



HAL
open science

Étude du méthylome de la lysine méthyltransférase PR-Set7 avec des paires d'enzyme et de cofacteur chimiquement modifié

Alexandre Désert

► To cite this version:

Alexandre Désert. Étude du méthylome de la lysine méthyltransférase PR-Set7 avec des paires d'enzyme et de cofacteur chimiquement modifié. Biochimie, Biologie Moléculaire. Sorbonne Université, 2021. Français. NNT : 2021SORUS031 . tel-04132277

HAL Id: tel-04132277

<https://theses.hal.science/tel-04132277v1>

Submitted on 19 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sorbonne Université

ED 406 Chimie moléculaire

Laboratoire des Biomolécules – UMR7203 – Sorbonne Université/ENS/CNRS

Etude du méthylome de la lysine méthyltransférase PR-Set7 avec des paires de mutant d'enzyme et de cofacteur chimiquement modifié

Par Alexandre Désert

Thèse de doctorat de chimie biologique

Dirigée par Fabienne Burlina et Dominique Guianvarc'h
en collaboration avec Raphaël Margueron

Présentée et soutenue publiquement le 18 juin 2021

Devant un jury composé de :

Laurent Salmon	Professeur des universités, Université Paris Saclay	Rapporteur
Boris Vauzeilles	Directeur de Recherche, ICSN	Rapporteur
Raphaël Rodriguez	Directeur de Recherche, Institut Curie	Examinateur
Frédérique Peronnet	Directrice de Recherche, Sorbonne Université	Examinatrice
Dominique Guianvarc'h	Professeure des universités, Université Paris Saclay	Co-directrice de thèse
Fabienne Burlina	Directrice de recherche, Sorbonne Université	Directrice de thèse
Raphaël Margueron	Directeur de recherche, Institut Curie	Encadrant et membre invité

Table des matières

Abréviations	5
I. Introduction.....	9
I.1 Epigénétique	9
I.1.1 Définition	9
I.1.2 Les marques épigénétiques	9
<i>I.1.2.A La méthylation de l'ADN</i>	<i>9</i>
<i>I.1.2.B Les modifications post-traductionnelles des histones</i>	<i>10</i>
<i>I.1.2.C Les inter-régulations entre marques épigénétiques.....</i>	<i>16</i>
I.2 La méthylation des histones et des protéines non-histones.....	19
I.2.1 Les histone lysine méthyltransférases.....	19
<i>I.2.1.A G9a et GLP, deux H3K9 méthyltransférase.....</i>	<i>21</i>
<i>I.2.1.B PR-Set7, une H4K20 méthyltransférase.....</i>	<i>22</i>
I.2.2 Méthylation des protéines non-histones : la face cachée de l'iceberg ?	26
I.3. Méthodologies pour le profilage de la méthylation sur des lysines	28
1.3.1 Les puces à peptides et puces à protéines	29
1.3.2 Les anticorps anti pan-méthyl-lysines couplés à la spectrométrie de masse.....	30
1.3.3 L'affinité de domaine couplée à la spectrométrie de masse	31
1.3.4 L'enrichissement par voie chimique couplé à la spectrométrie de masse.....	31
Projet de thèse	35
II. Conception et mise au point des outils moléculaires et enzymatiques pour la stratégie de marquage des substrats de PR-Set7	41
II.1 Conception d'une série d'analogues de SAM	41
II.1.1. La S-adénosyl-L-méthionine : un cofacteur enzymatique clé	41
II.1.2. Panorama des analogues de SAM modifiés au niveau du méthyle	42
II.1.3. Les différentes méthodes d'accès aux analogues de SAM alkylés sur le soufre.....	47
II.1.4. Synthèse des analogues de SAM alkylés utilisés dans cette étude	50
II.2 Conception et production de mutants de PR-Set7	55
II.2.1 Paramètres structuraux pour la fixation des ligands de PR-Set7	55
II.2.2 Mécanisme de mono-méthylation de PR-Set7 WT	57
II.2.3 Conception des mutants de PR-Set7	58
II.2.4 Production des mutants de PR-Set7.....	59
II.2.5. Évaluation de l'activité enzymatique des mutants de PR-Set7.....	61
II.2.6. Préparation des substrats contenant l'histone H4 entière.....	63
II.2.7. Evaluation de l'activité enzymatique des mutants de PR-Set7 en présence de complexes d'histones	65
II.3 Conclusion	66
III. Criblage des analogues de SAM envers les mutants de PR-Set7	71

III.1. Etude préliminaire réalisée avec l'HKMT GLP	73
III.1.1 Test sur substrat peptidique	73
III.1.2 Test sur complexe d'histone : le tétramère	76
III.2. Etude réalisée avec PR-Set7 et ses mutants	78
III.2.1 Test sur substrat peptidique	78
III.2.2 Test sur complexe d'histone : le nucléosome reconstitué	79
III.3. Etude en présence d'un analogue de cofacteur SAM sélénié minimaliste : ProSeAM	82
III.3.1. Synthèse et caractérisation de la ProSeAM	84
III.3.2. Evaluation de la ProSeAM sur PR-Set7 et ses mutants	87
III.4. Mise au point de nouvelles techniques pour faciliter l'analyse de la fonctionnalisation des histones	89
III.4.1. Les réactions de chimie click sur membrane de PVDF	89
III.4.2. L'analyse directe de la fonctionnalisation des histones par spectrométrie de masse MALDI-TOF après transfert sur membrane de nitrocellulose	92
III.5. Conclusion	93
IV. De la protéomique aux premières étapes de validation biologique	97
IV.1. Mise en place et optimisation du protocole de capture des protéines	99
IV.1.1. Validation de la spécificité de position de fonctionnalisation de PR-Set7	99
IV.1.2. Optimisation de la capture sur billes de streptavidine	100
IV.2. Stratégie mise en œuvre pour l'approche de protéomique	102
IV.3. Réalisation de l'expérience de protéomique	103
IV.3.1 Expérience préliminaire.....	103
IV.3.2 Expérience principale	104
IV.4. Interprétation des résultats de la protéomique	109
IV.4.1. Exploitation « non biaisée » des résultats de l'étude protéomique	110
IV.4.2. Premières étapes de la validation biologique des candidats substrats	112
<i>Conclusion et perspectives de la validation biologique des candidats substrats de PR-Set7.....</i>	<i>117</i>
IV.4.3. Présentation de la méthode d'interprétation « biaisée » des résultats de l'étude protéomique.....	118
<i>IV.4.3.A Etude des données par recherche de site consensus.....</i>	<i>120</i>
<i>IV.4.3.B Etude de l'interactome des potentiels substrats de PR-Set7</i>	<i>123</i>
IV.5 Etude bibliographique du lien entre les HKMT à domaine SET et les voies de signalisation impliquant des kinases.....	125
IV.5.1 La voie des MAP Kinase.....	125
IV.5.2 La voie Pi3k/AKT	126
IV.5.3 La régulation de facteurs de transcription par les MTases.....	127
IV.6 Conclusion	127
V. Analyse structurale - Rationalisation du design des paires enzyme mutée / analogue modifié.....	131
V.1. Contexte et objectifs de l'étude de docking.....	131

V.2 Mise au point et validation de la stratégie de docking	132
V.2.1 Principes généraux du protocole de simulation	132
V.2.2. Préparation des ligands et de la protéine	133
V.2.3. Validation du protocole de docking	135
V.3 Etude des paramètres structuraux essentiels à la prise en charge de ProSeAM par PR-Set7	136
V.4 Etude des paramètres structuraux essentiels à la prise en charge de ProSeAM par GLP	140
V.5 Etude des paramètres structuraux clés dans la prise en charge d’analogues de SAM à longue chaîne alkyle	142
V.6 Proposition de paires d’enzyme et d’analogue de 2 ^{ème} génération	146
V.6.1 Conception de mutants et étude par docking	146
V.6.2 Evaluation de l’activité des nouveaux mutants	149
V.7 Conclusion	151
Conclusion générale et perspectives	155
Annexes	163
Matériel et méthodes	177
Références	201

Abréviations

5mC	5-méthylcytosine
ACN	Acétonitrile
ADN	Acide désoxyribonucléique
ARN	Acide ribonucléique
ATP	Adénosine triphosphate
BER	Réparation par excision de base
BSA	Albumine de sérum bovin
CHARMm	<i>Chemistry at HARvard Macromolecular Mechanics</i>
CHCA	Acide α -cyano-4-hydroxycinnamique
CIDA	5'-chloro-5'-désoxyadénosine
CuAAC	Cycloaddition alcyne-azoture catalysée par le cuivre (I)
CpG	Dinucléotide CG
DCM	Dichlorométhane
DIBO	Dibenzocyclooctyne
DNMT	ADN méthyltransférase
Domaine MBD	<i>Methyl-CpG binding domain</i>
DTT	Dithiothréitol
EDTA	Éthylènediaminetétraacétique
eq.	équivalent
FDA	5'-fluoro-5'-désoxyadénosine
FDAS	5'-fluoro-5'-désoxyadénosine synthase
FDR	Taux de faux positif
GO	<i>Gene Ontology</i>
H3K9	Lysine 9 de l'histone H3
H4K20	Lysine 20 de l'histone H4
HAT	Histone acétyltransférase
HDAC	Histone déacétylase
HKMT	Histone lysine méthyltransférase
HPLC	Chromatographie liquide haute performance
HRMT	Histone arginine méthyltransférase

HRP	<i>Horseradish peroxidase</i>
HSL	L-homosérine lactone
IPTG	Isopropyl β -D-1-thiogalactopyranoside
LC-MS / MS	Chromatographie en phase liquide couplée à la spectrométrie de masse en tandem
MAT	Méthionine adénosyltransférase
MALDI-TOF	<i>Matrix assisted laser desorption ionization-Time of flight</i>
NER	Réparation par excision de nucléotides
PBST	Phosphate Buffered Saline with Tween 20
PCR	Amplification en chaîne par polymérase
PDB	<i>Protein Data Bank</i>
PEG	Polyéthylène glycol
PMSF	Fluorure de phénylméthylsulfonyle
PVDF	Polyfluorure de vinylidène
Rf	Facteur de rétention
RP	<i>Reverse phase</i>
Rt	<i>Retention time</i>
SA1L	5'-chloro-5'-désoxyadénosine synthase
SAH	S-Adénosyl-L-homocystéine
SAM	S-Adénosyl-L-Méthionine
SET	Su(var)3-9, Enhancer-of-zeste and Trithorax
SDS	Dodécylsulfate de sodium
SDS-PAGE	Electrophorèse dénaturante en gel de polyacrylamide en présence de SDS
SPAAC	<i>Strain-promoted azide-alkyne cycloaddition</i>
TAMRA	Tétraméthylrhodamine
TFA	Acide trifluoroacétique
TIC	Total Ion Chromatogram
THF	Tétrahydrofurane
THPTA	Tris-hydroxypropyltriazolylmethylamine
WT	wild-type
XIC	<i>Extracted Ion Chromatogram</i>

Introduction

I. Introduction

I.1 Epigénétique

I.1.1 Définition

Le terme épigénétique a pour racine le préfixe provenant du grec ancien “ἐπι” (epi-) signifiant « au-delà de » et du mot grec “γένος” (genos) référant aux gènes. Historiquement, le terme a été introduit par Conrad Waddington et décrivait la branche de la biologie étudiant les interactions entre les gènes et leurs produits amenant à la construction d’un phénotype¹. Le terme s’est précisé avec le temps et les découvertes scientifiques et désigne maintenant les processus moléculaires responsables de changements héréditaires dans l’expression des gènes qui ne sont pas causés par des modifications de la séquence d’ADN^{2,3}. Ces processus moléculaires incluent notamment la méthylation de l’ADN, les modifications des histones, les ARN non codants⁴, et participent à différents phénomènes biologiques parmi lesquels nous citerons les gènes soumis à l’empreinte parentale⁵, l’inactivation du chromosome X⁶ ou l’activation – l’inactivation des gènes homéotiques⁷. Ces phénomènes épigénétiques sont responsables d’une modulation de l’état transcriptionnel des gènes, passant par le recrutement de facteurs de transcription, et contribuent à la génération de différents types cellulaires à partir d’un même génome.

I.1.2 Les marques épigénétiques

I.1.2.A La méthylation de l’ADN

La méthylation de l’ADN désigne une modification chimique pouvant être introduite par des enzymes appelées **ADN méthyltransférases (DNMT)** sur certains acides nucléiques. La modification la plus importante est la méthylation de la cytosine pour former la 5-méthylcytosine (5mC). Cette base modifiée est souvent surnommée la cinquième base de l’ADN. Elle est très répandue chez de nombreux organismes⁸ (plantes, champignons, animaux) et joue un rôle dans la régulation de l’expression des gènes (Figure 1). Elle fait partie à ce titre, des nombreuses marques épigénétiques. En effet, la méthylation des cytosines dans le contexte des îlots CpG⁹, zones riches en dinucléotide CG très présents dans les régions promotrices des gènes, a un effet de suppression de la transcription génique grâce à deux mécanismes. Le premier est causé par la présence du groupement méthyle sur la cytosine en un site spécifique de la séquence d’ADN qui peut directement bloquer le site de reconnaissance de certains

facteurs de transcription. Le second, et probablement le plus important, se base sur la reconnaissance préférentielle de l'ADN méthylé par des protéines possédant un **domaine MBD** (methyl-CpG-binding domain)¹⁰. L'interaction de l'ADN avec ces protéines MBD permet le recrutement des partenaires protéiques impliqués dans la modification des histones telles que des histones déacétylases (HDAC). L'action de ces enzymes étant associée à la condensation de la chromatine, elle induit généralement la répression de la transcription des gènes.

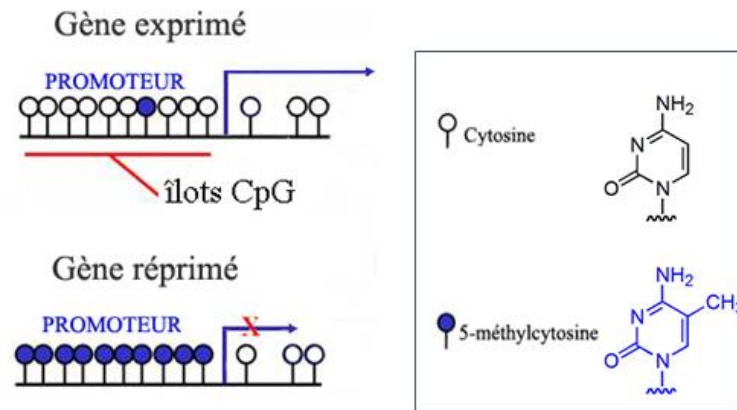


Figure 1. Rôle de la 5-méthylcytosine

La méthylation de l'ADN interagit aussi avec d'autres marques épigénétiques, notamment les modifications d'histones pour recruter des partenaires protéiques impliqués dans le remodelage de la chromatine. Des régions chromatiniennes plus compactées sont ainsi moins accessibles à des facteurs de transcription et les gènes sont par conséquent moins exprimés. Inversement, dans des régions plus accessibles aux facteurs de transcription, les gènes sont davantage exprimés.

La méthylation de l'ADN est considérée comme une marque épigénétique stable et persistante qui est généralement déposée mais pas enlevée. La méthylation de l'ADN au niveau des îlots CpGs induit donc généralement une répression sur le long terme. Deux vagues de déméthylation de l'ADN ont toutefois lieu lors de l'embryogénèse, une après les toutes premières étapes du développement embryonnaire et l'autre au cours de la gamétogénèse¹¹. Ces processus amènent à la perte presque totale de la marque de méthylation sur les cytosines.

I.1.2.B Les modifications post-traductionnelles des histones

Les histones sont des protéines localisées dans le noyau des cellules eucaryotes. Parmi elles, quatre histones H2A, H2B, H3, et H4 peuvent s'assembler sous la forme d'un hétéro-octamère qui, une fois associé à de l'ADN, forme une structure appelée nucléosome¹². Ces nucléosomes

sont les unités organisationnelles de la chromatine. Ils assurent le premier niveau de compaction de l'ADN au sein du noyau, formant la structure dite en « collier de perles » (Figure 2). En contrôlant l'accessibilité au double brin d'ADN, le nucléosome est impliqué dans la régulation de la transcription, de la réplication de l'ADN et dans les processus de réparation des dommages à l'ADN. Ils interviennent aussi dans la régulation de l'expression des gènes par l'intermédiaire des modifications d'histones.

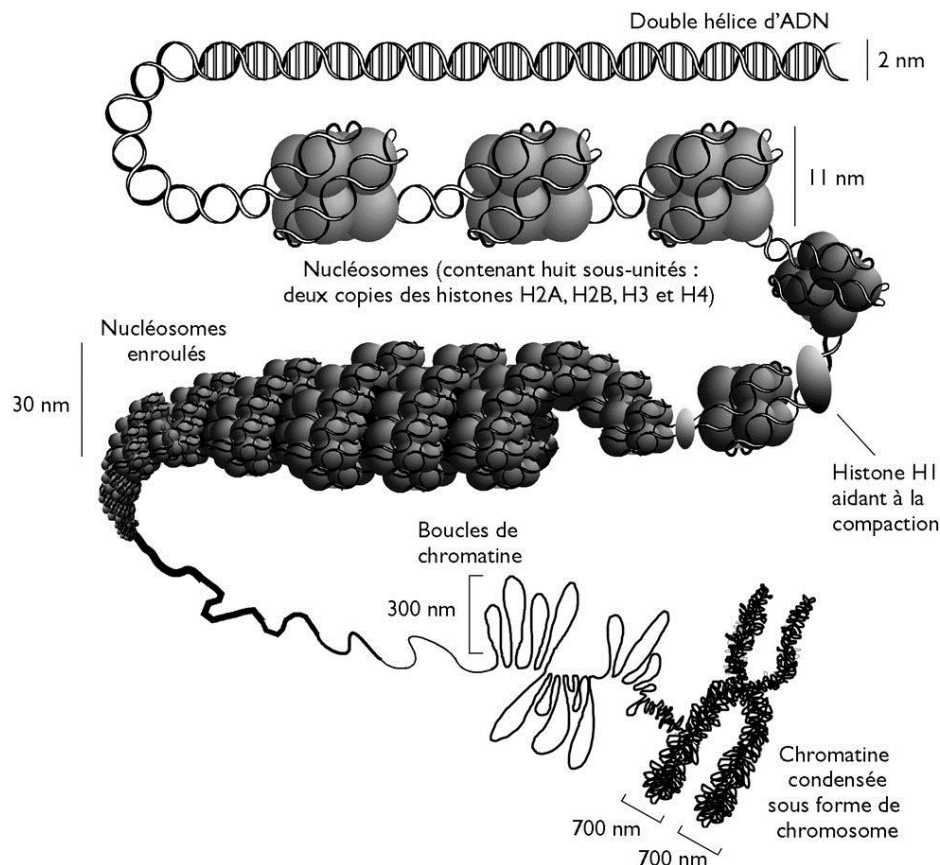


Figure 2. Les différents niveaux de compaction de la chromatine. D'après J. Brodeur et al. 2007.¹³

On appelle **modifications post traductionnelles (PTM)** des histones, les modifications chimiques covalentes au niveau de chaînes latérales d'acides aminés survenant après la traduction par le ribosome et le plus souvent introduites par des enzymes spécifiques¹⁴. Elles sont de nature extrêmement variées et concernent plusieurs types d'acides aminés ; notamment les lysines, les arginines et les sérines qui peuvent chacun être affecté par plusieurs types de modifications (Figure 3). La méthylation, l'acétylation, la phosphorylation et l'ubiquitinylation en sont les principales représentantes mais il en existe de nombreuses autres comme la citrullination, la SUMOylation, ou l'*O*-GlcNAcylation¹⁵.

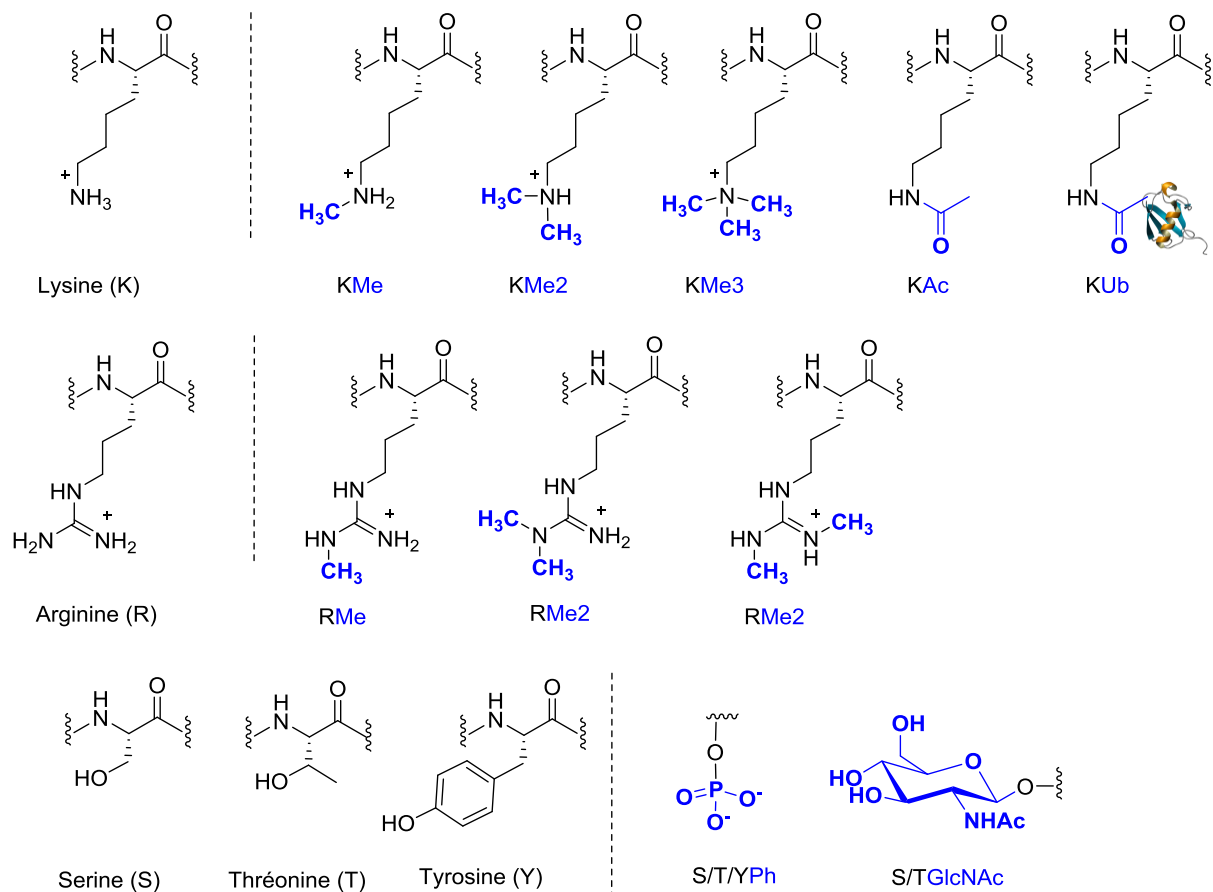


Figure 3. Principales modifications d'histones. Me : monométhylation, Me2 : diméthylation, Me3 : triméthylation, Ac : acétylation, Cr : crotonylation, Ub : ubiquitinylation, Ph : phosphorylation, GlcNAc : O-GlcNAcylation.

Ces modifications sont particulièrement enrichies sur les extrémités *N*-terminales des histones et régulent le degré de compaction de la chromatine mais sont également présentes dans le cœur des histones¹⁶ (Figure 4). Cette régulation peut être directe si la modification, de par ses caractéristiques physicochimiques, induit un changement de l'état chromatinien, ou indirecte, si la régulation passe par le recrutement de protéines effectrices capables de reconnaître la modification¹⁷.

Selon son emplacement, la modification d'un acide aminé d'une histone peut influencer de manière synergique ou antagoniste la modification d'un autre. Ces marques sont donc présentes de façon plus ou moins associées ce qui forme des combinaisons de marques, des motifs. Il a été montré qu'il existait un lien étroit entre ces différentes combinaisons de marques épigénétiques et l'état transcriptionnel des gènes permettant de prédire à partir de certains contextes épigénétiques, le statut transcriptionnel d'un gène¹⁸. C'est sur cette hypothèse que se base la théorie du « **code histone**¹⁹ » : un code qui serait une surcouche du code génétique et

qui permettrait une modulation fine de l'expression des gènes en fonction de différents contextes biologiques.

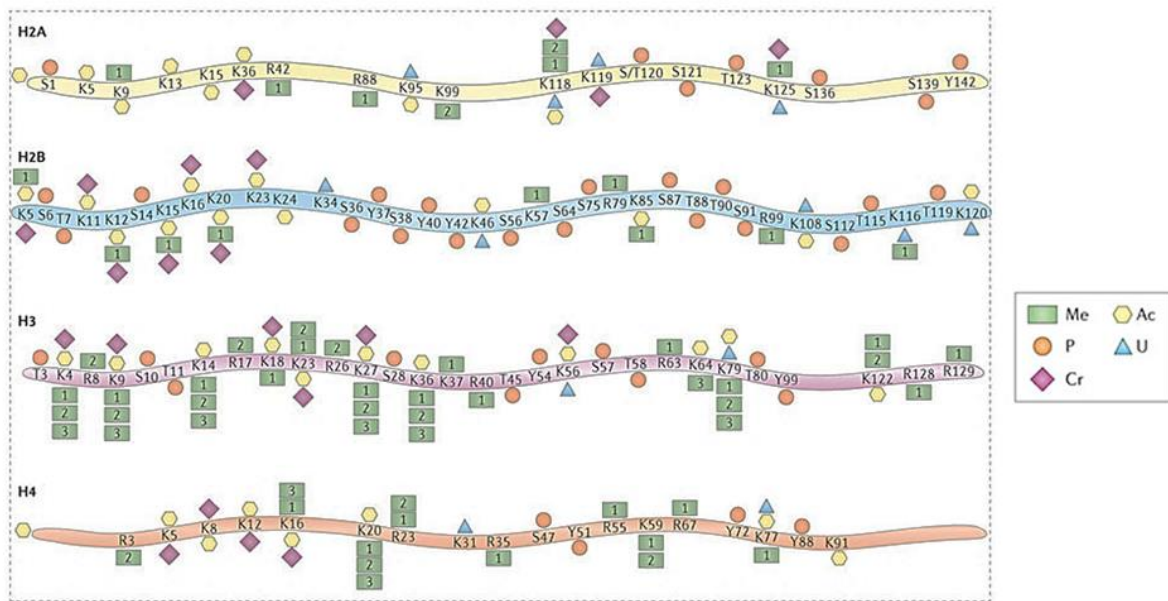


Figure 4. Diversité de positionnement des modifications introduites sur les histones. Me : méthylation, P : phosphorylation, Cr : crotonylation, Ac : acétylation, U : ubiquitinylation. D'après J. Gräff. 2013¹⁷

Le contrôle du positionnement des modifications des histones est assuré grâce à l'action de trois catégories de protéines :

- Les protéines « writer », des enzymes capables d'apposer les marques sur des résidus spécifiques des histones ;
- Les protéines « reader », disposant d'un domaine de reconnaissance d'une marque. Elles permettent souvent le recrutement d'autres protéines. Les protéines recrutées peuvent être par exemple des protéines capables d'apposer ou d'enlever d'autres marques épigénétiques pour constituer le motif épigénétique. Ces protéines peuvent aussi être des facteurs de transcription agissant donc directement sur l'état transcriptionnel des gènes au voisinage de la modification épigénétique ;
- Les protéines « eraser », des enzymes capables d'effacer les marques sur les histones.

Dans l'ensemble, et en comparaison avec la méthylation de l'ADN, les modifications des histones sont des modifications qui peuvent être rapidement ajoutées ou enlevées par l'action d'enzymes. Une fois déposées, les modifications sont lues par des protéines effectrices capables de déclencher des cascades d'événements amenant à une plus ou moins grande activité de transcription. Associés ensemble, les motifs constitués à partir des modifications des histones

et la méthylation de l'ADN assurent un contrôle finement régulé et très dynamique de l'expression de gènes (Figure 5) selon le contexte biologique et permettent de répondre à différents stress environnementaux.

Dans ce manuscrit nous aborderont essentiellement le sujet de la méthylation des histones mais aussi celle de protéines non-histones. Occasionnellement, une comparaison avec l'acétylation des histones pourra aussi être abordée.

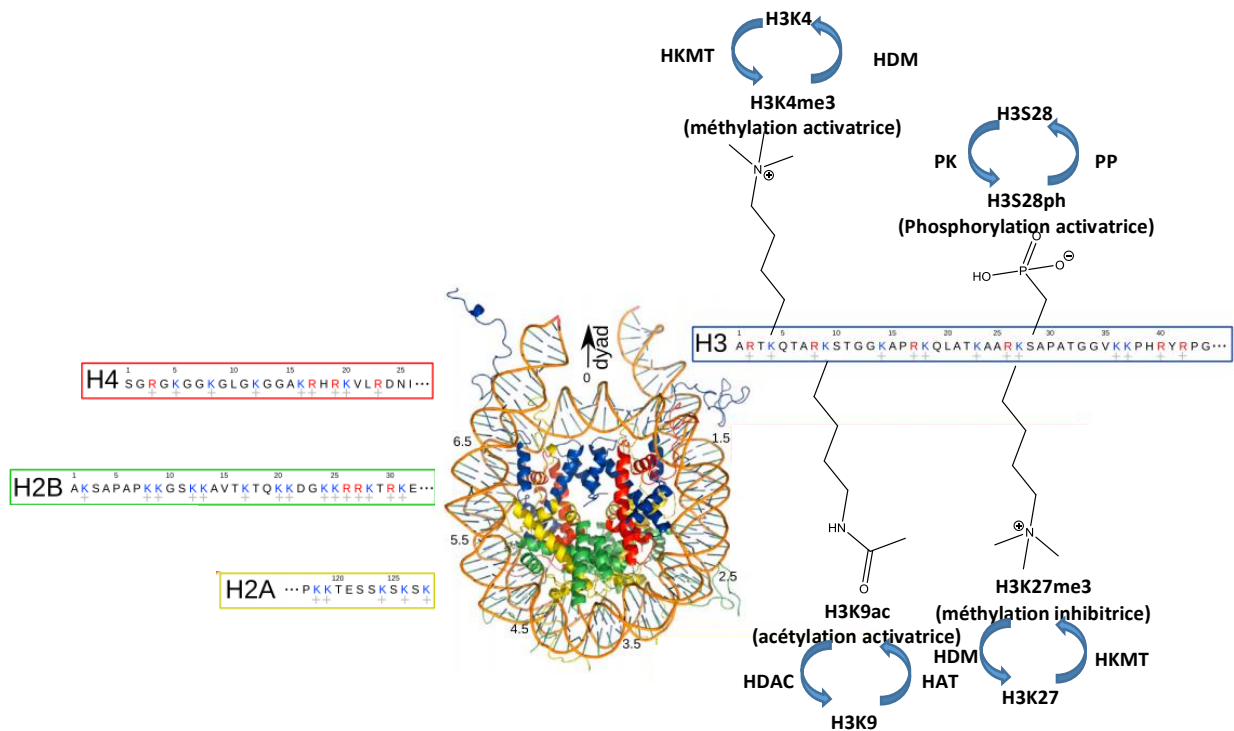


Figure 5. Nucléosome constitué de 147 pb d'ADN enroulés autour d'un cœur d'histone (H3 en bleu, H4 en rouge, H2A en jaune, H2B en vert). La séquence des queues d'histones est indiquée et des modifications d'histones illustrant la diversité des marques épigénétiques et leur effet sur la régulation de la transcription sont renseignées pour l'histone H3 à titre d'exemple. HKMT : histone lysine méthyltransférase ; HDM : histone déméthylase ; HAT histone acétyltransférase ; HDAC : histone déacétylase ; PK : protéine kinase ; PP : protéine phosphatase.

La méthylation des histones

La méthylation des histones est une modification chimique des lysines ou des arginines par l'action d'enzymes nommées **histone lysine méthyltransférases (HKMT)** ou histone arginine méthyltransférases (HRMT) utilisant la **S-adénosyl-L-méthionine** comme cofacteur. Selon les enzymes, les lysines peuvent être méthylées jusqu'à trois fois et les arginines, deux fois. De plus, selon la position et le degré de méthylation, l'effet sur l'expression génique peut être soit activateur soit répresseur. L'effet de la méthylation est donc entièrement dépendant du contexte et il sera largement influencé par la présence ou l'absence d'autres marques épigénétiques ou par la présence de protéines effectrices. La méthylation ne modifiant pas la charge électronique du résidu (lysine ou arginine) et étant une petite modification, ses effets passent essentiellement par l'action de protéines effectrices reconnaissant le site et son état de méthylation. Ces protéines « readers » doivent donc contenir un domaine de reconnaissance de la méthylation. En ce qui concerne la méthylation des lysines, ces domaines sont : le domaine PHD, le chromodomaine, les domaines, tudor, PWWP, WD40, BAH, ADD, MBT, zn-CW et les répétitions ankyrine²⁰. La suppression de la méthylation est assurée par des lysines et arginine déméthylases²¹.

L'acétylation des histones

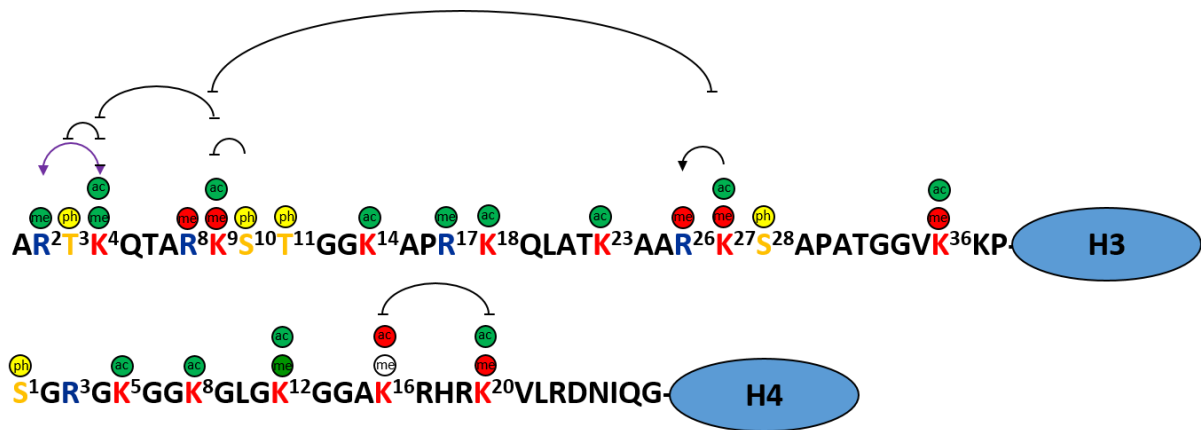
L'acétylation des histones est une modification introduite par l'action **d'histone acétyltransférases (HAT)**. Elles catalysent la réaction d'acétylation du groupement amine de la chaîne latérale de certaines lysines en utilisant l'acétyl-coenzyme A (acétyl-CoA) comme cofacteur. L'acétylation de lysines est généralement associée à un état décompacté de la chromatine (euchromatine) et à l'activation transcriptionnelle des gènes. En effet la perte de la charge positive du groupement ammonium induit une diminution des interactions entre les histones (chargées positivement) et l'ADN (chargé négativement). Ceci s'accompagne donc d'une meilleure accessibilité potentielle de l'ADN aux facteurs de transcription. Comme pour la méthylation, les effets liés à l'acétylation des histones peuvent passer par le recrutement de protéines *readers* reconnaissant les protéines acétylées. Plusieurs domaines protéiques en sont capables, le plus connu étant le bromodomaine. Mais contrairement à la méthylation, l'acétylation ayant pour conséquence une perte de charge sur l'histone, induit par nature, une conformation de la chromatine plus favorable à la transcription des gènes. Enfin, ces marques sont effacées par l'action des histones déacétylases (HDACs).

I.1.2.C Les inter-régulations entre marques épigénétiques

Régulation entre modifications d'histones

De façon intéressante, il a été montré à de nombreuses occasions que le dépôt d'une modification d'histone, pouvait influencer positivement ou négativement le dépôt d'une autre limitant ainsi le nombre de combinaisons possibles. Cette tendance peut s'expliquer de plusieurs façons. La première, s'explique par une augmentation de l'encombrement stérique généré par l'ajout de la marque par rapport au résidu non modifié qui pourrait interférer avec la capacité d'une enzyme - ou d'une protéine recrutant une enzyme – de se fixer sur l'histone cible. La deuxième raison est expliquée par le recrutement d'un partenaire sur un site fonctionnalisé qui empêcherait une autre enzyme d'aller modifier un résidu voisin. La Figure 6 a pour objectif de renseigner les principales inter-régulations connues à ce jour sur les histones H3 et H4 de *Homo sapiens*.

Afin d'illustrer cette inter-régulation, on peut citer un exemple précis impliquant la méthylation de l'histone H4K20 qui est au centre de mon étude : l'inter-régulation qui a lieu entre la lysine 20 méthylée (H4K20me1) et la lysine 16 acétylée (H4K16ac). En effet, la méthylation de la lysine 20, prévient l'acétylation de la lysine 16 par la KAT p300²². Inversement, l'acétylation de la lysine 16 empêche la méthylation de la lysine 20. Les marques H4K20me1 et H4K16ac sont donc mutuellement antagonistes²². Dans ce cas, l'exclusivité des marques peut s'expliquer avec la structure cristallographique de PR-Set7 lié à l'histone H4²³. Dans le site actif de PR-Set7, H4K16 est entourée des résidus Ala-342, Ala-346 et His-347 et le groupement imidazole de cette dernière interagit avec le squelette de H4K16 par liaison hydrogène. Ainsi la modification de H4K16 pourrait affecter la liaison de PR-Set7 avec l'histone H4 pour méthyler H4K20. En effet, l'His-347 de PR-Set7 fait partie d'un réseau d'empilement π important pour la reconnaissance de H4K20 par l'enzyme. Nous y reviendrons dans le chapitre dédié à l'enzyme PR-Set7 du chapitre II.



Inter-régulation	Caractéristique	Ref
H3R2me2a - H3K4me3	Exclusion mutuelle	24
H3R2me2s - H3K4me3	Agoniste mutuel	25
H3K4me1 - H3K9me2/3	Exclusion	26
H3K4me3 - H3K9me3	Exclusion	27–29
H3T3ph – H3K4me3	Exclusion mutuelle	30,31
H3K9me3 – H3K27me3	Généralement mutuellement exclusives	32
H4K16ac - H4K20me	Exclusion mutuelle	22

Figure 6. Inter-régulations sur les extrémités *N*-terminales des histones H3 et H4. Les séquences d'acides aminés sont annotées pour souligner les résidus modifiés importants. Les lignes noires représentent les événements d'inter-régulation existants chez Homo Sapiens. La flèche violette représente une superposition d'événements d'inter-régulation. Le tableau résume l'ensemble des interactions publiées. Les ronds verts symbolisent des marques associées à l'activation transcriptionnelle, les ronds rouges à la répression transcriptionnelle. Les ronds restent blanc si l'effet n'est pas déterminé. Le jaune est attribué à la phosphorylation des sérines et des thréonines.

Les travaux de recherche autour des différentes associations de marques tendent à montrer que ces motifs épigénétiques permettent une régulation dans le temps et selon le contexte biologique de l'expression de gènes et il faut garder en tête que pour chaque marque existe une inter-régulation très fine qui est encore loin d'être complètement élucidée.

Régulation entre modifications d'histones et méthylation de l'ADN

La méthylation de l'ADN et la modification des histones sont donc chacune impliquées dans la mise en place de la régulation de l'expression génique. La méthylation des histones est associée à une régulation positive ou négative de l'expression génique selon l'histone ciblée, la position et le niveau de méthylation et, est facilement réversible. Cela contraste avec la méthylation de l'ADN qui est plutôt associée à une répression génique stable sur le long terme. Toutefois, il a été récemment montré que les voies de signalisation dépendantes de la méthylation de l'ADN et celles dépendantes de la méthylation des histones étaient étroitement liées et qu'elles pouvaient dépendre l'une de l'autre. En effet, les inter-régulations entrant en jeu peuvent être médiées par les interactions existantes entre les ADN méthyltransférases et les histones méthyltransférases. En guise d'exemple, il a été montré dans différents articles que ce type d'inter-régulations avait des implications dans le développement³³, la reprogrammation cellulaire de cellules somatiques en cellules pluripotentes³⁴ ou dans des mécanismes de tumorigenèse³⁵. La Figure 7, reprise de la revue de Cedar et Bergman de 2009³⁶ illustre ces propos. Dans les cellules embryonnaires des gènes responsables de la pluripotence ont leurs îlots CpG non méthylés (cercles violets sur la Figure 7). Ces gènes qui doivent être transcrits sont aussi associés à des histones qui ont leurs lysines acétylées et H3K4 est méthylée. Avec le début de la différenciation cellulaire, l'HKMT G9a est recrutée avec une histone déacétylase qui va enlever l'acétylation sur les histones. H3K4 est parallèlement déméthylée. A ce stade, toutes les marques épigénétiques classiquement associées à la transcription ont été enlevées. Au stade suivant, G9a est recrutée sur H3K9 et méthyle cette position qui est ensuite reconnue par la protéine HP1 (heterochromatin protein 1). Ceci induit localement la formation d'hétérochromatine. Enfin, G9a recrute les ADN méthyltransférases DNMT3a et DNMT3B qui vont procéder à la méthylation des îlots CpG. (cercles bleu foncé sur la Figure 7) ce qui induira l'inhibition sur le long terme des gènes liés à la pluripotence.

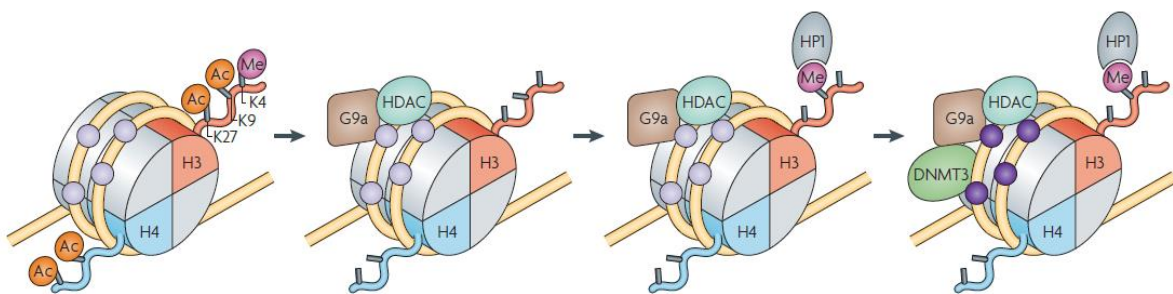


Figure 7. Mécanisme d'extinction de la pluripotence des gènes. D'après Cedar et Bergman, 2009³⁶. Les cercles violets symbolisent les îlots CpG non méthylés, les cercles bleus foncés, ceux méthylés par DNMTA et DNMT3B.

Dans l'ensemble, ces connexions établies entre la méthylation et acétylation des histones et la méthylation de l'ADN illustrent bien la complexité de la régulation génique exercée par l'ensemble des marques épigénétiques et révèle une grande régulation fine des différents processus biologiques pour le maintien de l'homéostasie cellulaire.

Dans la suite de ce chapitre, nous focaliserons notre attention sur la méthylation des histones au cœur de mon projet de thèse.

I.2 La méthylation des histones et des protéines non-histones

I.2.1 Les histone lysine méthyltransférases

Les histone lysine méthyltransférases (HKMT) sont des enzymes catalysant la réaction de transfert d'un groupement méthyle sur le groupement amine en position ε de lysines (Figure 8). Ces enzymes sont réparties en deux grandes familles : les HKMT à domaine SET et les HKMT à 7 brins β, toutes deux utilisant la *S*-adényl-L-méthionine (SAM) comme donneur de groupement méthyle. Dans cette thèse, nous étudierons exclusivement les HKMT possédant un domaine SET.

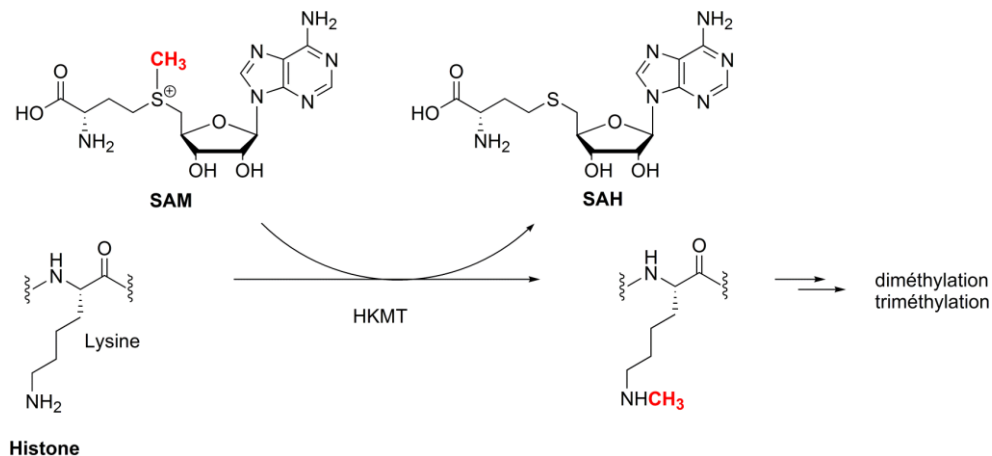


Figure 8. Réaction de méthylation catalysée par les HKMT

Le **domaine SET** est un domaine protéique minimal d'approximativement 130 acides aminés qui possède généralement une activité méthyltransférase. Il a été découvert en 1998 comme étant une séquence conservée dans trois protéines **Su(var)3-9**, **Enhancer of zeste** et **Trithorax** chez *Drosophila melanogaster* et tire son nom de l'acronyme formé par les initiales de ces protéines³⁷. La structure du domaine SET est caractérisée par le repliement de 3 séquences

(n-SET, SET et c-SET) sous la forme d'un pseudo-nœud de brins β . Dans la Figure 9, sont présentées les structures de trois exemples de protéines à domaine SET, Set7/9, GLP et PR-Set7. Cette structure particulière permet le rapprochement du site de liaison de la SAM et de celui du substrat pour former un site actif permettant le transfert du groupement méthyle³⁸.

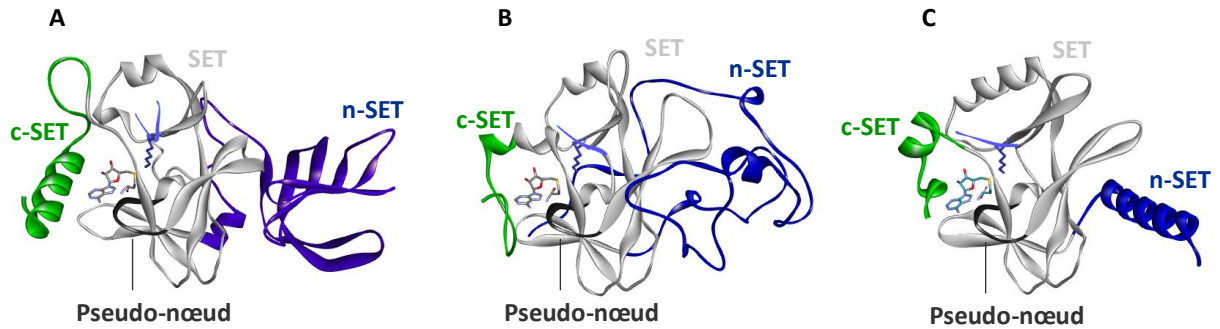


Figure 9. Structures de protéines contenant un domaine SET. A. SET7/9 (code PDB : 1O9S). B. GLP (code PDB : 2RFI). C. PR-Set7 (code PDB : 1ZKK). Les domaines n-SET, SET et c-SET des enzymes sont mentionnés. Sont également représentés, la structure secondaire de pseudo-nœud (en gris foncé) formée par le rapprochement dans l'espace de deux motifs conservés du domaine SET, le peptide substrat de chacune des enzymes (en bleu clair) et le co-produit de la réaction enzymatique : la SAH.

Depuis la découverte de ce domaine, des protéines à domaine SET ont été mises en évidence chez tous les organismes eucaryotes. A ce jour, on en dénombre près de 60³⁹. La Figure 10 extraite de l'article de Wu *et al.* de 2011⁴⁰ représente sous la forme d'un arbre phylogénétique la proximité des domaines SET de l'ensemble des méthyltransférases possédant un domaine SET (arginines méthyltransférases incluses).

Dans la suite de ce manuscrit, nous focaliserons notre attention sur deux membres de cette famille d'enzyme à domaine SET : PR-Set7 qui a fait l'objet de mes travaux et GLP que j'ai utilisée tout au long de mon étude comme enzyme modèle.

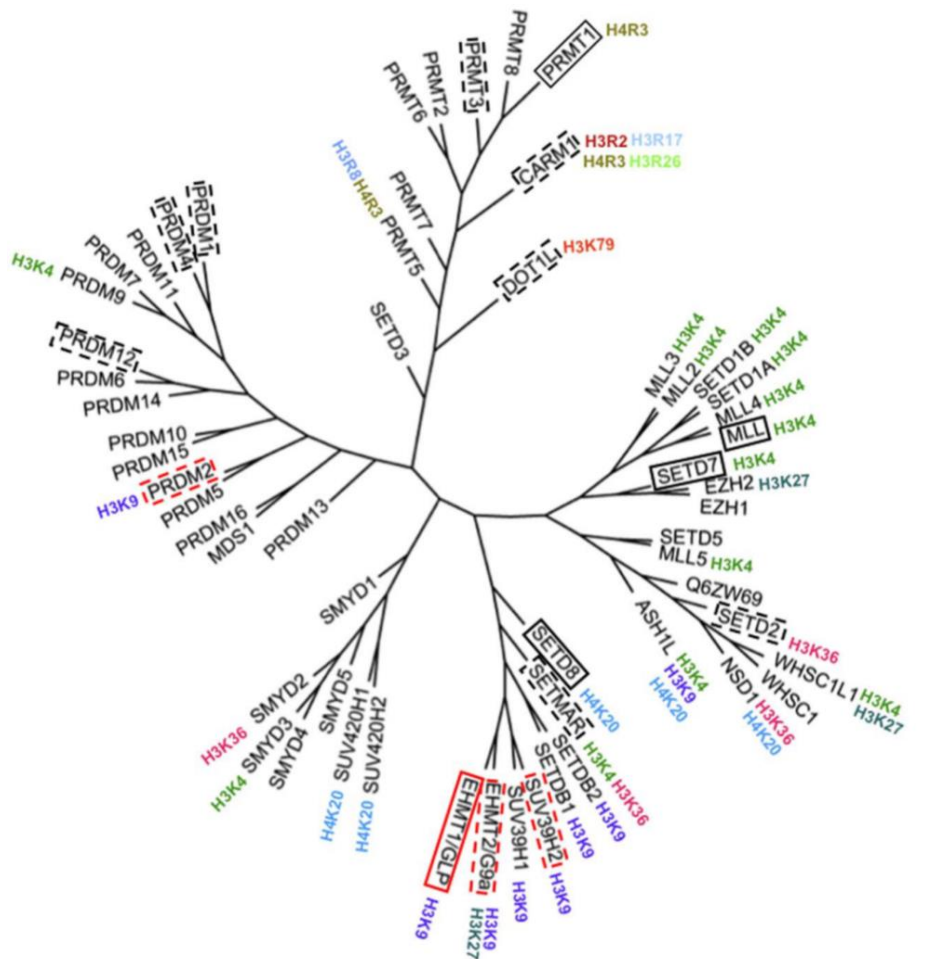


Figure 10. Phylogénie basée sur l’alignement de séquence multiple des méthyltransférases possédant un domaine SET (d’après Wu et al. 2011⁴⁰). Les enzymes encadrées sont des protéines dont la structure en complexe avec le peptide ont été résolues. Les enzymes encadrées en pointillés ont aussi leur structure résolue mais en absence du peptide substrat. En rouge, figurent les enzymes résolues par les auteurs de l’article.

I.2.1.A G9a et GLP, deux H3K9 méthyltransférase

Après la découverte de la première lysine méthyltransférase contenant un domaine SET, Suv39H1 en 2000, démontrant que le domaine SET était le domaine catalytique permettant l’activité de méthylation⁴¹, de nombreuses HKMT ont été identifiées chez de nombreux organismes multicellulaires mais aussi chez des procaryotes et des virus⁴². Parmi ces HKMT, G9a (aussi nommée KMT1C, ou EHMT2) et GLP (G9a like protein, aussi nommée KMT1D ou EHMT1) sont deux HKMT hyperactives, structurellement très proches, méthylant spécifiquement la lysine 9 de l’histone H3⁴³. G9a en particulier est responsable de l’essentiel de l’activité de diméthylation sur H3K9. Essentielle au développement, sa forte surexpression est associée au développement de nombreux cancers (du poumon, de la gorge, du sein...) et à des pronostics vitaux défavorables pour les patients atteints de cancer⁴⁴.

La méthylation sur H3K9 (H3K9me) est reconnue par des protéines « reader » associées à un état compacté de la chromatine et à la répression de la transcription des gènes. Parmi elles on retrouve les protéines HP1, G9a et GLP elles-mêmes et plusieurs autres protéines⁴⁵.

Substrats non histone de GLP et G9a

La répression transcriptionnelle passant par la méthylation de H3K9 est une fonction importante de GLP et G9a mais ces deux enzymes sont aussi capables de méthyler d'autres substrats que les histones. Le premier exemple étant l'événement d'auto-méthylation de G9a sur sa lysine 239⁴⁶. Tout comme H3K9me, la lysine auto-méthylée de GLP agit comme une plateforme permettant le recrutement de protéines affines pour la méthylation. En effet il a été montré que, les protéines HP1 pouvaient se lier à cette lysine auto-méthylée ce qui, par conséquent, pourrait contribuer à la répression de transcription⁴⁷. Par la suite, le site consensus « RK » de G9a a été défini grâce à une approche de puces à peptides dont nous reviendrons sur le principe au paragraphe suivant⁴⁸. Grâce à cette approche, l'équipe de Jeltsch réussit à identifier trois nouveaux candidats non-histone et retrouva en plus G9a, validant ainsi leur méthode. Ainsi la découverte de nouveaux substrats de G9a suggère un rôle important de GLP et G9a dans la méthylation de protéines non histone même si les effets induits par ces modifications restent encore mal compris. Enfin parmi les substrats communs de GLP et G9a, on retrouve par exemple la protéine suppresseur de tumeur p53, méthylée sur sa Lysine 373 (p53K373) et l'ADN méthyltransférase DNMT3a méthylée sur sa lysine K44(DNMT3AK44) dont les effets sont au moins partiellement redondants avec celui induit par la méthylation de H3K9, la méthylation de ces deux substrats ayant été associée à une répression de la transcription des gènes^{49,50}.

En résumé, en plus de la méthylation de H3K9, il a été montré que GLP et G9a méthylaient plusieurs protéines non histones et que les effets de la méthylation de ces nouveaux substrats sont au moins partiellement redondants avec celle de H3K9.

I.2.1.B PR-Set7, une H4K20 méthyltransférase

Au sein des HKMT, PR-Set7 (aussi nommée SETD8, SET8 ou KMT5A) a la particularité d'être la seule HKMT connue pour catalyser la réaction de mono-méthylation de la lysine 20 de l'histone H4 (H4K20me1)^{22,51}. La di- et la tri-méthylation de ce résidu est catalysée par les HKMT Suv4-20h1 et Suv4-20h2⁵², et nécessite une mono-méthylation préalable par PR-Set7. La méthylation de H4K20 est donc totalement altérée en l'absence de PR-Set7⁵³.

PR-Set7 appartient, comme GLP, à la famille des protéines contenant un domaine SET³⁸. Comme indiqué précédemment, il est responsable du transfert du groupe méthyle du cofacteur SAM. Le repliement de 3 séquences, n-SET, I-SET et c-SET permet un rapprochement du site de liaison de la SAM et de celui de la lysine pour former un site actif permettant le transfert du groupement méthyle sur H4K20 dans le cas de PR-Set7 (Figure 11).

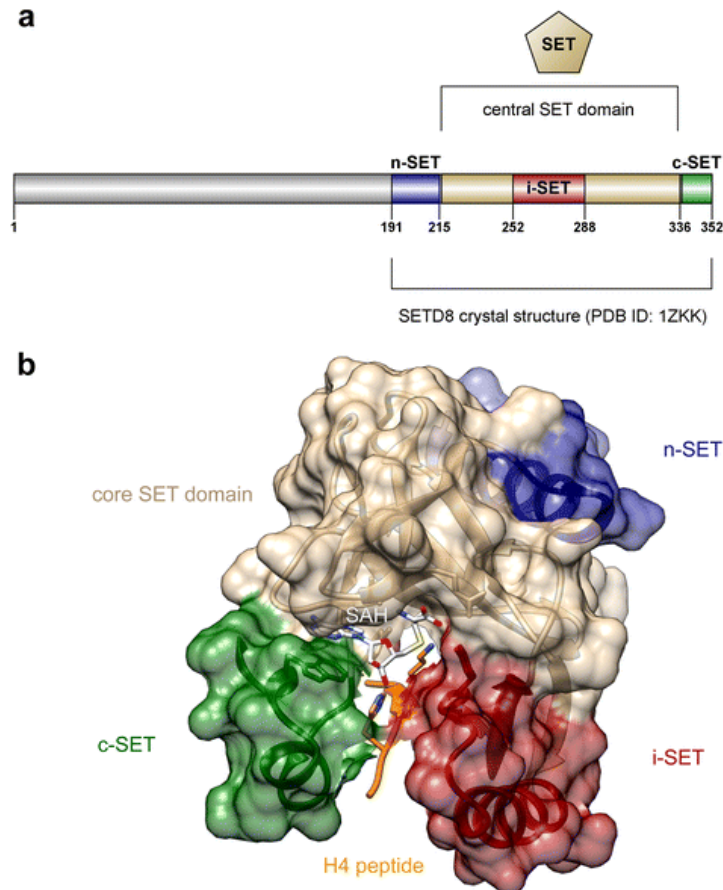


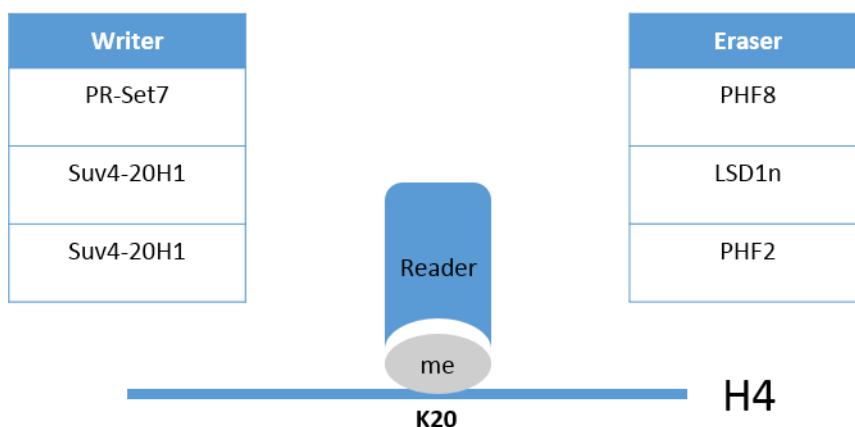
Figure 11. Structure de la protéine PR-Set7 d'après Milite et al. 2016⁵⁴. a) Schéma montrant la localisation des régions n-SET, I-SET et c-SET en C-terminal de la protéine. b) Structure cristalline du domaine SET de PR-Set7 montrant la lysine 20 du peptide de l'histone H4 (orange), la SAH (blanc), les domaines n-SET (bleu), i-SET (rouge) et c-SET (vert).

H4K20me1 est une marque qui est généralement associée à la répression de la transcription des gènes et à un état condensé de la chromatine chez les mammifères^{55,56}. Elle joue aussi un rôle dans les mécanismes de réplifications de l'ADN⁵⁷, dans les mécanismes de réparations des dommages à l'ADN⁵⁸⁻⁶⁰ ou encore dans certaines voies de signalisations liées au cancer^{61,62}

PR-Set7 joue un rôle dans divers processus cellulaires incluant la progression de la phase S⁶³, la mitose⁶⁴ et les mécanismes de contrôle des dommages à l'ADN⁶⁵. La méthylation de H4K20 par PR-Set7 est la marque apposée par cette enzyme la plus étudiée, mais il est possible que ce ne soit pas la plus importante. En effet, la contribution relative de la marque H4K20 peut

maintenant être évaluée en déterminant si le phénotype résultant de la suppression de PR-Set7 est obtenu également par la mutation de son résidu cible sur l'histone. Ainsi, des études de *knockout* ont montré que le mutant H4A20 engendrait un phénotype différent de celui observé suite à la délétion de PR-Set7⁶⁶. Chez la Drosophile, la délétion de l'enzyme est létale passée la troisième étape du développement de la larve⁶⁷ tandis que chez les mammifères elle provoque des dommages de l'ADN et des défauts du cycle cellulaire^{53,68} et entraîne la mort de l'embryon entre les étapes de développement embryonnaire 4 et 8 cellules⁵¹. L'activité essentielle de PR-Set7 ne pouvant être expliquée par la seule modification de H4K20 il semble fort probable que cette enzyme agit sur des protéines non-histones encore inconnues et que cette activité a un rôle essentiel.

Pour finir, la marque H4K20 dont la première protéine *writer* est PR-Set7, possède également diverses protéines *reader* et *eraser*. Ainsi plusieurs enzymes ont montré une capacité à déméthylater H4K20. De façon intéressante, la déméthylation de ce résidu par ces enzymes est dépendante du degré de méthylation. Ainsi, PHF8, première déméthylase de H4K20 découverte n'est capable de déméthylater que la monométhyl-lysine^{69,70}. LSD1n, une isoforme de l'enzyme LSD1 qui est exclusivement exprimée dans les neurones, ne déméthylerait aussi, *in vivo*, que la forme mono-méthylée⁷¹. Enfin l'enzyme PHF2 est capable de déméthylater la triméthyl-lysine⁷². L'ensemble des protéines connues à ce jour pour interagir avec H4K20 méthylée sont inventoriées Figure 12, dans l'objectif d'avoir une vision d'ensemble des différents interactants qui posent, qui lisent et qui suppriment la modification d'histone. Pour chacune des protéines *readers*, la fonction de la protéine effectrice est mentionnée, soulignant un rôle crucial de la méthylation de H4K20 en tant que régulatrice des mécanismes de réponse liés aux dommages à l'ADN.



Reader			
Name (modification)	Domaine	Fonction	Ref
TP53BP1 (H4K20me1/2)	Tudor	Réparation des dommages à l'ADN	58
L3MBTL1 (H4K20me1/2)	MBT	Répression de transcription	73
L3MBTL2 (H4K20me1/2)	MBT	Répression de transcription	74
MSL3 (H4K20me1)	Chromo-domaine	Répression de transcription, remodelage de la chromatine	75
JMJD2A (H4K20me2/3)	Tudor	Déméthylation, régulation transcriptionnelle, dommages à l'ADN	76,77
ORC1 (H4K20me2)	BAH	Licensing de réplication de l'ADN	78
LBR (H4K20me2)	Tudor	Formation de l'hétérochromatine à la membrane nucléaire interne	79
PHF20 (H4K20me2)	Tudor		80
ORCA (H4K20me3)	WD40	Licensing de réplication de l'ADN	81

Figure 12. Protéines interagissant avec H4K20 méthylée chez *Homo sapiens*.

Substrats non histone de PR-Set7

A ce jour plusieurs protéines autres que l'histone H4 sont connues pour être méthylées par PR-Set7. La méthylation de p53 sur la lysine 382 (p53K382me1) permet notamment la suppression de l'activité de transcription dépendante de p53 dans les cellules cancéreuses⁶². PR-Set7 méthyle aussi PCNA, un marqueur de la progression cancéreuse, sur sa lysine 248 (PCNAK248me1). Cette modification a un effet stabilisateur en inhibant sa poly-ubiquitinylation ce qui favoriserait la tumorigénèse⁶¹. La protéine Numb a aussi été proposée comme substrat méthylé par PR-Set7 sur les lysines en positions 158 et 163. La protéine Numb favorise l'apoptose de façon dépendante de p53. Sa méthylation abolirait l'interaction de Numb et p53 induisant un accroissement du taux d'ubiquitinylation de p53 et engendrerait sa dégradation⁸². Dans le cas de ce substrat, notons que le site de méthylation diffère fortement du site consensus de l'enzyme et que sa méthylation par PR-Set7 est sujet à controverse⁸³.

Encore d'autres rôles de PR-Set7 ont été montrés sans pour autant mettre en évidence de méthylation de substrats spécifique comme dans la progression du cycle cellulaire ou les mécanismes de réparation des dommages à l'ADN⁵⁴.

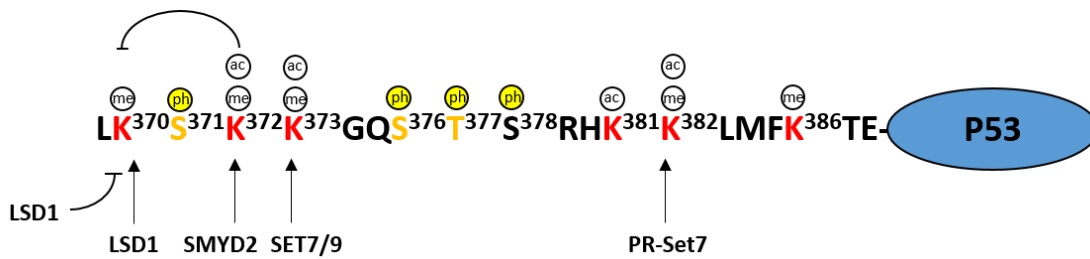
1.2.2 Méthylation des protéines non-histones : la face cachée de l'iceberg ?

Ainsi que nous avons commencé à le voir avec les exemples de PR-Set7 et GLP, certaines HKMT sont capables de méthyler des protéines non-histones. On peut légitimement se demander dans quelle mesure les histones représentent la cible principale de ces enzymes. En réalité, la caractérisation des HKMT initialement en tant qu'enzymes modifiant les histones est liée à la plus simple caractérisation de ces dernières. En effet l'abondance des histones dans la cellule et le nombre important de résidus basiques dans leur séquence, explique qu'elles aient été les premiers substrats découverts des HKMT. Ensuite, les approches biaisées de recherche de nouveaux substrats, accompagnées des difficultés de développement d'outils de protéomique permettant une exploration à grande échelle des substrats d'HKMT, ont largement retardé l'identification de nouvelles cibles. Il faut en effet souligner que contrairement à d'autres modifications post-traductionnelles, comme l'acétylation qui change la charge globale ou la phosphorylation qui peut être étudié de par sa réactivité chimique et pour laquelle des méthodes d'enrichissement sont possibles, la méthylation est plus complexe à étudier, ne changeant pas la charge et le méthyle étant un groupement peu réactif chimiquement.

Avec le développement de techniques comme les puces à peptides⁸⁴ ou l'utilisation d'anticorps spécifiques⁸⁵ couplés à la grande sensibilité de la spectrométrie de masse, le nombre

d'événements de méthylation de lysines identifiés a explosé permettant d'appréhender cette PTM, au même titre que l'acétylation ou la phosphorylation, comme une marque extrêmement répandue^{86,87}. De plus, la découverte de protéines *readers* et *erasers* capables de reconnaître les lysines méthylées de substrats non histone suggère une fine régulation de l'activité des cibles méthylées (comme dans le cas des histones décrit précédemment). C'est notamment le cas de la protéine suppresseur de tumeur p53 dont le mode de régulation *via* la méthylation rappelle fortement la régulation des queues d'histones (Figure 13). En effet, p53 a été identifiée comme étant méthylée par 5 HKMT : SMYD2⁸⁸, SET7/9⁸⁹, PR-Set7⁶², G9a et GLP^{50,90}. De plus, chaque méthylation d'une lysine induit un mécanisme réalisé grâce à l'action de différents partenaires. Nous avons déjà cité l'exemple de PR-Set7 qui monométhyle p53K382. Cette marque recrute la protéine L3MBTTL1 pour réprimer les gènes cibles de p53⁹¹. Un deuxième exemple est le recrutement de TP53BP1 sur p53K370 diméthylé par SETDB1, qui promeut à l'inverse l'expression des gènes cibles de p53⁹². Par ailleurs, comme avec la méthylation des histones, l'état de méthylation peut influencer sur la voie biologique déclenchée. Ainsi la mono- et la di-méthylation de p53K370 induisent respectivement une inhibition ou une activation de la transactivation des gènes cibles de p53^{88,93}. Enfin, une inter-régulation a été mise en évidence entre deux méthylations de p53. En effet, il a été montré que la mono-méthylation de p53K372 inhibait celle sur K370 en diminuant l'affinité de SMYD2 pour p53, inhibant par conséquent l'effet inhibiteur de p53K370me1. Ce qui revient à activer p53⁸⁸.

L'ensemble de ces régulations de p53 dévoile une grande complexité de la régulation de la protéine médiée par la méthylation des lysines. La participation de partenaires et les inter-régulations entre marques épigénétiques présentes sur les protéines existent ainsi de façon semblable à ce qui a été observé plus tôt chez les histones et conceptualisé sous le nom de code histone. L'ensemble des études sur la méthylation des lysines des protéines non-histones suggèrent en outre que ce processus est aussi un moyen de régulation fine de l'activité des protéines et de modulation de leur stabilité, de leur localisation cellulaire et de leur interaction avec d'autres protéines, que celles-ci soient en lien ou non avec la transcription génique.



Inter-régulation	Caractéristique	Ref	Site de méthylation	HKMT	Fonction	Ref
P53k370me1	Exclusion	88	K370me2	SETDB1	P53 activation	88
P53k372me1	Dans le sens K372 sur K370		K370me1	SMYD2 méthylation, LSD1 déméthylation	P53 inhibition	88,93
			K372me1	Set7/9	Augmentation de la stabilité de p53. Activation transcriptionnelle	89
			K373me2	G9a/GLP	Activation de transcription des gènes cibles de p53	50,90
			K382me1	PR-Set7	P53 inhibition	62

Figure 13. Modifications post-traductionnelles introduites sur la partie C-terminale de p53. Les séquences d'acides aminés sont annotées pour souligner les résidus modifiés importants. Les lignes noires représentent les événements d'inter-régulation existants chez Homo Sapiens. Les tableaux résumés l'ensemble des interactions publiées ainsi que l'effet de chacun des états de méthylation connus au sein de ce domaine de p53.

I.3. Méthodologies pour le profilage de la méthylation sur les lysines

Avant que ne soient développées des méthodes de détection des événements de méthylation à grande échelle, les études de méthylation des histones et autres substrats se heurtaient à un obstacle méthodologique majeur, l'essentiel de ces études utilisant des approches basées sur des candidats supposés. Cependant, des travaux récents se sont centrés sur la problématique de l'identification de la marque de méthylation à grande échelle. Ainsi différentes méthodologies

permettant un profilage de la méthylation ont successivement vu le jour. Nous en présentons ci-dessous les grandes lignes de quelques-unes.

1.3.1 Les puces à peptides et puces à protéines

Les puces à peptides font partie des premières innovations et ont permis d'identifier bon nombre de substrats d'HKMT⁹⁴ et à définir un site consensus pour une enzyme donnée. Cette approche consiste à immobiliser un grand nombre de peptides de séquences distinctes à la surface d'un support solide plan (plaque de verre, membrane de nitrocellulose...). Ces peptides peuvent alors mimer efficacement des régions non structurées de protéines⁹⁴. Ces supports ainsi fonctionnalisés sont incubés en présence d'une HKMT et de SAM. Une méthode de détection appropriée, par exemple, la radioactivité avec l'utilisation de SAM tritié sur le méthyle, permet de caractériser un peptide marqué, de remonter à sa séquence et au site de méthylation. A titre d'exemple, on pourra citer les travaux d'A. Jeltsch qui en utilisant des puces à peptides, a défini un site consensus pour l'enzyme G9a (méthyltransférase de H3K9) : il s'agit du motif court « RK »⁴⁸. Par la suite, en cherchant dans des bases de données de protéines, de nouveaux candidats substrats de G9a ont pu être identifiés, testés puis validés. Des substrats précédemment caractérisés par des approches basées sur de potentiels candidats (ex : autométhylation de G9a) ont ainsi pu être retrouvés confirmant la robustesse de la méthode^{46,47}. La même méthode a aussi permis la caractérisation des sites consensus des enzymes Set7/9⁹⁵ et PR-Set7⁸⁴ et Dim-5⁹⁶ ainsi que d'identifier de nouveaux substrats de Set7/9.

Bien que cette méthodologie soit relativement simple à mettre en place, elle possède plusieurs limites. Un des inconvénients des puces à peptides est qu'elles n'exposent à leur surface que des séquences minimales méthylables par une HKMT, qui ne contiennent donc pas de potentiels motifs de reconnaissance pouvant être nécessaires à l'enzyme ou qui peuvent ne pas être correctement structurées du fait de leur fixation sur un support solide. Elles sont par conséquent aussi éloignées de conditions natives qu'on puisse l'être. Les puces à protéines permettent de pallier ce problème, même si l'on reste dans des conditions *in vitro*. Malgré tout, de telles approches ne sont pas idéales car elles nécessitent la fixation des protéines sur une surface (la puce), cette contrainte pourrait notamment rendre inaccessible le site de méthylation ou empêcher la protéine de se replier correctement.

Le site consensus de l'enzyme GLP tel qu'il a été déterminé par l'équipe de Jeltsch a pour point de départ la séquence de l'enzyme H3 contenant la lysine 9⁴⁸. Il a pu être identifié par les auteurs en synthétisant toutes les combinaisons de peptides variant d'un unique résidu par rapport à la

séquence de départ. Le choix des peptides immobilisés sur le support est particulièrement important lors de la génération des puces et une génération plus aléatoire des séquences des peptides pourrait être cruciale dans le cas d'étude d'HKMT avec une séquence consensus plus grande. En effet, l'équipe de Jeltsch, a défini pour PR-Set7, le long consensus RH(RKY)K(VILFY)(LFY)R sans pour autant pouvoir valider de nouveaux substrats. Cette contrainte dans la génération des puces pourrait alors ne pas permettre d'explorer toute la spécificité de substrat d'une enzyme donnée.

1.3.2 Les anticorps anti pan-méthyl-lysines couplés à la spectrométrie de masse

La spectrométrie de masse est un formidable outils pour étudier les PTM à l'échelle du protéome. Ceci étant, elle présente quelques inconvénients. Elle a tendance à permettre plus facilement l'identification des protéines les plus abondantes dans un échantillon, ce qui peut générer un biais. De plus, étant donné la faible abondance des marques épigénétiques, les événements de méthylation induits par les HKMT à domaine SET sont difficilement détectables. Afin de contourner cette difficulté, une étape d'enrichissement des protéines méthylées est nécessaire. Ceci peut notamment être effectué grâce à l'utilisation d'anticorps spécifiques d'un état de méthylation indépendamment de la séquence entourant la lysine méthylée. Ce sont les anticorps anti pan-methyl-lysines. L'utilisation de méthodes utilisant la capacité d'anticorps de reconnaître un antigène spécifique présente alors plusieurs avantages. Elle permet d'identifier le site ainsi que l'état et les taux de méthylation. Cette méthode permet alors de mesurer l'activité d'une HKMT dans des conditions physiologiques. Cette stratégie découle des stratégies pour le profilage d'autre marques : phosphorylation⁹⁷, ubiquitinylation⁹⁸, acétylation⁹⁹ mais il est difficile de l'adapter à la méthylation en raison d'une difficulté de production d'anticorps spécifiques à la méthylation des lysines. Pour l'exemple, dans une des premières études du genre, les anticorps ne faisaient pas la distinction avec des arginines méthylées⁸⁵. De plus, une étude de la spécificité de plusieurs anticorps commerciaux anti pan-méthyl-lysines utilisant des puces sur lesquelles ont été greffés des peptides possédant des marques épigénétiques variées, a révélé que ces anticorps pouvaient être affins des marques autres que la méthylation¹⁰⁰. Ce type de technique utilisant des anticorps peu spécifiques comporte donc certaines faiblesses même si elle a permis à ce jour l'identification de nombreux substrats méthylés¹⁰¹.

1.3.3 L'affinité de domaine couplée à la spectrométrie de masse

Une troisième approche pour identifier les événements de méthylation et finalement assez similaire à celle utilisant les anticorps, repose sur l'utilisation de protéines affines pour les méthyl-lysines pour enrichir les cibles méthylées par immunoprécipitation. Il s'agit donc d'utiliser des protéines *reader* possédant des domaines de reconnaissance spécifiques. C'est avec cette méthode et en utilisant une des protéines de la famille HP1 (HP1 β) possédant un chromodomaine que l'équipe de Liu parvint à identifier 29 substrats méthylés interagissant avec cette protéine *reader*¹⁰². Peu de temps après, l'équipe de Moore, utilisa les domaines MBT de la protéine L3MBTL1 et établit que cette protéine se liait exclusivement aux lysines mono- ou diméthylées¹⁰³. Les auteurs identifièrent par la suite plusieurs centaines de protéines potentiellement méthylées et en validèrent 26¹⁰³.

Mises ensemble les expériences utilisant ces méthodes ont permis d'attribuer un rôle important de la méthylation dans des phénomènes tels que la transcription, le remodelage de la chromatine ou l'activité hélicase¹⁰¹⁻¹⁰⁴. Néanmoins, l'utilisation de ces protéines affines de la méthylation des lysines ne permet d'avoir qu'un petit aperçu de l'ensemble du méthylome. De plus, ce n'est pas une méthode qui permet de savoir par quelle enzyme, la marque a été déposée.

1.3.4 L'enrichissement par voie chimique couplé à la spectrométrie de masse

Au cours des deux dernières décennies, de nouvelles méthodes basées sur la chimie ont permis de faire des avancées dans le domaine de l'étude des PTM. L'étude des substrats de diverses transférases, grâce à l'utilisation de petites molécules de synthèse, couplée aux méthodes de protéomique a été développée ces dernières années, et les méthyltransférases ont bénéficié de ces avancées en s'appuyant sur les réactions bioorthogonales.

Les réactions bioorthogonales

En 2001, Kolb, Finn et Sharpless ont défini le concept de « chimie click » et un ensemble de règles qui seront une des bases du développement d'outils chimiques particulièrement bien adaptée à la bio-conjugaison. L'idée est d'utiliser un petit nombre de « blocs », qui vont pouvoir être combinés de manière la plus simple possible et avec les critères suivants : la réaction doit être rapide, doit présenter de très bons rendements, les sous-produits doivent être non réactifs, non toxiques, et pouvoir être éliminés facilement, la réaction doit être régiosélective, les

conditions réactionnelles doivent être simples et faciles à mettre en œuvre, les solvants doivent être non-toxiques (idéalement l'eau)¹⁰⁵.

Depuis une vingtaine d'années, ce concept est utilisé dans de nombreuses réactions réalisées dans des systèmes biologiques via des réactions dites bioorthogonales, terme introduit par C. Bertozzi¹⁰⁶. Les réactions bioorthogonales sont des réactions chimiques qui ont lieu dans des milieux biologiques complexes de façon sélective entre des fonctions chimiques absentes dans les molécules biologiques. Ces fonctions chimiques très réactives et sélectives entre elles présentent une faible réactivité vis-à-vis des fonctions chimiques présentes dans les molécules biologiques et répondent aux critères de la chimie click. Lorsque la réaction est effectuée *in vivo*, tous les composants impliqués doivent être non toxiques, ou au moins inertes vis-à-vis des milieux biologiques.

Les deux dernières décennies ont vu naître toute une série de réactions chimiques bioorthogonales basées entre autres sur des réactions plus anciennes et très utiles pour l'étude de processus biologiques. Parmi ces réactions chimiques, la cycloaddition 1,3-dipolaire entre une fonction azoture et un alcyne vraie catalysée par le cuivre (I) (**CuAAC**, Figure 14) et conduisant sélectivement au dérivé 1,4-triazole a été découverte indépendamment par les équipes de K.B. Sharpless et M. Meldal en 2002^{107,108}. La même année, le groupe de Sharpless applique cette réaction de cycloaddition entre un azoture et un alcyne à la formation *in situ* d'inhibiteurs d'une enzyme cible,¹⁰⁹ l'enzyme servant de *template* pour l'assemblage des inhibiteurs (effet de proximité), l'ajout de cuivre (I) n'est ici pas requis, mais cette stratégie n'est pas généralisable pour la bioconjugaison. Quelques années plus tard l'équipe de C. Bertozzi propose une évolution de la réaction chimique s'affranchissant du catalyseur au cuivre, potentiellement toxique pour une utilisation *in vivo*. Elle utilise alors la réaction entre une cyclooctyne et un groupement azoture dans laquelle la réaction est favorisée par la contrainte du cycle (**SPAAC**, Figure 14)¹¹⁰.

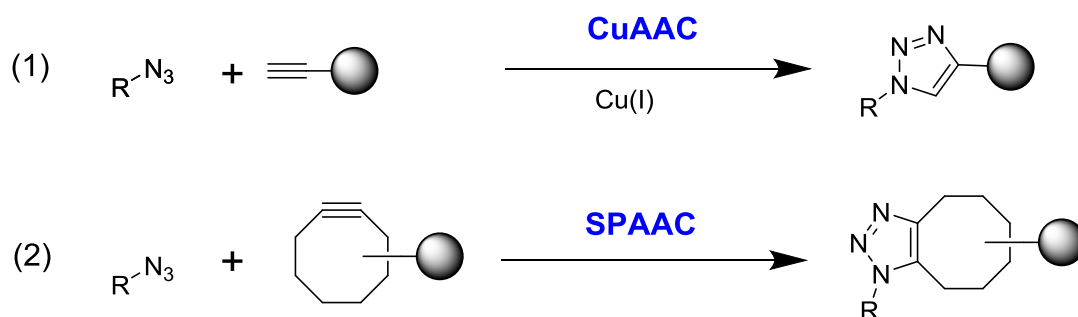


Figure 14. (1) Cycloaddition alcyne azoture catalysée par le cuivre (I) (CuAAC). (2) Cycloaddition alcyne azoture favorisée par l'alcyne contraint (SPAAC)

Appliquée à l'étude de la méthylation des lysines, l'utilisation de ces réactions a permis l'émergence d'analogues de SAM porteurs d'une fonction chimique bioorthogonale (un alcyne ou un azoture) pouvant être transféré en lieu et place d'un groupement méthyle par l'action d'une méthyltransférase¹¹¹. En effet, le méthyle est un petit groupement chimique, inerte qui ne peut être fonctionnalisé chimiquement pour y introduire une sonde exogène. En introduisant via l'enzyme étudiée, non pas un méthyle sur le substrat mais un substitut de méthyle fonctionnalisable et bioorthogonal on contourne cette contrainte. Une sonde d'affinité complémentaire du groupement chimique introduit peut, au cours d'une seconde étape, être utilisée afin de permettre une étape d'enrichissement compatible avec une analyse par MS.

Dans ce contexte l'équipe de Minkui Luo a décrit une stratégie utilisant des analogues de SAM porteur d'une fonction alcyne transférable par une HKMT et a réussi à identifier plusieurs centaines de substrats des HKMT G9a et GLP puis à en valider biologiquement¹¹². En comparaison des autres techniques, cette méthode a deux gros avantages. Le premier réside dans la capacité de pouvoir identifier les substrats d'HKMT avec certitude malgré la grande redondance des HKMT entre elles. De plus, la fonction transférée échappe aux voies de déméthylation et permet de contourner la réversibilité de la méthylation des protéines. C'est dans ce cadre que ce projet de thèse s'inscrit et l'application de cette stratégie à notre enzyme d'intérêt, PR-Set7 fera l'objet du chapitre II.

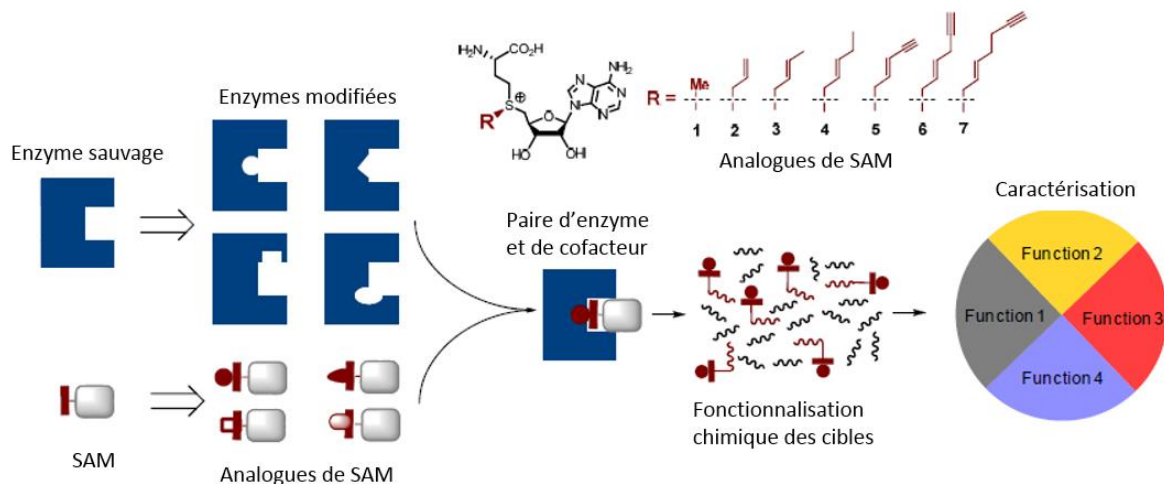


Figure 15. Schéma descriptif de la stratégie d'identification de nouveaux substrats d'HKMT. D'après Luo, 2013¹¹². Des HKMT modifiées et des analogues de SAM porteurs d'un groupement bioorthogonal (ici un alcyne) forment une paire capable de fonctionnaliser les substrats de l'HKMT. Ce qui permet in fine de caractériser ces derniers.

Pour résumer :

Méthode	Avantages	Inconvénients
Puces à peptides et Puces à protéines	<ul style="list-style-type: none"> • Génération de sites consensus de méthylation • Méthode simple • Possible utilisation de protéines entières 	<ul style="list-style-type: none"> • Conditions non physiologiques • L'enzyme peut nécessiter la présence de cofacteur dans les conditions <i>in vitro</i> • Les enzymes n'ont pas nécessairement d'activité sur peptide, • Séquence protéique incomplète ou protéines mal structurées •
Anticorps anti pan-methyl-lysines - MS	<ul style="list-style-type: none"> • Taux de méthylation physiologiques 	<ul style="list-style-type: none"> • Anticorps non fiables car peu spécifiques
Affinité de domaine - MS	<ul style="list-style-type: none"> • Taux de méthylation physiologiques • Reproductible, réactif d'affinité monoclonal modifiable 	<ul style="list-style-type: none"> • Identification de peu de sites de modifications post-traductionnelles
Affinité chimique - MS	<ul style="list-style-type: none"> • Taux de méthylation physiologiques • Possibilité d'identifier un substrat même en cas d'activité redondante entre les HKMT • Contourne le problème de la réversibilité de la méthylation 	<ul style="list-style-type: none"> • L'enzyme doit pouvoir accepter le cofacteur modifié

Projet de thèse

La méthylation des histones sur leurs lysines par des méthyltransférases (HKMT) est une modification post-traductionnelle qui joue un rôle crucial dans de nombreux processus biologiques. Néanmoins, une dérégulation de l'activité de ces enzymes est souvent associée à des pathologies telles que des cancers ou des maladies neurodégénératives. Les HKMT sont de ce fait, des cibles thérapeutiques intéressantes. La recherche de petites molécules inhibitrices de certaines HKMT est d'ailleurs un champ de recherche très actif et certains inhibiteurs ont atteint des stades d'études cliniques mais il faut souligner que l'on est encore loin de connaître tous les impacts de telles molécules sur d'autres cibles que les histones elles-mêmes et que la caractérisation du méthylome de ces enzymes est un enjeu important

Parmi les HKMT, PR-Set7 est une mono-méthyltransférase de la lysine 20 de l'histone H4, H4K20me1 étant associée à un état compacté de la chromatine. Plusieurs études ont montré que la délétion de PR-Set7 interrompait le développement embryonnaire très précocement alors que la seule absence de méthylation de H4K20 engendrait un retard de croissance mais n'était pas létale. Sachant que PR-Set7 est la seule méthyltransférase pouvant monométyler H4K20 et que son activité est requise par d'autres HKMT hyperméthylant cette même lysine, ce phénotype plus sévère observé en l'absence de l'enzyme suggère que des substrats non-histones de PR-Set7 restent à découvrir. L'objectif principal de cette thèse consiste donc à mettre en place une approche permettant d'identifier de nouveaux substrats de PR-Set7.

Comme nous l'avons vu précédemment, de nombreuses méthodes ont permis jusqu'ici d'identifier de nombreuses cibles d'HKMT de nature non-histone. Ces études s'accumulant, elles tendent à montrer que la méthylation de lysines de protéines est une marque répandue et que les HKMT n'ont rarement que les histones comme substrat. Toutefois, chaque approche comporte ses avantages et inconvénients et il n'existe pas de méthode universelle permettant de caractériser de façon exhaustive tous les substrats d'une HKMT particulière. Par ailleurs, il faut souligner que certaines études ont conduit à l'identification de substrats qui n'ont jamais été validés *in vivo*.

Ainsi, nous avons mentionné que quelques substrats de PR-Set7 ont déjà été identifiés par des approches de puces à peptides notamment. Il nous a semblé intéressant de mettre en œuvre ici une approche qui n'a pas été explorée jusqu'à présent pour cette enzyme. Dans ce contexte et dans l'objectif de caractériser de nouveaux substrats de PR-Set7, nous avons mis en œuvre une approche s'inspirant de la stratégie de M. Luo, dite « bump and hole »^{111,113} basée sur

l'utilisation d'outils chimiques et enzymatiques pour étiqueter et identifier les substrats protéiques de cette enzyme.

Mon projet se décompose en 4 grands axes :

- Le premier correspond à la conception et la synthèse d'analogues du cofacteur naturel de l'enzyme comportant une fonction chimique bioorthogonale, ainsi qu'à la conception et la production de mutants de PR-Set7 susceptibles d'accueillir ces molécules.
- Le second axe est le criblage de la série d'analogues de SAM envers les différents mutants de PR-Set7 conçus en vue d'identifier une paire cofacteur-enzyme fonctionnelle. Pour cela, le modèle d'étude se base sur le principal substrat connu de PR-Set7 : l'histone H4.
- Le troisième axe porte sur la mise en œuvre de l'approche de fonctionnalisation et d'enrichissement des substrats de PR-Set7 à l'aide du couple enzyme-cofacteur le plus approprié, afin d'identifier des protéines candidates lors d'une étude de protéomique par spectrométrie de masse.
- Pour finir, le dernier axe vise à apporter des données structurales clés permettant de rationaliser les résultats et utilise la modélisation moléculaire, en tenant compte des résultats obtenus au cours de cette thèse.

La figure 16 schématise l'approche mise en œuvre au cours de cette thèse et sera présentée plus en détail dans les quatre chapitres suivants.

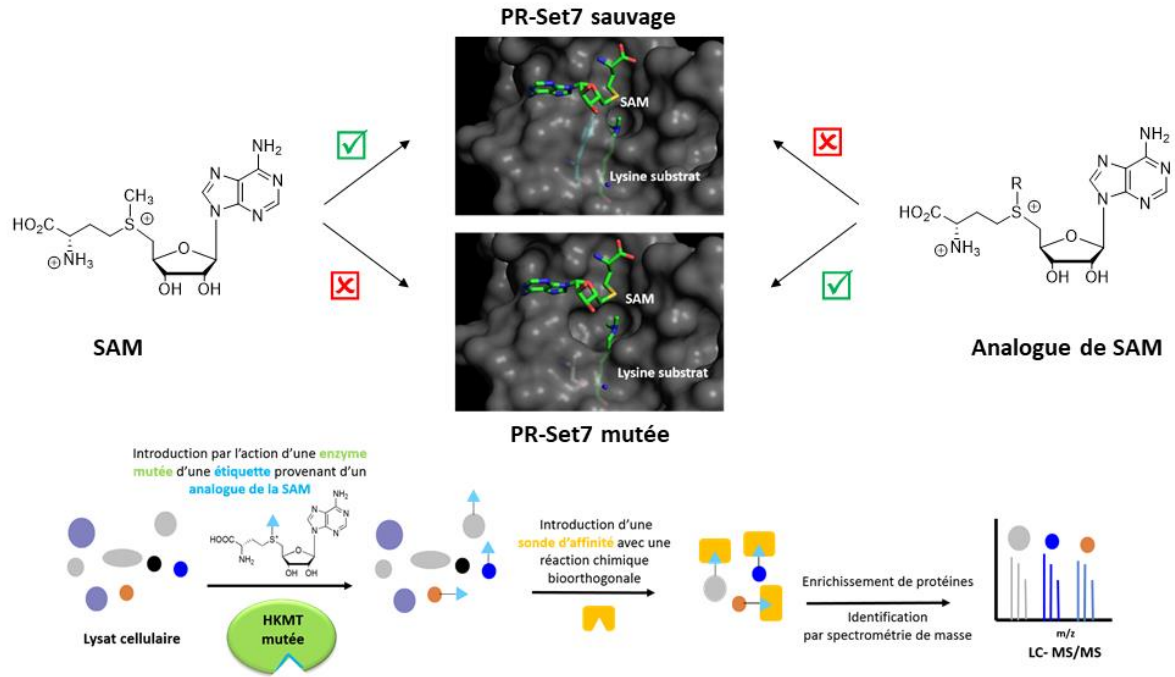


Figure 16. Schéma récapitulatif du projet de thèse. Un analogue du cofacteur SAM devra être pris en charge par un mutant de PR-Set7 approprié. La paire mutant d'enzyme - analogue de cofacteur efficace sera alors utilisée pour fonctionnaliser enzymatiquement les substrats de PR-Set7 présents dans des extraits cellulaires. Puis, une sonde d'affinité sera introduite par une réaction bioorthogonale uniquement sur les protéines ainsi étiquetées. Cette étiquette permettra l'enrichissement puis l'identification des substrats de PR-Set7 par spectrométrie de masse.

Chapitre 2

Conception et mise au point des outils moléculaires et enzymatiques pour la stratégie de marquage des substrats de PR-Set7

II. Conception et mise au point des outils moléculaires et enzymatiques pour la stratégie de marquage des substrats de PR-Set7

La stratégie de marquage chémoenzymatique des substrats de PR-Set7 que nous avons choisie nécessite *i)* la synthèse d'analogues de SAM comportant, au niveau du sulfonium, un substitut de méthyle porteur d'une fonction bioorthogonale, *ii)* la conception et production de mutants de PR-Set7 susceptibles de prendre en charge ces cofacteurs. Dans ce chapitre, je décrirai ces deux aspects préliminaires à la mise en place de l'approche de marquage chémoenzymatique.

II.1 Conception d'une série d'analogues de SAM

II.1.1. La *S*-adénosyl-L-méthionine : un cofacteur enzymatique clé

La ***S*-adénosyl-L-méthionine (SAM ou AdoMet)** est le deuxième substrat enzymatique le plus courant après l'ATP chez les organismes vivants¹¹⁴. Elle est biosynthétisée à partir de la L-méthionine et de l'ATP par une réaction de substitution nucléophile catalysée par la méthionine adénosyltransférase, MAT, aussi appelée SAM synthétase¹¹⁵. Dans cette réaction, l'attaque du soufre de la méthionine entraîne le départ du groupement triphosphate de l'ATP conduisant à la formation de la SAM (Figure 17). Cette réaction est stéréospécifique et génère uniquement la configuration (*S*) au niveau du sulfonium. Dans la cellule, seul l'isomère (*S*) de la SAM est actif et pris en charge par les méthyltransférases^{116,117}.

La SAM est un acteur clé dans les réactions de méthylation qui ont une grande importance biologique avec notamment des implications dans la régulation épigénétique, la modulation de la fonction des protéines, la synthèse et la dégradation des métabolites et des molécules de signalisation. Il s'agit du principal donneur de méthyle de la cellule agissant sur différents substrats nucléophiles azotés, oxygénés, soufrés, carbonés, en présence de diverses méthyltransférases. Ce cofacteur va permettre la méthylation de substrats variés tels que les protéines, l'ADN, l'ARN, les lipides, mais également de petites molécules. La réaction s'effectue par un mécanisme de type S_N2 avec une attaque nucléophile sur le groupe méthyle adjacent au centre sulfonium¹¹⁸. Seul l'épimère (*S*) naturel est utilisé préférentiellement par la plupart des enzymes.

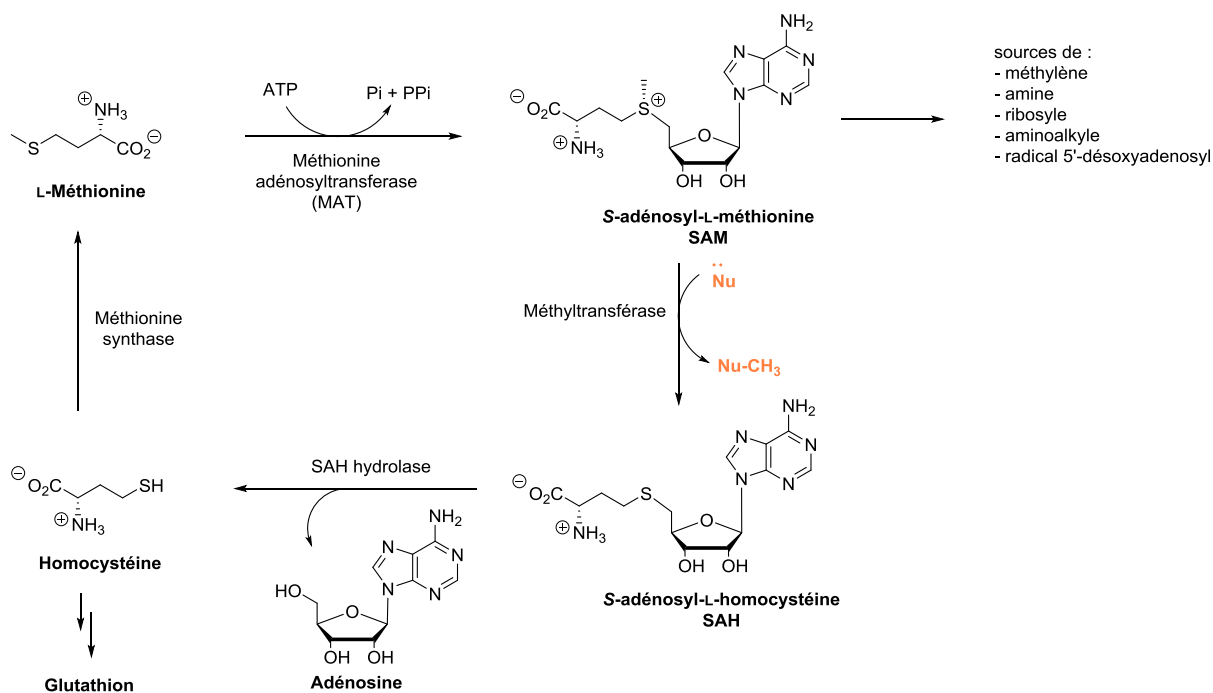


Figure 17. Biosynthèse de la SAM, utilisation par les méthyltransférases, autres rôles de la SAM et recyclage.

Outre le produit méthylé, les méthyltransférases génèrent la *S*-adénosyl-L-homocystéine (AdoHcy ou SAH). La SAH est hydrolysée par la SAH hydrolase pour donner l'adénosine et l'homocystéine qui est, soit convertie en glutathion, soit méthylée pour générer de la méthionine, par une réaction catalysée par la méthionine synthase dépendante de la méthylcobalamine¹¹⁹. Il convient de préciser par ailleurs que la SAM est un cofacteur multifonction impliqué dans un grand nombre de mécanismes enzymatiques que nous ne détaillerons pas ici. On peut citer son rôle comme source de groupes méthylène, amino, ribosyle, et aminopropyle. En outre, la SAM, après réduction à un électron, est une source de radicaux désoxyadénosyle qui initient de nombreuses réactions métaboliques et voies de biosynthèse par l'abstraction d'atomes d'hydrogène. Cette réactivité est due au caractère électrophile des carbones adjacents au soufre chargé¹¹⁷.

II.1.2. Panorama des analogues de SAM modifiés au niveau du méthyle

Les premiers analogues de la SAM, comportant un groupement éthyle ou propyle à la place du méthyle, ont été synthétisés par L. Parks¹²⁰ et F. Schlenk¹²¹ *via* une synthèse enzymatique. Bien que la prise en charge de ces analogues par les ADN méthyltransférases¹¹¹ et dans une moindre mesure les arginines méthyltransférases¹²² ait été démontrée, il a aussi été montré que le taux de méthylation diminuait avec la taille de la chaîne alkyle à transférer. Les résultats obtenus

avec ces molécules illustrent alors le fait que le transfert de chaînes alkyles longues quel que soit le substrat n'est pas chose aisée.

En 2006, l'équipe de Weinhold *et al.* synthétisa une série d'analogues de la SAM (Figure 18) par voie chimique. Ils obtinrent des molécules où le méthyle est remplacé par des fonctions alkyles insaturées, notamment un analogue avec un groupement allyle (Allyl-SAM) qu'ils testèrent vis-à-vis d'ADN méthyltransférases. Dans cette étude, il a été montré que la vitesse de la réaction pouvait être maintenue malgré l'encombrement stérique induit par la longueur de la chaîne alkyle¹¹¹ (Figure 18). Cette caractéristique est expliquée par une stabilisation de l'état de transition de la double liaison en position β du sulfonium. En effet le changement d'hybridation du carbone en α entraîne la formation d'une orbitale p qui peut se conjuguer avec l'orbitale antiliante π du carbone sp^2 adjacent (Figure 19). Cet effet de stabilisation est alors suffisant pour inverser l'effet défavorable de l'encombrement stérique lié à la longueur de la chaîne et permettre la transalkylation du substrat.

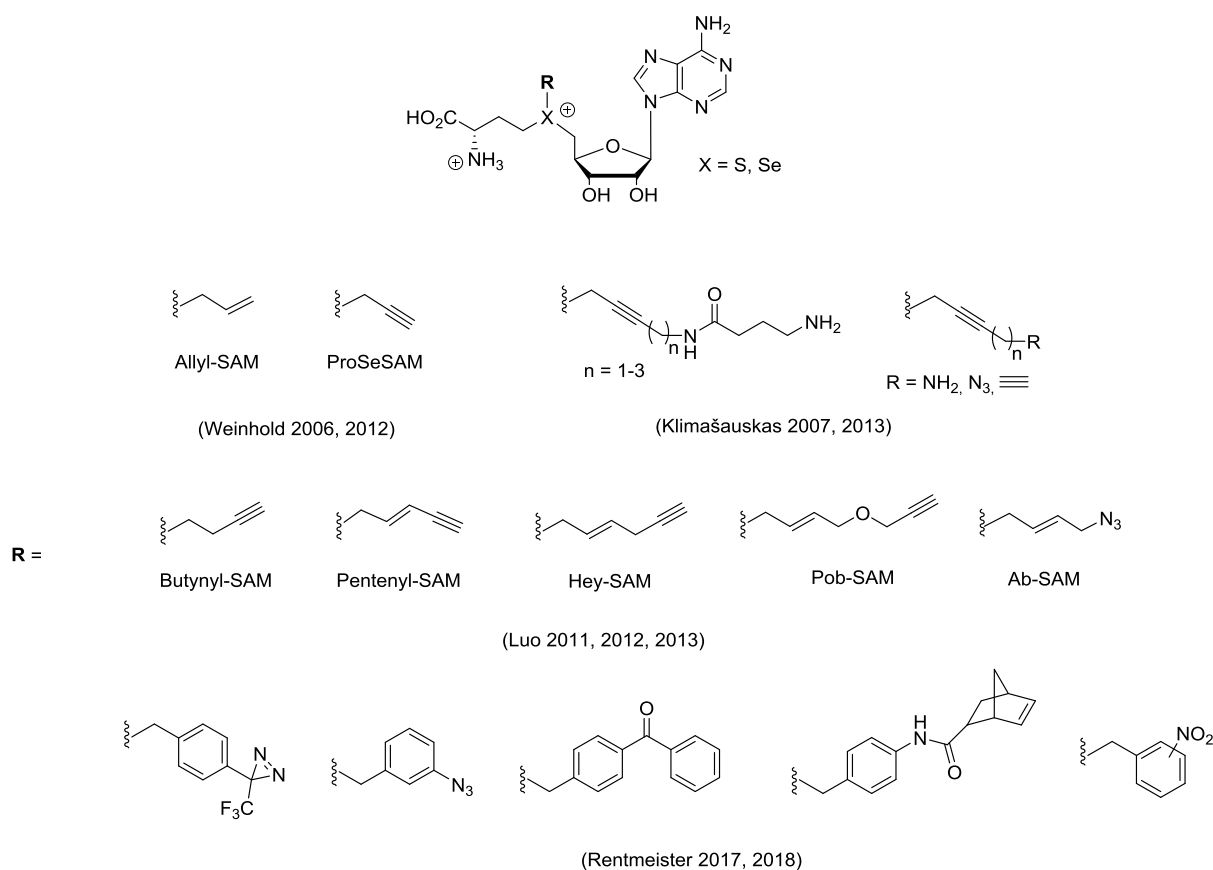


Figure 18. Principaux analogues de SAM

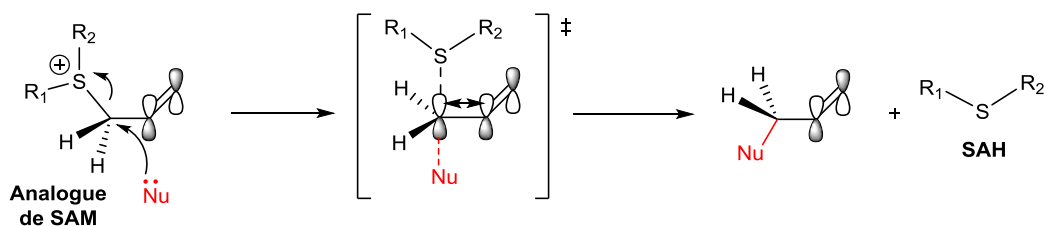


Figure 19. Stabilisation de l'état de transition en présence d'une double liaison en position β du sulfonium (ici avec l'analogue Allyl-SAM). Une triple liaison dans la même position induit le même effet. (d'après C. Dalhoff, 2006 ¹¹¹)

La preuve de concept de l'utilisation de ce type d'analogue de SAM étendu avec des lysine méthyltransférases est amenée en 2010 par l'équipe de E. Weinhold¹²³. Celle-ci montre que la méthyltransférase DIM-5 sauvage de *Neurospora crassa* est capable de transférer un groupement pent-2-en-4-ynyl sur la chaîne latérale de la lysine 9 de l'histone H3. L'analogue de SAM, donneur du groupement transféré sur K9 est nommé Pentenyl-SAM et a l'avantage d'avoir une fonction chimique bioorthogonale, l'alcyne terminal (Figure 18). Ce groupement peut alors servir à l'introduction d'une sonde détectable (biotine ou fluorophore) grâce à une réaction de cycloaddition alcyne-azoture catalysée par le cuivre (I) (CuAAC). L'utilisation d'une sonde biotine autorisant un enrichissement des substrats d'HKMT, l'utilisation de ce type d'analogue rend possible une nouvelle approche de caractérisation de leur spécificité de substrat.

Toutefois la réactivité de cette HKMT, comparée à d'autres envers cette molécule relève davantage de l'exception que de la règle. En effet, bien que la double liaison en position β du sulfonium favorise la catalyse, ce type d'analogue de cofacteur reste très volumineux par rapport à la SAM (Figure 19). Ainsi, les analogues de cofacteur à chaîne alkyle étendue ne sont quasiment jamais pris en charge par les HKMT.

Dans ce contexte, l'équipe de Minkui Luo a utilisé la stratégie dite « **bump and hole** » pour concevoir des mutants d'HKMT capables de prendre en charge ce type d'analogue de cofacteur. Initialement cette approche a été utilisée pour concevoir des inhibiteurs spécifiques de kinases d'une même famille ayant des sites actifs très conservés entre elles¹²⁴. Elle repose sur l'introduction de mutations ponctuelles au sein du site actif d'enzymes **E** naturellement inhibées par des molécules peu spécifiques **L**. Les mutants **E*** produits deviennent sensibles à l'inhibition par des petites molécules plus spécifiques **L*** analogues de **L** qui n'inhibaient initialement pas l'enzyme sauvage (Figure 20). L'approche a ensuite été utilisée sur une arginine méthyltransférase mutée et des analogues de SAM substitués par des groupements alkyles volumineux sur la position N⁶ constituant un premier système orthogonal incluant une

méthyltransférase¹²⁵. Dans le cas des HKMT, la grande similarité de séquence au sein du domaine SET des HKMT a permis à M. Luo, d'identifier un résidu conservé au sein de cette famille, dont la mutation abolit la prise en charge de la SAM mais crée une nouvelle fonction de prise en charge de différents analogues de SAM encombrés, en libérant suffisamment de place dans le site de fixation du cofacteur. Avec cette stratégie, en utilisant des mutants des enzymes GLP et G9a en combinaison avec l'analogue de SAM nommé **Hey-SAM**, il parvint à identifier plusieurs centaines de potentiels substrats des deux enzymes GLP et de G9a. Dans cette longue liste de candidats, il réussit à valider 11 nouveaux substrats des enzymes sauvages, plus ou moins sélectifs de l'une des deux enzymes¹¹².

Finalement, la non prise en charge d'analogues de cofacteur par les enzymes sauvages devient un avantage pour l'étude des substrats d'HKMT car il permet de s'affranchir de la redondance d'activité entre les méthyltransférases. Ainsi, dans un mélange complexe de protéines (e.g. extrait cellulaire), la totalité des protéines fonctionnalisées en vue d'une caractérisation seraient uniquement des cibles de l'HKMT mutée et par extension de l'enzyme sauvage si l'on admet que la spécificité de substrat n'est pas modifiée. Inversement, le design de mutants d'enzyme perdant leur activité avec la SAM est aussi un avantage car les lysines cibles de l'HKMT sont moins rapidement méthylées (par l'activité de cette HKMT mutée et de la SAM endogène) et restent donc libres plus longtemps pour une fonctionnalisation par le groupement rapporteur.

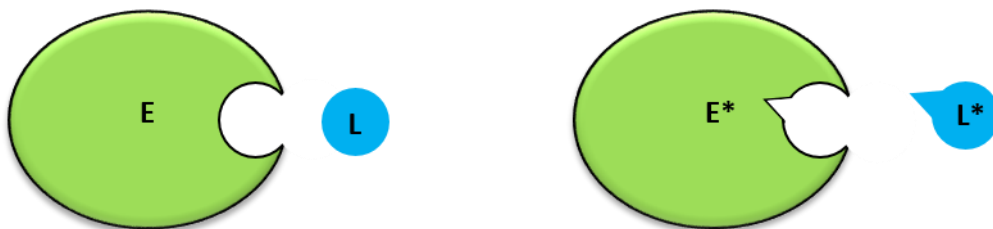


Figure 20. Représentation schématique de la stratégie « Bump and hole »
(E : Enzyme, L : Ligand, E* : Enzyme mutée, L* : Ligand modifié chimiquement).

Pour compléter le tour d'horizon des analogues de SAM, de nouveaux analogues ont également été proposés ces dernières années pour diverses applications sur des méthyltransférases d'ADN ou d'ARN plus permissives et donc non modifiées.

L'équipe de Klimasauskas *et al.* a développé des cofacteurs transférables par des enzymes ADN ou ARN méthyltransférases avec en position β du sulfonium une triple liaison^{126,127}. Cette triple liaison favorise la stabilité des analogues et le transfert de groupements alkyle clickables. Les *linkers* développés par cette équipe ont une chaîne alkyle longue contenant au moins 3 carbones

entre la triple liaison et la fonction clickable ce qui permet d'éviter une réaction secondaire d'hydrolyse de la triple liaison. Substitués ainsi, ces cofacteurs sont suffisamment stables pour fonctionnaliser la chaîne alkyle avec un fluorophore ou une biotine directement sur le cofacteur sous sa forme sulfonium (Figure 18). Cette stratégie a permis de transférer et marquer des séquences ADN avec une grande spécificité en une ou deux étapes. Cette technologie a aussi été utilisée pour fonctionnaliser et marquer de façon ciblée de l'ARN¹²⁸.

Plus récemment le groupe de Neely *et al.* a introduit un motif hydrazide ou alkoxyamine sur ces chaînes alcynes ce qui a permis de faire des analogues versatiles pouvant être dérivatisés par la suite selon une réaction réversible en utilisant des aldéhydes. Une fois transférés sur leurs cibles grâce aux méthyltransférases, cette stratégie permet de marquer de l'ADN de manière réversible avec différents fluorophores ou marqueurs d'affinité. Dernièrement ce même groupe a utilisé ce type de cofacteurs alcynes, pour développer un test FRET afin de quantifier l'activité d'ADN méthyltransférase en présence d'inhibiteurs¹²⁹.

Enfin, ces dernières années de nouveaux analogues ont été proposés par l'équipe d'A. Rentmeister. Ces analogues utilisent un groupement aromatique benzyle en β du sulfonium. Sur ce modèle, des analogues photoactivables ou photolabiles ont été synthétisés pour diverses applications sur des méthyltransférases d'ARN et d'ADN (Figure 18)¹³⁰⁻¹³². Ces nouveaux cofacteurs de type aryle cétone, ont été utilisés pour cager des bases nucléiques de façon spécifique. Ces groupements une fois introduits sur les bases nucléiques sont déprotégés sous irradiation ce qui permet, par exemple, d'étudier des processus biologiques d'interactions ARN-protéines ou ADN-protéines en temps réel. Ces analogues synthétisés par voie chimique avec de faibles rendements sont encombrés et restent peu stables dans les conditions de biocatalyse, ce qui a conduit le groupe de A. Rentmeister à travailler sur la structure de la *Cryptosporidium hominis* MAT synthétase pour la production de ces analogues. Ce travail a été publié récemment et donne de bonnes perspectives pour la production de ces composés encombrés à l'aide d'enzymes *in cellula*¹³³.

La Figure 18 récapitule les principaux analogues de SAM décrits à ce jour. La conception de nouveaux analogues de SAM est un champ de recherche très actif actuellement. Au laboratoire, de nouveaux analogues sont également en cours de synthèse mais ils ne feront pas l'objet de la présente étude et lors de la mise en place de ce projet, nous nous sommes principalement appuyés sur des analogues déjà décrits.

II.1.3. Les différentes méthodes d'accès aux analogues de SAM alkylés sur le soufre

Avant de présenter les analogues utilisés dans ce travail, une brève présentation des méthodes d'obtention des analogues de SAM alkylés sur le soufre décrits dans la littérature est présentée ci-dessous. Elles reposent sur deux approches : une synthèse chimique et une approche enzymatique. Des analogues sélénés ont également été obtenus selon des approches similaires.

• Synthèse chimique basée sur l'alkylation de la *S*-adénosyl-L-homocystéine

La synthèse de ce type d'analogues de SAM repose sur l'alkylation du soufre de la *S*-adénosyl-L-homocystéine avec un halogénure d'alkyle ou un triflate d'alkyle dans des conditions acides qui permettent de maintenir l'amine nucléophile de la SAH sous forme protonée, en présence ou en absence de perchlorate d'argent comme catalyseur de la réaction (Figure 21A). L'utilisation de cette méthode aboutit à un mélange des épimères de configuration *R* ou *S* au niveau du sulfonium qu'il est parfois possible de séparer par HPLC. Les rendements synthétiques typiques sont assez faibles car, comme nous le préciserons plus tard, ces composés sont instables et il faut trouver un juste milieu entre un temps de réaction permettant de consommer la SAH et un temps assez court pour éviter la dégradation du produit final. Par ailleurs, la SAH et les analogues de SAM, produits polaires, sont difficiles à séparer mais il est pourtant nécessaire d'effectuer soigneusement cette séparation par HPLC en phase inverse car la SAH est un inhibiteur puissant des méthyltransférases et des résidus de ce produit présents dans les analogues de SAM seraient préjudiciables aux réactions enzymatiques. Même si cette méthode chimique ne permet pas facilement de monter en échelle dans la synthèse de ces analogues, elle constitue la stratégie de synthèse privilégiée pour la préparation d'un grand nombre d'analogues de SAM alkylés. C'est la méthode utilisée lors des travaux pionniers de Weinhold et Klimasauskas dans le domaine^{126,134}.

• Synthèses enzymatiques

Deux méthodes utilisant les enzymes ont été employées : l'une utilisant les méthionine adénosyltransférases, l'autre utilisant les halogénases. Ces méthodes possèdent pour avantage *i*) la stéréospécificité de la réaction permettant d'accéder à l'épimère naturel, *ii*) la possibilité de coupler directement la production d'analogues de SAM à des réactions ultérieures et de contourner ainsi les problèmes de purification fastidieuse et de stabilité évoqués plus bas.

La première voie d'accès enzymatique aux analogues de SAM est assez ancienne et s'appuie sur la voie de biosynthèse de la SAM naturelle utilisant les *S*-adénosylméthionine synthétases (MAT) qui catalysent la formation de SAM à partir d'ATP et de L-Met. Cette méthode a ainsi

permis d'obtenir des analogues tritiés ou deutérés de la SAM pour l'étude de différents mécanismes enzymatiques. Dans une étude plus récente, Singh *et al.* ont étudié les capacités d'un ensemble représentatif de MAT bactériennes, d'archées et de mammifère de type sauvage avec différents analogues de L-Met alkylés sur le soufre ou le sélénium. Certaines MAT ont ainsi été identifiées comme très permissives^{135,136}. Dans d'autres travaux, des MAT ont été mutées afin de les rendre plus aptes à accommoder la méthionine modifiée par des groupements alkyles volumineux (Figure 21B)^{113,137}.

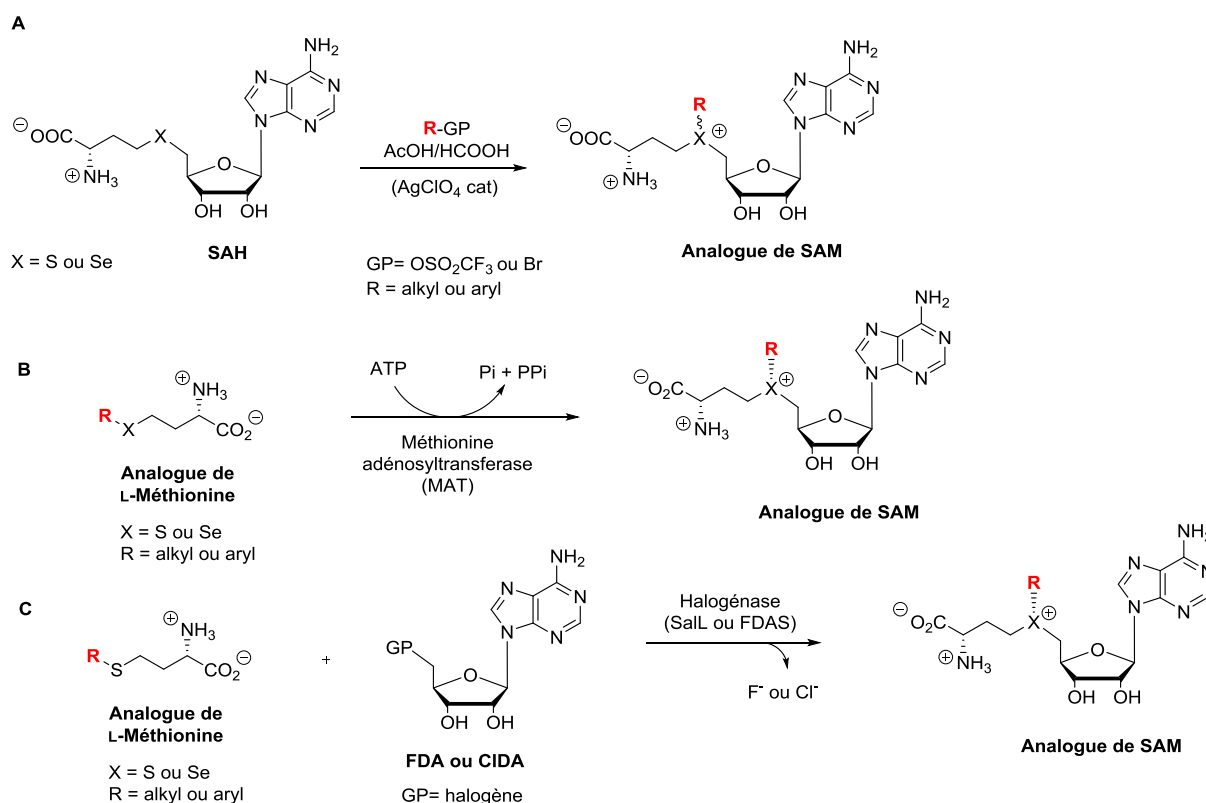


Figure 21. Les différentes voies d'accès aux analogues de SAM alkylés sur le soufre ou sélénium. A. Méthode chimique. B. Méthode enzymatique basée sur la voie de biosynthèse de la SAM naturelle. C. Méthode enzymatique basée sur une halogénase bactérienne.

Plus récemment, deux halogénases microbiennes, la 5'-chloro-5'-désoxyadénosine synthase (SaL) et la 5'-fluoro-5'-désoxyadénosine synthase (FDAS) ont été utilisées pour la synthèse d'analogues de SAM S-alkylés. Ces enzymes catalysent la formation réversible de L-Méthionine et de 5'-chloro ou 5'-fluoro-5'-désoxyadénosine (CIDA ou FDA, respectivement) à partir d'AdoMet et de chlorure ou de fluorure, respectivement (Figure 21C)¹³⁸⁻¹⁴¹. Une étude a permis de montrer que des SaL et FDAS sauvages ou mutées catalysent la production de différents analogues de SAM S-alkylés à partir d'analogues de L-Met et de CIDA ou de FDA¹²².

Ces méthodes biocatalytiques deviennent de plus en plus attractives. Très récemment une nouvelle étude utilisant l'évolution dirigée propose une autre alternative enzymatique permettant d'obtenir des analogues de SAM diastéréoisomériquement purs avec des rendements de l'ordre de 70 % pour des chaînes alkyles courtes et basée sur une activité halogénure méthyltransférase¹⁴².

• **Considérations autour de l'instabilité du cofacteur SAM et de ses analogues**

Bien que ces analogues de SAM soient fondamentalement des outils intéressants pour l'étude du méthylome de certaines enzymes, leur application reste contrainte du fait de leur temps de vie relativement court de l'ordre de 1 à 3 heures en milieu aqueux selon les composés. En effet, les caractéristiques structurales qui expliquent la réactivité de ce cofacteur en font également un composé instable. L'analyse *in vitro* de la stabilité de la SAM montre que cette molécule (et ses analogues) est sujette à 3 voies de dégradation (Figure 22)^{143,144}. L'épimérisation au niveau du soufre est l'une des principales transformations pouvant engendrer la formation de l'analogue inactif. La déprotonation au niveau de l'atome de carbone 5' peut engendrer la formation de l'adénine et de la *S*-pentosylméthionine. Enfin, l'attaque intramoléculaire du groupe carboxylate sur le carbone γ de la méthionine forme la L-homosérine lactone (HSL) et la 5'-désoxy-5'-méthylthioadénosine (MTA). Ces réactions, même si elles sont relativement rapides, ne posent pas de problème dans le contexte cellulaire mais doivent être prises en compte dès lors que l'on manipule ces composés *in vitro* et peuvent être un inconvénient pour les longues manipulations synthétiques et les étapes de purification, la quantité du composé actif disponible diminuant rapidement. Leur conservation nécessite également des précautions particulières : les produits sont échantillonnés et conservés à -20°C.

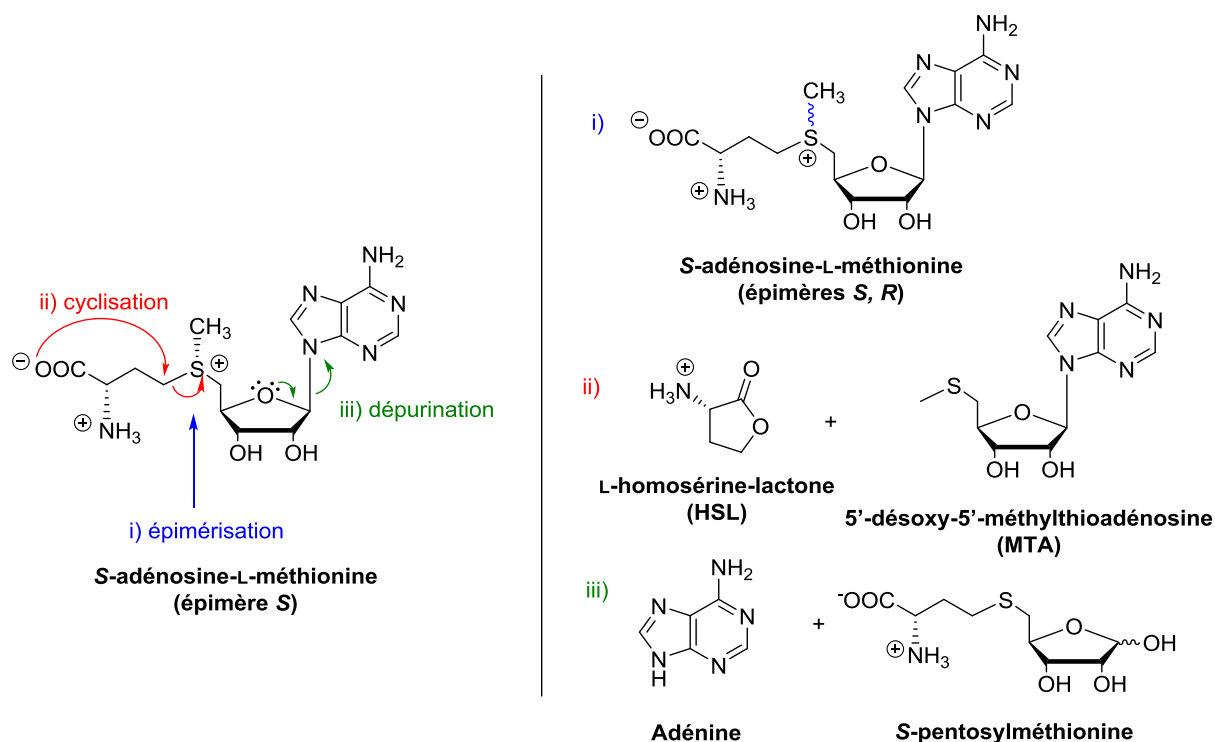


Figure 22. Schéma général représentant les voies de dégradation de la SAM ainsi que de ses dérivés, en milieu aqueux via i) une épimérisation au niveau du soufre, ii) une cyclisation produisant la HSL et la MTA et iii) une dépurination. Les mêmes voies de dégradation peuvent être observées pour les analogues soufrés et séléniés de la SAM.

II.1.4. Synthèse des analogues de SAM alkylés utilisés dans cette étude

Pour notre part, dans le cadre de notre étude, nous souhaitons accéder à des analogues alkylés sur le soufre et comportant un motif bioorthogonal alcyne ou azoture en vue d'une fonctionnalisation ultérieure par réaction CuAAC. Au vu de ce qui a été dit précédemment, on peut donc résumer les différents prérequis pour la conception des analogues de la SAM utiles à notre étude dans la Figure 23.

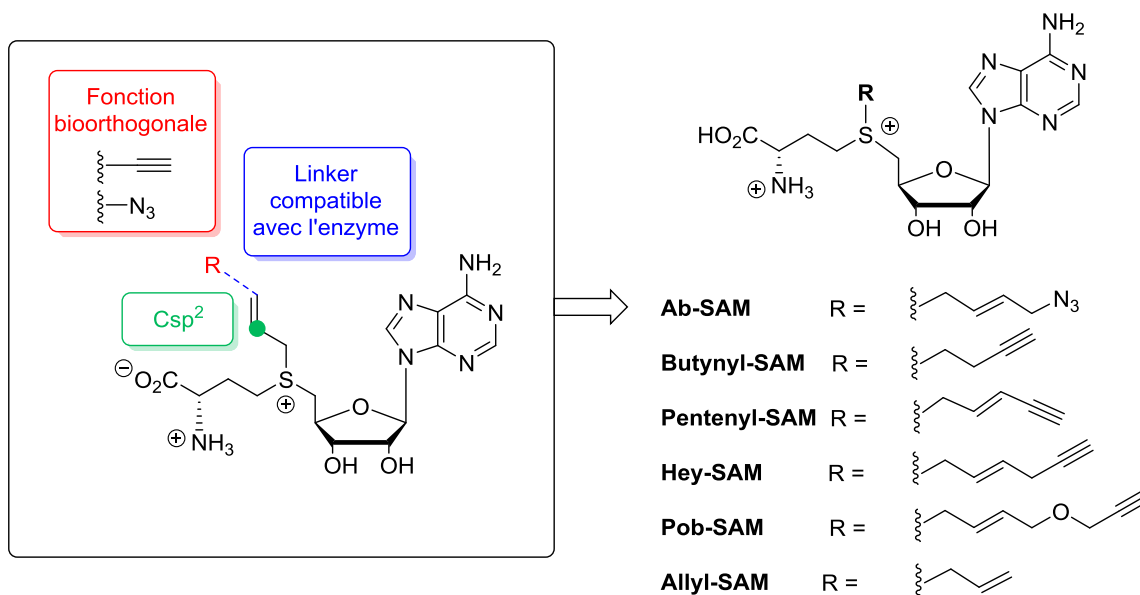


Figure 23. Critères généraux pour les analogues de SAM utilisés dans cette étude.

Le Dr. Karine Guitot, post-doctorante au laboratoire avait déjà entamé une étude sur la lysine méthyltransférase EZH2 à l'aide de cofacteurs modifiés décrits dans la littérature mais également de composés nouveaux et l'approche par synthèse chimique a été privilégiée. Au début de ce projet, les approches enzymatiques n'étaient pas aussi avancées qu'aujourd'hui et nécessitaient la production de SAM synthétases modifiées et dans tous les cas, la synthèse de précurseurs analogues de L-méthionine nécessitant plusieurs étapes.

Je me suis appuyé sur ses résultats préliminaires pour choisir les analogues qui semblaient les plus appropriés pour mon étude, à savoir **Ab-SAM**, **Pob-SAM** (4-propargyloxy-but-2-enyl SAM), **Hey-SAM** ((*E*)-hex-2-en-5-ynyl SAM) et **Pentenyl-SAM** ((*E*)-pent-2-en-4-ynyl SAM) (Figure 23). La synthèse d'un analogue plus court, le **Pro-SAM** comportant le motif propargyl, s'est révélée infructueuse car, comme nous en reparlerons plus tard, cet analogue a une faible stabilité et une très grande tendance à l'hydrolyse au cours de sa synthèse et de son utilisation par la suite en présence d'enzymes. Le **Butynyl-SAM** a donc été choisi comme analogue clickable minimaliste bien qu'il ne possède pas de carbone sp² en β du sulfonium. Un analogue minimaliste intéressant comportant une chaîne alkyle courte, l'**Allyl-SAM** a également été sélectionné pour les tests de mise au point, étant plus stable et bien pris en charge par plusieurs de nos enzymes (Figure 23). Il est ensuite possible de fonctionnaliser le motif allyle une fois introduit sur la chaîne latérale de lysines en présence de biaryle tétrazole sous irradiation à 360 nm et cette réaction a été réalisée au laboratoire sur des peptides modèles. Cependant nous n'avons pas retenu cette réaction de bioconjugaison dans la suite de notre étude

car plus délicate à mettre en œuvre que les réactions de CuAAC et SPAAC, et, comportant également des inconvénients en terme de spécificité et de réactivité^{145,146}.

Karine Guitot m'a régulièrement fourni les analogues de SAM qu'elle a synthétisés au laboratoire selon l'approche chimique et j'ai également participé à la synthèse de certains d'entre eux. Ces synthèses se font toutes sur un schéma commun en deux étapes : d'une part la préparation des agents alkylants, et d'autre part l'alkylation de la SAH.

a) Synthèse des agents alkylants

L'ensemble des composés **1** à **3** a été synthétisé à partir du même composé de départ : le (*E*)-1,4-dibromobut-2-ène commercial. Ce composé symétrique est fonctionnalisé par un groupement clickable en substituant l'un des atomes de brome de la molécule. Ainsi, le (*E*)-1-azido-4-bromobut-2-ène **1**, a été synthétisé à partir du (*E*)-1,4-dibromobut-2-ène en présence d'azoture de sodium (Schéma 1) selon un protocole décrit avec un rendement de 28 %^{135,147}. L'alcyne **2** a été obtenu à partir du (*E*)-1,4-dibromobut-2-ène et de l'alcool propargylique en présence de NaH avec un rendement de 32 %^{148,149}. Le (*E*)-6-bromo-hex-4-en-1-yne **3** a été obtenu par addition du bromure d'éthynyl magnésium au (*E*)-1,4-dibromobut-2-ène en présence de cuivre (I) dans des conditions anhydres (30 % de conversion). Les faibles rendements autour de 30 % de ces réactions s'expliquent par l'obtention, en plus du produit désiré, de composés disubstitués, et par la volatilité du composé. L'alcool activé **4** a été obtenu à partir du but-3-yn-1-ol et d'anhydride trifluorométhane sulfonique (95 % de conversion). Le (*E*)-pent-2-en-4-yn-1-ol a été activé sous forme de mésylate en présence de chlorure de mésyle conduisant à la formation du composé **5** avec 37 % de rendement¹²³. Ces électrophiles, peu stables, ont le plus souvent été engagés sans purification préalable dans les réactions suivantes pour fonctionnaliser la SAH et former les cofacteurs.

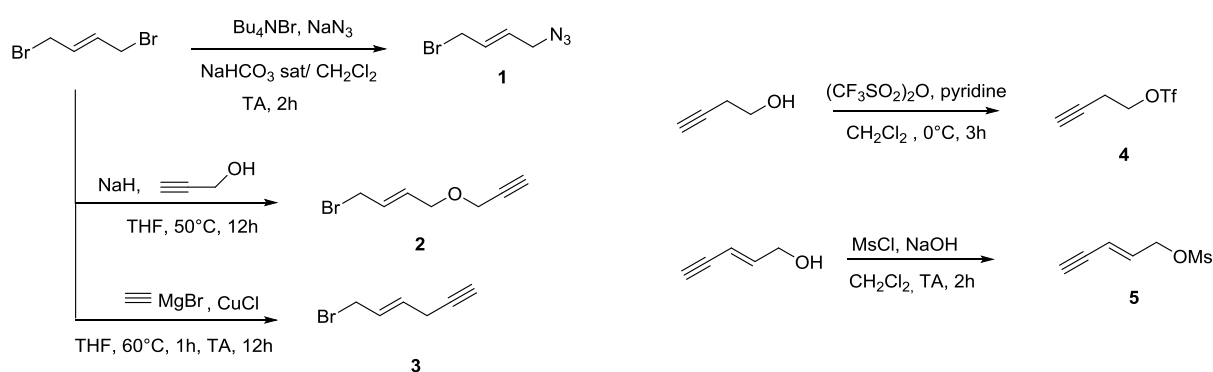


Schéma 1. Synthèse des différents précurseurs utilisés pour l'alkylation de la SAH.

b) Alkylation de la SAH

La synthèse des différents analogues de SAM a ensuite été effectuée selon l'approche décrite dans la Figure 21A par réaction d'alkylation du soufre de la *S*-adénosyl-L-homocystéine (SAH) en milieu acide dans les conditions décrites par Weinhold¹¹¹ en présence des agents alkylants préparés précédemment ou d'un réactif commercial, le bromure d'allyle dans le cas de la préparation de l'**Allyl-SAM**. Dans le cas des composés halogénés, du perchlorate d'argent est additionné ce qui conduit à une accélération notable de la réaction. Le suivi des réactions est effectué par HPLC et il est parfois nécessaire d'arrêter la réaction avant consommation totale de la SAH lorsque le produit formé commence à se dégrader. La purification des analogues est effectuée par HPLC en phase inverse et les deux épimères d'analogues de SAM ne sont pas séparés même si dans certaines conditions, il est possible de les distinguer en HPLC analytique. Nous devons en tenir compte dans les concentrations lors des réactions enzymatiques car les méthyltransférases ne prennent en charge qu'un seul des deux épimères mais ne semblent pas inhibées par l'autre.

La synthèse du dérivé 4-azidobut-2-enyl de la SAM, l'**Ab-SAM 5** est présentée comme exemple représentatif des différents analogues. Le composé **1** est engagé dans l'alkylation de la SAH dans un mélange acide acétique/acide formique en présence de perchlorate d'argent (Figure 24) à température ambiante. Le suivi de la réaction par HPLC analytique montre la diminution de la SAH et la formation d'un produit moins polaire que la SAH (Figure 24). Au bout de 26 h, la réaction n'évolue plus. Après purification par HPLC semi-préparative et lyophilisation, nous avons obtenu l'**Ab-SAM** avec un rendement de 12 % malgré les taux de conversion assez bons. De façon générale, nous avons obtenu les composés avec des rendements compris entre 5 et 15 %. Ces rendements faibles sont à comparer avec ceux de la littérature qui sont du même ordre de grandeur et s'expliquent notamment par l'instabilité des composés, et les échelles de synthèse très petites (quelques mg) sur lesquelles nous avons travaillé. Malgré ces rendements, l'obtention de quelques mg de produit est largement suffisante pour réaliser un grand nombre d'essais enzymatiques.

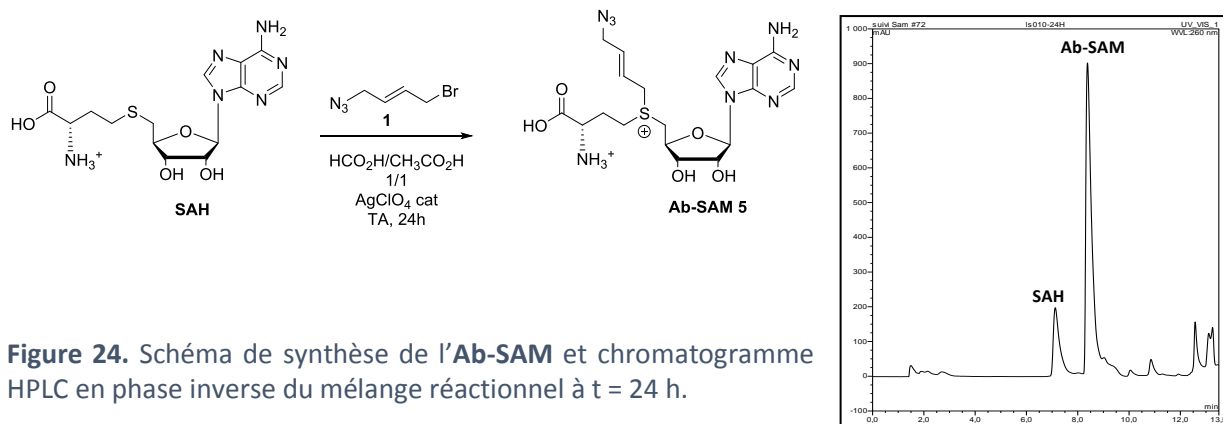


Figure 24. Schéma de synthèse de l'**Ab-SAM** et chromatogramme HPLC en phase inverse du mélange réactionnel à $t = 24$ h.

La Figure 25 récapitule les molécules synthétisées dans le cadre de cette étude selon un protocole similaire conduisant aux analogues butynyl-SAM **6**, pentenyl-SAM **7**, (*E*)-hex-2-en-5-ynyl-SAM (Hey-SAM) **8**, 4-propargyloxy-but-2-enyl-SAM (Pob-SAM) **9**, et allyl-SAM **10**.

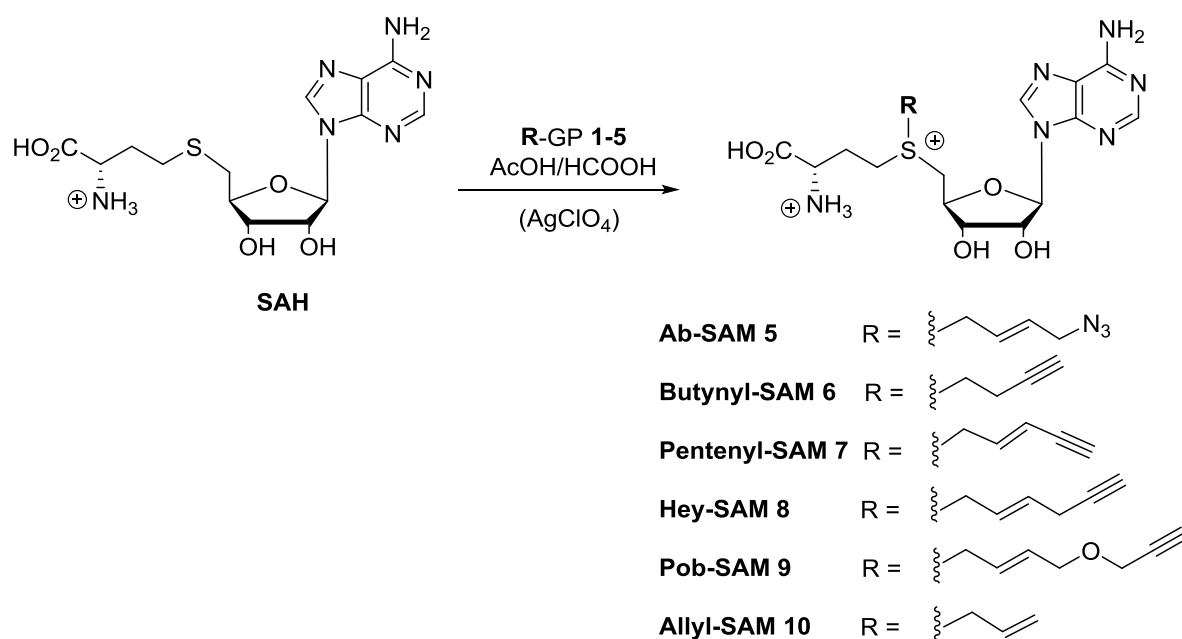


Figure 25. Alkylation de la SAH et préparation des analogues de SAM utilisés dans cette étude.

II.2 Conception et production de mutants de PR-Set7

Dans un deuxième temps, nous avons conçu et préparé différents mutants de l'enzyme PR-Set7 susceptibles de prendre en charge les analogues de SAM pour la mise en œuvre de la stratégie d'étiquetage des substrats non-histone. La conception a été réalisée sur la base d'analyses structurales, puis nous avons réalisé le clonage et la production des protéines mutées correspondantes.

II.2.1 Paramètres structuraux pour la fixation des ligands de PR-Set7

Dans le but de comprendre les paramètres structuraux conférant à PR-Set7 son activité de mono-méthyltransférase, une analyse du site actif de la structure cristallisée est nécessaire. La structure cristalline de l'enzyme complexée à la SAH et à un décapeptide constitué des acides aminés 15 à 24 de l'histone H4 est disponible sur la Protein Data Bank. (Code PDB : 1ZKK)²³

Dans son site de fixation, la SAH adopte une conformation dite en « U » (ou en « fer à cheval »). L'adénine du cofacteur est principalement flanquée par le cycle indole du Trp-349 et de la chaîne aliphatique de la Lys-226 et les atomes N^6 et N^7 de la purine forment respectivement des liaisons H avec les groupements carbonyle et amide de l'His-299 (Figure 26A). Côté « méthionine », l'amine chargée forme un réseau de liaisons H important avec les Lys-226, Arg-228 et Asn-298 ; quant au carboxylate, il forme un pont salin avec le groupement guanidinium d'Arg-228 ainsi qu'une liaison H avec l'hydroxyle de la Tyr-271 (Figure 26A). Dans son ensemble, le mode de fixation du cofacteur est très conservé chez les HKMT à domaine SET et il sert à orienter le groupement méthyle vers le port pour la réaction enzymatique. Il est à noter que contrairement à ce qu'il se passe dans d'autres HKMT, le groupement 3'-OH du ribose forme une liaison H avec l'His-18 du peptide (Figure 26A). Plus couramment, les hydroxyles du ribosyle sont soit directement exposés au solvant¹⁵⁰ soit forment des interactions avec des résidus de l'enzyme, comme observé chez GLP⁴⁰, prévenant ainsi le contact direct entre l'histone et le cofacteur.

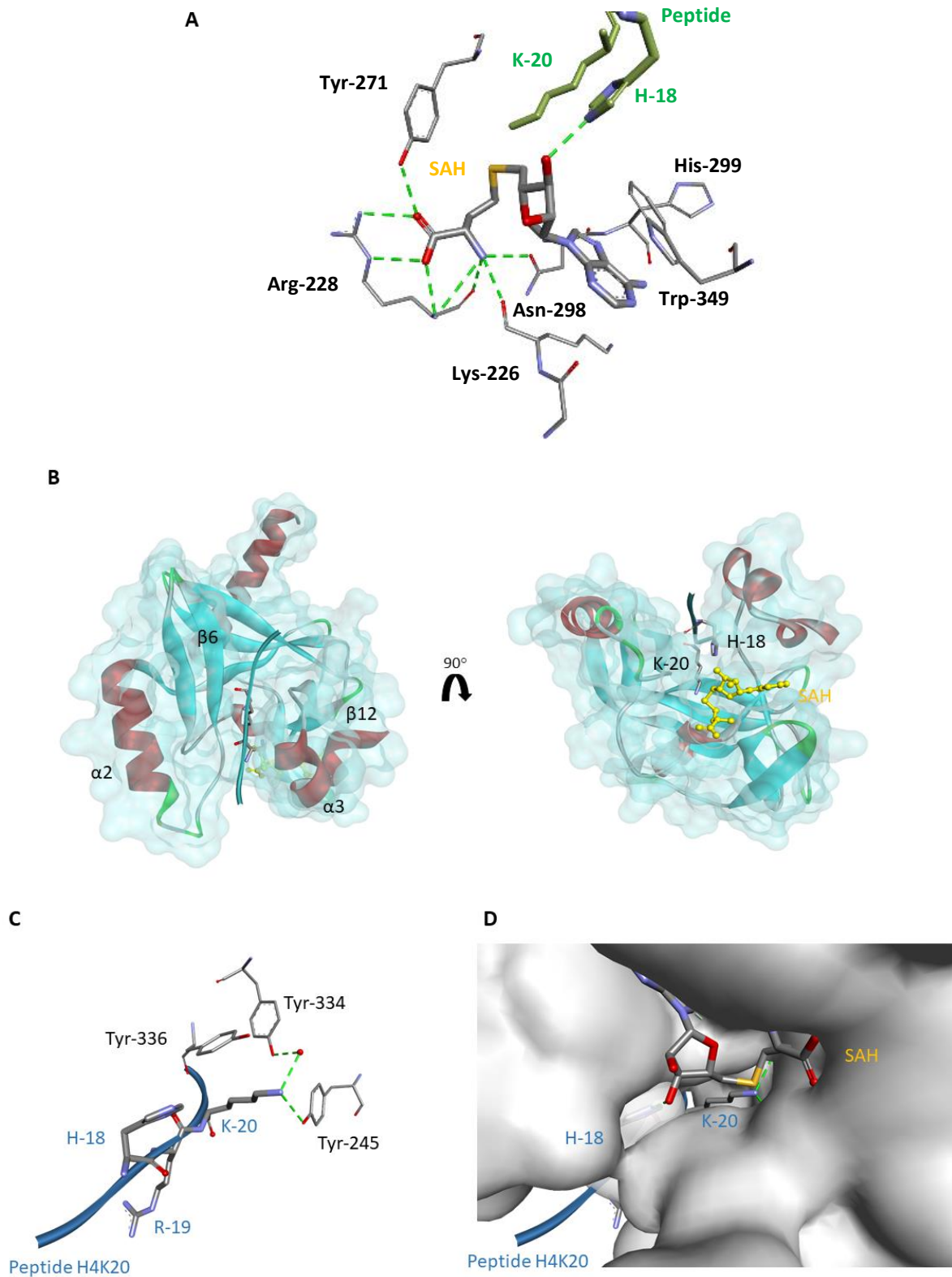


Figure 26. Caractéristiques structurales chez PR-Set7 des sites de fixation au cofacteur et au substrat. A. Interactions des résidus du site catalytique de PR-Set7 avec la SAH. B. Représentation tridimensionnelle du domaine SET de PR-Set7 accommodant à la fois la SAH et le peptide substrat. C et D. Interactions des résidus du site catalytique de PR-Set7 avec le peptide substrat respectivement sans ou avec la représentation de la surface de la protéine dans le but d’observer le canal reliant les sites de fixation au cofacteur et au substrat par lequel passe la lysine à méthyler. L’histidine 18 du peptide formant une liaison H avec la SAM apparaît par transparence.

Le site de fixation du peptide substrat est une profonde fente au sein de l'enzyme formée par le brin $\beta 6$ et l'hélice $\alpha 2$ du domaine i-SET d'une part, et la boucle connectant le brin $\beta 12$ et l'hélice $\alpha 3$ du domaine c-SET d'autre part. Le squelette du peptide se lie alors dans cette fente en formant un feuillet β parallèle au brin $\beta 6$. (Figure 26B). La lysine 20 se retrouve au centre du brin β nouvellement formé et passe par un canal composé de 4 résidus aromatiques : Tyr-245, Tyr-334, Tyr-336 de l'enzyme et His-18 du peptide. (Figure 26C-D). Ce canal connecte le site de fixation du cofacteur SAM et celui du peptide permettant à la lysine de se placer dans sa conformation entièrement étendue, prête à être méthylée.

La chaîne latérale de l'histidine 18 du peptide forme aussi des interactions avec les résidus de la protéine. En plus de la liaison H faite avec le 3'-OH du ribose de la SAH, la chaîne latérale de His-18 forme un empilement π à la fois avec le cycle indole de Trp-349 et le cycle imidazole de His-347. Au-delà de l'interaction entre cette histidine 18 et le ribose du cofacteur, il semble que le réseau d'empilement π soit essentiel pour une bonne reconnaissance du peptide par son enzyme. En effet l'étude de Jean-François Couture²³ révèle que la mutation H18F n'entraîne qu'une très faible diminution de l'affinité de l'enzyme pour H4 quand H18A l'abolit complètement. De façon cohérente, avec la mutation H347A l'efficacité catalytique de l'enzyme chute d'un facteur 5 quand elle augmente d'un facteur 25 avec la mutation H347F.

II.2.2 Mécanisme de mono-méthylation de PR-Set7 WT

Parmi les résidus au voisinage de la lysine, plusieurs sont directement impliqués dans le bon fonctionnement du mécanisme de transméthylation de la fonction amine de la lysine substrat. La Tyr-245 et la Tyr-334 en font partie. Extrêmement conservées chez les HKMT, leur groupement hydroxyle participent directement ou par l'intermédiaire d'une molécule d'eau au bon positionnement de la fonction amine. Il a entre autres été montré que la mutation de Tyr-245 en une phénylalanine ou une alanine réduisait fortement l'affinité de l'enzyme pour H4 et abolissait l'activité méthyltransférase²³. Le cas de la Tyr-334 est un peu différent. Alors que la mutation en une alanine réduit aussi l'affinité vis-à-vis du substrat, la mutation en une phénylalanine n'a que très peu d'effet sur l'efficacité catalytique de l'enzyme. Le mutant Y334F acquiert en revanche la possibilité de di-méthyliser l'histone H4, on parle de mutant gain de fonction. Ces données concordent avec l'idée générale que chez les HKMT à domaine SET, la tyrosine conservée en une position analogue à la Tyr-245 est essentielle à l'activité de l'enzyme grâce à l'établissement d'une liaison H stabilisante avec la lysine déprotonée ; la Tyr ou la Phe en une position analogue à la position 334, contrôle la spécificité de substrat, l'absence de

groupement hydroxyle induisant systématiquement un plus grand degré de méthylation. Enfin, un résidu est absolument invariant chez toutes les HKMT à domaine SET connues : la Tyr-336. Le groupement hydroxyle de la chaîne latérale de ce résidu participe en effet à la stabilisation de l'état de transition de la réaction de transméthylation en formant une liaison hydrogène de type C-H...O avec le groupement méthyle à transférer.

Les contraintes imposées par ces 3 résidus au voisinage direct de la lysine permettent ensemble une réaction de type S_N2 en plaçant l'amine dans un angle favorable pour l'attaque du nucléophile sur le carbone en α du sulfonium de la SAM, ou d'une molécule analogue.

II.2.3 Conception des mutants de PR-Set7

En 2013 l'équipe de Minkui Luo décrit que chez la lysine méthyltransférase GLP, histone méthyltransférase qui méthyle de façon multiple la lysine 9 de l'histone H3, la tyrosine 1211 du site actif agit comme un résidu gardien d'une cavité hydrophobe et que la mutation de ce résidu entraîne une accessibilité de cette cavité coïncidant avec la prise en charge de plusieurs analogues de SAM de façon plus ou moins efficace. Ils identifient notamment le couple formé par le mutant Y1211A et la molécule Hey-SAM, particulièrement efficace. Avec une efficacité catalytique du mutant seulement divisée par deux par rapport à l'enzyme sauvage envers la SAM, ils parviennent à utiliser ce couple pour la caractérisation des substrats de GLP. Ils dévoilent ainsi une liste de 500 candidats potentiels de GLP et parviennent ensuite à en valider 11 *in vitro*¹¹².

J'ai conçu différents mutants de PR-Set7 en me basant sur la grande similarité de séquence des domaines SET des histones lysines méthyltransférases. La séquence de ce domaine étant également très conservée chez PR-Set7 et GLP, un alignement de structures devrait permettre d'identifier en première approche, les résidus clefs à muter qui pourraient donner lieu à des mutants fonctionnels similaires au modèle GLP.

J'ai donc effectué un alignement structural avec le logiciel PyMOL entre deux structures cristallographiques, PR-Set7 en complexe ternaire avec la SAH et le peptide H4 (code PDB 1ZKK,²³), GLP en complexe ternaire avec la SAH et le peptide H3K9 (code PDB 2RFI⁴⁰) (Figure 27). Cette superposition révèle que les résidus aux alentours de la lysine à méthyler et de la SAH sont effectivement très conservés. Dans ses travaux, Luo avait proposé les mutations Y1211A et F1209A pour agrandir le site de fixation de la SAM dans l'enzyme GLP. Cette enzyme et le mutant Y1211A ont été produits au laboratoire par le Dr. Karine Guitot et elle a

en effet pu reproduire ces résultats et observer une bonne prise en charge de différents analogues de SAM. Nous avons donc cherché les résidus analogues chez PR-Set7 et proposé en conséquence les mutations des Tyr-336 et Tyr-334 de PR-Set7, en alanine qui sont parfaitement superposées aux résidus Phe-1209 et Tyr-1211 de GLP respectivement (Figure 27). La mutation de la tyrosine 334 en une phénylalanine est aussi proposée étant donné qu'elle est suffisante pour induire une plus grande spécificité de substrat par rapport à l'enzyme sauvage. En la comparant à la structure de GLP, l'introduction de cette mutation revient à obtenir un analogue de GLP sauvage qui contient naturellement une phénylalanine en cette position. Pour la même raison, le double mutant de PR-Set7 Y336A et Y334F est suggéré pour être le mutant le plus proche de la structure du mutant Y1211A de GLP. Il devient alors le candidat le plus prometteur pour le transfert des différentes chaînes alkyle des analogues de SAM

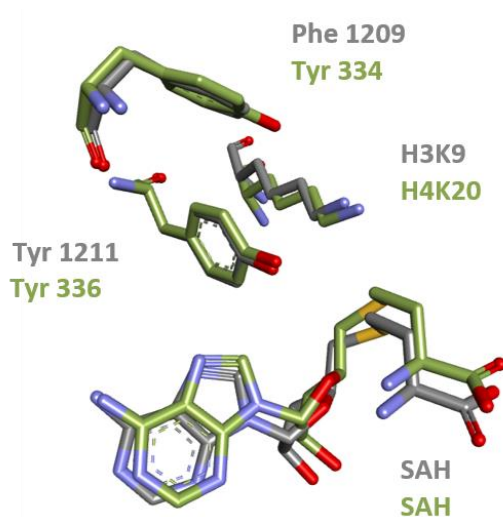


Figure 27. Superposition des domaines SET de GLP (code PDB 2RFI, en gris) et de PR-Set7 (code PDB 1ZKK, en vert). Les résidus aromatiques montrés sont les propositions de résidus à muter par mutagenèse dirigée.

II.2.4 Production des mutants de PR-Set7

J'ai donc préparé quatre mutants de PR-Set7 par mutagenèse dirigée en introduisant les mutations précédemment décrites sur un plasmide Pet-28a contenant la séquence de la protéine entière PR-Set7 flanquée de deux étiquettes, FLAG en C-ter et 6-His en N-ter. L'intégrité des gènes et l'incorporation des mutations ont été vérifiées par séquençage d'ADN. Des souches d'*E. coli* BL21(DE3) ont ensuite été transformées par les différents plasmides mutés et le plasmide sauvage. Chaque protéine PR-Set7 mutée ou sauvage a été surexprimée grâce au système d'expression T7 pour être ensuite purifiée sur colonne IMAC (Immobilized Metal Affinity Chromatography) saturée au nickel. Après lavage de la colonne, la protéine d'intérêt

est éluée en présence d'imidazole et les fractions d'intérêt sont analysées sur gel de polyacrylamide 12 % en conditions dénaturantes (Figure 28).

La présence de la protéine recombinante est confirmée par le gel car sa masse théorique est de 45 kDa, en accord avec les marqueurs de taille (Figure 28A). La figure 12A présente à titre d'exemple le gel d'électrophorèse réalisé lors de la purification de PR-Set7 sauvage. Nous notons également la présence d'une quantité variable de protéines co-purifiées avec les enzymes. Il est à noter que ces impuretés pourraient éventuellement être éliminées grâce à une étape de chromatographie échangeuse d'ion ou une chromatographie d'exclusion stérique, toutefois nous avons jugé ces fractions d'éluion suffisamment pures pour procéder au criblage de la collection d'analogues de SAM. A l'issue de la purification d'affinité nous avons obtenu entre 2 et 10 mg d'enzyme sauvage ou mutée à partir de 0,5 L de milieu de culture. Les fractions éluées sont conservées en présence de 50 % glycérol à -80°C. Les fractions purifiées pour chaque enzyme sont présentées dans la Figure 28B.

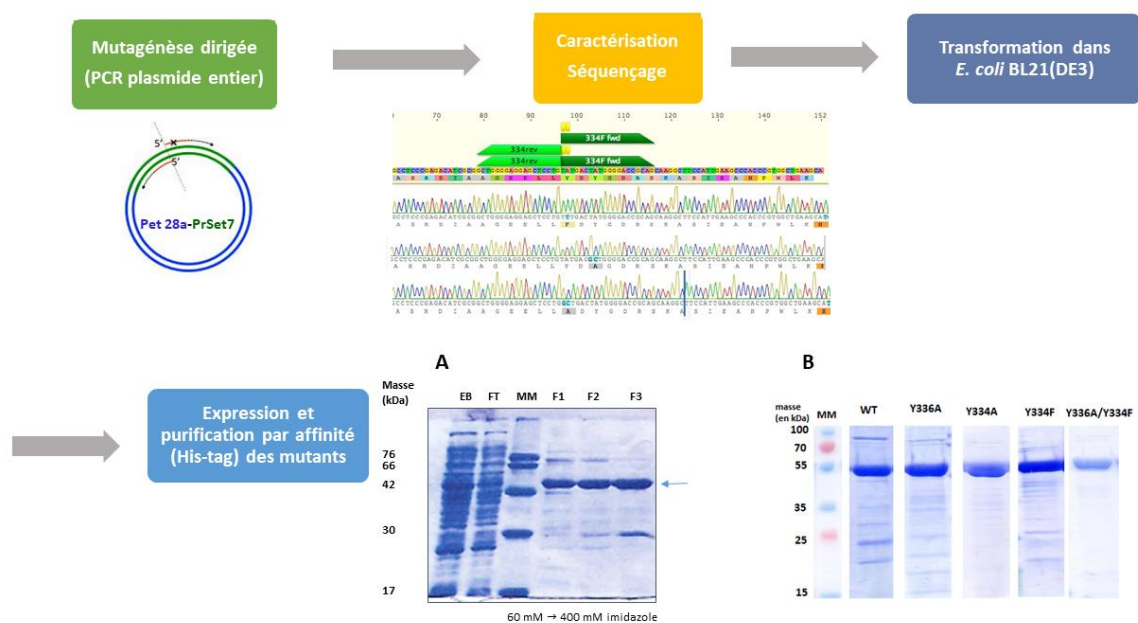


Figure 28. Schéma récapitulatif des différentes étapes de préparation des différents mutants de PR-Set7. A. Un exemple de gel SDS-PAGE 12 % d'acrylamide révélé au bleu de Coomassie des différentes fractions d'éluion de PR-Set7 sauvages. EB : extraits bruts, FT : Protéines non fixées sur la colonne, MM : marqueur, F1 à F3 : Fractions éluées 1 à 3. B. Gel SDS-PAGE 12 % d'acrylamide des enzymes PR-Set7 recombinantes sauvages et mutées révélées au bleu de Coomassie.

II.2.5. Évaluation de l'activité enzymatique des mutants de PR-Set7

Afin d'évaluer les capacités catalytiques de nos différents lots de PR-Set7, nous avons tout d'abord mis en œuvre un test simple sur peptide en présence du cofacteur SAM, en utilisant un court peptide contenant les acides aminés 11 à 28 de l'histone H4, la lysine 20 étant en position centrale :

H-G(11)KGGAKRHRK**V**LRDNIQG(28)-NH₂ (Figure 29).

Différentes conditions de réactions ont été testées notamment trois tampons permettant l'activité d'autres HKMT, un tampon 50 mM Tris-HCl pH 8, un tampon 50 mM Hepes pH 8 et un tampon 50 mM glycine pH 9,8. Les tests ont aussi été effectués à différentes températures, 20°C, 25°C, 30°C et 37°C. Nous avons défini par la suite que les conditions optimales pour les tests d'activité *in vitro* de PR-Set7 sont : un tampon glycine à pH 9,8, la réaction de méthylation étant réalisée à 25°C, pendant 3 h.

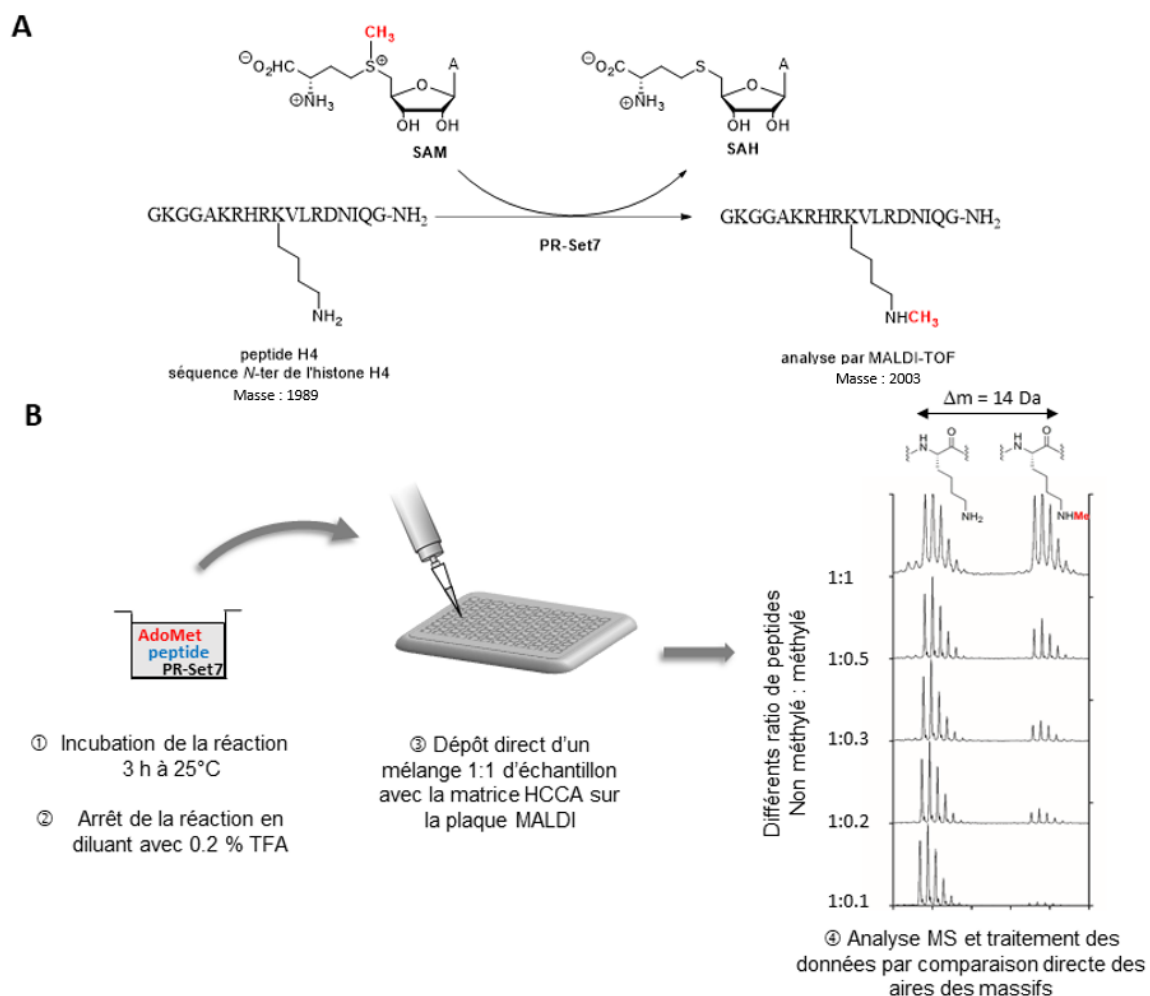


Figure 29. A. Réaction catalysée par PR-Set7 sur un peptide substrat peptidique fragment de l'histone H4. B. Principe du test enzymatique MALDI-TOF MS.

Le principe du test utilisé repose sur le dosage du taux de méthylation du peptide après son analyse par spectrométrie de masse MALDI-TOF. Un protocole établi précédemment au laboratoire, illustré dans la Figure 29 a été utilisé^{151,152}. Concrètement, le peptide issu du mélange réactionnel peut être analysé directement sans étape de dessalage et il a été montré que la méthylation ne modifiait pas les propriétés d'ionisation/désorption du peptide, ce qui facilite la quantification puisque le taux de méthylation est calculé en faisant le rapport des aires des massifs isotopiques de chaque espèce peptidique méthylée ou non.

En contrôle, nous avons également analysé la méthylation d'un peptide fragment de l'histone H3 (peptide H3 = fragment 1-19 de l'extrémité *N*-ter de H3) par l'enzyme GLP dont l'activité est robuste *in vitro*. Alors que le taux de méthylation du peptide H4 sauvage ne dépasse pas 20 % pour PR-Set7 (Figure 30A) après 3 h d'incubation, l'activité de GLP sauvage est bien plus robuste et plus de 90 % du peptide H3 est di ou tri-méthylé (Figure 30C) après seulement 30 min d'incubation. Nous avons ensuite évalué l'activité enzymatique de nos quatre mutants en présence du cofacteur SAM et nous avons observé que seul le simple mutant Y334F réussissait à méthyler le peptide H4 dans des proportions équivalentes à l'activité de PR-Set7 sauvage (Figure 30B) ce qui reste conforme avec ce qui est rapporté dans la littérature. Nous n'observons toutefois aucune trace de di-méthylation sur le peptide alors que l'on s'attendait du fait de cette mutation à obtenir un mutant gain de fonction. Pris ensemble, ces résultats sont cohérents avec le rôle connu de chacun des résidus mutés. Le très faible taux de méthylation observé avec l'enzyme WT et le mutant Y334F de PR-Set7 est cependant un obstacle en comparaison avec les excellents résultats observés avec GLP. En effet, nous pouvons légitimement anticiper qu'un analogue de la SAM puisse être bien moins pris en charge par un mutant de PR-Set7. Dès lors, l'activité basale de l'enzyme envers son cofacteur risque de poser un problème par la suite pour la détection des modifications introduites sur H4 par l'enzyme. En somme, ces résultats suggèrent que l'enzyme a une activité catalytique moins robuste que GLP et que le test enzymatique utilisant un peptide comme substrat risque d'être limitant dans l'étude de criblage ultérieure.

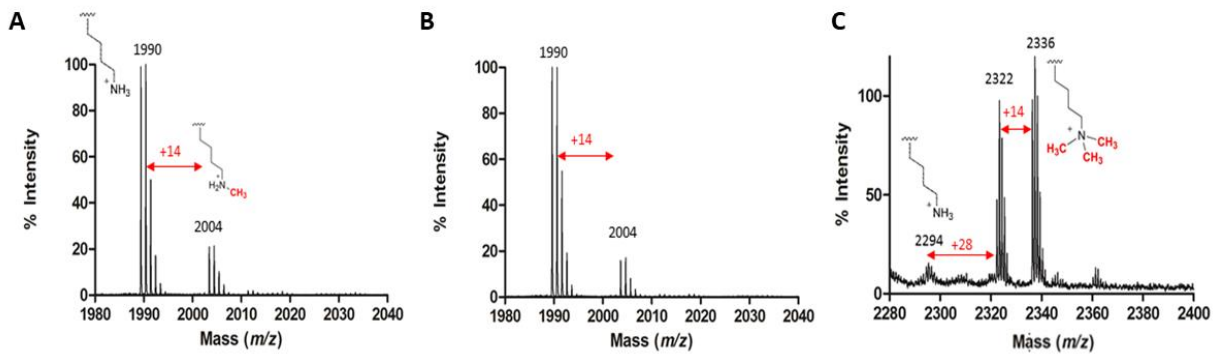


Figure 30. Spectres de masse MALDI-TOF des peptides substrats après réaction enzymatique A. Peptide H4 en présence de PR-Set7 sauvage et de SAM (3 h de réaction) B. Peptide H4 en présence de PR-Set7 Y334F et de SAM (3 h de réaction) C. Peptide H3 en présence de GLP sauvage et de SAM (30 minutes de réaction)

II.2.6. Préparation des substrats contenant l'histone H4 entière

Un des objectifs de notre étude étant d'observer la prise en charge d'un analogue du cofacteur par l'enzyme, on peut légitimement supposer qu'il s'accommodera moins efficacement dans le site actif que le cofacteur naturel sur l'enzyme sauvage. La faible activité enzymatique de l'enzyme sauvage observée précédemment vis-à-vis de son substrat peptidique pose donc problème. Afin d'anticiper de potentielles difficultés de détection, nous avons décidé d'étudier l'activité de l'enzyme sur un complexe d'histones entières de façon à nous rapprocher du substrat naturel de l'enzyme : le nucléosome.

Quatre histones recombinantes (H3, H4, H2A, H2B) produites chez *E. coli*, fournies par l'équipe de Raphaël Margueron, ont été utilisées pour préparer deux types de complexes : octamère d'histones, et nucléosome reconstitué. Le protocole que j'ai utilisé pour l'obtention de ces complexes est adapté de la méthode de reconstitution de particules cœur de nucléosome à partir d'histones recombinantes et d'ADN plasmidique¹⁵³. Cette méthode a pour avantage de permettre la production de complexes non méthylés qui pourront donc être substrats de notre enzyme. Brièvement, les différentes histones recombinantes, conservées en conditions dénaturantes sont rassemblées avec un léger excès molaire en faveur de H3 et H4 et mises à dialyser. Une chromatographie d'exclusion stérique permet de séparer les différents complexes d'histones que sont : le dimère H2A/H2B, l'hétérotétramère H3/H4 et l'octamère constitué de deux exemplaires de chacune des quatre histones (Figure 31A). Par la suite, la stœchiométrie du complexe est évaluée sur gel SDS-PAGE 15 % (Figure 31B). Pour terminer, la concentration de protéine a été déterminée par la méthode de Bradford.

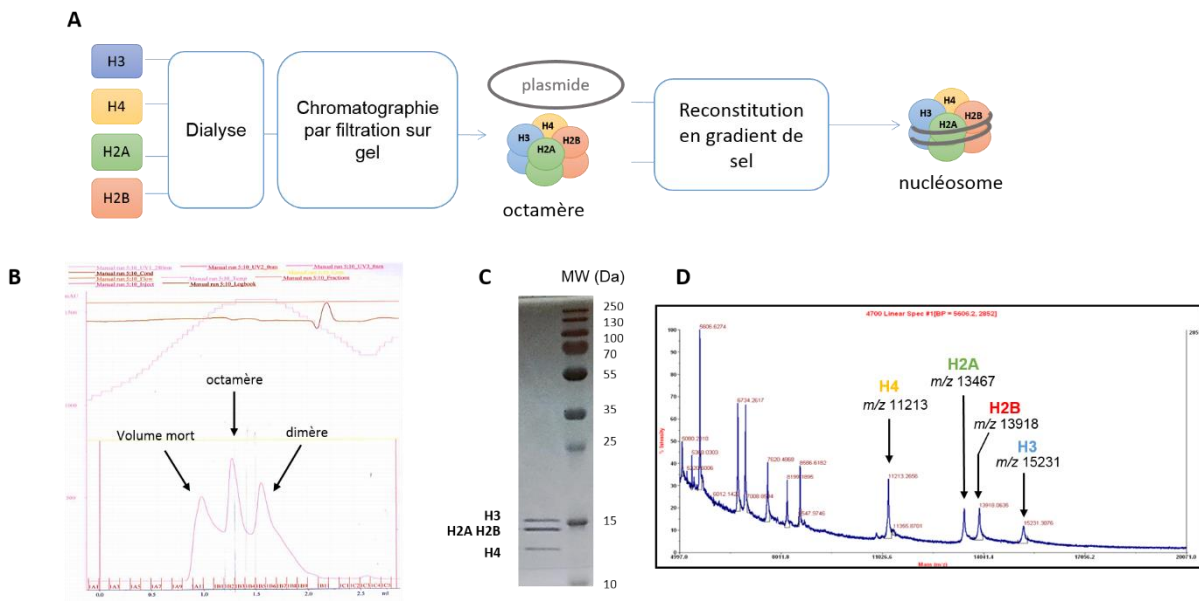


Figure 31. A. Principe général de la préparation de nucléosomes reconstitués. B. Chromatogramme d'exclusion stérique révélant les pics correspondant à l'octamère et au dimère d'histone. C. Gel SDS-PAGE 15 % d'acrylamide permettant de séparer les sous unités de l'octamère d'histone, H2A et H2B étant confondus. D. Caractérisation des protéines constituant le nucléosome par spectrométrie de masse MALDI-TOF en mode linéaire.

L'étape de séparation sur la colonne Superdex 200 montre la séparation de l'octamère d'histones des agrégats et du dimère formé pendant l'incubation des 4 types d'histone. Le gel d'électrophorèse coloré au bleu de Coomassie reflète l'abondance relative des différentes histones dans l'octamère, la bande correspondant à H2A H2B paraît approximativement deux fois plus intense que celle de H3 ou H4, donc le pic 2 du chromatogramme contient essentiellement de l'octamère. Il devrait aussi contenir du tétramère d'histone, les pics n'étant pas bien résolus. Le pic correspondant au tétramère H3/H4 n'apparaît pas sur le chromatogramme ce qui est cohérent avec le fait que nous avons introduit en excès molaires les histones H3 et H4 de façon à favoriser la création des complexes utilisant ces sous-unités et que son temps de rétention est très proche de celui de l'octamère d'histone.

La reconstitution de nucléosome est effectuée dans un second temps à partir de l'octamère précédemment formé et d'ADN plasmidique. Les deux produits sont ajoutés en quantité équimolaire à un tampon à forte concentration en sel (Tris-HCl 10 mM, 1mM EDTA, 2 M NaCl, 1mM β -mercaptoéthanol). La diminution progressive par dialyse de la concentration en sel favorise les interactions ADN / protéines permettant d'obtenir le nucléosome reconstitué. Dans un milieu sans sels, le nucléosome ainsi reconstitué est plutôt instable et se dégrade rapidement, ainsi il doit toujours être utilisé dans la semaine où il est produit. Le nucléosome est finalement caractérisé en spectrométrie de masse (Figure 31C), confirmant la présence de

chacun des histones et permettant d'observer les pics bien distincts de H2A et H2B respectivement à 13,7 et 13,9 kDa qui ne pouvaient être observés séparément par un gel d'électrophorèse en condition dénaturantes.

II.2.7. Evaluation de l'activité enzymatique des mutants de PR-Set7 en présence de complexes d'histones

L'activité de PR-Set7 et de ces quatre mutants ponctuels a été testée grâce au suivi de l'incorporation d'un groupement méthyle radiomarqué au tritium porté par la SAM*, avec l'aide d'Audrey Michaud dans le laboratoire du Dr. Raphaël Margueron. Ainsi, nous avons étudié deux substrats potentiels de PR-Set7, l'octamère d'histone et le nucléosome (Figure 32). Nous avons ainsi pu observer que seul le nucléosome était substrat de l'enzyme (pistes 1 à 6) tandis qu'aucun signal n'est observé en présence d'octamère (pistes 7 à 11). Ce résultat suggère que l'enzyme est ADN dépendante ce qui est tout à fait cohérent avec la littérature ²². Ceci explique au moins en partie la faible activité de l'enzyme observée sur substrat peptidique. De plus, de façon cohérente avec les tests sur peptide H4, nous observons que seuls l'enzyme WT et le mutant Y334F méthylent H4 et ce, dans des proportions équivalentes (pistes 1 et 3) alors que nous n'observons aucun signal dans les pistes correspondant aux autres mutants (pistes 2, 4, 5, 6). Cette observation correspond parfaitement à l'activité attendue pour chacun des mutants, la tyrosine 336 et le résidu aromatique en position 334 étant tous deux nécessaires pour la réaction enzymatique. Notons enfin, l'absence de méthylation pour le double mutant de PR-Set7. La structure de ce mutant étant la plus proche du mutant de GLP actif envers les analogues de SAM, nous chercherons à déterminer par la suite si une restauration de l'activité catalytique est possible en présence d'un analogue de SAM porteur d'une double liaison en β du sulfonium.

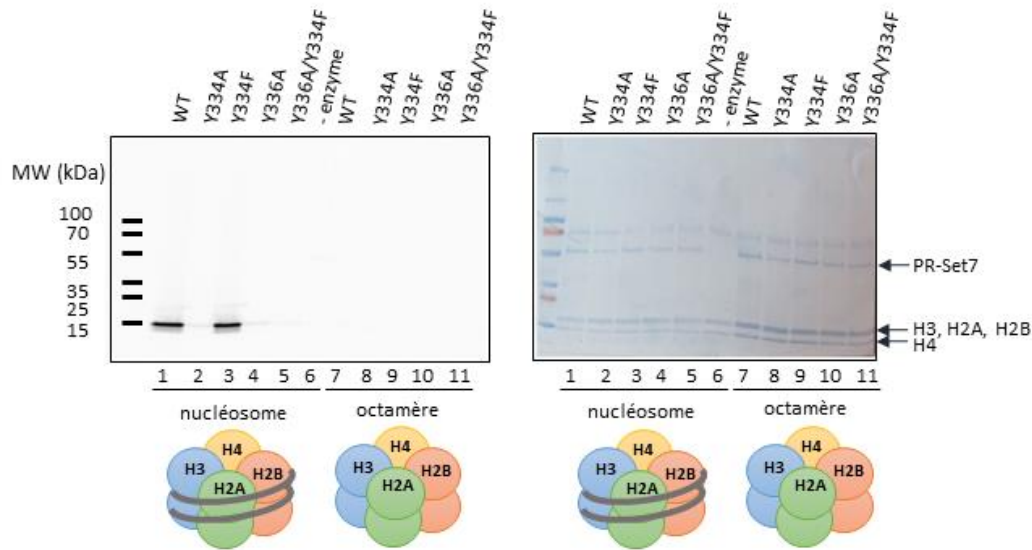


Figure 32. Analyse sur gel SDS PAGE de tests enzymatiques des différents mutants de PR-Set7 en présence de SAM tritiée au niveau du groupement méthyle et de nucléosome reconstitué (pistes 1 à 6) ou d'octamère d'histones (pistes 7 à 11) révélé par autoradiographie (à gauche) et contrôle de charge coloré au bleu de Coomassie (à droite).

II.3 Conclusion

Dans ce premier chapitre, nous avons donc mis en place tous les outils nécessaires à notre étude. Avec l'aide du Dr. Karine Guitot, j'ai synthétisé trois analogues de SAM comportant une fonction bioorthogonale transférable par une méthyltransférase ; trois autres analogues de SAM utilisés au cours de cette étude ont été synthétisés par Karine Guitot..

J'ai conçu les mutants de PR-Set7 en effectuant une superposition de structures de l'enzyme avec la lysine méthyltransférase GLP ce qui m'a permis d'identifier deux positions à muter dans le site actif et de générer quatre mutants ponctuels. J'ai ensuite mis au point les conditions de production et de purification de PR-Set7 permettant d'en produire des quantités satisfaisantes. Les mutants ont été produits et purifiés de la même façon. Enfin, j'ai produit des complexes d'histones pouvant servir de substrat de PR-Set7.

Après avoir produits l'ensemble des outils (enzyme, substrat et cofacteur) j'ai testé l'activité de méthylation des lots d'enzymes sauvage et mutées en utilisant un peptide fragment de l'histone H4. Cette caractérisation par spectrométrie de masse MALDI-TOF a révélé que le peptide n'était pas un bon substrat pour l'enzyme. Nous avons donc testé l'activité des enzymes sur des substrats plus complexes afin d'être plus proches de conditions physiologiques. Un test de

méthylation par autoradiographie a révélé que PR-Set7 fonctionnait exclusivement sur l'histone H4 au sein d'un nucléosome. L'enzyme est donc ADN dépendante conformément à ce qui est décrit dans la littérature²². C'est donc le nucléosome qui sera utilisé comme substrat dans la suite de l'étude.

Munis de ces outils, j'ai donc pu entreprendre le criblage des enzymes avec les différents cofacteurs en vue d'identifier un couple approprié pour le marquage chémoenzymatique de cibles de PR-Set7.

Chapitre 3

Criblage des analogues de SAM envers les mutants de PR-Set7

III. Criblage des analogues de SAM envers les mutants de PR-Set7

Les différents mutants de PR-Set7 et les analogues de SAM ayant été conçus sur des bases théoriques, il est nécessaire de cribler chacun des mutants avec chacune des petites molécules de façon à identifier des paires d'enzymes et de cofacteur permettant un transfert de la chaîne alkyle servant de rapporteur chimique. Nous avons testé les différentes combinaisons à l'aide de deux types de tests *in vitro* et deux types de substrats reposant *i)* sur le suivi en spectrométrie de masse MALDI-TOF de la fonctionnalisation d'un substrat peptidique représentatif de la séquence de l'histone méthylée (Figure 33A) *ii)* sur le suivi par western blot d'un substrat plus complexe, un nucléosome reconstitué qui, après marquage par le rapporteur chimique est ensuite fonctionnalisé par une réaction de chimie click avec un agent de détection fluorescent ou une biotine (Figure 33B).

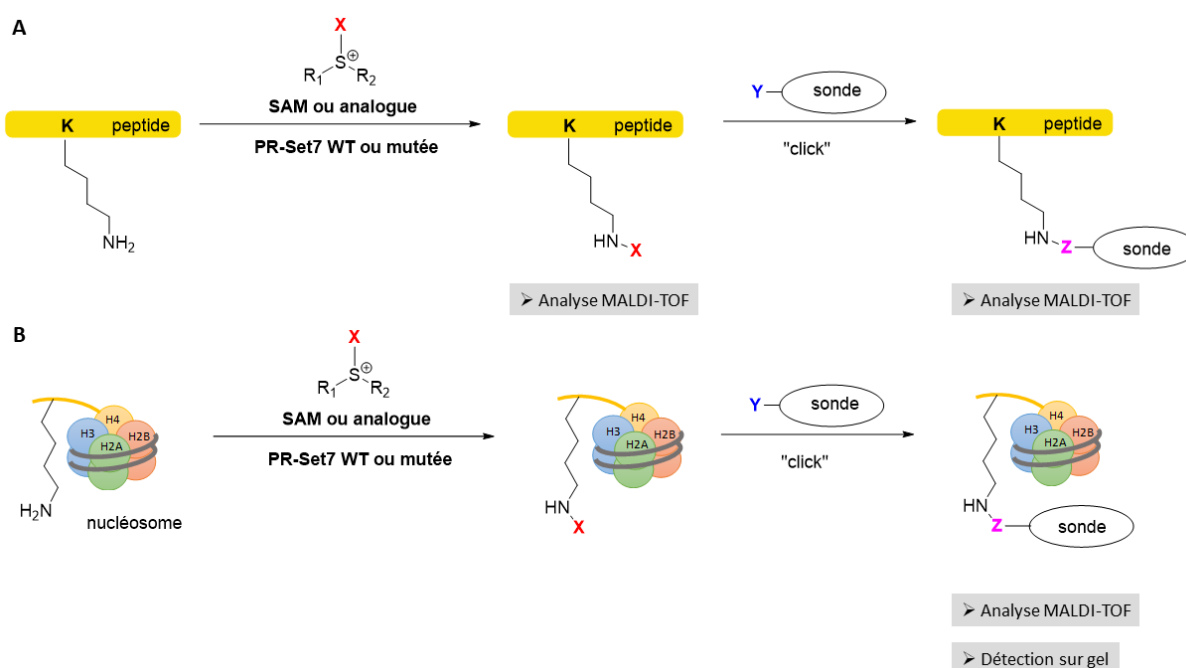


Figure 33. Principe des tests enzymatiques réalisés pour le criblage des paires enzyme/cofacteur. A. Test sur peptide et analyse en spectrométrie de masse MALDI-TOF. B. Test sur nucléosome et analyse par western blot ou spectrométrie de masse MALDI-TOF.

Les sondes utilisées dans cette étude sont des sondes commerciales possédant *i)* comme agent d'affinité ou de détection soit une biotine soit la tétraméthylrhodamine (TAMRA) pour une détection en fluorescence sur gel, *ii)* comme fonction chimique bioorthogonale soit un alcyne

vrai, soit un azoture pour la réaction de CuAAC, ou bien une cyclooctyne pour la réaction de SPAAC, *iii*) un linker de type PEG entre les deux. Elles sont présentées dans le Tableau 1.

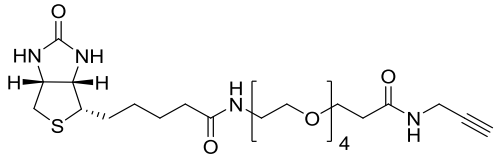
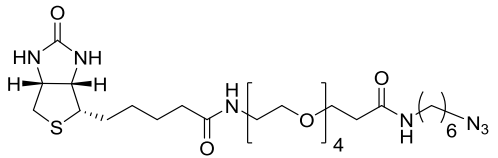
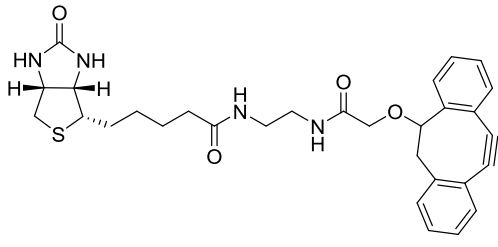
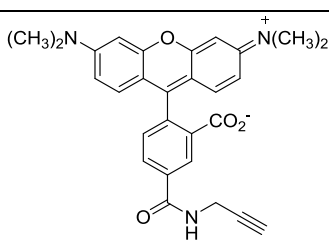
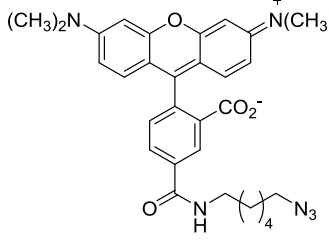
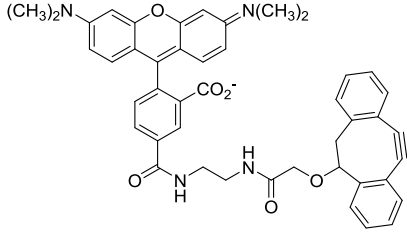
Structure	Nom de la sonde	Analogue de cofacteur compatible
	Biotine-PEG4- alcyne	Ab-SAM
	Biotine-PEG4- azide	Butynyl-SAM, Pentenyl-SAM, Hey-SAM, Pob-SAM
	Biotine-DIBO	Ab-SAM
	TAMRA-alkyne	Ab-SAM
	TAMRA-azide	Butynyl-SAM, Pentenyl-SAM, Hey-SAM, Pob-SAM
	DIBO TAMRA	Ab-SAM

Tableau 1. Structures chimiques des sondes commerciales utilisées pour les réactions de chimie click.

Avant de présenter les résultats obtenus pour PR-Set7, nous allons présenter les résultats préliminaires obtenus sur une enzyme modèle, l'HKMT GLP, utilisée comme contrôle positif dans tous nos tests.

III.1. Etude préliminaire réalisée avec l'HKMT GLP

Dans une étude préliminaire réalisée au laboratoire par le Dr. Karine Guitot, l'histone lysine méthyltransférase GLP sauvage ainsi que son mutant Y1211A ont été produits afin de disposer d'un contrôle positif dans tout le processus de mise en place de l'approche. Ces deux enzymes modèles ont été testées, comme dans l'article originel¹⁵⁴ vis-à-vis de la collection d'analogues de SAM décrite précédemment. Ce système modèle a également été utilisé pour tester de nouvelles molécules non présentées dans le cadre de cette étude.

III.1.1 Test sur substrat peptidique

Succinctement, l'expérience consiste en l'incubation pendant 3 h de l'enzyme ou du mutant, d'un peptide fragment de l'histone H3 (**peptide H3** = fragment 1-19 de l'extrémité *N*-ter de H3¹) et de l'analogue de SAM à tester, puis, après arrêt de la réaction, un échantillon du mélange réactionnel est analysé par spectrométrie de masse MALDI-TOF (Figure 34A). La Figure 34B présente des résultats caractéristiques obtenus en présence des différentes combinaisons GLP sauvage/mutée, cofacteur SAM naturel/Ab-SAM ou Allyl-SAM. Les résultats obtenus avec les autres analogues de SAM sont résumés dans le Tableau 2.

¹ Le peptide H3 utilisé dans cette étude comporte une biotine en *C*-ter comme précisé dans la partie Matériel et Méthodes. Il s'agit d'un peptide synthétisé et utilisé dans une étude précédente au laboratoire. Dans la présente étude, la présence de la biotine n'a pas été exploitée pour des étapes d'enrichissement, nous avons donc fait le choix de ne pas la représenter sur les schémas illustrant les études par spectrométrie de masse dans un souci de simplification de lecture.

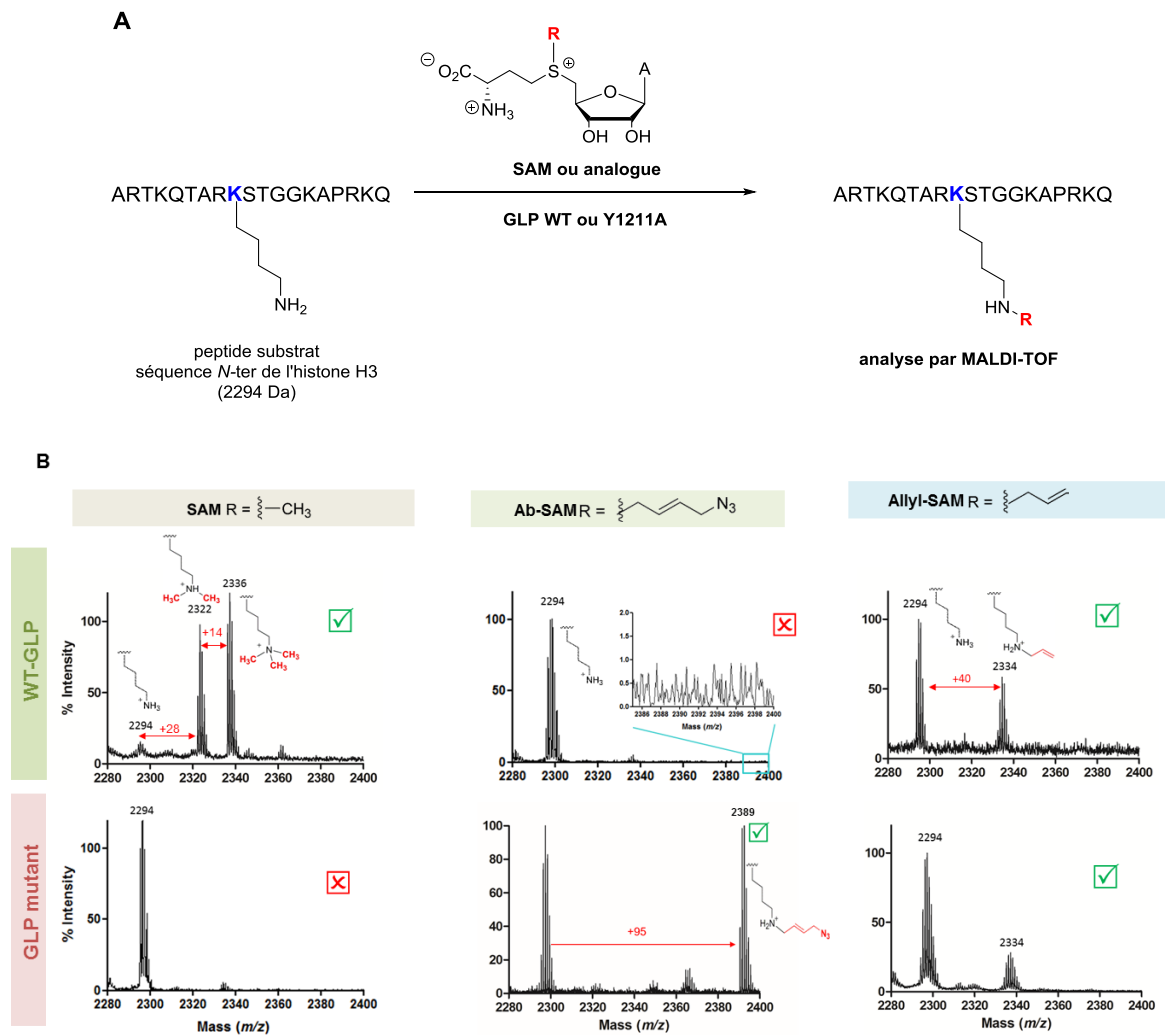


Figure 34. A. Tests enzymatique sur le peptide H3. B. Spectres de masse MALDI-TOF en mode réflecteur obtenus en présence des différentes combinaisons GLP sauvage/mutée, cofacteur SAM naturel/Ab-SAM ou Allyl-SAM.

L'analyse par spectrométrie de masse MALDI-TOF en mode réflecteur révèle que le mutant Y1211A de GLP est capable de prendre en charge avec plus ou moins d'efficacité chacun des différents analogues de SAM, sauf le Butynyl-SAM et le Pentenyl-SAM et de fonctionnaliser le peptide fragment de l'histone H3 avec des rapporteurs différents (Figure 34 et Tableau 2). En raison de sa taille et de l'absence de la double liaison en β du sulfonium il n'est pas étonnant que le Butynyl-SAM ne soit pas pris en charge. Le cas du Pentenyl-SAM est un peu plus délicat. Il est le deuxième plus petit analogue de SAM de notre série après Butynyl-SAM et possède la double liaison en β du sulfonium et pourtant, nous n'observons aucune fonctionnalisation de notre peptide H3. Par ailleurs, M. Luo décrit bien une prise en charge modérée de cet analogue par le mutant Y1211A de GLP¹¹². Cette divergence de résultats concernant la prise en charge

de l'analogue Pentenyl-SAM pourrait venir de la séquence du peptide substrat, notre peptide H3 et le peptide utilisé par le groupe de M. Luo ayant des séquences différentes.

Par ailleurs, à l'exception de l'analogue Allyl-SAM, qui est pris en charge à la fois par l'enzyme sauvage et dans une moindre mesure l'enzyme mutée, les différents analogues testés vis-à-vis de GLP sauvage n'ont montré aucune fonctionnalisation enzymatique. Contrairement à l'analyse quantitative possible sur les peptides méthylés par spectrométrie de masse MALDI-TOF^{151,152}, il convient de préciser que l'analyse est ici purement qualitative car, en fonction du groupement alkyle introduit sur le peptide, l'ionisation des peptides est probablement différente, notamment dans le cas du peptide comportant la fonction azide.



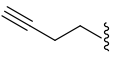
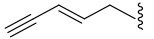
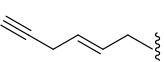
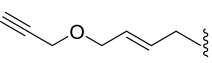
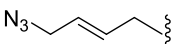
Nom	Cofacteur		Enzyme	
	R	Δm (Da)	GLP WT	GLP Y1211A
SAM		+ 14	+++	-
Allyl-SAM		+40	+	+
Butynyl-SAM		+52	-	-
Pentenyl-SAM		+64	-	-
Hey-SAM		+78	-	+
Pob-SAM		+108	-	+
Ab-SAM		+95	-	+

Tableau 2. Résultat du criblage des différents analogues de SAM testés sur les enzymes GLP sauvage et mutée. Le tableau indique qualitativement les couples qui ont donné lieu à l'observation d'une fonctionnalisation du peptide H3 dans les conditions du test enzymatique. Les incréments de masse observés sont indiqués.

Ainsi nos résultats sont en accord avec les résultats publiés dans l'article d'Islam *et al.*¹⁵⁴ et ce modèle a constitué un contrôle positif précieux pour vérifier la validité de nos conditions expérimentales. Nous avons également validé les conditions de réalisation de l'étape de fonctionnalisation ultérieure par réaction de CuAAC. La Figure 35 présente un exemple de spectre de masse MALDI-TOF en mode linéaire du peptide H3 après réaction enzymatique en présence de GLP mutée et d'Ab-SAM suivie d'une réaction de CuAAC avec la biotine-PEG4-

alkyne où l'on voit clairement l'incrément de masse correspondant à l'introduction de la chaîne alkyle ($\Delta m/z$ 95) et de la sonde biotine ($\Delta m/z$ 623).

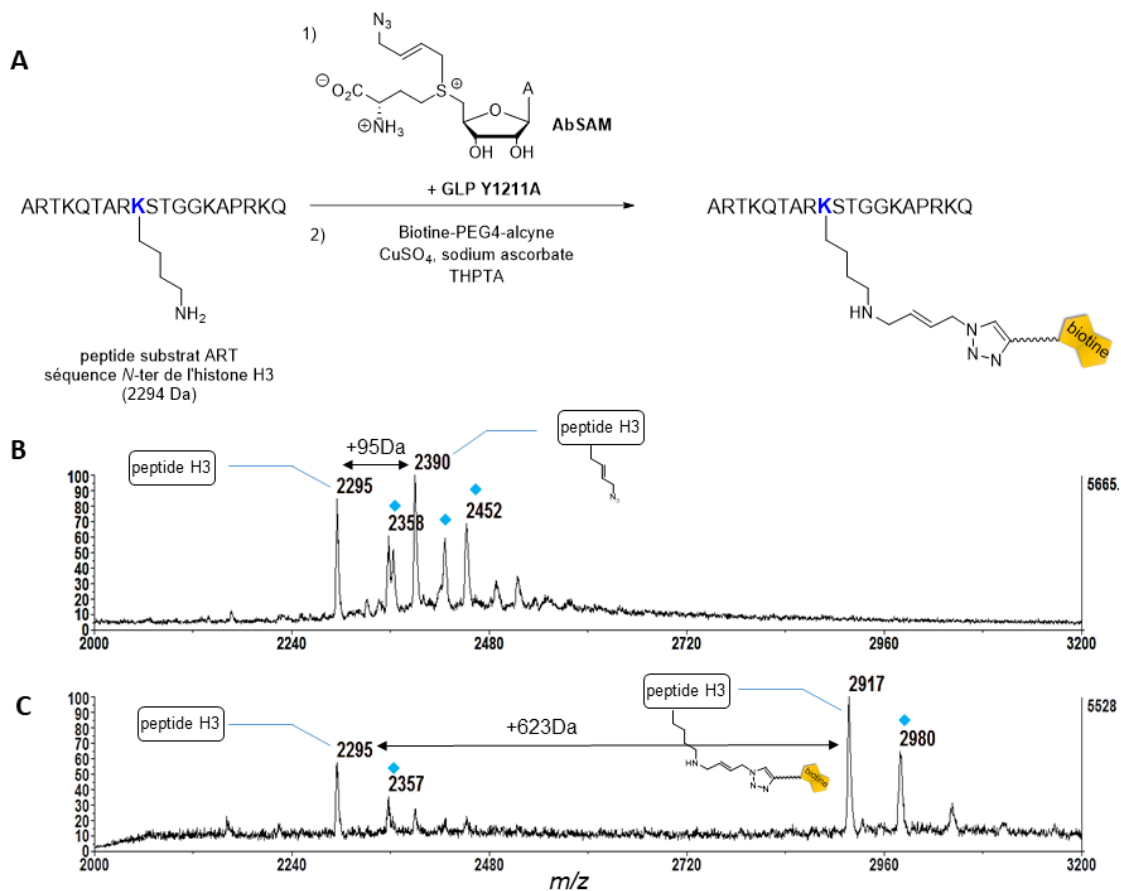


Figure 35. A. Test enzymatique sur le peptide H3 (extrémité N-ter de H3) en présence d'Ab-SAM et de GLP Y1211A, suivi de réaction de CuAAC en présence de biotine-PEG4-alkyne. Spectres de masse MALDI-TOF en mode linéaire d'échantillons d'essais enzymatiques suivis de réaction de CuAAC à $t = 0$ (B) et $t = 2$ h (C); Les signaux annotés ▲ sont des adduits au cuivre.

III.1.2 Test sur complexe d'histone : le tétramère

Les mises aux points de protocoles ont également été effectuées sur des substrats plus proches du substrat naturel à savoir un tétramère d'histones comportant deux histones H3 et deux histones H4 qui est un bon substrat pour cette enzyme. Après réaction enzymatique, la caractérisation directe par spectrométrie de masse n'est pas possible à moins d'envisager une digestion trypsique particulière assez lourde étant donné le haut contenu en résidus basiques des histones et ce type d'analyse n'est pas adapté dans un contexte de criblage. En effet l'incrément de masse après transfert enzymatique seul est trop petit pour une visualisation directe sur l'histone entière. En revanche, après fonctionnalisation par chimie click avec une sonde biotine comportant un bras PEG qui augmente la masse, il est possible de mettre en

évidence la fonctionnalisation de l'histone H3. La Figure 36B montre un exemple de spectre de masse MALDI-TOF en mode linéaire de l'histone H3 après réaction enzymatique en présence d'Ab-SAM et de GLP Y1211A suivie d'une fonctionnalisation par la biotine-DIBO où l'incrément de masse attendu est retrouvé ($\Delta m/z$ 641).

Par ailleurs, la détection par western blot a pu être reproduite à chaque fois dans les différents essais. Après réaction enzymatique avec les analogues de SAM, une étape de réaction CuAAC ou SPAAC est réalisée avec la sonde biotinylée ou fluorescente appropriée selon l'étiquette bioorthogonale introduite. La détection de la fonctionnalisation est ensuite effectuée après migration sur gel d'électrophorèse par western blot ou par fluorescence. La Figure 36C montre un western blot réalisé après réaction avec l'Ab-SAM et la sonde biotine-DIBO puis révélation par la HRP-streptavidine montrant un signal spécifique de l'histone H3 en présence d'Ab-SAM.

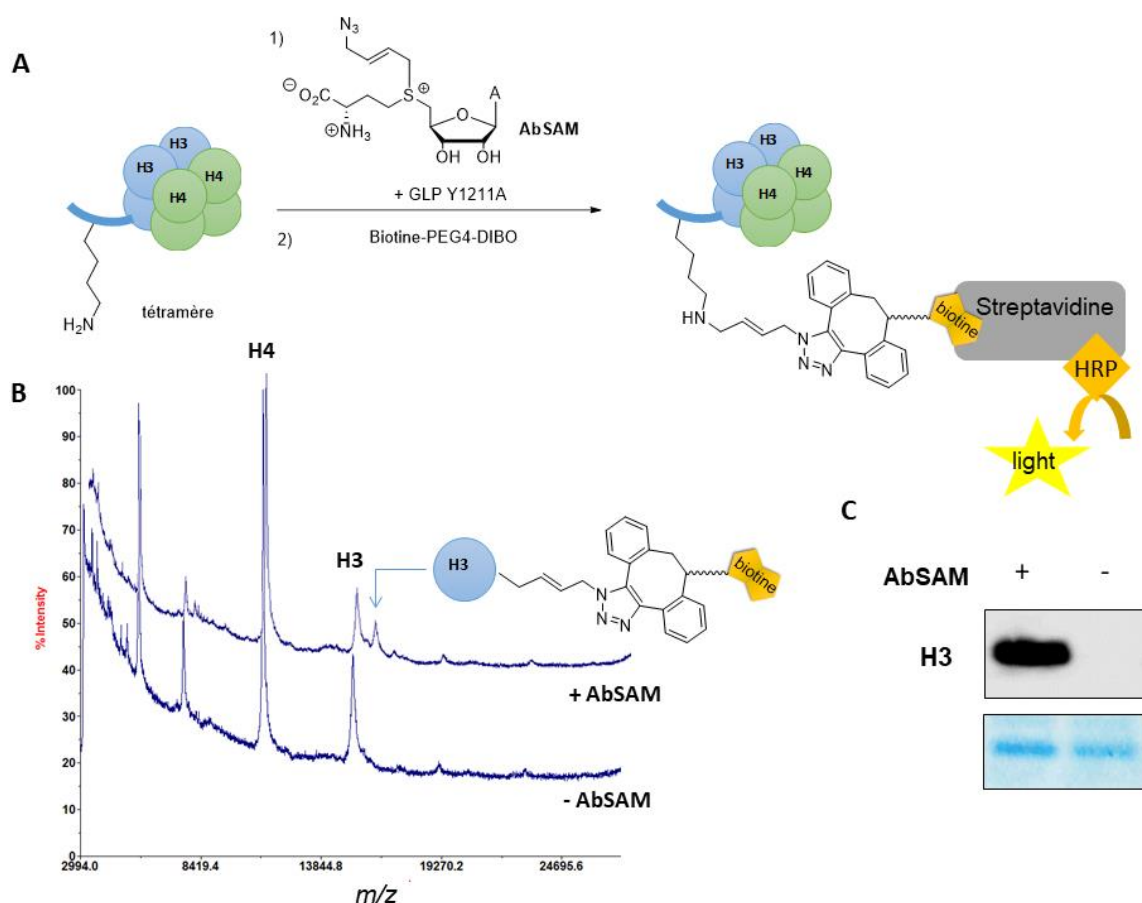


Figure 36. A. Test enzymatique sur tétramère d'histone en présence d'Ab-SAM et de GLP Y1211A, suivi de réaction de SPAAC en présence de biotine-DIBO. B. Spectres de masse MALDI-TOF en mode linéaire d'échantillons d'essais enzymatiques suivis de réaction de SPAAC. C. Western blot.

J'ai pour ma part systématiquement introduit dans toutes mes expériences avec PR-Set7 un contrôle positif impliquant GLP que ce soit en spectrométrie de masse ou dans les expériences de détection par western blot.

III.2. Etude réalisée avec PR-Set7 et ses mutants

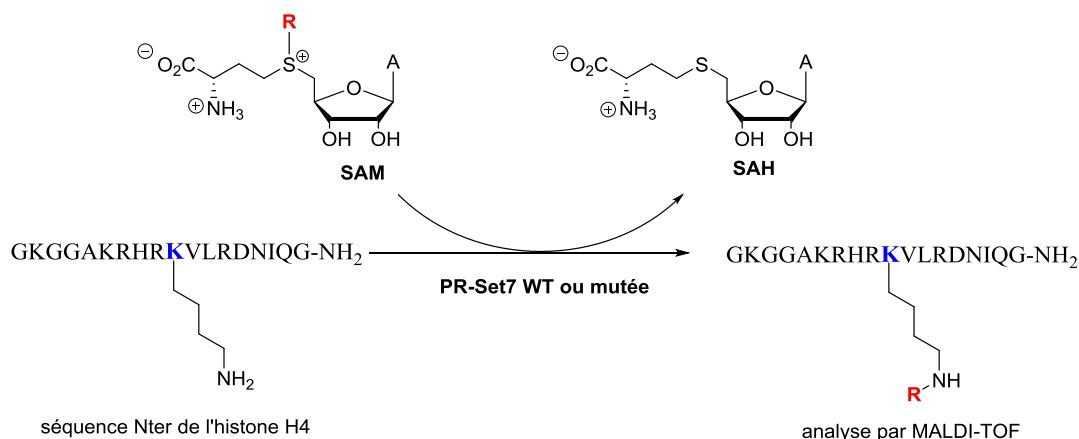
Etant assuré de maîtriser les protocoles dans ce type d'approche, j'ai alors testé la collection d'analogues de SAM sur chacun des mutants de PR-Set7 que j'ai précédemment produit à l'aide des deux tests enzymatiques présentés ci-dessus légèrement adaptés aux conditions de PR-Set7.

III.2.1 Test sur substrat peptidique

Sur la base du même protocole, mais en adaptant les conditions du test enzymatique optimisé dans le paragraphe précédent, nous avons testé la série d'analogues de SAM sur PR-Set7 et chacun de ses mutants en utilisant notre peptide fragment de l'histone H4 contenant la lysine 20 (peptide H4 : résidus 11 à 28 de l'extrémité *N*-ter de H4). Malheureusement, ni l'enzyme sauvage ni les mutants n'ont montré d'activité envers les différents analogues de cofacteurs (Tableau 3).

Ces résultats sont décevants mais assez cohérents avec nos précédentes observations pour le test enzymatique en conditions natives où nous avons déjà remarqué une faible activité enzymatique mesurée pour PR-Set7 sauvage lorsqu'elle était incubée avec la SAM (Figure 34). Nous avons également essayé de faire cette détection par spectrométrie de masse MALDI-TOF sur peptide mais cette fois après les deux étapes de fonctionnalisation (réaction enzymatique + réaction de chimie click) avec l'espoir que l'introduction de la sonde permettrait une meilleure désorption/ionisation des échantillons que ce que l'on peut observer avec par exemple les peptides comportant la fonction azoture. Malheureusement, aucune fonctionnalisation n'a pu être détectée.

Contrairement à GLP, PR-Set7 a clairement une activité catalytique moins robuste dans ce type de test *in vitro* utilisant un peptide comme substrat et bien que cette méthode de criblage nous aurait largement simplifié les choses, nous avons dû abandonner l'idée d'identifier les couples mutants/analogues par cette stratégie.



Cofacteur		GLP WT	GLP Y1211A	PR-Set7 WT	PR-Set7 Y336A	PR-Set7 Y334A	PR-Set7 Y334F	PR-Set7 Y336A /Y334F
Nom	R	Peptide Histone H3		Peptide Histone H4				
SAM		+++	-	+	-	-	+	-
Allyl-SAM		+	+	-	-	-	-	-
Butynyl-SAM		-	-	-	-	-	-	-
Pentenyl-SAM		-	-	-	-	-	-	-
Hey-SAM		-	+	-	-	-	-	-
Pob-SAM		-	+	-	-	-	-	-
Ab-SAM		-	+	-	-	-	-	-

Tableau 3. Résultat du criblage des différents analogues de SAM testés sur les enzymes PR-Set7 sauvage et mutées.

III.2.2 Test sur complexe d'histone : le nucléosome reconstitué

Par la suite nous avons donc tenté d'évaluer la prise en charge des analogues du cofacteur par PR-Set7 en utilisant des complexes d'histones comme substrat, ici, le nucléosome reconstitué. Le principe du criblage reste identique à celui décrit pour GLP. Après réaction enzymatique avec les analogues de SAM (Tableau 2), une étape de réaction CuAAC ou SPAAC est réalisée avec différents agents biotinylés ou fluorescents. La détection de la fonctionnalisation est ensuite effectuée après migration sur gel d'électrophorèse par western blot ou par fluorescence.

Ainsi pour évaluer l'efficacité de prise en charge des différents analogues nous avons testé, un par un, chacun des cofacteurs avec l'ensemble des mutants de PR-Set7. Etant donné que le mutant Y1211A de GLP prend en charge différents analogues de cofacteur, pour chaque expérience, nous ajoutons systématiquement en tant que contrôle positif, la paire constituée de GLP Y1211A et de l'analogue à tester.

La Figure 38 est caractéristique du type de résultats que nous avons obtenus. Elle montre les résultats d'un criblage des différents mutants de PR-Set7 avec l'analogue Ab-SAM. Dans la Figure 38A, après l'incubation des différents lots d'enzyme avec Ab-SAM, la réaction de CuAAC a été effectuée avec la sonde biotine-PEG4-alcyne. Les différents échantillons ont alors été analysés par western blot en utilisant un conjugué HRP-streptavidine. Dans cette figure, le contrôle positif avec le mutant GLP Y1211A montre, comme attendu, la fonctionnalisation de l'histone H3 (piste 2). En revanche, aucun signal évident n'a pu être détecté dans la zone de l'histone H4 montrant une absence de prise en charge d'Ab-SAM par les différents mutants de PR-Set7. D'autre part, nous observons des signaux intenses à la distance de migration de PR-Set7 et dans des zones correspondant à des impuretés co-purifiées avec PR-Set7. Ce type de marquage a été observé à de nombreuses reprises en reproduisant les expériences mais nous ne sommes pas parvenus à l'expliquer. Bien que la thèse de l'auto-fonctionnalisation soit envisageable, puisqu'il est connu que les PKMT sont capables de s'autométhyliser, nous privilégions celle de signaux artefactuels qui ont d'ailleurs déjà été rapportés dans la littérature^{155,156}. En effet, il semble que l'utilisation d'analogues de SAM soit souvent accompagnée de signaux non spécifiques. Et bien que l'origine de ce signal ne soit pas expliquée, il est recommandé d'optimiser pour chaque analogue, sa concentration et son temps d'incubation avec l'enzyme pour minimiser ce signal^{155,156}. Un second argument laissant penser à un signal non-spécifique est l'apparition de signal au niveau des histones dans la condition en absence d'enzyme (Figure 38A, piste 3 et annexe S1).

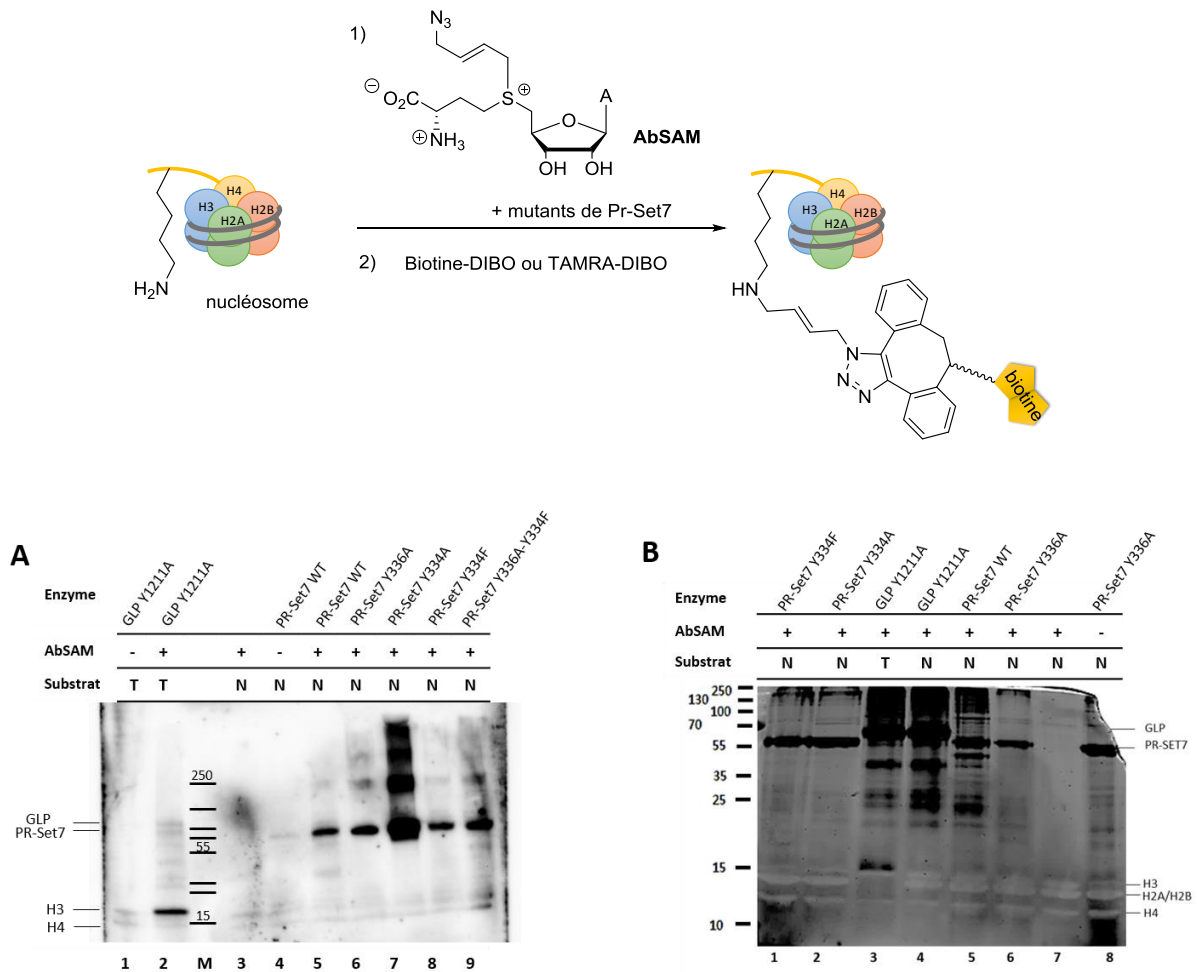


Figure 38. Analyse sur gel SDS-PAGE de la fonctionnalisation des histones par l’analogue Ab-SAM. A. Par Western Blot. B. Par Fluorescence (N : nucléosome reconstitué, T : tétramère d’histone).

Dans la Figure 38B, les différents lots d’enzyme ont aussi été incubés avec l’Ab-SAM et la réaction de SPAAC a été effectuée avec la sonde DIBO-TAMRA. Après migration sur gel SDS-PAGE, les différents échantillons ont été analysés par l’émission de la fluorescence dans le gel. Dans cette figure, nous constatons à nouveau la fonctionnalisation de l’histone H3 (piste 3) en présence de l’enzyme GLP Y2011A et d’Ab-SAM. Malheureusement, là encore, aucun signal dans la zone de l’histone H4 n’a permis de conclure qu’un mutant de PR-Set7 pouvait prendre en charge l’analogue Ab-SAM. Nous observons à nouveau ici de nombreux signaux à la distance de migration des enzymes (GLP et PR-Set7) et des impuretés co-purifiées avec elles. Enfin, un nouveau type de signal non spécifique apparaît, dans la condition contrôle en absence de cofacteur. Dans ces conditions, le signal observé pourrait être expliqué par une réaction secondaire où l’alcyne contraint de la sonde réagirait avec les groupements thiols de cystéines libres¹⁵⁷. Bien que cette réaction secondaire puisse théoriquement être prévenue par l’alkylation des cystéines, l’ajout d’étapes supplémentaires alors que nous n’avons pas résolu l’étape de

transfert enzymatique, ne nous a pas semblé opportun. De plus, nous avons observé que la sonde TAMRA était la source d'un très fort bruit de fond. Pour ces raisons, la réaction de type SPAAC a été mise de côté au profit de la cycloaddition catalysée par le cuivre (CuAAC).

Au cours de cette phase de criblage, toutes les combinaisons d'enzymes et d'analogues de cofacteur ont été testées. Nous avons aussi testé différentes conditions expérimentales du test enzymatique. Pour résumer l'influence de différents tampons (Tris, Hepes, Glycine), de pH plus ou moins basiques (7,4 à 9,8), de différentes durées d'incubation (1 h à 15 h), de différentes températures (20 °C à 37 °C) a été évaluée. Et, bien que certaines conditions soient légèrement plus efficaces (Tampon glycine pH 9,8, 3 h d'incubation à 25°C) que d'autres pour promouvoir l'activité de méthylation par l'enzyme, aucune n'a permis de révéler de signaux renseignant sur une prise en charge d'un des analogues de SAM par un quelconque mutant de PR-Set7. De manière générale, nous avons identifié lors de nos premiers essais enzymatiques que PR-Set7 est bien moins active que GLP et nous attendions à avoir des résultats moins spectaculaires que dans le cas de GLP mais nous espérons trouver une combinaison efficace pour mettre en œuvre l'approche.

De plus, les nombreuses sources de signaux non spécifiques diminuent fortement la résolution de nos images et empêchent la détection de potentiels faibles signaux. Ces sources de signaux non-spécifiques liées soit à l'analogue de SAM soit aux conditions de la réaction de chimie click réduisaient nos chances d'identification d'une paire d'enzyme – cofacteur efficace pour étiqueter les substrats de PR-Set7. Nous avons essayé d'y remédier en mettant en œuvre d'autres protocoles techniques que nous présenterons dans le chapitre III.3.

En l'absence de résultats avec cette série d'analogues de SAM, nous avons souhaité tester l'approche avec un analogue de SAM moins volumineux pour rentrer dans le site catalytique de PR-Set7. Nous nous sommes donc intéressés à la synthèse d'un analogue de SAM avec une chaîne alkyle transférable plus courte, tout en respectant les contraintes de conception des analogues de SAM : double ou triple liaison en position β du sulfonium.

III.3. Etude en présence d'un analogue de cofacteur SAM sélénié minimaliste : ProSeAM

De par sa petite taille, l'analogue de SAM porteur d'un groupement propargyle (ProSAM) est le plus petit analogue de SAM comportant une fonction bioorthogonale répondant aux critères nécessaires à une réaction de transalkylation par une HKMT. Toutefois, c'est aussi un composé

extrêmement instable à pH physiologique avec une demi-vie de l'ordre de la minute, ce qui le rend peu propice à son utilisation dans un contexte biologique. En plus des 3 voies de dégradation des analogues de la SAM décrites dans la partie II expliquant leur demi-vie de l'ordre de l'heure, dans le cas du ProSAM, une quatrième voie de dégradation s'ajoute et s'avère être la principale faisant drastiquement chuter la demi-vie de ce petit analogue de SAM par rapport à de plus grosses molécules. Cette voie supplémentaire de dégradation passe transitoirement par la formation d'un allène et aboutit à la formation de la cétone par hydratation de l'allène¹⁵⁸ (Figure 39).

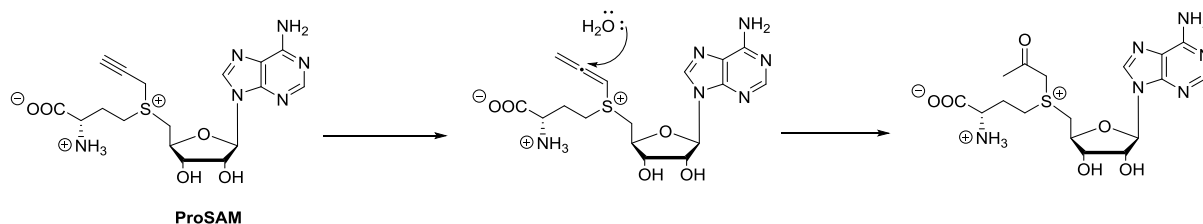


Figure 39. Voie de dégradation majoritaire du ProSAM

La synthèse de cet analogue a néanmoins été entreprise au laboratoire mais conformément aux travaux décrits, les tentatives se sont révélées décevantes car le ProSAM est très difficile à isoler et ne se conserve pas. Dans le même article décrivant la stabilité de cette molécule, il a été montré que la substitution de l'atome de soufre par un atome de sélénium augmentait nettement la stabilité de l'alcyne. En effet, la demi-vie de ce nouvel analogue, est approximativement 60 fois plus longue que celle de ProSAM. En effet, l'ion sélénonium réduit l'acidité des protons des carbones sp^3 adjacents par comparaison à son analogue sulfonium, ce qui permet d'éviter la formation de l'allène et par conséquent la perte de la fonction alcyne indispensable à la réaction de CuAAC. En outre, il apparaît que dans certaines réactions enzymatiques, l'analogue sélénié du SAM, sélénoSAM ou SeAM, est un meilleur donneur de méthyle que son homologue SAM¹⁵⁹. En effet, la liaison Se-C formée est plus faible que la liaison S-C, ce qui rend la SeAM plus réactive vis-à-vis de la réaction de transméthylation. Notons que ce dernier point est généralisable aux versions séléniées de l'ensemble des analogues de SAM et qu'ils sont généralement considérés comme des cofacteurs plus adaptés pour les réactions de transalkylation effectuées par les HKMT, facilitant *in fine* l'identification des protéines substrats de ces enzymes. La synthèse de ces composés est toutefois plus délicate que celle des analogues soufrés et la synthèse d'une série d'analogues séléniés de la SAM a fait l'objet du stage de M2 de Joana Dagorret sous la supervision du Dr. Karine Guitot à l'ICMMO.

Parmi les analogues synthétisés, la ProSeAM a été préparée et m'a été fournie pour effectuer la suite de mon travail.

III.3.1. Synthèse et caractérisation de la ProSeAM

La synthèse est basée sur une alkylation de la *Se*-adénosyl-L-sélénohomocystéine (SeAH) similaire à celui décrit pour la synthèse des analogues de SAM. Le SeAH est lui-même obtenu à partir de la L-méthionine et de la 5'-chloro-5'-désoxyadénosine. La synthèse du ProSeAM a été réalisée selon le schéma 2 en apportant quelques modifications à des protocoles décrits dans la littérature¹⁶⁰ (schéma 2).

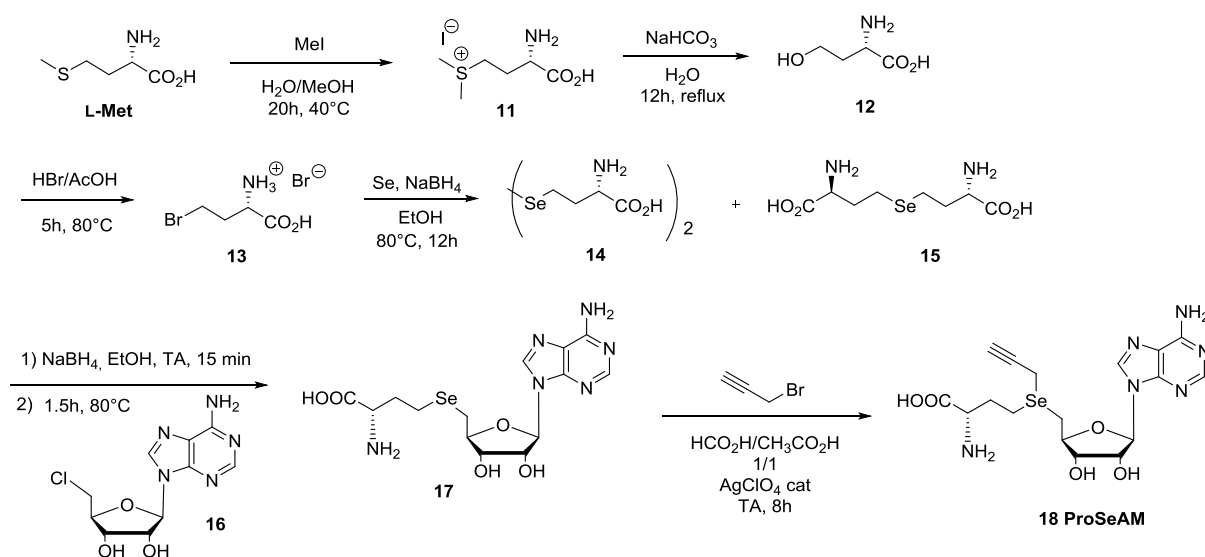


Schéma 2. Synthèse de l'analogue sélénié minimaliste ProSeAM

Dans un premier temps, la L-méthionine est traitée avec de l'iodure de méthyle dans de l'eau et du méthanol, pour former le sulfonium **11** isolé avec un rendement de 97 %. Le produit **11** est ensuite engagé dans une réaction de substitution du sulfure de diméthyle par un ion hydroxyde en milieu basique. Après précipitation, la L-homosérine **12** pure est obtenue avec un rendement de 80 %. Une réaction de bromation en présence d'acide bromhydrique est ensuite réalisée sur la L-homosérine **12** pour conduire à l'acide (2*S*)-amino-4-bromobutanoïque **13** avec 76 % de rendement.

Celui-ci est traité sous atmosphère inerte par Na₂Se₂, généré *in situ* à partir d'une poudre de sélénium et de NaBH₄. Après traitement acide et tentative de purification sur résine échangeuse d'ions, le produit **14** (67 %) est obtenu en mélange avec le sélénioéther **15** comme produit secondaire. Ce mélange sera engagé dans l'étape suivante. Afin d'obtenir la SeAH **17**, le pont diséléniure est réduit en présence de NaBH₄ et l'anion séléniure ainsi formé est engagé

directement dans une réaction avec la 5'-chloro-5'-désoxyadenosine **16**. Après purification par HPLC en phase inverse, la SeAH **17** est obtenue avec 53 % de rendement. Celle-ci est ensuite alkylée en présence de bromure de propargyle dans un mélange acide acétique/acide formique en présence de perchlorate d'argent pendant 8 h. De façon intéressante, la réaction, suivie par HPLC analytique, est beaucoup plus rapide que dans le cas des analogues soufrés. Après purification par HPLC en phase inverse, le ProSeAM **18** est obtenu avec 8 % de rendement et il a été possible de séparer partiellement les deux épimères (Figure 40).

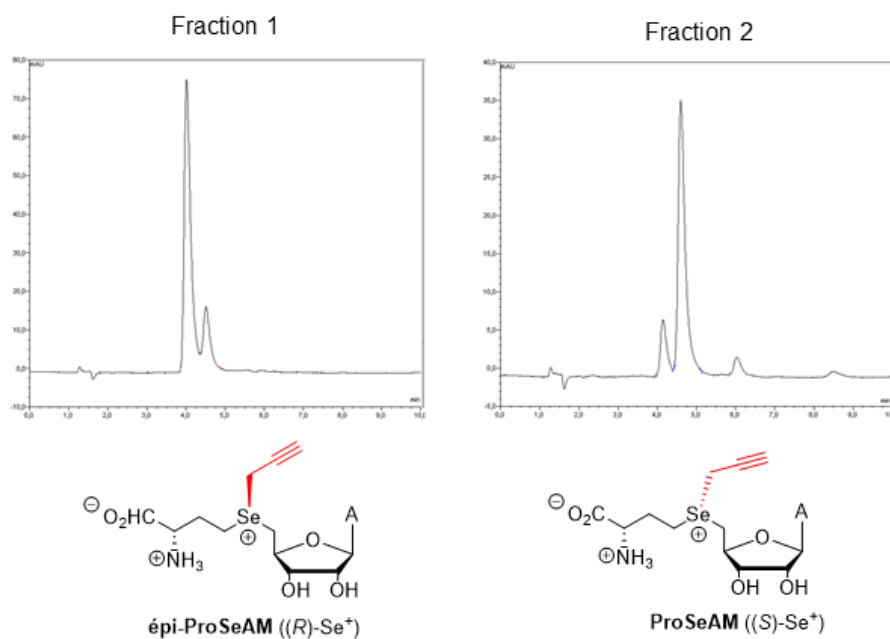


Figure 40. Les deux fractions de ProSeAm analysées par HPLC en phase inverse. L'identification de chaque épimère a été effectuée sur la base des tests enzymatiques décrits ci-après.

De façon intéressante, les deux épimères ont pu être caractérisés par leur activité enzymatique. Ainsi, dans le test enzymatique sur le peptide H3, les spectres de masse MALDI-TOF montrent que le peptide substrat est mono et di-alkylé par l'enzyme GLP sauvage en présence du ProSeAM de la fraction 2, preuve qu'il s'agit de la forme (*S*) alors que l'activité enzymatique est très faible avec la fraction 1, indiquant que le diastéréoisomère majoritaire de cette fraction est (*R*). La faible activité résiduelle s'explique probablement par le fait que les deux épimères n'ont pas été parfaitement séparés (Figure 41).

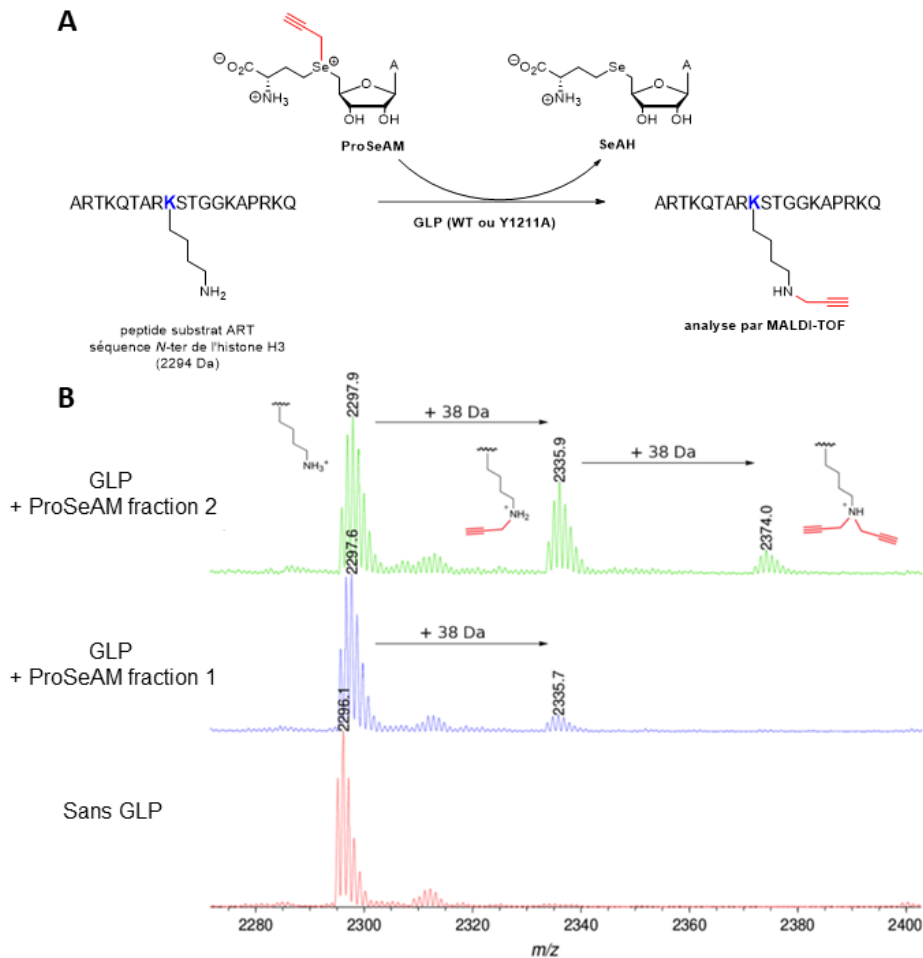


Figure 41. A. Réaction d'alkylation catalysée par l'enzyme GLP WT B. Spectres de masse MALDI-TOF du peptide substrat (peptide H3) après réaction enzymatique en présence des fractions 1 et 2 de l'analogue de cofacteur ProSeAM.

Ce résultat a également été confirmé par western blot avec le test sur octamère (Figure 42) en présence de l'enzyme GLP sauvage et mutée. En effet, on observe que le ProSeAM de la deuxième fraction d'élution F2 est systématiquement mieux prise en charge que celui de la fraction F1 indiquant que le diastéréoisomère (*S*) pris en charge par GLP se trouve effectivement dans la fraction F2.

Dans cet essai, nous remarquons également que ProSeAM est bien mieux pris en charge par l'enzyme sauvage que l'enzyme mutée. Ceci est cohérent avec le résultat observé lors du test sur peptide de la prise en charge de l'allylSAM par GLP sauvage et mutée (Figure 34).

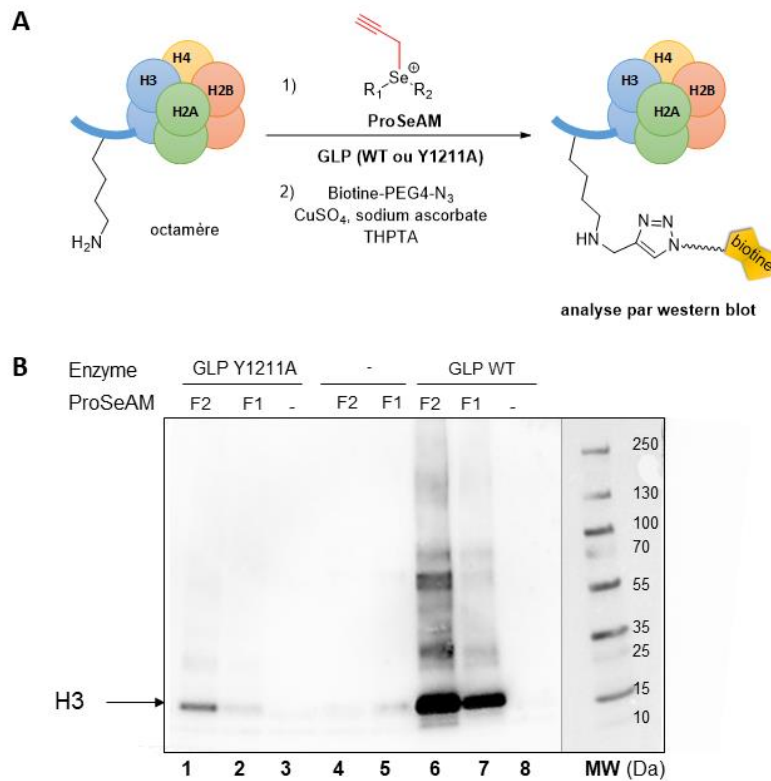


Figure 42. A. Réaction de propargylation catalysée par l'enzyme GLP WT ou mutée. B. Analyse par western blot (HRP-streptavidine) de l'octamère substrat après réaction enzymatique en présence de chaque fraction du cofacteur, puis réaction de CuAAC avec la biotine-PEG4-azide.

Finalement, la fraction 2 récoltée en HPLC correspond à l'épimère fonctionnel de l'analogue de cofacteur, à savoir celui ayant la configuration (*S*) au niveau du sélénium (Figure 40).

III.3.2. Evaluation de la ProSeAM sur PR-Set7 et ses mutants

Le nouvel analogue de cofacteur, ProSeAM a donc été testé en présence de PR-Set7 et ses mutants. Malgré la faible activité de PR-Set7 sur substrat peptidique (Figure 34), nous avons tenté d'évaluer ce nouvel analogue en utilisant le peptide H4 mais aucune fonctionnalisation n'a pu être observée. Le ProSeAM a donc été évalué sur le nucléosome reconstitué. Après une incubation de 3 h du cofacteur et de l'enzyme en présence de nucléosome reconstitué, la réaction de chimie click en solution a été effectuée avec la sonde biotine-PEG4-azide et les histones ont été séparés sur gel SDS-PAGE et analysés par western blot à l'aide d'HRP-streptavidine (Figure 43A.). La Figure 43B. correspond au contrôle de charge et la Figure 43C est une superposition des deux western blot. On observe un signal au niveau de la distance de migration de l'histone H4 dans les conditions où la ProSeAM a été incubée avec l'enzyme sauvage et le mutant Y334F, à savoir le mutant gain de fonction qui avait montré de l'activité en présence de SAM tritiée et dans le test peptidique (Figure 43A pistes 3 et 5). Ce signal est

absent en absence de cofacteur et en absence d'enzyme (Figure 10A pistes 1 et 2), ce qui prouve que le signal est bien dépendant de l'activité de l'enzyme.

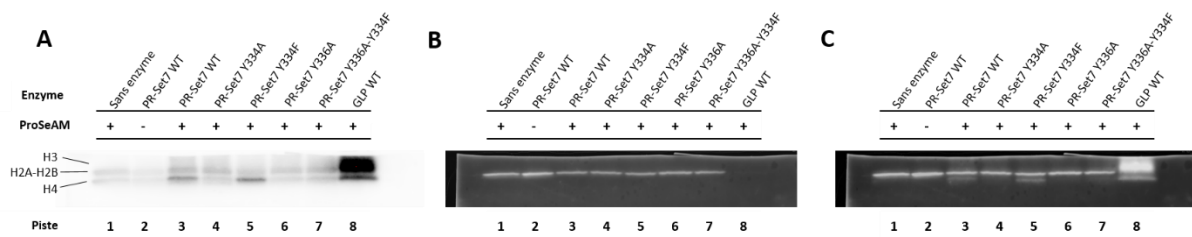


Figure 43. Résultat du test de fonctionnalisation par PR-Set7 de l'histone H4 en présence de ProSeAM. A. Western blot anti-biotine. B. Western blot anti-H2B servant de contrôle de charge. C. Superposition des western blots anti-biotine et anti-H2B.

Afin de vérifier que le signal observé n'était pas artéfactuel, la réaction a été reproduite et le ProSeAM a été incubé avec deux quantités différentes des différents mutants de PR-Set 7 et pour chacune de ces conditions, 5 équivalents de SAM qui va entrer en compétition avec ProSeAM, ont été ajoutés. Le western blot est présenté dans la Figure 44. On observe à nouveau un marquage de l'histone H4 dans le cas du mutant PR-Set7 Y334F mais pas dans le cas des deux autres mutants testés. Par ailleurs, ce signal disparaît en présence du compétiteur ce qui conforte la spécificité du marquage de H4 par transfert enzymatique. Nous observons aussi un effet de la quantité d'enzyme mise ce qui va aussi dans le sens d'une fonctionnalisation de H4 par voie enzymatique. Pour finir, le fait que le signal détecté sur le gel au niveau d'autres protéines (impuretés co-purifiées avec l'enzyme) ne diminue pas en présence de l'inhibiteur confirme nos soupçons d'un très fort signal non spécifique, qu'il faudra prendre en compte lors du passage à l'étude protéomique.

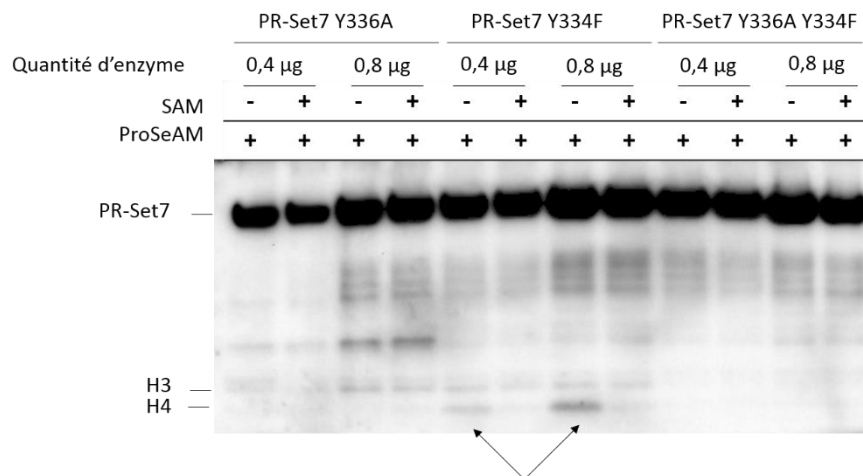


Figure 44. Caractérisation du transfert enzymatique du groupement propargyle par le mutant Y334F de PR-Set7. Des flèches noires montrent le signal correspondant à la fonctionnalisation de l'histone H4.

Même si le marquage n'est pas aussi important que pour GLP, le ProSeAM est donc le seul analogue de SAM pris en charge par un mutant de PR-Set7. La taille plus restreinte du propargyle pourrait expliquer ce résultat. C'est donc cet analogue qui sera choisi pour la suite de cette étude.

III.4. Mise au point de nouvelles techniques pour faciliter l'analyse de la fonctionnalisation des histones

Une des grosses difficultés de notre protocole résidait dans le fait que les étapes de test enzymatique et de fonctionnalisation par chimie click sont indissociables lorsque l'on travaille sur nucléosome contrairement à l'analyse d'un substrat peptidique simple. Ceci entraînait donc nécessairement une impossibilité de déterminer si l'absence de signal venait plutôt de la non prise en charge du cofacteur ou d'un problème lié à l'étape suivante de fonctionnalisation par une sonde. Afin de résoudre ce problème nous avons développé deux méthodes permettant de déconnecter les deux étapes.

III.4.1. Les réactions de chimie click sur membrane de PVDF

Au cours de la stratégie de marquage chémoenzymatique suivi de la réaction de chimie click, plusieurs facteurs principaux nous ont semblé pouvoir être à la source des signaux non spécifiques observés en western blot. Le premier viendrait de l'étape de fonctionnalisation enzymatique et serait dû à l'excès d'analogue de cofacteur incubé avec l'enzyme. En admettant que l'analogue de cofacteur ait une légère réactivité intrinsèque, il pourrait être responsable de ce signal non spécifique, et ce d'autant plus s'il est mis en large excès. Notons cependant que lors des expériences avec le modèle peptidique, aucune fonctionnalisation n'a été observée, par spectrométrie de masse, en présence de cofacteur seul ou bien en présence de cofacteur et de GLP sauvage, ce qui conforte dans l'idée d'un transfert médié par l'enzyme convenablement mutée.

Par ailleurs, au cours de la réaction de chimie click, en plus des protéines fonctionnalisées enzymatiquement, le cofacteur restant libre en solution réagit avec la sonde biotine. En admettant que, pour des raisons que nous ignorons, le cofacteur « clické » à la sonde biotine adhère aux protéines et ne soit pas éliminé au cours de la migration électrophorétique, ceci pourrait aussi expliquer le signal non spécifique. Enfin, la sonde biotine étant aussi introduite en excès, pour palier l'excès d'analogue de SAM, elle contribuerait aussi à augmenter ce signal non spécifique. Dans les expériences avec la sonde fluorescente, nous avons procédé à une

étape de précipitation des protéines afin d'essayer de nous affranchir de ces signaux non spécifiques. Dans le cas de la sonde biotine, et étant donné les faibles quantités de protéines manipulées, cette étape de précipitation s'est révélée trop délicate à mettre en œuvre.

En soi, ce problème de marquage non-spécifique est handicapant dans le cas d'une enzyme peu active comme c'est le cas pour PR-Set7 et dans lequel il est difficile de distinguer ce qui relève d'un signal spécifique ou d'un signal non spécifique. Le problème ne se pose pas pour GLP par exemple.

Afin d'essayer de nous affranchir de toutes ces contraintes nous avons mis en place une stratégie permettant de supprimer l'excès d'analogue de SAM à la fin de l'étape du test enzymatique lors de ces expériences de criblage. La même stratégie a aussi pour but de limiter les quantités de sonde complémentaire utilisées lors de l'étape de chimie click. Cette stratégie consiste à séparer les protéines sur gel dénaturant après l'étape de fonctionnalisation enzymatique et permet d'éliminer l'excès de cofacteur au cours de la migration électrophorétique. Les protéines sont transférées sur membrane de PVDF (poly(fluorure de vinylidène) et la réaction de chimie click est effectuée à la surface de cette membrane. L'application de cette stratégie dérive de celle initiée par Ohata *et al.* en 2015¹⁶¹. Initialement utilisée avec des sondes pro-fluorescentes, il a été montré que la cycloaddition alcyne-azoture était permise à la surface d'une membrane et conduisait à une visualisation de protéines étiquetées de façon spécifiques et avec une sensibilité équivalente aux méthodes de visualisation classiques.

Dans notre étude nous montrons que cette méthode permet aussi la visualisation de protéines par western blot. Pour ces essais de mise au point, nous avons utilisé le ProSeAM et la GLP WT capable de prendre en charge cet analogue. Après réaction enzymatique, les échantillons sont séparés sur gel d'électrophorèse SDS-PAGE, puis les protéines sont transférées sur membrane de PVDF. La réaction de chimie click est effectuée dans une boîte de Petri en verre à la surface de la membrane avec la sonde biotine-PEG4-azide et les réactifs de click, puis, cette dernière a été lavée plusieurs fois au DMSO de façon à éliminer la sonde biotine n'ayant pas réagi. Après cela, la membrane est réactivée au méthanol et le protocole standard de western blot utilisant le conjugué HRP-streptavidine peut être repris à partir de l'étape de blocage à la BSA. La Figure 45A montre le résultat de cet essai.

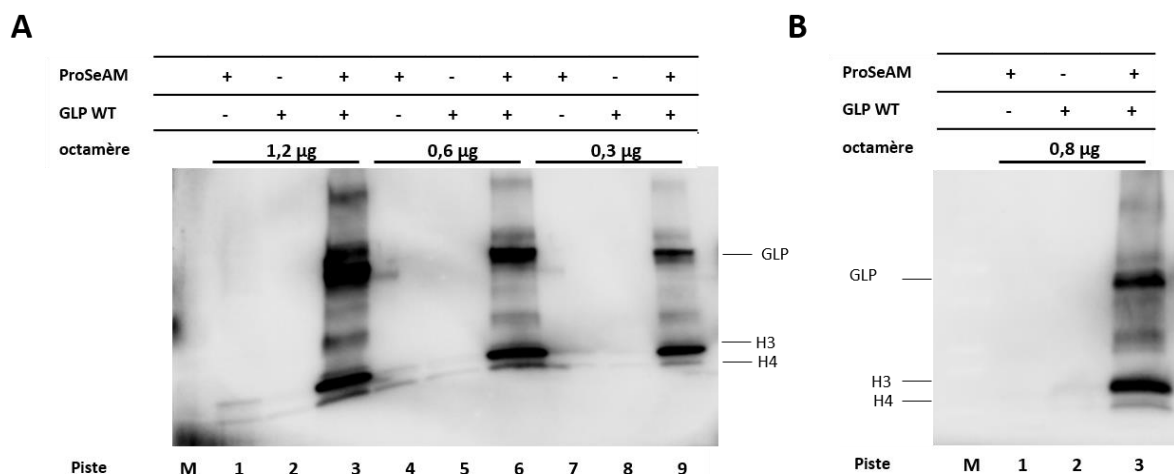


Figure 45. Western blot comparant l'efficacité de A. la réaction de la chimie click sur membrane et de B. la réaction de chimie click en solution.

La Figure 12B constitue une expérience contrôle dans laquelle GLP WT est mise à incuber avec ProSeAM et où la réaction de chimie click est effectuée en solution, il s'agit donc du protocole classique.

Pour tester la sensibilité du protocole de click sur membrane, des quantités croissantes du mélange réactionnel ont été déposées sur gel. Ainsi les quantités d'octamère d'histone déposées sur les pistes vont de 0,3 µg à 1,2 µg selon les conditions. Dans les essais en présence d'enzyme GLP et de cofacteur, nous observons la fonctionnalisation de l'histone H3 quelle que soit la quantité d'octamère déposée (pistes 3, 6 et 9). Dans les conditions sans enzymes ou sans cofacteur nous n'observons que très peu voire pas du tout de signal, indépendamment du protocole de chimie click. De plus, la comparaison des résultats du protocole de click sur la membrane et de la chimie click en solution nous laisse penser que la sensibilité de détection est équivalente dans les deux protocoles. Par ailleurs, dans les deux cas, en présence d'enzyme et de cofacteur, de nombreux signaux sont observés ailleurs que sur l'histone H3, au niveau de GLP, et des impuretés protéiques co-purifiées avec elle. Etant donné que l'excès d'analogue de cofacteur est éliminé à l'issue de l'étape du transfert enzymatique grâce à la migration électrophorétique et le transfert sur membrane, il ne peut pas expliquer ce signal. De plus, le signal non spécifique étant observé dans les mêmes proportions avec les deux protocoles de chimie click, il ne peut pas non plus être expliqué par une obscure réaction secondaire liée à la chimie click et à l'excès de réactifs puisqu'ils sont ici éliminés. Il semble donc que, parmi les hypothèses que nous avons émises, il ne reste plus que celle d'une réactivité intrinsèque de l'analogue du cofacteur pour expliquer le signal non spécifique même si nous ne comprenons pas le mécanisme.

Finalement, étant donné que l'histone H3 est fonctionnalisée de la même manière par les deux protocoles, cette expérience permet de conclure que les excès d'analogues de cofacteur et de sondes qui ne sont pas éliminés avec le protocole standard que nous avons effectué en solution ne sont pas si handicapants que l'on aurait pu le penser et ne sont pas la cause principale du marquage non spécifique que nous observons.

III.4.2. L'analyse directe de la fonctionnalisation des histones par spectrométrie de masse MALDI-TOF après transfert sur membrane de nitrocellulose

Nous avons également exploré la mise en place d'une méthode pour analyser la fonctionnalisation des histones en spectrométrie de masse. L'analyse par MALDI-TOF n'était jusqu'alors pas privilégiée car la présence importante de sels provenant du tampon et de cuivre dans les échantillons n'était pas compatible avec une telle analyse. Notons que nous avons tenté de dessaler les échantillons en utilisant de la résine C-18 mais les histones, en raison de leur nature très hydrophile ne sont que très peu retenues par la résine. Sur la base des protocoles décrits dans l'article de Luque-Garcia *et al.* de 2006¹⁶², nous avons essayé de voir s'il était possible de détecter les histones et leur fonctionnalisation après des étapes de migration sur gel de polyacrylamide et transfert sur membrane de nitrocellulose. La migration électrophorétique permettrait alors de purifier chacune des protéines histones et réduirait la complexité de l'échantillon à analyser. Succinctement, les échantillons à analyser sont préparés en dissolvant 2 mm² de membrane de nitrocellulose contenant une des histones dans 20 µL d'un mélange Méthanol/Acétonitrile/TFA contenant de la matrice HCCA. 1 µL de cette préparation est alors déposée sur plaque MALDI pour analyse.

La Figure 46 montre ainsi l'analyse d'un échantillon d'1 µg d'octamère d'histone après l'application de ce protocole. Nous observons ainsi les spectres de masse de chacune des histones constituant l'octamère d'histone (masses attendues de H2A : 13960 Da, H2B : 13774 Da, H3 : 15273 Da, H4 : 11236 Da). Notons que même si elle n'a pas été utilisée à cette fin, la migration électrophorétique permet une réduction de la complexité de l'échantillon en assurant la séparation de trois des quatre histones.

Nous avons tenté de déterminer si cette méthode pouvait permettre l'identification d'une fonctionnalisation introduite sur une des histones. Cependant l'analyse d'échantillons contenant H3 fonctionnalisée par Ab-SAM et ayant subi la réaction de chimie click en solution n'a pas permis d'obtenir des spectres de masses exploitables, possiblement à cause du cuivre utilisé pendant la réaction de chimie click. Notons qu'avec la SPAAC ce problème ne se poserait pas (Figure 46)

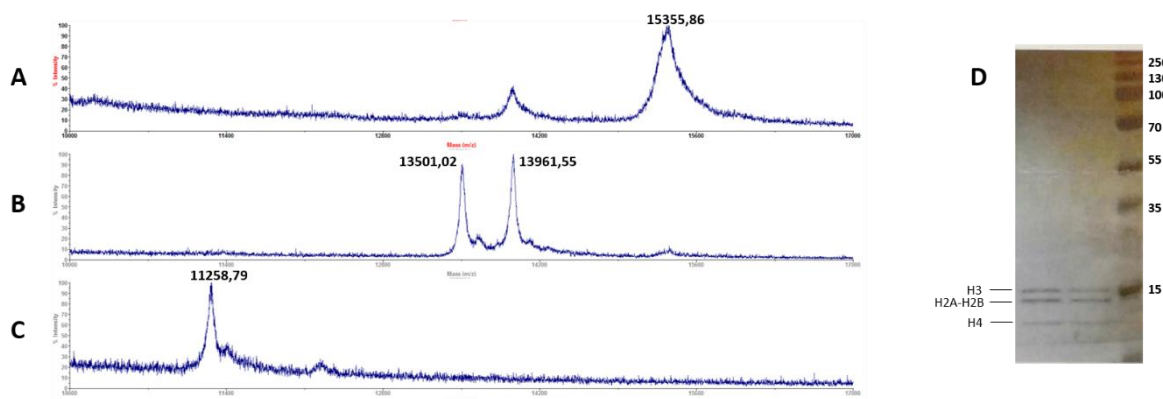


Figure 46. Analyse par MALDI-TOF en mode linéaire de la composition d'un octamère d'histone après purification sur gel d'électrophorèse et transfert sur membrane de nitrocellulose. A. Spectre de l'histone H3. B. Spectre des histones H2A et H2B. C. Spectre de l'histone H4. D. Histones transférées sur membrane de nitrocellulose colorée au Ponceau avant la découpe de chacune des bandes protéiques.

Le protocole de chimie click sur membrane ayant été développé plus tard pendant la thèse, nous n'avons pas encore pu le rendre compatible avec ce protocole d'analyse des protéines par spectrométrie de masse MALDI-TOF. Associée à des lavages efficaces de la membrane pour éliminer le surplus de réactifs et de sondes, nous pourrions probablement identifier la fonctionnalisation sur les histones.

Nous envisageons à terme de coupler cette approche avec le protocole de chimie click sur membrane ce qui constituerait une méthode complémentaire aux stratégies de détection de la fonctionnalisation par western blot ou par fluorescence et surtout permettrait d'avoir dans un même temps une analyse indirecte (western blot) couplée à une analyse directe (masse) d'un même échantillon.

III.5. Conclusion

Le criblage de notre série d'analogues de SAM sur le mutant Y1211A de GLP utilisé comme contrôle positif révèle que l'ensemble des analogues de SAM disposant d'une double liaison stabilisatrice de l'état de transition de la réaction enzymatique peut être utilisé par le mutant pour fonctionnaliser l'histone H3. Ceci a été observé à la fois sur substrat peptidique et protéique. Ces résultats sont parfaitement cohérents avec les résultats obtenus précédemment par l'équipe de M. Luo¹¹². Malheureusement, nous n'avons pas pu observer de prise en charge de ces analogues par nos mutants de l'enzyme PR-Set7 y compris avec le mutant correspondant au mutant Y1211A de GLP par superposition de structure. Ceci peut s'expliquer par une trop

faible activité intrinsèque *in vitro* de l'enzyme PR-Set7, activité qu'il est difficile de distinguer du bruit de fond souvent présent sur les gels d'électrophorèses et dont nous n'arrivons pas à déterminer la source. Une autre hypothèse serait que ces molécules sont trop encombrantes pour rentrer dans le site actif contraint de PR-Set7, malgré les mutations introduites. En tenant compte de ces hypothèses nous avons choisi de synthétiser ProSeAM, correspondant au plus petit analogue de SAM répondant aux critères désirés (double ou triple liaison en β du sulfonium et fonction chimique bioorthogonale). En raison de la présence de l'atome de sélénium, cet analogue de SAM est aussi plus réactif.

La synthèse de ce nouvel analogue a été réalisée par Joana Dagoret et Karine Guitot. Je l'ai ensuite testé vis-à-vis de GLP et du mutant Y1211A, révélant dans les deux cas une prise en charge, même si elle est accrue avec l'enzyme sauvage. Je l'ai ensuite testé vis-à-vis de PR-Set7, montrant qu'il est pris en charge par l'enzyme WT et le mutant Y334F de façon équivalente.

Bien que l'analogue ProSeAM ne soit pas totalement satisfaisant en terme d'orthogonalité PR-Set7 sauvage/mutant, et étant donné que c'est le seul analogue de SAM qui nous ait permis de fonctionnaliser de façon relativement sélective l'histone H4, nous avons entrepris l'étude protéomique avec cet analogue et cette étude est l'objet du chapitre IV.

Enfin, nous avons mis au point de nouveaux protocoles permettant à terme d'effectuer une analyse indirecte (western blot) couplée à une analyse directe (masse) d'un même échantillon, ce qui dans ce type d'approche nous a semblé très important pour distinguer ce qui relève d'un signal dû à une fonctionnalisation spécifique ou à d'autres facteurs artefactuels. Cette méthode pourrait être étendue à l'analyse d'autres protéines fonctionnalisées par des rapporteurs chimiques. Cette méthode est complémentaire aux trois analyses que nous utilisons jusque-là : par western blot, par fluorescence sur gel et par spectrométrie de masse MALDI-TOF de peptides fonctionnalisés mais n'a pas été exploitée par manque de temps.

Chapitre 4

De la protéomique aux premières étapes de la validation biologique

IV. De la protéomique aux premières étapes de validation biologique

L'étude présentée dans le chapitre précédent a permis de montrer que seul, l'analogue de SAM sélénié ProSeAM, pourrait être utilisé pour réaliser notre étude de protéomique visant à étiqueter spécifiquement les protéines substrats de PR-Set7. Ce cofacteur, associé au mutant Y334F permet en effet d'observer un marquage sur l'histone H4. Nous avons par ailleurs observé un important signal non spécifique, détecté lors d'analyses par western blot. Indépendamment de l'origine de ce signal, une méthode doit être trouvée afin de soustraire ce signal au marquage spécifique et identifier un maximum de potentiels substrats de PR-Set7 avec une confiance la plus grande possible.

Le principe de l'étude que nous avons réalisé est présenté dans la Figure 47.

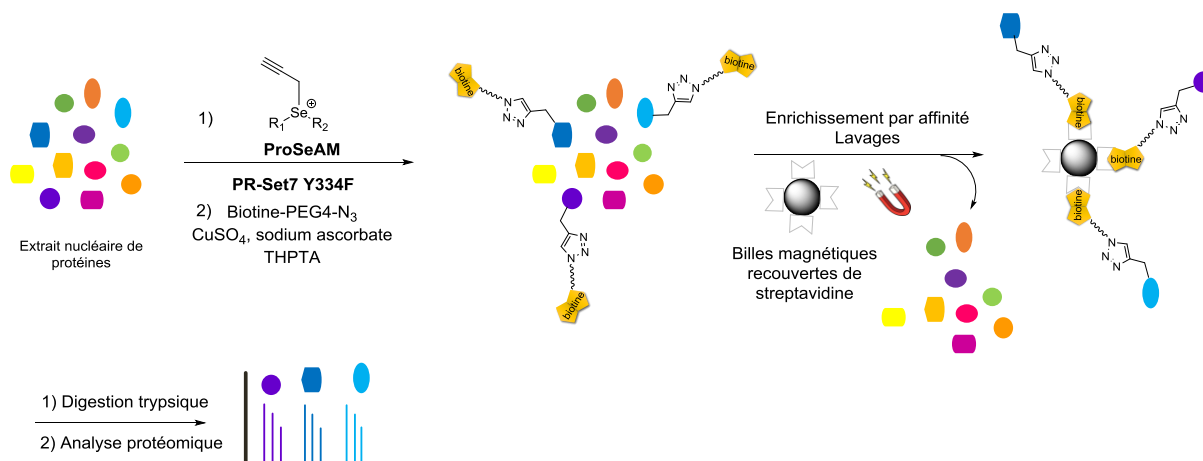


Figure 47. Principe de la stratégie de protéomique pour la caractérisation des cibles de PR-Set7

PR-Set7 étant une protéine nucléaire essentiellement exprimée pendant la mitose, nous avons cherché de potentiels nouveaux substrats de l'enzyme au sein d'extraits nucléaires de cellules en utilisant les HEK293T comme modèle. L'objectif de la stratégie consiste donc à étiqueter les protéines substrats de PR-Set7 présentes dans un extrait nucléaire par voie enzymatique à l'aide du ProSeAM, puis d'utiliser le motif alcyne introduit sur les protéines pour les fonctionnaliser par une biotine après réaction de CuAAC. L'identification des cibles de PR-Set7 est alors réalisée lors d'une étude de protéomique se basant sur la purification par affinité des protéines étiquetées d'une biotine avec des billes magnétiques recouvertes de streptavidine. L'interaction biotine-streptavidine fait partie des interactions non covalentes les plus fortes, avec une constante de dissociation K_d de l'ordre de 10^{-14} mol.L⁻¹. Cette très forte interaction permet, après capture des protéines biotinylées, d'appliquer des lavages stringents visant à

éliminer les protéines liées de façon non spécifiques à la surface des billes. Les protéines enrichies sont ensuite soumises à une digestion à la trypsine.

L'analyse quantitative des protéines enrichies sur les billes est alors effectuée à l'aide du logiciel développé par l'Institut Curie myProMS¹⁶³ v3.0, qui effectue à la fois la validation des identifications obtenues via des moteurs de recherche, l'évaluation du taux de faux positif (FDR), la quantification des protéines à l'aide du logiciel MassChroQ¹⁶⁴, l'analyse statistique et la visualisation des données. Ceci est réalisé en deux étapes.

- **Une analyse qualitative** qui permet l'identification des protéines enrichies à l'aide du logiciel Proteome Discoverer (Sequest HT). La banque de données utilisée est la banque humaine UniProt (2019).
- **Une analyse quantitative des protéines enrichies.** Cette quantification a été effectuée selon une stratégie « label free », par la méthode de l'extraction du chromatogramme ionique (XIC) *via* le logiciel MassChroQ.

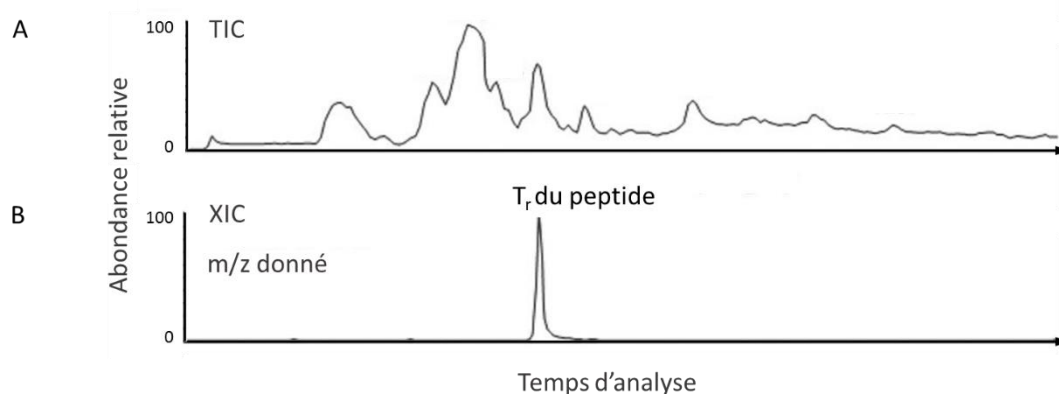


Figure 48. Représentation d'un chromatogramme correspondant à un peptide extrait par la méthode XIC (B) par rapport au chromatogramme correspondant à l'analyse de l'échantillon totale (A)

Le XIC est une méthode de quantification basée sur l'intensité du signal MS des peptides caractéristiques d'une protéine. Il correspond à la production d'un chromatogramme d'ion extrait qui va permettre de connaître la contribution d'un ion (à un m/z donné) et donc d'un peptide donné au signal MS total. A titre d'exemple, la Figure 48 représente le TIC (courant ionique total) correspondant à l'analyse d'un échantillon d'une des conditions (A) ainsi que le XIC pour un peptide de m/z donné (B).

L'aire du pic du XIC (B) peut être utilisée pour une quantification relative de ce peptide dans l'échantillon considéré. Pour ce faire, la méthode du XIC est appliquée sur tous les peptides

identifiés comme appartenant à une protéine et la quantité relative de cette protéine est calculée à l'aide de la moyenne géométrique des aires des pics des XIC obtenus. Cette approche est ensuite reproduite sur toutes les protéines identifiées au sein d'un échantillon, permettant ainsi d'obtenir la quantité relative de toutes les protéines pour chaque échantillon. La quantification étant relative, elle se fera toujours entre plusieurs échantillons. La comparaison de la quantité d'une protéine au sein de deux échantillons différents correspond au rapport entre les moyennes géométriques des différents peptides de cette protéine. Ce rapport est appelé taux d'enrichissement (FC). Un enrichissement d'une protéine dans une analyse par rapport à une autre est dit significatif si son $FC \geq 2$ et si sa $p\text{-value} < 0,05$.

IV.1. Mise en place et optimisation du protocole de capture des protéines

IV.1.1. Validation de la spécificité de position de fonctionnalisation de PR-Set7

Avant de réaliser l'expérience de capture qui nous permettra de caractériser les potentiels substrats de PR-Set7, nous avons vérifié par spectrométrie de masse quels étaient les résidus de l'histone H4 propargylés par PR-Set7 sauvage et son mutant Y334F en présence de la ProSeAM dans les conditions de l'étude protéomique. Ainsi, après avoir effectué un test enzymatique en présence de nucléosome recombinant, de ProSeAM et de PR-Set7 (mutée ou non), les protéines ont été déposées sur gel de polyacrylamide en condition dénaturante. Après coloration du gel au bleu de Coomassie, la bande migrant à la distance de migration de l'histone H4 a été découpée pour analyse. Les protéines ont ensuite été soumises à un traitement classique au 1,4-dithiothréitol (DTT) afin de réduire les ponts disulfures et un traitement à l'iodoacétamide (IAA) permettant d'alkyler les cystéines libres. La protéine a été digérée *in gel* par action de la trypsine. Pour rappel, la trypsine est une protéase qui hydrolyse spécifiquement les liaisons peptidiques après des résidus basiques de type lysine et arginine. La vérification manuelle de l'introduction de la modification a été effectuée, par Bérangère Lombard (Institut Curie), par extraction des XIC des peptides propargylés, nous avons ainsi recherché la présence d'adduits de masse de + 38 correspondants à la propargylation et avons validé que la propargylation n'était introduite par l'enzyme sauvage et le mutant Y334F que sur la lysine 20 de l'histone H4 par analyse MS/MS (Figure 49 et analyse complète en annexe S2). De cette façon, nous concluons que la spécificité de position de PR-Set7 semble respectée en présence de notre mutant et de la ProSeAM conformément à son activité rapportée de monométhyltransférase de H4K20.

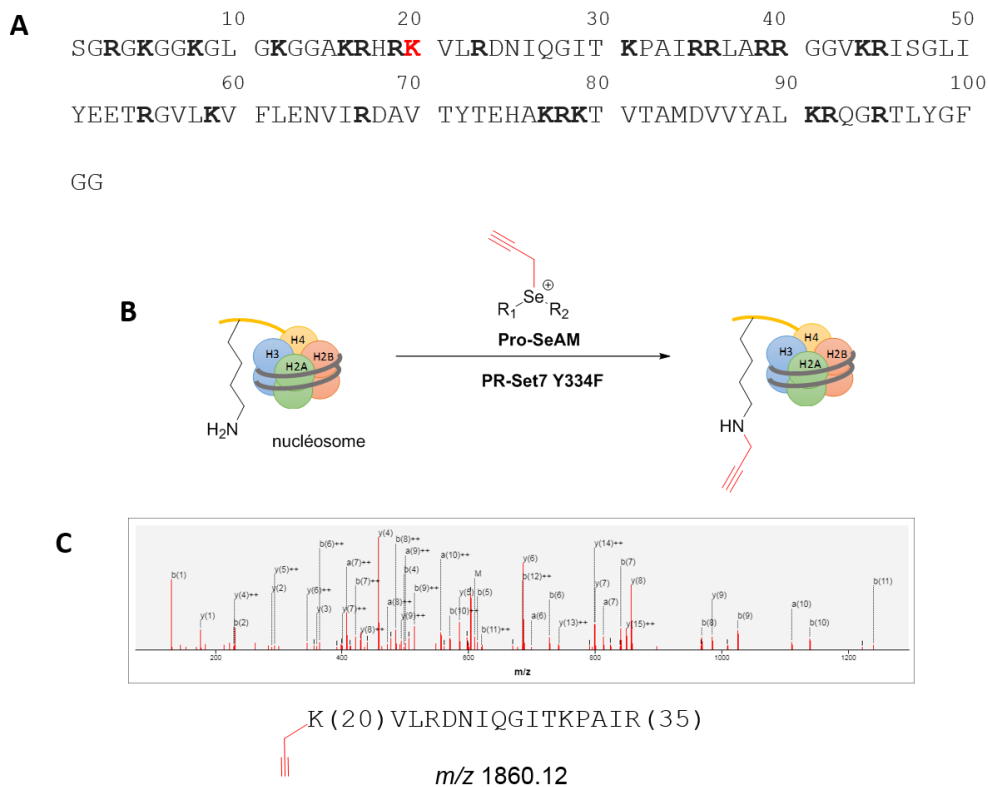


Figure 49. Caractérisation de la propargylation sur H4K20. A. Séquence de l’histone H4, les résidus basiques sont en gras, la lysine 20 fonctionnalisée figure en rouge. B. Principe du marquage chémoenzymatique en présence de ProSeAM et du mutant Y334F de ProSeAM. C. Caractérisation MS / MS de la propargylation sur H4K20.

IV.1.2. Optimisation de la capture sur billes de streptavidine

L’expérience sur nucléosome recombinant a permis de valider la spécificité de la fonctionnalisation sur H4, mais nous sommes encore loin des conditions complexes de l’étude protéomique. Dans ce contexte et dans la continuité du raisonnement vu précédemment de découplage des différentes étapes de notre stratégie d’étiquetage/fonctionnalisation, nous avons optimisé notre protocole initial dans le but de réduire au minimum les protéines non biotinylées adsorbées à la surface des billes de streptavidine.

Lors de la première étape de la stratégie consistant dans le transfert enzymatique de la chaîne propargyle de ProSeAM vers la lysine substrat, le cofacteur est introduit en large excès dans le mélange réactionnel. La fonctionnalisation ultérieure de l’alcyne par la sonde biotine-PEG4-azide nécessite alors une concentration importante de cette dernière pour fonctionnaliser les protéines propargylées, la sonde réagissant également avec l’analogue de SAM libre en solution. Nous avons donc ajouté une étape de précipitation des protéines au méthanol dans le but d’éliminer l’excès d’analogue de SAM, soluble dans le méthanol, avant de poursuivre avec

la réaction de chimie click. Cette étape de précipitation est ici plus facile à mettre en œuvre que dans nos précédents essais de par la plus grande quantité de protéines utilisées par rapport à nos tests de mise au point.

Le protocole mis en place peut ainsi être résumé dans la Figure 50.

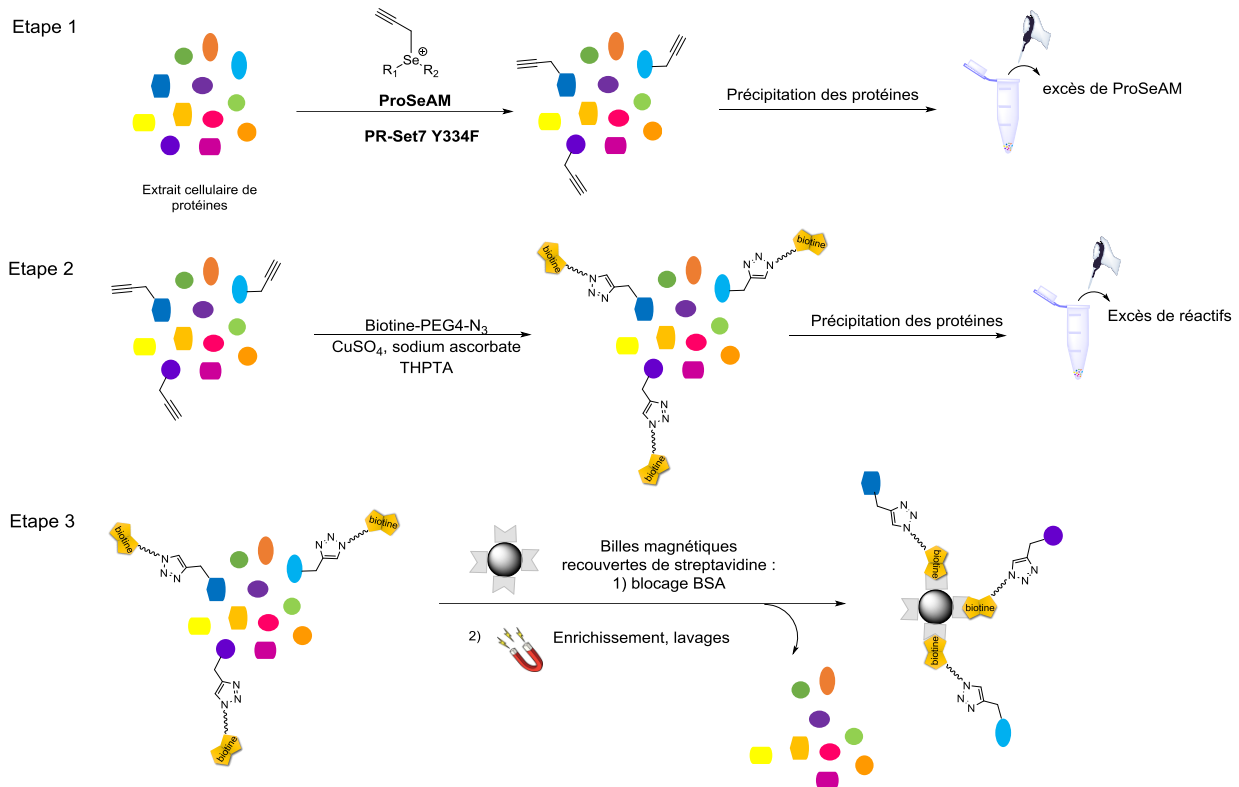


Figure 50. Principe des différentes étapes du protocole permettant d'augmenter les taux de fonctionnalisation des protéines et de capture spécifique à la surface des billes recouvertes de streptavidine.

Lors de l'étape 1, des extraits nucléaires de cellules HEK293T sont mis à incuber en présence de PR-Set7 et de ProSeAM. Les protéines sont ensuite précipitées par le méthanol à froid. Lors de l'étape 2, après re-solubilisation des protéines dans un tampon de resuspension, la réaction de chimie click avec la sonde biotine-PEG4-azide est effectuée. Là encore, cette sonde est introduite en excès par rapport aux groupements alcynes. Une seconde étape de précipitation au méthanol a donc été ajoutée pour s'affranchir de l'excès de sonde biotine.

Lors de l'étape 3, la capture des potentiels substrats de PR-Set7 à présent étiquetés avec une biotine est effectuée. Afin de limiter, la fixation indésirée de protéines non fonctionnalisées sur les billes de streptavidine, une étape préliminaire de blocage en présence de BSA a été effectuée en parallèle de la précipitation au méthanol. S'en suit l'incubation des billes avec les protéines étiquetées resolubilisées. Enfin, de nombreux lavages successifs et très stringents des billes permettent d'éliminer tout ce qui n'est pas lié aux billes par l'interaction biotine/streptavidine.

IV.2. Stratégie mise en œuvre pour l'approche de protéomique

Bien que nous ayons mis en place une stratégie de diminution du bruit de fond, étant donné l'activité enzymatique modeste de PR-Set7 *in vitro*, nous anticipons que la seule recherche de protéines enrichies dans la condition avec le mutant Y334F par rapport à celle avec l'enzyme sauvage ne permettra pas un enrichissement d'un nombre important de candidats substrats. Nous pourrions alors passer à côté de cibles intéressantes de notre enzyme. Pour pallier ce problème, nous prévoyons deux contrôles négatifs supplémentaires : le premier en présence du mutant perte de fonction Y334A, le second en absence d'enzyme exogène. La Figure 51 présente la stratégie globale adoptée afin de repérer les protéines réellement enrichies par la méthode de marquage chémoenzymatique et de s'affranchir des enrichissements artéfactuels et non liés à la propargylation des lysines par PR-Set7.

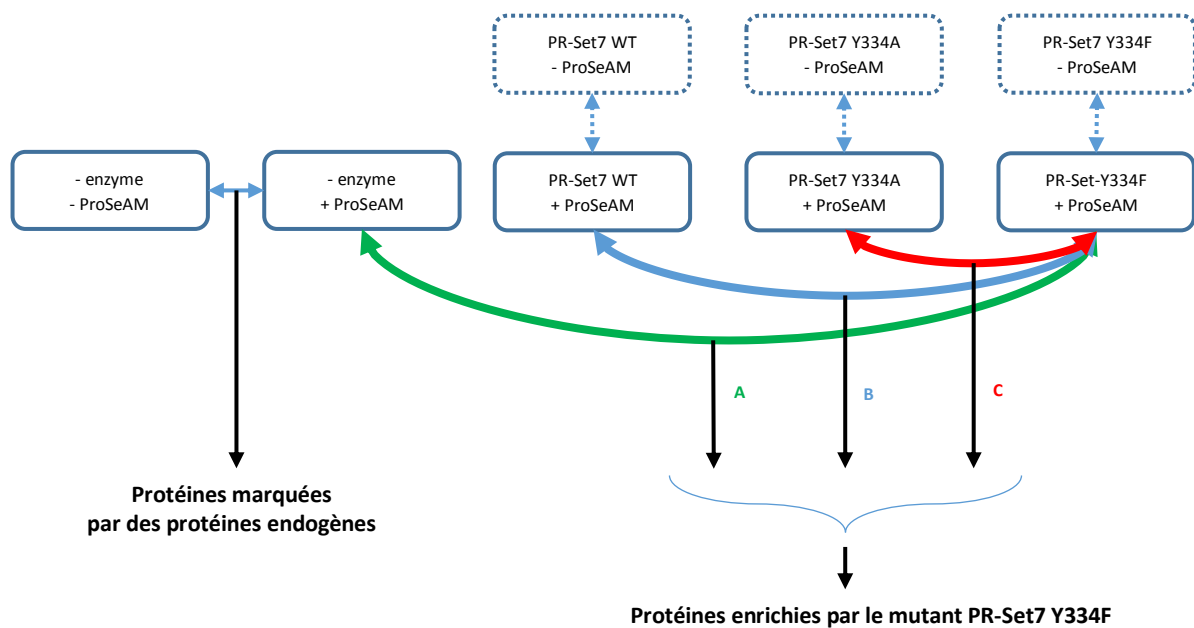


Figure 51. Schéma de la stratégie de protéomique

L'ensemble des quatre conditions en présence d'analogue de cofacteur accompagnées de leur contrôle respectif en son absence permet de répondre à deux questions essentielles :

- Quelle est la contribution des enzymes méthyltransférases endogènes dans la fonctionnalisation de leurs cibles ?

La réponse à cette question repose essentiellement sur la comparaison des deux conditions (avec et sans ajout de ProSeAM) en absence d'enzyme exogène. Des résultats équivalents pourraient être retrouvés en comparant les deux conditions (avec et sans ajout de ProSeAM) en présence du mutant Y334A non fonctionnel.

- Quelles sont les candidats substrats de PR-Set7 ? Deux méthodes s'offrent à nous pour répondre à cette question :

1) Par détermination des taux d'enrichissement des protéines dans les conditions en présence d'analogue de cofacteur contre les conditions en absence de l'analogue. Puis en comparant les quatre listes de protéines enrichies. Nous nous attendrions avec cette méthode, à des listes de protéines enrichies plus importantes dans les conditions avec ajout de l'enzyme sauvage ou du mutant Y334F qu'avec le mutant Y334A ou sans enzyme exogène ;

2) Par détermination des taux d'enrichissement en comparant, à la condition du mutant Y334F en présence de ProSeAM, chacune des trois autres conditions également en présence de l'analogue. Avec cette méthode et étant donné que nous avons vu que l'enzyme sauvage prenait en charge ProSeAM, nous nous attendrions à des taux d'enrichissement plus forts en comparant la condition avec le mutant « gain de fonction » Y334F au mutant perte de fonction Y334A ou à la condition sans enzyme exogène qu'avec la comparaison contre l'enzyme sauvage. Pour des raisons détaillées aux paragraphes suivants, nous avons utilisé cette deuxième méthode.

IV.3. Réalisation de l'expérience de protéomique

IV.3.1 Expérience préliminaire

Dans le but de vérifier que l'ensemble des mesures d'optimisations permettent effectivement une identification de cibles de PR-Set7, une expérience préliminaire à l'expérience de protéomique a été effectuée. Dans cette expérience, les 8 conditions détaillées plus haut ont été

effectuées en 1 seul exemplaire afin de vérifier que le protocole permettait d'enrichir des substrats connus de PR-Set7. Etant donné l'absence de répliques, les résultats doivent être pris avec prudence. Toutefois, nous nous en sommes servis comme une indication de l'efficacité des mesures prises.

Outre l'histone H4, trois autres substrats de PR-Set7 connus de la littérature ont été recherchés : la protéine Numb^{82,83}, la protéine PCNA⁶¹ et la protéine p53⁶².

Cette expérience a permis effectivement d'enrichir l'histone H4 dans la condition du mutant Y334F par rapport aux conditions avec le mutant Y334A, la condition sans enzyme ou la condition sans analogue de cofacteur de façon très significative (Figure 52). Par ailleurs comparer les conditions du mutant gain de fonction avec l'enzyme sauvage semble montrer une tendance d'enrichissement en faveur du mutant même si la p-value ne passe pas le seuil de 0,05. Aucune autre protéine n'a été significativement enrichie par les billes. Ces résultats sont plutôt encourageants. En effet, même si l'histone H4 est la seule protéine enrichie, c'est surtout la seule qui franchi le seuil de la p-value. Ceci est, entre autre, la conséquence du calcul de la p-value qui accorde une grande importance au nombre de peptides détectés et au nombre de répliques. Ainsi le seuil sera d'autant plus vite franchi si le nombre de peptides détectés par le spectromètre et le nombre de répliques est grand. Etant donné l'abondance des histones dans des extraits nucléaires, le fait que l'histone H4 soit enrichie en présence du mutant « gain de fonction » alors que les autres histones ne le sont pas est encourageant pour la suite de la méthode. Enfin, l'augmentation du nombre de répliques pourrait faire franchir à d'autres protéines le seuil statistique.

1 proteins	Gene	Y334F +coF/-coF		Y334F/sans enzyme +coF		Y334F/WT +coF		Y334F/Y334A +coF		MW kDa	
		Ratio	Pep. used	Ratio	Pep. used	Ratio	Pep. used	Ratio	Pep. used		
P62805	HIST1H4A	9.43 ↑	40	18.33 ↑	28	6.62 ↑	44	9.57 ↑	42	11.4	Histone H4 <i>Homo sapiens</i>

Figure 52. Identification des protéines enrichies lors de l'expérience préliminaire à l'expérience de protéomique

En conclusion, cette expérience préliminaire permet de valider notre protocole et la méthode à utiliser puisqu'elle permet l'enrichissement de l'histone H4 substrat connu de PR-Set7 au sein d'extraits de protéines. De plus, nous confirmons que les comparaisons deux à deux de la condition du mutant Y334F face aux trois contrôles négatifs vont bien dans le même sens.

IV.3.2 Expérience principale

Nous avons alors mis en œuvre le protocole complet de protéomique avec l'ensemble des contrôles et 5 répliques pour chaque condition. L'analyse par spectrométrie de masse a été

réalisée par Bérengère Lombard et Damarys Loew du laboratoire de spectrométrie de masse et de protéomique (LSMP) de l'institut Curie. J'ai pour ma part effectué l'analyse des résultats après avoir été formé sur la plateforme à l'analyse des données.

Les résultats des expériences d'enrichissement sont présentés sous forme de volcano plot. Pour être enrichie, une protéine, doit avoir un facteur d'enrichissement (FC) supérieur à 2 et une p-value inférieure à 0,05. Une protéine est dite « unique » si elle n'a été identifiée que dans une condition. Graphiquement, le volcano plot montre un nuage de point où chaque point correspond à une protéine, la taille des points étant proportionnelle au nombre de peptides utilisés pour l'identification de la protéine en question. Les seuils d'enrichissement sont représentés par des barres verticales rouges, le seuil de la p-value est représenté par une ligne horizontale rouge. Les points correspondant aux protéines uniques sont distincts des protéines enrichies car leur caractère unique empêche tout calcul de taux d'enrichissement et de p-value. La confiance dans l'identification d'une protéine ne peut alors augmenter qu'avec le nombre de peptides protéotypiques détectés par le spectromètre de masse.

Enzyme Cofacteur	-PR-Set7 -ProSeAM	-PR-Set7 +ProSeAM	+PR-Set7 WT -ProSeAM	+PR-Set7 WT +ProSeAM	+PR-Set7 Y334A -ProSeAM	+PR-Set7 Y334A +ProSeAM	+PR-Set7 Y334F -ProSeAM	+PR-Set7 Y334F +ProSeAM
	Réplicat 1	2805	4390	2972	4348	3046	4489	3415
Réplicat 2	2935	4363	3142	4413	3798	4325	3390	4127
Réplicat 3	3314	4293	3140	4193	3910	4336	3056	4240
Réplicat 4	557	3929	2702	3615	2807	4568	2478	4281
Réplicat 5	2353	4013	2618	3739	3092	4190	2688	3959
Moyenne	2852	4198	2915	4062	3331	4382	2805	4151
Pourcentage de variation	+47%		+39%		+32%		+48%	

Tableau 4. Nombre de protéines identifiées pour chaque condition expérimentale. Pour un des réplicas de la première condition, la quantité de protéine identifiée par le spectromètre est aberrante, le réplica a été exclu de l'analyse

Les résultats de ces expériences révèlent en premier lieu une très grande différence de quantité de protéines identifiées dans les conditions en présence de ProSeAM par rapport à celles où il est absent (Tableau 4). Bien que ce résultat soit cohérent, ceci est problématique car la méthode de quantification employée nécessite une normalisation qui devient impossible (Figure 53). En effet, l'étalonnage n'est possible que si l'on admet qu'une grande majorité des protéines de l'échantillon ne voient pas leur quantité varier d'une condition à l'autre. Ceci est très courant dans des expériences plus classiques d'immunoprécipitation dans lesquelles l'adsorption de

protéines de manière non spécifique à la surface de billes entraîne ce résultat. Dans notre cas, il est possible qu'une grande quantité de protéines fonctionnalisées, par PR-Set7 ou des HKMT endogènes induise une saturation des sites de streptavidine des billes (fixation spécifique) qui, associée aux lavages des billes très stringents éliminant correctement les protéines adsorbées sur les billes, mène à cette situation de non équivalence de la quantité de peptides détectée dans les conditions en présence et en absence de l'analogue du cofacteur. Ce résultat est donc cohérent mais soulève la nécessité de traiter les données de façon particulière.

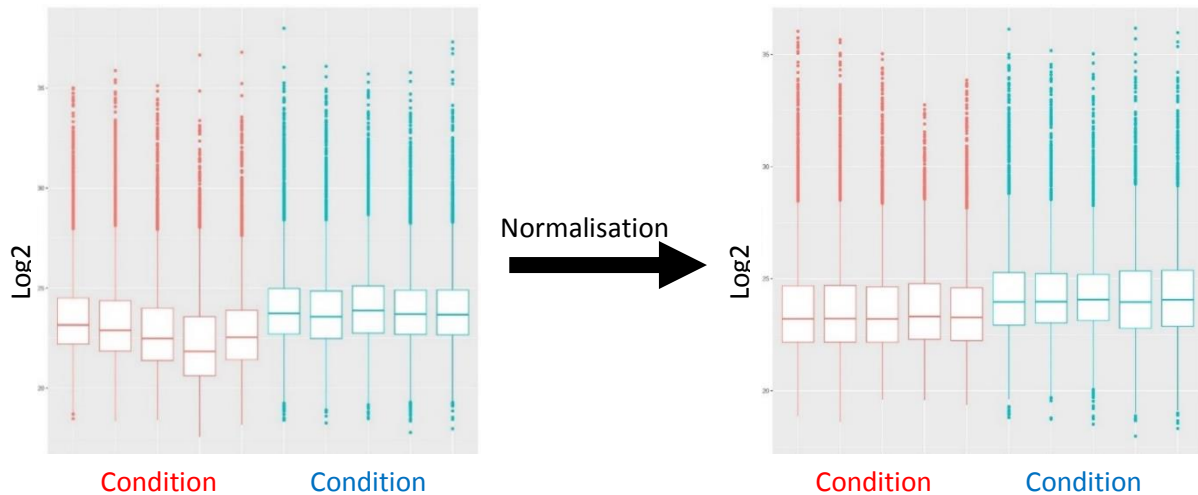


Figure 53. Schéma de la normalisation de la distribution de l'intensité des peptides en fonction des répliquas d'une condition. Dans cet exemple, la condition 1 (en rouge) correspond à « PR-Set7 Y334F - ProSeAM », la condition 2 (en bleu) à « PR-Set7 Y334F + ProSeAM »

Dans ces conditions, les contrôles en absence de l'analogue de cofacteur deviennent inutilisables et l'évaluation de la contribution des protéines endogènes dans la consommation du ProSeAM ne peut être effectuée. Pour la même raison, la 1^{ère} méthode de détermination des candidats substrats discutée plus haut ne sera pas retenue pour établir notre liste de protéines. Le résultat de la comparaison de deux conditions en absence ou en présence de ProSeAM est représenté à titre indicatif par le volcano plot ci-dessous (Figure 54). On observe alors, sans surprise, qu'une très grande proportion des protéines est enrichie en présence de ProSeAM mais il est impossible de déterminer lesquelles le sont par l'action de PR-Set7 par cette comparaison.

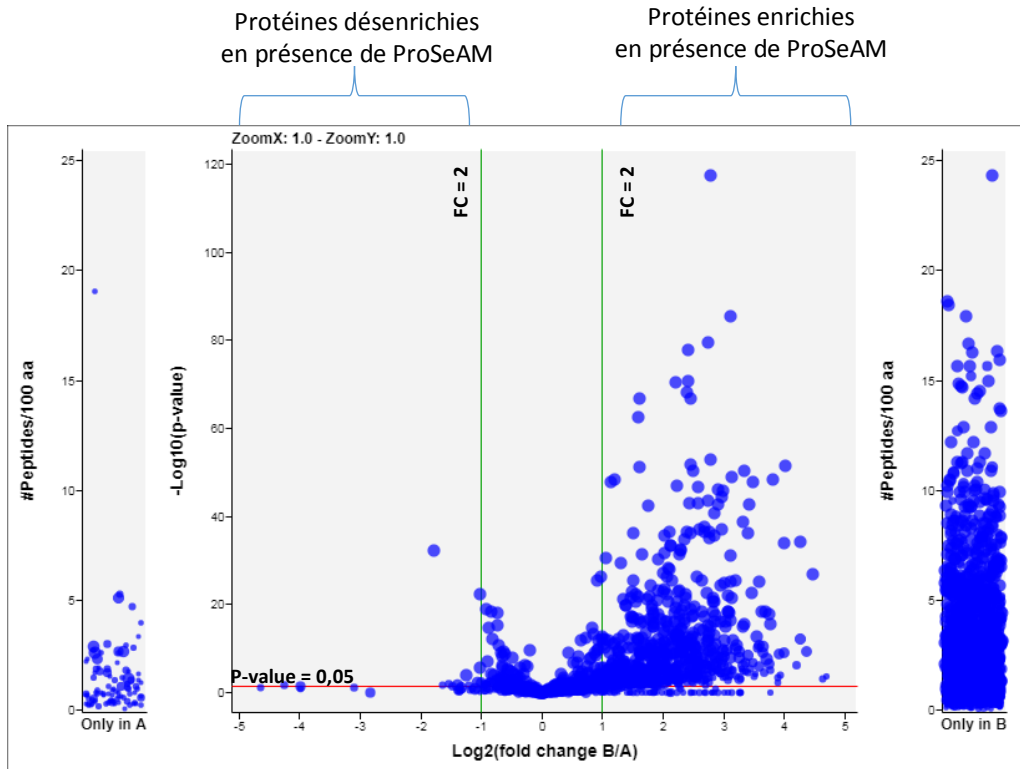


Figure 54. Volcano plot représentant la quantification des protéines enrichies dans les conditions sans enzyme, en absence (A) ou en présence (B) de l'analogue du cofacteur ProSeAM.

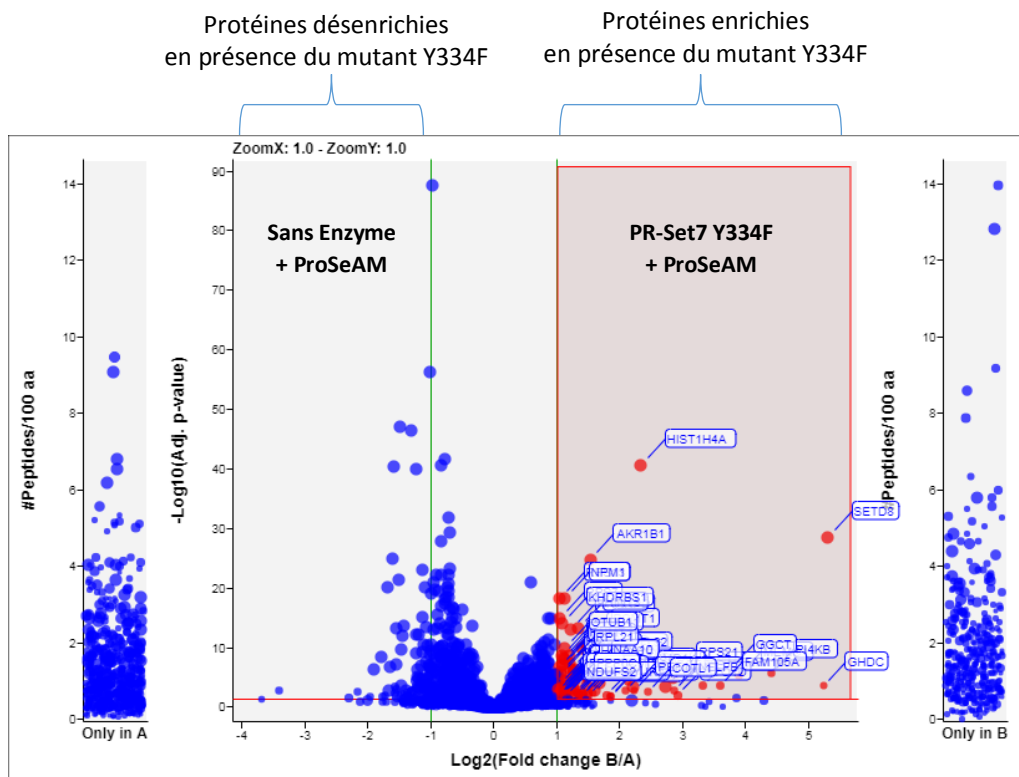


Figure 55. Volcano plot représentant la quantification des protéines enrichies en présence de ProSeAM dans la condition sans enzyme (A) comparée à la condition du mutant Y334F (B).

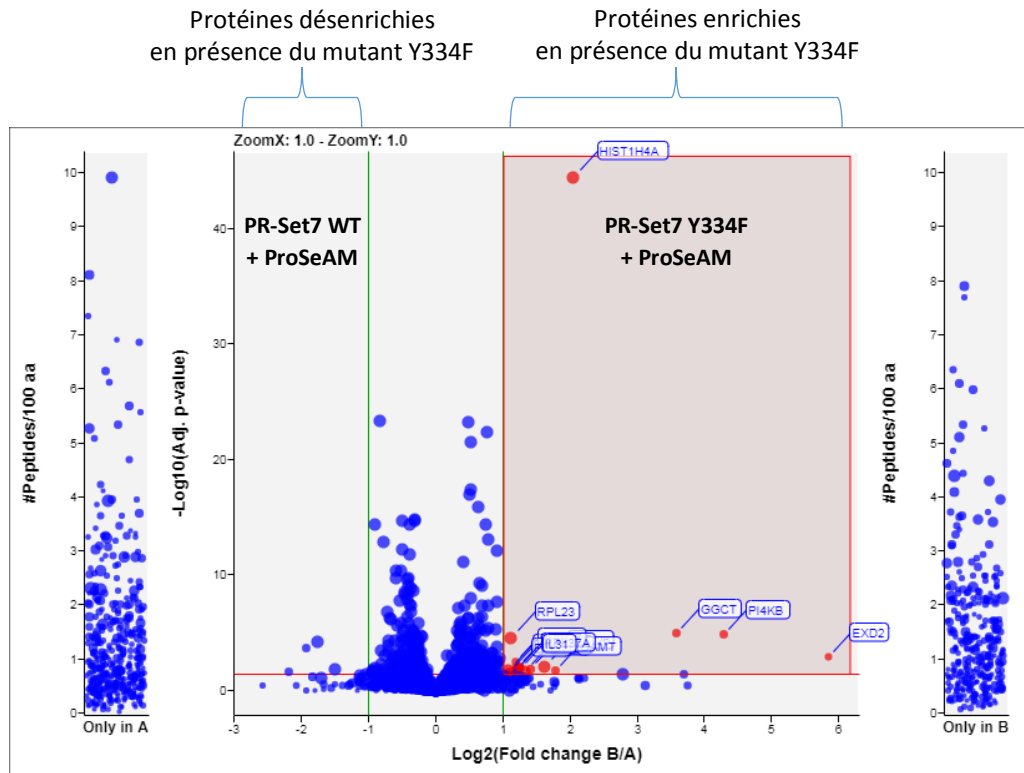


Figure 56. Volcano plot représentant la quantification des protéines enrichies en présence de ProSeAM dans la condition sauvage (A) comparée à la condition du mutant Y334F (B).

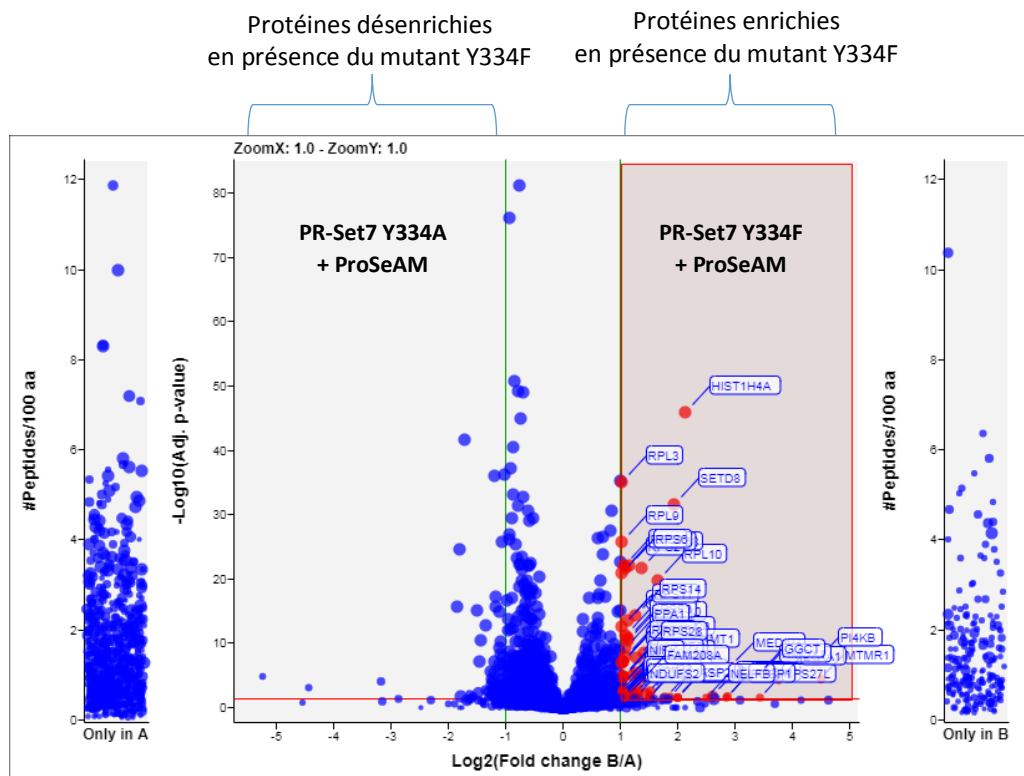


Figure 57. Volcano plot représentant la quantification des protéines enrichies en présence de ProSeAM dans la condition du mutant Y334A perte de fonction (A) comparée à la condition du mutant Y334F gain de fonction (B).

L'analyse comparée des données des volcano-plots de la condition du mutant Y334F contre les 3 autres conditions en présence de l'analogue du cofacteur permet de générer trois listes de protéines (Figures 55 à 57). La **liste de protéines enrichies A** rassemble les protéines enrichies grâce à l'activité du mutant gain de fonction vis à vis de la condition où aucune enzyme n'a été ajoutée. De même les **listes de protéines enrichies B et C** rassemblent les protéines enrichies par l'action du mutant Y334F vis-à-vis, respectivement, de l'enzyme sauvage et du mutant Y334A. Notons que l'histone H4 est systématiquement présente dans chacune des listes, ce qui n'est pas le cas des autres histones.

Les listes de protéines enrichies A, B, C contiennent respectivement 88, 16 et 67 protéines et sont partiellement redondantes. Après suppression des doublons, la fusion de ces trois listes permet la création d'une **liste de protéines enrichies D de 169 protéines** dont le contenu est détaillé en annexe S4.

Parallèlement, nous avons aussi listé les **protéines uniques A, B et C** représentant toutes les protéines uniquement présentes dans la condition du mutant Y334F vis-à-vis des trois autres conditions. Ainsi les listes de protéines uniques A, B, C contiennent respectivement 371, 322 et 205 protéines. A nouveau, les listes étant redondantes, nous avons créé une **liste de protéines uniques D de 545 protéines**.

En conclusion, cette expérience de protéomique a permis de mettre en évidence un enrichissement de protéines en présence d'un mutant capable de prendre en charge ProSeAM. Par ailleurs, le fait que le volcano plot de la Figure 56 soit moins élargi que dans les autres cas, reflète le moindre enrichissement de protéines vis-à-vis de la condition de l'enzyme sauvage. Cela reste cohérent par rapport aux résultats précédents de prise en charge du ProSeAM par l'enzyme sauvage et conforte l'idée que la spécificité de substrat n'est pas modifiée par la mutation de la Tyr-334 en une phénylalanine.

IV.4. Interprétation des résultats de la protéomique

En vue d'une validation biologique des substrats de PR-Set7, une sélection de protéines doit ensuite être effectuée parmi les deux listes de protéines enrichies et uniques. Pour cela deux stratégies d'exploitation des données brutes sont possibles.

La première consiste en une exploitation « non biaisée » ou sans a priori des données. Dans cette approche, les protéines ne sont sélectionnées que sur la base des taux d'enrichissement les plus

importants (pour les protéines enrichies) ou sur la base du nombre de peptide détectés (pour les protéines uniques). Cette méthode se voulant la plus objective a pour intérêt de mettre en avant les protéines candidates les plus facilement méthylées par PR-Set7. Ceci faciliterait donc l'étape de validation *in vitro*.

La seconde exploitation possible, « biaisée » comparée à la première, consiste à prendre en compte les caractéristiques connues des différentes protéines enrichies (ou uniques). Une surreprésentation d'une localisation cellulaire, d'une fonction moléculaire, d'un processus biologique particulier serait informatif quant aux rôles non connus de PR-Set7 dans la cellule et dirigerait notre intérêt sur des protéines clés de différents mécanismes physiologiques ou pathologiques. Dans cette approche, la présence ou non du site consensus de PR-Set7 connu à ce jour^{23,84} serait informatif du potentiel d'un nouveau candidat à être méthylé par l'enzyme. Cette méthode a pour avantage de diriger efficacement notre recherche vers des candidats biologiquement intéressants mais pourrait être incompatible avec des taux élevés de fonctionnalisation, ce qui rendrait la validation biologique plus laborieuse.

Dans un premier temps, nous avons choisi la première méthode pour la sélection des candidats de PR-Set7 à valider biologiquement avec l'objectif de valider rapidement quelques nouveaux substrats de PR-Set7. Cette stratégie du choix des protéines les plus enrichies permet aussi d'identifier celles qui sont distinctes des protéines identifiées comme contaminantes (e.g. kératines) dans la plupart des expériences de protéomique publiées dans la littérature scientifique constituant le CRAPome¹⁶⁵.

Bien que l'approche non biaisée nous ait servi de fil conducteur jusqu'à l'étape de validation biologique *in vitro*, une analyse approfondie du reste de l'ensemble des protéines enrichies est proposée à la fin de cette section.

IV.4.1. Exploitation « non biaisée » des résultats de l'étude protéomique

La présence de 5 répliquas par condition expérimentale et le nombre de peptides détectés par le spectromètre de masse rendent nos résultats suffisamment fiables. Nous avons effectué un tri sur la liste des protéines enrichies et sur la liste des protéines uniques.

La sélection des protéines les plus enrichies parmi la liste D des **169 protéines enrichies** est un facteur de confiance supplémentaire. Par ailleurs, parmi ces protéines, seules celles ayant été enrichies dans chacune des trois listes A, B et C ont été conservées pour la validation

biologique. Une liste de 15 protéines uniques a ainsi été sélectionnée avec les taux d'enrichissements présentés dans le Tableau 5.

candidat	Liste A : Y334F/- Enzyme	Liste B : Y334F /WT	Liste C : Y334F / Y334A
Myo1D	7,6	$+\infty$	7.3
PI4KB	21	19	21
EXD2	$+\infty$	57	71
GGCT	13	12	10
SMARCC1	4	6	6
ETAA1	$+\infty$	4	13
NMT1	2,6	1,8	3.8
PPA1	2,6	1,4	2.2
RPL23	2,8	2,1	2.1
RPS21	7	2,1	1.7
RPS26	1,7	3	2.6
LYAR	2	2,6	2
PPL	$+\infty$	2	2
SNRPE	4,6	2,3	1.2
TK1	1,7	2	2.4

Tableau 5 : Liste des candidats de PR-Set7 sur la base de leurs taux d'enrichissement dans les comparaisons de la condition du mutant Y334F face aux trois conditions contrôles

La liste D de **545 protéines uniques** est aussi porteuse d'informations. Les protéines de cette liste ne sont retrouvées que dans la condition du mutant Y334F. Bien qu'il ne soit pas possible de calculer ni taux d'enrichissement ni p-value pour les protéines uniques, les protéines retrouvées de façon récurrentes dans les trois listes A, B et C de protéines uniques et avec un nombre le plus élevé possible de peptides protéotypiques restent de solides candidats substrats de PR-Set7. Nous avons donc élaboré une méthode de façon à réduire au maximum cette liste de protéines de façon à ajouter quelques candidats pour lesquels procéder à la validation biologique.

Pour ce faire, un premier critère de sélection a été appliqué : « 1 peptide protéotypique au moins doit être trouvé dans au moins 4 répliquas sur les 5 ». Les protéines satisfaisant cette condition sont ajoutées à une liste E de protéines. Ce critère de sélection favorise la reproductibilité entre

les répliquas d'une même condition et permet de restreindre le nombre de candidats à un nombre de 106 protéines (annexe S6). Un deuxième critère de sélection a été appliqué de façon à affecter un poids plus important aux protéines avec un nombre élevé de peptides et ce, dans les 3 listes A, B et C de protéines uniques. Fixer arbitrairement un seuil d'au moins 7 peptides utilisés pour l'analyse quantitative nous a permis de restreindre la liste E à seulement 8 protéines que sont : ANAPC16, BRAP, EPS8L1, FAM208A, ISCU, NDC80, PNISR et SRFBP1 (Tableau 6).

Nombre de peptides	Liste A	Liste B	Liste C	Couverture de séquence
ANAPC16	7	7	7	26,4 %
BRAP	8	10	9	6,4 %
EPS8L1	7	7	7	5,7 %
FAM208A	19	8	9	7,1 %
ISCU (NIFUN)	10	10	10	11,4 %
NDC80	15	15	15	5,8 %
PNISR	9	7	7	6,1 %
SRFBP1	9	9	9	7 %

Tableau 6 Tableau du nombre de peptides protéotypiques utilisés pour la quantification

IV.4.2. Premières étapes de la validation biologique des candidats substrats

Cette liste de protéines en main, nous avons cherché à valider leur pertinence biologique. Pour cela, une première étape consiste à produire les protéines d'intérêt et à vérifier grâce à un test enzymatique *in vitro* si l'on observe effectivement un transfert de méthyle sur l'un de ces candidats, puis à caractériser la position de la méthylation. Par la suite, si une protéine est validée, il faudrait produire des anticorps appropriés pour réaliser une étude biologique plus poussée qui ne pourrait pas rentrer dans le cadre de cette thèse. J'ai cependant eu la possibilité d'entamer les premières étapes de cette validation biologique.

L'association des sélections des **15 protéines enrichies** et des **8 protéines uniques** a constitué notre panel de candidats de PR-Set7. Pour chacun de ces candidats, nous avons d'abord cherché si des constructions génétiques contenant le gène codant la protéine entière, et permettant une expression en bactérie, en cellule d'insecte ou en cellule de mammifère existaient dans la banque de plasmide Addgene. Cet organisme référençant des plasmides ayant déjà été utilisés par d'autres laboratoires est une garantie que la construction permet effectivement l'expression

de la protéine d'intérêt. Nous avons ainsi pu nous procurer les plasmides codant pour NMT1, PI4KB pour une production en bactérie et SMARCC1 pour une production en cellule d'insecte. Les constructions pour les autres protéines ont été commandées dans la banque de plasmide de DNASU. Bien que les plasmides n'aient que rarement été utilisés, cette banque a l'avantage de proposer un très grand nombre de constructions géniques permettant théoriquement une expression dans n'importe quel système.

Etant donné le grand nombre de candidats à exprimer, nous avons utilisé la stratégie Gateway® de Invitrogen (Figure 58). Cette technique de biologie moléculaire est une technique assimilable à du clonage reposant sur l'usage de séquences de recombinaisons propriétaires nommées « Gateway Att ». Ces séquences flanquant un gène présent dans un plasmide « d'entrée » peuvent recombiner avec les séquences d'un plasmide « de destination » pour former deux nouveaux plasmides dont un plasmide « d'expression » permettant l'expression d'un gène dans un système d'expression. Le second plasmide généré contient le gène CcdB, utilisé comme un marqueur de sélection positive, il entraîne la mort de la cellule qui le contient. Dans notre contexte, l'utilisation de cette stratégie a deux gros avantages. Le premier réside dans le fait de pouvoir cloner chacun des 20 gènes dans un même vecteur de destination pour former des vecteurs d'expressions similaires. Ainsi les éléments induisant la transcription puis la traduction du gène d'intérêt sont strictement les mêmes d'un candidat à l'autre et autorisent l'utilisation de moyens de production et de purification des protéines exprimées équivalents entraînant un gain de temps considérable mais limitant les opportunités d'optimisation de protocole. Le second avantage est la possibilité d'exprimer un même gène dans des systèmes différents, le système permettant de passer très rapidement d'une construction génétique maîtresse (plasmide d'entrée) à des constructions génétiques permettant une expression dans tout système (bactérien, insecte, mammifère, ...). Ainsi une protéine dont nous aurions constaté qu'elle n'est pas exprimée dans un système bactérien pourrait être très facilement exprimée dans un modèle d'insecte ou de mammifère.

En somme, cette stratégie simplifie drastiquement la méthode de production des nombreux candidats substrats de PR-Set7.

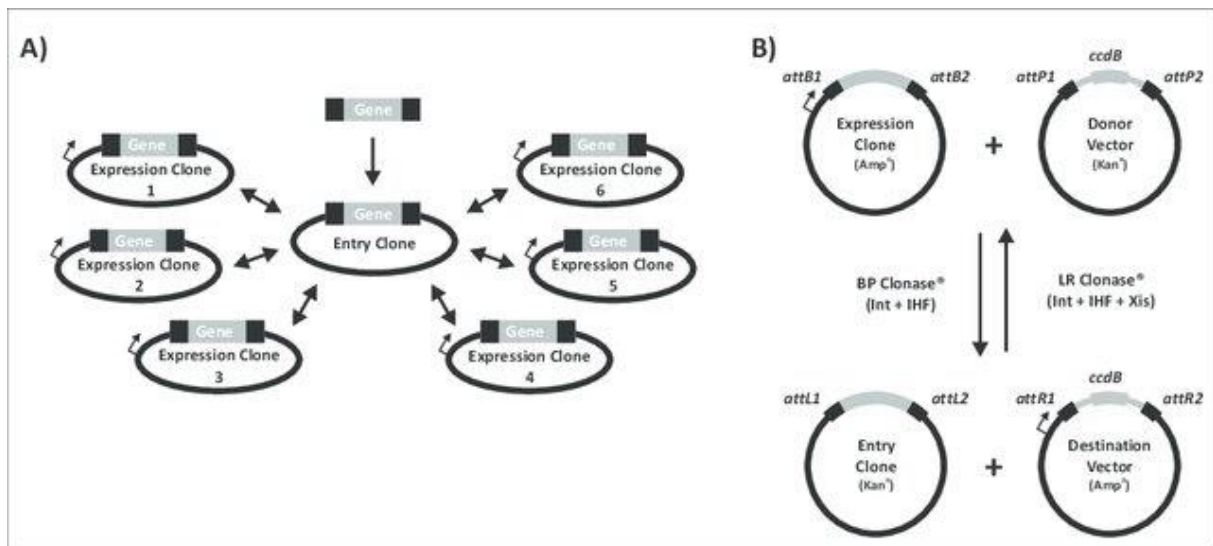


Figure 58. Technique de clonage Gateway® de Invitrogen. A) Schéma représentant l'obtention de multiples vecteurs d'expression à partir d'un même vecteur d'entrée. B) Schéma de l'obtention du plasmide d'expression à partir d'un plasmide donneur (entry clone) et d'un plasmide receveur (destination vector) ainsi que la réversibilité de la réaction.

Les détails concernant les conditions de production, d'expression des protéines et les résultats des tests de méthylation par autoradiographie sont compilés dans le Tableau 7. Pour cette étude, j'ai réalisé l'ensemble des constructions génétiques permettant d'obtenir les vecteurs d'expressions bactériens et en cellule d'insecte. Pour les productions de protéines, j'ai été aidé par Audrey Michaud de l'équipe de recherche sur les mécanismes de répression par les protéines Polycomb (dirigée par mon encadrant de l'Institut Curie, Raphaël Margueron). Audrey Michaud a par ailleurs réalisé l'ensemble des tests *in vitro* de méthylation qui doivent être révélés par autoradiographie.

En résumé, parmi les 23 candidats substrats sélectionnés de l'expérience de protéomique, nous avons pu nous procurer les constructions génétiques codant pour les protéines de 20 d'entre eux. Parmi celles-ci, nous avons d'abord tenté de les exprimer dans un système bactérien en transformant les plasmides d'expression bactériens dans des bactéries BL21(DE3)pLysS. Les bactéries ont été cultivées à 37°C jusqu'à atteindre une DO de 0,6 et la surexpression des candidats a été induite par ajout de 0,1 mM d'IPTG pendant 3 h à 37°C. Des tests de solubilité ont en parallèle, été effectués en prélevant des aliquotes ayant subi ou non l'induction à l'IPTG et après dépôt sur gel dénaturant d'une fraction. Ainsi les protéines au moins partiellement solubles ont été purifiées par affinité sur colonne de Nickel.

Tableau 7. Synthèse de l'état d'avancement de la production et de la validation biologique *in vitro* des candidats substrats de PR-Set7. ND : non déterminé, Pas concerné signifie que l'évaluation de l'expression, de la purification, de la méthylation n'a pu être effectuée en raison d'un blocage sur une étape en amont (Ex : une protéine non purifiée à cause de son insolubilité)

Protéine	Système d'expression	étiquettes	Exprimée (oui/non/peu)	Purifiée (oui/non)	Validation de la méthylation <i>in vitro</i> (oui/non)	Commentaires
ANAPC16	ND	ND	ND	ND	ND	Plasmide non livré
BRAP	ND	ND	ND	ND	ND	Plasmide non livré
EPS8L1	Bactérien	6-His	Peu	Oui	Non	
	Baculovirus	6-His-TEV	Oui	Oui	Non	
ETAA1	Bactérien	6-His	Peu	Non	Pas concerné	Protéine insoluble
	Baculovirus	6-His-TEV	Non	Non	Pas concerné	
EXD2	Bactérien	6-His	Oui	Non	Pas concerné	Protéine insoluble
FAM208A	Bactérien	6-His	Non	Pas concerné	Pas concerné	
	Baculovirus	6-His-TEV	Oui	Oui	Non	
GGCT	Bactérien	6-His	Oui	Oui	Non	
NIFUN	Bactérien	6-His	Oui	Oui	Non	
	Baculovirus	6-His-TEV	Oui	Oui	Non	
LYAR	Bactérien	6-His	Non	Pas concerné	Pas concerné	
	Baculovirus	6-His-TEV	Non	Pas concerné	Pas concerné	
MYO1D	ND	ND	ND	ND	ND	Plasmide non livré
NDC80	Bactérien	6-His	Oui	Non	ND	
NMT1	Bactérien (DNASU)	6-His	ND	ND	ND	
	Bactérien (Addgene)	6-His	Oui	Oui	Non	Très soluble !
PI4KB	Bactérien	myc	ND	ND	ND	

Protéine	Système d'expression	étiquettes	Exprimée (oui/non/peu)	Purifiée (oui/non)	Validation de la méthylation <i>in vitro</i> (oui/non)	Commentaires
PNISR	Bactérien	6-His	Non	Pas concerné	Pas concerné	
	Baculovirus	6-His-TEV	Oui	Oui	Non	
PPA1	Bactérien	6-His	Oui	Oui	Non	Très soluble !
PPL	Bactérien	6-His	Peu	Oui	Non	
	Baculovirus	6-His-TEV	Non	Non	Pas concerné	
RPL23	Bactérien	6-His	Oui	Oui	Non	Protéine peu soluble
RPS21	Bactérien	6-His	Oui	Oui	Non	Protéine peu soluble
RPS26	Bactérien	6-His	Oui	Oui	Non	Protéine peu soluble
SMARCC1	Bactérien	6-His	Peu	Oui	Non	Protéine peu soluble
	Baculovirus (DNASU)	6-His-TEV	Oui	Oui	Non	
	Baculovirus (Addgene)	Flag	ND	ND	ND	
SNRPE	Bactérien	6-His	Peu	Oui	Non	Protéine insoluble
	Baculovirus	6-His-TEV	Non	Non	Pas concerné	Purifiée en conditions dénaturantes
SRFBP1	Bactérien	6-His	Oui	Non	Pas concerné	Protéine peu soluble
	Baculovirus	6-His-TEV	Oui	Oui	Non	
TK1	Bactérien	6-His	Oui	Non	Pas concerné	Protéine insoluble

Ainsi à partir des 20 constructions d'expression bactérienne, 16 induisent effectivement une expression avec des efficacités plus ou moins grandes. Parmi celles-ci, 10 ont pu être purifiées en condition native et 1 en condition dénaturante. Les 5 protéines restantes n'ont pas pu être purifiées en condition native et la faisabilité d'une purification en condition dénaturante reste à montrer.

Dans un second temps, les constructions génétiques permettant une expression en cellule d'insecte ont été réalisées pour les protéines n'ayant pas montré d'expression chez la bactérie ainsi que pour les quelques protéines peu solubles. Ainsi, parmi les 10 constructions permettant théoriquement une expression en cellules d'insecte, 6 nous ont permis de produire les protéines associées avec succès dont 3 qui n'avaient initialement pas pu être purifiées à la suite de l'expression bactérienne.

Des tests de méthylation reposant sur l'incubation de chacun des 14 candidats substrats avec de la SAM radioactive et PR-Set7 sauvage ont été effectués avec, en contrôle positif, un test sur le nucléosome reconstitué. Après réaction enzymatique, les protéines sont séparées sur SDS-PAGE et les protéines méthylées peuvent être visualisées par autoradiographie. A ce jour la méthylation n'a pu être confirmée, quel que soit le candidat testé.

Conclusion et perspectives de la validation biologique des candidats substrats de PR-Set7

Bien que les résultats de production et de purification de protéines aient pu mener à un nombre conséquent de candidat purifiés, les résultats de la validation biologique ne sont guère à la hauteur de nos espoirs. Néanmoins, des hypothèses pourraient expliquer cette absence d'observation de méthylation *in vitro* et l'hypothèse que ces candidats seraient méthylés par PR-Set7 dans un contexte plus proche de la réalité cellulaire ne peut pour le moment pas être exclue. Parmi ces hypothèses, on peut proposer :

- Une mauvaise structure tridimensionnelle du substrat liée aux conditions de production et/ou de purification. Ces conditions n'ayant pu être optimisées pour chacun des candidats, il est fort probable que la conformation des protéines en soit impactée. PR-Set7 ayant besoin d'un long motif, non structuré pour reconnaître efficacement son substrat, un repliement natif de ce dernier est absolument essentiel ;
- Une absence de protéines partenaires ou d'acides nucléiques assurant la bonne conformation du candidat, sa stabilité ou bien assurant le bon recrutement de PR-Set7 par l'intermédiaire ou non d'un contexte épigénétique ;

- Une absence de partenaire de PR-Set7, notamment d'acides nucléiques. Nous avons en effet montré que la présence d'ADN était nécessaire pour la méthylation de l'histone H4 par PR-Set7. Inversement il a été montré que la méthylation de p53 était possible sans ADN. Qu'en est-il alors de substrats inconnus ?

De manière générale, l'absence de complexité de nos conditions *in vitro* semble limitante dans notre recherche des substrats méthylés de PR-Set7. Dans ce contexte, une connaissance accrue de ces candidats est nécessaire pour optimiser les conditions de production de ces protéines ainsi que les conditions réactionnelles du test enzymatique.

D'autre part, notons que bien que les taux d'enrichissement des candidats de PR-Set7 sélectionnés soient statistiquement fiables, ils sont loin d'être extrêmement importants et seules, quelques rares protéines franchissent de peu le ratio d'enrichissement de 2. Cette étroite distribution des taux d'enrichissement des protéines est très probablement une conséquence directe de la consommation de ProSeAM par des enzymes endogènes, fonctionnalisant leurs propres substrats résultant en une capture de protéines sur les billes de streptavidine même dans les contrôles négatifs. Précisons finalement que notre choix des conditions contrôles nous assure que les protéines franchissant le seuil d'enrichissement, le sont du fait de l'activité du mutant gain de fonction de PR-Set7.

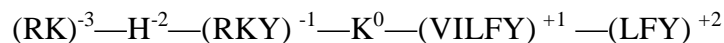
IV.4.3. Présentation de la méthode d'interprétation « biaisée » des résultats de l'étude protéomique

Une autre approche pour l'identification des candidats de PR-Set7 nous a semblé intéressante à mettre en œuvre pour essayer de trouver une certaine logique dans les résultats obtenus. Elle consiste en une analyse des résultats en tenant compte de la réalité biologique décrite dans la littérature pour chacune des protéines enrichies ou uniques. La probabilité de méthylation d'un candidat a aussi été évaluée sur la base des contraintes structurales connues pour PR-Set7 concernant le site de fonctionnalisation.

D'une part, une analyse de type Gene Ontology (GO) a été effectuée sur l'ensemble formé par les 169 protéines enrichies et les 106 protéines uniques. Chacune des protéines de cet ensemble étant associé à un certain nombre d'annotations, il est possible de déterminer s'il existe une surreprésentation d'une localisation cellulaire, d'une fonction moléculaire ou d'un processus biologique en particulier. Les analyses d'enrichissement des annotations permettent d'obtenir une p-value. Elle caractérise la significativité de la surreprésentation des termes GO renvoyés

par l'analyse en comparaison de ce qui serait obtenu pour des protéines choisies au hasard dans la base de données.

D'autre part, nous nous sommes appuyés sur les données de séquences consensus connues à ce jour. L'article de S. Kudithipudi de 2012 étudie la spécificité de substrat par une approche utilisant des puces à peptides⁸⁴. Sur la base d'un peptide fragment de l'histone H4 et en mutant chacun des acides aminés composant la séquence consensus de méthylation de PR-Set7, les auteurs ont prédit une flexibilité du site de reconnaissance par PR-Set7. Ainsi, ils proposent le site consensus suivant :



Une recherche stricte de site utilisant sur toutes les protéines humaines ne leur permit d'identifier que 4 protéines et une recherche relaxée du même site autorisant n'importe quel résidu dans les positions -1, +1 et +2 permit de trouver 59 potentiels substrats de PR-Set7. Pour chacun de ces candidats, ils produisirent des peptides d'une vingtaine de résidus contenant le site consensus qu'ils testèrent en tant que substrat de PR-Set7. Ils parvinrent à en méthyler 21 avec une efficacité variable allant de 60 % à 140 % du taux de méthylation pour le peptide fragment de l'histone H4 (Tableau 8)

Rappelons finalement que la présence d'une phénylalanine en position -2 par rapport à la lysine à méthyler n'a pas d'impact significatif sur la reconnaissance de l'enzyme par son substrat comme nous en discutons dans le chapitre II dans les paramètres structuraux essentiels pour la fixation des ligands dans le site catalytique.

Nom de la protéine	Séquence de la protéine	Lysine cible	N° Swissprot
Adénylate cyclase type 9	KINPKQLSSNSHPKHCKYSI	K44	O60503
Kinetochore-associated protein	DASMDSAKRRHPKLLAKALE	K1493	NP_055523.1
TBC1 domain family member 10A	NNWDKWMAKKHKKIRLRCQK	K103	Q9BXI6
Zinc finger protein 471	SFSKNSMVIKHKVYVGGKLF	K198	Q9BX82
Endonuclease G like 1	SKIMGDADRKHCKFKPDPNI	K113	Q9Y2C4
Hypermethylated in cancer 1	LVALCKKRLKRHGKYCHLRG	K154	Q14526
ADP-ribosyltransferase like 1	PELRLSKRKHKIPFSKRKM	K1242	Q9UKK3
Endonuclease VIII-like 3	RKAGLALSKEYKVKRPNCG	K247	Q8TAT5
Phosphoinositide 3-Kinase-C2-beta	FLCRHEKIFHPNKGYYVVK	K1377	O00750
P53_HUMAN	SKKGQSTSRHKKLMFKTEGP	K382	P04637
PR domain zinc finger protein 16	SKLDLRRHKKYTCGSVGAAL	K250	Q9HAZ2
BMP-2-inducible protein kinase	TYRTPERARRHKVGRDSQ	K1023	Q9NSY1
Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B alpha isoform	LCDRHSKLFEEPEDPSNRSF	K267	P63151
HERV-K_5q33.3 provirus ancestral Pol protein	GENQLPVWLPTRHLKFYNP	K855	P10266.2
PR domain zinc finger protein 5	VQVVHERHKKYRCELCNKAF	K460	Q9NQX1
Exportin-7	EINQADTTHPLTKHRKIASS	K185	Q9UIA9
DNA-directed RNA polymerase I subunit RPA2	APGIADSLRHFKVLREKRIP	K582	Q9H9Y6
F-box only protein 11	IRTNSCPIVRHNKIHDGQH	K504	Q86XK2
Transcription initiation factor TFIID subunit 11	VRRLLKSKGQIPNSKHKKIIF	K207	Q15544
Zinc finger protein Helios	QKGNLLRHIIKLHSGEKPFKC	K160	Q9UKS7
Zinc finger protein 505	SSTLIKHKKIHTREKPYKCE	K529	P35789

Tableau 2. Liste des peptides fragments de différentes protéines dont la méthylation par PR-Set7 a été montrée dans l'article de Kudithipudi de 2012⁸⁴

IV.4.3.A Etude des données par recherche de site consensus

Dans ce contexte, nous avons fait une recherche de site consensus ci-dessous parmi toutes les protéines humaines, la liste étant ensuite croisée avec nos listes de 169 protéines enrichies et 106 protéines uniques.

$$(RK)^{-3}-(HF)^{-2}-(X)^{-1}-K^0-(X)^{+1}-(X)^{+2}$$

où X correspond à tout résidu excepté chargé négativement

Notons que malgré cette recherche peu restrictive comparée au site consensus décrit pour PR-Set7, plusieurs critères généraux sont recherchés pour prédire la méthylation par PR-Set7. Notamment, dans l'état actuel des connaissances, la lysine à modifier devrait se trouver dans un environnement largement chargé positivement, ainsi nous admettons qu'une région riche en lysine est plutôt favorable tandis que la présence de résidus aspartate ou glutamate dans un voisinage proche d'une lysine élimine d'emblée cette dernière comme position méthylable par

PR-Set7. De plus, les positions +1 et +2 par rapport à la lysine sont généralement occupés par de petits résidus non polaires, toutefois la présence de résidus plus polaires ou de résidus aromatiques n'est pas rare. Parmi les protéines intéressantes retrouvées par cette approche nous avons identifié 3 kinases ainsi qu'une phosphatase pour lesquelles nous proposons un site de méthylation, une courte description et un bref argumentaire sur la pertinence en termes de structure et fonction.

- **MAP2K6** (Lys-141, environnement du site de méthylation envisagé : TSLDKKFYKQVUDK)

MAP2K6 est une MAP-kinase essentielle de la voie de transduction du signal médiée par la MAP kinase p38. MAP2K6 agit avec MAP3K3 pour phosphoryler les thréonines et tyrosines de plusieurs MAP Kinases p38, activant la voie de signalisation en réponse à des stress environnementaux (stress osmotique, choc thermique, etc.). MAP2K6 est connue pour phosphoryler et activer les MAP Kinases p38 en réponse à un stress inflammatoire ou un stress environnemental^{166,167}. Cette protéine est impliquée dans des processus tels que l'arrêt du cycle cellulaire induit par des stress, l'activation de transcription ou certains processus apoptotiques.

D'un point de vue structural la lysine que nous proposons comme position de méthylation est exposée en surface de la protéine d'après les données structurales connues sur cette protéine¹⁶⁸. Elle semble donc accessible, toutefois elle se situe au milieu d'une hélice α . Il est généralement admis que les HKMT à domaine SET ont besoin d'accéder à une région déstructurée de leur substrat pour apposer la méthylation. C'est particulièrement évident sur les sites de méthylation que représentent les queues d'histone ou la partie N-ter de p53 qui sont les sites les plus étudiés. Nous avons cependant relevé que pour une autre HKMT à domaine SET nommée SMYD2, cette lysine 141 avait déjà été validée comme position méthylée¹⁶⁹ bien qu'elle devrait être contrainte par la présence de cette structure secondaire en hélice α . Ajoutons que le lien entre la structure tridimensionnelle de substrats autre que les histones ou p53, et la position de méthylation par les HKMT à domaine SET n'est que très peu étudié. Ainsi, la méthylation de cette protéine à cette position est tout à fait envisageable et réaliste.

- **PRKCZ** (Lys-124, environnement du site de méthylation envisagé : RRGARRWRKLYRA)

Il s'agit d'une sérine/thréonine protéine kinase impliquée dans de très nombreuses voies de signalisation, telles que la voie PI3K/Akt et la voie des MAP Kinases¹⁷⁰. Elle est aussi impliquée dans l'activation de NF- κ B, la prolifération cellulaire, la polarisation de la cellule, la réponse à l'inflammation, etc...^{170,171} Dans la voie de réponse à l'inflammation médiée par NF- κ B, PRKCZ peut atténuer la répression dépendante de l'HKMT SETD6 de la répression des

gènes cibles de NF- κ B. Dans ce mode de fonctionnement, le facteur de transcription NF- κ B est régulé par SETD6 qui mono-méthyle la lysine 310 de la sous unité RelA du facteur de transcription. La lysine méthylée est reconnue par l'HKMT GLP dont l'activité basale de méthylation des histones a pour conséquence une compaction de la chromatine contenant des gènes cibles de RelA. La phosphorylation de la Sérine 311 de RelA bloque la reconnaissance de la lysine monométhylée voisine par GLP atténuant la compaction chromatinienne localement et diminue la répression génique des cibles du facteur de transcription (Figure 59).^{172,173}

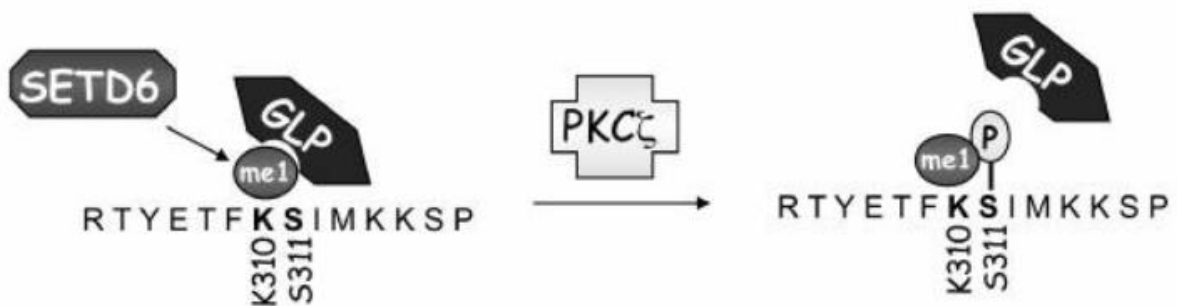


Figure 59. Interrégulation entre RelA_{K310me1} et RelA_{S311ph} exercée par l'activité de la protéine kinase de PRKC ζ (nommée PKC ζ dans la figure) et l'HKMT SETD6

Cet exemple illustre bien la complexité de la régulation introduite par les marques épigénétiques (ici méthylation et phosphorylation), et un des rôles de la kinase PRKC ζ . Une régulation de l'activité de PRKC ζ liée à une méthylation par PR-Set7, par l'intermédiaire ou non d'une restriction d'interaction de la kinase avec ces cibles ajouterait un nouveau degré de complexité. Aucune structure n'étant disponible pour cette protéine, nous ne pouvons rien conclure quant à l'accessibilité de la lysine 124 que nous proposons comme substrat de PR-Set7.

- **CDK9** (Lys-40, environnement du site de méthylation envisagé : VFKARHRKTGGQK)

CDK9 est une kinase cycline-dépendante participant à la régulation de la transcription. La lysine à fonctionnaliser serait au niveau d'un tournant β exposé à l'extérieur de la protéine¹⁷⁴. La question de l'accessibilité de cette lysine pour subir la méthylation par PR-Set7 est là encore à discuter.

- **PTPRA** (Lys-178, environnement du site de méthylation envisagé : VLYMLRFKKYKQA)

C'est une protéine transmembranaire ayant une activité tyrosine phosphatase. Le site de méthylation est situé sur une portion intra-cytoplasmique de la protéine¹⁷⁵. Ce récepteur est impliqué dans la formation des plaques d'adhérence médiées par les intégrines. Après

activation, il recrute des partenaires au niveau des plaques d'adhérence promouvant la phosphorylation de ses cibles et l'activation de petites GTPases¹⁷⁶.

IV.4.3.B Etude de l'interactome des potentiels substrats de PR-Set7

Un interactome constitué des protéines enrichies et uniques a été réalisé avec l'outil de formation de réseaux d'interaction de protéines STRING. Ce type de représentation est utile pour représenter le réseau d'association directe (physique) ou indirecte (association fonctionnelle) de nombreuses protéines. Ainsi la formation du réseau de nos 169 protéines enrichies et 106 protéines uniques révèle un réseau d'interactions avec des zones où les protéines sont très liées entre elles donc suggérant un haut degré d'association fonctionnelle (Figure 60). Ces zones où le nombre de liaisons entre chaque nœud est élevé est appelé hub et leur existence présuppose souvent une surreprésentation des termes de Gene Ontology.

Dans notre réseau nous visualisons essentiellement 3 hubs repérés grâce à des cercles en pointillés noirs sur la Figure 60. Une analyse *Gene Ontology* a été effectuée et nous a permis de déterminer les fonctions moléculaires et les processus biologiques surreprésentés au sein de ces hubs.

Le hub A correspond à des protéines dont la fonction moléculaire surreprésentée est la fixation à des acides nucléiques et leurs processus biologiques sont liés à la régulation de la transcription génique ou à des processus de maturation des ARNm.

Le hub B est composé essentiellement de sous-unités ribosomiques. Les processus biologiques surreprésentés sont par conséquent liés à la traduction.

Enfin, le hub C, dont les interactions sont moins denses que dans les deux premiers hubs a pour annotations surreprésentées la progression du cycle cellulaire, les mécanismes de réponse au stress et la régulation négative des processus métaboliques. Nous observons aussi que pour bon nombre des protéines de ce hub, elles sont impliquées dans différentes voies de signalisation comme la voie des MAP Kinases. De façon intéressante nous notons que nous retrouvons justement dans ce hub deux des trois protéines kinases identifiées comme potentiellement substrat de PR-Set7 plus tôt (MAP2K6 et PRKCZ). La troisième protéine apparaît en dehors mais proche des *hubs* A et C suggérant une certaine proximité fonctionnelle avec les protéines contenues dans ces hubs. La phosphatase PTPRA ne figurant en revanche pas dans les listes de 169 et 108 protéines (elle fait néanmoins partie de la liste élargie des 545 protéines uniques), elle n'apparaît pas dans cet interactome.

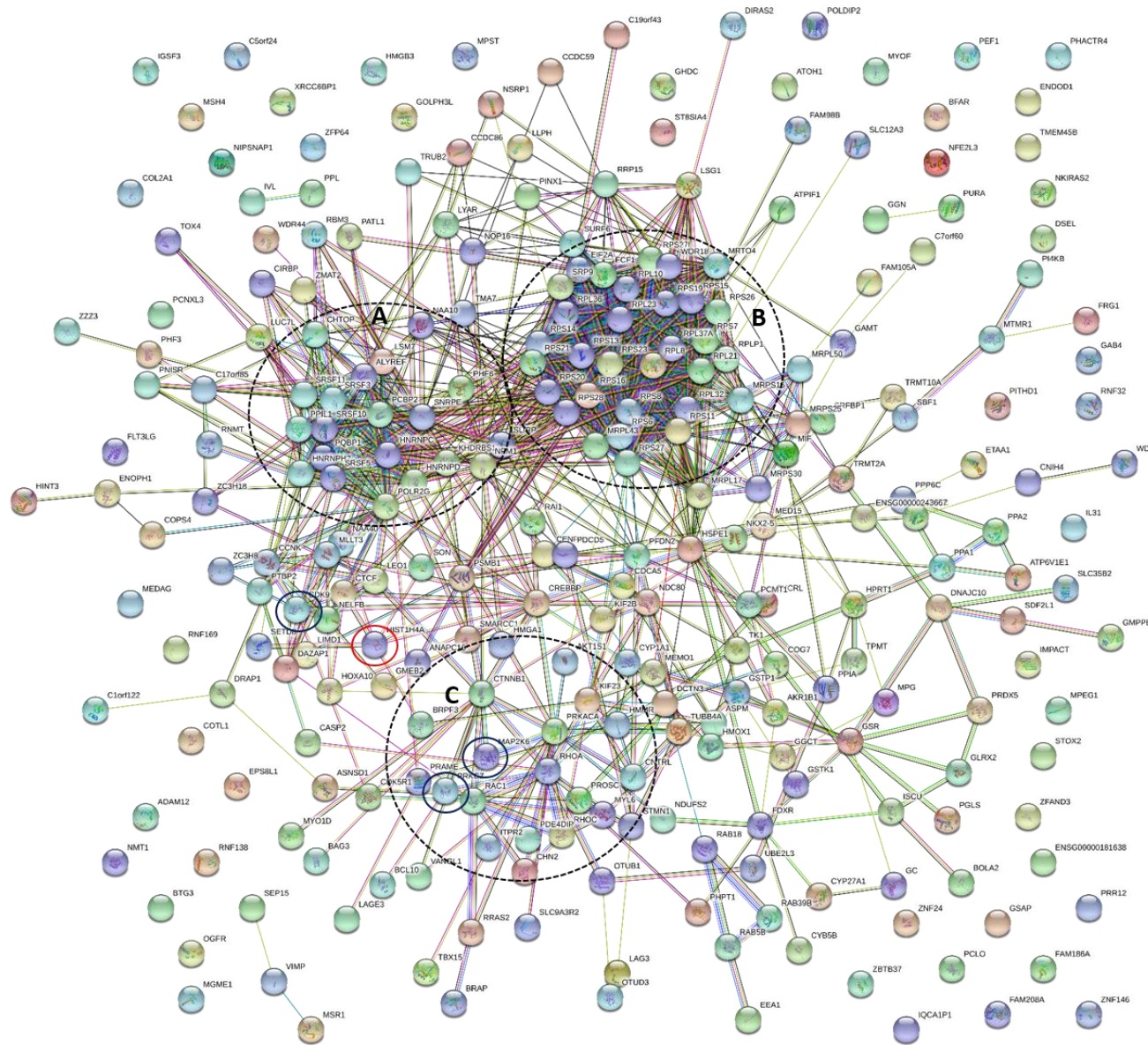


Figure 60. Interactome de la sélection des 169 protéines enrichies et des 106 protéines uniques réalisé. Les *hubs* sont encerclés de pointillés noirs, le cercle rouge correspond à l'histone H4. Les cercles bleu foncé mettent en avant 3 kinases, identifiées comme potentiels substrats de PR-Set7 par notre approche tenant compte de la structure et de la conformation des protéines.

IV.5 Etude bibliographique du lien entre les HKMT à domaine SET et les voies de signalisation impliquant des kinases

Dans cette section, nous présentons succinctement de quelle manière PR-Set7 pourrait éventuellement réguler des voies de signalisations importantes dans la cellule. Certaines voies ayant déjà été montrées comme régulées par PR-Set7, il serait intéressant d'étudier si notre enzyme d'étude peut avoir ou non un rôle majeur. Cette fin de section est entièrement spéculative mais permet de mettre en perspective nos analyses par rapport à des analyses d'autres HKMT dont les substrats non-histones et les effets biologiques induits par leur méthylation sont mieux connus.

IV.5.1 La voie des MAP Kinase

La voie des MAP Kinase est une voie permettant une adaptation rapide de la cellule à des stimuli externes. En 2014, il a été montré que l'HKMT SMYD3 triméthylait la lysine 260 de MAP3K2 entraînant une inhibition de l'interaction de cette protéine avec la phosphatase PP2a menant à une activation constitutive de MAP3K2 phosphorylée induisant une stimulation constitutive des kinases en aval de la voie de signalisation¹⁷⁷. Cette activité de SMYD3 a permis d'expliquer le rôle de cette HKMT dans certains cancers du poumon et du pancréas induit par l'oncogène RAS. Un second exemple est la monométhylation de la lysine 355 de MAPKAPK3 par SMYD2. Cette MAP Kinase est une kinase en aval de la voie médiée par la MAP Kinase p38 en réponse à un stress. Il a été montré que cet événement de méthylation par SMYD2 de cette MAP Kinase influençait la réponse inflammatoire et la production de cytokines dans des cellules cancéreuses et que l'inhibition de SMYD2 contribuait à traiter certains cancers du pancréas soulignant le rôle crucial de l'HKMT SMYD2 dans ce type de cancer¹⁷⁸.

Certaines méthyltransférases étant clairement des régulateurs de voies de signalisation des MAP Kinases, PR-Set7 pourrait tout à fait être une méthyltransférase impliquée dans la régulation de cette voie. De façon intéressante le nouveau candidat MAP2K6, s'il était effectivement méthylé par PR-Set7 agirait bien plus en amont que SMYD2 dans la voie de signalisation dépendante de p38.

L'accumulation de preuves de méthylation de MAP Kinase par des HKMT correspond à autant d'indications de l'importance des méthyltransférases en tant que protéines régulatrices de cette voie.

IV.5.2 La voie Pi3k/AKT

La voie Pi3k/AKT est une voie de signalisation intracellulaire souvent activée par la stimulation d'un récepteur membranaire. Elle joue un rôle important dans de nombreux processus cellulaires (maintien de l'homéostasie, régulation de l'apoptose, cycle cellulaire). Elle joue aussi un rôle clé dans l'oncogénèse étant donné que bon nombre des protéines la constituant ont des effets suppresseurs de tumeur ou proto-oncogène. Une des cibles de cette voie est la protéine mTOR, une kinase occupant un rôle crucial dans de nombreux processus physiologiques et métaboliques.

A nouveau, il a été montré que cette voie de signalisation pouvait être régulée par des modifications post-traductionnelles et notamment grâce à la méthylation. Pour exemple, il a été montré que l'activité de méthylation par SMYD3 de Akt1 était essentielle à son activité¹⁷⁹. Dans la même étude il est montré que cette activité de méthylation était importante pour la phosphorylation de mTOR par AKT et que l'inhibition de cet événement de méthylation prévenait la prolifération de cellules cancéreuses.

Parmi la liste des protéines enrichies, **Akt1S1** et **BMT2** sont toutes deux impliquées dans cette voie de signalisation. La première est une sous unité du complexe mTORC1. Ce complexe, formé de mTOR, mLST8, Akt1S1, RPTOR et DEPTOR, régule la croissance cellulaire et est impliqué dans la survie en réponse à des nutriments ou des hormones¹⁸⁰. Dans mTORC1, AKT1S1 régule négativement l'activité du complexe d'une manière dépendante notamment de son état de phosphorylation. BMT2 est un inhibiteur de mTORC1 grâce à des interactions établies avec les complexes GATOR1 et KICSTOR, et agit comme un senseur de la présence de SAM, lui-même indicateur de la quantité de méthionine. En présence de méthionine, BMT2 se lie à la SAM et rompt ses interactions avec GATOR1 et KICSTOR promouvant ainsi la voie de mTORC1. A l'inverse, en absence de méthionine, les taux de SAM sont réduits et BMT2 s'associe à GATOR1 et KICSTOR inhibant la voie mTORC1. Une recherche de site consensus de PR-Set7 dans la protéine BMT2 pourrait laisser penser que la portion : **KSVHRRLRKKYREV**, contenant la Lys-49, pourrait éventuellement être une cible de méthylation. Bien qu'il n'y ait pas de résidu aromatique en position -2 de la lysine 49, le site reste très chargé positivement. Aucune structure 3D de cette protéine n'est connue à ce jour et aucune modification post-traductionnelle dans cette région de la protéine n'a été référencée.

Ces 2 potentiels substrats de PR-Set7 découverts dans cette étude pourraient suggérer un rôle de PR-Set7 en aval de la voie de signalisation de Pi3K/AKT et en amont de la voie de mTOR.

IV.5.3 La régulation de facteurs de transcription par les MTases

Au cours de ces dernières années de plus en plus de facteurs de transcription ont été découverts en tant que substrats de méthylation par des HKMT à domaine SET. Le plus connu étant p53, méthylé essentiellement en sa partie C-terminale par de multiples HKMT dont notamment SMYD2, Set7/9, G9a, PR-Set7⁶². Les différents événements de méthylation, combinés à d'autres modifications post-traductionnelles (acétylation, phosphorylation, ubiquitinylation...) introduites par d'autres enzymes sur les mêmes lysines ou sur des résidus voisins, révèle une grande complexité de régulation de l'activité de p53 et a été un premier pas dans la généralisation du concept de code histone étendu à des protéines non-histones¹⁸¹.

IV.6 Conclusion

Pour conclure, notre étude du méthylome révèle de nombreuses cibles potentielles de PR-Set7. Parmi elles, nous avons retrouvé des protéines, rassemblées au sein d'un premier hub, impliquées notamment dans la transcription génique. Ceci est cohérent avec le rôle connu de PR-Set7 dans la compaction de la chromatine et le maintien de la répression génique médiée par la méthylation de H4K20.

Les autres potentiels substrats suggèrent notamment un rôle de PR-Set7 dans la progression du cycle cellulaire et en tant que protéine régulatrice de voies de signalisations majeures dans la cellule. La validation biologique de certains de ces candidats pourrait résulter en une meilleure compréhension du rôle essentiel occupé par PR-Set7 au cours du développement embryonnaire et pourrait permettre d'expliquer la différence de gravité entre les phénotypes exprimés en absence de l'enzyme contre celui observé lorsque H4K20 est mutée en une alanine.

Malgré les points faibles de notre couple PR-Set7 Y334F / ProSeAM relevés dans le chapitre III qui sont essentiellement liés à la faible activité enzymatique de l'enzyme *in vitro* et conduisant à une difficulté de discerner les marquages spécifiques, nous avons pu mettre en œuvre la stratégie d'étiquetage initialement envisagée en prenant en compte au préalable un certain nombre de précautions pour être en mesure d'obtenir des résultats fiables. Nous avons validé le fait que le couple enzyme/cofacteur introduisait bien l'étiquette alcyne exclusivement sur la lysine 20 de l'histone H4 lors d'un test enzymatique sur nucléosome, ce qui valide l'approche et exclu l'idée que la mutation pourrait changer la spécificité de substrat.

L'étude protéomique sur des extraits nucléaires de cellules HEK293T a par ailleurs permis de retrouver parmi les protéines enrichies l'histone H4, déjà connue pour être substrat de PR-Set7 mais pas les autres histones. Le contraire aurait largement remis en cause la méthode et ce résultat a été un résultat très encourageant pour la validité de notre étude. Les répliques au nombre de 5 mis en œuvre et la méthodologie employée pour définir les protéines enrichies de façon spécifiques ont permis de sortir une liste de protéines fiable et nous sommes convaincus que les protéines enrichies sont bien liées à l'intervention du couple enzyme/cofacteur.

Nous avons eu la chance de pouvoir aller plus loin que l'obtention d'une liste de protéines grâce à notre collaboration avec le Dr. Raphaël Margueron qui nous a aidé à produire et évaluer un certain nombre de protéines candidates grâce à un test enzymatique *in vitro* utilisant PR-Set7 sauvage et de la SAM tritiée. De façon frustrante, aucun candidat n'a pour le moment pu être validé, les tests ne révélant pas le candidat méthylé en présence de SAM radioactive. Cela ne les exclue pas définitivement comme potentiels substrats car comme nous l'avons vu pour l'histone H4, PR-Set7 n'effectue cette méthylation *in vitro* que sur un substrat ayant un environnement proche de celui de la cellule, le nucléosome reconstitué, tandis que la méthylation par la SAM radioactive n'est pas observée avec l'octamère pris comme substrat. L'hypothèse d'une prise en charge de ces substrats dans un environnement plus complexe, notamment en présence d'ADN ou d'autres partenaires cellulaires ne doit donc pas être exclue. Nous creuserons d'ailleurs cette piste afin de finaliser notre validation biologique.

De plus, notre exploitation des résultats de protéomique tenant compte de la présence du site consensus au sein de protéines et de différentes caractéristiques structurales nous a amené à envisager quatre nouveaux substrats de PR-Set7 que nous essaierons aussi de valider.

En parallèle de cette étude, nous nous sommes questionnés sur la possibilité d'obtenir des mutants plus efficaces pour obtenir un enrichissement plus important et augmenter nos chances de trouver des protéines candidates. Nous avons initié notre travail de conception des mutants sur la base de superposition de structures. Au regard de tout ce que nous avons appris lors des études de criblage, il nous a semblé approprié de terminer nos travaux par une étude de docking afin de mieux comprendre les paramètres structuraux guidant la prise en charge des différents analogues et de proposer des mutants potentiellement plus performants et satisfaisant mieux les critères d'orthogonalité si important dans cette stratégie de marquage chémoenzymatique.

Chapitre 5

Analyse structurale - Rationalisation du design des paires enzyme mutée / analogue de cofacteur

V. Analyse structurale - Rationalisation du design des paires enzyme mutée / analogue modifié

V.1. Contexte et objectifs de l'étude de docking

Les résultats des expériences de criblage sur PR-Set7 ont permis de révéler que l'enzyme sauvage mise à part, seul le mutant Y334F est capable de prendre en charge l'analogue de cofacteur ProSeAM pour propargyler l'histone H4. Aucune des autres molécules de la série d'analogues de SAM n'a montré une quelconque activité vis-à-vis des mutants de notre méthyltransférase d'intérêt dans les conditions testées. A l'inverse, nous avons montré que la totalité de ces analogues pouvaient être utilisés par le simple mutant Y1211A de GLP. Le double mutant de PR-Set7 Y336A/Y334F semblait être le mutant le plus prometteur au sens où il mimait au mieux la structure du mutant de GLP Y1211A de par les superpositions de structures que nous avons effectuées, mais il n'a pas montré non plus d'activité pour les différentes petites molécules. Ces résultats corrént bien avec le fait que le site de fixation du cofacteur est beaucoup plus contraint chez PR-Set7 que chez GLP mais restent quelque peu décevants car ils ne permettent pas d'exploiter tout le potentiel de cette stratégie d'étiquetage des protéines, en tout cas d'étendre facilement les données d'une enzyme à une autre pourtant structuralement proche. En particulier, étant donné que ProSeAM est pris en charge par de nombreuses enzymes sauvages^{158,182,183} (SETDB1, G9a, GLP, SUV39H2, Set7/9, PRMT1, Clr4, DIM-5), les protéines fonctionnalisées par le mutant gain de fonction de PR-Set7 ne représentent qu'une partie des protéines fonctionnalisées dans un mélange complexe (e.g. extrait nucléaire) déjà présentes dans les contrôles négatifs. Ce biais donne comme résultat, dans l'expérience de protéomique, des taux d'enrichissements corrects mais qui restent relativement faibles. Ces taux auraient été vraisemblablement bien plus importants si l'analogue de SAM utilisé n'était pas pris en charge par les enzymes endogènes et que nous avions disposé d'un système parfaitement orthogonal comme dans le cas de GLP Y1211A. Une compréhension plus approfondie du rôle des résidus impliqués dans l'accommodation des analogues du cofacteur dans sa poche, voire de ceux impliqués dans la catalyse, permettrait éventuellement de concevoir :

- De nouveaux analogues de SAM, plus encombrants que ProSeAM, qui seraient pris en charge par un des mutants de PR-Set7 sans l'être par d'autres enzymes non mutées ;
- De nouveaux mutants de PR-Set7 plus permissifs que les mutants déjà conçus.

Dans ce contexte, nous avons mis en œuvre une méthode de modélisation moléculaire pouvant répondre à ces questions. La stratégie choisie utilise le logiciel *Discovery Studio*, qui permet de simuler des systèmes composés de macromolécules et de petites molécules. Ce logiciel est une interface qui inclut notamment des outils permettant le docking de ligands dans des récepteurs, le raffinement des poses générées et leur *scoring*. Appliqué à notre cas, le docking a pour objectif de prédire la conformation et l'orientation préférentielle de deux ligands, l'analogue de SAM mais aussi le substrat, dans leur site de fixation respectif. Ainsi il nous permettra de rationaliser les résultats obtenus lors du criblage des analogues de SAM envers les mutants de PR-Set7. Il devrait également servir d'outil de prédiction pour la conception de nouvelles paires d'enzyme-cofacteur plus efficaces.

V.2 Mise au point et validation de la stratégie de docking

V.2.1 Principes généraux du protocole de simulation

Le docking est une technique *in silico* qui place une petite molécule dans un site de liaison d'une protéine selon les paramètres autorisés par un champ de force et estime l'affinité de liaison du complexe stable. Au cours d'une étape d'échantillonnage, l'algorithme génère un ensemble de poses du ligand en question à la surface d'une protéine. Idéalement, cette étape doit être capable de générer au moins une pose reproduisant le mode d'interaction optimal, ce qui implique de connaître ce mode d'interaction sur la base de données structurales. Par la suite, une fonction de score permet de hiérarchiser les résultats parmi l'ensemble des poses générées en associant à chaque pose, une énergie d'interaction.

Etant donné la méthode de conception des analogues de SAM, nous avons fait l'hypothèse raisonnable qu'il existe des conformations du récepteur (l'enzyme) aptes à reconnaître les différents ligands. Cette hypothèse nous permet de sélectionner une stratégie de docking dite « semi-flexible ». Ce type de docking, dans lequel la protéine reste « rigide » et où seul le ligand est « flexible » permet d'échantillonner les degrés de libertés conformationnels du ligand en plus des degrés de libertés translationnels et rotationnels explorés par le docking dit rigide. Par ailleurs, un docking complètement flexible, prenant en compte la flexibilité du récepteur en plus de celle du ligand, n'est pas envisageable car il impliquerait un temps de calcul trop important. Afin de proposer une simulation plus réaliste du système, des étapes de solvatation, de minimisation et de dynamique sont ajoutées aux meilleures poses générées par le docking semi-flexible. Ainsi, la minimisation optimise la géométrie des poses générées, notamment en

« relâchant » la protéine bloquée par l'algorithme de docking, simulant correctement le mode d'interaction du ligand et de son récepteur, et amène le système à un minimum d'énergie local. Quant à la dynamique moléculaire, elle permet de simuler l'évolution du système au cours du temps selon des conditions thermodynamiques définies en prenant en compte l'ensemble des degrés de liberté du ligand, du récepteur mais aussi ceux de solvant. Elle permet en outre une exploration d'un plus grand nombre de poses que la minimisation en autorisant le franchissement de barrières énergétiques plus élevées.

Les grandes étapes de notre stratégie de modélisation moléculaire sont schématisées avec le diagramme de décision ci-dessous (Figure 61). L'objectif initial de l'approche étant de rationaliser des résultats expérimentaux, une attention particulière a été apportée pour éviter l'introduction de biais de confirmation. En effet le processus a été réalisé de façon à éviter les étapes de tris subjectifs. Ainsi les étapes de tri de poses 1, 2 et 3 de l'arbre de décision interviennent essentiellement pour limiter la complexité superflue de l'analyse – qui augmenterait le temps de calcul – et sont justifiées par la conservation des complexes les plus bas en énergie.

V.2.2. Préparation des ligands et de la protéine

Préparation des protéines PR-Set7 et GLP : Les structures initiales des domaines catalytiques de PR-Set7 et GLP en complexe ternaire avec la SAH et leur peptide substrat respectif proviennent de structures aux rayons X disponibles sur la Protein Data Bank (entrées 1ZKK et 2RFI respectivement). Les fichiers des complexes contiennent tous deux des molécules d'eau potentiellement importantes pour la fixation du cofacteur (SAH) dans son site, voire nécessaires à la catalyse. Les valeurs de déplacement isotropique reflétant la mobilité de chacun des atomes, sont alors investiguées de façon à ne conserver que les molécules d'eau ayant le déplacement isotropique le plus bas et par conséquent les plus résolues dans la structure cristallographique. Par la suite, l'ensemble des mutants ponctuels de GLP et PR-Set7 sont créés en simulant 50 modèles pour chacune des mutations introduites. Seul le modèle le plus bas en énergie est conservé. Les différentes enzymes, ou mutants, sont préparées de façon à ajouter les atomes d'hydrogène, optimiser les liaisons H, supprimer les éventuels clashes stériques, optimiser la conformation de la chaîne latérale des acides aminés et protoner les résidus sur la base des pKs calculés par le logiciel pour une protéine mise dans un tampon à pH 8 et contenant 150 mM de NaCl.

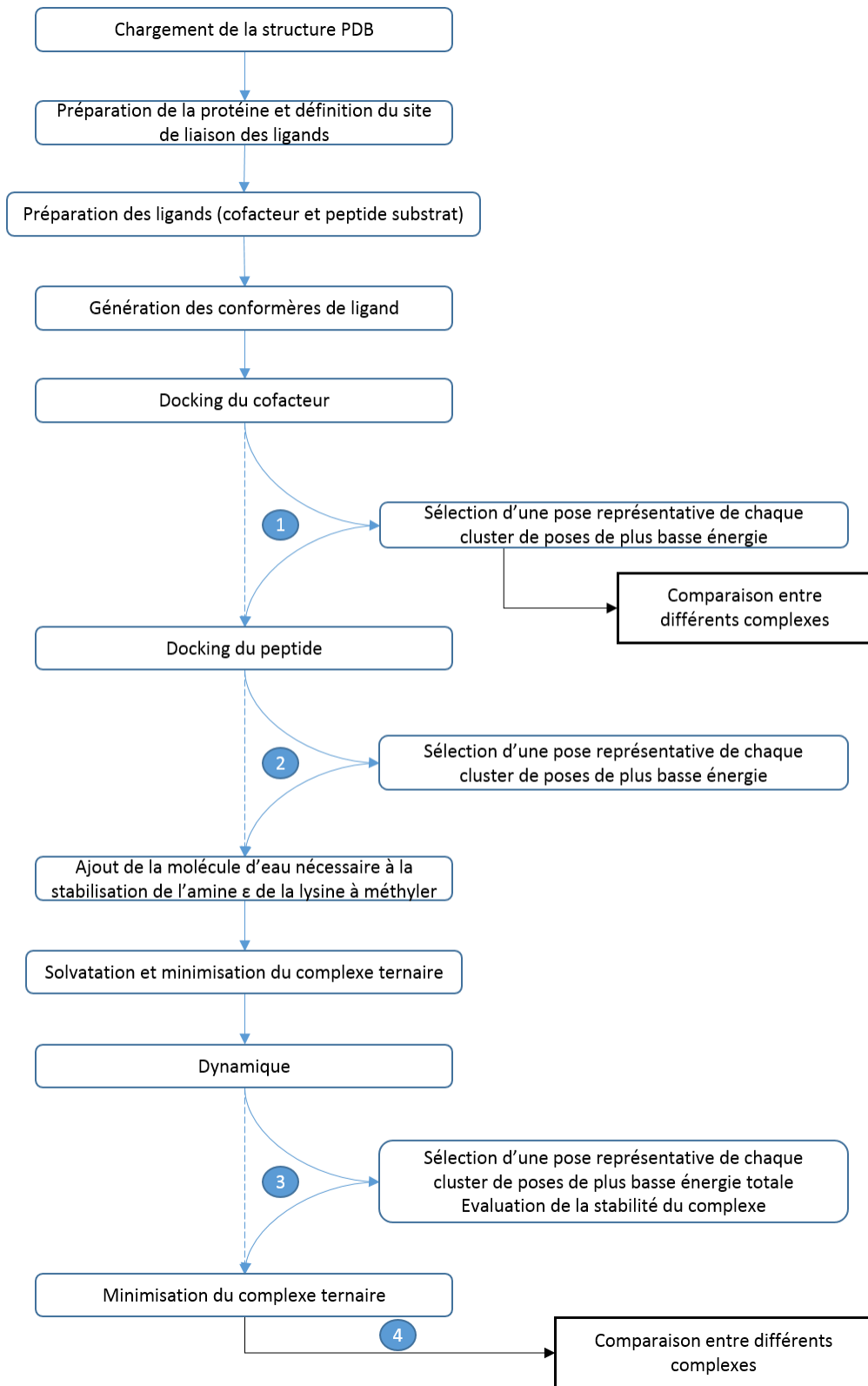


Figure 61. Diagramme de décision schématisant le processus de modélisation moléculaire.

Une sphère de 15 Å de rayon centré sur la SAH cristallisée est créée, limitant l'espace à explorer pour le docking du premier ligand : le cofacteur. De la même façon, une seconde sphère de 20 Å centrée sur le peptide est créée. Pour finir, les deux ligands sont supprimés, permettant au logiciel l'exploration des sites de fixations des deux ligands.

Chaque analogue de SAM est dessiné avec le logiciel *Chemdraw* puis importé dans *Discovery Studio*. Pour chaque molécule, tous les isomères, les tautomères et les états d'ionisation sont générés sur une gamme de pH allant de 6 à 10. La génération des deux diastéréoisomères sur le groupement sulfonium (ou le sélénonium) est particulièrement pertinente étant donné que la grande majorité de nos protocoles de synthèse d'analogues de SAM conduit à la formation de diastéréoisomères non séparables pendant l'étape de purification. Pour chaque analogue de cofacteur considéré dans cette étude, jusqu'à 250 conformères ont été générés en utilisant l'algorithme BEST de *Discovery Studio*.

L'algorithme de docking sélectionné est CDocker¹⁸⁴. Pour initier le protocole de docking, les paramètres par défaut du logiciel ont été utilisés permettant la génération de 10 poses pour chaque conformère de ligand. La qualité du docking est évaluée grâce à deux valeurs d'énergie : l'énergie CDOCKER, et l'énergie d'interaction CDOCKER. La première est calculée en se basant sur l'énergie de déformation interne du ligand et de l'énergie d'interaction récepteur-ligand. La seconde énergie renseigne sur l'énergie d'interaction non covalente qui existe entre la protéine et le ligand. Dans les deux cas, plus les valeurs d'énergie CDOCKER et d'énergie d'interaction CDOCKER sont faibles plus elles reflètent une liaison favorable du ligand avec son récepteur. Parmi, les poses de plus basse énergie, nous identifions des clusters, isolons une molécule représentative et interprétons les résultats en regard des structures cristallographiques existantes (Figure 61, 1^{ère} étape de tri).

V.2.3. Validation du protocole de docking

Dans le but de tester la validité du protocole de docking et scoring, nous avons procédé au docking de la SAH dans son site de fixation chez PR-Set7 WT, l'objectif étant d'obtenir un cluster de pose au sein duquel la SAH dockée se superpose parfaitement au positionnement de la SAH cristallisée dans le site actif. Au cours du processus de docking, 14390 poses ont pu être générées à partir de 1447 conformères de la SAH. Sur les poses générées, les 400 poses avec les valeurs de *CDocker energy* et *CDocker interaction energy* les plus basses ont été sélectionnées et ont révélé 4 clusters. Ces clusters se distinguent par les combinaisons de 2 conformations possibles de la partie méthionine de l'analogue et de 2 isomères de

conformations du ribose. Ces molécules possédant des énergies proches reflètent ainsi une certaine mobilité de ces portions de la molécule. Cette flexibilité du ligand mise de côté, ces poses sont parfaitement superposables à la SAH cristallisée dans PR-Set7.

De la même façon, nous avons re-docké la SAM dans le site de notre enzyme et avons comparé les clusters de poses générés au positionnement de la SAM co-cristallisée dans PR-Set7 avec le peptide H4 (code PDB : 5TEG). A nouveau, la superposition nous a renseigné sur le bon fonctionnement du protocole. D'autre part les contrôles similaires ont été effectués avec les systèmes GLP/SAH et GLP/SAM validant définitivement le protocole.

V.3 Etude des paramètres structuraux essentiels à la prise en charge de ProSeAM par PR-Set7

Le protocole de docking ayant été validé, nous avons dans un premier temps cherché à expliquer pourquoi, de tous les mutants de PR-Set7, seul le mutant Y334F était capable de prendre en charge ProSeAM. Pour cela, ProSeAM a été docké dans le site catalytique de l'enzyme sauvage et de chacun des mutants ponctuels (Figure 62). La superposition des structures des complexes de ProSeAM dans chacune des enzymes montre que l'orientation du groupement propargyle de l'analogue de SAM diffère dans le cas des mutants Y336A et du double mutant Y336A/Y334F par rapport aux 3 autres enzymes. En effet une rotation de 75° autour du carbone situé en α du sélénium a lieu. Cette observation est corrélée avec l'absence de la tyrosine en position 336 suggérant que ce résidu impose la conformation de la chaîne alkyle de l'analogue. Cette observation pourrait justifier la non prise en charge de ProSeAM par les deux enzymes mutées en position 336. En effet une telle orientation de la chaîne alkyle pourrait interférer avec le positionnement de la lysine du substrat à modifier. Au cours de cette étape, aucune observation ne permet de comprendre la différence de traitement de ProSeAM entre les enzymes dont la Tyr-336 n'est pas mutée (à savoir : PR-Set7 WT et Y334F vs Y334A).

Nous avons par la suite docké le peptide substrat, contenant les résidus 15 à 24 de l'histone H4 et ajouté la molécule d'eau au voisinage du groupement amine de la Lys-20 et du groupement hydroxyle de la Tyr-245 (cf. paragraphe II.2.1 sur les paramètres structuraux permettant la fixation du ligand). A cette étape, nous avons, comme pour l'analogue de SAM, vérifié que le logiciel permettait de re-docker le peptide dans une conformation comparable à celle du peptide cristallisé. Le docking du peptide réussit dans les enzymes WT et les mutants Y334A et Y334F mais échoua dans Y336A et Y336A Y334F, confirmant que l'orientation du propargyle dans

ces deux derniers mutants n'est pas compatible avec le placement de la lysine substrat. Lorsque le peptide pouvait être docké, nous avons systématiquement observé parmi les poses de plus basse énergie (Figure 61, 2^{ème} étape de tri), une majorité de poses du peptide très proches de celle du peptide cristallisé. Dans le cas des systèmes basés sur l'enzyme GLP, le même travail a été effectué en re-dockant le peptide contenant les résidus 4 à 11 de l'histone H3.

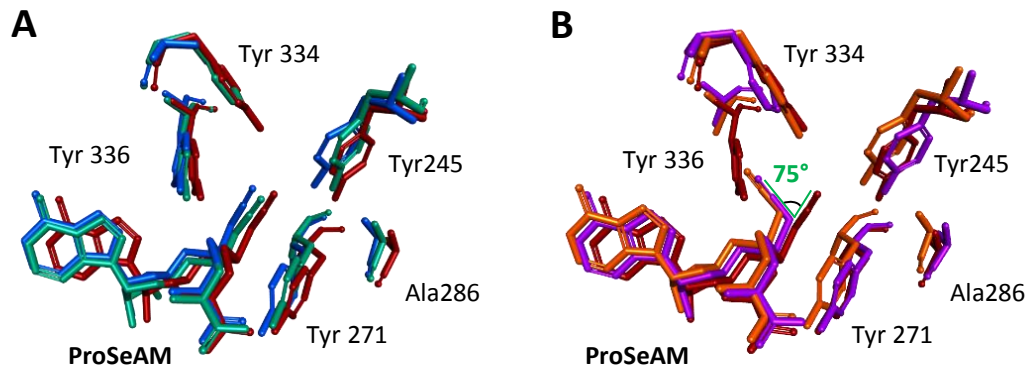


Figure 62. Superposition des poses de plus basse énergie résultant du docking de ProSeAM dans PR-Set7 sauvage ou ses mutants A. Superposition de ProSeAM dans l'enzyme sauvage (en rouge), dans le mutant Y334A (en bleu) et dans le mutant Y334F (en cyan). B Superposition de ProSeAM dans le mutant Y334F, dans le mutant Y336A (en orange) et dans le double mutant Y336A/Y334F (en violet).

Après avoir généré une « boîte d'eau » contenant 150 mM de NaCl simulant les complexes en solution tamponnée, nous avons minimisé nos systèmes pour trouver la conformation du système la plus stable possible. C'est au cours de ces étapes que, pour la première fois depuis le début du protocole, les résidus de la protéine peuvent se réarranger au voisinage de l'analogue de SAM et du peptide substrat nouvellement introduit. A nouveau, cette étape ne nous a pas permis d'observer de différence significative entre les systèmes contenant PR-Set7 WT et ses deux mutants en position 334. A partir des systèmes minimisés nous avons lancé une dynamique d'une durée de 1 ns, enregistrant des poses toutes les 10 ps pour observer les changements d'orientation des atomes au cours du temps. Nous avons finalement sélectionné les 10 dernières poses de la simulation correspondant aux 100 dernières ps, les avons superposées et clusterisées si nécessaire. Une pose représentative du cluster majoritaire a été conservée pour chacun des complexes. Ces poses ont finalement été comparées entre elles. Après superposition (Figure 63), nous observons les caractéristiques suivantes dans le complexe qui avait montré une activité *in vitro* :

- Le basculement du groupement méthyle de l'alanine 286 dans le cas du mutant Y334F
- Le décalage de la tyrosine 245 vers l'espace libéré par la mutation de la tyrosine 334 en une phénylalanine
- Le groupement propargyle de ProSeAM effectue une rotation de 60° vers Ala-286 par rapport aux deux autres mutants
- La phénylalanine 334 du mutant Y334F et la tyrosine dans la même position de l'enzyme sauvage se superposent mal

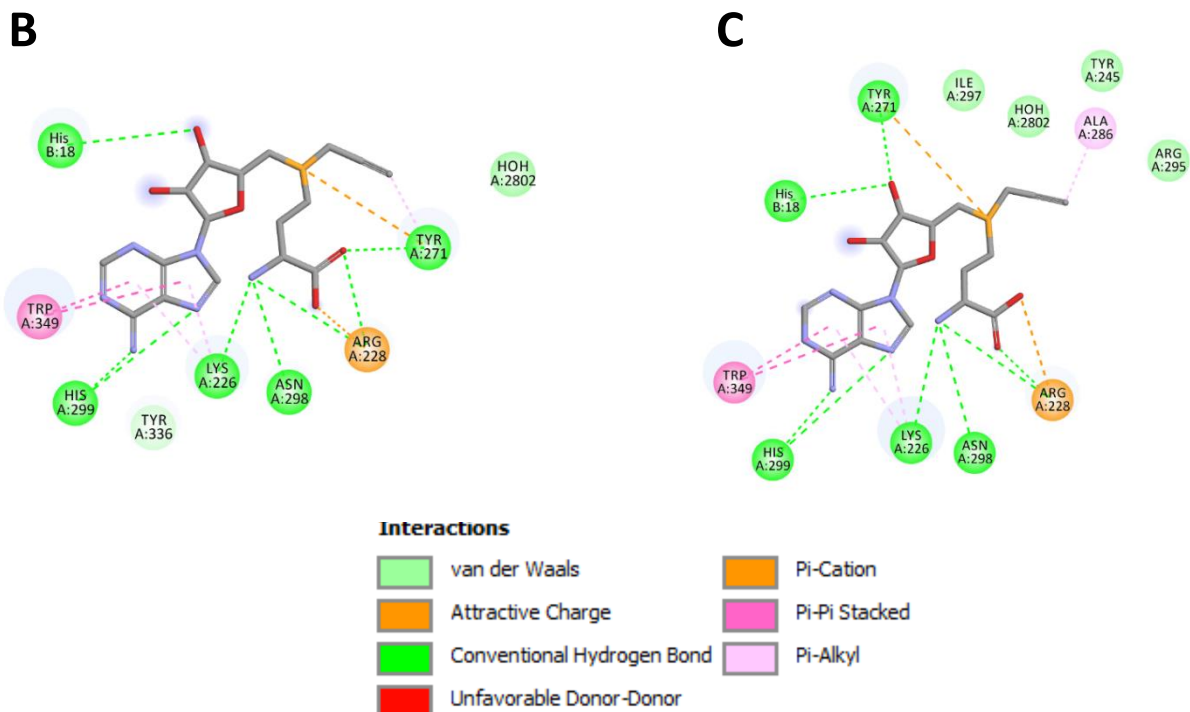
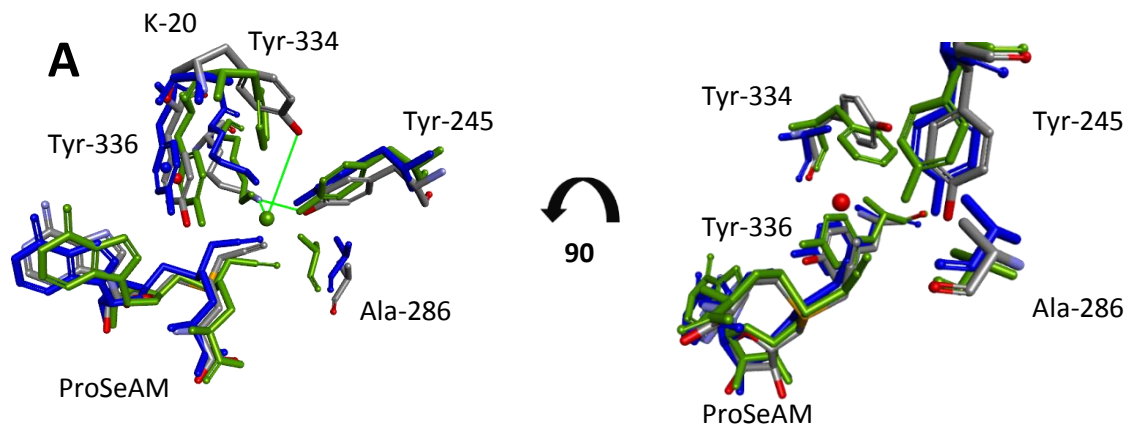


Figure 63. Interactions entre les acides aminés du site de fixation du cofacteur de PR-Set7 et ProSeAM
 A. Superposition de ProSeAM dans l'enzyme sauvage (en gris), dans le mutant Y334A (en bleu) et dans le mutant Y334F (en vert). B. Diagramme 2D des interactions de ProSeAM dans l'enzyme sauvage. C. Diagramme 2D des interactions de ProSeAM dans le mutant Y334F.

Il est difficile d'effectuer un lien de causalité entre chacun des déplacements des résidus étant donné le nombre de groupements impactés par la simple mutation d'un résidu. Cependant nous proposons l'hypothèse suivante :

La mutation de la tyrosine 334 en une phénylalanine induit le basculement de la Tyr-245 qui n'est plus contrainte que par la gêne stérique exercée par l'Ala-286. Le groupement méthyle de cette alanine peut alors établir une interaction hydrophobe avec le dernier carbone de la chaîne alkyle de ProSeAM, induisant son décalage. Ensemble, ces petites modifications de conformation dans le site catalytique pourraient être à l'origine de la prise en charge de cet analogue de SAM par le mutant Y334F. Notons, qu'il n'est pas du tout certain que le mouvement de cette alanine soit représentatif de ce qu'il se passe *in vivo* en présence du cofacteur naturel de l'enzyme. En effet l'interaction entre le résidu Ala-286 et la chaîne alkyle de la SAM ne pouvant avoir lieu, il est plutôt attendu que l'alanine reste à sa place. Cette hypothèse est en accord avec le positionnement de Ala-286 au sein des complexes cristallisés de PR-Set7 Y334F avec la SAH et le peptide non méthylé, monométhylé et diméthylé sur la lysine 20 (code PDB respectifs : 3F9W, 3F9Y et 2F9X).

Nous avons ensuite comparé les positionnements des résidus en position 334 à l'issue de la dynamique et avons remarqué que la phénylalanine du mutant Y334F se superposait parfaitement avec la tyrosine de l'enzyme sauvage complexée à la SAH. Or ce n'est pas le cas avec l'enzyme sauvage complexée à ProSeAM où l'on observe un recul, du cycle aromatique de la tyrosine chez le mutant contrastant avec le positionnement du cycle de la phénylalanine de l'enzyme sauvage. Nous en avons conclu, qu'en présence de ProSeAM, le positionnement de la tyrosine tel qu'observé dans la structure cristallographique est incompatible avec la présence de la molécule d'eau dès lors que le cofacteur est plus encombrant que la SAM. Cela s'explique par une augmentation de la gêne stérique au voisinage de la chaîne alkyle de l'analogue qui a pour conséquence manifeste de repousser la chaîne latérale de la tyrosine. Toutefois, nous savons que l'enzyme sauvage prend en charge l'analogue du cofacteur donc cet éloignement de la tyrosine 334 n'a probablement pour conséquence qu'une diminution de la stabilisation de la lysine à fonctionnaliser ce qui ferait chuter son taux de fonctionnalisation.

Notons ici que nous avons fait face à une limitation de la méthode car l'étape de solvatation ne permet pas de simuler la présence d'une molécule d'eau à cet endroit bien qu'elle soit présente dans la structure cristallographique. Nous supposons que dans les conditions « naturelles », la

molécule d'eau ne peut entrer au cœur de la protéine s'il y a à la fois une tyrosine en position 334 et le ProSeAM en tant que cofacteur. L'absence de la molécule d'eau ne permettrait alors pas un positionnement idéal de la lysine à modifier, ce qui, *in fine*, diminuerait le taux de fonctionnalisation.

Finalement, nous observons que la conformation des atomes du squelette de la lysine 20 varie de façon non négligeable dans les 3 complexes conduisant à une distance différente entre l'azote N ϵ de la lysine 20 et le carbone en α du sulfonium sur le cofacteur

$$\begin{aligned} \text{Distance N}\epsilon\text{-C}\alpha\text{WT} &= 4 \text{ \AA} \\ \text{Distance N}\epsilon\text{-C}\alpha\text{Y334A} &= 4.2 \text{ \AA} \\ \text{Distance N}\epsilon\text{-C}\alpha\text{Y334F} &= 3.8 \text{ \AA} \end{aligned}$$

Ainsi, on observe que la distance entre l'amine de la lysine substrat et le carbone α du cofacteur est la plus faible dans le cas du mutant Y334F, le plus actif. Un éloignement de 0,2 \AA est observé dans le cas de l'enzyme sauvage, ce qui est cohérent avec la déstabilisation du positionnement de la lysine induite par l'éloignement de la Tyr-334. Enfin la distance la plus grande est obtenue avec le mutant Y334A. Une perte complète du rôle stabilisateur du résidu situé en position 334 combinée au départ de la molécule d'eau, pourrait expliquer pourquoi le mutant Y334A est incapable de fonctionnaliser H4 en présence de ProSeAM.

V.4 Etude des paramètres structuraux essentiels à la prise en charge de ProSeAM par GLP

Le même protocole de docking a été tenté vis-à-vis de notre système enzymatique modèle, l'enzyme GLP et son mutant ponctuel Y1211A. Contrairement au cas de PR-Set7, le peptide substrat de l'enzyme, c'est-à-dire le fragment de l'histone H3, a pu être docké à la suite de ProSeAM à la fois dans l'enzyme sauvage et dans le mutant Y1211A (données non montrées). Nous avons conclu précédemment que la Tyr-336 de PR-Set7 contraignait le positionnement de la chaîne propargyle de ProSeAM. Cela semble être aussi le cas dans le système de GLP avec la rotation du groupement propargyle en fonction de la présence du résidu analogue la Tyr-1211. Néanmoins, alors que dans le système de PR-Set7, l'orientation du propargyle dans les mutants Y336A et Y336A Y334F, empêchait le docking du peptide substrat dans leur site de fixation respectif, ce ne fut pas le cas avec le simple mutant Y1211A de GLP. En effet, le site actif de GLP est moins contraint que celui de PR-Set7, ce qui a permis au logiciel de

proposer des poses de docking du peptide substrat. Ainsi le protocole de docking a pu être mené à son terme.

A l'issue de la dynamique, la superposition de GLP sauvage et de son mutant Y1211A révèle ainsi deux orientations possibles de la chaîne propargyle de l'analogue. La première, correspondant à l'orientation du propargyle dans GLP WT, est superposable à son équivalent dans le site catalytique de PR-Set7 Y334F. La seconde n'a pas d'équivalent chez PR-Set7 mais la chaîne propargyle occupe le vide créé par la mutation de la Tyr-1211 en une alanine. Ces observations sont essentielles car les expériences de criblage enzymatique nous ont permis d'établir que GLP sauvage et son mutant Y1211A sont tous deux capables de prendre en charge ProSeAM, le mutant étant moins efficace. Et, rappelons que c'était aussi le cas de Allyl-SAM mieux pris en charge par GLP WT que le mutant Y1211A (Fig 34 Chapitre III).

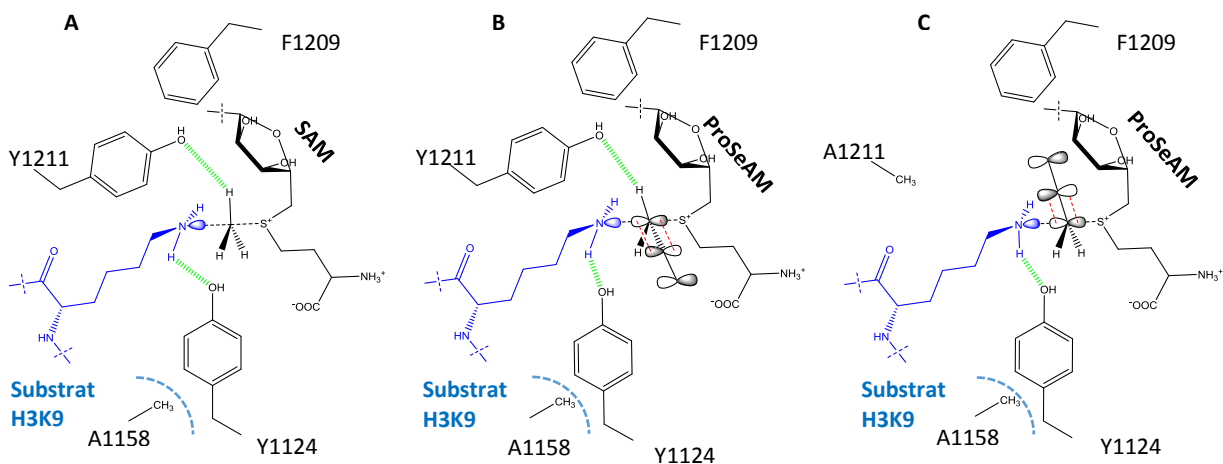


Figure 64. Modèle proposé des différents effets de stabilisation de l'état de transition de la transalkylation par l'enzyme GLP ou son mutant en position 1211. A. Stabilisation de l'état de transition de la S_N2 linéaire pour le couple GLP WT / SAM. B. Stabilisation de l'état de la réaction de transpropargylation par le couple GLP WT / ProSeAM C. Stabilité de la réaction de transpropargylation avec le couple GLP Y1211A / ProSeAM.

En sachant que le mécanisme permettant la réaction de transalkylation diffère selon si la Tyr-1211 est présente ou non, plusieurs hypothèses sont envisageables pour expliquer la façon dont ProSeAM peut être pris en charge par ces deux enzymes :

Dans le cas de l'enzyme sauvage, le mécanisme de stabilisation de la transalkylation de la lysine passe par l'établissement d'une liaison H hydrogène de type C-H...O avec la Tyr-1211. La Figure 64A. schématise cet état de transition pour GLP WT en complexe avec la SAM. Lorsque GLP WT prend en charge ProSeAM, une seule orientation de la chaîne alkyle est possible,

contrainte par la présence de la Tyr-1211. Dans cette position, deux effets stabilisateurs pourraient entrer en jeu : l'effet de stabilisation par liaison H pourrait être conservé et suppléé grâce au recouvrement orbitalaire lié à la présence de la triple liaison en position β du sulfonium (Figure 64B.). Enfin, dans le cas du mutant Y1211A, la chaîne alkyle se réorienterait en direction de l'espace libéré par l'absence de la Tyr-1211. Avec une telle orientation, l'effet stabilisateur de la tyrosine est absent expliquant la moindre activité de l'enzyme (Figure 64C).

En somme, selon si le ProSeAM se situe dans GLP sauvage ou son mutant Y1211A, sa chaîne propargyle se conformera différemment selon la contrainte exercée par le résidu en position 1211. Nous expliquons ensuite la meilleure prise en charge de cet analogue de cofacteur par deux effets stabilisateurs possible dans le cas de l'enzyme WT contre un seul dans le cas du mutant.

Mises ensemble, ces hypothèses sont autant d'arguments permettant d'expliquer pourquoi ProSeAM est utilisé ou non en tant que cofacteur par l'une ou l'autre des formes de l'enzyme. Le parallèle avec le système GLP laisse cependant une question en suspens. Nous avons vu que les formes sauvages et le mutant Y1211A (équivalent à Y336A/Y334F dans le système PR-Set7) permettaient, tous les deux, la prise en charge de ProSeAM, même si l'efficacité catalytique du mutant est bien plus faible que celle de l'enzyme sauvage. Aussi, le simple mutant de GLP permet la prise en charge d'analogues bien plus encombrants que ProSeAM. Dès lors, et en admettant que la comparaison entre les deux systèmes se tienne, il semble étonnant que le double mutant de PR-Set7 soit aussi peu permissif. Nous avons donc tenté d'évaluer par le docking les paramètres structuraux pouvant expliquer cette fois, la non-prise en charge d'analogues de SAM très encombrants (à chaînes alkyle longues).

V.5 Etude des paramètres structuraux clés dans la prise en charge d'analogues de SAM à longue chaîne alkyle

Nous avons désigné la molécule Hey-SAM comme molécule représentative de ces analogues très encombrants non pris en charge par les mutants de PR-Set7. En effet, avec sa chaîne de 6 carbones, c'est le 3^{ème} analogue plus volumineux après Pob-SAM et Ab-SAM. Utilisée par de nombreuses lysine et arginine méthyltransférases mutées¹⁶⁰, il a été montré à de multiples occasions qu'il n'était pas pris en charge par des HKMT non modifiées et est, par conséquent, un candidat idéal pour appréhender les contraintes existantes dans PR-Set7 qui n'existent pas dans ces autres systèmes. Dans la suite de cette partie nous expliquerons pourquoi les

paramètres initiaux de conception des mutants n'étaient pas adaptés pour leur permettre de prendre en charge des analogues de SAM encombrants et proposerons des solutions.

Le processus de génération des poses de Hey-SAM dans PR-Set7, GLP et leurs mutants respectifs est strictement le même qu'avec ProSeAM. Après avoir dynamisé le système, la superposition de GLP Y1211A et de PR-Set7 Y334F/Y336A nous a renseigné sur les contraintes supplémentaires exercées dans le site catalytique du double mutant de PR-Set7 comparé au mutant de GLP dont on sait qu'il est capable d'accommoder la molécule (Figure 65).

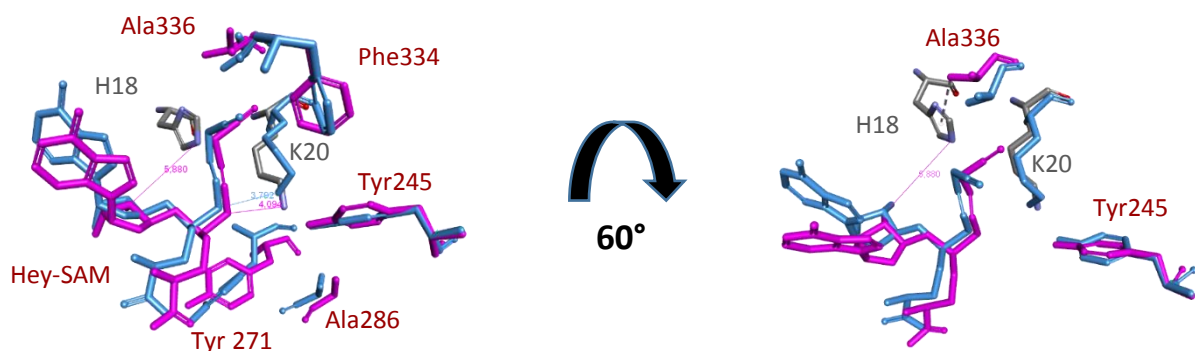


Figure 65. Superposition des poses de plus basse énergie résultant du docking de Hey-SAM dans GLP Y1211A (en bleu clair) et dans le double mutant Y334F/Y336A de PR-Set7 (en violet).

Nous remarquons dans un premier temps qu'il est possible de docker la molécule dans son site de fixation. La mutation de la Tyr-336 en une alanine, permet donc bien de libérer une cavité hydrophobe de taille suffisante pour permettre l'entrée de la chaîne alkyle du Hey-SAM. Ensuite, la superposition des deux systèmes révèle une relativement bonne superposition des cofacteurs dans leur site de fixation. Cependant nous avons identifié trois paramètres essentiels les différenciant.

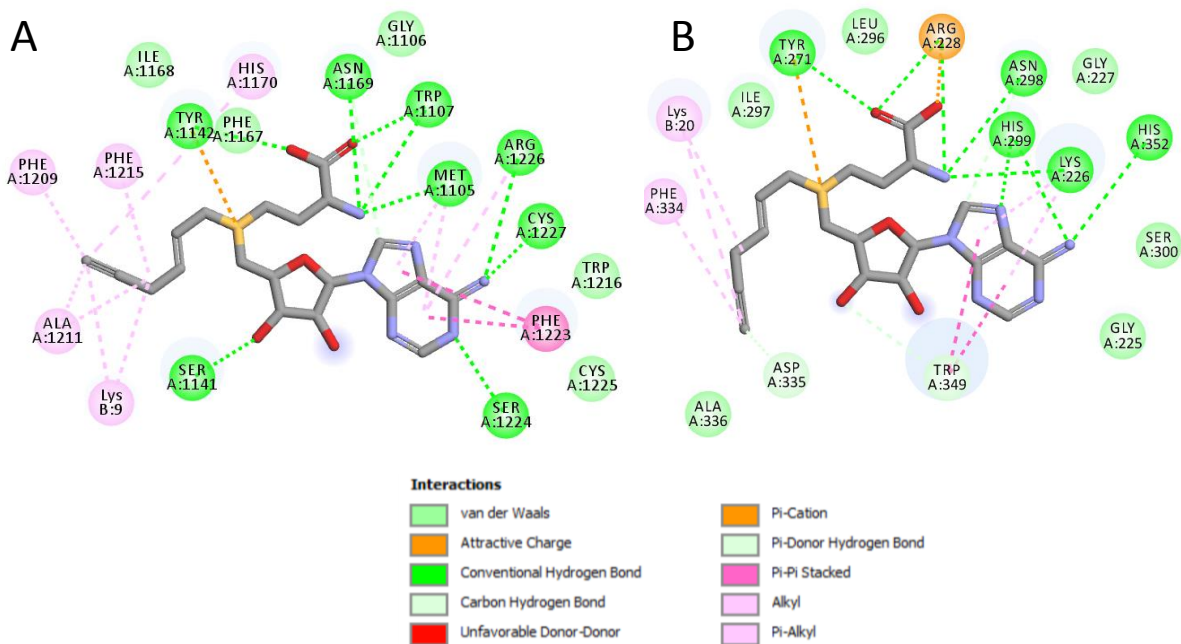


Figure 66. A. Diagramme 2D des interactions de Hey-SAM GLP Y1211A. B. Diagramme 2D des interactions de Hey-SAM dans PR-Set7 Y336A/Y334F.

Nous avons tout d'abord écarté le paramètre de l'orientation de la partie adénine du cofacteur. Formant des interactions avec des résidus des mutants de GLP ou de PR-Set7, l'orientation de cette partie du cofacteur est tout à fait cohérente avec ce qui est observé dans les enzymes non mutées et n'explique en rien la différence de prise en charge de notre analogue de cofacteur.

Nous avons également observé une parfaite superposition des chaînes alkyles des analogues jusqu'au carbone en position δ . Une comparaison des diagrammes d'interaction 2D des deux systèmes révèle cependant des réseaux d'interactions différents des deux groupements alcyne (Figure 66). Dans le modèle GLP mutée, l'alcyne forme des interactions π -alkyl avec l'His-1170 (His-299 chez PR-Set7) de la protéine et la Phe-1209 (Phe-334 chez PR-Set7). L'alcyne forme aussi des interactions hydrophobes avec le groupement méthyle d'Ala-1211 (Ala-336 chez PR-Set7) et la chaîne latérale de la lysine à modifier. Dans le système PR-Set7, ce même groupement ne forme plus qu'une interaction π -alkyl avec la Phe-334, une interaction hydrophobe avec la lysine substrat et, de façon moins attendue, interagit aussi avec la chaîne latérale de l'Asp-335. Bien que ces différences d'interaction impliquent une rotation de 61° des deux derniers carbones de la chaîne alkyle de l'analogue dans le site de PR-Set7 doublement mutée par rapport au modèle GLP Y1211A, la non prise en charge de Hey-SAM par ce mutant de l'enzyme ne peut être justifiée ainsi. En effet, cette variation de positionnement de l'alcyne terminal ne semble pas impacter le placement de la lysine à modifier dont l'amine terminale reste à bonne distance du carbone en α du sulfonium pour être alkylée.

Enfin, le dernier paramètre que nous avons examiné est le positionnement de l'histidine 18 du peptide. Il n'existe pas d'équivalent de cette histidine chez GLP ou dans son peptide substrat. Toutefois nous avons vu plus tôt que cette histidine forme un réseau d'interaction π avec les His-347 et Trp-349 de la protéine PR-Set7 et surtout qu'une liaison hydrogène avec le 3'-OH du ribose du cofacteur était essentielle à une bonne affinité de l'enzyme pour son substrat (cf. partie II.2.1 sur les paramètres structuraux permettant la fixation des ligands de PR-Set7). Or, la distance mesurée entre ce groupement OH de l'analogue de SAM et l'His-18 du substrat docké dans le double mutant de PR-Set7 est de 5,9 Å contre 2,7 Å dans le système cristallisé SAH / PR-Set7 WT / peptide H4 et 3 Å dans le système issu du docking SAM / PR-Set7 WT / peptide H4. Sachant que nous fixons un seuil à 4 Å pour admettre la possibilité d'établissement d'une liaison H, il est impossible qu'une telle interaction ait lieu dans le cas du double mutant de PR-Set7. L'absence de cette interaction aurait alors pour conséquence directe une diminution (probablement importante) de l'affinité de l'enzyme pour son substrat bien que l'analogue du cofacteur entre dans son site de fixation. **Ainsi nous expliquons l'absence d'activité observée avec la ProSeAM en présence de PR-Set7 Y336A/Y334F, non pas par une impossibilité d'accommodation du cofacteur dans le site catalytique mais par une mauvaise, voire une impossible, reconnaissance du substrat.**

Pour terminer, cette absence d'interaction entre l'histidine du peptide et le sucre du cofacteur, est nécessairement liée à la question de l'orientation des derniers carbones de la chaîne alkyle de l'analogue. Le lien de causalité n'est toutefois pas évident. Nous montrons que dans le double mutant et contrairement à l'enzyme sauvage, l'histidine 18 du peptide interagit avec le groupement méthyle de l'alanine 1211 (mutée à partir de la tyrosine), l'éloignant du centre du site actif et l'empêchant d'établir une interaction avec le groupement alcyne, comme observé avec le système modèle.

Ensemble, ces résultats permettent de concevoir d'un point de vue théorique, pourquoi le Hey-SAM ne serait pas pris en charge par le double mutant de PR-Set7 grâce à une meilleure appréhension des paramètres structuraux. Ceci nous a permis par la suite, d'émettre différentes hypothèses quant à la conception de mutants de 2^{ème} génération qui devraient être capables de prendre en charge ce type d'analogues du cofacteur.

V.6 Proposition de paires d'enzyme et d'analogue de 2^{ème} génération

L'étude de docking, nous a permis de comprendre en partie *a posteriori* les résultats enzymatiques obtenus au cours de nos travaux, de les rationaliser et de proposer des explications sur la grande différence entre ces systèmes enzymatiques pourtant si proches du point de vue structurale. Mais est-ce que ces données pourraient permettre également de concevoir de nouveaux mutants plus performants sur la base de cette connaissance fine du site catalytique s'appuyant sur résultats expérimentaux et docking ? C'est ce que nous avons essayé d'étudier pour clôturer ce travail. Pour cela nous avons émis plusieurs hypothèses.

V.6.1 Conception de mutants et étude par docking

La 1^{ère} hypothèse a été émise sur la base des résultats obtenus avec l'analogue ProSeAM. Nous avons précédemment conclu que la tyrosine 245 et l'alanine 286 jouaient de concert et devaient permettre une meilleure accommodation de ProSeAM dans le site catalytique grâce au contrôle du positionnement de la tyrosine 245 par le couple de résidus Ala-286 et Tyr-334.

Cette alanine fait partie des résidus très conservés au sein des HKMT à domaine SET (annexe S2). Il a notamment été montré dans la littérature que le résidu équivalent chez l'HKMT EZH2 (Ala-677), stabilisait le positionnement de la tyrosine 641 (équivalent de notre Tyr-245) formant une liaison hydrogène avec la lysine substrat, H3K27^{185,186}. Ceci a pu être montré grâce à l'existence du mutant A677G identifié dans des lymphomes et pour lequel il a été établi que la mutation accroissait l'efficacité de méthylation par rapport à l'enzyme sauvage.

Par analogie, nous avons choisi de générer, *in silico*, un nouveau mutant ayant la mutation A286G en plus de la double mutation. Nous avons ainsi obtenu le triple mutant A286G/Y334F/Y336A auquel nous avons appliqué le processus de docking afin de déterminer si la molécule Hey-SAM et le peptide substrat se plaçaient correctement à l'intérieur du site actif (Figure 67). La superposition du triple mutant avec le double mutant révèle que cette nouvelle mutation est effectivement associée à une légère réorientation de la Tyr-245. Nous notons par ailleurs que la lysine substrat adopte sa conformation la plus détendue, permettant au groupement amine de se placer au plus proche possible du carbone en α du sulfonium. Cette possibilité de placement de la lysine, plus propice à l'alkylation est probablement la conséquence directe de la réorientation de la Tyr-245. Ce résultat est donc parfaitement

cohérent avec les hypothèses émises sur le rôle des résidus Ala-286 et Tyr-245 de PR-Set7 ainsi qu'avec le rôle des résidus analogues d'autres HKMT décrit dans la littérature.

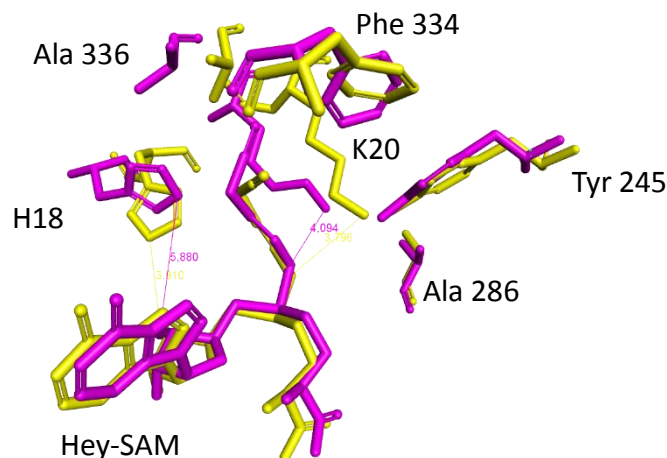


Figure 67. Superposition des poses de plus basse énergie résultant du docking de Hey-SAM dans PR-Set7 Y334F Y336A (en violet) ou dans le triple mutant A286G Y334F Y336A (en jaune).

De façon plus surprenante nous observons aussi une rotation de la chaîne alkyle de l'analogue de SAM autour du carbone δ de 56° qui résulte en une conformation parfaitement superposable au système GLP Y1211A / Hey-SAM fonctionnel. Une analyse des interactions montre que le dernier carbone de la chaîne alkyle forme des interactions alkyl- π avec l'His-299 et la Phe-334. Elle forme aussi une interaction hydrophobe avec la chaîne latérale de la lysine 20 du peptide H4. Ce réseau d'interactions est donc lui aussi parfaitement conforme avec celui observé au niveau du système GLP. Enfin, nous constatons que l'histidine-18 est à une distance de 3,9 Å du 3'-OH de Hey-SAM, rendant limite mais possible une interaction de type liaison H. La restauration de ce dernier paramètre devrait donc assurer à ce nouveau mutant une reconnaissance de son substrat.

La seconde hypothèse est basée sur les observations de la non-prise en charge de Hey-SAM dans le double mutant de départ. En effet, nous avons déduit précédemment que l'interaction hydrophobe entre la chaîne latérale de l'Ala-336 de l'enzyme et l'histidine du peptide devait nécessairement diminuer la capacité de PR-Set7 à reconnaître son substrat. Bien que la proposition du triple mutant semble résoudre ce problème, cette solution fut tout à fait fortuite. En effet, nous n'avons pas de raison de penser que l'éloignement de la lysine fût dépendant du

problème de positionnement de cette histidine. Nous avons donc cherché à supprimer la source de l'interaction en mutant virtuellement la tyrosine 336 en une glycine plutôt qu'en une alanine. Ainsi nous avons généré un nouveau double mutant : PR-Set7 Y336G/Y334F et avons repris les études de docking comme vu précédemment.

La superposition des deux doubles mutants PR-Set7 Y336A/Y334F et Y336G/Y334F révèle en premier lieu que l'interaction hydrophobe est bien inhibée en présence d'une glycine en position 336 (Figure 68). La conséquence directe en est le placement de l'histidine 18 de façon à former la liaison H avec le ribose du Hey-SAM. Là encore, de façon inattendue nous observons la réorientation de l'alcyne de l'analogue de SAM et un placement idéal de la lysine à alkyler (distance $N\epsilon-C\alpha = 3,7 \text{ \AA}$). Dans ces conditions, il est aussi attendu de ce mutant qu'il puisse prendre en charge Hey-SAM ce qui validerait notre hypothèse.

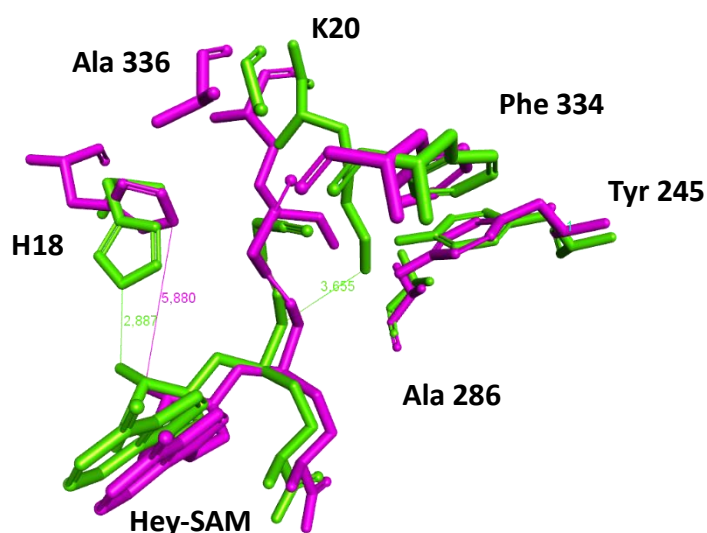


Figure 68. Superposition des poses de plus basse énergie résultant du docking de Hey-SAM dans PR-Set7 Y336A/Y334F (en violet) ou dans le double mutant Y336G/Y334F (en vert).

Finalement, la superposition des deux nouveaux mutants montre que les propositions issues de ces deux hypothèses aboutissent à un résultat très similaire (Figure 69). Le fait que les problématiques de positionnement de l'histidine, nécessaire à une bonne affinité de l'enzyme pour son substrat, et de placement de la lysine à alkyler soient connectées reflète bien l'exiguïté du site catalytique de PR-Set7 et est à la source des difficultés de conception de mutants actifs envers notre série d'analogues de SAM.

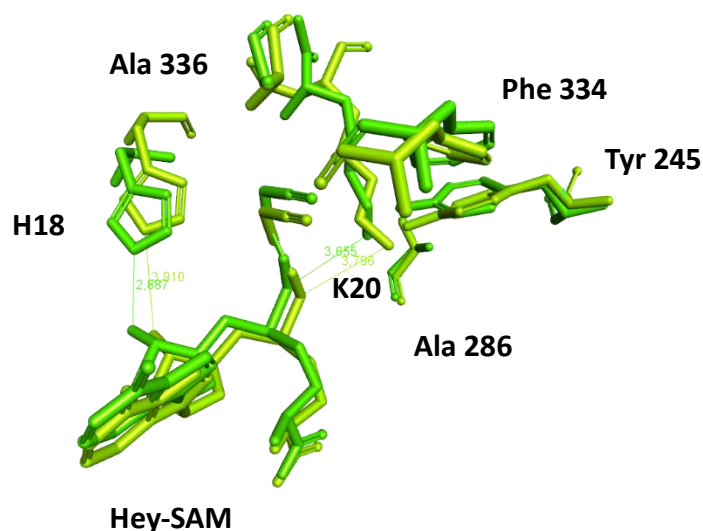


Figure 69. Superposition des poses de plus basse énergie résultant du docking de Hey-SAM dans PR-Set7 Y336G/Y334F (en vert) ou dans le triple mutant Y336A/Y334F/A286G (en jaune).

V.6.2 Evaluation de l'activité des nouveaux mutants

Après, avoir conçu théoriquement les deux nouveaux mutants, qui, nous l'espérons, prendront en charge des analogues de SAM plus volumineux, nous avons décidé de les produire et de les purifier. Etant donné que nous allons créer un triple mutant, nous avons considéré le problème de l'accumulation de mutations qui pourrait éventuellement altérer la spécificité de substrat de l'enzyme. Nous avons donc produit deux autres mutants ponctuels de l'enzyme afin de vérifier l'impact de la mutation en position 286 de PR-Set7. Ces 2 mutants sont : le mutant A286G et le double mutant Y334F/A286G. En comparant notre enzyme à d'autres HKMT on peut prédire l'impact de chacune des mutations sur l'effet catalytique de l'enzyme.

Dans le cas du double mutant Y336G/Y334F et du triple mutant Y336A/Y334F/A286G, étant donné l'absence de la tyrosine 336 essentielle à la réaction, nous ne nous attendons pas à observer de méthylation. Nous nous attendons ensuite à restaurer une activité catalytique mais seulement envers les analogues de SAM pourvus d'une double liaison stabilisatrice.

Le résidu alanine en position 286 est conservé chez les HKMT, toutefois, dans certaines enzymes, il peut être muté en une glycine. C'est le cas chez DIM-5, une tri-méthyltransférase de H3K9 chez *Neurospora crassa* mais aussi chez un mutant naturel de EZH2 dans lequel la mutation équivalente à notre A286G est A677G. Cette marque est associée à de l'hyper-triméthylation de H3K27, son substrat principal ¹⁸⁶

Nous nous attendons donc par analogie avec des enzymes proches que la mutation A286G soit gain de fonction. De plus la mutation Y334F étant aussi gain de fonction, le nouveau double mutant Y334F/A286G devrait donc l'être aussi.

Enfin, étant donnée la petite taille de ProSeAM, nous nous attendons à ce que tous les mutants capables de méthylation soient aussi capables de propargylation, le mécanisme stabilisateur de l'état de transition étant conservé.

Après purification, le test d'activité des 4 nouveaux lots d'enzymes a donc été effectué en incubant ces enzymes avec de la SAM radioactive en présence de nucléosome reconstitué. La Figure 70 est équivalente à la précédente figure 32 de la partie 2 dans laquelle nous testions l'activité de Y336A, Y334A, Y334F et Y336A/Y334F. Nous constatons ainsi que PR-Set7 A286G et Y334F/A286G sont plus actifs que l'enzyme sauvage, conformément à nos attentes. Le double mutant Y336G/Y334F et le triple mutant sont quant à eux perte de fonction ce qui était aussi attendu.

Cette étude de docking alimentée par les données expérimentales obtenues dans les Chapitres 2 et 3, a donc permis d'obtenir de nouveaux résultats expérimentaux très prometteurs qui nous laissent penser que cette étude structurale est pertinente.

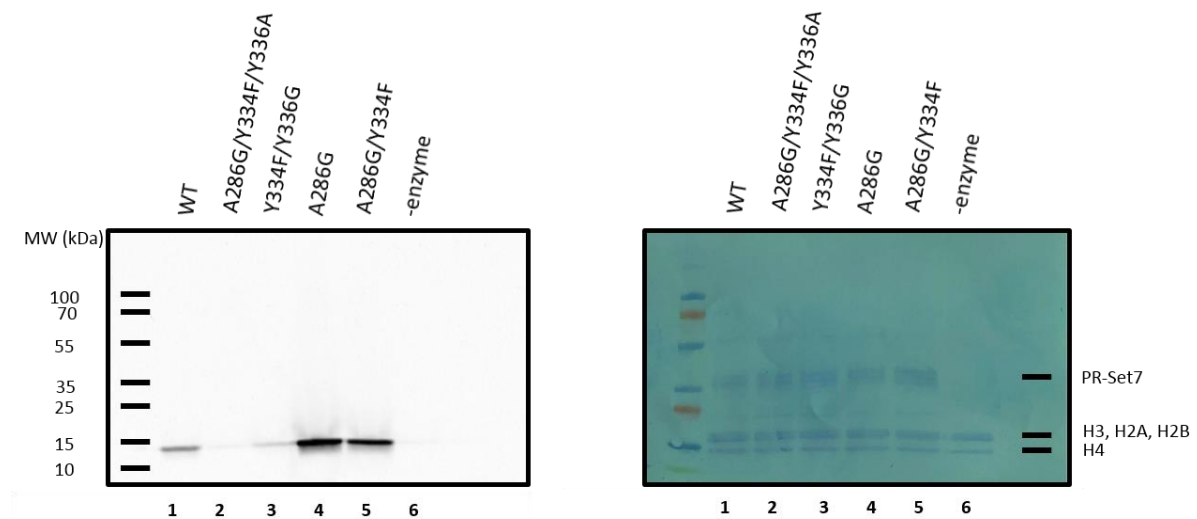


Figure 70. Analyse sur gel SDS PAGE de tests enzymatiques des nouveaux mutants de PR-Set7 en présence de SAM tritiée au niveau du méthyle. Le gel est révélé par autoradiographie (à gauche) ; Le contrôle de charge coloré au bleu de Coomassie (à droite). Pour rappel, parmi les précédents mutants, le mutant Y334F méthylait l'histone H4 dans les mêmes proportions que l'enzyme sauvage.

V.7 Conclusion

En tenant compte de nos résultats expérimentaux et de l'état des connaissances sur les différentes HKMT à domaine SET, nous avons mené une étude de modélisation moléculaire afin de comprendre pourquoi l'analogue ProSeAM était pris en charge par certains lots d'enzyme et pas d'autre. Nous cherchions aussi à déterminer pourquoi le double mutant Y334F/Y336A de PR-Set7, spécialement conçu par analogie avec l'enzyme GLP mutée, était incapable de prendre en charge de gros analogues de cofacteur.

Notre étude a permis de mettre en lumière un ensemble de caractéristiques structurales de PR-Set7 influençant directement l'activité de l'enzyme et nous avons ainsi pu rationaliser les résultats observés au cours de ce travail. Les résultats que nous en avons tirés nous ont permis de concevoir de nouveaux mutants de l'enzyme susceptibles de prendre en charge nos analogues de SAM volumineux. Pour le moment, les nouveaux mutants de PR-Set7 se comportent comme le prédit la théorie vis-à-vis de l'activité de méthylation. Ils seront testés prochainement afin de vérifier leur activité avec les analogues de SAM.

Ainsi, le docking a été ici utile non seulement comme un outil servant à la rationalisation de nos résultats mais aussi comme outil puissant pour réaliser des prédictions. En effet, contrairement à l'alignement structural de structures aux rayons X complètement figés, qui nous ont servi à concevoir les premiers mutants, il permet de travailler sur un système plus dynamique et d'en tirer des informations assez subtiles sur des différences importantes entre les sites actifs de cette famille d'enzyme dont les structures sont pourtant proches.

Enfin, rappelons que c'est parce que nous avons collecté un certain nombre de données expérimentales que ce modèle a pu faire de nouvelles prédictions. Il aurait été difficile de prévoir ceci avant de réaliser les expériences. Le docking est ainsi particulièrement intéressant car il est à la jonction entre les travaux expérimentaux faits et ceux à venir sur la base de nouvelles prédictions. Ajoutons pour finir que nous nous attendons aussi à ce que ce modèle puisse prédire d'autres types d'analogues de SAM qui pourraient rentrer dans le site de fixations de mutants d'enzymes.

Conclusion générale et perspectives

Conclusion générale et perspectives

Au cours de cette thèse nous nous sommes donc appliqués à rechercher de nouveaux substrats de l'histone lysine méthyltransférase PR-Set7. Jusque-là plusieurs méthodes permettaient de caractériser les substrats non-histones des HKMT mais avaient l'inconvénient de ne pas refléter des conditions très physiologiques ou ne permettaient d'identifier que peu de nouveaux substrats. Dans ce contexte, nous avons cherché à adapter une stratégie initialement mise en place par l'équipe de Minkui Luo pour identifier les substrats d'une autre HKMT : GLP. En effet, dans un article décrivant une preuve de concept, cette équipe décrit l'utilisation d'un mutant ponctuel de GLP capable d'accommoder un analogue de la *S*-adénosyl-L-méthionine porteur d'un groupement bioorthogonal. Ce mutant d'enzyme et cet analogue du cofacteur naturel de l'enzyme leur a ainsi permis de mettre en place une stratégie de marquage chémoenzymatique en deux étapes permettant *in fine* l'identification des substrats de l'HKMT hyperactive GLP.

Nous avons cherché à transposer cette stratégie sur une enzyme de la même famille (l'HKMT à domaine SET), connue pour son activité de monométhyltransférase de la lysine 20 de l'histone H4 (H4K20) : PR-Set7, dont l'activité dérégulée est associée à différentes pathologies telles que des cancers. Il a été montré par différentes études que la délétion de l'enzyme entraînait une létalité précoce au cours du développement embryonnaire chez les mammifères ne pouvant être expliquée par la seule absence de méthylation sur H4K20. Dans ce contexte, il semble évident qu'il existe des substrats non-histone de PR-Set7 qu'il est crucial de caractériser afin de mieux comprendre le rôle biologique de cette enzyme.

Ainsi dans le premier chapitre présentant les résultats de ma thèse, le Chapitre 2, j'ai présenté la façon dont nous avons conçu les différents outils nécessaires à la stratégie de marquage chémoenzymatique. La synthèse de 6 analogues de SAM porteurs pour la majorité d'une double liaison en position β du sulfonium et d'un groupement bioorthogonal, a été réalisée avec l'aide du Dr. Karine Guitot. Parallèlement, sur la base d'une grande similarité de séquence entre les HKMT à domaine SET, 4 mutants ponctuels de l'enzyme PR-Set7 (Y336A, Y334A, Y334F et Y336A/Y334F) ont été conçus puis produits dans un système bactérien. De plus afin de disposer d'un contrôle positif, l'enzyme GLP et le mutant ponctuel décrit dans la littérature comme capable de prendre en charge les analogues de SAM ont aussi été produits. Enfin, différents

complexes contenant les histones entières ont aussi été produits de façon à effectuer les tests enzymatiques dans des conditions plus physiologiques que sur des substrats minimaux de natures peptidiques. Les différents mutants ont été testés avec de la SAM tritiée sur du nucléosome reconstitué afin d'évaluer leur activité et nous avons pu montrer que seul, le mutant gain de fonction Y334F possédait une activité de transfert de méthyle.

Munis de nos outils, nous avons réalisé dans le Chapitre 3, un criblage de l'ensemble des analogues de SAM envers les différentes enzymes sauvage ou mutées. Afin de disposer d'un contrôle positif dans cette stratégie complexe, nous avons mis au point les protocoles sur l'enzyme GLP dont l'activité enzymatique robuste permettait de s'assurer de la validité des protocoles. Ainsi, le criblage sur un substrat peptidique contenant la lysine 9 de l'histone H3, son substrat, nous a permis d'observer le transfert des chaînes alkyles d'analogues de SAM possédant une double liaison en position β du sulfonium permettant ainsi un marquage de la lysine 9 avec la fonctionnalité bioorthogonale alcyne ou azide, alors que l'enzyme sauvage en est incapable. Nous avons également pu mettre en place et valider les conditions d'analyse par spectrométrie de masse et western blot suite à la réaction bioorthogonale de type CuAAC ou SPAAC avec une sonde complémentaire comportant une biotine ou un agent fluorescent que ce soit sur peptide ou en utilisant les complexes d'histone.

Nous avons ensuite mis en œuvre des expériences équivalentes avec l'enzyme PR-Set7 et ses mutants. Malheureusement, aucune prise en charge de ces analogues de SAM n'a pu être observée. Etant donné qu'il nous était difficile de savoir si l'absence de signal était due à une trop faible activité de l'enzyme ou à un trop fort bruit de fond nous avons tenté différentes stratégies d'optimisation des protocoles. Ces stratégies, ne nous ayant pas permis d'obtenir davantage de résultats, nous avons conclu que le site catalytique de PR-Set7 était vraisemblablement trop contraint pour accommoder ces analogues de SAM. Un analogue de SAM plus petit a alors été synthétisé : le ProSeAM. Cet analogue a la particularité d'avoir un atome de sélénium à la place du soufre, ce qui lui confère une stabilité supérieure à son homologue soufré vis-à-vis d'une voie de dégradation ; il est aussi plus réactif, la liaison Sélénium-Carbone étant plus faible que la liaison Soufre-Carbone, un grand avantage étant donné que PR-Set7 a une efficacité catalytique *in vitro* bien moins robuste que GLP. Le criblage de ProSeAM envers PR-Set7 et ses mutants nous a permis de mettre en évidence que seules l'enzyme WT et son mutant gain de fonction Y334F pouvaient le prendre en charge. Munis de

mutants actifs et non-actifs de PR-Set7 nous avons pu procéder à l'étude protéomique de PR-Set7.

Dans le Chapitre 4, nous avons pu vérifier que le mutant Y334F propargyle exclusivement la lysine 20 de l'histone H4 validant ainsi que la spécificité de position de fonctionnalisation par l'enzyme est conservée. Par la suite, nous avons vérifié que nos outils permettent l'enrichissement de l'histone H4 à partir de mélanges complexes (extraits nucléaires de cellules HEK293T). Passées ces vérifications nécessaires, la recherche de nouveaux substrats de PR-Set7 a pu être mise en oeuvre grâce à une expérience de protéomique où chacune des conditions est menée en quintuplicat, assurant une grande fiabilité des résultats.

Cette étude a abouti à l'identification de 275 potentiels substrats de PR-Set7. Notons que nous avons anticipé la prise en charge de ProSeAM par des enzymes présentes dans les extraits de cellules et que nos contrôles permettent d'attribuer exclusivement à PR-Set7, les substrats que nous trouvons enrichis. Une sélection des meilleurs candidats a ensuite été effectuée sur la base des taux d'enrichissement réduisant cette liste de 275 protéines à 23 candidats dont la validation biologique a été initiée. Sur ces 23 candidats, nous avons pu en produire 14. Cependant aucune méthylation par PR-Set7 n'a été observée à ce jour sur ces candidats dans nos conditions *in vitro*. Bien que décevants ces résultats restent encore à explorer. En effet, notre rapide tentative de validation de ces substrats ne pouvait pas explorer toutes les conditions expérimentales favorables à une bonne activité de PR-Set7 alors que nous nous attendons à ce que l'enzyme ait besoin d'acides nucléiques, de partenaires ou d'un substrat particulièrement bien conformé.

En l'absence de substrats validés, une analyse plus approfondie des résultats de la protéomique a révélé, pour certaines des protéines enrichies, une surreprésentation de certains processus biologiques importants et connus comme étant régulés par des HKMT autres que PR-Set7. Enfin, certains candidats possèdent dans leur séquence primaire une suite de résidus très proche du site consensus de PR-Set7. Ces protéines n'ont pas encore été produites et pourraient constituer des pistes intéressantes de substrats potentiels qu'il faudra essayer de valider prochainement.

Enfin, dans le Chapitre 5, nous avons cherché une méthode permettant de rationaliser nos résultats expérimentaux obtenus avec PR-Set7, GLP et leurs mutants par une étude de docking. L'interprétation de cette étude de modélisation moléculaire nous a permis de conclure quant aux paramètres structuraux essentiels autorisant ou interdisant la prise en charge de nos

différents analogues de SAM, malgré une contrainte stérique élevée à l'intérieur du site catalytique de PR-Set7. En particulier, nous avons compris que l'absence de prise en charge de gros analogues de SAM par le double mutant Y336A/334F qui nous semblait le plus prometteur n'était probablement pas dû à une impossibilité d'accommodation de la molécule dans le site actif de l'enzyme mais par une impossibilité de l'enzyme à reconnaître son substrat. De plus, des hypothèses émises à partir des sites catalytiques d'autres HKMT de la même famille (GLP, PRC2 et DIM-5) ont permis la conception de nouveaux mutants de PR-Set7 devant être capables de prendre en charge les analogues de SAM les plus volumineux de notre série d'analogues. A ce jour, l'activité de méthylation de ces nouveaux mutants ponctuels a été évaluée et les résultats du test sont parfaitement cohérents avec les prédictions faites par le docking et l'analyse bibliographique. L'activité de ces mutants envers les analogues de SAM de grande taille reste à évaluer et cette étude sera réalisée prochainement. Par ailleurs, une prise de recul sur les résultats du travail de modélisation moléculaire permet de mettre en évidence que les mécanismes de fonctionnalisation des méthyltransférases à domaine SET semblent très similaires d'une HKMT à une autre. En effet, la transposition directe de mutations de résidus conservés du domaine SET d'une HKMT quelconque vers le domaine SET de PR-Set7 a souvent permis d'observer des conséquences similaires sur les réseaux d'interaction.

Enfin, parmi les conclusions générales de cette étude de docking nous envisageons 3 cas distincts de prise en charge des analogues de SAM selon leur taille. Les analogues de petites tailles, dont ProSeAM est le représentant, seraient utilisés par les HKMT de la même façon que la SAM avec un état de transition de la réaction de transalkylation qui passerait par l'établissement d'une liaison H avec la Tyr-336 (ou résidu équivalent chez une autre HKMT). Des analogues un peu plus gros comme le Pentenyl-SAM pourraient aussi être pris en charge de cette façon à condition que l'alanine 286 soit mutée en une glycine (mutation existante chez l'HKMT DIM-5 naturellement et PRC2 dans certaines cellules cancéreuses). Dans le cas où la Tyr-336 serait mutée, la chaîne alkyle de l'analogue de SAM effectuerait une rotation de 180° par rapport au premier cas de façon à occuper l'espace libéré par l'absence de la chaîne latérale de la tyrosine. Pentenyl-SAM disposerait donc aussi de deux conformations possibles à l'intérieur du site catalytique des HKMT compatible avec la catalyse. Enfin, le dernier cas de prise en charge des analogues de SAM est celui des analogues de tailles importantes, dont le Hey-SAM est un représentant. De par leur taille, leur accommodation dans le site catalytique est conditionnée par l'absence de la Tyr-336 et de l'Ala-286.

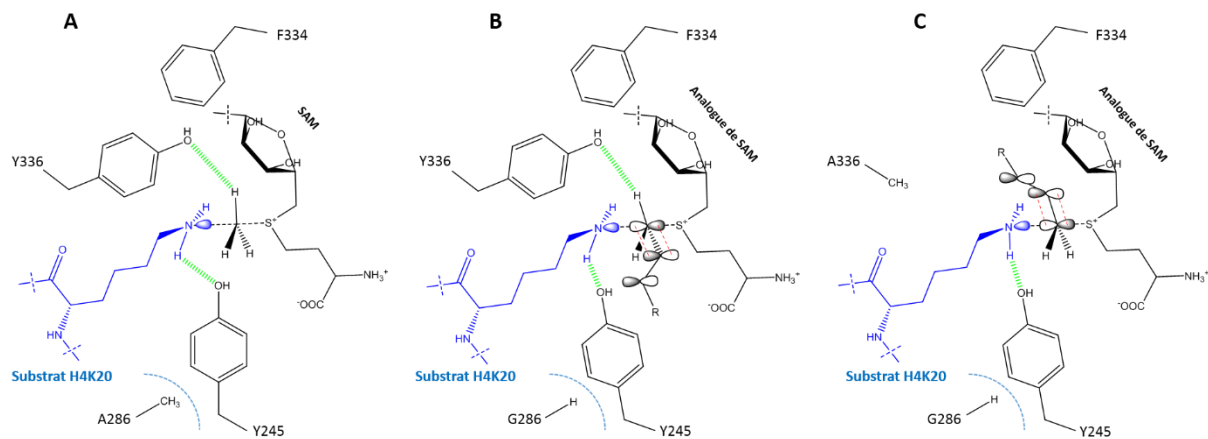


Figure 71. Proposition des conformations et de modes de stabilisations des états de transition de différentes catégories d’analogues de SAM. A. cas de la SAM et de petits analogues de SAM stabilisés par l’établissement d’une liaison H de type C-H----O. B. Cas d’analogues de SAM de taille intermédiaire, la mutation en position 286 induisant un basculement de la Tyr-245 laissant davantage de place à une chaîne alkyle de taille modérée. C. Cas de gros analogues de SAM dont la chaîne alkyle ne peut rentrer dans le site catalytique que si la Tyr-336 est mutée. Dans cette position, la stabilisation de l’état de transition passe par un recouvrement orbitalaire.

En tenant compte de ces conclusions et du fait que PR-Set7 figure probablement parmi les HKMT dont les contraintes stériques à l’intérieur du site actif sont les plus importantes (Tyr-334 au lieu de Phe-334), on comprend pourquoi le ProSeAM a été le seul fonctionnel dans notre étude. Cet analogue constitue à la fois un atout mais aussi un inconvénient majeur. Un atout en premier lieu, l’accès au méthylome devenant simple puisqu’il devient facultatif de concevoir des mutants gain de fonction prenant en charge le cofacteur, les enzymes sauvages, le faisant sans difficulté. Mais il s’agit aussi d’un inconvénient puisque la large utilisation de ProSeAM par les HKMT d’un mélange complexe, comme des extraits de cellule, induit un bruit de fond puissant qui diminue les ratios d’enrichissement déterminés par la protéomique et par conséquent diminue le nombre de substrats d’HKMT caractérisés.

Au cours de ma thèse, j’ai aussi eu l’occasion de travailler quelques temps sur une troisième HKMT, EZH2 qui est la sous unité catalytique du complexe PRC2 (Polycomb Repressive Complex 2). Ce complexe est capable de tri-méthyler la lysine 27 de l’histone H3 et est impliqué dans le maintien de la répression de transcription et nous cherchions à identifier, comme pour PR-Set7, ses substrats non-histones. Toutefois la méconnaissance du mode de fonctionnement de l’enzyme, le fait qu’il s’agisse d’un complexe multi-protéique difficile à produire et sa faible activité enzymatique nous ont poussés à travailler sur le système plus « simple » qu’est PR-Set7. C’est néanmoins grâce à nos travaux sur cette enzyme que nous avons eu l’idée de la mutation supplémentaire à introduire dans le site catalytique de PR-Set7

qui était probablement la dernière barrière à lever pour une prise en charge d'analogues de SAM à longues chaînes alkyles. Les données obtenues grâce à notre étude vont maintenant pouvoir être transposées sur EZH2 pour laquelle nous disposons déjà de mutants appropriés.

Enfin, nous envisageons, d'étendre notre stratégie à une échelle plus physiologique en synthétisant des analogues de SAM capable de passer la membrane ou bien en les faisant synthétiser via la machinerie cellulaire. Combiné à des lignées cellulaires capables d'exprimer le mutant d'enzyme prenant en charge l'analogue de cofacteur, la stratégie permettrait un marquage des cibles d'HKMT dans les mêmes conditions et dans les mêmes taux qu'à lieu la méthylation des lysines.

Depuis une dizaine d'années, la conception d'outils chimiques pour l'étude des modifications post-traductionnelles des protéines permet d'apporter de nouvelles stratégies complémentaires aux méthodes plus traditionnelles pour appréhender la complexité de ces marques et de leur rôle. Ainsi, les avancées dans le domaine de la ligation chimique native, la fonctionnalisation sélective de protéines ont, par exemple, permis d'obtenir des informations importantes sur le rôle de certaines PTM. L'utilisation de marqueurs d'affinité analogues de substrat ont également permis d'aider à caractériser des enzymes impliquées dans ces PTM.

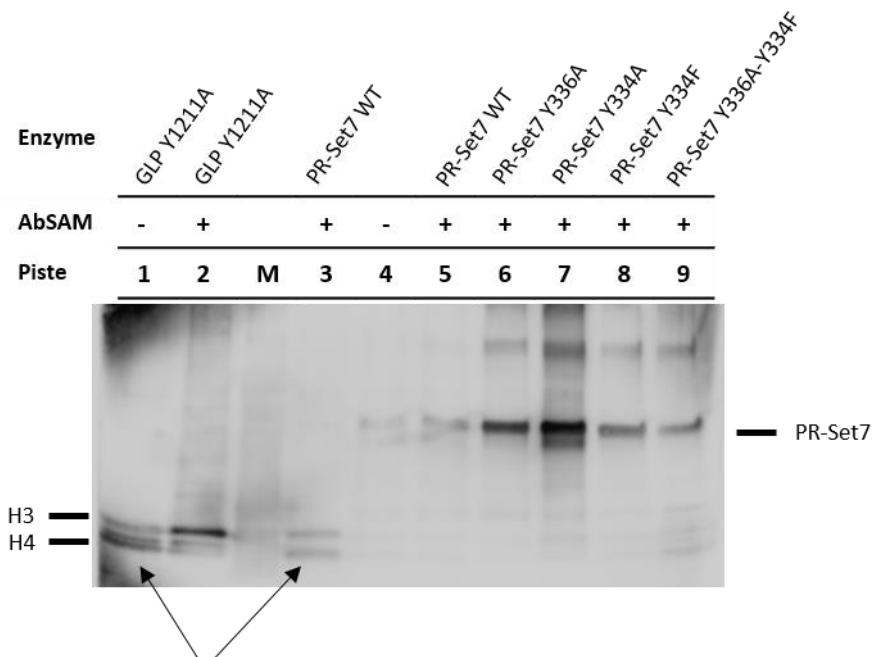
A terme, ces méthodes participeront sûrement à décoder le « code non-histone », qui, à l'instar, du code histone cherchant à établir des liens de causalité entre les marques épigénétique et leur impact sur l'état chromatinien, relie les modifications post-traductionnelles introduites sur les protéines à leur conséquence biologique (régulation d'une voie métabolique ou d'une voie de signalisation particulière).

Le travail effectué au cours de cette thèse est finalement un pas de plus vers la caractérisation de la spécificité de substrat de PR-Set7 et plus largement vers la compréhension du rôle de la méthylation des lysines, une des marques épigénétiques les plus communes.

Annexes

Annexes

Annexe 1



Annexe S1. Analyse sur gel SDS-PAGE de la fonctionnalisation des histones par l'analogue Ab-SAM par Western Blot. Les flèches noires révèlent des signaux dans des pistes où le cofacteur est absent donc où théoriquement, il ne devrait pas y avoir de signal.

Annexe S2 : Alignement structural de nombreuses HKMT à domaine SET dont PR-Set7, GLP, G9a, DIM-5 et EZH2 dont nous discutons les liens tout au long de cette thèse.

CLUSTAL O(1.2.4) multiple sequence alignment

```

AtKYP      446  FNLEVFR-SAKKGVAVRSWEYIPAGSPVCEYIGVVRRTADVDTI-----SDNEYIFE
HsSETDB1  803  VRLQLFK-TQNKGVIRCLDDIAKGSFVCIYAGKILTDDFADKEGLEM-----GDEYFAN
NcDIM-5   162  VPLQIFR-TKDRGVKCPVNIKRGQFVDRYLGEIITSEEADRRRAESTIARRKDVYLF
DmASH1    1389 APGVERFMTADKGVVVRTKLPVIAKGTIILEYVGEVVTEKEFKQRM-ASIYLNDRTHHYCLH
AtSUVH6   616  LPLEIFK-TKSRGVVRLKSIPIGSFICEYVGELELSEAEERRI-----GNDEYLF
SpCLR4    328  LPLEIFK-TKEKGVVRSRFRFAPAGTFITCYLGEVITSAEAAKRDK--NYDDDGITVLF
HsGLP     1095  ARLQLYR-TRDMGVVRSRQDIPPGTFVCEYVGELELSDSEADVR-----EEDSYLF
HsG9a     1038  VRLQLYR-TAKMGVVRALQTIPOGTFICEYVGELELSDAEADVR-----EDDSYLF
HsSUV39H1 343  YDLCIFRTDDGRGVVRLTLEKIRKNSFVMEYVGEIITSEEAEERRGQ--IYDRQGATVLF
HsSUV39H2 250  YSLCIFRTSNGRCGVKTLVKIKRMSFVMEYVGEVITSEEAEERRGQ--FYDNKGITVLF
HsPr-Set7 216  EGMKID-LIDGKGRVIATKQFSRQDFVVEYHGDLEITDAKKREALYAQDPSTGQYMY
ScSET2    120  APIAIF-KTKHKGYGVRAEQDIEANQFIYEYKGEVIEEMEFDRDLIDYDQRHFHFYFMM
HsEZH2    612  KHLLLA-PSDVACWIFIKDPVQKNEFISEYCGEIIISQDEADRGRKVYDKYM--CSELFN
HsMLL1    3829  EAVGVY-RSPIHGRGLFCKRNIDAGEMVIEYAGNVIRSITQTDKREKYYSKIGIG-CYMF
ScSET1    938  KPVMFA-RSAIHNWGLYALDSIAAKEMIEYVGERIRQVPAEMREKRYLKNIGISSYLF
      : . . . : * * : :
  
```

```

AtKYP      IDCQQTMOQLGGRQRRRLRDVAVPMNNGVSSQSEEDENAP-----EFCIDAGST
HsSETDB1   LDHIESVENFKEGYESDAPCSSDSSGVLDKQEDGNSGTEDPEESN (304) CYI IDAKLE
NcDIM-5    LDKFSDPDSLDP-----LAGQPLEVDGEYM
DmASH1     LH-----LDG-----GLVIDGQRM
AtSUVH6    IGNRYDNSLAQGMSELMLGTQAGRSMAEG-----DESSGFTIDAASK
SpCLR4     LDMF-----DDASEYTVDAQNY
HsGLP      LDNK-----DGEVYCIDARFY
HsG9a      LDNK-----DGEVYCIDARYY
HsSUV39H1  LDYV-----EDVYTVDAAYY
HsSUV39H2  LDYE-----SDEFTVDAARY
HsPr-Set7  FQY-----LSKTYCVDATRE
ScSET2     -----LQNGEFIDATIK
HsEZH2     -----LNNDFFVDATRK
HsMLL1     -----IDDSEVVDATMK
ScSET1     -----VDENTVIDATKK
      : * .
  
```

```

      Pseudo-noeud
      /-----\
     /         \
    /           \
   /             \
  /               \
 /                 \
/                   \
G-NFARFINHSCENLFLVQCVLSSHQDIRLARVVLFAADNISPMQELTYDYGYALDSV
HsSETDB1  G-NLGRYLNHSCSPNLFVQNVFVDTHDLRFPWVAFFASKRIRAGTELTDWYNYEVS-
NcDIM-5   S-GPTRFINHSCDPNMAIFARVGDHADKHIHDLALFAIKDIPKGTELTDYVNGLTGL
DmASH1    G-SDCRFVNHSCENCEMQKWSVN-GLSRM--V-LFAKRAIEEGEELTYDYNFSLFNP
AtSUVH6   G-NVGRFINHSCSPNLYAQNVLYDHEDSRIPHVMFFAQDNIPPLQELTYDYNALDQV
SpCLR4    G-DVSRFFNHSCSPNIAIYSAVRNHGFRTIYDLAFAIKDIQPLEELTDYAGAKDFS
HsGLP     G-NVSRFINHSCENLVPVRVFMHQDLRFPRIAFFSTRLEAGEQLGFDYGERFWDI
HsG9a     G-NISRFINHLCDPNIIIPVRVFMHLQDLRFPRIAFFSSRDIRTGEEIGFDYGDRFWDI
HsSUV39H1 G-NISHFVNHSCDPNLQVYNVFDIDNLDLRLPRIAFFATRTIRAGEELTDYDNMQVDPV
HsSUV39H2 G-NVSHFVNHSCDPNLQVFNVFDIDNLDTRLPRIALFSTRINAGEELTDYQMKGSGD
HsPr-Set7 TNRLGRLINHSKCGNCQTKLHDIDG----VPHLILIASRDIAAGEELTYDYGDRSKAS
ScSET2    G-SLARFCNHSCSPNAYVNKVVVDK----KLRMGIFAQRKILKGEELTYDYNVDRYGA
HsEZH2    G-NKIRFANHSVNPNCYAKVMVNG----DHRIGIFAKRAIQTGEELTDYRYSQADA
HsMLL1    G-NAARFINHSCENPCYSRVINIDG----QKHIVIFAMRKIYRGEELTYDYKFPEDA
ScSET1    G-GIARFINHSCDPNCTAKIKVGG----RRRIVIYALRDIAAGEELTYDYKFEREKD
      : ** * : : * : : **
  
```

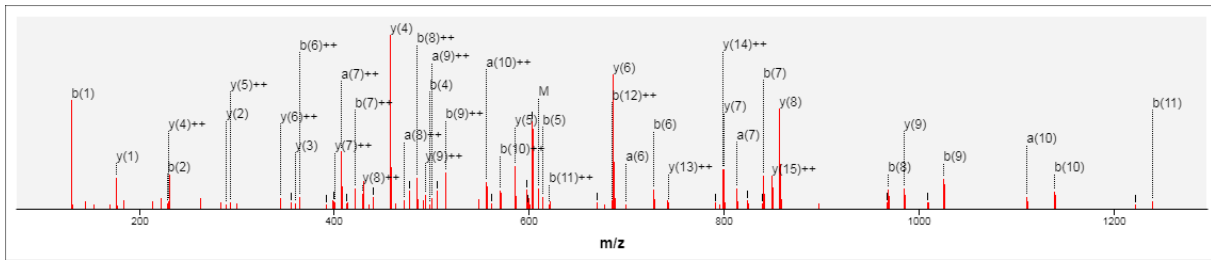
: Substitution conservative (aa de meme groupe et score >0)

* Identité de sequence = correspondance

. Substitution semi-conservative (aa de même groupe)

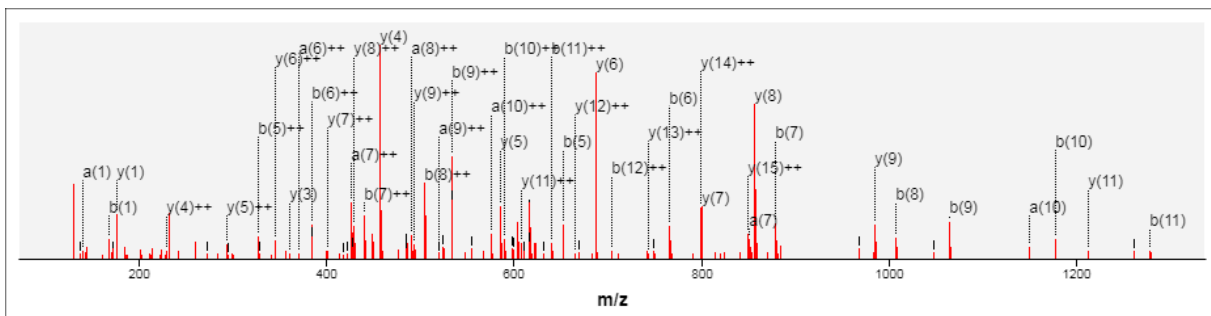
Annexe S3

Caractérisation MS/MS du peptide K(20)VLRDNIQGITKPAIR(35) propargylé



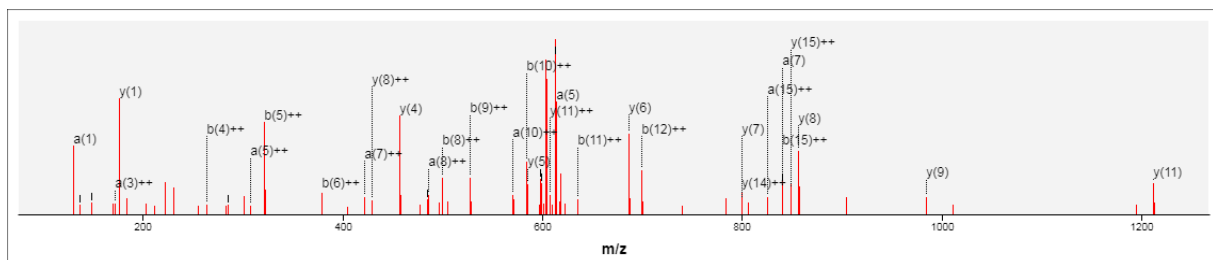
#	aminoacid	#	y ⁺	y ²⁺	y ³⁺	y ^{*+}	y ^{*2+}	y ^{*3+}	y ^{o+}	y ^{o2+}	y ^{o3+}
1	K	16	1822.10	911.56	608.04	1805.08	903.04	602.36	1804.09	902.55	602.04
2	V	15	1694.01	847.51	565.34	1676.98	838.99	559.67	1676.00	838.50	559.34
3	L	14	1594.94	797.97	532.32	1577.91	789.46	526.64	1576.93	788.97	526.31
4	R	13	1481.86	741.43	494.62	1464.83	732.92	488.95	1463.84	732.43	488.62
5	D	12	1325.75	663.38	442.59	1308.73	654.87	436.91	1307.74	654.38	436.59
6	N	11	1210.73	605.87	404.25	1193.70	597.35	398.57	1192.72	596.86	398.24
7	I	10	1096.68	548.85	366.23	1079.66	540.33	360.56	1078.67	539.84	360.23
8	Q	9	983.60	492.30	328.54	966.57	483.79	322.86	965.59	483.30	322.54
9	G	8	855.54	428.27	285.85	838.52	419.76	280.18	837.53	419.27	279.85
10	I	7	798.52	399.76	266.85	781.49	391.25	261.17	780.51	390.76	260.84
11	T	6	685.44	343.22	229.15	668.41	334.71	223.48	667.43	334.22	223.15
12	K	5	584.39	292.70	195.47	567.36	284.18	189.79			
13	P	4	456.29	228.65	152.77	439.27	220.14	147.09			
14	A	3	359.24	180.12	120.42	342.21	171.61	114.74			
15	I	2	288.20	144.61	96.74	271.18	136.09	91.06			
16	R	1	175.12	88.06	59.05	158.09	79.55	53.37			

Caractérisation MS/MS du peptide K(20)VLRDNIQGITKPAIR(35) propargylé



#	aminoacid	#	y ⁺	y ²⁺	y ³⁺	y ^{*+}	y ^{*2+}	y ^{*3+}	y ^{o+}	y ^{o2+}	y ^{o3+}
1	K*	16	1860.12	930.56	620.71	1843.09	922.05	615.04	1842.11	921.56	614.71
2	V	15	1694.01	847.51	565.34	1676.98	838.99	559.67	1676.00	838.50	559.34
3	L	14	1594.94	797.97	532.32	1577.91	789.46	526.64	1576.93	788.97	526.31
4	R	13	1481.86	741.43	494.62	1464.83	732.92	488.95	1463.84	732.43	488.62
5	D	12	1325.75	663.38	442.59	1308.73	654.87	436.91	1307.74	654.38	436.59
6	N	11	1210.73	605.87	404.25	1193.70	597.35	398.57	1192.72	596.86	398.24
7	I	10	1096.68	548.85	366.23	1079.66	540.33	360.56	1078.67	539.84	360.23
8	Q	9	983.60	492.30	328.54	966.57	483.79	322.86	965.59	483.30	322.54
9	G	8	855.54	428.27	285.85	838.52	419.76	280.18	837.53	419.27	279.85
10	I	7	798.52	399.76	266.85	781.49	391.25	261.17	780.51	390.76	260.84
11	T	6	685.44	343.22	229.15	668.41	334.71	223.48	667.43	334.22	223.15
12	K	5	584.39	292.70	195.47	567.36	284.18	189.79			
13	P	4	456.29	228.65	152.77	439.27	220.14	147.09			
14	A	3	359.24	180.12	120.42	342.21	171.61	114.74			
15	I	2	288.20	144.61	96.74	271.18	136.09	91.06			
16	R	1	175.12	88.06	59.05	158.09	79.55	53.37			

Caractérisation MS/MS du peptide K(20)VLRDNIQGITKPAIR(35) diméthylé



#	aminoacid	#	y ⁺	y ²⁺	y ³⁺	y ^{*+}	y ^{*2+}	y ^{*3+}	y ^{o+}	y ^{o2+}	y ^{o3+}
1	K*	16	1850.13	925.57	617.38	1833.11	917.06	611.71	1832.12	916.57	611.38
2	V	15	1694.01	847.51	565.34	1676.98	838.99	559.67	1676.00	838.50	559.34
3	L	14	1594.94	797.97	532.32	1577.91	789.46	526.64	1576.93	788.97	526.31
4	R	13	1481.86	741.43	494.62	1464.83	732.92	488.95	1463.84	732.43	488.62
5	D	12	1325.75	663.38	442.59	1308.73	654.87	436.91	1307.74	654.38	436.59
6	N	11	1210.73	605.87	404.25	1193.70	597.35	398.57	1192.72	596.86	398.24
7	I	10	1096.68	548.85	366.23	1079.66	540.33	360.56	1078.67	539.84	360.23
8	Q	9	983.60	492.30	328.54	966.57	483.79	322.86	965.59	483.30	322.54
9	G	8	855.54	428.27	285.85	838.52	419.76	280.18	837.53	419.27	279.85
10	I	7	798.52	399.76	266.85	781.49	391.25	261.17	780.51	390.76	260.84
11	T	6	685.44	343.22	229.15	668.41	334.71	223.48	667.43	334.22	223.15
12	K	5	584.39	292.70	195.47	567.36	284.18	189.79			
13	P	4	456.29	228.65	152.77	439.27	220.14	147.09			
14	A	3	359.24	180.12	120.42	342.21	171.61	114.74			
15	I	2	288.20	144.61	96.74	271.18	136.09	91.06			
16	R	1	175.12	88.06	59.05	158.09	79.55	53.37			

liste D de protéines enrichies en présence de PR-Set7 Y334F (N = 169 protéines)
(en vert sont surlignées les protéines évoquées dans le corps du manuscrit)

#	Identifiant	Name in myProMS	Mol. weight (Da)	Matching peptides in best analysis	Coverage (%) (best Analysis)
1	Q9NQR1	SETD8	42889.9	87	79,6
2	P18583	SON	263830	33	15,3
3	O60437	PPL	204747	29	24,3
4	P04350	TUBB4A	49585.8	28	50,5
5	P06748	NPM1	32575	25	56,1
6	P22061	PCMT1	24636.4	20	71,4
7	P62805	HIST1H4A	11367.3	20	62,1
8	P07476	IVL	68478.5	19	37,3
9	Q07666	KHDRBS1	48227.4	19	38,6
10	P27635	RPL10	24603.9	18	53,7
11	P62753	RPS6	28680.7	17	31,7
12	P62910	RPL32	15859.8	17	53,3
13	P62917	RPL8	28024.7	17	54,5
14	P62280	RPS11	18430.7	16	46,2
15	P62937	PPIA	18012.5	16	50,9
16	P46778	RPL21	18564.8	15	60,0
17	P62241	RPS8	24205.2	15	41,3
18	P62249	RPS16	16445.3	15	56,2
19	Q14103	HNRNPD	38434.2	15	31,0
20	Q15366	PCBP2	38580.1	15	45,2
21	P07910	HNRNPC	33670	13	35,3
22	P15121	AKR1B1	35853.4	13	51,6
23	Q05519	SRSF11	53542.2	13	28,5
24	P09211	GSTP1	23355.8	12	62,4
25	P55795	HNRNPH2	49263.6	12	29,8
26	P62081	RPS7	22126.9	12	61,9
27	P41227	NAA10	26458.5	11	60,0
28	P62277	RPS13	17222.3	11	46,4
29	P62829	RPL23	14865.5	11	42,1
30	P00492	HPRT1	24579.4	10	45,0
31	P39019	RPS19	16060.5	10	62,8
32	Q15181	PPA1	32660.1	10	44,3
33	Q9BY44	EIF2A	64990.1	10	25,6
34	O75494	SRSF10	31300.6	9	32,4
35	P62263	RPS14	16272.7	9	41,7
36	P84103	SRSF3	19329.6	9	44,5
37	Q96FW1	OTUB1	31284.1	9	32,5
38	Q9BV38	WDR18	47405.3	9	19,9
39	O75223	GGCT	21007.7	8	30,3
40	P25325	MPST	33178.5	8	32,7
41	P61020	RAB5B	23706.8	8	42,3
42	P63220	RPS21	9111.38	8	60,2
43	P68036	UBE2L3	17861.6	8	62,3
44	P98179	RBM3	17170.4	8	63,7

45	Q86V81	ALYREF	26887.9	8	47,1
46	Q8IWS0	PHF6	41290.5	8	13,4
47	Q92922	SMARCC1	122867	8	11,2
48	Q9H2U2	PPA2	37920.2	8	35,0
49	Q9Y3U8	RPL36	12253.7	8	25,7
50	P29372	MPG	32868.7	7	29,9
51	P60660	MYL6	16930.1	7	51,0
52	P60866	RPS20	13372.7	7	42,9
53	P62304	SNRPE	10803.6	7	44,6
54	Q9H7D7	WDR26	72124	7	15,7
55	Q9Y2S7	POLDIP2	42033.3	7	32,3
56	O43148	RNMT	54844.1	6	17,6
57	O95336	PGLS	27546.9	6	34,1
58	P20618	PSMB1	26489.4	6	32,8
59	P22570	FDXR	53836.8	6	13,6
60	P42677	RPS27	9461.13	6	32,1
61	P51580	TPMT	28180.3	6	29,0
62	P52564	MAP2K6	37492.4	6	19,2
63	P61586	RHOA	21768.1	6	18,1
64	P62266	RPS23	15807.6	6	30,8
65	P62857	RPS28	7841.03	6	44,9
66	Q13243	SRSF5	31263.8	6	29,0
67	Q9BQ61	C19orf43	18419.4	6	56,8
68	Q9NQ29	LUC7L	43728	6	17,5
69	Q9NQE9	HINT3	20361.3	6	42,9
70	Q9Y3Y2	CHTOP	26396.6	6	18,1
71	P00390	GSR	56257	5	12,8
72	P35222	CTNNB1	85496.6	5	6,3
73	P61604	HSPE1	10931.7	5	62,7
74	P62854	RPS26	13015.4	5	33,9
75	Q02241	KIF23	110059	5	8,3
76	Q14011	CIRBP	18648	5	46,5
77	Q71UM5	RPS27L	9477.17	5	32,1
78	Q9BT78	COPS4	46268.8	5	17,5
79	Q9NX58	LYAR	43614.8	5	15,3
80	Q9NZT2	OGFR	73324.7	5	10,9
81	Q9Y3B9	RRP15	31484.2	5	18,8
82	Q9Y3C1	NOP16	21188.4	5	32,0
83	O15347	HMGB3	22980	4	28,0
84	O60613	SELENOF	17941.8	4	30,9
85	O60828	PQBP1	30472.3	4	31,3
86	O75683	SURF6	41450.4	4	20,2
87	P04183	TK1	25468.7	4	22,2
88	P14174	MIF	12476.3	4	40,9
89	P30419	NMT1	56806.2	4	10,3
90	P49458	SRP9	10111.7	4	32,6
91	P62070	RRAS2	23399.6	4	21,6
92	P63000	RAC1	21450.1	4	20,8
93	Q00577	PURA	34910.8	4	19,6
94	Q52LJ0	FAM98B	37190.9	4	17,9
95	Q9GZP4	PITHD1	24177.9	4	20,4
96	Q9UBF8	PI4KB	91378.9	4	4,7

97	Q9UKA9	PTBP2	57490.8	4	13,7
98	Q9UKD2	MRTO4	27560.5	4	21,3
99	Q9Y5P6	GMPPB	39834.4	4	16,4
100	O00743	PPP6C	35144.3	3	11,5
101	O75306	NDUFS2	52545.7	3	9,9
102	P08134	RHOC	22006.4	3	15,0
103	P17028	ZNF24	42155.4	3	13,6
104	P17612	PRKACA	40589.7	3	10,0
105	P30044	PRDX5	22086.4	3	25,7
106	P42575	CASP2	50685.3	3	11,3
107	Q14353	GAMT	26318.2	3	16,1
108	Q14571	ITPR2	308064	3	1,4
109	Q86TB9	PATL1	86850.1	3	5,8
110	Q86UY6	NAA40	27193.9	3	15,2
111	Q86VM9	ZC3H18	106379	3	5,8
112	Q8N5P1	ZC3H8	33575.8	3	13,1
113	Q8TBZ6	TRMT10A	39718.8	3	13,9
114	Q8WX92	NELFB	65697.4	3	7,8
115	Q92576	PHF3	229481	3	3,0
116	Q96EP5	DAZAP1	43383.4	3	15,2
117	Q9BQP7	MGME1	39421	3	17,4
118	Q9H3K6	BOLA2	10116.5	3	39,5
119	Q9NP72	RAB18	22977.1	3	14,6
120	Q9NRX4	PHPT1	13832.5	3	36,8
121	Q9UHV9	PFDN2	16647.8	3	29,9
122	Q9UHY7	ENOPH1	28932.7	3	18,0
123	Q9Y316	MEMO1	33733.2	3	11,4
124	O43169	CYB5B	16332.2	2	32,2
125	O75330	HMMR	84100.5	2	5,7
126	Q94832	MYO1D	116202	2	3,1
127	P16949	STMN1	17302.5	2	12,8
128	P17096	HMGA1	11676	2	23,4
129	P31260	HOXA10	42414.2	2	4,6
130	P50750	CDK9	42777.5	2	6,7
131	P62841	RPS15	17040.1	2	27,6
132	Q14331	FRG1	29172.4	2	9,7
133	Q14657	LAGE3	14803.8	2	23,8
134	Q1RMZ1	BMT2	46323.9	2	11,6
135	Q53H96	PYCR3	28663.3	2	8,8
136	Q6ZSJ8	C1orf122	11470.8	2	33,6
137	Q8IZ69	TRMT2A	68725.6	2	5,1
138	Q8TB61	SLC35B2	47514.8	2	5,1
139	Q96B36	AKT1S1	27383.3	2	10,5
140	Q9BPW8	NIPSNAP1	33310	2	8,8
141	Q9BRT6	LLPH	15225.2	2	29,5
142	Q9GZT3	SLIRP	12349	2	20,2
143	Q9NP92	MRPS30	50364.7	2	7,3
144	Q9NS18	GLRX2	18051.7	2	21,3
145	Q9NVH0	EXD2	70352.7	2	4,5
146	Q9NYR9	NKIRAS2	21508.5	2	12,6
147	Q9P2X3	IMPACT	36476.5	2	13,4
148	Q9UBV8	PEF1	30380.9	2	8,5

149	Q9UGP4	LIMD1	72190.2	2	4,1
150	Q9Y2S6	TMA7	7066.3	2	20,3
151	Q9Y3D3	MRPS16	15344.8	2	19,0
152	O75909	CCNK	64239.8	1	4,1
153	P02458	COL2A1	141785	1	0,7
154	P05386	RPLP1	11513.9	1	14,0
155	P61513	RPL37A	10275.3	1	19,6
156	Q05513	PRKCZ	67660	1	2,4
157	Q13613	MTMR1	74677.8	1	1,1
158	Q14019	COTL1	15945	1	11,3
159	Q5VYS4	MEDAG	34190.1	1	2,3
160	Q7Z6I8	C5orf24	20131.8	1	5,3
161	Q8IZU8	DSEL	139238	1	0,7
162	Q96BK5	PINX1	37034.7	1	4,9
163	Q96RN5	MED15	86753.5	1	1,4
164	Q9NPA5	ZFP64	74643.5	1	1,5
165	Q9NUU6	FAM105A	42195.9	1	1,7
166	Q9NY74	ETAA1	103440	1	1,6
167	Q9P003	CNIH4	16093.4	1	14,4
168	Q9UII2	ATP5IF1	12248.8	1	10,4
169	Q9Y6V0	PCLO	560700	1	0,2

liste E de protéines uniques en présence de PR-Set7 Y334F (N = 106 protéines)
(en vert sont surlignées les protéines évoquées dans le corps du manuscrit)

#	Identifiant	Name in myProMS	Mol. weight (Da)	Matching peptides in best analysis	Coverage (%) (best Analysis)
1	Q9H089	LSG1	75225.4	5	8,5
2	Q9UK61	FAM208A	189032	5	3,7
3	Q8TF01	PNISR	92576.8	4	6,1
4	O95248	SBF1	208443	3	3,1
5	Q8TE68	EPS8L1	80251.1	3	5,7
6	Q14777	NDC80	73912.7	2	4,0
7	O94919	ENDOD1	55016.6	2	4,4
8	P09601	HMOX1	32818.6	2	13,5
9	P49711	CTCF	82785.4	2	2,8
10	P52952	NKX2-5	34917.7	2	9,3
11	Q15075	EEA1	162466	2	1,7
12	Q5JSH3	WDR44	101366	2	6,4
13	Q7Z569	BRAP	67304.6	2	3,2
14	Q8N4N8	KIF2B	76253.8	2	3,9
15	Q8N983	MRPL43	23430.8	2	8,4
16	Q8NEF9	SRFBP1	48633.9	2	7,0
17	Q92793	CREBBP	265351	2	0,7
18	Q96DE5	ANAPC16	11667	2	26,4
19	Q9BQE4	SELENOS	21013	2	16,9
20	Q9H1K1	ISCU	17999	2	11,4
21	Q9UKD1	GMEB2	56421.3	2	2,8
22	Q9Y324	FCF1	23369.6	2	12,1
23	Q9Y3C6	PPIL1	18236.8	2	16,9
24	A0A2R8YEI5	CHN2	62440.3	1	1,5
25	A4D1B5	GSAP	97801.5	1	0,8
26	A6NCM1	IQCA1L	95061.2	1	1,1
27	A6NE01	FAM186A	262758	1	0,5
28	O14737	PDCD5	14285.1	1	10,4
29	O15457	MSH4	104756	1	2,7
30	O43184	ADAM12	99542.2	1	0,9
31	O75054	IGSF3	135196	1	0,6
32	O75935	DCTN3	21119.4	1	4,3
33	O94842	TOX4	66194.8	1	1,4
34	O94903	PROSC	30343.9	1	4,7
35	O95817	BAG3	61594.7	1	2,1
36	O95900	TRUB2	36694.5	1	3,0
37	O95999	BCL10	26251.6	1	9,0
38	P02774	GC	52917.6	1	1,7
39	P04798	CYP1A1	58165.3	1	1,6
40	P18627	LAG3	57448.6	1	1,1
41	P21757	MSR1	49762.1	1	1,8
42	P36543	ATP6V1E1	26145.3	1	6,2
43	P42568	MLLT3	63351.4	1	3,2
44	P49771	FLT3LG	26415.8	1	3,4

45	P55017	SLC12A3	113139	1	1,7
46	P62487	POLR2G	19294.3	1	6,4
47	P78395	PRAME	57889.7	1	1,2
48	P82663	MRPS25	20116.4	1	6,9
49	P83436	COG7	86343.9	1	1,6
50	Q02318	CYP27A1	60234.8	1	1,9
51	Q03188	CENPC	106834	1	1,0
52	Q14201	BTG3	29116.4	1	4,0
53	Q14919	DRAP1	22349.8	1	5,9
54	Q15072	ZNF146	33307.8	1	4,1
55	Q15078	CDK5R1	34060.2	1	2,3
56	Q15599	SLC9A3R2	37413.8	1	3,9
57	Q2M385	MPEG1	78587.1	1	0,8
58	Q2WGN9	GAB4	62367.5	1	1,2
59	Q53F19	C17orf85	70592.6	1	2,1
60	Q5T2D3	OTUD3	45124.1	1	2,8
61	Q5TC79	ZBTB37	56054.9	1	2,4
62	Q5VU43	PDE4DIP	265103	1	0,7
63	Q6EBC2	IL31	18204.9	1	6,1
64	Q7Z5J4	RAI1	203352	1	0,4
65	Q7Z7A1	CNTRL	268886	1	0,5
66	Q86UU5	GGN	66698.5	1	1,5
67	Q8IXB1	DNAJC10	91079.7	1	1,4
68	Q8IYH5	ZZZ3	102023	1	1,1
69	Q8IZ21	PHACTR4	78211.2	1	1,9
70	Q8IZT6	ASPM	409800	1	0,3
71	Q8N2G8	GHDC	57523	1	2,8
72	Q8N5N7	MRPL50	18325	1	12,0
73	Q8N8Y5	ZFP41	22806	1	5,1
74	Q8NCN4	RNF169	77194.3	1	2,3
75	Q8NDF8	TENT4B	63266.9	1	2,8
76	Q8TAA9	VANGL1	59975	1	4,8
77	Q8WVC0	LEO1	75404.1	1	2,1
78	Q8WVD3	RNF138	28192.6	1	3,7
79	Q92187	ST8SIA4	41295	1	3,6
80	Q92858	ATOH1	38160.4	1	3,1
81	Q96B21	TMEM45B	31826.2	1	3,3
82	Q96DA2	RAB39B	24622.2	1	7,5
83	Q96FF9	CDCA5	27600.6	1	6,0
84	Q96HU8	DIRAS2	22484.9	1	5,5
85	Q96MX6	WDR92	39740.5	1	2,5
86	Q96NC0	ZMAT2	23611.9	1	4,5
87	Q96SF7	TBX15	65757.5	1	2,5
88	Q9H0A6	RNF32	41516.2	1	1,9
89	Q9H0G5	NSRP1	66390.1	1	2,3
90	Q9H4A5	GOLPH3L	32767.4	1	3,2
91	Q9H6A9	PCNX3	222039	1	1,1
92	Q9H6F5	CCDC86	40235.6	1	3,1
93	Q9H8U3	ZFAND3	25184.1	1	5,7
94	Q9HCN8	SDF2L1	23598.3	1	3,2
95	Q9NRX2	MRPL17	20050.2	1	8,6
96	Q9NWL6	ASNSD1	72079.9	1	3,1

97	Q9NZM1	MYOF	234709	1	0,5
98	Q9NZS9	BFAR	52738	1	1,6
99	Q9P031	CCDC59	28669.8	1	4,1
100	Q9P2F5	STOX2	102671	1	1,4
101	Q9UK45	LSM7	11602.4	1	8,7
102	Q9ULD4	BRPF3	135745	1	0,6
103	Q9ULL5	PRR12	129991	1	2,2
104	Q9Y2Q3	GSTK1	25496.9	1	5,3
105	Q9Y4A8	NFE2L3	76154.2	1	2,4
106	Q9Y6H3	ATP23	28080.6	1	9,3

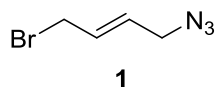
Matériel et Méthodes

Matériel et méthodes

Solvents and most of the starting materials were purchased from Sigma Aldrich, TCI and Carlo Erba. All solvents were of analytical grade. ^1H and ^{13}C NMR spectra were measured with a Bruker Avance 300 (300 MHz) and Avance 400 (360 MHz). Chemical shifts (δ) are given in parts per million (ppm) relative to D_2O (4.79 ppm) and CHCl_3 (7.24 ppm). The following abbreviations are used: s (singlet), d (doublet), dd (doublet of doublet), t (triplet), m (multiplet). Mass spectra were obtained with an Esquire LC Bruker spectrometer. Optical rotations were measured on an Anton Paar MCP 150 Polarimeter (c in g/100 mL). TLC were performed on Merck precoated silica gel 60F plates. Merck silica gel 60 (230-400 mesh) was used for column chromatography. Analytical reverse phase high performance liquid chromatography (RP-HPLC) analyses were performed with a Dionex UltiMate 3000 machine on a Nucleodur HTec C18 column (250 mm x 3 mm); particles size: 5 μm with UV detection at 260 nm. Unless otherwise stated, a linear gradient of ACN (0 to 10 %) in 0.1 % aqueous TFA in 20 minutes (flow rate of 1 $\text{mL}\cdot\text{min}^{-1}$) was used for analysis. Semi Preparative RP-HPLC was performed on a JASCO LC-1500 system equipped with a Nucleodur HTec C18 column (250 mm x 10 mm); particles size: 5 μm and UV detection at 260 nm. A linear gradient of ACN (0 to 10 %) in 0.1 % aqueous TFA in 20 minutes (flow rate of 4 $\text{mL}\cdot\text{min}^{-1}$) was used for purifications. UV spectra were recorded on a Nanodrop-2000c spectrophotometer (Thermo Scientist). High Resolution Mass Spectra (HRMS) were recorded at the Institut Parisien de Chimie Moléculaire, Sorbonne Université (electrospray source) and at Institut de Chimie Moléculaire et des Matériaux d'Orsay. MALDI-TOF mass spectra of peptides were recorded at the plateforme de spectrométrie de masse (IBPS, Sorbonne Université) on a ABI Voyager DE-Pro or on a 4700 Proteomix MALDI-TOF MS/MS mass spectrometer (Applied Biosystems) and at the Institut de biologie intégrative de la cellule on a Shimadzu Axima Performance time-of-flight mass spectrometer system in the positive-ion reflector mode. α -Cyano-4-hydroxycinnamic acid (CHCA) was used as a matrix. It was dispersed at 10 mg/mL in a $\text{H}_2\text{O}/\text{ACN}$ (1/1) mixture containing 0.1 % TFA. External calibration was performed using the pepmix 4 peptide mixture as calibrants (proteomix, laser Biolabs).

1. SAM analogues synthesis

(*E*)-1-azido-4-bromobut-2-ene¹³⁵

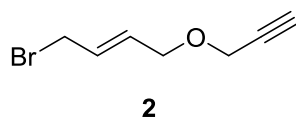


The (*E*)-1,4-dibromobut-2-ene (1 g, 4.7 mmol) was diluted in CH₂Cl₂ (20 mL), followed by the addition of nBu₄NBr (756 mg, 2.35 mmol) and NaN₃ (300 mg, 4.7 mmol). A saturated solution of NaHCO₃ (20 mL) was added and the mixture was stirred for 2 h. The organic phase was then washed with water (20 mL), and brine (20 mL) and dried with Na₂SO₄. The solvent was evaporated (P=350 mBar, T_{bath}=40°C) and the obtained crude product was purified by flash chromatography (petroleum ether). 234 mg of product **1** was obtained as a transparent liquid (28 % yield).

R_f = 0.18 in petroleum ether

¹H-NMR (400 MHz, CDCl₃) δ: 6.07-5.95 (m, 1H), 5.93-5.81 (m, 1H), 3.99 (d, *J* = 7.6 Hz, 2H), 3.84 (d, *J* = 6.0 Hz, 2H).

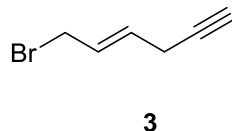
(*E*)-1-bromo-4-(prop-2-ynoxy)but-2-ene¹⁴⁹



Propargyl alcohol (760 μL, 13 mmol) was added dropwise to a suspension of NaH (300 mg, 13 mmol) in anhydrous THF at 0°C. The solution was stirred for 30 minutes at 0°C and (*E*)-1,4-dibromobut-2-ene (3 g, 1.1 eq) previously diluted in 13 mL of dry THF was added via a cannula. The reaction was heated at 50°C overnight. The reaction was quenched with a mixture of diethyl ether and HCl 2 M. The solution was extracted with diethyl ether and the organic phase was washed with water, brine and dried over Na₂SO₄. The solvent was removed under reduced pressure and the crude product was purified by flash chromatography (cyclohexane/EtOAc, 6/1). Compound **2** was isolated as a colorless oil (32 % yield).

¹H-NMR (400 MHz, CDCl₃) δ (ppm): 6.03-5.95 (m, 1H), 5.93-5.81 (m, 1H), 4.16 (d, *J* = 2.48 Hz, 2H), 4.10 (d, *J* = 7.40 Hz, 2H), 3.96 (d, *J* = 7.68 Hz, 2H), 2.44 (t, *J* = 2.40 Hz, 1H).

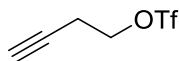
(*E*)-6-Bromo-hex-4-en-1-yne¹⁴⁷



(*E*)-1,4-Dibromobut-2-ene (2 g, 9.35 mmol), and CuCl (60 mg, 1 eq), were degassed under argon and diluted with 4 mL of anhydrous THF. 15 mL of ethynylmagnesium bromide 0.5 M in THF was then added dropwise over 1 h. The reaction was stirred for 1 h at 60°C, and left overnight at room temperature. The reaction was quenched with 20 mL of 5 % HCl. The product was then extracted with DCM, the organic phase was washed with water and brine and dried over Na₂SO₄. After filtration, the solvent was evaporated. The crude mixture containing (*E*)-6-bromo-hex-4-en-1-yne **3** (20 %, calculated from the NMR spectrum) and dialkylated product was used directly without further purification in the next reaction.

¹H-NMR (400 MHz, CDCl₃) δ (ppm): 6.05-5.99 (m, 1H), 5.78-5.92 (m, 1H), 3.96-3.98 (m, 2H), 2.96-3.01 (m, 2H), 2.16 (t, 1H, *J* = 2.95 Hz).

But-3-yn-1-yl trifluoromethanesulfonate¹⁸⁷

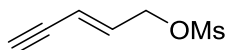


4

But-3-yn-1-ol (350 mg, 5 mmol) was diluted in anhydrous CH₂Cl₂ (35 mL) and dry pyridine (0.4 mL, 1 eq) under argon. Trifluoromethanesulfonic anhydride (1.1 mL, 1.3 eq) was then slowly added at -15°C. The mixture was stirred for 2 h at 0°C followed by 1 h at room temperature. The reaction was quenched with water, extracted with CH₂Cl₂ and the organic phase dried over Na₂SO₄. After concentration under vacuum, the crude product containing **4** (80% calculated from the NMR spectrum) was directly used in the subsequent reaction.

¹H-NMR (400 MHz, CDCl₃) δ (ppm): 4.60 (t, *J* = 7.0 Hz, 2H, H₁), 2.76 (dt, *J* = 7.0 Hz, ⁴*J* = 2.4 Hz, 2H, H₂), 2.14 (t, ⁴*J* = 2.4 Hz, 1H, H₄)

(*E*)-Pent-2-en-4-yn-1-methanesulfonate¹²³



5

NaOH (290 mg, 7.25 mmol) was added and stirred in anhydrous CH₂Cl₂ (8 mL). Mesyl chloride was then added (0.5 mL, 6.45 mmol) to the solution. At 0°C, distilled (*E*)-pent-2-en-4-yn-1-ol (500 mg, 6.1 mmol) was then slowly added and the mixture was allowed to warm to room temperature. After 2 h stirring, the reaction was quenched with a saturated solution of NaHCO₃ (15 mL). Extraction with CH₂Cl₂ was performed and the combined organic phases were dried over Na₂SO₄. The solvent was removed under vacuum (P=650 mBar, T_{bath}=40°C). The crude mixture containing (*E*)-pent-2-en-4-yn-1-oximesyle **5** (37 %, calculated from the NMR spectrum) and unreacted starting material was used directly without further purification in the next reaction.

R_f = 0.6 in CH₂Cl₂

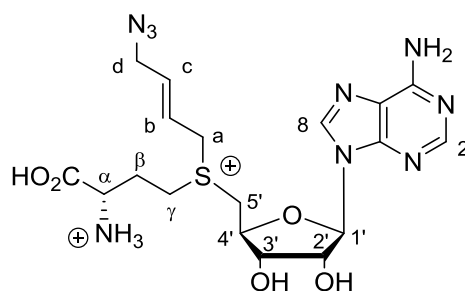
¹H-NMR (400 MHz, CDCl₃) δ: 6.20 (dt, *J* = 15.92 Hz, *J* = 6.12 Hz, 1H), 5.84 (m, *J* = 15.88 Hz, 1H), 4.75 (dd, *J* = 6.16 Hz, *J* = 1.48 Hz, 2H), 3.03 (s, 3H), 3.01 (d, *J* = 2.72 Hz, 1H).

General procedure for the synthesis of SAM and SeAM analogues

SeAH (15 mg, 34 μmol) or SAH (10 mg, 26 μmol) was dissolved into a mixture of 1:1 formic acid and acetic acid (1 mL) and cooled on ice. The alkyl bromide derivative (50 eq, 1.7 mmol) was then slowly added. After addition of AgClO₄ (7 mg, 1 eq), in the case of alkyl bromide, the reaction was allowed to stand at room temperature for 6 to 8 h before being quenched with 5 mL of distilled water containing 0.1 % TFA (v/v). The aqueous phase was washed three times with diethyl ether and passed through 0.2 μm syringe filter. The crude product was purified by semi-preparative RP-HPLC at a flow rate of 4 mL/min with a gradient of ACN in water containing 0.1 % of TFA. Depending on the compounds, the diastereoisomers were separated or collected together. After lyophilization, the purity of the products was evaluated by analytical HPLC using a gradient of 0 to 10 % ACN using a C18 column, 5 μm; 100x4.6 mm and the analogues were characterized by mass spectrometry and ¹H-NMR when quantities were sufficient. The concentration of SAM analogues was determined by its UV absorption (ε₂₆₀ =

15400 M⁻¹.cm⁻¹) and gave the final yields. Into the resultant concentrated solution, a small amount of TFA was added to adjust pH to ~2 and the stock solution was stored at -20 °C.

Ab-SAM



5

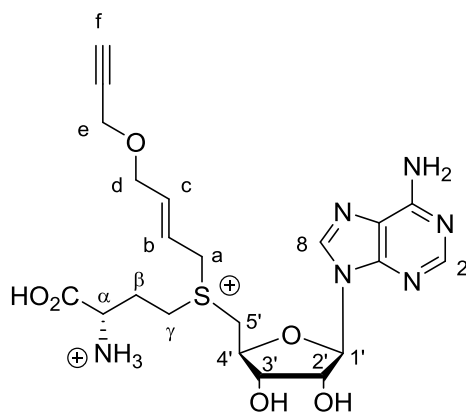
The compound **5** was synthesized as described previously¹⁸⁸, using (*E*)-1-azido-4-bromobut-2-ene **1** and *S*-adenosyl-L-homocysteine. The compound was purified by HPLC. The collected fractions were lyophilized to give **5** as a white powder (12 % yield).

¹H-NMR (400 MHz, D₂O) δ (ppm): 8.51 (m, 2H, H₂, H₈), 6.22 (m, 1H, H_{1'}), 6.20-6.15 (m, 0.5H, H_b), 6.07-6.00 (m, 0.5H, H_b), 5.91-5.83 (m, 0.5H, H_c), 5.79-5.71 (m, 0.5H, H_c), 4.79 (m, 1H, H_{2'}), 4.71-4.67 (m, 1H, H_{3'}), 4.59 (m, 1H, H_{4'}), 4.25 (d, *J* = 8 Hz, 1H, H_a), 4.22 (d, *J* = 8 Hz, 1H, H_a), 3.99-3.83 (m, 5H, H_α, H_{5'}, H_d), 3.68-3.47 (m, 2H, H_γ), 2.42-2.36 (m, 2H, H_β).

Rt = 11 min (gradient 2-10% ACN in 10 min).

MS (EI) *m/z* = 480.18 [M]⁺, 379.13 [5'-(1-azido-but-2-ensulfonio)-5'-deoxyadenosine + H]⁺

Pob-SAM



6

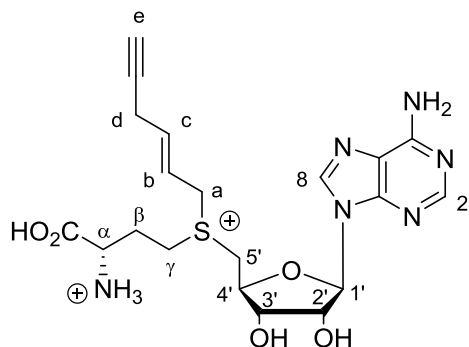
The compound **6** was synthesized as described previously¹⁴⁸, using (*E*)-1-bromo-4-(prop-2-ynyloxy)but-2-ene **2** and *S*-adenosyl-L-homocysteine. The compound was purified by HPLC. The collected fractions were lyophilized to give **6** as a white powder (3 % yield).

¹H-NMR (400 MHz, D₂O + 0.1 % TFA-d) δ (ppm): 8.49-8.50 (m, 2H, arom), 6.20-6.22 (m, 1H, H_{1'}), 6.15-6.18 (m, *J* = 15.6 Hz, 0.5H, H_b), 5.99-6.05 (dt, *J* = 5.2 Hz, 1H, H_b), 5.80-5.88 (m, 0.5H, H_c), 5.70-5.77 (m, 0.5H, H_c), 4.85-4.87 (m, 1H, H_{2'}), 4.66-4.71 (m, 1H, H_{3'}), 4.55-4.61 (m, 1H, H_{4'}), 4.19-4.25 (m, 6H, H_d, H_e, H_a), 4.10 (d, *J* = 6.41 Hz, 1H, H_{4'}), 3.88-3.95 (m, 3H, H_α, H_{5'}), 3.59-3.67 (m, 1H, H_γ), 3.47-3.55 (m, 1H, H_γ), 2.94 (t, 1H, H_f), 2.36-2.42 (m, 2H, H_β).

Rt = 9 min (gradient: 0-10 % ACN in 0.1 % aqueous TFA in 15 min).

MS (ESI, positive, *m/z*): 493.51[M]⁺, 392.13 [5'-(4-propargyloxy-but-2-enyl)-*S*-adenosine+H]⁺

Hey-SAM



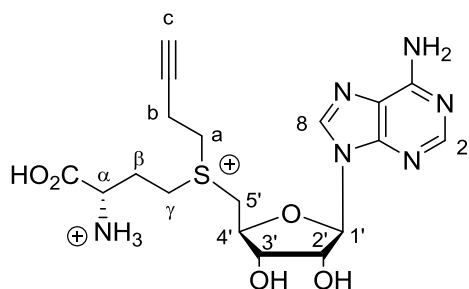
7

The compound **7** was synthesized as described previously¹⁵⁴, using (*E*)-6-bromo-hex-4-en-1-yne **3** and *S*-adenosyl-L-homocysteine. The compound was purified by HPLC. The collected fractions were lyophilized to give **7** as a white powder (2 % yield).

Rt = 14 min (gradient: 0-10 % in 0.1 % aqueous TFA in 15 min).

MALDI TOF MS, $m/z = 463.68$ [M]⁺

Butynyl-SAM



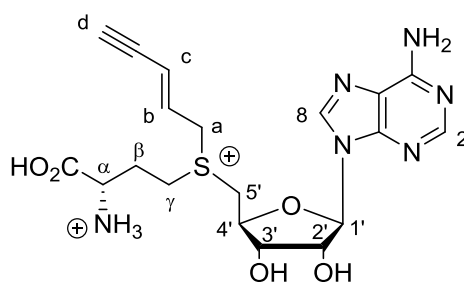
8

The compound was synthesized as described previously¹⁵⁴, using activated butynyl alcohol **4** and *S*-adenosyl homocysteine. The compound was purified by HPLC. The collected fractions were lyophilized to give **8** as a white powder (3 % yield).

RMN ¹H (400 MHz, D₂O) δ (ppm): 8.48-8.50 (m, 2H, H₂,H₈), 6.21 (s, 1H, H_{1'}), 4.96 (m, 0.5H), 4.88 (m, 0.5H), 4.63-4.69(m, 1H, H_{3'}), 4.58-4.65 (m, 1H, H_{4'}), 4.06-4.10 (m, 2H, H_{5'}), 3.92-3.98 (m, 1H, H _{α}), 3.66-3.80 (m, 4H, H _{a} , H _{γ}), 2.80-2.87 (m, 2H, H _{b}), 2.47-2.53 (m,1H, H _{c}), 2.38-2.48 (m, 2H, H _{β}).

Rt = 2.9 min (gradient: 2-10 % ACN in 0.1 % aqueous TFA in 10 min, then 100 % A in 5 min).

Pentenyl-SAM



9

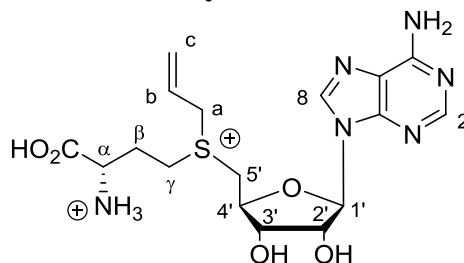
The compound **9** was synthesized as described previously¹²³, using activated pentenyl alcohol **5** and *S*-adenosyl homocysteine. The compound was purified by HPLC. The collected fractions were lyophilized to give **9** as a white powder (5 % yield).

¹H-RMN (400 MHz, D₂O) δ : 8.45 (d, 1H, H_{arom}), 8.43 (s, 1H, H_{arom}), 6.15 (d, $J = 2.92$ Hz, 1H, H_{1'}), 6.15-6.11 (m, 0.5H, H_b), 6.05 (dt, $J = 7.60$ Hz, $J = 15.60$ Hz, 0.5H, H_b), 5.89 (d, $J = 16.84$ Hz, 0.5H, H_c), 5.70 (d, $J = 16$ Hz, 0.5H, H_c), 4.65 (t, $J = 5.8$ Hz, 1H, H_{3'}), 4.53 (t, $^3J = 5.64$ Hz, 1H, H_{4'}), 4.22 (d, $J = 8$ Hz, 1H, H_a), 4.17 (d, $J = 8$ Hz, 1H, H_a), 3.93-3.89 (m, 1H, H _{α}), 3.86-3.80 (m, 2H, H_{5'}), 3.59-3.46 (m, 3H, H _{γ} , H_d), 2.36 (m, 2H, H _{β}).

MS (ESI, positive, m/z): 449.16034 [M]⁺, 348.11 [5'-(pent-2-enyl)thio-5'-deoxyadenosine + H]⁺

Rt = 2.25 min (gradient: 0-10 % ACN in 0.1% aqueous TFA in 15 min).

Allyl-SAM



10

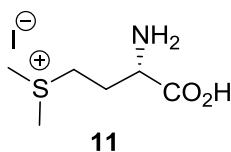
The compound **10** was synthesized as described previously¹¹¹, using activated allyl bromide and *S*-adenosyl homocysteine. The compound was purified by HPLC. The collected fractions were lyophilized to give **10** as a white solid (45 % yield).

¹H-RMN (400 MHz, D₂O) δ (ppm): 8.31 (s, 2H, H Arom), 6.03 (d, $J = 6.2$ Hz, 1H, H_{1'}), 5.74 (m, 1H, H_b), 5.57 (m, $J = 10$ Hz, 17 Hz, 1H, H_c), 5.40 (d, $J = 17$ Hz, 1H, H_c), 4.84 (m, 1H, H_{2'}), 4.52 (q, $J = 8$ Hz, 1H, H_{3'}), 4.45 (q, $J = 8$ Hz, 1H, H_{4'}), 4.02-4.07 (m, 2H, H_a), 3.74-3.82 (m, 3H, H _{α} , H_{5'} et H_{5''}) 3.33-3.50 (m, 2H, 2H _{γ}), 2.25 (q, $J = 8$ Hz, 2H, H _{β}).

Rt = 2.7-3.8 min (gradient 2-10 % ACN in 0.1% aqueous TFA in 10 min, then 100 % A in 5 min).

MS (EI) $m/z = 425.2$ [M]⁺, 324.1 [5'-(prop-2-enyl)thio-5'-deoxyadenosine + H]⁺.

(2*S*)-Methionine methylsulfonium iodide



Methyl iodide (0.25 mL, 1.8 mmol) was added to a solution of L-methionine (244.4 mg, 1.11 mmol) in water (3 mL) and the mixture was stirred vigorously for at least 20 h at 40 °C. Solvent was then evaporated and the resulting white solid was dissolved in a minimum amount of water (0.4 mL) by gentle heating. Ethanol (2.8 mL) was then slowly added, and the suspension was allowed to stand for 12 h in order to obtain maximum precipitate. The product was then filtered to give (2*S*)-methionine methylsulfonium iodide **11** as a white solid (314.8 mg, 97 %).

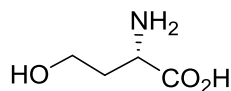
Rf = 0,08 (butanol/H₂O/acetic acid; 6/1.5/2.5)

$[\alpha_D]^{20} = +16.0$ (c 1.0 H₂O) [Lit. ; $[\alpha_D]^{20} = +17.0$ (c 1.0 H₂O)]¹⁸⁹

¹H-NMR (300 MHz, D₂O) δ (ppm): 2.36 (m, 2H, H _{β}), 2.95 (s, 6H, S(CH₃)₂), 3.37-3.57 (m, 2H, H _{γ}), 3.87 (dd, 1H, $J = 6,6$ Hz, $J = 6,6$ Hz, H _{α}). ¹³C-NMR (75 MHz, D₂O) δ (ppm) 24.7 (S(CH₃)₂), 24.9 (C _{β}), 39.4 (C _{γ}), 52.7 (C _{α}), 172.5 (COOH).

MS (ESI, positive, m/z) : 164.07 [M]⁺

(*S*)-2-Amino-4-hydroxybutanoic acid (L-homoserine)



(2*S*)-Methionine methylsulfonium iodide **11** (3.02 g, 10.4 mmol) was dissolved in water (9.5 mL) and the mixture was heated to reflux. The solution prepared by dissolving sodium bicarbonate (893.0mg, 10.6 mmol) in water (13.5 mL) was then added dropwise. The reaction mixture was stirred and heated at reflux for 12 h before being cooled and concentrated to give a clear oil. The crude product was then dissolved in water (1.6 mL) and ethanol (3.2 mL), and treated with acetone (48.2 mL). The resulting white precipitate was then filtered and dissolved in water (6.5 mL) before being freeze-dried, giving (*S*)-2-amino-4-hydroxybutanoic acid **12** as a white solid (998.1 mg, 80 %).

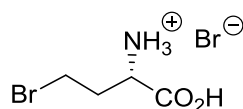
Rf = 0.21 (butanol/H₂O/acid acetic; 6/1.5/2.5)

$[\alpha_D]^{20} = -8.7$ (c 1.0 H₂O) [Lit. $[\alpha_D]^{20} = -8.9$ (c 2.0 H₂O)].¹⁹⁰

¹H-NMR (360 MHz, D₂O) δ (ppm): 1.84-1.92 (m, 1H, H _{β}), 1.98-2.09 (m, 1H, H _{β}), 3.62-3.66 (m, 2H, H _{γ}), 3.69 (dd, 1H, $J = 7.5$ Hz, $J = 4.9$ Hz, H _{α}); ¹³C-NMR (90 MHz, D₂O) δ (ppm): 32.3 (C _{β}), 53.4 (C _{α}), 58.6 (C _{γ}), 174.8 (COOH).

MS (ESI, positive, m/z) : 120.07 [M+H]⁺

α -Amino-4-bromobutanoic acid hydrobromide



(*S*)-2-amino-4-hydroxybutanoic acid **12** (300 mg, 2.5 mmol) was suspended in 7 mL hydrobromic acid (33 % w/v in acetic acid) in a glass high-pressure reaction vessel. The mixture was heated at 80°C for at least 2 days and then allowed to cool down to room temperature before being filtered. The resulting yellow product was washed with diethyl ether, filtered and

dried under high vacuum to give α -amino-4-bromobutanoic acid hydrobromide **13** (505 mg, 76 %).

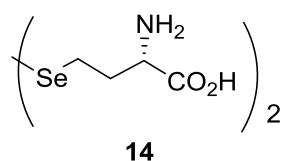
Rf = 0.58 (butanol/H₂O/acid acetic; 6/1.5/2.5)

$[\alpha_D]^{20} = +5.116$ (c, 1.0 H₂O) [Lit. $[\alpha_D]^{20} = +16.0$ (c 1.0 MeOH)].

¹H-NMR (360 MHz, D₂O) δ (ppm): 2.35-2.45 (m, 1H, H _{β}), 2.52-2.62 (m, 1H, H _{β}), 3.56-3.68 (m, 2H, H _{γ}), 4.25 (dd, 1H, $J = 6.8$ Hz, $J = 6.8$ Hz, H _{α}); ¹³C-NMR (90 MHz, D₂O) δ (ppm): 28.0 (C _{γ}), 32.7 (C _{β}), 51.4 (C _{α}), 171.3 (COOH).

MS (ESI, positive, m/z): 181.98 [M+H]⁺

Selenohomocystine

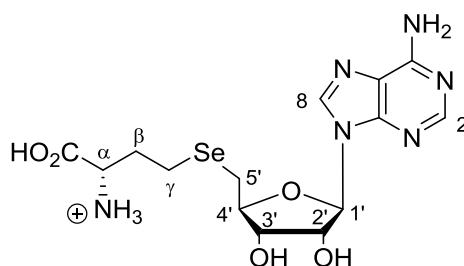


200-mesh selenium powder (114.0 mg, 1.4 mmol) was suspended in 3.5 mL anhydrous ethanol. Sodium borohydride (38.5 mg, 1.0 mmol) was then slowly added and the mixture was heated to reflux at 80 °C for 2 h. The solution became dark brown. After 2 h, α -amino-4-bromobutanoic acid hydrobromide **13** (113.5 mg, 0.43 mmol) was added, resulting in the rapid formation of an opaque yellow mixture. The reaction was kept under reflux at 80°C overnight and quenched with 2 mL 2 M HCl. The solvent was then evaporated and the resulting residue was mixed with 3 mL 2 M HCl. This aqueous solution was washed three times with 1 mL diethyl ether and filtered. A yellow solid was obtained after removing bulk solvent. This crude product was then dissolved in 1 mL 1 M HCl and purified over ion exchange resin. After removing the solvent, the obtained yellow solid was dried under high vacuum to afford selenohomocystine **14**.

Rf = 0.30 (butanol/H₂O/acid acetic; 6/1.5/2.5).

¹H-NMR (300 MHz, D₂O) δ (ppm): 2.28-2.52 (m, 2H, H _{β}), 3.00 (dd, 2H, $J = 7.9$ Hz, $J = 7.9$ Hz, H _{γ}), 4.17 (dd, 1H, $J = 6.6$ Hz, $J = 6.6$ Hz, H _{α}); ¹³C-NMR (75 MHz, D₂O) δ (ppm): 23.6 (C _{γ}), 32.7 (C _{β}), 54.7 (C _{α}), 175.3 (COOH).

Se-adenosyl-L-homocysteine (SeAH)



Selenohomocystine **14** (51.0 mg, 0.14 mmol, 1 eq) was suspended in anhydrous EtOH and NaBH₄ (45.0 mg, 1.18 mmol, 10 eq) was slowly added under a steady flow of argon. The reaction was cooled on ice. The white mixture was stirred vigorously for 15 min at room temperature before adding 5'-chloro-5'-deoxyadenosine (75.4 mg, 0.26 mmol, 2.4 eq). The reaction was allowed to stand at reflux for 1.5 h at 80 °C under a steady flow of argon. The evaporated solvent was periodically replaced before being removed under reduced pressure.

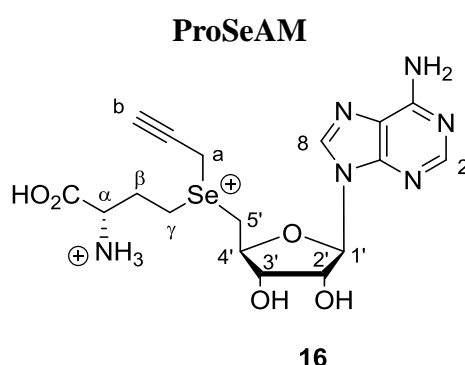
The crude product was then diluted with HCl (5 %, 5 mL) and the resulting solution was filtered using a syringe filter before being purified by semi-preparative RP-HPLC (detection at 260 nm). Fractions containing the desired product were collected and lyophilized, yielding *Se*-adenosylselenohomocysteine **15** as a white solid (75.9 mg, 53 %).

Rf = 0.39 (butanol/H₂O/acetic acid; 6/1.5/2.5).

¹H-NMR (300 MHz, D₂O + 0.1 % TFA-d) δ (ppm): 1.31-1.51 (m, 2H, H_β), 1.86 (dd, 2H, *J* = 7.7 Hz, *J* = 7.7 Hz, H_γ), 2.14-2.26 (m, 2H, H_{5'}), 3.28 (t, 1H, *J* = 6.4 Hz, H_α), 3.47-3.53 (m, 1H, H_{4'}), 3.55-3.58 (dd, 1H, *J* = 5.0 Hz, *J* = 5.0 Hz, H_{3'}), 4.77 (m, 1H, H_{2'}), 5.27 (d, 1H, *J* = 4.8 Hz, H_{1'}), 8.29 (s, 1H, H₂), 8.37 (s, 1H, H₈); ¹³C-NMR (75 MHz, D₂O + 0.1 % TFA-d) δ (ppm): 18.6 (C_γ), 25.3 (C_{5'}), 30.5 (C_β), 52.4 (C_α), 72.8 (C_{3'}), 73.5 (C_{2'}), 84.0 (C_{4'}), 88.3 (C_{1'}), 118.1 (C_q), 142.8 (C₈), 144.4 (C₂), 148.2 (C_q), 149.8 (C_q), 171.3 (COOH).

HPLC Rt = 13.2 min (gradient from 2 to 10% ACN 0.1% TFA in 30 minutes)

MS (ESI, positive, *m/z*) : 433. 07 [M+H]⁺



The compound **16** was synthesized following the general protocol above and as described previously, using allyl bromide and *Se*-adenosyl-L-homocysteine **15**. After HPLC purification and lyophilization, Pro-*Se*AM **16** was obtained with 8 % yield.

¹H-NMR (300 MHz, D₂O + 0.1 % TFA-d) δ (ppm): 2.33-2.42 (m, 2H, H_β), 3.06 (m, 1H, H_b), 3.44-3.63 (m, 2H, H_γ), 3.72-3.88 (m, 3H, 2H_{5'} et H_α), 4.05-4.10 (m, 2H, H_a), 4.49-4.60 (m, 2H, H_{4'} and H_{3'}), 4.72-4.77 (m; 1H; solvent overlap; H_{2'}), 6.10 (d, 1H, *J* = 3.4 Hz, H_{1'}), 8.39-8.40 (m, 2H, H_{arom}).

HPLC retention time for both diastereoisomers = 4.01 min and 4.51 min (gradient 0-10% ACN in 0.1% aqueous TFA in 20 min)

MS (ESI, positive, *m/z*) : 471.09 [M]⁺, 370.04 [5'-(propargyl)-selenoadenosine+H]⁺.

2. Substrates preparation

2.1. Histone H3 and histone H4 peptides used in the methyltransferase assays

The histone H3 peptide (*H*-ART**K**QTARKSTGGKAPRK(Biotin)Q-NH₂, residues 1-19 of H3) was previously synthesized at the LBM¹⁵¹. The lysine residue methylated by GLP is shown in bold in the peptide sequence.

The histone H4 peptide (G(11)KGGAKRHR**K**VLRDNIQG(28)-NH₂, residues 11-28 of H4) was synthesized by Christophe Piesse at the Plateforme Protéomique - Synthèse peptidique, IBPS, Sorbonne Université. The lysine residue methylated by PR-Set7 is shown in bold in the peptide sequence.

Both peptides were characterized by MALDI-TOF MS (Voyager-DE PRO or 4700 Proteomix MALDI-TOF MS/MS, Applied Biosystems) in the reflector positive ion mode using a saturated solution of α -cyano-4-hydroxycinnamic acid (CHCA) matrix in H₂O/ acetonitrile/ TFA (50:50:0.1). The m/z values of the protonated peptides (monoisotopic peak) are: H3 peptide m/z = 2295.2 (calculated m/z = 2295.2); H4 peptide m/z = 1989 (calculated m/z = 1989.3).

2.2. Histone complexes production (tetramer and octamer)

The reconstitution of histone octamer (or tetramer) was performed as in Luger et al. 1997¹⁹¹. Briefly, the four histone proteins H2A, H2B, H3 and H4 of *Xenopus laevis* were expressed individually in *Escherichia coli*. The proteins are completely insoluble and accumulate as inclusion bodies. Consequently, the histone monomers are purified under denaturing conditions (8 M guanidine hydrochloride, 20 mM Tris-HCl pH 8) by gel filtration chromatography using a size-exclusion Superdex 200 column 10/300 (GE Healthcare). The histone monomers were mixed with a H2A:H2B:H3:H4 1,2/1,2/1/1 ratio and dialyzed from the denaturing solution to a Refolding buffer (2 M NaCl, 10 mM Tris-HCl pH 8, 1 mM EDTA, 2 mM β -mercaptoethanol) to form either the heterotetramer complex H3-H4, the dimer complex H2A-H2B or the histone octamer. The mix was concentrated after dialysis and loaded onto a size-exclusion Superdex 200 column 10/30 (GE Healthcare) equilibrated in dialysis buffer (2 M NaCl, 10 mM Tris-HCl pH 8, 1 mM EDTA, 2 mM β -mercaptoethanol). The elution profile and the SDS-PAGE analysis confirmed dimerization of H2A-H2B, and octamerization of H3-H4-H2A-H2B. The purity and the stoichiometry of the different complexes were evaluated with an acrylamide gel stained with Coomassie Blue.

Histone tetramer was obtained mixing only H3 and H4 with a 1/1 ratio and dialyzed from the denaturing solution to the same Refolding buffer. The purification protocol was the same as for octamer purification. The elution profile and the SDS-PAGE analysis confirmed the tetramerization of H3-H4. Histone tetramer had the same running behaviour on size exclusion chromatography as histone octamer.

2.3. Nucleosome core particles

The reconstitution of nucleosome core particle from histone octamer and a defined sequence of DNA was performed as in Dyer et al. 2004¹⁵³. The Figure 1, extracted from this article summarizes the approach. Briefly, the histone octamer and a small sequence of 146 pb of DNA were mixed in a 1/1 ratio. After 3 dialysis to a Reconstitution buffer (10 mM Tris-HCl pH 8, 1 mM EDTA, 2 mM β -mercaptoethanol), the salts were slowly removed and the native interactions between the DNA and the histones were established. This protocol allowed the production of small quantities of recombinant nucleosome (30-80 μ g). These histone-DNA complexes are not very stable and were freshly prepared before the methyltransferase experiments (less than one week before).

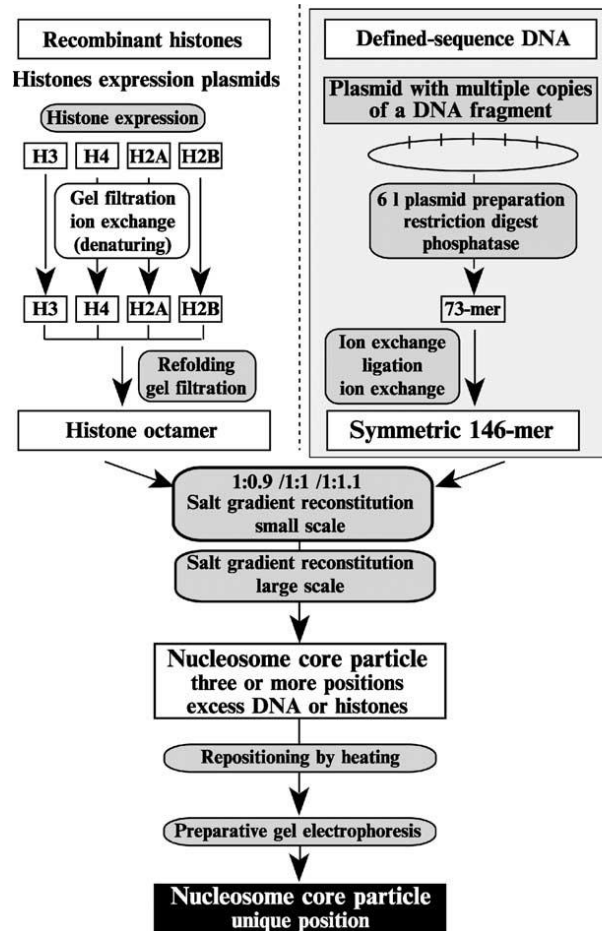


Figure 14. Flow chart of methods used for preparation of components for nucleosomes as described in Dyer et al. 2004 ¹⁵³

3. GLP, PR7-Set WT and mutants expression and site directed mutagenesis

3.1. General procedures

- Preparation of stock antibiotics

Table 1. Stock antibiotics

Antibiotic	Stock Concentration
Chloramphenicol (Cam)	25 mg/mL (in ethanol)
Ampicillin (Amp)	100 mg/mL (in sterile water)
Kanamycin (Kan)	50 mg/mL (in sterile water)

Five milliliters of chloramphenicol, ampicillin and kanamycin were prepared in the above concentrations, separated into 1 mL eppendorf tubes and stored at -20 °C. The ampicillin was filtered (syringe filter Minisart 0.2 µm from Sartorius Stedim biotech) before being distributed into the eppendorf tubes.

- Preparation of culture media

Table 2. Culture media composition

Medium	Component	Concentration
LB (Luria-Bertani)	Tryptone	10 g/L
	Yeast extract	5 g/L
	NaCl	10 g/L
	Agar ^a	15 g/L
SOC	Tryptone	10 g/L
	Yeast extract	5 g/L
	NaCl	10 g/L
	MgCl ₂ ^b	2 M
	Glucose ^b	20 %

^aAgar added only in the preparation of LB agar, a solid LB medium.

^bAdded just before use.

Each component of the LB medium was added to 950 mL of distilled water. The mixture was agitated until all components had dissolved. The pH was adjusted to 7, using NaOH. Distilled water was then added to adjust the volume to one liter. These methods were carried-out on larger scales to obtain sufficient stocks of LB medium. The LB medium was separated into recipients of various sizes and sterilized in an autoclave.

The SOC medium was prepared and sterilized in the same way as the LB medium. The last two components, MgCl₂ and glucose, were added just before SOC use.

3.2. Plasmids

Plasmid pGEX-C-ter-GLP for GLP expression was provided by Prof. Albert Jeltsch (Institute of Biochemistry and Technical Biochemistry, University of Stuttgart). Plasmid Pet28b-KMT5A-Flag-6-His for PR-Set7 expression was provided by R. Margueron (originally from Dr D. Reinberg, NYU, New York, USA). From the PR-Set7 and GLP expression plasmids, all point mutants were generated by Gibson assembly.

Table 3. Protein, cell, plasmid and corresponding antibiotics

Protein	Tag	<i>E. coli</i> cell	Plasmid	Antibiotic	Source
C-ter-mouse GLP (1002-1297)	GST	BL21(DE3)pLysS competent cells (Novagen)	pGEX	Chloramphenicol Ampicillin	Prof. Albert Jeltsch, Institute of Biochemistry and Technical Biochemistry University of Stuttgart
PR-Set7 full length	6-His and Flag	BL21(DE3)pLysS competent cells	pET28b	Chloramphenicol Kanamycin	Dr D. Reinberg, NYU, New York, USA

3.3. Competent cells transformation

For the transformation of PR-Set7, 100 µL of *E. coli* BL21(DE3)pLysS (Novagen®) were grown at 37°C in 100 mL of LB medium, pH 7, supplemented with 25 µg/mL chloramphenicol and 50 µg/mL Kanamycin. For the transformation of PR-Set7, 100 µL of *E. coli* BL21(DE3)pLysS (Novagen®) were grown at 37°C in 100 mL of LB medium, pH 7,

supplemented with 25 µg/mL chloramphenicol and 50 µg/mL Ampicillin. The following protocol was employed for both transformations.

Plasmids were diluted in 5 µL of H₂O milli-Q (Thermo Scientific Barnstead Easypure II); 1 µL of this solution (at approximately 0.5 µg/µL) was added to 100 µL of BL21(DE3)pLysS competent cells (Invitrogen) and put in ice for almost 30 minutes, then, transferred in a water heater thermostat at 42°C during 50 seconds and then again put into ice for 2 minutes. After that time 800 µL of SOC medium (LB medium supplemented with MgCl₂ 10 mM, Glucose 20%) preheated at 42°C were added and the obtained mixture was incubated at 37°C for one hour before the spreading on a Petri dish containing LB Agar (supplemented with the appropriate antibiotics).

3.4. Protein expression

For the bacterial cultures, 50 mL of LB-CAM medium were prepared with the appropriate antibiotics: chloramphenicol 25 µg/mL and 50 µg/mL kanamycin in the case of the culture for the enzyme PR-Set7, chloramphenicol 25 µg/mL and 50 µg/mL ampicillin in the case of the culture for GLP; a bacteria colony was taken from both Petri dishes and added to the respective medium solution. The pre-cultures were then incubated overnight at 37°C with stirring (200 rpm). Next, 10 mL of the pre-cultures were added to 500 mL of LB-CAM containing the same antibiotic as in the precultures.

In the case of PR7-Set, the solution was incubated at 37°C with stirring (180 rpm) until a value of OD₆₀₀ ≈ 0.4; the temperature was then shifted to 22°C until OD₆₀₀ ≈ 0.6. Overexpression was induced by 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) overnight. In the case of GLP, the solution was incubated at 37°C with stirring (180 rpm) a value of OD₆₀₀ ≈ 0.6; Overexpression was induced by 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) overnight for 3 h.

The cells were collected by centrifugation at 6000 rpm, 4°C for 15 minutes and resuspended in 25 mL of binding and washing Buffer 1 for PR-Set7 and Buffer 2 for GLP. Cells were then centrifuged once again at 6000 rpm, 4°C for 20 minutes and resuspended in 25 mL of the same Binding and washing buffer. The cells were disrupted by sonication (Bioblock Scientific® Vibra-Cell), 10 times during 15 seconds with pauses of 45 seconds. The solution was centrifuged a third time at 20000 rpm, 4°C for one hour and finally passed through the purification column (equilibrated with their respective binding and washing buffer) with a rate flow of 1 mL/min on a Ni-column (His-Trap FF column Cytiva Life Science) in the case of PR-Set7 or on an immobilized glutathione column (GSTrap FF column Cytiva Life Science) in the case of GLP.

In the case of PR7-Set, the bound enzymes were then eluted with Elution buffer 1 and the fractions were dialysed by means of a semi-permeable membrane (Spectra/Por® Dialysis Membrane, MWCO = 6000-8000 Da) in 1 L of Dialysis buffer 1 for 1 h then in Storage buffer 1 overnight. In the case of GLP, the GST-tag protein was eluted with Elution Buffer 2, dialyzed 1 h in Dialysis buffer 2 and overnight in Storage buffer 2. The purity of the eluted protein is verified with the loading of 5 µL of the sample on a 12 % acrylamide gel.

The concentration of each fraction was measured with a NanoDrop (Thermo Scientific, NanoDrop 2000c, absorbance reading at 280 nm). The dialysed fractions were then aliquoted in portions of 100 µL and stored at -80°C.

Mutation	Primers	Matrix
Y336A	Primer Fwd 1 Primer Rev 1	<i>Pet28b-KMT5A-Flag-6-His</i>
Y334A	Primer Fwd 2 Primer Rev2	<i>Pet28b-KMT5A-Flag-6-His</i>
Y334F	Primer Fwd 3 Primer Rev2	<i>Pet28b-KMT5A-Flag-6-His</i>
Y336A/ Y334F	Primer Fwd 4 Primer Rev2	<i>Pet28b-KMT5A-Y336A-Flag-6-His</i>
Y336A/Y334F/A286G	Gibs PRSet7 EcoRI F Gibs PRSet7 A286G R Gibs PRSet7 A286G F Gibs pET28b XhoI R	<i>pET28b PR-SET7 Y334F / Y336A</i>
A286G	Gibs PRSet7 EcoRI F Gibs PRSet7 A286G R Gibs PRSet7 A286G F Gibs pET28b XhoI R	<i>pET28b PR-SET7 WT</i>
Y334F/A286G	Gibs PRSet7 EcoRI F Gibs PRSet7 A286G R Gibs PRSet7 A286G F Gibs pET28b XhoI R	<i>pET28b PR-SET7 Y334F</i>
Y336G/Y334F	Gibs PRSet7 EcoRI F Gibs PRSet7 Y334FY336G F Gibs PRSet7 Y334FY336G R Gibs pET28b XhoI R	<i>pET28b PR-SET7 WT</i>

3.6. Plasmid purification

The plasmid DNA was extracted from 2 mL of suspended bacteria, using the Nucleospin plasmid kit (Macherey Nagel), following the provider's recommendations. The protocol of the Miniprep kit is as follows: the bacteria are condensed via centrifugation (10000g, 10 min) and are then treated with an alkaline lysate, which neutralizes and forms a precipitate, which is eliminated via centrifugation. The lysate is deposited onto a resin, to which the plasmids stick. The resin is washed and the plasmids are eluted with milliQ water.

4. Enzymes assays

4.1. Methyltransferase assay on peptide substrates

For the screening, a 20- μ L reaction mixture contains 2 μ g HKMT (GLP, PR-Set7 or their mutants), 50 μ M histone peptide (H3 or H4) and 100 μ M cofactor (or analogue) in the reaction buffer (50 mM glycine, pH 9.8, 2 mM dithiothreitol, 25 μ g/mL BSA, 10 % glycerol). In the case of PR7-Set, the reaction was carried out for 3 h at 30°C. In the case of GLP, the reaction was carried out for 30 min at 30°C. Then, the reaction was either quenched by adding to the enzymatic mixture 180 μ L of acidified water (H₂O, 0.2 % TFA) for MALDI-TOF MS analysis or the sample was subjected to the click chemistry protocol.

4.2. MALDI-TOF MS analysis

To analyze the peptide samples, 1 μ L of reaction mixture was mixed with 1 μ L of a saturated solution of α -cyano-hydroxycinnamic acid (CHCA) in H₂O/ACN/TFA (50/50/0.1) (Sigma). The resultant sample was deposited on a MALDI plate and dried at room temperature. MALDI-TOF MS analyses of these samples were performed with delayed extraction in a positive ion mode on a DE-PRO mass spectrometer (Applied Biosystems). External calibration Proteomix 4 was used for mass calibration (LaserBio Labs Sofia-Antipolis). Desorption/Ionization was

obtained by using a 337-nm pulsed nitrogen laser (20 Hz). Laser power was adjusted to slightly above threshold to obtain good resolution and signal/noise ratios. All data were processed using the Data Explorer software package (Applied Biosystems).

4.3. Methyltransferase assay on histone complexes

For the reactions containing histone complexes, 5 μg of PR-Set7 (or 1 μg of GLP) was incubated with 0.5 μg histone H4 in complex (i.e. 2.5 μg histone tetramer, 2.5 μg histone octamer or 5 μg nucleosome core particles) and 100 μM cofactor analogue for 3 h at 30°C in a glycine reaction buffer (50 mM glycine, pH 9.8, 2 mM dithiothreitol, 25 $\mu\text{g}/\text{mL}$ BSA, 10 % glycerol). Then the enzymatic reaction was either quenched by adding 5X Laemmli buffer to apply the on-membrane click chemistry protocol, or the sample was directly subjected to the in-solution click chemistry protocol. For the competition assay, in addition to the 100 μM of cofactor analogue used during the standard methyltransferase assay, 500 μM of SAH was added.

4.4. In-solution click chemistry reactions

Cu(I)-catalyzed alkyne-azide cycloaddition (CuAAC):

After the methyltransferase assay, the enzymatic reaction mixture was subjected to the click reaction by adding 100 μM of a probe functionalized with TAMRA or biotin, 2 mM CuSO_4 , 4 mM sodium ascorbate, 400 μM Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA). The mixture was gently shaken (300 rpm) for 2 h at 30 °C in the dark. The reaction was stopped by adding Laemmli buffer and the samples were deposited on a polyacrylamide gel.

Strain-promoted azide alkyne cycloaddition (SPAAC):

To the enzymatic reaction mixture, 100 μM of the cyclooctyne probe functionalized with TAMRA (or biotin) was added. The mixture was gently shaken (300 rpm) for 2 h at 30 °C in the dark. The reaction was quenched with Laemmli buffer and the samples were subjected to the In-gel fluorescence protocol or Western blot analysis.

4.5. SDS PAGE analysis

20 μL of the samples were mixed with 5 μL of a 5 X loading buffer (Tris-HCl 0.25 M, pH 6.8, SDS 10 %, glycerol 50 %, Bromophenol blue 0.25 %, β -mercaptoethanol 10 %). The samples were then heated in a water thermostat at 95°C for 5 minutes and 25 μL of the samples were deposited on an electrophoresis gel SDS-PAGE 12 % (**Table 5**), 0.75 mm thick. The potential was regulated at 120 V for the time the samples went across the stacking portion of the gel; at 200 V (Apelex® Electrophoresis Power Supply ST606T) for the time they went across the resolving portion.

Table 5. The composition of the SDS-PAGE 12%

Component of resolving solution (for 2 gels)	Quantity	Component of stacking solution (for 2 gels)	Quantity
Acrylamide / Bisacrylamide 40 % 37,5/1	5.4 mL	Acrylamide / Bisacrylamide 40 % 37.5/1	0.8 mL
TRIS 1M (pH = 8.8)	6.75 mL	TRIS 1M (pH 6.8)	1 mL
SDS 10 %	0.18 mL	SDS 10 %	0.08 mL
Ammonium Persulfate (APS) 10 %	0.18 mL	Ammonium Persulfate (APS) 10 %	0.08 mL
Tetramethylethylenediamine (TEMED) (added just before the use)	18 μ L	Tetramethylethylenediamine (TEMED) (added just before the use)	8 μ L
H ₂ O milliQ	5.47 mL	H ₂ O milliQ	6,03 mL

4.6. In-gel Fluorescence

After the click chemistry reaction introducing a fluorescent probe, the proteins were precipitated with a standard chloroform-methanol protocol.

To 50 μ L of the sample, 200 μ L of methanol and 50 μ L of chloroform were added. After mixing, 150 μ L of H₂O were added and the proteins precipitate. The sample was centrifuged for 5 min at 14000g. The aqueous phase was removed carefully to not touch the interface containing the proteins. Then, 200 μ L of methanol was added, the sample was mixed and centrifuged for 5 min at 14000g. The methanol was removed and the pellet was dried 5 min with air. The pellet was solubilized in Laemmli buffer without bromophenol blue and heated 10 min at 70°C. Then the proteins were separated on a polyacrylamide gel. The gel was washed at least 5 times with water and scanned with a Typhoon Tri Imaging System (GE Healthcare) to measure the emitted fluorescence of the TAMRA probe ($\lambda_{ex} = 550$ nm, $\lambda_{em} = 575$ nm). In case the fluorescence signal saturates the detector, the gel is washed to obtain a better signal/noise ratio.

4.7. Western blot (standard protocol)

After the samples were mixed with Laemmli buffer, they were boiled for 5 min. The proteins were separated on a 4-15 % polyacrylamide denaturing gel (Biorad) then protein transfer was performed on a PVDF membrane using the Trans-Blot Turbo Transfer System (Biorad). After verification of the transfer with Ponceau staining, the membrane was rinsed once with PBST (phosphate buffered saline, 0.1 % Tween) buffer and then incubated in blocking buffer (3 % BSA in PBST) at room temperature for 1 h on an orbital shaker. The membrane was rinsed once with PBST and then incubated in antibody solution I (PBST supplemented with the diluted antibody) for 1 hour with gentle agitation. The membrane was rinsed three times for 10 min with PBST and then incubated in antibody solution II (PBST supplemented with the diluted antibody) coupled with a fluorophore for 1 h on an orbital shaker. Finally, the membrane was washed three times for 10 min with PBST.

Antibody used for antibody solution I: H4 mAb (Active Motif) 1:2000 ; H2B mAb (Active Motif) ; His-tag antibody (Genscript)

Antibody used for antibody solution II: Mouse Starbright 700 fluorescent antibody (Biorad) 1:10000 ; Goat Starbright 700 fluorescent antibody (Biorad) 1:10000.

In the case of Anti-Biotin Western blots : The antibody solution I is replaced by a solution of Streptavidin-HRP conjugate (1:5000). After a 1 h incubation step, the membrane was revealed with the standard ECL western blotting protocol from ThermoScientific (SuperSignal West Pico PLUS Chemiluminescent substrate).

The fluorescence or the chemiluminescence was visualized with a ChemiDoc Imaging System (Biorad).

4.8. *On-membrane click chemistry reaction*

After electroblotting, the membrane was rinsed twice with distilled water. Then, the membrane was placed into a Petri dish and incubated in the click reaction solution (75 μ M copper sulfate, 1.5 mM sodium ascorbate, 100 μ M THPTA and 10 μ M Biotin-PEG4-azide probe in 50 % water, 50 % dimethyl sulfoxide) for 45 min at room temperature under gentle agitation. The membrane was alternately washed 3 times for 5 min with methanol and DMSO. Finally, the membrane was rinsed once with methanol and once with 70 % ethanol.

The membrane was then revealed by Western blot (resume the protocole from the blocking step)

4.9. *Electroblotting coupled with MALDI-TOF*

After protein separation by SDS-PAGE and transfer to a nitrocellulose membrane, as described in the standard western blot protocol, the membrane was stained with Ponceau (0.5 % Ponceau-S in 1 % acetic acid) then partially destained in distilled water until the bands became visible. The protein bands (4 mm²) were excised, stored individually and completely destained with distilled water. The bands were dissolved in 40 μ L of a saturated solution of CHCA Matrix (69.5 % acetonitrile, 29.5 % methanol and 1 % trifluoroacetic acid). Solutions were sonicated for 10 min and 1 μ L spotted onto the MALDI plate for MS analysis of intact proteins.

4.10. *Radiolabeling*

200 ng of PR-Set7 was incubated with 2 μ g nucleosome core particles and 1.5 μ L ³H-SAM (Perkin Elmer ref NET155001MC) for 1 h at 30°C in the glycine reaction buffer. The samples were then loaded on a 15 % SDS-PAGE gel. Electrophoresis was performed for 15 min at room temperature. After electrophoresis, the gel was exposed to a phosphorimager screen. After being exposed overnight, the screen was scanned with the Phosphorimager TYPHOON Trio instrument (GE Healthcare).

5. Characterization of H4 specific functionalization by mass spectrometry²

5.1. *Protein in-gel digestion*

Proteins were separated by 15 % SDS-PAGE gels and stained with colloidal Coomassie blue (LabSafe Gel Blue™, AGRO-BIO) reagent. Histone bands were excised and digested as described in standard protocols. Briefly, following the SDS-PAGE and washing of the excised gel slices, proteins were reduced by adding 10 mM dithiothreitol (DTT, Sigma Aldrich) prior

² Collaboration with Bérangère Lombard and Damarys Loew from the mass spectrometry and proteomic facility (LSMP), Institut Curie, PSL Research University

to alkylation with 55 mM iodoacetamide (Sigma Aldrich). After washing and shrinking of the gel pieces with 100 % acetonitrile (ACN), digestion was done overnight in 25 mM ammonium bicarbonate at 30°C, by adding 0.10 µg trypsin (Promega). Peptides were extracted from the gel using 60/40 ACN/H₂O + 5 % formic acid, vacuum concentrated at room temperature to dryness and reconstituted in 0.1 % formic acid. Samples were then loaded onto a homemade C18 StageTips packed with Empore™ C18 Extraction Disk (3M™ Disc 2215) for desalting. Peptides were eluted using 40/60 ACN/H₂O + 0.1 % formic acid, vacuum concentrated to dryness and reconstituted in 10 µL 0.3 % trifluoroacetic acid (TFA) prior to liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis.

5.2. Liquid chromatography- tandem mass spectrometry (LC-MS/MS) and data analysis

LC was performed with an RSLCnano system (Ultimate 3000, Thermo Scientific) coupled on line to a Q Exactive HF-X mass spectrometer (MS, Thermo Scientific). Peptides were first trapped onto a C18 column (75 µm inner diameter × 2 cm; nanoViper Acclaim PepMap™ 100, Thermo Scientific) with buffer A (98/2 H₂O/ACN + 0.1 % formic acid) at a flow rate of 2.5 µL/min over 4 min. Separation was performed on a 50 cm x 75 µm C18 column (nanoViper C18, Acclaim PepMap™ RSLC, 2 µm, 100Å, Thermo Scientific) regulated to a temperature of 50°C and with a linear gradient of 2 % to 35 % buffer B (100 % acetonitrile in 0.1 % formic acid) at a flow rate of 300 nL/min over 91 min. MS full scans were performed in the ultrahigh-field Orbitrap mass analyzer in ranges *m/z* 300–1500 with a resolution of 120 000 at *m/z* 200. The maximum injection time (MIT) in MS was 50 ms and the automatic gain control (AGC) was set to 3×10^6 . The top 20 intense ions were subjected to Orbitrap for further fragmentation via high energy collision dissociation (HCD) activation and a resolution of 15 000 with the intensity threshold kept at 1.3×10^5 . We selected ions with charge state from 2+ to 6+ for screening. Normalized collision energy (NCE) was set at 27. For each MS/MS scan, the AGC was set at 1×10^5 , the MIT was 60 ms and the dynamic exclusion of 15s.

For protein identification, the data were searched against the *Xenopus Laevis* Swiss-Prot database using Mascot™. Carbamidomethylation of cysteines, oxidation of methionines, methylation, dimethylation and trimethylation of lysines were set as variable modifications for biological condition with SAM as cofactor. Carbamidomethylation of cysteines, oxidation of methionines, mono and di propargylation on lysines, mono and dipropargylation on lysines were set as variable modifications for biological conditions with ProSeAM. Specificity of trypsin or ArgC digestion was set and a maximum of five-missed cleavage sites was allowed. Maximum allowed mass deviation was set to 10 ppm for monoisotopic precursor ions and 0.02 Da for MS/MS peaks. The resulting Mascot files were further processed using myProMS v3.9.2¹⁶³. The maximum false discovery rate (FDR) calculation was set to 5 % at the peptide level for the whole study (QUALITY algorithm) and the spectra were checked manually.

6. Identification of potential targets for PR-Set7 – Proteomic study

6.1. Cell lines

HEK-293T (Invitrogen) cell lines were grown according to the manufacturer's instructions. Cell lines were tested for the absence of mycoplasma every month.

6.2. Nuclear extract

For nuclear extract preparation, cells were incubated with 200 µL buffer A (10 mM HEPES pH 7.9, 2.5 mM MgCl₂, 0.25 M sucrose, 0.1 % NP-40, 0.5 mM DTT, 1 mM PSMF) for 10 min on

ice, centrifuged at 7000g for 10 min, resuspended in 200 μ L buffer B (25 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 700 mM NaCl, 0.5 mM DTT, 0.1 mM EDTA, 20 % glycerol), sonicated and centrifuged at 21,000g during 15 min.

6.3. Enzymatic assay and click chemistry reaction

For this enzymatic assay the different conditions were done in 5 replicates. 20 μ g of PR-Set7 was incubated with 200 μ g nuclear extract and 100 μ M ProSeAM for 3 h at 30°C in 200 μ L reaction buffer (50 mM glycine, pH 9.8, 2 mM dithiothreitol, 25 μ g/mL BSA, and 10 % glycerol). Then the reaction was quenched by adding 4 volumes of cold methanol. Proteins were precipitated overnight at -80°C. The pellets were dissolved in resuspension buffer 1 (PBS, 0, 5 % SDS).

After the methyltransferase assay, the enzymatic reaction mixture was subjected to the “click” reaction by adding 100 μ M biotine-PEG4-azide, 2 mM $CuSO_4$, 4 mM sodium ascorbate and 400 μ M THPTA. The mixture was gently shaken (300 rpm) for 2 h at 30°C in the dark. Then the reaction was quenched by adding 4 volumes of cold methanol. Proteins were precipitated overnight at -80°C. The pellets were dissolved in resuspension buffer 2 (PBST, 3 % SDS, 10 mM DTT, 1 mM EDTA).

6.4. On beads protein enrichment

Simultaneously with the protein precipitation step following the click chemistry reaction, Dynabeads M-280 magnetic beads coupled with streptavidin (ThermoFisher Scientific) were washed once with PBST and the surfaces were blocked with PBST, 3 % BSA overnight and washed twice with PBST. After enzymatic functionalization and chemical labeling with the biotin probe, the precipitated proteins were resuspended in 100 μ L PBST supplemented with 3 % SDS, 10 mM DTT, 1 mM EDTA. 200 μ L of the magnetic beads suspension (5 mg.mL⁻¹) were incubated with the solubilized proteins, 2 h at 4°C with up and down agitation. The magnetic beads were then successively washed three times with each of the following buffers: PBST; PBS 500 mM NaCl; PBS 1 % NP-40, PBS 4 M Urea. Finally, the beads were washed once with PBST, once with PBS and twice with a 25 mM NH_4HCO_3 buffer.

6.5. Protein on beads digestion

After resuspending beads in 200 μ L of 25 mM NH_4HCO_3 , samples were digested on beads by adding 0.2 μ g of trypsin/LysC (Promega) for 1 hour at 37°C. The resulting peptide mixtures were then loaded onto a homemade C18 StageTips for desalting. Peptides were eluted from beads using 40/60 ACN/ H_2O + 0.1 % formic, vacuum concentrated to dryness and reconstituted in 10 μ L 0.3 % TFA prior to LC-MS/MS analysis.

6.6. LC-MS/MS and data analysis

LC was performed with an RSLCnano system coupled online to an Orbitrap Exploris 480 mass spectrometer (Thermo Scientific). Peptides were trapped on a C18 column (75 μ m inner diameter \times 2 cm; nanoViper Acclaim PepMapTM 100, Thermo Scientific) with buffer A at a flow rate of 3.0 μ L/min over 4 min. Separation was performed on a 50 cm \times 75 μ m C18 column (nanoViper Acclaim PepMapTM RSLC, 2 μ m, 100Å, Thermo Scientific) regulated to a temperature of 40°C with a linear gradient of 3 % to 29 % buffer B at a flow rate of 300 nL/min over 91 min. MS full scans were performed in the ultrahigh-field Orbitrap mass analyzer in ranges m/z 375–1500 with a resolution of 120 000 at m/z 200. The top 20 most intense ions were subjected to Orbitrap for further fragmentation via HCD activation and a resolution of 15 000 with the AGC target set to 100 %. We selected ions with charge state from 2+ to 6+ for screening. NCE was set to 30 % and the dynamic exclusion to 40s.

For protein identification, the data were searched against Homo sapiens one gene one protein (UP000005640_9606) UniProt database using Sequest HT through proteome discoverer (version 2.2). Enzyme specificity was set to trypsin and a maximum of two miss cleavages sites were allowed. Oxidized methionine, *N*-terminal acetylation and loss of *N*-terminal methionine were set as variable modifications. Maximum allowed mass deviation was set to 10 ppm for monoisotopic precursor ions and 0.02 Da for MS/MS peaks.

The resulting files were further processed using myProMS v3.9.2. FDR calculation used Percolator and was set to 1% at the peptide level for the whole study. The label free quantification was performed by peptide Extracted Ion Chromatograms (XICs) computed with MassChroQ version 2.2.1¹⁶⁴. For protein quantification, XICs from proteotypic peptides shared between compared conditions (TopN matching) and missed cleavages were allowed. Median and scale normalization was applied on the total signal to correct the XICs for each biological replicate. To estimate the significance of the change in protein abundance, a statistical test based on a linear model adjusted on peptides and biological replicates was performed and *p*-values were adjusted with a Benjamini–Hochberg FDR procedure with a control threshold set to 0.05.

7. Expression of the candidates

7.1. Plasmids

All the pDONR and pDEST plasmids for the hypothetical targets of PR-Set7 expression come from DNASU.

For EPS8L1, ETAA1, EXD2, FAM208A, GGCT, NIFUN, LYAR, NDC80, NMT1, PNISR, PPA1, PPL, RPL23, RPS21, RPS26, SMARCC1, SNRPE, SRFBP1, TK1, all the plasmids for a bacterial or a baculovirus expression were generated from the pDONR plasmids using the Gateway LR Clonase kit (ThermoFisher).

Table 6 : List of plasmid for bacterial or baculovirus expression of PR-Set7 potential targets

Donor plasmids (DNASU)	pENTR223-EPS8L1, pDONR221-ETAA1, pENTR223-EXD2, pDONR221-FAM208A, pENTR223-GGCT, pDONR221-NIFUN, pENTR223-LYAR, pENTR223-NDC80, pENTR223-NMT1, pDONR221-PNISR, pDONR221-PPA1, pDONR221-PPL, pENTR223-RPL23, pDONR221-RPS21, pENTR223-RPS26, pDONR221-SMARCC1, pENTR223-SNRPE, pENTR223-SRFBP1, pENTR223-TK1
Destination vector for bacterial expression	pDEST17 vector (ThermoFischer Scientific)
Destination vector for baculovirus expression	pDEST10 vector (ThermoFischer Scientific)

7.2. Protein expression

Expression in *E. coli*

Recombinant EPS8L1, ETAA1, EXD2, GGCT, NIFUN, NDC80, NMT1, PPA1, PPL, RPL23, RPS21, RPS26, SMARCC1, SNRPE, SRFBP1, TK1 were overexpressed in *E.coli*

BL21(DE3)pLysS (Novagen) induced by 0,1 mM isopropyl-1-thio-D-galactopyranoside (IPTG) at 37°C for 3 h. The protocols were essentially the same as for PR-Set7 expression and purification except for the antibiotics added to the LB medium. 25 µg/mL chloramphenicol and 50 µg/mL Spectinomycin was used for EPS8L1, EXD2, GGCT, NDC80, NMT1, RPL23, RPS26, SNRPE, SRFBP1 and TK1 expression ; 25 µg/mL chloramphenicol and 50 µg/mL Kanamycin was used for ETAA1, NIFUN, PPA1, PPL, RPS21 and SMARCC1 expression. Lysates were resuspended in Binding and washing Buffer 1 (composition in Table 4), sonicated and clarified by centrifugation before incubation with Ni-NTA resin (Qiagen) and eluted with imidazole using Elution Buffer 1 (composition in Table 4). Purified proteins were aliquoted and stored at -80°C

Expression in SF-9 insect cells

The baculoviruses were produced accordingly to Bac-to-Bac Baculovirus Expression Systems (Invitrogen) after cDNA cloning into pDEST vectors using the Gateway LR Clonase kit (ThermoFischer).

Recombinant EPS8L1, ETAA1, FAM208A, NIFUN, PNISR, SMARCC1, SRFBP1 were overexpressed in SF-9 insect cells after infection with the corresponding baculoviruses by adding each baculovirus to $800 \cdot 10^6$ of SF-9 cells. After 3 days the infected SF-9 cells were spined down and the pellets resuspended in His buffer (30mM NaHPO4 pH7,2, 500mM NaCl, 20mM imidazole, 10% glycerol) then sonicated and clarified by centrifugation before incubation with Ni-NTA resin (Qiagen). eluted with imidazole using Elution Buffer 1. Purified proteins were aliquoted and stored at -80°C

8. Docking study³

Molecular Dynamics

3D models of the mutants of PR-Set7 and GLP coupled to a cofactor (SAH, SAM or SAM analogs) and their respective peptide substrate (residue 16 to 23 of histone H4 and residues 2 to 9 of the histone H3) were generated based on a 3D structure of PR-Set7 (PDB entry: 1ZKK) and GLP (PDB entry: 2RFI) and refined by several cycles of minimization and equilibrated by molecular dynamics simulations (1 ns runs) using the CHARMM force field. Bonds involving hydrogen atoms were constrained using the SHAKE algorithm¹⁹². A solvent box with a periodic boundary of 10 Å was added to the model. Solvation was completed with 0.150 mM NaCl using the solvation protocol implemented in Discovery Studio (Dassault Systèmes BIOVIA, Discovery Studio, 2019, San Diego: Dassault Systèmes, 2019).

³Collaboration with Nicolas Pietranscosta (Laboratoire des Biomolécules (LBM) and Neuroscience Paris Seine laboratory (NPS), IBPS, Sorbonne Université)

Références

Références

1. Waddington, C. H. The Strategy of the Genes. *Yale J Biol Med* **30**, 470–471 (1958).
2. Goldberg, A. D., Allis, C. D. & Bernstein, E. Epigenetics: a landscape takes shape. *Cell* **128**, 635–638 (2007).
3. Deans, C. & Maggert, K. A. What do you mean, ‘epigenetic’? *Genetics* **199**, 887–896 (2015).
4. Lee, J. T. Epigenetic regulation by long noncoding RNAs. *Science* **338**, 1435–1439 (2012).
5. Razin, A. & Cedar, H. DNA methylation and genomic imprinting. *Cell* **77**, 473–476 (1994).
6. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33 Suppl**, 245–254 (2003).
7. Ernst, P. *et al.* Definitive hematopoiesis requires the mixed-lineage leukemia gene. *Dev Cell* **6**, 437–443 (2004).
8. Feng, S., Jacobsen, S. E. & Reik, W. Epigenetic reprogramming in plant and animal development. *Science* **330**, 622–627 (2010).
9. Bird, A. P. CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209–213 (1986).
10. Moore, L. D., Le, T. & Fan, G. DNA methylation and its basic function. *Neuropsychopharmacology* **38**, 23–38 (2013).
11. Oswald, J. *et al.* Active demethylation of the paternal genome in the mouse zygote. *Curr Biol* **10**, 475–478 (2000).
12. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
13. Brodeur, J. & Toussaint, M. *Biologie moléculaire: Concepts . Techniques . Applications*. (Centre collégial de développement de matériel didactique (CCDMD), 2007).
14. Chan, J. C. & Maze, I. Nothing Is Yet Set in (Hi)stone: Novel Post-Translational Modifications Regulating Chromatin Function. *Trends Biochem Sci* **45**, 829–844 (2020).
15. Nadal, S., Raj, R., Mohammed, S. & Davis, B. G. Synthetic post-translational modification of histones. *Curr Opin Chem Biol* **45**, 35–47 (2018).
16. Farooq, Z., Banday, S., Pandita, T. K. & Altaf, M. The many faces of histone H3K79 methylation. *Mutat Res Rev Mutat Res* **768**, 46–52 (2016).
17. Gräff, J. & Tsai, L.-H. Histone acetylation: molecular mnemonics on the chromatin. *Nat Rev Neurosci* **14**, 97–111 (2013).
18. Gibney, E. R. & Nolan, C. M. Epigenetics and gene expression. *Heredity* **105**, 4–13 (2010).
19. Janssen, K. A., Sidoli, S. & Garcia, B. A. Recent Achievements in Characterizing the Histone Code and Approaches to Integrating Epigenomics and Systems Biology. *Methods Enzymol* **586**, 359–378 (2017).
20. Hyun, K., Jeon, J., Park, K. & Kim, J. Writing, erasing and reading histone lysine methylations. *Exp. Mol. Med.* **49**, e324 (2017).
21. Agger, K., Christensen, J., Cloos, P. A. C. & Helin, K. The emerging functions of histone demethylases. *Curr Opin Genet Dev* **18**, 159–168 (2008).
22. Nishioka, K. *et al.* PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. *Mol. Cell* **9**, 1201–1213 (2002).
23. Couture, J.-F., Collazo, E., Brunzelle, J. S. & Trievel, R. C. Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase. *Genes Dev* **19**, 1455–1465 (2005).
24. Guccione, E. *et al.* Methylation of histone H3R2 by PRMT6 and H3K4 by an MLL complex are mutually exclusive. *Nature* **449**, 933–937 (2007).
25. Yuan, C.-C. *et al.* Histone H3R2 symmetric dimethylation and histone H3K4 trimethylation are tightly correlated in eukaryotic genomes. *Cell Rep* **1**, 83–90 (2012).
26. Wang, H. *et al.* Purification and functional characterization of a histone H3-lysine 4-specific methyltransferase. *Mol Cell* **8**, 1207–1217 (2001).

27. Delaval, K. *et al.* Differential histone modifications mark mouse imprinting control regions during spermatogenesis. *The EMBO Journal* **26**, 720–729 (2007).
28. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
29. Regha, K. *et al.* Active and Repressive Chromatin Are Interspersed without Spreading in an Imprinted Gene Cluster in the Mammalian Genome. *Molecular Cell* **27**, 353–366 (2007).
30. Eswaran, J. *et al.* Structure and functional characterization of the atypical human kinase haspin. *Proc Natl Acad Sci U S A* **106**, 20198–20203 (2009).
31. Southall, S. M., Wong, P.-S., Odho, Z., Roe, S. M. & Wilson, J. R. Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. *Mol. Cell* **33**, 181–191 (2009).
32. Rougeulle, C. *et al.* Differential histone H3 Lys-9 and Lys-27 methylation profiles on the X chromosome. *Mol Cell Biol* **24**, 5475–5484 (2004).
33. Feldman, N. *et al.* G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat Cell Biol* **8**, 188–194 (2006).
34. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49–55 (2008).
35. Viré, E. *et al.* The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, 871–874 (2006).
36. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nature Reviews Genetics* **10**, 295–304 (2009).
37. Jenuwein, T., Laible, G., Dorn, R. & Reuter, G. SET domain proteins modulate chromatin domains in eu- and heterochromatin. *Cell Mol Life Sci* **54**, 80–93 (1998).
38. Dillon, S. C., Zhang, X., Trievel, R. C. & Cheng, X. The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol.* **6**, 227 (2005).
39. Petrossian, T. C. & Clarke, S. G. Uncovering the human methyltransferasome. *Mol Cell Proteomics* **10**, M110.000976 (2011).
40. Wu, H. *et al.* Structural biology of human H3K9 methyltransferases. *PLoS ONE* **5**, e8570 (2010).
41. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
42. Nwasike, C., Ewert, S., Jovanovic, S., Haider, S. & Mujtaba, S. SET domain-mediated lysine methylation in lower organisms regulates growth and transcription in hosts. *Ann NY Acad Sci* **1376**, 18–28 (2016).
43. Tachibana, M., Sugimoto, K., Fukushima, T. & Shinkai, Y. Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *J Biol Chem* **276**, 25309–25317 (2001).
44. Casciello, F., Windloch, K., Gannon, F. & Lee, J. S. Functional Role of G9a Histone Methyltransferase in Cancer. *Front Immunol* **6**, 487 (2015).
45. Yun, M., Wu, J., Workman, J. L. & Li, B. Readers of histone modifications. *Cell Res* **21**, 564–578 (2011).
46. Chin, H. G. *et al.* Automethylation of G9a and its implication in wider substrate specificity and HP1 binding. *Nucleic Acids Res* **35**, 7313–7323 (2007).
47. Sampath, S. C. *et al.* Methylation of a histone mimic within the histone methyltransferase G9a regulates protein complex assembly. *Mol. Cell* **27**, 596–608 (2007).
48. Rathert, P. *et al.* Protein lysine methyltransferase G9a acts on non-histone targets. *Nat. Chem. Biol.* **4**, 344–346 (2008).
49. Chang, Y. *et al.* MPP8 mediates the interactions between DNA methyltransferase Dnmt3a and H3K9 methyltransferase GLP/G9a. *Nat Commun* **2**, 533 (2011).
50. Huang, J. *et al.* G9a and Glp methylate lysine 373 in the tumor suppressor p53. *J Biol Chem* **285**, 9636–9641 (2010).
51. Oda, H. *et al.* Monomethylation of histone H4-lysine 20 is involved in chromosome structure and stability and is essential for mouse development. *Mol. Cell. Biol.* **29**, 2278–2295 (2009).

52. Schotta, G. *et al.* A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev.* **18**, 1251–1262 (2004).
53. Schotta, G. *et al.* A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. *Genes Dev.* **22**, 2048–2061 (2008).
54. Milite, C. *et al.* The emerging role of lysine methyltransferase SETD8 in human diseases. *Clin Epigenetics* **8**, 102 (2016).
55. Sims, J. K. & Rice, J. C. PR-Set7 establishes a repressive trans-tail histone code that regulates differentiation. *Mol Cell Biol* **28**, 4459–4468 (2008).
56. Congdon, L. M., Houston, S. I., Veerappan, C. S., Spektor, T. M. & Rice, J. C. PR-Set7-mediated monomethylation of histone H4 lysine 20 at specific genomic regions induces transcriptional repression. *J Cell Biochem* **110**, 609–619 (2010).
57. Yin, Y., Yu, V. C., Zhu, G. & Chang, D. C. SET8 plays a role in controlling G1/S transition by blocking lysine acetylation in histone through binding to H4 N-terminal tail. *Cell Cycle* **7**, 1423–1432 (2008).
58. Botuyan, M. V. *et al.* Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell* **127**, 1361–1373 (2006).
59. Dulev, S., Tkach, J., Lin, S. & Batada, N. N. SET8 methyltransferase activity during the DNA double-strand break response is required for recruitment of 53BP1. *EMBO Rep.* **15**, 1163–1174 (2014).
60. Beck, D. B., Oda, H., Shen, S. S. & Reinberg, D. PR-Set7 and H4K20me1: at the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription. *Genes Dev.* **26**, 325–337 (2012).
61. Takawa, M. *et al.* Histone lysine methyltransferase SETD8 promotes carcinogenesis by deregulating PCNA expression. *Cancer Res.* **72**, 3217–3227 (2012).
62. Shi, X. *et al.* Modulation of p53 function by SET8-mediated methylation at lysine 382. *Mol. Cell* **27**, 636–646 (2007).
63. Huen, M. S. Y., Sy, S. M.-H., van Deursen, J. M. & Chen, J. Direct interaction between SET8 and proliferating cell nuclear antigen couples H4-K20 methylation with DNA replication. *J Biol Chem* **283**, 11073–11077 (2008).
64. Rice, J. C. *et al.* Mitotic-specific methylation of histone H4 Lys 20 follows increased PR-Set7 expression and its localization to mitotic chromosomes. *Genes Dev* **16**, 2225–2230 (2002).
65. Tardat, M., Murr, R., Herceg, Z., Sardet, C. & Julien, E. PR-Set7-dependent lysine methylation ensures genome replication and stability through S phase. *J Cell Biol* **179**, 1413–1426 (2007).
66. McKay, D. J. *et al.* Interrogating the function of metazoan histones using engineered gene clusters. *Dev. Cell* **32**, 373–386 (2015).
67. Karachentsev, D., Sarma, K., Reinberg, D. & Steward, R. PR-Set7-dependent methylation of histone H4 Lys 20 functions in repression of gene expression and is essential for mitosis. *Genes Dev.* **19**, 431–435 (2005).
68. Oda, H. *et al.* Regulation of the histone H4 monomethylase PR-Set7 by CRL4(Cdt2)-mediated PCNA-dependent degradation during DNA damage. *Mol. Cell* **40**, 364–376 (2010).
69. Liu, W. *et al.* PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. *Nature* **466**, 508–512 (2010).
70. Qi, H. H. *et al.* Histone H4K20/H3K9 demethylase PHF8 regulates zebrafish brain and craniofacial development. *Nature* **466**, 503–507 (2010).
71. Wang, J. *et al.* LSD1n is an H4K20 demethylase regulating memory formation via transcriptional elongation control. *Nat Neurosci* **18**, 1256–1264 (2015).
72. Stender, J. D. *et al.* Control of proinflammatory gene programs by regulated trimethylation and demethylation of histone H4K20. *Mol Cell* **48**, 28–38 (2012).
73. Min, J. *et al.* L3MBTL1 recognition of mono- and dimethylated histones. *Nat Struct Mol Biol* **14**, 1229–1230 (2007).
74. Guo, Y. *et al.* Methylation-state-specific recognition of histones by the MBT repeat protein L3MBTL2. *Nucleic Acids Res* **37**, 2204–2210 (2009).
75. Kim, D. *et al.* Corecognition of DNA and a methylated histone tail by the MSL3 chromodomain. *Nat Struct Mol Biol* **17**, 1027–1029 (2010).

76. Lee, J., Thompson, J. R., Botuyan, M. V. & Mer, G. Distinct binding modes specify the recognition of methylated histones H3K4 and H4K20 by JMJD2A-tudor. *Nat Struct Mol Biol* **15**, 109–111 (2008).
77. Mallette, F. A. *et al.* RNF8- and RNF168-dependent degradation of KDM4A/JMJD2A triggers 53BP1 recruitment to DNA damage sites. *EMBO J* **31**, 1865–1878 (2012).
78. Kuo, A. J. *et al.* The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. *Nature* **484**, 115–119 (2012).
79. Hirano, Y. *et al.* Lamin B receptor recognizes specific modifications of histone H4 in heterochromatin formation. *J Biol Chem* **287**, 42654–42663 (2012).
80. Kim, J. *et al.* Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep* **7**, 397–403 (2006).
81. Beck, D. B. *et al.* The role of PR-Set7 in replication licensing depends on Suv4-20h. *Genes Dev.* **26**, 2580–2589 (2012).
82. Dhami, G. K. *et al.* Dynamic methylation of Numb by Set8 regulates its binding to p53 and apoptosis. *Mol. Cell* **50**, 565–576 (2013).
83. Weirich, S., Kusevic, D., Kudithipudi, S. & Jeltsch, A. Investigation of the methylation of Numb by the SET8 protein lysine methyltransferase. *Sci Rep* **5**, 13813 (2015).
84. Kudithipudi, S., Dhayalan, A., Kebede, A. F. & Jeltsch, A. The SET8 H4K20 protein lysine methyltransferase has a long recognition sequence covering seven amino acid residues. *Biochimie* **94**, 2212–2218 (2012).
85. Ong, S.-E., Mittler, G. & Mann, M. Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat Methods* **1**, 119–126 (2004).
86. Moore, K. E. & Gozani, O. An Unexpected Journey: Lysine Methylation Across the Proteome. *Biochim Biophys Acta* **1839**, 1395–1403 (2014).
87. Lukinović, V., Casanova, A. G., Roth, G. S., Chuffart, F. & Reynoird, N. Lysine Methyltransferases Signaling: Histones are Just the Tip of the Iceberg. *Curr. Protein Pept. Sci.* (2020) doi:10.2174/1871527319666200102101608.
88. Huang, J. *et al.* Repression of p53 activity by Smyd2-mediated methylation. *Nature* **444**, 629–632 (2006).
89. Chuikov, S. *et al.* Regulation of p53 activity through lysine methylation. *Nature* **432**, 353–360 (2004).
90. Zhang, J. *et al.* G9a stimulates CRC growth by inducing p53 Lys373 dimethylation-dependent activation of Plk1. *Theranostics* **8**, 2884–2895 (2018).
91. West, L. E. *et al.* The MBT repeats of L3MBTL1 link SET8-mediated p53 methylation at lysine 382 to target gene repression. *J Biol Chem* **285**, 37725–37732 (2010).
92. Kachirskaia, I. *et al.* Role for 53BP1 Tudor Domain Recognition of p53 Dimethylated at Lysine 382 in DNA Damage Signaling. *J Biol Chem* **283**, 34660–34666 (2008).
93. Huang, J. *et al.* p53 is regulated by the lysine demethylase LSD1. *Nature* **449**, 105–108 (2007).
94. Mauser, R. & Jeltsch, A. Application of modified histone peptide arrays in chromatin research. *Arch. Biochem. Biophys.* **661**, 31–38 (2019).
95. Dhayalan, A., Kudithipudi, S., Rathert, P. & Jeltsch, A. Specificity analysis-based identification of new methylation targets of the SET7/9 protein lysine methyltransferase. *Chem. Biol.* **18**, 111–120 (2011).
96. Rathert, P., Zhang, X., Freund, C., Cheng, X. & Jeltsch, A. Analysis of the substrate specificity of the Dim-5 histone lysine methyltransferase using peptide arrays. *Chem Biol* **15**, 5–11 (2008).
97. Zhang, Y. *et al.* Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol Cell Proteomics* **4**, 1240–1250 (2005).
98. Kim, W. *et al.* Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol Cell* **44**, 325–340 (2011).
99. Choudhary, C. *et al.* Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834–840 (2009).

100. Levy, D. *et al.* A proteomic approach for the identification of novel lysine methyltransferase substrates. *Epigenetics & Chromatin* **4**, 19 (2011).
101. Cao, X.-J., Arnaudo, A. M. & Garcia, B. A. Large-scale global identification of protein lysine methylation in vivo. *Epigenetics* **8**, 477–485 (2013).
102. Liu, H. *et al.* A method for systematic mapping of protein lysine methylation identifies new functions for HP1 β in DNA damage repair. *Mol Cell* **50**, 723–735 (2013).
103. Moore, K. E. *et al.* A general molecular affinity strategy for global detection and proteomic analysis of lysine methylation. *Mol Cell* **50**, 444–456 (2013).
104. Bremang, M. *et al.* Mass spectrometry-based identification and characterisation of lysine and arginine methylation in the human proteome. *Mol Biosyst* **9**, 2231–2247 (2013).
105. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew Chem Int Ed Engl* **40**, 2004–2021 (2001).
106. Prescher, J. A. & Bertozzi, C. R. Chemistry in living systems. *Nat Chem Biol* **1**, 13–21 (2005).
107. Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. A stepwise Huisgen cycloaddition process: copper(I)-catalyzed regioselective ‘ligation’ of azides and terminal alkynes. *Angew Chem Int Ed Engl* **41**, 2596–2599 (2002).
108. Tornøe, C. W., Christensen, C. & Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J Org Chem* **67**, 3057–3064 (2002).
109. Lewis, W. G. *et al.* Click chemistry in situ: acetylcholinesterase as a reaction vessel for the selective assembly of a femtomolar inhibitor from an array of building blocks. *Angew Chem Int Ed Engl* **41**, 1053–1057 (2002).
110. Agard, N. J., Prescher, J. A. & Bertozzi, C. R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J Am Chem Soc* **126**, 15046–15047 (2004).
111. Dalhoff, C., Lukinavicius, G., Klimasauskas, S. & Weinhold, E. Direct transfer of extended groups from synthetic cofactors by DNA methyltransferases. *Nat. Chem. Biol.* **2**, 31–32 (2006).
112. Islam, K. *et al.* Defining efficient enzyme-cofactor pairs for bioorthogonal profiling of protein methylation. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 16778–16783 (2013).
113. Wang, R. *et al.* Profiling genome-wide chromatin methylation with engineered posttranslation apparatus within living cells. *J. Am. Chem. Soc.* **135**, 1048–1056 (2013).
114. Cantoni, G. L. Biological Methylation: Selected Aspects. *Annual Review of Biochemistry* **44**, 435–451 (1975).
115. Markham, G. D., Hafner, E. W., Tabor, C. W. & Tabor, H. S-Adenosylmethionine synthetase from *Escherichia coli*. *J Biol Chem* **255**, 9082–9092 (1980).
116. Borchardt, R. T., Shiong, Y., Huber, J. A. & Wycpalek, A. F. Potential inhibitors of S-adenosylmethionine-dependent methyltransferases. 6. Structural modifications of S-adenosylmethionine. *J Med Chem* **19**, 1104–1110 (1976).
117. Fontecave, M., Atta, M. & Mulliez, E. S-adenosylmethionine: nothing goes to waste. *Trends Biochem Sci* **29**, 243–249 (2004).
118. Ho, D. K., Wu, J. C., Santi, D. V. & Floss, H. G. Stereochemical studies of the C-methylation of deoxycytidine catalyzed by HhaI methylase and the N-methylation of deoxyadenosine catalyzed by EcoRI methylase. *Arch Biochem Biophys* **284**, 264–269 (1991).
119. Drennan, C. L., Matthews, R. G. & Ludwig, M. L. Cobalamin-dependent methionine synthase: the structure of a methylcobalamin-binding fragment and implications for other B12-dependent enzymes. *Curr Opin Struct Biol* **4**, 919–929 (1994).
120. Parks, L. W. S-Adenosylethionine and ethionine inhibition. *J Biol Chem* **232**, 169–176 (1958).
121. Schlenk, F. & Dainko, J. L. The S-n-propyl analogue of S-adenosylmethionine. *Biochim Biophys Acta* **385**, 312–323 (1975).
122. Thomsen, M., Vogensen, S. B., Buchardt, J., Burkart, M. D. & Clausen, R. P. Chemoenzymatic synthesis and in situ application of S-adenosyl-L-methionine analogs. *Org Biomol Chem* **11**, 7606–7610 (2013).

123. Peters, W. *et al.* Enzymatic site-specific functionalization of protein methyltransferase substrates with alkynes for click labeling. *Angew Chem Int Ed Engl* **49**, 5170–5173 (2010).
124. Bishop, A. C., Buzko, O. & Shokat, K. M. Magic bullets for protein kinases. *Trends Cell Biol.* **11**, 167–172 (2001).
125. Lin, Q., Jiang, F., Schultz, P. G. & Gray, N. S. Design of allele-specific protein methyltransferase inhibitors. *J. Am. Chem. Soc.* **123**, 11608–11613 (2001).
126. Lukinavičius, G., Tomkuvienė, M., Masevičius, V. & Klimašauskas, S. Enhanced Chemical Stability of AdoMet Analogues for Improved Methyltransferase-Directed Labeling of DNA. *ACS Chem. Biol.* **8**, 1134–1139 (2013).
127. Lukinavičius, G. *et al.* Targeted labeling of DNA by methyltransferase-directed transfer of activated groups (mTAG). *J. Am. Chem. Soc.* **129**, 2758–2759 (2007).
128. Mickutė, M. *et al.* Animal Hen1 2'-O-methyltransferases as tools for 3'-terminal functionalization and labelling of single-stranded RNAs. *Nucleic Acids Research* **46**, e104–e104 (2018).
129. Long, Y., Ubych, K., Jagu, E. & Neely, R. K. FRET-Based Method for Direct, Real-Time Measurement of DNA Methyltransferase Activity. *Bioconjugate Chem.* **32**, 192–198 (2021).
130. Ovcharenko, A., Weissenboeck, F. P. & Rentmeister, A. Tag-Free Internal RNA Labeling and Photocaging Based on mRNA Methyltransferases. *Angewandte Chemie International Edition* **60**, 4098–4103 (2021).
131. Anhäuser, L., Muttach, F. & Rentmeister, A. Reversible modification of DNA by methyltransferase-catalyzed transfer and light-triggered removal of photo-caging groups. *Chem. Commun. (Camb.)* **54**, 449–451 (2018).
132. Muttach, F., Mäsing, F., Studer, A. & Rentmeister, A. New AdoMet Analogues as Tools for Enzymatic Transfer of Photo-Cross-Linkers and Capturing RNA-Protein Interactions. *Chemistry* **23**, 5988–5993 (2017).
133. Michailidou, F. *et al.* Engineered SAM Synthetases for Enzymatic Generation of AdoMet Analogs with Photocaging Groups and Reversible DNA Modification in Cascade Reactions. *Angewandte Chemie International Edition* **60**, 480–485 (2021).
134. Masevičius, V., Nainytė, M. & Klimašauskas, S. Synthesis of S-Adenosyl-L-Methionine Analogs with Extended Transferable Groups for Methyltransferase-Directed Labeling of DNA and RNA. *Curr Protoc Nucleic Acid Chem* **64**, 1.36.1-1.36.13 (2016).
135. Singh, S. *et al.* Facile chemoenzymatic strategies for the synthesis and utilization of S-adenosyl-(L)-methionine analogues. *Angew Chem Int Ed Engl* **53**, 3965–3969 (2014).
136. Wang, F. *et al.* Understanding molecular recognition of promiscuity of thermophilic methionine adenosyltransferase sMAT from *Sulfolobus solfataricus*. *FEBS J* **281**, 4224–4239 (2014).
137. Dippe, M. *et al.* Rationally engineered variants of S-adenosylmethionine (SAM) synthase: reduced product inhibition and synthesis of artificial cofactor homologues. *Chem. Commun. (Camb.)* **51**, 3637–3640 (2015).
138. Eustáquio, A. S., Pojer, F., Noel, J. P. & Moore, B. S. Discovery and characterization of a marine bacterial SAM-dependent chlorinase. *Nat Chem Biol* **4**, 69–74 (2008).
139. Zhu, X., Robinson, D. A., McEwan, A. R., O'Hagan, D. & Naismith, J. H. Mechanism of Enzymatic Fluorination in *Streptomyces cattleya*. *J. Am. Chem. Soc.* **129**, 14597–14604 (2007).
140. Deng, H. *et al.* The Fluorinase from *Streptomyces cattleya* Is Also a Chlorinase. *Angewandte Chemie International Edition* **45**, 759–762 (2006).
141. Deng, H. & O'Hagan, D. The fluorinase, the chlorinase and the duf-62 enzymes. *Current Opinion in Chemical Biology* **12**, 582–592 (2008).
142. Tang, Q. *et al.* Directed Evolution of a Halide Methyltransferase Enables Biocatalytic Synthesis of Diverse SAM Analogs. *Angewandte Chemie International Edition* **60**, 1524–1527 (2021).
143. Hoffman, J. L. Chromatographic analysis of the chiral and covalent instability of S-adenosyl-L-methionine. *Biochemistry* **25**, 4444–4449 (1986).

144. Iwig, D. F. & Booker, S. J. Insight into the polar reactivity of the onium chalcogen analogues of S-adenosyl-L-methionine. *Biochemistry* **43**, 13496–13509 (2004).
145. Holstein, J. M., Stummer, D. & Rentmeister, A. Enzymatic modification of 5'-capped RNA with a 4-vinylbenzyl group provides a platform for photoclick and inverse electron-demand Diels–Alder reaction. *Chem. Sci.* **6**, 1362–1369 (2015).
146. Zhang, Y. *et al.* In vivo protein allylation to capture protein methylation candidates. *Chem. Commun.* **52**, 6689–6692 (2016).
147. Muttach, F. & Rentmeister, A. A Biocatalytic Cascade for Versatile One-Pot Modification of mRNA Starting from Methionine Analogues. *Angew Chem Int Ed Engl* **55**, 1917–1920 (2016).
148. Wang, R., Zheng, W., Yu, H., Deng, H. & Luo, M. Labeling Substrates of Protein Arginine Methyltransferase with Engineered Enzymes and Matched S-Adenosyl-L-methionine Analogues. *J. Am. Chem. Soc.* **133**, 7648–7651 (2011).
149. Goeta, A., Salter, M. M. & Shah, H. New indium-mediated cyclisation reactions of tethered haloynes in aqueous solvent systems. *Tetrahedron* **62**, 3582–3599 (2006).
150. Xiao, B. *et al.* Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature* **421**, 652–656 (2003).
151. Guitot, K. *et al.* Label-free measurement of histone lysine methyltransferases activity by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal. Biochem.* **456**, 25–31 (2014).
152. Guitot, K. *et al.* A direct label-free MALDI-TOF mass spectrometry based assay for the characterization of inhibitors of protein lysine methyltransferases. *Anal Bioanal Chem* **409**, 3767–3777 (2017).
153. Dyer, P. N. *et al.* Reconstitution of nucleosome core particles from recombinant histones and DNA. *Meth. Enzymol.* **375**, 23–44 (2004).
154. Islam, K., Zheng, W., Yu, H., Deng, H. & Luo, M. Expanding Cofactor Repertoire of Protein Lysine Methyltransferase for Substrate Labeling. *ACS Chem. Biol.* **6**, 679–684 (2011).
155. Blum, G., Bothwell, I. R., Islam, K. & Luo, M. Profiling Protein Methylation with Cofactor Analogue Containing Terminal Alkyne Functionality. *Curr Protoc Chem Biol* **5**, 67–88 (2013).
156. Blum, G., Islam, K. & Luo, M. Using Azido Analogue of S-Adenosyl-L-methionine for Bioorthogonal Profiling of Protein Methylation (BPPM). *Curr Protoc Chem Biol* **5**, 45–66 (2013).
157. van Geel, R., Pruijn, G. J. M., van Delft, F. L. & Boelens, W. C. Preventing thiol-yne addition improves the specificity of strain-promoted azide-alkyne cycloaddition. *Bioconjug Chem* **23**, 392–398 (2012).
158. Bothwell, I. R. *et al.* Se-adenosyl-L-selenomethionine cofactor analogue as a reporter of protein methylation. *J. Am. Chem. Soc.* **134**, 14905–14912 (2012).
159. Iwig, D. F., Grippe, A. T., McIntyre, T. A. & Booker, S. J. Isotope and elemental effects indicate a rate-limiting methyl transfer as the initial step in the reaction catalyzed by Escherichia coli cyclopropane fatty acid synthase. *Biochemistry* **43**, 13510–13524 (2004).
160. Bothwell, I. R. & Luo, M. Large-scale, protection-free synthesis of Se-adenosyl-L-selenomethionine analogues and their application as cofactor surrogates of methyltransferases. *Org. Lett.* **16**, 3056–3059 (2014).
161. Ohata, J., Vohidov, F. & Ball, Z. T. Convenient analysis of protein modification by chemical blotting with fluorogenic 'click' reagents. *Mol Biosyst* **11**, 2846–2849 (2015).
162. Luque-Garcia, J. L., Zhou, G., Sun, T.-T. & Neubert, T. A. Use of nitrocellulose membranes for protein characterization by matrix-assisted laser desorption/ionization mass spectrometry. *Anal. Chem.* **78**, 5102–5108 (2006).
163. Poulet, P., Carpentier, S. & Barillot, E. myProMS, a web server for management and validation of mass spectrometry-based proteomic data. *Proteomics* **7**, 2553–2556 (2007).
164. Valot, B., Langella, O., Nano, E. & Zivy, M. MassChroQ: a versatile tool for mass spectrometry quantification. *Proteomics* **11**, 3572–3577 (2011).
165. Mellacheruvu, D. *et al.* The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat. Methods* **10**, 730–736 (2013).

166. Stein, B., Brady, H., Yang, M. X., Young, D. B. & Barbosa, M. S. Cloning and characterization of MEK6, a novel member of the mitogen-activated protein kinase kinase cascade. *J Biol Chem* **271**, 11427–11433 (1996).
167. Sturchler, E., Feurstein, D., McDonald, P. & Duckett, D. Mechanism of oxidative stress-induced ASK1-catalyzed MKK6 phosphorylation. *Biochemistry* **49**, 4094–4102 (2010).
168. Matsumoto, T. *et al.* Crystal structure of non-phosphorylated MAP2K6 in a putative auto-inhibition state. *J Biochem* **151**, 541–549 (2012).
169. Olsen, J. B. *et al.* Quantitative Profiling of the Activity of Protein Lysine Methyltransferase SMYD2 Using SILAC-Based Proteomics*. *Molecular & Cellular Proteomics* **15**, 892–905 (2016).
170. Schönwasser, D. C., Marais, R. M., Marshall, C. J. & Parker, P. J. Activation of the mitogen-activated protein kinase/extracellular signal-regulated kinase pathway by conventional, novel, and atypical protein kinase C isoforms. *Mol Cell Biol* **18**, 790–798 (1998).
171. Monick, M. M., Carter, A. B., Flaherty, D. M., Peterson, M. W. & Hunninghake, G. W. Protein kinase C zeta plays a central role in activation of the p42/44 mitogen-activated protein kinase by endotoxin in alveolar macrophages. *J Immunol* **165**, 4632–4639 (2000).
172. Levy, D. *et al.* SETD6 lysine methylation of RelA couples GLP activity at chromatin to tonic repression of NF- κ B signaling. *Nat Immunol* **12**, 29–36 (2011).
173. Duran, A., Diaz-Meco, M. T. & Moscat, J. Essential role of RelA Ser311 phosphorylation by ζ PKC in NF- κ B transcriptional activation. *EMBO J* **22**, 3910–3918 (2003).
174. Baumli, S. *et al.* The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation. *EMBO J* **27**, 1907–1918 (2008).
175. Wen, Y. *et al.* RPTP α phosphatase activity is allosterically regulated by the membrane-distal catalytic domain. *J Biol Chem* **295**, 4923–4936 (2020).
176. Sun, G., Cheng, S. Y. S., Chen, M., Lim, C. J. & Pallen, C. J. Protein tyrosine phosphatase α phosphotyrosyl-789 binds BCAR3 to position Cas for activation at integrin-mediated focal adhesions. *Mol Cell Biol* **32**, 3776–3789 (2012).
177. Mazur, P. K. *et al.* SMYD3 links lysine methylation of MAP3K2 to Ras-driven cancer. *Nature* **510**, 283–287 (2014).
178. Reynoird, N. *et al.* Coordination of stress signals by the lysine methyltransferase SMYD2 promotes pancreatic cancer. *Genes Dev* **30**, 772–785 (2016).
179. Yoshioka, Y. *et al.* SMYD3-mediated lysine methylation in the PH domain is critical for activation of AKT1. *Oncotarget* **7**, 75023–75037 (2016).
180. Huang, B. & Porter, G. Expression of proline-rich Akt-substrate PRAS40 in cell survival pathway and carcinogenesis. *Acta Pharmacol Sin* **26**, 1253–1258 (2005).
181. Sims, R. J. & Reinberg, D. Is there a code embedded in proteins that is based on post-translational modifications? *Nat Rev Mol Cell Biol* **9**, 815–820 (2008).
182. Binda, O. *et al.* A Chemical Method for Labeling Lysine Methyltransferase Substrates. *ChemBioChem* **12**, 330–334 (2011).
183. Willnow, S., Martin, M., Lüscher, B. & Weinhold, E. A selenium-based click AdoMet analogue for versatile substrate labeling with wild-type protein methyltransferases. *Chembiochem* **13**, 1167–1173 (2012).
184. Wu, G., Robertson, D. H., Brooks, C. L. & Vieth, M. Detailed analysis of grid-based molecular docking: A case study of CDOCKER-A CHARMM-based MD docking algorithm. *J Comput Chem* **24**, 1549–1562 (2003).
185. Wu, H. *et al.* Structure of the catalytic domain of EZH2 reveals conformational plasticity in cofactor and substrate binding sites and explains oncogenic mutations. *PLoS ONE* **8**, e83737 (2013).
186. McCabe, M. T. *et al.* Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). *Proc Natl Acad Sci U S A* **109**, 2989–2994 (2012).
187. Hanack, M., Dehesch, T., Hummel, K. & Nierth, A. Cyclobutanone via Solvolytic Cyclization. (2003) doi:10.1002/0471264180.OS054.19.

188. Islam, K. *et al.* Bioorthogonal profiling of protein methylation using azido derivative of S-adenosyl-L-methionine. *J Am Chem Soc* **134**, 5909–5915 (2012).
189. Jamieson, A. G. *et al.* Positional Scanning for Peptide Secondary Structure by Systematic Solid-Phase Synthesis of Amino Lactam Peptides. *J. Am. Chem. Soc.* **131**, 7917–7927 (2009).
190. Davidson, D. & Bernhard, S. A. The Structure of Meldrum's Supposed β -Lactonic Acid. *J. Am. Chem. Soc.* **70**, 3426–3428 (1948).
191. Luger, K., Rechsteiner, T. J., Flaus, A. J., Waye, M. M. & Richmond, T. J. Characterization of nucleosome core particles containing histone proteins made in bacteria. *J. Mol. Biol.* **272**, 301–311 (1997).
192. Ciccotti, G. & Ryckaert, J. P. Molecular dynamics simulation of rigid molecules. *Computer Physics Reports* **4**, 346–392 (1986).

La régulation de la méthylation des lysines par les histone lysine méthyltransférases joue un rôle important dans le fonctionnement normal de la cellule et une dérégulation de la méthylation est régulièrement associée à différentes pathologies telles que des cancers ou des maladies neurodégénératives. Historiquement, les histones ont été identifiées comme les cibles principales des HKMT, la méthylation de leur queue d'histone étant directement liée à l'état de compaction de la chromatine et le niveau d'expression des gènes. Depuis plusieurs années, un nombre croissant de substrats non-histones ont été découverts pour ces méthyltransférases faisant apparaître de nouvelles fonctions biologiques pour ces enzymes. Alors que les HKMT deviennent des cibles thérapeutiques intéressantes, il devient crucial de caractériser leur spécificité de substrat.

Dans ce projet, nous avons exploré la capacité de mutants de PR-Set7 à accommoder des analogues du cofacteur naturel de l'enzyme : la SAM, dans le but de leur faire transférer un groupement chimique bioconjugué. L'introduction par voie enzymatique d'une étiquette porteuse d'un groupement alcyne ou azoture permet alors une seconde étape de fonctionnalisation par chimie click d'une sonde biotine. Cette fonctionnalisation en deux étapes permet, *in fine*, un enrichissement des cibles de PR-Set7. Lors de cette thèse, une petite librairie d'analogues de SAM a été synthétisée et des mutants ponctuels de PR-Set7 ont été conçus et produits dans un système bactérien. Le criblage des molécules de la librairie a mis en évidence la prise en charge par un des mutants de PR-Set7 du plus petit des analogues de cofacteur produit : ProSeAM. Une étude du méthylome de PR-Set7 a été menée aboutissant à l'identification de 23 potentiels substrats de PR-Set7 qu'il reste à valider biologiquement. Enfin une étude de modélisation moléculaire a été effectuée permettant d'expliquer la prise en charge de ce cofacteur par PR-Set7 et de proposer de nouveaux mutants de l'enzyme apte à accommoder de nouveaux analogues de SAM plus encombrants.

Mots clés : HKMT, PR-Set7, Modifications post-traductionnelles, Chimie click, Protéomique, Docking

Study of the methylome of PR-Set7 lysine methyltransferase with enzyme and chemically modified cofactor pairs

The regulation of histone lysine methylation by methyltransferases (HKMTs) plays an important role in the basic cell functions and a deregulation is often associated with pathologies like cancers or neurodegenerative diseases. Historically, histone proteins have been identified as the major targets of these HKMTs and the methylation of their N-terminal tails have been associated to the regulation of chromatin compaction and gene transcription. However, for several years, an increasing number of HKMTs have shown their ability to methylate non-histone proteins. While the potential of these enzymes grows as interesting therapeutic targets, it becomes crucial to characterize their substrate specificity.

In this project we explored the ability of PR-Set7 mutants to accommodate various SAM analogs. The enzymatic transfer of an alkyne or an azide tag on PR-Set7 substrates followed by a bioorthogonal functionalization with a biotin probe leads to the enrichment of PR-Set7 targets. During this thesis, a small library of SAM analogs have been produced and several PR-Set7 mutants have been designed and expressed in bacteria. The screening of the SAM analogs library demonstrated the ability of one of PR-Set7 mutant to accommodate the smallest synthesized SAM analog: ProSeAM. A proteomics study using this molecule allowed the characterization of 23 potential substrates of PR-Set7 which remain to be biologically validated. Finally, a molecular modeling study was carried out to rationalize how PR-Set7 accommodate specifically this analog. This study also allowed to define new conditions to accommodate bulkier SAM analogs.

Keywords : HKMT, PR-Set7, Post-translational modification , Click chemistry, Proteomics, Docking

La régulation de la méthylation des lysines par les histone lysine méthyltransférases joue un rôle important dans le fonctionnement normal de la cellule et une dérégulation de la méthylation est régulièrement associée à différentes pathologies telles que des cancers ou des maladies neurodégénératives. Historiquement, les histones ont été identifiées comme les cibles principales des HKMT, la méthylation de leur queue d'histone étant directement liée à l'état de compaction de la chromatine et le niveau d'expression des gènes. Depuis plusieurs années, un nombre croissant de substrats non-histones ont été découverts pour ces méthyltransférases faisant apparaître de nouvelles fonctions biologiques pour ces enzymes. Alors que les HKMT deviennent des cibles thérapeutiques intéressantes, il devient crucial de caractériser leur spécificité de substrat.

Dans ce projet, nous avons exploré la capacité de mutants de PR-Set7 à accommoder des analogues du cofacteur naturel de l'enzyme : la SAM, dans le but de leur faire transférer un groupement chimique bioconjugué. L'introduction par voie enzymatique d'une étiquette porteuse d'un groupement alcyne ou azoture permet alors une seconde étape de fonctionnalisation par chimie click d'une sonde biotine. Cette fonctionnalisation en deux étapes permet, *in fine*, un enrichissement des cibles de PR-Set7. Lors de cette thèse, une petite librairie d'analogues de SAM a été synthétisée et des mutants ponctuels de PR-Set7 ont été conçus et produits dans un système bactérien. Le criblage des molécules de la librairie a mis en évidence la prise en charge par un des mutants de PR-Set7 du plus petit des analogues de cofacteur produit : ProSeAM. Une étude du méthylome de PR-Set7 a été menée aboutissant à l'identification de 23 potentiels substrats de PR-Set7 qu'il reste à valider biologiquement. Enfin une étude de modélisation moléculaire a été effectuée permettant d'expliquer la prise en charge de ce cofacteur par PR-Set7 et de proposer de nouveaux mutants de l'enzyme apte à accommoder de nouveaux analogues de SAM plus encombrants.

Mots clés : HKMT, PR-Set7, Modifications post-traductionnelles, Chimie click, Protéomique, Docking

Study of the methylome of PR-Set7 lysine methyltransferase with enzyme and chemically modified cofactor pairs

The regulation of histone lysine methylation by methyltransferases (HKMTs) plays an important role in the basic cell functions and a deregulation is often associated with pathologies like cancers or neurodegenerative diseases. Historically, histone proteins have been identified as the major targets of these HKMTs and the methylation of their N-terminal tails have been associated to the regulation of chromatin compaction and gene transcription. However, for several years, an increasing number of HKMTs have shown their ability to methylate non-histone proteins. While the potential of these enzymes grows as interesting therapeutic targets, it becomes crucial to characterize their substrate specificity.

In this project we explored the ability of PR-Set7 mutants to accommodate various SAM analogs. The enzymatic transfer of an alkyne or an azide tag on PR-Set7 substrates followed by a bioorthogonal functionalization with a biotin probe leads to the enrichment of PR-Set7 targets. During this thesis, a small library of SAM analogs have been produced and several PR-Set7 mutants have been designed and expressed in bacteria. The screening of the SAM analogs library demonstrated the ability of one of PR-Set7 mutant to accommodate the smallest synthesized SAM analog: ProSeAM. A proteomics study using this molecule allowed the characterization of 23 potential substrates of PR-Set7 which remain to be biologically validated. Finally, a molecular modeling study was carried out to rationalize how PR-Set7 accommodate specifically this analog. This study also allowed to define new conditions to accommodate bulkier SAM analogs.

Keywords : HKMT, PR-Set7, Post-translational modification , Click chemistry, Proteomics, Docking