 Da). RMN ^{1}H (d-DMSO, 360 MHz, δ en ppm/TMS) : 4.35 (s, 2H) 8.21 (dd, J = 8.3, 4.9, 2H), 8.31 (s, 1H), 8.57 (s, 1H), 9.05 (d, J = 7.7, 1H), 9.15 (d, J = 7.5, 1H), 9.21 (d, J = 5.0, 1H), 9.33 (d, J = 3.6, 1H), 10.90 (s, 1H).

I.1.2 Synthèse du ligand (N-[2-((2,2':6',2''-terpyridin)-4'-yloxy)éthyl]maléimide

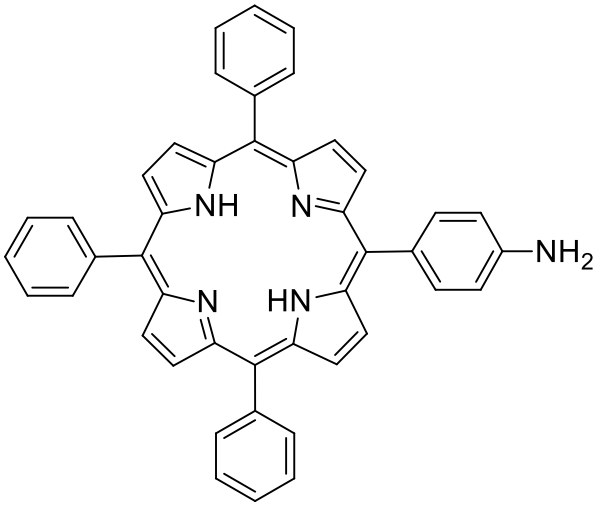


3

219 mg de NaHCO3 et 238 mg de N-méthoxycarbonyl-maléimide sont dissous dans 10 mL de mélange acétone/eau (2/1) à 4°C. 136,4 mg de 2((2,2':6',2''-terpyridin)-4'-yloxy))éthylamine sont dissous dans 10 mL d'acétone/eau (2/1). Cette solution est ensuite ajoutée goutte à goutte à la solution précédente et le mélange réactionnel est laissé 1h30 sous agitation à 4°C. 3 mL d'eau distillée sont alors ajoutés au mélange réactionnel, qui est ensuite laissé sous agitation à température ambiante pendant 2h30. Le précipité obtenu est récupéré par filtration sur Büchner et lavé avec un mélange acétone/eau (4/1). Le produit obtenu est un solide blanc (m = 158,2 mg, Rendement = 80 %) ESI-micrOToF-QII : m/z = 373.129 Da (calculée pour M^+ : 373.128 Da) RMN ^{1}H (CDCl3, 360 MHz, δ en ppm/TMS) : 4,065 (t, J=5,6 Hz, 2H), 4,435 (t, J=5,6 Hz, 2H), 6,778 (s, 2H), 7,34 (dd, J=4,9 Hz, 1H), 7,38 (dd, J=4,7 Hz, 1H), 7,88 (td, J=7,7 Hz, 2H), 8,011 (s, 2H), 8,61 (dt, J=8, 2H), 8,74 (d, J=4,8 Hz, 2H).

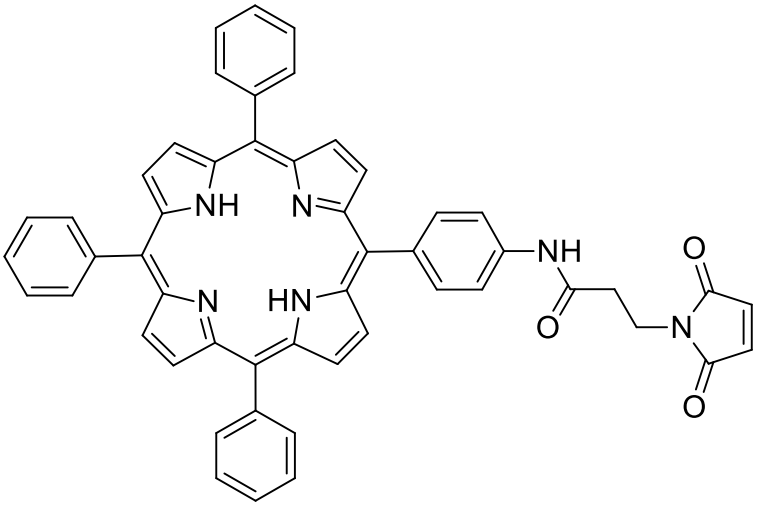
I.1.3 Porphyrine MnTPP-mal

I.1.3.1 5-(4-Aminophényl)-10,15,20-triphénylporphyrine



La 5-(4-Aminophényl)-10,15,20-triphénylporphyrine est préparée en deux étapes par nitration puis réduction de la tetraphénylporphyrine (TPPH2). 315 mg (0,5 mmol, 1 équiv.) de tetraphénylporphyrine sont solubilisés dans 12 mL d'acide trifluoroacétique (TFA), puis 1,4 équiv. de NaNO2 sont ajoutés à température ambiante sous agitation. Après avoir laissé la réaction s'opérer durant 3 minutes seulement afin d'empêcher les réactions de polynitration, 60 mL d'eau sont ajoutés pour arrêter la réaction. La solution est extraite au dichlorométhane (DCM) puis la phase organique est lavée 3 fois avec une solution saturée en NaHCO3 et 2 fois avec une solution saturée en NaCl. Après évaporation sous pression réduite, on obtient un mélange de TPPH2 et de 5,10,15-trisphényl-20-(4-nitrophényl)porphyrine (TPpNO2PH2). Le produit brut (317 mg, ~0,4 mmol, 1 équiv.) est dissout dans 40 mL de HCl 37% et 10 mL de DCM sous atmosphère d'argon. Le mélange est agitée à 0°C pendant 5 minutes et 660 mg (5 mmol, ~12,5 équiv.) de SnCl2 sont ajoutés lentement. La solution est agitée pendant 24h à reflux. Le mélange est refroidi dans un mélange glace-eau-sel et l'acide est neutralisé par ajout progressif de pastilles de KOH jusqu'à ce que la solution prenne une couleur bordeaux. La solution neutralisée est extraite dans du dichlorométhane puis lavée deux fois avec une solution saturée en NaCl. La phase organique est enfin évaporée. Le produit brut est séparé sur une colonne de silice en utilisant un mélange DCM/Cyclohexane de rapport (v/v) allant de 95/5 à 100/0 par paliers de 1%. On obtient 75 mg de cristaux violets correspondant au produit 5-(4-Aminophényl)-10,15,20-triphénylporphyrine (24% de rendement). Rf = 0.7 dans 95:5 DCM:cyclohexane. UV–visible dans DCM: λmax (nm) = 420 (Soret), 518, 555, 592, 649. RMN 1H (CDCl3, 360 MHz, δ en ppm/TMS): 8.91 (d, J = 4.8 Hz, 2H); 8.78 (m, 6H); 8.28 (m, 6H); 8.4 (d, J = 8.3 Hz, 2H); 7.79 (m, 9H); 8.61 (d, J = 8.3 Hz, 2H).

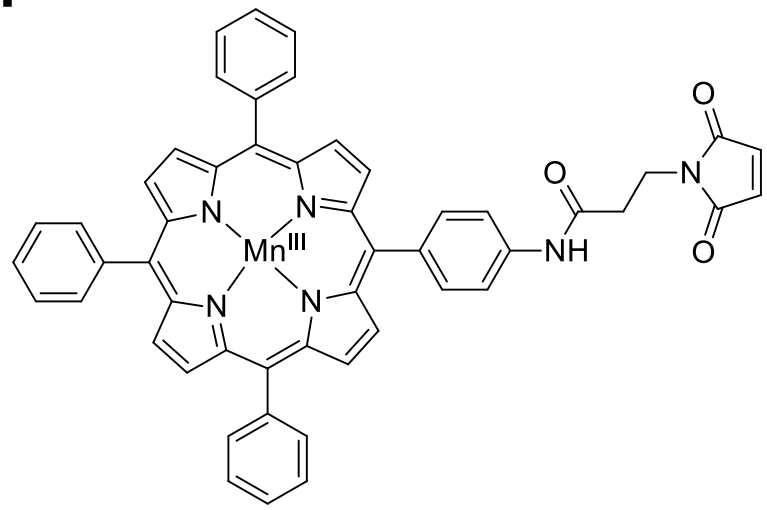
I.1.3.2 5,10,15-triphényl-20-(4-(maléimidopropylamido)phényl)porphyrine



Dans un ballon de 10 mL, 16 mg (25 µmol, 1 équiv.) de (4-Aminophényl)-10,15,20-triphénylporphyrine, 55 mg (325 µmol, 13 équiv.) de d'acide 3-maléimidopropionique et 66 mg (320 µmol, 12,8 équiv.) de 1-éthyl-3-(3-diméthylaminopropyl)carbodiimide sont dissous dans 2 mL de diméthylformamide (DMF). La solution est mélangée durant 18h à température ambiante puis le solvant est évaporé. Le produit obtenu est dissous dans 50 mL de chloroform et la solution est lavée 3 fois avec de l'eau distillée. La phase organique est séchée au sulfate de sodium (Na_2SO_4) puis évaporée. Le produit est purifié sur colonne de silice 60 ACC 70-200 µm avec un mélange DCM/AcOEt de 20/1 (v/v). Les fractions sont collectées pour récolter le produit sous la forme d'un solide violet. RMN ^{1}H (CDCl3, 500 MHz, δ en ppm/TMS): −2.79 (bs, 2H); 2.88 (t, 2H, J = 6.85 Hz), 4.05 (t, 2H, J = 6.85 Hz), 6.78 (s, 2H), 7.75 (m, 9H), 7.87 (d, 2H, J = 8.40 Hz), 8.15 (d, 2H, J = 8.40 Hz), 8.20 (d, 6H, J = 6.15 Hz), 8.84 (bs, 8H) MALDI-ToF, α-CHCA: m/z = 780.8 Da (calculé pour [M + H]+: 781.3 Da).

I.1.3.3 Insertion du manganèse

P1



La base libre de 5,10,15-triphényl-20-(*4*-(maléimidopropylamido)phényl) porphyrine a été traitée avec un large excès de $Mn(OAc)_2,4H_2O$ dans un mélange MeOH/DCM à température ambiante, conduisant au complexe Manganèse(III) 5,10,15-trisphényl-20-(*4*-(maléimidopropylamido)phényl)porphyrine (**P1**). Ce complexe a été caractérisé en spectrophotométrie UV-Visible dans le DCM avec des maximums d'absorption à 376, 401, 474, 578 et 609 nm, caractéristiques d'un complexe Mn(III).

I.2 Synthèse des substrats et authentiques

I.2.1 Transfert hydrolytique de liaison phosphoester

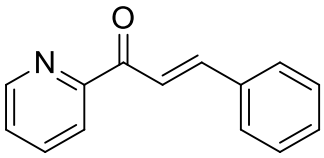
I.2.1.1 2-hydroxypropyl-4-nitrophenyl phosphate (HPNP)

La synthèse du 2-hydroxypropyl-4-nitrophenyl phosphate a été effectuée selon la méthode déjà décrite dans la littérature.[1] 1,32g (5 mmoles) de disodium p-nitrophenyl phosphate (Sigma) est dissout dans 10mL d'eau distillée puis la solution est passée à travers une colonne acide échangeuse d'ions. La solution est ensuite basifiée jusqu'à pH 8 avec une solution aqueuse d'ammoniaque. 20mL de 1,2-Epoxypropane sont ajoutés et le mélangé est placé sous agitation à 35°C pendant 40h. L'époxyde qui n'a pas réagi est évaporé et le mélange est passé dans une colonne acide échangeuse d'ions. La solution est neutralisée sans dépasser un pH de 7 avec une solution d'hydroxyde de barium puis est concentrée jusqu'à un volume d'environ 10 mL. 2 volumes d'EtOH sont ajoutés et le mélange est filtré. Le filtrat est concentré puis 300mL d'un mélange EtOH/Acétone (10/90) est ajouté goutte à goutte. Un précipité blanc se forme qui est filtré sur Büchner.

I.2.2 Diels-Alder

I.2.2.1 Azachalcone

I.2.2.2 (*E*)-3-phényl-1-(pyridine-2-yl)prop-2-ène-1-one (Azachalcone)



3 g de benzaldéhyde (28 mmol, 1.1 éq.) sont mélangés avec 3 g de 2-acétylpyridine (25 mmol, 1 éq.) dans un ballon (50 mL) contenant une solution de soude à 10 % (20 mL), muni d'une agitation magnétique et placé sous argon à l'abri de la lumière dans un bain de glace pendant 4 h.

Le solide formé est collecté par filtration et purifié par chromatographie sur colonne de silice (cyclohexane : acétate d'éthyle = 70 : 30 volumes) pour obtenir 4 g du produit recherché sous forme de solide blanc. Rendement : 77 %.

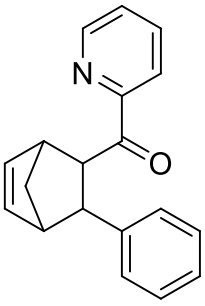
^{1}H RMN ($CDCl_3$, 300 MHz) : 8,73 (ddt, *J* = 4,6 Hz, *J* = 1,7 Hz, *J* = 0,8 Hz, 1H), 8,3 (dd, J_{AB} = 16,1 Hz, *J* = 0,8 Hz, 1H), 8,18 (dq, *J* = 7,8 Hz, *J* = 0,8 Hz, 1H) 7,93 (d, J_{AB} = 16,1 Hz, 1H), 7,86 (tdd, *J* = 7,8 Hz, *J* = 1,7 Hz, *J* = 0,8 Hz, 1H), 7,75 – 7,68 (m, 2H), 7,50 – 7,37(m, 4H) δ ppm.

I.2.2.3 Cyclopentadiène

50 ml de dimère de cyclopentadiène et 100 mg de limaille de fer sont chauffés à 75 C dans un ballon (100 mL) muni d'une colonne de vigreux (30 cm) suivie d'un condensateur droit réfrigéré et d'un collecteur de fraction. Le cyclopentadiene collecté sous forme de liquide transparent est conservé à -20 C.

^{1}H RMN ($CDCl_3$, 300 MHz) : 6,18 – 5,88 (m, 2H), 5,62 – 5,43 (m, 2H), 4,20 – 3,85 (m, 2H) δ ppm.

I.2.2.4 (3-phenylbicyclo[2.2.1]hept-5-en-2-yl)(pyridin-2-yl)methanone



418 mg d'azachalcone (2 mmol, 1 éq.) et 54 mg de chlorure d'aluminium (0,4 mmol, 0,2 éq.) sont dissous dans du dichlorométhane sec (25 mL) dans un ballon (50 mL) sous argon muni d'une agitation magnétique et placé dans un bain de glace. Une solution de cyclopentadiène à 0,5 M dans du dichlorométhane (12 mL, 6 mmol, 3 éq.) y est ajouté goutte à goutte pendant 15 min. Après 2 h le mélange réactionnel est lavé avec de l'eau (3 x 20 mL) et de la saumure (1 x 30 mL), séché au sulfate de magnésium et concentré. Les produits isomères sont purifiés du reste du brut par chromatographie sur colonne de silice (cyclohexane : acétate d'éthyle =

90 : 10 volumes) pour obtenir 506 g du produit recherché sous forme de solide blanc. Rendement : 92 %.

^{1}H RMN ($CDCl^{3}$, 400 MHz) : 8,67 (m, 1H), 8,01 (m, 1H), 7,81 (m, 1H), 7,45 (m, 1H), 7,29 (m, 4H), 7,16 (m, 1H), 6,49 (m, 1H), 5,82 (m, 1H), 4,53 (m, 1H), 3,54 (s, 1H), 3,45 (d, 1H, *J* = 4 Hz), 3,09 (s, 1H), 2,06 (d, 1H, *J* = 8.4 Hz), 1,62 (m, 1H) δ ppm.

I.3 Cristallographie du complexe $Cu(II)[(phen)2(NO_3)]^+$

2,42 mg (10 mmol) de $Cu(NO_3)_2,3H_2O$ (Sigma) sont dilués dans 20 mL d'eau distillée. 1 mL de cette solution (1 mmol, 1 équiv.) est transféré dans un ballon rond de 5 mL. Parallèlement, 0,54 mg (3 mmol) de 1,10-Phénanthroline sont dilués dans 3 mL d'éthanol pur. 1 mL de cette solution (2 équiv.) est mélangé à la solution chaude de Cu(II). Le mélange est agité à reflux jusqu'à ce qu'un précipité vert apparaisse. Le précipité est lavé sur Hirsch de 5 avec une solution de méthanol glacé. Le produit est ensuite recristallisé dans un minimum de méthanol chaud.

Les mesures de cristallographie aux rayons X ont été effectuées sur un diffractomètre Kappa-APEX II (Bruker) à 100K. Affinement des F^2 contre toutes les réflexions, les R-factor pondérés wR et la qualité de fit S sont basés sur F^2, le R-factor conventionnel est basé sur F, avec F mis à zéro pour les F^2 négatifs. Les R-factors basés sur F^2 sont statistiquement deux fois plus grands que ceux basés sur F, et les R-factors basés sur toutes les données sont encore plus grand. Composition de l'unité asymétrique : C_{24} H_{20} Cu N_6 O_6, groupe d'espace : P_{-1}, longueurs de maille : **a** 7.4170(4) **b** 10.9701(6) **c** 15.4896(9), angles de maille : α 98.994(2) β 103.746(2) γ 102.446(2), volume de maille : 1166.48 nm^3.

Chapitre II - Biologie

II.1 Constructions génétiques

II.1.1 Souche bactérienne

Les bactéries utilisées dans les expériences de biologie moléculaire sont issues d'une souche XL1blue MRF' (Stratagene) de génotype : Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac [F' proAB lacIqZΔM15 Tn10 (Tetr)]. Elles sont transformées comme il suit :

1 à 10 µL de plasmide sont ajoutés à 100 µL de bactéries électrocompétentes XL1blue MRF' préalablement décongelées sur glace. Après un mélange doux et une minute sur glace, les bactéries sont transférées dans des cuvettes d'électroporation stériles *E.coli* Pulser 0,2 cm (BioRad). Après un choc électrique (25 µFd / 2,5 volts / 400 Ohms), 400 µL de milieu 2YT à température ambiante sont immédiatement ajoutés puis les bactéries sont mises à incuber 1h à 37°C à 220 RPM. 200 µL de bactéries diluées 10x sont enfin étalés sur boite 2YT (AMRESCO) et Agar 1,5% contenant de l'Ampicilline (269 µM) (pour la transformation des vecteurs d'expression pQE), Tetracycline (28 µM), Glucose (1%) et incubées à 37°C O/N.

II.1.2 Synthèse du gène A3s

La séquence du gène A3 synthétique a été conçue à l'aide de l'algorithme d'optimisation génétique proposé par Genscript puis modifié manuellement afin de différencier au maximum les codons identiques entre motifs répétés. Le gène a été synthétisé par Genscript dans un vecteur pUC57. Le vecteur est donc appelé pUC57[*A3s*]

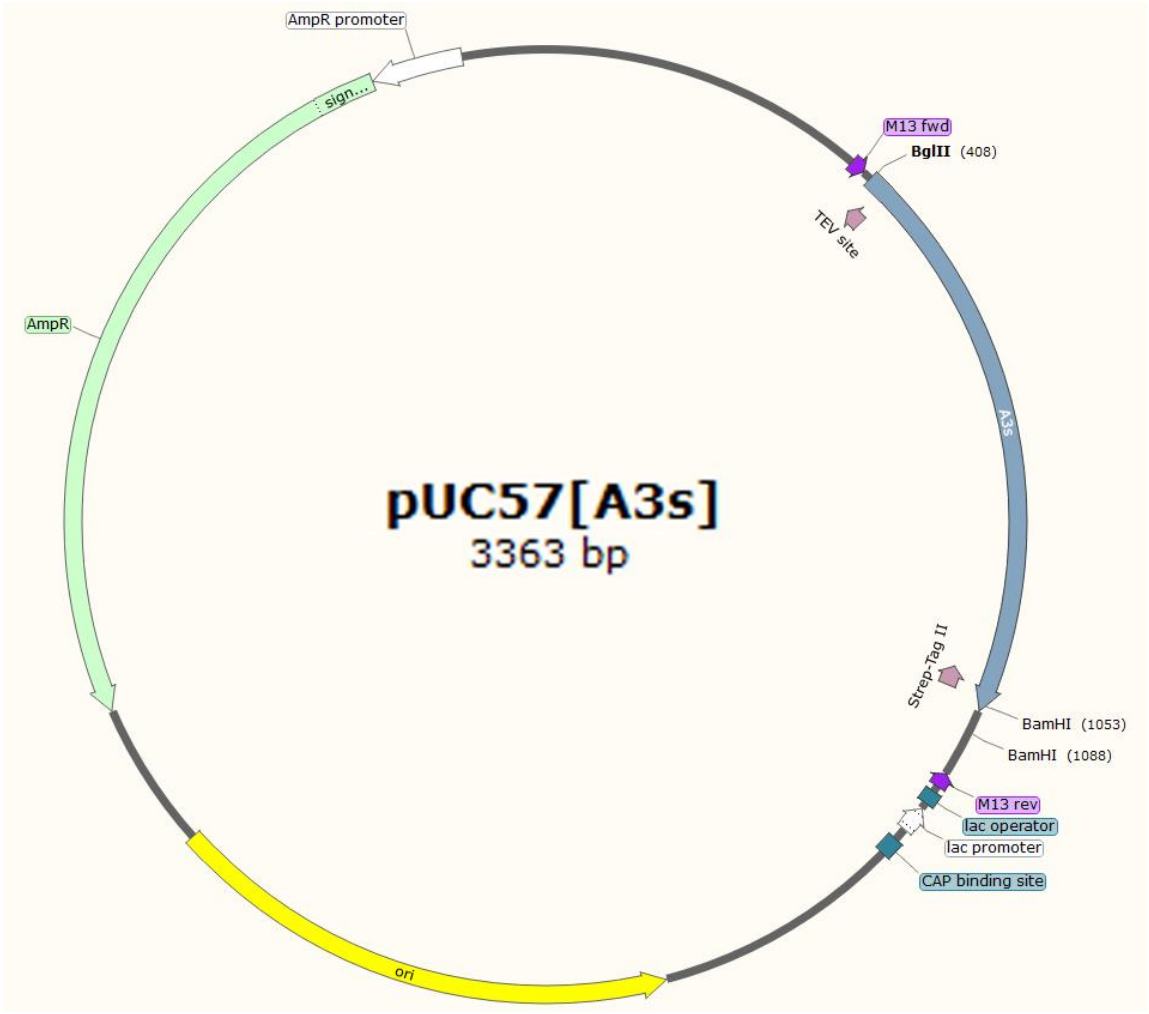


Figure 158 : Carte du plasmide de clonage pUC57[A3s]

II.1.3 Obtension des vecteurs d'expression des protéines A3s

II.1.3.1 Sous clonage pUC57 vers pQE70

La phase ouverte de lecture du gène A3s a été transférée dans le plasmide commercial pQE70 (QIAGEN) entre les sites de restriction BamHI et BglII. : 5 µg du plasmide pQE70 sont digérés par double restriction avec 2 µL de chacune des deux enzymes FastDigest (ThermoFisher Scientific) dans 40 µL de réaction en présence du tampon FastDigest GreenBuffer. Le mélange de digestion est laissée 1h à 37°C puis le produit de digestion est purifié sur gel d'agarose à l'aide du kit NucleoSpin Gel and PCR Clean-Up (Macherey-Nagel) selon les instructions du kit, l'élution finale est effectuée avec 15 µL de tampon fourni chauffé à 50°C. La concentration de l'éluat est mesurée au nanodrop2000 (ThermoFisher Scientific). Les embouts de digestion de BamHI et BglII étant compatibles, 1 µg de plasmide est traité avec 1 µL (5U) d'Antartic Phosphatase (NEB) dans 20 µL de réaction en présence du tampon fourni. Le mélange est mis à 37°C pendant 1h puis le plasmide déphosphorylé est purifié avec le kit MinElute PCR Purification Kit (QIAGEN) selon les instructions du kit, l'élution finale est effectuée avec de l'eau milliQ (Millipore) à 70°C. En parallèle, 7,8 µg de plasmide pUC57[*A3sWT*] sont digérés avec les enzymes BamHI et BglII FastDigest FastDigest (ThermoFisher Scientific) dans 80 µL de réaction en présence du tampon FastDigest GreenBuffer. Le mélange de digestion est laissé 1h à 37°C puis le produit de digestion est purifié sur gel d'agarose 1% à l'aide du kit NucleoSpin Gel and PCR Clean-Up (Macherey-Nagel) selon les instructions du kit, l'élution finale est effectuée avec de l'eau milliQ (Millipore) chauffé à 50°C. La concentration de l'éluat est mesurée au nanodrop2000 (ThermoFisher Scientific). Insert et plasmide accepteur sont enfin mis en présence avec un rapport molaire insert:plasmide de 10:1 puis incubés sur la nuit à 16°C en présence de T4 DNA ligase (ThermoScientific) dans un tampon adéquat, pour un volume final de 20 µL. Le produit de ligation est purifié grâce au MinElute PCR Purification Kit (Qiagen) et élué dans 10 µL de tampon EB au 1/10e. Ces 10 µL sont ensuite utilisés pour transformer 100 µL de bactéries XL1blue MRF' électrocompétentes. Les produits de ligation sont vérifiés par séquençage. Le produit de ligation est appelé pQE70[*A3s*]. Les gènes mutants sont effectués dans cette construction selon le protocole de mutagénèse dirigée décrite.

II.1.3.2 Sous clonage pQE70 vers pQE80N

Figure 159 : Schéma de la phase ouverte de lecture de la construction pQE80N[A3s]

La construction pQE80N est utilisée pour produire les variants A3s munis d'un HisTag clivable (figure 159). 5 ng des plasmides pQE70 contenant les gènes *A3s* sauvages ou mutants sont amplifiés par PCR dans 50 µL de réaction final à l'aide de l'enzyme Phusion polymerase (ThermoFisher Scientific) dans le tampon HF fourni.

Cycles PCR :

95°C 5 min ; 10x(95°C 30s ; 60°C 30s ; 72°C 1 min) ; 20x(95°C 30s ; 72°C 1 min) ; 72°C 5 min.

Amorces :

5'-catcacGGATCCgaaaatctctatttccagggtGCCGCAGATCCGG-3'
5'-ctaattAAGCTTatttctcaaactgcgaccagagctcAGAAATCAGGGATTTATGAGTTTCCAGG-3'

1 µL de chaque enzyme Dpn1, BamHI et HindIII FastDigest (ThermoFisher Scientific) sont directement ajoutés dans la solution d'amplification. Le mélange est mis à incuber à 37°C sur la nuit. 40 µL du mélange est ensuite purifié sur gel d'agarose 1% à l'aide du kit NucleoSpin Gel and PCR Clean-Up (Macherey-Nagel) selon les instructions du kit, l'élution finale est effectuée avec 20 µL de tampon d'élution fourni chauffé à 50°C. Parallèlement 5 µg de plasmide pQE80L commercial (QIAGEN) est digéré à l'aide des enzymes BamHI et HindIII FastDigest (ThermoFisher Scientific) et purifié sur gel d'agarose 1% à l'aide du kit NucleoSpin Gel and PCR Clean-Up (Macherey-Nagel) selon les instructions du kit, l'élution finale est effectuée avec 20 µL de tampon d'élution fourni chauffé à 50°C.

Insert et plasmide accepteur sont enfin mis en présence avec un rapport molaire insert:plasmide de 10:1 puis incubés sur la nuit à 16°C en présence de T4 DNA ligase (ThermoScientific) dans un tampon adéquat, pour un volume final de 20 µL. Le produit de ligation est purifié grâce au MinElute PCR Purification Kit (Qiagen) et élué dans 10 µL de tampon EB au 1/10e. Ces 10 µL sont ensuite utilisés pour transformer 100 µL de bactéries XL1blue MRF' électrocompétentes. Les produits de ligation sont vérifiés par PCR, digestion et séquençage de la phase ouverte de lecture. Le produit de ligation est appelé pQE80N[*A3s*].

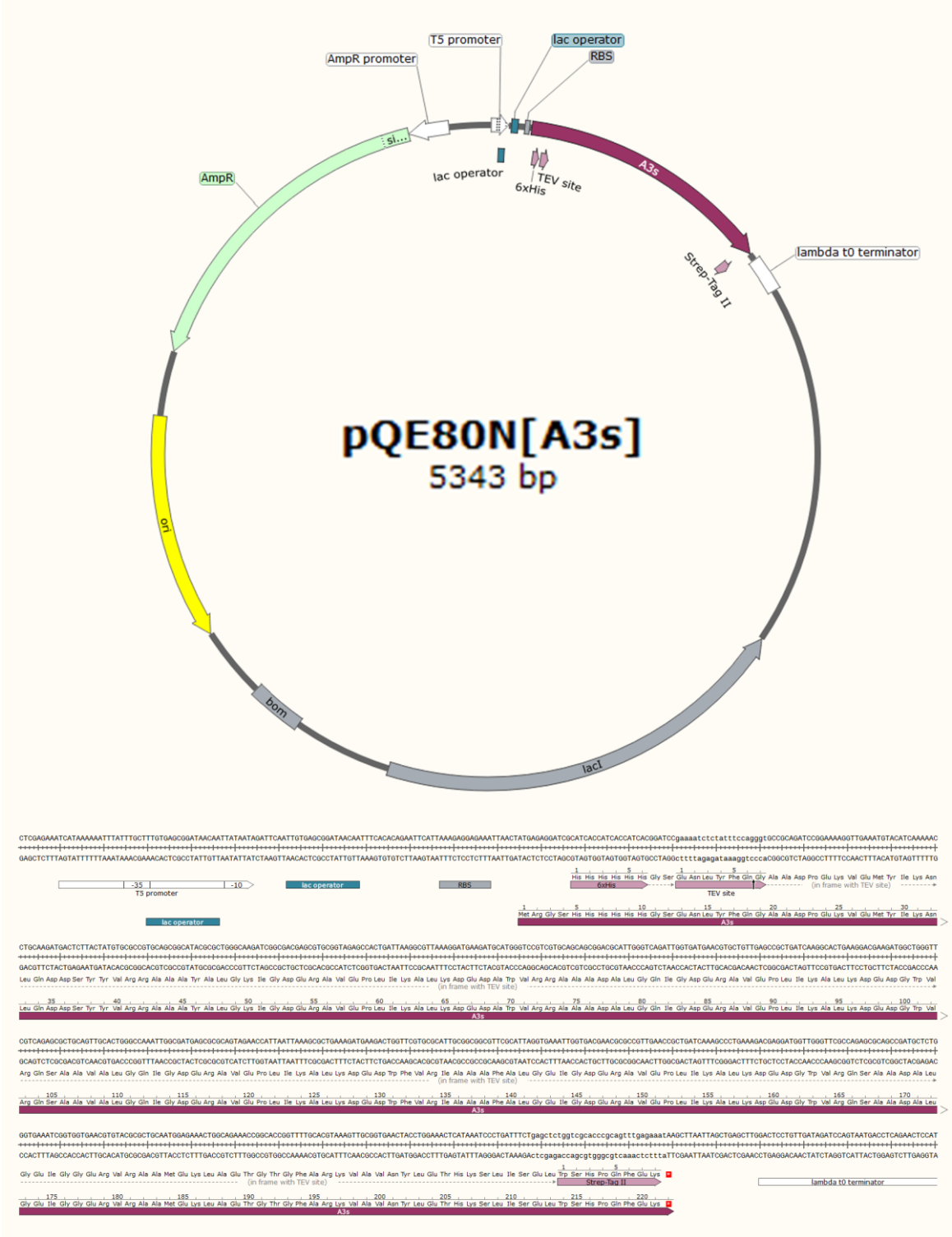


Figure 160 : Carte du plasmide d'expression pQE80N utilisé pour la production des variants A3s munis d'un HisTag clivable.

II.1.3.3 Construction du vecteur pQE80S

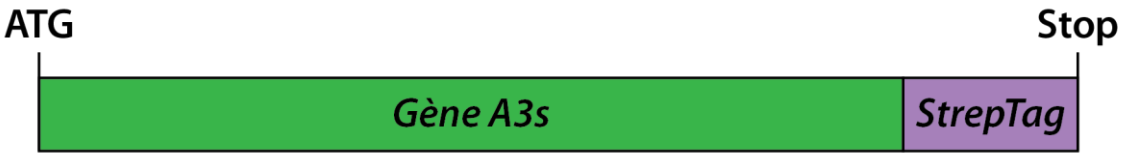


Figure 161 : Schéma de la phase ouverte de lecture de la construction pQE80S[A3s]

La construction pQE80S est utilisée pour produire les variants A3s sans HisTag (figure 161). 5 ng de plasmide pQE80N contenant les gènes *A3s* sauvage ou mutants sont amplifiés par PCR dans 50 µL de réaction final à l'aide de l'enzyme Phusion polymerase (ThermoFisher Scientific) dans le tampon HF fourni.

Cycles PCR :

98°C 1 min ; 25x(98°C 20s ; 66°C 15s , 72°C 2 min 40s) ; 72°C 10 min.

Amorces :

5'-GGTGCCGCAGATCCG-3'
5'-CGATCCTCTCATAGTTAATTTCTCCTCTTT -3'

10 µL de produit d'amplification sont directement utilisés pour effectuer une ligation à bouts francs en présence de 5% de PEG 4000 et 5U de T4 DNA ligase (ThermoFisher Scientific) en présence du tampon d'activité fourni. Le mélange est incubé 1h à 22°C puis la ligase est inactivée à 70°C pendant 5 minutes. 2 µL de mélange sont directement utilisés pour transformer des bactéries *E. coli* XL1blue électrocompétentes. L'ADN plasmidique de 5 clones indépendants est extrait grâce à un kit NucleoSpin Plasmid (Macherey-Nagel) et analysés par digestion BamHI. La phase ouverte de lecture des clones positifs est séquencée pour vérifier l'absence de mutation.

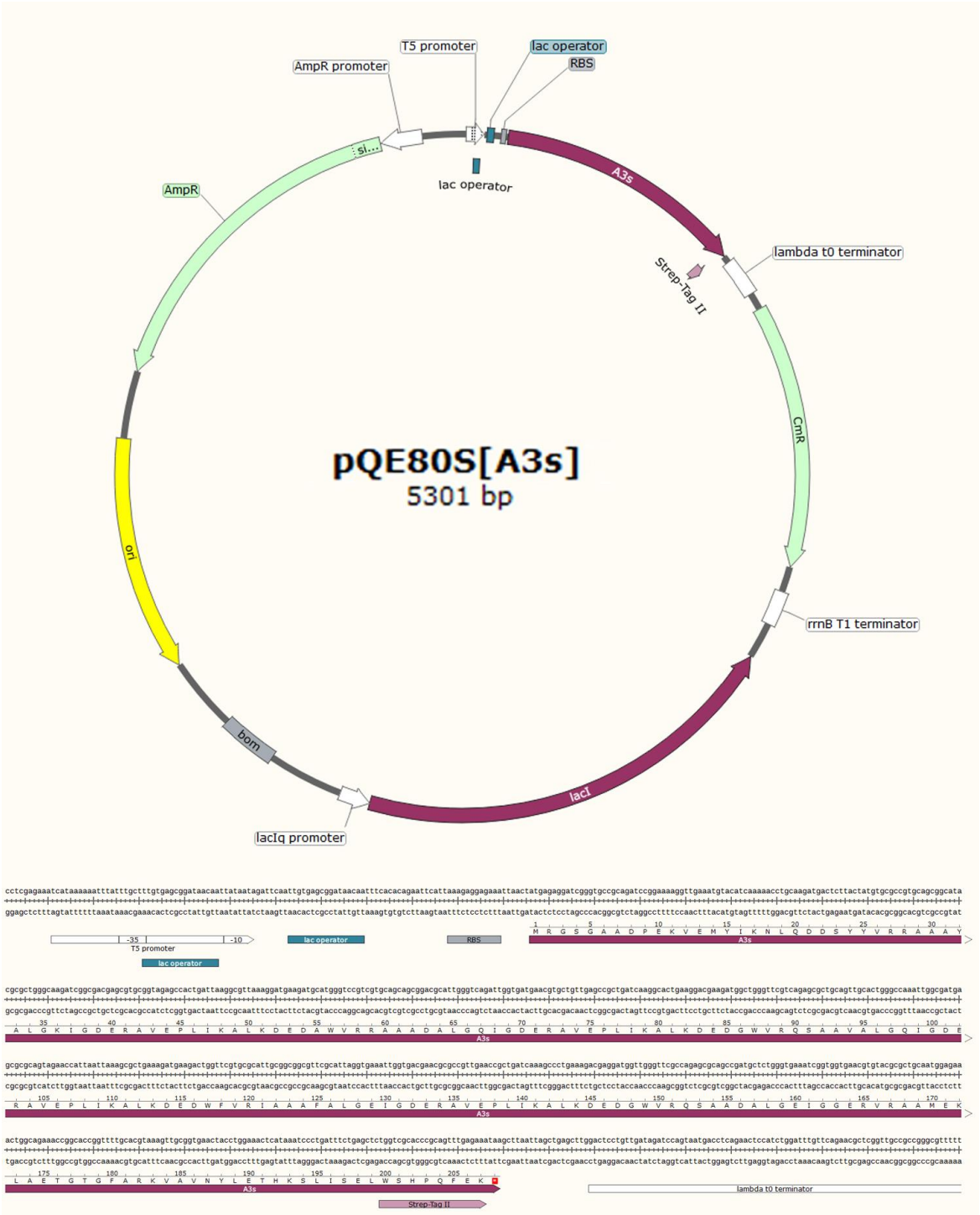


Figure 162 : Carte du plasmide d'expression pQE80S utilisé pour la production des variants A3s.

II.1.4 Obtention du vecteur d'expression des protéines A3sA3

Figure 163 : Schéma de la phase ouverte de lecture de la construction pQE81[A3sA3]

Le plasmide contenant le gène *A3sA3* a été conçu par CPEC par insertion du gène A3s provenant du vecteur pUC57 vers un plasmide pQE81 (QIAGEN) contenant la protéine A3 répétée (rapport stage ERASMUS de Conor Nelson). Les plasmides ont été mutés par la méthode de mutagénèse dirigée décrite.

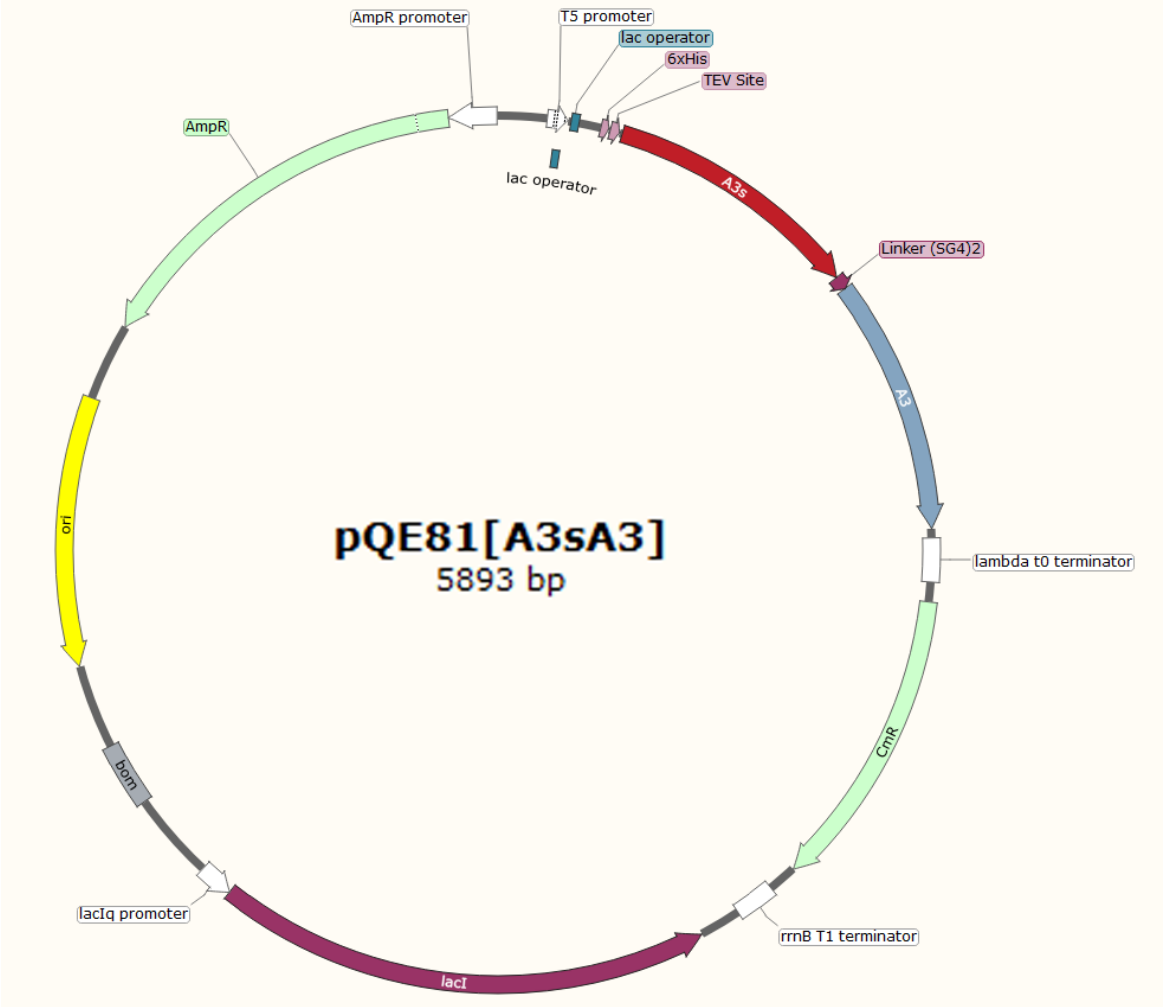


Figure 164 : Carte du plasmide d'expression pQE81 utilisé pour la production des variants A3sA3.

II.1.5 Mutagénèse dirigée

La méthode QuikChange (Stratagene) est utilisée avec le kit enzymatique Phusion polymerase (ThermoFisher Scientific). Dans un volume total de 50 µL, 10 µL de tampon Phusion HF (5X), 1 µL de solution de dNTP à 10 mM chacun, 2,5 µL de chaque amorce à 10 µM, 1µL d'ADN matrice à 10 ng/µL, 0,5 µL de Phusion polymerase HF. Cycles PCR: 98°C 1 min ; 20x(98°C 20s ; 65°C 15s ; 72°C 3 min) ; 72°C 10 min. Après amplification, 1 µL de Dpn1 FastDigest (ThermoFisher Scientific) est ajouté directement au produit de PCR et le mélange est incubé 1h à 37°C. 2 µL produits bruts sont directement utilisés pour transformer 100 µL de bactéries électrocompétentes *E. coli* XL1blue MRF' (Agilent technologies), suivant la procédure standard. Trois clones indépendants sont utilisés pour effectuer une extraction d'ADN plasmidique selon les instructions du Kit NucleoSpin Plasmid (Macherey-Nagel). Les solutions plasmidiques sont séquencées par méthode Sanger (Eurofins) avec les amorces commerciales pQEfor (et si nécessaire pQErev) afin de couvrir toute la phase ouverte de lecture.

Mutation	Amorce Sens	Amorce Antisens
Y26C	5'-GCAGCGGCATgCGCGCTGGG-3'	5'-CCCAGCGCGcATGCCGCTGC-3'
K30C	5'-GCGCTGGGCtgcATCGGCGAC-3'	5'-GTCGCCGATgcaGCCCAGCGC-3'
F119C	5'-CGGCGGCGTgCGCATTAGG-3'	5'-CCTAATGCGcACGCCGCCG-3'
D150C	5'-GAGCGCAGCCtgTGCTCTGGG-3'	5'-CCCAGAGCAcaGGCTGCGCTC-3'

II.2 Obtention des protéines

II.2.1 Souche bactérienne

Les bactéries utilisées pour la production des protéines sont issues d'une souche M15 (QIAGEN) de génotype F-, Φ80ΔlacM15, thi, lac-, mtl-, recA+ , KmR. Elles sont transformées comme il suit : Une quantité variable d'ADN plasmidique n'excèdant pas 10% du volume de bactéries est ajouté à un aliquote de 100 à 200 µL de bactéries compétentes préalablement décongelées sur la glace. Après 30 min d'incubation, les bactéries subissent un choc thermique de 45 sec à 42°C suivi d'un retour immédiat sur glace. Après 2 min de repos sur glace, 300 µL de 2YT est ajouté puis les bactéries sont incubées 1h à 37°C à 220 RPM. Les bactéries sont enfin étalées sur boîte de 2YT-Glucose 1% avec les antibiotiques appropriés. (Ampicilline pour maintenir les plasmides pQE et en plus Tetracycline pour XL1blue, Chloramphenicol pour Rosetta et Kanamycine pour M15)

II.2.2 Protéines possédant un HisTag

II.2.2.1 Production

Un clone issu d'une transformation fraiche de *E.coli* M15 portant le plasmide pQE80N ou pQE81 du variant désiré est mis en culture sur la nuit à 37°C et 220RPM dans 20 mL de milieu 2YT (AMRESCO) supplémenté avec 269 µM d'ampiciline, 103 µM de kanamycine et 1% de glucose stérile. 1L de milieu 2YT supplémenté avec 269 µM d'ampiciline, 103 µM de kanamycine est inoculé avec 1/100^{e} de préculture et séparé en deux erlenmeyers à ergots de 2L contenant chacun 500 mL de culture. Les bactéries sont mises en culture à 37°C et 220 RPM en incubateur Novotron HT (Infors) jusqu'à une densité optique à 600 nm de 0,7-0,9. La production de protéines est induite par ajout d'isopropyl β-D-1-thiogalactopyranoside (IPTG) pour une concentration finale de 500 µM. Après 5h, les cultures sont centrifugées pendant 20 min à 5000g à 4°C dans une centrifugeuse Avanti J-E (Beckman Coulters). Les culots sont repris chacun dans 15 mL de tampon Na-Phosphate pH 8, 300 mM NaCl, avec 20 mM Imidazole, rassemblés et congelés à -80°C.

II.2.2.2 Purification

Les culots correspondant à 1L de culture sont décongelés et une tablette d'inhibiteurs de protéases PIC EDTA-free (Roche) est ajoutée. Les bactéries sont lysées à

l'aide d'une sonde de sonication Sonifier 250 (Branson) muni d'une sonde 1/2", sur glace avec quatre séries de pulses (réglages : 50% Duty Cycle, 40% output). 1µL de DNAse (thermo) est ajouté et le lysat est centrifugé 40 min à 10 000 g à 4°C. La fraction soluble est séparée du culot et injectée sur colonne HisTrap Crude FF 5 mL (GE Healthcare) à l'aide d'un appareil ÄKTA pure muni d'une pompe échantillon. Tampon de lavage : Na-Phosphate 50 mM pH 8, 300 mM NaCl, 20 mM Imidazole. Tampon d'élution : Na-Phosphate 50 mM pH 8, 300 mM NaCl, 300 mM Imidazole. La protéine est éluée en fractions de 2 mL. Les fractions sont analysées au nanodrop 2000 (ThermoFisher Scientific) et les fractions rassemblant plus de 90% des protéines totales éluées sont réunies (typiquement 6 mL).

II.2.2.3 Digestion par la TEV protéase

Les fractions sont diluées dans du tampon Na-Phosphate 50 mM pH 8, 0 mM NaCl, 0 mM Imidazole pour atteindre une concentration finale d'imidazole de 50 mM. 1/20^{e} de tampon TEV 20X (10 mM EDTA, 20 mM DTT dilués dans de l'eau milliQ (Millipore)) est ajouté ainsi que la TEV protéase pour un rapport massique TEV/Protéine de 1/100. Le mélange réactionnel est mis à incuber à 8°C sur la nuit.

II.2.2.4 Contre-purification

Le mélange est dilué par un volume égal de tampon Na-Phosphate 50 mM, 150 mM NaCl, 0 mM Imidazole pour atteindre une concentration finale d'imidazole de 25 mM. La solution est ensuite déposée par gravité sur colonne HisTrap Crude FF 5 mL (GE Healthcare) préalablement équilibrée par du tampon de lavage (Na-Phosphate 50 mM pH 8, 300 mM NaCl, 20 mM Imidazole) et la fraction non retenue, contenant la protéine clivée, est conservée. 5 mL (1 CV) de tampon de lavage sont à nouveau ajoutés à l'aide d'une seringue afin de récupérer le liquide restant dans la colonne, qui est ajouté à la fraction non retenue. La fraction retenue, contenant la TEC protéase ainsi que la protéine non clivée, est éluée par 10 mL (2 CV) de tampon Na-Phosphate 50 mM pH 8, 300 mM NaCl, 300 mM Imidazole.

II.2.2.5 Concentration et changement de tampon

Deux tampon Na-Phosphate 50 mM pH 7,75 sont préparés, l'un contenant 0 mM NaCl (tampon A) et l'autre contenant 1M NaCl (tampon B). Les tampons sont agités au moins 30 min en présence de 1 g/L de CHELEX100 (Bio-Rad), filtrés sur unité de filtration 0,45 µm Standart Line (VWR) et dégazés sous vide de 100 mbar à l'aide d'une pompe à membrane ME1 (Vacuubrand). La fraction non retenue issue de la contre-purification est

diluée avec le tampon sans sel pour une concentration finale de NaCl d'au moins 80 mM. La fraction non retenue diluée est et injectée sur colonne HiTrap Q HF 5 mL (GE Healthcare) à l'aide d'un appareil ÄKTA pure muni d'une pompe échantillon. L'élution est effectuée par un gradient linéaire de 0 à 50% de tampon B durant 20 CV. Les fractions contenant la protéine sont déposées sur gel SDS-PAGE, rassemblées et analysées par test d'Ellman avant couplage du cofacteur. Une production typique de 1L de culture donne environ 70 mg de protéines disponibles pour le couplage.

II.2.3 Protéines possédant un StrepTag

II.2.3.1 Production

Un clone issu d'une transformation fraiche de *E.coli* M15 portant le plasmide pQE80S du variant désiré désiré est mis en culture sur la nuit à 37°C et 220RPM dans 20 mL de milieu 2YT (AMRESCO) supplémenté avec 269 µM d'ampiciline, 103 µM de kanamycine et 1% de glucose stérile. Le lendemain 1L de milieu 2YT supplémenté avec 269 µM d'ampiciline et 103 µM de kanamycine est inoculé avec 1/100^{e} de préculture et séparé en deux erlenmeyers à ergots de 2L contenant chacun 500 mL de culture. Les bactéries sont mises en culture à 37°C et 220 RPM en incubateur Novotron HT (Infors) jusqu'à une densité optique à 600 nm de 0,7-0,9. La production de protéines est induite par ajout d'isopropyl β-D-1-thiogalactopyranoside (IPTG) pour une concentration finale de 500 µM. Après 5h, les cultures sont centrifugées pendant 20 min à 5000g à 4°C dans une centrifugeuse Avanti J-E (Beckman Coulters). Les culots sont repris chacun dans 15 mL de tampon Tris/HCl 100 mM pH 8, 150 mM NaCl, 1 mM EDTA, rassemblés et congelés à -80°C.

II.2.3.2 Purification

Les culots correspondant à 1L de culture sont décongelés et une tablette d'inhibiteurs de protéases PIC EDTA-free (Roche) est ajoutée. Les bactéries sont lysées à l'aide d'une sonde de sonication Sonifier 250 (Branson) muni d'une sonde 1/2", sur glace avec quatre séries de pulses (réglages : 50% Duty Cycle, 40% output). 1µL de DNAse (thermo) est ajouté et le lysat est centrifugé 40 min à 10 000 g à 4°C. La fraction soluble est séparée du culot et injectée sur colonne StrepTrap HP 5 mL (GE Healthcare) à l'aide d'un appareil ÄKTA pure muni d'une pompe échantillon. Tampon de lavage Tris/HCl 100 mM pH 8, 150 mM NaCl, 1 mM EDTA. Tampon d'élution : 2,5 mM d-Desthiobiotin (Sigma) ajoutés au tampon de lavage. La protéine est éluée en fractions de 1 mL. Les fractions sont analysées au nanodrop 2000 (ThermoFisher Scientific) et les fractions rassemblant plus

de 90% des protéines totales éluées sont réunies, concentrées si nécessaire sur Vivaspin 5 kDa MWCO (Sartorius). Le concentrât est injecté sur colonne Superdex S75 16/600 (GE Healthcare) équilibrée avec du tampon Na-Phosphate 50 mM pH 7,75, 150 mM NaCl préalablement agité au moins 30 min en présence de 1 g/L de CHELEX100 (Bio-Rad), filtré sur unité de filtration 0,45 µm Standart Line (VWR) et dégazé sous vide de 100 mbar à l'aide d'une pompe à membrane ME1 (Vacuubrand). Les fractions contenant la protéine sont déposées sur gel SDS-PAGE, rassemblées et analysées par test d'Ellman avant couplage du cofacteur. Une production typique de 1L de culture donne environ 20 mg de protéines disponibles pour le couplage.

II.2.4 TEV-protéase

II.2.4.1 Production

La souche E. coli BL21(DE3) CodonPlus RIL (Stratagene) contenant déjà le plasmide d'expression pRK793 (Addgene #8827) nous a été fournie par le Dr. Yann Ferrandez du laboratoire LBPA de l'ENS Cachan. Une colonie unique est mise à incuber sur la nuit à 37°C, 220 RPM dans 25 mL d'une solution de 2YT (AMRESCO) supplémenté en ampicilline (269 µM), chloramphénicol (108 µM) et glucose (1%). Le lendemain, 1L de milieu 2YT supplémenté avec 269 µM d'ampiciline et 108 µM de chloramphénicol est inoculé avec 1/100^{e} de préculture et séparé en deux erlenmeyers à ergots de 2L contenant chacun 500 mL de culture. Les bactéries sont mises en culture à 37°C et 220 RPM en incubateur Novotron HT (Infors) jusqu'à une densité optique à 600 nm de 0,7-0,9. Les solutions sont ensuite refroidies dans de la glace pendant 10 min. La production de protéines est induite par ajout d'isopropyl β-D-1-thiogalactopyranoside (IPTG) pour une concentration finale de 350 µM et les cultures sont mises à 17°C sur la nuit. Le lendemain, les cultures sont centrifugées pendant 20 min à 5000g à 4°C dans une centrifugeuse Avanti J-E (Beckman Coulters). Les culots sont repris chacun dans 15 mL de tampon de lyse, rassemblés et congelés à -80°C.

II.2.4.2 Purification

La TEV protéase est purifiée dans des tampons particuliers.

Tampon de Lyse : Tris 20 mM pH 8, NaCl 500 mM, Imidazole 20 mM (0.681 g @ 68.08 g/mol) Glycerol 10%, β-Mercaptoéthanol 10 mM, MgSO4 20 mM

Tampon A : Tris 20 mM pH 8, NaCl 500 mM, Imidazole 20 mM

Tampon B : Tris 20 mM pH 8, NaCl 500 mM, Imidazole 500 mM

Tampon C : Tris 20 mM pH 8, NaCl 1500 mM, Imidazole 20 mM

Tampon D : Tris 50 mM pH 8, NaCl 200 mM, Glycerol 10%, EDTA 2 mM, DTT 10 mM

Les bactéries sont décongelées et une tablette d'antiprotéase PIC EDTA-free (Roche) est ajoutée. Les bactéries sont lysées à l'aide d'une sonde de sonication Sonifier 250 (Branson) muni d'une sonde 1/2", sur glace avec trois séries de pulses (réglages : 50% Duty Cycle, 40% output). 1µL de DNAse (thermo) est ajouté et le lysat est centrifugé 40 min à 10 000 g à 4°C. La fraction soluble est séparée du culot et déposée sur une colonne contenant 5 mL de résine Ni-NTA (Macherey-Nagel) préalablement équilibrée avec 10 CV de tampon A. La purification Ni-NTA est effectuée en chambre froide. Après dépôt de la fraction soluble, la résine est lavée par du tampon C jusqu'à ce que l'absorbance à 260 nm et 280 nm en sortie de colonne cesse de baisser, puis avec 5 CV de tampon A. La protéine est enfin éluée avec du tampon C en fractions de 2 mL. Les fractions contenant plus de 90% de protéine totale sont rassemblées et injectées sur une colonne de tamis moléculaire préparative Superdex S75 16/600 (GE Healthcare) préalablement équilibrée avec du tampon D. Les fractions d'élution sont rassemblées et mélangée avec du tampon D supplémenté de glycérol de manière à avoir une concentration de protéine de 1 mg/mL et une concentration de glycérol de 25%. La protéine est enfin aliquotée en fractions de 1 mL et congelée dans de l'azote liquide. Les stocks sont conservés à -80°C.

II.2.5 SDS-PAGE et Western Blot

Les échantillons sont mélangés avec du tampon de charge 4X Laemmli Sample Buffer (Bio-Rad), chauffés à 95°C pendant 10 min sur un thermobloc Dry Bath FB 1501 (Fisher Scientific) et déposés sur des gels de polyacrylamide 12% TGX Stain Free FastCast (Bio-Rad). Ils sont migrés dans un bloc Mini Protean Tetra Cell (Bio-Rad) immergé dans du tampon Tris-Glycine SDS-Buffer (BioRad) avec un transformateur EPS 301 (Amersham) pour une intensité limitante de 40 mA par gel jusqu'à ce que le front de migration atteigne le bas du gel. Deux gels avec des dépôts identiques sont préparés, l'un est directement coloré par 20 mL de colorant Instant Blue (Expedeon), l'autre est transféré sur membrane de nitrocellulose de 0,45 µm (Bio-Rad). Le transfert est fait entre deux filtres MiniTransBlot (Bio-Rad) ainsi qu'entre deux éponges, immergé dans du tampon Tris 25 mM / glycine 200 mM pH 8,3, ethanol 20%, pendant 1h à 60 mA limitant. Les étapes suivantes sont effectuées à température ambiante sous agitation : la membrane est bloquée par une incubation de 1h dans 20 mL de tampon TBS, tween 0,1%, BSA 3% ; lavée trois fois 5 min avec 20 mL de tampon TBS, Tween 0,1 % ; incubée 1h avec 10 mL d'un anticorps de souris anti-pentaHistidine (QIAGEN) dilué au 1/10000 dans du tampon TBS, tween 0,1%, BSA 1% ; et lavée trois fois 5 min avec 20 mL de tampon TBS, Tween 0,1 %. Les étapes suivantes sont effectuées dans le noir : la membrane est incubée 1h avec 10

mL d'un anticorps de chèvre anti-souris couplé à Alexa fluor 680 (Stratagene) dilué au 1/10000 dans du tampon TBS, tween 0,1%, BSA 1% ; lavée trois fois 5 min avec 20 mL de tampon TBS, Tween 0,1 % puis lavée deux fois 5 min avec 20 mL de tampon TBS seul. La membrane est enfin révélée à l'aide d'un appareil LiCor Odyssey (BioAgilytix) réglé sur la fluorescence de l'Alexa 680.

Chapitre III - Biohybride

III.1 Couplage des cofacteurs

III.1.1 Conditions de couplage

Tous les couplages sont conduit dans du tampon Na-Phosphate pH 7,75, 150 mM NaCl, 4°C sur la nuit. Les cofacteurs sont ajoutés respectivement à la concentration en thiols libres déterminée par test d'Ellman (100 µM pour les couplages *phen* et *terpy*, 5 µM pour les couplages porphyrine). Les cofacteurs sont solubilisés dans du DMSO pour une concentration finale dans le milieu réactionnel de 1%, 1% et 10% pour *phen*, *terpy* et porphyrine respectivement. Les équivalents molaires de cofacteur ajouté sont 1.2, 2 et 5 pour *phen*, *terpy* et porphyrine respectivement.

III.1.2 Arrêt des couplages

Les couplages *phen* et *terpy* sont stoppés par ajout d'1/20^{e} d'une solution de 10 mM EDTA, 20 mM DTT suivis d'une injection sur une colonne de tamis moléculaire Superdex S75 16/600 (GE Healthcare) préalablement équilibrée avec le tampon MOPS 20 mM pH 7, 150 mM NaCl chélexé et dégazé. Les couplages porphyrine sont stoppés par extraction selon la méthode de Teale modifiée.

III.1.3 Méthode de Teale modifiée

Un volume équivalent de méthyl éthyl cétone (butanone) à 0°C est ajouté aux solutions de couplage, le mélange est ensuite agité pendant 30s. Les tubes sont ensuite centrifugés 5 min à 5000g à 4°C dans des *swing-buckets*. La fraction supérieure (butanone) est récupérée par succion à l'aide d'une pipette pasteur reliée à une fiole à vide. La fraction aqueuse est séparée du culot de protéine et mise à dialyser si nécessaire contre du tampon MOPS 20 mM pH 7, 150 mM NaCl. Le culot est resuspendu dans 1/5^{e} du volume de couplage avec du tampon MOPS 20 mM pH 7, 150 mM NaCl par agitation

douce. Le culot resuspendu est filtré avec un filtre 0,2 µM (Minisart, Sartorius) puis injecté sur une colonne de tamis moléculaire équilibrée en tampon MOPS 20 mM pH 7, 150 mM NaCl.

III.2 Caractérisation

III.2.1 Test d'Ellman

Les solutions suivantes sont préparées : **Tampon d'Ellman :** 100 mM tampon phosphate sodium pH 8, 1 mM EDTA. **Solution de réactif :** 4 mg de DTNB (Sigma) pour 1 mL de tampon (10 mM). **Protéine** de concentration connue (estimé par l'absorbance à 280 nm). La lecture d'absorbance à 412 nm est réalisée sur un spectromètre Cary 100 UV/Vis (Varian) en mode cinétique dans des cuves en quartz de 500 µL. Du tampon d'Ellman est tout d'abord utilisé pour effectuer le zero de mesure. La cinétique est lancée, 5 µL de solution de réactif (concentration finale 100 µM) sont ajoutés et la solution est homogénéisée à la pipette pasteur. L'absorbance de base (DO_{base}) est notée, 50 µL de solution de protéines sont ajoutés et la solution est homogénéisée avec la même pipette pasteur. L'absorbance est suivie jusqu'à atteindre un plateau (DO_{signal}). Le calcul de la concentration en cystéines libres dans la solution mère se fait grâce à l'équation : $[\boldsymbol{Thiols}](\boldsymbol{M}) = \frac{\boldsymbol{DO_{signal} - DO_{base}}}{\mathbf{14150}} \times \mathbf{10}$. Le temps écoulé entre l'ajout de la protéine et l'apparition de la moitié de l'absorbance au plateau (absorbance de base soustraite) représente le t½.

III.2.2 Tamis moléculaire analytique

Toutes les expériences de tamis moléculaire ont été menées sur les appareils ÄKTA Purifier ou ÄKTA Pure (GE Healthcare) contre le tampon spécifié. Colonne préparative : Superdex S75 16/600 (GE Healthcare), débit standard 1 mL/min. Colonne analytique : Superdex S75 10/300 (GE Healthcare), débit standard 0,7 mL/min. Pour une analyse précise des volumes d'élution et calcul des poids moléculaires apparents, seulement 100 µL de protéines à 60 µM sont injectés sur la colonne de tamis moléculaire analytique et les volumes d'élutions sont comparés à la droite d'étalonnage. La courbe d'étalonnage a été effectuée à l'aide de protéines globulaires du kit MWGF-70 (Sigma) en traçant le rapport volume d'élution/Volume mort (mesuré avec une solution de Blue Dextran présente dans le kit) en fonction du logarithme du poids moléculaire.

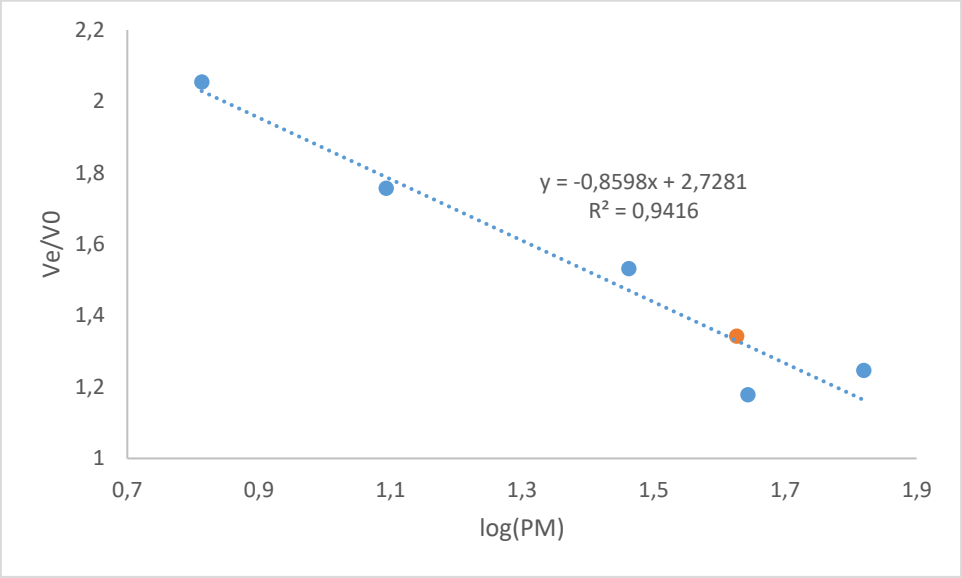


Figure 165 : Droite d'étalonnage de la colonne de tamis moléculaire analytique. En bleu en ordre croissant de LogP : Aprotinin (6,5 kDa) Cytochrome C (12,4 kDa), Anhydrase carbonique (29 kDa), Ovalbumine (44 kDa), Albumine (66 kDa). En orange est représenté un dimère A3 (mesuré : 40 kDa, théorique : 44 kDa).

III.2.3 Dichroïsme Circulaire

Les protéines sont diluées pour une concentration finale de 5 µM de monomère et 5 mM de NaCl, dans du tampon phosphate 25 mM pH 7. Les spectres d'ellipticité sont mesurés dans des cuves en quartz suprasil de 1 mm de trajet optique sur un appareil J-810 (Jasco). Le signal du tampon seul est soustrait. Les ellipticités (en °) sont converties en valeurs moyennes par résidu (MRW) par l'équation :

$$\theta_{MRE}\ (deg.cm^2.dmol^{-1}) = \frac{0{,}1}{l \times C \times N \times 3298 \times 10^{-6}}$$

avec l en cm, C en µM, N = nombre de liaisons peptidiques.

III.2.4 Cristallographie

III.2.4.1 Cristallogénèse

Les biohybrides métallés sont concentrés dans du tampon HEPES 20 mM pH 7, 150 mM NaCl ou MOPS 20 mM pH 7, 150 mM NaCl sont concentrés à hauteur de 20 mg/mL et déposés dans des gouttes de 100 nL par la méthode de goutte assise dans des boites de kit de cristallographie (Hampton Research) avec un robot de pipetage Cartesian Microsys (Proteigen). Les gouttes sont photographiées à intervalles réguliers par un robot

Rock Imager 182 (Formulatrix). Les conditions présentant des pistes de cristallisation ont été reproduites manuellement.

III.2.4.2 Analyse

Les clichés de diffraction ont été acquis sur les lignes PROXIMA du synchrotron SOLEIL ou les lignes ID-23 et ID-29 du synchrotron de l'ESRF. Les clichés sont traités avec la suite logicielle XDS.[2] Les modèles cristallographiques sont effectués grâce à la suite logicielle CCP4[3] par remplacement moléculaire de la protéine A3 (PDB : 3LTJ) avec le programme Phaser[4] et affinés grâce aux programmes Refmac[5] et *Coot*[6].

III.2.5 Ultracentrifugation analytique

Toutes les expériences de vitesse de sédimentation sont effectuées sur une ultracentrifugeuse analytique XL (Beckam Coulter) avec un rotor An-50 Ti à 20°C équipé d'un système de détection de fluorescence (AVIV Biomedical). Les expériences de vitesse de sédimentation sont effectuées dans des pièces centrales en Epon de deux canaux de 12 mm de trajet optique. La protéine A3 est marquée avec un kit Monolith NT.115 Protein Labeling Kit BLUE (Nanotemper Technologies) suivant le guide d'instruction. Les échantillons sont préparés par dilution des solutions stock avec du tampon MOPS 20 mM pH 7, 150 mM NaCl contenant 0,1 mg/mL d'albumine sérique bovine (BSA). Les échantillons contenant entre 1 et 60 nM de protéine sont réalisés par dilution de la protéine marquée, contrairement aux solutions dont les concentrations sont plus élevées pour lesquelles un mélange de protéine marquée et non marquée est utilisé pour atteindre la concentration totale désirée. 400 µL d'échantillon sont centrifugés à 42 000 RPM (128 297 g). Les profils de sédimentation sont collectés toutes les 5 min et les scans sont analysés grâce au logiciel SEDFIT [7]. Les distributions c(s) résultantes sont transférées dans le logiciel GUSSI [8] en utilisant le mode vitesse de sédimentation isotherme. Pour chaque concentration le coefficient de sédimentation moyen pondéré par le signal (s_w) est calculé par intégration et tracé sur graphique.[9] La courbe de s_w isotherme est rentrée dans le programme SEDPHAT [10] et un modèle monomère-dimère est utilisé pour déterminer le K_D. La densité et la viscosité du tampon à 20°C a été déterminée par le logiciel SEDNTERP (T. Laue, Royal Society of Chemistry, Cambridge). Le volume partiel spécifique des variants A3s est prédit en se basant sur la composition en acides aminés grâce à ce programme.

III.2.6 Spectrométrie de masse MALDI-ToF

Les échantillons sont dessalés avec une colonne Zeba Spin (ThermoScientific) contre un tampon de bicarbonate d'ammonium 10 mM. Les échantillons sont ensuite mélangés à différent rations avec une solution d'acétone saturée d'acide sinapinique à différent puis mis à évaporer sur une plaque de MALDI. Les échantillons sont analysés par MALDI-ToF en mode linéaire positif sur un spectromètre UltrafleXtreme (Bruker Daltonique) afin de vérifier la masse de la protéine avec ou sans cofacteur.

III.2.7 Spectrométrie de masse ICP-MS

Le biohybride est métallé dans un tampon MOPS 20 mM pH 7, 150 mM NaCl sur la nuit. Le lendemain, l'échantillon est dessalé sur colonne analytique Superdex S75 10/300 (GE Healthcare) préalablement équilibrés avec un tampon phosphate 10 mM pH 7, 10 mM NaCl. Ce tampon a préalablement été traité avec 1g/L de CHELEX 100 (Bio-Rad) pendant au moins 30 min, puis filtré sur membrane 0,45 µm et dégazé. Le biohybride est dosé par nanodrop 2000 (Thermo). L'échantillon est dilué 1000 fois dans une solution d'acide nitrique. Le tampon est traité de la même manière, la valeur de contenu en cuivre du tampon étant soustraite à celle de l'échantillon. Les deux échantillons sont analysés sur un spectromètre ICP-MS quadripolaire Thermoelectron X7 (ThermoFisher Scientific) et comparés à une calibration de Cu(II) encadrant la valeur de concentration des échantillons (de 0,5 à 4 ppb) dans la même solution d'acide nitrique.

III.2.8 Résonnance Paramagnétique Électronique

Les différents mutants sont métallés avec 1 équiv. de Cu(II) (en rapport avec la concentration de monomère) par ajout de 2 µL de solution concentrée de nitrate de cuivre dans une solution de protéines à 200 µM de monomère. Après une incubation sur la nuit, les protéines sont désallées par tamis moléculaire analytique S75 10/300 (GE Healthcare) et concentrés sur vivaspin 5000 kDa MWCO (Sartorius) jusqu'à une concentration de 200 µM de monomère. 100 µL de solution sont injectés dans un tube RPE et soumis à 5 cycles de vide/hélium dans le but de purger l'oxygène des échantillons. Les tubes sont ensuite immergés dans un bain d'éthanol fondant et conservés à 80 K avant analyse. Les spectres de bande X sont enregistrés sur un spectromètre ELEXSYS 500 (Bruker) équippé avec un résonnateur en bande X ER4119HS (Bruker), un cryostat en flux continu ESR 900 (Oxford Instrument) et un système de contrôle de température. Les conditions d'acquisition sont : T=100 K, MW freq. = 3,38 GHz, MW power = 0,80 mW, MA = 6 G, Mod. Freq. = 100 kHz, Gain = 60 db. Les simulations spectrales ont été effectuées grâce au logiciel XSophe.

III.3 Catalyse

III.3.1 Transfert hydrolytique de liaison phosphoester

III.3.1.1 Étalonnage

III.3.1.1.1 Courbe pH Nitrophénol

Une gamme de mélanges de tampon HEPES 20 mM pH 8, 150 mM NaCl et MES 20 mM pH 6, 150 mM NaCl est effectuée et le pH résultant mesuré. Une concentration connue de 4-nitrophénol (Sigma) est diluée à chacun des pH, la valeur d'absorbance mesurée au Cary150 (Varian) est convertie en coefficient d'extinction molaire ($M^{-1}.cm^{-1}$). La courbe résultante est comparée à une courbe calculée à partir des coefficients reportés des formes nitrophénol et nitrophénolates (18400 et 200 M-1.cm-1 respectivement) ainsi que son pKa de 7,15 à 25°C.[11] Les deux courbes étant proches, la courbe théorique a été utilisée.

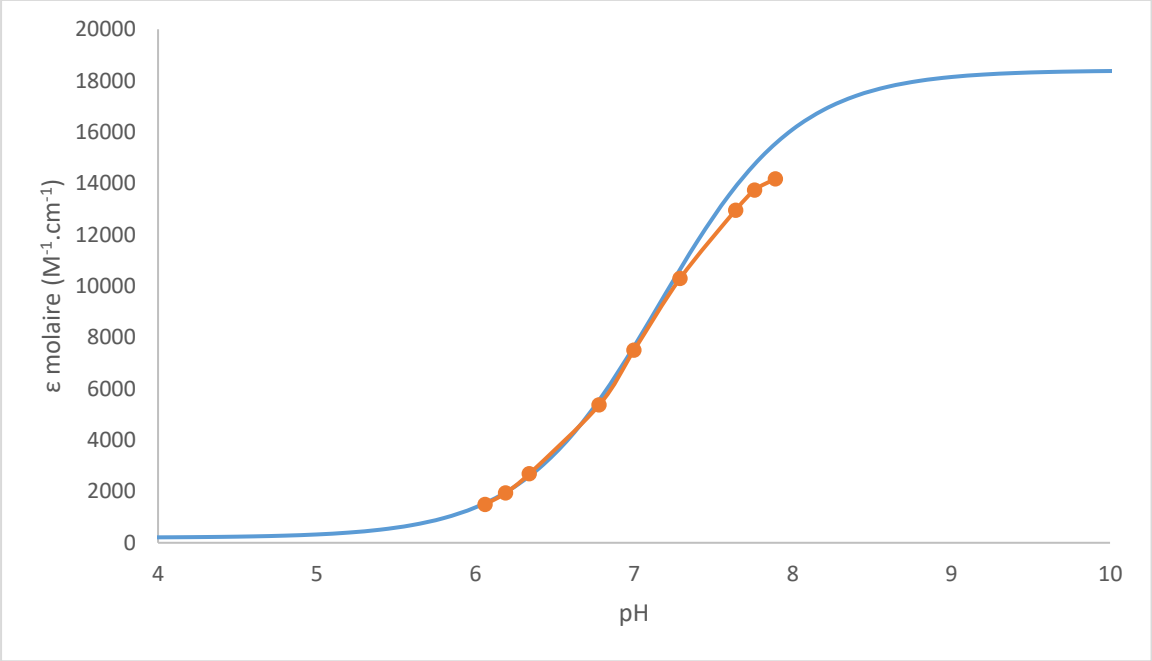


Figure 166 : Courbe bleue : epsilon molaire calculé du paranitrophénol en fonction du pH, courbe orange : epsilon molaire mesuré dans le tampon.

III.3.1.1.2 Courbe dose-réponse microplaques

Le lecteur de microplaques infinite pro (TECAN) n'ayant pas de trajet optique constant il a fallu comparer la réponse du lecteur avec une gamme d'aborbance connue, la courbe de pH du paranitrophénol a été utilisée. Un volume de 200 µL ou 250 µL de la gamme

mesurée au spectrophotomètre Cary150 (Varian) a été déposé dans le lecteur et l'absorbance a été mesurée avec un filtre à 405 nm. Les droites de régression sont :

250 µL : $A_{cary} = (A_{tecan}-0,103)*1,3219$; $R^2 = 0,9993$

200 µL : $A_{cary} = (A_{tecan}-0,103)*1,3219$; $R^2 = 0,9987$

III.3.1.1.3 Tampons d'activité

Les tampons d'activité des expériences de catalyse de Diels-Alder sont préparés à base d'un tampon HEPES 20 mM pH 7, 150 mM NaCl agité avec 1 g/L de résine CHELEX100 (Bio-Rad), ajusté aux pH 6,9 ; 7,3 ; 7,7 et 8,1. Le pH final résultant du mélange du tampon d'activité et du tampon de la protéine (HEPES 20 mM pH 7, 150 mM NaCl) est calculé mathématiquement considérant un pKa du tampon HEPES de 7,55.

III.3.1.2 Protocole de catalyse

Les protéines non métallées sont transférées dans du tampon HEPES 20 mM pH 7, 150 mM NaCl par dessalage sur une colonne de tamis moléculaire Superdex S75 10/300 (GE Healthcare) puis sont portées à une concentration de 50 µM à l'aide d'un concentrateur Vivaspin 5 kDa MWCO (Sartorius). Une solution de HPNP est préparée extemporanément dans de l'eau milliQ (Millipore) pour une concentration finale de 50 mM. Une solution de Zn(NO3)2 ou Cu(NO3)2 est également préparée extemporanément dans de l'eau milliQ (Millipore) pour une concentration finale de 165 µM. A l'aide d'un robot pipeteur de type Tecan Genesis (TECAN), les réactifs sont ajoutées dans des plaques 96 puits MicroWell Delta Surface (Nunclon) dans cet ordre : 120 µL de tampon d'activité, 20 µL de solution de Cu(II) ou Zn(II) (ou H_2O milliQ pour les témoins sans métal), 60 µL de solution de protéine (ou de tampon HEPES 20 mM pH 7, 150 mM NaCl pour les témoins sans protéine), 20 µL de solution de HPNP. Les concentrations finales sont : 5 mM de HPNP, 16.5 µM de métal, 15 µM de protéine. Les cinétiques sont suivies à 405 nm dans un robot Infinite F Pro (TECAN) équipé d'un filtre d'absorbance à 405 nm $\pm$ 10 nm.

III.3.2 Diels-Alder

III.3.2.1 Courbe d'étalonnage

Une gamme d'un mélange des produits authentiques est diluée dans de l'acétonitrile et diluée dans 500 µL de tampon MOPS pH 7, 150 mM NaCl pour des concentrations finales allant de 0 à 1 mM, représentant 0 à 100% de rendement théorique dans les conditions standard de la réaction de Diels-Alder. L'étalon interne diphénylsulfone est dilué dans un volume minimal d'acétonitrile et ajouté dans chaque tube pour une concentration finale de 1,87 mM. Le protocole standard d'extraction et d'analyse est effectué pour chacune des solutions. L'aire des produits est calculée en

termes de pourcentage de l'aire de l'étalon interne. La droite de régression permet de déterminer le rendement de la réaction.

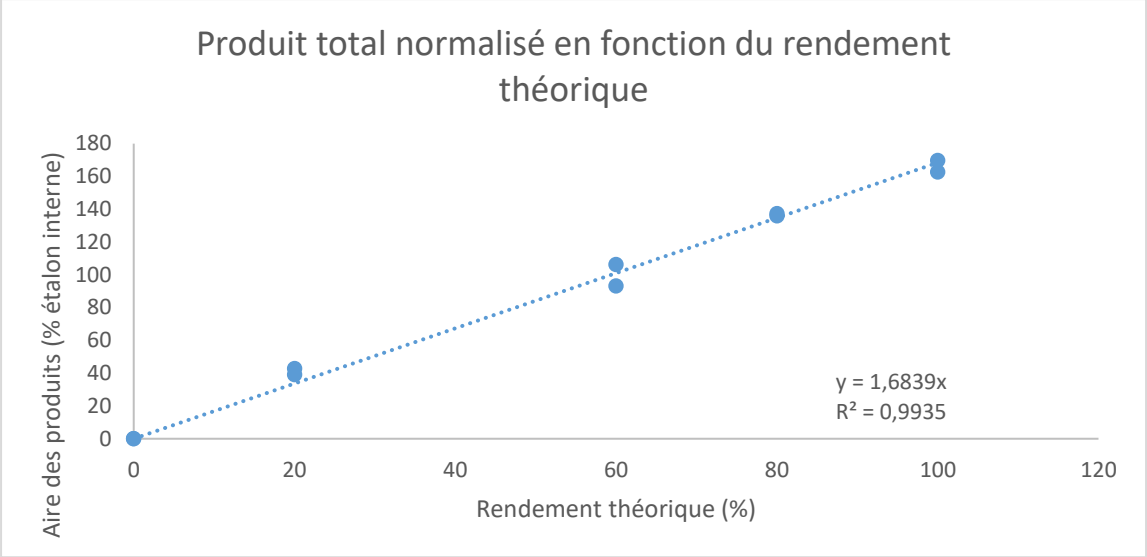


Figure 167 : Droite d'étalonnage du produit extrait. Le rendement théorique de la peut être relié au rapport de l'aire des produits sur l'aire de l'étalon interne.

III.3.2.2 Métallation des biohybrides

Les protéines sont à une concentration de 100 µM environ dans du tampon MOPS 20 mM pH 7, 150 mM NaCl. Du nitrate de cuivre ($Cu(NO_3)_2,3H_2O$) est pesé et dilué dans de l'eau distillée pour une concentration de $100xN_{eq}$ la concentration des biohybrides. 1/100^e de solution de cuivre (II) est ajouté dans les solutions de protéine qui sont immédiatement et doucement resuspendues à la pipette. Les biohybrides sont laissés sur glace en chambre froide sur la nuit. Pour les conditions avec dessalage (Neq = 1,5), les biohybrides sont injectés sur colonne de tamis moléculaire analytique Superdex S75 10/300 GL (GE Healthcare) à l'aide d'une machine ÄKTA Purifier ou ÄKTA Pure (GE healthcare) contre du tampon MOPS 20 mM pH 7, 150 mM NaCl.

III.3.2.3 Protocole de catalyse

Protocole standard, les composants et concentrations et temps sont les suivants, sauf indiqués autrement. Les catalyseurs protéiques sont dans le tampon de métallation, le cuivre (II) $Cu(NO_3)_2,3H_2O$ est pesé et dilué dans un volume minimal d'eau distillée. Sur glace, dans un tube de microcentrifuge 2 mL (VWR) à fond conique contenant du tampon MOPS 20 mM pH 7, 150 mM NaCl (QsP 500µL total) sont ajoutés dans cet ordre : quantité variable pour 30 µM de catalyseur (1 équiv.), 5 µL de solution mère acétonitrile de substrat diénophile (concentration finale 1 mM azachalcone, 33 équiv. et 1,87 mM étalon interne diphénylsulfone) puis 1,4 µL de cyclopentadiène pur fraichement distillé (34 mM, 1133

équiv.). Les tubes sont placés dans un agitateur Thermomixer compfort (eppendorf) réglé à 4°C, dans une chambre froide pendant 48h. 500 µL d'acétate d'éthyle sont ensuite ajoutés, agité vigoureusement 20 fois puis centrifugé dans une centrifugeuse Centrifuge 5418R (Eppendorf) à vitesse maximum pendant 5 min à 4°C. 400 µL d'acétate d'éthyle sont récupérés, la procédure est répétée deux autres fois avec 400 µL d'acétate d'éthyle ajoutés. Les extraits sont rassemblés et séchés à l'aide de sulfase de magnésium ($MgSO_4$) puis à nouveau centrifugés dans les mêmes conditions. 500 µL sont prélevés et 6 µL sont injectés sur colonne HPLC chirale (ODH Chiralpak IC (4,6*250 mm), 5 µm) à un débit de 0,5 mL/min, mélange n-hexane:isopropanol 99:1, 25°C. Les pics d'étalon interne et des produits sont intégrés à une longueur d'onde de 270nm, le substrat est intégré à une longueur d'onde de 300nm.

III.3.3 Oxydation du TMB

Les réactifs sont préparés extemporanément. Le réactif TMB est dissout dans de l'eau distillée pour une concentration finale de 5 mM. Une solution de H_2O_2 stock à 11,6 M (Sigma) est diluée dans de l'eau distilée pour une concentration finale de 5 mM. Les protéines sont transférées dans du tampon phosphate 10 mM pH 7, 0 mM NaCl avec une colonne de desalage ZebaSpin (ThermoFisher Scientific). Les protéines et substrats sont ajoutés dans des plaques 96 puits MicroWell DeltaSurface (Nunclon) pour une concentration finale de : 5 µM de A3sA3Y26TPPMn ; 0, 5 ou 25 µM de α2 ; 250 µM de substrat TMB ; 250 µM de H_2O_2. Les cinétiues sont suivies grâce à un robot Infinite M Pro (TECAN) réglé sur 652 nm $\pm$ 10 nm. La quantité de substrat produit est calculée grâce à l'absorbance du complexe de transfert de charge à 652 nm (39 000 $M^{-1}.cm^{-1}$), l'absorbance étant convertie grâce à la courbe de dose-réponse calculée avec le *para*-nitrophénol (III.3.1.1.2) selon l'équation :

$$A_{cary} = (A_{tecan}-A_{tampon})*1,3219$$

Chapitre IV - Bio-informatique

IV.1 Représentations moléculaires

Toutes les inspections représentations moléculaires de protéines ont été menées sur le logiciel PyMOL (The PyMOL Molecular Graphics System, Version 1.7.5 Schrödinger, LLC.) Les représentations moléculaires des structures cristallographiques des petites molécules ont été faites à l'aide du logiciel Mercury[12] Les formules chimiques sont dessinées avec le logiciel ChemDraw (Perkin Elmer).

IV.2 Modélisation de docking moléculaire

Le *docking* moléculaire a été effectué avec le logiciel AutoDock Vina[13] suivant une procédure standard. Les structures utilisées de la protéine A3 (PDB : 3LTM) et de la porphyrine TMPyP (CCDC : JOGCOX) ont été débarrassées des molécules d'eau et de l'ion métallique. Les liaisons aux positions *meso* de la porphyrine ont été laissées libres de rotation. La zone d'analyse a été soit limitée à la crevasse, soit étendue à la protéine entière. Énergie de liaison mesurée pour le site de fixation dans la poche : -7,5 kcal/mol. Énergie de liaison mesurée pour le site extérieur : -9 kcal/mol.

IV.3 PDB-Mining des protéines binucléaires à zinc

Toutes les structures de protéines contenant du zinc disponibles sur la Protéine Data Bank ont été téléchargées (8921 structures au total en 2014). Dans le but d'éliminer le biais dû aux multiples structures de protéines trop proches, les protéines ayant plus de 90% d'homologie de séquence ont été regroupées grâce à l'algorithme blastclust. La première protéine de chacun de ces groupes a alors été utilisée comme protéine représentante du groupe (3215 protéines au total). A l'aide d'un script PERL, toutes ces structures ont été analysées et la position des différents atomes de zinc présent a été relevée. Quand le fichier contenait différents modèles structuraux, seul le premier modèle était pris en compte. Les distances entre les différents zincs (quand il y en avait plus d'un) ont été mesurées et seule la plus petite distance a été retenue si elle était inférieure à 10 Å (264 protéines au total). Les enzymes ont été identifiées grâce au numéro EC présent dans le fichier de structure, les protéines ne présentant pas cette information ont été considérées comme non catalytiques bien que de faux négatifs aient été observés (total : 120 enzymes, 144 autres).

Bibliographie

(1) Brown, D. M., et Usher, D. A. (1965) Hydrolysis of hydroxyalkyl phosphate esters: effect of changing ester group. *J. Chem. Soc.* 6558.

(2) Kabsch, W. (2010) Xds. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 125-132.

(3) Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., et Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr. 67*, 235-242.

(4) McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., et Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr. 40*, 658-674.

(5) Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F., et Murshudov, G. N. (2004) REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. Sect. D Biol. Crystallogr. 60*, 2184-2195.

(6) Emsley, P., Lohkamp, B., Scott, W. G., et Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr. 66*, 486-501.

(7) Schuck, P. (1950) Analytical Biochemistry , in press ON THE ANALYSIS OF PROTEIN SELF-ASSOCIATION BY SEDIMENTATION VELOCITY ANALYTICAL ULTRACENTRIFUGATION. *Bioengineering*.

(8) http://biophysics.swmed.edu/MBR/software.html.

(9) Zhao, H., Lomash, S., Glasser, C., Mayer, M. L., et Schuck, P. (2013) Analysis of high affinity self-association by fluorescence optical sedimentation velocity analytical ultracentrifugation of labeled proteins: Opportunities and limitations. *PLoS One 8*.

(10) Schuck, P. (2003) On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem. 320*, 104-124.

(11) Bowers, G. N., McComb, R. B., Christensen, R. G., et Schaffer, R. (1980) High-purity 4-nitrophenol: purification, characterization, and specifications for use as a spectrophotometric reference material. *Clin. Chem. 26*, 724-729.

(12) Macrae, C. F., Edgington, P. R., McCabe, P., Pidcock, E., Shields, G. P., Taylor, R., Towler, M., et van de Streek, J. (2006) Mercury : visualization and analysis of crystal structures. *J. Appl. Crystallogr. 39*, 453-457.

(13) Trott, O., et Olson, A. J. (2009) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem. 31*, 455-461.

Partie D - Conclusion

Chapitre I - La protéine αRep A3 : une architecture prometteuse pour la génération de nouveaux biocatalyseurs hybrides.

L'étude sur les variants fonctionnalisés protéine A3s présentée dans cette thèse nous a permis de prouver que l'architecture de la protéine A3 était suffisamment rigide pour accepter la modification des acides aminés variables en un résidu NPH sans affecter sa structure tertiaire. L'étude de l'activité catalytique pour la réaction de Diels-Alder entre l'azachalcone et le cyclopentadiène a permis de déduire que le biohybride A3sF119NPH contenant du Cu (II) catalysait cette réaction de manière énantiosélective, présentant un excès énantiomérique de la forme *endo* jusqu'à 60%, faisant de ce biohybride un des catalyseurs les plus sélectifs pour cette réaction à ce jour. En approfondissant l'étude de ce catalyseur, les différentes données de RPE, ICP-MS et catalyse nous ont permis de proposer que le centre catalytique pouvait être un complexe bis-phénanthroline de cuivre (II). Par ailleurs, les différentes études structurales nous ont montré que la structure quaternaire du dimère pouvait être modifiée par la présence du complexe, compliquant la rationalisation de l'ingénierie de ce type de biohybride.

L'étude de la protéine bi-domaine A3sA3 fonctionnalisée avec des ligands terpyridine et phénanthroline nous a permis d'atteindre des énantiosélectivités pour la réaction de Diels-Alder d'environ 20 % avec les protéines A3sA3F119Tpy et A3sA3Y26NPH. L'utilisation de ces protéines comportant cette fois un seul complexe métallique, nous semblait une approche prometteuse, car dans ces nouveaux biohybrides, la cavité devenait totalement asymétrique. Or, les résultats plutôt décevants en terme d'énantiosélectivité, ont montré des rendements plus élevés que pour les biohybrides dimériques et bi-fonctionnalisés. *A posteriori*, au vu des résultats de cristallographie, nous pouvons interpréter les résultats obtenus en termes d'encombrement stérique au sein de la crevasse: une seule cystéine fonctionnalisée dans la cavité permettrait un meilleur accès aux substrats et produits de la réaction, conduisant un meilleur rendement ; par contre, le complexe métallique unique au sein de cette crevasse serait également plus mobile, ce qui induirait un environnement moyen isotrope et réduirait l'énantiosélectivité.

De manière intéressante, l'utilisation de ligands différents à la même position induit des sélectivités différentes, que nous n'avons cependant pas pu clairement rationnaliser. Au vu de la structure, la présence de chaines latérales pouvant entrainer une plus grande stabilisation conformationnelle du complexe au sein de la crevasse, soit complétant la sphère de coordination du métal, soit en favorisant des interactions avec le ligand. La taille

et la rigidité du bras utilisé pour le greffage du ligand pourraient également être un déterminant de cette sélectivité. L'ensemble de ces observations confirme l'utilité de tester différents complexes pour obtenir des sélectivités différentes.

Chapitre II - Plasticité du dimère et dynamique entre les sous-unités

La première partie de l'étude nous a enseigné que le dimère A3, tout en présentant une grande stabilité et conservant une structure dimérique, pouvait donner lieu à un glissement entre les deux monomères. La mobilité des deux sous-unités nécessite d'être analysée.

La particularité de la protéine A3 est qu'il s'agit d'un dimère fortuit : cette protéine n'a pas été sélectionnée par l'évolution (naturelle ou dirigée) pour présenter cette aptitude à dimériser. Il est probable que l'instabilité du dimère découle de cette caractéristique propre à la protéine A3. En effet les études bio-informatiques indiquaient que la ceinture hydrophobe était la principale source de l'énergie de dimérisation, les liaisons hydrogène entre les modules terminaux étant labiles. Bien qu'étant *de facto* source d'une grande affinité (constante de dissociation de l'ordre de 40 nM pour le dimère A3), ces interactions sont peu spécifiques. De plus, cette ceinture hydrophobe est située sur la « tranche » de la protéine, sans complémentarité de forme particulière. Les interfaces des complexes de protéines naturelles sont, elles, généralement sélectionnées pour être complémentaires [1]. Par exemple, le régulateur de transcription LmrR, dimère utilisé par l'équipe de Roelfes pour la création d'une Diels-Alderase par couplage covalent de phénanthrolines de cuivre, présente une grande complémentarité de forme par entrelacement des hélices α (figure 168). A première vue, ces différents points (complémentarité, type d'interactions, dimérisation fortuite) peuvent expliquer le manque de spécificité de l'interaction. Des études expérimentales ont également été menées pour tenter de comprendre les déterminants de cette dimérisation.

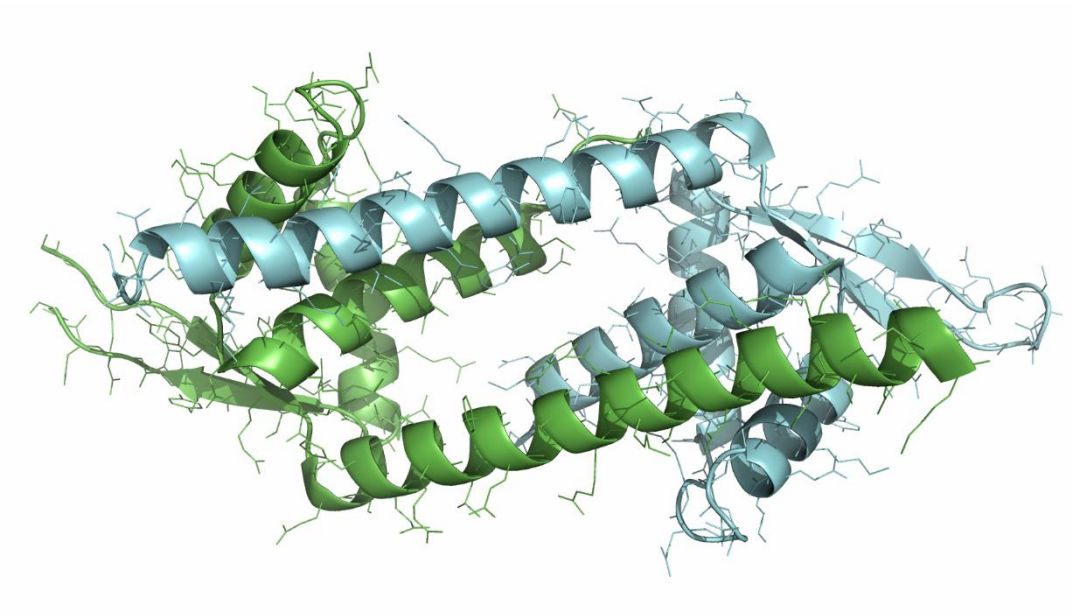


Figure 168 : Structure cristallographique du régulateur de la transcription LmrR (PDB : 3F8F), protéine dimérique utilisée par le groupe de Roelfes pour la création d'une Diels-Alderase artificielle. On remarque une grande complémentarité des deux interfaces, où la grande hélice α de chaque monomère s'enchasse entre deux hélices α du partenaire.

Deux protéines dérivées de l'αRep A3 ont été récemment créées pour tenter de déterminer expérimentalement les déterminants de dimérisation : (i) La protéine A3-noCeinture est un variant de la protéine A3 où les acides aminés de la ceinture hydrophobe ont été remplacés par de petits aminés. Cette protéine ne pourrait *a priori* former un dimère qu'uniquement par le biais d'interactions entre les modules terminaux. (ii) La protéine A3-noCap est un variant de la protéine A3 où le module C-terminal (C-noCap) a été retiré. Le quatrième module variable de cette dernière protéine, devenant de fait le nouveau C-noCap, a été modifié dans la composition en acides aminés de sa face nouvellement exposée (exoface) afin de la rendre plus hydrophile et donc limiter l'agrégation.

Les résultats préliminaires obtenus avec ces deux nouvelles protéines montrent que les deux protéines présentent un volume d'élution sur tamis moléculaire analytique correspondant à celui d'un dimère, elles semblent donc toutes deux capables de dimériser en solution. Des expériences d'ultracentrifugation analytique effectuées sur la protéine A3-noCap semblent indiquer une affinité encore plus grande entre les deux sous unités comparé au dimère sauvage. Deux hypothèses peuvent être émises à la vue de ces résultats surprenants. (i) L'interaction entre les modules terminaux peut finalement être aussi un déterminant de la dimérisation et (ii) l'affinité accrue du dimère A3-noCap sous-entendrait que la présence des modules terminaux serait aussi une gêne pour la dimérisation induite par la ceinture hydrophobe, en effet la courbure de la protéine et sa rigidité tendent à l'éloignement des acides aminés de cette dernière. Cependant,

connaissant à présent la capacité du dimère à adopter des conformations inattendues, il est nécessaire d'attendre des informations structurales pour les deux variants pour pouvoir confirmer ces hypothèses.

Bien que partiels, ces résultats appuient l'hypothèse selon laquelle les interfaces du dimère de la protéine A3 ne sont pas aussi complémentaires que pour une protéine dimérique issue d'une sélection naturelle ou artificielle.

L'étude structurale de la protéine A3sA3F119NPH indique que la protéine bi-domaine ne présente pas de phénomène de glissement. Cette approche semble la plus prometteuse pour la création d'une banque basée sur l'architecture αRep pour l'évolution dirigée de catalyseurs biohybrides énantiosélectifs.

Il est néanmoins toujours possible que la protéine puisse conduire soit à un glissement de domaine comme vu sur notre biohybride (partie I des résultats), soit à une ouverture complète de la protéine conduisant à un échange de domaines, comme il a été suspecté lors du couplage d'une porphyrine dans la dernière partie des résultats. Plusieurs solutions (additives) présentées sur la figure 169 sont alors envisageables pour tenter de réduire au maximum le glissement des sous-unités et maintenir la chimère dans une conformation fermée présentant une crevasse adéquate pour de la catalyse enzymatique.

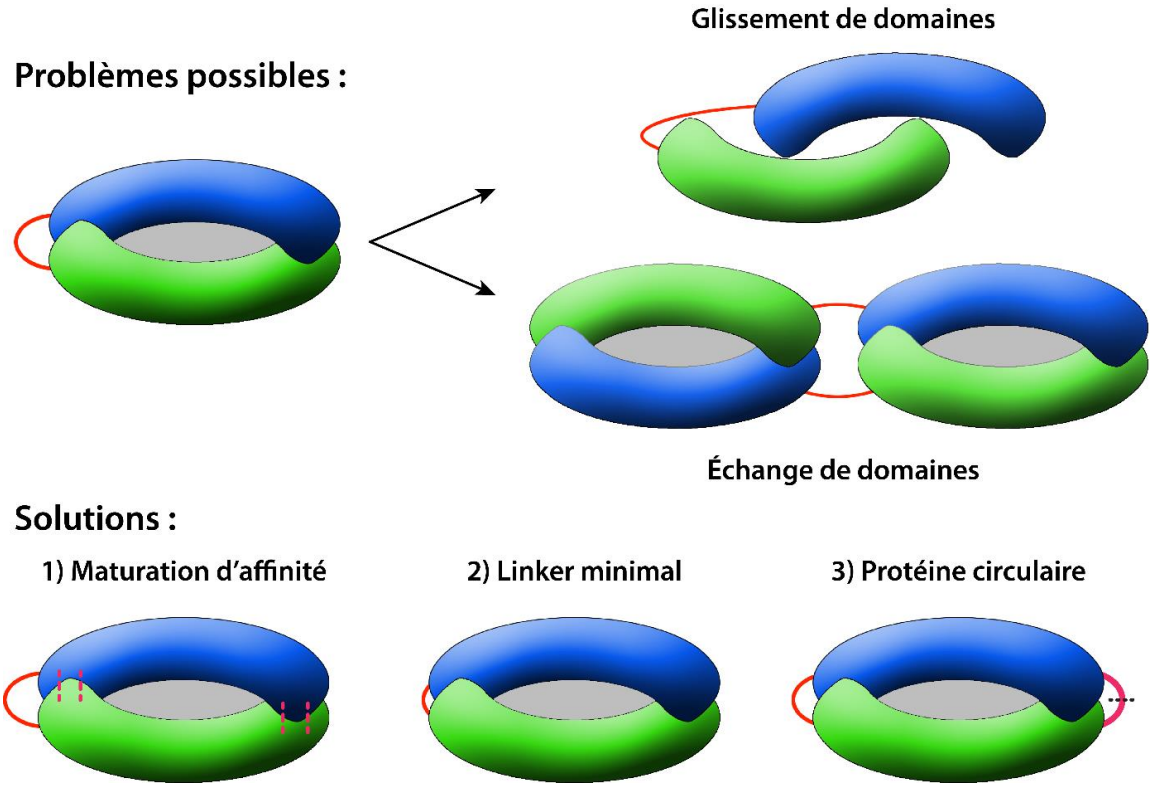


Figure 169 : Bien que la stabilité du dimère soit accrue grâce à l'ajout d'un linker, il n'est pas exclu que l'ajout d'un ou plusieurs complexes au sein de la crevasse mènent à un glissement ou un échange de domaines. Trois solutions sont possibles : la maturation d'affinité pour augmenter la stabilité et la spécificité du dimère, la diminution ou la rigidification du linker pour diminuer les mouvements, ou la synthèse de protéine circulaire à l'aide de la « split-intein », d'un pont disulfure ou de chimie click. Les trois solutions sont additives.

La première solution pour pallier à l'instabilité du dimère serait d'augmenter l'affinité entre les sous-unités et la spécificité de la dimérisation par des méthodes de sélection. La maturation d'affinité d'anticorps par des méthodes de bio-panning telle que le phage display ont déjà été mis en œuvre avec succès[2] et pourraient être utilisée en concevant une banque de protéine αRep basée sur le monomère A3 et dirigée contre A3. Une telle sélection a déjà été effectuée avec la banque αRep 1.0 a déjà permis de sélectionner la protéine α2 comme vu dans le dernier chapitre de cette thèse. Cette méthode présente cependant le risque de sélectionner des αRep reconnaissant la protéine avec une orientation différente, par exemple de manière perpendiculaire, faisant alors disparaitre la crevasse.

Une deuxième solution serait de rigidifier la région charnière qu'est le linker. La structure cristallographique de la protéine A3sA3 nous a en effet enseigné que ce linker était flexible, ce qui était attendu pour ce type de séquence $((GS_4)_2)$.[3] Un linker plus petit ou

plus rigide (par exemple structuré en hélice α) pourrait alors limiter ou éliminer le phénomène de glissement de domaine.[3,4] L'échange de domaines pourrait cependant toujours avoir lieu, notamment lors de l'utilisation de méthodes d'extraction déstabilisantes comme l'utilisation de la butanone lors du greffage de ligands de type porphyrinique.

Une dernière solution consisterait à créer une protéine circulaire par liaison covalente des deux extrémités. La présence de deux linkers empêcherait alors toute possibilité d'échange de domaines et augmenterait la stabilité de la protéine. Plusieurs options sont possibles pour arriver à cette fin : Un pont disulfure est la solution la plus simple à mettre en place, la protéine serait cependant sensible au pouvoir réducteur du milieu (intracellulaire ou réactionnel). L'utilisation de la chimie click à l'aide d'acides aminés non naturels tels que le couple azoture/cyclooctyne, décrites dans la dernière partie de l'introduction, peut permettre d'effectuer une liaison covalente dans un milieu complexe tel que le cytosol. Cela nécessiterait cependant l'utilisation de deux codons non naturels. Dans ces deux cas un choix devrait être fait entre le type de chimie utilisée pour circulariser la protéine et pour effectuer le lien covalent avec le cofacteur métallique. Enfin, la dernière solution serait l'utilisation de la « split-intein » et de la méthode SICLOPPS (« Split Intein Circular Ligation Of Peptides and Proteins », figure 170) ou dérivés.[5] Les intéines sont des enzymes catalysant le réarrangement de leur propre liaison peptidique. Elles sont constituées dans leur forme naturelle d'un domaine catalytique entouré de deux autres chaines peptidiques (appelées « extéines ») séparées par un peptide signal. Le domaine catalytique effectue sa propre excision puis la reformation de la chaine polypeptidique au niveau de la séquence spécifique. Il a été remarqué que le domaine catalytique pouvait être couplé en deux parties N et C terminales (d'où le terme « split-intein »). Les deux parties N et C terminales peuvent ainsi être placées autour d'une séquence polypeptidique (telle que A3) et ainsi catalyser la circularisation de la protéine. [5] Cette méthode a été utilisée avec succès pour la cyclisation de plusieurs protéines et enzymes, augmentant leur thermostabilité ainsi que leur résistance à l'agrégation. [6,7]

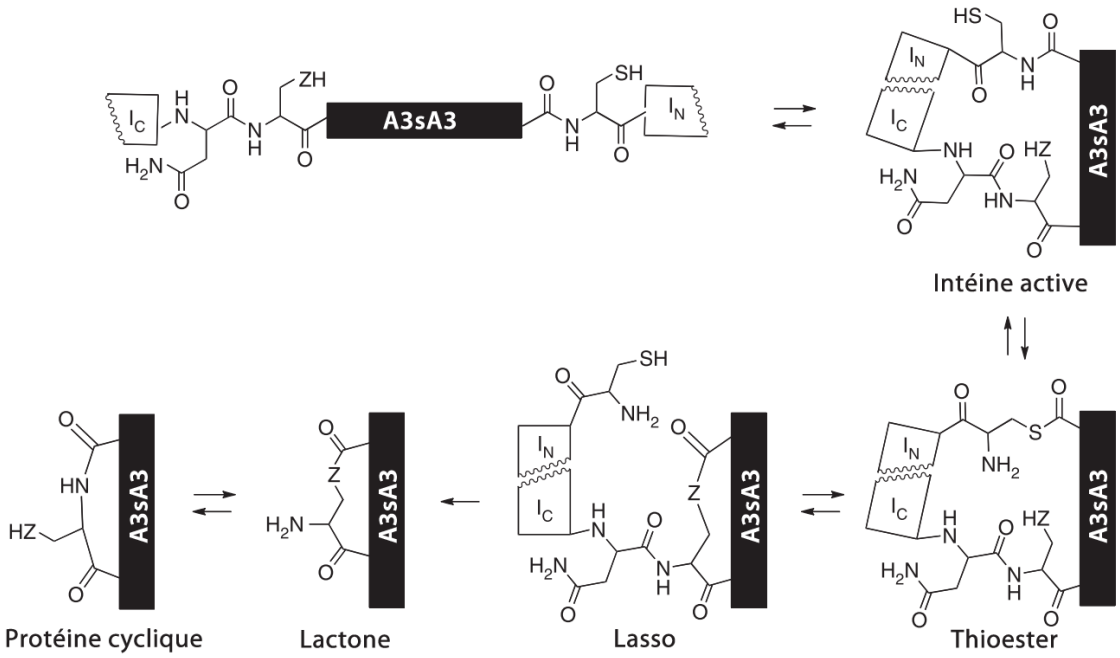


Figure 170 : Méthode SICLOPPS utilisant les split-intein. La protéine à circulariser (ici A3sA3) est placée entre les deux fragments de l'intéine, séparés par deux séquences spécifiques (AsnSer ou AsnCys en amont et Cys en aval). Une fois synthétisée l'intéine peut se reconstituer en solution et catalyser le réarrangement de la chaine peptidique. Un transfert acyl N-vers-S au niveau du fragment amont de la protéine cible produit un thioester, qui effectue une transestérification avec la chaine latérale nucléophile de l'acide aminé Serine ou Cystéine (Z = O ou S) pour former un intermédiaire en forme de lasso (par analogie à la procédure d'épissage génétique). Le résidu asparagine est ensuite clivé libérant la protéine cyclisée sous forme de lactone qui bascule enfin vers la forme lactame in vivo.

Chapitre III - Poursuite du projet

III.1 Vers une stratégie d'évolution dirigée basée sur A3sA3

Cette étude permet d'envisager concernant la création d'une banque de protéines pour l'évolution dirigée du bi-domaine αRep A3sA3 ? Différentes stratégies peuvent être mises en place, qui sont discutées ici. Le but de cette thèse était d'étudier les possibilités de création d'un catalyseur biohybride à base de la protéine αRep A3. Le principe des protéines biohybrides étant d'adapter la poche de la protéine pour fixer et orienter spécifiquement le substrat relativement au métal, il faut maintenant étudier une stratégie afin de faire varier la poche de la protéine A3sA3. Deux stratégies sont rendues possibles avec les αRep : la création d'une banque de grande taille ou de petite taille.

III.1.1 Création d'une banque de grande taille

Le grand avantage des protéines αRep est la facilité avec laquelle de grandes banques de protéines peuvent être créées, par la polymérisation de micro-gènes dégénérés codant chacun un motif. Ainsi les banques αRep contiennent communément plus de 10^9 protéines différentes. La protéine A3sA3 est cependant un peu différente, car une position doit nécessairement être laissée intacte pour le greffage ciblé d'un cofacteur. Il est alors nécessaire d'utiliser la combinaison de deux stratégies.

Prenant avantage du caractère non répété du gène *A3s*, il est possible de modifier les positions variables de ce gène de manière spécifique grâce aux techniques de mutagénèse dirigée. Ce faisant, une petite banque de gènes *A3s* peut être créée, sans risquer de modifier la position sélectionnée pour le greffage du cofacteur.

Le gène *A3* codant l'autre domaine de la protéine peut, lui, être complètement re-synthétisé grâce à la méthode utilisée pour concevoir la banque αRep avec un schéma de mutagénèse spécifique, créant très efficacement une banque de très grande taille.[8] Une méthode plus simple bien que moins efficace serait d'importer les séquences depuis une banque déjà existante (la banque des gènes à 4 motifs depuis la banque 1.0 par exemple).

On peut résumer la création d'une banque de grande taille contenant un acide aminé prévu pour le greffage d'un cofacteur avec la figure 171.

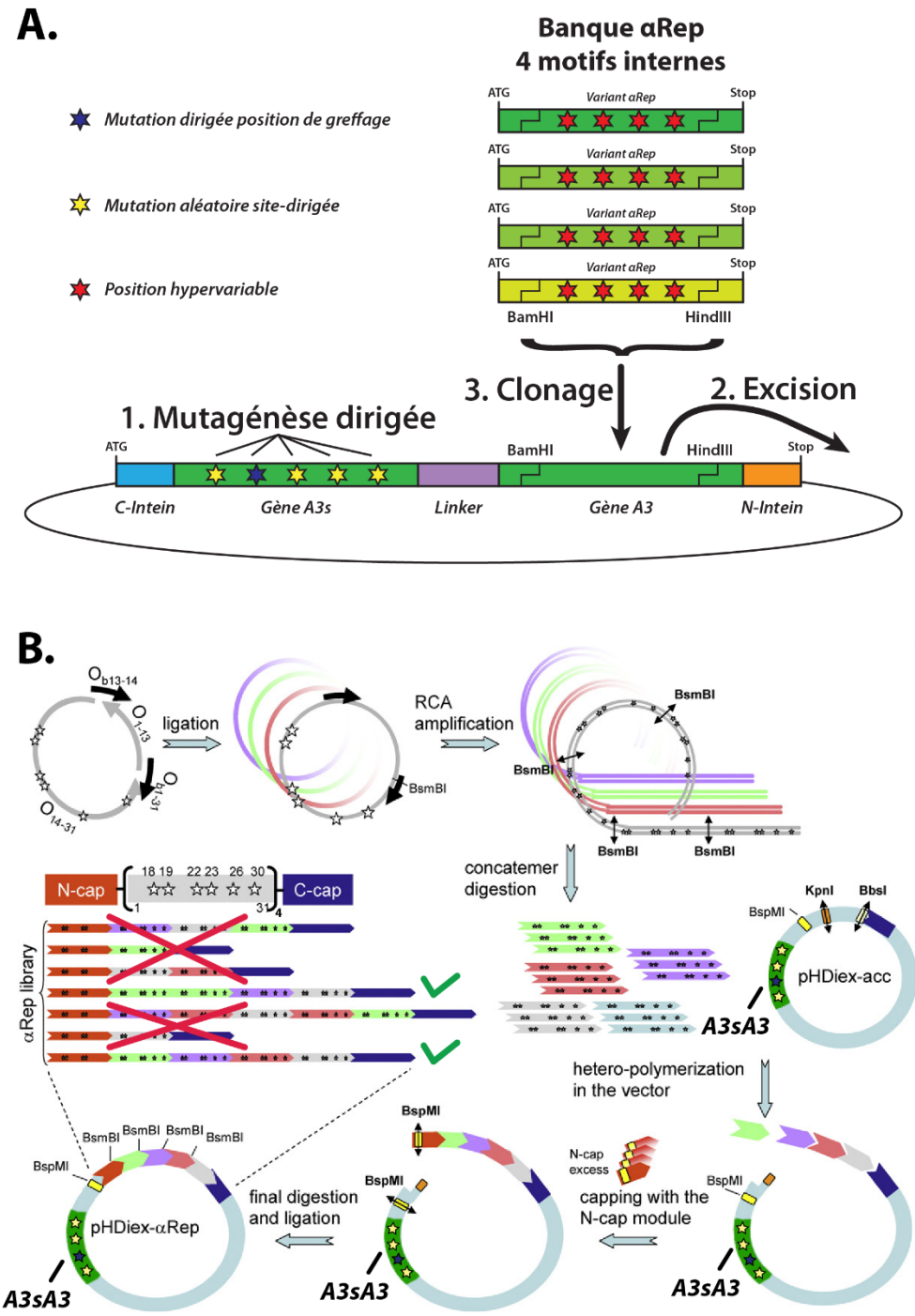


Figure 171 : Stratégies de création d'une banque de crevasses basée sur la protéine αRep A3sA3 en conservant une position spécifique pour le greffage d'un cofacteur métallique. A. 1. Le gène A3s est muté par la méthode QuikChange afin de randomiser sélectivement certaines positions de la crevasse tout en gardant la ceinture hydrophobe et la position choisie pour le greffage du cofacteur métallique. 2. La portion contenant le gène A3 est excisé par digestion enzymatique. 3. Les gènes d'une banque αRep contenant 4 motifs internes sont insérés au sein du vecteur A3sA3 à la place du gène A3 d'origine. Le gène A3sA3 est entouré de la split-intein pour circulariser la protéine après expression. B. Modification de la stratégie classique des banques αRep où le gène A3s est présent en amont de l'insertion des microgènes. Seuls les gènes contenant 4 microgènes insérés sont sélectionnés sur gel d'agarose.

III.1.2 Création d'une banque de petite taille

Cette banque de grande taille est utile si un système de crible haut débit est possible. Dans le cas où la méthode de criblage ne permet que de tester qu'un faible nombre de variants, une stratégie entièrement basée sur la mutagénèse dirigée peut être effectuée, en utilisant deux gènes A3s dé-répétés. Ainsi, des stratégies de type CAST ou ISM, décrite dans l'introduction, sont possibles.

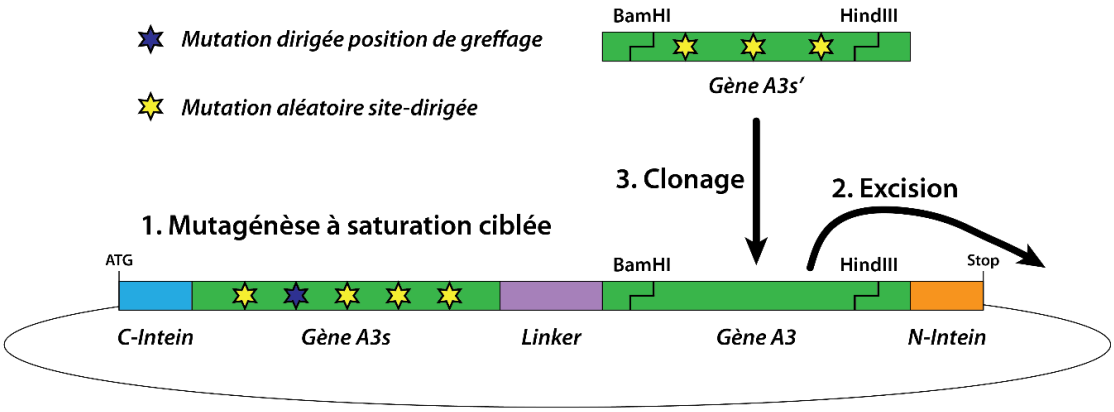


Figure 172 : Stratégie de création d'une banque de petite taille basée sur A3sA3. L'utilisation de deux gènes dé-répétés A3s et A3s' permet d'introduire des mutations à certaines positions très spécifiques, permettant une stratégie CAST ou ISM.

Ces considérations prises en compte, il faut à présent se pencher sur le problème qu'est le greffage du cofacteur métallique. En effet pour permettre un criblage à grande échelle, l'association entre protéine et cofacteur doit se produire dans des conditions les plus simples possibles car il n'est pas possible d'envisager de longues étapes de purification. Deux stratégies sont alors possibles.

III.1.3 Stratégies de liaison du cofacteur

La première option est une liaison par chimie bio-orthogonale (« click ») présentée dans l'introduction. La chimie bio-orthogonale permet un couplage covalent dans un milieu très complexe tel que le cytosol d'une cellule, les groupements fonctionnels communément employés étant le couple azoture/cyclooctyne car ils ne nécessitent pas de catalyseur au cuivre I (figure 173). [9] Le triplet codant la cystéine employé dans cette étude peut être modifié en codon de type ambre transcrit en un acide aminé « click » par une souche bactérienne modifiée. Ainsi il est possible d'effectuer le couplage sans étape de purification, alors que les ligands possédant un bras haloacétamide fonctionnaliseraient tous les nucléophiles forts. Il faut cependant prendre en compte qu'une solution

complexe inclut également des traces d'ions métalliques en solution. Il est alors nécessaire soit de clarifier la solution par une étape d'immobilisation en microplaque (comme utilisé par l'équipe de Ward avec des billes biotinylées [10]) à l'aide d'un marquage par biotine par exemple, soit d'utiliser un ligand préalablement métallé.

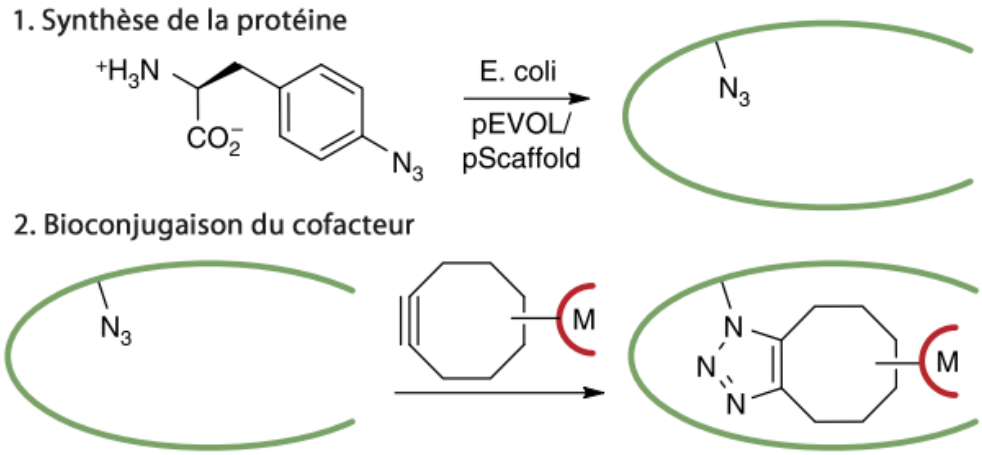


Figure 173 : Stratégie de couplage d'un cofacteur métallique par acide aminé « click ». 1. La protéine est produite dans une bactérie possédant le système orthogonal pour l'insertion d'un acide aminé de type azoture. 2. Le cofacteur est couplé covalemment à l'aide d'un groupement cyclooctyne (selon [9]).

La deuxième option repose également sur des acides aminés non naturels. Pour s'affranchir de la phase de couplage, l'utilisation d'un acide aminé chélatant directement les métaux de transition tel que l'acide aminé bipyridine-alanine utilisé avec succès par l'équipe de Roelfes[11] peut être une solution satisfaisante. La bipyridine étant un ligand bidenté peu spécifique à l'instar de la phénanthroline, il est également nécessaire de prendre en compte la contamination par des traces de métaux. Dans l'étude citée ci-dessus, une étape de purification était effectuée. Une immobilisation en microplaque peut également être envisagée après biotinylation comme évoqué ci-dessus.

III.1.4 Méthodes de criblage

La réussite de toute approche d'évolution dirigée repose sur la mise au point d'une méthode de criblage robuste et simple à mettre en œuvre. Le choix d'une méthode de criblage appropriée dépend de deux critères. Le premier est la possibilité de couplage et catalyse dans un milieu plus ou moins complexe. Comme évoqué ci-dessus, si le système étudié est sensible aux traces de contaminations de métaux de transition il est alors nécessaire de passer par une étape d'immobilisation en microplaque, après expression de la protéine dans le surnageant de culture. Si le système est plus robuste, il est alors

possible d'envisager une expression périplasmique comme récemment proposé par Ward [12] et ainsi bénéficier du lien génotype-phénotype offert par la bactérie (voir figure 174).

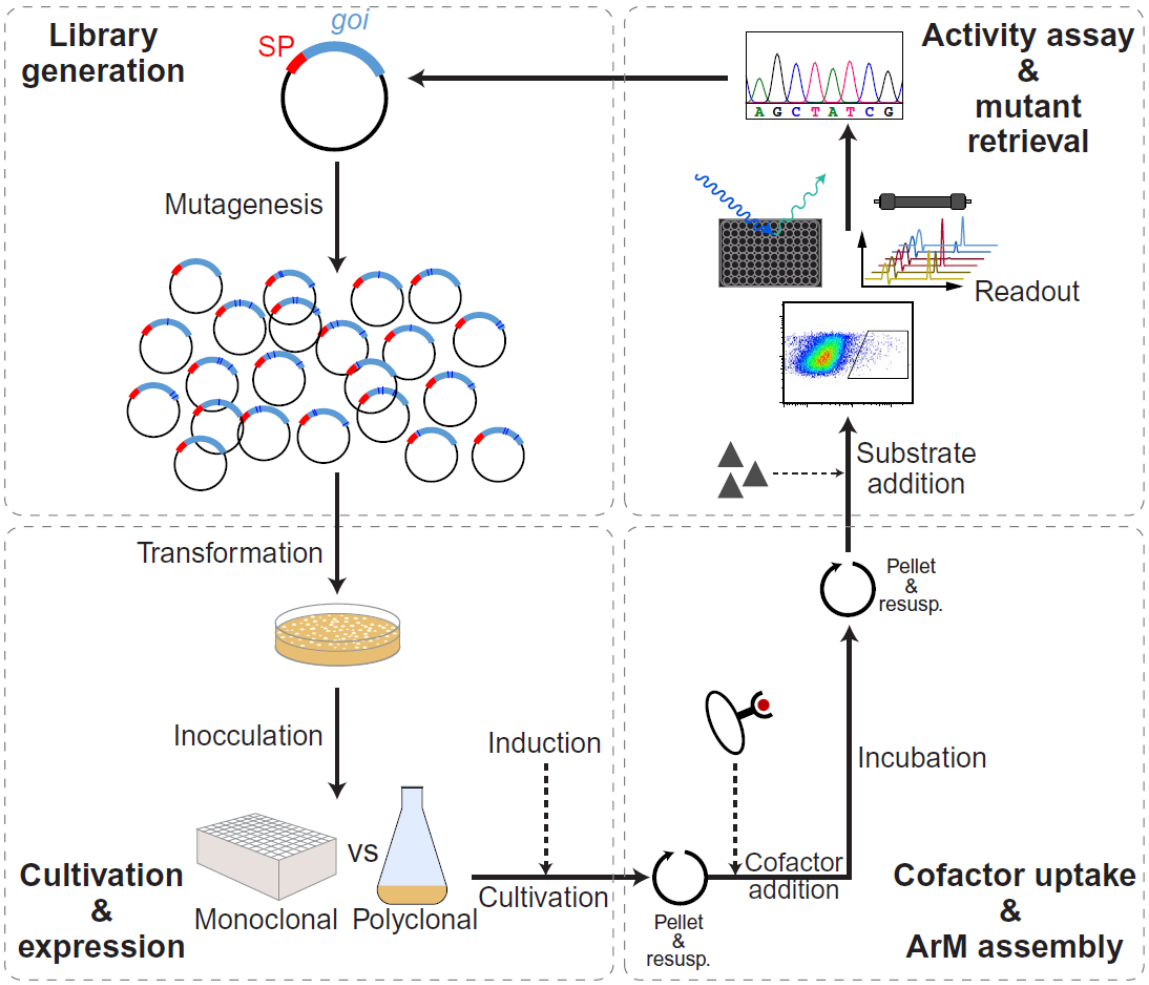


Figure 174 : Stratégie d'évolution dirigée d'un biohybride proposée par Ward, avec expression périplasmique de la banque de protéines. Cette stratégie est adaptable à l'évolution d'un biohybride avec ligand lié covalemment à l'aide de chimie click

Le deuxième critère dans la mise au point d'une méthode de criblage, et encore aujourd'hui le critère limitant, est celui de la disponibilité d'une méthode de suivi de la réaction. Comme abordé dans le premier chapitre de l'introduction, les substrats fluorogènes ou chromogènes sont les plus adaptés à un criblage haut débit. Le suivi peut alors se faire en microplaque avec robot de lecture, ou la mise en place d'un système microfluidique avec méthode de tri peut être envisagé, permettant un criblage très haut débit d'une banque de variants. Dans ce cas, une grande banque de protéine pourra être employée.

Si l'énantiosélectivité est le critère de sélection, il n'est pas possible de discriminer par fluorescence ou absorption. Les méthodes de criblage associées sont donc plus bas débit,

avec la spectrométrie de masse (si possible) et les techniques de chromatographie gazeuse ou liquide. Dans ce cas, une petite banque de protéine devra être employée.

La méthode de criblage est donc le dernier frein technologique à la mise en place d'un système d'évolution dirigée à grande échelle.

Est-il cependant nécessaire de cribler une banque très grande pour obtenir de hauts rendements et énantiosélectivités ? L'équipe de John F. Hartwig, poursuivant les travaux sur les protéines à porphyrine d'iridium abordés dans l'introduction[13], tend à prouver le contraire. Cette équipe a très récemment publié les résultats d'une étude d'évolution dirigée d'un variant de cytochrome P450 de *S. solfataricus* (une archée hyperthermophile), et notamment un variant catalysant l'insertion de carbène intramoléculaire au sein de liaison C-H non activée avec 98% d'excès énantiométique et des paramètres cinétiques équivalents à ceux d'une enzyme naturelle (k_{cat} = 45,8 min^{-1}, k_{cat}/K_M= 269 $min^{-1}.mM^{-1}$).[14] Ce variant a été découvert après un crible en plaque suivi d'extraction manuelle des produits et analyse en chromatographie gazeuse de « seulement » 162 mutants.

La construction de banques de grande taille pour la création de catalyseurs biohybrides est-elle une vaine croisade ? La réponse est sans doute au cas-par-cas. L'exemple des travaux de l'équipe d'Hartwig est très encourageante mais reste pour le moment encore une exception, de plus la haute réactivité de la porphyrine d'iridium est également responsable de la grande efficacité de catalyse.[13] L'utilisation de métaux non nobles et accessibles en grande quantité est un enjeu important dans la chimie catalytique. Développer des réactions avec des complexes moins efficaces nécessitera donc une activation plus poussée du substrat par la protéine, alors une possibilité d'évolution dirigée à grande échelle sera indispensable.

III.2 Utilisation des αRep pour une autre stratégie : la liaison directe du métal

Cette thèse s'est axée sur l'utilisation des αRep dans la stratégie du couplage covalent d'un complexe métallique de synthèse. Nous avons cependant également envisagé un autre type d'approche qu'est la complexation directe du métal par les chaines latérales des acides aminés.

Nous avons vu dans l'introduction que l'ingénierie d'un nouveau site de fixation d'un métal au sein d'une protéine pouvait se révéler particulièrement complexe, et qu'il est avantageux de prendre exemple sur une protéine homologue possédant déjà une affinité pour un métal de transition. Or il existe une protéine homologue aux αRep et aux propriétés particulièrement intéressantes : la DesOxyHypusine Hydroxylase (DOHH). L'analyse de la séquence de cette protéine indique qu'il s'agit d'une protéine à motifs répétés de type PBS-HEAT repeat, tout comme les protéines αRep. Les informations structurales de cette protéine au moment où nous avons commencé à nous y intéresser étaient peu précises car aucune structure cristallographique n'existait. Il était cependant proposé par différentes approches que cette protéine était sous la forme d'un bi-domaine, et formait en son sein un centre bi-nucléaire 1,2µ-perroxo Fe (III)[15], semblable à celui de la méthane monooxygénase hydroxylase soluble (sMMOH). Cet intermédiaire formé, reconnaissable à sa couleur bleue, est stable plusieurs heures même à température ambiante chez la DOHH, ce qui est unique.[15]

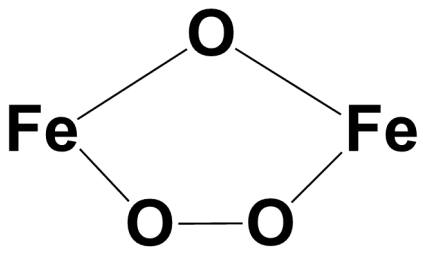


Figure 175 : Structure du coeur bi-Fe(III) µ-1,2 peroxo de la DOHH, l'oxygène pontant provient d'une molécule d'eau.

Comme son nom l'indique, cette protéine catalyse l'hydroxylation de l'hypusine, un acide aminé issu de la modification d'une lysine par une spermidine par la DesoxyHypusine Synthase. Cette protéine catalyse donc une hydroxylation extrêmement regio-sélective d'une chaine carbonée saturée.

Il semble tout à fait intéressant de combiner la robustesse et la capacité d'évolution des αRep avec le cœur catalytique de la DOHH. J'ai au cours de ma thèse tenté d'obtenir la structure de cette protéine. Sachant que dans plusieurs cas, des αRep ont permis d'obtenir des structure cristallographiques des protéines difficiles à cristalliser, un

partenaire protéique αRep a pour cela été sélectionné contre la DOHH humaine par un crible utilisant une complémentation fonctionnelle dans les bactéries (PCA) afin de servir de partenaire de cristallisation. Cependant, la structure a été publiée entre-temps.[16]

La structure confirme que la protéine est effectivement une protéine à deux domaines de type HEAT repeat, une superposition avec les modules d'A3 montre la très forte ressemblance entre les deux protéines (figure 176).

Le but serait alors de modifier une αRep pour y insérer le même site de coordination de la DOHH. Deux stratégies sont possibles : la modification progressive de la DOHH vers des séquences consensus des αRep, ou la modification des chaines variables d'une αRep afin de présenter le même site de coordination du métal. Il faut pour cela concevoir une αRep faite de deux domaines à 4 motifs avec le même linker, ayant les mêmes résidus aux positions variables que la DOHH. Il serait alors intéressant d'observer si la géométrie des deux domaines est équivalente. Une telle étude pourrait engendrer de nouvelles connaissances concernant les deux protéines : d'une part apporter des informations sur les facteurs responsables de la dimérisation des αRep et de leur potentielle aptitude à fixer de manière directe les métaux de transition. D'autre part, la protéine DOHH étant une protéine peu commode (agrégation aux hautes concentrations, faible stabilité thermique), l'utilisation de l'architecture stable des αRep permettrait de comprendre plus en détail cette protéine.

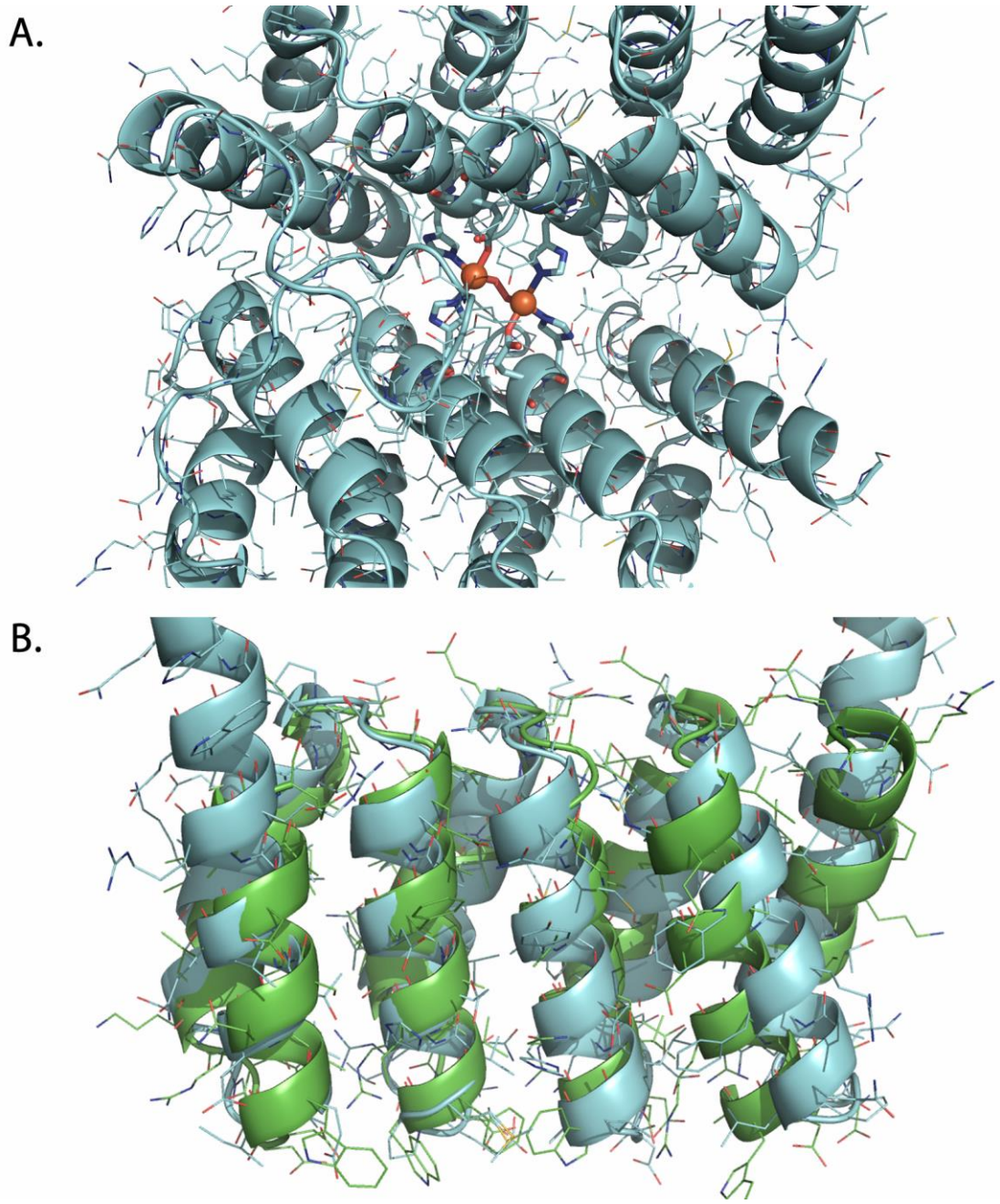


Figure 176 : A. Structure cristallographique de la DOHH humaine. On peut observer les deux ions Fe (III) coordonnés par deux histidines et un glutamate chacun, avec un ligand μ-1,2 peroxo pontant entre les deux (selon [16]). B. Superposition d'un domaine de la hBOHH (bleu) avec les 4 modules de la protiéine A3, on peut constater la très forte similitude entre les deux structures, toutes deux appartenant à la famille des protéines HEAT repeat.

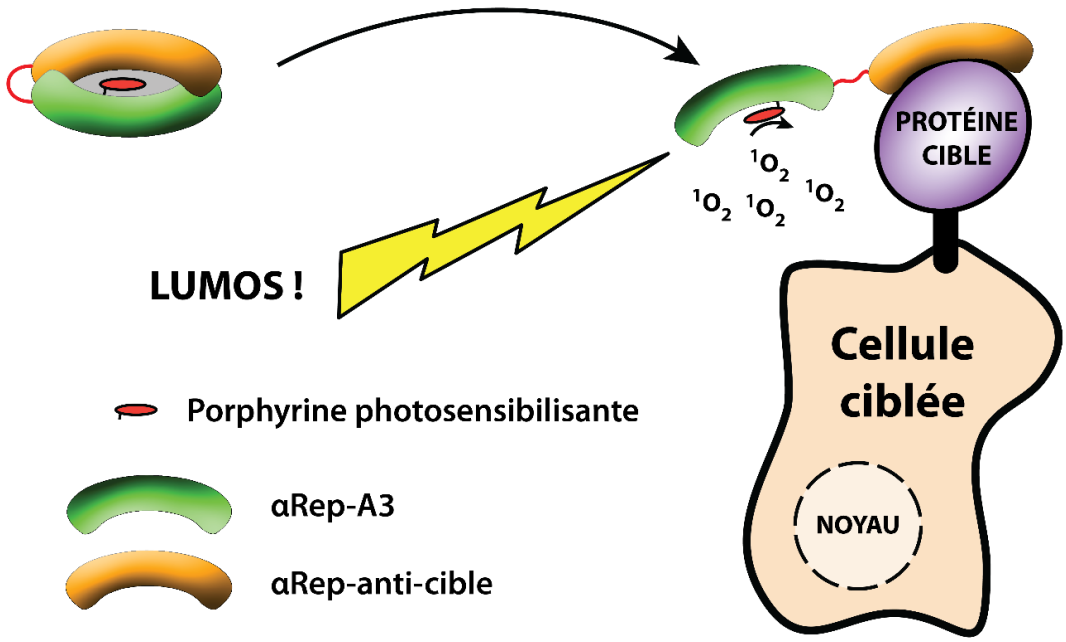


Figure 177 : Stratégie de photosensibilisation de cellules spécifiques par l'utilisation d'un catalyseur commutable. L'αRep orange est un variant dirigé contre une protéine surexprimée par une cellule en particulier. En absence du partenaire protéique, le couple d'αRep piège la porphyrine, la solubilise et la protège de l'oxydation. L'association de l'αRep avec le partenaire protéique ouvre la crevasse et libère la porphyrine qui peut alors catalyser des réactions, comme par exemple la production d'espèces réactives d'oxygène dans le cas de la destruction d'une cellule.

III.3 Utilisation d'un catalyseur commutable

Le greffage covalent d'une porphyrien de type Mn(III)TPP au sein de la crevasse du variant A3sA3Y26C a permis d'observer un effet d'encapsulation de la porphyrine par la protéine. Cette encapsulation protège la porphyrine des attaques oxydatives, mais inhibe dans le même temps sa réactivité. L'ouverture de la chimère par l'utilisation d'un partenaire protéique du domaine non modifié a permis de recouvrer son activité catalytique. Ces résultats préliminaires semblent indiquer que ce biohybride pourrait devenir la première métalloenzyme artificielle commutable par interaction protéique connue à ce jour.

Ces travaux ouvrent un nouveau champ d'application dans le domaine de la biologie cellulaire, à visée thérapeutique : un autre aspect de notre approche consisterait à exploiter la capacité de la protéine bi-domaine à s'ouvrir et à libérer une activité catalytique de manière commutable par l'ajout d'un partenaire protéique. Des travaux effectués par Corentin Léger, un autre doctorant du laboratoire, semblent montrer une tendance générale des αRep à adopter une forme fermée si elles sont liées par un linker. La protéine ne s'ouvrirait donc qu'en cas d'interaction avec un partenaire, comme il est visible avec A3sA3 greffé avec la porphyrine. Dans la chimère A3sA3, le domaine A3 peut alors être remplacé par une autre αRep reconnaissant une protéine spécifique, il est très probable que le phénomène d'encapsulation s'opère également. L'activité de la porphyrine pourra donc être révélée uniquement en présence du partenaire protéique, surexprimé à la surface d'une cellule cancéreuse par exemple (figure 177). La tétraphénylporphyrine (TPP) peut effectivement être utilisée dans des traitements anticancéreux par thérapie photodynamique, la ZnTPP provoquant la libération de dérivés réactifs de l'oxygène tels que l'oxygène singulet après irradiation.[17,18]

Pour conclure, les protéines αRep, qui avaient déjà démontré leur efficacité comme protéines alternatives aux anticorps et partenaires de cristallisation [19–22] sont également une architecture robuste et versatile ouvrant tout un panel de possibilités pour l'ingénierie de nouvelles métalloenzymes artificielles. Les avantages et limites du dimère αRep-A3 ont pu être appréciés au cours de cette thèse, les nouvelles perspectives de création d'une banque pour l'évolution dirigée d'un catalyseur biohybride sont maintenant à portée de main.

Bibliographie

(1) Chakravarty, D., Janin, J., Robert, C. H., et Chakrabarti, P. (2015) Changes in protein structure at the interface accompanying complex formation. *IUCrJ 2*, 643-652.

(2) Hu, D., Hu, S., Wan, W., Xu, M., Du, R., Zhao, W., Gao, X., Liu, J., Liu, H., et Hong, J. (2015) Effective optimization of antibody affinity by phage display integrated with high-throughput DNA synthesis and sequencing technologies. *PLoS One 10*, 1-17.

(3) Chen, X., Zaro, J. L., et Shen, W. C. (2013) Fusion protein linkers: Property, design and functionality. *Adv. Drug Deliv. Rev. 65*, 1357-1369.

(4) van Dongen, E. M. W. M., Evers, T. H., Dekkers, L. M., Meijer, E. W., Klomp, L. W. J., et Merkx, M. (2007) Variation linker length in ratiometric fluorescent sensor protein allows rational tuning of Zn(II) affinity in the pM to fM range [CFP Atox1 FRET to YFP]. *J. Am. Chem. Soc. 129*, 3494-3495.

(5) Tavassoli, A., et Benkovic, S. J. (2007) Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in E. coli. *Nat. Protoc. 2*, 1126-1133.

(6) Iwai, H., Lingel, A., et Plückthun, A. (2001) Cyclic Green Fluorescent Protein Produced in Vivo Using an Artificially Split PI-PfuI Intein from Pyrococcus furiosus. *J. Biol. Chem. 276*, 16548-16554.

(7) Waldhauer, M. C., Schmitz, S. N., Ahlmann-Eltze, C., Gleixner, J. G., Schmelas, C. C., Huhn, A. G., Bunne, C., Büscher, M., Horn, M., Klughammer, N., Kreft, J., Schäfer, E., Bayer, P. A., Krämer, S. G., Neugebauer, J., Wehler, P., Mayer, M. P., Eils, R., et Di Ventura, B. (2015) Backbone circularization of Bacillus subtilis family 11 xylanase increases its thermostability and its resistance against aggregation. *Mol. Biosyst. 11*, 3231-43.

(8) Urvoas, A., Guellouz, A., Valerio-Lepiniec, M., Graille, M., Durand, D., Desravines, D. C., van Tilbeurgh, H., Desmadril, M., et Minard, P. (2010) Design, production and molecular structure of a new family of artificial alpha-helicoidal repeat proteins (αRep) based on thermostable HEAT-like repeats. *J. Mol. Biol. 404*, 307-27.

(9) Srivastava, P., Yang, H., Ellis-Guardiola, K., et Lewis, J. C. (2015) Engineering a dirhodium artificial metalloenzyme for selective olefin cyclopropanation. *Nat. Commun. 6*, 7789.

(10) Creus, M., Pordea, A., Rossel, T., Sardo, A., Letondor, C., Ivanova, A., LeTrong, I., Stenkamp, R. E., et Ward, T. R. (2008) X-ray structure and designed evolution of an artificial transfer hydrogenase. *Angew. Chemie - Int. Ed. 47*, 1400-1404.

(11) Drienovská, I., Rioz-Martínez, A., Draksharapu, A., et Roelfes, G. (2014) Novel artificial metalloenzymes by in vivo incorporation of metal-binding unnatural amino acids. *Chem. Sci. 6*, 770-776.

(12) Jeschek, M., Panke, S., et Ward, T. R. (2016) Periplasmic Screening for Artificial Metalloenzymes, in *Methods in enzymology*.

(13) Key, H. M., Dydio, P., Clark, D. S., et Hartwig, J. F. (2016) Abiological catalysis by artificial haem proteins containing noble metals in place of iron. *Nature 534*.

(14) Dydio, P., Key, H. M., Nazarenko, A., Rha, J. Y.-E., Seyedkazemi, V., Clark, D. S., et Hartwig, J. F. (2016) An artificial metalloenzyme with the kinetics of native enzymes. *Science (80-.). 354*.

(15) Vu, V. V, Emerson, J. P., Martinho, M., Kim, Y. S., Münck, E., Park, M. H., et Que, L. (2009) Human deoxyhypusine hydroxylase, an enzyme involved in regulating cell growth, activates O2 with a nonheme diiron center. *Proc. Natl. Acad. Sci. U. S. A. 106*, 14814-9.

(16) Han, Z., Sakai, N., Böttger, L. H., Klinke, S., Hauber, J., Trautwein, A. X., et Hilgenfeld, R. (2015) Crystal Structure of the Peroxo-diiron(III) Intermediate of Deoxyhypusine Hydroxylase, an Oxygenase Involved in Hypusination. *Structure* 882-892.

(17) Nakamura, H., Liao, L., Hitaka, Y., Tsukigawa, K., Subr, V., Fang, J., Ulbrich, K., et Maeda, H. (2013) Micelles of zinc protoporphyrin conjugated to N-(2-hydroxypropyl) methacrylamide (HPMA) copolymer for imaging and light-induced antitumor effects in vivo. *J. Control. Release 165*, 191-198.

(18) Yao, X., Chen, L., Chen, X., Xie, Z., Ding, J., He, C., Zhang, J., et Chen, X. (2015) PH-responsive metallo-supramolecular nanogel for synergistic chemo-photodynamic therapy. *Acta Biomater. 25*, 162-171.

(19) Figueroa, M., Sleutel, M., Vandevenne, M., Parvizi, G., Attout, S., Jacquin, O., Vandenameele, J., Fischer, A. W., Damblon, C., Goormaghtigh, E., Valerio-Lepiniec, M., Urvoas, A., Durand, D., Pardon, E., Steyaert, J., Minard, P., Maes, D., Meiler, J., Matagne, A., Martial, J. A., et Van de Weerdt, C. (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.

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

(21) Gurunatha, K. L., Fournier, A. C., Urvoas, A., Valerio-Lepiniec, M., Marchi, V., Minard, P., et Dujardin, E. (2016) Nanoparticles Self-Assembly Driven by High Affinity Repeat Protein Pairing. *ACS Nano 10*, 3176-3185.

(22) Tiouajni, M., Durand, D., Blondeau, K., Graille, M., Urvoas, A., Valerio-Lepiniec, M., Guellouz, A., Aumont-Nicaise, M., Minard, P., et Van Tilbeurgh, H. (2014) Structural and functional analysis of the fibronectin-binding protein FNE from Streptococcus equi spp. equi. *FEBS J. 281*, 5513-5531.

Partie E - Annexes

Tables cristallographiques

Title: A3sY41NPH
Date: 10–OCT–15
PDB code: xxxx

Crystal

Cell parameters:

a: 78.14 A b: 46.53 A c: 110.10 A
α: 90.00 β: 91.61 γ: 90.00

Space group: C 1 2 1

Model

3330 atoms (214 water molecules)

Number of chains:	9
Volume not occupied by model:	29.3 %
<B> (for atomic model):	27.4 A^2
σ(B):	11.48 A^2
Matthews coefficient:	2.15
Corresponding solvent % :	42.23

Refinement

Program:	REFMAC 5.8.0158
Nominal resolution range:	40.0 – 1.99 A
Reported R–factor:	0.185
Number of reflections used:	25848
Reported Rfree:	0.24
Sigma cut–off:	N.A.

Structure Factors

Input

Nominal resolution range:	40.0 – 1.99 A
Reflections in file:	27207
Unique reflections above 0:	27207
above 1σ:	26474
above 3σ:	16390

SFCHECK

Nominal resolution range:	40.0 – 1.99 A
\05max. from input data, min. from author\05	
Used reflections:	27203
Reflections out of resolution:	4
Completeness:	99.2 %
R_stand(F) = <σ(F)>/<F> :	0.080
Anisotropic distribution of Structure Factors	
ratio of eigen values: 0.6376 0.8036 1.0000	
B_overall (by Patterson):	26.A^2
Optical resolution:	1.53 A
Expected opt. resol. for complete data set:	1.53 A
Estimated minimal error:	0.064 A

Model vs. Structure Factors

R–factor for all reflections:	0.217
Correlation factor:	0.933
R–factor:	0.221
for F > 2.0 σ	
nom. resolution range:	39.97 – 1.99A
reflections used:	26494
Rfree:	0.268
Nfree:	1319
R–factor without free–refl.:	0.218
Non free–reflections:	25175
<u> (error in coords by Luzzati plot):	0.236 A
Estimated maximal error:	0.177 A
DPI:	0.165 A

Scaling

Scale:	0.017
Bdiff:	–3.50
Anisothermal Scaling (Beta):	
0.0332 –0.2245 2.7817 –0.0000 0.1326 0.0000	
Solvent correction – Ks,Bs:	0.900 218.924

Title: A3sD165NPH
Date: 20–MAR–15
PDB code: xxxx

Crystal

Cell parameters:

a: 67.03 A b: 67.03 A c: 183.27 A
α: 90.00 β: 90.00 γ: 120.00

Space group: P 65

Model

3218 atoms (224 water molecules)

Number of chains:	8
Volume not occupied by model:	40.9 %
<B> (for atomic model):	42.6 A^2
σ(B):	14.31 A^2
Matthews coefficient:	2.64
Corresponding solvent % :	53.01

Refinement

Program:	REFMAC 5.8.0158
Nominal resolution range:	49.0 – 2.00 A
Reported R–factor:	0.188
Number of reflections used:	29709
Reported Rfree:	0.23
Sigma cut–off:	N.A.

Structure Factors

Input

Nominal resolution range:		49.0 – 1.80 A
Reflections in file:		42841
Unique reflections	above 0:	42841
	above 1σ:	40648
	above 3σ:	20727

SFCHECK

Nominal resolution range:	49.0 – 2.00 A
\05max. from input data, min. from author\05	
Used reflections:	31269
Reflections out of resolution:	11572
Completeness:	99.6 %
R_stand(F) = <σ(F)>/<F> :	0.058
Anisotropic distribution of Structure Factors	
ratio of eigen values: 0.4166 0.4166 1.0000	
B_overall (by Patterson):	32.A^2
Optical resolution:	1.63 A
Expected opt. resol. for complete data set:	1.63 A
Estimated minimal error:	0.024 A

Model vs. Structure Factors

R–factor for all reflections:	0.229
Correlation factor:	0.935
R–factor:	0.233
for F > 2.0 σ	
nom. resolution range: 49.04 – 2.00A	
reflections used: 31143	
Rfree:	0.271
Nfree:	1564
R–factor without free–refl.:	0.231
Non free–reflections:	29579
<u> (error in coords by Luzzati plot):	0.275 A
Estimated maximal error:	0.121 A
DPI:	0.115 A

Scaling

Scale:	0.013
Bdiff:	–6.40

Anisothermal Scaling (Beta):
–1.5493 –1.5493 7.1488 –0.7745 –0.0000 –0.0000
Solvent correction – Ks,Bs: 0.900 431.061

Title: A3sA3WT
Date: 30–MAR–15
PDB code: xxxx

Crystal

Cell parameters:

a: 89.87 A	b: 46.68 A	c: 63.34 A
α: 90.00	β: 128.91	γ: 90.00

Space group: C 1 2 1

Model

1530 atoms (56 water molecules)

Number of chains:	10
Volume not occupied by model:	37.4 %
<B> (for atomic model):	45.6 A^2
σ(B):	14.75 A^2
Matthews coefficient:	2.41
Corresponding solvent % :	48.63

Refinement

Program:	REFMAC 5.8.0135
Nominal resolution range:	44.7 – 1.94 A
Reported R–factor:	0.175
Number of reflections used:	14285
Reported Rfree:	0.22
Sigma cut–off:	N.A.

Structure Factors

Input

Nominal resolution range:	44.7 – 1.94 A
Reflections in file:	15038
Unique reflections above 0:	15038
above 1σ:	14718
above 3σ:	8680

SFCHECK

Nominal resolution range:	44.7 – 1.94 A
\05max. from input data, min. from author\05	
Used reflections:	15036
Reflections out of resolution:	2
Completeness:	98.3 %
R_stand(F) = < σ(F)>/<F> :	0.072
Anisotropic distribution of Structure Factors	
ratio of eigen values: 0.4806 1.0000 0.9826	
B_overall (by Patterson):	30.A^2
Optical resolution:	1.58 A
Expected opt. resol. for complete data set:	1.58 A
Estimated minimal error:	0.068 A

Model vs. Structure Factors

R–factor for all reflections:	0.214
Correlation factor:	0.936
R–factor:	0.220
for F > 2.0 σ	
nom. resolution range:	44.69 – 1.94A
reflections used:	14752
Rfree:	0.272
Nfree:	735
R–factor without free–refl.:	0.217
Non free–reflections:	14017
<u> (error in coords by Luzzati plot):	0.225 A
Estimated maximal error:	0.300 A
DPI:	0.139 A

Scaling

Scale:	0.019
Bdiff:	–10.98

Anisothermal Scaling (Beta):
1.8117 1.7471 1.6767 0.0000 2.8588 0.0000
Solvent correction – Ks,Bs: 0.900 261.790

Title: A3sA3F134NPH
Date: 25–AUG–17
PDB code: xxxx

Crystal

Cell parameters:

a: 90.04 A b: 46.90 A c: 64.15 A
α: 90.00 β: 127.43 γ: 90.00

Space group: C 1 2 1

Model

1526 atoms (79 water molecules)

Number of chains:	7
Volume not occupied by model:	39.5 %
<B> (for atomic model):	47.2 A^2
σ(B):	13.19 A^2
Matthews coefficient:	2.52
Corresponding solvent % :	50.75

Refinement

Program:	REFMAC 5.8.0158
Nominal resolution range:	44.7 – 2.00 A
Reported R–factor:	0.207
Number of reflections used:	13769
Reported Rfree:	0.27
Sigma cut–off:	N.A.

Structure Factors

Input

Nominal resolution range:		44.7 – 1.80 A
Reflections in file:		19826
Unique reflections	above 0:	19826
	above 1σ:	15105
	above 3σ:	6904

SFCHECK

Nominal resolution range: 44.7 – 2.00 A
\05max. from input data, min. from author\05

Used reflections:	14488
Reflections out of resolution:	5338
Completeness:	99.7 %
R_stand(F) = <σ(F)>/<F> :	0.089

Anisotropic distribution of Structure Factors
ratio of eigen values: 0.5067 1.0000 0.9368

B_overall (by Patterson):	33.A^2
Optical resolution:	1.64 A
Expected opt. resol. for complete data set:	1.64 A
Estimated minimal error:	0.106 A

Model vs. Structure Factors

R–factor for all reflections:	0.240
Correlation factor:	0.934
R–factor:	0.244

for F > 2.0 σ
nom. resolution range: 44.71 – 2.00A
reflections used: 13875

Rfree:	0.328
Nfree:	692
R–factor without free–refl.:	0.239
Non free–reflections:	13183
<u> (error in coords by Luzzati plot):	0.278 A
Estimated maximal error:	0.364 A
DPI:	0.137 A

Scaling

Scale:	0.015
Bdiff:	–9.03

Anisothermal Scaling (Beta):
3.1768 2.9751 1.3843 –0.0000 3.0380 –0.0000
Solvent correction – Ks,Bs: 0.900 182.037

Article

ChemPubSoc Europe

DOI: 10.1002/chem.201701518

CHEMISTRY A European Journal Full Paper

Chemical Biology

αRep A3: A Versatile Artificial Scaffold for Metalloenzyme Design

Thibault Di Meo,[a, b] Wadih Ghattas,[a] Christian Herrero,[a] Christophe Velours,[b] Philippe Minard,[b] Jean-Pierre Mahy,*[a] Rémy Ricoux,*[a] and Agathe Urvoas*[b]

Abstract: αRep refers to a new family of artificial proteins based on a thermostable α-helical repeated motif. One of its members, αRep A3, forms a stable homo-dimer with a wide cleft that is able to accommodate metal complexes and thus appears to be suitable for generating new artificial biocatalysts. Based on the crystal structure of αRep A3, two positions (F119 and Y26) were chosen, and independently changed into cysteine residues. A phenanthroline ligand was covalently attached to the unique cysteine residue of each protein variant, and the corresponding biohybrids were purified and characterized. Once mutated and coupled to phenanthroline, the protein remained folded and dimeric. Copper(II) was specifically bound by the two biohybrids with two different binding modes. Furthermore, the holo-biohybrid A3F119NPH was found to be capable of enantioselectively catalyzing Diels–Alder (D-A) cycloadditions with up to 62% *ee*. This study validates the choice of the αRep A3 dimer as a protein scaffold and provides a promising new route for the design and production of new enantioselective biohybrids based on entirely artificial proteins obtained from a highly diverse library.

Introduction

Enzyme-catalyzed chemical reactions represent the best eco-friendly scenario in organic synthesis.[1] Enzymes are natural catalysts that can be used in aqueous solution, at room temperature, and under atmospheric pressure and aerobic conditions, while avoiding the formation of undesirable side-products. More than 100 biotransformation processes have already been developed at the industrial scale.[2,3] However, industrial application of natural enzymes is limited by their lack of stability outside of their natural environment and their poor activity or selectivity towards non-natural substrates.

Among the different approaches for obtaining improved enzyme activities, the construction of artificial metalloenzymes or "artzymes" has emerged as one of the most promising.[4–7] Such hybrid biocatalysts combine the efficiency and wide reaction scope of synthetic catalysts with the high selectivity of natural enzymes and their ability to function under mild conditions. These catalysts combine a synthetic metal cofactor that is responsible for the catalytic activity with a protein scaffold that provides protection against degradation. Furthermore, in the assembled hybrid catalysts, the active metal center is bound within a defined and folded protein structure, which provides the metal cofactor with a chiral environment.

Different strategies have been described for obtaining "artzymes".[8] Among these strategies, the first involves the direct insertion of a metal ion into proteins,[9,10] sometimes aided by computational design.[11] Other approaches involve the supramolecular anchoring of metal cofactors into proteins, either by using a binding site of the protein, such as the heme binding site in hemoproteins ("host–guest strategy") or by grafting the metal complex onto a small natural molecule that spontaneously binds to the protein ("Trojan Horse strategy").[12,13] The success of such approaches depends greatly on the affinity between the protein and its anchored ligand. The extremely high stability of the biotin streptavidin conjugate, coupled with the development of new screening approaches of protein variants, has allowed the development of various biocatalysts by Ward and co-workers.[14,15] A third approach involves the covalent attachment of a metal cofactor to a protein, most often through a covalent bond with a cysteine residue.[16–18] This approach does not a priori require any specific interaction between the protein scaffold and the anchored metal complex and can potentially generate interesting hybrid biocatalysts for various reactions. The aforementioned strategies have allowed the generation of different "artzymes", which have been shown to catalyze various reactions, such as hydrogenation, ketone reduction, alcohol oxidation, sulfoxidation, dihydroxylation, peroxidation, hydrolysis, hydroformylation, transamination, and Diels–Alder cycloaddition (for a review, see ref. [19]). For catalysts designed to perform the Diels–Alder reaction, only a few

[a] T. Di Meo, Dr. W. Ghattas, Dr. C. Herrero, Prof. J.-P. Mahy, Dr. R. Ricoux
Institut de Chimie Moléculaire et des Matériaux d'Orsay (ICMMO)
UMR 8182, CNRS, Univ. Paris Sud, Université Paris-Saclay
Bât. 420, rue du Doyen Georges Poitou, 91405 Orsay cedex (France)
E-mail: jean-pierre.mahy@u-psud.fr
remy.ricoux@u-psud.fr

[b] T. Di Meo, Dr. C. Velours, Prof. P. Minard, Dr. A. Urvoas
Institute for Integrative Biology of the Cell (I2BC)
CEA, CNRS, Univ. Paris-Sud, Université Paris-Saclay
91198, Gif-sur-Yvette cedex (France)
E-mail: agathe.urvoas@u-psud.fr

Supporting information and the ORCID identification number for the author of this article can be found under https://doi.org/10.1002/chem.201701518.

© 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

ChemPubSoc Europe

CHEMISTRY A European Journal Full Paper

examples of "artzymes" have hitherto been described in the literature.[20,21] All protein-based "artzymes" have been prepared using natural proteins, and only one of them has been based on a dimeric protein, that is, the transcription factor Lactococcal multidrug resistance Regulator (LmrR).[22] Indeed, Roelfes and co-workers recently created a new class of artificial metalloenzymes by covalently anchoring a copper(II)-phenanthroline complex within the active site formed at the dimer interface of LmrR. The resulting artificial metalloenzyme was found to be able to catalyze the Diels–Alder (D-A) reaction with up to 97% *ee*[22] as well as hydration and the asymmetric vinylogous Friedel–Crafts alkylation of indoles in water, giving rise to *ee* values up to 94%.[23]

Extending the "artzyme" approach to artificial proteins would open a new route towards the generation of entirely artificial protein-based metalloenzymes. In this respect, recent progress in protein engineering should allow the development of new families of proteins that could be easily produced and purified in high yields, which should display appropriate stability, high folding efficiency, and three-dimensional structure for the insertion of metal complexes. Herein, we describe the first generation of artificial biohybrid catalysts based on a recently reported family of non-natural engineered repeat proteins named αReps.[24]

Proteins from the αRep library (1.7×10^9 variants) display the same overall elongated structure and comprise a variable number of repeated motifs.[24,25] Each repeated motif (31 amino acids) forms two roughly antiparallel helices, and consecutive motifs mutually stack to impart the whole protein with a right-handed elongated solenoid shape. The sequence of repeated motifs was defined by consensus design based on a sub-class of HEAT-like repeats commonly found in thermoresistant microorganism proteins. All repeats have the same sequence, except at six defined positions that are naturally highly tolerant to sequence variations. The first and last repeats of each protein, referred to as N- and C-cap repeats, have different sequences in order to shield the hydrophobic interface. In each internal repeat, six hyper-variable positions were randomized according to the natural diversity observed in the homologues of the natural protein, and the series of degenerate positions at the surface of each concatenated motif resulted in a diversified protein surface. Proteins from the library can be efficiently expressed in *E. coli* (10–100 mg per L of culture). They are stable, deprived of cysteine residues, and remain folded throughout a wide range of pH values and temperatures (T_m >70 °C). These favorable biophysical properties, together with the high tolerance to sequence variation on the binding surface, suggest that αRep proteins could be suitable candidates in an "artzyme" engineering approach. Among the numerous αRep proteins, the αRep A3 variant was found to be homodimeric in its crystal structure as well as in solution.[24] This protein contains the capping motifs (N-cap, which contains six variable positions, and the constant C-cap) and four internal variable repeats. Two different X-ray structures of the αRep A3 dimer have previously been resolved and are available in the Protein Data Bank (Figure 1 A).[24] Both structures revealed two monomers facing each other in a head-to-tail fashion. One structure was in a PEG-bound form (3LTM), while the other was empty in a slightly more compact form (3LTJ). These two forms emphasize the flexibility of the αRep A3 dimer, which suggests that it could be well suited to accommodate in its pocket various kinds of metal cofactors and substrates for enzyme design purposes. We attempted to take advantage of this fortuitous wide and amphiphilic pocket as a receptacle for catalytic metal complexes in order to design artificial metalloenzymes. We envisaged the insertion of copper(II) complexes based on the planar aromatic phenanthroline ligand inside this crevice, which would provide a chiral microenvironment through the amino acid side chains surrounding the ligand.

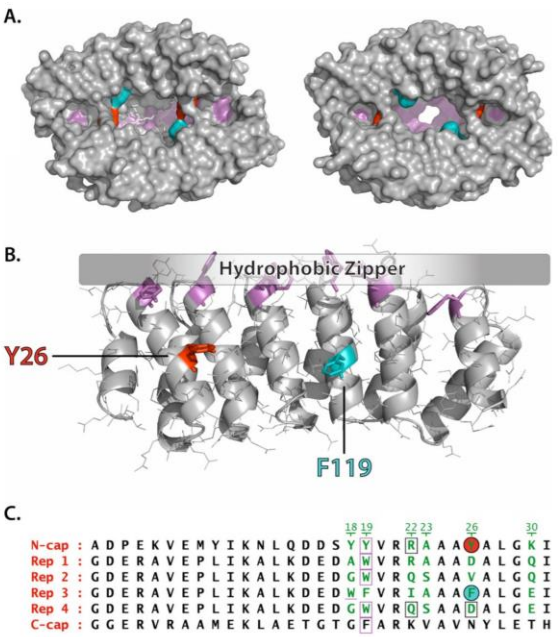


Figure 1. Design of biohybrids based on the αRep protein scaffold. A) Crystallographic structures of the dimer αRep A3: Crevice in the PEG-bound form (left; PDB ID: 3LTM) with the PEG molecule visible in sticks and in the empty form (right; PDB ID: 3LTJ). The hydrophobic zipper is shown in purple. Residues Y26 and F119 are shown in red and cyan, respectively. B) Structure of a monomer of αRep A3: The hydrophobic zipper is shown at the uppermost positions. Residues Y26 and F119 are shown in red and cyan, respectively. C) Sequence alignment of αRep A3 repeats. Variable positions from the αRep library are shown in green; main residues involved in dimer interaction (hydrophobic zipper in purple) are framed; the W111 residue interacting only in the open form is underlined; residues that are candidates for cysteine mutation are circled. The two .pdb files were uploaded into the Robetta Server and a virtual alanine scanning was performed. Framed residues present the highest computed $\Delta\Delta G$ values and are located at the interface of the dimer.

Our approach involved mutating a unique residue in the αRep A3 sequence into cysteine, to which a phenanthroline ligand was covalently attached. Taking into account the dimeric structure of the protein, a single mutation into cysteine would result in the presence of two phenanthroline ligands at symmetric positions in the crevice. Recent results[22] suggested that a generic crevice at the interface between two subunits of

© 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

ChemPubSoc Europe

CHEMISTRY A European Journal Full Paper

a homo-dimeric protein would be suitable for covalent coupling of an organometallic cofactor (copper(II)-phenanthroline) and could induce enantioselective catalysis of Diels–Alder cycloaddition. We thus studied this model reaction and approach as a first step to explore the capacity of αRep A3 for conversion into an "artzyme".

Results and Discussion

Biohybrid synthesis

Structural study of the dimer

At first glance, one may note that in both crystal structures of αRep A3 the dimer presents a large cavity with a wide aperture. The bottom of the cavity is closed by a wall of hydrophobic residues, forming a so-called "aromatic zipper" (purple residues, Figure 1 A, B). The "PEG-bound" form is slightly asymmetrical and the "empty" form is in a strictly C_2 symmetrical geometry, in which the hydrophobic zipper at the rear side of the protein is discontinuous, as shown by the hole visible in Figure 1 A. The cavity dimensions are roughly width 23 Å, depth 15 Å, and height 13 Å, and by approximating the cavity to a truncated rectangular-based pyramid, we calculated the volume to be about 3600 Å^3.

In order to choose positions within the dimer suitable for mutagenesis, we identified, in silico, which amino acids were responsible for dimerization (Figure 1 C, framed amino acids) and thus unavailable for mutagenesis and ligand grafting. One can see that the interacting residues at the interface between subunits were consistently at variable positions in the randomization scheme of the αRep library (Figure 1 C, green). The large hydrophobic zipper composed of residues of variable position 19 of each repeat is thought to be principally responsible for the interaction energy.[24] Residue Trp111 (underlined in Figure 1 C) only interacts in the open form and is responsible for the perforation in the zipper in the closed form. The residue Arg22 participates in an intermolecular H-bond with either residue Gln146 in the open form or residue Asp150 in the closed form.

Choice of the positions for mutation into cysteine residues

The positions most suitable for mutations were chosen according to these structural analyses, taking into account the fact that a single mutation into cysteine in a monomer will result in the presence of two thiol groups in the dimer and thus the coupling of two copper(II)-phenanthroline groups at symmetric positions in the crevice.

As amino acids in variable positions of each motif are non-essential for the folding properties of αRep proteins, mutation of any of the 30 variable residues for a monomer into a cysteine residue is a priori possible. Among the variable positions, 18 and 19 were disregarded for each repeat because of their implication in the dimerization. Similarly, residues at position 22 were also disregarded because they were either involved in intermolecular hydrogen bonding or too close to the hydrophobic zipper (Figure 1 C). Position 23 for each repeat was overlooked because it could only preferentially accommodate small residues, as reported in the design of the αRep library.[24,25] Residues at position 30 were considered to be too exposed to solvent molecules. Among the remaining residues, Tyr26 (position 26 in the N-cap) and Phe119 (position 26 in repeat 3) appeared to be good candidates as they are located quite deep within the inter-subunit crevice. Furthermore, the distance between the two β-carbon atoms of symmetrical Tyr26 residues (27.3 Å in the PEG-bound form and 28.9 Å in the empty form) is longer than that between the two β-carbon atoms of symmetrical Phe119 residues (21.3 Å and 18.3 Å, respectively). Models of two biohybrids based on these positions, αRep A3F119C-phenanthroline (A3-F119NPH) and αRep A3Y26C-phenanthroline (A3-Y26NPH), suggested that the dimer could accommodate two phenanthroline ligands in the crevice (Figure 2 B). In order to investigate the suitability of the

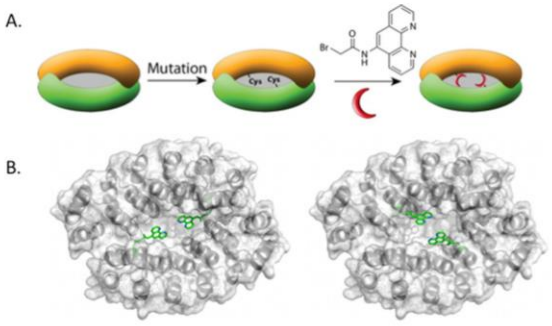


Figure 2. αRep A3-derived biohybrids obtained after functionalization with phenanthroline. A) Biohybrid design approach: In a first step, a single point mutation into cysteine introduces two thiol groups in the crevice; in a second step, phenanthroline (red crescent) is covalently attached to each thiol group of the A3 dimer. B) Models obtained for the biohybrids after covalent attachment of phenanthroline on the variants A3Y26C (left) and A3F119C (right). The models based on the crystallographic structure of the open form were obtained using PyMOL software.

crevice for providing the chiral environment needed for enantioselective catalysis of the Diels–Alder reaction by the corresponding copper complexes, the biohybrids were obtained as follows: either position Phe119 or Tyr26 was first independently mutated into cysteine, and then chemical coupling of a phenanthroline moiety to each purified protein variant followed by copper(II) coordination led to the copper(II)-biohybrid complexes (Figure 2 A).

Production of the protein variants

Genes encoding αRep proteins are derived from the concatenation of partially randomized DNA fragments for each repeat. These DNA sequences therefore have internal repetitions in their sequences that can interfere with PCR amplification. In order to minimize this drawback and to simplify the introduction of directed mutations, we designed a synthetic gene coding for αRep A3 protein, namely *A3s*, which denotes "A3 synthetic". This DNA sequence encoded the αRep A3 protein

© 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

ChemPubSoc Europe

CHEMISTRY A European Journal Full Paper

using different codons at conserved positions in each repeat. The *A3s* gene was cloned into an expression vector derived from the commercial pQE80 vector, allowing its expression with the hexa-histidine sequence (HisTag) at the N-terminal position, and then subjected to the Tobacco Etch Virus cleavage sequence. Two independently mutated genes could then be easily obtained by site-directed mutagenesis, that is, *A3sF119C* and *A3sY26C* codings for proteins A3F119C and A3Y26C, respectively. After production in *E. coli*, purification, and cleavage of the HisTag, pure proteins were obtained with a typical yield of 70 mg purified protein per liter of culture.

The reactivity and accessibility of the cysteine residues were assessed through a time-resolved Ellman's assay using the sulfhydryl reagent 5,5-dithio-bis(2-nitrobenzoic acid) (DTNB). The DTNB reacts with sulfhydryl groups, releasing 2-nitro-5-thiobenzoate (TNB^-), which displays a maximum absorption at 412 nm. Firstly, for each of the two variants, the labeled cysteine concentration at the end of the reaction was equal to the monomer concentration. This indicated that all cysteine residues were reactive and did not form disulfide bonds, even after prolonged storage without any reducing agent. Secondly, the kinetics of the reaction was different for the two variants: A3Y26C showed very fast labeling kinetics and led immediately to a maximum stable absorbance at 412 nm, whereas A3F119C reacted more slowly with DTNB and the absorbance at 412 nm reached a plateau value only after 10 min. This denoted either a sterically hindered environment or a shift in the pK_a of the cysteine residue at position 119. Thus, the thiol function appeared to be more accessible or reactive in A3Y26C than in A3F119C.

Protein functionalization

The two A3Y26C and A3F119C variants were then functionalized by incubation with 2-bromo-*N*-(1,10-phenanthrolin-5-yl)-acetamide hydrobromide (1.2 equiv), overnight, at 4 °C, resulting in the cysteine-methylene-carbamoyl-1,10-phenanthroline (NPH) biohybrids in their apo-forms: A3Y26NPH and A3F119NPH. Complete functionalization was confirmed by Ellman's assay, which indicated a complete loss of thiol reactivity (data not shown). MALDI-ToF mass spectrometric (MS) analysis showed peaks at 22490 and 22505 Da for A3Y26NPH and A3F119NPH, respectively (Figure 3A), consistent with the covalent anchoring of one phenanthroline moiety per monomer (calculated masses 22489 Da and 22505 Da, respectively). When an analogous functionalization experiment was performed on A3 protein bearing no cysteine, the MALDI-ToF MS analysis showed only a peak at 22301 Da, corresponding to the non-coupled protein, which indicated that the coupling reaction only occurred on the cysteine residues in the variants and that no other reactive group was modified (Figure 3A).

Biohybrid characterization

Protein folding in the biohybrids

Circular dichroism (CD) experiments were carried out to analyze the effect of chemical modifications on the secondary structure content of the protein (Figure 3B). For the two *apo*-biohybrids, the far-UV-CD spectra showed a signal characteristic of all-α protein structures, with two negative bands at 222 and 208 nm and a positive band at around 190 nm.[26] This CD signal is almost the same as that obtained for αRep A3 at the same concentration. This experiment indicated that the content of α-helix secondary structures was conserved in the biohybrids, supporting the hypothesis that the proteins remained folded upon cysteine mutation followed by NPH functionalization. The addition of metal ions to the solution did not affect the CD spectra.

Protein dimer characterization in the biohybrids

In order to ensure that phenanthroline grafting on the protein A3 did not disrupt its dimeric structure in solution, the oligomeric states of functionalized proteins were investigated by analytical gel filtration (Figure 3C). The elution volumes observed for the biohybrids (9.58 mL for A3F119NPH; 9.87 mL for A3Y26NPH) were very close to the elution volume of unmodified A3 (9.91 mL). The apparent molecular weights calculated from a calibration curve obtained for globular proteins (42 kDa, 38 kDa, and 37 kDa for A3F119NPH, A3Y26NPH, and A3, respectively) were consistent with dimeric forms in solution. This result showed that the dimeric form was retained for both biohybrids. The slight shift towards smaller elution volumes suggested that the conversion of free proteins (A3Y26C and mainly A3F119C) into biohybrids (A3Y26NPH and A3F119NPH) was accompanied by a small increase in the hydrodynamic radius. For more precise insight into the effect of the NPH residue on the dimerization, analytical ultracentrifugation analyses were performed on both biohybrids (Figure S3). The sedimentation distribution profiles of the two functionalized proteins were very similar to that of A3 at a concentration of 2 μM (total protein concentration), indicating similar hydrodynamic properties. This result suggested that the proteins remained dimeric at micromolar concentrations.

The effect of copper(II) on the biohybrids was also assessed through analytical SEC experiments. One molar equivalent of copper(II) per monomer was added prior to injection, and no change in the elution profile was observed (Figure 3C, red and blue curves). Taken together, these results indicated that the presence of copper(II) did not disrupt the quaternary structure of the biohybrids.

Metal content in the biohybrids

Metallation of the biohybrids was performed by overnight incubation with one molar equivalent of copper(II) nitrate with respect to phenanthroline (200 μM copper(II) with 200 μM protein monomer), followed by gel filtration to remove unbound or weakly bound copper(II) ions.

To assess whether the metal ion was present only on the phenanthroline ligand (NPH), samples were submitted to a gel filtration step on a Superdex™ S75 column, and then EPR experiments were performed on frozen solutions of unmodified protein (A3) and on the biohybrids (A3F119NPH and

© 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

ChemPubSoc Europe
CHEMISTRY A European Journal Full Paper

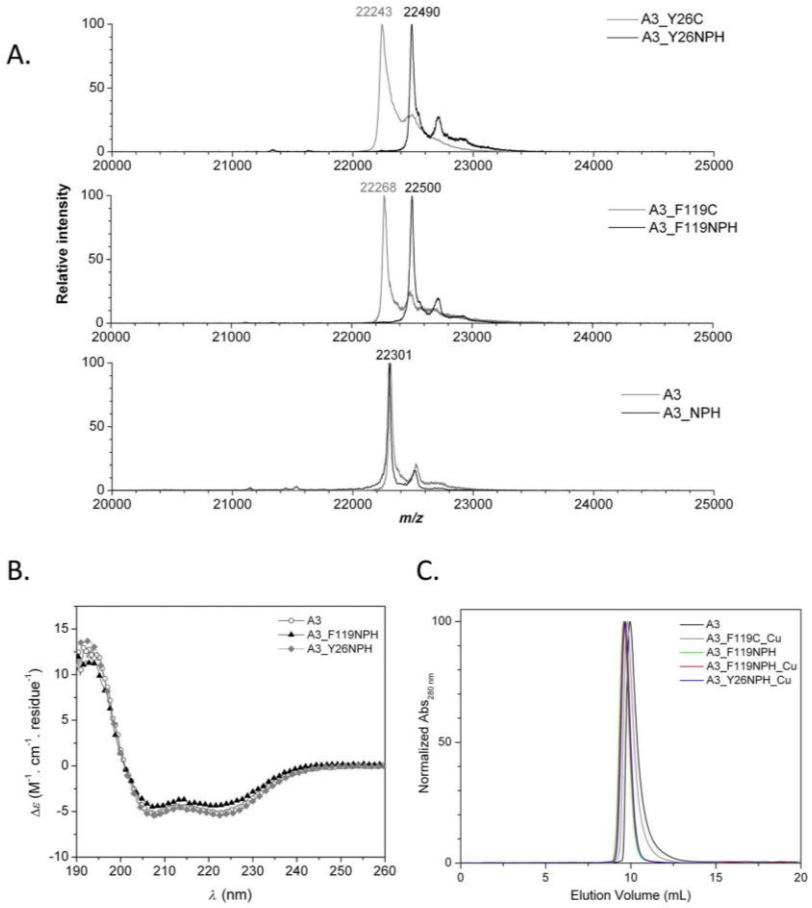


Figure 3. Biophysical characterization of the biohybrids. A) MALDI-ToF spectra of the protein variants measured before (grey) and after (black) functionalization with compound **1**. Spectra corresponding to the proteins A3, A3F119c, and A3Y41C are shown from bottom to top. B) Far-UV circular dichroism spectra of the functionalized proteins A3F119NPH (filled triangles) and A3Y26NPH (grey diamonds) at 5 μM in 50 mM phosphate buffer (pH 7). Unmodified A3 was used as a control (open circles). C) Size-exclusion profiles of the biohybrids. Solutions containing 100 μL of 100 μM sample proteins or standard globular proteins used as a control were injected on a Superdex™ 75 10/300 GL column equilibrated in 20 mM MOPS; 150 mM NaCl, pH 7, and elution profiles were recorded by monitoring the absorbance at 280 nm. For metallated samples, an excess of copper(II) per protein monomer was pre-incubated before injection on the column. Normalized elution profiles and elution volumes (V_e) are shown for A3 (black, V_e = 9.95 mL), A3F119C-Cu (grey, V_e = 9.79 mL), A3F119NPH (green, 9.58 mL), A3F119NPH-Cu (red, V_e = 9.62 mL), and A3Y26NPH-Cu (blue, V_e = 9.67 mL). Protein standards were injected for comparison of their elution volumes: ovalbumin (43 kDa, V_e = 8.47 mL), carbonic anhydrase (29 kDa, V_e = 11.01 mL), and cytochrome *c* (12.4 kDa, V_e = 12.63 mL).

A3Y26NPH) in MOPS buffer (20 mM, pH 7; 150 mM NaCl) at 80 K. Under these conditions, no signal was observed with the A3 protein (Figure 4, bottom). This result indicated that even if copper(II) ions were capable of binding with the A3 protein, they were bound at low-affinity, non-specific binding sites and were fully removed after gel filtration. Such weak copper binding sites have previously been described in proteins.[27] The biohybrids showed different behavior, as in both cases the copper(II) EPR signal remained after the samples had been submitted to gel filtration, proving that the copper(II) was indeed bound by the phenanthroline moiety (Figure 4, top and middle). In both cases, an anisotropic EPR signal with axial symmetry was observed, that could be fitted (Figure S2). The signal obtained for A3F119NPH could be fitted with *g* values ($g_\perp = 2.06$, $g_{//} = 2.25$), hyperfine coupling constants of $A_\perp = 10^{-3}\,cm^{-1}$ and $A_{//} = 1.7 \times 10^{-2}\,cm^{-1}$, and line widths of $W_\perp = 37$ G and $W_{//} = 44$ G. The signal observed for A3Y26NPH could also be fitted with similar *g* values ($g_\perp = 2.05$, $g_{//} = 2.26$), hyperfine coupling constants $A_\perp = 10^{-3}\,cm^{-1}$ and $A_{//} = 1.7 \times 10^{-2}\,cm^{-1}$, and line widths $W_\perp = 47$ G and $W_{//} = 30$ G. Peisach–Blumberg analysis[28] of the parameters corresponding to the parallel region ($g_{//}$ and $A_{//}$) did not allow unequivocal determination of the coordination sphere of the copper center. Nonetheless, integration of the EPR spectra of both holo biohybrids showed that the copper(II) concentration in A3F119NPH was half of that in A3Y26NPH, which suggested that in A3F119NPH only one copper(II) ion was actually complexed by the two phenanthroline residues of the dimer, whereas in A3Y26NPH

© 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

ChemPubSoc Europe
CHEMISTRY A European Journal Full Paper

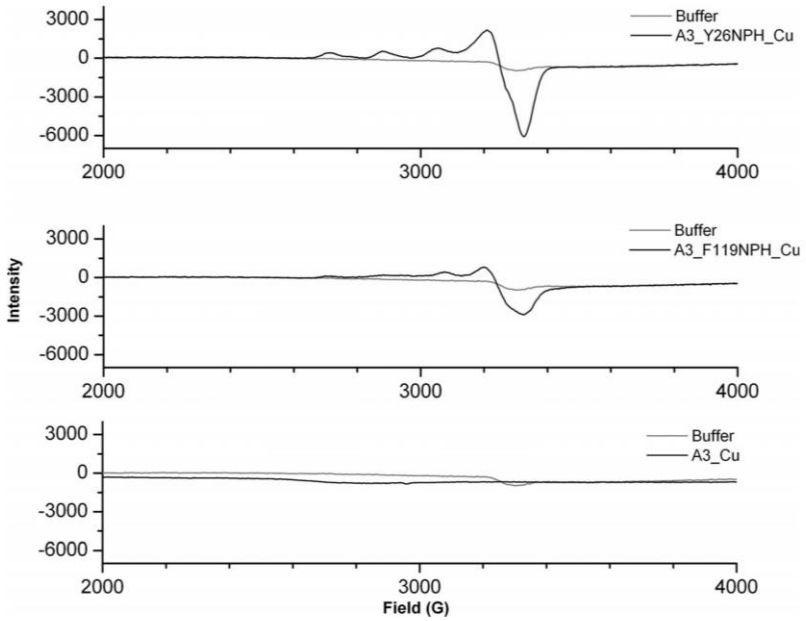
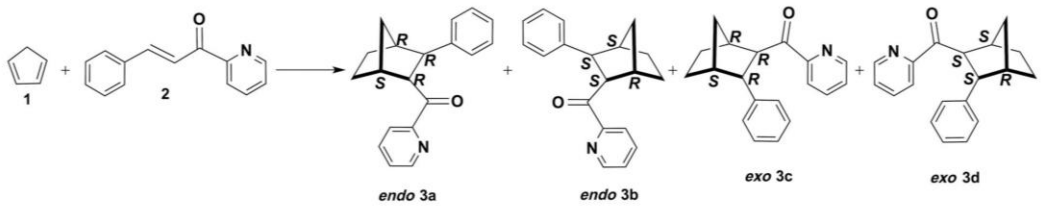


Figure 4. EPR spectra of biohybrids measured after metal incorporation. EPR spectra were recorded using 200 μM of protein solutions prepared in 20 mM MOPS pH 7, 150 mM NaCl. Samples were prepared by incubation 200 μM of functionalized proteins with 1 equiv $Cu(NO_3)_2$ per protein monomer followed by a gel filtration on a Superdex™ 75 column in 20 mM MOPS pH 7, 150 mM NaCl and concentration by using a Vivaspin™ concentrator. Spectra corresponding to the proteins A3, A3F119NPH, and A3Y41NPH, respectively, are shown from bottom to top. The spectrum of the buffer containing 200 μM of $Cu(NO_3)_2$ is shown in grey.

one copper(II) ion was complexed by each phenanthroline residue. This assumption was confirmed through ICP-MS measurements, whereby a 1:2 copper(II) to monomer ratio was observed in the case of A3F119NPH. This unexpected result could be explained by the closer proximity of the two mutation points of each monomer in A3F119C, as shown in Figure 2B, combined with a plasticity of the dimer. In comparison, the EPR signal of a Cu^{II}-phenanthroline complex recorded under the same conditions, but in the absence of protein, was characterized by a signal that could be fitted with g values of $g_{\perp} = 2.06$ and $g_{//} = 2.30$, hyperfine coupling constants of $A_{\perp} = 10^{-3}\,cm^{-1}$ and $A_{//} = 1.4 \times 10^{-2}\,cm^{-1}$, and line widths of $W_{\perp} = 38$ G and $W_{//} = 40$ G. This signal is clearly different from those observed for both A3F119PNH- and A3Y26NPH-Cu complexes and is in agreement with an N_2O_2 coordination sphere, which is consistent with the binding of one phenanthroline ligand and two water molecules to the Cu^{II} ion. Thus, the change in the signal observed for the Cu^{II} center in A3F119PNH- and A3Y26NPH-Cu is in line with its insertion in the hydrophobic pocket of the protein dimer and possibly the release, or at least a weakening of the binding, of its water ligands.

Catalysis of a Diels–Alder reaction

The Diels–Alder (D-A) cycloaddition is a very useful carbon-carbon bond-forming reaction for the synthesis of organic building blocks. Few enzymes are known to catalyze this reaction; therefore, artificial enzymes with this activity have high potential for synthetic applications.[29] As a consequence, in the last decade, many studies have been dedicated to the development of artificial enzymes capable of catalyzing D-A reactions by using as model starting materials 2-azachalcone derivatives that are water-soluble and possess a wide range of biological activities that depend on their substitution patterns. We thus focused on the Diels–Alder cyclization reaction between cyclopentadiene (**1**) and 2-azachalcone (**2**) (Scheme 1) cata-



Scheme 1. Diels–Alder reaction.

© 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

ChemPubSoc Europe

CHEMISTRY A European Journal Full Paper

lyzed by either the unmodified αRep A3 protein or by the biohybrids obtained after phenanthroline functionalization of the A3F119C and A3Y26C variants (A3F119PNH and A3Y26NPH), in the presence of copper(II). It is noteworthy that this reaction can result in up to four products: two major *endo* enantiomers (**3a** and **3b**) and two minor *exo* enantiomers (**3c** and **3d**, Scheme 1). Secondary orbital effects favor formation of the *endo* isomers; however, because of their protein components, artificial Diels–Alderases are able to shift the *endo*/*exo* ratio as well as the enantiomeric balance for each isomer.[30]

Here, the A3F119NPH biohybrid was used to explore the influence of various physicochemical parameters on the catalysis. Cyclopentadiene (**1**; 34 mM, 1133 equiv) and 2-azachalcone (**2**; 1 mM, 33 equiv) were allowed to react in the presence of 30 μM protein monomer (A3F119NPH, 1 equiv) and 15 μM $Cu(NO_3)_2$ (0.5 equiv). After 48 h, the products were extracted with ethyl acetate and the reaction yields and enantiomeric excesses (with respect to *endo* products (**3**)) were determined by chiral HPLC as described in the Materials and Methods section. The reaction was first performed in the presence of three kinds of buffers: 20 mM MOPS, 20 mM HEPES, and 50 mM phosphate buffer (pH 7.0), respectively, each containing 150 mM NaCl. It appeared that, irrespective of the buffer used, no significant changes in the yields (7–12%) or enantiomeric excesses (about 50%) occurred, nor in the proportions of *exo*/*endo* products (around 1:9). The effect of a co-solvent to aid solubilization of the reactant **1** was then examined. Three co-solvents, in 5 and 10% (v/v) proportions, were used: acetonitrile, isopropanol, and *tert*-butanol. The results indicated that, irrespective of the nature and the proportion of co-solvent, no significant changes in the yield of **3**, and in the enantiomeric excess (around 50%), were induced. Finally, the reaction was performed at three different temperatures in MOPS buffer: 4, 15, and 25 °C. It was found that when the temperature was increased from 4 to 25 °C, the yield decreased from 12 to 3%, while the enantiomeric excess remained significant. This effect could be attributed to the thermal dimerization of cyclopentadiene, as described previously.[31]

Considering the aforementioned results, it was decided to perform the D-A reaction under the following optimal conditions: 34 mM **1** and 1 mM **2** were allowed to react at 4 °C in 20 mM MOPS buffer (pH 7) containing 150 mM NaCl, and no co-solvent was added other than the substrate solution (1% acetonitrile), in the presence of 30 μM protein monomer A3, A3F119NPH, or A3Y26NPH and 0–1 equiv of copper(II) nitrate as catalyst (Table 1).

Under these conditions, in the absence copper(II) nitrate, very low yields were obtained, *endo* products **3** being obtained in about 3% yield, even in the presence of A3 or of the A3F119NPH and A3Y26NPH biohybrids (Table 1). It is noteworthy, however, that a slight enantiomeric excess was observed in the presence of both A3F119NPH and A3Y26NPH biohybrids (10% and 11%, respectively). This may have been due to a template effect of the protein that could position the two reactants, **1** and **2**, in an appropriate mutual arrangement for formation of the *endo* products **3**. Copper(II) nitrate alone was found to be capable of catalyzing the D-A reaction, with 13%

Table 1. Comparison of *ee*, yields, and *endo*/*exo* ratios measured after the D-A reaction of cyclopentadiene **1** with 2-azachalcone **2** in MOPS buffer catalyzed by A3, A3Y41NPH, or A3F119NPH biohybrids in the presence of copper(II) nitrate. In all experiments, 30 μM ligand (monomer protein or phenanthroline) was used (1 equiv) and incubated with different copper(II) concentrations.

Ligand	Catalyst $[Cu^{II}]$/[protein monomer] ratio	Yield[a] [%]	*ee* [%]	*endo*/*exo* [%:%]
none	0	3	3±1	N/A[b]
	1:0	13±7	0±2	92:8
	2:0	21±6	1±1	92:8
phenanthroline	1:2	5±3	2±1	90:10
	2:2	6±1	1±1	90:10
A3Y26NPH	0	3±2	9.5±3	N/A[b]
	2:2	16±2	7.5±1	92:8
A3F119NPH	0	4±1	11±4	N/A[b]
	1:2	15±1	40±5	86:14
	2:2	26±11	3±1	92:8
	excess Cu^{II} + GF[c]	**11±1**	**62±3**	**86:14**
A3	excess Cu^{II} + GF[c]	3±1	3±2	N/A[b]

[a] Yield represents the recovered product and was calculated with a standard curve. [b] N/A: not available. [c] Reaction catalyzed after gel filtration of the samples on a Superdex™ 75 column.

and 21% yields of products **3** being obtained with 0.5 equiv (15 μM) and 1 equiv (30 μM) of catalyst, respectively, but almost no enantiomeric excess was observed. In the presence of phenanthroline, copper(II) also catalyzed the D-A reaction, albeit in lower yields (6%) than copper(II) nitrate alone. Moreover, a similar yield was obtained for 1:1 and 1:2 copper(II)/phenanthroline ratios, which suggested the formation of a $Cu^{II}(Phen)_2$ complex as an active catalytic species. Indeed, the structure of such a complex, obtained after reaction of copper(II) nitrate with phenanthroline, has been previously characterized by X-ray diffraction analysis.[32] Here again, no significant enantiomeric excess could be observed. On the contrary, reactions catalyzed by either A3F119NPH or A3Y26NPH in the presence of copper(II) were enantioselective. In all cases described below, as *exo* products were negligible with respect to *endo* products, *ee* values were calculated for the *endo* products. Using A3Y26NPH in the presence of 1 equiv of copper(II) per monomer as catalyst, a higher yield of product **3** (16%) than with A3Y26NPH alone (3%) was obtained, but the enantiomeric excess was about the same in both cases (8–10% *ee*). This suggested that the reaction occurred in an isotropic environment with moderate control of the relative positioning of both complex and substrates, leading to a low *ee*. When A3F119NPH was used as catalyst, in the presence of 0.5 equiv copper(II) per monomer, the yield of product **3** reached 15% and the enantiomeric excess increased to 40%. Addition of a further 0.5 equiv of copper(II) per monomer of A3F119NPH led to an increase in the yield of **3**; it reached a value of 26%, even higher than that obtained with copper(II) nitrate alone, but the enantiomeric excess decreased to 3%. A3F119NPH was then incubated with an excess of copper(II) per monomer, submitted to a gel filtration step on a Superdex™ S75 column, and used for catalysis. The obtained yield and enantiomeric excess were similar to those achieved in the presence of

© 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

ChemPubSoc Europe

CHEMISTRY A European Journal Full Paper

0.5 equiv of copper(II) per monomer of A3F119NPH (11% yield and 62% *ee*). This suggested that A3F119NPH was able to bind 0.5 equiv of copper(II) per monomer (i.e., 1 equiv per dimer), and that in the presence of an excess of copper(II) the reaction was catalyzed by the copper(II) ions that remained outside of the designed binding site, leading to a loss of enantioselectivity. This result was consistent with the aforementioned EPR and ICP-MS results, which showed a 1:2 copper(II)/A3F119NPH stoichiometry. Taken together, these results suggest the formation of a $Cu^{II}(Phen)_2$ complex in the biohybrid, which seems to be responsible for the enantioselective catalysis. The closer proximity of the two mutation points in A3F119C could favor the formation of this complex through coordination of one copper(II) ion by the two phenanthroline moieties of A3F119NPH. The dual covalent attachment of such a complex could increase its stability inside the crevice of the protein, leading to an anisotropic environment for the substrates. The effect of a dual anchoring on the efficiency of the enantioselectivity has been described previously.[37]

The characteristics of the reaction catalyzed by copper(II)-A3F119NPH, that is, yield, *endo/exo* ratio, and enantiomeric excess, are compared in Table 2 with those reported to date for the D-A reaction between **1** and **2** catalyzed by various artificial metalloenzymes. As regards the yield of the reaction, copper(II)-A3F119NPH furnishes a lower yield of product **3** compared to those reported for other copper(II)-biohybrid catalysts, which range from 15% in the case of some FhuA-Tpy-copper(II) complexes[36] to 98% in the case of an SB-Lys-GTL*/196-Phen-copper(II) complex.[33]

Table 2. Comparison of yields and *ee* measured after the D-A reaction of cyclopentadiene **1** with 2-azachalcone **2** in the presence of various copper(II)-biohybrid catalysts reported in the literature.

Biohybrid catalyst/ protein–ligand[a]	Conversion/ yield [%]	*endo/exo* [%]	*ee* [%]	Ref.
LmrR_M89C-Phen[b]	93	95:5	97	[22]
SB-Lys-GTL*/196-Phen[b]	98	93:7	93.5	[33]
HSA-phthalocyanine	89	91:9	85	[34]
A3F119C-Phen[b] (A3F119NPH)	11	86:14	62	this work
NCS-Phen[b]	84	84:16	0	[21]
NB-Tpy[c]	22	90:10	90	[35]
FhuA-Tpy[c]	15	66:34	ND[e]	[36]
FhuA-NHC[d]	62	90:10	ND[e]	

[a] Nomenclature for the various proteins used: LmrR, *Lactococcal* multidrug resistance Regulator; SB-Lys-GTL*/196, Sepabeads Lys196 mutant of *Geobacillus thermocatenulatus* lipase; HSA, human serum albumin; A3, αRep A3 protein; SCP-2L, sterol carrier protein type 2-like domain; NCS, neocazinostatin; NB, nitrobindin; FhuA, ferric hydroxamate uptake protein component A. [b] Phen, phenanthroline. [c] Tpy, terpyridine. [d] NHC, N-heterocyclic carbene. [e] Not determined.

This result could be an indication of either restricted access of the substrates to the active site (due to moderate hydrophobicity of the crevice), or inhibition of the reaction by either the substrate or the product. Finally, the chiral environment provided by the *apo*-A3F119NPH around the copper(II) ion induced good enantioselectivity in the formation of the *endo* product **3**, characterized by an *ee* of 62%. This is lower than those reported in the case of copper(II) complexes of LmrR_M89C-Phen (97%),[22] SB-Lys-GTL*/196-Phen (93.5%),[33] HSA-phthalocyanine (85%),[34] and NB-Tpy (90%),[35] but higher than those reported for copper(II) complexes of SCP-2LV83C-Phen (25%)[38] and NCS-Phen (no *ee*),[21] no enantiomeric excesses being reported for the copper(II) complexes of FhuA-Tpy and FhuA-NHC biohybrids.[36]

Conclusions

Covalent coupling of copper(II)-phenanthroline to the protein dimer allowed conversion of the variant αRep A3F119C into an enantioselective Diels–Alderase that catalyzed reactions with up to 62% *ee*. This result suggests the presence of a copper(II) bis-phenanthroline as a catalytic species, buried in the crevice formed by the protein dimer. However, it would be possible for access of both substrates to the catalytic center to be very constrained, in particular in the case of A3F119NPH, in which copper(II) is coordinated by the two phenanthroline ligands in a face-to-face arrangement, which could account for the low yields obtained. A possible improvement of the catalytic properties of our biohybrid catalysts might then be obtained by combining, in a single chain protein, an A3F119NPH domain with a non-mutated A3 domain. This would both ensure a 1:1 copper(II)-phenanthroline complex as the catalytic center and enlarge the catalytic pocket in order to accommodate both D-A substrates.

Taken together, these encouraging results indicate that the selected αRep is a promising protein scaffold to accommodate other organometallic complexes. The association of αRep proteins with different metal ions and organic ligands will allow further exploration of the catalysis of a variety of chemical reactions under mild aqueous conditions.

To the best of our knowledge, this study presents the first instance of the conversion of an artificial protein into an entirely artificial metalloenzyme. It opens the way to a wide range of exploration routes to develop a new class of αRep-based biohybrid catalysts. The intrinsic modular nature of αReps paves the way to directed-evolution approaches based on αRep A3 dimer to enhance both catalytic efficiency and substrate specificity. Furthermore, the protein scaffold used in this study is a single variant among a highly diverse library containing stable and folded proteins of different sizes and sequences. Further exploration of αRep variants will thus be a means of exploring new tailored scaffolds that could each be engineered into a class of biohybrids.

Experimental Section

Plasmid construction

The genetic sequence coding for the protein αRep A3 is issued from a combinatorial library of concatenated, repeated, and degenerate sequences. In order to simplify molecular biology processes, the sequence was recoded using unrepeated codons by means of the GenScript® algorithm dedicated to codon optimization for

© 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

ChemPubSoc Europe

CHEMISTRY A European Journal Full Paper

E. coli expression. Additional manual modifications were used to increase the differences between repeats, taking into account the Codon Adaptation Index using the JCat tool (http://www.jcat.de/). The gene was named A3s, synthesized by GenScript® and cloned into PQE80N, which is the pQE80L (Qiagen®) expression vector modified to include a Tobacco Etch Virus (TEV) protease cleaving site after the N-terminal hexa-histidine tag.

Directed mutagenesis

Gene codings for the variants A3sF119C and A3sY26C were obtained by site-directed mutagenesis using the QuikChange® protocol with the following primers: F119C-FOR: 5′-CGGCGGCGtgcGCATTAGG; F119C-REV: 5′-CCTAATGCgcaCGCCGCCG; Y41C-FOR: 5′-GCAGCGGCAtgcGCGCTGGG, and Y41C-REV: 5′-CCCAGCGCgcaTGCCGCTG. The vector containing the A3s gene was amplified by PCR with Phusion high-fidelity DNA polymerase (Thermo Scientific™) as follows: initial denaturation at 98 °C for 1 min; 18 cycles including a denaturation step at 98 °C for 20 s, an annealing step at 68 °C for 15 s, and an extension step at 72 °C for 3 min. The final extension step was performed at 72 °C for 10 min. The PCR product was treated with Dpn I and directly used for transformation in *E. coli* XL1-Blue electrocompetent cells. For each mutagenesis, plasmids from five independent clones were purified using a Nucleospin® plasmid extraction kit (Macherey–Nagel). The sequences of the variants were confirmed by sequencing both DNA strands.

Protein production

Chemocompetent *E. coli* M15 (Qiagen) bacteria were transformed with the desired plasmid and grown on 2YT agar plates supplemented with Ampicillin (100 $\mu g\,mL^{-1}$), Kanamycin (50 $\mu g\,mL^{-1}$), and 1% glucose. One clone was selected and grown overnight at 37 °C in liquid 2YT medium supplemented with Ampicillin (100 $\mu g\,mL^{-1}$) and Kanamycin (50 $\mu g\,mL^{-1}$). The overnight pre-culture was used to inoculate 1 L of fresh medium distributed between two 2 L flasks. Cells were grown at 37 °C until $OD_{600\,nm}$ reached 0.8, protein expression was induced by the addition of 500 μM IPTG, and cultures were further incubated at 37 °C for 4 h. Cells were harvested by centrifugation for 20 min at 5000 *g* at 4 °C and resuspended in washing buffer (40 mL; 50 mM sodium phosphate, pH 8; 300 mM NaCl; 20 mM imidazole). Suspended cells were stored at −80 °C.

Protein purification

Frozen cells from 1 L of bacterial culture were thawed and a tablet of EDTA-free antiprotease cocktail (Roche) was added. Cells were lysed by three 30 s cycles of sonication on ice and centrifuged for 30 min at 11 000 *g* at 4 °C. His-tagged proteins were purified from the crude supernatant by means of nickel-affinity chromatography on a 4 mL Ni-NTA column (Macherey–Nagel). The column was washed with washing buffer (150 mL; 50 mM sodium phosphate, pH 8; 300 mM NaCl; 20 mM imidazole) and proteins were eluted in six 1 mL fractions in the elution buffer (50 mM sodium phosphate, pH 8; 300 mM NaCl; 300 mM imidazole). A typical yield of about 100 mg of protein per liter of culture was obtained. Eluted fractions were analyzed by SDS-PAGE and pure fractions of proteins were pooled, diluted with buffer (50 mM sodium phosphate, pH 8; 150 mM NaCl; 50 mM imidazole), and finally cleaved with His-tagged TEV protease as described for the pRK793 vector (Addgene): His-tagged TEV protease and purified αRep protein with its cleavable His tag were mixed in a 1:100 ratio and incubated, overnight, at 10 °C in the presence of 0.5 mM EDTA and 1 mM DTT. The mixture was then diluted with buffer (50 mM sodium phosphate, pH 8; 150 mM NaCl) leading to a final imidazole concentration of 30 mM, and loaded onto another 4 mL Ni-NTA column. His-tagged TEV protease and uncleaved His-tagged αRep were retained on the column, while the eluate containing cleaved αRep protein was recovered and concentrated, resulting in a 5 mL solution that was loaded on a gel filtration column (HiLoad 16/600 Superdex™ 75, GE Healthcare) equilibrated with filtered coupling buffer (50 mM sodium phosphate, pH 7.75; 150 mM NaCl). The coupling buffer had previously been incubated for 1 h with Chelex® 100 resin (Bio-Rad) to remove any traces of divalent cations. Elution fractions after gel filtration were pooled and stored at 4 °C prior to chemical coupling. Purity was checked by SDS-PAGE and concentrations were calculated from the absorbance at 280 nm measured on a Nanodrop spectrophotometer (Thermo Scientific) using calculated extinction coefficients at 280 nm ($34\,950\,M^{-1}cm^{-1}$ for A3s and A3sF119C; $33\,460\,M^{-1}cm^{-1}$ for A3sY26C). After the cleavage step, a typical yield of 70 mg of pure protein ready for the coupling reaction was obtained.

Monitoring of accessibility by thiols

Accessibility of the thiol groups of Cys119 and Cys26 in the A3s variants was monitored using Ellman's reagent.[39] In a standard assay, 445 μL of reaction buffer (0.1 M sodium phosphate, pH 8.0; 1 mM EDTA) was mixed with 5 μL of reagent solution (10 mM 5,5′-dithio-bis(2-nitrobenzoic acid)) in a quartz cuvette. Absorbance at 412 nm was monitored using a JASCO spectrophotometer. Absorbance was recorded in kinetic mode after the addition of 50 μL of a 100 μM protein solution until complete reaction. The final absorbance obtained at the plateau was noted as A_{eq} and the initial absorbance measured before protein addition was noted as A_0. The reduced cysteine content was then calculated from the difference signal corrected by the dilution factor using $\varepsilon_M^{412\,nm} = 14\,150\,M^{-1}cm^{-1}$.

Synthesis of 2-bromo-*N*-(1,10-phenanthrolin-5-yl)acetamide hydrobromide (4)

Compound **4** was synthesized according to the literature.[22] 1,10-Phenanthroline-5-amine (195 mg, 1 mmol) was dissolved in anhydrous CH_2Cl_2 (30 mL) in a 50 mL two-necked round-bottomed flask equipped with a magnetic stirrer and a condenser. Bromoacetyl bromide (104 μL, 1.2 mmol) was diluted with CH_2Cl_2 (10 mL) and added dropwise to the phenanthroline solution. The mixture was then heated under reflux overnight and the resulting suspension was filtered. The precipitate was recrystallized from MeOH, affording the HBr salt as brownish crystals (yield 70%). The structure of **4** was confirmed by its ^1H NMR spectrum, which was identical to that reported previously.[22]

Phenanthroline coupling

The concentration of thiol groups calculated by Ellman's test was used as the free cysteine concentration available for phenanthroline coupling. Compound **4** (1.2 equiv) was diluted with DMSO and mixed with a 100 μM protein solution in coupling buffer (50 mM sodium phosphate, pH 7.75; 150 mM NaCl) to give a final DMSO concentration of 1% in the desired volume. The mixture was incubated overnight at 4 °C. The coupling reaction was then stopped by the addition of 0.5 mM EDTA and 1 mM DTT. The protein solution was concentrated to about 5 mL using a Vivaspin™ concentrator and loaded on a gel filtration column (HiLoad 16/600 Superdex™ 75, GE Healthcare) equilibrated in buffer A (MOPS 20 mM, pH 7; 150 mM NaCl) that had been previously incubated with

© 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

ChemPubSoc Europe

CHEMISTRY A European Journal Full Paper

Chelex® 100 resin and filtered. The coupling efficiency was measured by means of Ellman's test and MALDI-ToF MS. The molar extinction coefficient was calculated by adding the computed molar extinction coefficient of the protein variants to that measured for the phenanthroline moiety. The latter was determined in a 70% ethanol solution containing a known concentration of 1,10-phenanthroline, giving a coefficient of 22428 $M^{-1}cm^{-1}$ at 280 nm.

Mass spectrometry analysis

The protein solutions were desalted on a Zeba Spin column (Thermo) against 10 mM NH_4HCO_3 buffer (pH 8), diluted in a sinapinic acid matrix, and analyzed by MALDI-ToF MS in linear mode with a MALDI-ToF/ToF UltrafleXtreme spectrometer (Bruker Daltonique).

Circular dichroism measurements

Circular dichroism (CD) spectra were recorded as previously reported[24] from 185 nm to 260 nm using quartz cells with a 1 mm path length, on a JASCO J810 dichrograph. Each spectrum was recorded three times and averaged. The CD signal was corrected by buffer subtraction and converted to delta epsilon. Data were acquired from 5 μM samples in 50 mM sodium phosphate buffer (pH 7.5).

Analytical size-exclusion chromatography

Analytical size-exclusion chromatography was performed essentially as described previously[24] with an ÄKTA Purifier (GE Healthcare) system on a Superdex™ 75 10/300 GL column (flow rate 1 mL min^{-1}) equilibrated in 20 mM MOPS and 150 mM NaCl (pH 7). Solutions containing 100 μL of sample proteins or standard globular proteins used as a control were injected and elution profiles were recorded by monitoring the absorbance at 280 nm.

Analytical ultracentrifugation

Sedimentation velocity experiments were performed at 42 000 rpm (128, 297 *g*) and 20 °C using a Beckman Coulter XL-I analytical ultracentrifuge (AUC) equipped with an Aviv fluorescence detection system (AU-FDS) with an An-50Ti rotor. Sample solution (400 μL) was loaded in a 12 mm charcoal-epon double-sector centerpiece equipped with sapphire windows. Proteins were labeled according to the manufacturer's instructions with a monolith NT-115 labeling kit (Nanotemper Technology). Labeled protein (60 nM) was added to unlabeled protein to reach 2 μM total protein concentration. A3Y and A3F proteins were diluted in 20 mM MOPS, pH 7.0; 150 mM NaCl; 0.1 mg mL^{-1} BSA, whereas A3 wild-type was diluted in 50 mM NaP, pH 7.0; 150 mM NaCl; 0.1 mg mL^{-1} BSA. Fluorescence scans were collected every 5 min; sedimentation velocity data were analyzed using SEDFIT software (https://sedfitsedphat.nibib.nih.gov/) and plotted with GUSSI (http://biophysics.swmed.edu/MBR/software.html). Partial specific volumes were calculated using the SEDNTERP program, and the viscosities and densities of buffers were measured using Anton–Paar Lovis 2000 ME and DMA 4500 instruments.

EPR measurements

The different protein biohybrids were incubated in 20 mM MOPS, pH 7; 150 mM NaCl with 1 equiv of $Cu(NO_3)_2$ (with respect to the phenanthroline concentration) by pipetting concentrated aqueous $Cu(NO_3)_2$ solution (2 μL) into 200 μM protein solutions. After incubation overnight, the proteins were desalted by size-exclusion chromatography on a gel filtration column (Superdex™ 75 10/300 GL, GE Healthcare) in 20 mM MOPS, pH 7; 150 mM NaCl and concentrated using a Vivaspin™ concentrator (cut-off 5000 Da satorius, GE Healthcare). The resulting 200 μM protein solution (calculated in terms of monomer concentration) was transferred to EPR tubes. X-band EPR spectra were recorded on a Bruker ELEXSYS 500 spectrometer equipped with a Bruker ER4119HS X-band resonator, an Oxford Instruments continuous-flow ESR 900 cryostat, and a temperature control system. The EPR tubes were subjected to five cycles of vacuum/helium in order to purge oxygen from the samples, and immersed into liquid nitrogen for cooling to 80 K before being introduced into the EPR spectrometer. EPR spectra were recorded under the following conditions $T = 100$ K, MW frequency = 9.38 GHz, MW power = 0.80 mW, MA = 6 G, mod. frequency = 100 kHz, gain = 60 dB. Spectra were simulated using the simulation software XSophe. The simulation parameters are given in the text and the simulated spectra are shown in Figure S2.

Catalysis of the Diels–Alder reaction

Metallation of the *apo*-biohybrids: Solutions of protein A3, A3F119NPH, and A3Y26NPH (each at 100 μM) were prepared in 20 mM MOPS buffer, pH 7; 150 mM NaCl. Aliquots (0.5 μL or 1 μL) of a 10 mM stock solution of $Cu(NO_3)_2{\cdot}3H_2O$ in distilled water were added to portions (100 μL) of these solutions, and rapid mixing was aided by pipetting. In this way, 0.5:1 and 1:1 copper(II)/monomer ratios were obtained, respectively. The solutions were kept on ice in a cold room overnight and then the proteins were desalted by size-exclusion chromatography on a gel filtration column (Superdex™ 75 10/300 GL, GE Healthcare) in 20 mM MOPS buffer, pH 7; 150 mM NaCl, using an ÄKTA purifier (GE Healthcare).

General procedure for the catalysis of Diels–Alder cyclization reactions: All solutions were kept on ice, and reactions were performed in duplicate at 4 °C in 2 mL Eppendorf conical tubes on a ThermoMixer (Eppendorf) at 500 rpm. The tubes were successively charged with 344 μL of 20 mM MOPS buffer (pH 7) containing 150 mM sodium chloride (in sufficient amount for 500 μL), 150 μL of the 100 μM solution of copper(II)-biohybrid prepared previously (final concentration 30 μM), 5 μL of a stock solution of 100 mM azachalcone **2** (final concentration 1 mM, 33 equiv), and diphenylsulfone (187 mM in acetonitrile) used as an internal standard (final concentration 1.87 mM), and finally 1.4 μL of pure freshly distilled cyclopentadiene (final concentration 34 mM, 1133 equiv). The tubes were then placed on the ThermoMixer and stirred at 500 rpm for 48 h at 4 °C. Ethyl acetate (3 × 500 μL) was used for extraction. After centrifugation, the organic layers were collected, dried over $MgSO_4$, and analyzed by HPLC.

HPLC analyses were performed at 20 °C on an Agilent Technologies 1260 Infinity apparatus equipped with a Chiralpak OD-H column (250 × 4.6 mm, 5 μm; Diacel) and *n*-hexane/isopropanol (99:1) as the mobile phase at a flow rate of 0.5 mL min^{-1} (diode array detection: 270 nm for **3** and diphenylsulfone, 300 nm for **2**). Retention times: 11.9 and 13.0 min (**3** *exo* isomers), 14.4 and 18.3 min (**3** *endo* isomers), 20.1 min (starting material, **2**), and 37.2 min (internal standard, diphenylsulfone).

Abbreviations

HEAT repeat: subfamily of repeated motifs found in Huntingtin, Elongation factor 3, protein phosphatase 2A, and yeast kinase TOR; NPH: cysteine-methylene-carbamoyl-1,10-phenanthroline; equiv: equivalent; D-A: Diels–Alder; *ee*: enantiomeric excess; CD: circular dichroism; EPR: electron paramagnetic resonance; Phen: phenanthroline.

© 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

ChemPubSoc Europe

CHEMISTRY A European Journal Full Paper

Acknowledgements

This work was supported by the CNRS (Centre National de Recherche Scientifique), by Univ. Paris-Sud, and by the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INBS-05. We also thank the "Initiative Doctorale Interdisciplinaire IDEX" program of University Paris-Saclay for the financial support of T.D.M.

Conflict of interest

The authors declare no conflict of interest.

Keywords: artificial enzymes · Diels–Alder reaction · enantioselective catalysis · green chemistry · protein engineering

[1] A. Neville, J. Iniesta, J. M. Palomo, *Molecules* **2016**, *21*, 1680.
[2] A. Schmid, J. S. Dordick, B. Hauer, A. Kiener, M. Wubbolts, B. Witholt, *Nature* **2001**, *409*, 258–268.
[3] A. J. J. Straathof, S. Panke, A. Schmid, *Curr. Opin. Biotechnol.* **2002**, *13*, 548–556.
[4] J. M. Palomo, M. Filice, *Biotechnol. Adv.* **2015**, *33*, 605–613.
[5] J. C. Lewis, *Curr. Opin. Chem. Biol.* **2015**, *25*, 27–35.
[6] F. Nastri, M. Chino, O. Maglio, A. Bhagi-Damodaran, Y. Lu, A. Lombardi, *Chem. Soc. Rev.* **2016**, *45*, 5020–5054.
[7] P. Dydio, H. M. Key, A. Nazarenko, J. Y.-E. Rha, V. Seyedkazemi, D. S. Clark, J. F. Hartwig, *Science* **2016**, *354*, 102–106.
[8] T. Heinisch, T. R. Ward, *Curr. Opin. Chem. Biol.* **2010**, *14*, 184–199.
[9] K. Okrasa, R. J. Kazlauskas, *Chem. Eur. J.* **2006**, *12*, 1587.
[10] A. Fernández-Gacio, A. Codina, J. Fastrez, O. Riant, P. Soumillion, *ChemBioChem* **2006**, *7*, 1013.
[11] T. Heinisch, M. Pellizzoni, M. Dürrenberger, C. E. Tinberg, V. Köhler, J. Klehr, D. Häussinger, D. Baker, T. R. Ward, *J. Am. Chem. Soc.* **2015**, *137*, 10414–10419.
[12] J.-P. Mahy, J.-D. Maréchal, R. Ricoux, *Chem. Commun.* **2015**, *51*, 2476–2494.
[13] T. Heinisch, T. R. Ward, *Acc. Chem. Res.* **2016**, *49*, 1711–1721.
[14] H. Mallin, M. Hestericová, R. Reuter, T. R. Ward, *Nat. Protoc.* **2016**, *11*, 835–852.
[15] M. Jeschek, S. Panke, T. R. Ward, *Methods Enzymol.* **2016**, *580*, 539–556.
[16] P. Haquette, B. Talbi, L. Barilleau, N. Madern, C. Fosse, M. Salmain, *Org. Biomol. Chem.* **2011**, *9*, 5720–5727.
[17] A. Onoda, K. Fukumoto, M. Arlt, M. Bocola, U. Schwaneberg, T. Hayashi, *Chem. Commun.* **2012**, *48*, 9756–9758.
[18] J. Bos, G. Roelfes, *Curr. Opin. Chem. Biol.* **2014**, *19*, 135–143.
[19] T. K. Hyster, T. R. Ward, *Angew. Chem. Int. Ed.* **2016**, *55*, 7344–7357; *Angew. Chem.* **2016**, *128*, 7468–7482.
[20] M. T. Reetz, *Chem. Rec.* **2012**, *12*, 391–406.
[21] W. Ghattas, L. Cotchico-Alonso, J.-D. Maréchal, A. Urvoas, M. Rousseau, J.-P. Mahy, R. Ricoux, *ChemBioChem* **2016**, *17*, 433–440.
[22] J. Bos, F. Fusetti, A. J. M. Driessen, G. Roelfes, *Angew. Chem. Int. Ed.* **2012**, *51*, 7472–7475; *Angew. Chem.* **2012**, *124*, 7590–7593.
[23] J. Bos, W. R. Browne, A. J. M. Driessen, G. Roelfes, *J. Am. Chem. Soc.* **2015**, *137*, 9796–9799.
[24] A. Urvoas, A. Guellouz, M. Valerio-Lepiniec, M. Graille, D. Durand, D. C. Desravines, H. van Tilbeurgh, M. Desmadril, P. Minard, *J. Mol. Biol.* **2010**, *404*, 307–327.
[25] A. Guellouz, M. Valerio-Lepiniec, A. Urvoas, A. Chevrel, M. Graille, Z. Fourati-Kammoun, M. Desmadril, H. van Tilbeurgh, P. Minard, *PloS One* **2013**, *8*, e71512.
[26] R. W. Woody, *Methods Enzymol.* **1995**, *246*, 34–71.
[27] J. Podtetenieff, A. Taglieber, E. Bill, E. J. Reijerse, M. T. Reetz, *Angew. Chem. Int. Ed.* **2010**, *49*, 5151–5155; *Angew. Chem.* **2010**, *122*, 5277–5281.
[28] W. E. Blumberg, J. Peisach, *Arch. Biochem. Biophys.* **1974**, *162*, 502–512.
[29] E. M. Stocking, R. M. Williams, *Angew. Chem. Int. Ed.* **2003**, *42*, 3078–3115; *Angew. Chem.* **2003**, *115*, 3186–3223.
[30] A. Draksharapu, A. J. Boersma, W. R. Browne, G. Roelfes, *Dalton Trans.* **2015**, *44*, 3656–3663.
[31] J. Krupka, *Pet. Coal* **2010**, *52*, 290–306.
[32] K. Lacková, I. Potočňák, *Acta Crystallogr. Sect. A* **2012**, *68*, m1553–m1554.
[33] M. Filice, O. Romero, J. Gutiérrez-Fernández, B. de Las Rivas, J. A. Hermoso, J. M. Palomo, *Chem. Commun.* **2015**, *51*, 9324–9327.
[34] M. T. Reetz, N. Jiao, *Angew. Chem. Int. Ed.* **2006**, *45*, 2416–2419; *Angew. Chem.* **2006**, *118*, 2476–2479.
[35] T. Himiyama, D. F. Sauer, A. Onoda, T. P. Spaniol, J. Okuda, T. Hayashi, *J. Inorg. Biochem.* **2016**, *158*, 55–61.
[36] H. Osseili, D. F. Sauer, K. Beckerle, M. Arlt, T. Himiyama, T. Polen, A. Onoda, U. Schwaneberg, T. Hayashi, J. Okuda, *Beilstein J. Org. Chem.* **2016**, *12*, 1314–1321.
[37] J. R. Carey, S. K. Ma, T. D. Pfister, D. K. Garner, H. K. Kim, J. A. Abramite, Z. Wang, Z. Guo, Y. Lu, *J. Am. Chem. Soc.* **2004**, *126*, 10812–10813.
[38] P. J. Deuss, G. Popa, A. M. Z. Slawin, W. Laan, P. C. J. Kamer, *ChemCatChem* **2013**, *5*, 1184–1191.
[39] P. Eyer, F. Worek, D. Kiderlen, G. Sinko, A. Stuglin, V. Simeon-Rudolf, E. Reiner, *Anal. Biochem.* **2003**, *312*, 224–227.

Manuscript received: April 5, 2017
Accepted manuscript online: May 22, 2017
Version of record online: July 10, 2017

© 2017 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Titre : Ingénierie de l'architecture protéique artificielle αRep : élaboration de catalyseurs biohybrides par couplage covalent de complexes métalliques

Mots clés : metalloenzyme artificielle, biohybride, catalyse énantiosélective, chimie verte

Résumé : Le développement d'une nouvelle génération de catalyseurs dits biohybrides est basé sur l'association d'un complexe métallique et d'une protéine. D'un côté, le complexe métallique est responsable de l'activité catalytique ; de l'autre côté, la protéine protège le complexe métallique vis-à-vis de la dégradation en milieu aqueux et fournit également un environnement chiral propice à une catalyse énantiosélective. Ces catalyseurs fonctionnant de manière sélective en milieu aqueux s'inscrivent tout à fait dans les préceptes de la chimie verte.
Une nouvelle famille de protéines artificielles, nommée αRep, a été récemment décrite. Toutes les protéines de la bibliothèque αRep présentent le même repliement en solénoïde incurvé, mais diffèrent à la fois en taille (nombre de motifs répétés) et dans la nature de 5 acides aminés par motif répété. Une surface variable est ainsi générée sur la surface concave du solénoïde. Ces protéines sont extrêmement stables et modifiables. La modularité de ces protéines ouvre la voie à un panel varié d'ingénierie des protéines, notamment la conception de catalyseurs artificiels. Au sein de la bibliothèque αRep, le variant αRep-A3 est une protéine homodimérique pour laquelle les surfaces concaves de chaque monomère génèrent une crevasse. Les résidus formant cette crevasse peuvent être modifiés sans affecter la structure tridimensionnelle de la protéine. Le but de cette thèse a été d'évaluer la capacité de la protéine αRep-A3 à procurer une architecture rigide pour l'incorporation de complexes de métaux de transition. Pour cela, différents ligands de métaux de transition (phénanthroline, terpyridine, porphyrine) ont été couplés covalemment à des variants de αRep-A3 à différentes positions. Des résultats encourageants concernant la réaction de Diels-Alder entre azachalcone et cyclopentadiène (60% *ee*) suggèrent que ce type d'architecture pourrait fournir une base intéressante pour la création de nouvelles classes de métalloenzymes entièrement artificielles. Des pistes pour l'amélioration des catalyseurs basés sur les αRep par des méthodes d'évolution dirigée sont alors avancées sur la base de ces résultats.

Title : Engineering of the artificial protein architecture αRep: development of biohybrid catalysts by covalent coupling of metal complexes

Keywords : artificial metalloenzyme, biohybrid, enantioselective catalysis, green chemistry

Abstract : The development of a new generation of so-called biohybrid catalysts is based on the association of a metal complex and a protein. On the one hand, the metal complex is responsible for the catalytic activity; On the other hand, the protein protects the metal complex from degradation in aqueous medium and also provides a chiral environment conducive to enantioselective catalysis. These catalysts, which function selectively in an aqueous medium, fit perfectly into the precepts of green chemistry.
A new family of artificial proteins, called αRep, has recently been described. All proteins in the αRep library exhibit the same curved solenoid folding, but differ in size (number of repeating units) and in the nature of 5 amino acids per repeat unit. A variable surface is thus generated on the concave surface of the solenoid. These proteins are extremely stable and modifiable. The modularity of these proteins paves the way for a varied panel of protein engineering, including the design of artificial catalysts.
Within the αRep library, the variant αRep-A3 is a homodimeric protein for which the concave surfaces of each monomer generate a crevice. The residues forming this crevice can be modified without affecting the three-dimensional structure of the protein. The aim of this thesis has been to evaluate the ability of the αRep-A3 protein to provide a rigid scaffold for the incorporation of transition metal complexes. To this end, various transition metal ligands (phenanthroline, terpyridine, porphyrin) have been covalently coupled to variants of αRep-A3 at different positions. Encouraging results regarding the Diels-Alder reaction between azachalcone and cyclopentadiène (60% *ee*) suggest that this type of scaffold could provide an interesting basis for the creation of new classes of fully artificial metalloenzymes. From these results, lines of improvement for αRep-based catalysts by means of directed evolution are then advanced.