Institutional Repository - Research Portal
Dépôt Institutionnel - Portail de la Recherche
researchportal.unamur.be
University of Namur

THESIS / THÈSE

MASTER EN SCIENCES CHIMIQUES

Etude structurale du domaine PWWP des DNMT3A et DNMT3B

Dal Maso, Thomas

Award date:
2015

Awarding institution:
Universite de Namur

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 03. Jul. 2025

UNIVERSITE DE NAMUR

Faculté des Sciences

Etude structurale du domaine PWWP des DNMT3A et DNMT3B

Mémoire présenté pour l'obtention

du grade académique de Master Chimie «Chimie du Vivant et des Nanomatériaux» : Finalité Approfondie

Thomas Dal Maso

Janvier 2015

Université de Namur

FACULTE DES SCIENCES

Enseignement du Département de Chimie

Rue de Bruxelles 61 - 5000 NAMUR

Téléphone: + 32(0)81.72.54.44 - Téléfax: + 32(0)81.72.54.40

E-mail: enseignement.chimie@unamur.be - http://www.unamur.be

Etude structurale du domaine PWWP de la DNMT3A et DNMT3B

Thomas Dal Maso

Résumé :

Depuis quelques années, un nombre important de recherches se sont intéressées à l'épigénétique et son rôle dans le développement de l'individu mais également de maladies comme le cancer. La méthylation de l'ADN, catalysée par les ADN méthyltransférases (DNMTs), est la modification épigénétique la plus étudiée et permet de réguler l'expression des gènes. Dans le cancer, ce processus est dérégulé et conduit à l'inactivation des gènes suppresseurs de tumeurs permettant aux cellules cancéreuses de proliférer.

Les DNMTs sont donc des cibles thérapeutiques intéressantes. Jusqu'à présent, les études portaient sur le domaine catalytique des DNMTs et ont mené au développement d'inhibiteurs nucléosidiques. Ceux-ci présentent cependant une forte toxicité et la recherche actuelle se dirige vers d'autres inhibiteurs plus sélectifs et moins nocifs pour les cellules saines.

Parmi les isoformes des DNMTs, les DNMT3A et DNMT3B sont surexprimées dans certains cancers et peuvent être ciblées spécifiquement. En effet, elles sont les seules à posséder un domaine PWWP qui reconnait un marqueur épigénétique essentiel se trouvant sur la queue des histones. En inhibant l'interaction protéine-protéine du domaine PWWP avec cette histone, la méthylation ne se ferait plus sur les séquences des gènes suppresseurs de tumeurs. Récemment, en 2011, une étude a pu résoudre pour la première fois la structure cristallographique du domaine PWWP des DNMT3A/3B en complexe avec un ligand, le bis-tris.

Le but de ce mémoire est donc d'étudier l'interaction de ce domaine PWWP des DNMT3A/3B avec l'histone, et également de mettre en évidence le mode de liaison d'analogues du bis-tris pouvant inhiber cette interaction protéine-protéine. Pour ce faire, les protéines seront surexprimées, purifiées et cristallisées en présence des différents ligands. Les cristaux obtenus seront étudiés par diffraction de rayons X, et les structures résolues conduiront à la rationalisation du mode d'interaction de la protéine à la fois avec le marqueur épigénétique et les ligands potentiels.

Mémoire de Master en Chimie (orientation Chimie du Vivant et des Nanomatériaux) :

Finalité Approfondie

Janvier 2015

Promoteur : Prof. J. Wouters

Encadrant : G. Rondelet

Pour commencer, je tiens à remercier le Professeur Johan Wouters pour m'avoir accueilli et m'avoir permis d'enrichir mes connaissances au sein de son laboratoire.

Je voudrais également remercier Grégoire Rondelet pour son encadrement et son aide dans la réalisation de ce travail.

Mes remerciements vont aussi aux membres de l'Unité de Recherche en Biologie Moléculaire (URBM) pour l'accès au laboratoire pour la composante biologique de ce mémoire ainsi que les membres du CBS pour leur bonne humeur et les conseils prodigués tout au long de ce travail.

Enfin, je voudrais remercier ma famille pour son soutien et plus particulièrement ma mère et Ghislaine Davreux pour la relecture et la correction de l'orthographe.

Table des matières

Liste des 20 acides aminés naturels 1
Liste des abréviations 2
1 Introduction 3
1.1 Le cancer 3
1.2 Epigénétique 6
1.3 Modifications épigénétiques 6
1.3.1 Méthylation de l'ADN 7
1.3.2 Modifications des histones 9
1.3.3 Interdépendance histones et méthylation de l'ADN 10
1.4 Les ADN (cytosine-5) méthyltransférases 13
1.4.1 Méthylation de l'ADN 13
1.4.2 Isoformes des méthyltransférases de l'ADN 15
1.4.3 Inhibiteurs des méthyltransférases de l'ADN 16
1.5 DNMT3A/DNMT3B 19
1.6 Domaine PWWP 21
2 Objectifs et stratégie 24
3 Résultats expérimentaux et discussion 27
3.1 Etude structurale préliminaire 27
3.2 Surexpression et purification 29
3.2.1 Principes 29
3.2.2 Partie expérimentale 36
3.3 Cristallogenèse 40
3.3.1 Principes 40
3.3.2 Co-cristallisation du domaine PWWP de la DNMT3A et de la DNMT3B avec un peptide 41
3.3.3 Co-cristallisation du domaine PWWP de la DNMT3B avec des ligands 44
4 Analyses des données cristallographiques et interprétations 47
4.1 Principes 47
4.2 Structure obtenue 49
4.3 Analyses des différentes structures du domaine PWWP de la DNMT3B 52
4.4 Analyse du mode d'interaction et comparaison avec d'autres structures 54
4.4.1 Analyse des structures 54

5 Conclusion 65
6 Matériel et méthodes 67
6.1 Plasmide codant pour le domaine PWWP de la DNMT3A et DNMT3B 67
6.2 Surexpression et purification du domaine PWWP de la DNMT3A et DNMT3B 67
6.3 SDS-PAGE 71
6.4 TEV protéase 74
6.5 Essais de cristallisation 76
6.6 Détermination de la structure 79
7 Références 80

Liste des 20 acides aminés naturels

Noms	Code à 3 lettres	Code à 1 lettre
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cystéine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Méthionine	Met	M
Phénylalanine	Phe	F
Proline	Pro	P
Sérine	Ser	S
Thréonine	Thr	T
Tryptophane	Trp	W
Tyrosine	Tyr	Y
Valine	Val	v

Liste des abréviations

ADD	ATRX-DNMT3-DNMT3L
ADN	Acide désoxyribonucléique
ARN	Acide ribonucléique
Bis-tris	2-Bis(2-hydroxyethyl)amino-2-(hydroxymethyl)-1,3-propanediol
BRPF1	Bromodomaine and PHD finger containing, 1
CpG	Dinucléotide cytosine-guanine
DNMT	Méthyltransférase de l'ADN
DRX	Diffraction des rayons X
FDA	Food and Drug Administration
H3K36me3	Lysine 36 triméthylée de l'histone 3
HAT	Histone acéthyltransférase
HDAC	Histone déacéthylase
His-tag	Queue de polyhistidines
HMT	Histone méthyltransférase
IMAC	Immobilized Metal Affinity Chromatography
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
JMJD2A	Jumonji domain containing 2A
KDM2A	Lysine-specific demethylase 2A
MBDP ou MBD	Methyl-CpG-binding domain protein
MBT	Malignan brain tumor
PDB	Protein Data Bank
PEG	Poléthylène glycol
PHF1	PHD finger protein 1
pI	Point isolélectrique
RMSD	Root mean square deviation
SAH	S-adénosyl-L-homocystéine
SAM	S-adénosyl-L-méthionine
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TEMED	Tétra-méthyl-éthylènediamine

1 Introduction

1.1 Le cancer

Au cours des derniers siècles, l'amélioration des conditions socio-économiques, de l'industrie et de la science en général (ex : médecine) ont modifié nos milieux de vie (bien-être, santé,…) et ont entrainé une augmentation de l'espérance de vie. Cependant, certaines transformations de nos modes et milieux de vie (pollution,…) ont montré des impacts néfastes sur la santé et pourraient, selon des données récentes, conduire à une inflexion de cette tendance d'évolution de l'espérance de vie. Les cancers, en particulier, sont des pathologies traçantes des risques majeurs de l'environnement sur la santé (Jones and Baylin 2002). A tel point que les répercussions des modifications de l'environnement et des modes de vie (exposition professionnelle, tabagisme,…) sur l'augmentation de l'incidence de certains cancers sont devenues une préoccupation majeure de santé publique. On estime que le nombre de cancers a augmenté de 50% en 50 ans (Institut de Veille Sanitaire 2013). Cette augmentation a donc poussé les scientifiques à se pencher sur cette maladie.

Un cancer est une maladie caractérisée par une prolifération cellulaire (tumeur) anormalement importante au sein d'un tissu de l'organisme (Hanahan and Weinberg 2000). Ces cellules dérivent toutes d'un même clone, cellule initiatrice du cancer qui a acquis certaines caractéristiques (Hallmarks) lui permettant d'acquérir des avantages (voir Fig 1). Au cours de l'évolution de la maladie, certaines cellules peuvent migrer de leur lieu de production et former des métastases.

Figure 1 : Capacités distinctives des cancers ("Hallmarks") (Hanahan and Weinberg 2011).

Ces capacités distinctives, décrites pour la première fois par Hanahan et Weinberg en 2000, étaient au nombre de six. Par la suite, quatre nouvelles capacités y ont été ajoutées. Ces différentes capacités, acquises par les cellules cancéreuses pour leur donner un avantage face aux cellules saines et aux mécanismes d'élimination naturels du corps humain, sont : l'autosuffisance en signaux de croissance, l'insensibilité aux signaux inhibiteurs de la croissance, la capacité à éviter l'apoptose, la capacité de se répliquer indéfiniment, l'induction de l'angiogenèse, la capacité à former des métastases, la dérégulation du métabolisme énergétique cellulaire, la capacité à éviter une destruction par le système immunitaire, l'inflammation favorisant les tumeurs et l'instabilité génomique (mutation).

L'acquisition de multiples capacités par une cellule est souvent due à une succession d'altérations dans le génome de la cellule. Ainsi, certains mutants auront un avantage et pourront le transmettre à la lignée de cellules suivantes. En conséquence, la progression de la tumeur peut être décrite comme une succession d'expansions clonales, dont chacune est déclenchée par l'acquisition fortuite d'un génotype mutant favorable. Cette propension à

accumuler les mutations chez les cellules ayant commencé un processus tumoral va dépendre de l'augmentation de la sensibilité aux agents mutagènes mais aussi, d'une rupture dans le fonctionnement de la machinerie cellulaire de maintenance de l'intégrité du génome. Dans les cellules normales, cette machinerie est « composée » de gardiens du génome, avec comme molécule centrale p53. Elle comprend des entités capables de reconnaitre les dommages subis par l'ADN, de réparer l'ADN et d'inactiver les molécules mutagènes avant qu'elles n'atteignent l'ADN (Nelson and Cox 2010). Les gènes codant les différents éléments de la machinerie cellulaire de maintenance de l'intégrité du génome sont donc considérés comme des gènes ralentissant la propagation des cancers et sont appelés « gènes suppresseurs de tumeurs ».

Ainsi, quelques expansions clonales peuvent être déclenchées par des changements non mutationnels affectant la régulation de l'expression des gènes. L'extraordinaire capacité des systèmes de maintenance du génome de détecter et résoudre les défauts de l'ADN garantit que les taux de mutation spontanée sont généralement très faibles au cours de chaque génération cellulaire. Dans le but d'acquérir les gènes mutants nécessaires pour orchestrer la carcinogenèse, les cellules cancéreuses sont capables d'augmenter le taux de mutation. Un des moyens mis en place par les cellules cancéreuses est la dérégulation de la méthylation de l'ADN (Tollefsbol 2009) (Kulis and Esteller 2010).

1.2 Epigénétique

Le terme épigénétique est introduit pour la première fois par C. Waddington en 1942 dans un article paru dans « International Journal of Epidemiology ». Il le définit comme étant la branche de la biologie qui concerne les interactions entre les gènes et leurs produits (Waddington 1942). Ces interactions conduisent à des modulations du phénotype. De nos jours, la définition du terme épigénétique a été revue pour devenir l'ensemble des modifications chimiques réversibles et héritables qui modifient l'expression des gènes mais sans toucher à la séquence nucléotidique de l'ADN (Tollefsbol 2009). L'expression du phénotype n'est donc plus exclusivement liée au génotype mais aussi à un autre facteur, l'épigénétique. Pendant longtemps, l'impact des modifications épigénétiques a été considéré comme limité à la division cellulaire. Cependant, diverses études ont montré que l'information épigénétique pouvait être transmise d'une génération à la suivante (Chong and Whitelaw 2004). Des dérèglements épigénétiques sont à la base de certains cancers et d'autres maladies touchant le système immunitaire et nerveux (Baylin and Herman 2000). L'étude des phénomènes épigénétiques parait dès lors importante pour le traitement de maladies telles que le cancer.

Alors que la séquence de l'ADN conditionne la nature des produits associés aux gènes, les régulations épigénétiques, quant à elles, jouent un rôle dans le niveau de transcription des gènes. Ces principales régulations correspondent à des modifications chimiques réversibles de l'ADN et des queues des histones, impactant directement la forme de la chromatine (Huen, Huang et al. 2010).

1.3 Modifications épigénétiques

Les modifications épigénétiques ont une place centrale dans le développement des cellules saines mais se retrouvent dérégulées dans le processus tumoral (Jin, Li et al. 2011). En effet, la méthylation d'un certain nombre de séquences induit une instabilité génomique qui a pour conséquence un niveau d'expression variable de certains gènes. On observera ainsi, dans une cellule tumorale, une répression de la transcription des gènes suppresseurs de

tumeurs et une augmentation de la transcription d'oncogènes. Les principales modifications épigénétiques se trouvent au niveau de l'ADN (méthylation) et des histones (méthylation, acétylation et phosphorylation) (Esteller 2007) (Hughes, Wiggins et al. 2007).

1.3.1 Méthylation de l'ADN

Chez les eucaryotes, la méthylation de l'ADN se déroule sur les cytosines (Goll and Bestor 2005). Bien que cette modification épigénétique soit détournée par les cancers, pour acquérir un avantage certain et se développer, elle est nécessaire au cours du développement de la cellule (Stec, Nagl et al. 2000). Elle intervient notamment dans la différenciation cellulaire, l'inactivation du chromosome X, la régulation des gènes et la carcinogenèse (Klose and Bird 2006) (Reik 2007).

Cette méthylation des cytosines se passe plus spécifiquement au niveau des régions riches en dinucléotides CpG, aussi appelées « îlots CpG ». Dans les cellules normales, seulement 3 à 6% des cytosines sont méthylées alors que 76% des gènes contiennent des îlots CpG dans leur région promotrice. Dans les cellules normales, les séquences génomiques répétées sont, quant à elles, fortement méthylées pour protéger l'intégrité génomique (stabilité chromosomique, translocation, perturbation des gènes,…) (Esteller and Almouzni 2005). Le statut de méthylation d'un certain nombre de séquences est souvent reconnu comme prédictif de la survenue ou de l'issue d'un cancer (Esteller 2005).

I. Hyperméthylation de l'ADN

Dans les cellules cancéreuses, l'hyperméthylation de l'ADN se déroule au niveau des gènes suppresseurs de tumeurs et entraine leur répression transcriptionnelle (voir Fig 2). Les principaux gènes visés sont ceux intervenant dans la régulation du cycle cellulaire, la réparation de l'ADN, la signalisation cellulaire, l'apoptose, la reconnaissance immune ou encore l'angiogenèse. Ces altérations épigénétiques agissent donc toutes sur les capacités distinctives des cancers tel que décrit par Hanahan et Weinberg. Les conséquences de cette dérégulation sont la résistance aux médicaments anti-cancéreux (provoquant l'apoptose), l'insensibilité aux signaux d'anti-croissance (répression des inhibiteurs de croissance cellulaire), la réplication sans limite, un défaut dans la réparation de l'ADN,… (Tollefsbol 2009)

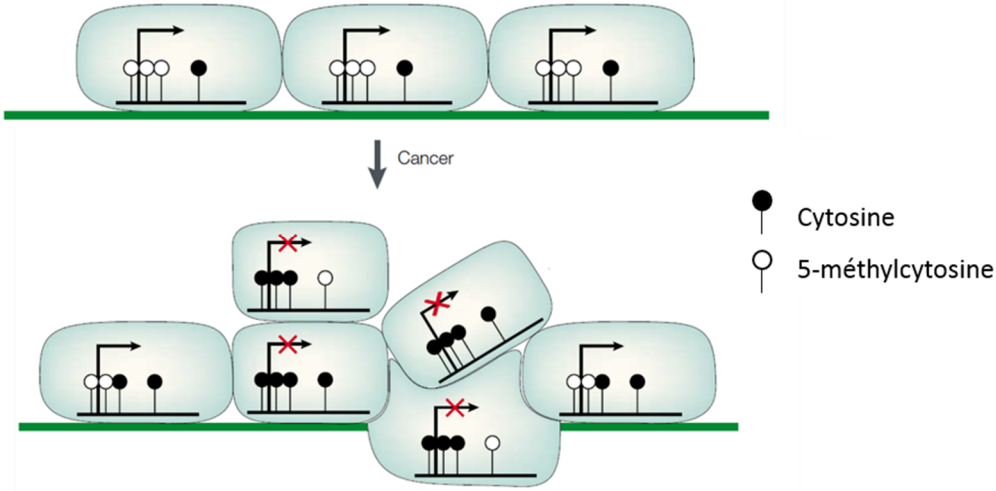


Figure 2 : Altération du pattern de méthylation influençant la transcription des gènes et conduisant à une prolifération anarchique des cellules cancéreuses. Image adaptée de (Jones and Baylin 2002).

II. Hypométhylation de l'ADN

L'hypométhylation est souvent observée dans la région promotrice des oncogènes et dans les séquences répétées de l'ADN. Ces oncogènes sont des gènes favorisant, d'une manière ou d'une autre, le processus tumoral comme, par exemple, en stimulant le cycle cellulaire des cellules cancéreuses, en favorisant le processus de métastase, en diminuant l'adhésion cellulaire ou en empêchant la reconnaissance des cellules tumorales par le système immunitaire (Gaudet, Hodgson et al. 2003).

Ces séquences surtout présentes au niveau du centromère et dans les régions promotrices de certains gènes, sont composées d'éléments génétiques mobiles (séquences satellites) qui sont méthylés pour permettre de maintenir l'intégrité du génome. Dans le cadre de certains cancers, il a été constaté que ces séquences subissent une déméthylation de façon à réactiver certains éléments génétiques mobiles. Cette réexpression entraine donc une instabilité génomique et des mutations (erreur lors de la réplication, translocation,…).

1.3.2 Modifications des histones

Les histones, de par leur rôle central dans la structuration de l'ADN, sont des composants intégraux de la machinerie responsable de la régulation des gènes. En plus des modifications de l'ADN, on retrouve également des modifications au niveau des histones. En effet, ces dernières peuvent subir des modifications post-traductionnelles sur leur partie N-terminale (queue de l'histone) telles que l'acétylation, la phosphorylation, la méthylation, l'ubiquitination et la sumoylation (Müller and Muir 2014). Ces différentes modifications perturbent l'interaction des histones avec l'ADN et donc, par conséquent, la structure de la chromatine. Par exemple, les résidus lysines, qui sont chargés positivement à pH physiologique (point isoélectrique de 9,8), interagissent avec l'ADN, chargé négativement par la présence de groupements phosphates. Il en résulte donc une forme condensée de l'ADN (hétérochromatine). Au contraire, si les « Histones deacetylases » transfèrent un groupement acétyle sur l'azote de la lysine, celle-ci va perdre sa charge positive et n'interagira plus avec l'ADN de la même manière. Cela conduira donc à la forme décondensée de l'ADN (euchromatine) (Barski, Cuddapah et al. 2007).

L'ensemble des modifications présentes sont réversibles et catalysées par des enzymes. Ainsi, les modifications les plus étudiées et celles qui sont reliées au cancer sont l'acétylation/déacétylation et la méthylation des queues des histones. Ces modifications sont catalysées par des enzymes appelées « Histones deacetylases » (HDACs), « Histones acetyltransferases » (HATs) et « Histones methyltansferase » (HMTs) (Shahbazian and Grunstein 2007).

1.3.3 Interdépendance histones et méthylation de l'ADN

Il existe une interdépendance entre la méthylation de l'ADN et la méthylation/acétylation des queues des histones (voir Fig 3).

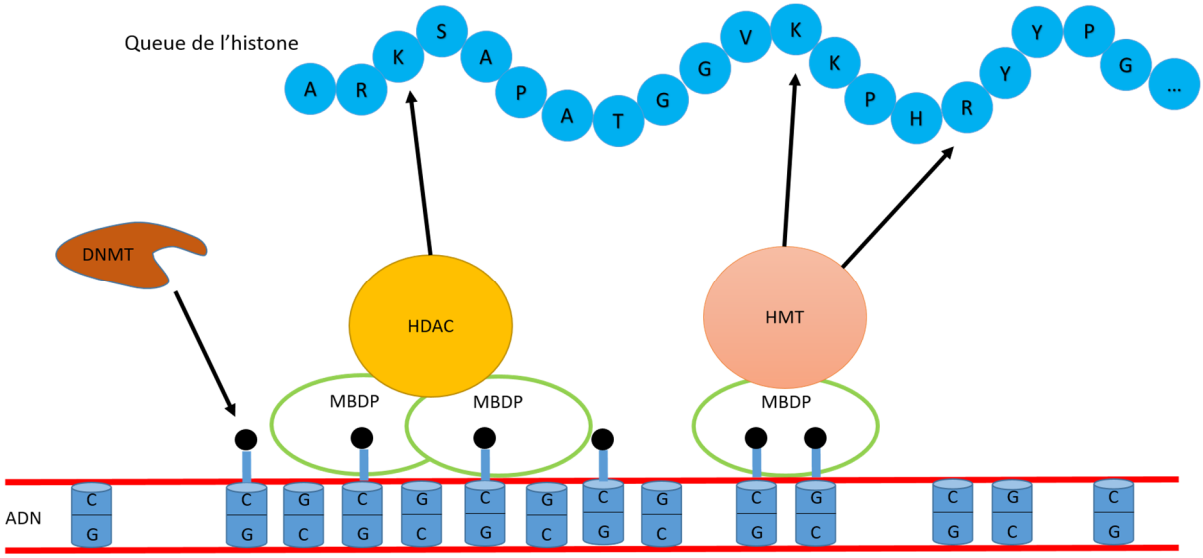


Figure 3 : Schéma de la régulation épigénétique par la méthylation de l'ADN et l'acétylation/méthylation des queues des histones.

La région méthylée de l'ADN est reconnue par des protéines appelées « methyl-CpG-binding domain proteins » (MBDPs) qui sont elles-mêmes capables de recruter des HDACs et HMTs pour acétyler ou méthyler la queue des histones impliquées. On parlera ainsi de marqueurs épigénétiques pour la méthylation de l'ADN et de « lecteurs » de marqueurs épigénétiques pour ces autres protéines (Adams-Cioaba and Min 2009).

L'ADN, au sein du noyau, est empaqueté sous forme de chromatine. Ceci permet à la cellule de pouvoir stocker une grande quantité d'ADN dans une entité aussi petite qu'une cellule. L'ADN double brin s'enroule donc autour d'un octamère d'histones appelé nucléosome. Cette unité se répète et donne lieu à des structures plus organisées qui aboutissent à la chromatine (voir Fig 4). Dans les cellules eucaryotes, l'initiation de la transcription des gènes est contrôlée par la force intrinsèque des promoteurs, par les facteurs de transcription spécifiques des promoteurs mais également par l'accès de l'ADN à la machinerie basale de transcription. Cette accessibilité de l'ADN est par conséquent gouvernée par la structure de la chromatine. Les histones composant le nucléosome possèdent une

queue qui peut subir diverses modifications chimiques comme expliqué précédemment. Ensemble, la méthylation de l'ADN et les modifications des histones régulent l'organisation de la structure de la chromatine et donc l'expression des gènes (Li, Mao et al. 2013).

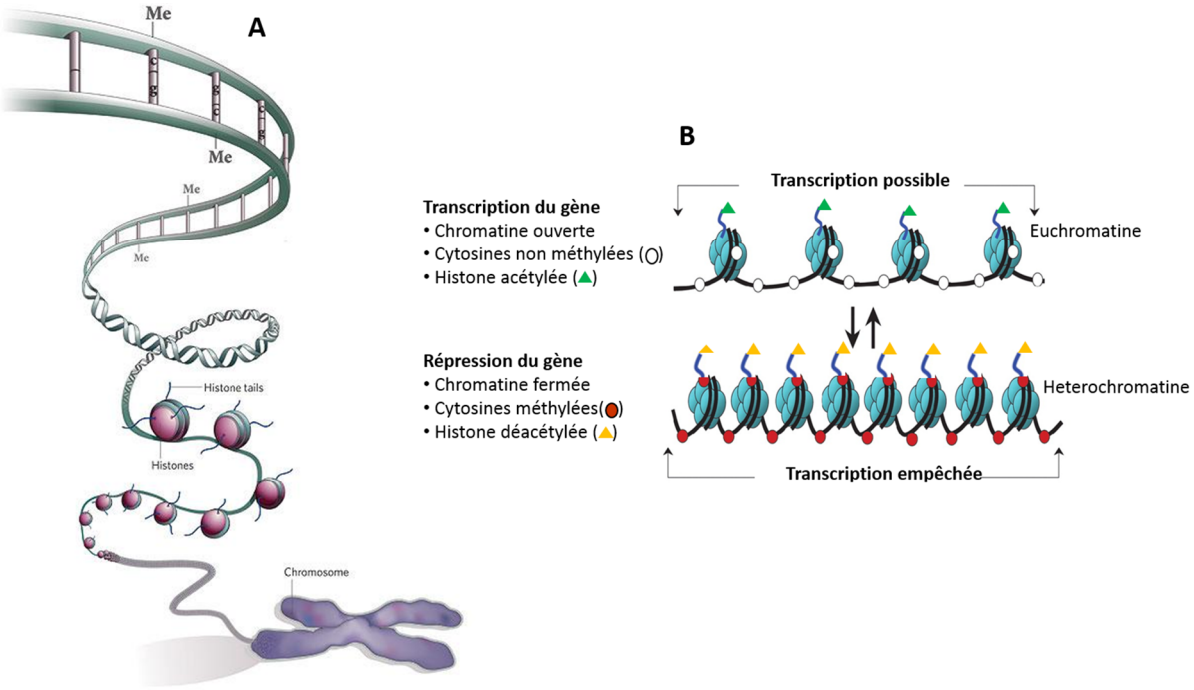


Figure 4 : (A) Schéma des différents niveaux de structure de la chromatine. (B) Les deux formes de la chromatine (l'euchromatine et l'hétérochromatine) et les modifications chimiques relatives à ces structures (Jones and Liang 2012) (Liêu 2014).

Les modifications post-traductionnelles des histones provoquent un changement conformationnel de la chromatine et, via une interdépendance avec d'autres modifications telles que la méthylation de l'ADN, elles peuvent contrôler la transcription du gène impliqué. Lorsque les histones sont acétylées, la chromatine est sous forme ouverte (euchromatine) permettant l'accès à la machinerie basale de transcription, facilitant la réparation de l'ADN et sa réplication. Lorsque l'ADN est méthylé, la chromatine est sous forme fermée (hétérochromatine). Cette forme permet un meilleur maintien dans l'espace cellulaire, une protection contre les dommages et facilite la transmission de l'ADN aux cellules filles lors de la réplication (Qin and Min 2014). Lorsque la chromatine est sous forme fermée, la machinerie basale de transcription n'a plus accès à l'ADN. Il en résulte donc une répression dans la transcription des gènes associés.

Tous les résidus composant la queue de l'histone ne peuvent être modifiés chimiquement (voir Fig 5). Ainsi seulement les Lys/ Arg seront méthylées, les Lys acétylées et les Ser/Thr/Tyr phosphorylées. L'emplacement du résidu sur la queue de l'histone est également important. En fonction du résidu modifié, l'effet sur la cellule sera différent (Strahl and Allis 2000).

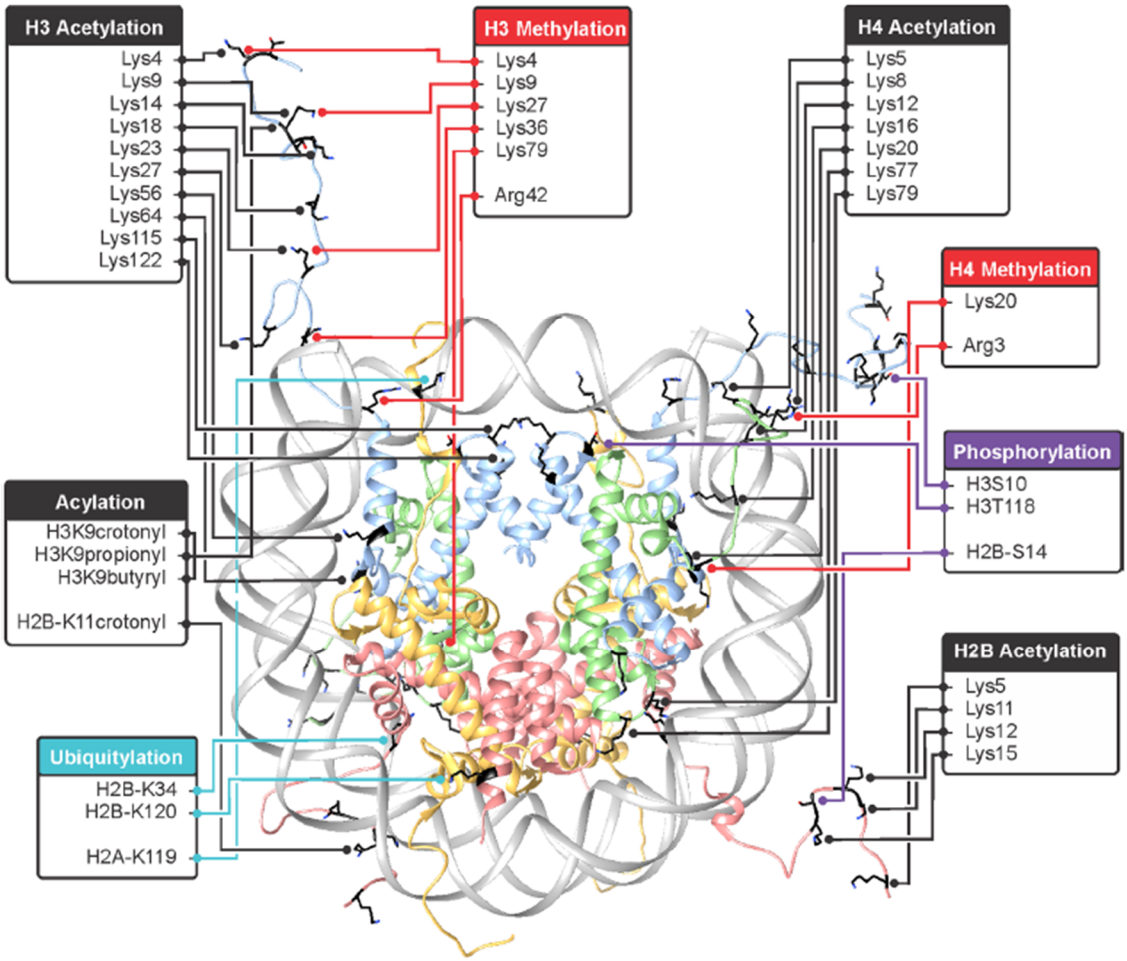


Figure 5 : Représentation des différentes modifications possibles sur les résidus de la queue des histones (Müller and Muir 2014).

Comme le montre cette représentation du nucléosome, ce sont principalement les résidus Lys qui subissent des modifications. Ainsi, l'acétylation de la lysine 16 de l'histone 4 va jouer un rôle dans le repliement de la chromatine et par conséquent sur la forme de cette dernière alors que la lysine 5 et 12 de cette même histone ont un rôle dans l'assemblage des histones et de l'ADN.

La modification la plus étudiée est la méthylation des résidus lysine et arginine. Il existe différents niveaux de méthylation de ces résidus. En effet, la lysine peut être monométhylée, diméthylée ou triméthylée et l'arginine peut subir une méthylation (diméthylation) symétrique ou asymétrique. Cette modification ne va pas changer la charge du résidu et donc n'affectera pas directement la structure de la chromatine. Cependant, les groupements méthyles vont permettre de recruter d'autres protéines qui impacteront sur la forme ouverte ou fermée de la chromatine (Filippakopoulos and Knapp 2014). Toutes ces modifications définissent le « code histone ». Quelques résidus ont déjà été étudiés et, pour certains, leur rôle a pu être établi. Ainsi, les lysines 4 et 16 triméthylées de l'histone 3 (H3K4me3 et H3K36me3), sont des marqueurs (« hallmark ») de gènes transcriptionnellement actifs alors que les lysines triméthylées 27 de cette même histone (H3K27me3) sont des marqueurs de gènes transcriptionnellement inactifs (inhibition). Le niveau de méthylation est aussi à prendre en compte. Dans le cas des H3K9, si le résidu est diméthylé (H3K9me2) ou triméthylé (H3K9me3), la transcription du gène sera réprimée. A l'inverse, la lysine 9 monométhylée (H3K9me1) est associée à une activation transcriptionnelle. Bien que tous ces marqueurs ne soient pas encore bien connus, de plus en plus d'études y sont consacrées de par leur place dans la compréhension des évènements épigénétiques (Dhayalan, Rajavelu et al. 2010).

1.4 Les ADN (cytosine-5) méthyltransférases

1.4.1 Méthylation de l'ADN

Les ADN méthyltransférases (DNMTs) catalysent la réaction de méthylation (transfert de méthyle) de la cytosine sur le carbone 5 (voir Fig 6). Les intervenants sont l'enzyme, l'ADN et le cofacteur S-adénosyl-L-méthionine (SAM) (Chedin 2010).

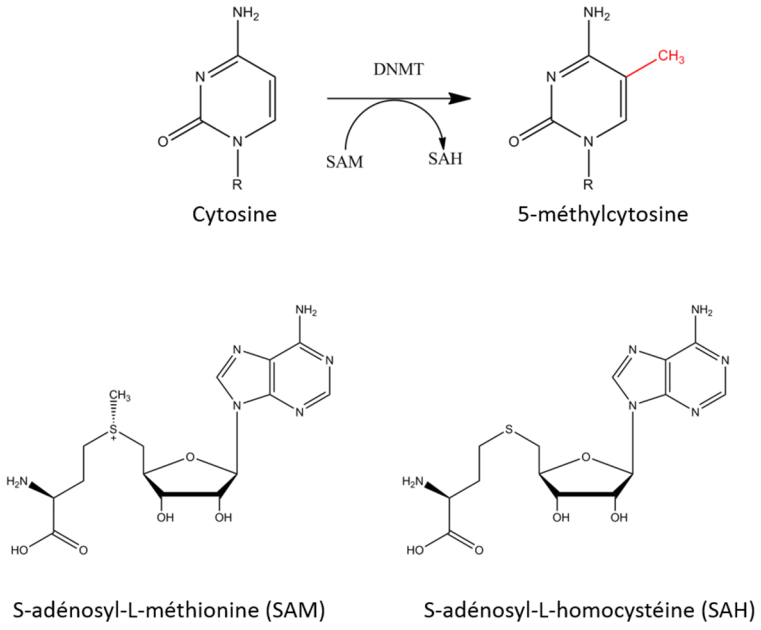


Figure 6 : Méthylation d'une cytosine à la position 5 par les DNMTs. Structure des cofacteurs SAM et SAH.

Des études réalisées sur les méthyltransférases bactériennes ont montré que la méthylation des bases cytosines de l'ADN se faisait grâce à un phénomène de « base flipping ». Ce phénomène, considéré comme universel chez les DNMTs, consiste en une rupture des ponts hydrogène et du π-stacking entre bases. Il en résulte que la base cytosine sort de la double hélice de l'ADN, se place dans le site actif et va pouvoir être méthylée par l'enzyme (Klimasauskas, Kumar et al. 1994).

Le mécanisme catalytique a été décrit pour la première fois par Cheng et Blumenthal en 2008 (voir Fig 7).

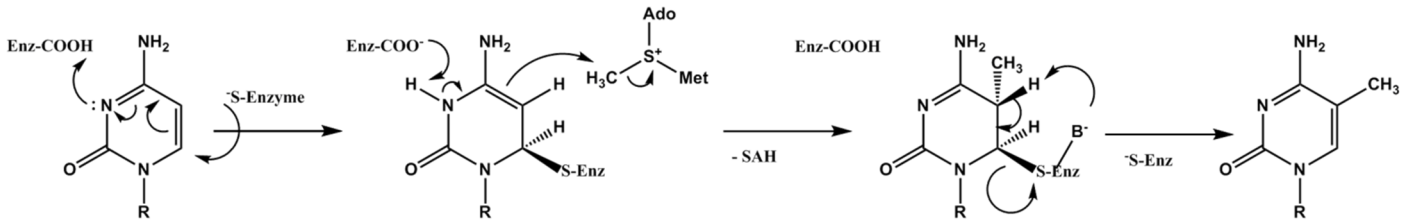


Figure 7 : Mécanisme de méthylation des cytosines de l'ADN pour les DNMTs. R = désoxyribose (Cheng and Blumenthal 2008).

1.4.2 Isoformes des méthyltransférases de l'ADN

Les DNMTs sont classées en 3 catégories : les DNMT1, DNMT2 et DNMT3. La DNMT1, première méthyltransférase découverte, est impliquée dans le maintien du pattern de méthylation lors de la réplication cellulaire en agissant sur l'ADN hémiméthylé. La DNMT2, quant à elle n'agit pas sur l'ADN mais sur l'ARN de transfert (Goll, Kirpekar et al. 2006). Les DNMT3 sont réparties en 3 sous-classes : les DNMT3A, DNMT3B et DNMT3L. Les DNMT3A et DNMT3B sont responsables de la méthylation *de novo* de l'ADN et établissent donc le pattern de méthylation (Chedin 2010). La DNMT3L régule l'activité de la DNMT3A et DNMT3B en formant un complexe avec celles-ci (voir Fig 8) (Liao, Tai et al. 2012). La DNMT3L ne possède pas d'activité catalytique intrinsèque mais sert plutôt dans la régulation de l'établissement du pattern de méthylation par les DNMT3A et DNMT3B. En effet, la DNMT3L possède un domaine régulateur capable d'interagir avec ces deux dernières de façon à réguler leurs activités (Zhang, Jurkowska et al. 2010). On la retrouve donc, la plupart du temps, en complexe (Otani, Nankumo et al. 2009).

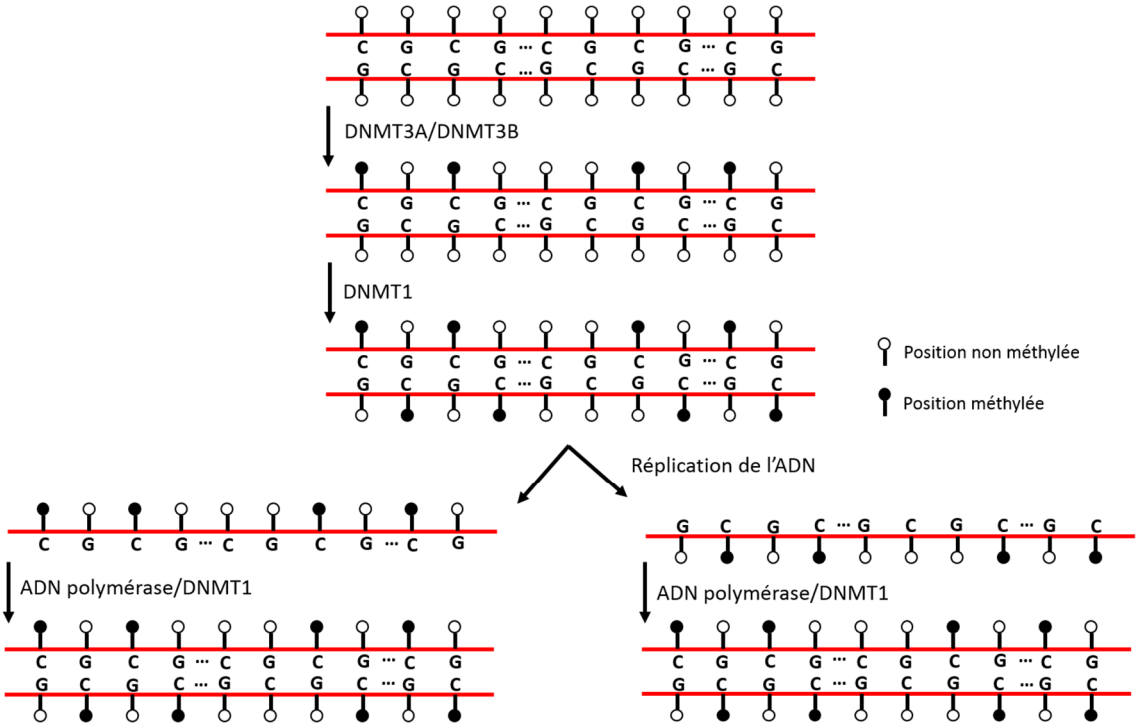


Figure 8 : Méthylation de novo et maintenance de la méthylation de l'ADN.

Les principaux domaines présents dans ces enzymes sont un domaine régulateur variable du côté N-terminal et un domaine catalytique du côté C-terminal. Seule la DNMT2 ne possède pas de domaine régulateur mais seulement un domaine catalytique (voir Fig 9).

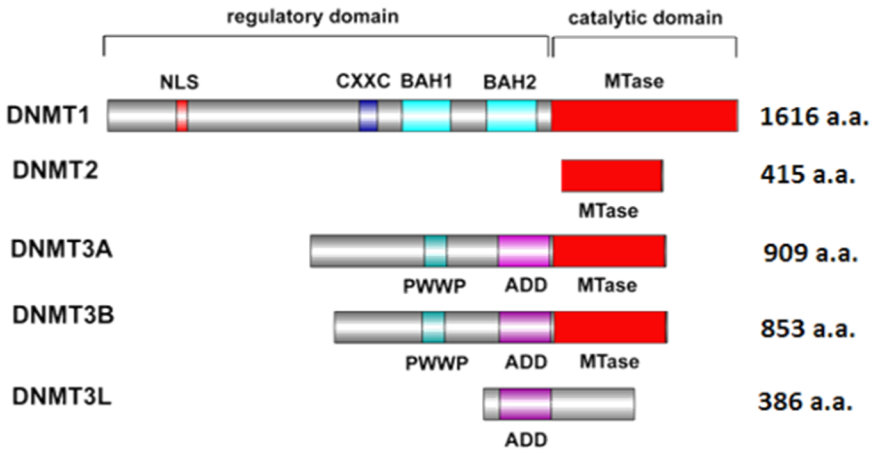


Figure 9 : Les différentes isoformes des DNMTs présentées avec leurs principaux domaines (Yuan-Yeh, Li-Yu et al. 2013).

1.4.3 Inhibiteurs des méthyltransférases de l'ADN

Comme montré précédemment, les modifications épigénétiques jouent un rôle important dans la transcription des gènes et dans les divers processus tumoraux. Dès lors, un nombre important d'études visent le développement d'inhibiteurs du domaine catalytique des DNMTs en vue de rétablir l'expression des gènes (voir Fig 10 A) (Peedicayil 2006). Si peu d'inhibiteurs s'appuient sur le cofacteur, la similarité avec le substrat a été fortement étudiée. Cela conduit donc à la première génération d'inhibiteurs qui sont des inhibiteurs nucléosidiques. Suite à leur importante toxicité, les études plus récentes tentent de cibler des composés non nucléosidiques qui présentent une plus faible toxicité. Cette nouvelle catégorie d'inhibiteurs fait partie de la seconde génération de molécules testées.

Les inhibiteurs nucléosidiques sont des analogues de la cytosine modifiée en position 5 généralement (voir Fig 10 B). Ces analogues vont remplacer les cytosines naturelles au sein de l'ADN et former un complexe covalent (inhibiteurs suicides) avec l'enzyme lors de la réaction de transfert du méthyle. Une déméthylation de l'ADN sera donc observée après

plusieurs cycles de réplication cellulaire (American Society of Clinical Oncology 2003).

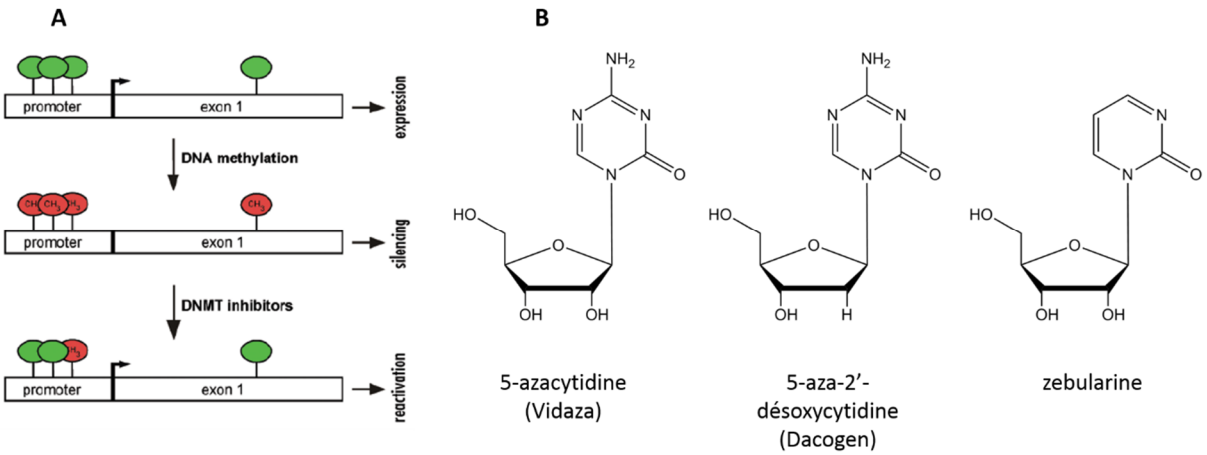


Figure 10 : **A :** Inhibition de la méthylation de l'ADN par des inhibiteurs. **B :** Présentation de quelques inhibiteurs nucléosidiques (Mund, Brueckner et al. 2006).

Le premier inhibiteur (Vidaza®) est incorporé au sein de l'ARN et interfère lors de la traduction de protéine alors que le second (Dacogen®) agit directement au niveau de l'ADN. La zébularine, quant à elle, permet l'apoptose des cellules cancéreuses et l'arrêt de leur cycle cellulaire. Ces composés permettent donc de réactiver l'expression des gènes précédemment réprimés et de stopper la propagation du cancer. Cependant, ces inhibiteurs présentent de gros défauts notamment une mauvaise stabilité et une forte cytotoxicité. Il a donc fallu passer à un autre type de médicaments que sont les inhibiteurs non-nucléosidiques (Brueckner and Lyko 2004).

Contrairement aux inhibiteurs nucléosidiques, les non-nucléosidiques ne s'incorporent pas dans l'ADN. En effet, ils sont capables d'occuper le site catalytique des DNMTs ou d'interagir avec les sites riches en îlots CpG. L'enzyme ne pourra donc plus se lier. N'interférant donc plus avec l'ADN, ils vont présenter une toxicité moindre que leurs homologues nucléosidiques. L'epigallocatechin-3-gallate, molécule dérivée du thé vert, agit sur le site actif de la DNMT1 alors que le composé RG108 agit sur le site catalytique des DNMTs en général (voir Fig 11). Ces molécules entrainent une déméthylation de l'ADN et une réactivation des gènes tout en n'affectant pas le profil de méthylation des séquences satellites

centromériques. Ces composés semblent donc être de bien meilleurs médicaments car ces derniers ont comme capacité de cibler toutes les DNMTs car le site catalytique est conservé (Mund, Brueckner et al. 2006).

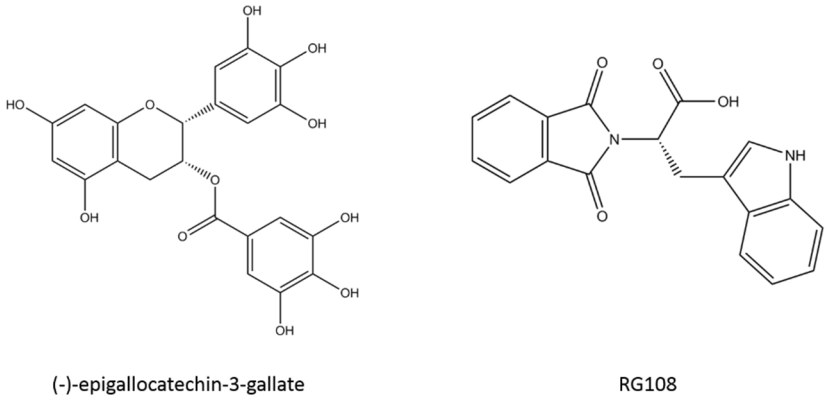


Figure 11 : Structures de deux inhibiteurs non-nucléosidiques.

1.5 DNMT3A/DNMT3B

Dans le cadre de ce mémoire, une étude plus approfondie sur les méthyltransférases 3A et 3B de l'ADN a été réalisée. Ces DNMT3A et 3B se différencient des autres DNMTs par la présence d'un domaine ADD (ATRX-DNMT3-DNMT3L) régulateur du domaine catalytique et d'un domaine PWWP qui est un domaine de reconnaissance de la queue des histones (voir Fig 12) (Bestor 2000). Celui-ci sera abordé plus en détail au point suivant. C'est grâce à ces différents domaines que les DNMT3A et 3B peuvent lire les marques épigénétiques et plus particulièrement la méthylation/non-méthylation de certains résidus lysines sur les queues des histones (Vezzoli, Bonadies et al. 2010). Comme nous l'avons vu auparavant, ces DNMTs sont responsables de la méthylation de novo de l'ADN (Chen, Tsujimoto et al. 2004). Des études ont montré que la DNMT3A était impliquée dans la fixation de l'empreinte parentale et la DNMT3B dans la méthylation des péricentromères répétés des chromosomes (Ge, Pu et al. 2004).

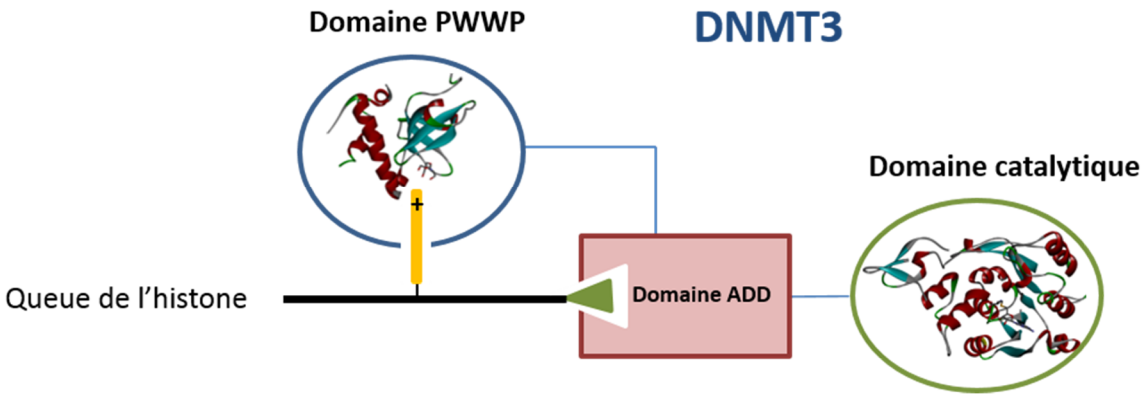


Figure 12 : Représentation schématique de la liaison d'une queue d'histone avec une DNMT3 (Themans 2014).

Les DNMT3 sont des enzymes très proches en terme de structure. Elles sont codées par les gènes localisés en 2p23.3 pour la DNMT3A et en 20q11.21 pour la DNMT3B (Robertson, Uzvolgyi et al. 1999). Il existe différentes isoformes pour chacune de ces DNMTs. En effet, la DNMT3A possède trois isoformes dont deux sont catalytiquement actives. On en trouve six pour la DNMT3B dont deux sont également actives (voir Fig 13). Ces différentes sous isoformes apparaissent à différents moments au cours du développement de l'organisme.

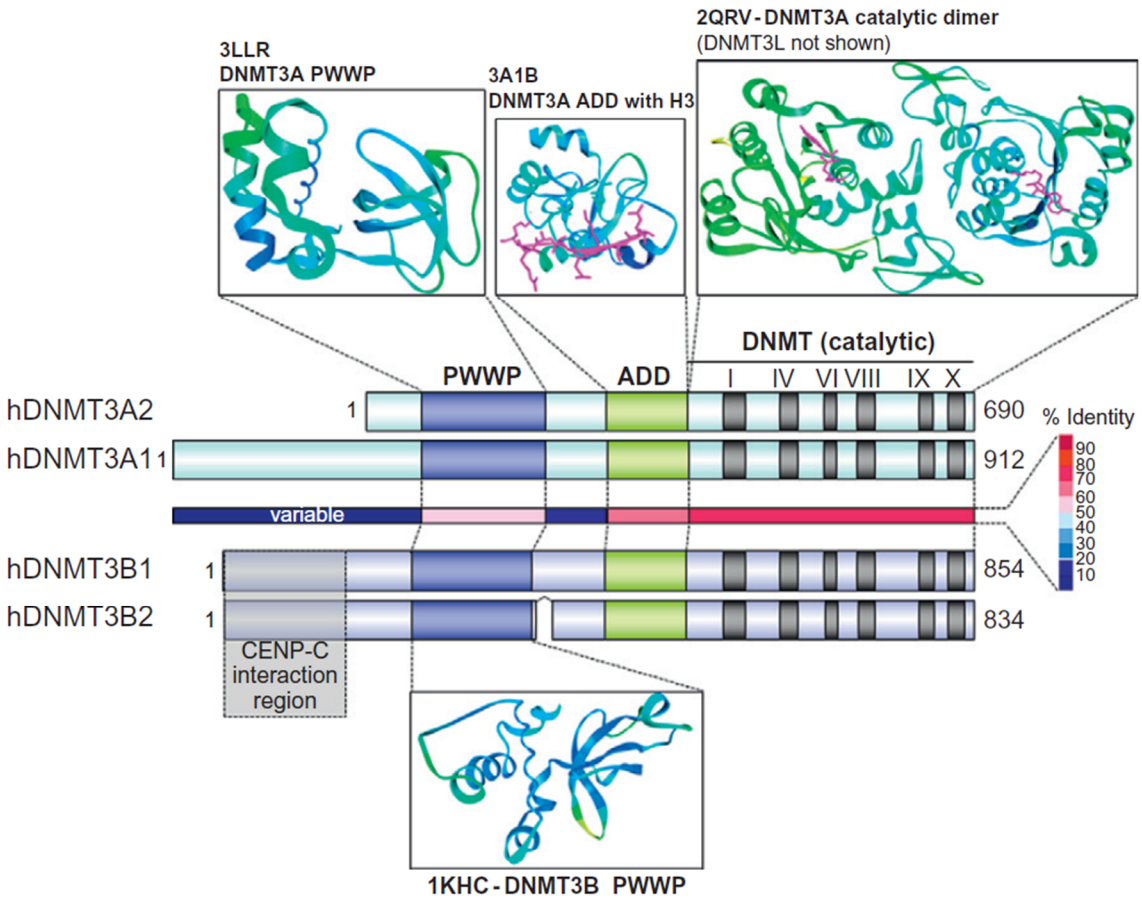


Figure 13 : Présentation des deux isoformes de la DNMT3A et DNMT3B avec leurs domaines respectifs et leur pourcentage d'identité (Chedin 2010).

Les enzymes de la sous-classe DNMT3 possèdent globalement la même organisation avec une région N-terminale régulatrice et une région C-terminale catalytique. Du point de vue du domaine catalytique, DNMT3A et DNMT3B présentent une identité de séquence de 90% et 6 motifs catalytiques essentiels qui sont hautement conservés (Jeltsch 2002). Un second domaine ATRX-DNMT3-DNMT3L (ADD) permet de réguler l'activité du domaine catalytique (Guo, Wang et al. 2014). Celui-ci est composé de 132 acides aminés et est riche en résidus cystéine. Il lie spécifiquement la lysine 4 de l'histone 3 non méthylée (H3K4) (Otani, Nankumo et al. 2009). Enfin, on abordera le domaine PWWP au point suivant (Zhang, Jurkowska et al. 2010).

1.6 Domaine PWWP

Le domaine PWWP appartient à la « Famille Royale » Tudor qui est composée des domaines Tudor, chromodomain, malignan brain tumor (MBT) et PWWP (Maurer-Stroh, Dickens et al. 2003). Le domaine PWWP a été nommé ainsi par la présence du motif Trp-Pro-Pro-Trp qui a été retrouvé dans plusieurs enzymes (Qiu, Sawada et al. 2002). Cette famille est caractérisée par le fait de ne reconnaitre que les histones méthylées, plus particulièrement la lysine 36 triméthylée de l'histone et ce, via une cage de résidus aromatiques (Kim, Daniel et al. 2006). Ce domaine est défini comme un domaine « reader » car il est capable de lire le « code histone ». Les deux enzymes (DNMT3A et DNMT3B) présentent chacune un domaine appelé PWWP malgré que le motif soit SWWP. Les principales informations concernant ces deux domaines sont reprises dans la table suivante (voir Table 1).

	DNMT3A	**DNMT3B**
Nombre de résidus	150	150
Motif conservé	SWWP	SWWP
Point isoélectrique	5,45	8,83
Poids moléculaire (Da)	16 883,1	16 676,8
Structure PDB	3LLR	3QKJ, 3FLG et 1KHC

Tableau 1 : Informations importantes sur le domaine PWWP de la DNMT3A et DNMT3B.[1]

Le point isoélectrique et le poids moléculaire ont été calculés au moyen du programme ProtParam (Swiss Institute of Bioinformatics 2014).

Récemment (en 2011), une équipe a pu résoudre la structure cristallographique du domaine PWWP de la DNMT3A (PDB : 3LLR) et DNMT3B (PDB : 3QKJ), tous deux en complexe avec une molécule de bis-tris (voir Fig 14).

[1] Ces valeurs seront utiles lors des mesures de concentration et de la vérification sur gel SDS-PAGE.

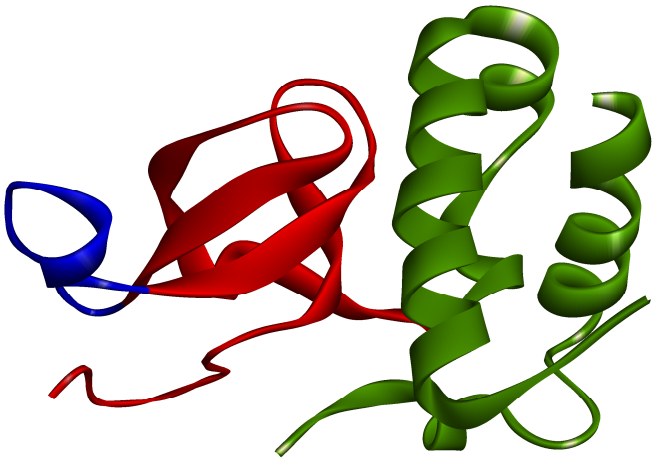


Figure 14 : Structure cristallographique du domaine PWWP de la DNMT3B (PDB : 3QKJ) mettant en évidence les motifs présents dans cette enzyme.

La structure montre trois motifs importants : un tonneau β (en rouge) avec un motif d'insertion (en bleu) et des hélices α (en vert). Le tonneau β est une structure protéique composée de feuillets β, le tout s'enroulant pour former un tonneau. La séquence SWWP est localisée sur le feuillet β numéro 4. Les résidus, formant le reste de la cage d'aromatique pour la reconnaissance de la lysine triméthylée, se trouvent à une position bien définie répartie sur les autres feuillets β ou sur le motif d'insertion (voir Fig 15).

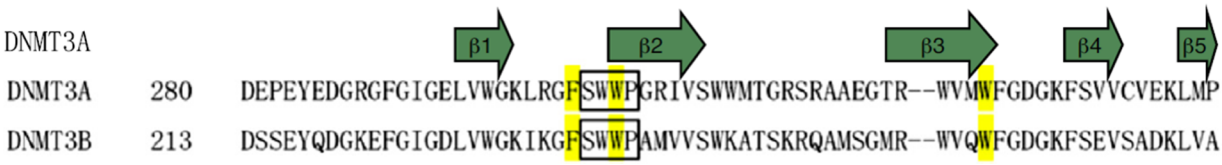


Figure 15 : Morceau de la séquence du domaine PWWP de la DNMT3A et DNMT3B avec mise en évidence du motif SWWP et des résidus intervenant dans la formation de la cage d'aromatique (Qin and Min 2014).

La cage d'aromatique comprend trois résidus dont un Trp (Trp 32) composant le motif SWWP, les deux autres résidus sont la Phe 29 et le Trp 56 du feuillet β numéro 3. Les médicaments visaient jusqu'alors le site catalytique des DNMTs. Mais les avancées dans l'étude de ces enzymes, notamment la résolution de la structure du domaine PWWP, ouvrent donc de nouvelles voies de recherche. Il était déjà connu que le domaine PWWP était capable

de reconnaitre le « code histone », plus particulièrement la lysine 36 triméthylée de l'histone 3 (H3K36me3) (van Nuland, van Schaik et al. 2013) (Wagner and Carpenter 2012). La résolution de la structure par Wu et al. de l'enzyme en complexe avec le bis-tris, suggère la possibilité d'inhiber l'interaction protéine-protéine (queue de l'histone-domaine PWWP) (Guo, Wang et al. 2014). Cette inhibition serait donc susceptible d'empêcher la méthylation des îlots CpG des séquences promotrices des gènes suppresseurs de tumeurs. La conception de nouveaux composés anti-cancéreux, ciblant le domaine PWWP, dans la thérapie de certains cancers pourrait donc être une nouvelle approche (Wu and Zhang 2014).

2 Objectifs et stratégie

L'incidence des phénomènes épigénétiques a été démontrée dans de nombreuses recherches comme étant un facteur clé dans le développement de certaines maladies comme les cancers. La modification épigénétique la plus étudiée est la méthylation de l'ADN jouant un rôle important dans la régulation de la transcription des gènes. En effet, une hyperméthylation est souvent observée au niveau des promoteurs des gènes suppresseurs de tumeurs conduisant à la répression de l'expression de ces gènes. Cela a donc conduit au développement d'inhibiteurs ciblant l'enzyme et permettant la réexpression des gènes associés.

Premièrement, des inhibiteurs nucléosidiques (suicides) ciblant le domaine catalytique des DNMTs ont été testés et approuvés par la FDA. Cependant, ceux-ci présentent une forte toxicité, et la recherche se tourne actuellement vers des inhibiteurs non-nucléosidiques qui montrent une toxicité plus faible.

En 2011, la résolution de la structure du domaine PWWP des DNMT3 par *Wu et al.* a ouvert de nouvelles perspectives de recherche sur le développement de composés interagissant avec ce domaine de reconnaissance de ces enzymes. La mise en évidence de ligands tels que le bis-tris (en complexe avec le domaine PWWP de la DNMT3A/DNMT3B) et la lysine triméthylée 36 de l'histone (peptide en complexe avec le domaine PWWP de la *Bromodomain and PHD finger containing 1*) permet d'envisager d'inhiber l'interaction protéine-protéine entre la queue de l'histone et l'enzyme. Les nouveaux composés qui en résulteraient (analogues du bis-tris et de la H3K36me3) ne cibleraient donc plus le site catalytique mais le domaine de reconnaissance et ce, de façon sélective pour la DNMT3A et DNMT3B. Le développement de pareils composés passe avant tout par une étude structurale de l'enzyme et des interactions possibles entre le site de reconnaissance et les ligands.

L'objectif de ce mémoire est donc de réaliser une étude structurale du domaine PWWP de la DNMT3A et DNMT3B en présence de ligands de façon à pouvoir comprendre et rationnaliser le mode d'interaction protéine-ligand.

La stratégie suivie est présentée à la figure suivante (voir Fig 16).

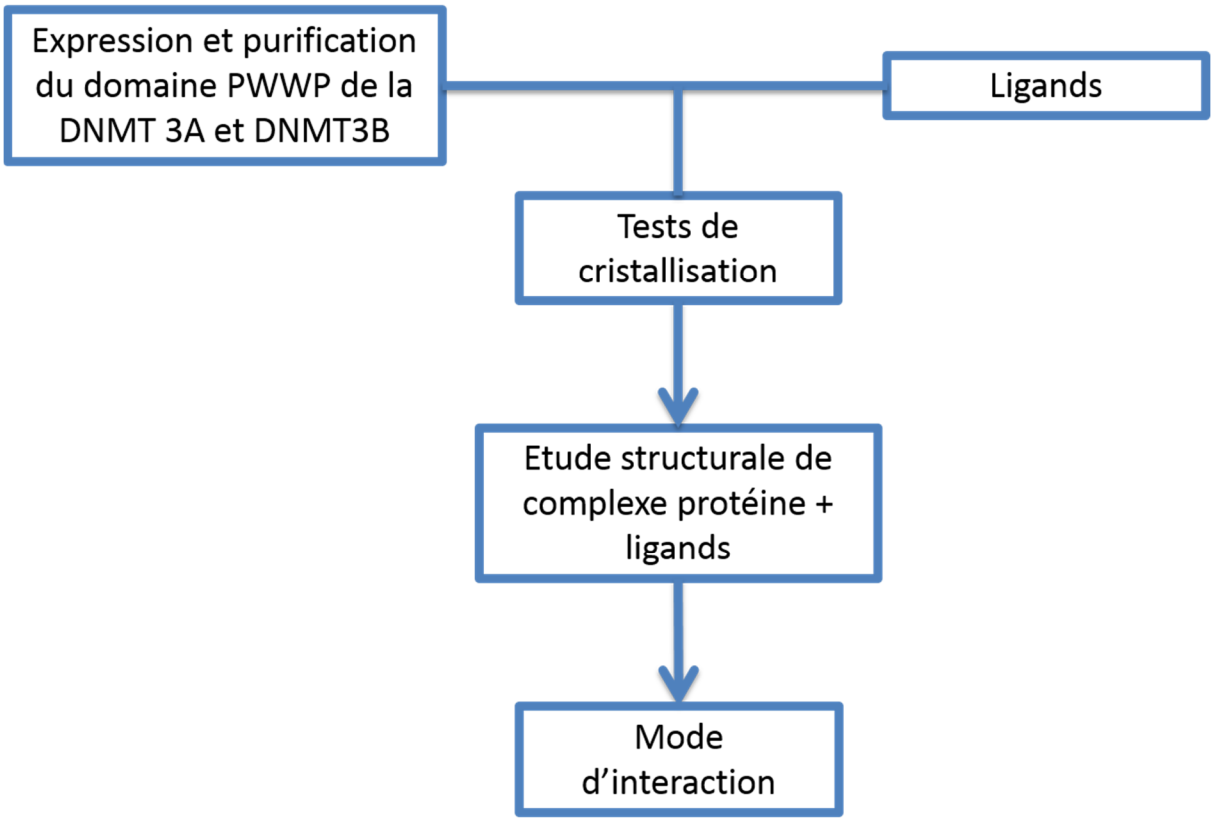


Figure 16 : Stratégie employée pour l'étude du mode d'interaction protéine-ligand.

La stratégie employée consiste en la production et purification du domaine PWWP de la DNMT3A et DNMT3B en vue d'obtenir le matériel de base nécessaire aux tests cristallographiques. Les ligands utilisés sont un peptide (contenant la H3K36me dont la séquence est la suivante : SAPATGGVT{Lys(me)$_3$}KPHRYR) et des analogues du bis-tris identifiés au sein du laboratoire et la structure de ceux qui seront testés sont repris à la figure suivante (voir Fig 17).

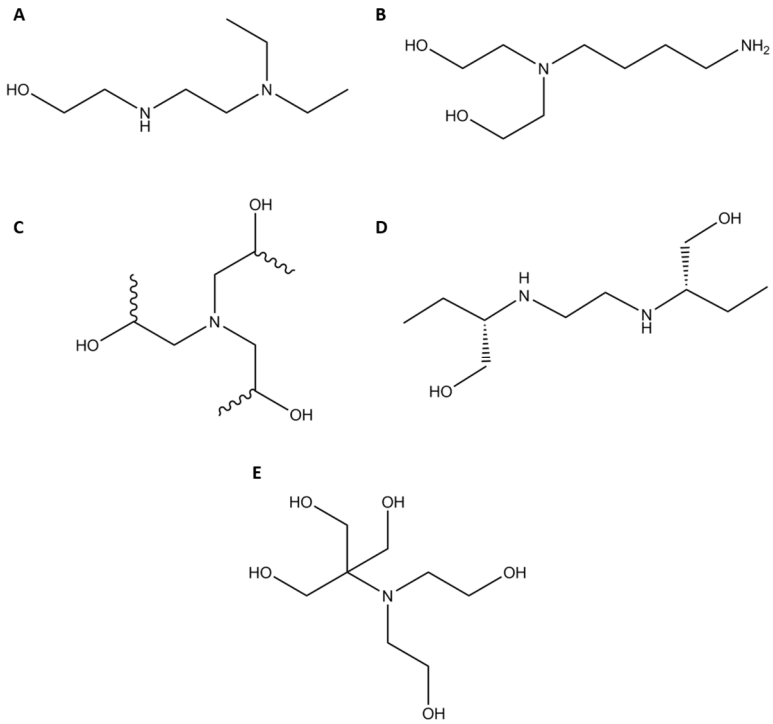


Figure 17 : Ligands utilisés pour les tests de cristallisation. A: 2-(2-diethylaminoethylamino)-éthanol, B : 2,2'-(4-aminobutylazanediyl)-diéthanol, C : Triisopropanolamine, D : (2S,2'S)-2-[2-(1-hydroxymethyl-propylamino)-ethylamino]-butan-1-ol (éthambutol) et E : Bis-tris.

Une fois le matériel de base obtenu (protéine purifiée et ligands), des tests de co-cristallisation vont pouvoir être réalisés. Les cristaux obtenus pourront être étudiés par diffraction de rayons X (DRX) et la structure pourra être résolue, ce qui conduira à une étude du mode d'interaction entre la protéine et le ligand.

3 Résultats expérimentaux et discussion

3.1 Etude structurale préliminaire

Au cours de ce travail, le domaine PWWP des DNMT3A et DNMT3B sera étudié. Cette étude passe par la surexpression, la purification, la cristallisation et le traitement des données de diffraction. Pour ce faire, une étude structurale préliminaire est nécessaire. Celle-ci sert de base pour l'établissement d'une stratégie permettant un bon choix de ligands pour les tests cristallographiques.

Des tests in vitro ont montré que le domaine PWWP était capable de reconnaitre spécifiquement la queue de l'histone 3 et plus particulièrement la lysine 36 triméthylée (H3K36me3) (Guo, Wang et al. 2014). L'équipe de recherche de Wu et al. a pu mettre en évidence le domaine PWWP de la DNMT3A (3LLR) et DNMT3B (3QKJ) en complexe avec une molécule de bis-tris (molécule tampon présente dans le milieu de cristallisation). Le domaine PWWP est donc capable de reconnaitre certaines molécules en plus de son ligand naturel. Bien que le domaine PWWP puisse interagir avec la H3K36me3 (morceau de la queue de l'histone 3 contenant la lysine triméthylée), aucune structure cristallographique de ce complexe n'a encore pu être obtenue pour les domaines PWWP des ADN méthyltransférases. Cela nous conduit donc à réaliser une première étude structurale permettant de mettre en évidence les interactions du bis-tris au sein de la poche de reconnaissance du domaine PWWP de la DNMT3B/3A et du peptide (H3K36me3) en complexe avec un domaine PWWP d'une autre protéine, la BRPF1 (PDB : 3MO8). Les principales interactions intervenant dans la reconnaissance du ligand par la protéine sont les interactions π-cation. En effet, le bis-tris (pKa 6,5) est chargé positivement au pH du milieu (pH 6,0) tout comme la lysine triméthylée du peptide H3K36me3. Les interactions particulières du peptide avec la protéine seront investiguées par la suite dans ce travail.

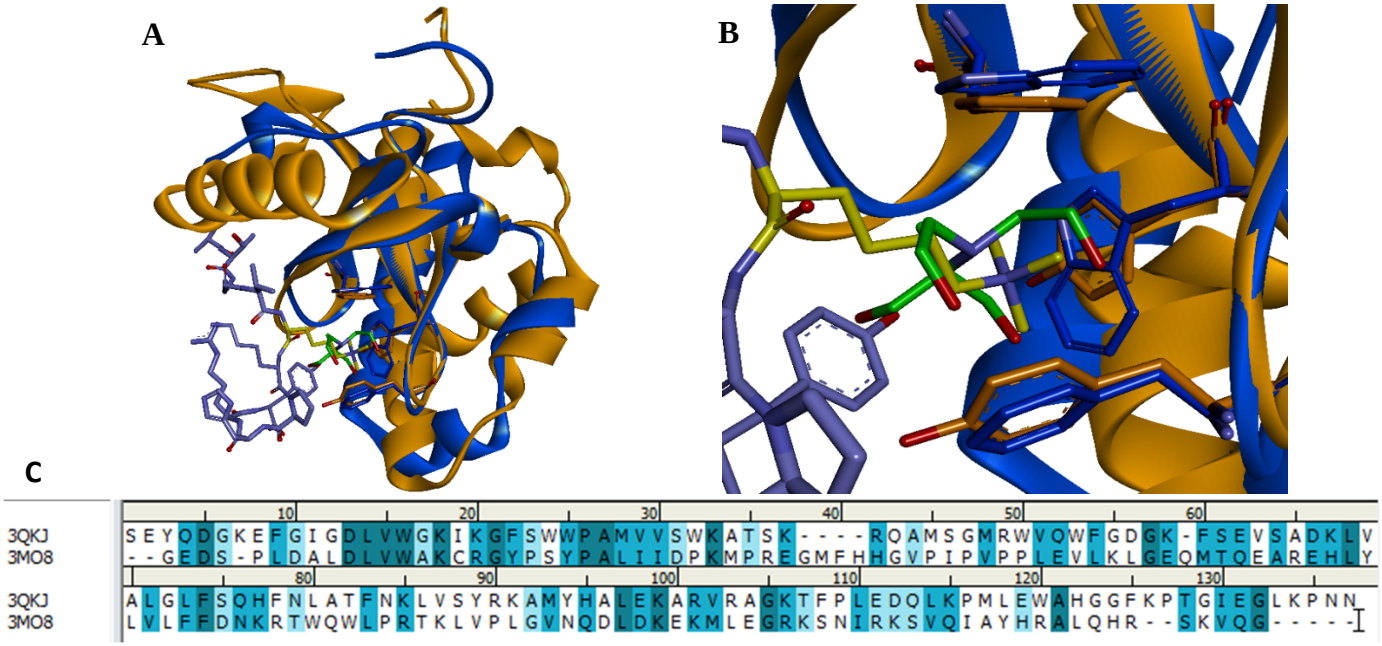


Figure 18 : **A** Superposition des complexes entre le domaine PWWP de la DNMT3B (orange) avec le bis-tris (vert) et le domaine PWWP de la BRPF1 (bleue) avec le peptide (cyan) contenant la lysine triméthylée (jaune). **B** Zoom sur la cage aromatique constituant le site. **C** Alignement de séquence du domaine PWWP de la DNMT3B (3QKJ) et du domaine PWWP de la BRPF1 (3MO8) (Accelrys 2013). En couleurs sont repris les résidus identiques et similaires entre les deux protéines.

Une superposition du domaine PWWP de la DNMT3B en complexe avec le bis-tris et du domaine PWWP de la BRPF1 en complexe avec le peptide H3K36me3 a pu être réalisée (voir Fig 18) et montre que le bis-tris reconnait bien le même site que la lysine triméthylée. Le RMSD (déplacement quadratique moyen) est de 0,44 Å pour les C_α du site de reconnaissance et l'alignement de séquence présente une identité de 14% et une similarité de 43%. La Fig 18 indique que la cage d'aromatique et le tonneau β se recouvrent correctement. Il est donc normal de penser que le peptide (H3K36me3) pourrait être observé au sein du site de reconnaissance du domaine PWWP de la DNMT3B/3A. La superposition de structure a été effectuée sur le domaine PWWP de la DNMT3B mais ce raisonnement semble parfaitement plausible dans le cas du domaine PWWP de la DNMT3A car ces deux protéines présentent une identité de séquence de 56,8% et une similarité de séquence de 71,2%. Elles ont donc une structure proche l'une de l'autre.

Cette étude montre que la lysine triméthylée et le bis-tris sont reconnus au sein du même site. Il pourrait donc être intéressant d'étudier le mode d'interaction du domaine PWWP de la DNMT3B en complexe avec le peptide H3K36me3. En effet, ces deux protéines (DNMT3B et BRPF1) ne présentent pas les mêmes résidus en surface et montreront donc une affinité différente pour le peptide.

La lysine triméthylée et le bis-tris étant reconnus, il pourrait être intéressant de se baser sur ces deux molécules pour le design et l'étude de nouveaux composés présentant une meilleure affinité pour le domaine PWWP. En effet, l'inhibition de la reconnaissance de marqueurs cancéreux (interaction protéine-protéine), comme la H3K36me3 par la protéine est un domaine de recherche en pleine expansion en cancérologie. Cela pourrait conduire à l'inactivation des DNMTs et par conséquent au maintien de la transcription des gènes suppresseurs de tumeurs. La compréhension de ces mécanismes passe par l'étude du mode d'interaction entre le peptide et le domaine PWWP.

Cette étude préliminaire pose donc les bases pour l'orientation et les choix de recherche de ce travail. Pour parvenir à étudier le mode d'interaction de la protéine, celle-ci va devoir être produite, purifiée et cristallisée en présence de ligands. Des comparaisons pourront être faites avec d'autres protéines possédant elles aussi un domaine PWWP.

3.2 Surexpression et purification

3.2.1 Principes

I. Culture et surexpression

La surexpression des protéines se déroule au sein de bactéries possédant un plasmide codant pour le gène d'intérêt. Les plasmides contenant le gène pour le domaine PWWP de la DNMT3A et de la DNMT3B ont été obtenus auprès de la société Addgene (plasmide 25328 et 32044 respectivement). Ils ont été transformés dans la bactérie *Escherichia coli Rosetta 2* dérivé de *BL21 (DE3)*. Celle-ci possède les codons rares (AGA, AGG, CUA, GGA, CCC et CGG) permettant l'expression d'une protéine humaine (Novagen 2003). Les bactéries contiennent

un gène de résistance au chloramphénicol et le plasmide, un gène de résistance à la kanamycine (voir Fig 19 A) . La combinaison de gènes de résistance aux antibiotiques permet de sélectionner les bactéries ayant intégré le plasmide et d'empêcher la prolifération d'autres bactéries contaminant éventuellement le milieu.

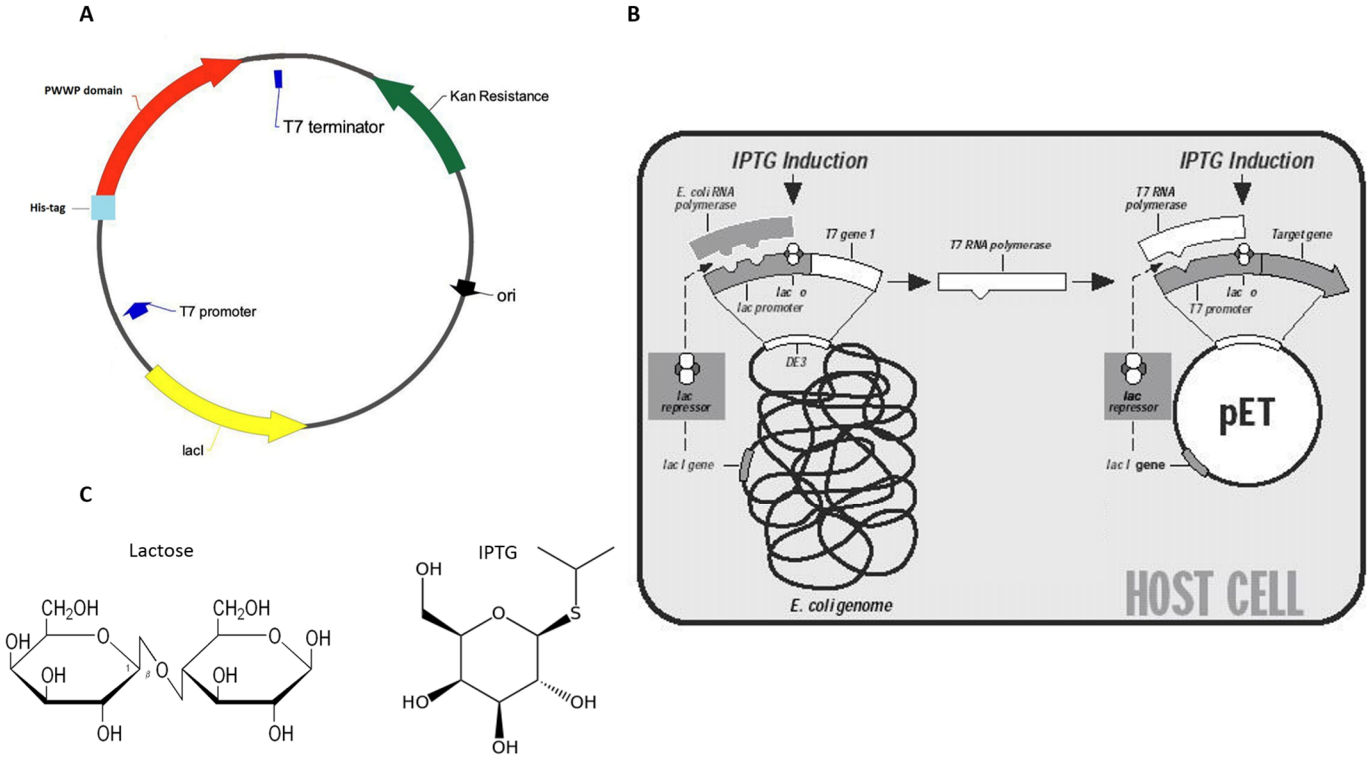


Figure 19 : **A** Schéma du plasmide codant pour le domaine PWWP de la DNMT3A et DNMT3B avec les principaux gènes intervenant dans la résistance et l'induction. **B** Schéma du contrôle du promoteur T7 pour l'induction à l'IPTG (Novagen 2010). **C** Structure du lactose et de l'isopropyl-β-D-1thiogalactopyranoside (IPTG).

Une culture cellulaire est donc réalisée et la division cellulaire des bactéries est stoppée lorsque la densité optique à 600 nm atteint 1,5 (voir Matériel et Méthode). A cette densité optique, le milieu contient assez de matériel biologique pour effectuer la surexpression de la protéine.

L'expression du gène codant pour les domaines PWWP des DNMT3A et DNMT3B est induite par l'IPTG (isopropyl-β-D-1-thiogalactopyranoside) sous le contrôle du promoteur T7 (voir Fig 19 B). Dans ce système de contrôle, le répresseur lac vient se lier sur les promoteurs

lac et T7 au niveau de l'opérateur lac O, ce qui empêche à la fois l'expression de l'ARN polymérase T7 et l'expression de la protéine d'intérêt. Le répresseur lac est codé par le gène lac I présent au niveau du génome de la bactérie *Escherichia coli* et du plasmide (double régulation). Pour induire l'expression de la protéine, il est donc nécessaire d'empêcher la liaison du répresseur lac au niveau des promoteurs lac et T7. Cela se fait au moyen de lactose qui est capable de se lier au répresseur lac et donc d'empêcher sa liaison au lac 0. Pratiquement, le lactose n'est pas utilisé mais son analogue l'IPTG (voir Fig 19 C) car le lactose, utilisé au sein de la bactérie, est dégradé (Novagen 2003).

Le même principe est utilisé pour la surexpression de la TEV protéase qui sera expliqué plus en détail par la suite. Les seules modifications sont les deux antibiotiques utilisés (voir Matériel et méthodes).

II. Lyse cellulaire

Une fois la culture cellulaire et la surexpression terminées, il faut récupérer la protéine d'intérêt. Pour cela, la membrane cellulaire doit être brisée au moyen d'une lyse cellulaire. Cette dernière consiste en une sonication (utilisation d'ultrasons) qui détruit la membrane des bactéries libérant ainsi tout le contenu cytosolique dans le milieu. Une centrifugation est ensuite essentielle de façon à séparer le surnageant contenant la protéine de tous les débris cellulaires.

III. Purification

Lors de l'expression de la protéine, celle-ci est produite avec une queue de six histidines (his-tag) à son extrémité N-terminale, permettant de purifier la protéine par chromatographie d'affinité. La chromatographie utilisée est l'IMAC (Immobilized metal ion affinity chromatography) chargée au Ni^{2+}. Ce type de colonne permet de complexer la queue histidine avec le nickel de façon à retenir la protéine au sein de la colonne alors que toutes les autres protéines seront éluées. La protéine d'intérêt sera détachée à l'aide d'un gradient en imidazole, ce dernier entrant en compétition avec les histidines pour la complexation à la colonne.

La protéine obtenue, une dialyse et un clivage de la queue histidine sont réalisés. La dialyse sert à diminuer la concentration en imidazole et le clivage de l'His-tag se fait au moyen de la TEV protéase (Tobacco etch virus) qui provient d'un virus contaminant la plante du tabac. Cette enzyme (cystéine protéase) reconnait une séquence d'acides aminés bien particulière et catalyse le clivage de la queue histidine de la protéine d'intérêt (voir Fig 20).

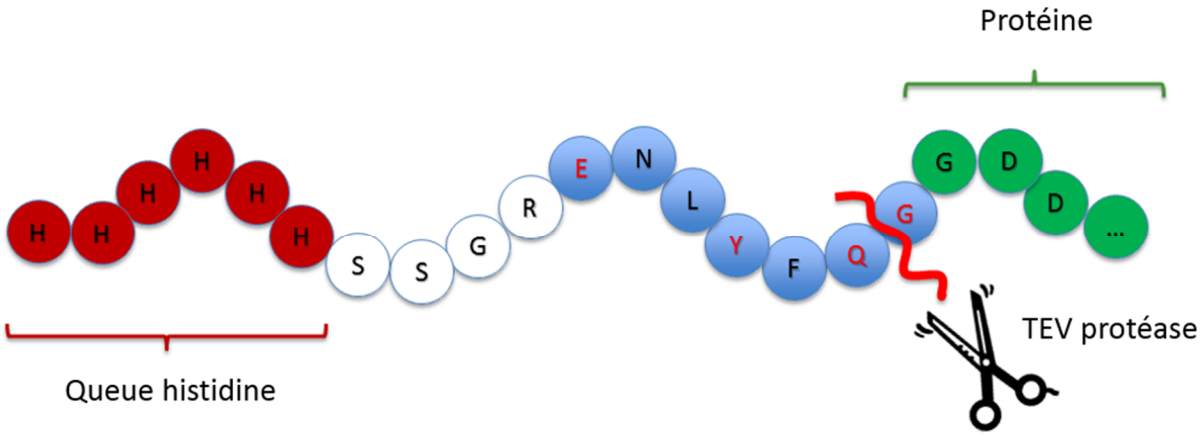


Figure 20 : Schéma de la queue histidine et de la séquence reconnue par la TEV protéase.

La TEV protéase reconnait le motif E-XX-Y-X-QG et clive entre les résidus glutamine (Gln, Q) et glycine (Gly, G).

Ensuite, une seconde purification IMAC est effectuée. Cette étape permet de retenir par la colonne l'His-tag clivé, la protéine non clivée et la TEV protéase qui possède elle aussi une queue histidine.

La dernière étape consiste en une chromatographie échangeuse d'ions qui permet d'homogénéiser la solution pour obtenir de meilleurs résultats lors des essais de cristallographie. La table 2 reprend les informations importantes (PI, poids moléculaire,…) concernant le domaine PWWP de la DNMT3A et DNMT3B nécessaire pour le choix du type de colonne et le pH du milieu de travail.

	PWWP DNMT3A		**PWWP DNMT3B**	
	Avec His-tag	Sans His-tag	Avec His-tag	Sans His-tag
Numéro plasmide (*Addgene*)	25328		32044	
Nombre de résidus	168	150	168	150
Motif conservé	SWWP		SWWP	
Point isoélectrique (PI)	6,23	5,45	8,72	8,83
Poids moléculaire (Da)	19 076,4	16 883,1	18 870,2	16 676,8
Coefficient d'extinction molaire ($M^{-1}.cm^{-1}$) (eau)	51 575	50 085	44 460	42 970

Tableau 2 : Informations importantes sur le domaine PWWP de la DNMT3A et DNMT3B. Les PI, poids moléculaire et coefficient d'extinction molaire ont été calculé à partir du programme ExPASy.

Le PI du domaine PWWP de la DNMT3A étant de 5,45 (sans His-tag) et le pH de la solution de 8,0, celui-ci est chargé négativement. Il est donc nécessaire d'utiliser une colonne échangeuse d'anions (Ressource 30Q). Dans le cas du domaine PWWP de la DNMT3B, le PI de la protéine (sans His-tag) est de 8,83 et le pH des solutions de 6,5. La protéine est dès lors chargée positivement et une colonne échange de cations (ressource S) doit être utilisée. Les protéines sont toutes les deux éluées au moyen d'un gradient en NaCl.

Dans le cadre de la purification de la TEV protéase, une complexation de l'ADN doit être réalisée après la lyse cellulaire. Une chromatographie d'affinité (IMAC) permet ensuite de récupérer la TEV protéase. La dernière étape consiste en une chromatographie d'exclusion de taille (tamis moléculaire Superdex S 75) pour récupérer la protéine d'intérêt.

IV. Méthode de caractérisation

- SDS-PAGE

La technique SDS-PAGE, pour Sodium dodecyl sulfate polyacrylamide gel electrophoresis, est utilisée pour déterminer la pureté d'une protéine. Elle permet de séparer les protéines selon leur poids moléculaire. Pour cela, un gel de polyacrylamide est utilisé. Tout d'abord, la protéine est dénaturée à l'aide d'un agent tensioactif, le SDS. Ce dernier va également permettre de charger la protéine négativement (une molécule de SDS pour 2 acides aminés). La migration des protéines dans la matrice est donc induite par l'application d'un potentiel. Les paramètres influençant la migration sont la taille de la protéine et le niveau de réticulation du gel. En effet, une protéine de grande taille migrera moins vite qu'une autre de plus petite taille. En jouant sur le degré de réticulation (et donc du pourcentage en acrylamide), il est possible d'optimiser le processus pour différentes tailles de protéines. Par exemple, un gel 15% en acrylamide sépare mieux les protéines entre 15 et 25 kDa alors qu'un gel 12% est plus efficace pour une gamme de protéines entre 25 et 80 kDa. Cette technique n'étant pas « native », il est nécessaire de dénaturer la protéine au moyen du SDS et de chauffer. La révélation des protéines sur le gel peut se faire de deux manières. La première consiste à utiliser le bleu de Coomassie qui se lie aux protéines et plus particulièrement aux acides aminés basiques et aromatiques par physisorption. La seconde est une coloration au nitrate d'argent. Les groupes thiols et carboxylique des chaines latérales des protéines vont interagir avec l'ion argent et le formaldéhyde présent dans le milieu le réduira. Ceci colorera donc les zones où les protéines sont présentes. L'utilisation d'une de ces deux techniques va surtout dépendre de la sensibilité voulue. Ainsi, la coloration au nitrate d'argent est environ 1000 fois plus sensible (nM) que la coloration au bleu de Coomassie (µM). La masse moléculaire de la protéine peut être déterminée au moyen d'un marqueur de poids moléculaire qui contient un mélange de protéines de masses connues.

- Absorbance à 280 nm

Les acides aminés aromatiques absorbent à une longueur d'onde de 280 nm (UV). En utilisant la loi de Beer-Lambert (voir équation 1), il est possible de déterminer la concentration d'une protéine.

$$A = \varepsilon . C . l \ (1)$$

Où A est l'absorbance, ε le coefficient d'extinction molaire, C la concentration et l la longueur de la cellule. Les différentes grandeurs utiles à la mesure de la concentration de la protéine sont reprises à la table 2.

3.2.2 Partie expérimentale

La Figure 21 présente les étapes qui ont été réalisées pour produire et purifier le domaine PWWP de la DNMT3A et DNMT3B.

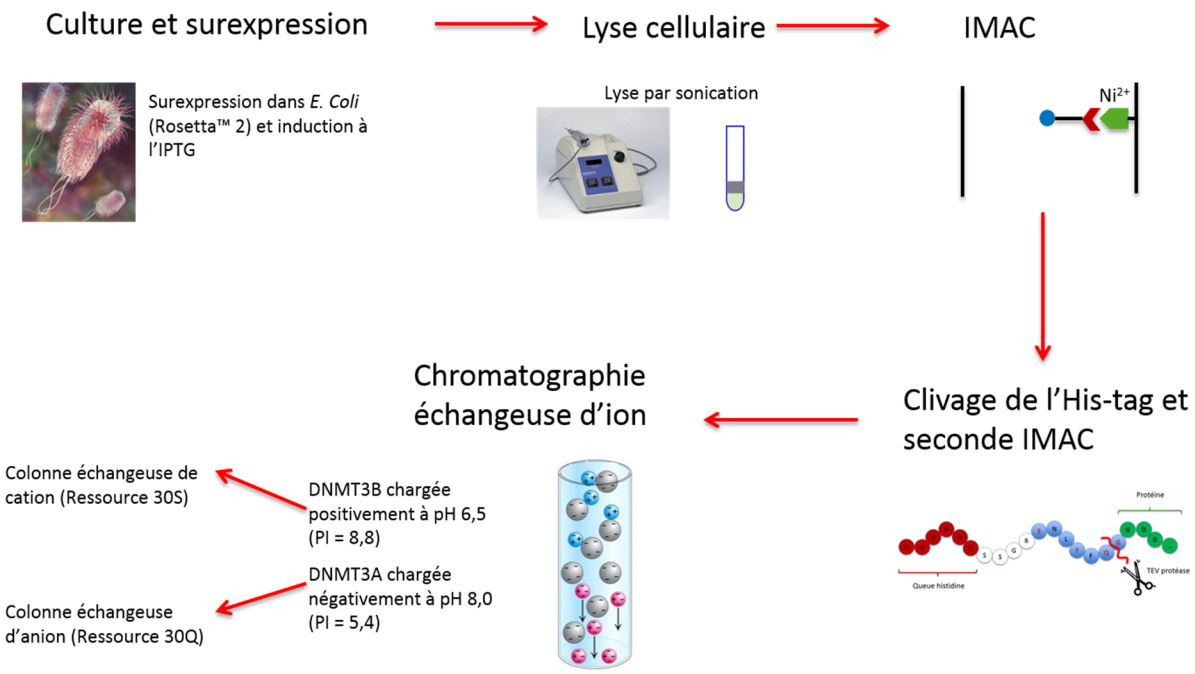


Figure 21 : Schéma présentant les différents procédés utilisés pour obtenir le domaine PWWP de la DNMT3A et DNMT3B purifié.

I. Surexpression du domaine PWWP de la DNMT3A et de la DNMT3B

Les différentes étapes de purification ont été suivies par SDS-PAGE (voir Fig 22). Les deux domaines ont bien été exprimés comme montré dans le puits numéro 2. On peut constater qu'après la lyse, la majorité de la protéine est récoltée dans le surnageant (puits 4). Cette dernière fraction est gardée pour les étapes ultérieures de purification. Par soucis de clarté, seul le gel de la DNMT3B sera présenté.

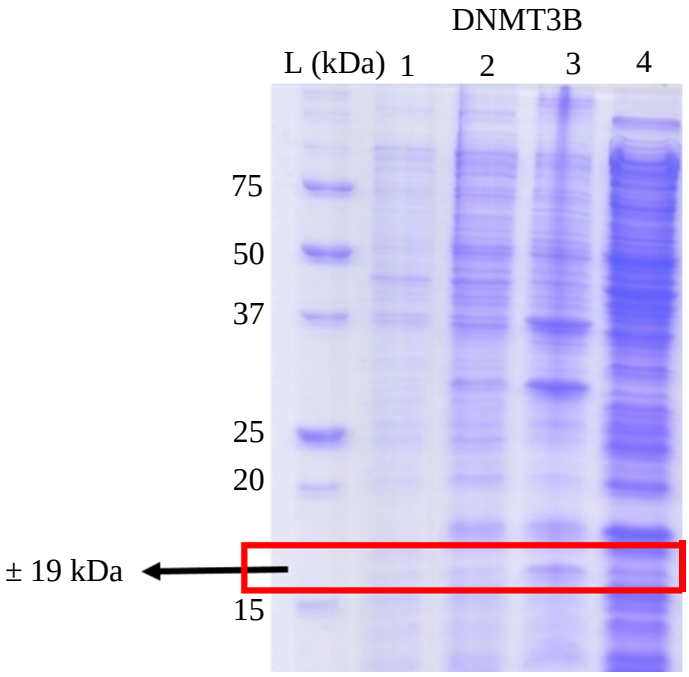


Figure 22 : Gel d'électrophorèse reprenant les résultats de la surexpression du domaine PWWP de la DNMT3B (± 19 kDa avec l'His-tag) : L : marqueur de poids moléculaire ; 1 : avant induction ; 2 : après induction ; 3 : culot après lyse ; 4 : surnageant après lyse.

II. Purification de la DNMT3A et de la DNMT3B

Une première purification a été réalisée au moyen d'une chromatographie d'affinité (Immobilized ion metal affinity chromatography) au nickel (voir Fig 23). Un gradient linéaire d'imidazole a permis d'éluer la protéine d'intérêt. Après cette première étape de purification (voir Fig 23 A et B), les deux domaines PWWP sont purs comme constaté par SDS-PAGE (Figure 23 C puit 4). Ensuite, la queue histidine a été clivée par une TEV protéase. Une perte de poids de 2 kDa, correspondant à l'His-tag, est observée au puits 5 (Fig 23 C). La bande de la protéine est déplacée vers le bas et correspond à 17 kDa indiquant que le clivage s'est bien réalisé. Ensuite, une seconde IMAC est effectuée dans le but de retenir dans la colonne les His-tag clivés, la TEV protéase et la protéine d'intérêt éventuellement non clivée. Le puits 6 de la Figure 23 C montre que la protéine obtenue à la fin du procédé est bien pure.

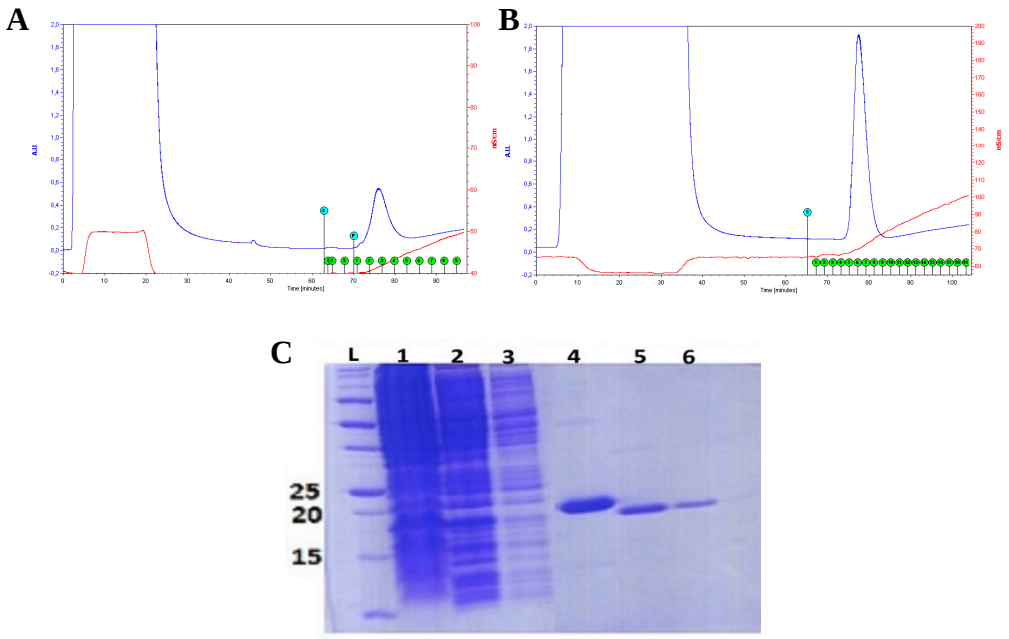


Figure 23 : Chromatogramme de purification du domaine PWWP de la DNMT3A et DNMT3B. **A** et **B** : première purification sur une colonne Hitrap (IMAC) pour la DNMT3A et DNMT3B respectivement. Conductivité en rouge et absorbance (à 280 nm) en bleu. **C** SDS-PAGE du domaine PWWP de la DNMT3B lors de la production et purification. L : marqueur de poids moléculaire ; 1 : culot après lyse ; 2 : surnageant après lyse ; 3 : « flow throught » IMAC1 ; 4 : élution IMAC1 ; 5 : après dialyse et clivage ; 6 : IMAC2.

L'étape suivante consiste en une colonne échangeuse d'ions. Suite au calcul des PI pour les deux protéines (voir Table 2) et de la valeur du pH des solutions de tampon utilisées, les deux protéines seront chargées différemment. Une colonne échangeuse d'anions (voir Fig 24 A) devra être utilisée pour le domaine PWWP de la DNMT3A et une colonne échangeuse de cations (voir Fig 24 B) pour le domaine PWWP de la DNMT3B. Les deux protéines sont éluées au moyen d'un gradient linéaire en NaCl allant de 0 à 1M. Il a été constaté que dans le cas particulier du domaine de la DNMT3A, le pic n'était pas correctement résolu (voir Fig 24 A).

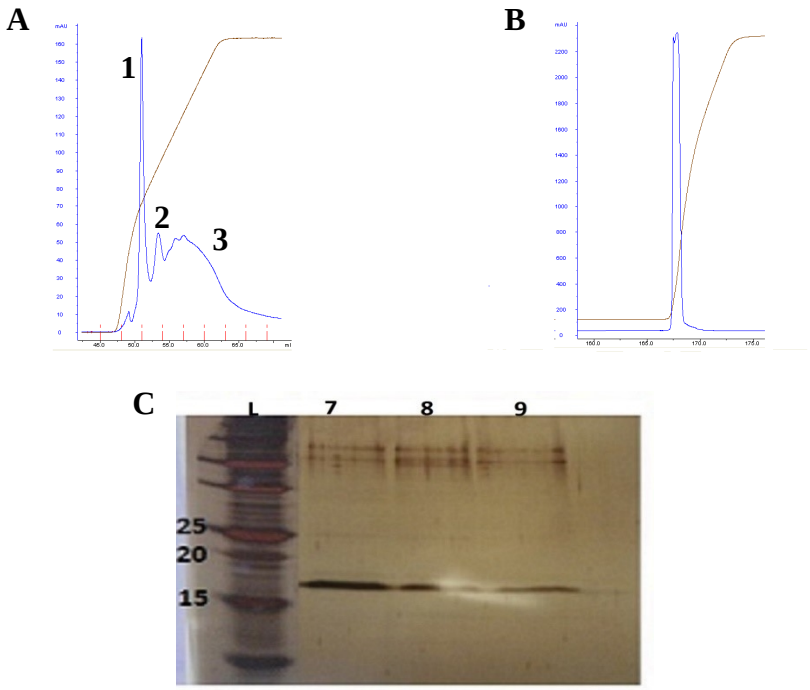


Figure 24 : Dernière étape de purification sur une colonne Ressource Q pour la DNMT3A (**A)** et sur une Ressource S pour la DNMT3B (**B)**. La conductivité est représentée en rouge et l'absorbance à 280 nm en bleu. **C** : SDS-PAGE du domaine PWWP purifié de la DNMT3A dans le cas d'une colonne échangeuse d'anions. L, marqueur de poids moléculaire (kDa) ; 1-3, fractions des différents pics.

La pureté des différents pics observés dans le cas de la DNMT3A a été vérifiée au moyen d'une coloration au nitrate d'argent (puits 1, 2, 3 de la Fig 24 C). Il peut être constaté que la protéine est bien présente et est pure en condition dénaturante. Cependant, le chromatogramme du domaine PWWP de la DMT3A laisse suggérer une inhomogénéité au sein de la population de la protéine. Chacune des trois populations (correspondant aux trois pics observés) présenterait un repliement légèrement différents et donc une charge en surface variable. Cela expliquerait donc les différents pics et la pureté observés.

A la fin de la purification, un rendement en protéines par litre de culture peut être calculé. Celui-ci est de 6 mg par litre de culture dans le cas du domaine PWWP de la DNMT3A

et de 30 mg par litre de culture pour le domaine PWWP de la DNMT3B. Cette différence de rendement peut être expliquée en partie par la grande différence de solubilité entre ces deux protéines. En effet, le domaine PWWP de la DNMT3B est soluble jusqu'à 45,5 mg/mL alors que dans le cas de la DNMT3A, la solubilité n'est pas supérieure à 5,8 mg/mL. Le domaine PWWP de la DNMT3A précipitera donc plus facilement et une plus faible quantité pourra être récupérée. Une fois la protéine purifiée, elle peut être utilisée pour les essais de cristallographie.

3.3 Cristallogenèse

3.3.1 Principes

Plusieurs techniques existent pour la formation de cristaux. Celle utilisée au cours de ce travail est la technique de la diffusion de vapeur (voir Fig 25). Elle peut être appliquée en goutte pendante et goutte assise. Cette dernière présente l'avantage de pouvoir travailler avec des volumes de goutte de solutions protéiques plus élevés. Pour la technique de la goutte pendante (voir Fig 25 A), un réservoir contenant une certaine concentration en agent précipitant est disposé dans le fond du puits alors qu'une petite quantité (1 µL) de solution contenant la protéine est placée sur une lame en verre. A cette goutte de protéine est ajouté un volume équivalent (1 µL) de la solution du réservoir. La lame de verre est finalement retournée (de façon à ce que la goutte (2 µL) soit dans le puits), le système est rendu hermétique au moyen de graisse de laboratoire. Dans le cas de la goutte assise (voir Fig 25 B), un support est disposé à l'intérieur de réservoir de façon à être surélevé par rapport à ce dernier. La goutte contenant la protéine (4 µL) sera déposée dans ce support et la solution concentrée en agents précipitants dans le réservoir. Le système sera rendu hermétique.

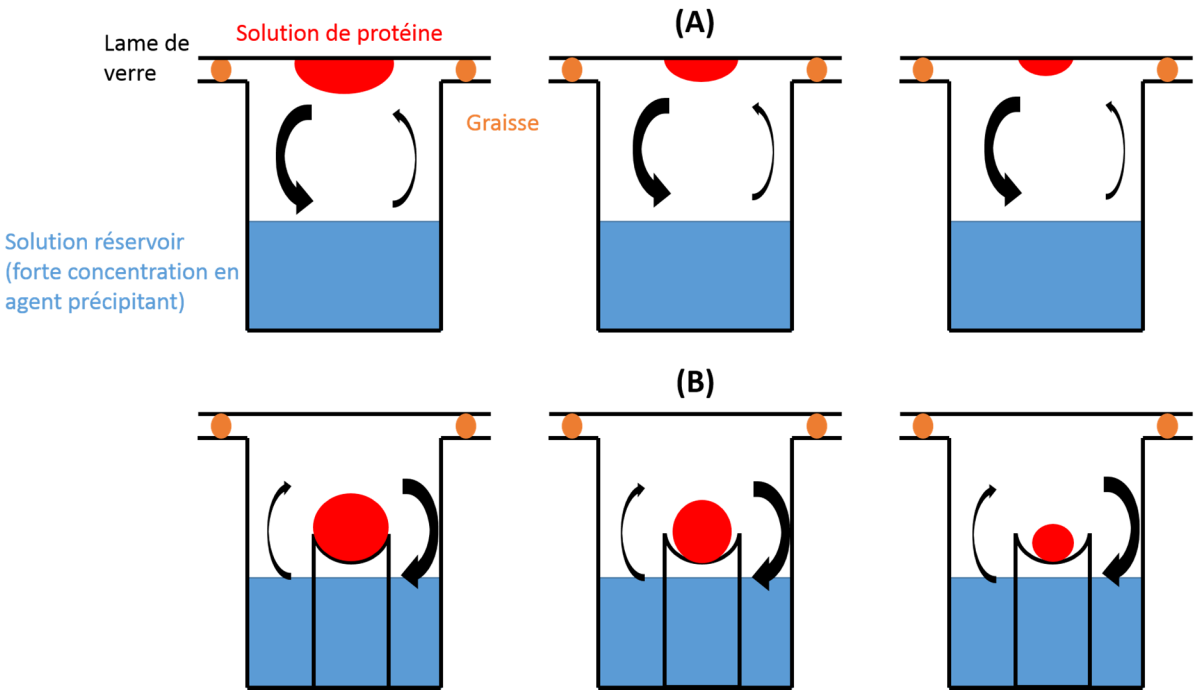


Figure 25 : Schéma représentant la diffusion de vapeur dans le cas de la goutte pendante (A) et de la goutte assise (B).

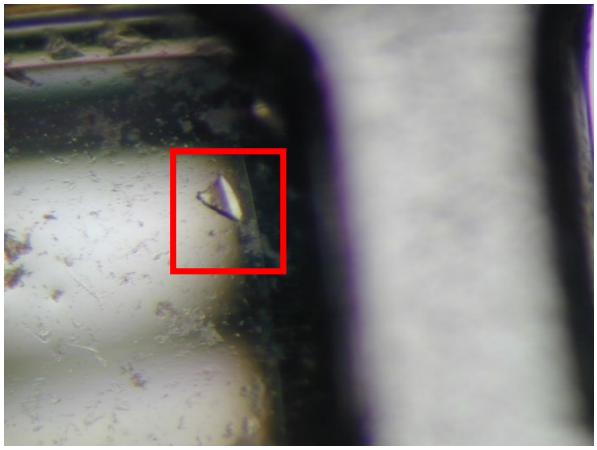
La goutte étant moins concentrée en agents précipitants que le réservoir, un phénomène de diffusion de vapeur va avoir lieu, la tendance étant de concentrer la protéine dans la goutte par diffusion du solvant vers le réservoir, laissant le temps aux protéines de s'organiser et de s'ordonner en un cristal.

3.3.2 Co-cristallisation du domaine PWWP de la DNMT3A et de la DNMT3B avec un peptide

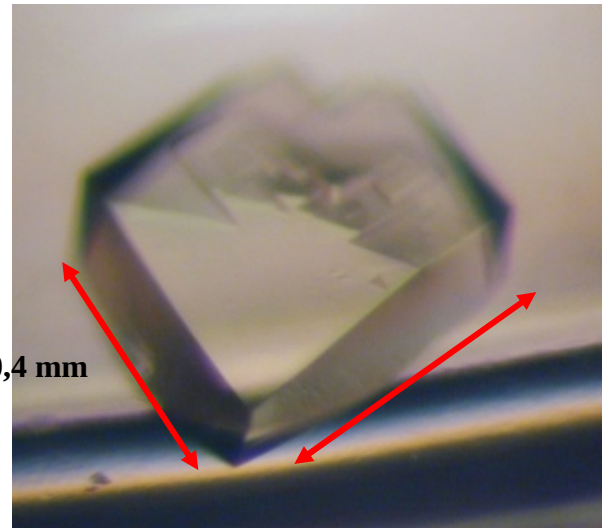
Le domaine PWWP de la DNMT3B et DNMT3A a déjà pu être cristallisé par l'équipe de Wu et al., (Wu, Zeng et al. 2011) les conditions de cristallisation sont donc déjà connues. Cependant, celles-ci présentent un désavantage : elles contiennent du bis-tris (molécule tampon) qui est capable de se placer au sein du site de reconnaissance du domaine PWWP (voir point 3.1 Etude préliminaire). Cette molécule peut donc entrer en compétition avec d'autres ligands potentiels qui seront testés. Ces conditions seront donc reprises et modifiées

de façon à enlever ou à diminuer la concentration en bis-tris du milieu. Le but de ces essais est d'observer le complexe protéine-ligands et de mettre en évidence le mode d'interaction.

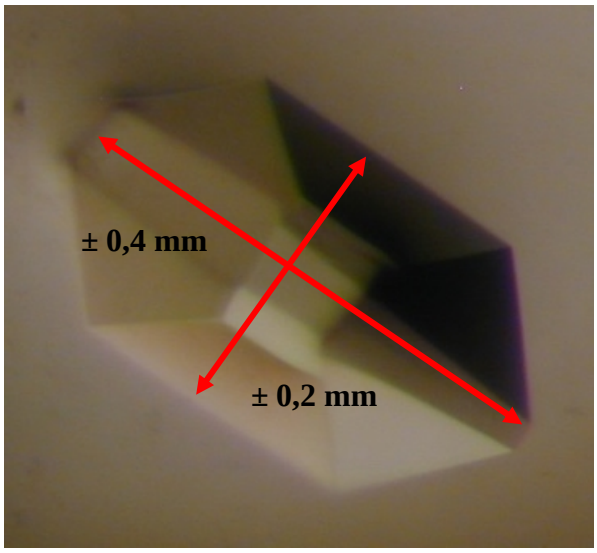
Le peptide utilisé est un segment de la queue de l'histone 3 contenant la lysine 36 triméthylée. La séquence en acides aminés est la suivante : SAPATGGV{Lys(me)$_3$}KPHRYR. Les conditions testées sont reprises dans la partie « Matériel et méthodes ». Ci-dessous (voir Fig 26) sont reprises les conditions ayant conduit à l'obtention de cristaux pour le domaine PWWP de la DNMT3B en complexe avec le ligand.



PWWP DNMT3B + peptide
[Prot] = 22,0 mg/mL
Ratio 1 : 5
Citrate de sodium tribasique dihydrate 1,6 M pH 6,5



PWWP DNMT3B + peptide
[Prot] = 22,0 mg/mL
Ratio 1 : 5
Citrate de sodium tribasique dihydrate 1,6 M pH 6,5



PWWP DNMT3B + peptide
[Prot] = 22,0 mg/mL
Ratio 1 : 5
Tartrate de potassium-sodium tétrahdrate 0,2 M
Sulfate d'ammonium 2,0 M

Figure 26 : Images des différents cristaux obtenus et leurs conditions de cristallisation.

Les quelques cristaux obtenus sont relativement de grosse taille et peu nombreux. Les deux derniers cristaux obtenus ont été analysés par diffraction des rayons X au synchrotron SOLEIL à Paris. Le second cristal n'ayant pas diffracté correctement, seul le pattern de diffraction du premier sera présenté et analysé au point suivant « Analyse des données cristallographiques et interprétations ».

Bien que des tests cristallographiques aient été effectués sur le domaine PWWP de la DNMT3A, aucun cristal n'a encore pu être obtenu. Cela s'explique notamment par la solubilité de la protéine qui nous oblige à travailler à une concentration protéique beaucoup plus faible que dans le cas de la DNMT3B et pourrait être également liée à l'inhomogénéité de la protéine comme expliqué précédemment lors de la purification sur colonne échangeuse d'ions.

3.3.3 Co-cristallisation du domaine PWWP de la DNMT3B avec des ligands

Wu et al. ont montré que le bis-tris était un ligand du site de reconnaissance du domaine PWWP des DNMTs. Des analogues du bis-tris ont donc été testés pour tenter d'obtenir des cristaux du complexe protéine-analogues du bis-tris. Ces ligands ont été identifiés au sein du laboratoire. Les ligands testés sont repris ci-dessous (voir Fig 27).

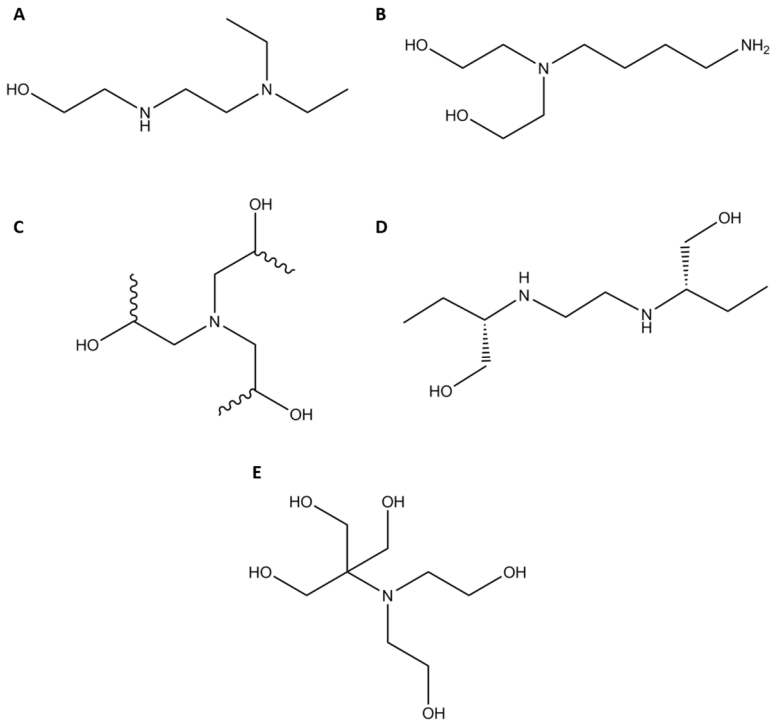
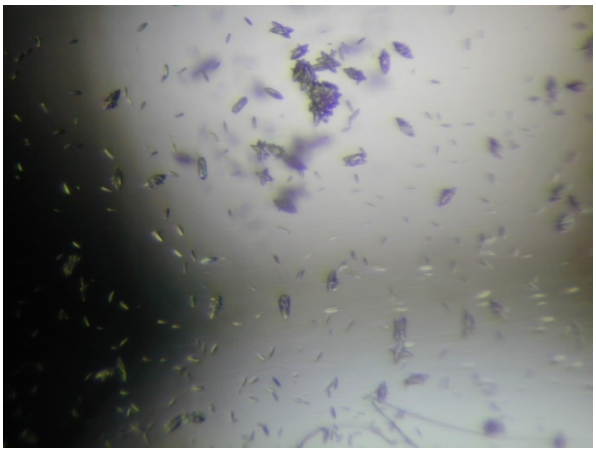
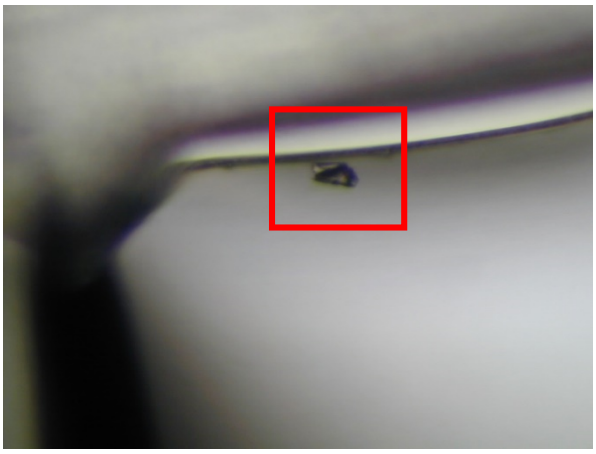


Figure 27 : Ligands utilisés pour les tests de cristallisation. A: 2-(2-diethylaminoethylamino)-éthanol, B : 2,2'-(4-aminobutylazanediyl)-diéthanol, C : Triisopropanolamine, D : (2S,2'S)-2-[2-(1-hydroxymethyl-propylamino)-ethylamino]-butan-1-ol (éthambutol) et E : Bis-tris.

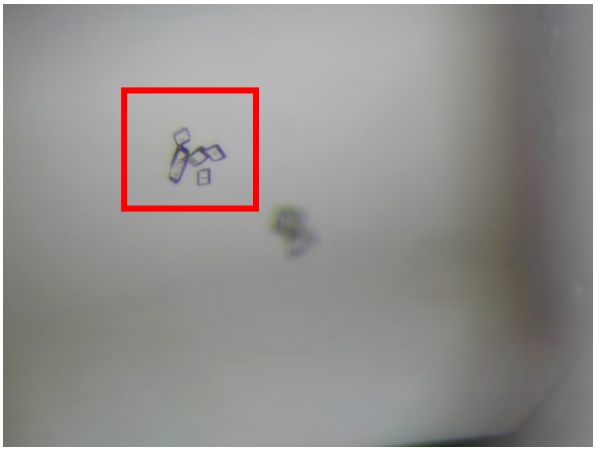
Les tests de co-cristallisation ont été effectués dans les conditions de recherche de Hampton Research pour tester la plus grande gamme de conditions cristallographiques possibles. Les conditions de Hampton Research reprennent une grande gamme de conditions cristallographiques dans lesquelles les protéines cristallisent en général. Des cristaux ont ainsi pu être obtenus pour certains de ces ligands dans des conditions particulières. Les photos et conditions sont reprises page suivante (voir Fig 28).



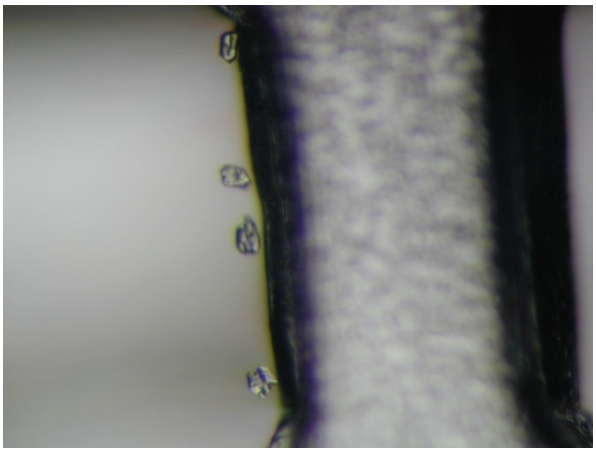
PWWP DNMT3B + 2,2'-(4-aminobutylazanediyl)-diethanol
[Prot] = 22,0 mg/mL
Ratio 1 : 5
Chlorure de nickel (II) hexahydrate 0,01 M
Tris 0,1 M pH 8.5
Sulfate de sodium monohydrate 1,0 M



PWWP DNMT3B + 2,2'-(4-aminobutylazanediyl-diethanol
[Prot] = 22,0 mg/mL
Ratio 1 : 5
Sulfate de zinc heptahydrate 0,01 M
MES monohydrate 0,1 M pH 6.5
Polyéthylène glycol monométhyl éther 550
25% v/v



PWWP DNMT3B + éthambutol
[Prot] = 22,0 mg/mL
Ratio 1 : 5
Sulfate d'ammonium 0,2 M
Cacodylate de sodium trihydrate 0,1 M
pH 6,5
Polyéthylène glycol 8 000 30% m/v



PWWP DNMT3B + 2-(2-diethylaminoethylamino)-ethanol
[Prot] = 22,0 mg/mL
Ratio 1 : 5
Sulfate d'ammonium 0,2 M
Cacodylate de sodium trihydrate 0,1 M
pH 6,5
Polyéthylène glycol 8 000 30% m/v

Figure 28 : Images des différents cristaux obtenus et leurs conditions de cristallisation.

Par rapport aux cristaux observés pour le domaine PWWP de la DNMT3B avec le peptide, les cristaux des complexes protéines-ligands sont beaucoup plus petits et beaucoup plus nombreux. Les cristaux obtenus n'ont pas encore pu être analysés par diffraction des rayons X. Certains promettent cependant d'être intéressants notamment les cristaux du complexe entre domaine PWWP de la DNMT3B et de l'éthambutol. En effet, cette molécule est déjà connue comme antituberculeux et pourrait donc être un médicament de premier choix si son action sur le domaine PWWP pouvait être avérée.

4 Analyses des données cristallographiques et interprétations

4.1 Principes

I. Diffraction des rayons X et synchrotron

La diffraction des rayons X est une méthode d'analyse physico-chimique fondée sur la diffraction des rayons X sur la matière cristalline. Les rayons X, comme toutes les ondes électromagnétiques, provoquent un déplacement du nuage électronique par rapport au noyau dans les atomes. Ces oscillations induites provoquent une réémission d'ondes électromagnétiques de même fréquence dans toutes les directions de l'espace : ce phénomène est appelé diffusion Rayleigh. Une des conditions nécessaire à la diffraction est que la longueur d'onde des rayons X utilisés doit être du même ordre de grandeur que la distance inter-atomique soit de l'ordre de l'Ångström (Å). On observera ainsi des interférences constructives et destructives selon les directions de diffusion. Les directions qui donnent lieu à des interférences constructives produisent un spot de diffraction et peuvent être trouvées via la construction de la sphère d'Ewald ou de la relation de Bragg :

$$2d\ \sin\theta = n.\lambda$$

Où d est la distance inter-réticulaire ; θ, le demi angle de diffraction ; n, un nombre entier et λ, la longueur d'onde des rayons X.

Les rayons X sont obtenus par transition électronique dans les couches internes atomiques et dans le spectre du rayonnement électromagnétique émis par accélération ou décélération de particules. Dans le premier cas, les éléments le plus souvent utilisés sont le cuivre et le molybdène dont la raie Kα vaut respectivement 1,54 Å et 0,7 Å.

Dans le second cas, le dispositif (synchrotron) accélère des particules (électrons) et modifie leur trajectoire par un jeu de champs magnétiques. Ainsi, toute particule chargée soumise à une accélération (ou une décélération) émet un rayonnement électromagnétique. Les principales propriétés du rayonnement synchrotron sont :

- Une continuité spectrale (de l'infrarouge lointain aux rayons X),
- Une excellente focalisation des faisceaux,
- Une intensité du rayonnement de plusieurs ordres de grandeur supérieure à celle des générateurs de rayons X des laboratoires.

Il est dès lors possible d'utiliser une source de rayons X plus puissante et ayant une gamme de longueurs d'ondes modulable. Cela permet d'obtenir une bonne résolution pour les structures cristallographiques.

II. Programmes d'analyse des données

Le diffractomètre récolte une série d'images correspondant aux intensités diffractées. Ces tâches de diffraction sont indexées et intégrées au moyen du programme XDS. L'utilisation d'un module de *PHENIX* (*Xtriage*) (PHENIX 2010) permet de déterminer la qualité des données collectées. Le module *phaser-MR* permet de faire du remplacement moléculaire à partir d'une structure déjà existante de la protéine pour résoudre le problème de phase et ainsi calculer la carte de densité électronique. Ensuite, un affinement de la structure est réalisé grâce au module auto-build de *PHENIX* qui est une automatisation de toute une série d'affinement dont un « *rigid body* ». Les étapes d'affinements ultérieures seront effectuées via le module *phenix.refine*. Le peptide n'étant pas repris dans la structure PDB de la protéine, il est primordial de le construire et de le placer au sein de la structure. Cette opération est réalisée via *Coot* et un fichier reprenant les différentes contraintes pour la triméthyle lysine est généré dans *JLigand*. Plusieurs étapes d'affinement et de validation de la structure sont nécessaires pour obtenir la structure finale de la protéine. Plusieurs programmes de visualisation peuvent être utilisés tels que *Pymol*, *DS Visualizer* ou encore *Coot*.

4.2 *Structure obtenue*

Les cristaux obtenus ont été analysés par DRX (diffraction des rayons) et une image de diffraction est reprise à la Figure 29.

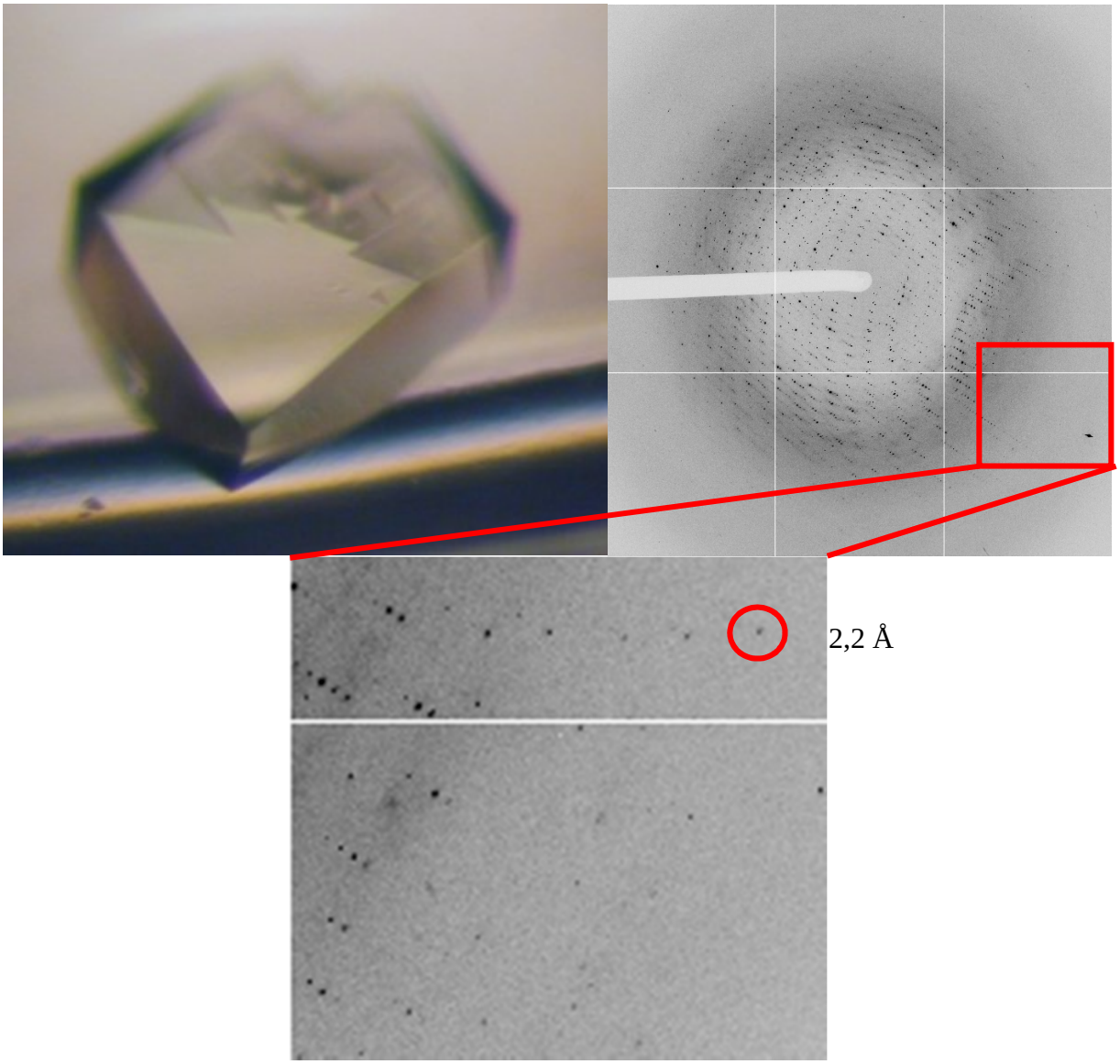


Figure 29 : Diffractogramme du cristal du complexe entre le domaine PWWP de la DNMT3B et le peptide dans les conditions suivantes : [Prot] = 22,0 mg/mL, ratio protéine-peptide 1 : 5, 1.6 M Sodium citrate tribasic dihydrate pH 6.5. La taille du cristal est de plus ou moins 0,4 mm sur 0,2 mm. Zoom sur sur une partie de l'image pour montrer la résolution maximum.

L'image de diffraction obtenue est bien celle d'une protéine car les spots observés sont proches les uns des autres. Cela correspond à des points proches dans le réseau réciproque, ce qui indique une distance inter-réticulaire élevée et une maille de grande taille (caractéristiques des macromolécules selon la loi de Bragg).

Ci-dessous se trouve la table (voir Table 3) des données cristallographiques correspondant à la structure obtenue.

Crystal Data	
Space group	P 32 2 1
Unit cell	a=b=73.437, c=158.2, α=β=90°, ϒ=120°
Subunit per asymmetric unit	3
Data set	
Resolution range (Å)	36.72 - 2.2 (2.279 - 2.2)
Unique reflections	25773
Completeness (%)	99.74 (99.92)
R_{merge} (%)	7.5 (52.9)
I/σ (I)	17.66 (2.56)
Redundancy	6.4
Refinement	
R_{factor}	0.2301 (0.3184)
R_{free}	0.2524 (0.3334)
RMS(bonds) (Å)	0.010
RMS(angles) (°)	1.32
Average B-factor ($Å^2$)	56.00
Ramachandran favored (%)	96
Ramachandran outliers (%)	1.9

Tableau 3 : Données de collection et d'affinement.

Il est intéressant de remarquer que les paramètres de maille obtenus sont identiques à ceux du domaine PWWP de la DNMT3B présent dans la littérature. A la fin du protocole d'analyse des données cristallographiques, de la densité inoccupée reste observable au niveau du site de reconnaissance de la protéine. D'abord, une lysine a été placée dans la poche de densité correspondante. Après une étape d'affinement, de la densité supplémentaire (densité verte) était encore observée au niveau de l'azote du résidu, indiquant clairement la présence des trois méthyles (voir Fig 30).

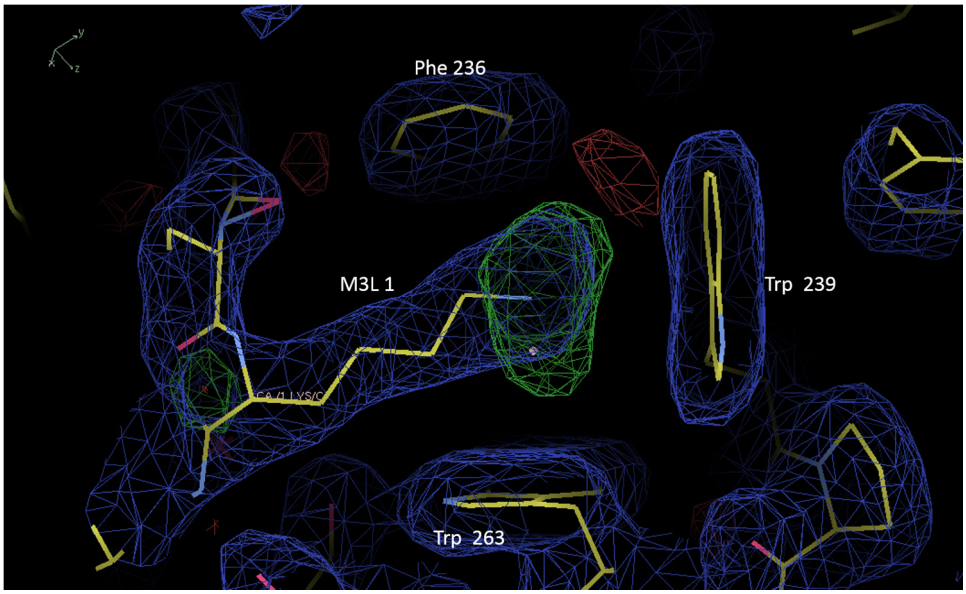


Figure 30 : Observation du peptide dans la poche de reconnaissance du domaine PWWP de la DNMT3B. Carte de densité électronique (bmeu, 2Fo-Fc). De la densité verte résiduelle (Fo-Fc) est observée au niveau de l'azote du résidu lysine, ce qui correspond à la densité électronique des 3 méthyles.

A partir de cette lysine, le restant du peptide a pu être construit et la structure du domaine PWWP de la DNMT3B en complexe avec le peptide a ainsi été obtenue (voir Figure 31 B). Cependant, tout le peptide n'était pas correctement défini, il manquait encore les extrémités de ce dernier (voir Fig 31 A). Sur les 15 acides aminés présents, seuls 7 ont pu être identifiés avec précision et seront par conséquent détaillés. Les résidus qui ont été observés sont les suivants (en rouge dans la séquence complète du peptide) : SAPATGGV{Lys(me)$_3$}KPHRYR. L'obtention du complexe protéine-ligand nous conduit à une étude du mode d'interaction entre le domaine PWWP et le peptide.

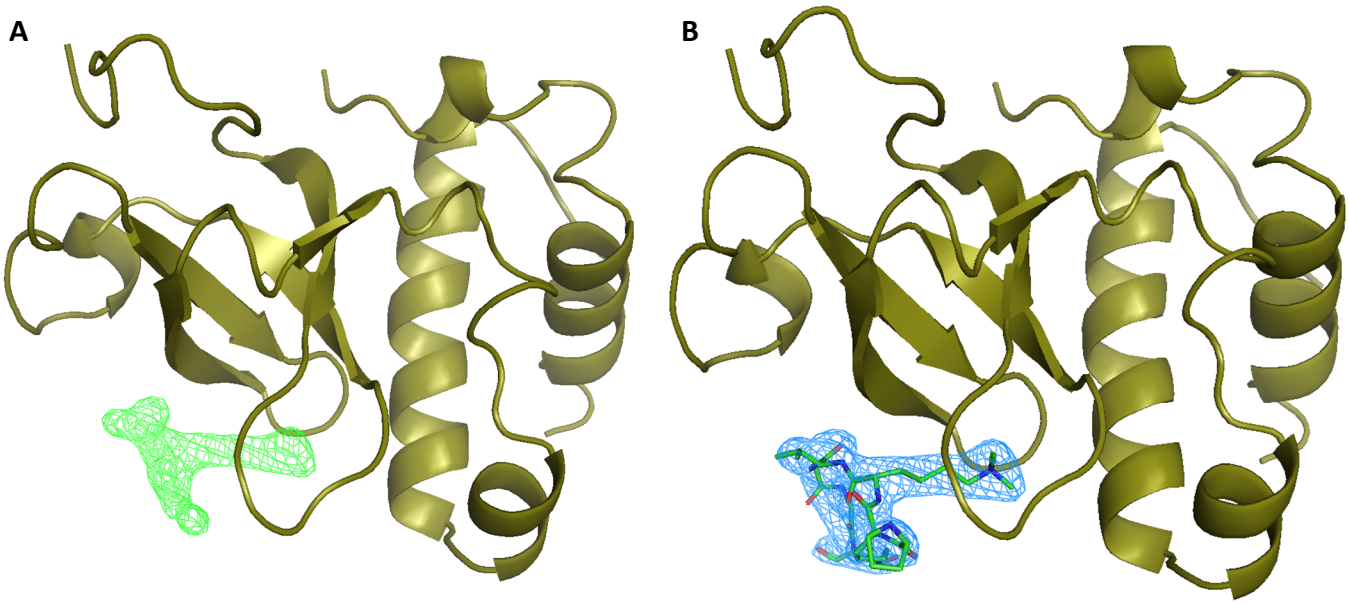


Figure 31 : Structure du domaine PWWP de la DNMT3B en complexe avec le peptide H3K36me3. **A** La carte Fo-Fc (en vert) représente la densité liée au peptide H3K36me. **B** En bleu est représentée la carte 2Fo-Fc du peptide H3K36me.

4.3 Analyses des différentes structures du domaine PWWP de la DNMT3B

Une superposition des trois structures du domaine PWWP de la DNMT3B (seule, avec Bis-tris et avec le peptide H3K36me3) a été réalisée dans le but d'observer des changements au niveau de la protéine lors de la reconnaissance du peptide. La superposition a mis en évidence (voir Fig 32) qu'un des trois acides aminés aromatiques, la Phe 236, faisait une rotation de 90° lors de la reconnaissance d'un ligand et se déplaçait d'environ 1,2 Å lors de l'interaction avec le peptide.

Le déplacement du résidu Phe 266 vers la lysine est dû à la position et à l'encombrement de la lysine triméthylée et plus particulièrement à la position de l'azote. Il faut également tenir compte de l'encombrement stérique du bis-tris au sein de la cage aromatique car ce ligand, bien que plus petit que le peptide, est plus encombrant que la lysine. De plus, comme ce sont surtout les interactions π-cation qui interviennent dans la reconnaissance du ligand, on peut

donc conclure que les interactions π-cation de la lysine triméthylée avec la Phe 266 sont plus importantes que dans le cas du bis-tris.

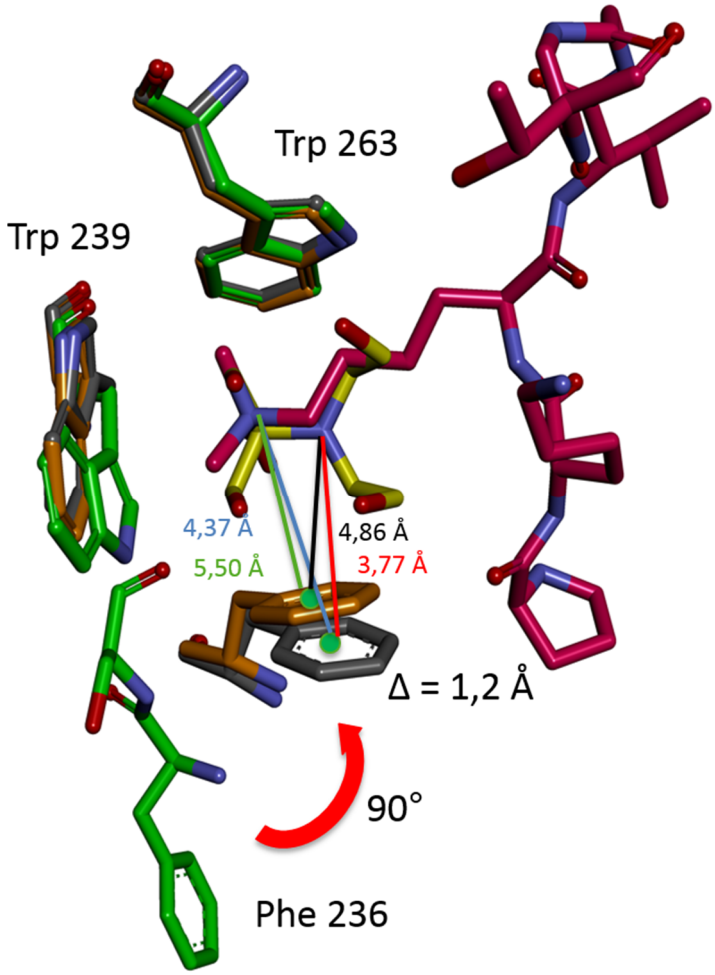


Figure 32 : Visualisation de la modification du site de liaison lors de l'interaction avec la lysine triméthylée du peptide. Les résidus du domaine PWWP de la DNMT3B seule sont représentés en vert, en complexe avec le bis-tris (jaune) en gris et en complexe avec le peptide H3K36me3 (rose) en orange.

4.4 Analyse du mode d'interaction et comparaison avec d'autres structures

Le but de cette partie consiste à relever les interactions observées entre le domaine PWWP de différentes protéines reconnaissant le peptide H3K36me3 (DNMT3B, BRPF1, PHF1 et PHF19) et un même peptide, la H3K36me3, contenue dans la séquence suivante : SAPATGGV{Lys(me)3}KPHRYR. Pour ce faire, une recherche a été effectuée dans la Protein Data Bank et diverses structures cristallographiques de protéines en complexe avec le peptide de la H3K36me3 ont été trouvées. Ces différentes protéines ont toutes en commun un domaine capable de reconnaitre la lysine triméthylée 36 de l'histone 3. Dès lors, il est possible de comparer les différentes structures (et de mettre en évidence des éléments de structures secondaires récurrents) de la protéine et du peptide. La séquence du peptide utilisé dans ces différentes structures correspond à celle utilisée dans ce travail. Cette analyse permettra donc d'avoir un regard plus global sur la structure du peptide et des interactions possibles entre ce dernier et les protéines le reconnaissant.

4.4.1 Analyse des structures

La structure du complexe et les interactions protéine-ligand ont été étudiées au moyen du programme *DS Visualiser* (Accelrys 2013). Les images présentées sont générées par *DS Visualiser* et *Pymol*. Les structures cristallographiques proviennent de la *PDB* (Protein Data Bank) ou de notre recherche.

I. PWWP DNMT3B-H3K36me3

Pour le domaine PWWP de la DNMT3B en complexe avec le peptide, plusieurs éléments de structures secondaires sont observés. On retrouve ainsi cinq hélices α et quatre feuillets β pour le domaine PWWP, ces derniers s'enroulant pour former une structure plus complexe appelée tonneau β. Cette structure est observable à la Figure 31 où l'on retrouve les feuillets β (schématisés par des flèches). Comme pour la structure développée par Wu et al., celle obtenue au cours de ce travail présente une partie manquante au niveau d'une des boucles de la protéine. En effet, ces boucles protéiques sont très mobiles et sont par conséquent difficilement observables par DRX.

Les interactions protéine-peptide ont été investiguées par le module « *Interaction ligand-receptor* » du programme *DS Visualiser* (voir Fig 33) et sont reprises dans un tableau (voir Table 4). Le site de reconnaissance est une cage d'aromatiques formée par les résidus Phe 236, Trp 239 et Trp 263. Les résidus du peptide interagissant avec la protéine sont M3L 36 (lysine triméthylée), Val 35 et Gly 33.

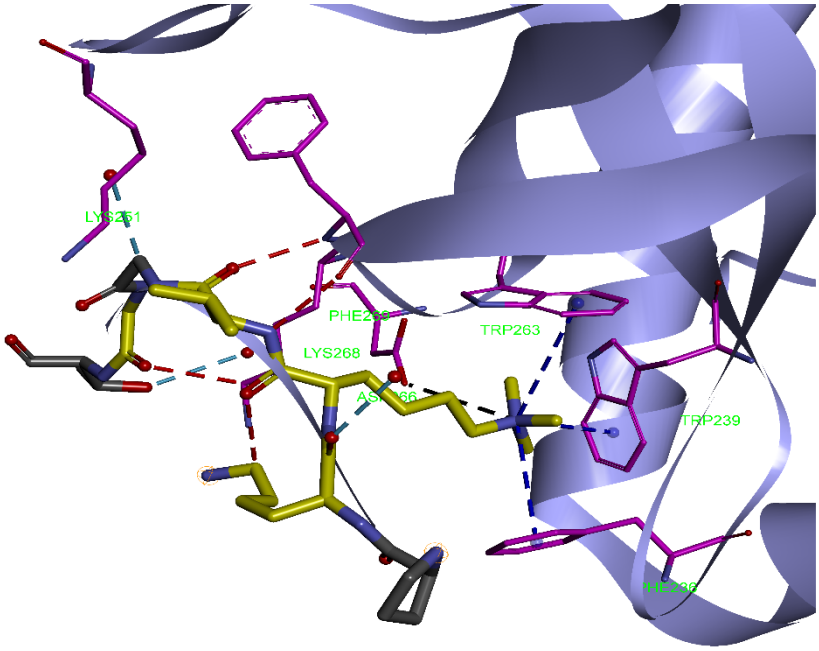


Figure 33 : Interactions entre le domaine PWWP de la DNMT3B et le peptide. Les résidus en jaune sont les acides aminés du peptide entrant en interactions avec ceux de la protéine (en rose). Les traits en pointillés représentent les interactions : π-cation (bleu foncé), pont H avec l'eau (bleu clair), pont H intermoléculaire (rouge) et électrostatique (noir).

Résidu 1	Résidu 2	Type d'interaction	Distance (Å) (H…A)	Distance (Å) (D…A)	Angle (°)
M3L 36	Phe 236	π-cation		4,37	
M3L 36	Trp 263	π-cation		4,97	
M3L 36	Trp 239	π-cation		4,35	
M3L 36	Asp 266	Électrostatique		4,09	
Val 35	Lys 251	Effet hydrophobe		5,32	
Val 35	Phe 269	Pont H (O…H-N)	1,79	2,78	169,3
Val 35	Phe 269	Pont H (N-H…O)	1,68	2,72	166,0
Gly 33	Lys 268	Pont H (O…H-N)	2,04	3,12	173,6

Tableau 4 : Tableau récapitulatif des interactions protéine-peptide dans le cadre du domaine PWWP de la DNMT3B. Dans le cas des ponts H, le premier terme de la parenthèse correspond à la partie du résidu 1 et un code couleur indique leur force : rouge (fort), bleu (modéré) et vert (faible).

Les interactions ayant le plus de poids au niveau énergétique sont les interactions électrostatiques, π-cation et pont H. Les ponts H étant des interactions directionnelles, il est important d'indiquer l'angle entre le donneur, l'hydrogène et l'accepteur (D-H…O). Plus cet angle est proche de 180°, plus le pont H sera fort (caractère covalent). Dans le cas présent, tous les ponts H sont modérés car la distance est importante et l'angle est inférieur à 175°. Il est important de préciser que la position des hydrogènes de la protéine est calculée par le module *BIOVIA* du programme *Discovery Studio (Accelrys 2013)*. Il ne s'agit donc pas d'une mesure expérimentale mais bien d'une valeur « calculée ». De plus, l'utilisation du terme « effet hydrophobe » doit être explicitée. Ce dernier comprend des interactions de type de Van der Walls et un effet entropique.

Au moyen des données reprises à la Table 4, une homologie entre les résidus du domaine PWWP de la DNMT3B entrant en interaction avec le peptide et ces mêmes résidus au sein du domaine PWWP de la DNMT3A a pu être réalisée. Il a pu être remarqué que ces derniers étaient bien conservés, ce qui renforce donc l'hypothèse de leur rôle dans la reconnaissance du peptide. La même observation a été faite quant à la position et l'orientation de ces mêmes résidus par rapport à ceux du domaine PWWP de la DNMT3B en complexe avec le bis-tris. Le peptide peut présenter plusieurs conformations en solution suite à sa petite taille et ses nombreux degrés de libertés. Il peut donc en être conclu que le peptide

va adapter préférentiellement sa conformation au site de reconnaissance de l'enzyme plutôt que l'inverse (tel qu'observé au sein de la cage d'aromatiques lors de la reconnaissance d'un ligand).

II. PWWP BRPF1-H3K36me3

La BRPF1 (bromodomain and PHD finger containing 1) est une protéine régulatrice associée aux HATs (histone acétyl transférases). On comprend dès lors pourquoi cette protéine est capable de reconnaitre la H3K36me3. La structure du domaine PWWP de la protéine (PDB 3MO8, résolution 1,8 Å) est présentée à la Figure 34 en superposition avec la structure du domaine PWWP de la DNMT3B. L'alignement de séquences des deux protéines (voir Fig 18 C) donne une identité de 13% et une similarité de séquence de 45%. Cette dernière valeur explique en grande partie la ressemblance entre les deux structures.

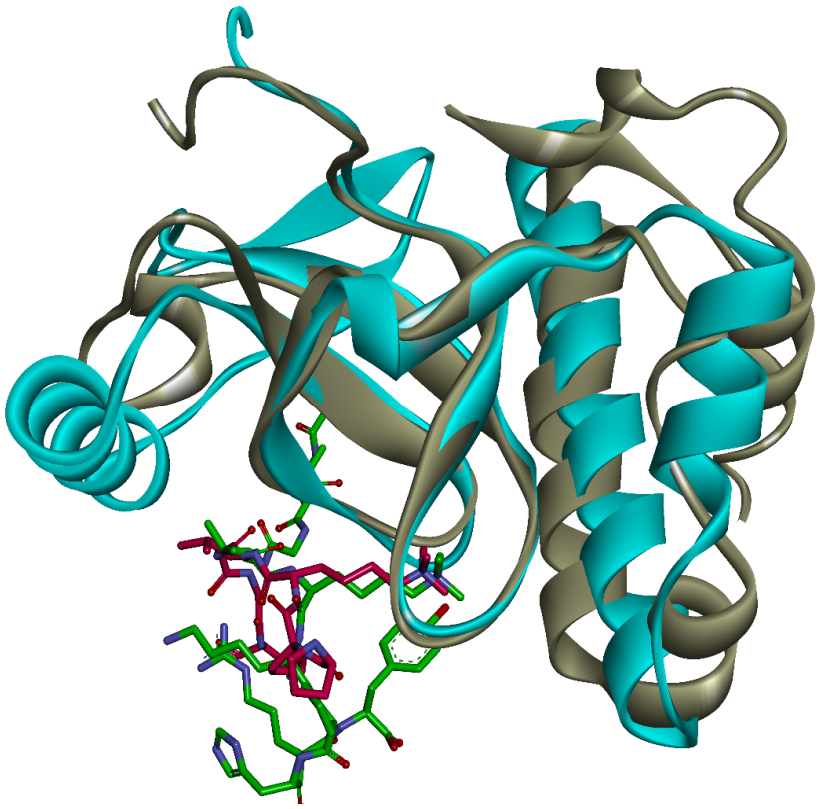


Figure 34 : Présentation de la superposition des structures du complexe domaine PWWP de la DNMT3B (gris)-peptide (rose) et du complexe BRPF1(bleu)-peptide (vert).

Comme il peut l'être constaté sur la figure 34, les deux protéines présentent bien une similarité dans leur structure. Elles sont toutes les deux composées d'un tonneau β et d'hélices α. Cependant, la BRPF1 présente seulement 4 hélices α. La lysine triméthylée se positionne bien de la même manière dans le site de reconnaissance. Le motif conservé dans le cas de la BRPF1 est PSYP et la cage d'aromatique est formée par les résidus Tyr 1096, Tyr 1099 et Phe 1147.

La figure 35 et la Table 5 reprennent les interactions observées entre le peptide et la protéine. Les principales interactions relevées sont les interactions π-cation, pont H et effet hydrophobe, bien que ce dernier intervienne le moins au niveau de l'énergie.

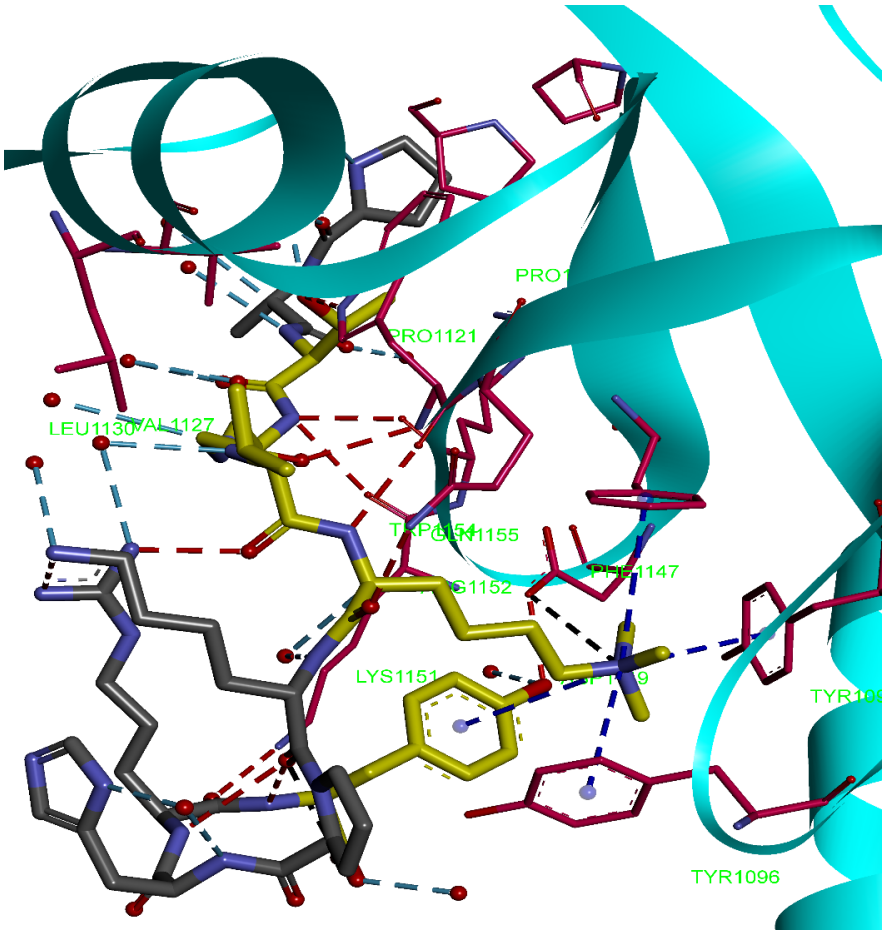


Figure 35 : Interactions BRPF1-peptide. Les résidus en jaune sont les acides aminés du peptide entrant en interaction avec ceux de la protéine (en rose). Les traits en pointillés représentent les interactions : π-cation (bleu foncé), pont H avec l'eau (bleu clair), pont H intermoléculaire (rouge) et électrostatique (noir).

Dans la Table 5, on observe que les résidus du peptide entrant en interaction avec la protéine sont les mêmes que dans le cas du domaine PWWP de la DNMT3B. Dans ce cas de figure, le nombre de ponts H est beaucoup plus important (suite à la conformation différente du peptide), laissant supposer une interaction plus forte entre la protéine et le peptide. On observe également des ponts H intramoléculaires au sein du peptide, ce qui n'était pas le cas pour le peptide en complexe avec la DNMT3B. Une autre observation importante a été faite. Il s'agit de la fermeture de la cage d'aromatique par la tyrosine (Tyr 12) du peptide. Cette dernière entre donc en interaction au niveau du site de reconnaissance. Cela n'a pas pu être observé dans le cas du peptide de la DNMT3B car la structure du peptide n'était pas suffisamment définie aux extrémités. Cependant, la superposition des deux structures du peptide (voir Fig 34) montre très peu de variation sur les 7 acides aminés identifiés (TGGV{Lys(me)$_3$}KP). Toutes les interactions sont reprises dans la Table 5.

Résidu 1	Résidu 2	Type d'interaction	Distance (Å) (H…A)	Distance (Å) (D…A)	Angle (°)
Pro 30	Pro 1119	Effet hydrophobe		4,99	
Pro 30	Pro 1121	Effet hydrophobe		4,45	
Thr 32	Trp 1154	Pont H (O-H…O)	1,94	2,90	148,7
Gly 33	Arg 1152	Pont H (N-H…O)	1,84	2,84	155,0
Gly 33	Lys 1151	Pont H (N-H…O)	2,55	3,08	109,9
Gly 34	Trp 1154	Pont H (O…H- N)	1,89	2,87	150,0
Val 35	Leu 1130	Effet hydrophobe		4,95	
Val 35	Val 1127	Effet hydrophobe		5,36	
Val 35	Lys 8	Effet hydrophobe		5,48	
Val 35	Arg 11	Pont H (O…H-N)	1,78	2,73	146,5
M3L 36	Tyr 12	π-cation		3,69	
M3L 36	Tyr 1096	π-cation		3,97	
M3L 36	Tyr 1099	π-cation		3,74	
M3L 36	Phe 1147	π-cation		3,89	
M3L 36	Tyr 1154	Pont H (N-H…O)	1,75	2,80	166,7
M3L 36	Gln 1155	Pont H (O…H-N)	1,87	2,87	153,5
M3L 36	Asp 1149	Electrostatique		4,54	
Lys 37	His 10	Effet hydrophobe		4,66	
Lys 37	Tyr 12	Pont H (O…H-N)	2,18	3,07	139,6
Lys 37	Arg 11	Pont H (O…H-N)	2,25	3,27	161,2
Tyr 41	Lys 1151	Effet hydrophobe		5,38	

Tableau 5 : Tableau récapitulatif des interactions protéine-peptide dans le cadre du domaine PWWP de la BRPF1. Dans le cas des ponts H, le premier terme de la parenthèse correspond à la partie du résidu 1. Un code couleur indique leur force : rouge (fort), bleu (modéré) et vert (faible).

Les résidus du peptide qui interagissent avec la protéine sont la Pro 1, Thr 3, Gly 4, Gly 5, Val 6, M3L 7 et Tyr 12. On retrouve donc bien la Gly 4, la Val 6 et la M3L 7 comme dans le cas de l'interaction du peptide avec le domaine PWWP de la DNMT3B.

III. PHF1-H3K36me3

La PHF1 (PHD finger protein 1) est une protéine régulatrice de la transcription. Cette protéine fait partie du complexe méthyltransférase spécifique de la lysine 27 de l'histone 3. Elle affecte l'expression des gènes en modifiant la structure de la chromatine. Cette protéine doit donc être capable de reconnaitre les marqueurs épigénétiques pour pouvoir en placer de nouveaux. En effet, elle fait partie des domaines Tudor qui sont de la même famille que les domaines PWWP.

La Figure 36 montre la superposition de la structure de la PHF1 en complexe avec la H3K36me3 (PDB 4HCZ, résolution 1,8 Å) et du domaine PWWP de la DNMT3B en complexe avec la H3K36me3. Cette protéine est beaucoup plus petite que le domaine PWWP de la DNMT3B ou que la BRPF1. Elle est seulement composée de 5 feuillets β (formant un tonneau β) et ne présente qu'une petite hélice α. Au niveau du site de reconnaissance, la cage d'aromatique est composée des résidus Trp 41, Tyr 47 et Phe 65.

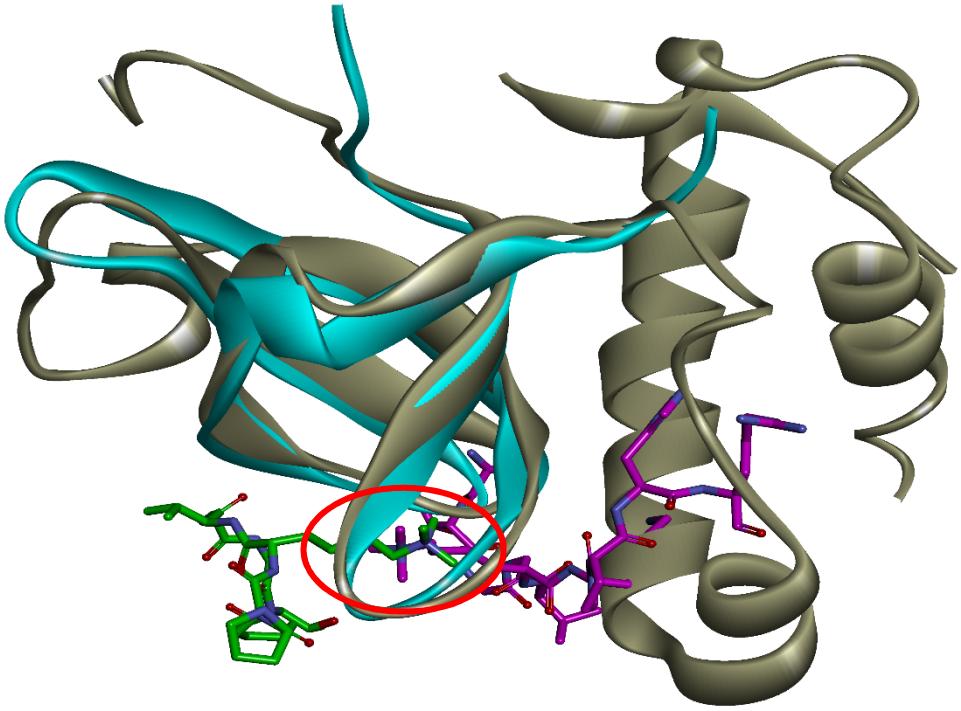


Figure 36 : Présentation de la superposition des structures du complexe domaine PWWP de la DNMT3B (gris)-peptide (rose) et du complexe PHF1 (bleu)-peptide (vert).

Comme la Figure 36 le montre, les deux protéines présentent une structure proche en ce qui concerne le tonneau β alors que les deux peptides ont une orientation complètement différente. Cela peut être expliqué notamment par la taille de la protéine et l'absence d'hélice α pour empêcher l'approche. En effet, la présence des hélices α entraine des contraintes pour l'approche et la reconnaissance du peptide par le site, ce qui limite considérablement ses conformations possibles. Cependant, il peut être remarqué que les deux azotes des lysines triméthylées occupent la même place au sein de la structure et que tous les deux sont reconnus par trois acides aminés aromatiques. La liste des interactions est reprise dans la Table 6. Celles-ci sont représentées à la Figure 37.

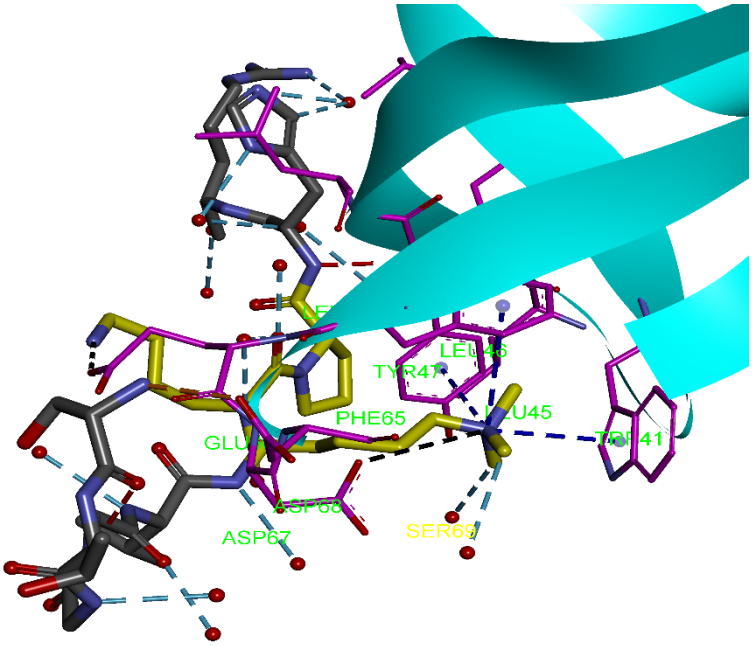


Figure 37 : Interactions PHF1-peptide. Les résidus en jaune sont les acides aminés du peptide entrant en interaction avec ceux de la protéine (en rose). Les traits en pointillés représentent les interactions : π-cation (bleu foncé), pont H avec l'eau (bleu clair), pont H intermoléculaire (rouge) et électrostatique (noir).

Les résidus du peptide entrant en interaction avec la protéine sont la Ser 31, la M3L 36, la Lys 37, la Pro 38 et la His 39. Dans ce cas-ci, seuls deux des trois acides aminés observés dans le cas du domaine PWWP de la DNMT3B (Gly, Val et M3L) interagissent. Cependant, il est difficile d'interpréter la structure du peptide car sa conformation n'est pas bloquée par des hélices α comme dans le cas du peptide de la DNMT3B. Il peut être également intéressant

de remarquer que le peptide n'établit que deux ponts H avec la protéine : il possède donc une plus grande mobilité que dans le cas de la H3K36me3 en interaction avec la DNMT3B.

Résidu 1	Résidu 2	Type d'interaction	Distance (Å) (H…A)	Distance (Å) (D…A)	Angle (°)
Ser 31	Asp 68	Pont H (N-H…O)	2,06	3,06	156,2
Ser 31	Gly 34	Pont H (O…H-N)	2,23	3,20	150,7
M3L 36	Trp 41	π-cation		4,24	
M3L 36	Tyr 47	π-cation		4,50	
M3L 36	Phe 65	π-cation		4,46	
M3L 36	Asp 67	Electrostatique		5,32	
Lys 37	Glu 66	Electrostatique		3,04	
Pro 38	Tyr 47	Effet hydrophobe		4,69	
Pro 38	Leu 45	Effet hydrophobe		5,00	
His 39	Leu 46	Pont H (N-H…O)	1,70	2,76	170,1
His 39	Leu 48	Effet hydrophobe		5,29	
His 39	Leu 38	Effet hydrophobe		3,50	

Tableau 6 : Tableau récapitulatif des interactions protéine-peptide dans le cadre de la PHF1. Dans le cas des ponts H, le premier terme de la parenthèse correspond à la partie du résidu 1. Pour les ponts H, un code couleur indique leur force : rouge (fort), bleu (modéré) et vert (faible).

Une structure de la PHF19, qui est une isoforme de la PHF1, a également été trouvée dans la PDB. L'analyse détaillée de cette structure en complexe avec la H3K36me3 ne sera pas présentée car les interactions sont identiques au cas de la PHF1 excepté dans le cas de la Tyr 41 qui entre en interaction avec la Tyr 55 de la protéine (T-shape). Le T-shape est une interaction quadrupolaire entre deux groupements aromatiques dont l'un se présente de manière perpendiculaire par rapport au plan du second.

IV. Autres structures contenant le peptide H3K36me3

D'autres structures en complexe avec la H3K36me3 ont été trouvées dans la PDB. Il s'agit de la KDM2A-H3K36me3 (PDB 4QXB, résolution 1,6 Å), la JMJD2A-H3K36me3 (PDB 2YBP, résolution 2,0 Å) et la PHF19 (PDB 4BD3, résolution 2,5 Å). Les deux premières protéines, KDM2A (pour « Lysine-specific demethylase 2A ») et JMJD2A (pour « Jumonji domain containing 2A »), sont toutes les deux des déméthylases de la queue des histones. Elles sont donc capables de reconnaitre la lysine 36 triméthylée de l'histone 3 pour effectuer

leur réaction. Cependant, ces deux protéines ne peuvent pas être analysées au même titre que le domaine PWWP de la DNMT3B car ce dernier ne possède aucune activité catalytique, bien que des ressemblances au niveau structural soient observables (tonneau β, hélices α). Il serait dès lors peu approprié de comparer des protéines avec des fonctions aussi différentes. Cependant, certains éléments de structures secondaires ont pu être observés tels que la présence d'un tonneau β composé de 13 feuillets β et la présence d'hélices α pour bloquer certaines conformations du peptide au sein du site catalytique. Le cas de la PHF19 a été abordé précédemment lors de l'analyse de la PHF1.

5 Conclusion

Les DNMT3 sont des enzymes intervenant dans l'établissement de la méthylation *de novo* de l'ADN et sont composés de trois domaines : PWWP-ADD-domaine catalytique. La plupart des recherches se sont d'abord intéressées au domaine catalytique de ces enzymes. Mais depuis quelques années, suite à la découverte du domaine PWWP et de son rôle au sein de la protéine, de plus en plus de chercheurs l'ont investigué pour étudier la reconnaissance de « marqueurs » cancéreux et l'établissement du pattern de méthylation. Ces études visent plus particulièrement l'inhibition de l'interaction protéine-protéine (domaine PWWP-histone) pour rétablir le pattern de méthylation et permettre la réexpression des gènes associés.

Le domaine PWWP joue un rôle central dans le recrutement des méthyltransférases au niveau des emplacements cibles de la chromatine. Il est maintenant bien établi que le domaine PWWP est un « reader » des marques épigénétiques comme la H3K36me3 et qu'il permet ainsi la méthylation de l'ADN. Les protéines possédant des domaines PWWP sont impliquées dans de nombreux processus biologiques (la différenciation cellulaire, l'inactivation du chromosome X,… et bien sûr la régulation des gènes). Le mauvais fonctionnement de ces protéines est à la base de nombreuses maladies humaines. Du fait de leur rôle important dans la régulation des gènes, des études récentes se tournent vers ces domaines pour développer des inhibiteurs capables d'empêcher la reconnaissance protéine-protéine (queue de l'histone-domaine PWWP).

Sur base de la structure cristallographique du domaine PWWP de la DNMT3A et de la DNMT3B en complexe avec le bis-tris et de la structure du complexe domaine PWWP de la BRPF1-H3K36me3, il a été montré que le domaine PWWP des DNMTs était susceptible de reconnaitre la lysine triméthylée. Dans la littérature, des tests *in vitro* ont démontré que le domaine PWWP de la DNMT3A était capable de reconnaitre cette lysine mais aucune évidence expérimentale n'est rapportée pour le cas de la DNMT3B.

Pour mettre en évidence le mode d'interaction de ces domaines PWWP, il a fallu les surexprimer au sein de bactéries et ensuite purifier la protéine au moyen de chromatographie d'affinité et échangeuse d'ions. Le domaine PWWP de la DNMT3B a été obtenu avec un bon

rendement (30 mg par litre de culture) alors que le domaine PWWP de la DNMT3A a été obtenu avec un rendement moindre (6 mg par litre de culture).

Une fois le domaine purifié, des tests cristallographiques ont été effectués pour obtenir des cristaux de complexe domaine PWWP-ligand, le ligand était soit, un peptide contenant la lysine triméthylée soit, un analogue du bis-tris. Des cristaux du complexe protéine-peptide ont pu être obtenus dans plusieurs conditions et analysés par diffraction des rayons X, conduisant à la résolution d'une structure cristallographique du complexe du domaine PWWP de la DNMT3B avec le peptide H3K36me3. Les cristaux de la protéine avec les analogues du bis-tris sont en cours d'analyse.

Sur base de la structure obtenue (première structure du domaine PWWP de la DNMT3B en complexe avec H3K36me3) et de celles d'autres protéines capables de reconnaitre la lysine triméthylée, une étude de la structure du peptide a été réalisée montrant les interactions de celui-ci avec la protéine. Il a pu en être déduit que l'approche du peptide était fortement conditionnée par la présence d'hélices α au sein de la protéine et que sa structure de ce dernier était fort similaire d'un cas à l'autre. De plus, un déplacement de la Phe 266 de la cage aromatique a été montré dans le cas de la reconnaissance de la lysine triméthylée. Il en ressort que les principales interactions entrant en jeu dans la reconnaissance du peptide sont les interactions π-cation.

La résolution de la structure du domaine PWWP en complexe avec le peptide et l'obtention de cristaux dans le cadre des tests cristallographiques entre la protéine et les analogues du bis-tris pourraient conduire à une meilleure compréhension du mode de reconnaissance des sites spécifiques qui sont méthylés au niveau de l'ADN.

En perspective, l'analyse des cristaux obtenus, l'identification de nouveaux ligands et la détermination de la constante d'affinité entre le peptide et le domaine PWWP de la DNMT3B pourraient conduire au développement de médicaments anti-cancéreux jouant sur le processus de déméthylation de l'ADN qui conduisent à la réexpression des gènes suppresseurs de tumeurs.

6 Matériel et méthodes

6.1 Plasmide codant pour le domaine PWWP de la DNMT3A et DNMT3B

Les plasmides codant pour le domaine PWWP de la DNMT3A (résidus 247-427, plasmide n° 25328) et de la DNMT3B (résidus 206-355, plasmide n° 32044) humaine ont été obtenus auprès de la firme Addgene. Ces plasmides ont été intégrés au sein de la bactérie *Escherichia coli* Rosetta 2.

6.2 Surexpression et purification du domaine PWWP de la DNMT3A et DNMT3B

Les protocoles d'expression et de purification ont été adaptés de la publication de *Wu et al*. (Wu, Zeng et al. 2011)

I. Surexpression

La partie de surexpression de la protéine est commune aux deux domaines PWWP.

D'abord, une préculture est réalisée dans 5 mL de milieu de culture Terrific Broth (TB). Ce milieu est composé de 1,2% de tryptone, 2,4% d'extrait de levure, 0,4% de glycérol, 0,72 M de K_2HPO_4, 0,17 M de KH_2PO_4, 50 µg/mL de Kanamycine et 34 µg/mL de Chloramphénicol. 50 µL de bactéries contenant le plasmide d'intérêt y sont ajoutés et le tout est incubé durant une nuit à 37°C sous agitation.

La préculture sert de base pour l'ensemencement d'une culture cellulaire à plus gros volume. Les cultures sont réalisées dans 500 mL de milieu TB contenant deux antibiotiques : la Kanamycine (Cf = 50 µg/mL) et le Chloramphénicol (Cf = 34 µg/mL). Les milieux sont incubés à 37°C et des mesures de contrôle de turbidité sont effectuées jusqu'à ce que la densité optique (DO) à 600 nm atteigne 1,5. Une fois la DO voulue atteinte, la protéine est induite à 18°C durant une nuit par addition d'isopropyl-1-thio-D-galactopyranoside (IPTG, Cf = 1 mM) sous agitation. Le culot bactérien est ensuite récupéré le lendemain par centrifugation au moyen d'une « Beckman Coulter Avanti J-E centrifuge » à 7000 RPM, 4°C pendant 30 minutes.

II. Purification

Domaine PWWP de la DNMT3A : la première étape de purification consiste en une lyse cellulaire permettant de récupérer la protéine d'intérêt. Pour ce faire, les cellules récoltées ont été resuspendues dans un tampon de lyse : 1X PBS (137 mM NaCl, 2,7 mM KCl, 10 mM Na_2HPO_4, 1,76 mM KH_2PO_4), pH 7,4, 250 mM NaCl, 2 mM β-mercaptoéthanol, 5% glycérol et 0,1% CHAPS. Des anti-protéases « protease inhibitor cocktail complete Mini EDT-free » (1 tablette pour 10 mL de solution) doivent également être ajoutées pour éviter la dégradation de la protéine lors de la lyse cellulaire par des protéases. Les cellules sont lysées sur glace au moyen d'un sonicateur « Branson sonifier 150 » dans une gamme allant de 10 à 15 MHz (10 cycles de lyse de 30 secondes en alternance avec 30 secondes de pause). L'extrait brut obtenu est ensuite centrifugé à 9000 RPM, 4°C pendant 30 minutes pour séparer le culot du surnageant. Ce dernier (contenant la protéine d'intérêt) est filtré sur 0,45 et 0,22 µm.

Une première purification est réalisée afin d'éliminer la plus grande partie des impuretés. Cette étape consiste en une colonne IMAC (immobilized metal ion affinity chromatography) chargée au Ni^{2+} (colonne : HisTrap FF crude 5mL – GE Healthcare). La colonne est équilibrée et lavée (10 volumes de colonne) avec un « tampon de lavage » composé de 20 mM Tris-HCl, pH 8,0, 250 mM NaCl, 5% glycérol et 50 mM imidazole. La protéine est ensuite éluée avec un gradient linéaire (pendant 30 minutes) de 20 mM Tris-HCl, pH 8,0, 250 mM NaCl, 5% glycérol et 1 M imidazole. Les fractions contenant la protéine d'intérêt sont rassemblées.

Dans un second temps, la protéine purifiée est dialysée pendant 2 heures afin de diminuer la concentration en imidazole dans un tampon 20 mM Tris-HCl, pH 8,0 et 150 mM NaCl. Durant cette étape, une TEV protéase est également ajoutée afin de cliver la queue poly-histidine de la protéine (50 µg/mg de protéine).

Une seconde colonne IMAC est réalisée afin de séparer la protéine clivée de l'His-tag et de la TEV-protéase précédemment ajoutée. En effet, cette TEV-protéase possède elle aussi une queue poly-histidine mais n'est pas capable de cliver son propre His-tag. La protéine est chargée sur une colonne HisTrap FF crude 5mL chargée au Ni^{2+} et est lavée avec 10 volumes

de colonnes de tampon de lavage (20 mM Tris-HCl, pH 8,0, 250 mM NaCl, 5% glycérol et 50 mM imidazole). Les impuretés sont éluées (sans gradient) au moyen d'un tampon composé de 20 mM Tris-HCl, pH 8,0, 250 mM NaCl, 5% glycérol et 1 M imidazole. Les fractions contenant la protéine sont rassemblées.

Ensuite, la protéine est dialysée dans un tampon 20 mM Tris-HCl, pH 8,0. Cette opération est effectuée 3 fois pendant 30 minutes.

Finalement, la protéine est purifiée sur une colonne échangeuse d'anions « Source 30Q » équilibrée avec un tampon 20 mM Tris-HCl, pH 8,0 et éluée avec un gradient linéaire de NaCl jusqu'à une concentration de 1 M. Pour terminer, la protéine est concentrée au moyen « d'amicon » à 5 mg/mL pour les tests de cristallisation.

Domaine PWWP de la DNMT3B : la première étape de purification consiste en une lyse cellulaire permettant de récupérer la protéine d'intérêt. Pour ce faire, les cellules récoltées ont été resuspendues dans un tampon de lyse : 50 mM HEPES, pH 7,4, 500 mM NaCl, 2 mM β-mercaptoéthanol, 5% glycérol et 0,1% CHAPS. Des anti-protéases « protease inhibitor cocktail complete Mini EDT-free » (1 tablette pour 10 mL de solution) doivent également être ajoutées pour éviter le clivage de la protéine lors de la lyse cellulaire. Les cellules sont lysées sur glace au moyen d'un sonicateur « Branson sonifier 150 » dans une gamme allant de 10 à 15 MHz (10 cycles de lyse de 30 secondes en alternance avec 30 secondes de pause). L'extrait brut obtenu est ensuite centrifugé à 9000 RPM, 4°C pendant 30 minutes pour séparer le culot du surnageant. Ce dernier (contenant la protéine d'intérêt) est filtré sur 0,45 et 0,22 µm.

Une première purification est réalisée afin d'éliminer la plus grande partie des impuretés. Cette étape consiste en une colonne IMAC (immobilized metal ion affinity chromatography) chargée au Ni^{2+} (colonne : HisTrap FF crude 5mL – GE Healthcare). La colonne est équilibrée et lavée (10 volumes de colonne) avec un « tampon de lavage » composé de 20 mM HEPES, pH 7,4, 500 mM NaCl, 5% glycérol et 50 mM imidazole. La protéine est ensuite éluée avec un gradient linéaire (pendant 30 minutes) de 20 mM HEPES, pH 7,4, 250 mM NaCl, 5% glycérol et 1 M imidazole. Les fractions contenant la protéine d'intérêt sont rassemblées.

Dans un second temps, la protéine purifiée est dialysée pendant 2 heures afin de diminuer la concentration en imidazole dans un tampon 20 mM HEPES, pH 7,4 et 250 mM NaCl. Durant cette étape, une TEV protéase est également ajoutée afin de cliver la queue poly-histidine de la protéine (50 µg/mg de protéine).

Une seconde colonne IMAC est réalisée afin de séparer la protéine clivée de l'His-tag et de la TEV-protéase précédemment ajoutée. En effet cette TEV-protéase possède elle aussi une queue poly-histidine mais n'est pas capable de cliver son propre His-tag. La protéine est chargée sur une colonne HisTrap FF crude 5mL chargée au Ni^{2+} et est lavée avec 10 volumes de colonnes de tampon de lavage (20 mM HEPES, pH 7,4, 500 mM NaCl, 5% glycérol et 50 mM imidazole). Les impuretés sont éluées (sans gradient) au moyen d'un tampon composé de 20 mM HEPES, pH 7,4, 500 mM NaCl, 5% glycérol et 1 M imidazole. Les fractions contenant la protéine sont rassemblées.

Ensuite, la protéine est dialysée dans un tampon 20 mM PIPES, pH 6,5. Cette opération est effectuée 3 fois pendant 30 minutes.

Finalement, la protéine est purifiée sur une colonne échangeuse de cations « Source 30S » équilibrée avec un tampon 20 mM PIPES, pH 6,5 et éluée avec un gradient linéaire de NaCl jusqu'à une concentration de 1 M. Pour terminer, la protéine est concentrée à 25 mg/mL au moyen d'amicon pour les tests de cristallisation.

6.3 SDS-PAGE

Les différentes étapes de productions et de purifications ont été suivies par électrophorèse sur gel (SDS-PAGE) afin de vérifier la présence et la pureté de la protéine. Le gel est constitué de deux parties : un resolving gel et un stacking gel. Lors de la préparation du resolving gel, le TEMED (catalyseur) doit être ajouté au dernier moment. Lorsque le gel est coulé dans la plaque, il doit être recouvert d'éthanol et laissé au repos durant 45 minutes. Ensuite, l'éthanol est retiré et le stacking gel peut être coulé. La polymérisation prend environ 30 minutes.

Les échantillons, qui sont analysés, sont préparés différemment en fonction de leur nature (culot et surnageant). Avant et après induction, les échantillons sont centrifugés 12 secondes afin d'éliminer le milieu TB de la culture cellulaire. Le culot est resuspendu dans 45 µL d'eau et 15 µL de tampon d'échantillon. Le surnageant après lyse est centrifugé pendant 12 secondes. 30 µL de tampon d'échantillon sont ajoutés à 10 µL de surnageant alors qu'une pointe de tips du culot est mélangé à 30 µL d'eau et 30 µL de tampon d'échantillon. Quant à la protéine pure, seulement 4 µL sont ajoutés à 16 µL de tampon.

Tous les échantillons précédemment préparés sont ensuite chauffés à 100°C durant 5 minutes. Les plaques sont ensuite montées verticalement dans un dispositif approprié. Le fond de la plaque doit plonger dans une solution de tampon d'électrophorèse. Un marqueur de poids moléculaire (Bio-Rad) est placé dans le premier puits et les échantillons dans les puits suivants. La cuve est ensuite branchée à un générateur et un courant de 200 V est maintenu durant la migration de l'échantillon. Lorsque le front de migration a atteint le fond de la plaque, le courant est arrêté. Une étape de révélation est nécessaire à l'observation des différents spots. Deux techniques de coloration ont été utilisées : la coloration au bleu de Coomassie et la coloration au nitrate d'argent. Pour la première technique, le gel est plongé dans une solution de bleu de Coomassie (sous agitation). Après 5 minutes, le gel est retiré et décoloré dans une solution de décoloration. La composition de ces solutions est fournie à la Table 7.

Resolving gel (15%)	Stacking gel
5,1 mL H_2O distillée 9,5 mL Acrylamide mix (30%) 5,0 mL Tris-HCl 1,5 M pH 8,8 200 µL SDS 10% 200 µL APS 10% 8 µL TEMED	2,77 mL H_2O distillée 0,83 mL Acrylamide mix (30%) 1,3 mL Tris-HCl 0,5 M pH 6,8 50 µL SDS 10% 50 µL APS 10% 5 µL TEMED
Tampon d'échantillon SDS-PAGE	**Tampon d'électrophorèse (10X)**
3,55 mL H_2O distillée 1,25 mL Tris-HCl 0,5 M pH 6,8 2,5 mL glycérol 2,0 mL SDS 10% 200 µL bleu de bromophénol 0,5%	Volume : 1L 30,35g Tris 44g glycine 10g SDS
Solution de bleu de Coomassie	**Solution de décoloration**
Volume : 100 mL 0,25g bleu de Coomassie 45 mL méthanol 10 mL acide acétique	Volume : 1L 300 mL méthanol 100 mL acide acétique glacial 600 mL eau

Tableau 7 : Compositions du resolving et du stacking gel et des différentes solutions utilisées pour la coloration et décoloration.

Le protocole pour une coloration au nitrate d'argent (voir Table 8) est beaucoup plus contraignant et nécessite l'utilisation d'eau Milli Q pour éviter toute contamination à base d'ions étrangers.

#	Etape	Solution		Temps
0	Lavage	25 mL	Eau	120'
1	Fixation	25 mL 100 mL 125 mL	Acide acétique Méthanol (1) Eau	15'
2	Rinçage	250 mL	Eau	1'
3	Fixation	25 mL 100 mL 125 mL	Acide acétique Méthanol Eau	15'
4	Rinçage	250 mL	Eau	1'
5	Sensibilisation	75 mL 165 mL 500 mg 17 g	Méthanol Eau Thiosulfate de sodium Acétate de sodium	30'
6	Rinçage	250 mL	Eau	2'
7	Rinçage	250 mL	Eau	2'
8	Rinçage	250 mL	Eau	2'
9	Marquage	250 mL 350 mg	Eau Nitrate d'argent (2)	20'
10	Rinçage	250 mL	Eau	1'
11	Rinçage	250 mL	Eau	1'
12	Développement	250 mL 6,25 g 100 µL	Eau Carbonate de sodium Formaldéhyde (3)	?
13	Arrêt	250 mL 1,65 g	Eau EDTA (4)	10'
14	Rinçage	250 mL	Eau	2'
15	Rinçage	250 mL	Eau	2'
16	Rinçage	250 mL	Eau	2'

Tableau 8 : Protocoles et compositions des solutions pour la coloration aux nitrates d'argent.

(1) Possibilité d'utiliser de l'éthanol à la place du méthanol.

(2) Récupérer dans une tourie.

(3) À ajouter au dernier moment.

(4) Ajouter quelques gouttes de NaOH 5M pour faciliter la solubilisation.

6.4 TEV protéase

I. Plasmide codant pour la TEV protéase

Le plasmide codant pour la TEV-protéase (plasmide n° 8827) a été obtenu auprès de la firme Addgene. Ce plasmide a été intégré au sein de la bactérie *Escherichia coli* BL21 (DE3) Codon Plus-RIL.

II. Surexpression et purification de la TEV protéase

Les protocoles d'expression et de purification sont basés et ont été adaptés de la publication de *Tropea et al*. (Tropea, Cherry et al. 2009)

Inoculer 5 mL de milieu de culture *Lysogeny Broth* (LB) contenant 50 µg/mL d'Ampicilline et 34 µg/mL de Chloramphénicol dans un falcon. Pour la préparation du milieu LB, 20g de milieu LB sont ajoutés à 1 L d'eau. 50 µL de bactéries contenant le plasmide d'intérêt y sont ajoutés et le tout est laissé incuber à 37°C durant une nuit sous agitation.

La préculture sert de base pour l'ensemencement d'une culture cellulaire à plus gros volume. Les précultures sont ajoutées à chacun des différents erlens bafflés contenant 500 mL de milieu LB, 50 µg/mL d'Ampicilline et 34 µg/mL de Chloramphénicol. Les milieux sont incubés à 37°C et des mesures de contrôle de turbidité sont effectuées jusqu'à ce que la densité optique (DO) à 600 nm atteigne 0,5. Une fois la DO voulue atteinte, la protéine est induite à 32°C durant 4 heures par addition d'isopropyl-1-thio-D-galactopyranoside (IPTG, C_f = 1 mM) sous agitation. Le culot bactérien est ensuite récupéré le lendemain par centrifugation au moyen d'une « *Beckman Coulter Avanti J-E centrifuge* » à 4000 RPM, 4°C pendant 30 minutes.

Le plasmide contenant le gène de la TEV-protéase possède également un « insert » codant pour une queue polyhistidine. La His_6-TEV-protéase peut donc être purifiée et homogénéisée en deux étapes : une chromatographie d'affinité (IMAC) chargée au Ni et une chromatographie d'exclusion (Superdex 75).

La première étape consiste en une lyse cellulaire permettant de récupérer la protéine d'intérêt. Les cellules récoltées ont été resuspendues dans un « tampon de lyse » : 100 mM NaCl, 50 mM KH_2PO_4, 10% glycérol, 25 mM imidazole et pH 8. Des anti-protéases *« protease inhibitor cocktail complete Mini EDT-free »* (1 tablette pour 10 mL de solution) doivent également être ajoutées pour éviter la destruction de la protéine lors de la lyse cellulaire. Les cellules sont lysées au moyen d'un sonicateur *« Branson sonifier 150 »* sur glace à 20 000 psi (10 cycles de lyse de 30 secondes en alternance avec 30 secondes de pause). L'extrait brut obtenu est ensuite centrifugé à 9000 RPM, 4°C pendant 30 minutes pour séparer le culot du surnageant. De la polyéthylèneimine est ajoutée à ce dernier pour atteindre une concentration finale de 0,1%. La solution est mélangée lentement par inversion et centrifugée à 10 000 RPM, 4°C pendant 30 minutes. Le surnageant est conservé et le culot est jeté.

Une première purification est réalisée afin d'éliminer la plus grande partie des impuretés. Cette étape consiste en une colonne *IMAC* (immobilized metal ion affinity chromatography) chargée au Ni^{2+} (colonne : HisTrap FF crude 5mL – GE Healthcare). La queue poly-histidine (His-tag) va pouvoir interagir avec le nickel et la protéine sera chélatée au niveau de la colonne. La colonne est équilibrée et lavée (10 volumes de colonne) avec le « tampon de lyse ». La protéine est ensuite éluée avec un gradient linéaire (pendant 30 minutes) de 50 mM KH_2PO_4, 100 mM NaCl, 10% glycérol, 500 mM imidazole et pH 8. 1 mL de « tampon de dilution » (50 mM K_2HPO_4, 200 mM NaCl, 10% glycérol et pH 8,0) doit être ajouté dans chacun des tubes de collecte afin de diluer l'imidazole présent. Les fractions contenant la protéine d'intérêt sont rassemblées et de l'EDTA et DTT y sont ajoutés pour atteindre une concentration finale de 1 mM (des deux composés). L'échantillon doit ensuite être concentré pour atteindre un volume de 5 mL (volume maximal de charge de la colonne S75) par centrifugation : amicon 15 kDa, 3500 RPM, 4°C et cycles de 5 minutes.

Finalement, la protéine est purifiée et homogénéisée sur un tamis moléculaire (exclusion de taille) *« HiLoad 16/60 Superdex 75 – GE Healthcare »* équilibré avec un tampon 25 mM K_2HPO_4, 100 mM NaCl, 10% glycérol et pH 7,5. Pour terminer, la protéine est concentrée à 1 mg/mL pour être stockée à -80°C dans des eppendorfs de 500 µL.

6.5 Essais de cristallisation

Les conditions de cristallisation pour les tests sont basées sur la littérature et le criblage des conditions de Hampton Research.

Le domaine PWWP purifié de la DNMT3A/3B a été cristallisé par la méthode de la diffusion de vapeur en goutte assise (g.a.) ou en goutte pendante (g.p.) à 20°C. Pour cela, 1 µL de solution de protéine (5 mg/mL pour la DNMT3A et 22 mg/mL pour la DNMT3B) a été mélangé à 1 µL de la solution du réservoir. Le volume final des puits est de 700 µL. Les conditions des différentes boites sont reprises ci-dessous.

Boite 1 (g.a.) : domaine PWWP DNMTA et peptide H3K36me$_3$ (ratio 1 : 5)

- 0,1 M Tris-HCl (pH 7,0 à 8,5 par incrément de 0,5)
- 0,01 M NaCl
- 1,5 à 4 M formiate de sodium (incrément de 0,5)
- H_2O jusque 700 µL

Boite 2 (g.a.) : domaine PWWP DNMTA et peptide H3K36me$_3$ (ratio 1 : 5)

- 0,1 M Tris-HCl (pH 7,0 à 8,5 par incrément de 0,5)
- 1,5 à 4 M formiate de sodium (incrément de 0,5)
- H_2O jusque 700 µL

Boite 3 (g.a.) : domaine PWWP DNMT3A (2,5 mg/mL) et peptide H3K36me$_3$ (ratio 1 : 5)

- 0,1 M Tris-HCl (pH 7,0 à 8,5 par incrément de 0,5)
- 1,5 à 4 M formiate de sodium (incrément de 0,5)
- H_2O jusque 700 µL

Boite 4 (g.p.) : domaine PWWP DNMT3A seul

- 0,1 M bis-tris (pH 6,0 à 7,2 par incrément de 0,4)
- 0,1 M sulfate d'ammonium
- 24 à 34% PEG 3350
- H_2O jusque 700 µL

Boite 5 (g.a.) : domaine PWWP DNMT3B et 3 ligands (ratio 1 : 5). **A**: 2-(2-diethylaminoethylamino)-ethanol, **B** : 2,2'-(4-aminobutylazanediyl)-diethanol et **C** : Triisopropanolamine.

- Boite 96 puits (1 µL de solution de ligand et 1 µL de la solution réservoir)
- Conditions HR2-100 et HR2-112 (*Hampton Research*) (volume 200 µL)

Boite 6 (g.p.) : domaine PWWP DNMT3B seul

- 0,025 M bis-tris (pH 6,0 à 6,3 par incrément de 0,1)
- 0,2 M Li_2SO_4
- 28 à 33% PEG 3350
- H_2O jusque 700 µL

Boite 7 (g.a.) : domaine PWWP DNMT3B et peptide H3K36me_3 (ratio 1 : 5)

- Boite 96 puits (1 µL de solution de ligand et 1 µL de la solution réservoir)
- Conditions HR2-100 et HR2-112 (*Hampton Research*) (volume 200 µL)
    - Des cristaux ont été obtenus pour deux conditions HR2-112 14 et HR2-112 28

Boite 8 (g.a.) : domaine PWWP DNMT3B et 3 ligands (ratio 1 : 5). **D :** (2S,2'S)-2-[2-(1-hydroxymethyl-propylamino)-ethylamino]-butan-1-ol (éthambutol), **A**: 2-(2-diethylaminoethylamino)-ethanol, **B** : 2,2'-(4-aminobutylazanediyl)-diethanol

- Boite 96 puits (1 µL de solution de ligand et 1 µL de la solution réservoir)
- Conditions HR2-100 et HR2-112 (*Hampton Research*) (volume 200 µL)

La structure des analogues est présentée à la figure ci-dessous (voir Fig 38) et la séquence du peptide est la suivante : SAPATGGV{Lys(me)$_3$ }KPHRYR (poids moléculaire : 1669,8 Da).

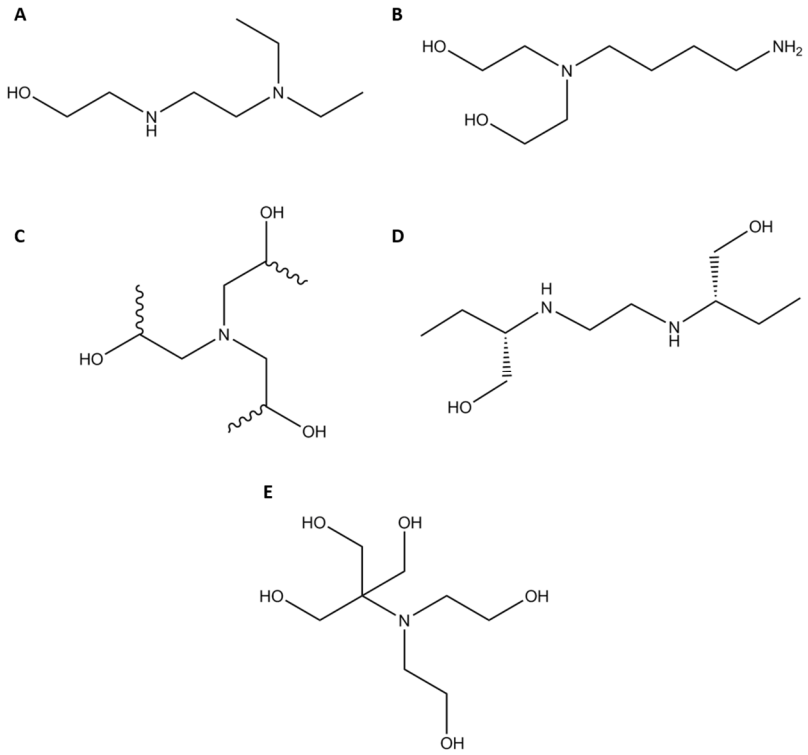


Figure 38 : Ligands utilisés pour les tests de cristallisation. A: 2-(2-diethylaminoethylamino)-éthanol, B : 2,2'-(4-aminobutylazanediyl)-diéthanol, C : Triisopropanolamine, D : (2S,2'S)-2-[2-(1-hydroxymethyl-propylamino)-ethylamino]-butan-1-ol (éthambutol) et E : Bis-tris.

6.6 Détermination de la structure

Les données ont été collectées sur la ligne *Proxima 2* du synchrotron Soleil à Paris, France, à une longueur d'onde de 0,980 Å. Les données ont été indexées et intégrées au moyen du programme *XDS* qui a généré un fichier .mtz (contenant toutes les réflexions). Le module de *PHENIX* (*Xtriage*), permet de déterminer la qualité des données collectées, les axes et centres de symétrie et de proposer le système cristallin le plus probable parmi les sept existants (ainsi que le groupe d'espace). L'orientation et la position des complexes DNMT3B-peptide ont été premièrement déterminées par remplacement moléculaire en utilisant l'unité monomérique de la DNMT3B (présente dans la PDB sous la référence 3QKJ) comme modèle de recherche dans le programme *phaser-MR*. Le modèle résultant obtenu a été raffiné au moyen des packages de *PHENIX* (PHENIX 2010). Le peptide a ensuite été ajouté manuellement dans la densité électronique correspondante au moyen de *Coot* et les contraintes liées à la structure ont été générées dans un fichier .cif par le programme *JLigand* (Lebedev, Young et al. 2012). La structure générale a subi plusieurs étapes de raffinement grâce aux modules *refinement* et *auto-build* de *PHENIX*. La qualité du modèle a pu être vérifiée par le module *phenix.validate*. Ainsi, la carte de Ramachandran de cette structure montre que 96 % de résidus sont dans une région favorable, 2,4 % sont dans une région permise et 1,6 % dans une région interdite. Les programmes de visualisation utilisés sont *Pymol*, *DS Visualizer* ou encore *Coot*.

7 Références

Accelrys (2013). Discovery Studio.

Adams-Cioaba, M. A. and J. Min (2009). "Structure and function of histone methylation binding proteins1." Biochem. Cell Biol **87**: 93-105.

American Society of Clinical Oncology (2003). "American Society of Clinical Oncology policy statement update: genetic testing for cancer susceptibility." Journal of Clinical Oncology **21**(12): 2397.

Barski, A., S. Cuddapah, K. Cui, T.-Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev and K. Zhao (2007). "High-resolution profiling of histone methylations in the human genome." Cell **129**(4): 823-837.

Baylin, S. B. and J. G. Herman (2000). "DNA hypermethylation in tumorigenesis: epigenetics joins genetics." Trends in Genetics **16**(4): 168-174.

Bestor, T. H. (2000). "The DNA methyltransferases of mammals." Human Molecular Genetics **9**(16): 2395-2402.

Brueckner, B. and F. Lyko (2004). "DNA methyltransferase inhibitors: old and new drugs for an epigenetic cancer therapy." Trends in Pharmacological Sciences **25**(11): 551-554.

Chedin, F. (2010). "The DNMT3 family of mammalian de novo DNA methyltransferases." Progress in Molecular Biology and Translational Science **101**: 255-285.

Chen, T., N. Tsujimoto and E. Li (2004). "The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin." Molecular and Cellular Biology **24**(20): 9048-9058.

Cheng, X. and R. M. Blumenthal (2008). "Mammalian DNA methyltransferases: a structural perspective." Structure **16**(3): 341-350.

Chong, S. and E. Whitelaw (2004). "Epigenetic germline inheritance." Current Opinion in Genetics & Development **14**(6): 692-696.

Dhayalan, A., A. Rajavelu, P. Rathert, R. Tamas, R. Z. Jurkowska, S. Ragozin and A. Jeltsch (2010). "The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation." Journal of Biological Chemistry **285**(34): 26114-26120.

Esteller, M. (2005). "Aberrant DNA methylation as a cancer-inducing mechanism." Annual Review of Pharmacology and Toxicology **45**: 629-656.

Esteller, M. (2007). "Cancer epigenomics: DNA methylomes and histone-modification maps." Nature Reviews Genetics **8**(4): 286-298.

Esteller, M. and G. Almouzni (2005). "How epigenetics integrates nuclear functions." EMBO reports **6**(7): 624-628.

Filippakopoulos, P. and S. Knapp (2014). "Targeting bromodomains: epigenetic readers of lysine acetylation." Nature Reviews Drug Discovery **13**(5): 337-356.

Gaudet, F., J. G. Hodgson, A. Eden, L. Jackson-Grusby, J. Dausman, J. W. Gray, H. Leonhardt and R. Jaenisch (2003). "Induction of tumors in mice by genomic hypomethylation." Science **300**(5618): 489-492.

Ge, Y.-Z., M.-T. Pu, H. Gowher, H.-P. Wu, J.-P. Ding, A. Jeltsch and G.-L. Xu (2004). "Chromatin targeting of de novo DNA methyltransferases by the PWWP domain." Journal of Biological Chemistry **279**(24): 25447-25454.

Goll, M. G. and T. H. Bestor (2005). "Eukaryotic cytosine methyltransferases." Annu. Rev. Biochem. **74**: 481-514.

Goll, M. G., F. Kirpekar, K. A. Maggert, J. A. Yoder, C.-L. Hsieh, X. Zhang, K. G. Golic, S. E. Jacobsen and T. H. Bestor (2006). "Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2." Science **311**(5759): 395-398.

Guo, X., L. Wang, J. Li, Z. Ding, J. Xiao, X. Yin, S. He, P. Shi, L. Dong and G. Li (2014). "Structural insight into autoinhibition and histone H3-induced activation of DNMT3A." Nature **0**: 1-20.

Hanahan, D. and R. A. Weinberg (2000). "The Hallmarks of Cancer." Cell **100**: 57-70.

Hanahan, D. and R. A. Weinberg (2011). "Hallmarks of cancer: the next generation." Cell **144**(5): 646-674.

Huen, M. S., J. Huang, J. W. Leung, S. M.-H. Sy, K. M. Leung, Y.-P. Ching, S. W. Tsao and J. Chen (2010). "Regulation of chromatin architecture by the PWWP domain-containing DNA damage-responsive factor EXPAND1/MUM1." Molecular cell **37**(6): 854-864.

Hughes, R. M., K. R. Wiggins, S. Khorasanizadeh and M. L. Waters (2007). "Recognition of trimethyllysine by a chromodomain is not driven by the hydrophobic effect." Proceedings of the National Academy of Sciences **104**(27): 11184-11188.

Insitut de Veille Sanitaire (2013). Evolution de l'incidence et de la mortalité par cancer en France entre 1980 et 2012.

Jeltsch, A. (2002). "Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases." Chembiochem **3**(4): 274-293.

Jin, B., Y. Li and K. D. Robertson (2011). "DNA Methylation Superior or Subordinate in the Epigenetic Hierarchy?" Genes & cancer **2**(6): 607-617.

Jones, P. A. and S. B. Baylin (2002). "The fundamental role of epigenetic events in cancer." Nature **3**(6): 415-428.

Jones, P. A. and G. Liang (2012). The Human Epigenome. Epigenetic Epidemiology. Springer, Springer**:** 5-20.

Kim, J., J. Daniel, A. Espejo, A. Lake, M. Krishna, L. Xia, Y. Zhang and M. T. Bedford (2006). "Tudor, MBT and chromo domains gauge the degree of lysine methylation." Nature **7**(4): 397-403.

Klimasauskas, S., S. Kumar, R. J. Roberts and X. Cheng (1994). "Hhal methyltransferase flips its target base out of the DNA helix." Cell **76**(2): 357-369.

Klose, R. J. and A. P. Bird (2006). "Genomic DNA methylation: the mark and its mediators." Trends in biochemical sciences **31**(2): 89-97.

Kulis, M. and M. Esteller (2010). "DNA methylation and cancer." Journal of Clinical Oncology **70**: 27-56.

Lebedev, A. A., P. Young, M. N. Isupov, O. V. Moroz, A. A. Vagin and G. N. Murshudov (2012). "JLigand: a graphical tool for the CCP4 template-restraint library." Acta Crystallographica Section D **68**(4): 431-440.

Li, F., G. Mao, D. Tong, J. Huang, L. Gu, W. Yang and G.-M. Li (2013). "The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSα." Cell **153**(3): 590-600.

Liao, H.-F., K.-Y. Tai, W. S.-C. Chen, L. C. W. Cheng, H.-N. Ho and S.-P. Lin (2012). "Functions of DNA methyltransferase 3-like in germ cells and beyond." Biology of the Cell **104**(10): 571-587.

Liêu, T. (2014). "Basic Principles of Genetics." Retrieved 1st June 2014, from https://voer.edu.vn/m/basic-principles-of-genetics.

Maurer-Stroh, S., N. J. Dickens, L. Hughes-Davies, T. Kouzarides, F. Eisenhaber and C. P. Ponting (2003). "The Tudor domain 'Royal Family': tudor, plant agenet, chromo, PWWP and MBT domains." Trends in Biochemical Sciences **28**(2): 69-74.

Müller, M. M. and T. W. Muir (2014). "Histones : At the Crossroads of Peptide and Protein Chemistry." Chemical reviews.

Mund, C., B. Brueckner and F. Lyko (2006). "Reactivation of epigenetically silenced genes by DNA methyltransferase inhibitors: basic concepts and clinical applications." Epigenetics **1**(1): 8-14.

Nelson, D. and M. Cox (2010). Lehninger Principles of Biochemistry.

Novagen (2003). pET system manual.

Novagen (2010). Induction & Purification GST-pET System Proteins.

Otani, J., T. Nankumo, K. Arita, S. Inamoto, M. Ariyoshi and M. Shirakawa (2009). "Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L domain." EMBO reports **10**(11): 1235-1241.

Peedicayil, J. (2006). "Epigenetic therapy-a new development in pharmacology." Indian Journal of Medical Research **123**(1): 17.

PHENIX (2010). Python-based Hierarchical ENvironment for Integrated Xtallography.

Qin, S. and J. Min (2014). "Structure and function of the nucleosome-binding PWWP domain." Trends in Biochemical Sciences **39**(11): 536-547.

Qiu, C., K. Sawada, X. Zhang and X. Cheng (2002). "The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds." Nature **9**(3): 217-224.

Reik, W. (2007). "Stability and flexibility of epigenetic gene regulation in mammalian development." Nature **447**(7143): 425-432.

Robertson, K. D., E. Uzvolgyi, G. Liang, C. Talmadge, J. Sumegi, F. A. Gonzales and P. A. Jones (1999). "The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors." Nucleic acids research **27**(11): 2291-2298.

Shahbazian, M. D. and M. Grunstein (2007). "Functions of site-specific histone acetylation and deacetylation." Annual Review of Biochemistry **76**: 75-100.

Stec, I., S. B. Nagl, G.-J. B. van Ommen and J. T. den Dunnen (2000). "The PWWP domain : a potential protein-protein interaction domain in nuclear proteins influencing differentiation?" FEBS Letters **473**(1): 1-5.

Strahl, B. D. and C. D. Allis (2000). "The language of covalent histone modifications." Nature **403**(6765): 41-45.

Swiss Institute of Bioinformatics (2014). ProtParam.

Themans, Q. (2014). Conception de composés anti-cancéreux ciblant le domaine PWWP des DNMT3A et DNMT3B. UNAMUR. Namur.

Tollefsbol, T. (2009). Cancer epigenetics, Taylor & Francis Group.

Tropea, J. E., S. Cherry and D. S. Waugh (2009). Expression and purification of soluble His6-tagged TEV protease. High Throughput Protein Expression and Purification. Springer**:** 297-307.

van Nuland, R., F. van Schaik, M. Simonis, S. van Heesch, E. Cuppen, R. Boelens, H. M. Timmers and H. van Ingen (2013). "Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain." Epigenetics Chromatin **6**(1): 1-12.

Vezzoli, A., N. Bonadies, M. D. Allen, S. M. V. Freund, C. M. Santiveri, B. T. Kvinlaug, B. J. P. Huntly, B. Göttgens and M. Bycroft (2010). "Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1." Nature Structural & Molecular Biology **17**(5): 617-619.

Waddington, C. (1942). "The epigenotype." International Journal of Epidemiology **41**: 10-13.

Wagner, E. J. and P. B. Carpenter (2012). "Understanding the language of Lys36 methylation at histone H3." Nature reviews Molecular cell biology **13**(2): 115-126.

Wu, H., H. Zeng, R. Lam, W. Tempel, M. F. Amaya, C. Xu, L. Dombrovski, W. Qiu, Y. Wang and J. Min (2011). "Structural and histone binding ability characterizations of human PWWP domains." Plos one **6**(6): e18919.

Wu, H. and Y. Zhang (2014). "Reversing DNA Methylation: Mechanisms, Genomics and Biological Functions." Cell **156**(1): 45-68.

Yuan-Yeh, K., L. Li-Yu and T. Hwei-Fang. (2013). "DNMT3A (DNA (cytosine-5-)-methyltransferase 3 alpha)." Retrieved 16th November 2014, from http://atlasgeneticsoncology.org.

Zhang, Y., R. Jurkowska, S. Soeroes, A. Rajavelu, A. Dhayalan, I. Bock, P. Rathert, O. Brandt, R. Reinhardt and W. Fischle (2010). "Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail." Nucleic Acids Research **38**(13): 4246-4253.