



**HAL**  
open science

# Compréhension structurale des complexes de PRMT et de leur régulation du processus de méthylation des arginines

Enes Dursun

► **To cite this version:**

Enes Dursun. Compréhension structurale des complexes de PRMT et de leur régulation du processus de méthylation des arginines. Chimie thérapeutique. Université de Strasbourg, 2022. Français. NNT : 2022STRAF046 . tel-04343690

**HAL Id: tel-04343690**

**<https://theses.hal.science/tel-04343690v1>**

Submitted on 14 Dec 2023

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

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

**ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES**

**Laboratoire de Conception et Application de Molécules Bioactives**

**THÈSE** présentée par :

**Enes DURSUN**

soutenue le : **13 décembre 2022**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Chimie Biologique et Thérapeutique

**COMPREHENSION STRUCTURALE DES COMPLEXES  
DE PRMT ET DE LEUR REGULATION DU PROCESSUS  
DE METHYLATION DES ARGININES**

**THÈSE dirigée par :**

**M CHAUBET Guilhem** Chargé de recherche, UMR 7199, Université de Strasbourg

**THÈSE dirigée par :**

**M WAGNER Alain** Directeur de recherche, UMR 7199, Université de Strasbourg

**RAPPORTEURS :**

**Mme LOPEZ Marie** Chargé de recherche, UMR 5247, Université de Montpellier

**M AGOURIDAS Vangelis** Maître de conférences, UMR 9017, Centrale Lille – ENSCL

**EXAMINATEURS :**

**M SMIETANA Michael** Professeur, UMR 5247, Université de Montpellier

**M CURA VINCENT** Maître de conférences, UMR 7104, Université de Strasbourg



# REMERCIEMENTS

Ces travaux de thèse ont été réalisés au laboratoire de Conception et Applications de Molécules Bioactives à la faculté de pharmacie d'Illkirch Graffenstaden, au sein de l'équipe Chimie BioFonctionnelle (BFC) sous la direction du Dr Guilhem Chaubet et co-direction du Dr Alain Wagner.

Tout d'abord, je tiens à remercier Guilhem pour m'avoir encadré au cours de ces trois années. Je te remercie pour tes idées, ton soutien et pour avoir partagé tes multiples connaissances en chimie. Je te remercie également pour toutes les corrections que tu as apportées à ce manuscrit. Je remercie également Alain qui m'a permis de faire cette thèse dans son équipe. Merci pour tes conseils et ta disponibilité.

Je souhaite également remercier les différents membres du Jury, le Dr Vangelis Agouridas, Dr Marie Lopez, Pr Michael Smietana et Dr Vincent Cura pour avoir accepté de juger ces travaux. Je tiens à témoigner toute ma gratitude au Dr Vincent Cura pour son implication dans mon projet, sa disponibilité et ses conseils et à toute l'équipe du Pr Jean Cavarelli pour m'avoir donné l'opportunité de travailler sur ce projet et l'Agence National de Recherche (ANR) pour le financement.

Je tiens également remercier Marie de l'équipe du Dr Sarah Cianférani pour avoir analysé mes échantillons et Pascal de l'IGBMC pour sa gentillesse et pour la synthèse des peptides. Un grand merci au service d'analyse de la Faculté de Pharmacie, en particulier à Delphine, pour sa grande gentillesse et pour l'analyse RMN de mes produits.

Je tiens ensuite à remercier les membres de l'équipe BFC pour leurs conseils et leurs aides précieuses au cours de cette thèse, en particulier à Safae (aussi pour tes cakes et ta bonne humeur), Indréalie, Valentine (aussi de toujours perdre aux jeux de cartes), Robin, Lorenzo, Ilias (aussi d'avoir montré que la Moussaka Grecque pouvait être aussi bon que la version Turque), Ketty, Layal, Michel, Jessica, Héroïse, Marc, Isabelle et Jean-Serge. Un merci tout particulier à Julien pour son amitié et son humour sublime sans quoi les journées de manip auraient manqué un peu de piquant. C'était un plaisir de partager le même labo avec toi, même si je n'aimais pas trop tes musiques.

Un grand merci aux trois alumni de notre équipe : Charlotte, Tony, et Victor. J'ai vraiment apprécié les deux années que nous avons passé ensemble. Merci Charlotte de m'avoir appris à jouer au tarot et pour tes conseils. Merci Tony d'avoir été toujours là quand il s'agissait de sciences. Merci Victor de m'avoir accompagné le week-end au labo, pour ton amitié et nos échanges au coin de la paillasse ou autour d'un café.

A Mickael pour son amitié, son soutien mais aussi pour m'avoir prêté pleins de catalyseurs et ligands pour mes tests de Suzuki.

A Madame Tirmarche, notre secrétaire qui gère nos commandes et règle tous nos problèmes du quotidien.

Je ne peux pas finir ces remerciements sans mentionner Vezir et Muhafiz qui ont augmenté mon temps de travail et démontré que dormir après 6 h est un acte insignifiant et littéralement interdit lorsque le bol de croquette est vide. Merci.

Le dernier remerciement va à ma famille, en particulier à mon père, ma mère et mon frère pour leur amour et leur soutien qui a amélioré mon quotidien. Un grand merci également à mon cousin Talha qui a toujours su être là pour m'encourager.

Pour finir, un grand merci à mon épouse pour son amour et son soutien.

*« Dès que tu avances sur le chemin, le chemin apparaît »*

*Rûmi*

*A « ruh-û revanim »*

## ABBREVIATIONS

<b>aaRS</b>	Aminoacyl-ARNt synthétase
<b>ARNt</b>	ARN de transfert
<b>Acétyl-coA</b>	Acétyl-coenzyme A
<b>AD1</b>	Activation domain 1
<b>ADN :</b>	Acide DésoxyriboNucléique
<b>ADP</b>	Adénosine diphosphate
<b>ARNm</b>	ARN messenger
<b>ATM</b>	Ataxia telangiectasia mutated
<b>ATR</b>	AT RAD3-related
<b>BER</b>	Base excision repair
<b>Boc</b>	tert-butyloxycarbonyl
<b>BTP</b>	Bis-Tris Propane
<b>CARM1</b>	Coactivator-associated arginine méthyltransférase 1
<b>CBD</b>	Cassure double brin
<b>CpG :</b>	<i>Cytosine</i> -phosphate-guanine
<b>CREB</b>	cAMP response element-binding protein.
<b>CRTC2</b>	CREB Regulated Transcription Coactivator 2
<b>Cryo-EM</b>	Cryomicroscopie électronique
<b>CuAAC</b>	Copper-catalyzed azide-alkyne cycloaddition
<b>DMF</b>	Diméthylformamide
<b>DMSO</b>	Diméthylsulfoxyde
<b>DNMT</b>	DNA methyltransferase
<b>DPPF</b>	1,1'Bis(diphénylphosphino)ferrocène
<b>DPPE</b>	1,2-bis(diphénylphosphino)éthane
<b>DPPM</b>	1,1-Bis(diphénylphosphino)méthane
<b>DPPP</b>	1,3-Bis(diphénylphosphino)propane
<b>Era</b>	Nuclear estrogen receptor
<b>EC<sub>50</sub></b>	Concentration efficace médiane
<b>Et<sub>3</sub>N</b>	Triéthylamine
<b>FAD</b>	Flavine adénine dinucléotide
<b>Fmoc-Cl</b>	Chlorure de fluorénylméthoxycarbonyle
<b>GAR</b>	Glycine-Arginine Rich
<b>HAT</b>	Histones acétyltransférases
<b>HDAC</b>	Histones désacétylases
<b>HKMT</b>	Histones lysines méthyltransférases
<b>IBX</b>	Acide 2-iodoxybenzoïque
<b>IC<sub>50</sub></b>	Concentration inhibitrice médiane
<b>IL-8</b>	Interleukine humaine 8

<b>JABP1</b>	Jak Binding Protein 1
<b>JAK2</b>	Janus Kinase 2
<b>JHDM</b>	Jumonji C domain-containing histone demethylase
<b>JmjC</b>	Jumonji C
<b>KAT</b>	Lysines acétyltransférases
<b>K<sub>d</sub></b>	Constante de dissociation
<b>KDMT</b>	Les lysines déméthylases
<b>LC-MS</b>	Chromatographie en phase liquide-spectrométrie de masse
<b>LSD</b>	Lysine-specific histone demethylases
<b>MeOH</b>	Méthanol
<b>MEP50</b>	Methylosome protein 50
<b>MMA-N<sup>δ</sup></b>	N <sup>δ</sup> -monométhylarginine
<b>MMA-N<sup>ω</sup>,N<sup>ω'</sup></b>	MMA-N <sup>ω</sup> N <sup>ω'</sup> -monométhylarginine
<b>MPAA</b>	4-mercaptophénylacétique
<b>NCL</b>	Ligation Chimique Native
<b>sDMA-N<sup>ω</sup>,N<sup>ω'</sup></b>	Diméthylarginine symétrique
<b>aDMA-N<sup>ω</sup>,N<sup>ω'</sup></b>	Diméthylarginine asymétrique
<b>PABP1</b>	Polyadenylate-binding protein 1
<b>PAD</b>	Peptidylarginine déiminases
<b>pb</b>	Paire de bases
<b>PGM</b>	Proline-glycine-méthionine
<b>pH</b>	Potentiel hydrogène
<b>PRMT</b>	Protéine arginines méthyltransférases
<b>PTM</b>	Modifications post-traductionnelles
<b>RMN</b>	Résonance magnétique nucléaire
<b>RSF1</b>	Repressor Splicing Factor 1
<b>SAM</b>	S-adénosyl-L-méthionine
<b>SAP145</b>	Spliceosome-associated protein 145
<b>SAP49</b>	Spliceosome-associated protein 49
<b>SMN</b>	Survival of Motor Neuron
<b>SPhos</b>	2-Dicyclohexylphosphino-2',6'-dimethoxybiphenyl
<b>Xphos</b>	2-Dicyclohexylphosphino-2',4',6'-triisopropylbiphenyl
<b>SPPS</b>	Synthèse sur support solide
<b>TBTA</b>	( <i>tris</i> [(1-benzyl-1 <i>H</i> -1,2,3-triazol-4-yl)methyl]amine)
<b>TCDI</b>	1,1'-thiocarbonyldiimidazole
<b>TFA</b>	Acide trifluoroacétique
<b>THF</b>	Tétrahydrofurane
<b>THPTA</b>	( <i>tris</i> (3-hydroxypropyltriazolylmethyl)amine)
<b>TIPS</b>	Triisopropylsilane
<b>T<sub>m</sub></b>	Température de fusi

## Table des matières

<b>I. Introduction</b> .....	<b>11</b>
I.1 L'épigénétique .....	11
I.1.1 L'ADN: le support de l'information génétique .....	11
I.1.2 Le génome .....	13
I.1.3 De l'ADN aux chromosomes.....	14
I.1.3.1 Euchromatine et l'hétérochromatine .....	15
I.1.4 La méthylation de l'ADN .....	16
I.1.5 Les modifications post-traductionnelles des histones.....	17
I.1.5.1 Phosphorylation des histones .....	18
I.1.5.2 L'acétylation des lysines.....	18
I.1.5.3 Méthylation des lysines .....	19
I.1.5.4 La méthylation des arginines.....	20
I.2 Les Protéine Arginines Méthyltransférases .....	23
I.2.1 Rôles et Fonctions biologiques des PRMT .....	23
I.2.1.1 PRMT de Type I.....	24
I.2.1.1.1 PRMT1 :.....	24
I.2.1.1.2 PRMT2 :.....	24
I.2.1.1.3 PRMT3 :.....	25
I.2.1.1.4 PRMT4/CARM1 :.....	25
I.2.1.1.5 PRMT6 :.....	25
I.2.1.1.6 PRMT8 :.....	26
I.2.1.2 PRMT de Type II.....	26
I.2.1.2.1 PRMT5 :.....	26
I.2.1.2.2 PRMT9 :.....	26
I.2.1.3 PRMT de Type III.....	27
I.2.1.3.1 PRMT7 :.....	27
<b>II. Objectifs</b> .....	<b>28</b>
<b>III. Synthèses d'arginines modifiées et incorporation régiosélective dans des protéines d'intérêt</b> .....	<b>30</b>
III.1 Introduction .....	30
III.2 Première approche: Incorporation de l'arginine modifiée (R*) sur la queue N-terminale (R7) .....	35
III.2.1 Introduction.....	35
III.2.2 La ligation chimique native.....	35

III.2.3	L'objectif de l'approche .....	36
III.2.4	Résultats et discussions :.....	37
III.2.5	Conclusion.....	47
III.3	Deuxième approche : Incorporation de l'arginine modifiée (R**) sur la partie structurée de RSF1 (R24) .....	49
III.3.1	Introduction.....	49
III.3.2	La chimie click .....	49
III.3.3	La CuAAC.....	50
III.3.4	Incorporation d'acides aminés non naturels au sein de protéines .....	51
III.3.4.1	Incorporation d'acides aminés non naturels par approche site-spécifique : 52	
III.3.5	Objectif de l'approche .....	53
III.3.6	Résultats et discussion .....	54
III.3.7	Conclusion.....	55
<b>IV.</b>	<b>Conception et synthèse d'inhibiteurs de PRMT .....</b>	<b>56</b>
IV.1	Introduction .....	56
IV.2	Inhibiteurs analogues du SAM ciblant la poche du SAM et du substrat.....	57
IV.2.1	Inhibiteurs sélectifs de PRMT1, PRMT4 et PRMT6 .....	57
IV.2.2	Inhibiteurs bi substrats de PRMT1 et PRMT4 .....	58
IV.2.3	La cytosine : un potentiel isostère de la guanidine.....	60
IV.3	Inhibiteurs non-analogues du SAM en phase clinique : .....	63
IV.3.1	Collaboration des laboratoires pharmaceutiques pour la mise au point d'inhibiteurs de PRMT5 .....	63
IV.3.2	Collaboration des laboratoires pharmaceutiques pour la mise au point d'inhibiteurs de PRMT4 .....	63
IV.4	Synthèse d'inhibiteurs de PRMT .....	64
IV.4.1	Introduction.....	64
IV.4.2	Résultats et discussions.....	66
IV.4.3	Conclusion.....	71
<b>V.</b>	<b>Conclusion et Perspectives .....</b>	<b>72</b>
<b>Experimental Part.....</b>	<b>74</b>	
Material and Methods .....	74	
Chemical characterizations .....	75	
Bionconjugation procedures :.....	110	
Inhibition assays for modified arginines .....	110	
<b>Références .....</b>	<b>113</b>	



# I. Introduction

## I.1 L'épigénétique

### I.1.1 L'ADN: le support de l'information génétique

L'ADN est un polymère qui est constitué par l'assemblage de 4 nucléotides (A-T-C-G) qui eux-mêmes sont constitués de trois éléments principaux : un groupement phosphate, un sucre (désoxyribose) et une base azotée (Figure 1). Ces quatre bases azotées sont classées en deux types : les purines Adénine (A) et Guanine (G) et pyrimidines Cytosine (C) et Thymine (T) (Figure 1).

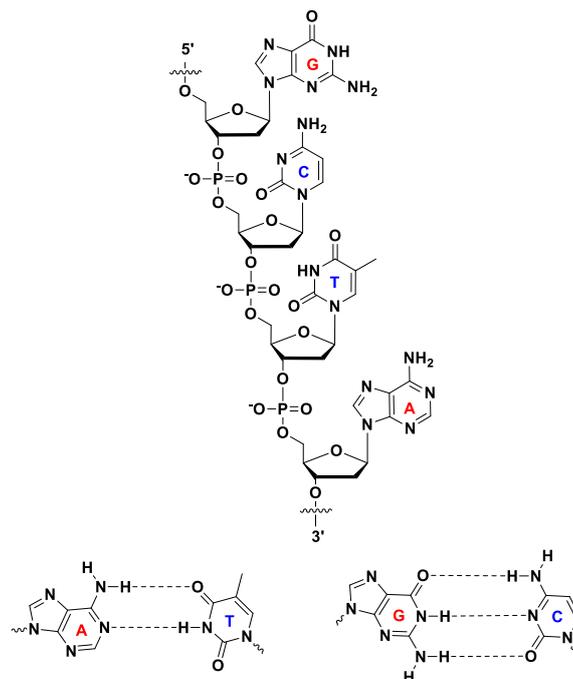


Figure 1: Structure de l'ADN

Notre ADN est constitué de deux brins antiparallèles et complémentaires, assemblés en une double hélice. Cette structure est maintenue par des liaisons hydrogènes entre bases azotées « inter-brins » dictant la complémentarité d'appariement de ces dernières : adénine et thymine – via deux liaisons hydrogènes –, guanine et cytosine – via trois liaisons hydrogènes (Figure 1).<sup>1,2</sup>

On parle dans ce cas d'interactions de type Watson-Crick qui permettent aux bases de s'empiler les uns sur les autres sous la forme d'une hélice hélicoïdale comme l'ont décrit en 1953 Watson et Crick à partir des données expérimentales obtenues par Rosalind Franklin.<sup>1,3</sup>

Il faut savoir qu'il existe plusieurs conformères de l'ADN (ADN A, ADN B, ADN Z), en fonction de la séquence, des modifications chimiques présentes sur les bases azotées qui le composent, de la salinité, et le taux d'hydratation (Figure 2).<sup>2,4</sup>

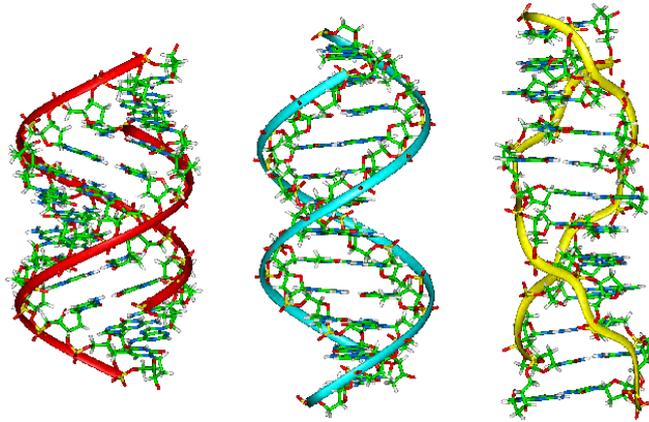


Figure 2: De gauche à droite les conformères de l'ADN : ADN A, ADN B, ADN Z.

La conformation B est la forme qui prédomine dans les conditions physiologiques.<sup>5</sup> Elle est structurée en hélice droite et un tour complet d'hélice fait 3.4 nm. En d'autres termes, la double hélice fait un tour complet autour de son axe en moyenne toutes les 10.5 paires de bases (pb). Ainsi, un tour d'hélice comporte environ 21 nucléotides chacun perpendiculaire à l'axe de rotation. Les bases sont orientées en *anti* par rapport à l'ose (pour des raisons d'encombrement stérique) via la rotation de la liaison N- $\beta$ -osidique et les riboses présentent un plissement C2'-*endo* (Figure 3).<sup>6,7</sup>

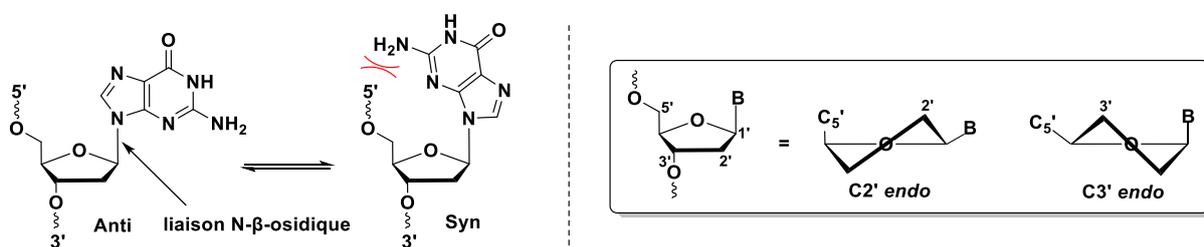


Figure 3: Conformations des nucléotides<sup>7</sup>

La conformation A est observée à des taux d'hydratation faible alors que la topologie Z est aperçue quand la séquence de l'ADN est riche en bases G et C.<sup>8,9</sup> D'autres formes de l'ADN ont été observés telles que les G-quadruplexes, i-motifs et plusieurs d'autres mais elles ne seront pas définies ni discutées dans ce manuscrit.<sup>10,11,12</sup>

La complémentarité observée ont permis d'émettre l'hypothèse que l'ADN est le support de l'information génétique et que la séparation des deux brins pourrait être derrière l'héritabilité.

Cette théorie a été démontrée en 1958 par Matthew Meselson et Franklin Stahl en établissant le modèle de répllication semi-conservative. Ainsi, les scientifiques se sont interrogés sur le code génétique et son déchiffrement pour comprendre comment un brin d'ADN pourrait être traduit en protéines. Dans ce contexte, Nirenberg et Matthaei démontrèrent en 1962 que la plus petite partie des gènes codant l'information génétique est constituée d'un triplet de nucléotides.<sup>13</sup> Depuis lors, des méthodes de séquençage de l'ADN ont été développées comme la méthode de Sanger, qui a permis d'initier à la fin du 20<sup>e</sup> siècle le projet de séquençage du génome humain qui s'acheva en 2004, révélant l'organisation des gènes, la complexité de leur architecture et leur fonctionnement.<sup>14,15</sup>

### **I.1.2 Le génome**

Les quatre nucléotides constitutifs de l'ADN étant les mêmes pour tous les individus (ATCG), la diversité du code génétique est constituée par leur agencement. Cette séquence d'ADN forme ce qu'on appelle un gène qui correspond à un segment (triplet de nucléotides) de cette ADN qui va coder pour la synthèse d'une protéine spécifique. Cette ADN est stockée dans la cellule sous la forme de chromosomes (que nous verrons plus loin) où l'ADN est condensé et enroulé autour de protéines afin de rentrer dans le noyau (où il se situe) des cellules eucaryotes. Le nombre de gènes varie compte tenu des espèces. Pour l'homme, on dénombre entre 20-25000 gènes présents sur les 46 chromosomes, qui représentent un total de 6 milliards de paires de bases.<sup>15,16</sup> Cet ADN est le même dans toutes les cellules somatiques d'un individu mais les cellules n'en font pourtant pas toutes le même usage. Ainsi, la seule séquence et organisation de l'ADN ne permet pas d'expliquer les différences observées dans différentes cellules, tissus et organes. Cette diversité dans l'expression de l'information génétique est contrôlée par l'épigénétique.<sup>17</sup>

L'épigénétique est la discipline de la biologie qui étudie les changements héréditaires dans les fonctions des gènes sans modifier la séquence de l'Acide DésoxyriboNucléique (ADN).<sup>18,19</sup> Elle détermine comment une information supplémentaire va être transmise à la cellule pour lui dire d'activer ou de réprimer les gènes. Le terme d'épigénétique a été introduit au début des années 1940 par Conrad Waddington qui le définit comme « la branche de la biologie qui étudie les relations de cause à effet entre les gènes et leurs produits, lesquelles donnent naissance au phénotype ». <sup>20</sup> En d'autres termes, des mécanismes moléculaires (ou modifications chimiques – qu'on verra plus loin) qui convertissent l'information génétique en traits ou phénotypes observables.<sup>21</sup>

### I.1.3 De l'ADN aux chromosomes

A la fin du XIX<sup>ème</sup> siècle, Walther Flemming a découvert dans les noyaux cellulaires des substances s'organisant en structures filamenteuses lors de la division cellulaire, auxquelles il donne le nom de chromatine.<sup>22</sup> La chromatine fait référence à un mélange d'ADN et de protéines, essentiellement des histones, formant les chromosomes présents dans les cellules eucaryotes. Les histones maintiennent la quantité massive d'ADN d'un génome sous une forme très compacte pouvant tenir dans le noyau cellulaire : les nucléosomes. Ces derniers sont constitués de 145 à 147 pb d'ADN enroulés autour d'unités protéiques composées d'un dimère de quatre histones (H2A, H2B, H3 et H4) assemblées en octamères.<sup>23</sup>

Ces nucléosomes se répètent et forment ensuite une structure plus organisée donnant l'apparence d'un collier de perles.<sup>24</sup> Cette structure originale constituée de l'enchaînement des nucléosomes possède un diamètre de 11 nm et permet d'obtenir ainsi un premier niveau de compactage de l'ADN. À la suite de l'enroulement de l'ADN autour des huit histones et la répétition des nucléosomes, le compactage peut être accru grâce à la présence de l'histone H1, stabilisant les structures d'ordre supérieur de la chromatine et induisant la formation de la fibre chromatinienne de 30 nm (Figure 4).<sup>25</sup> En effet, l'histone H1 se lie d'un côté au nucléosome et de l'autre à l'ADN dit de liaison (séquence d'ADN reliant deux nucléosomes). On retrouve ainsi une molécule de H1 par nucléosome, située à l'extérieur du complexe ADN-histones et positionnée aux points d'entrée et de sortie de l'ADN afin de sceller le nucléosome et le stabiliser (Figure 4). Le complexe nucléoprotéique ainsi construit s'appelle le chromatosome qui s'organise et se compacte sous forme de fibre chromatinienne de 30 nm (Figure 4).<sup>25,26,27</sup>

Des repliements successifs de la fibre de chromatine sur elle-même permet d'avoir un autre ordre de compression comprenant des boucles de chromatines ayant un diamètre 300 nm, qui elles-mêmes se replient pour devenir encore plus dense et atteindre un diamètre de 700 nm et ce jusqu'à la formation du chromosome. Il est clair que devant une telle condensation, la lecture de l'information génétique est soumise à des régulations très pointues et à la nécessité de décompacter localement ces agencements macromoléculaires.

Cette dynamique de la chromatine (condensée ou décondensée) est au cœur des phénomènes épigénétiques et définit l'accessibilité de l'ADN à la machinerie transcriptionnelle. On distingue alors deux types de chromatine en fonction de leur degré de compaction : l'euchromatine lorsque l'ADN est relâché autour des histones, permettant la liaison des facteurs de transcription et son activation ; l'hétérochromatine lorsque l'ADN est compacté autour des histones, réduisant son accès à la machinerie transcriptionnelle et réprimant l'expression génétique.<sup>28,29</sup>

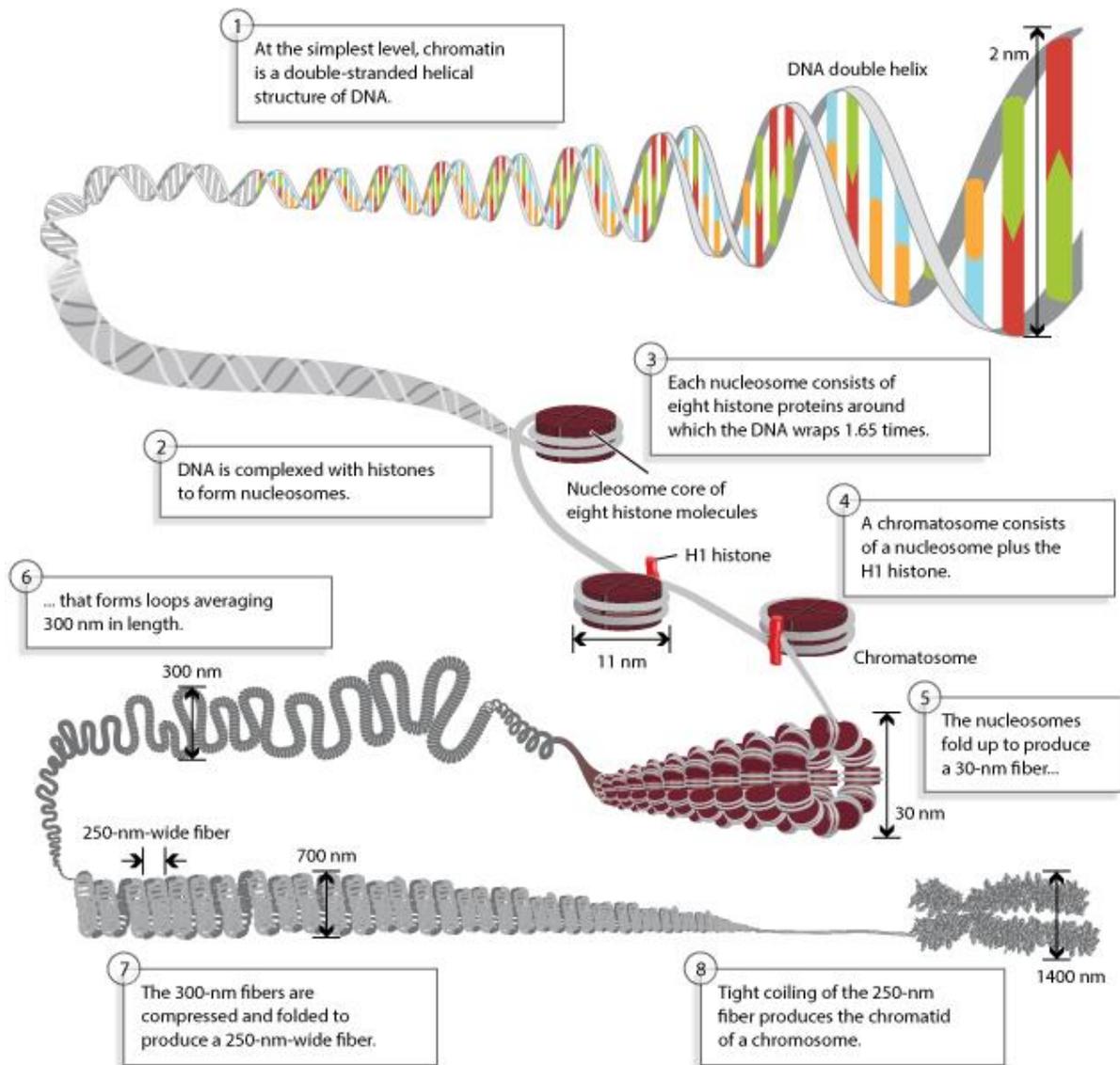


Figure 4: L'enroulement de l'ADN autour des histones

Source de l'image: Annunziato, A. (2008) DNA Packaging: Nucleosomes and Chromatin. Nature Education 1(1):26

### I.1.3.1 Euchromatine et l'hétérochromatine

La formation d'euchromatine et d'hétérochromatine dicte donc l'expression génétique et se distingue par des compositions en protéines ainsi que des modifications chimiques de l'ADN et des histones les constituant qui leur sont exclusives. Ainsi, l'euchromatine, la région de la chromatine où l'activité transcriptionnelle est intense, est reliée à une peracétylation (des lysines) des queues d'histone (les extrémités *N*-terminales des histones qui pointent vers l'extérieur des nucléosomes) H3 et H4, tandis que l'hétérochromatine, lieu d'une activité transcriptionnelle faible, se rattache à un enrichissement en histone H1 et en une méthylation des cytosines (Figure 5).<sup>28,30,31</sup>

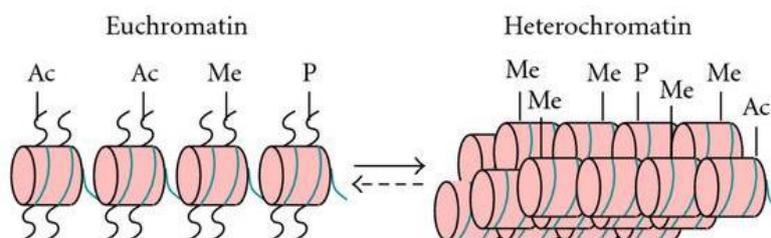


Figure 5: L'euchromatine et l'hétérochromatine<sup>32</sup>

Ces modifications, qu'elles soient acétylations ou méthylations, sont des « interrupteurs épigénétiques » qui définissent l'état de la chromatine et module donc l'accessibilité de l'ADN. Elles sont sous le contrôle d'enzymes dites « writers » (méthyltransférases, acétyltransférases, ...) et « erasers », catalysant le processus inverse. En outre, on parle de « readers » pour les protéines possédant la capacité à reconnaître et à se fixer à des histones ou de l'ADN modifiés et à recruter les complexes protéiques impliqués dans l'activation et la répression de l'activité.

Méthylation et acétylation sont donc des processus clés de modification temporaire de l'ADN et des protéines qui lui sont associées. Ces modifications chimiques sont finement contrôlées et ne concernent que certaines régions bien particulières de ces biomolécules.

### I.1.4 La méthylation de l'ADN

La méthylation de la cytosine est la principale modification touchant l'ADN.<sup>33</sup> Ce processus consiste en l'addition régiosélective d'un groupement méthyle sur le carbone 5 (Figure 6) de l'hétérocycle et est catalysé par les enzymes de la famille des méthyltransférases de l'ADN (DNMT).<sup>34</sup> Ces enzymes présentent au sein de leur structure une poche de liaison au cofacteur S-adénosyl-L-méthionine (SAM), le transporteur et donneur de méthyle naturel. Cette transformation affecte principalement les séquences d'ADN riches en dinucléotides CG, nommées « îlots de CpG », le p représentant le groupement phosphate reliant ces deux nucléotides.<sup>35,36</sup>

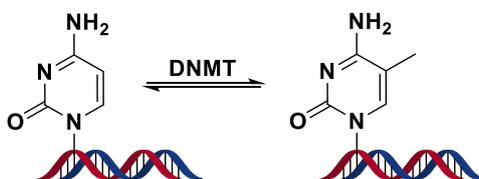


Figure 6: La méthylation de la cytosine

Il existe deux grandes classes de DNMT dans les cellules de mammifères: les méthyltransférases de l'ADN *de maintenance* (DNMT1) et celles *de novo*.<sup>37</sup>

Plusieurs études ont démontré que DNMT1 avait une meilleure affinité pour l'ADN dit hémiméthylé, où les îlots de CpG d'un seul brin sont méthylés, en comparaison à l'ADN non-méthylé.<sup>38,39,40,41</sup> Ainsi, DNMT1 est plutôt décrite comme une enzyme de maintenance, qui interviendrait pendant la réplication de l'ADN en méthylant le brin fille nouvellement synthétisé (et donc dépourvu de méthylation) afin de maintenir le schéma de méthylation sur l'ensemble de la double hélice.<sup>42,43,44,45</sup> En plus de son principal rôle de maintenance, autrement dit de transmission du profil de méthylation du brin parent vers le brin fille, l'activité de DNMT1 semble également requise pour la méthylation *de novo*.<sup>46,47</sup> Celle-ci est portée par la famille des DNMT3, qui comprend trois membres : DNMT3a, DNMT3b et DNMT3L. Contrairement à DNMT1, DNMT3a et DNMT3b sont essentiellement responsables de la méthylation sur un site CpG d'une molécule ADN double-brin non méthylée (*de novo*) et ont été identifiées comme responsable de l'établissement des schémas de méthylation lors du développement embryonnaire.<sup>48</sup> De plus, elles ne présentent aucune préférence significative entre l'ADN hémiméthylé et non méthylé.<sup>49,50</sup>

Plus tard, il a été démontré qu'elles jouent également un rôle dans le maintien de la méthylation de l'ADN dans les régions hétérochromatiques.<sup>51,52</sup> A l'inverse, DNMT3L est catalytiquement inactive mais aide DNMT3a et DNMT3b à moduler leurs activités et fonctionne donc comme un facteur de régulation dans les cellules germinales.<sup>53</sup> Les modifications épigénétiques étant d'une manière générale réversibles, la déméthylation de l'ADN est un processus qui a été largement étudié. Celui-ci est plus complexe que le mécanisme de méthylation et implique plusieurs voies et mécanismes : la première étape consiste en une oxydation de la 5-méthylcytosine en 5-hydroxyméthylcytosine par les protéines de la famille TET, suivie par d'autres étapes d'oxydation conduisant à la formation de 5-formylcytosine et de 5-carboxylcytosine. Chacun de ces produits peuvent être converti indépendamment en cytosine selon divers mécanismes. La restauration de la cytosine (non modifiée) est réalisée par l'action des glycosylases de la thymine et se termine par l'excision de la base et réparation de l'ADN (BER).<sup>54,55</sup>

### **I.1.5 Les modifications post-traductionnelles des histones**

La plupart des modifications post-traductionnelles (PTM) des histones ont lieu en général sur leurs chaînes *N*-terminales. Les résidus d'acides aminés localisés dans les queues des histones pointant vers l'extérieur des nucléosomes sont les cibles privilégiées de multiples

modifications qui vont moduler l'expression génétique. A ce titre, nous décrivons les trois modifications les plus importantes : méthylation, acétylation et phosphorylation.<sup>56,57</sup>

### I.1.5.1 Phosphorylation des histones

La phosphorylation est une modification épigénétique qui a lieu majoritairement sur les résidus sérine, thréonine et tyrosine.<sup>56</sup> Ce processus est catalysé par les kinases et permet l'ajout d'un phosphate sur l'hydroxyle de la chaîne latérale des résidus concernés (Figure 7).<sup>58</sup>

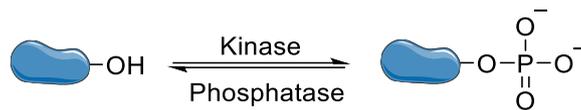


Figure 7: Phosphorylation des hydroxyles

Cette modification ajoute deux charges négatives qui influent sur l'affinité des histones pour l'ADN et jouent donc sur la structure de la chromatine.<sup>56</sup> La phosphorylation des histones est impliquée dans plusieurs processus biologiques tels que le remodelage de la chromatine et l'expression des gènes. Une des modifications la plus étudiée est la phosphorylation de la sérine 139 (dans les mammifères) du variant d'histone H2A.X par deux kinases intitulées Ataxia telangiectasia mutated (ATM) et AT RAD3-related (ATR).<sup>59,60</sup> Cette modification a lieu en réponse à certains dommages (comme la cassure de double brin (CBD) par exemple) de l'ADN et qui permet le recrutement de la machinerie de réparation de l'ADN.<sup>60</sup>

Le processus d'hydrolyse de ces phosphates quant à lui est contrôlé par les phosphatases, permettant de restaurer l'hydroxyle originel.

### I.1.5.2 L'acétylation des lysines

L'acétylation des histones a été rapportée pour la première fois en 1964.<sup>61</sup> Son degré est contrôlé par deux familles d'enzymes, aux effets inverses : les histones acétyltransférases (HAT) (ou les KAT (lysines acétyltransférases) d'après la nouvelle nomenclature) et les histones désacétylases (HDAC) (Figure 8).<sup>62,63</sup>

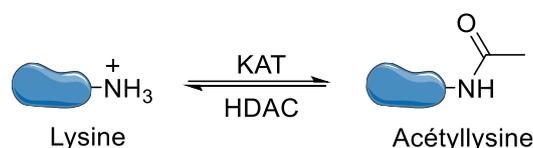


Figure 8: L'acétylation des lysines

Les KAT manipulent le cofacteur acétyl-coenzyme A (acétyl-coA) et catalysent le transfert d'un groupement acétyl vers la chaîne latérale des lysines, formant ainsi une liaison amide. Il en découle une neutralisation de la charge positive des lysines entraînant une diminution de l'affinité des lysines aux phosphates (chargés négativement) de l'ADN. Cette absence d'interactions électrostatiques entre l'ADN et les histones a pour conséquence la relaxation de la structure de la chromatine autour de la zone acétylée qui devient plus accessible aux facteurs de transcription.<sup>56</sup> De plus, la liaison amide devient le site de nouvelles interactions protéine-protéine permettant la régulation génétique. En particulier, les protéines déployant des bromodomains (séquences d'environ 110 acides aminés - qualifié de « reader ») qui reconnaissent spécifiquement les lysines acétylées.<sup>64</sup> Ces bromodomains recrutent ensuite d'autres protéines régulatrices. En effet, les acétyllysines sont souvent liées par des KAT ou des complexes de remodelage de la chromatine (qui arborent un bromodomaine dans leurs structures).<sup>65</sup> Par exemple, Swi2/Snf2 contient un bromodomaine qui se lie sur les histones acétylées. A son tour, ce dernier recrute le complexe de remodelage SWI/SNF dont sa fonction est de décondenser la chromatine.<sup>56,66,67</sup>

### I.1.5.3 Méthylation des lysines

La méthylation détient un rôle conséquent dans la régulation des gènes. Les histones sont très riches en lysine qui peuvent être méthylées à différents degrés : mono-, di-, ou triméthylées (Figure 9). Le transfert du méthyle provenant du SAM vers l'amine de la chaîne latérale de la lysine est catalysé par les enzymes nommés histones lysines méthyltransférases (HKMT).<sup>68,69</sup>

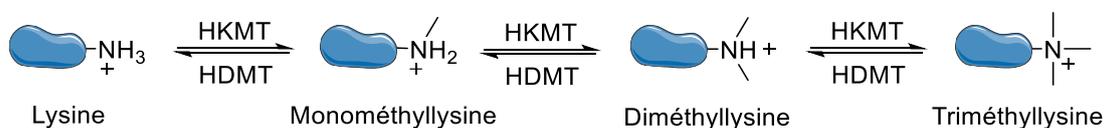


Figure 9: La méthylation des lysines

Contrairement à l'acétylation, la méthylation n'altère pas la charge globale de la lysine modifiée et donc n'a pas de répercussion directe sur la structure de la chromatine. Au lieu de cela, les méthyles agissent comme des sites de fixation à d'autres protéines qui vont compacter les nucléosomes ou bien lier des protéines régulatrices.<sup>70</sup> Ces protéines régulatrices vont reconnaître un signal bien spécifique dépendant du degré de méthylation (mono di ou tri méthylation) et de la séquence d'acides aminés environnante.<sup>70</sup> C'est pour cette raison que le rôle biologique de la méthylation est complexe et difficile à élucider : selon le résidu méthylé et le nombre de méthyle ajoutés, des effets opposés peuvent être constatés. A titre

d'illustration, la méthylation des lysines 4, 36 et 79 de l'histone 3 (H3K4, H3K36, H3K79) entraîne une activation de la transcription tandis que la méthylation des lysines 9 et 27 de cette même histone (H3K9, H3K27) provoque sa répression.<sup>71, 68</sup>

Les méthyles des lysines sont éliminés par les lysines déméthylases (KDMT), réparties en deux familles : la première comprend les déméthylases à domaine Jumonji (JHDM, Jumonji C (JmjC) domain-containing histone demethylase), dont l'activité de déméthylation nécessite l'ion  $Fe^{2+}$  et l' $\alpha$ -cétoglutarate comme cofacteurs ; la seconde comprend les déméthylases dites *Lysine-specific histone demethylases* (LSD) nécessitant la flavine adénine dinucleotide (FAD) comme cofacteur pour exercer leur rôle.<sup>70,72,73</sup>

#### I.1.5.4 La méthylation des arginines

Les arginines peuvent se prêter à plusieurs modifications telles que la citrullination (voir plus loin) et l'ADP ribosylation mais la méthylation est la plus importante et la plus répandue et sera le sujet de cette thèse.<sup>74,75</sup> Le processus de méthylation est catalysé par les membres de la famille des Protéines arginines méthyltransférases (PRMT) via le transfert d'un méthyle provenant du donneur, le SAM, vers la chaîne latérale des résidus d'arginines (Figure 10).<sup>76</sup> Les arginines possèdent une fonction guanidine sur leur chaîne latérale, dont les trois atomes d'azote - à savoir  $N^\delta$ ,  $N^{\omega'}$  et  $N^{\omega''}$  - peuvent être méthylés. Les PRMT sont ainsi classiquement divisées en quatre classes,<sup>3</sup> en fonction de l'atome d'azote qu'elles méthylent (Figure 10).

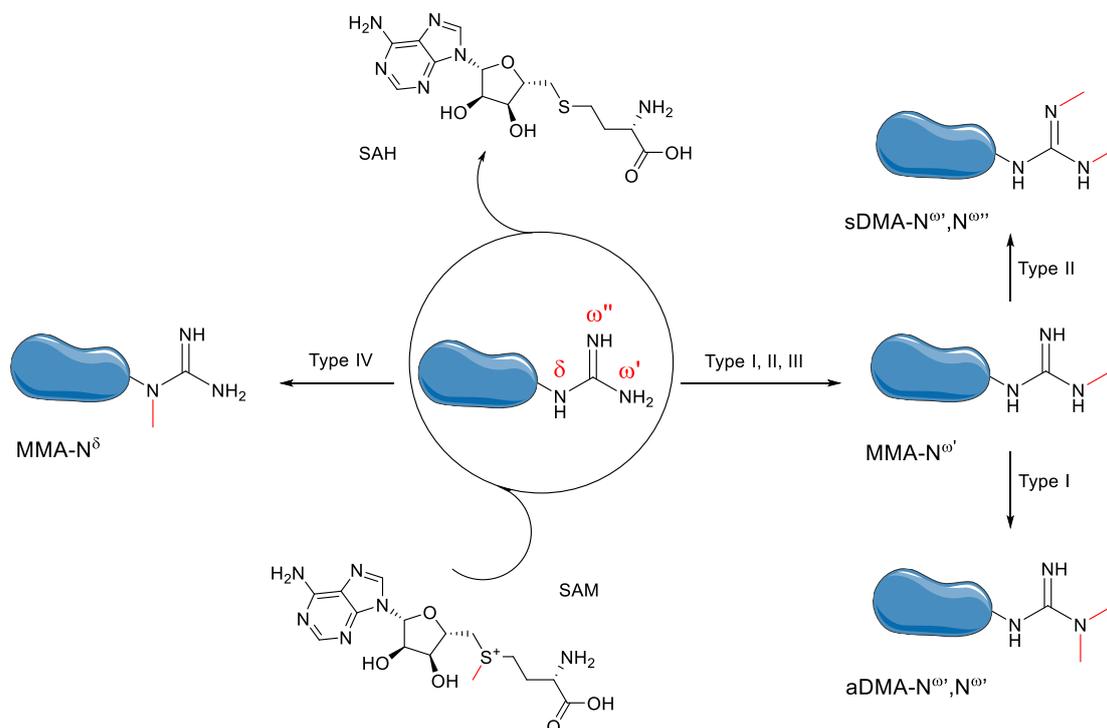


Figure 10: La méthylation des arginines catalysées par les PRMT

Les PRMT de type I, II et III catalysent la monométhylation de la position  $\omega$ , conduisant à la  $N^{\omega}$ -monométhylarginine (MMA- $N^{\omega}$ ), tandis que les PRMT de type IV catalysent la méthylation  $\delta$  et la formation de  $N^{\delta}$ -monométhylarginine (MMA- $N^{\delta}$ ). Bien que les PRMT de type III (souvent appelées PRMT7) s'arrêtent à la monométhylation, les PRMTs de type I et II peuvent catalyser une deuxième étape de méthylation si les processus cellulaires l'exigent.

La diméthylation peut alors se produire sur l'une des deux positions  $\omega$  :

- sur le même azote  $\omega$  que la première méthylation, conduisant alors à la  $N^{\omega},N^{\omega}$ -diméthylarginine asymétrique (aDMA- $N^{\omega},N^{\omega}$ ), un processus catalysé par les PRMTs de type I – une classe de PRMT comprenant PRMT1, PRMT2, PRMT3, PRMT4 (également connue sous le nom d'arginine méthyltransférase 1 associée au coactivateur ou CARM1), PRMT6 et PRMT8
- sur l'autre position  $\omega$ , conduisant alors à la  $N^{\omega},N^{\omega'}$ -diméthylarginine symétrique (sDMA- $N^{\omega},N^{\omega'}$ ) via un processus catalysé par les PRMT5 et 9, membres de la classe des PRMT de type II (Figure 10).

Etant donné que le  $pK_a$  du groupement guanidinium est déjà relativement élevé ( $\sim 13.8$ ), la méthylation ne modifie pas la charge nette des arginines à pH physiologique. Cependant, elle va influencer sur l'hydrophobicité du groupement tout en accroissant sa taille, réduisant ainsi sa capacité à faire des liaisons hydrogènes. Toutes ces caractéristiques suggèrent que les motifs de méthylarginines sont reconnus selon des mécanismes différents étant donné que les interactions protéines-protéines sont différentes.<sup>7,77</sup>

### ➤ La déméthylation des arginines :

Nous avons vu pour les modifications épigénétiques précédentes que leur réversibilité était essentielle car elle permettait de corriger le taux d'expression des gènes en fonction du besoin de la cellule. Dans le cas de la déméthylation des arginines cependant, l'identité des enzymes dites « erasers » responsables reste inconnue. Plusieurs pistes ont été néanmoins explorées afin d'identifier ces protéines. Les peptidylarginine déiminases (PAD), catalysant la réaction de déimination des arginines (non méthylés) afin de les convertir en citrulline, ont ainsi été suggérées comme pouvant être responsables de la conversion des méthylarginines en citrulline (Figure 11).<sup>78,79,80</sup>

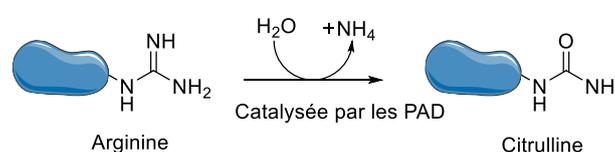


Figure 11: La citrullination des arginines

Cependant, cette hypothèse s'est vu infirmer deux ans plus tard avec la preuve que seuls des peptides synthétiques portant des résidus d'arginine étaient des substrats efficaces des enzymes humaines PAD2, PAD3, et PAD4 et non ceux porteurs de leurs analogues méthylés.<sup>81</sup>

Il a également été démontré que quelques lysines déméthylases (telle que KDM3A, KDM4E, KDM5C et KDM6B qui sont des protéines à domaine jmjC) étaient pourvues d'une certaine activité sur les arginines diméthylés *in vitro* mais que leur influence *in vivo* était négligeable.<sup>82</sup> Une autre étude publiée en 2017<sup>83</sup> a mis en évidence que deux protéines à domaine jmjC (JMJD5, JMJD7) ne déméthylent pas les méthylarginines mais qu'elles sont pourvus d'une activité de protéase qui leur confère la capacité de cliver les queues des histones H2, H3 et H4 possédant des méthylarginines. Cependant, en 2018 une autre équipe suggère que JMJD5 aurait plutôt une activité d'hydroxylation et non un rôle protéolytique.<sup>84</sup>

Bien que les données rapportées jusqu'ici ne permettent pas de lever le voile sans ambiguïté sur l'identité des enzymes responsables de leur déméthylation, les processus de méthylation des arginines – et donc les PRMT – sont essentiels à quantité de fonctions biologiques.

## I.2 Les Protéine Arginines Méthyltransférases

### I.2.1 Rôles et Fonctions biologiques des PRMT

Depuis que la méthylation des arginines a été constatée, celle-ci a été étudiée pendant de nombreuses années afin de comprendre son fonctionnement. Il a été démontré que neuf PRMT étaient indispensables chez les mammifères et intervenaient sur les processus cellulaires gouvernant l'activité des gènes.<sup>85,86,87</sup> Les PRMT partagent un cœur catalytique remarquablement conservé (partie en bleu, Figure 12), indispensable pour leur activité méthyltransférase. A ce domaine central s'ajoutent d'autres domaines sur les parties *N*- ou *C*-terminales qui viennent donner à la PRMT en question sa spécificité et sélectivité (Figure 12).<sup>85,86,87</sup>

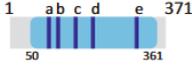
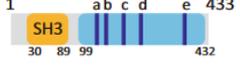
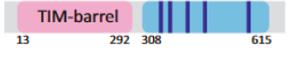
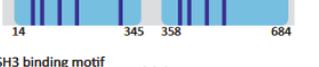
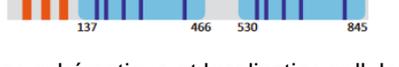
Name	Domain structure	Classification	Cellular localization
PRMT1		type I	cytoplasm/nucleus
PRMT2		type I	cytoplasm/nucleus
PRMT3		type I	cytoplasm
CARM1		type I	cytoplasm/nucleus
PRMT5		type II	cytoplasm/nucleus
PRMT6		type I	nucleus
PRMT7		type III	cytoplasm/nucleus
PRMT8		type I	plasma membrane
PRMT9		type II	cytoplasm

Figure 12: Représentation schématique et localisation cellulaire des neuf PRMT caractérisées chez les mammifères<sup>85</sup>

Nous verrons ainsi brièvement la fonction de chaque PRMT et les mécanismes dans lesquels elles sont impliquées pour réguler les processus cellulaires.

### **I.2.1.1 PRMT de Type I**

#### **I.2.1.1.1 PRMT1 :**

PRMT1 est une arginine méthyltransférase de type I, et c'est l'enzyme prédominante dans la famille des PRMT, étant responsable à elle seule de plus de 90%<sup>88,85</sup> de la méthylation des arginines dans les cellules de mammifères et plus précisément de la méthylation des aDMA. On la retrouve aussi bien dans le noyau que dans le cytoplasme et elle reconnaît un grand nombre de substrats (histones ou pas). Comme toutes les autres PRMT de type I (sauf PRMT4), elle catalyse la diméthylation des résidus d'arginines se trouvant dans les régions riches en glycine et arginine (GAR, pour *glycine-arginine rich*).<sup>89,90</sup>

L'une des principales cibles de PRMT1 est l'arginine R3 portée par l'histone H4 (H4R3), qui une fois méthylée facilite l'acétylation des résidus lysines des queues de H4 par l'enzyme p300, permettant ainsi l'activation de la transcription.<sup>91,92,93</sup> PRMT1 est également impliquée dans la réparation de l'ADN via la méthylation du motif GAR de 53BP1 (p53 Binding Protein 1).<sup>94</sup> Dans le cytoplasme, elle méthyle le domaine de liaison à l'ADN (DBD) du récepteur des œstrogènes alpha (E $\alpha$ ). Cela permet de contrôler la survie et la prolifération des cellules en régulant les voies de signalisation, notamment en intervenant sur le mécanisme de signalisation rapide des œstrogènes. Ce récepteur (E $\alpha$ ) est hyperméthylé dans certains cancers du sein, faisant de cette PRMT une cible étudiée en oncologie.<sup>95</sup>

#### **I.2.1.1.2 PRMT2 :**

PRMT2 catalyse également la méthylation des aDMA et se trouve aussi bien dans le cytoplasme que dans le noyau. PRMT2 a été identifiée comme coactivateur de plusieurs récepteurs nucléaires, tels que les récepteurs aux œstrogènes et androgènes.<sup>96,97</sup> De plus, le lien entre PRMT2 et les récepteurs aux œstrogènes aurait là aussi une influence dans le développement du cancer du sein.<sup>98</sup> Cette enzyme utilise d'autres moyens pour réguler la transcription, notamment en méthylant l'arginine R8 de l'histone H3 (H3R8), un rôle essentiel dans l'établissement de l'architecture de la chromatine lors du développement embryonnaire.<sup>99</sup> Elle se distingue des autres PRMT par sa partie N-terminale qui comporte un domaine SH3. Le premier substrat non-histone de PRMT2 a été identifié en 2017 par l'équipe de Jean-Cavarelli. C'est le sujet de cette thèse et sera discuté plus en détail dans le deuxième chapitre.

#### I.2.1.1.3 PRMT3 :

PRMT3 est la seule PRMT de type I qui ne présente pas de localisation nucléaire et se trouve essentiellement dans le cytoplasme.<sup>100</sup> Il a été initialement proposé que PRMT3 était une protéine de liaison à PRMT1. Cependant, il a été démontré par la suite que celui-ci s'associe comme la plupart des PRMT sous la forme d'homodimères.<sup>100,76</sup> Elle se différencie des autres PRMT par la présence d'un domaine à doigt de zinc situé à son extrémité *N*-terminale, et qui semble essentiel pour la reconnaissance des substrats.<sup>101</sup> Plusieurs substrats ont été identifiés (*in vitro*) et attribués à PRMT3 mais seul la protéine rpS2 (ribosomal protein S2) a été rapportée comme substrat *in vivo* de PRMT3.<sup>102,103</sup>

#### I.2.1.1.4 PRMT4/CARM1 :

PRMT4 est la PRMT la plus étudiée suite à sa découverte en 1999 par l'équipe de Stallcup.<sup>104</sup> Elle fut alors nommée *Coactivator-associated arginine méthyltransférase 1* (CARM1) due à son association à la famille de coactivateurs p160 (SRC-1, SRC-2, SRC-3), qui coactivent la transcription des gènes en recrutant d'autres coactivateurs possédant une activité histone-acétyltransférase ou -méthyltransférase. *In vivo*, PRMT4 diméthyle asymétriquement l'histone H3, plus précisément les positions R17 et R26 (H3R17me2a, H3R26me2a), des marquages qui sont associés à l'activation de la transcription. De plus, PRMT4 méthyle tout un ensemble d'autres protéines impliquées dans multiples mécanismes,<sup>105</sup> telle que la *Polyadenylate-binding protein 1* (PABP1),<sup>106</sup> une protéine se liant à la partie poly A des ARN messagers (ARNm). Bien que l'importance de sa méthylation n'ait pas encore été expliquée et comprise exactement, PABP1 a plusieurs rôles comme le transport de l'ARNm hors du noyau. PRMT4 possède ainsi toute une panoplie de substrats, qui ne seront pas tous énumérés ici mais il convient de mentionner que ceux-ci ne possèdent pas de motif GAR, PRMT4 méthylant préférentiellement des motifs PGM (proline-glycine-méthionine) suggérant ainsi un mode de reconnaissance différent pour cette PRMT. L'implication de PRMT4, allant de l'activation de la transcription (autant par sa co-activation des récepteurs nucléaires que par sa méthylation de l'histone H3) jusqu'à la régulation de l'épissage alternatif et la maturation des ARNm, montre l'importance de cette dernière et la place au cœur de la vie cellulaire.

#### I.2.1.1.5 PRMT6 :

PRMT6 est principalement localisée dans le noyau et méthyle les motifs GAR. Elle est responsable de la méthylation de la position R2 de l'histone H3 (H3R2me2a).<sup>107,108,109</sup> La méthylation de H3R2 antagonise la modification H3K4 (associé à l'activation de la transcription), ce qui en fait un répresseur transcriptionnel. De plus, il a été démontré qu'elle

méthylerait l'arginine 29 de l'histone H2A, un marquage qui est associé à une répression de l'activité transcriptionnelle.<sup>110</sup> En plus de cela, PRMT6 jouerait également un rôle de coactivateur de la transcription dans les cellules hépatiques où elle méthylerait le coactivateur CRTC2 (CREB Regulated Transcription Coactivator 2) du facteur de transcription CREB (cAMP response element-binding protein).<sup>111</sup>

#### **I.2.1.1.6 PRMT8 :**

PRMT8 a été identifié en 2005 par homologie de séquence par rapport aux autres PRMT. En effet, celle-ci affiche un niveau élevé d'identité de séquence avec PRMT1 mais diffère selon deux points. Premièrement, son extrémité *N*-terminale est myristoylée ce qui entraîne sa fixation à la membrane plasmique, bien qu'une étude plus récente suggère qu'elle serait plutôt une enzyme nucléaire spécifique des neurones. Cette localisation dans les neurones a à voir avec la deuxième propriété particulière de PRMT8 : son expression dans le cerveau uniquement.<sup>112,113</sup>

### **I.2.1.2 PRMT de Type II**

#### **I.2.1.2.1 PRMT5 :**

PRMT5 est l'arginine méthyltransférase de type II prédominante chez les mammifères. Elle a d'abord été identifiée comme une protéine de liaison avec la tyrosine kinase JAK2 (Janus Kinase 2) d'où son nom initial de Jak Binding Protein 1 (JABP1).<sup>114</sup> Sa fonction de méthyltransférase n'a été mise en évidence que plus tardivement, avec l'identification des histones H2A, H3 et H4 comme substrats préférentiels.<sup>115</sup> Elle est recrutée au sein de plusieurs facteurs de transcription et de répression et elle est beaucoup plus impliquée dans la répression que l'activation.<sup>116</sup> Elle est essentiellement cytoplasmique même si elle se trouve également dans le noyau. Elle rejoint les autres PRMT par sa forte expression dans les cellules cancéreuses et est impliquée dans la tumorigenèse.<sup>117,105</sup>

#### **I.2.1.2.2 PRMT9 :**

Elle a été découverte en même temps que PRMT8, par homologie de séquence par rapport aux autres PRMT. En 2005, il a été mis en évidence qu'elle diméthylait les arginines symétriquement, tout comme PRMT5.<sup>118</sup> Plus précisément, il a été démontré que PRMT9 interagissait avec les facteurs d'épissage SAP145 (spliceosome-associated protein 145) et SAP49 (spliceosome-associated protein 49) pour former un complexe permettant la diméthylation de SAP145 au niveau de R408. Cette diméthylation génère alors un site de

liaison pour le domaine de Tudor de la protéine SMN (Survival of Motor Neuron), dont le rôle est de faciliter l'assemblage du spliceosome. Ainsi, il a été mis en évidence que l'atténuation de l'expression PRMT9 affectait le mécanisme d'épissage de l'ARN, ne pouvant plus méthylé SAP145.<sup>118</sup>

### **I.2.1.3 PRMT de Type III**

#### **I.2.1.3.1 PRMT7 :**

PRMT7 est unique car c'est une PRMT de type III – elle s'arrête ainsi à la monométhylation et est incapable de catalyser la diméthylation. Cependant, elle a été l'objet de vifs débats : une étude publiée en 2005<sup>119</sup> a proposé que PRMT7 catalysait la formation de sDMA. Cependant, des travaux antérieurs démontraient l'inverse et suggéraient qu'elle s'arrêtait en fait à la monométhylation.<sup>120</sup> Le débat a été tranché par deux études ultérieures<sup>121,122</sup> prouvant que PRMT7 s'arrêtait bien à la monométhylation, confortant donc les conclusions de l'étude initiale (il s'est avéré par la suite que les résultats de la seconde étude étaient dues à la contamination de PRMT7 par de la PRMT5). La structure cristallographique de PRMT7 murine a été décrite<sup>123</sup> par l'équipe de Jean Cavarelli, démontrant que celle-ci était composée de deux modules catalytiques, dont l'un était inactif. PRMT7 est impliquée dans divers processus tels que la régulation de transcription, la réparation des dommages à l'ADN, l'épissage de l'ARN etc.

## II. Objectifs

L'épigénétique est une branche de la biologie assez récente et en pleine croissance permettant de dévoiler les mécanismes de la régulation des gènes. Ces mécanismes épigénétiques et spécialement la méthylation des arginines est une cible thérapeutique émergente et particulièrement important dans les cancers. Cependant, l'émergence d'inhibiteurs a été limité par le manque d'informations structurales sur la reconnaissance des substrats des PRMT. De nombreuses recherches sont menées pour apporter davantage de données biochimiques et structurales sur les complexes incorporant des PRMT. Cette connaissance est d'une grande importance car elle permet non seulement de comprendre les bases fonctionnelles de la méthylation des arginines mais aussi d'accéder à des composés à fort pouvoir inhibiteurs pour cibler ces enzymes qui régule aussi bien la fonction des protéines histone que non-histone. Mes travaux de thèse ont portés sur cette thématique et se décompose en deux axes:

- Synthèse de ligands mimant l'état de transition de la première ou seconde étape de méthylation des différents types de PRMT et l'incorporation de ces mimes de l'état de transition au sein des substrat entier de PRMT
- Synthèse d'inhibiteurs sélectifs de CARM1

Ainsi, le premier axe a pour ambition de résoudre les difficultés liées à l'identification et caractérisation de complexe PRMT-substrat. Elle repose sur une approche consistant à lier de manière covalente une molécule analogue du SAM (mime de l'état de transition de la réaction de méthylation) à l'arginine cible d'un substrat. Cette approche favoriserait l'ancrage du substrat à la surface du PRMT et de cette manière de stabiliser le complexe, afin d'aplanir les difficultés liées à l'identification de complexes PRMT-substrat peu stables.

Dans ce cadre on s'est focalisé sur le facteur d'épissage RSF1 et PRMT2 qui forme un complexe ternaire constitué d'un dimère de PRMT2 fixant une molécule de RSF1. Ainsi, ce premier axe repose sur la stabilisation de ce complexe par des mimes de l'état de transition de la réaction de méthylation dont leurs synthèses seront décrites en détail dans ce manuscrit. Cette première approche permettra à long terme de:

- Comprendre comment les PRMT reconnaissent et se lient à leurs substrats et appose le méthyle du SAM sur les arginines (de manière sélective) sur ces dernières

- Concevoir des inhibiteurs plus sélectifs en étudiant les modes de reconnaissance PRMT-substrat

Le deuxième axe de ce projet a pour ambition de rechercher de nouveaux inhibiteurs de CARM1. Ainsi, la deuxième partie de ma thèse porte sur la synthèse d'un inhibiteur dont la structure a été sélectionnée via une approche *in silico* et basé sur des travaux précédents réalisées par le Dr Samira Ajebar et Dr Julie Schmitt. L'originalité de notre approche (ayant comme ambition de fournir un inhibiteur sélectif d'un seul PRMT) repose sur le fait que le candidat inhibiteur cible une nouvelle région (appelé poche SATT) qui possède une séquence très variable parmi les PRMT.

# III. Synthèses d'arginines modifiées et incorporation régiosélective dans des protéines d'intérêt

## III.1 Introduction

La méthylation de l'arginine est une PTM très répandue, et est un mécanisme épigénétique majeur de la régulation des fonctions protéiques dans les cellules eucaryotes.<sup>124</sup> En effet, ces modifications jouent sur les propriétés biophysiques des protéines, affectant ainsi leur stabilité, leur localisation et leurs interactions et donc leur rôle dans la cellule.<sup>87</sup>

Dans le premier chapitre, nous avons décrit que plusieurs facteurs de transcription, de coactivateurs et une variété d'enzymes ont des arginines méthylées *in vivo*. Ces méthylations ont une répercussion sur tous les aspects de la biologie cellulaire, y compris le contrôle de la transcription,<sup>91</sup> la cascade de signalisation des récepteurs nucléaires, l'épissage et maturation des ARNm ou la réparation de l'ADN.<sup>106,55</sup> De plus, il a été démontré que la privation de PRMT1 chez les souris transgéniques a été associée à la sénescence cellulaire ainsi qu'à un vieillissement prématuré.<sup>125</sup> Ainsi, on peut voir que les PRMT sont associées à d'innombrables processus cellulaires primordiaux. C'est pour cette raison que de nombreuses études ont associé la dérégulation du processus de la méthylation des arginines au développement de nombreuses maladies comme les cancers mais aussi les maladies chroniques. Dans ces maladies, le niveau d'expression des PRMT est généralement élevé classant les PRMT comme des cibles potentielles pour traiter ces pathologies. Plusieurs groupes et entreprises pharmaceutiques (GSK, Pfizer...) se penchent sur cette problématique pour trouver des inhibiteurs à visée thérapeutique. De nombreux inhibiteurs des PRMT efficaces ont été développés mais aucun n'est encore utilisé dans un cadre thérapeutique : ce n'est que récemment que des inhibiteurs de PRMT5 sont entrés dans des essais cliniques de phase 1 (voir Chapitre III. Conception et synthèse d'inhibiteurs de PRMT). Ainsi, le développement d'inhibiteurs de PRMT représente un réel besoin et l'émergence d'inhibiteurs a été refrénée par le manque de données structurales sur la reconnaissance des substrats des PRMT.

Par conséquent, caractériser les complexes PRMT-Substrat/Cible d'un point de vue structural et déchiffrer le mécanisme précis de la méthylation des arginines fournira des informations clés sur la reconnaissance du substrat, la sélectivité et la régulation du processus de méthylation des PRMT. Ces informations obtenues via des études cristallographiques approfondies pourraient inspirer la conception de petites molécules à fort pouvoir inhibiteur.

De nombreux progrès ont été réalisés dans la découverte des fonctions de chaque PRMT et l'identification de leurs cibles. On sait par exemple que les PRMT sont des protéines modulaires qui sont construites autour d'un module catalytique hautement conservé constitué

de 300 acides aminés où la fonction méthyltransférase des PRMT est assurée. C'est dans cette partie centrale que se trouvent les sites de fixation du SAM et de l'arginine cible du substrat. D'autres domaines viennent s'ajouter à cette partie centrale et donne à la PRMT sa spécificité structurale et affecte l'activité enzymatique, la liaison au substrat et la régulation.<sup>85,86,87</sup>

Actuellement, les structures des neuf PRMT caractérisées chez les mammifères ont été résolues. De plus, d'autres structures de PRMT liées à des ligands (analogue du SAM) ou des dérivés peptidiques ont été résolues, apportant chacune leur contribution à la compréhension du mécanisme de méthylation des arginines.<sup>126</sup> Cependant, aucune structure cristallographique n'a été rapportée pour les PRMT avec leurs substrats entiers à l'exception de PRMT5 dont la structure cristallographique du complexe PRMT5:MEP50 a été rapporté (PDB 4GQB).<sup>127</sup>

Par conséquent, en dépit de la multitude de données structurales disponibles, les éléments mécanistiques exacts permettant la reconnaissance des substrats et la sélectivité de la réaction de méthylation restent mal compris. La difficulté à obtenir des complexes entre les PRMT et leurs substrats peut s'expliquer par le manque de partenaire venant le stabiliser.<sup>128</sup> En effet, les PRMT font communément partie de complexes plus grands qui incluent d'autres protéines, comme les « readers » que nous avons introduits dans le premier chapitre : pour interagir avec leurs substrats et les recruter sur les sites de méthylation, les PRMT font généralement appel à d'autres partenaires essentiels au recrutement du substrat. Ainsi, le complexe PRMT-substrat peut être moins stable en absence de ces protéines partenaires essentielles. Un des exemples le mieux connu concerne PRMT5. Il a été démontré que PRMT5 doit d'abord se lier à MEP50 (Methylosome protein 50) avant de se fixer à son substrat Sm (une famille de protéines se liant à l'ARN). MEP50 agit en tant que protéine permettant le recrutement du substrat Sm, mais également en tant que stabilisateur de PRMT5, augmentant ainsi son activité de méthyltransférase.<sup>129</sup>

Un autre exemple bien connu concerne l'implication de CARM1 et PRMT1 avec les protéines des familles p160 et p300. La famille de protéines p160 recrute par son domaine AD1 (Activation domain 1) la famille de protéines p300 (responsable de l'acétylation de H3K18) alors que son domaine AD2 reconnaît CARM1 et PRMT1 (responsable respectivement de la méthylation de l'histone H3 et l'histone H4).

Ces différentes modifications agissent en synergie, puisqu'il a été montré que la méthylation de H4R3 par PRMT1 facilite l'acétylation de H3K18 par p300 et que cette acétylation est nécessaire à la méthylation de H3R17 par CARM1.<sup>130,131</sup>

Ainsi, les PRMT opèrent de concert en milieu biologique avec d'autres partenaires et se trouvent en même temps stabilisées par ces derniers. Cependant, le fonctionnement

transitoire de ces complexes ainsi que leur stœchiométrie restent très mal connus. Même si les PRMT partagent quelquefois certains substrats, chaque PRMT possède sa propre panoplie de substrat et ce malgré une structure des sites de liaison aux arginines hautement conservée. Par conséquent, trouver des stratégies pour résoudre les difficultés liées à l'identification de complexes PRMT-Substrat peu stables est essentiel.

Trois problématiques majeures peuvent donc être dégagées, dépendantes les unes des autres :

- Comprendre comment les PRMT reconnaissent et se lient à leurs substrats.
- Décrire comment les PRMT appose le méthyle du SAM sur les arginines (de manière sélective) sur différentes protéines.
- Caractériser les complexes peu stables impliquant PRMT-Substrat ou PRMT-Substrat-Partenaires.

C'est dans ce contexte que s'inscrit ce projet porté par l'équipe du Professeur Jean Cavarelli (IGBMC, Département de Biologie Structurale Intégrée, CNRS UMR 7104, Inserm U964, Illkirch, France) et financé par l'agence national de recherche (ANR).

En 2017, l'équipe de Jean Cavarelli a identifié RSF1 (Repressor Splicing Factor 1) comme le substrat de PRMT2.<sup>132</sup> Dans cette étude, la stœchiométrie et l'homogénéité du complexe ont été déterminées. Il a été démontré par plusieurs méthodes biophysiques qu'un complexe asymétrique se formait *in vitro* au sein duquel une molécule de RSF1 se trouvait complexée par un dimère de PRMT2. Au total, six sites de méthylation ont été identifiés sur RSF1 : quatre résidus arginine sont monométhylés (R24, R100, R111 et R139) et deux sont diméthylés (R7 et R120).

De surcroît, une reconstruction par cryomicroscopie électronique (Cryo-EM) a été réalisée confirmant les résultats précédents. Cependant, l'analyse des données Cryo-EM a également révélé la présence de plusieurs conformations et une certaine flexibilité interne, expliquant la faible résolution observée (Figure 13)

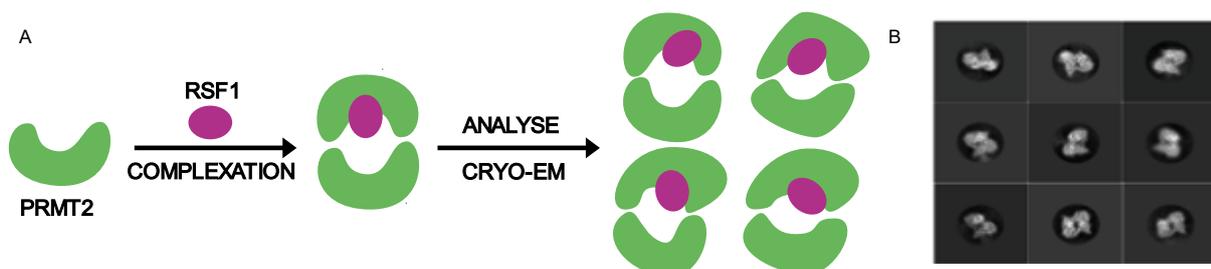


Figure 13: A) Représentation schématique de la complexation de PRMT2/RSF1 et son instabilité visualisée via la Cryo-EM. B) Analyse Cryo-EM et la représentation 2D des résultats obtenus avec le complexe PRMT2/RSF1.

De nombreux essais de cristallisation du complexe PRMT2/RSF1 ont été réalisés mais la collecte de données des cristaux obtenus a révélé qu'ils étaient composés uniquement de PRMT2. Il semblerait que la flexibilité du complexe PRMT2-RSF1 qui a été mise en évidence pourrait être la raison pour laquelle RSF1 n'est pas retenue et se dissocie durant le processus de cristallisation. La flexibilité observée peut être due à plusieurs facteurs, comme l'absence de partenaires qui viendrait stabiliser le complexe.<sup>128</sup>

Pour réduire cette hétérogénéité conformationnelle et la flexibilité observée, nous avons entrepris de stabiliser davantage le complexe PRMT2/RSF1 en insérant un mime de l'état de transition aux différents sites de méthylation de RSF1. En effet, notre approche pour résoudre les limitations actuelles d'identification et de visualisation de complexes peu stables consiste à modifier de manière irréversible l'arginine d'un substrat de telle sorte à ce qu'elle ressemble à un état de transition possiblement impliqué dans le mécanisme de méthylation des arginines. Pour ce faire, il est possible de lier de manière covalente une molécule analogues du SAM à l'arginine cible d'un substrat. Cet analogue du SAM est un mime de l'état de transition de la réaction de méthylation.

La preuve de concept a été faite par l'équipe de Jean Cavarelli, en collaboration avec l'équipe du Dr. Nathaniel I. Martin (Université d'Utrecht, Pays Bas).<sup>126</sup> Ils ont démontré que l'insertion d'un mime de l'état de transition au sein d'un peptide augmentait son affinité envers les PRMT. Pour ce faire, ces deux équipes ont remplacé l'arginine naturelle cible au sein d'un peptide **1**, par le mime de l'état de transition **2**, ce qui a permis de capturer et stabiliser des complexes PRMT-peptide.<sup>126</sup> Ce mime est composé d'un dérivé d'adénosine destiné à mimer le SAM (Figure 14).

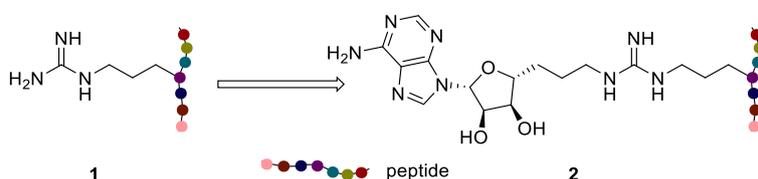


Figure 14: Remplacement de l'arginine naturelle dans les dérivés peptidiques

Dans leur étude, ils ont déterminé qu'un linker à trois atomes entre le groupement guanidine et le fragment adénosine était optimal pour mimer la géométrie de l'état de transition (Figure 15). C'est cette longueur de chaîne carbonée qui permet de positionner correctement l'adénosine et la guanidine dans le site catalytique. Par conséquent, la guanidine se trouve bien impliquée dans trois liaisons hydrogènes avec les résidus cruciaux (partagés par tous les PRMT) pour la catalyse de la réaction de méthylation. La guanidine forme deux liaisons

hydrogènes avec E258 et E267 pour former la fameuse double boucle de glutamate et une liaison hydrogène avec H415. Les deux glutamates permettent d'orienter la guanidine pour permettre la méthylation, tandis que l'histidine joue plutôt un rôle stabilisateur.<sup>126</sup> Pour montrer la preuve de concept, cette arginine modifiée (**R\***) a été incorporée en remplaçant une arginine d'intérêt au sein des séquences peptidiques dérivées de PABP1 (substrat bien défini de CARM1) pour générer des mimes de l'état de transition à base de peptide **2** (Figure 14). Ces peptides modifiés ont donné lieu à des complexes stables avec CARM1, apportant ainsi un outil très précieux qui pourra être utilisé dans la stabilisation des complexes PRMT-substrat. Dans le cadre notre projet, nous avons entrepris d'étendre cette stratégie pour stabiliser le complexe PRMT2/RSF1 en insérant un mime de l'état de transition dans la séquence de RSF1. Ainsi, les sites de méthylation de RSF1 seront remplacés soit par une **R\*** permettant l'accès à la RSF1 artificielle **3**, soit par un pseudo-analogue d'arginine modifiée **R\*** ici nommée **R\*\*** donnant lieu à la RSF1 artificielle **4** dont la capacité à entrer dans la poche catalytique reste à confirmer (Figure 15).

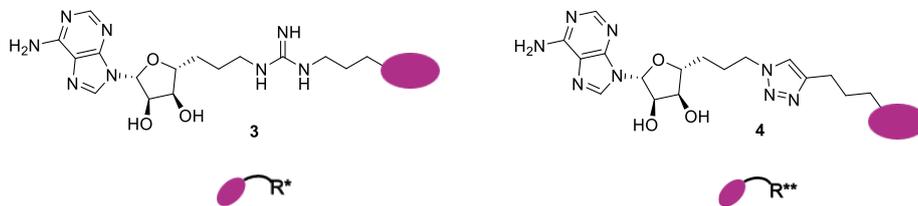


Figure 15: Modification de RSF1 par l'arginine modifiée (**R\***) ou l'analogue du SAM (**R\*\***)

Ainsi, la formation *in vitro* du complexe sera renforcée par un deuxième point d'ancrage donné par **R\*** ou **R\*\*** (Figure 16) afin de :

- Produire un échantillon de meilleure qualité pour l'étude structurale (Cryo-EM et cristallisation).
- Congeler un complexe biologique pertinent qui serait bloqué sur une conformation où PRMT2 se positionnerait possiblement juste avant le transfert du méthyle.

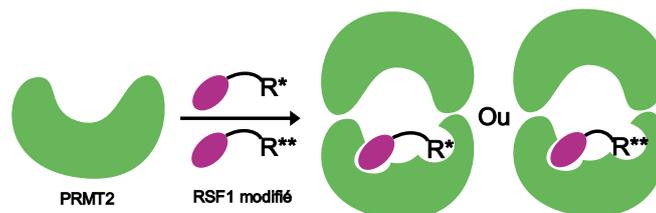


Figure 16: Stabilisation du complexe RSF1/PRMT2 par **R\*** ou **R\*\***

Ainsi, nous avons entrepris de développer deux approches différentes (indépendante l'une de l'autre) pour lier de manière covalente **R\*** ou **R\*\*** dans la séquence de RSF1. Dans un premier

temps, nous avons choisi deux sites de méthylation où incorporer nos mimes de l'état de transition:

- **R\*** à la place de l'arginine en position 7 (R7), sur la queue N-terminale.
- **R\*\*** en position 24 (R24), positionné sur la partie structurée, « foldée », de RSF1.

La complexation de PRMT2 avec RSF1 modifiée soit sur R7 soit R24 devrait fournir des informations structurales sur deux complexes de méthylation fonctionnels différents, notamment les positions relatives de PRMT2 et RSF1, ainsi que les interactions cruciales pour la réaction de méthylation.

## III.2 Première approche: Incorporation de l'arginine modifiée (R\*) sur la queue N-terminale (R7)

### III.2.1 Introduction

La première approche que nous avons entreprise dans le cadre de la stabilisation du complexe PRMT2-RSF1 est de remplacer l'arginine naturelle R7 de RSF1 par notre arginine modifiée (**R\***). Pour ce faire, nous avons choisi d'utiliser la Ligation Chimique Native (NCL).

### III.2.2 La ligation chimique native

La NCL a été décrite pour la première fois en 1994 par Kent et ses collègues et appliquée à la synthèse totale de l'interleukine humaine 8 (IL-8).<sup>133</sup> Elle est couramment utilisée de nos jours pour la synthèse de longs peptides ou de protéines à partir de peptides plus courts.<sup>134</sup> Dans la NCL, un thioester C-terminal **5** et une cystéine N-terminale **6** réagissent ensemble dans une réaction de transthioestérification liant ainsi les deux peptides. Cela conduit à la formation d'un intermédiaire thioester **7** qui subit un réarrangement avec un transfert S,N-acyle intramoléculaire, conduisant à la formation d'une liaison amide native **8** comme le nom de la réaction l'indique (Schéma 1).

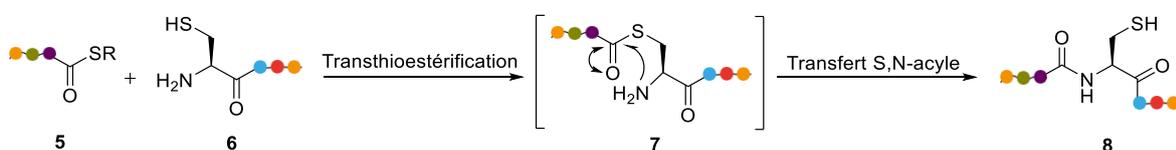


Schéma 1 : Ligation chimique native

La particularité de cette méthode réside dans son excellente régio- et chimiosélectivité, ne nécessitant pas de protection des chaînes latérales des autres résidus. Cette réaction est réalisable dans l'eau et donne généralement de très bon rendements, bien que ceux-ci dépendent fortement du type de résidu C-terminal portant la fonction thioester.<sup>135,136,137</sup> Les acides aminés porteurs de chaînes latérales encombrées telles que la valine, l'isoleucine ou la thréonine sont corrélés à des réactions plus lentes aboutissant à des rendements faibles. Il en est de même pour les prolines C-terminales qui adoptent une conformation trans **10** où les deux carbonyles se trouvent « face à face » (Figure 17).

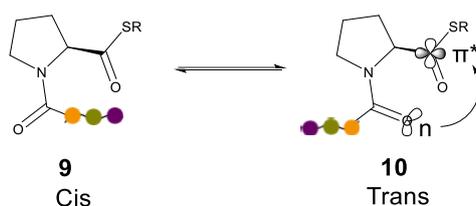


Figure 17: Représentation Cis-Trans de la proline et l'interaction  $n \rightarrow \pi^*$

Dans ce cas il existe une interaction  $n \rightarrow \pi^*$  provenant de la délocalisation d'une des paires d'électrons ( $n$ ) du carbonyle de l'amide dans l'orbitale antiliante ( $\pi^*$ ) du carbonyle du thioester, augmentant ainsi sa densité électronique et réduisant par conséquent son électrophilie.<sup>138</sup>

De plus, le type de thioester utilisé est lui aussi décisif dans cette réaction, les thioesters d'aryle étant plus réactifs du fait de la délocalisation électronique de la charge négative du thiolate sur le cycle aromatique, et doivent donc être privilégiés par rapport aux thioesters d'alkyle.<sup>139</sup>

### III.2.3 L'objectif de l'approche

Dans le cadre de notre projet, nous avons entrepris d'exploiter la NCL afin de pouvoir introduire notre arginine modifiée au sein de RSF1.

Ainsi, dans le but de générer une telle RSF1 artificielle, l'équipe de Jean Cavarelli a ainsi exprimé une RSF1 tronquée et amputée de sept résidus à son extrémité *N*-terminale et portant une cystéine *N*-terminale. L'heptapeptide amputé possède une arginine *C*-terminale (R7) qui pourrait être remplacée par notre arginine modifiée. Une fois synthétisée, notre arginine modifiée sera ainsi incorporée dans l'heptapeptide **11**, qui sera relié à la RSF1 artificielle via une NCL, afin d'obtenir la RSF1 modifiée attendue (Figure 18).

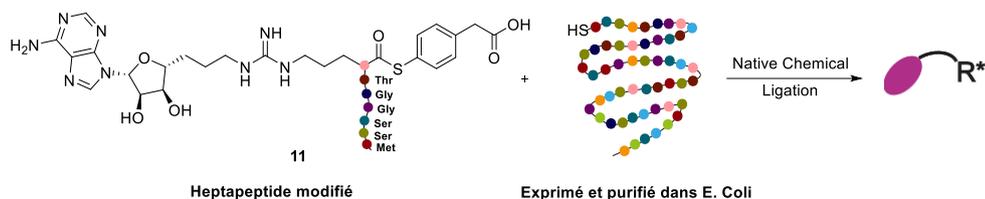


Figure 18: La ligation chimique native entre la RSF1 et l'heptapeptide portant l'arginine modifiée

Pour atteindre cet objectif, nous avons entrepris de synthétiser trois types d'arginine modifiée différents pour évaluer l'influence de l'étape de méthylation (en noir, Figure 19) sur la stabilité des complexes RSF1/PRMT2 (Figure 19):

- L'analogue d'arginine non méthylé (12a) mimant l'état de transition de monométhylation partagé par tous les PRMTs ( $\text{MMA-N}^\omega$ )
- Les analogues d'arginines méthylés (Figure 19):
  - Composé (12b) mimant l'état de transition des PRMTs de type II ( $\text{sDMA-N}^{\omega'}$ ,  $\text{N}^{\omega''}$ )
  - Composé (12c) mimant l'état de transition des PRMTs de type I ( $\text{aDMA-N}^{\omega'}$ ,  $\text{N}^{\omega''}$ )

Les fonctions *N*-terminales des arginines modifiées sont protégées sous forme de Fmoc pour faciliter la synthèse peptidique en phase solide (SPPS).

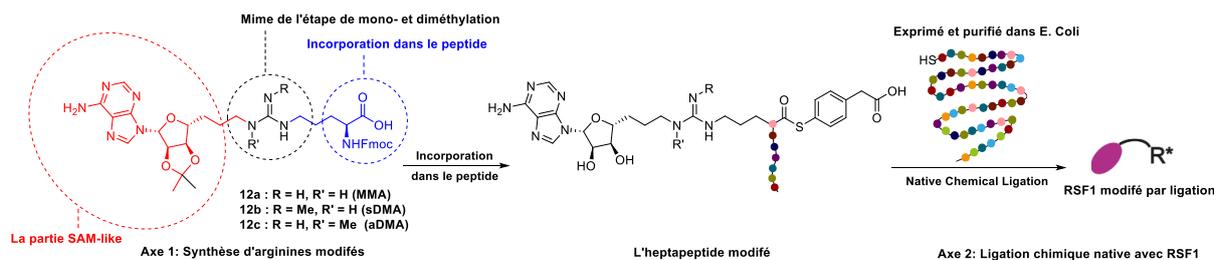


Figure 19: Les deux axes de la première approche

Ainsi, une partie de ma thèse a été consacrée à la synthèse de ces arginines modifiées **12a-c** (Axe 1 du projet, Figure 19), et à la fixation de l'heptapeptide modifié à l'aide de la NCL (Axe 2 du projet, Figure 19). La voie de synthèse déjà décrite dans la littérature pour le dérivé **2** n'étant pas adaptée pour accéder aux trois différentes arginines modifiées, nous avons dû mettre en place une nouvelle voie de synthèse.

### III.2.4 Résultats et discussions :

L'analyse rétrosynthétique de la synthèse des trois arginines modifiées est présentée au schéma 1. L'étape clé est le couplage entre les amines **13a** et **13b** et les thiourées **14a-b** dont

leur synthèse est prévue à partir de 2',3'-O-isopropylidèneadénosine **15** et Fmoc-Ornithine(Boc)-OH **16**, respectivement (Schéma 2).

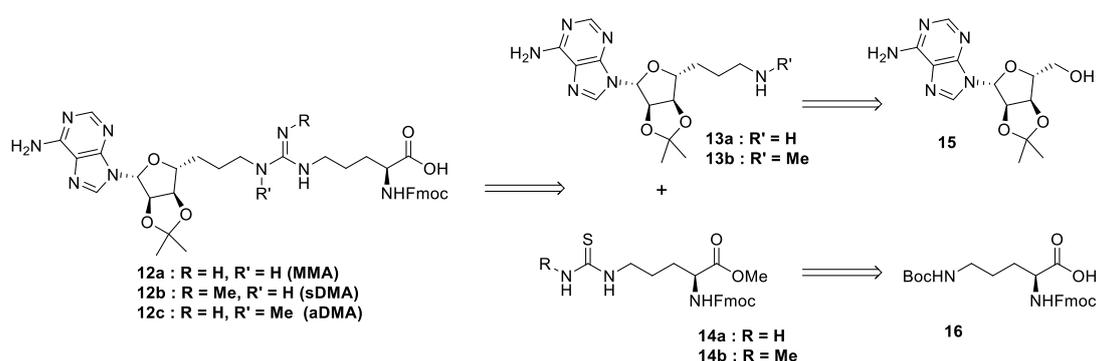


Schéma 2: Rétrosynthèse initiale des arginines modifiées

La préparation des amines primaire **13a** et secondaire **13b** commence à partir de l'alcool commercial 2',3'-O-isopropylidèneadénosine **15**. L'ester éthylique insaturé **18** a été préparé grâce à une réaction domino consistant en une oxydation de l'alcool primaire avec de l'acide 2-iodoxybenzoïque (IBX) suivi d'une réaction de Wittig avec un ylure de phosphonium commercial, délivrant l'ester avec un rendement de 77%. Ensuite, l'hydrogénation catalytique de la double liaison suivie de la réduction de l'ester avec du Dibal-H nous a permis d'isoler l'alcool correspondant **19** qui a été activé sous forme de nucléoside tosylé **20** en présence de pyridine et de chlorure de *p*-toluènesulfonyl (Schéma 3).

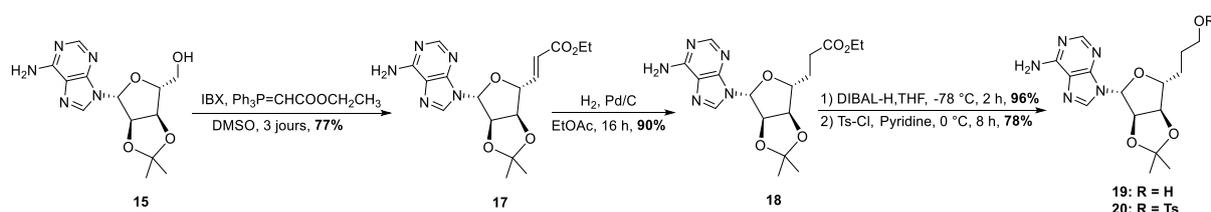


Schéma 3: Synthèse du nucléoside tosylé permettant l'accès aux deux amines

Ce composé tosylé a donné accès aux dérivés aminés. D'une part, une réaction de substitution du tosyloxy par une fonction azoture, suivie d'une hydrogénation catalytique a permis d'obtenir l'amine primaire **13a**. D'autre part, le tosyloxy a été substitué par de la méthylamine permettant de délivrer l'amine secondaire **13b**, le faible rendement s'expliquant par la faible nucléophilie de la méthylamine (Schéma 4).

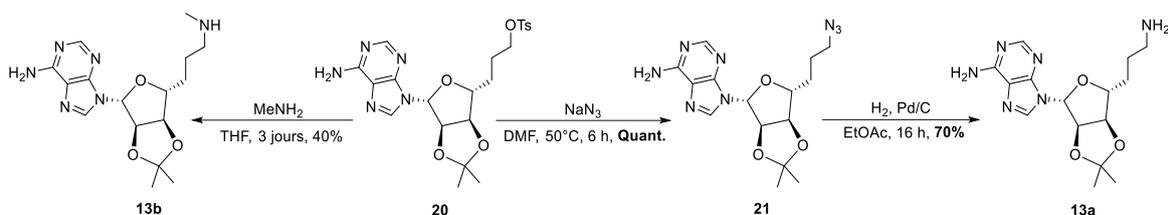


Schéma 4: Synthèse de l'amine primaire et secondaire

Dans un premier temps, la thiourée **14b** a été synthétisée en trois étapes à partir du Fmoc-Ornithine(Boc)-OH **16**, avec un rendement global de 78% (Schéma 5). La fonction acide carboxylique a été protégée sous forme d'ester méthylique en présence de triméthylsilyldiazométhane puis le groupement Boc porté par l'amine de la chaîne latérale a été déprotégé dans des conditions douces, en présence de trifluorométhanesulfonate de triméthylsilyle. Une fois l'amine déprotégée, elle a été mise en présence de 1,1'-thiocarbonyldiimidazole (TCDI) puis traitée avec une solution de méthylamine afin d'obtenir la thiourée **14b**.

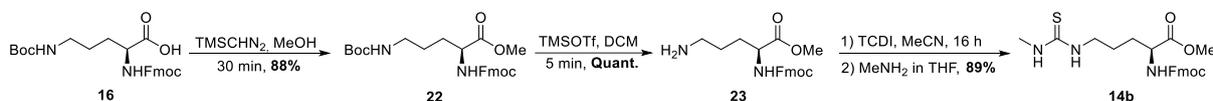


Schéma 5: Synthèse de thiourée Fmoc protégée

Ayant cette thiourée en main, nous nous sommes alors tournés vers la réaction de couplage générant les guanidines cibles. Nous avons tenté cette réaction de guanylation avec deux réactifs très utilisés dans la synthèse de guanidines : le chlorure de mercure ( $\text{HgCl}_2$ ) ou le réactif de Mukaiyama (Schéma 6). Cependant, dans ces conditions, la réaction de guanylation a échoué, causant également une rapide déprotection du groupement Fmoc ( $t_{1/2} = \sim 10$  min) par l'amine **13a** (Schéma 6).

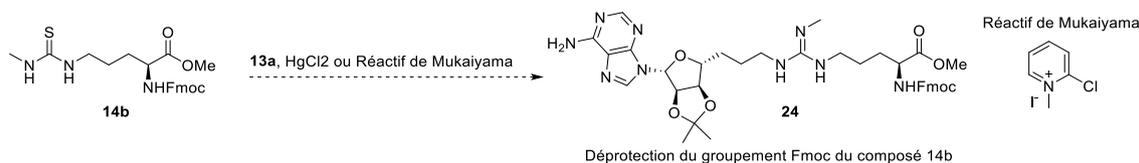


Schéma 6: Essais de guanylation avec la thiourée Fmoc protégée

Ainsi, notre stratégie alternative envisagée (Schéma 7) a été de protéger la chaîne principale de cette ornithine sous forme d'ester et carbamate benzyle. Ainsi, après l'étape de couplage avec les amines **13a** et **13b**, ces deux groupements pourront être déprotégés dans une réaction d'hydrogénéolyse. Par la suite, ces arginines modifiées pourront être protégées sous forme de Fmoc permettant l'accès aux composés **12a-c** désirés.

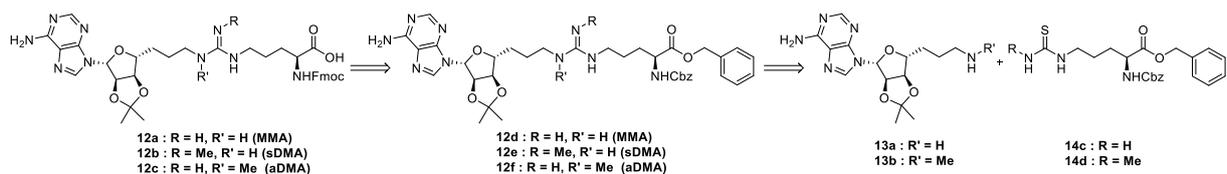


Schéma 7: Rétrosynthèse révisée des arginines modifiées

La synthèse de cette deuxième série de thiourées **14c-d** commence à partir du Cbz-Ornithine-OH commercial **25**. La réaction entre le TCDI et l'amine primaire de l'ornithine suivie d'un traitement avec de la méthylamine (ou de l'ammoniac) génère les thiourées asymétriques **26a** et **26b**. L'acide carboxylique a été ensuite protégé sous forme d'ester benzyle afin de délivrer les composés **14c** et **14d** avec un rendement global allant de 39 à 60% (Schéma 8).

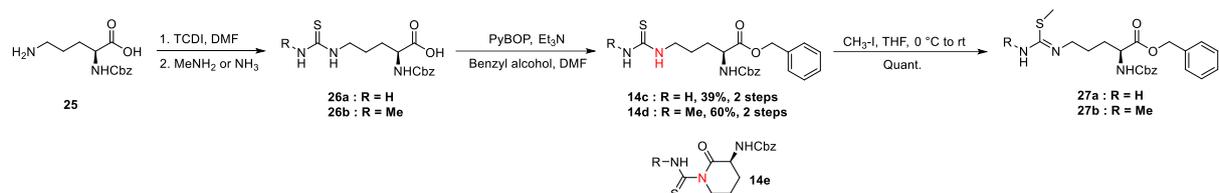


Schéma 8: Synthèse des isothiourées

Le rendement modeste pour la synthèse de la thiourée **14c** s'explique par deux phénomènes. Premièrement, pour la première étape de formation de thiourée **26a**, l'ammoniaque est moins nucléophile que la méthylamine, ce qui se reflète sur le rendement. Deuxièmement, pour l'étape de couplage, nous avons observé une réaction secondaire issue d'une cyclisation intramoléculaire qui survient via l'attaque d'un azote (en rouge, Schéma 8) de la thiourée lors de l'activation de la fonction acide carboxylique, donnant ainsi lieu au  $\delta$ -lactame **14e** et affectant là aussi le rendement.

Ces deux thiourées **14c** et **14d** ont été engagées dans des réactions de guanylation. Cependant, ces composés se révélant peu réactifs dans les conditions évaluées ( $\text{HgCl}_2$ , Réactif de Mukayama), nous avons dû procéder à une dernière étape de S-méthylation en présence d'iodométhane pour activer ces thiourées sous la forme d'isothiourées **27a** et **27b** (Schéma 8).

Dans un premier temps, afin de vérifier la réactivité de **27a** et **27b**, une réaction de guanylation a été réalisée en utilisant la benzylamine comme amine modèle. L'arginine correspondante **28** a été obtenue avec un rendement de 64% en présence de  $\text{HgCl}_2$  dans du diméthylformamide et de la triéthylamine. Une réaction d'hydrogénolyse nous a ensuite permis de libérer les fonctions acides carboxylique et amine, cette dernière ayant ensuite été protégée sous la forme d'un groupement Fmoc, par action du chlorure de fluorénylméthoxycarbonyle (Fmoc-Cl)

en présence de pyridine, nous permettant d'isoler le composé **30**. Bien que le rendement global de 30% n'ait pas été optimisé, cette séquence conduite sur un composé modèle nous a permis de valider notre approche synthétique (Schéma 9).

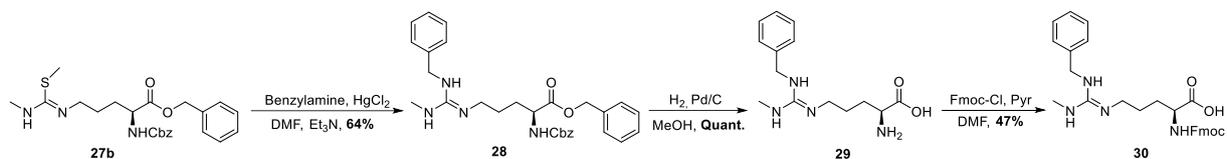


Schéma 9: Synthèse de l'arginine et protection sous forme de Fmoc

Après avoir établi la preuve de concept, nous nous sommes alors tournés vers la synthèse des arginines modifiées cibles **12d-f** (Schéma 10). L'addition de l'amine primaire **13a** ou secondaire **13b** sur les isothiourées **27a** ou **27b** a été réalisée dans les mêmes conditions que précédemment décrites, en présence de chlorure de mercure dans du DMF et de la triéthylamine, mais avec des rendements beaucoup plus faibles cependant. Le faible rendement peut s'expliquer d'une part par l'encombrement stérique de notre amine qui est plus important que celui de la benzylamine. D'autre part, par l'absence de groupements attracteurs (i.e., Cbz, Boc, etc.) sur le fragment isothiourée qui sont habituellement utilisés dans ce type de réaction comme activateurs supplémentaires.<sup>140</sup>

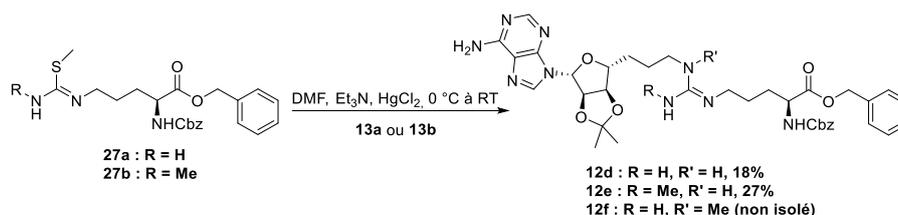


Schéma 10: Hydrogénolyse des arginines modifiés

Afin d'incorporer ces composés dans des séquences peptidiques, nous avons ensuite entrepris la déprotection des composés **12d-f**, pour les protéger sous forme de Fmoc **12a-c** (Schéma 7). Dans un premier temps nous avons procédé à l'hydrogénolyse de l'arginine modifiée **12d** en utilisant les conditions mises au point précédemment (i.e. atmosphère d'hydrogène, 0.1 equiv., Pd/C 10 wt.% (Sigma), MeOH ; Schéma 11). Cependant, après plusieurs essais, aucune déprotection n'a été constatée. Ceci est d'autant plus surprenant que, dans des conditions exactement similaires, l'arginine modifiée **12e** a conduit au produit d'hydrogénolyse **32** avec un rendement quantitatif (Schéma 11).

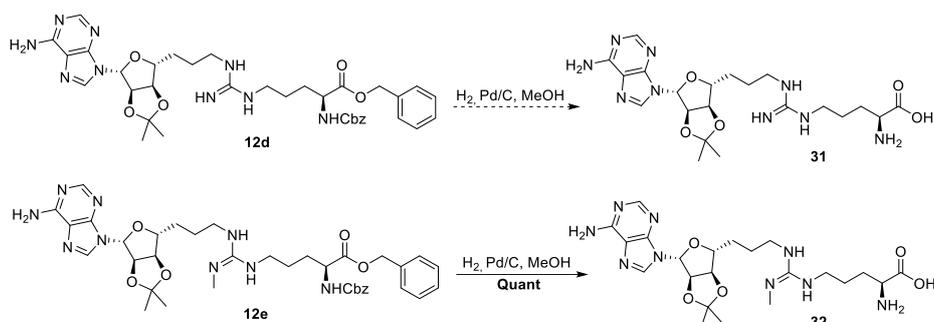


Schéma 11: Déprotection de la chaîne principale des arginines modifiées

Afin de mieux comprendre l'origine de cette différence de réactivité, nous avons entrepris de jouer sur la source de palladium, le type de solvant et sur l'acidité du milieu réactionnel. Dans un premier temps, nous avons remplacé le Pd/C 10 wt.% (Sigma) par son équivalent chez Alfa Aesar, mais aucune déprotection n'a été observée. Nous avons alors décidé d'utiliser le catalyseur de Pearlman ( $\text{Pd}(\text{OH})_2$  20 wt.%), qui est décrit dans la littérature comme efficace pour les réactions d'hydrogénolyse et est utilisé dans la déprotection des groupements Cbz.<sup>141</sup> Cependant, nous avons là encore constaté une absence totale de réaction malgré plusieurs essais (Schéma 12). Conscients que des catalyseurs à base de palladium pourraient ne pas être le choix le plus opportun, nous avons également évalué le Pt/C 10 wt.%, qui s'est montré tout aussi inefficace. Suspectant un problème de solubilité, nous avons remplacé le méthanol par de l'éthanol absolu, tout en gardant le Pd/C de chez Sigma comme source de palladium, ce dernier ayant montré son efficacité lors de la déprotection du composé **12e**.

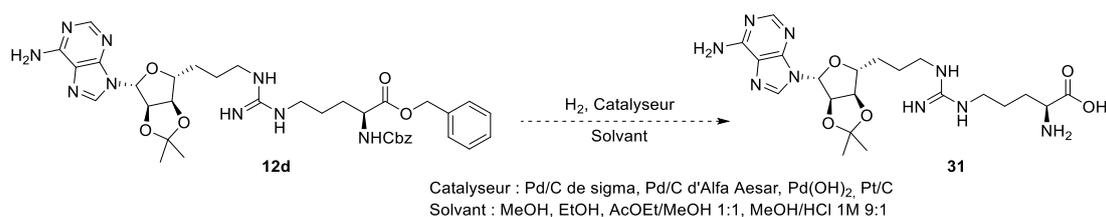


Schéma 12: Screening de conditions pour la déprotection de la chaîne principale de l'arginine modifiée non méthyliée.

A nouveau, aucune conversion n'a pu être détectée et, plus inquiétant, nous avons systématiquement remarqué tout au long de ces essais que nous perdions en outre des petites quantités de notre réactif de départ. Par conséquent, nous nous sommes demandé si la présence de la guanidine non substituée présente au sein de notre molécule **12d** pouvait chélater et empoisonner notre catalyseur. De ce fait, nous avons utilisé une solution diluée d'HCl afin de protoner la guanidine et empêcher cette possible chélation du palladium.

Evaluant des conditions d'hydrogénolyse dans une solution de MeOH/HCl 1.0 M, 9:1 (v/v), sous atmosphère d'hydrogène et en présence de Pd/C (Sigma), **12d** s'est révélé parfaitement inerte, aucune hydrolyse de la fonction acétal n'étant même constatée.

Remplacer les solvants protiques évalués jusqu'ici par de l'acétate d'éthyle s'est révélé inefficace, notre composé **12d** étant insoluble dans ce solvant. Cependant, l'utilisation d'un mélange MeOH/AcOEt, 1:1 (v/v) nous a permis d'observer une déprotection partielle, cependant avec une conversion variable oscillant entre 15 et 40%, et limitée à l'ester benzylique uniquement (Schéma 12). Actuellement, la déprotection de l'arginine modifiée **12d** est toujours en cours d'optimisation afin de permettre la déprotection totale de la molécule et de pouvoir accéder à un acide aminé prêt à être couplé. Pour cela plusieurs possibilités s'offrent à nous comme l'utilisation d'autres catalyseurs (Ni de Raney, Rh), l'emploi d'autres solvants ou de conditions réactionnelles plus drastiques, par exemple sous pression élevée d'hydrogène.

Afin de valoriser ces arginines modifiées et d'évaluer leur potentiel en tant que ligands de PRMT, des études de co-cristallisation ont été menées par le Dr Vincent Cura au sein de l'IGBMC. Dans cette optique, nous avons réalisé la déprotection complète des groupements protecteurs de l'arginine **12e** (i.e., Cbz, ester benzylique et acétal) en deux étapes mais aucune structure cristallographique n'a pu être obtenue entre le produit de déprotection **33** et les PRMT, probablement du fait d'une trop grande polarité de ces composés (Schéma 13).

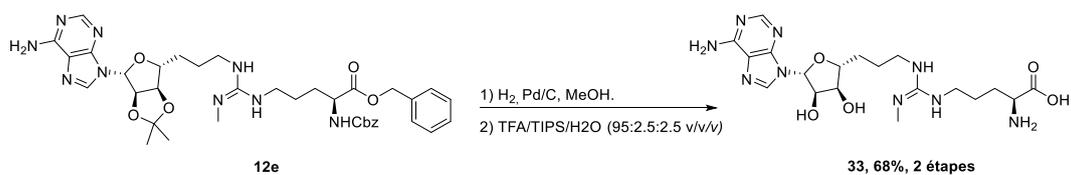


Schéma 13 : Déprotection complète de l'arginine modifié mime des PRMT de type sDMA

C'est pourquoi nous avons alors opté pour une stratégie de déprotection partielle du groupement acétone uniquement, réalisée grâce à une solution de TFA/TIPS/H<sub>2</sub>O (95:2.5:2.5 v/v/v), permettant l'isolement des trois arginines modifiées, **34a**, **34b**, et **34c** (Schéma 14). Ces composés ont été co-cristallisés dans un premier temps avec PRMT4.

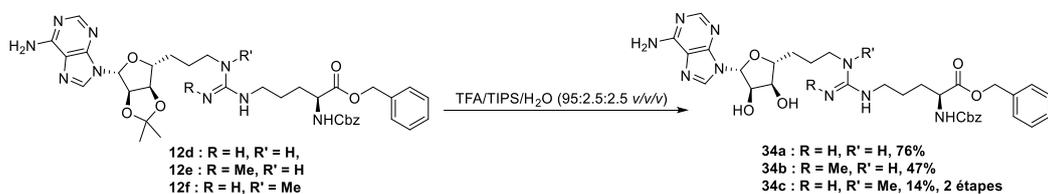


Schéma 14: Déprotection de l'acétal des trois arginines

Les trois structures cristallographiques obtenues avec PRMT4 révèlent que ces composés occupent partiellement le site du SAM et que la guanidine se place bien dans la poche arginine. On observe des différences d'affinités qui sont uniquement liées au groupement méthyle qui est plus ou moins bien accommodé dans la poche arginine. Il apparaît que le composé le moins affiné (c'est-à-dire où le méthyle n'est pas bien accommodé dans la poche arginine) est l'arginine modifiée **34b** (Figure 22) qui mime l'état de transition de la diméthylation symétrique, qui ne se fait pas dans PRMT4. Les structures cristallographiques obtenues avec l'arginine non méthylée **34a** (Figure 20) mimant la première étape de méthylation et l'arginine modifiée **34c** (Figure 21) mimant la diméthylation asymétrique s'accrochent plutôt bien dans la poche arginine.

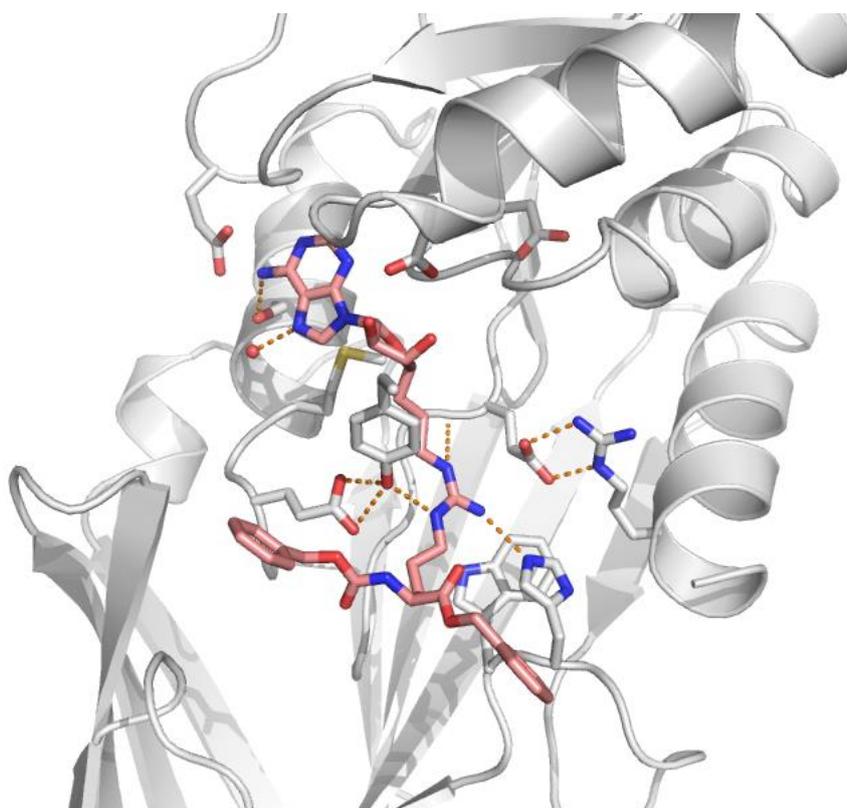


Figure 20: Complexe **34a**-PRMT4

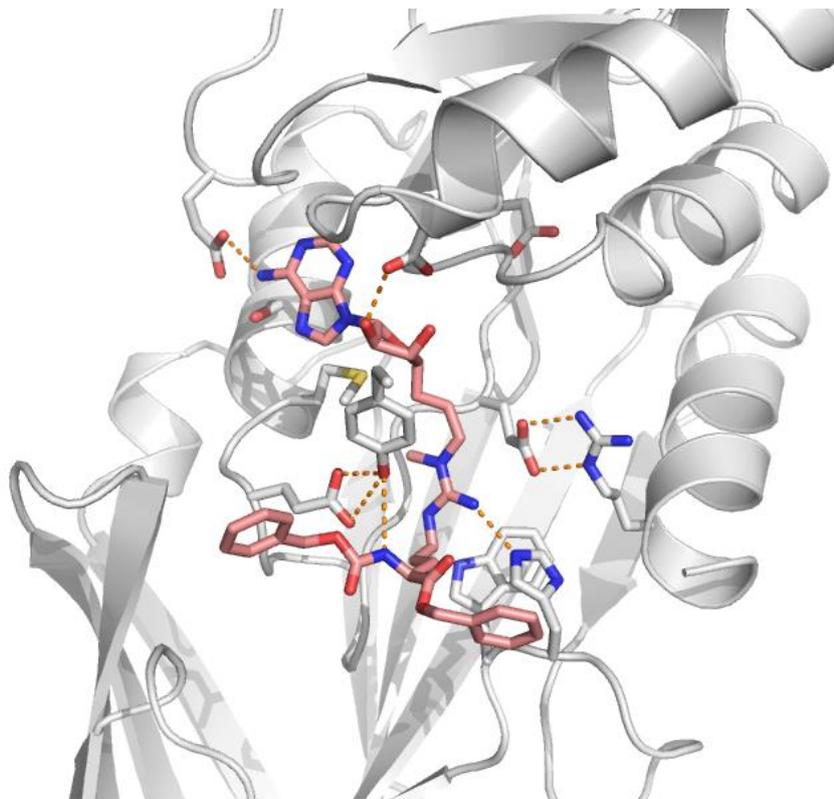


Figure 21: Arginine modifié **34c** mimant l'étape de diméthylation asymétrique complexé avec PRMT4

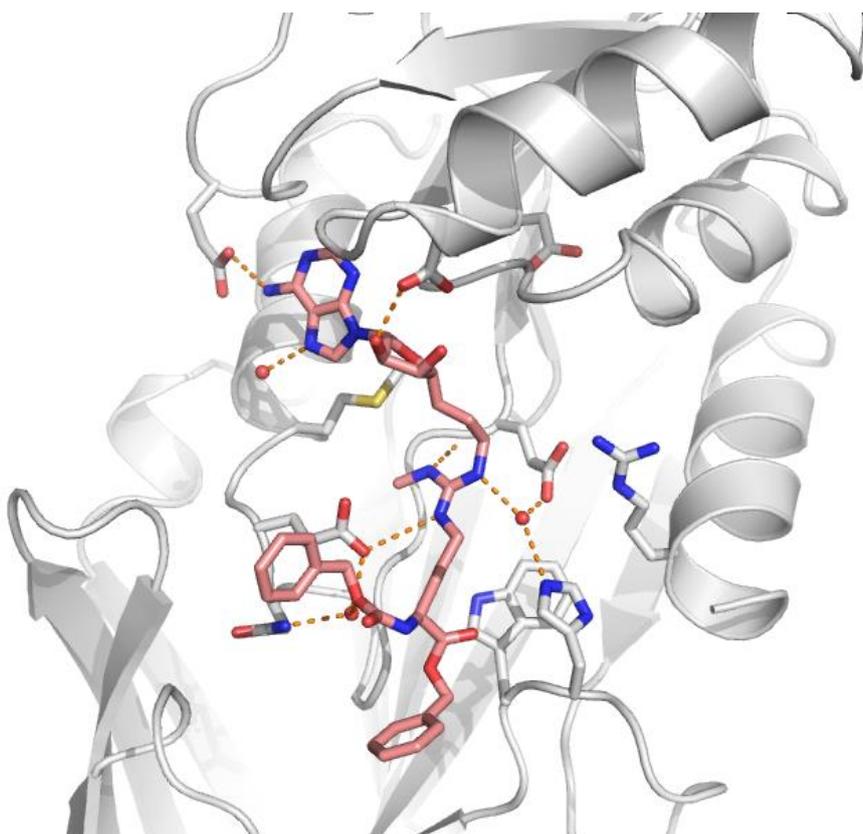


Figure 22: Arginine modifié **34b** mimant l'étape de diméthylation symétrique complexé avec PRMT4

Ces arginines modifiées occupent les quatre modules catalytiques et comme le taux d'occupation est de manière générale corrélé à l'IC<sub>50</sub>, ces dernières ont été testées pour leur pouvoir inhibiteur. Le SAH, l'inhibiteur connu des PRMT a été également inclus comme référence. Les tests d'inhibitions montrent que l'arginine modifiée non méthylée **34a** inhibe l'activité enzymatique de PRMT4 avec une IC<sub>50</sub> de 330 nM, étant dix fois plus actif que le SAH. L'inhibition chute drastiquement (180 fois par rapport à **34a**) pour le composé **34b** ayant une efficacité d'inhibition de 60 µM. Le composé **34c** en termes d'efficacité se place entre les deux avec une IC<sub>50</sub> de 790 nM (Figure 23). Ces résultats doivent cependant être pris avec précaution et restent à confirmer en triplicat.

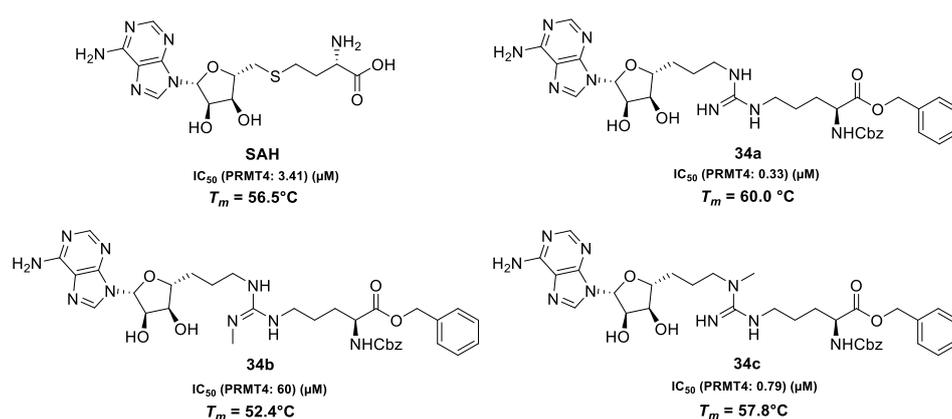


Figure 23: Evaluation des arginines modifiées contre PRMT4

Il faut savoir qu'il a été démontré que l'affinité de la protéine pour un ligand est corrélée à la stabilisation de la protéine par ce ligand.<sup>142</sup> Ainsi, afin de mesurer la fixation directe de ces composés à PRMT4 des mesures de stabilité thermique ont été réalisées. Celle-ci est défini par la valeur de  $T_m$  (température de fusion) qui est la température à laquelle la moitié de la protéine est dénaturée.

Les variations des  $T_m$  comparées à la forme apo (enzyme seule,  $T_m = 50.7$  °C) ont démontré une stabilisation comparable de PRMT4 en présence des trois arginines modifiées (mesurée en triplicat) et les valeurs de  $T_m$  sont en corrélation avec les valeurs d'IC<sub>50</sub> (Figure). L'inhibiteur le plus efficace (composé **34a**) a un fort effet stabilisateur pour PRMT4 comparé à la forme apo avec un  $T_m$  de 60.0 °C ( $\Delta T_m = +9.3$  °C vs apo), tandis qu'une faible stabilisation est observée avec l'inhibiteur le moins efficace (composé **34b**,  $\Delta T_m = +1.7$  °C vs apo) (Figure 23). Ainsi, les structures cristallographiques, les tests d'inhibition et de stabilité thermique sont en parfaites corrélation entre elles. Les structures cristallographiques obtenu dans ce contexte sont essentielles et permettent de dévoiler les bases structurales des différences d'affinités observés avec les arginines modifiées.

Par ailleurs, des essais de co-cristallisation ont été réalisés avec PRMT2 qui a permis de résoudre deux structures cristallographiques entre ce dernier et l'arginine modifiée **34a** et l'arginine modifiée mimant l'étape de diméthylation asymétrique **34c**. Aucune structure cristallographique n'a été obtenue avec l'arginine modifiée mimant l'étape de diméthylation symétrique **34b** et PRMT2, qui ne catalyse que la réaction de diméthylation asymétrique. Ces résultats suggèrent que les arginines modifiés (du moins deux d'entre elles) peuvent être utilisés pour stabiliser le complexe PRMT2/RSF1.

En parallèle de la synthèse des arginines modifiées, nous avons aussi développé et optimisé les conditions de NCL qui seront essentielles à la formation ultérieure des RSF1 modifiées (Axe 2 du projet, Figure 19). Pour cela, nous avons utilisé un peptide modèle possédant une arginine naturelle en position C-terminale dont l'acide  $\alpha$ -carboxylique peut être converti en un thioester. Comme les peptides C-terminal-thioester sont difficiles à isoler, nous avons utilisé le dérivé hydrazide du peptide **35**, qui a été converti *in situ* en thioester **37** via la formation de pyrazole **36** suivi d'une thioestérification avec de l'acide 4-mercaptophénylacétique (MPAA).<sup>143</sup> L'ajout de RSF1 (tronquée et portant une cystéine en N-ter et une étiquette hexahistidine en C-ter) et la neutralisation du pH du milieu réactionnel ont conduit à la formation de RSF1 synthétique **38**, produit de ligature attendu, dans des conditions très douces. Une rapide optimisation a été nécessaire pour améliorer cette étape de couplage, en jouant sur la concentration des différents réactifs, conduisant à une conversion satisfaisante de 65% déterminée par LC-MS (Schéma 15). Il convient de signaler à ce stade que les 35% restants correspondent à de la RSF1 tronquée résiduelle, ce qui suggère qu'un accroissement de cette conversion serait encore possible.

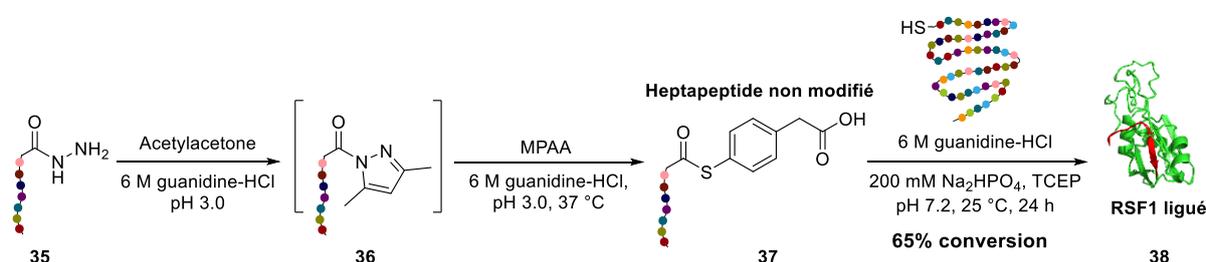


Schéma 15: La NCL entre RSF1-Cys et l'heptapeptide comportant une arginine naturelle

### III.2.5 Conclusion

En conclusion, nous avons constitué une petite librairie de ligands mimant l'état de transition de la première ou de la seconde étape de méthylation des différents types de PRMT. Les réactions de guanylation permettant leur obtention sont les étapes limitantes de notre voie de

synthèse. Ces dernières restent donc à optimiser afin d'accroître les rendements et monter en échelle en vue de permettre la synthèse des heptapeptides prévus. De plus, l'étape de déprotection de la molécule **12d** reste également à optimiser en vue de permettre la protection des arginines modifiés par des groupements Fmoc. De manière intéressante, la présence d'un méthyle sur la partie guanidine a un profond impact sur la réactivité de nos composés, que ce soit lors de l'étape de guanylation ou lors des étapes de déprotection, bien que les raisons puissent être différentes selon les cas.

La preuve de concept et l'utilisation de l'arginine modifiée en tant que mime de l'état de transition avait déjà été démontré dans le cadre de CARM1.<sup>126</sup> Dans notre cas nous avons obtenu également le complexe **34a**-PRMT2 qui permet de confirmer que la géométrie du mime de l'état de transition est aussi adaptée pour PRMT2 et qu'une fois qu'il sera incorporé au sein de RSF1, il stabilisera probablement avec une bonne aisance le complexe RSF1-PRMT2.

Ces dernières ont inhibé l'activité enzymatique de PRMT4 avec une efficacité allant du micromolaire au sub-micromolaire. Les tests d'inhibition doivent être réalisés en triplicat et étendu aux autres PRMT pour évaluer leurs sélectivités parmi les PRMT.

L'optimisation de l'étape de ligation chimique native nous a permis d'atteindre une conversion satisfaisante de 65% entre la RSF1 tronquée, comportant une cystéine en *N*-ter (et une étiquette hexahistidine en *C*-ter qui servira pour la purification de la protéine) et notre peptide modèle contenant une arginine naturelle.

Les prochaines étapes clés de ce projet sont la synthèse et l'incorporation de l'arginine modifiée **34a** dans l'heptapeptide.

## III.3 Deuxième approche : Incorporation de l'arginine modifiée (R\*\*) sur la partie structurée de RSF1 (R24)

### III.3.1 Introduction

La deuxième approche que nous avons entreprise de mettre en place dans le cadre de la stabilisation du complexe PRMT2-RSF1, est l'incorporation de l'analogue du SAM R\*\* sur la position 24 de RSF1. Contrairement à l'approche précédente, la structure et la géométrie de cet analogue du SAM ne sont pas basées sur une étude antérieure mais sur une approche *in silico*. Ainsi, si la géométrie de R\*\* permet théoriquement de rentrer dans la poche de fixation de l'arginine, il reste à le confirmer de manière expérimentale, par une étude de co-cristallisation afin de valider le potentiel de R\*\* en tant que ligand de PRMT. Dans cette deuxième approche, avons pris l'initiative d'exploiter une réaction de chimie click catalysée au cuivre(I) entre un azoture et un alcyne terminal – la CuAAC (pour *Copper-catalyzed azide-alkyne cycloaddition*) – afin d'introduire l'arginine modifiée R\*\* au sein de RSF1. Pour cela nous devons tout d'abord incorporer au sein de RSF1 un acide aminé non naturel, comportant sur sa chaîne latérale une fonction capable de réagir dans une réaction de CuAAC. Nous avons décidé d'opter pour une méthode d'incorporation site-spécifique, exploitant la machinerie cellulaire, afin d'incorporer cet acide aminé non naturel à la place de l'arginine naturelle R24 au sein de RSF1. Ainsi, après avoir introduit le terme de chimie click et la CuAAC nous allons voir les bases de la synthèse protéique pour mieux comprendre les méthodes permettant l'incorporation d'acides aminées non naturels.

### III.3.2 La chimie click

Le concept de chimie click a été décrit pour la première fois en 2001 par K. B. Sharpless et ses collègues qui le définissent comme une famille de réactions « idéales » qui opéreraient dans des conditions aqueuses, à l'air libre, avec une cinétique rapide et une très bonne chimiosélectivité, conduisant à des hauts rendements, et ne produisant pas de sous-produits.<sup>144,145,146</sup>

Les réactions de chimie click ont été utilisées pour un large éventail d'applications, notamment dans le domaine de la chimie, de la biologie, des matériaux, des polymères et de la biotechnologie.<sup>147</sup> Parmi les réactions de chimie click les plus notables, la CuAAC est l'une des plus utilisées car elle s'est avérée très efficace, robuste et fiable, notamment pour la modification des protéines.<sup>148</sup>

### III.3.3 La CuAAC

Les cycloadditions 1,3-dipolaires de Huisgen sont des processus d'association exothermiques qui réunissent deux réactifs et permettent un accès rapide à une énorme variété d'hétérocycles à cinq chaînons. La cycloaddition entre azotures **39** et alcynes **40** pour donner des triazoles s'est avérée être la réaction la plus efficace de cette famille. Son application à la chimie click a nécessité le développement de conditions réactionnelles douces. En effet, sous seul contrôle thermique, cette réaction conduit à un mélange de deux régioisomères de triazoles : le 1,4-disubstitué **41** et le 1,5-disubstitué **42** (Figure 24).

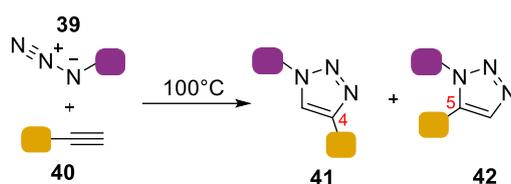
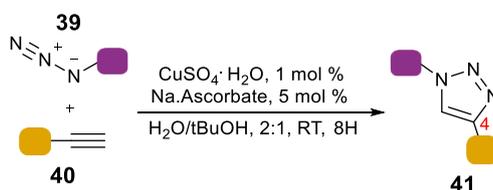


Figure 24: Cycloaddition 1,3-dipolaire de Huisgen entre un azide et un alcyne

En 2002, les travaux menés en parallèle par les groupes de Meldal et Sharpless, ont permis de s'affranchir à la fois des problèmes de régiosélectivité mais aussi des températures élevées nécessaires au bon déroulement de la réaction, en catalysant la réaction au cuivre (I).<sup>149,150</sup> Plusieurs sources de cuivre (I) peuvent être utilisées, mais l'équipe de Sharpless a constaté qu'une réduction *in situ* par l'ascorbate de sodium des sels de cuivre (II) génère un catalyseur performant et peu coûteux.<sup>150</sup> Avec cette méthode, on obtient ainsi une large variété de triazoles-1,4-disubstitués, avec de très hauts rendements et dans des conditions douces, compatibles avec l'utilisation de biomolécules (Figure 25).



Pour la modification de protéines : CuSO<sub>4</sub>·H<sub>2</sub>O, THPTA, Na.Ascorbate Aminoguanidine, DMSO/H<sub>2</sub>O (10% DMSO max)

Figure 25 : La cycloaddition de Huisgen catalysée au cuivre (I)

La réaction semble être très tolérante et ne nécessite aucune précaution particulière. Elle est généralement complète en 6 à 36 heures à température ambiante et dans un mélange d'eau et de *tert*-butanol, conditions décrites comme universelles pour ce type de réaction. Cependant, divers co-solvants peuvent être utilisés comme l'éthanol, le DMSO, ou le THF,

selon la nécessité et la solubilité des réactifs. Il est recommandé de réaliser la catalyse à pH neutre mais elle tolère un intervalle de pH allant de 4.0 à 12.0, rendant donc ce procédé catalytique extrêmement robuste.<sup>147,151</sup> La cinétique de cette réaction peut être accélérée par l'utilisation de ligand à base de triazole, tels que THPTA (*tris*(3-hydroxypropyltriazolylmethyl)amine), ou TBTA (*tris*[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine).

Pour la modification de protéines, l'utilisation du DMSO comme co-solvant (~10% max) semble privilégiée, de même que les ligands à base de triazole (Figure 25).<sup>148</sup> Dans ce cas, ces ligands détiennent un double rôle : en plus d'accélérer la cinétique de réaction en coordonnant et maintenant l'état d'oxydation du cuivre (I), ils protègent les chaînes latérales des acides aminés constitutifs des protéines (telles que les histidines) contre les espèces réactives de l'oxygène en venant intercepter les peroxydes produits par le mélange O<sub>2</sub>/Cu/ascorbate.<sup>152</sup> L'utilisation d'aminoguanidine est elle aussi recommandée pour intercepter les sous-produits délétères de l'ascorbate (déhydroascorbate), qui viendraient à priori oxyder les chaînes latérales des arginines (principalement).<sup>152</sup>

La CuAAC se prête donc extrêmement bien à la modification de protéines, sous réserve qu'elles contiennent au moins un groupement azoture ou alcyne. Ces fonctionnalités étant absentes au sein des biomolécules, il a fallu développer des méthodes permettant l'incorporation de tels acides aminés non naturels au sein des protéines.

### III.3.4 Incorporation d'acides aminés non naturels au sein de protéines

Toute une gamme de méthodes chimiques (SPPS, NCL) ont été utilisées pour la synthèse de protéines modifiées portant des groupements fonctionnels exogènes sur leur chaîne latérale. La principale limitation de ces méthodes concerne leur montée en échelle car ce sont des méthodes coûteuses et peu efficaces. La SPPS ne permet pas techniquement la production de protéines de grosse taille et la synthèse devient très difficile pour les protéines présentant plus de 50 acides aminés, malgré l'utilisation de synthétiseurs automatisés.<sup>153</sup> Ceci est principalement dû à des problèmes de solubilité et de repliement de la chaîne en cours d'élongation, rendant la partie *N*-terminale moins accessible pour les étapes de déprotection et couplage.<sup>153</sup>

La NCL ne souffre pas de ces limitations et c'est donc la méthode de choix pour la synthèse de protéines supérieures à 50 acides aminés. Cependant, en plus des limitations présentées au chapitre 2 (Section II.2.2. Ligation Chimique Native), la NCL souffre du problème d'insolubilité des peptides lipophiles dans les tampons utilisés.<sup>153</sup>

Pour surmonter cela, d'autres méthodes ont été développées pour compléter les méthodes chimiques, exploitant la machinerie cellulaire afin de permettre l'incorporation d'acides aminés non naturels dans les protéines produites en cellules vivantes. Plus précisément, elles sont basées sur la modification de l'acide aminé au sein d'un ARNt et l'utilisation d'un codon unique, pour qu'au cours de la traduction, le ribosome réalise son incorporation au sein de la protéine en cours de synthèse.

Dans ce contexte deux méthodes ont été développées :

- La première méthode permet l'incorporation d'acides aminés non naturels de manière spécifique à un résidu.
- La deuxième méthode permet l'incorporation d'acides aminés non naturels de manière spécifique à un site.

Nous ne détaillerons que la méthode que nous avons retenu pour notre synthèse, celle permettant une approche site-spécifique.

#### **III.3.4.1 Incorporation d'acides aminés non naturels par approche site-spécifique :**

L'incorporation d'acides aminés de manière site-spécifique est une méthode puissante et élégante qui permet de remplacer un acide aminé naturel par un acide aminé non naturel. Cette approche nécessite une paire aaRS/ARNt pour l'insertion des résidus non naturels dans la chaîne polypeptidique de la protéine en cours de synthèse. Cette combinaison unique aaRS/ARNt doit être orthogonale au système de traduction endogène pour que l'ARNt qui contient l'anticodon vierge ne soit pas un substrat pour les aaRS endogène et que l'aaRS orthogonale ne reconnaisse ni les ARNt endogènes, ni les acides aminés naturels (Figure 26).<sup>154,155,156</sup>

C'est cette aminoacyl-ARNt orthogonale qui sera utilisée par le ribosome pour synthétiser la protéine modifiée. Pour cela, il faut que l'aminoacyl-ARNt soit reconnu par un codon au sein de l'ARNm grâce à un anticodon unique, qui n'est pas utilisé pour l'incorporation d'un autre acide aminé et écarter toute incorporation à une position non désirée (Figure 26). Pour cela, l'ARNt doit avoir un anticodon vierge, qui va être modifié afin de coder spécifiquement pour l'incorporation de l'acide aminé non naturel, en général l'un des trois codons « stop » qui est généralement le codon ambre (UAG). Ce codon ambre est généré au sein de la séquence de l'ARNm par mutagenèse dirigée qui a pour but de muter une séquence d'ADN à un endroit précis. Ainsi, la séquence de l'ARNm correspondante sera mutée à la position voulue et permettra la traduction de la protéine modifiée (Figure 26).

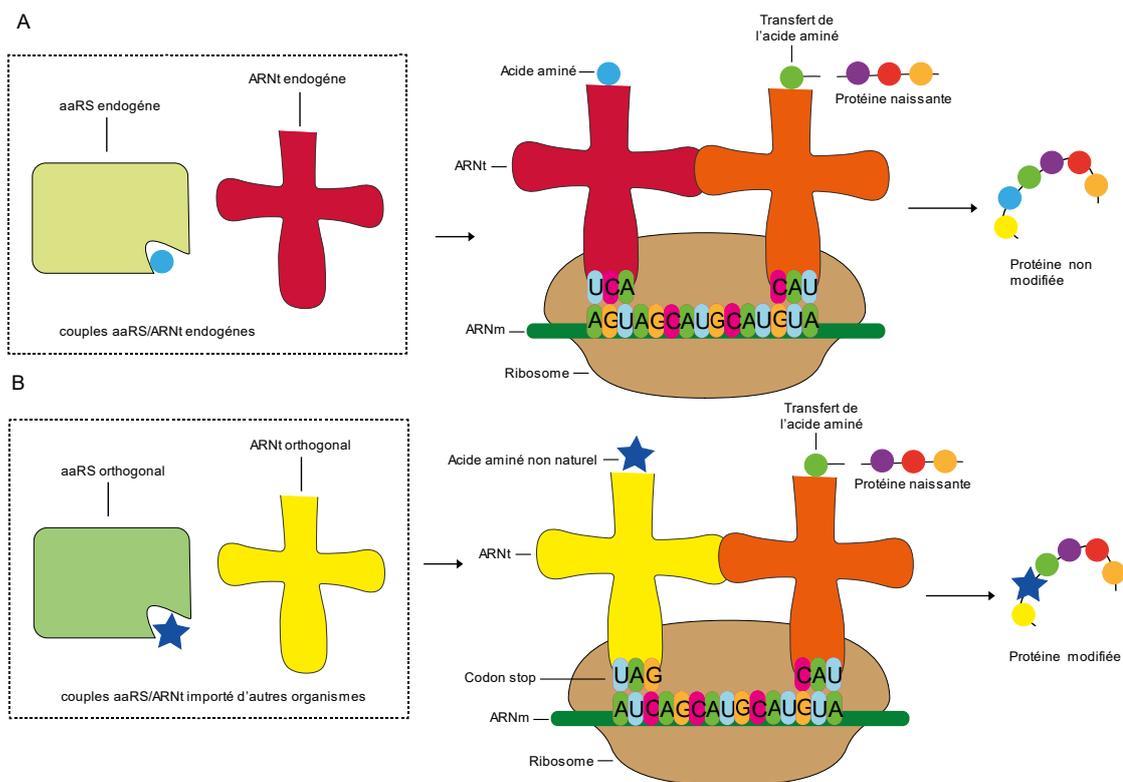


Figure 26: Synthèse de protéines modifié grâce à l'expansion du code génétique A: Synthèse protéique endogène, B: Détournement du codon UAG pour l'incorporation de l'acide aminée non naturel

La production des protéines est de manière générale réalisée dans *Escherichia coli* (*E. Coli*) et les couples aaRS/ARNT orthogonaux sont importés d'autres organismes car la spécificité moléculaire et structurale diffère entre les organismes. Ainsi, cette stratégie permet l'incorporation d'acides aminés non naturels à des positions stratégiques comme les sites de PTM. Ces sites peuvent ensuite être marqués chimio- et régio-sélectivement afin d'accéder à des conjugués de protéines porteurs de diverses fonctions non-naturelles.

### III.3.5 Objectif de l'approche

La synthèse de ligands mimes de l'état de transition des PRMT constitue un défi de notre projet, autant sur le plan de la synthèse organique que sur le plan biologique. Les méthodes chimiques permettant de modifier orthogonalement les protéines étant limitées, aucune réaction de chimie click ne nous permet actuellement de modifier chimio- et régio-sélectivement une arginine bien précise au sein d'une chaîne protéique. L'incorporation au cœur de la séquence protéique des arginines modifiées présentées précédemment étant impossible (pour les raisons évoquées ci-dessus), nous avons donc dû nous tourner vers un analogue du SAM ( $R^{**}$ ) **4** (Figure 27), sélectionné *in silico*, et pouvant être généré sur une

protéine recombinante (dont la rétrosynthèse est présentée à la figure 27), exprimant un résidu de L-bishomopropargylglycine **44** en lieu et place de l'arginine R24 de RSF1.

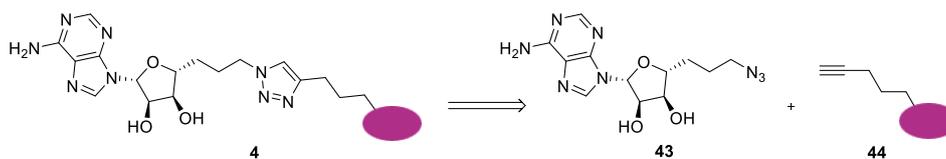


Figure 27 : Rétrosynthèse de RSF1 modifiée par **R\*\***

Ce résidu alcyne a été sélectionné sur la base de travaux publiés en 2021 et rapportant son incorporation au sein de protéines.<sup>157</sup> La synthèse de l'azoture **43** ayant été déjà décrite dans la première approche (Section II.2.4. Résultats et discussion) sous sa forme protégée **21**, une simple étape de déprotection devrait nous fournir le partenaire de CuAAC nécessaire à la génération du dérivé **R\*\*** sur RSF1 mutée donnant lieu à la protéine modifiée **4**.

### III.3.6 Résultats et discussion

Avant d'incorporer notre alcyne au sein de RSF1, nous avons tout d'abord voulu évaluer l'affinité de ce candidat mime de l'état de transition vis-à-vis des PRMT. Pour cela, nous avons entrepris la synthèse du mime de l'état de transition **47** pour pouvoir tenter de le co-cristalliser avec PRMT2. Après une courte optimisation, nous avons pu développer des conditions douces, permettant d'accéder au triazole **46** avec un rendement de 72% en utilisant un mélange DMSO/H<sub>2</sub>O 8:1, en présence de sulfate de cuivre (II), de THPTA et d'ascorbate de sodium (Schéma 16), suivie d'une étape de déprotection de la fonction acétone pour conduire à l'acide aminé non naturel **47**.

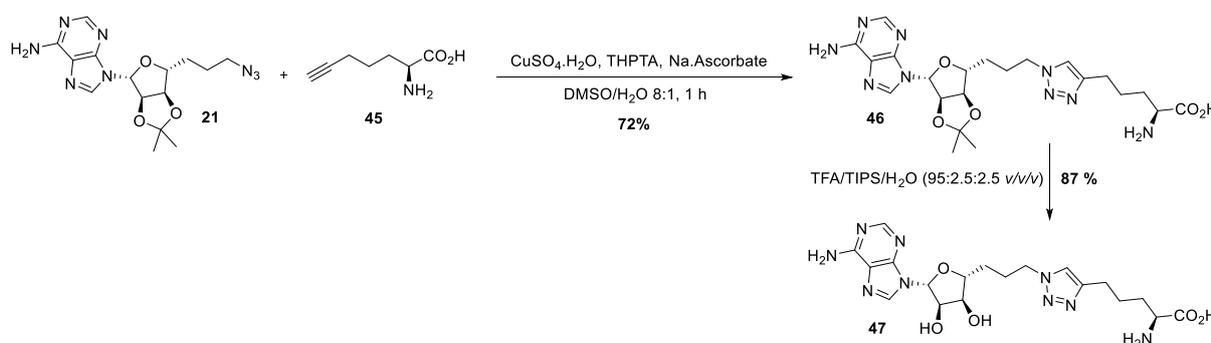


Schéma 16: Synthèse du candidat mime de l'état de transition

Cependant aucune co-cristallisation avec PRMT2 n'a été constatée. Comme dans le cas des arginines modifiées, nous suspectons que ce dernier ne cristallise pas du fait de sa grande polarité. Ainsi, pour atténuer sa polarité, nous avons entrepris de protéger la chaîne principale avec des groupements benzyle comme dans le cas des arginines modifiées, en partant de la Cbz-L-bishomopropargylglycine **48** commerciale (Schéma 17).

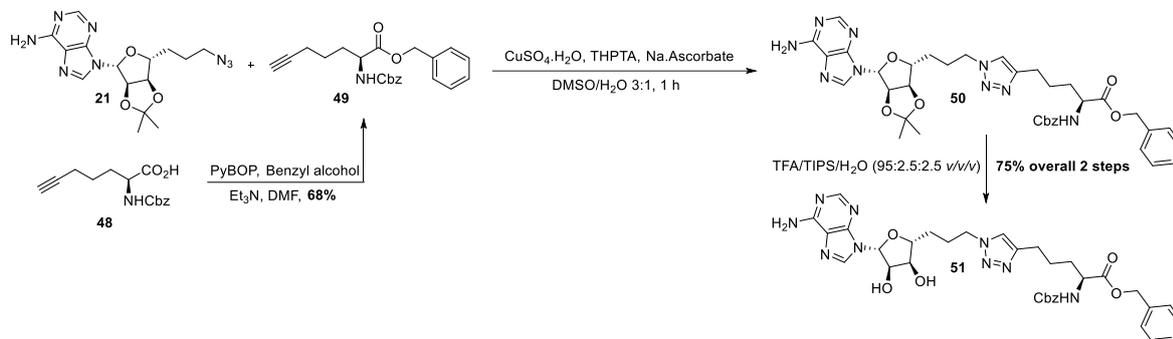


Schéma 17: Synthèse du candidat mime de l'état de transition protégé sur sa chaîne principale

La partie acide carboxylique a été protégée en présence de PyBOP, d'alcool benzyle, de triéthylamine et dans du DMF permettant l'isolement de l'alcyne **49**. Ce dernier a été engagé dans la réaction de CuAAC puis la fonction acétal a été déprotégée grâce à une solution de TFA/TIPS/H<sub>2</sub>O (95:2.5:2.5 v/v/v), permettant la synthèse du ligand désiré **51** (Schéma 17). Ce ligand a été délivré aux biologistes pour des essais de co-cristallisation pour déterminer si le triazole pourrait être un isostère de la guanidine et s'il s'accommoderait suffisamment bien dans la poche arginine pour être utilisé en tant que mime de l'état de transition.

### III.3.7 Conclusion

En conclusion nous avons synthétisé le candidat mime de l'état de transition **47**. Cependant, celui-ci n'a donné aucune structure cristallographique avec les PRMT. En nous basant sur notre expérience sur les arginines modifiées, nous avons décidé de synthétiser sa forme protégée qui a été délivrée aux biologistes pour des études cristallographiques. Ces études nous permettront de comprendre si cette absence de cristallisation est liée à la géométrie ou à la polarité de **R\*\***. Si les études de co-cristallisation permettent de valider que ce dernier est bien un mime de l'état de transition, l'alcyne sera incorporé au sein de RSF1 par nos collaborateurs pour donner lieu à la protéine modifiée **44** qui pourra réagir dans une réaction de CuAAC nous permettant d'insérer **R\*\*** sur RSF1.

## IV. Conception et synthèse d'inhibiteurs de PRMT

### IV.1 Introduction

De nombreux travaux ont été réalisés au niveau international pour découvrir des anticancéreux sélectifs de PRMT et ouvrir de nouvelles voies thérapeutiques. Un des obstacles majeurs au développement de nouveaux inhibiteurs notamment pour les composés dérivés du SAM découlent de leur difficulté à traverser les membranes cellulaires pour opérer sur leurs cibles. Récemment, des études menés en parallèle ont révélé de considérables progrès dans l'identification d'inhibiteurs cliniquement pertinents et les essais cliniques mettent en évidence que le ciblage des PRMT pourrait être un véritable succès.<sup>158,159</sup>

Nos travaux de recherche d'inhibiteurs de PRMT ont débuté en 2009 lors d'une collaboration avec l'équipe du Prof. Jean Cavarelli. L'idée initiale était de trouver des mimes de l'état de transition analogues du SAM qui permettraient d'étudier le mode de reconnaissance des complexes PRMT-Substrat. De manière fortuite, les composés synthétisés dans ce contexte ont démontré une capacité à inhiber sélectivement CARM1 à une concentration micromolaire voir submicromolaire. Ces composés doivent être encore optimisés premièrement pour augmenter leur efficacité mais aussi leurs propriétés pharmacologiques pour avoir l'effet souhaité sur des lignées cancéreuses. Ainsi, notre démarche pour trouver de nouveaux inhibiteurs de PRMT s'inscrit dans la continuité de ce projet, initié en 2009. Par conséquent, ce dernier chapitre ambitionne de présenter les inhibiteurs ciblant les PRMT.

La première approche consiste à explorer la poche du SAM et du substrat simultanément. Les inhibiteurs synthétisés dans ce contexte ont des structures proches du SAM (en particulier le groupe adénosine) et permet d'être compétitif de la fixation du SAM et du substrat, en occupant totalement ou partiellement les deux poches.

La deuxième approche explore la poche de fixation du substrat (où se produit la réaction de méthylation) avec des composés non-analogues du SAM. Pour cette approche, seuls les inhibiteurs en phase clinique seront décrits.

Une troisième approche implique les inhibiteurs allostériques qui se fixent à distance des poches de fixation du SAM et du substrat et provoquent des changements conformationnels qui empêche la fixation des substrats mais ne seront pas décrits dans ce manuscrit.

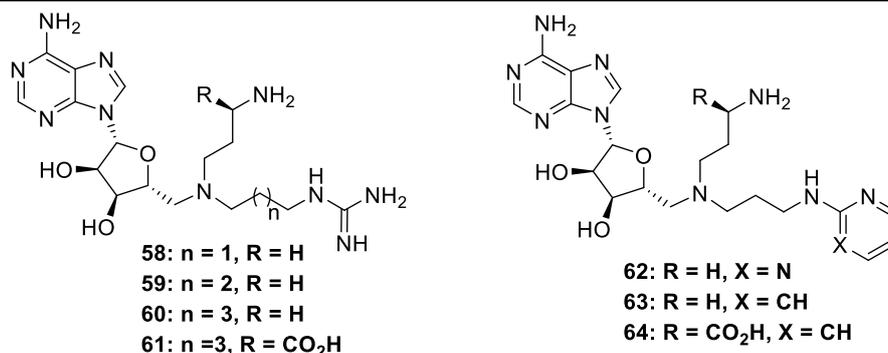


En conclusion, ces analogues inhibiteurs de la SAM montrent une spécificité bien plus élevée pour les PRMT que pour la méthyltransférase G9a et ont des IC<sub>50</sub> allant du micromolaire au sub-micromolaire pour un panel de PRMT. Les auteurs ont réalisé des études de docking avec les composés **52–54** et PRMT4. Ces études montrent le mode de liaison possible avec les inhibiteurs et les PRMT, qui pourrait expliquer l'inhibition observée. Cependant, le guanidinium ne semble pas interagir simultanément avec les chaînes latérales du glutamate (liaisons essentielles permettant d'orienter le guanidinium pour que la méthylation puisse avoir lieu), suggérant que ces trois composés pourraient être optimisés afin d'améliorer ces interactions en ajoutant des substituants donneurs de liaisons hydrogènes sur le fragment guanidine, tels des groupements *N*-hydroxy- ou *N*-amino.

#### IV.2.2 Inhibiteurs bi substrats de PRMT1 et PRMT4

En 2020, Emma A. Gunnell *et al.* ont développé des inhibiteurs de PRMT1 et PRMT4. Leurs travaux rejoignent les travaux de Van Haren *et al.* avec la conception d'inhibiteurs bi substrat qui ciblent simultanément le site du SAM et du substrat.<sup>161</sup> Une partie de leurs recherches consiste à remplacer le motif guanidinium de ces bi substrats par un isostère afin de pallier les problèmes de perméabilité cellulaire.

Les inhibiteurs **58**, **59** et **60** arborant des amines (tertiaire et primaire) et un groupe guanidinium ont été évalués pour leur potentiel à se lier et à inhiber PRMT1 et PRMT4 et les valeurs d'IC<sub>50</sub> résultantes sont présentées dans le tableau 2. L'inhibition contre PRMT4 diminue de manière modeste avec l'augmentation de la longueur du linker qui sépare l'adénosine et le guanidinium. Cependant, aucune tendance de ce type n'est observée contre PRMT1, bien que les valeurs d'IC<sub>50</sub> soient similaires à celles observées pour PRMT4.<sup>161</sup> Par ailleurs, les valeurs K<sub>d</sub> de cette série d'inhibiteurs montre un profil similaire à l'IC<sub>50</sub> pour PRMT1 et PRMT4, décrivant ainsi l'affinité directe de chaque inhibiteur. Il a été démontré que l'inhibiteur **61** (analogue acide aminé de **60**) se lie et inhibe à la fois PRMT1 et PRMT4 avec la même efficacité (Tableau 2). Les inhibiteurs **62**, **63** et **64** ont été synthétisés pour remplacer le groupe guanidinium par des isostères moins polaires, tels les motifs 2-aminopyrimidine (**62**) ou 2-aminopyridine (**63**, **64**) afin de résoudre les problèmes de perméabilité cellulaire constatés dans la famille de composés précédentes.

**Tableau 2:** Valeurs de  $K_d$  et d' $IC_{50}$  pour la série d'inhibiteurs (**7–13**) et SAH contre PRMT1 et CARM1

Inhibiteur	$K_d \pm$ erreur standard ( $\mu M$ )		Inhibiteur	$IC_{50}$ (Intervalle de confiance : 95%) ( $\mu M$ )	
	PRMT1	PRMT4		PRMT1	PRMT4
<b>58</b>	$10.2 \pm 0.996$	$3.32 \pm 0.335$	<b>58</b>	7.2 (3.5–15.0)	2.3 (1.8–3.0)
<b>59</b>	$4.02 \pm 0.453$	$7.18 \pm 0.557$	<b>59</b>	11.8 (9.0–15.5)	8.7 (6.9–10.8)
<b>60</b>	$9.99 \pm 1.89$	$9.80 \pm 1.83$	<b>60</b>	11.6 (4.9–27.2)	17.7 (13.5–23.2)
<b>61</b>	$32.2 \pm 3.53^1$	$38.1 \pm 9.61$	<b>61</b>	22.0 (14.4–33.5)	24.1 (15.4–37.6)
<b>62</b>	$10.2 \pm 4.21$	$2.30 \pm 0.162$	<b>62</b>	22.0 (15.7–30.8)	0.7 (0.5–1.0)
<b>63</b>	$15.3 \pm 1.57$	$1.14 \pm 0.411$	<b>63</b>	11.1 (5.6–21.8)	0.3 (0.1–1.1)
<b>64</b>	$43.7 \pm 14.9^2$	$1.11 \pm 0.132$	<b>64</b>	25.3 (17.3–37.0)	0.8 (0.6–1.0)
<b>SAH</b>	$0.709 \pm 0.057$	$0.775 \pm 0.111$	<b>SAH</b>	0.5 (0.2–1.0)	0.2 (0.1–0.4)

Les affinités de liaison de ces inhibiteurs avec PRMT4 oscillent autour de la valeur obtenue pour le composé **58**. Les tests d'inhibitions révèlent que ces trois composés sont plus actifs que leurs analogues guanidinium **58–60** pour PRMT4 et possèdent des valeurs d' $IC_{50}$  de l'ordre du sub-micromolaire. La meilleure inhibition est obtenue avec l'isostère 2-aminopyridine avec une  $IC_{50}$  de 0.3  $\mu M$ . Cependant, les mêmes molécules (**62–64**) sont des inhibiteurs environ trente fois moins actifs de PRMT1. Ainsi, la substitution du groupe guanidinium par l'isostère 2-aminopyridine améliore la liaison à CARM1 mais pas à PRMT1 (Tableau 2).

Pour mieux comprendre les interactions et les modes de liaison, les analogues **62**, **63** et **64** ont été co-cristallisés avec PRMT4 et trois structures cristallographiques ont été résolues avec ces inhibiteurs. Les cartes de densité électronique des sites actifs ont montré que la position du groupement adénosine est la même dans tous les cas, aussi bien dans les structures portant le guanidinium que le groupement isostérique. Les structures cristallographiques révèlent que les groupements aminés et aromatiques des composés **62** et **63** occupent simultanément la poche du substrat et du cofacteur. Cependant, l'introduction d'un groupe carboxylique (comme dans **64**) conduit à une nette préférence pour une conformation interagissant avec le site actif du cofacteur, où probablement le carboxylate de l'inhibiteur **64** mime le carboxylate du cofacteur.

En conclusion, en plus d'obtenir l'hydrophobicité nécessaire pour traverser les membranes cellulaires, les auteurs ont réussi à augmenter la puissance d'inhibition envers PRMT4 en

utilisant des isostères hydrophobes de la guanidine. Ce type d'approche pourra être utilisée pour trouver des inhibiteurs pour d'autres PRMT.

### IV.2.3 La cytosine : un potentiel isostère de la guanidine

En 2018, Halby *et al.* ont reporté des inhibiteurs analogues du SAM pharmacologiquement actifs sur les PRMT.<sup>162</sup> Ces composés ont été conçus initialement pour mimer l'état de transition de la réaction de méthylation de la cytosine et donc étaient plutôt destinés à être des inhibiteurs des ADN et ARN méthyltransférases. Cependant, ces inhibiteurs se sont révélés totalement inactifs ou très faiblement actifs lors des tests d'inhibition réalisés sur un panel d'ADN et d'ARN méthyltransférases humaines et virales mais efficaces contre les PRMT.

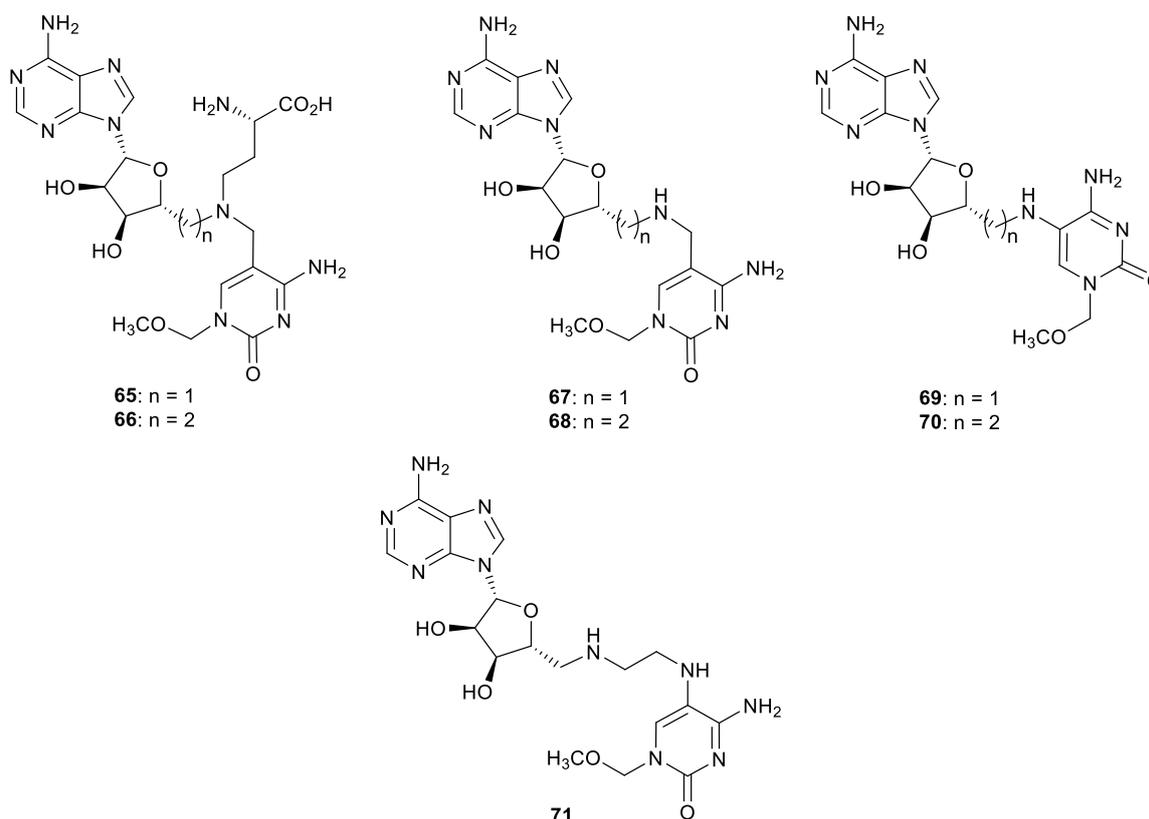


Figure 28: Inhibiteurs de PRMT portant un dérivé cytosine

Les composés **65–71** ont été testés contre quatre cibles : PRMT1, PRMT4, PRMT6 et PRMT7. Deux d'entre eux ont une partie acide aminé (**65–66**) mimant le SAM et les cinq autres n'ont que la partie adénosine et cytosine (**67–71**). Les tests d'inhibition ont montré que ces composés sont de potentiels inhibiteurs sélectifs de PRMT4, PRMT7 et à un niveau moindre de PRMT6. Deux composés, **66** et **68**, se démarquent des autres et sont beaucoup plus actifs

et sélectifs, inhibant notamment PRMT4 avec des  $IC_{50}$  de 14  $\mu$ M et 1.5  $\mu$ M respectivement. Le composé **66** diffère de son analogue **68** par la présence de la fraction acide aminé, ce qui se traduit par une diminution de l'inhibition d'environ neuf fois (Figure 28). Plusieurs structures cristallographiques ont été obtenues entre les PRMT et ces composés. Les quatre structures cristallographiques de PRMT4 (PRMT4–68, PRMT4–67, PRMT4–71, PRMT4–65) et les deux structures cristallographiques de PRMT6 (PRMT6–66, PRMT6–65) ont révélé que ces composés se fixaient selon le mode de fixation attendu, à savoir qu'ils occupaient simultanément le site du cofacteur et du substrat.

La structure cristallographique obtenue entre le composé **68** et PRMT4 montre que celui-ci occupe tous les sites actifs et déplace complètement la molécule de SAH naturellement liée sur PRMT4. Cependant, les structures cristallographiques obtenues avec les composés **65**, **67** et **71** ne déplacent que partiellement le SAH et les tests d'inhibition démontrent une faible activité sur PRMT4. L'étude cristallographique montre que cette perte d'inhibition peut s'expliquer par une longueur de linker qui est trop courte pour que la cytosine puisse pénétrer suffisamment dans le site de liaison du substrat arginine.

L'efficacité des composés **66** et **68** a été mesurée sur des lignées cancéreuses mais aucun effet cytotoxique n'a pu être observé avec ces deux inhibiteurs. Cette absence d'activité est probablement due à la faible capacité de ces composés à traverser les membranes cellulaires, ce qui a été confirmé par des tests de perméabilité cellulaire avec le composé **68**.

En conclusion, le composé **68** s'est révélé l'inhibiteur le plus efficace contre PRMT4 parmi les composés testés dans cette étude et apporte des informations essentielles pour comprendre les bases structurales de l'inhibition des PRMT. Ces informations pourront servir pour optimiser cette structure de manière à améliorer ses propriétés pharmacologiques ou bien de concevoir de nouveaux inhibiteurs non-analogues du SAM avec de meilleures propriétés pharmacologiques et une meilleure efficacité.

En 2017 *Brehmer et al.* ont identifié grâce à une étude de SAR un nouvel inhibiteur analogue du SAM sélectif de PRMT5.<sup>163</sup> Cet inhibiteur appelé **JNJ64619178** (Figure 29) s'est révélé être très efficace *in vitro* ( $IC_{50}$  = 0.13 nM) sur diverses lignées cellulaires cancéreuses notamment pulmonaires, mammaires, pancréatiques et hématologiques. Pour confirmer davantage l'activité de celui-ci, les évaluations ont été étendu à des études *in vivo*. Pour cela des xénogreffes (type : NCI-H1048) dérivées de patients atteints de tumeurs ont été greffés sur des souris immunodéficientes. L'inhibiteur a été administré par voie intraveineuse et a conduit à une inhibition et une régression efficace de la croissance tumorale sur plusieurs modèles de xénogreffe ( $EC_{50}$  = 0.25 nM). Celui-ci a été engagée en essai clinique de phase I (NCT03573310) pour des patients atteints de tumeurs solides avancées, de lymphomes non hodgkiniens et de syndrome myélodysplasique à faible risque.<sup>163,164</sup> Aucun résultat n'a été

publié à ce jour et la date d'achèvement estimée de l'étude sur les 114 participants est le 29 décembre 2023.<sup>165</sup>

En 2018, *Bonday et al.* ont rapporté un nouvel inhibiteur sélectif de PRMT5 qui se lie spécifiquement dans la poche SAM de PRMT5. Ce composé appelé LLY-283 (Figure 29) présente une bonne activité *in vitro* (IC<sub>50</sub> de 22 ± 3 nM) et *in vivo* (EC<sub>50</sub> de 25 ± 1 nM).<sup>166</sup> Ce dernier présente une bonne perméabilité cellulaire et pharmacocinétique bien définie (disponibilité orale, clairance plasmatique, ...). Ce composé a montré une activité antitumorale dans les xénogreffes de souris de type A375 lorsqu'il est administré par voie orale. Par ailleurs, une étude de SAR a permis de mettre au point le composé **PF-06855800** étant plus efficace avec une activité *in vitro* de 1.4 nM. Une caractéristique partagée par ces trois inhibiteurs est le cycle déazapurine, qui améliore les propriétés physico-chimiques et fait gagner une interaction hydrophobie supplémentaire. En effet l'azote en position 7 peut être retiré du cycle adénine sans compromettre l'efficacité de l'inhibiteur (Figure 29).<sup>167</sup>

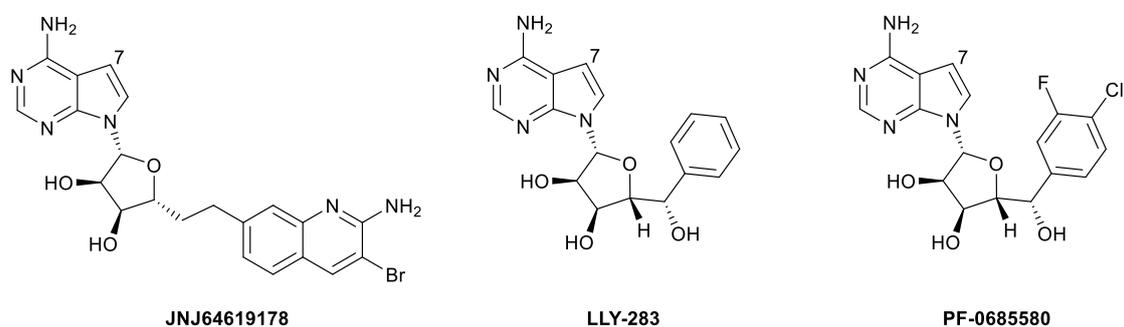


Figure 29: Les inhibiteurs analogues du SAM en essais cliniques

Récemment, dans la continuité de ces travaux le composé **PF-06939999** (structure non dévoilée) est rentré en essai clinique chez des patients atteints de tumeurs solides avancées ou métastatiques. Le but de cette étude était d'évaluer la pharmacocinétique et pharmacodynamique de cet inhibiteur chez 54 participants, en administrant (voie orale) des doses croissantes de celui-ci. L'étude a abouti à son terme en avril 2022 et **PF-06939999** a montré des toxicités dose-dépendantes avec quelques effets indésirables (fatigue, nausée, thrombocytopenie, anémie) gérables. De plus, des réponses tumorales objectives (permet de retarder la progression tumorale et d'améliorer la survie) ont été observées chez les patients atteints de carcinome épidermoïde de la tête et du cou et de cancer du poumon non à petites cellules. Celui-ci va entamer des essais de clinique de phase II prochainement.<sup>168</sup>

## IV.3 Inhibiteurs non-analogues du SAM en phase clinique :

### IV.3.1 Collaboration des laboratoires pharmaceutiques pour la mise au point d'inhibiteurs de PRMT5

En 2015 d'autres inhibiteurs ont été rapportés par criblage à haut débit dans le cadre d'une collaboration entre plusieurs laboratoires pharmaceutiques. Le premier Hit **72** qui a été rapporté représente une  $IC_{50}$  de 0.33  $\mu$ M contre PRMT5. Par la suite, une campagne d'optimisation par une étude de SAR a dispensé l'identification de l'inhibiteur **73** ayant une  $IC_{50}$  de 22 nM.<sup>169,170</sup> Le remplacement de l'oxétane présent sur cet inhibiteur par un cyclobutane aboutissant à l'inhibiteur **74** permet d'atteindre une  $IC_{50}$  de 11 nM (Figure 30).<sup>85</sup>

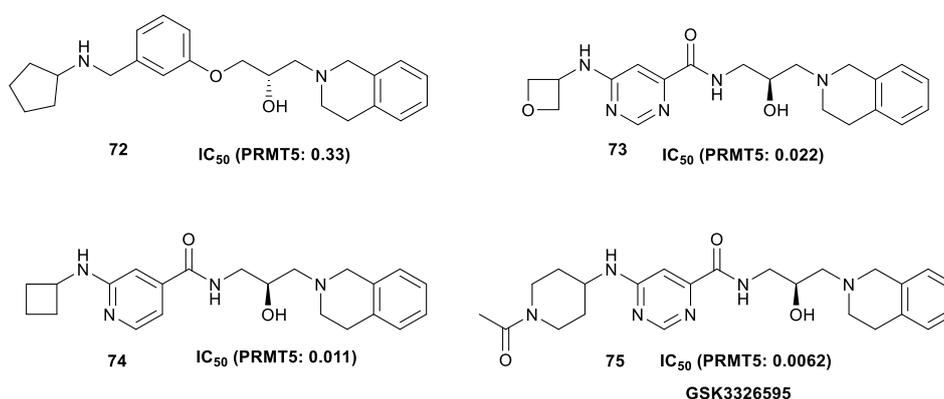


Figure 30: Inhibiteurs efficace et sélectif de PRMT5

Ce dernier a une très haute sélectivité pour PRMT5 comparé à une vingtaine d'enzymes responsables de PTM. Dans la foulée le dérivé **75** (appelé GSK3326595) a également été synthétisé et évalué contre PRMT5 et l'inhibe avec une  $IC_{50}$  de 6.2 nM. Celui-ci est entré en essai clinique de phase I (NCT03886831, NCT04089449 et NCT02783300) pour le développement de thérapies contre les lymphomes non hodgkinien, les tumeurs solides et les cancers avancés où les patients ont épuisés tous les traitement disponibles.<sup>85, 171</sup>

### IV.3.2 Collaboration des laboratoires pharmaceutiques pour la mise au point d'inhibiteurs de PRMT4

Une autre collaboration entre entreprises pharmaceutiques a permis de d'identifier un inhibiteur de PRMT4 puissant de l'ordre du nanomolaire.<sup>172</sup> Cet inhibiteur est le fruit d'une conception rationnelle (basés sur les structures cristallographiques obtenues avec d'autres inhibiteurs de PRMT) combinés à une étude SAR. En effet, cette collaboration a permis de

mettre au point le composé **76** qui inhibe l'activité enzymatique de PRMT4 avec une  $IC_{50}$  de 3 nM. De plus celui-ci est sélectif de PRMT4 comparé à une large panoplie de méthyltransférases (Figure 31).

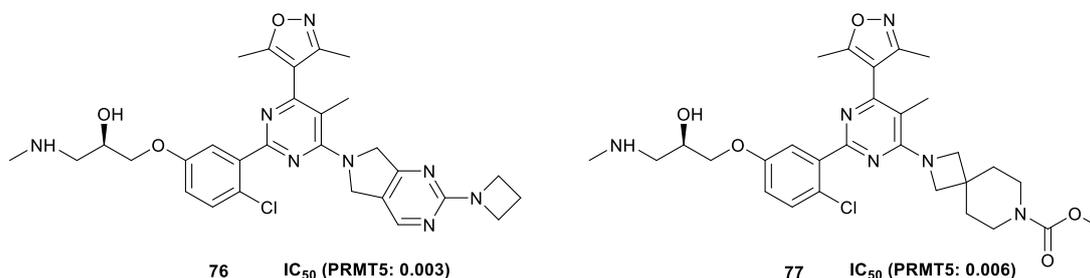


Figure 31: Une autre série d'inhibiteurs sélectifs de PRMT5 de l'ordre du nanomolaire

Même si celui s'est avéré efficace pour des études *in vitro*, son profil pharmacocinétique n'était pas satisfaisant chez la souris et le rat. Ainsi, une amélioration de la structure a permis l'accès à l'inhibiteur **77** qui est légèrement moins efficace avec une  $IC_{50}$  de 6 nM mais possède une meilleure stabilité plasmatique, le plaçant dans une meilleure position pour une utilisation *in vivo*. De plus, celui-ci maintient son profil de sélectivité par rapport aux autres histones méthyltransférases et bloque la prolifération sur plusieurs lignées cancéreuses à la fois *in vitro* et *in vivo*.<sup>172</sup>

## IV.4 Synthèse d'inhibiteurs de PRMT

### IV.4.1 Introduction

Une partie de ma thèse a été consacrée à la recherche de nouveaux inhibiteurs de CARM1. Comme annoncé précédemment, la surexpression des PRMT dans les cellules cancéreuses est un fait établi donc il est nécessaire d'atténuer leur fonction et de contrôler l'équilibre épigénétique en inhibant ces acteurs épigénétiques.

Dans ce contexte, nous envisageons de synthétiser la molécule **79** (Figure 34), dont la structure a été sélectionnée via une approche *in silico*, basée sur et dans la continuité des résultats précédents obtenues avec des analogues du SAM, en collaboration avec l'équipe de Jean Cavarelli.<sup>128</sup> En effet, au sein de notre équipe, le Dr Samira Ajebbar a synthétisé l'inhibiteur **SA678** dérivé du SAM qui inhibe CARM1 avec une  $IC_{50}$  de 5  $\mu$ M (Figure 32).

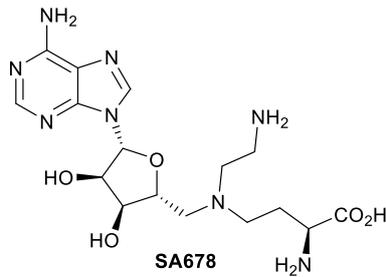


Figure 32: L'inhibiteur SA678

Celui-ci a été co-cristallisé avec CARM1 (PDB 5NTC) au sein de l'équipe de Jean Cavarelli. La structure cristallographique a mis en évidence la présence de BTP (Bis-Tris Propane, un additif utilisé pour la cristallisation des protéines) provenant du milieu de cristallisation. Ce BTP se trouve dans une poche adjacente à la poche SAM et se trouve stabilisé en formant plusieurs liaisons hydrogènes avec des résidus de cette poche et des molécules d'eau. De plus, le BTP forme une liaison hydrogène avec l'azote en position 7 du groupement adénine de l'inhibiteur SA678 (Figure 33).

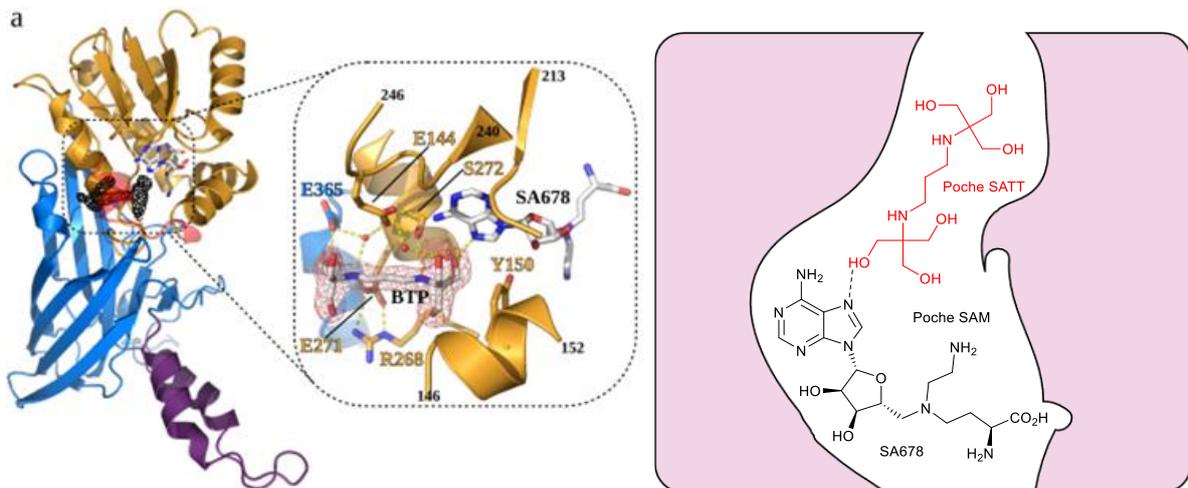


Figure 33: A gauche : structure cristallographique du module de CARM1 de souris en complexe avec l'inhibiteur SA678. A droite: Représentation schématique du positionnement de la molécule de BTP (en rouge) dans la poche catalytique

Ainsi, nous avons estimé que cette poche (appelé poche SATT) où se fixe le BTP pourrait être une région à cibler pour de futurs inhibiteurs de PRMT. Contrairement aux sites SAM et arginine extrêmement conservés dans la famille des PRMT, ce site SATT possède une séquence très variable. Cela suggère que des inhibiteurs se liant à la poche SATT permettraient potentiellement d'avoir des approches plus sélectives.

Cette découverte nous a conduit à développer une nouvelle famille d'inhibiteurs fixant cette poche SATT. Ainsi, au sein de notre équipe, le Dr. Julie Schmitt a synthétisé le composé **78** qui inhibe sélectivement PRMT1 et CARM1 à une concentration sub-micromolaire (Figure 34). Cependant, des études de co-cristallisation ont démontré une certaine flexibilité autour du

groupement phénylpropanediol (en rouge), nous poussant à développer un nouvel inhibiteur moins flexible et moins polaire, le dérivé quinazoline **79**.

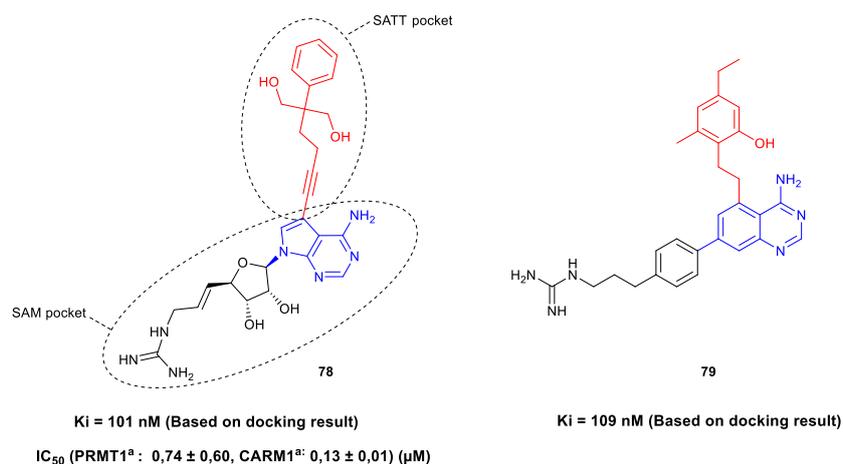


Figure 34: Conception de nouveaux inhibiteurs contre CARM1

#### IV.4.2 Résultats et discussions

Afin de synthétiser l'inhibiteur retenu, nous avons développé une approche rétrosynthétique en le divisant tout d'abord en trois synthons, le corps central **80**, le boronate **81** et l'alcyne correspondant **82**, dont la synthèse est prévue à partir de la dibromoaniline **83**, alcool propargylique **84** et cyclohexènone **85**, respectivement. Le guanidinium porté par le boronate **81** sera introduit sur le corps central **80** via un couplage de Suzuki-Miyaura, tandis que le dernier synthon **82** le sera par un couplage de Sonogashira (Figure 35). A première vue les étapes de couplages risquent d'éprouver un problème de régiosélectivité. Dans un premier cas on pourrait espérer que la pseudo-aniline du corps central agisse comme un « groupement directeur » où elle viendrait chélater le palladium et l'approcher du brome environnant. Dans un deuxième cas, pour s'assurer de la régiosélectivité, on pourrait essayer de protéger la pseudo-aniline du corps central par un groupement Boc pour encombrer le voisinage d'un des bromures et favoriser l'attaque sur le deuxième brome.

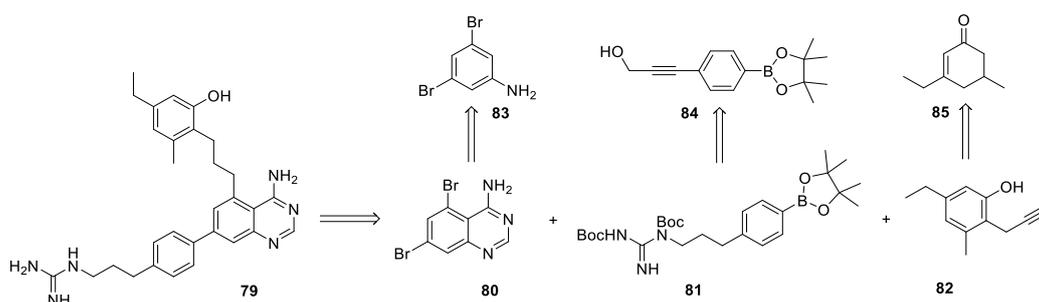


Figure 35: Analyse rétrosynthétique du candidat inhibiteur

La synthèse du boronate a été réalisée en deux étapes à partir de l'ester pinacolique commercial **84**. Celui-ci a été engagé dans une réaction d'hydrogénation catalytique pour réduire l'alcyne de l'alcool propargylique permettant l'obtention du composé **86** avec un rendement quantitatif. La guanidine a été introduite par une réaction de Mitsunobu en présence de 1,3-bis(*tert*-butoxycarbonyl)guanidine conduisant au boronate attendu avec un rendement de 65% (Schéma 18).

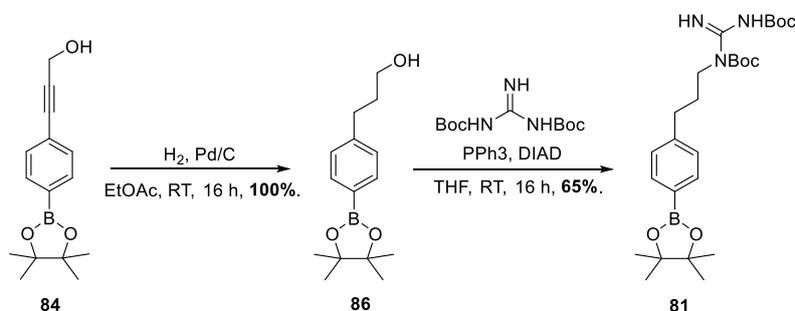


Schéma 18: Synthèse du boronate

La synthèse du corps central emploie la 3,5-dibromoaniline commerciale **83**, convertie en isatine **88** grâce à la méthodologie de Sandmeyer, à savoir une condensation initiale de l'hydrate de chloral sur l'aniline en présence d'hydroxylamine afin d'isoler l' $\alpha$ -isonitrosoacetanilide **87** correspondant qui, sous l'action de l'acide sulfurique subit une déshydratation puis une cyclisation afin de conduire à l'isatine attendue avec un rendement global de 92%.<sup>173</sup> Une oxydation en présence d'eau oxygénée et de soude permet d'ouvrir le cycle et de le convertir en acide anthranilique **89** correspondant. Une estérification a ensuite été réalisée en présence de triméthylsilyldiazométhane conduisant à l'ester méthylique **90**. Ce dernier a été mis en présence d'acide formique chauffé à 135°C pendant 10 h dans du formamide afin de générer le motif 4-quinazolinone **91**, ayant une structure proche du corps central (appelé ci-après pseudo corps central) (Schéma 19).

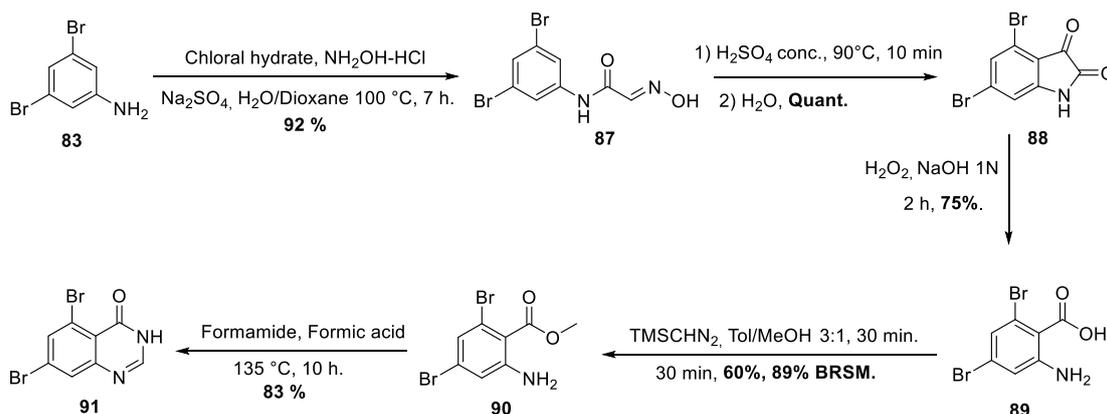


Schéma 19: Synthèse de la quinazolinone

Possédant désormais un intermédiaire avancé entre les mains, nous avons voulu évaluer le couplage de Suzuki-Miyaura entre le pseudo corps central **91** et le boronate **81** précédemment décrit. Nous voulions en effet avoir une idée préalable de la faisabilité de cette voie ainsi que de sa régiosélectivité, quitte à générer le motif 4-aminoquinazoline dans un second temps.

Cependant, ce couplage de Suzuki-Miyaura s'est révélé absolument inefficace malgré tous nos efforts, conduisant systématiquement à l'obtention d'un mélange de pseudo corps central n'ayant pas réagi et le produit d'hydrolyse du boronate **81**.

Afin de mieux identifier l'élément problématique de ce couplage, nous nous sommes tournés vers l'utilisation d'un système modèle plus simple, composé du boronate **81** et du bromobenzène **92** en espérant mettre au point des conditions optimales. Dans un système réactionnel standard utilisant l'acétate de palladium et le phosphate de potassium dans un mélange THF/eau, nous nous sommes concentrés sur l'évaluation de divers ligands (Schéma 20).

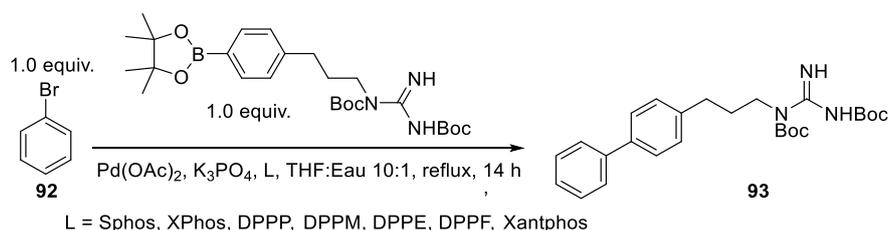


Schéma 20 : Optimisation des conditions de couplage avec le bromobenzène

Parmi les sept ligands testés, la meilleure conversion (déterminée par LC-MS) a été observée avec Xphos, nous permettant d'obtenir un rendement isolé pour notre produit de couplage de 91%. Cependant, l'application de ces conditions optimales au pseudo corps central **91** s'est à nouveau révélée inefficace, aucune conversion n'ayant été constatée.

Nous avons alors décidé de nous tourner vers le précurseur ester méthylique **90** afin d'évaluer si le motif quinazolinone n'était pas responsable de l'absence de réactivité. Deux séries de réactions ont été réalisées en faisant varier soit la source de palladium, soit le type de ligand.

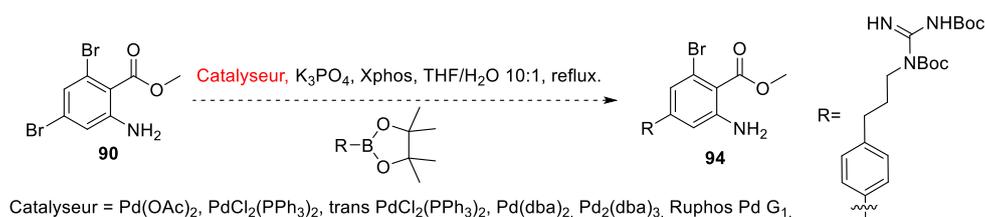


Schéma 21: Couplage de Suzuki avec différents catalyseurs

Ainsi, dans la première série de réactions nous avons choisi comme ligand le Xphos, le ligand le plus efficace précédemment identifié. Parmi les six catalyseurs testés, nous n'avons

observé aucune conversion, avec une récupération complète des réactifs de départ (Schéma 21). Dans une deuxième série de réactions, nous avons alors choisi de passer en revue divers ligands, tout en conservant l'acétate de palladium comme source de palladium. Pour les ligands monodentates (SPhos et Xphos), aucune conversion n'a été constatée. De manière surprenante, avec les ligands bidentates DPPF, DPPE, DPPM, DPPPP nous avons observé un couplage hautement régiosélectif, uniquement sur la position 3 du phényle. A l'inverse, en présence de Xantphos, un mélange de régioisomères **95** (15% de rendement isolé) et **94** (21% de rendement isolé) a été obtenu (Schéma 22).

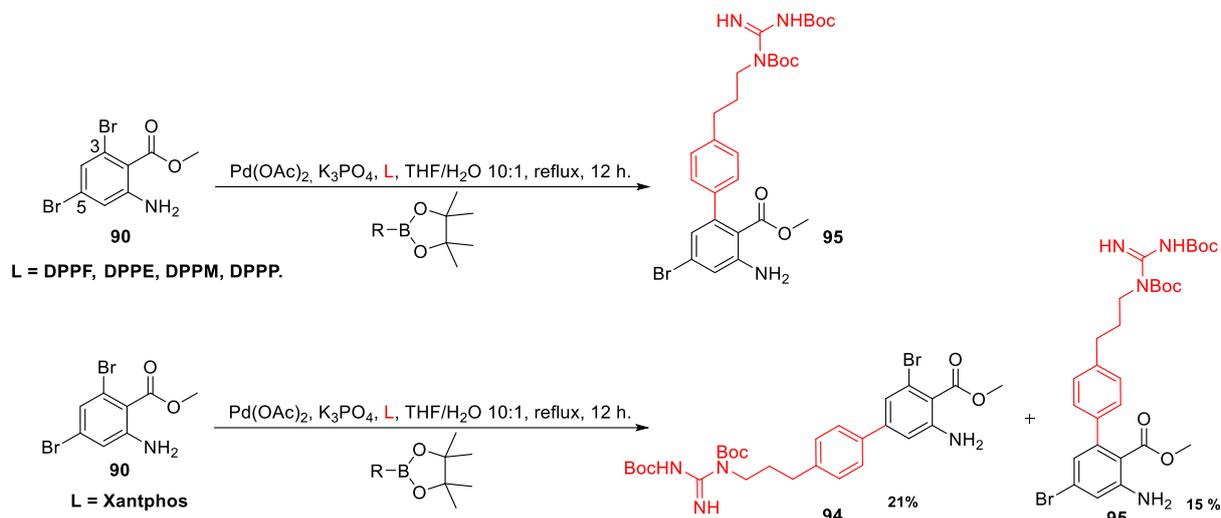


Schéma 22: Couplage de Suzuki avec différents ligands

A la suite de ces résultats encourageants, nous avons alors entrepris la cyclisation du composé **94** en présence de formamide et d'acide formique comme précédemment décrit, mais qui s'est cette fois-ci accompagnée d'une conversion médiocre. Utilisant de l'acide paratoluènesulfonique monohydraté (APTS.H<sub>2</sub>O) à la place de l'acide formique, 15% de la quinazolinone **96** correspondante a pu être obtenue, avec un faible rendement s'expliquant essentiellement par la dégradation du réactif de départ, notamment via une réaction de débromation donnant lieu au composé **97** (confirmé par LC-MS et RMN) (Schéma 23).

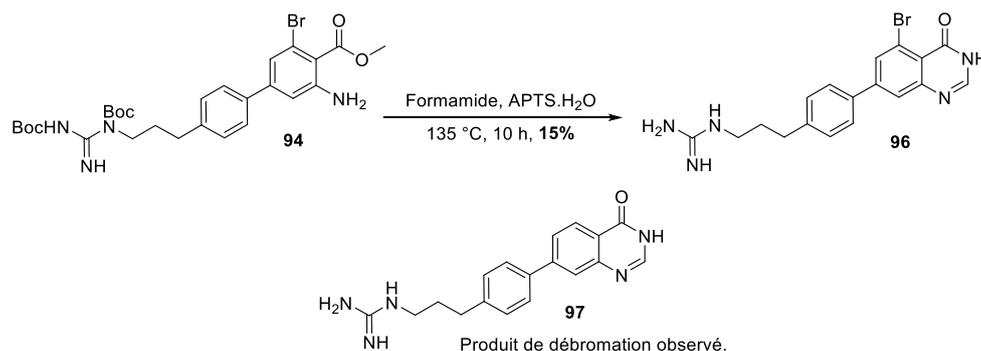


Schéma 23: Cyclisation et voies envisagées pour l'obtention de l'inhibiteur

Bien que ces résultats soient encourageants en l'état, nous n'avons pas pu poursuivre cette voie de synthèse plus avant par manque de temps.

En parallèle de cette approche, nous avons en outre travaillé sur une voie alternative, permettant l'accès au dérivé aminoquinazoline **80** d'intérêt. Celle-ci commence à partir de l'ester méthylique précédemment décrit, réduit en deux étapes en benzaldéhyde **99** (Schéma 24).

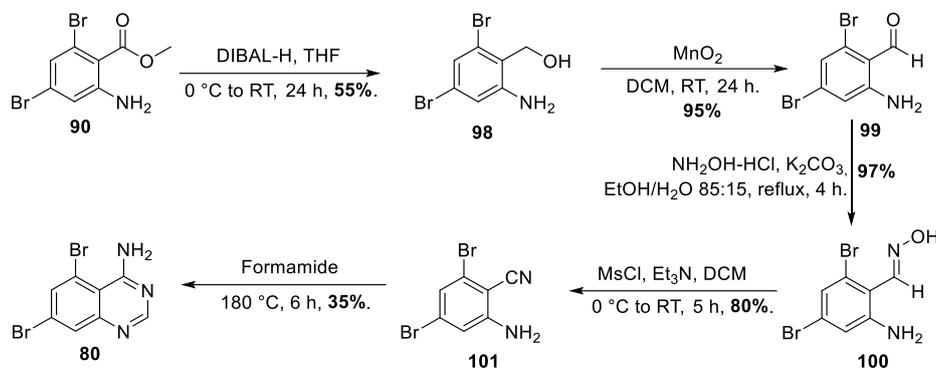


Schéma 24: Voie alternative pour la synthèse du corps central

Afin d'introduire la fonction aminoquinazoline, cet aldéhyde a dû être converti en nitrile **101**, via la formation de l'oxime intermédiaire **100** en présence d'hydroxylamine et carbonate de potassium (K<sub>2</sub>CO<sub>3</sub>). Cette oxime a été déshydratée en présence de chlorure de méthanesulfonyle, permettant l'accès au benzonitrile **101** correspondant avec un bon rendement. Une condensation du formamide sur l'aniline suivi d'une cyclisation nous a ensuite conduit à l'hétérocycle désiré **80** (Schéma 24). Malheureusement, cette aminoquinazoline s'est révélée tout aussi inerte aux conditions de couplage de Suzuki-Miyaura que son analogue quinazolinone **91**. Comme précédemment, nous nous sommes alors tournés vers un intermédiaire précédent, le benzonitrile **101** en appliquant les conditions mises au point dans le cas de l'analogue ester méthylique **90**. Cette fois-ci, à notre grande surprise, nous avons observé le couplage uniquement sur la position désirée, donnant lieu au composé **102** avec un rendement de **36%** (Schéma 25). Ainsi, l'application de nos conditions optimales sur le benzonitrile **101** nous a permis d'avoir cette fois-ci une très bonne régiosélectivité.

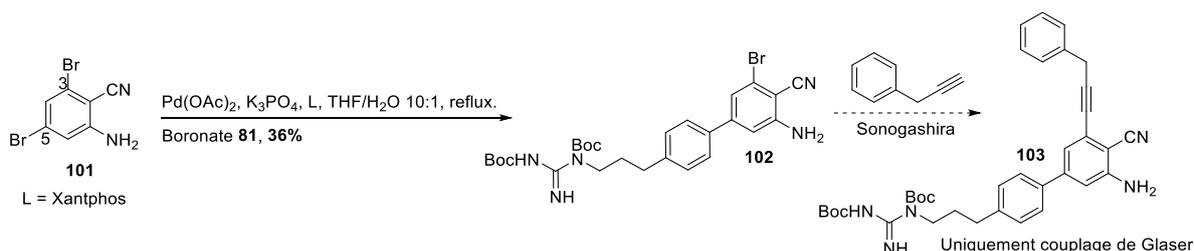


Schéma 25: Couplage de Suzuki avec le benzonitrile et la voie envisagée pour l'obtention de l'inhibiteur

Mus par cet excellent résultat, nous avons alors entrepris le couplage de Sonogashira avec un alcyne simplifié (Schéma 25). Cependant, seul le produit de Glaser (produit d'homodimérisation de l'alcyne) a été obtenu, suggérant un dégazage de nos solvants perfectible. Par manque de temps, nous n'avons cependant pas pu aller plus loin dans la synthèse, mais cette voie de synthèse s'avère plus prometteuse que la précédente.

#### IV.4.3 Conclusion

En conclusion, le boronate **81** a été synthétisé en deux étapes avec un rendement global de 65%. Le corps central **80** a été synthétisé en 9 étapes avec un rendement global de 6% et le pseudo corps central **91** en 5 étapes avec un rendement global de 34%. Les couplages entre les deux corps centraux et le boronate se sont révélés inopérants. Nous avons conjecturé que ce manque de réactivité pourrait être lié à une quelconque chélation du catalyseur par le corps central, ce qui a été partiellement confirmé par le succès de ces couplages sur des précurseurs acycliques. De plus, nous avons démontré que parmi les ligands testés, uniquement le Xantphos nous permettait d'obtenir la régiosélectivité souhaitée. Les prochaines étapes clés de cette voie de synthèse sont l'introduction du dernier synthon par un couplage de Sonogashira et la cyclisation sous forme d'aminoquinazoline. Deux étapes supplémentaires d'hydrogénation catalytique et de déprotection des groupements Boc seront nécessaires pour aboutir à l'inhibiteur.

## V. Conclusion et Perspectives

Les travaux réalisés au cours de cette thèse avaient deux objectifs principaux :

- Comprendre le processus de régulation de la réaction de méthylation des protéines
- Découvrir des inhibiteurs sélectifs de PRMT4 à fort pouvoir inhibiteur et possédant des bonnes propriétés pharmacologiques pour un usage thérapeutique

Pour parvenir au premier objectif nous avons synthétisé plusieurs composés analogues du SAM. Dans ce contexte, nous avons constitué une petite librairie de ligands mimant l'état de transition de la première ou de la seconde étape de méthylation des différents types de PRMT. Les structures cristallographiques obtenues avec ces composés et les PRMT sont toujours à l'étude. Les données obtenues doivent être encore étudiées de manière approfondie pour essayer de comprendre le mécanisme de reconnaissance entre PRMT-ligand. Les informations collectées avec ces complexes pourront donner quelques pistes pour expliquer pourquoi les PRMT n'adoptent pas le même mécanisme de reconnaissance et comment elles peuvent méthyler sélectivement plusieurs sites sur une même protéine. De plus, les structures cristallographiques obtenues avec ces composés et PRMT4 révéleront des informations capitales pour guider une étude de SAR afin d'augmenter le pouvoir inhibiteur de ce composé mais aussi de concevoir d'autres inhibiteurs à action pharmacologique. Une publication sur ce travail est en préparation.

Les tests d'inhibition réalisés avec ces composés analogues du SAM sont très encourageants, notamment pour le composé **34a** qui présente une  $IC_{50}$  de 330 nM (Figure 36).

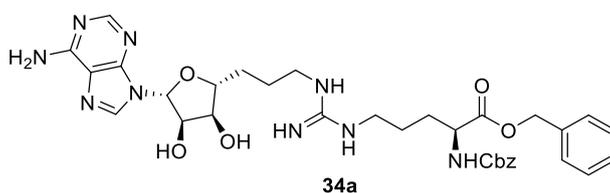


Figure 36: Arginine modifiée non méthylé

La fixation directe de ces composés aux PRMT ont été mesurés grâce à des mesures de stabilité thermique. Une variation significative a été obtenue avec PRMT4 en présence de ces composés, démontrant que ces composés stabilisent PRMT4. Ainsi, l'incorporation de ces analogues au sein des substrats de PRMT permettra la stabilisation des complexes PRMT-substrat qui représente un grand défi dans ce domaine

C'est dans ce contexte que nous avons entrepris d'utiliser la ligation chimique native pour introduire ces analogues pour stabiliser le complexe RSF1-PRMT2. Ainsi, nous avons atteint une conversion satisfaisante de 65% entre la RSF1 tronquée, comportant une cystéine en N-ter et notre peptide modèle contenant une arginine naturelle.

Ces analogues du SAM une fois incorporé dans le peptide d'intérêt ligué à la RSF1 en appliquant les conditions mis au point. Cependant, les réactions de guanylation permettant leur obtention sont les étapes limitantes de notre voie de synthèse. Ces dernières restent donc à optimiser afin d'accroître les rendements et monter en échelle en vue de permettre la synthèse du peptide prévus.

Pour parvenir au deuxième objectif, nous avons conçu un nouvel inhibiteur par une approche *in silico* qui se présente plus adaptés pour une utilisation pharmacologique. Premièrement, elle présente une structure assez hydrophobe qui lui permettrait de traverser les membranes cellulaires. Deuxièmement elle détient le motif quinazoline qui revient fréquemment dans les anticancéreux.

A ce jour, nous avons réussi à introduire la partie guanidine **81** sur le benzonitrile **101**, précurseur du motif quinazoline central, via une réaction de couplage de Suzuki-Miyaura avec un rendement de 36% avec du Xantphos qui s'est avérée l'idéal parmi les ligands testés car elle permet d'avoir la régiosélectivité souhaitée. Le plus difficile étant fait, nous nous engageons nos efforts pour l'introduction de la partie phénylpropane et la cyclisation pour obtenir le candidat inhibiteur.

# Experimental Part

## Material and Methods

All dry solvents and reagents were purchased from commercial suppliers and were used without further purification. Thin layer chromatography (TLC) analyses were carried out on 60 F-254 aluminium sheets. The purifications by chromatography on silica gel were carried out on a stationary phase of silica 60 (Merck, Geduran 60, 40-63  $\mu\text{m}$ ). NMR experiments were recorded at 23 °C on Bruker - 400 MHz / 500 MHz / 700 MHz spectrometers. Chemical shifts are expressed in parts per million (ppm) relative to the residual solvent signal:  $\text{CDCl}_3$  ( $\delta\text{H} = 7.26$ ,  $\delta\text{C} = 77.36$ ),  $\text{CD}_3\text{OD}$  ( $\delta\text{H} = 3.31$ ,  $\delta\text{C} = 49.00$  ( $\text{CH}_3$ )), DMSO ( $\delta\text{H} = 2.5$ ,  $\delta\text{C} = 39.52$ ).  $^1\text{J}$  values are in Hz.

**Analytical LC-MS analyses** were carried out on Waters 2695 separations module equipped with Waters2487 UV detector, Waters Acquity QDa mass detector and CORTECS, 2.7  $\mu\text{m}$ , C18, 50 x 4.6 mm column. The flow rate was 1 mL/min, and the solvent system was composed as follows: solvent A: 0.05% TFA in water; solvent B: acetonitrile. Unless indicated otherwise, the gradient run was 0 to 5 min. – 5% to 95% B; 5 to 6 min. – 95% B; 6 to 7 min. – 5% B. Mass detector was operated in positive MS Scan mode with 600 °C probe temperature, 1.5 kV capillary voltage and 10 V cone voltage.

**High-resolution mass spectra** (HRMS) were obtained using an Agilent Q-TOF (time of flight) 6520

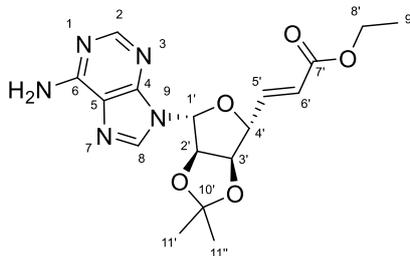
**Purification by preparative HPLC** were carried out on a Waters 600 controller system (pumps: Waters Delta 600; detector: Waters 2489 UV/Vis) equipped with a SunFire™ Prep C18 OBD 5  $\mu\text{m}$  19x150 mm column (Waters), using water (0.1% TFA, solvent A) and acetonitrile (solvent B) as a solvent system with a flow rate of 17 mL.min<sup>-1</sup>.

**IR spectra** were recorded in a Thermo-Nicolet FT/IR-380 spectrometer. Spectra were interpreted with OMNIC 9 software and are reporter in  $\text{cm}^{-1}$ .

**Melting points** were carried out on a melting point apparatus SMP3 from Stuart Scientific.

## Chemical characterizations

### Ethyl (E)-3-((3aR,4R,6R,6aR)-6-(6-amino-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)acrylate (17)



2',3'-O-isopropylideneadenosine **15** (1.0 equiv., 16.28 mmol, 5.0 g) was dissolved in dimethylsulfoxide (DMSO) (50 mL) and 2-iodoxybenzoic acid (IBX) (2.5 equiv., 40.70 mmol, 11.4 g) and  $\text{Ph}_3\text{P}=\text{CHCOOC}_2\text{H}_5$  (3.0 equiv., 48.8 mmol, 17 g) were added. The mixture was stirred at room temperature for 72 h. Water (200 mL) was added and the mixture was extracted with EtOAc (5 x 75 mL). The combined layers were washed with brine (5 x 50 mL), dried with  $\text{Na}_2\text{SO}_4$ , and concentrated. The crude product was purified by flash chromatography (4% MeOH in EtOAc) and concentrated to afford the title compound **17** (4829 mg, 12.86 mmol, 79%) as a white powder.

$R_f$  0.30 (96:4 EtOAc/MeOH)

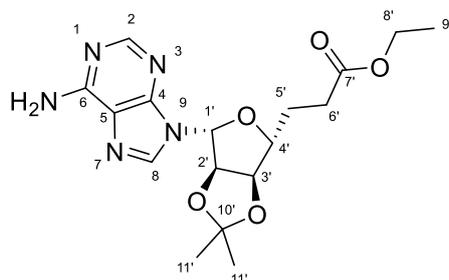
mp: 74 °C

$^1\text{H NMR}$  (500 MHz,  $\text{CDCl}_3$ )  $\delta$  8.32 (s, 1H, H2), 7.86 (s, 1H, H8), 6.96 (dd,  $J = 15.7, 5.6$  Hz, 1H, H5'), 6.13 (d,  $J = 1.7$  Hz, 1H, H1'), 5.88-5.78 (m, 3H, C6-NH2, H6'), 5.55 (dd,  $J = 6.2, 1.7$  Hz, 1H, H2'), 5.13 (dd,  $J = 6.2, 3.6$  Hz, 1H, H3'), 4.81 (m, 1H, H4'), 4.12 (q,  $J = 7.2$  Hz, 2H, H8'), 1.62 (s, 3H, H11''), 1.40 (s, 3H, H11'), 1.22 (t,  $J = 7.2$  Hz, 3H, H9').

$^{13}\text{C NMR}$  (125 MHz,  $\text{CDCl}_3$ )  $\delta$  165.7 (C7'), 155.7 (C5), 153.3 (C2), 149.6 (C4), 143.5 (C5'), 140.1 (C8), 122.7 (C6'), 120.4 (C6), 114.9 (C10'), 90.7 (C1'), 86.4 (C4'), 84.5 (C3'), 84.1 (C2'), 60.7 (C8'), 27.2 (C11''), 25.5 (C11'), 14.3 (C9').

LR-ESI-MS ( $\text{M}+\text{H}^+$ ) calc. for  $\text{C}_{17}\text{H}_{22}\text{N}_5\text{O}_5$ : 376.16; found: 376.16

**Ethyl 3-((3aR,4R,6R,6aR)-6-(6-amino-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)propanoate (18)**



**Ethyl (E)-3-((3aR,4R,6R,6aR)-6-(6-amino-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)acrylate 17** (1.0 equiv., 1.47 mmol, 550 mg) was dissolved in 50 mL of EtOAc. A flow of argon is carried out before the addition of palladium on carbon (0.1 equiv., 0.147 mmol, 10% wt, 155 mg). A second flow of argon is carried out before putting the solution under an H<sub>2</sub> atmosphere. The reaction medium is stirred overnight at room temperature. The solution is filtered through celite, washed EtOAc (3 x 25 mL) and concentrated under reduced pressure affording the title compound **18** as a white foam (549 mg, 1.45 mmol, Quant.).

**R<sub>f</sub>** 0.27 (96:4 Dichloromethane/MeOH)

**v<sub>max</sub>/cm<sup>-1</sup>**: 3225 (br) 3170 (br), 2984 (br), 2935 (br), 1729 (s), 1643 (s), 1597 (s), 1476 (m), 1374 (w), 1208 (m), 868 (w), 650(w).

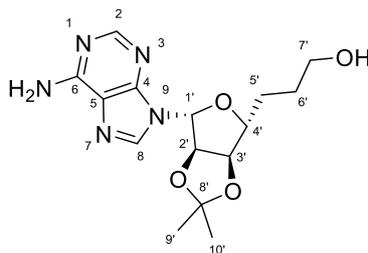
**mp**: 64 °C

**<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)** δ 8.32 (s, 1H, H2), 7.87 (s, 1H, H8), 6.29 (bs, 2H, C6-NH2), 6.02 (d, *J* = 2.4 Hz, 1H, H1'), 5.46 (dd, *J* = 6.5, 2.4 Hz, 1H, H2'), 4.86 (dd, *J* = 6.5, 3.9 Hz, 1H, H3'), 4.23 – 4.14 (m, 1H, H4'), 4.05 (qd, *J* = 7.2, 1.6 Hz, 2H, H8'), 2.35 (dd, *J* = 8.1, 6.9 Hz, 2H, H5'), 2.09 – 1.95 (m, 2H, H6'), 1.57 (s, 3H, H11''), 1.36 (s, 3H, H11'), 1.17 (td, *J* = 7.2, 1.6 Hz, 3H, H9').

**<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>)** δ 172.8 (C7'), 155.9 (C5), 153.1 (C2), 149.4 (C4), 139.9 (C8), 120.4 (C6), 114.8 (C10'), 90.3 (C1'), 86.0 (C4'), 84.1 (C2', C3'), 60.6 (C8'), 30.5 (C5'), 28.6 (C6'), 27.3 (C11''), 25.5 (C11'), 14.2 (C9').

**LR-ESI-MS (M+H<sup>+</sup>)** calc. for C<sub>17</sub>H<sub>24</sub>N<sub>5</sub>O<sub>5</sub>: 378.17; found: 378.18

**3-((3aR,4R,6R,6aR)-6-(6-amino-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)propan-1-ol (19)**



To a solution of ethyl (E)-3-((3aR,4R,6R,6aR)-6-(6-amino-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4 d][1,3]dioxol-4-yl)acrylate **18** (1.0 equiv., 0.4907 mmol, 185 mg) in THF (15 mL) at -78°C was added DIBAL-H (10.0 equiv., 0.203 mmol, 1.0 M in THF, 6 mL) dropwise. The reaction mixture was stirred for 1 hours at rt, then it was quenched with cold methanol (5 mL), then with a solution of rock salt (10 mL, aq., sat.). After stirring 1 hour, the aqueous phase was extracted with EtOAc (3 x 10 mL). The combined layers were washed (brine), dried (Na<sub>2</sub>SO<sub>4</sub>) and concentrated under reduced pressure affording the title compound **19** as a white foam.

**R<sub>f</sub>** 0.31 (9:1 Dichloromethane/MeOH)

**v<sub>max</sub>/cm<sup>-1</sup>**: 3225 (br), 3188 (br), 2928 (br), 1644 (s), 1600 (s), 1582 (m), 1475 (w), 1422 (w), 1296(w), 1256 (m), 1210 (s), 1076 (s), 910 (w), 867 (w).

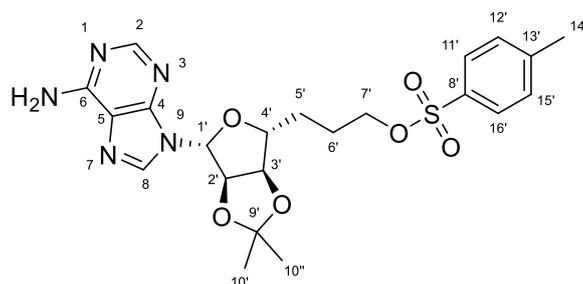
**mp**: 57 °C

**<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)** δ 8.31 (s, 1H, H2), 7.88 (s, 1H, H8), 6.25 (bs, 2H, C6-NH2), 6.03 (d, *J* = 2.4 Hz, 1H, H1'), 5.47 (dd, *J* = 6.5, 2.4 Hz, 1H, H2'), 4.83 (dd, *J* = 6.5, 3.7 Hz, 1H, H3'), 4.22 (td, *J* = 7.0, 3.7 Hz, 1H, H4'), 3.60 (t, *J* = 6.2 Hz, 2H, H7'), 1.82 – 1.72 (m, 2H, H5'), 1.69 – 1.59 (m, 2H, H6'), 1.59 (s, 3H, H9'), 1.37 (s, 3H, H10').

**<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>)** δ 155.7 (C5), 153.1 (C2), 149.5 (C4), 139.9 (C8), 120.3 (C6), 114.8 (C8'), 90.5 (C1'), 87.0 (C4'), 84.3 - 84.2 (C2', C3'), 62.2 (C7'), 30.1 (C5'), 28.9 (C6'), 27.3 (C9''), 25.55 (C9').

**HR-ESI-MS (M+Na<sup>+</sup>)** calc. for C<sub>15</sub>H<sub>21</sub>N<sub>5</sub>O<sub>4</sub>Na: 358.1491; found: 358.1497

**3-((3aR,4R,6R,6aR)-6-(6-amino-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)propyl 4-methylbenzenesulfonate (20)**



3-((3aR,4R,6R,6aR)-6-(6-amino-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)propan-1-ol (**19**) (1.0 equiv., 0.9035 mmol, 303 mg) was dissolved in dry pyridine (4 mL) and the solution was warmed to 50°C for 10 mins while stirring. Then, the mixture was cooled to 0°C and *p*-Toluenesulfonyl chloride (1.5 equiv., 1.36 mmol, 258 mg) was added by portion and the mixture was kept at 0°C for 10 hours. The solution was then diluted with EtOAc (100 mL), washed with 10% CuSO<sub>4</sub>.H<sub>2</sub>O (3 x 50 mL), saturated EDTA solution (3 x 50 mL), brine (50 mL), dried (Na<sub>2</sub>SO<sub>4</sub>) and concentrated under reduced pressure affording the tosylated compound **20** as white powder (410 mg, 0.8375 mmol, 93%) which is used in the next step without any further purification.

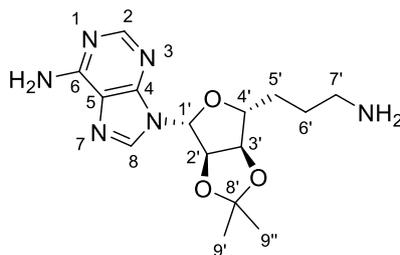
R<sub>f</sub> 0.25 (96:4 DCM/MeOH)

<sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 8.33 (s, 1H, H2), 7.88 (s, 1H, H8), 7.77 – 7.70 (m, 2H, H11', H16'), 7.34 – 7.28 (m, 2H, H12', H15'), 6.17 (s, 2H, C6-NH<sub>2</sub>), 5.99 (d, *J* = 2.5 Hz, 1H, H1'), 5.42 (dd, *J* = 6.5, 2.5 Hz, 1H, H2'), 4.80 (dd, *J* = 6.5, 4.0 Hz, 1H, H3'), 4.09 (ddd, *J* = 7.6, 5.9, 4.0 Hz, 1H, H4'), 4.05 – 3.92 (m, 2H, H7'), 2.42 (s, 3H, H14'), 1.74 (m, 4H, H5', H6'), 1.58 (s, 3H, H10''), 1.36 (s, 3H, H10').

<sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>) δ 155.3 (C5), 152.3 (C2), 149.4 (C4), 144.9 (C13'), 140.2 (C8), 133.2 (C8'), 130.0 (C12', C15'), 128.0 (C11', C16'), 120.4 (C6), 115.0 (C9'), 90.3 (C1'), 86.1 (C4'), 84.1 (C2', C3'), 69.9 (C7'), 29.4 (C5'), 27.3 (C10''), 25.5 (C10'), 25.3 (C6'), 21.8 (C14')

HR-ESI-MS (M+H<sup>+</sup>) calc. for C<sub>22</sub>H<sub>28</sub>N<sub>5</sub>O<sub>6</sub>S: 490.1760; found: 490.1763

**9-((3aR,4R,6R,6aR)-6-(3-aminopropyl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)-9H-purin-6-amine (13a)**



Adenosine azide **21** (1.0 equiv., 3.5 mmol, 1260 mg) was dissolved in 75 mL of EtOAc. A flow of argon is carried out before the addition of palladium on carbon (0.1 equiv., 0.350 mmol, 10% wt, 372 mg). A second flow of argon is carried out before putting the solution under an H<sub>2</sub> atmosphere. The reaction medium is stirred overnight at room temperature. The solution is filtered through celite, washed EtOAc (50 mL) and MeOH (50 mL) and concentrated under reduced pressure affording the title compound **13a** as a slightly yellow hygroscopic foam (1110 mg, 3.32 mmol, 95%).

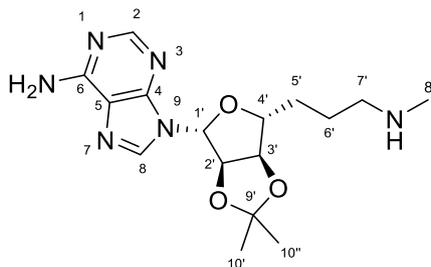
$\nu_{\max}/\text{cm}^{-1}$ : 3325 (br), 3152 (br), 2986 (br), 2933 (br), 1643 (s), 1596 (s), 1473 (m), 1421 (w) 1074 (s), 732 (s).

mp: 62 °C

**<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)**  $\delta$  8.33 (s, 1H, H2), 7.90 (s, 1H, H8), 6.03 (d,  $J = 2.5$  Hz, 1H, H1'), 5.97 (bs, 2H, probablement C6-NH<sub>2</sub>), 5.47 (dd,  $J = 6.5, 2.5$  Hz, 1H, H2'), 4.82 (dd,  $J = 6.5, 3.8$  Hz, 1H, H3'), 4.18 (td,  $J = 7.0, 3.8$  Hz, 1H, H4'), 2.67 (t,  $J = 7.0$  Hz, 2H, H7'), 1.76 – 1.64 (m, 2H, H5'), 1.59 (s, 3H, H9''), 1.53 (ddd,  $J = 14.6, 7.0, 3.3$  Hz, 2H, H6'), 1.37 (s, 3H, H9'). **<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>)**  $\delta$  155.6 (C6), 153.2 (C2), 149.4 (C4), 139.7 (C8), 120.3 (C5), 114.6 (C8'), 90.3 (C1'), 86.8 (C4'), 84.2 (C3'), 84.1 (C2'), 41.7 (C7'), 30.7 (C5'), 29.4 (C6'), 27.2 (C9''), 25.4 (C9')

**LR-ESI-MS (M+H<sup>+</sup>)** calc. for C<sub>15</sub>H<sub>23</sub>N<sub>6</sub>O<sub>3</sub>: 335.18; found: 335.18

**9-((3aR,4R,6R,6aR)-2,2-dimethyl-6-(3-(methylamino)propyl)tetrahydrofuro[3,4-d][1,3]dioxol-4-yl)-9H-purin-6-amine (13b)**



3-((3aR,4R,6R,6aR)-6-(6-amino-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)propyl 4-methylbenzenesulfonate **20** (1.0 equiv., 0.4085 mmol, 200 mg) was dissolved in dry THF (1 mL), MeNH<sub>2</sub> (20 equiv., 8.17 mmol, 2M in THF, 4.1 mL) was added and the mixture was stirred at 50 °C for 3 days. The mixture was concentrated under reduced pressure and purified by flash chromatography (82/18/0.01 DCM/MeOH/Et<sub>3</sub>N) affording the title compound **13b** as an oily foam (57 mg, 0.1636 mmol, 40%).

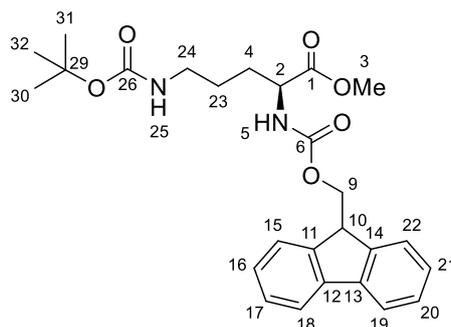
**R<sub>f</sub>** 0.05 (8:2 DCM/MeOH)

**<sup>1</sup>H NMR (400 MHz, MeOD)** δ 8.43 (s, 1H, H8), 8.38 (s, 1H, H2), 6.21 (d, *J* = 2.6 Hz, 1H, H1'), 5.42 (dd, *J* = 6.5, 2.6 Hz, 1H, H2'), 4.93 – 4.90 (m, 1H, H3'), 4.19 (td, *J* = 6.5, 4.1 Hz, 1H, H4'), 2.97 (t, *J* = 7.5 Hz, 2H, H7'), 2.65 (s, 3H), 1.87 – 1.69 (m, 4H, H5', H6'), 1.59 (s, 3H, H10''), 1.38 (s, 3H, H10').

**<sup>13</sup>C NMR (101 MHz, MeOD)** δ 153.70 (C6), 149.85 (C4), 148.02 (C2), 143.65 (C8), 120.78 (C5), 116.11 (C9'), 91.18 (H1'), 87.38 (H4'), 85.43 (C2'), 85.18 (C3'), 49.94 (C7'), 33.48 (C8'), 31.20 (C5'), 27.46 (C10''), 25.55 (C10'), 23.60 (C6').

**HR-ESI-MS (M+H<sup>+</sup>)** calc. for C<sub>16</sub>H<sub>25</sub>N<sub>6</sub>O<sub>3</sub>: 349.1988; found: 349.1996

**Methyl(S)-2-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-5-((tertbutoxycarbonyl)amino)pentanoate (22)**



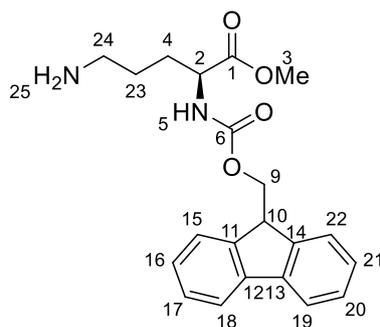
Fmoc-L-Ornithine(Boc)-OH **16** (1.0 equiv., 4.41 mmol, 2.0 g) was dissolved in a mixture of toluene-methanol (50 ml, 4:1, v/v). Thereafter, TMSCHN<sub>2</sub> 2M in hexanes (2.4 equiv., 10.58 mmol, 5.29 mL) was added dropwise. The reaction medium is left under stirring for 30 min. The solution was quenched with EtOAc (10 mL) and water (25 mL). The aqueous phase is extracted with EtOAc (3 x 50 mL). The combined organic layers were washed with brine (3 x 50 mL), dried (Na<sub>2</sub>SO<sub>4</sub>) and concentrated under reduced pressure affording the title compound **22** as white powder (2.0 g, 4.27 mmol, 97%).

**<sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>)** δ 7.76 (d, *J* = 7.6 Hz, 2H, Haromatique), 7.60 (dd, *J* = 7.6, 4.0 Hz, 2H, Haromatique), 7.44 – 7.37 (m, 2H, Haromatique), 7.32 (tt, *J* = 7.6, 1.4 Hz, 2H, Haromatique), 5.50 – 5.41 (m, 1H, H5), 4.58 (s, 1H, H25), 4.40 (dd, *J* = 15.3, 6.2 Hz, 3H, H2, H9), 4.22 (t, *J* = 6.9 Hz, 1H, H10), 3.75 (s, 3H, H3), 3.23 – 3.07 (m, 2H, H24), 1.88 (ddt, *J* = 15.3, 10.9, 5.5 Hz, 1H, H4'), 1.74 – 1.67 (m, 1H, H4''), 1.53 (dt, *J* = 13.4, 9.1 Hz, 2H, H23), 1.45 (s, 9H, H30, H31, H32)

**<sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>)** δ 172.9, 156.1, 144.0, 143.9, 141.5, 127.9, 127.2, 125.2, 120.1, 79.5, 67.1, 53.8, 52.6, 47.4, 40.1, 30.0, 28.5, 26.2.

**HR-ESI-MS (M+H<sup>+</sup>)** calc. for C<sub>26</sub>H<sub>32</sub>N<sub>2</sub>O<sub>6</sub>: 468.2260; found: 468.2265

**methyl (S)-2-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-5-aminopentanoate (23)**



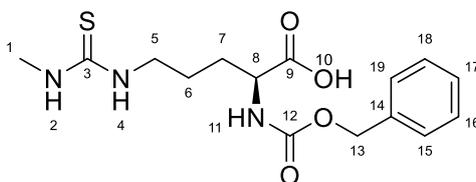
To a solution of methyl(S)-2-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-5-((tertbutoxy carbonyl)amino) pentanoate **22** (1.0 equiv., 0.487 mmol, 228 mg) in dry DCM (5 mL) was added TMSOTf (2.0 equiv., 176.1  $\mu$ L) at 0°C. The reaction mixture was stirred for 20 minutes at rt, then it was quenched with a solution of MeOH (2 mL) and concentrated below 30°C affording the title compound **23** as a white solid (179 mg, 0.487 mmol, Quant.) which was used in the next step without further purification.

**<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)**  $\delta$  7.78 (d,  $J$  = 7.5 Hz, 2H), 7.65 (t,  $J$  = 6.4 Hz, 2H), 7.38 (t,  $J$  = 7.5 Hz, 2H), 7.30 (tt,  $J$  = 7.5, 1.2 Hz, 2H), 4.43 – 4.27 (m, 2H), 4.23 – 4.15 (m, 2H), 3.35 (s, 3H), 2.94 (t,  $J$  = 7.1 Hz, 2H), 2.03 – 1.81 (m, 1H), 1.81 – 1.64 (m, 3H)

**<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>)**  $\delta$  173.9, 158.6, 145.1, 145.1, 142.5, 128.8, 128.1, 128.1, 126.2, 126.1, 123.3, 120.9, 120.1, 68.0, 54.8, 40.3, 40.2, 29.3, 25.0.

**LR-ESI-MS (M+H<sup>+</sup>)** calc. for C<sub>21</sub>H<sub>24</sub>N<sub>2</sub>O<sub>4</sub>: 368.17; found: 368.81

**(S)-2-(((benzyloxy)carbonyl)amino)-5-(3-methylthioureido)pentanoic acid (26b)**



A round bottom flask containing Z-Ornithine-OH **25** (1.0 equiv., 3.94 mmol, 1050 mg) and 1,1'-Thiocarbonyldiimidazole (1.0 equiv., 3.94 mmol, 703 mg) and a stirring bar was backfilled with argon. Then, DMF (50 mL) was added and the mixture was stirred for 16 hours. To the reaction medium, methylamine was added and stirred 20 min. The reaction was quenched with EtOAc (300 mL) and washed with HCl 1 M solution (50 mL x 3). Then, the organic layers were washed with saturated NaHCO<sub>3</sub> aqueous solution (50 mL x 3). The combined aqueous phases were acidified dropwise with cold 3 M HCl solution at 0°C, until we see a consistent white precipitate (pH~2). This latter was extracted with EtOAc (50 mL x 3) and concentrated under reduced pressure affording the title compound **26b** as a hygroscopic white foam (1204 mg, 3.55 mmol, 90%).

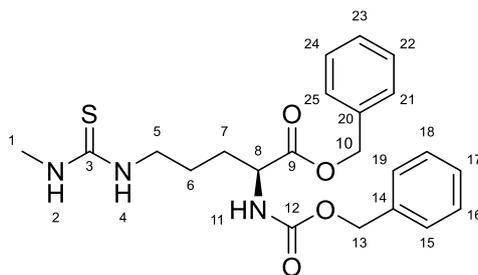
$\nu_{\max}/\text{cm}^{-1}$  : 3323 (br), 2929 (br), 2359 (w), 2359 (w), 1699 (s), 1558 (br, s), 1340 (m), 1259 (br,s), 1050 (m), 737 (w), 697 (w).

<sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>, 373 K)  $\delta$  7.41 – 7.19 (m, 5H, H15, H16, H17, H18, H19), 7.16 – 6.97 (m, 3H, H2, H4, H11), 5.06 (s, 2H, H13), 4.01 (td,  $J$  = 8.2, 5.1 Hz, 1H, H8), 3.38 (q,  $J$  = 6.3 Hz, 2H, H5), 2.86 (bs,  $J$  = 2.0 Hz, 3H, H1), 1.82 – 1.72 (m, 1H, H7), 1.69 – 1.54 (m, 3H, H6, H7).

<sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>)  $\delta$  183.0 (C3), 172.8 (C9), 155.4 (C12), 136.6 (C14), 127.7 – 126.9 (C15, C16, C17, C18, C19), 65.0 (C13), 53.5 (C8), 42.8 (C5), 29.9 (C1), 28.2 (C7), 25.1 (C6).

HR-ESI-MS ( $M+H^+$ ) calc. for C<sub>15</sub>H<sub>22</sub>N<sub>3</sub>O<sub>4</sub>S: 340.1331; found: 340.1326

**benzyl (S)-2-(((benzyloxy)carbonyl)amino)-5-(3-methylthioureido)pentanoate (14d)**



To the (S)-2-(((benzyloxy)carbonyl)amino)-5-(3-methylthioureido)pentanoic acid **26b** (1.0 equiv., 1.86 mmol, 630 mg), benzyl alcohol (2.0 equiv., 3.71 mmol, 384  $\mu$ L) was added under argon. In the wake, DMF (10 mL) and Et<sub>3</sub>N (3.0 equiv., 5.57 mmol, 776  $\mu$ L) were added. Subsequently, PyBOP (2.0 equiv., 3.71 mmol, 1932 mg) was added and the reaction mixture was stirred 40 hours at rt. Then, the reaction was quenched with EtOAc (300 mL) and washed with HCl 1M (50 mL x 3), Brine (50 mL x 5), dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated to dryness. The crude product was purified by flash chromatography (4% MeOH in DCM) affording the title compound **14d** as a foam (534 mg, 1.24 mmol, 67%).

**R<sub>f</sub>** 0.26 (96:4 DCM/MeOH)

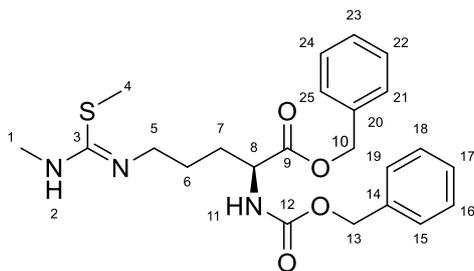
**v<sub>max</sub>/cm<sup>-1</sup>**: 3306 (br), 2945 (br), 1704 (s), 1556 (br, s), 1454 (m), 1259 (br, s), 1213 (br,s), 1050 (m), 739 (m), 697 (s).

**<sup>1</sup>H NMR (500 MHz, DMSO, 373K)**  $\delta$  7.37 – 7.27 (m, 10H, H15, H16, H17, H18, H19, H21, H22, H23, H24, H25), 7.16 – 6.99 (m, 2H, H2, H4), 5.14 (s, 2H, H10), 5.11 – 5.01 (m, 1H, H13), 4.20 – 4.11 (m, 1H, H8), 3.42 – 3.32 (m, 2H, H5), 2.86 (d, *J* = 4.7 Hz, 3H, H1), 1.85 – 1.53 (m, 4H, H6, H7).

**<sup>13</sup>C NMR (126 MHz, DMSO)**  $\delta$  183.0 (C3), 171.4 (C9), 155.4 (C12), 136.5 (C14), 135.5 (C20), 127.8 – 127.0 (C15, C16, C17, C18, C19, C21, C22, C23, C24, C25), 65.5 (C10), 65.2 (C13), 53.7 (C8), 42.7 (C5), 29.9 (C1), 28.0 (C7), 25.0 (C6).

**HR-ESI-MS (M+Na<sup>+</sup>)** calc. for C<sub>22</sub>H<sub>27</sub> N<sub>3</sub>O<sub>4</sub>SNa: 452.1625; found: 452.1619

**benzyl (S,Z)-2-(((benzyloxy)carbonyl)amino)-5 (((methylamino)(methylthio)methylene) amino) pentanoate (27b)**



To a solution of benzyl (S)-2-(((benzyloxy)carbonyl)amino)-5-(3-methylthioureido)pentanoate **14d** (1.0 equiv., 0.3958 mmol, 170 mg) solubilized in THF (5 mL), CH<sub>3</sub>-I (1.5 equiv., 0.5937mmol, 37  $\mu$ L) was added at 0°C and stirred for 2 hours and let smoothly warm to rt. After stirring 14 hours, the reaction was quenched with MeOH (2 mL) and concentrated under reduced pressure affording the title compound **27b** as a yellow oil (175 mg, 0.3958 mmol, Quant.) which is used in the next step without further purification.

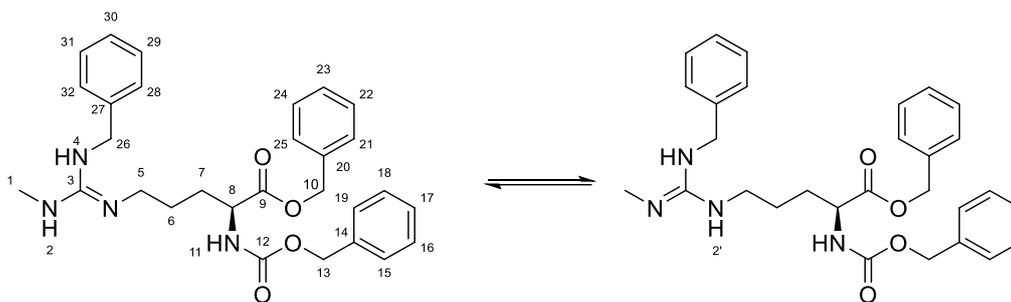
$\nu_{\max}/\text{cm}^{-1}$ : 3233 (br), 3032 (br), 2951(br), 1708 (s), 1609 (s), 1519 (m), 1454 (m), 1259 (br, s), 1211 (s), 1054 (m), 740 (m), 698 (m).

<sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>)  $\delta$  9.05 (s, 1H, H2), 7.85 (d,  $J$  = 7.8 Hz, 1H, H11), 7.41 – 7.24 (m, 10H, H15, H16, H17, H18, H19, H21, H22, H23, H24, H25), 5.18 – 5.08 (m, 2H, H10), 5.08 – 4.96 (m, 2H, H13), 4.16 – 4.08 (m, 1H, H8), 3.42 – 3.26 (m, 1H, H5), 2.91 (s, 3H, H1), 2.61 (s, 3H, H4), 1.86 – 1.74 (m, 1H, H7), 1.73 – 1.56 (m, 3H, H6, H7).

<sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>)  $\delta$  172.1 (C9), 165.7 (C3), 156.2 (C12), 136.9 (C14), 135.9 (C20), 128.4 – 127.3 (C15, C16, C17, C18, C19, C21, C22, C23, C24, C25), 66.0 (C10), 65.5 (C13), 53.7 (C8), 44.4 (C5), 31.0 (C1), 27.7 (C7), 25.6 (C6), 13.8 (C4).

**HR-ESI-MS (M+H<sup>+</sup>)** calc. for C<sub>23</sub>H<sub>30</sub>N<sub>3</sub>O<sub>4</sub>S: 444.1952; found: 444.1963

**benzyl (Z)-Nw-benzyl-N2-((benzyloxy)carbonyl)-Nw'-methyl-L-argininate (28)**



benzyl (S,Z)-2-(((benzyloxy)carbonyl)amino)-5(((methylamino)(methylthio)methylene) amino) pentanoate **27b** (1.0 equiv., 0.09324 mmol, 40 mg) benzylamine (1.1 equiv., 0.1026 mmol, 11.2  $\mu$ L) and  $\text{Et}_3\text{N}$  (1.0 equiv., 0.09324 mmol, 13  $\mu$ L) were dissolved in DMF (1 mL). The mixture was cooled in an ice-bath.  $\text{HgCl}_2$  (1.0 equiv., 0.09324 mmol, 25 mg) was added and the mixture was stirred for 20 min before it was warmed smoothly to rt and stirred 24 hours. The reaction medium was quenched with EtOAc (5 mL) and filtered through celite with additional EtOAc (5 mL) and MeOH. The filtrate was concentrated and the precipate formed was again filtered through celite and washed with EtOAc (5 mL), DCM (5 mL) and MeOH (5 mL). The filtrate was again concentrated to dryness and purified by flash chromatography (15% MeOH in DCM) to give the corresponding guanidine containing compound **28** (30 mg, 0.05969 mmol, 64%).

**R<sub>f</sub>** 0.29 (9:1 DCM/MeOH)

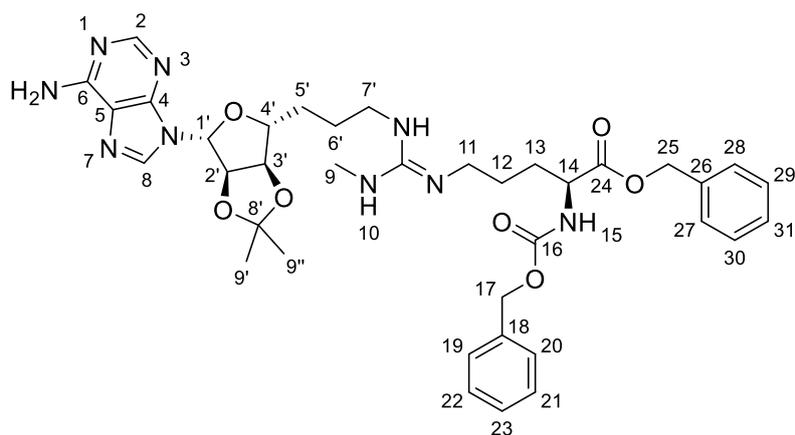
**v<sub>max</sub>/cm<sup>-1</sup>**: 3205 (br), 2928 (br), 1740 (s), 1626 (s), 1456 (w), 1370 (w), 1348 (w), 1246 (m), 1153 (m), 745 (w), 697 (m).

**<sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>)**  $\delta$  8.16 (t,  $J$  = 6.3 Hz, 1H, H4), 7.85 (d,  $J$  = 7.7 Hz, 1H, H11), 7.80 (q,  $J$  = 4.7 Hz, 1H, H2), 7.63 (t,  $J$  = 5.8 Hz, 1H, H2'), 7.41 – 7.22 (m, 15H, H15, H16, H17, H18, H19, H21, H22, H23, H24, H25, H28, H29, H30, H31, H32), 5.15 – 5.07 (m, 2H, H10), 5.07 – 4.97 (m, 2H, H13), 4.47 (d,  $J$  = 5.8 Hz, 2H, H26), 4.11 – 4.03 (m, 1H, H8), 3.16 (d,  $J$  = 4.7 Hz, 1H, H5), 2.76 (d,  $J$  = 4.0 Hz, 3H, H1), 1.70 (bs, 1H, H7), 1.64 – 1.46 (m, 3H, H6, H7).

**<sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>)**  $\delta$  172.1 (C9), 156.2 (C12), 154.6 (C3), 137.6 (C27), 136.9 (C14), 135.9 (C20), 128.7 – 127.0 (C15, C16, C17, C18, C19, C21, C22, C23, C24, C25, C28, C29, C30, C31, C32), 66.0 (C10), 65.6 (C13), 53.8 (C8), 43.7 (C26), 40.5 (C5), 28.3 (C1), 27.6 (C7), 25.2 (C6).

**HR-ESI-MS (M+H<sup>+</sup>)** calc. for  $\text{C}_{29}\text{H}_{35}\text{N}_4\text{O}_4$ : 503.2680; found: 503.2676

**benzyl (S)-5-(((Z)-((3-((3aR,4R,6R,6aR)-6-(6-amino-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl) propyl)amino)(methylamino)methylene)amino)-2-(((benzyloxy)carbonyl)amino)pentanoate (12e)**



benzyl (S,Z)-2-(((benzyloxy)carbonyl)amino)-5(((methylamino)(methylthio)methylene) amino) pentanoate **27b** (1.0 equiv., 0.5974 mmol, 265 mg), the corresponding primary amine (1.0 equiv., 0.5974 mmol, 200 mg) and Et<sub>3</sub>N (2.0 equiv., 1.19 mmol, 167  $\mu$ L) were dissolved in DMF (5 mL). HgCl<sub>2</sub> (1.0 equiv., 0.5974 mmol, 162 mg) was added to this solution with vigorous magnetic stirring and ice-bath cooling and left for 24 hours at room temperature. The reaction was quenched with EtOAc (10 mL) and the suspension was filtered through a pad of celite and washed with additional EtOAc (20 mL) and MeOH (30 mL). The filtrate was concentrated, diluted with MeOH (20 mL) and EtOAc (20 mL) and filtered through a pad of celite. The filtrate was concentrated under reduced pressure and purified by flash chromatography (15% MeOH in DCM) affording the title compound **12e** (120 mg, 0.1644 mmol, 27%).

R<sub>f</sub> 0.10 (95:15 DCM/MeOH)

$\nu_{\text{max}}/\text{cm}^{-1}$ : 3188 (br), 2952 (w), 1709 (s), 1633 (s), 1210 (s), 1075 (s), 1026 (w), 733 (w).

mp: 84°C

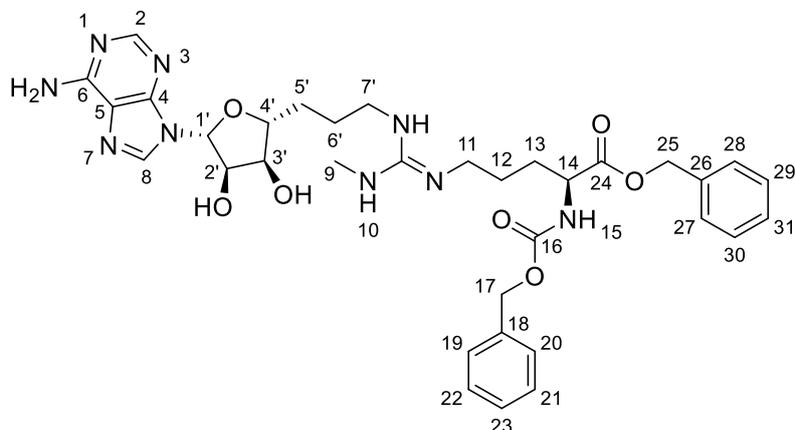
<sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>)  $\delta$  8.26 (s, 1H, H2), 8.08 (br, 1H, H8), 7.51 – 7.20 (m, 11H, H19, H20, H21, H22, H23, H27, H28, H29, H30, H31, H10), 7.14 (s, 1H, H10' ou H10''), 6.86 (s, 1H, H10' ou H10''), 6.63 (d, *J* = 7.8 Hz, 1H, H15), 6.55 (s, 2H, C6-NH<sub>2</sub>), 6.03 (t, *J* = 3.1 Hz, 1H, H1'), 5.43 (dd, *J* = 6.4, 2.3 Hz, 1H, H2'), 5.18 – 4.96 (m, 4H, H17, H25), 4.87 (dd, *J* = 6.5, 3.3 Hz, 1H, H3'), 4.30 (d, *J* = 11.1 Hz, 1H, H14), 4.15 (q, *J* = 5.6 Hz, 1H, H4'), 3.22 (s, 4H, H11', H7'), 2.76 (d, *J* = 4.6 Hz, 3H, H9), 1.90 – 1.50 (m, 11H, H5', H6', H12, H13, H9' ou H9''), 1.33 (s, 3H, H9' ou H9').

<sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>)  $\delta$  172.3 (C24), 156.8 (C16), 155.6 (C6), 155.1 (Cguanidine), 152.9 (C2), 149.3 (C4), 140.0 (C8), 136.1 (C18), 135.2 (C26), 128.6 – 127.9 (C19, C20, C21, C22, C23, C27, C28, C29, C30, C31), 119.7 (C5), 114.4 (C8'), 90.1 (C1'), 86.8 (C4'), 84.0 (C3'),

83.9 (C2'), 67.3 (C25), 67.0 (C17), 53.4 (C14), 41.7 – 41.3 (C11 + C7'), 30.3 (C5'), 29.7 (C13), 28.7 (C9), 27.1 (H9'), 25.4 (H9''), 25.2 – 25.0 (C6' + C12) ;

**HR-ESI-MS (M+H<sup>+</sup>)** calc. for C<sub>37</sub>H<sub>48</sub>N<sub>9</sub>O<sub>7</sub>: 730.3671; found: 730.3676

**benzyl(S)-5-(((Z)-((3-((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)propyl)amino)(methylamino)methylene)amino)-2-(((benzyloxy)carbonyl)amino)pentanoate (34b)**



To the protected adenosine **12e** (0.01370 mmol, 10 mg) a cocktail of TFA/H<sub>2</sub>O/TIPS (95:2.5:2.5 v/v, 1 mL) was added. The mixture was stirred for 1 hours at rt and quenched with MeCN (2 mL) and H<sub>2</sub>O (2 mL) and purified by preparative HPLC with a linear gradient of MeCN (0 – 100%) in 0.1 % aq. TFA as the mobile phase, affording the title compound **34b** (0.006440 mmol, 4.44 mg, 47%) as oil after being lyophilized.

**tr** = 10.20 min

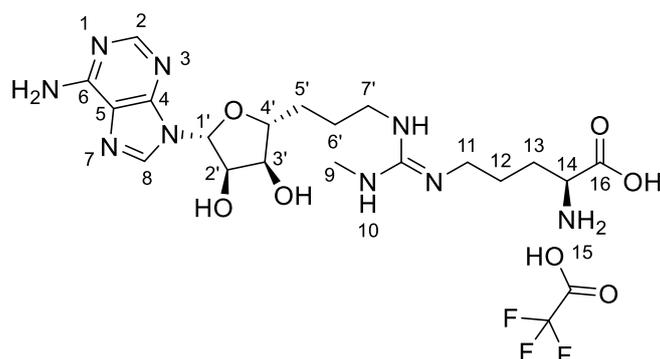
**v<sub>max</sub>/cm<sup>-1</sup>**: 3307 (br), 2359 (w), 2341 (w), 1683 (s), 1636 (s), 1203 (s), 1051 (s), 1026 (w), 1005 (w), 829 (w), 720 (w).

**<sup>1</sup>H NMR (700 MHz, DMSO-d<sub>6</sub>)** δ 8.43 (s, 1H, H8), 8.28 (s, 1H, H2), 8.06 (s, 2H), 7.79 (d, *J* = 7.9 Hz, 1H, H15), 7.47 – 7.17 (m, 14H), 5.87 (d, *J* = 5.1 Hz, 1H, H1'), 5.15 – 5.08 (m, 2H, H25), 5.07 – 4.96 (m, 2H, H17), 4.64 (t, *J* = 5.2 Hz, 1H, H2'), 4.10 (td, *J* = 8.7, 4.8 Hz, 1H, H14), 4.05 (t, *J* = 5.0 Hz, 1H, H3'), 3.87 (dt, *J* = 9.4, 5.0 Hz, 1H, H4'), 3.14 – 3.03 (m, 4H, C7', C11), 2.70 (d, *J* = 4.6 Hz, 3H, C9'), 1.77 – 1.45 (m, 8H, C5', C6', C12, C13).

**<sup>13</sup>C NMR (176 MHz, DMSO-d<sub>6</sub>)** δ 172.2 (C24), 156.3 (C16), 154.5 (gua), 154.1 (C6), 149.4 (C2), 149.1 (C4), 141.0 (C8), 136.9 (C18), 135.9 (C26), 128.7 – 127.7 (C19, C20, C21, C22, C23, C27, C28, C29, C30, C31), 119.1 (C5), 87.8 (C1'), 83.5 (C4'), 73.4 (C3'), 73.2 (C2'), 66.1 (C26), 65.7 (C18), 53.7 (C14), 40.9 – 40.4 (C11, C7'), 30.2 (C5'), 28.1 (C9), 27.8 (C13), 25.1 (C12, C6').

**HR-ESI-MS (M+H<sup>+</sup>)** calc. for C<sub>34</sub>H<sub>44</sub>N<sub>9</sub>O<sub>7</sub>: 690.3358; found: 690.3361

**(S)-2-amino-5-(((Z)-((3-((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)propyl)amino)(methylamino)methylene)amino)pentanoic acid (33)**



To the protected adenosine (0.01582 mmol, 8 mg) a cocktail of TFA/H<sub>2</sub>O/TIPS (95:2.5:2.5 v/v, 1 mL) was added. The mixture was stirred for 1 hours at rt and quenched with MeCN (3 mL) and H<sub>2</sub>O (3 mL) and purified by preparative HPLC with a linear gradient of MeCN (0 – 100%) in 0.1 % aq. TFA as the mobile phase, affording the title compound **33** (0.01076 mmol, 6.23 mg, 68%) as oil after being lyophilized.

**tr** = 1.82 min

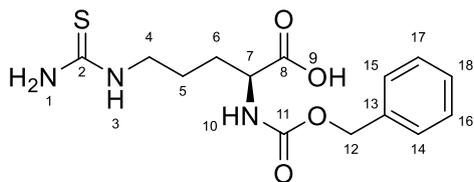
**v<sub>max</sub>/cm<sup>-1</sup>**: 3307 (br,s), 2919 (br), 1677 (s), 1635 (s), 1202 (s), 1134 (s), 1026 (w), 722 (w).

**<sup>1</sup>H NMR (700 MHz, DMSO-d<sub>6</sub>)** δ 8.38 (d, *J* = 16.0 Hz, 3H, H8, + H15), 8.23 (s, 1H, H2), 7.78 (s, 2H, C6-NH<sub>2</sub>), 7.55 – 7.37 (m, 3H, H10, C11-NH, C7'-NH), 5.86 (d, *J* = 5.2 Hz, 1H, H1'), 4.65 (t, *J* = 5.2 Hz, 1H, H2'), 4.06 (t, *J* = 5.2 Hz, 1H, H3'), 3.93 – 3.84 (m, 2H, H4', H14), 3.22 – 3.08 (m, 2H, C7', C11), 2.73 (d, *J* = 4.5 Hz, 3H, H9), 1.85 – 1.49 (m, 8H, C5', C6', C12, C13).

**<sup>13</sup>C NMR (176 MHz, DMSO-d<sub>6</sub>)** δ 171.0 (C16), 154.9 (C6), 154.5 (gua), 151.0 (C2), 149.2 (C4), 140.6 (C8), 119.2 (C5), 87.8 (C1'), 83.4 (C4'), 73.3 (C3'), 73.1 (C4'), 51.7 (C14), 40.9 – 40.2 (C7', C11), 30.2 (C5'), 28.1, 27.2, 25.1, 24.3.

**HR-ESI-MS (M+H<sup>+</sup>)** calc. for C<sub>19</sub>H<sub>32</sub>N<sub>9</sub>O<sub>5</sub>: 466.2526; found: 466.2518

### (S)-2-(((benzyloxy)carbonyl)amino)-5-thioureidopentanoic acid (**26a**)



A round bottom flask containing Z-Ornithine-OH **25** (1.0 equiv., 1.13 mmol, 300 mg) and 1,1'-Thiocarbonyldiimidazole (1.05 equiv., 1.18 mmol, 211 mg) and a stirring bar was backfilled with argon. Then, DMF (5 mL) was added and the mixture was stirred for 16 hours. To the reaction medium, NH<sub>3</sub> (18.70 equiv., 7N in MeOH, 3 mL) was added and stirred 2 days. The reaction was quenched with EtOAc (100 mL), washed with HCl 1 M solution (10 mL x 3), brine (20 mL x 3) dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated under reduced pressure affording the title compound **26a** as a hygroscopic white foam (274 mg, 3.55 mmol, 75%).

R<sub>f</sub> 0.26 (8:2 DCM/MeOH)

$\nu_{\text{max}}/\text{cm}^{-1}$ : 3308 (br, s), 2945 (w), 1698 (s), 1633 (m), 1537 (s), 1454 (w), 1345 (w), 1243 (m), 1054 (w), 739 (w), 697 (w).

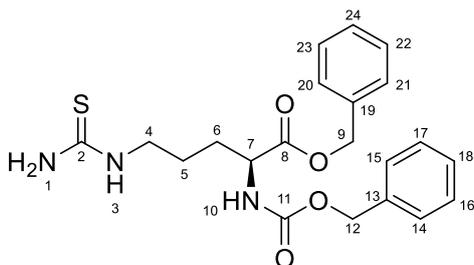
mp: 56 °C

<sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>)  $\delta$  7.44 – 7.26 (m, 5H, H14, H15, H16, H17, H18, H3), 7.14 – 6.95 (s, 1H, H10), 6.68 (s, 1.6 H, H1), 5.06 (s, 2H, H12), 4.05 – 3.97 (m, 1H, H7), 3.35 – 3.25 (m, 2H, H4), 1.83 – 1.70 (m, 1H, H6), 1.70 – 1.42 (m, 3H, H6, H5).

<sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>)  $\delta$  182.9 (C2), 172.8 (C8), 155.4 (C11), 136.6 (C13), 127.7 – 126.9 (C14, C15, C16, C17, C18), 65.0 (C12), 53.5 (C7), 43.0 (C4), 28.2 (C6), 24.9 (C5).

HR-ESI-MS (M+H<sup>+</sup>) calc. for C<sub>14</sub>H<sub>20</sub>N<sub>3</sub>O<sub>4</sub>S: 326.1169; found: 326.1165

### benzyl (S)-2-(((benzyloxy)carbonyl)amino)-5-thioureidopentanoate (**14c**)



To a solution of (S)-2-(((benzyloxy)carbonyl)amino)-5-thioureidopentanoic acid **26a** (1.0 equiv., 0.4610 mmol, 150 mg), benzyl alcohol (2.0 equiv., 0.9220 mmol, 95.9  $\mu$ L) and Et<sub>3</sub>N (3.0

equiv., 1.38 mmol, 193  $\mu$ L) in DMF (1.5 mL) was added PyBOP (2.0 equiv., 0.9220 mmol, 479.8 mg) and the mixture was stirred at room temperature for 16 hours. The reaction was then diluted with EtOAc (50 mL) before washing with HCl 1M (3 x 10 mL) and brine (3 x 25 mL). The combined layers were dried ( $\text{Na}_2\text{SO}_4$ ), filtered and concentrated under reduced pressure. The crude product was purified by flash chromatography (4% MeOH in DCM) affording the title compound **14c** (99 mg, 0.2397 mmol, 52%) as a slightly yellow foam

$R_f$  0.20 (96:4 DCM / MeOH)

$\nu_{\text{max}}/\text{cm}^{-1}$ : 3383 (br), 3315 (br), 3220 (br), 3034 (w), 2946 (w), 1698 (s), 1546 (m), 1348(m), 989 (s), 642 (s).

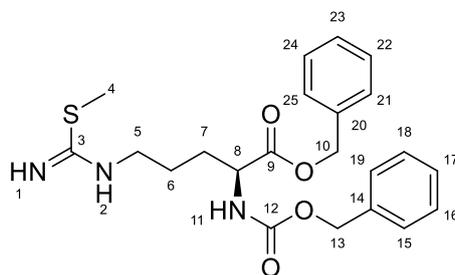
mp: 117  $^{\circ}\text{C}$

$^1\text{H NMR}$  (500 MHz,  $\text{DMSO-d}_6$ , 373K)  $\delta$  7.40 – 7.27 (m, 12H, H14, H15, H16, H17, H18, H20, H22, H22, H23, H24, H3, H10), 6.66 (bs, 1.34H, H1), 5.14 (s, 2H, H9), 5.11 – 5.00 (m, 2H, H12), 4.19 – 4.11 (m, 1H, H7), 3.30 (bs, 2H, H4), 1.87 – 1.75 (m, 1H, H6), 1.75 – 1.64 (m, 1H, H6), 1.63 – 1.51 (m, 2H, H5).

$^{13}\text{C NMR}$  (126 MHz,  $\text{DMSO-d}_6$ )  $\delta$  171.3 (C8), 155.5 (C11), 136.5 (C13), 135.5 (C19), 127.8 – 126.9 (C14, C15, C16, C17, C18, C20, C21, C22, C23, C24), 65.4 (C9), 65.1 (C12), 53.7 (C7), 42.9 (C4), 27.9 (C6), 24.8 (C5).

HR-ESI-MS ( $\text{M}+\text{Na}^+$ ) calc. for  $\text{C}_{21}\text{H}_{25}\text{N}_3\text{O}_4\text{SNa}$ : 438.1463; found: 438.1463

### benzyl (S)-2-(((benzyloxy)carbonyl)amino)-5-((imino(methylthio)methyl)amino) pentanoate (**27a**)



To a solution benzyl (S)-2-(((benzyloxy)carbonyl)amino)-5-thioureidopentanoate **14c** (1.0 equiv., 0.9627 mmol, 400 mg) solubilized in THF (15 mL),  $\text{CH}_3\text{I}$  (1.5 equiv., 1.44 mmol, 89.9  $\mu$ L) was added at  $0^{\circ}\text{C}$  and stirred for 2 hours and let smoothly warm to rt. After stirring 14 hours, the reaction was quenched with MeOH (5 mL) and concentrated under reduced pressure affording the title compound **27a** as a yellow oil (413 mg, 0.9627 mmol, Quant.) which is used in the next step without further purification.

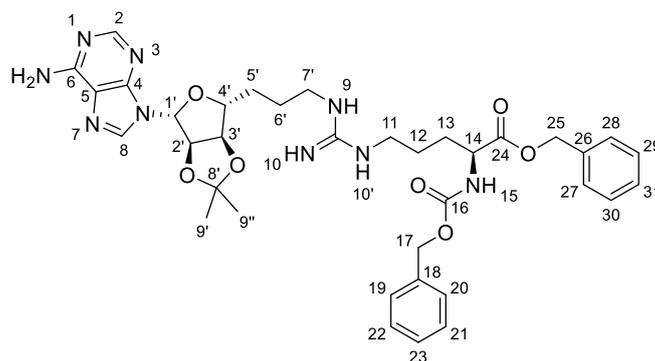
$\nu_{\text{max}}/\text{cm}^{-1}$ : 3272 (br), 3138 (br), 3033 (br), 1704 (s), 1639 (s), 1593 (m), 1520 (m), 1454 (w), 1260 (m), 1213 (s), 1059 (m), 737 (s), 698 (s).

**<sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>)** δ 9.40 (t, *J* = 5.6 Hz, 1H, H2), 9.05 (s, 1.5H, H1), 7.80 (d, *J* = 7.8 Hz, 1H, H11), 7.42 – 7.23 (m, 10H, H15, H16, H17, H18, H19, H21, H22, H23, H24, H25), 5.18 – 5.09 (m, 2H, H10), 5.08 – 4.98 (m, 2H, H13), 4.15 – 4.08 (m, 1H, H8), 3.34 – 3.22 (m, 2H, H5), 2.59 (s, 3H, H4), 1.83 – 1.73 (m, 1H, H7), 1.69 – 1.54 (m, 3H, H6, H7).

**<sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>)** δ 172.0 (C9), 167.4 (C3), 156.2 (C12), 136.8 (C20), 135.8 (C14), 128.4 – 127.8 (C15, C16, C17, C18, C19, C21, C22, C23, C24, C25), 66.0 (C10), 65.6 (C13), 53.6 (C8), 43.1 (H5), 27.7 (C7), 24.1 (C6), 13.6 (C4).

**HR-ESI-MS (M+H<sup>+</sup>)** calc. for C<sub>22</sub>H<sub>28</sub>N<sub>3</sub>O<sub>4</sub>S: 430.1795; found: 430.1807

**benzyl N<sup>w</sup>-(3-((3aR,4R,6R,6aR)-6-(6-amino-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)propyl)-N<sup>2</sup>-((benzyloxy)carbonyl)-L-argininate (12d)**



The benzyl (S)-2-(((benzyloxy)carbonyl)amino)-5-((imino(methylthio)methyl)amino) pentanoate **27a** (1.0 equiv., 0.7473 mmol, 321 mg), the corresponding primary amine (1.0 equiv., 250 mg, 0.7473 mmol) and Et<sub>3</sub>N (2.0 equiv., 1.44 mmol, 208 μL) were dissolved in DMF (5 mL). The mixture was cooled in an ice-bath. HgCl<sub>2</sub> (1.0 equiv., 0.7473 mmol, 203 mg) was added and the mixture was stirred for 20 min before it was warmed smoothly to rt and stirred 24 hours. The reaction was quenched with EtOAc (20 mL) and the suspension was filtered through a pad of celite and washed with additional EtOAc (30 mL) and MeOH (50 mL). The filtrate was concentrated, diluted with MeOH (30 mL) and EtOAc (30 mL) and filtered again through a pad of celite. The filtrate was concentrated under reduced pressure and purified by flash chromatography (18 %MeOH in DCM) concentrated under reduced pressure affording the title compound **12d** (96mg, 0.1345 mmol, 18%) as a foam.

**R<sub>f</sub>** 0.12 (DCM/MeOH 80:20)

**v<sub>max</sub>/cm<sup>-1</sup>**: 3308 (br), 3183 (br), 2937 (br), 1708 (s), 1633 (s), 1600 (w), 1246 (w), 1210 (w), 1026 (w), 735 (w), 698 (w).

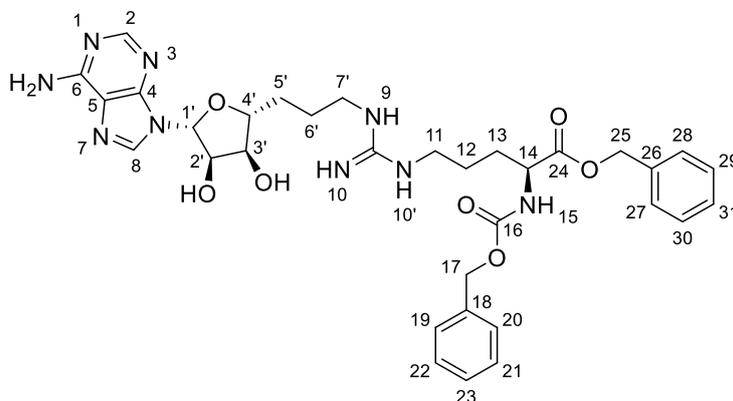
**mp**: 103 °C

**<sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>)** δ 8.29 (s, 1H, H2), 8.19 (s, 1H, H8), 7.34 – 7.16 (m, 10H, H19, H20, H21, H22, H23, H27, H28, H29, H30, H31), 6.04 (s, 1H, H1'), 5.41 – 5.34 (m, 1H, H2'), 5.19 – 4.96 (m, 4H, H17, H25), 4.91 – 4.85 (m, 1H, H3'), 4.36 – 4.32 (m, 1H, H14), 4.19 – 4.08 (m, 1H, H4'), 3.37 – 2.98 (m, 4H, H11, H7'), 1.92 – 1.42 (m, 11H, H5', H6', H12, H13, H9''), 1.32 (s, 1H, H9').

**<sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>)** δ 172.3 (C24), 156.9 (C16), 155.8 (C6, Cguanidine), 153.3 (C2), 149.4 (C4), 140.2 (C8), 136.1 (C18), 135.1 (C26), 128.8 – 127.9 (C19, C20, C21, C22, C23, C27, C28, C29, C30, C31), 119.7 (C5), 114.7 (C8'), 89.9 (C1'), 86.5 (C4'), 84.1 – 83.9 (C2', C3'), 67.6 (C25), 67.3 (C17), 53.6, 41.8, 41.4, 30.3 (C5'), 29.9 (C13), 27.2 (C9''), 25.5 (C9'), 25.0 (C6' + C12).

**HR-ESI-MS (M+H<sup>+</sup>)** calc. for C<sub>36</sub>H<sub>46</sub>N<sub>9</sub>O<sub>7</sub>: 716.3475; found: 716.3520

**benzyl N<sup>w</sup>-(3-((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)propyl)-N<sup>2</sup>-((benzyloxy)carbonyl)-L-argininate (34a)**



To benzyl N<sup>w</sup>-(3-((3aR,4R,6R,6aR)-6-(6-amino-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro [3,4-d][1,3]dioxol-4-yl)propyl)-N<sup>2</sup>-((benzyloxy)carbonyl)-L-argininate **12d** (0.01397 mmol, 10 mg) solubilized in MeOH (1 mL) a cocktail of TFA/H<sub>2</sub>O/TIPS (95:2.5:2.5 v/v, 1 mL) was added. The mixture was stirred for 4 hours at rt and quenched with H<sub>2</sub>O (5 mL) and MeCN (2 mL) and purified by preparative HPLC with a linear gradient of MeCN (0 – 100%) in 0.1 % aq. TFA as the mobile phase, affording the title compound **34a** (0.01062 mmol, 7.17 mg, 76%) as oil after being lyophilized.

**tr** = 10.78 min

**v<sub>max</sub>/cm<sup>-1</sup>**: 3336 (br), 2947 (br), 1674 (s), 1636 (s), 1456 (m), 1426(w), 1338 (w), 1201(w), 1134 (w), 721(m), 697(m).

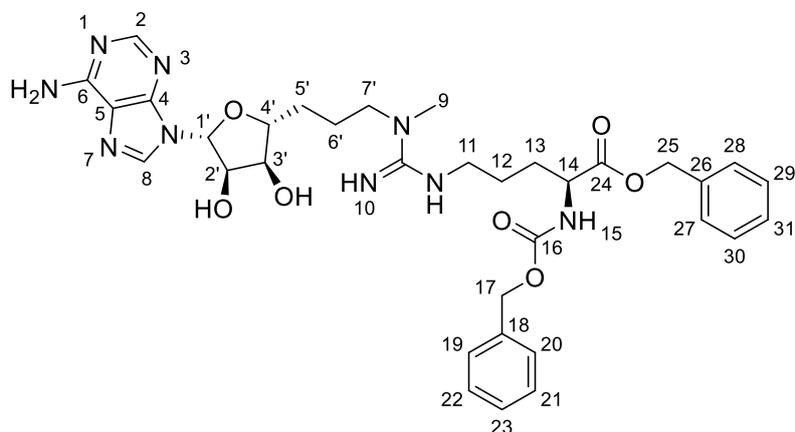
**<sup>1</sup>H NMR (500 MHz, MeOD)** δ 8.32 (s, 1H, H8), 8.29 (s, 1H, H2), 7.41 – 7.16 (m, 10H, H19, H20, H21, H22, H23, H27, H28, H29, H30, H31), 5.98 (d, *J* = 4.4 Hz, 1H, H1'), 5.21 – 5.10 (m,

2H, H25), 5.10 – 5.04 (m, 2H, H17), 4.73 – 4.67 (m, 1H, H2'), 4.26 (dd,  $J = 9.1, 5.0$  Hz, 1H, H14), 4.19 (t,  $J = 5.5$  Hz, 1H, H3'), 4.06 – 3.99 (m, 1H, H4'), 3.26 – 3.13 (m, 4H, H7', H11), 1.91 – 1.57 (m, 8H, H5', H6', H12, H13).

$^{13}\text{C}$  NMR (126 MHz, MeOD)  $\delta$  173.4 (C24), 158.8 (C16), 157.4 (Cgua), 154.9 (C6) 150.3 (C4), 149.7 (C2), 142.8 (C8), 138.1 (C18) 137.2 (C26), 129.6 – 128.8 (C19, C20, C21, C22, C23, C27, C28, C29, C30, C31), 120.7 (C5), 90.7 (C1'), 85.1 (C4'), 75.2 (C2'), 75.0 (C3'), 68.0 (C25), 67.8 (C17), 55.0 (C14), 42.3 – 41.8 (C7', C11), 31.5 (C5'), 29.7 (C13), 26.5 – 26.3 (C6', C12).

HR-ESI-MS ( $\text{M}+\text{H}^+$ ) calc. for  $\text{C}_{33}\text{H}_{42}\text{N}_9\text{O}_7$ : 676.3202; found: 676.3205

**benzyl  $\text{N}^w$ -(3-((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)propyl)- $\text{N}^2$ -((benzyloxy)carbonyl)- $\text{N}^w$ -methyl-L-argininate (34c)**



The benzyl (S)-2-(((benzyloxy)carbonyl)amino)-5-((imino(methylthio)methyl)amino) pentano - ate **27a** (1.0 equiv., 0.1196 mmol, 51.4 mg), the corresponding secondary amine (1.2 equiv., 0.1435 mmol, 50 mg) and  $\text{Et}_3\text{N}$  (2.0 equiv., 0.2392 mmol, 33.3  $\mu\text{L}$ ) were dissolved in DMF (2 mL). The mixture was cooled in an ice-bath.  $\text{HgCl}_2$  (2.0 equiv., 0.2392 mmol, 64.9 mg) was added and the mixture was stirred for 20 min before it was warmed smoothly to rt and stirred 20 hours and warmed to  $50^\circ\text{C}$  and stirred 2 hours. The reaction was quenched with EtOAc (5 mL) and the suspension was filtered through a pad of celite and washed with additional EtOAc (20 mL) and MeOH (50 mL). The filtrate was concentrated, diluted with MeOH (20 mL) and EtOAc (20 mL) and filtered again through a pad of celite. The filtrate was again concentrated to dryness, diluted with EtOAc (5 mL), DCM (5mL) and MeOH (5 mL) and was left 2 hours at  $0^\circ\text{C}$  and filtered through a pad of celite and washed with additional cold MeOH (10 mL). The filtrate was concentrated to dryness and a cocktail of TFA/ $\text{H}_2\text{O}$ /TIPS (95:2.5:2.5 v/v, 1 mL) was added. The mixture was stirred for 1 hours at rt and quenched with  $\text{H}_2\text{O}$  (6 mL) and MeCN (8 mL) and purified by preparative HPLC with a linear gradient of MeCN (0 – 100%) in 0.1 % aq.

TFA as the mobile phase, affording the title compound **34c** (13.9 mg, 0.02015 mmol, 14%) as oil after being lyophilized.

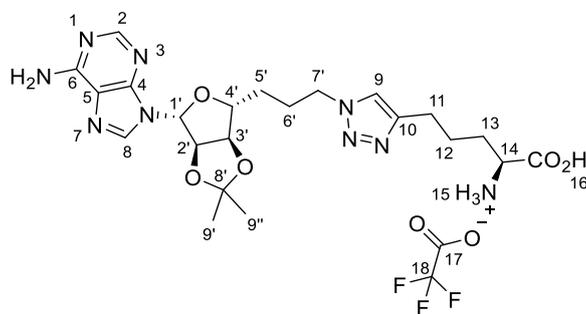
**tr** = 11.5 min.

**<sup>1</sup>H NMR (500 MHz, MeOD)** δ 8.38 (s, 1H, H8), 8.34 (s, 1H, H2), 7.37 – 7.24 (m, 10H), 6.01 (d, *J* = 4.3 Hz, 1H, H1'), 5.20 – 5.01 (m, 4H, H17, H25), 4.70 – 4.64 (m, 1H, H2'), 4.29 – 4.23 (m, 1H, H14), 4.20 (t, *J* = 5.5 Hz, 1H, H3'), 4.07 – 4.00 (m, 1H, H4'), 3.40 – 3.34 (m, 2H, H7'), 3.26 – 3.19 (m, 2H, H11), 2.97 (s, 3H, H9), 1.95 – 1.57 (m, 8H, H5', H6', H12, H13).

**<sup>13</sup>C NMR (126 MHz, MeOD)** δ 173.6 (C24), 158.7 (C16), 157.1 (Cgua), 153.5 (C6), 150.1 (C4), 147.7 (C2), 143.4 (C8), 138.1 (C18), 137.1 (C26), 129.6 – 128.8 (C18, C19, C20, C21, C22, C23, C27, C28, C29, C30, C31), 120.7 (C5), 90.8 (C1'), 85.1 (C4'), 75.3 (C2'), 75.0 (C3'), 68.0 (C25), 67.8 (C27), 55.1 (C14), 51.4 (C7'), 42.7 (C11), 36.8 (C9), 31.1 (C5'), 29.7 (C13), 26.3 (C12), 24.6 (C6').

**HR-ESI-MS (M+H<sup>+</sup>)** calc. for C<sub>34</sub>H<sub>44</sub>N<sub>9</sub>O<sub>7</sub>: 690.3364; found: 690.3383

**(S)-4-(1-(3-((3aR,4R,6R,6aR)-6-(6-amino-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)propyl)-1H-1,2,3-triazol-4-yl)-1-carboxybutan-1-aminium (46)**



The corresponding commercial alkyne **45** (1.0 equiv., 0.08325 mmol, 11.8 mg) solubilized in 5 mL of DMSO/H<sub>2</sub>O<sub>(milliQ)</sub> (4:1) and the corresponding azide (1.0 equiv., 0.08325 mmol, 30 mg) solubilized in DMSO (200 μL) was mixed together. Then, a premixed solution of THPTA (1.0 equiv., 0.08325 mmol, 44.2 mg, in 300 μL of H<sub>2</sub>O<sub>(milliQ)</sub>) and CuSO<sub>4</sub>·5H<sub>2</sub>O (0.5 equiv., 0.04162 mmol, 10.4 mg, in 200 μL of H<sub>2</sub>O<sub>(milliQ)</sub>) was added to the first solution, followed by freshly prepared sodium ascorbate (3.0 equiv., 0.2497 mmol, 49.5 mg) solution in H<sub>2</sub>O (200 μL, milliQ). The reaction was stirred 1 hour at rt and diluted with water (7 mL) and MeCN (3 mL) and purified by preparative HPLC with a linear gradient of MeCN (0 – 100%) in 0.1 % aq. TFA as the mobile phase, giving the clicked product **46** (36.9 mg, 0.05994 mmol, 72%) as white solide after being lyophilized.

**tr** = 8.38 min

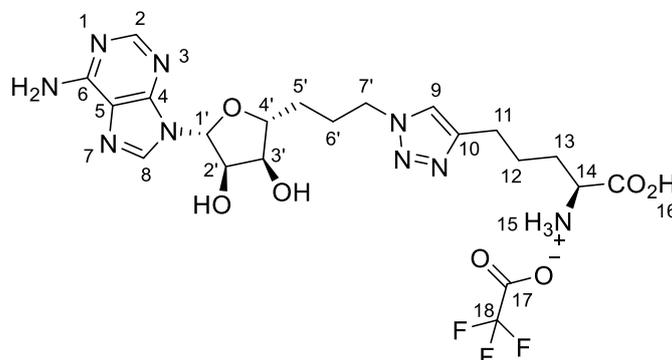
**v<sub>max</sub>/cm<sup>-1</sup>**: 2940 (br), 1681(s), 1457(w), 1200 (s), 1135 (s), 1098 (w), 1026 (w), 799 (w), 722 (w).

**<sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>)** δ 8.57 (bs, 2H, C6-NH<sub>2</sub>), 8.51 (s, 1H, H<sub>8</sub>), 8.36 (s, 1H, H<sub>2</sub>), 8.26 (bs, 3H, H<sub>15</sub>), 7.77 (s, 1H, H<sub>9</sub>), 6.15 (d, *J* = 2.6 Hz, 1H, H<sub>1'</sub>), 5.41 (dd, *J* = 6.4, 2.6 Hz, 1H, H<sub>2'</sub>), 4.84 (dd, *J* = 6.4, 3.8 Hz, 1H, H<sub>3'</sub>), 4.31 – 4.22 (m, 1H, H<sub>7'</sub>), 4.16 – 4.08 (m, 1H, H<sub>4'</sub>), 3.92 (s, 1H, H<sub>14</sub>), 2.65 – 2.56 (m, 2H, H<sub>11</sub>), 1.87 – 1.55 (m, 8H, H<sub>5'</sub>, H<sub>6'</sub>, H<sub>12</sub>, H<sub>13</sub>), 1.53 (s, 1H, H<sub>9''</sub>), 1.31 (s, 1H, H<sub>9'</sub>).

**<sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>)** δ 171.0 (C<sub>16</sub>), 158.6 – 158.4 (C<sub>17</sub>), 152.9 (C<sub>6</sub>), 148.5 – 148.3 (C<sub>2</sub>, C<sub>4</sub>), 146.0 (C<sub>10</sub>), 141.5 (C<sub>8</sub>), 121.7 (C<sub>9</sub>), 119.0 (C<sub>5</sub>), 113.8 (C<sub>8'</sub>), 88.6 (C<sub>1'</sub>), 85.3 (C<sub>4'</sub>), 83.3 (C<sub>2'</sub>, C<sub>3'</sub>), 51.9 (C<sub>14</sub>), 48.8 (C<sub>7'</sub>), 29.7 – 29.6 (C<sub>5'</sub>, C<sub>13</sub>), 27.0 (C<sub>9''</sub>), 26.0 (C<sub>6'</sub>), 25.2 (C<sub>9'</sub>), 24.5 (C<sub>11</sub>, C<sub>12</sub>).

**HR-ESI-MS (M+H<sup>+</sup>)** calc. for C<sub>22</sub>H<sub>32</sub>N<sub>9</sub>O<sub>5</sub><sup>+</sup>: 502.2521; found: 502.2524

**(S)-4-(1-(3-((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)propyl)-1H-1,2,3-triazol-4-yl)-1-carboxybutan-1-aminium (47)**



To **S)-4-(1-(3-((3aR,4R,6R,6aR)-6-(6-amino-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)propyl)-1H-1,2,3-triazol-4-yl)-1-carboxybutan-1-aminium 46** (0.01624 mmol, 10 mg) a cocktail of TFA/H<sub>2</sub>O/TIPS (95:2.5:2.5 v/v, 1 mL) was added. The mixture was stirred for 1 hours at rt and quenched with MeCN (3 mL) and H<sub>2</sub>O (3 mL) and purified by preparative HPLC with a linear gradient of MeCN (0 – 100%) in 0.1 % aq. TFA as the mobile phase, giving the title product **47** (0.01413 mmol, 8.1 mg, 87%) as oil after being lyophilized.

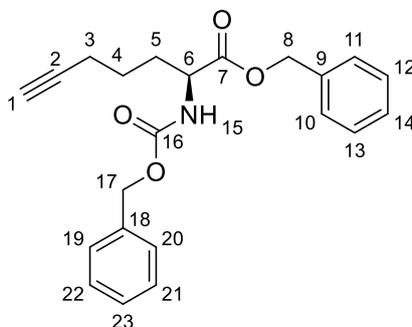
**tr** = 6.42 min

**<sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>)** δ 8.47 (s, 1H, H<sub>8</sub>), 8.31 (s, 1H, H<sub>2</sub>), 8.23 (s, 5H, C<sub>6</sub>-NH<sub>2</sub>, H<sub>15</sub>), 7.81 (s, 1H, H<sub>9</sub>), 5.88 (d, *J* = 5.1 Hz, 1H, H<sub>1'</sub>), 4.60 (t, *J* = 5.1 Hz, 1H, H<sub>2'</sub>), 4.32 (t, *J* = 7.1 Hz, 2H, H<sub>7'</sub>), 4.04 (t, *J* = 5.1 Hz, 1H, H<sub>3'</sub>), 3.96 – 3.85 (m, 2H, H<sub>4'</sub>, H<sub>14</sub>), 2.62 (t, *J* = 7.2 Hz, 2H, H<sub>11</sub>), 1.95 – 1.69 (m, 5H, H<sub>5'</sub>, H<sub>12</sub>, H<sub>13</sub>), 1.64 (p, *J* = 8.3 Hz, 3H, H<sub>12</sub>, H<sub>6'</sub>).

**<sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>)** δ 171.0 (C16), 158.5 – 158.0 (C17), 153.6 (C6), 149.5 (C2), 148.8 (C4), 146.0 (C10), 141.2 (C8), 121.8 (C9), 119.1 (C5), 87.9 (C1'), 83.3 (C4'), 73.2 (C2',C3'), 51.9 (C14), 49.0 (C7'), 29.9 (C13), 29.6 (C5'), 26.4 (C6'), 24.5 (C11, C12).

**HR-ESI-MS (M+H<sup>+</sup>)** for C<sub>19</sub>H<sub>28</sub>N<sub>9</sub>O<sub>5</sub><sup>+</sup>: 462.2213; found: 462.2205

**benzyl (S)-2-(((benzyloxy)carbonyl)amino)hept-6-ynoate (49)**



To a solution of the corresponding commercial acid **48** (1.0 equiv., 0.1090 mmol, 30 mg), benzyl alcohol (2.0 equiv., 0.2179 mmol, 22.7 μL) and Et<sub>3</sub>N (3.0 equiv., 0.3269 mmol, 45.6 μL) in DMF (1 mL) was added PyBOP (2.0 equiv., 0.2179 mmol, 113.4 mg) and the mixture was stirred at room temperature for 22 hours and quenched with H<sub>2</sub>O (5 mL), MeCN (6 mL) and MeOH (3 mL) purified by preparative HPLC with a linear gradient of MeCN (0 – 100%) in 0.1 % aq. TFA as the mobile phase, giving the alkyne benzyl ester product **49** (27.2 mg, 0.07443 mmol, 68%) as a slightly yellow oil.

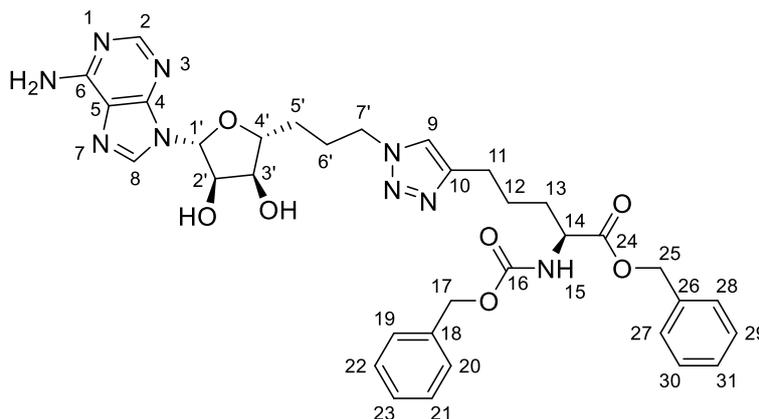
**tr** = 17.8 min

**<sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>)** δ 7.82 (d, J = 7.8 Hz, 1H, H15), 7.41 – 7.23 (m, 11H), 5.19 – 5.09 (m, 2H, C8), 5.09 – 4.97 (m, 2H, C17), 4.15 – 4.05 (m, 1H, H6), 2.78 (t, J = 2.6 Hz, 1H, H1), 2.18 – 2.11 (m, 2H, H3), 1.89 – 1.62 (m, 2H, H5), 1.58 – 1.41 (m, 2H, H4).

**<sup>13</sup>C NMR (126 MHz, DMSO-d<sub>6</sub>)** δ 172.1 (C7), 156.2 (C16), 136.9 (C18), 135.9 (C9), 128.4 – 127.70 (C10, C11, C12, C13, C14, C19, C20, C21, C22, C23), 83.9 (C2), 71.5 (C1), 65.9 (C8), 65.5 (C17), 53.6 (C6), 29.7 (C5), 24.4 (C4), 17.2 (C3).

**HR-ESI-MS (M+H<sup>+</sup>)** calc. for C<sub>22</sub>H<sub>24</sub>NO<sub>4</sub>: 366.1705; found: 366.1718

**benzyl (S)-5-(1-(3-((2R,3S,4R,5R)-5-(6-amino-9H-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)propyl)-1H-1,2,3-triazol-4-yl)-2-(((benzyloxy)carbonyl)amino)pentanoate (51)**



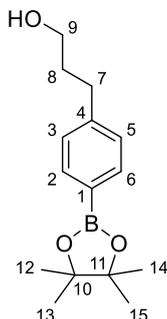
benzyl (S)-2-(((benzyloxy)carbonyl)amino)hept-6-ynoate **49** (1.0 equiv., 0.07215 mmol, 26.4 mg) solubilized in DMSO (1.6 mL) and the corresponding azide (1.0 equiv., 0.07215 mmol, 26 mg) solubilized in DMSO (740  $\mu$ L) was mixed together. Then, a premixed solution of THPTA (1.0 equiv., 0.07215 mmol, 38.3 mg, in 300  $\mu$ L of  $H_2O_{(milliQ)}$ ) and  $CuSO_4 \cdot 5H_2O$  (0.5 equiv., 0.03607 mmol, 9.0 mg, in 300  $\mu$ L of  $H_2O_{(milliQ)}$ ) was added to the first solution, followed by freshly prepared sodium ascorbate (3.0 equiv., 0.2164 mmol, 42.9 mg) solution in  $H_2O$  (300  $\mu$ L, milliQ). The reaction was stirred 30 min at rt and diluted with water (5 mL) and MeCN (4 mL) and purified by preparative HPLC with a linear gradient of MeCN (0 – 100%) in 0.1 % aq. TFA as the mobile phase, giving the clicked product. Then, a cocktail of TFA/ $H_2O$ /TIPS (95:2.5:2.5 v/v, 1 mL) was added to the lyophilized product. The mixture was stirred for 1 hours at rt and quenched with  $H_2O$  (5 mL) and MeCN (5 mL) and purified by preparative HPLC with a linear gradient of MeCN (0 – 100%) in 0.1 % aq. TFA as the mobile phase, giving the deprotected product **51** (37.1 mg, 0.05411 mmol, 75%) as white solid after being lyophilized.

**tr** = 12 min

**$^1H$  NMR (500 MHz, DMSO- $d_6$ )**  $\delta$  8.58 (s, 1H, H8), 8.40 (s, 1H, H2), 7.81 – 7.73 (m, 2H, H9, H15), 7.39 – 7.24 (m, 10H, H19, H20, H21, H22, H23, H27, H28, H29, H30, H31), 5.90 (d,  $J$  = 5.1 Hz, 1H, H1'), 5.16 – 4.98 (m, 4H, H17, H25), 4.58 (t,  $J$  = 5.1 Hz, 1H, H2'), 4.30 (t,  $J$  = 7.1 Hz, 2H, H7'), 4.15 – 4.07 (m, 1H, H14), 4.03 (t,  $J$  = 5.1 Hz, 1H, H3'), 3.94 – 3.87 (m, 1H, H4'), 2.66 – 2.52 (m, 2H, H11), 1.95 – 1.81 (m, 2H, H6'), 1.80 – 1.56 (m, 6H, H5', H12, H13).  **$^{13}C$  NMR (126 MHz, DMSO- $d_6$ )**  $\delta$  172.7 (C24), 156.7 (C16), 152.3 (C6), 149.0 (C4), 147.6 (C2), 146.8 (C10), 142.4 (C8), 137.4 (C18), 136.4 (C26), 128.9 – 128.2 (C19, C20, C21, C22, C23, C27, C28, C29, C30, C31), 122.1 (C9), 119.5 (C5), 88.5 (C1'), 84.0 (C4'), 73.9 (C2'), 73.6 (C3'), 66.4 (C25), 66.0 (C17), 54.3 (C14), 49.5 (C7'), 30.7 – 30.4 (C5', C13), 26.9 (C6'), 25.8 (C12), 25.0 (C11).

**HR-ESI-MS (M+H<sup>+</sup>)** for C<sub>34</sub>H<sub>40</sub>N<sub>9</sub>O<sub>7</sub>: 686.3051; found: 686.3058

**3-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)propan-1-ol (86)**



Commercially available **4-(3-Hydroxy-1-propynyl)benzeneboronic acid pinacol ester 84** (1.0 equiv., 0.193 mmol, 50 mg) was dissolved in EtOAc (2 mL) and the solution was flushed with argon before the addition of Pd/C (0.1 equiv., 0.0193 mmol, 10% wt, 20 mg). The reaction was flushed with argon a second time and the atmosphere was replaced by H<sub>2</sub>. The reaction mixture was stirred overnight at room temperature. The reaction mixture was filtered through celite, washed with EtOAc (3 x 10 mL) and the pooled liquid phase were concentrated under reduced pressure to afford the title compound **86** as a pale-yellow oil (51 mg, 0.194 mmol, 100%).

**R<sub>f</sub>** = 0.30 (cyclohexane/EtOAc 9:1)

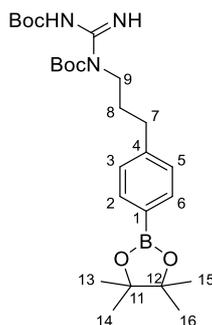
**v<sub>max</sub>/cm<sup>-1</sup>**: 2977 (w), 2934 (br), 2359 (w), 1611 (m), 1398 (s), 1359 (s), 1318 (m), 1270 (w), 1214 (w), 1143 (m), 1088 (m), 1020 (w), 962 (w), 909 (w), 858 (w), 731 (w), 658 (w)

**<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>)** δ 7.72 (2H, d, *J* = 7.0 Hz, H2, H6), 7.20 (2H, d, *J* = 7.0 Hz, H3, H5), 3.65 (2H, t, *J* = 6.0 Hz, H9), 2.71 (2H, t, *J* = 7.0 Hz, H7), 1.88 (2H, m, H8), 1.31 (12H, s, H12, H13, H14, H15)

**<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>)** δ 145.3, 134.9, 127.9, 83.7, 62.1, 34.1, 32.3, 24.8.

**HR-ESI-MS (M+H<sup>+</sup>)** calcd for C<sub>15</sub>H<sub>24</sub>BO<sub>3</sub>: 263.1819; found: 263.1828.

***tert*-butyl-N-(((*tert*-butoxy)carbonyl)(3-[4-(*t*-tétraméthyl-1,3,2-dioxaborolan-2-yl)phenyl]propyl)amino)methanimidoyl)carbamate (**81**)**



To a solution of 3-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)propan-1-ol **86** (1.0 equiv., 0.202 mmol, 51 mg), *tert*-butyl N-(((*tert*-butoxy) carbonyl] amino) methanimidoyl) carbamate (1.1 equiv., 0.222 mmol, 57 mg) and triphenylphosphine (1.1 equiv, 0.222 mmol, 58 mg) in THF (3 mL) was added DIAD (1.1 equiv., 0.222 mmol, 0.04 mL) dropwise. The reaction mixture was stirred overnight at room temperature, before being concentrated under reduced pressure. The residue was then purified by flash chromatography (cyclohexane/EtOAc 9:1) to afford the title compound **81** as a white solid (66 mg, 0.131 mmol, 65%).

$R_f = 0.43$  (cyclohexane/EtOAc 9:1)

$\nu_{\max}/\text{cm}^{-1}$ : 3382 (br), 2977 (br), 2930 (w), 2358 (w), 1713 (m), 1610 (m), 1509 (w), 1457, 1390 (w), 1361 (m), 1316 (w), 1275 (m), 1251 (w), 1146 (s), 1123 (w), 1090 (w), 1022 (w), 984 (w).

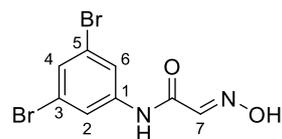
mp: 119 °C

$^1\text{H NMR}$  (500 MHz,  $\text{CDCl}_3$ )  $\delta$  9.39 (1H, br, NH) 9.21 (1H, br, NH), 7.73 (2H, d,  $J = 7.9$  Hz, H2, H6), 7.21 (2H, d,  $J = 7.9$  Hz, H3, H5), 4.01 – 3.83 (2H, m, H9), 2.65 (2H, t,  $J = 7.7$  Hz, H7), 2.02 – 1.77 (2H, m, H8), 1.51 (9H, s, *tert*butyl), 1.45 (9H, s, *tert*butyl), 1.33 (12H, s, H13, H14, H15, H16)

$^{13}\text{C NMR}$  (126 MHz,  $\text{CDCl}_3$ )  $\delta$  164.1, 160.8, 155.2, 145.3, 135.1, 127.9, 83.8, 78.9, 44.6, 33.4, 30.1, 28.5, 28.1, 25.0.

HR-ESI-MS ( $\text{M}+\text{H}^+$ ) for  $\text{C}_{26}\text{H}_{43}\text{BN}_3\text{O}_6$ : 504.3245; found: 504.3274

**(E)-N-(3,5-dibromophenyl)-2-(hydroxyimino)acetamide (87)**



3,5-dibromoaniline **83** (1.0 equiv., 1.99 mmol, 500 mg) was dissolved in a mixture of water (2.5 mL), concentrated HCl (37% w/w, 0.6 mL) and 1,4-dioxane (3 mL) and the mixture was heated to 50 °C for 1 h. A solution of chloral hydrate (1.5 equiv., 2.99 mmol, 494 mg) and Na<sub>2</sub>SO<sub>4</sub> (12.0 equiv., 23.91 mmol, 3.3 g) in 11 mL of water was prepared and heated to 50 °C. This second solution was then added to the first one, followed by a solution of hydroxylamine hydrochloride (2.9 equiv., 5.77 mmol, 401 mg) in 3 mL of water. The reaction mixture was brought to reflux for 7 h. Once the mixture has returned to room temperature, the reaction mixture was extracted with EtOAc (3 x 30 mL). The organic phases were combined, washed with brine (3 x 50 mL), dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated under reduced pressure. The crude residue was purified by column chromatography (cyclohexane/EtOAc 7:3) to afford the title compound **87** as a brown solid (589 mg, 1.829 mmol, 92%).

**R<sub>f</sub>** = 0.38 (cyclohexane/EtOAc 7:3)

**v<sub>max</sub>/cm<sup>-1</sup>** : 3389 (w), 2871 (br), 2747 (w), 1661 (m), 1531 (m), 1480 (s), 1409 (m), 1290 (m), 1268 (m), 1116 (m), 1029 (s), 905 (m), 664 (m), 615 (s), 585 (s), 575 (s), 513 (w), 435 (w) **mp**: 227 °C

**<sup>1</sup>H NMR** (400 MHz, DMSO-d<sub>6</sub>) δ 12.32 (1H, s, OH), 10.43 (1H, s, NH), 7.97 (2H, d, *J* = 1.7 Hz, H2, H6), 7.60 (1H, s, H7), 7.54 (1H, t, *J* = 1.7 Hz, H4).

**<sup>13</sup>C NMR** (101 MHz, DMSO-d<sub>6</sub>) δ 161.3, 144.0, 141.5, 128.7, 122.7, 121.6.

**HR-ESI-MS (M+H<sup>+</sup>)** calcd for C<sub>8</sub>H<sub>7</sub>Br<sub>2</sub>N<sub>2</sub>O<sub>2</sub>: 320.8874; found: 320.8863

### 4,6-dibromoindoline-2,3-dione (**88**)



To a concentrated H<sub>2</sub>SO<sub>4</sub> solution (30.0 equiv., 11.26 mmol, (98% w/w, 0.6 mL) cooled at 0 °C, (E)-N-(3,5-dibromophenyl)-2-(hydroxyimino)acetamide **87** (1.0 equiv., 0.372 mmol, 120 mg) was added using a spatula tip carefully and slowly so that the temperature should not exceed 50 °C. After addition, the black solution was heated to 90 °C for 10 min, then the mixture was allowed to return to room temperature. The reaction mixture was poured onto crushed ice (ten times the volume) dropwise, resulting in a yellow solution. The mixture was stirred at room temperature for 2 h until a precipitate appears. The reaction mixture was then extracted with EtOAc (3 x 10 mL), and then the organic layers were combined, washed with brine (3 x 30 mL), dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated under reduced pressure to afford the title compound **88** (112 mg, 0.37 mmol, 99%) as an orange solid

R<sub>f</sub> = 0.40 (cyclohexane/EtOAc 7:3)

$\nu_{\max}/\text{cm}^{-1}$ : 3239 (br), 1771 (s), 1731 (m), 1596 (s), 1566 (m), 1399 (w), 1295 (m), 1261 (w), 1217 (m), 1188 (w), 1152 (w), 1073 (m), 921 (s), 880 (m), 871 (m), 850 (w), 712 (s), 692 (m), 661 (m), 607 (m), 586 (w), 524 (w).

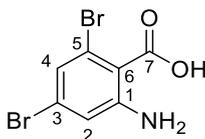
mp: 224 °C

<sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>)  $\delta$  11.28 (1H, s, NH), 7.51 (1H, d, *J* = 1.5 Hz, H2), 7.06 (1H, d, *J* = 1.5 Hz, H4).

<sup>13</sup>C NMR (101 MHz, DMSO-d<sub>6</sub>)  $\delta$  181.1, 159.1, 153.5, 131.8, 129.0, 120.6, 116.4, 114.6

HR-ESI-MS (M+H<sup>+</sup>) calcd for C<sub>8</sub>H<sub>4</sub><sup>79</sup>Br<sub>2</sub>NO<sub>2</sub>: 303.8609; found: 303.8596

### 2-amino-4,6-dibromobenzoic acid (**89**)



To a solution of 4,6-dibromoindoline-2,3-dione **88** (1.0 equiv., 0.367 mmol, 112 mg) in an aqueous solution of NaOH (1 N, 1.6 mL) was added H<sub>2</sub>O<sub>2</sub> (30% v/v, 0.23 mL) dropwise at room

temperature. The mixture was stirred for 2 h and then carefully acidified to pH 2 with concentrated HCl (37% w/w), and a light brown precipitate appears. The reaction mixture was then extracted with EtOAc (3 x 10 mL), then the organic layers were combined, washed with brine (3 x 10 mL), dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated under reduced pressure to afford the title compound **89** as an orange solid (81 mg, 0.24 mmol, 75%).

**R<sub>f</sub>** = 0.22 (cyclohexane/EtOAc 1:1)

**v<sub>max</sub>/cm<sup>-1</sup>**: 3354 (m), 2987 (br), 1650 (m), 1581 (m), 1526 (m), 1416 (m), 1301 (m), 1242 (m), 1130 (m), 1038 (br), 921 (w), 811 (w), 587 (w).

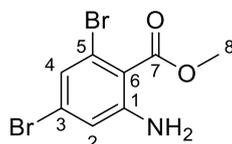
**mp**: 181 °C

**<sup>1</sup>H NMR** (400 MHz CDCl<sub>3</sub>) δ 7.16 (d, *J* = 1.8 Hz, 1H, H4), 6.83 (d, *J* = 1.8 Hz, 1H, H2).

**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>) δ 167.9, 149.8, 124.1, 121.7, 121.6, 117.9, 117.3

**HR-ESI-MS (M+H<sup>+</sup>)** calcd for C<sub>7</sub>H<sub>5</sub><sup>79</sup>Br<sub>2</sub>NO<sub>2</sub>: 293.8765; found: 293.8762

### Methyl-2-amino-4,6-dibromobenzoate (**90**)



2-amino-4,6-dibromobenzoic acid **89** (1.0 equiv., 5.08 mmol, 1500 mg) was dissolved in a mixture of toluene-methanol (3:1, v/v, 73 mL). Thereafter, trimethylsilyldiazomethane (1.2 equiv., 6 mmol, 2M in hexanes, 3.0 mL) was added dropwise. The reaction mixture was stirred for 30 min at room temperature. The solution was diluted with EtOAc (10 mL) and water (25 mL) and the aqueous phase was extracted with EtOAc (3 x 50 mL). The organic layers were combined, washed with brine (3 x 30 mL), dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated under reduced pressure affording the title compound **90** as a brown solid (950 mg, 3.07 mmol, 60%, BRSM 89%).

**R<sub>f</sub>** = 0.30 (cyclohexane/EtOAc 8:2)

**v<sub>max</sub>/cm<sup>-1</sup>** : 3486 (br), 3375 (br), 2950 (w), 1702 (s), 1583 (s), 1544 (m), 1433 (m), 1415 (m), 1292 (s), 1260 (m), 1190 (w), 1117 (s), 918 (w), 826 (w), 755 (w)

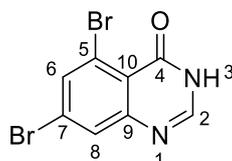
**mp**: 72 °C

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>) δ 7.08 (1H, d, *J* = 1,6 Hz, H4), 6.77 (1H, d, *J* = 1,6 Hz, H2), 4.90 (2H, s, NH<sub>2</sub>), 3.83 (3H, s, CH<sub>3</sub>).

**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>) δ 167.4, 149.4, 125.9, 125.3, 122.8, 118.0, 115.5, 52.1

**HR-ESI-MS (M+H<sup>+</sup>)** calcd for C<sub>8</sub>H<sub>8</sub><sup>79</sup>Br<sub>2</sub>NO<sub>2</sub>: 307.8922; found: 307.8932

### 5,7-dibromoquinazolin-4-ol (**91**)



A solution of methyl-2-amino-4,6-dibromobenzoate **90** (1.0 equiv., 0.174 mmol, 54 mg), formic acid (5.0 equiv., 0.873 mmol, 0.03 mL) in formamide (1 mL) was stirred at 135 °C for 10 h. The reaction was allowed to return to room temperature and was concentrated under reduced pressure. Water (1 mL) was added to the residue and the solution was stirred for 1 h, leading to the formation of a precipitate. The mixture was then extracted with EtOAc (3 x 10 mL). The organic phases were combined, washed with brine (3 x 10 mL), dried over MgSO<sub>4</sub>, filtered, and concentrated under reduced pressure. The crude product was purified by column chromatography (cyclohexane/EtOAc 1:1) to afford the title compound **91** as a white solid (44.1 mg, 0.145 mmol, 83%).

**R<sub>f</sub>** = 0.30 (cyclohexane/EtOAc 1:1)

**v<sub>max</sub>/cm<sup>-1</sup>** : 3208, 2360, 2341, 1688, 1620, 1584, 1527, 1425, 1372, 1283, 1110, 942, 861, 621, 501.

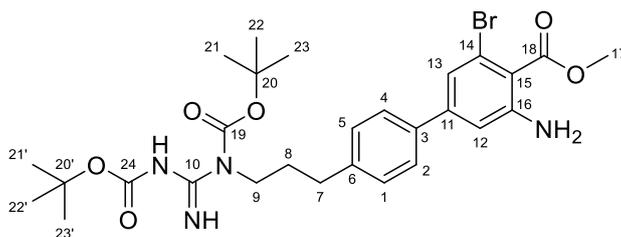
**mp**: 307 °C

**<sup>1</sup>H NMR (500 MHz, DMSO-d<sub>6</sub>)** δ 12.45 (1H, bs, NH), 8.13 (1H, s, H2), 7.94 (1H, d, *J* = 1.9 Hz, H6), 7.86 (1H, d, *J* = 1.9 Hz, H8)

**<sup>13</sup>C NMR (126 MHz, DMSO)** δ 158.8, 151.7, 147.4 (C2), 134.6, 129.7, 127.1, 121.4, 119.8

**HR-ESI-MS (M+Na<sup>+</sup>)** calcd for C<sub>8</sub>H<sub>4</sub><sup>79</sup>Br<sub>2</sub>N<sub>2</sub>ONa: 324.8588; found: 324.8589

**Methyl 3-amino-4'-(3-(1,3-bis(tert-butoxycarbonyl)guanidino)propyl)-5-bromo-[1,1'-bi phenyl]-4-carboxylate (94)**



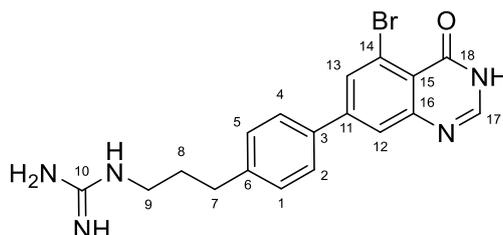
A screw-cap test tube containing a magnetic stir bar was charged with Pd(OAc)<sub>2</sub> (5.0 mol%, 0.081 mmol, 18.16 mg), Xantphos (10.0 mol%, 0.16 mmol, 93.62 mg), the aryl boronate ester **81** (1.0 equiv., 1.62 mmol, 815.5 mg), aryl halide **90** (1.0 equiv., 1.62 mmol, 500 mg) and K<sub>3</sub>PO<sub>4</sub> (2.0 equiv., 3.24 mmol, 687 mg). The tube was capped with a septum and then backfilled with argon (this sequence was repeated three times). Deionized water (500 μL) and THF (5.0 mL) and were added sequentially via syringe through the septum and the resulting mixture was stirred at room temperature for 5 min. The reaction mixture was heated at reflux with vigorous magnetic stirring for 12 h. The reaction mixture was then allowed to cool to room temperature and concentrated under reduced pressure. The crude product was purified by flash chromatography (9:1 Cyclohexane/EtOAc) to afford the desired product **94** (205 mg, 21%).

**<sup>1</sup>H NMR (500 MHz, MeOD)** δ 7.51 – 7.46 (m, 2H, H<sub>2</sub>, H<sub>4</sub>), 7.32 – 7.26 (m, 2H, H<sub>1</sub>, H<sub>5</sub>), 7.09 (d, *J* = 1.7 Hz, 1H, H<sub>13</sub>), 6.96 (d, *J* = 1.6 Hz, 1H, H<sub>12</sub>), 4.00 – 3.79 (m, 5H, H<sub>17</sub>, H<sub>9</sub>), 2.68 (t, *J* = 7.4 Hz, 2H, H<sub>7</sub>), 1.97 – 1.87 (m, 2H, H<sub>8</sub>), 1.45 (s, 9H, H<sub>21'</sub>, H<sub>22'</sub>, H<sub>23'</sub>), 1.49 (s, 9H, H<sub>21</sub>, H<sub>22</sub>, H<sub>23</sub>);

**<sup>13</sup>C NMR (126 MHz, MeOD)** δ 169.4 (C<sub>18</sub>), 162.1 (C<sub>10</sub>), 156.1 (C<sub>19</sub>), 150.8 (C<sub>24</sub>), 146.3 (C<sub>11</sub>), 143.4 (C<sub>6</sub>), 138.0 (C<sub>3</sub>), 130.1 (C<sub>1</sub>, C<sub>5</sub>), 128.0 (C<sub>2</sub>, C<sub>4</sub>), 122.8 (C<sub>14</sub>), 121.2 (C<sub>13</sub>), 116.9 (C<sub>15</sub>), 114.4 (C<sub>12</sub>), 85.1 (C<sub>20'</sub>), 79.9 (C<sub>20</sub>), 52.4 (C<sub>17</sub>), 45.6 (C<sub>9</sub>), 33.5 (C<sub>7</sub>), 31.2 (C<sub>8</sub>), 28.7 (C<sub>21'</sub>, C<sub>22'</sub>, C<sub>23'</sub>), 28.2 (C<sub>21</sub>, C<sub>22</sub>, C<sub>23</sub>).

**HR-ESI-MS (M+Na<sup>+</sup>)** calc. for C<sub>28</sub>H<sub>37</sub><sup>79</sup>BrN<sub>4</sub>O<sub>6</sub>Na: 627.1794; found: 627.1799

### 1-(3-(4-(5-bromo-4-oxo-3,4-dihydroquinazolin-7-yl)phenyl)propyl)guanidine (96)



The mixture of the corresponding methyl ester **94** (1.0 equiv., 0.826 mmol, 50 mg), APTS monohydrate (1.0 equiv., 0.826 mmol, 15.72 mg) and formamide (1 mL) was stirred at 135 °C for 8 h. The reaction is allowed to return to room temperature. Then, the crude was diluted with 4 mL of HPLC buffer and purified by preparative HPLC with a linear gradient of MeCN (0 – 100%) in 0.1 % aq. TFA as the mobile phase, giving the product **96** (4.96 mg, 0.0124 mmol, 15%) as white powder after being lyophilized.

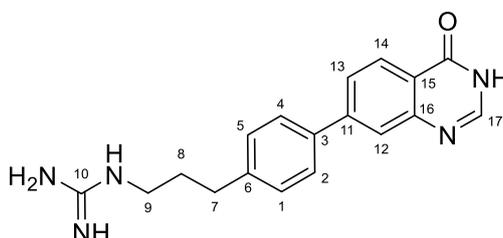
$\nu_{\text{max}}/\text{cm}^{-1}$  : 3350 (br), 3188 (br), 1666 (s), 1603 (s), 1201 (m), 1136 (m), 801 (w).

$^1\text{H NMR}$  (700 MHz, MeOD)  $\delta_{\text{H}}$  8.10 (s, 1H, H17), 8.03 (d,  $J = 1.8$  Hz, 1H, H13), 7.87 (d,  $J = 1.8$  Hz, 1H, H12), 7.70 – 7.66 (m, 2H, H2, H4), 7.39 (m, 2H, H1, H5), 3.23 (t,  $J = 7.0$  Hz, 2H, H9), 2.80 – 2.75 (t,  $J = 7.8$  Hz, 2H, H7), 2.00 – 1.93 (m, 2H, H8).

$^{13}\text{C NMR}$  (176 MHz, MeOD)  $\delta_{\text{C}}$  161.2 (C18), 158.7 (C10), 152.4 (C16), 148.3 (C11), 147.4 (C17), 143.8 (C6), 136.9 (C3), 133.1 (C13), 130.4 (C1, C5), 128.5 (C2, C4), 125.4 (C12), 122.6 (C14), 120.5 (C15), 41.9 (C9), 33.3 (C7), 31.4 (C8).

HR-ESI-MS ( $\text{M}+\text{H}^+$ ) calc. for  $\text{C}_{18}\text{H}_{19}^{79}\text{BrN}_5\text{O}$ : 400.0773; found: 400.0756

### 1-(3-(4-(4-oxo-3,4-dihydroquinazolin-7-yl)phenyl)propyl)guanidine (97)



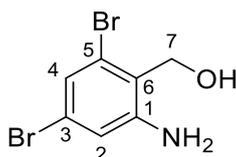
The mixture of the corresponding methyl ester **94** (1.0 equiv., 0.0165 mmol, 10 mg), APTS monohydrate (1.0 equiv., 0.0165 mmol, 3.14 mg) and formamide (200  $\mu\text{L}$ ) was stirred at 135 °C for 8 h. The reaction is allowed to return to room temperature. Then, the crude was diluted with 4 mL of HPLC buffer and purified by preparative HPLC with a linear gradient of MeCN (0

– 100%) in 0.1 % aq. TFA as the mobile phase, giving the product (0.75 mg, 0.00231 mmol, 14%) as white powder after being lyophilized.

**<sup>1</sup>H NMR (700 MHz, MeOD)** δ 8.29 (d, *J* = 8.3 Hz, 1H, H14), 8.14 (s, 1H, H17), 7.91 (d, *J* = 1.8 Hz, 1H, H12), 7.84 (dd, *J* = 8.3, 1.8 Hz, 1H, H13), 7.73 – 7.68 (m, 2H, H2, H4), 7.41 – 7.35 (m, 2H, H1, H5), 3.23 (t, *J* = 7.1 Hz, 2H, H9), 2.82 – 2.76 (m, 2H, H7), 1.97 (m, 2H, H8).

**HR-ESI-MS (M+H<sup>+</sup>)** calc. for C<sub>18</sub>H<sub>20</sub>N<sub>5</sub>O: 322.1668; found: 322.1654

### (2-amino-4,6-dibromophényl)methanol (**98**)



To a solution of Methyl-2-amino-4,6-dibromobenzoate **90** (1.0 equiv., 0.097 mmol, 30 mg) in THF (1 mL) at 0 °C was added DIBAL-H (2.1 equiv., 0.203 mmol, 1.0 M in THF, 0,20 mL) dropwise. The reaction mixture was stirred for 24 h at room temperature and was quenched with saturated solution of Rochelle salt (2 mL). After stirring 1 hour, the aqueous phase reaction mixture was extracted with EtOAc (3 x 10 mL). The organic layers were combined, washed with brine (3 x 10 mL), dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated under reduced pressure. The crude product was purified by chromatography on a silica gel column (cyclohexane/ EtOAc 8:2) to afford the title compound **98** as a white solid (13.6 mg, 0.04 mmol, 55 %).

**R<sub>f</sub>** = 0.30 (cyclohexane/EtOAc 8:2)

**v<sub>max</sub>/cm<sup>-1</sup>** : 3380, 3292, 3169, 3087, 2957, 2920, 2873, 1588, 1555, 1456, 1412, 1352, 1295, 1261, 1083, 1014, 977, 853, 842, 676.

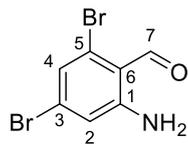
**mp**: 116 °C

**<sup>1</sup>H RMN** (500 MHz, CDCl<sub>3</sub>) δ 7.08 (d, *J* = 1.8 Hz, 1H, H4), 6.78 (d, *J* = 1.8 Hz, 1H, H2), 4.84 (s, 2H, H7).

**<sup>13</sup>C NMR** (126 MHz, CDCl<sub>3</sub>) δ 148.9, 124.9, 124.5, 122.9, 122.5, 118.0, 62.5

**HR-ESI-MS (M+H<sup>+</sup>)** calcd for C<sub>7</sub>H<sub>7</sub><sup>79</sup>Br<sub>2</sub>NO: 279.8973; found: 279.8976

### 2-amino-4,6-dibromobenzaldehyde (**99**)



To a solution of (2-amino-4,6-dibromophényl)methanol **98** (1.0 equiv., 6.05 mmol, 1700 mg) in DCM (80 mL) was added MnO<sub>2</sub> (10.0 equiv., 60.5 mmol, 5263 mg,) at room temperature. The reaction mixture was stirred for 24 h at room temperature and was filtered through Celite. The solvent was removed under reduced pressure to afford the title compound **99** as a yellow solid (1553 mg, 5.57 mmol, 92%).

R<sub>f</sub> = 0.25 (cyclohexane)

v<sub>max</sub>/cm<sup>-1</sup> : 3420, 3320, 1645, 1596, 1522, 1398, 1222, 933, 801.

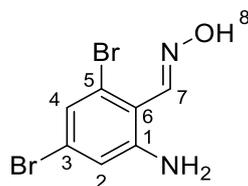
mp : 117 °C

<sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 10.31 (s, 1H, H7), 7.05 (d, J = 1.7 Hz, 1H, H4), 6.79 (dd, J = 1.8, 0.6 Hz, 1H, H2), 6.53 (s, 2H, NH<sub>2</sub>).

<sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>) δ 194.7, 152.2, 130.1, 130.0, 124.1, 119.0, 114.0

LR-ESI-MS (M+H<sup>+</sup>) calcd for C<sub>7</sub>H<sub>6</sub><sup>79</sup>Br<sub>2</sub>NO: 277.88; found: 277.70

### (E)-2-amino-4,6-dibromobenzaldehyde oxime (**100**)



To a solution of the 2-amino-4,6-dibromobenzaldehyde **99** (1.0 equiv., 0.4661 mmol, 130 mg) and K<sub>2</sub>CO<sub>3</sub> (2.0 equiv., 0.9321 mmol, 129 mg) in ethanol (8.5 mL) was added dropwise a solution of hydroxylamine hydrochloride (3.0 equiv., 1.40 mmol, 97 mg) in water (1.5 mL) at room temperature. The reaction was brought to reflux for 4 h and then allowed to cool to room temperature. The reaction was diluted with EtOAc (5 mL) and water (5 mL). The mixture was extracted with EtOAc (3 x 25 mL) and the organic layers were combined, washed with brine (1 x 10 mL), dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated under reduced pressure to afford the title compound **100** as a white solid (134 mg, 0.4521 mmol, 97%).

R<sub>f</sub> = 0.27 (cyclohexane/EtOAc 8:2)

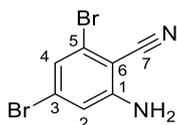
$\nu_{\max}/\text{cm}^{-1}$  : 3397, 3242, 1583, 1557, 1472, 1396, 1317, 1188, 985, 922, 827, 694

$^1\text{H NMR}$  (400 MHz,  $\text{CDCl}_3$ )  $\delta$  8.77 (s, 1H, H7), 7.19 (bs, 1H, H8), 7.09 (d,  $J = 1.8$  Hz, 1H, H4), 6.80 (dd,  $J = 1.9, 0.6$  Hz, 1H, H2).

$^{13}\text{C NMR}$  (101 MHz,  $\text{CDCl}_3$ )  $\delta_{\text{C}}$  153.0, 148.4, 126.6, 124.2, 123.7, 117.6, 111.5.

**HR-ESI-MS ( $\text{M}+\text{H}^+$ )** calcd for  $\text{C}_7\text{H}_6^{79}\text{Br}_2\text{N}_2\text{O}$ : 292.8925; found: 292.8918

## 2-amino-4,6-dibromobenzonitrile (**101**)



To a solution of (E)-2-amino-4,6-dibromobenzaldehyde oxime **100** (1.0 equiv., 0.4320 mmol, 127 mg) in DCM (5 mL) was added triethylamine (2.0 equiv., 0.8639 mmol, 120  $\mu\text{L}$ ) dropwise. After stirring for 15 min, the solution was cooled down to 0  $^{\circ}\text{C}$  and mesyl chloride (1.2 equiv., 0.5184 mmol, 40  $\mu\text{L}$ ) was added slowly. The solution was stirred for 5 h at room temperature and was diluted with EtOAc (5 mL) and water (5 mL). The reaction mixture was extracted with EtOAc (3 x 25 mL) and the organic layers were combined, washed with brine (1 x 10 mL), dried over  $\text{Na}_2\text{SO}_4$  and concentrated under reduced pressure to afford the title compound **101** as a slightly orange solid (95 mg, 0.3442 mmol, 80%)

$R_f = 0.37$  (cyclohexane/EtOAc 8:2)

$\nu_{\max}/\text{cm}^{-1}$  : 3362 (m), 2923 (w), 2218 (w), 1645 (m), 1621 (m), 1586 (s), 1558 (s), 1540 (s), 1, 1429 (m), 1272 (w).

**mp**: 170  $^{\circ}\text{C}$

$^1\text{H NMR}$  (400 MHz,  $\text{CDCl}_3$ )  $\delta$  7.11 (d,  $J = 1.6$  Hz, 1H, H4), 6.87 (d,  $J = 1.6$  Hz, 1H, H2), 4.61 (s, 2H,  $\text{NH}_2$ ).

$^{13}\text{C NMR}$  (101 MHz,  $\text{CDCl}_3$ )  $\delta$  151.6, 128.5, 125.6, 124.4, 116.5, 115.6, 98.7.

**HR-ESI-MS ( $\text{M}+\text{H}^+$ )** calcd for  $\text{C}_7\text{H}_5^{79}\text{Br}_2\text{N}_2$ : 274.8819; found: 274.8811

## Bionconjugation procedures :

The NCL was first optimized using two random peptides which led to the following protocol:

To a solution of peptide hydrazide MSSGGTRVYVG-NH-NH<sub>2</sub> (2.66 μmol, 3 mg, 1.4 eq, in 600 μL of 6M GdmCl at pH<sub>app</sub> 3.5), MPAA (145.86 μmol, 20.16 mg, 76.77 eq) was added to form a saturated MPAA emulsion. Then, a solution of acac (3.99 μmol, 8.16 μL, 2.1 eq, 5% v/v (10 μL in 200 μL of milli Q)) was added and leaved at 37°C for 5 hours by briefly vortexing periodically. Then a solution of peptide N-ter cystéine CGKPIPPLLGLDST (1.90 μmol, 2.9 mg, 1 eq, in 600 μL of 6M GdmCl, 200mM Na<sub>2</sub>HPO<sub>4</sub>, 50 mM TCEP at pH<sub>app</sub> 8.5) was added to the first solution at 4°C. The pH of this solution was adjusted to pH<sub>app</sub> 7.22 by maintaining the temperature at 4°C with the addition of 1M NaOH and the reaction was left 24 hours by gently agitating in a thermomixer at 700 rpm/min at 25°C. Then, the crude was diluted with 18 mL of HPLC buffer, TCEP (28.5 μmol, 7.1 mg, 15 eq) was added and purified by preparative HPLC with a linear gradient of MeCN (0 – 100%) in 0.1 % aq. TFA as the mobile phase, giving the NCL product MSSGGTRVYVCGKPIPPLLGLDST (4.38 mg, 1.67 μmol, 88%) as white powder after being lyophilized. BRMS (ESI<sup>+</sup>) calc. for C<sub>113</sub>H<sub>187</sub>N<sub>31</sub>O<sub>36</sub>S<sub>2</sub> ([M+H<sup>+</sup>]): 2618.32; found: 2619.33

### NCL between RSF1-Cys and the corresponding heptapeptide

In a solution of peptide MSSGGTR-NH-NH<sub>2</sub> (0.016252 μmol, 18.76 μL, 1 eq, 0.7 mg/mL in 6M GdmCl at pH 3.5), MPAA (4.16 μmol, 0.7 mg, 256 eq) was added to form a saturated MPAA emulsion. Then, a solution of acac (0.024378 μmol, 1 μL, 1.5 eq, 0.25% v/v (2.5 μL in 1000 μL of milli Q)) was added and leaved at 37°C for 5 hours by briefly vortexing periodically. Then a solution of H-Cys-RSF1 (0.0065 μmol, 30 μL, 0.4 eq, 3.4 mg/mL in 6M GdmCl, 200mM Na<sub>2</sub>HPO<sub>4</sub>, 50 mM TCEP, 100 mM NaCl at pH 7.8) was added to the first solution at 4°C. The pH of this solution was adjusted to pH<sub>app</sub> 7.25 by maintaining the temperature at 4°C with the addition of 1M NaOH (17 μL) and the reaction was left 24 hours by gently agitating in a thermomixer at 700 rpm/min at 25°C.

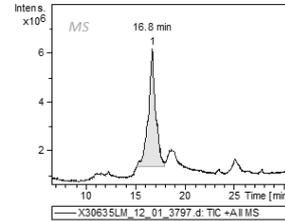
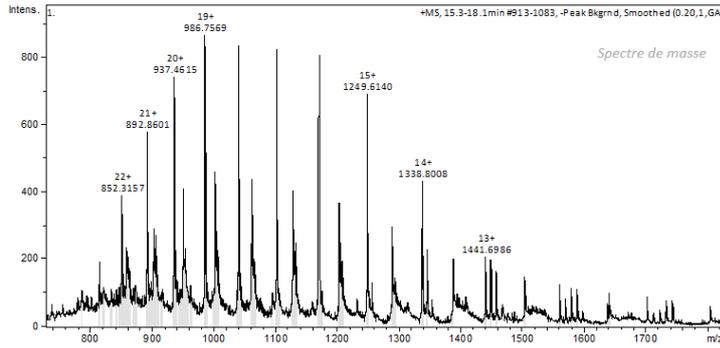
## Inhibition assays for modified arginines

All the compounds were dissolved in 20%DMSO and their concentration was determined by UV absorption ( $\epsilon_{260}=15400 \text{ M}^{-1}\text{cm}^{-1}$ ). The substrate and enzyme concentrations were adapted to ensure a linear accumulation of the product with *time during the methylation reaction*.

A solution containing 26  $\mu\text{M}$  of RSF1 (repressor of splicing factor 1), 10  $\mu\text{g/mL}$  of bovin serum albumin and 350 nM of CARM1 in 50 mM Tris HCl pH8, 200 mM NaCl, 0.5 mM EDTA, 0.5 mM DTT was incubated with increasing concentrations of Arginine-SAM analogues (20 nM to 125  $\mu\text{M}$ ) for 10 min at room temperature. The reaction was initiated by addition of 10  $\mu\text{M}$   $^{14}\text{C}$ -labeled SAM at 50 mCi.mmol<sup>-1</sup> (Perkin Elmer) at 25°C and stopped after 30 minutes by mixing to 3  $\mu\text{L}$  of SDS PAGE loading buffer. Samples were analysed on a 15% polyacrylamide denaturing gel. After electrophoresis, the gel was stained with Coomassie Brilliant Blue, destained, dried on a Whatman 3MM paper sheet and exposed for 24 h to 5 days against a BAS-IP MS 2025 imaging plate (Fujifilm). The amount of radioactivity incorporated specifically by RSF1 was quantified by phosphorimager analysis (Typhoon, Molecular Dynamic). IC<sub>50</sub> values were determined by fitting the data for sigmoidal dose–response curves using the R version 3.4.4 software package.

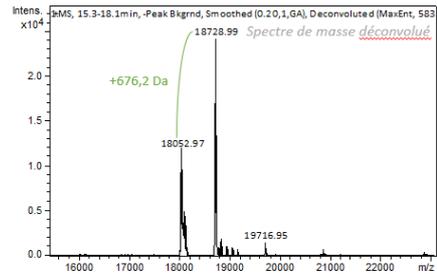
# Mass analysis : NCL between RSF1 and the corresponding heptapeptide

END264-2

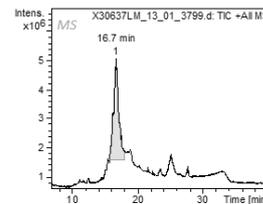
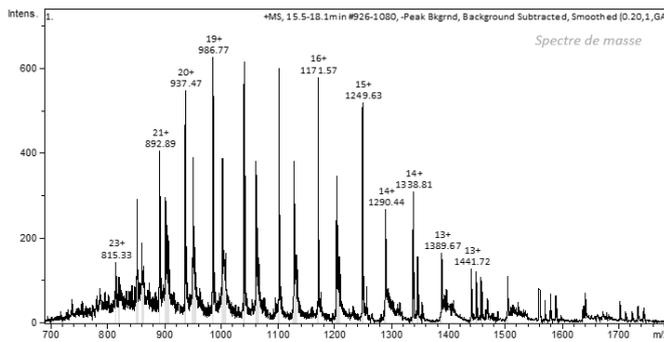


Masse exacte du peptide : 708,33 Da  
 Incrément de masse du à la ligation attendu : 676,3 Da

Masse du composé majoritaire (RSF1-ligué) : **18728,9 ± 0,2 Da (65%)**  
 Autres masses mesurée : 18053,0 ± 0,3 Da (31%) RSF1-cys  
 19716,9 ± 0,5 Da (4%)

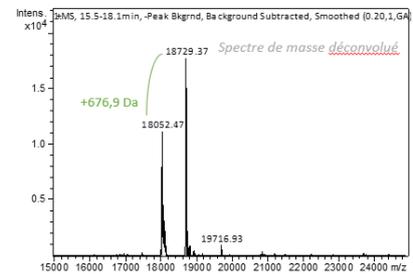


END264-3



Masse exacte du peptide : 708,33 Da  
 Incrément de masse du à la ligation attendu : 676,3 Da

Masse du composé majoritaire (RSF1-ligué) : **18729,4 ± 0,2 Da (60%)**  
 Autres masses mesurée : 18052,5 ± 0,5 Da (37%) RSF1-cys  
 19716,9 ± 0,7 Da (3%)



## Références

- (1) WATSON, J. D.; CRICK, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171* (4356), 737–738. <https://doi.org/10.1038/171737a0>.
- (2) Ghosh, A.; Bansal, M. A Glossary of DNA Structures from A to Z. *Acta Crystallogr. D Biol. Crystallogr.* **2003**, *59* (4), 620–626. <https://doi.org/10.1107/S0907444903003251>.
- (3) Klug, A. Rosalind Franklin and the Discovery of the Structure of DNA. *Nature* **1968**, *219* (5156), 808–810. <https://doi.org/10.1038/219808a0>.
- (4) Belmont, P.; Constant, J. F.; Demeunynck, M. Nucleic Acid Conformation Diversity: From Structure to Function and Regulation. *Chem. Soc. Rev.* **2001**, *30* (1), 70–81. <https://doi.org/10.1039/a904630e>.
- (5) Richmond, T. J.; Davey, C. A. The Structure of DNA in the Nucleosome Core. *Nature* **2003**, *423* (6936), 145–150. <https://doi.org/10.1038/nature01595>.
- (6) Wing, R.; Drew, H.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. Crystal Structure Analysis of a Complete Turn of B-DNA. *Nature* **1980**, *287* (5784), 755–758. <https://doi.org/10.1038/287755a0>.
- (7) Evich, M.; Spring-Connell, A. M.; Germann, M. W. Impact of Modified Ribose Sugars on Nucleic Acid Conformation and Function. *Heterocycl. Commun.* **2017**, *23* (3), 155–165. <https://doi.org/10.1515/hc-2017-0056>.
- (8) Wood, B. R. The Importance of Hydration and DNA Conformation in Interpreting Infrared Spectra of Cells and Tissues. *Chem Soc Rev* **2015**, *19*.
- (9) Herbert, A.; Rich, A. The Biology of Left-Handed Z-DNA. *J. Biol. Chem.* **1996**, *271* (20), 11595–11598. <https://doi.org/10.1074/jbc.271.20.11595>.
- (10) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: Sequence, Topology and Structure. *Nucleic Acids Res.* **2006**, *34* (19), 5402–5415. <https://doi.org/10.1093/nar/gkl655>.
- (11) Guéron, M.; Leroy, J.-L. The I-Motif in Nucleic Acids. *Curr. Opin. Struct. Biol.* **2000**, *10* (3), 326–331. [https://doi.org/10.1016/S0959-440X\(00\)00091-9](https://doi.org/10.1016/S0959-440X(00)00091-9).
- (12) Choi, J.; Majima, T. Conformational Changes of Non-B DNA. *Chem. Soc. Rev.* **2011**, *40* (12), 5893. <https://doi.org/10.1039/c1cs15153c>.
- (13) Nirenberg, M. Historical Review: Deciphering the Genetic Code – a Personal Account. *Trends Biochem. Sci.* **2004**, *29* (1), 46–54. <https://doi.org/10.1016/j.tibs.2003.11.009>.
- (14) Sanger, F.; Nicklen, S.; Coulson, A. R. DNA Sequencing with Chain-Terminating Inhibitors. *Proc. Natl. Acad. Sci.* **1977**, *74* (12), 5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>.
- (15) International Human Genome Sequencing Consortium. Finishing the Euchromatic Sequence of the Human Genome. *Nature* **2004**, *431* (7011), 931–945. <https://doi.org/10.1038/nature03001>.
- (16) Nurk, S.; Koren, S.; Rhie, A.; Rautiainen, M.; Bizikadze, A. V.; Mikheenko, A.; Vollger, M. R.; Altemose, N.; Uralsky, L.; Gershman, A.; Aganezov, S.; Hoyt, S. J.; Diekhans, M.; Logsdon, G. A.; Alonge, M.; Antonarakis, S. E.; Borchers, M.; Bouffard, G. G.; Brooks, S. Y.; Caldas, G. V.; Chen, N.-C.; Cheng, H.; Chin, C.-S.; Chow, W.; de Lima, L. G.; Dishuck, P. C.; Durbin, R.; Dvorkina, T.; Fiddes, I. T.; Formenti, G.; Fulton, R. S.; Fungtammasan, A.; Garrison, E.; Grady, P. G. S.; Graves-Lindsay, T. A.; Hall, I. M.; Hansen, N. F.; Hartley, G. A.; Haukness, M.; Howe, K.; Hunkapiller, M. W.; Jain, C.; Jain, M.; Jarvis, E. D.; Kerpedjiev, P.; Kirsche, M.; Kolmogorov, M.; Korlach, J.; Kremitzki, M.; Li, H.; Maduro, V.

- V.; Marschall, T.; McCartney, A. M.; McDaniel, J.; Miller, D. E.; Mullikin, J. C.; Myers, E. W.; Olson, N. D.; Paten, B.; Peluso, P.; Pevzner, P. A.; Porubsky, D.; Potapova, T.; RogaeV, E. I.; Rosenfeld, J. A.; Salzberg, S. L.; Schneider, V. A.; Sedlazeck, F. J.; Shafin, K.; Shew, C. J.; Shumate, A.; Sims, Y.; Smit, A. F. A.; Soto, D. C.; Sović, I.; Storer, J. M.; Streets, A.; Sullivan, B. A.; Thibaud-Nissen, F.; Torrance, J.; Wagner, J.; Walenz, B. P.; Wenger, A.; Wood, J. M. D.; Xiao, C.; Yan, S. M.; Young, A. C.; Zarate, S.; Surti, U.; McCoy, R. C.; Dennis, M. Y.; Alexandrov, I. A.; Gerton, J. L.; O'Neill, R. J.; Timp, W.; Zook, J. M.; Schatz, M. C.; Eichler, E. E.; Miga, K. H.; Phillippy, A. M. The Complete Sequence of a Human Genome. *Science* **2022**, 376 (6588), 44–53. <https://doi.org/10.1126/science.abj6987>.
- (17) Murrell, A.; Rakyán, V. K.; Beck, S. From Genome to Epigenome. *Hum. Mol. Genet.* **2005**, 14 (suppl\_1), R3–R10. <https://doi.org/10.1093/hmg/ddi110>.
- (18) Wu, C. -t.; Morris, J. R. Genes, Genetics, and Epigenetics: A Correspondence. *Sci. New Ser.* **2001**, 293 (5532), 1103–1105.
- (19) Dupont, C.; Armant, D.; Brenner, C. Epigenetics: Definition, Mechanisms and Clinical Perspective. *Semin. Reprod. Med.* **2009**, 27 (05), 351–357. <https://doi.org/10.1055/s-0029-1237423>.
- (20) Waddington, C. H. Towards a Theoretical Biology. *Nature* **1968**, 218 (5141), 525–527. <https://doi.org/10.1038/218525a0>.
- (21) Portela, A.; Esteller, M. Epigenetic Modifications and Human Disease. *Nat. Biotechnol.* **2010**, 28 (10), 1057–1068. <https://doi.org/10.1038/nbt.1685>.
- (22) Olins, D. E.; Olins, A. L. Chromatin History: Our View from the Bridge. *Nat. Rev. Mol. Cell Biol.* **2003**, 4 (10), 809–814. <https://doi.org/10.1038/nrm1225>.
- (23) Luger, K. Crystal Structure of the Nucleosome Core Particle at 2.8 Å Resolution. **1997**, 389, 10.
- (24) Olins, A. L.; Olins, D. E. Spheroid Chromatin Units (v Bodies). *Science* **1974**, 183 (4122), 330–332. <https://doi.org/10.1126/science.183.4122.330>.
- (25) Robinson, P. J.; Rhodes, D. Structure of the '30nm' Chromatin Fibre: A Key Role for the Linker Histone. *Curr. Opin. Struct. Biol.* **2006**, 16 (3), 336–343. <https://doi.org/10.1016/j.sbi.2006.05.007>.
- (26) Kornberg, R. D.; Lorch, Y. Chromatin Structure and Transcription. *Annu. Rev. Cell Biol.* **1992**, 8 (1), 563–587. <https://doi.org/10.1146/annurev.cb.08.110192.003023>.
- (27) Quina, A. S.; Buschbeck, M.; Di Croce, L. Chromatin Structure and Epigenetics. *Biochem. Pharmacol.* **2006**, 72 (11), 1563–1569. <https://doi.org/10.1016/j.bcp.2006.06.016>.
- (28) Babu, A.; Verma, R. S. Chromosome Structure: Euchromatin and Heterochromatin. In *International Review of Cytology*; Elsevier, 1987; Vol. 108, pp 1–60. [https://doi.org/10.1016/S0074-7696\(08\)61435-7](https://doi.org/10.1016/S0074-7696(08)61435-7).
- (29) Horvath, J. E. Lessons from the Human Genome: Transitions between Euchromatin and Heterochromatin. *Hum. Mol. Genet.* **2001**, 10 (20), 2215–2223. <https://doi.org/10.1093/hmg/10.20.2215>.
- (30) Ball, D. J.; Gross, D. S.; Garrard, W. T. 5-Methylcytosine Is Localized in Nucleosomes That Contain Histone H1. *Proc. Natl. Acad. Sci.* **1983**, 80 (18), 5490–5494. <https://doi.org/10.1073/pnas.80.18.5490>.
- (31) Hebbes, T. R.; Thorne, A. W.; Crane-Robinson, C. A Direct Link between Core Histone Acetylation and Transcriptionally Active Chromatin. *EMBO J.* **1988**, 7 (5), 1395–1402. <https://doi.org/10.1002/j.1460-2075.1988.tb02956.x>.

- (32) Hu, T.; Banzhaf, W. Evolvability and Speed of Evolutionary Algorithms in Light of Recent Developments in Biology. *J. Artif. Evol. Appl.* **2010**, *2010*, 1–28. <https://doi.org/10.1155/2010/568375>.
- (33) Singal, R.; Ginder, G. D. DNA Methylation. *Blood* **1999**, *93* (12), 4059–4070. <https://doi.org/10.1182/blood.V93.12.4059>.
- (34) Rottach, A.; Leonhardt, H.; Spada, F. DNA Methylation-Mediated Epigenetic Control. *J. Cell. Biochem.* **2009**, *108* (1), 43–51. <https://doi.org/10.1002/jcb.22253>.
- (35) Malygin, E. G.; Hattman, S. DNA Methyltransferases: Mechanistic Models Derived from Kinetic Analysis. *Crit. Rev. Biochem. Mol. Biol.* **2012**, *47* (2), 97–193. <https://doi.org/10.3109/10409238.2011.620942>.
- (36) Das, P. M.; Singal, R. DNA Methylation and Cancer. *J. Clin. Oncol.* **2004**, *22* (22), 4632–4642. <https://doi.org/10.1200/JCO.2004.07.151>.
- (37) Jin, B.; Robertson, K. D. DNA Methyltransferases, DNA Damage Repair, and Cancer. In *Epigenetic Alterations in Oncogenesis*; Karpf, A. R., Ed.; Advances in Experimental Medicine and Biology; Springer New York: New York, NY, 2013; Vol. 754, pp 3–29. [https://doi.org/10.1007/978-1-4419-9967-2\\_1](https://doi.org/10.1007/978-1-4419-9967-2_1).
- (38) Pradhan, S. Baculovirus-Mediated Expression and Characterization of the Full-Length Murine DNA Methyltransferase. *Nucleic Acids Res.* **1997**, *25* (22), 4666–4673. <https://doi.org/10.1093/nar/25.22.4666>.
- (39) Pradhan, S.; Bacolla, A.; Wells, R. D.; Roberts, R. J. Recombinant Human DNA (Cytosine-5) Methyltransferase. *J. Biol. Chem.* **1999**, *274* (46), 33002–33010. <https://doi.org/10.1074/jbc.274.46.33002>.
- (40) Goyal, R. Accuracy of DNA Methylation Pattern Preservation by the Dnmt1 Methyltransferase. *Nucleic Acids Res.* **2006**, *34* (4), 1182–1188. <https://doi.org/10.1093/nar/gkl002>.
- (41) Jeltsch, A. On the Enzymatic Properties of Dnmt1: Specificity, Processivity, Mechanism of Linear Diffusion and Allosteric Regulation of the Enzyme. *Epigenetics* **2006**, *1* (2), 63–66. <https://doi.org/10.4161/epi.1.2.2767>.
- (42) Bostick, M.; Kim, J. K.; Estève, P.-O.; Clark, A.; Pradhan, S.; Jacobsen, S. E. UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science* **2007**, *317* (5845), 1760–1764. <https://doi.org/10.1126/science.1147939>.
- (43) Arita, K.; Ariyoshi, M.; Tochio, H.; Nakamura, Y.; Shirakawa, M. Recognition of Hemi-Methylated DNA by the SRA Protein UHRF1 by a Base-Flipping Mechanism. *Nature* **2008**, *455* (7214), 818–821. <https://doi.org/10.1038/nature07249>.
- (44) Iida, T.; Suetake, I.; Tajima, S.; Morioka, H.; Ohta, S.; Obuse, C.; Tsurimoto, T. PCNA Clamp Facilitates Action of DNA Cytosine Methyltransferase 1 on Hemimethylated DNA: Interactions between Dnmt1 and PCNA. *Genes Cells* **2002**, *7* (10), 997–1007. <https://doi.org/10.1046/j.1365-2443.2002.00584.x>.
- (45) Prokhortchouk, E.; Defossez, P.-A. The Cell Biology of DNA Methylation in Mammals. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **2008**, *1783* (11), 2167–2173. <https://doi.org/10.1016/j.bbamcr.2008.07.015>.
- (46) Arand, J.; Spieler, D.; Karius, T.; Branco, M. R.; Meilinger, D.; Meissner, A.; Jenuwein, T.; Xu, G.; Leonhardt, H.; Wolf, V.; Walter, J. In Vivo Control of CpG and Non-CpG DNA Methylation by DNA Methyltransferases. *PLoS Genet.* **2012**, *8* (6), e1002750. <https://doi.org/10.1371/journal.pgen.1002750>.
- (47) Egger, G.; Jeong, S.; Escobar, S. G.; Cortez, C. C.; Li, T. W. H.; Saito, Y.; Yoo, C. B.; Jones, P. A.; Liang, G. Identification of DNMT1 (DNA Methyltransferase 1) Hypomorphs

- in Somatic Knockouts Suggests an Essential Role for DNMT1 in Cell Survival. *Proc. Natl. Acad. Sci.* **2006**, *103* (38), 14080–14085. <https://doi.org/10.1073/pnas.0604602103>.
- (48) Okano, M.; Bell, D. W.; Haber, D. A.; Li, E. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell* **1999**, *99* (3), 247–257. [https://doi.org/10.1016/S0092-8674\(00\)81656-6](https://doi.org/10.1016/S0092-8674(00)81656-6).
- (49) Okano, M.; Xie, S.; Li, E. Cloning and Characterization of a Family of Novel Mammalian DNA (Cytosine-5) Methyltransferases. *Nat. Genet.* **1998**, *19* (3), 219–220. <https://doi.org/10.1038/890>.
- (50) Gowher, H.; Jeltsch, A. Enzymatic Properties of Recombinant Dnmt3a DNA Methyltransferase from Mouse: The Enzyme Modifies DNA in a Non-Processive Manner and Also Methylates Non-CpA Sites. *J. Mol. Biol.* **2001**, *309* (5), 1201–1208. <https://doi.org/10.1006/jmbi.2001.4710>.
- (51) Liang, G.; Chan, M. F.; Tomigahara, Y.; Tsai, Y. C.; Gonzales, F. A.; Li, E.; Laird, P. W.; Jones, P. A. Cooperativity between DNA Methyltransferases in the Maintenance Methylation of Repetitive Elements. *Mol. Cell. Biol.* **2002**, *22* (2), 480–491. <https://doi.org/10.1128/MCB.22.2.480-491.2002>.
- (52) Chen, T.; Ueda, Y.; Dodge, J. E.; Wang, Z.; Li, E. Establishment and Maintenance of Genomic Methylation Patterns in Mouse Embryonic Stem Cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* **2003**, *23* (16), 5594–5605. <https://doi.org/10.1128/MCB.23.16.5594-5605.2003>.
- (53) Jurkowska, R. Z.; Jurkowski, T. P.; Jeltsch, A. Structure and Function of Mammalian DNA Methyltransferases. *ChemBioChem* **2011**, *12* (2), 206–222. <https://doi.org/10.1002/cbic.201000195>.
- (54) Bhutani, N.; Burns, D. M.; Blau, H. M. DNA Demethylation Dynamics. *Cell* **2011**, *146* (6), 866–872. <https://doi.org/10.1016/j.cell.2011.08.042>.
- (55) Wu, X.; Zhang, Y. TET-Mediated Active DNA Demethylation: Mechanism, Function and Beyond. *Nat. Rev. Genet.* **2017**, *18* (9), 517–534. <https://doi.org/10.1038/nrg.2017.33>.
- (56) Bannister, A. J.; Kouzarides, T. Regulation of Chromatin by Histone Modifications. *Cell Res.* **2011**, *21* (3), 381–395. <https://doi.org/10.1038/cr.2011.22>.
- (57) Li, Q.; Shortreed, M. R.; Wenger, C. D.; Frey, B. L.; Schaffer, L. V.; Scalf, M.; Smith, L. M. Global Post-Translational Modification Discovery. *J. Proteome Res.* **2017**, *16* (4), 1383–1390. <https://doi.org/10.1021/acs.jproteome.6b00034>.
- (58) Rubin, C. S.; Rosen, O. M. Protein Phosphorylation. *Annu. Rev. Biochem.* **1975**, *44* (1), 831–887. <https://doi.org/10.1146/annurev.bi.44.070175.004151>.
- (59) Downs, J. A.; Lowndes, N. F.; Jackson, S. P. A Role for *Saccharomyces Cerevisiae* Histone H2A in DNA Repair. *Nature* **2000**, *408* (6815), 1001–1004. <https://doi.org/10.1038/35050000>.
- (60) Rossetto, D.; Avvakumov, N.; Côté, J. Histone Phosphorylation: A Chromatin Modification Involved in Diverse Nuclear Events. *Epigenetics* **2012**, *7* (10), 1098–1108. <https://doi.org/10.4161/epi.21975>.
- (61) Allfrey, V. G.; Faulkner, R.; Mirsky, A. E. ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. *Proc. Natl. Acad. Sci.* **1964**, *51* (5), 786–794. <https://doi.org/10.1073/pnas.51.5.786>.
- (62) Allis, C. D.; Berger, S. L.; Cote, J.; Dent, S.; Jenuwien, T.; Kouzarides, T.; Pillus, L.; Reinberg, D.; Shi, Y.; Shiekhattar, R.; Shilatifard, A.; Workman, J.; Zhang, Y. New Nomenclature for Chromatin-Modifying Enzymes. *Cell* **2007**, *131* (4), 633–636. <https://doi.org/10.1016/j.cell.2007.10.039>.

- (63) Zhao, S.; Xu, W.; Jiang, W.; Yu, W.; Lin, Y.; Zhang, T.; Yao, J.; Zhou, L.; Zeng, Y.; Li, H.; Li, Y.; Shi, J.; An, W.; Hancock, S. M.; He, F.; Qin, L.; Chin, J.; Yang, P.; Chen, X.; Lei, Q.; Xiong, Y.; Guan, K.-L. Regulation of Cellular Metabolism by Protein Lysine Acetylation. *Science* **2010**, *327* (5968), 1000–1004. <https://doi.org/10.1126/science.1179689>.
- (64) Sanchez, R.; Zhou, M.-M. The Role of Human Bromodomains in Chromatin Biology and Gene Transcription. *Curr. Opin. Drug Discov. Devel.* **2009**, *12* (5), 659–665.
- (65) Mujtaba, S.; Zeng, L.; Zhou, M.-M. Structure and Acetyl-Lysine Recognition of the Bromodomain. *Oncogene* **2007**, *26* (37), 5521–5527. <https://doi.org/10.1038/sj.onc.1210618>.
- (66) Hassan, A. H.; Prochasson, P.; Neely, K. E.; Galasinski, S. C.; Chandy, M.; Carrozza, M. J.; Workman, J. L. Function and Selectivity of Bromodomains in Anchoring Chromatin-Modifying Complexes to Promoter Nucleosomes. *Cell* **2002**, *111* (3), 369–379. [https://doi.org/10.1016/S0092-8674\(02\)01005-X](https://doi.org/10.1016/S0092-8674(02)01005-X).
- (67) Choudhary, C.; Kumar, C.; Gnad, F.; Nielsen, M. L.; Rehman, M.; Walther, T. C.; Olsen, J. V.; Mann, M. Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science* **2009**, *325* (5942), 834–840. <https://doi.org/10.1126/science.1175371>.
- (68) Martin, C.; Zhang, Y. The Diverse Functions of Histone Lysine Methylation. *Nat. Rev. Mol. Cell Biol.* **2005**, *6* (11), 838–849. <https://doi.org/10.1038/nrm1761>.
- (69) Black, J. C.; Van Rechem, C.; Whetstine, J. R. Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. *Mol. Cell* **2012**, *48* (4), 491–507. <https://doi.org/10.1016/j.molcel.2012.11.006>.
- (70) Arrowsmith, C. H.; Bountra, C.; Fish, P. V.; Lee, K.; Schapira, M. Epigenetic Protein Families: A New Frontier for Drug Discovery. *Nat. Rev. Drug Discov.* **2012**, *11* (5), 384–400. <https://doi.org/10.1038/nrd3674>.
- (71) Sims, R. J.; Nishioka, K.; Reinberg, D. Histone Lysine Methylation: A Signature for Chromatin Function. *Trends Genet.* **2003**, *19* (11), 629–639. <https://doi.org/10.1016/j.tig.2003.09.007>.
- (72) Tsukada, Y.; Fang, J.; Erdjument-Bromage, H.; Warren, M. E.; Borchers, C. H.; Tempst, P.; Zhang, Y. Histone Demethylation by a Family of JmjC Domain-Containing Proteins. *Nature* **2006**, *439* (7078), 811–816. <https://doi.org/10.1038/nature04433>.
- (73) Shi, Y.; Lan, F.; Matson, C.; Mulligan, P.; Whetstine, J. R.; Cole, P. A.; Casero, R. A.; Shi, Y. Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell* **2004**, *119* (7), 941–953. <https://doi.org/10.1016/j.cell.2004.12.012>.
- (74) Bedford, M. T. Arginine Methylation at a Glance. *J. Cell Sci.* **2007**, *120* (24), 4243–4246. <https://doi.org/10.1242/jcs.019885>.
- (75) Lorton, B. M.; Shechter, D. Cellular Consequences of Arginine Methylation. *Cell. Mol. Life Sci.* **2019**, *76* (15), 2933–2956. <https://doi.org/10.1007/s00018-019-03140-2>.
- (76) Zhang, X.; Zhou, L.; Cheng, X. Crystal Structure of the Conserved Core of Protein Arginine Methyltransferase PRMT3. *EMBO J.* **2000**, *19* (14), 3509–3519. <https://doi.org/10.1093/emboj/19.14.3509>.
- (77) Chen, C.; Nott, T. J.; Jin, J.; Pawson, T. Deciphering Arginine Methylation: Tudor Tells the Tale. *Nat. Rev. Mol. Cell Biol.* **2011**, *12* (10), 629–642. <https://doi.org/10.1038/nrm3185>.
- (78) Wang, Y.; Wysocka, J.; Sayegh, J.; Lee, Y.-H.; Perlin, J. R.; Leonelli, L.; Sonbuchner, L. S.; McDonald, C. H.; Cook, R. G.; Dou, Y.; Roeder, R. G.; Clarke, S.; Stallcup, M. R.; Allis, C. D.; Coonrod, S. A. Human PAD4 Regulates Histone Arginine Methylation Levels

- via Demethylation. *Science* **2004**, *306* (5694), 279–283. <https://doi.org/10.1126/science.1101400>.
- (79) McGraw, W. T.; Potempa, J.; Farley, D.; Travis, J. Purification, Characterization, and Sequence Analysis of a Potential Virulence Factor from *Porphyromonas Gingivalis*, Peptidylarginine Deiminase. *Infect. Immun.* **1999**, *67* (7), 3248–3256. <https://doi.org/10.1128/IAI.67.7.3248-3256.1999>.
- (80) Mondal, S.; Thompson, P. R. Chemical Biology of Protein Citrullination by the Protein A Arginine Deiminases. *Curr. Opin. Chem. Biol.* **2021**, *63*, 19–27. <https://doi.org/10.1016/j.cbpa.2021.01.010>.
- (81) Kearney, P. L.; Bhatia, M.; Jones, N. G.; Yuan, L.; Glascock, M. C.; Catchings, K. L.; Yamada, M.; Thompson, P. R. Kinetic Characterization of Protein Arginine Deiminase 4: A Transcriptional Corepressor Implicated in the Onset and Progression of Rheumatoid Arthritis. *Biochemistry* **2005**, *44* (31), 10570–10582. <https://doi.org/10.1021/bi050292m>.
- (82) Walport, L. J.; Hopkinson, R. J.; Chowdhury, R.; Schiller, R.; Ge, W.; Kawamura, A.; Schofield, C. J. Arginine Demethylation Is Catalysed by a Subset of JmJc Histone Lysine Demethylases. *Nat. Commun.* **2016**, *7* (1), 11974. <https://doi.org/10.1038/ncomms11974>.
- (83) Liu, H.; Wang, C.; Lee, S.; Deng, Y.; Wither, M.; Oh, S.; Ning, F.; Dege, C.; Zhang, Q.; Liu, X.; Johnson, A. M.; Zang, J.; Chen, Z.; Janknecht, R.; Hansen, K.; Marrack, P.; Li, C.-Y.; Kappler, J. W.; Hagman, J.; Zhang, G. Clipping of Arginine-Methylated Histone Tails by JMJD5 and JMJD7. *Proc. Natl. Acad. Sci.* **2017**, *114* (37). <https://doi.org/10.1073/pnas.1706831114>.
- (84) Wilkins, S. E.; Islam, Md. S.; Gannon, J. M.; Markolovic, S.; Hopkinson, R. J.; Ge, W.; Schofield, C. J.; Chowdhury, R. JMJD5 Is a Human Arginyl C-3 Hydroxylase. *Nat. Commun.* **2018**, *9* (1), 1180. <https://doi.org/10.1038/s41467-018-03410-w>.
- (85) Hwang, J. W.; Cho, Y.; Bae, G.-U.; Kim, S.-N.; Kim, Y. K. Protein Arginine Methyltransferases: Promising Targets for Cancer Therapy. *Exp. Mol. Med.* **2021**, *53* (5), 788–808. <https://doi.org/10.1038/s12276-021-00613-y>.
- (86) Yang, Y.; Bedford, M. T. Protein Arginine Methyltransferases and Cancer. *Nat. Rev. Cancer* **2013**, *13* (1), 37–50. <https://doi.org/10.1038/nrc3409>.
- (87) Guccione, E.; Richard, S. The Regulation, Functions and Clinical Relevance of Arginine Methylation. *Nat. Rev. Mol. Cell Biol.* **2019**, *20* (10), 642–657. <https://doi.org/10.1038/s41580-019-0155-x>.
- (88) Tang, J.; Frankel, A.; Cook, R. J.; Kim, S.; Paik, W. K.; Williams, K. R.; Clarke, S.; Herschman, H. R. PRMT1 Is the Predominant Type I Protein Arginine Methyltransferase in Mammalian Cells. *J. Biol. Chem.* **2000**, *275* (11), 7723–7730. <https://doi.org/10.1074/jbc.275.11.7723>.
- (89) Zhang, X.; Cheng, X. Structure of the Predominant Protein Arginine Methyltransferase PRMT1 and Analysis of Its Binding to Substrate Peptides. *Structure* **2003**, *11* (5), 509–520. [https://doi.org/10.1016/S0969-2126\(03\)00071-6](https://doi.org/10.1016/S0969-2126(03)00071-6).
- (90) Wooderchak, W. L.; Zang, T.; Zhou, Z. S.; Acuña, M.; Tahara, S. M.; Hevel, J. M. Substrate Profiling of PRMT1 Reveals Amino Acid Sequences That Extend Beyond the “RGG” Paradigm. *Biochemistry* **2008**, *47* (36), 9456–9466. <https://doi.org/10.1021/bi800984s>.
- (91) Wang, H.; Huang, Z.-Q.; Xia, L.; Feng, Q.; Erdjument-Bromage, H.; Strahl, B. D.; Briggs, S. D.; Allis, C. D.; Wong, J.; Tempst, P.; Zhang, Y. Methylation of Histone H4 at Arginine 3 Facilitating Transcriptional Activation by Nuclear Hormone Receptor. *Science* **2001**, *293* (5531), 853–857. <https://doi.org/10.1126/science.1060781>.

- (92) Strahl, B. D.; Briggs, S. D.; Brame, C. J.; Caldwell, J. A.; Koh, S. S.; Ma, H.; Cook, R. G.; Shabanowitz, J.; Hunt, D. F.; Stallcup, M. R.; Allis, C. D. Methylation of Histone H4 at Arginine 3 Occurs in Vivo and Is Mediated by the Nuclear Receptor Coactivator PRMT1. *Curr. Biol.* **2001**, *11* (12), 996–1000. [https://doi.org/10.1016/S0960-9822\(01\)00294-9](https://doi.org/10.1016/S0960-9822(01)00294-9).
- (93) Huang, S.; Litt, M.; Felsenfeld, G. Methylation of Histone H4 by Arginine Methyltransferase PRMT1 Is Essential in Vivo for Many Subsequent Histone Modifications. *Genes Dev.* **2005**, *19* (16), 1885–1893. <https://doi.org/10.1101/gad.1333905>.
- (94) Boisvert, F.-M.; Rhie, A.; Richard, S.; Doherty, A. J. The GAR Motif of 53BP1 Is Arginine Methylated by PRMT1 and Is Necessary for 53BP1 DNA Binding Activity. *Cell Cycle* **2005**, *4* (12), 1834–1841. <https://doi.org/10.4161/cc.4.12.2250>.
- (95) Le Romancer, M.; Treilleux, I.; Leconte, N.; Robin-Lespinnasse, Y.; Sentis, S.; Bouchekioua-Bouzaghrou, K.; Goddard, S.; Gobert-Gosse, S.; Corbo, L. Regulation of Estrogen Rapid Signaling through Arginine Methylation by PRMT1. *Mol. Cell* **2008**, *31* (2), 212–221. <https://doi.org/10.1016/j.molcel.2008.05.025>.
- (96) Qi, C.; Chang, J.; Zhu, Y.; Yeldandi, A. V.; Rao, S. M.; Zhu, Y.-J. Identification of Protein Arginine Methyltransferase 2 as a Coactivator for Estrogen Receptor  $\alpha$ . *J. Biol. Chem.* **2002**, *277* (32), 28624–28630. <https://doi.org/10.1074/jbc.M201053200>.
- (97) Meyer, R.; Wolf, S. S.; Obendorf, M. PRMT2, a Member of the Protein Arginine Methyltransferase Family, Is a Coactivator of the Androgen Receptor. *J. Steroid Biochem. Mol. Biol.* **2007**, *107* (1–2), 1–14. <https://doi.org/10.1016/j.jsbmb.2007.05.006>.
- (98) Zhong, J.; Cao, R.-X.; Hong, T.; Yang, J.; Zu, X.-Y.; Xiao, X.-H.; Liu, J.-H.; Wen, G.-B. Identification and Expression Analysis of a Novel Transcript of the Human PRMT2 Gene Resulted from Alternative Polyadenylation in Breast Cancer. *Gene* **2011**, *487* (1), 1–9. <https://doi.org/10.1016/j.gene.2011.06.022>.
- (99) Blythe, S. A.; Cha, S.-W.; Tadjuidje, E.; Heasman, J.; Klein, P. S.  $\beta$ -Catenin Primes Organizer Gene Expression by Recruiting a Histone H3 Arginine 8 Methyltransferase, Prmt2. *Dev. Cell* **2010**, *19* (2), 220–231. <https://doi.org/10.1016/j.devcel.2010.07.007>.
- (100) Tang, J.; Gary, J. D.; Clarke, S.; Herschman, H. R. PRMT 3, a Type I Protein Arginine N-Methyltransferase That Differs from PRMT1 in Its Oligomerization, Subcellular Localization, Substrate Specificity, and Regulation. *J. Biol. Chem.* **1998**, *273* (27), 16935–16945. <https://doi.org/10.1074/jbc.273.27.16935>.
- (101) Frankel, A.; Clarke, S. PRMT3 Is a Distinct Member of the Protein Arginine N-Methyltransferase Family. *J. Biol. Chem.* **2000**, *275* (42), 32974–32982. <https://doi.org/10.1074/jbc.M006445200>.
- (102) Bachand, F.; Silver, P. A. PRMT3 Is a Ribosomal Protein Methyltransferase That Affects the Cellular Levels of Ribosomal Subunits. *EMBO J.* **2004**, *23* (13), 2641–2650. <https://doi.org/10.1038/sj.emboj.7600265>.
- (103) Swiercz, R.; Person, M. D.; Bedford, M. T. Ribosomal Protein S2 Is a Substrate for Mammalian PRMT3 (Protein Arginine Methyltransferase 3). *Biochem. J.* **2005**, *386* (1), 85–91. <https://doi.org/10.1042/BJ20041466>.
- (104) Chen, D. Regulation of Transcription by a Protein Methyltransferase. *Science* **1999**, *284* (5423), 2174–2177. <https://doi.org/10.1126/science.284.5423.2174>.
- (105) Di Lorenzo, A.; Bedford, M. T. Histone Arginine Methylation. *FEBS Lett.* **2011**, *585* (13), 2024–2031. <https://doi.org/10.1016/j.febslet.2010.11.010>.
- (106) Lee, J.; Bedford, M. T. PABP1 Identified as an Arginine Methyltransferase Substrate Using High-density Protein Arrays. *EMBO Rep.* **2002**, *3* (3), 268–273. <https://doi.org/10.1093/embo-reports/kvf052>.

- (107) Guccione, E.; Bassi, C.; Casadio, F.; Martinato, F.; Cesaroni, M.; Schuchlantz, H.; Lüscher, B.; Amati, B. Methylation of Histone H3R2 by PRMT6 and H3K4 by an MLL Complex Are Mutually Exclusive. *Nature* **2007**, *449* (7164), 933–937. <https://doi.org/10.1038/nature06166>.
- (108) Hyllus, D.; Stein, C.; Schnabel, K.; Schiltz, E.; Imhof, A.; Dou, Y.; Hsieh, J.; Bauer, U.-M. PRMT6-Mediated Methylation of R2 in Histone H3 Antagonizes H3 K4 Trimethylation. *Genes Dev.* **2007**, *21* (24), 3369–3380. <https://doi.org/10.1101/gad.447007>.
- (109) Iberg, A. N.; Espejo, A.; Cheng, D.; Kim, D.; Michaud-Levesque, J.; Richard, S.; Bedford, M. T. Arginine Methylation of the Histone H3 Tail Impedes Effector Binding. *J. Biol. Chem.* **2008**, *283* (6), 3006–3010. <https://doi.org/10.1074/jbc.C700192200>.
- (110) Waldmann, T.; Izzo, A.; Kamieniarz, K.; Richter, F.; Vogler, C.; Sarg, B.; Lindner, H.; Young, N. L.; Mittler, G.; Garcia, B. A.; Schneider, R. Methylation of H2AR29 Is a Novel Repressive PRMT6 Target. *Epigenetics Chromatin* **2011**, *4* (1), 11. <https://doi.org/10.1186/1756-8935-4-11>.
- (111) Han, H.-S.; Jung, C.-Y.; Yoon, Y.-S.; Choi, S.; Choi, D.; Kang, G.; Park, K.-G.; Kim, S.-T.; Koo, S.-H. Arginine Methylation of CRT2 Is Critical in the Transcriptional Control of Hepatic Glucose Metabolism. *Sci. Signal.* **2014**, *7* (314). <https://doi.org/10.1126/scisignal.2004479>.
- (112) Lee, J.; Sayegh, J.; Daniel, J.; Clarke, S.; Bedford, M. T. PRMT8, a New Membrane-Bound Tissue-Specific Member of the Protein Arginine Methyltransferase Family. *J. Biol. Chem.* **2005**, *280* (38), 32890–32896. <https://doi.org/10.1074/jbc.M506944200>.
- (113) Kousaka, A.; Mori, Y.; Koyama, Y.; Taneda, T.; Miyata, S.; Tohyama, M. The Distribution and Characterization of Endogenous Protein Arginine N-Methyltransferase 8 in Mouse CNS. *Neuroscience* **2009**, *163* (4), 1146–1157. <https://doi.org/10.1016/j.neuroscience.2009.06.061>.
- (114) Pollack, B. P.; Kotenko, S. V.; He, W.; Izotova, L. S.; Barnoski, B. L.; Pestka, S. The Human Homologue of the Yeast Proteins Skb1 and Hsl7p Interacts with Jak Kinases and Contains Protein Methyltransferase Activity. *J. Biol. Chem.* **1999**, *274* (44), 31531–31542. <https://doi.org/10.1074/jbc.274.44.31531>.
- (115) Branscombe, T. L.; Frankel, A.; Lee, J.-H.; Cook, J. R.; Yang, Z.; Pestka, S.; Clarke, S. PRMT5 (Janus Kinase-Binding Protein 1) Catalyzes the Formation of Symmetric Dimethylarginine Residues in Proteins. *J. Biol. Chem.* **2001**, *276* (35), 32971–32976. <https://doi.org/10.1074/jbc.M105412200>.
- (116) Fabbrizio, E.; Messaoudi, S. E.; Polanowska, J.; Paul, C.; Cook, J. R.; Lee, J.; Nègre, V.; Rousset, M.; Pestka, S.; Le Cam, A.; Sardet, C. Negative Regulation of Transcription by the Type II Arginine Methyltransferase PRMT5. *EMBO Rep.* **2002**, *3* (7), 641–645. <https://doi.org/10.1093/embo-reports/kvf136>.
- (117) Kim, H.; Ronai, Z. A. PRMT5 Function and Targeting in Cancer. *Cell Stress* **2020**, *4* (8), 199–215. <https://doi.org/10.15698/cst2020.08.228>.
- (118) Yang, Y.; Hadjikyriacou, A.; Xia, Z.; Gayatri, S.; Kim, D.; Zurita-Lopez, C.; Kelly, R.; Guo, A.; Li, W.; Clarke, S. G.; Bedford, M. T. PRMT9 Is a Type II Methyltransferase That Methylates the Splicing Factor SAP145. *Nat. Commun.* **2015**, *6* (1), 6428. <https://doi.org/10.1038/ncomms7428>.
- (119) Lee, J.-H.; Cook, J. R.; Yang, Z.-H.; Mirochnitchenko, O.; Gunderson, S. I.; Felix, A. M.; Herth, N.; Hoffmann, R.; Pestka, S. PRMT7, a New Protein Arginine Methyltransferase That Synthesizes Symmetric Dimethylarginine. *J. Biol. Chem.* **2005**, *280* (5), 3656–3664. <https://doi.org/10.1074/jbc.M405295200>.

- (120) Miranda, T. B.; Miranda, M.; Frankel, A.; Clarke, S. PRMT7 Is a Member of the Protein Arginine Methyltransferase Family with a Distinct Substrate Specificity. *J. Biol. Chem.* **2004**, *279* (22), 22902–22907. <https://doi.org/10.1074/jbc.M312904200>.
- (121) Zurita-Lopez, C. I.; Sandberg, T.; Kelly, R.; Clarke, S. G. Human Protein Arginine Methyltransferase 7 (PRMT7) Is a Type III Enzyme Forming  $\omega$ -N-Monomethylated Arginine Residues. *J. Biol. Chem.* **2012**, *287* (11), 7859–7870. <https://doi.org/10.1074/jbc.M111.336271>.
- (122) Feng, Y.; Maity, R.; Whitelegge, J. P.; Hadjikyriacou, A.; Li, Z.; Zurita-Lopez, C.; Al-Hadid, Q.; Clark, A. T.; Bedford, M. T.; Masson, J.-Y.; Clarke, S. G. Mammalian Protein Arginine Methyltransferase 7 (PRMT7) Specifically Targets RXR Sites in Lysine- and Arginine-Rich Regions. *J. Biol. Chem.* **2013**, *288* (52), 37010–37025. <https://doi.org/10.1074/jbc.M113.525345>.
- (123) Cura, V.; Troffer-Charlier, N.; Wurtz, J.-M.; Bonnefond, L.; Cavarelli, J. Structural Insight into Arginine Methylation by the Mouse Protein Arginine Methyltransferase 7: A Zinc Finger Freezes the Mimic of the Dimeric State into a Single Active Site. *Acta Crystallogr. D Biol. Crystallogr.* **2014**, *70* (9), 2401–2412. <https://doi.org/10.1107/S1399004714014278>.
- (124) Fulton, M. D.; Brown, T.; Zheng, Y. G. The Biological Axis of Protein Arginine Methylation and Asymmetric Dimethylarginine. *Int. J. Mol. Sci.* **2019**, *20* (13), 3322. <https://doi.org/10.3390/ijms20133322>.
- (125) Blanc, R. S.; Richard, S. Arginine Methylation: The Coming of Age. *Mol. Cell* **2017**, *65* (1), 8–24. <https://doi.org/10.1016/j.molcel.2016.11.003>.
- (126) van Haren, M. J.; Marechal, N.; Troffer-Charlier, N.; Cianciulli, A.; Sbardella, G.; Cavarelli, J.; Martin, N. I. Transition State Mimics Are Valuable Mechanistic Probes for Structural Studies with the Arginine Methyltransferase CARM1. *Proc. Natl. Acad. Sci.* **2017**, *114* (14), 3625–3630. <https://doi.org/10.1073/pnas.1618401114>.
- (127) Antonysamy, S.; Bonday, Z.; Campbell, R. M.; Doyle, B.; Druzina, Z.; Gheyi, T.; Han, B.; Jungheim, L. N.; Qian, Y.; Rauch, C.; Russell, M.; Sauder, J. M.; Wasserman, S. R.; Weichert, K.; Willard, F. S.; Zhang, A.; Emtage, S. Crystal Structure of the Human PRMT5:MEP50 Complex. *Proc. Natl. Acad. Sci.* **2012**, *109* (44), 17960–17965. <https://doi.org/10.1073/pnas.1209814109>.
- (128) Marechal, N. Étude Structurale Des Protéine Arginine Méthyltransférases : Reconnaissance Des Substrats et Conception Rationnelle de Modulateurs, Thèse de Doctorat En Science Du Vivant, Sous La Direction de Jean Cavarelli, Université de Strasbourg, 2018. <http://www.theses.fr/2018STRAJ048/document>.
- (129) Antonysamy, S. The Structure and Function of the PRMT5:MEP50 Complex. In *Macromolecular Protein Complexes*; Harris, J. R., Marles-Wright, J., Eds.; Subcellular Biochemistry; Springer International Publishing: Cham, 2017; Vol. 83, pp 185–194. [https://doi.org/10.1007/978-3-319-46503-6\\_7](https://doi.org/10.1007/978-3-319-46503-6_7).
- (130) Chen, D.; Huang, S.-M.; Stallcup, M. R. Synergistic, P160 Coactivator-Dependent Enhancement of Estrogen Receptor Function by CARM1 and P300. *J. Biol. Chem.* **2000**, *275* (52), 40810–40816. <https://doi.org/10.1074/jbc.M005459200>.
- (131) Lee, Y.-H.; Coonrod, S. A.; Kraus, W. L.; Jelinek, M. A.; Stallcup, M. R. Regulation of Coactivator Complex Assembly and Function by Protein Arginine Methylation and Demethylination. *Proc. Natl. Acad. Sci.* **2005**, *102* (10), 3611–3616. <https://doi.org/10.1073/pnas.0407159102>.
- (132) Cura, V.; Marechal, N.; Troffer-Charlier, N.; Strub, J.; Haren, M. J.; Martin, N. I.; Cianférani, S.; Bonnefond, L.; Cavarelli, J. Structural Studies of Protein Arginine

- Methyltransferase 2 Reveal Its Interactions with Potential Substrates and Inhibitors. *FEBS J.* **2017**, *284* (1), 77–96. <https://doi.org/10.1111/febs.13953>.
- (133) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. Synthesis of Proteins by Native Chemical Ligation. *Science* **1994**, *266* (5186), 776–779. <https://doi.org/10.1126/science.7973629>.
- (134) Conibear, A. C.; Watson, E. E.; Payne, R. J.; Becker, C. F. W. Native Chemical Ligation in Protein Synthesis and Semi-Synthesis. *Chem. Soc. Rev.* **2018**, *47* (24), 9046–9068. <https://doi.org/10.1039/C8CS00573G>.
- (135) Agouridas, V.; El Mahdi, O.; Diemer, V.; Cargoët, M.; Monbaliu, J.-C. M.; Melnyk, O. Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations. *Chem. Rev.* **2019**, *119* (12), 7328–7443. <https://doi.org/10.1021/acs.chemrev.8b00712>.
- (136) Cistrone, P. A.; Bird, M. J.; Flood, D. T.; Silvestri, A. P.; Hintzen, J. C. J.; Thompson, D. A.; Dawson, P. E. Native Chemical Ligation of Peptides and Proteins. *Curr. Protoc. Chem. Biol.* **2019**, *11* (1), e61. <https://doi.org/10.1002/cpch.61>.
- (137) Kulkarni, S. S.; Sayers, J.; Premdjee, B.; Payne, R. J. Rapid and Efficient Protein Synthesis through Expansion of the Native Chemical Ligation Concept. *Nat. Rev. Chem.* **2018**, *2* (4), 0122. <https://doi.org/10.1038/s41570-018-0122>.
- (138) Pollock, S. B.; Kent, S. B. H. An Investigation into the Origin of the Dramatically Reduced Reactivity of Peptide-Prolyl-Thioesters in Native Chemical Ligation. *Chem Commun* **2011**, *47* (8), 2342–2344. <https://doi.org/10.1039/C0CC04120C>.
- (139) Thapa, P.; Zhang, R.-Y.; Menon, V.; Bingham, J.-P. Native Chemical Ligation: A Boon to Peptide Chemistry. *Molecules* **2014**, *19* (9), 14461–14483. <https://doi.org/10.3390/molecules190914461>.
- (140) Levallet, C.; Lerpiniere, J.; Ko, S. Y. The HgCl<sub>2</sub>-Promoted Guanylation Reaction: The Scope and Limitations. *Tetrahedron* **1997**, *53* (14), 5291–5304. [https://doi.org/10.1016/S0040-4020\(97\)00193-2](https://doi.org/10.1016/S0040-4020(97)00193-2).
- (141) Bernotas, R. C.; Cube, R. V. The Use of Pearlman's Catalyst for Selective *N*-Debenzylation in the Presence of Benzyl Ethers. *Synth. Commun.* **1990**, *20* (8), 1209–1212. <https://doi.org/10.1080/00397919008052829>.
- (142) Baljinnyam, B.; Ronzetti, M.; Yasgar, A.; Simeonov, A. Applications of Differential Scanning Fluorometry and Related Technologies in Characterization of Protein–Ligand Interactions. In *Targeting Enzymes for Pharmaceutical Development*; Labrou, N. E., Ed.; Methods in Molecular Biology; Springer US: New York, NY, 2020; Vol. 2089, pp 47–68. [https://doi.org/10.1007/978-1-0716-0163-1\\_4](https://doi.org/10.1007/978-1-0716-0163-1_4).
- (143) Flood, D. T.; Hintzen, J. C. J.; Bird, M. J.; Cistrone, P. A.; Chen, J. S.; Dawson, P. E. Leveraging the Knorr Pyrazole Synthesis for the Facile Generation of Thioester Surrogates for Use in Native Chemical Ligation. *Angew. Chem. Int. Ed.* **2018**, *57* (36), 11634–11639. <https://doi.org/10.1002/anie.201805191>.
- (144) Wang, Q.; Chan, T. R.; Hilgraf, R.; Fokin, V. V.; Sharpless, K. B.; Finn, M. G. Bioconjugation by Copper(I)-Catalyzed Azide-Alkyne [3 + 2] Cycloaddition. 2.
- (145) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem. Int. Ed.* **2001**, *40* (11), 2004–2021. [https://doi.org/10.1002/1521-3773\(20010601\)40:11<2004::AID-ANIE2004>3.0.CO;2-5](https://doi.org/10.1002/1521-3773(20010601)40:11<2004::AID-ANIE2004>3.0.CO;2-5).
- (146) Meldal, M.; Tornøe, C. W. Cu-Catalyzed Azide–Alkyne Cycloaddition. *Chem. Rev.* **2008**, *108* (8), 2952–3015. <https://doi.org/10.1021/cr0783479>.

- (147) Li, L.; Zhang, Z. Development and Applications of the Copper-Catalyzed Azide-Alkyne Cycloaddition (CuAAC) as a Bioorthogonal Reaction. *Molecules* **2016**, *21* (10), 1393. <https://doi.org/10.3390/molecules21101393>.
- (148) Presolski, S. Modification of Protein Scaffolds via Copper-Catalyzed Azide–Alkyne Cycloaddition. In *Protein Scaffolds*; Udit, A. K., Ed.; Methods in Molecular Biology; Springer New York: New York, NY, 2018; Vol. 1798, pp 187–193. [https://doi.org/10.1007/978-1-4939-7893-9\\_14](https://doi.org/10.1007/978-1-4939-7893-9_14).
- (149) Tornøe, C. W.; Christensen, C.; Meldal, M. Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides. *J. Org. Chem.* **2002**, *67* (9), 3057–3064. <https://doi.org/10.1021/jo011148j>.
- (150) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective “Ligation” of Azides and Terminal Alkynes. *Angew. Chem. Int. Ed.* **2002**, *41* (14), 2596–2599. [https://doi.org/10.1002/1521-3773\(20020715\)41:14<2596::AID-ANIE2596>3.0.CO;2-4](https://doi.org/10.1002/1521-3773(20020715)41:14<2596::AID-ANIE2596>3.0.CO;2-4).
- (151) Hong, V.; Presolski, S.; Ma, C.; Finn, M. G. Analysis and Optimization of Copper-Catalyzed Azide–Alkyne Cycloaddition for Bioconjugation. *Angew. Chem. Int. Ed.* **2009**, *48* (52), 9879–9883. <https://doi.org/10.1002/anie.200905087>.
- (152) Presolski, S. I.; Hong, V. P.; Finn, M. G. Copper-Catalyzed Azide–Alkyne Click Chemistry for Bioconjugation. *Curr. Protoc. Chem. Biol.* **2011**, *3* (4), 153–162. <https://doi.org/10.1002/9780470559277.ch110148>.
- (153) Mueller, L. K.; Baumruck, A. C.; Zhdanova, H.; Tietze, A. A. Challenges and Perspectives in Chemical Synthesis of Highly Hydrophobic Peptides. *Front. Bioeng. Biotechnol.* **2020**, *8*, 162. <https://doi.org/10.3389/fbioe.2020.00162>.
- (154) Liu, D. R.; Magliery, T. J.; Pastrnak, M.; Schultz, P. G. Engineering a TRNA and Aminoacyl-TRNA Synthetase for the Site-Specific Incorporation of Unnatural Amino Acids into Proteins *in Vivo*. *Proc. Natl. Acad. Sci.* **1997**, *94* (19), 10092–10097. <https://doi.org/10.1073/pnas.94.19.10092>.
- (155) Lang, K.; Chin, J. W. Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chem. Rev.* **2014**, *114* (9), 4764–4806. <https://doi.org/10.1021/cr400355w>.
- (156) Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. A General Method for Site-Specific Incorporation of Unnatural Amino Acids into Proteins. *Science* **1989**, *244* (4901), 182–188. <https://doi.org/10.1126/science.2649980>.
- (157) Koch, N. G.; Goettig, P.; Rappsilber, J.; Budisa, N. Engineering Pyrrolysyl-TRNA Synthetase for the Incorporation of Non-Canonical Amino Acids with Smaller Side Chains. *Int. J. Mol. Sci.* **2021**, *22* (20), 11194. <https://doi.org/10.3390/ijms222011194>.
- (158) Smith, E.; Zhou, W.; Shindiapina, P.; Sif, S.; Li, C.; Baiocchi, R. A. Recent Advances in Targeting Protein Arginine Methyltransferase Enzymes in Cancer Therapy. *Expert Opin. Ther. Targets* **2018**, *22* (6), 527–545. <https://doi.org/10.1080/14728222.2018.1474203>.
- (159) Jarrold, J.; Davies, C. C. PRMTs and Arginine Methylation: Cancer’s Best-Kept Secret? *Trends Mol. Med.* **2019**, *25* (11), 993–1009. <https://doi.org/10.1016/j.molmed.2019.05.007>.
- (160) van Haren, M.; van Ufford, L. Q.; Moret, E. E.; Martin, N. I. Synthesis and Evaluation of Protein Arginine N-Methyltransferase Inhibitors Designed to Simultaneously Occupy Both Substrate Binding Sites. *Org. Biomol. Chem.* **2015**, *13* (2), 549–560. <https://doi.org/10.1039/C4OB01734J>.

- (161) Gunnell, E. A.; Al-Noori, A.; Muhsen, U.; Davies, C. C.; Dowden, J.; Dreveny, I. Structural and Biochemical Evaluation of Bisubstrate Inhibitors of Protein Arginine N-Methyltransferases PRMT1 and CARM1 (PRMT4). *Biochem. J.* **2020**, *477* (4), 787–800. <https://doi.org/10.1042/BCJ20190826>.
- (162) Halby, L.; Marechal, N.; Pechalrieu, D.; Cura, V.; Franchini, D.-M.; Faux, C.; Alby, F.; Troffer-Charlier, N.; Kudithipudi, S.; Jeltsch, A.; Aouadi, W.; Decroly, E.; Guillemot, J.-C.; Page, P.; Ferroud, C.; Bonnefond, L.; Guianvarc'h, D.; Cavarelli, J.; Arimondo, P. B. Hijacking DNA Methyltransferase Transition State Analogues to Produce Chemical Scaffolds for PRMT Inhibitors. *Philos. Trans. R. Soc. B Biol. Sci.* **2018**, *373* (1748), 20170072. <https://doi.org/10.1098/rstb.2017.0072>.
- (163) Brehmer, D.; Wu, T.; Mannens, G.; Beke, L.; Vinken, P.; Gaffney, D.; Sun, W.; Pande, V.; Thuring, J.-W.; Millar, H.; Poggesi, I.; Somers, I.; Boeckx, A.; Parade, M.; Heerde, E. van; Nys, T.; Yanovich, C.; Herkert, B.; Verhulst, T.; Jardin, M. D.; Meerpoel, L.; Moy, C.; Diels, G.; Viellevoye, M.; Schepens, W.; Poncelet, A.; Linders, J. T.; Lawson, E. C.; Edwards, J. P.; Chetty, D.; Laquerre, S.; Lorenzi, M. V. Abstract DDT02-04: A Novel PRMT5 Inhibitor with Potent in Vitro and in Vivo Activity in Preclinical Lung Cancer Models. *Cancer Res.* **2017**, *77* (13\_Supplement), DDT02-04-DDT02-04. <https://doi.org/10.1158/1538-7445.AM2017-DDT02-04>.
- (164) Brehmer, D.; Beke, L.; Wu, T.; Millar, H. J.; Moy, C.; Sun, W.; Mannens, G.; Pande, V.; Boeckx, A.; van Heerde, E.; Nys, T.; Gustin, E. M.; Verbist, B.; Zhou, L.; Fan, Y.; Bhargava, V.; Safabakhsh, P.; Vinken, P.; Verhulst, T.; Gilbert, A.; Rai, S.; Graubert, T. A.; Pastore, F.; Fiore, D.; Gu, J.; Johnson, A.; Philippar, U.; Morschhäuser, B.; Walker, D.; De Lange, D.; Keersmaekers, V.; Viellevoye, M.; Diels, G.; Schepens, W.; Thuring, J. W.; Meerpoel, L.; Packman, K.; Lorenzi, M. V.; Laquerre, S. Discovery and Pharmacological Characterization of JNJ-64619178, a Novel Small-Molecule Inhibitor of PRMT5 with Potent Antitumor Activity. *Mol. Cancer Ther.* **2021**, *20* (12), 2317–2328. <https://doi.org/10.1158/1535-7163.MCT-21-0367>.
- (165) Motolani, A.; Martin, M.; Sun, M.; Lu, T. The Structure and Functions of PRMT5 in Human Diseases. *Life* **2021**, *11* (10), 1074. <https://doi.org/10.3390/life11101074>.
- (166) Bonday, Z. Q.; Cortez, G. S.; Grogan, M. J.; Antonysamy, S.; Weichert, K.; Bocchinfuso, W. P.; Li, F.; Kennedy, S.; Li, B.; Mader, M. M.; Arrowsmith, C. H.; Brown, P. J.; Eram, M. S.; Szewczyk, M. M.; Barsyte-Lovejoy, D.; Vedadi, M.; Guccione, E.; Campbell, R. M. LLY-283, a Potent and Selective Inhibitor of Arginine Methyltransferase 5, PRMT5, with Antitumor Activity. *ACS Med. Chem. Lett.* **2018**, *9* (7), 612–617. <https://doi.org/10.1021/acsmchemlett.8b00014>.
- (167) Ferreira de Freitas, R.; Ivanochko, D.; Schapira, M. Methyltransferase Inhibitors: Competing with, or Exploiting the Bound Cofactor. *Molecules* **2019**, *24* (24), 4492. <https://doi.org/10.3390/molecules24244492>.
- (168) Rodon Ahnert, J.; Perez, C. A.; Wong, K. M.; Maitland, M. L.; Tsai, F.; Berlin, J.; Liao, K. H.; Wang, I.-M.; Markovtsova, L.; Jacobs, I. A.; Cavazos, N.; Li, M.; Tolcher, A. W. PF-06939999, a Potent and Selective PRMT5 Inhibitor, in Patients with Advanced or Metastatic Solid Tumors: A Phase 1 Dose Escalation Study. *J. Clin. Oncol.* **2021**, *39* (15\_suppl), 3019–3019. [https://doi.org/10.1200/JCO.2021.39.15\\_suppl.3019](https://doi.org/10.1200/JCO.2021.39.15_suppl.3019).
- (169) Duncan, K. W.; Rioux, N.; Boriack-Sjodin, P. A.; Munchhof, M. J.; Reiter, L. A.; Majer, C. R.; Jin, L.; Johnston, L. D.; Chan-Penebre, E.; Kuplast, K. G.; Porter Scott, M.; Pollock, R. M.; Waters, N. J.; Smith, J. J.; Moyer, M. P.; Copeland, R. A.; Chesworth, R. Structure and Property Guided Design in the Identification of PRMT5 Tool Compound EPZ015666. *ACS Med. Chem. Lett.* **2016**, *7* (2), 162–166. <https://doi.org/10.1021/acsmchemlett.5b00380>.

- (170) Chan-Penebre, E.; Kuplast, K. G.; Majer, C. R.; Boriack-Sjodin, P. A.; Wigle, T. J.; Johnston, L. D.; Rioux, N.; Munchhof, M. J.; Jin, L.; Jacques, S. L.; West, K. A.; Lingaraj, T.; Stickland, K.; Ribich, S. A.; Raimondi, A.; Scott, M. P.; Waters, N. J.; Pollock, R. M.; Smith, J. J.; Barbash, O.; Pappalardi, M.; Ho, T. F.; Nurse, K.; Oza, K. P.; Gallagher, K. T.; Kruger, R.; Moyer, M. P.; Copeland, R. A.; Chesworth, R.; Duncan, K. W. A Selective Inhibitor of PRMT5 with in Vivo and in Vitro Potency in MCL Models. *Nat. Chem. Biol.* **2015**, *11* (6), 432–437. <https://doi.org/10.1038/nchembio.1810>.
- (171) Copeland, R. A. Protein Methyltransferase Inhibitors as Precision Cancer Therapeutics: A Decade of Discovery. *Philos. Trans. R. Soc. B Biol. Sci.* **2018**, *373* (1748), 20170080. <https://doi.org/10.1098/rstb.2017.0080>.
- (172) Drew, A. E.; Moradei, O.; Jacques, S. L.; Rioux, N.; Boriack-Sjodin, A. P.; Allain, C.; Scott, M. P.; Jin, L.; Raimondi, A.; Handler, J. L.; Ott, H. M.; Kruger, R. G.; McCabe, M. T.; Sneeringer, C.; Riera, T.; Shapiro, G.; Waters, N. J.; Mitchell, L. H.; Duncan, K. W.; Moyer, M. P.; Copeland, R. A.; Smith, J.; Chesworth, R.; Ribich, S. A. Identification of a CARM1 Inhibitor with Potent In Vitro and In Vivo Activity in Preclinical Models of Multiple Myeloma. *Sci. Rep.* **2017**, *7* (1), 17993. <https://doi.org/10.1038/s41598-017-18446-z>.
- (173) Sandmeyer, T. Über Isonitrosoacetanilide Und Deren Kondensation Zu Isatinen. *Helv. Chim. Acta* **1919**, *2* (1), 234–242. <https://doi.org/10.1002/hlca.19190020125>.

## COMPREHENSION STRUCTURALE DES COMPLEXES DE PRMT ET DE LEUR REGULATION DU PROCESSUS DE METHYLATION DES ARGININES

### Résumé

La méthylation des arginines par les protéines arginine méthyltransférase (PRMT) est une modification post-traductionnelle très répandue et un mécanisme épigénétique majeur de la régulation des fonctions protéiques dans les cellules eucaryotes. Cette méthylation entraîne une répercussion sur tous les aspects de la biologie cellulaire, y compris le contrôle de la transcription, l'épissage et la maturation des ARNm ou le maintien de l'intégrité du génome. De nombreuses études ont associé la dérégulation du processus de la méthylation des arginines au développement de nombreuses maladies comme les cancers mais aussi les maladies chroniques. Cependant, le manque de données structurales concernant la reconnaissance des substrats des PRMT a freiné l'émergence de ces enzymes comme cible thérapeutique. Par conséquent, caractériser précisément la structure de ces complexes PRMT-substrat et déchiffrer le mécanisme précis de la méthylation des arginines devrait fournir des informations clés sur la reconnaissance du substrat, la sélectivité et la régulation du processus de méthylation des PRMT et déboucher sur l'émergence d'inhibiteurs sélectifs. Afin de répondre à ces attentes, nous avons développé trois approches principales: (i) Synthèse de ligands mimant l'état de transition de la première ou seconde étape de méthylation des différents types de PRMT. (ii) Incorporation des mimes de l'état de transition au sein des substrats entier de PRMT. (iii) Synthèse d'inhibiteurs sélectifs de CARM1.

**Mots-clés:** PRMT, Epigénétique, SAM-Analogues, Inhibiteurs, Ligation Chimique Native

### Summary

Arginine methylation by protein arginine methyltransferase (PRMT) is a widespread post-translational modification and a major epigenetic mechanism for the regulation of protein functions in eukaryotic cells. This methylation has an impact on all aspects of cell biology, including the control of transcription, splicing and maturation of mRNAs or the maintenance of genome integrity. Many studies have associated the deregulation of the arginine methylation process with the development of many diseases such as cancers but also chronic diseases. However, the lack of structural data concerning the recognition of PRMT substrates has hampered the emergence of these enzymes as therapeutic targets. Therefore, precisely characterizing the structure of these PRMT-substrate complexes and deciphering the precise mechanism of arginine methylation should provide key insights into substrate recognition, selectivity, and regulation of the PRMT methylation process and lead to the emergence of selective inhibitors. To meet these expectations, we have developed three main approaches: (i) Synthesis of ligands mimicking the transition state of the first or second methylation step of the different types of PRMT. (ii) Incorporation of transition state mimics into full-length PRMT substrates. (iii) Synthesis of selective CARM1 inhibitors.

**Keywords:** PRMT, Epigenetic, SAM-Analogs, Inhibitors, Native Chemical Ligation