THESIS / THÈSE

DOCTOR OF SCIENCES

Structural study of non-nucleoside inhibitors of DNA (cytosine-5) methyltransferases

Rondelet, Grégoire

Award date:
2016

Awarding institution:
University of Namur

Link to publication

**UNIVERSITY OF
NAMUR**

STRUCTURAL STUDY OF NON-NUCLEOSIDE INHIBITORS OF DNA (CYTOSINE-5) METHYLTRANSFERASES

Grégoire RONDELET

Thesis Supervisor
Prof. Johan WOUTERS
Laboratory of Structural Biological Chemistry

DECEMBER 2016

University of Namur
Faculty of Sciences
Department of Chemistry
Laboratory of Structural Biological Chemistry
61 Rue de Bruxelles
5000 Namur, Belgium

Structural study of non-nucleoside inhibitors of DNA (cytosine-5) methyltransferases

Essay by **Grégoire RONDELET**
to obtain the academic degree
of Doctorate in Chemical Sciences

Jury constitution

Prof. Johan Wouters (University of Namur, Thesis Supervisor)

Prof. Daniel Vercauteren (University of Namur, Jury President)

Prof. Paulette Charlier (University of Liège)

Prof. Luc Willems (University of Liège)

Prof. Xavier De Bolle (University of Namur)

Prof. Steve Lanners (University of Namur)

DECEMBER 2016

Remerciements

Au terme de cette thèse réalisée dans le laboratoire de Chimie Biologique Structurale (CBS), je tiens à remercier les personnes qui m'ont aidé et soutenu, contribuant ainsi à la réalisation de ce travail.

Je remercie le Prof. Johan Wouters, mon promoteur de thèse, de m'avoir encouragé à réaliser un doctorat au sein de son laboratoire. Je le remercie également pour tous les entretiens et les conseils prodigués pour mener à bien ce projet.

Mes remerciements vont également au Prof. Luc Willems du laboratoire d'Epigénétique Cellulaire et Moléculaire (GIGA-Cancer) de l'université de Liège et aux membres de son laboratoire, spécialement Alix de Brogniez, Jacques Jean-Rock et Hamaidia Malik pour leur accueil et leur aide lors de la réalisation des essais de cytotoxicité.

J'ai pu rencontrer, au cours des différents séjours au synchrotron SOLEIL (Gif-sur-Yvette, France), des personnes passionnées, disponibles et prodiguant de bons conseils. À ce titre, je tiens ici à remercier Andrew Thompson, Pierre Legrand et Leonard Chavas de la ligne de lumière PROXIMA 1 et William Shepard de la ligne de lumière Proxima 2. De plus, je tiens à mettre en avant le fait que les analyses réalisées à SOLEIL ont pu se faire grâce au financement du FNRS et au travail de coordination du Prof. Paulette Charlier de l'Université de Liège.

Je remercie l'Unité de Pharmacochimie de la Régulation Epigénétique du Cancer de Toulouse (ETaC) et notamment le Prof. Paola Arimondo pour la collaboration entamée au cours de cette thèse.

Je tiens également à remercier la Plateforme Technologique de calcul Intensif (PTCI), membre du Consortium des Équipements de Calcul Intensif (CÉCI), pour les ressources calculatoires mises à disposition. Mes remerciements vont, tout particulièrement, aux informaticiens Frédéric Wautelet et Laurent Demelenne pour leur aide précieuse.

Je remercie l'UNamur et le Télévie de m'avoir permis, financièrement, de réaliser cette thèse.

Je remercie les membres de mon jury d'avoir accepté de faire partie de mon comité de thèse, plus précisément le Prof. Johan Wouters, le Prof. Daniel Vercauteren, le Prof. Paulette Charlier, le Prof. Luc Willems, le Prof. Xavier De Bolle et le Prof. Steve Lanners.

Je tiens ici à souligner le travail réalisé par les stagiaires (Justine Beauquenne et Nathalie Magema Pono) ainsi que les mémorants (Quentin Thémans, Thomas Dal Maso et Antonin Maniquet) que j'ai encadrés au cours de cette thèse.

Mes remerciements vont également aux membres du laboratoire CBS, anciens et actuels, pour les discussions stimulantes et les bons moments passés ensemble.

Pour terminer, je remercie chaleureusement ma famille et mes amis pour leur soutien constant au cours de ces quatre années de thèse.

Table of Contents

Chapter 3. Structural basis for recognition of histone H3K36me3 nucleosome by human *de novo* DNA methyltransferases 3A and 3B..79

Chapter 4. Targeting PWWP domain of DNA methyltransferase 3B for epigenetic cancer therapy: Identification and structural characterization of new potential protein-protein interaction inhibitors 117

Preface

This PhD project was supported by the Belgian National Fund for Scientific Research (Grant F.R.S.-FNRS-Télévie 7.4.532.15.F.) thanks to the Télévie campaign, an operation begun in 1989 to raise funds for scientific cancer research in children and adults. This work will therefore focus on cancer research and, especially, deal with the epigenetic therapy of cancer. This issue was first addressed during my master's thesis at the University of Namur (UNamur) completed in 2012 in the Laboratory of Structural Biological Chemistry directed by Professor Johan Wouters.

Tumorigenesis and tumor maintenance are controlled by genetic events but also by epigenetic events. These events control the expression of genes by modifications on ("epi-") DNA and histones (proteins wrapping DNA into nucleosomes). As these modifications are reversible, it is possible to revert the cell to a normal state leading to an inhibition of cancer cell growth. There is a plethora of proteins involved in the epigenetic regulation of gene expression which may be considered as epigenetic targets for chemotherapy. Among these, we focused our research on human DNA (cytosine-5) methyltransferases (DNMTs) which catalyze the methylation of cytosine nucleotide. This epigenetic modification, 5-methylcytosine, can lead to the development and progression of cancer and, therefore, the discovery and optimization of novel epi-drugs targeting DNMTs are needed.

To this end, structural studies of complexes between DNMTs and non-nucleoside inhibitors, as well as an identified epigenetic mark guiding DNA methylation, were carried out. To achieve this goal, several techniques including, protein expression and purification, crystallography, molecular modeling and enzymology were used during this thesis. The following manuscript, containing results obtained during the four years of doctorate study, is presented as an article thesis. This consists of an introduction, three main chapters and a conclusion part including perspectives for future work.

Here, I would especially like to thank the Télévie for the grant, all authors of publications and the different collaborators met to achieve the intended goals of my thesis research.

I hope you enjoy your reading.   Grégoire Rondelet

Namur, December 2016

List of abbreviations

ADD	ATRX–DNMT3–DNMT3L
ADMET	Absorption, Distribution, Metabolism, and Excretion-Toxicity
ADSC	Area Detector Systems Corporation
AML	Acute Myeloid Leukemia
BET	Bromodomain and Extra-Terminal
CAPRI	Critical Assessment of PRedicted Interactions
CD	Catalytic Domain
CGI	CpG island
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CHARMm	Chemistry at HARvard Macromolecular mechanics
CMML	Chronic Myelomonocytic Leukemia
CpG	Cytosine-phosphate-Guanine
CXXC	Cysteine-rich
DMAP	DNA methyltransferase associated protein
DMSO	Dimethyl sulfoxide
DNMTi	DNA methyltransferase inhibitor
DNMT	DNA (cytosine-5) methyltransferase
DSF	Differential Scanning Fluorimetry
EDTA	Ethylenediaminetetraacetic acid
EGCG	(-)-Epigallocatechin-3-gallate
EMA	European Medicines Agency
FDA	Food and Drug Administration
FPLC	Fast protein liquid chromatography
GA	Genetic Algorithm
GLP	G9a-like protein
GST	Glutathione S-Transferase
H3K36me3	Trimethylated lysine 36 of histone 3
HADDOCK	High Ambiguity Driven biomolecular DOCKing
HDAC	Histone deacetylase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HGP	Human Genome Project
HPLC	High-performance liquid chromatography
ICF	Immunodeficiency, Centromeric instability and Facial abnormalities
IMAC	Immobilized Metal ion Affinity Chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
LB	Lysogeny Broth
LEDGF	Lens Epithelium-Derived Growth Factor
M.*Hha*I	*Haemophilus haemolyticus* DNA methyltransferase
M.*Sss*I	CpG Methyltransferase *Spiroplasma sp.* strain MQ1

MBP	Methyl-CpG-binding domain protein
MBT	Malignant Brain Tumor
MDS	Myelodysplastic syndromes
MMFF	Merck Molecular Force Field
MTS	Aqueous soluble tetrazolium/formazan
MWCO	Molecular weight cut-off
NCP	Nucleosome Core Particle
NLS	Nuclear Localization Signal
OD	Optical Density
PDB	Protein Data Bank
PEG	Polyethylene glycol
PHD	Plant homeodomain
PIPES	Piperazine-N,N'-bis(2-ethanesulfonic acid)
PMS	Phenazine methosulfate
PPI	Protein-rotein interaction
PTD	Protein transduction domain
PWWP	Pro–Trp–Trp–Pr
RFTS	Replication foci targeting sequence
RMSBB	Backbone root-mean-square deviation
RMSD	Root-mean-square deviation
SAH	S-adenosyl-L-homocysteine
SAM	S-adenosyl-L-methionine
SAR	Structure-activity relationship
SDF	System Data Format
SDS–PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SOLEIL	Source Optimisée de Lumière d'Énergie Intermédiaire du LURE
TB	Terrific Broth
TE	Tris-EDTA
TEV	Tobacco Etch Virus
TF	Transcription Factor
TPSA	Topological Polar Surface Area
TSG	Tumor Suppressor Gene
XDS	X-ray Detector Software
ΔTm	Variations of the melting temperature

Table of amino acids and their abbreviations

Full Name	Three letter code	One letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Résumé

Chez les eucaryotes, la méthylation de l'ADN est une importante modification épigénétique contrôlant l'expression des gènes. Cette réaction de méthylation est catalysée par des enzymes appelées « ADN (cytosine-5) méthyltransférases » (DNMTs). Dans les cellules normales, la méthylation de l'ADN est impliquée, entre autres, dans la différenciation cellulaire lors de l'embryogenèse, l'inactivation d'un des chromosomes X chez la femme et la stabilité chromosomique. Cependant, dans la plupart des cancers, ce processus de méthylation est dérégulé et conduit au développement et à la progression des cellules tumorales. Comme cette modification est réversible, il est possible de renverser le profil de méthylation en inhibant l'activité des DNMTs et, ainsi, arrêter la progression tumorale.

Cette thèse porte sur l'étude structurale d'inhibiteurs non-nucléosidiques des DNMTs. En particulier, nous avons étudié des dérivés maléimides du RG108 (RG108-1 et RG119-1). Ces composés ont le potentiel de réactiver des « gènes suppresseurs de tumeurs » réduits au silence par méthylation de leur promoteur. Les résultats ont montré une corrélation entre la cytotoxicité sur des cellules de mésothéliome et leur puissance inhibitrice envers les DNMTs. Des études théoriques, appuyées par des études cristallographiques (apo structure de la M.*Hha*I) et de fluorimétrie à balayage différentiel, ont permis de déterminer leur mode d'action. Nous nous sommes également intéressés à une modification post-traductionnelle des histones, la triméthylation de la lysine 36 de l'histone 3 (H3K36me3). Ce marqueur épigénétique permet le recrutement des DNMT3s, via leur domaine PWWP, à des régions génomiques impliquées dans l'oncogenèse. La résolution structurale de cette reconnaissance biologique (domaine PWWP de la DNMT3B avec le marqueur épigénétique H3K36me3) et la construction d'un modèle complet de la DNMT3A en complexe avec un dinucléosome nous ont permis de proposer un mécanisme de reconnaissance génomique des DNMT3s. À partir de ce complexe, nous avons pu développer des inhibiteurs potentiels pouvant perturber l'interaction entre le marqueur épigénétique H3K36me3 et le domaine PWWP des DNMT3s. L'inhibition de cette interaction protéine-protéine empêcherait la méthylation intragénique d'oncogènes, qui est responsable de leur transcription.

Abstract

In eukaryotes, DNA methylation is an important epigenetic modification involved in gene regulation. This methylation is catalyzed by DNA (cytosine-5) methyltransferases (DNMTs). In normal cells, DNA methylation is involved, *inter alia*, in cell differentiation, embryonic development, X-chromosome inactivation and maintenance of chromosomal stability. However, in most cancers, this methylation process is deregulated and leads to the development and progression of tumor cells. As this modification is a reversible process, inhibitors targeting DNMTs are promising anticancer agents to reverse the deregulated methylome to stop cancer proliferation.

This thesis deals with the structural study of non-nucleoside inhibitors of DNMTs. In particular, we studied maleimide derivatives of RG108 (RG108-1 and RG119-1) which have the potential to reactivate tumor suppressor genes. Findings demonstrated a correlation between cytotoxicity on mesothelioma cells of these compounds and their inhibitory potency against DNMTs. Non-covalent and covalent docking studies, supported by crystallographic (apo structure of M.*Hha*I) and differential scanning fluorimetry assays, provided detailed insights into their mode of action and revealed essential residues for the stabilization of such compounds inside DNMTs. We were also interested in a post-translational modification of histones, the trimethylation of lysine 36 of histone 3 (H3K36me3). This epigenetic mark allows the recruitment of DNMT3s, via their PWWP domain, for intragenic methylation of oncogenes. The structural resolution of this biological system (DNMT3B PWWP/H3K36me3) and the construction of a complete model of the DNMT3A in complex with a dinucleosome allowed us to propose a genomic recognition mechanism for DNMT3s. This complex led us to identify new potential inhibitors disrupting the interaction between the histone mark H3K36me3 and the DNMT3s PWWP domain. The inhibition of this protein-protein interaction would prevent the intragenic methylation of oncogenes, which is responsible for their transcription.

INTRODUCTION

Chapter 1. Human DNA (cytosine-5) methyltransferases: a functional and structural perspective for epigenetic cancer therapy

Grégoire Rondelet* and Johan Wouters

Department of Chemistry, NAmur MEdicine & Drug Innovation Center (NAMEDIC-NARILIS), University of Namur, 61 rue de Bruxelles, B-5000 Namur, Belgium

*Corresponding author: gregoire.rondelet@unamur.be

† **In preparation**

Personal contribution: Performed literature search, generated figures and tables, and wrote the manuscript.

Abstract

Epigenetic modifications modulate chromatin states to regulate gene expression. Among them, DNA methylation and histone modifications play a crucial role in the establishment of the epigenome. In cancer, these epigenetic events may act in concert to repress tumor suppressor genes or promote transcription of oncogenes. In the context of cancer initiation and progression, recruitment of DNA (cytosine-5) methyltransferases to specific genomic regions is mediated by histone epigenetic marks, transcription factors and co-regulators. This introductory chapter will review these mechanisms and present state-of-the-art of DNA methylation, treatment and development of epigenetic cancer therapies targeting this epigenetic modification.

Keywords: Cancer, DNA methylation, Promoter hypermethylation, Gene body methylation, DNMT, DNMT inhibitors, Epigenetic therapy

1.1 From genome to epigenome

Initiated in 1990, the Human Genome Project finished sequencing and mapping all the genes of Homo sapiens in 2003 (Consortium, 2004). This complete genome, ~3.10^9 DNA base pairs distributed among 23 chromosome pairs, contains all the genetic information essential for human life. With the complete genome sequence as a powerful tool, research tends to understand cancer by discovering genetic mutations located in certain genomic regions and the functions of various genes in order to establish new treatment strategies. However, genome sequence is just a part of a more complex system involved in gene expression regulation.

Indeed, DNA is a dynamic molecule organized in chromatin structure in which nucleosomes (histone octamer core around which DNA is wrapped) are the basic units. These units can be arranged to give an open (euchromatin) or a closed (heterochromatin) form of chromatin in local regions (Fig. 1). Euchromatin is associated with transcriptional activation, whereas heterochromatin blocks DNA accessibility to transcription machinery and conducts to inactivation of gene expression (Keshet et al., 1986). This chromatin remodeling is controlled essentially by chemical modifications on DNA (DNA methylation) and histone tails (methylation and/or acetylation) which define the epigenome (Fig. 1).
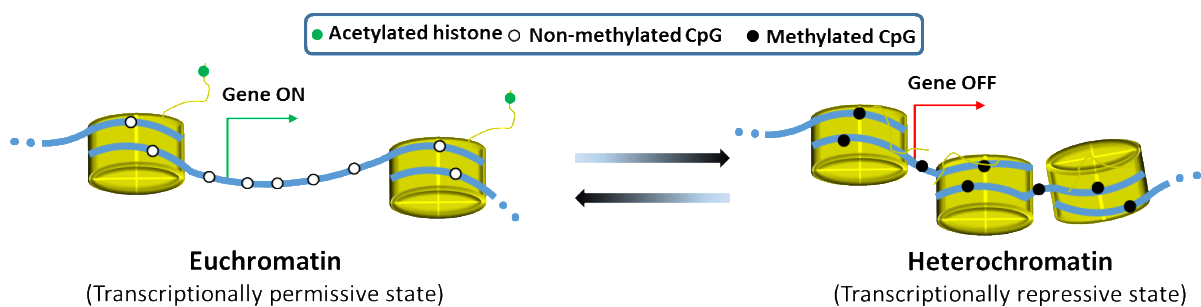
Figure 1. Representation of a local chromatin modification regulating gene expression. *Depending on the cellular context, the epigenetic modifications, such as DNA methylation and histone modifications, will modulate the chromatin state to control gene expression. Euchromatin and heterochromatin result from these modifications and, consequently, are associated with distinct epigenomic patterns. Euchromatin is transcriptionally permissive, while heterochromatin is transcriptionally repressive.*

These modifications, which occur without a change in DNA sequence, are heritable, reversible and response to external signals (Berger et al., 2009; Jaenisch and Bird, 2003; Wolffe and Matzke, 1999). Epigenetic events are essential for interpreting the genome in normal and abnormal cells. Recently, a great progress in the comprehensive analysis of the chromatin remodeling was reached with a paper published in PNAS (Almassalha et al., 2016). This study revealed a new imaging technique to study the chromatin dynamic and organization in real time from the nucleosomal (10 nm) to chromosomal (> 200 nm) length scales allowing, for example, to investigate the effect of a drug on cancer cells. As an overview, genome is the blueprint for coding proteins, and the epigenome controls the access to this genetic information. In 2008, an international project, named "Alliance for the Human Epigenome and Disease", was set up to decode the human epigenome, just as the Human Genome Project, to have a better understanding of disease mechanisms with direct implications in prevention, diagnosis and treatment of human diseases (Jones et al., 2008). In this spirit, the US National Institutes of Health Roadmap Epigenomics Mapping Consortium produces a public resource of epigenome maps and proposes standardization of protocols for the research community (Bernstein et al., 2010).

Modulation of local chromatin states to regulate gene expression is achieved by a plethora of chromatin-binding proteins assigned to different mechanisms such as editing the epigenome (writing or erasing it) or reading it. A nice example combining these mechanisms is the interplay between DNA methylation and histone modifications to silence genes: the "writers" DNA methyltransferases (DNMTs) deposit DNA methylation on cytosine in CpG (cytosine-phosphate-guanine) dinucleotides, then "readers" methyl-CpG-binding domain proteins (MBPs) recognise methylated DNA and recruit histone deacetylases (HDACs) as "erasers" of histone acetylation. Resulting hypoacetylated histones lead to chromatin condensation (heterochromatin), preventing DNA binding factors and, consequently, form a stable gene silencing (Fig. 1) (Bird, 2002; Klose and Bird, 2006).

1.2 Epigenome in diseases

Epigenetic modifications control the precise expression of cell genes and are then crucial for cell differentiation, embryonic development, X-chromosome inactivation and maintenance of chromosomal stability (Fig. 2) (Mohn and Schübeler, 2009; Riggs, 1975; Wutz, 2011). However, epigenetic dysregulation occurs in multiple diseases, including diabetes, cardiopulmonary diseases, neurological disorders, imprinting disorders, autoimmune diseases and cancer (Esteller, 2008; Jones, 2012; Jones et al., 2008; Shirohzu et al., 2002; Tatton-Brown et al., 2014; Wang et al., 2008). As epigenetic modifications are heritable and reversible in somatic cells, epigenomic patterns could be reversed to a normal cell phenotype by drug treatments. In this sense, Epi-drugs (chromatin-modulating agents) targeting DNA methylation, histone deacetylation, methylation and demethylation, or histone readers are developed to treat cancer (Erdmann et al., 2015; Falkenberg and Johnstone, 2014; Miranda et al., 2009; Zhao et al., 2013).

1.3 Cancer DNA methylome

To introduce this part, the following timeline (Fig. 2) presents the fundamental findings of DNA methylation, which is one of the most important epigenetic modifications.

Fundamental Findings of DNA methylation

DNA methylation affects the formation of active chromatin (Keshet et al. **1986**)

DNMT3A/3B are essential for *de novo* methylation (Okano et al. **1999**)

Relationship between nucleosome positioning and DNA methylation (Chodavarapu et al. **2010**)

X inactivation, differentiation, and DNA methylation (Riggs **1975**)

Hypermethylation in human neoplasia (de Bustros et al. **1988**)

DNMT3L stimulates activity of DNMT3A and DNMT3B (Suetake et al. **2004**)

DNA methylation linked to deacetylation of histones (Eden et al. **1998**)

Gene body methylation as therapeutic target in cancer (Yang et al. **2014**)

*Figure 2. Timeline of fundamental findings of DNA methylation. (Riggs, **1975**): DNA methylation explains the initiation and maintenance of mammalian X inactivation. (Keshet et al., **1986**): CpG methylation is linked with alterations in chromatin structure and gene silencing. DNA methylation affects interactions between DNA and nuclear proteins. (de Bustros et al., **1988**): One of the first example of hypermethylation of tumor suppressor genes in a human cancer associated with gene*

*inactivation. (Eden et al., **1998**): Proteins belong to the family of methyl-CpG-binding domain proteins (MBPs) recruit histone deacetylases (HDACs) to mediate transcriptional repression. (Okano et al., **1999**): DNMT3A and DNMT3B are required for de novo methylation. (Suetake et al., **2004**): DNMT3L (DNA methyltransferase 3-like protein) stimulates the activity of DNMT3A and DNMT3B for maternal methylation imprinting. (Chodavarapu et al., **2010**): DNMTs preferentially target nucleosome-bound DNA. (Yang et al., **2014**): Gene body methylation is correlated with gene transcription and represents a new therapeutic target.*

DNA methylation occurs on cytosine at the C5 position (5-methylcytosine, 5-mC) of CpG (Cytosine-phosphate-Guanine) dinucleotides and is catalyzed by DNA methyltransferases (DNMTs) (Fig. 3) (Bestor, 2000).
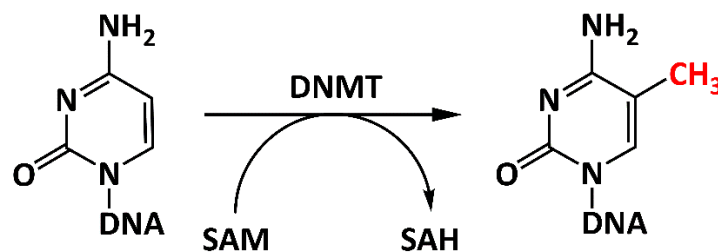
Figure 3. DNA methylation at the C5 position of cytosine catalyzed by DNMTs in the presence of the SAM cofactor as methyl group donor. *SAM: S-adenosyl-L-methionine, SAH: S-adenosyl-L-homocysteine*

The distribution of 5-methyl cytosine throughout the genome depends on the genomic and epigenomic contexts during development and disease (Hackett and Surani, 2013). As an important remark, the function of DNA methylation is related to CpG density and their localization on the genome. Thereby, methylation of promoter CpG islands (CGIs) is correlated with transcriptional silencing of the gene, whereas intragenic methylation, or so-called gene body methylation, at lower CpG density or certain CGIs is associated with gene expression (Fig. 4) (Ball et al., 2009; Rakyan, 2008; Deaton et al., 2011; Meissner et al., 2008; Nguyen et al., 2001; Weber et al., 2007).
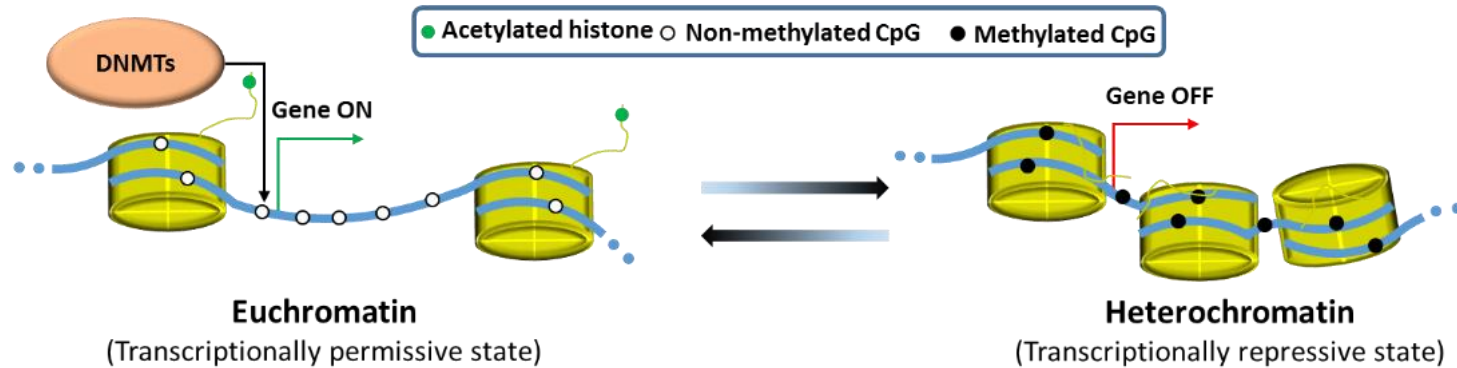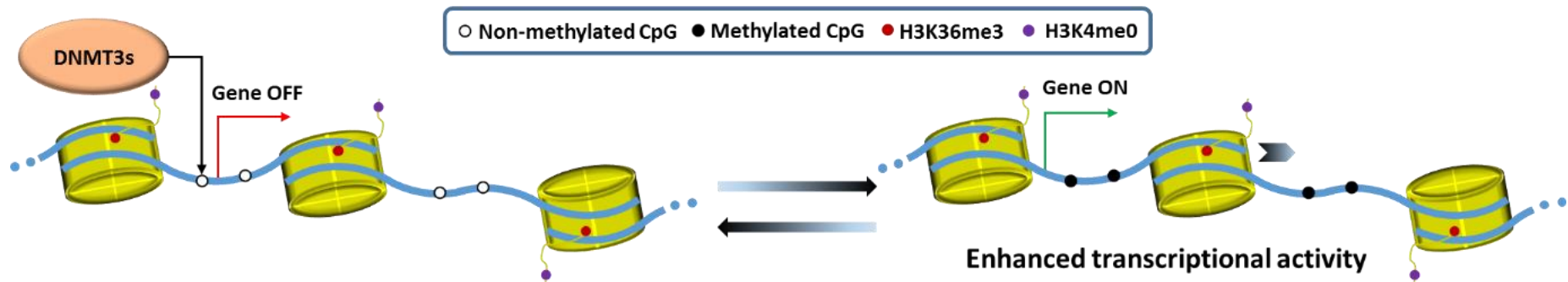
Figure 4. **Proposed model for dual function of DNA methylation in gene regulation.** *Promoter methylation (top): hypermethylation of CGIs in a gene promoter is associated with gene inactivation (de Bustros et al., 1988). Gene body methylation (bottom): In cancer, gene body CpGs are preferential sites for de novo methylation (Nguyen et al., 2001). This gene body methylation is associated with gene expression (Ball et al., 2009). Combination of permissive histone epigenetic marks H3K4me0 and H3K36me3 are necessary for gene body methylation (Stewart et al., 2015; Tomizawa et al., 2012).*

When DNA methylation is associated with transcriptional repression, the heterochromatin is formed and the compacted nucleosomes block RNA polymerase transcription (Jones and Liang, 2009). Conversely, gene body methylation promotes transcriptional elongation (Ball et al., 2009). Deregulation of these DNA methylation patterns contribute to tumorigenesis and tumor maintenance. Indeed, in cancer, promoter hypermethylation of CpG islands in tumor suppressor genes (TSGs) promoters (e.g., TP53 encoding p53 essential for apoptosis induction) leads to their inactivation (de Bustros et al., 1988; Hansen et al., 2011; Irizarry et al., 2009; Rideout et al., 1990). On the other hand, gene body methylation is associated with activation of oncoprotein-regulated genes as observed in colon cancer cell lines (Yang et al., 2014).

1.4 Human DNA (cytosine-5) methyltransferases

Human DNMTs are classified in two families (DNMT1 and DNMT3s) that are structurally and functionally distinct (Cheng and Blumenthal, 2008). DNMT3A and DNMT3B are *de novo* methyltransferases as they establish the initial CpG methylation pattern during embryogenesis (Li et al., 2007; Okano et al., 1999). They are coupled with the inactive DNMT3-Like (DNMT3L) to stimulate their activity (Chédin et al., 2002; Jia et al., 2007). The pattern of methylation is then maintained during chromosome replication by DNMT1 (Chen and Li, 2006).

Structurally, these enzymes (DNMT1 and DNMT3s) possess a N-terminal regulatory domain and a C-terminal catalytic domain (Fig. 5) (Jeltsch, 2002; Jurkowska et al., 2011). In mammals, the N-terminal domain guides their localization to chromatin by interacting with other proteins and DNA, and regulates their intrinsic activity. The C-terminal catalytic domain has six highly conserved motifs involved in DNA binding and catalysis of methyl transfer. Residues in motifs I and X are involved in the cofactor binding, motifs IV, VI and VIII are responsible for the target-base binding and the catalysis, and motif IX is involved in DNA recognition (Fig. 5 and Fig. S4 in chapter 2). DNMT3L lacks essentials motifs for DNA and cofactor binding (IX and X) and is catalytically inactive (Bourc'his et al., 2001).
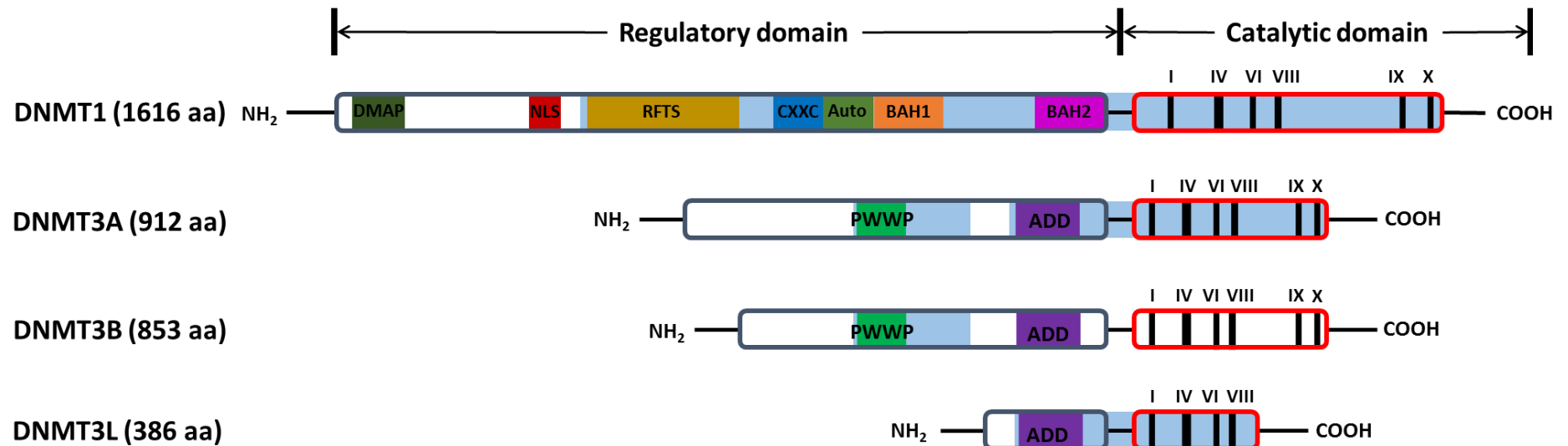
Figure 5. Schematic representation of the domain structure of mammalian DNMT isoforms. *Blue portions reveal domains structurally solved. I, IV, VI, IX and X are conserved motifs of the active C-terminal catalytic domain of DNMTs (red box). DNMT1 has a large regulatory domain comprising a DNA methyltransferase associated protein (DMAP), a nuclear localization signal (NLS), a replication foci targeting sequence (RFTS), a cysteine-rich (CXXC) DNA binding domain and bromo-adjacent homology domains 1 and 2 (BAH1 and BAH2). The autoinhibitory linker (Auto) between CXXC–BAH1 prevents de novo methylation (Song et al., 2011). N-terminal part of DNMT3A and DNMT3B have an ATRX–DNMT3–DNMT3L (ADD) domain and Pro–Trp–Trp–Pro (PWWP) domain for activity regulation and nucleosome recognition (Baubec et al., 2015; Guo et al., 2015). Catalytically inactive DNMT3L possesses only an ADD domain on the N-terminal part. This isoform stimulates the activity of DNMT3A and DNMT3B (Jia et al., 2007).*

1.5 Bacterial DNA methyltransferases - Their use as model in cancer research

Bacterial DNMTs possess only the C-terminal catalytic domain required for methyltransferase activity. In prokaryotes, DNA methylation (adenine or cytosine methylation) controls initiation of DNA replication, protects cells from foreign DNA through restriction-modification systems and plays an essential role in bacterial virulence (e.g., *Mycobacterium tuberculosis* and *Brucella*) (Low et al., 2001; Shell et al., 2013).

In cancer research, two C5-DNA methyltransferase bacteria (M.*Sss*I from *Spiroplasma* and M.*Hha*I from *Haemophilus haemolyticus*) serve as model to investigate the inhibition mechanism of human DNMTs. Indeed, they share a high structural homology and sequence identity with human DNMTs for the conserved motifs involved in DNA binding and catalysis of methyl transfer (see Fig. S4 in chapter 2). M.*Sss*I (376 aa) completely methylates CpG sequences in a processive manner and was validated for use in activity assays in epigenetic studies due to its high efficiency and fast reaction rate (Renbaum et al., 1990). Conversely, M.*Hha*I (325 aa) recognizes and methylates the sequence 5'-GCGC-3' in a non-processive manner but is largely studied in crystallography and constitutes an appropriate model to investigate the inhibition mechanism of human DNMTs.

1.6 Current treatments targeting DNA methylome

Actually, only two drugs received FDA (Food and Drug Administration) approval as demethylating agent for myelodysplastic syndromes, acute myeloid leukemia and chronic myelomonocytic leukemia: the cytosine analogs 5-azacytidine (Vidaza®) and decitabine (Dacogen®) (Fig. 6) (Christman, 2002a; Issa and Kantarjian, 2009; Jones and Taylor, 1980; Wijermans et al., 2008). An extension of the use of these drugs to the European market was applied in 2008 for Vidaza® and in 2012 for Dacogen® by the European Medicines Agency. These inhibitors are incorporated into the DNA, and they covalently trap DNMTs (Christman, 2002a; Issa and Kantarjian, 2009; Jones and Taylor, 1980). However, due to their high toxicity, low specificity, chemical instability and poor bioavailability during cancer therapy, development of a new generation of nucleoside drugs is ongoing (Christman, 2002b; Lyko and Brown, 2005). Among these, the prodrug SGI-110 derived from decitabine presents improvement in pharmaceutical properties and is under investigation in phase 2 clinical trials for the treatment of myelodysplastic syndromes and acute myeloid leukemia (Fig. 6) (Chuang

et al., 2010; Issa et al., 2015). Combinatorial therapy approach using nucleoside analogs is also used in clinical trials. In epigenetic therapy, combination of Vidaza® and Entinostat® has a synergistic therapeutic effect on solid tumors and abolished chemoresistance in cancer cells (Clozel et al., 2013; Juergens et al., 2011).

In parallel to these advances in research on nucleoside analogs, non-nucleoside compounds are developed to increase specificity and selectivity towards DNMTs, and less cytotoxicity by acting directly on DNMTs without incorporation into DNA. For example, drugs used for other indications (hydralazine (antihypertensive), procainamide (antiarrhythmic), and procaine (local anesthetic)), several natural compounds (parthenolide, nanaomycin A, (-)-epigallocatechin-3-gallate (EGCG, a polyphenol from green tea leaves), curcumin, genistein and laccaic acid A) and inhibitors discovered from virtual screening campaigns (e.g., RG108) have been identified as DNMT inhibitors (Fig. 6) (Erdmann et al., 2015).

As DNA methylation occurs in normal cells, selectivity towards DNMTs is a challenging task to personalize treatment for cancer, and so to decrease side effects. To achieve this goal, current studies tend to identify the contribution of each DNMT in various cancer cells. For example, DNMT3A mutations occur in acute myeloid leukemia (Ley et al., 2010; Shlush et al., 2014) and this isoform is overexpressed in melanoma cell lines (Deng et al., 2009), while DNMT3B is overexpressed in breast and colorectal cancers (Nosho et al., 2009; Roll et al., 2008). It was also shown that DNMT1 and DNMT3B cooperate to maintain gene silencing in cancer cells (Rhee et al., 2002). The development of selective inhibitors for a particular DNMT isoform is therefore an important issue. However, development of non-nucleoside compounds targeting the catalytic domain is slowed by lack of selectivity or structure–activity relationships information. The only complex reported so far is the one between bacterial C5-DNA methyltransferase M.*Hha*I and a nucleoside inhibitor, zebularine (PDB code: 1M0E) (Fig. 6) (Zhou et al., 2002).
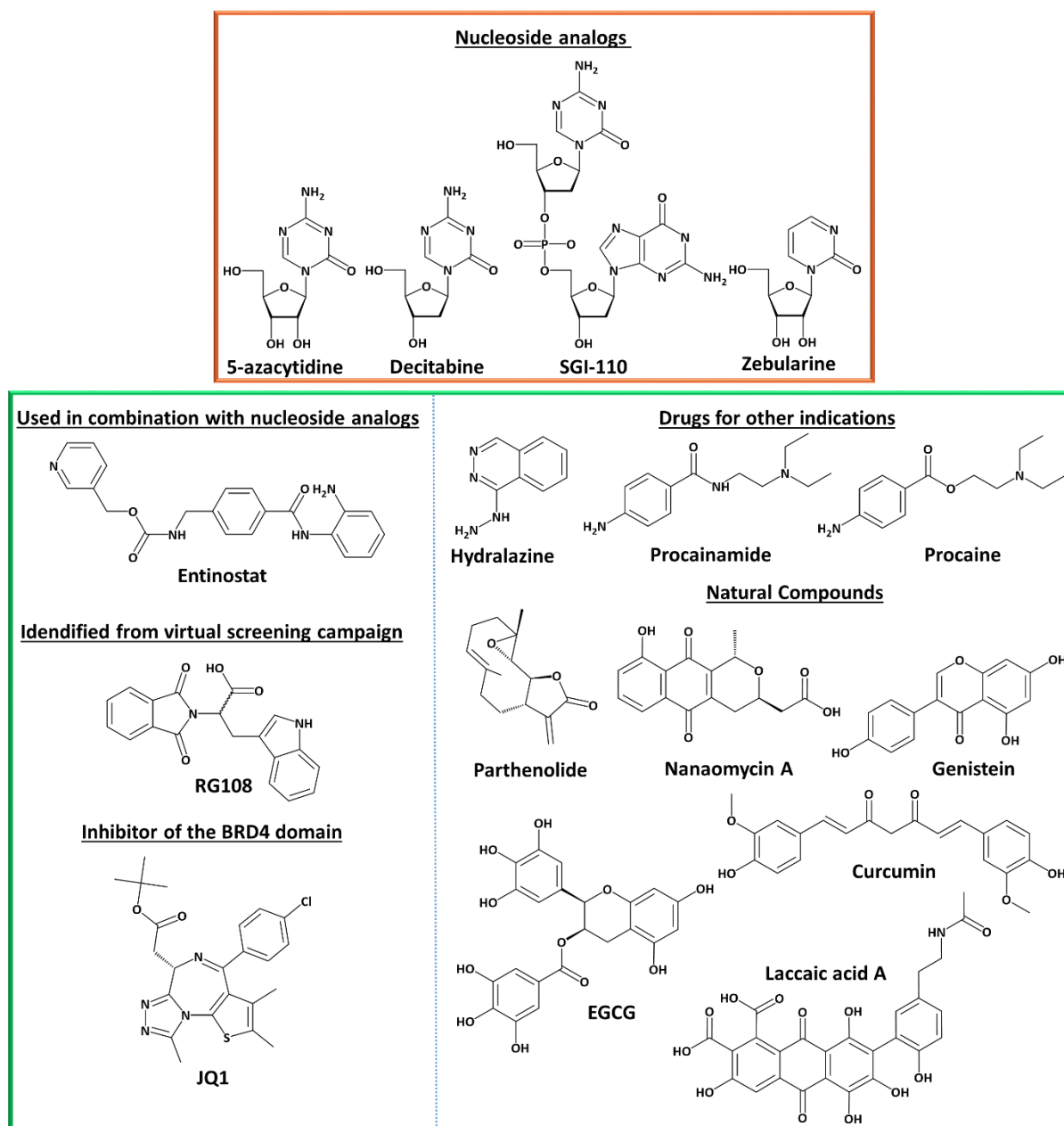
Figure 6. Chemical structures of DNA methyltransferase inhibitors and JQ1 molecule as protein-protein inhibitor example. *The first approved drugs for the treatment of myelodysplastic syndromes and acute myeloid leukemia are nucleoside analogs 5-azacytidine and Decitabine (orange top box). New studied non-nucleoside compounds (green bottom box) include drugs used for other indications, natural compounds and RG108, a compound identified from a virtual screening campaign.*

1.7 Targeting transcription regulation in cancer: DNMTs cooperativity

The following timeline (Fig. 7) presents all the different types of DNMTs complexes which will be discussed later (in parts 1.7.1 and 1.7.2).
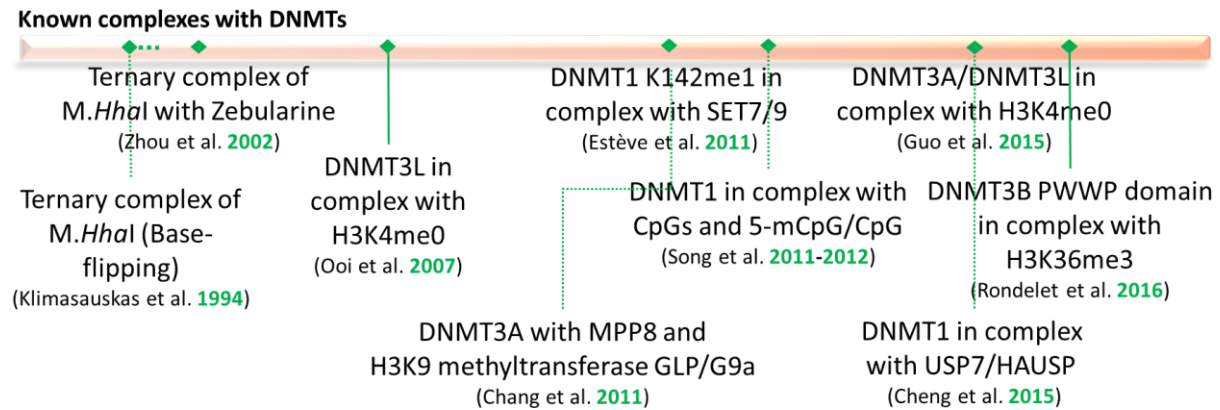
Figure 7. Timeline of known crystal complexes with DNMTs. *(Klimasauskas et al., **1994**): M.HhaI flips its target base out of the DNA helix (PDB code: 1MHT). (Zhou et al., **2002**): First structure of M.HhaI in complex with a nucleoside inhibitor (PDB code: 1M0E). (Ooi et al., **2007**): DNMT3L recognition of nucleosome via the ADD domain and the unmethylated histone H3K4 tail (PDB code: 2PVC). (Chang et al., **2011**): chromodomain of MPP8 recognizes the methylated DNMT3A and the self-methylated GLP to form a repressive complex (PDB code: 3SW9 and 3SWC). (Estève et al., **2011**): Lys142 methylation is a key signal for degradation of DNMT1 (PDB code: 3OS5). (Song et al., **2012**): Structural insights of the mechanism of DNA methylation maintenance by DNMT1. (Guo et al., **2015**): Regulation of DNMT3A activity by the unmethylated histone H3K4 tail and DNMT3L (PDB code: 4U7T and 4U7P). (Cheng et al., **2015**): USP7 (also known as HAUSP) interacts with DNMT1 and stimulates its activity (PDB code: 4YOC). (Rondelet et al., **2016**): DNMT3B PWWP recognizes the permissive histone epigenetic mark H3K36me3 (PDB code: 5CIU).*

Current inhibitors targeting the catalytic domain of DNMTs induce DNA demethylation across genomic regions. Here we will show that protein-protein interactions are important mediators for DNA methylation in cancer and can be targeted selectively and specifically for cancer treatment depending on the epigenomic features.

Transcription regulation is a complex process involving multiple proteins that target a specific gene. Indeed, synergistic mechanisms between epigenetic effectors conduct to the inactivation of tumor suppressor genes in cancers, and, therefore, their effective reactivation depends on inhibition of multiple epigenetic pathways. An important level of transcriptional

regulation is achieved through protein-protein interactions between transcription factors (e.g., c-Myc and UHRF1) and transcriptional co-regulators (e.g., DNMTs and HDACs) (Yan and Higgins, 2013). Another transcriptional control involves DNA methylation and/or histone modifications to recruit transcription factors/co-regulators or directly the regulators to the target gene. In cancers, this regulation controls the oncogenic pathway, and, therefore, transcription therapy targeting aberrant protein-protein interactions was proposed (Pandolfi, 2001; White et al., 2008).

We will review these different processes involving DNMTs recognition to regulate the transcription in the context of cancer initiation and progression. Understanding the recruitment mechanisms of DNMTs to specific genomic regions and identification of interaction sites between transcription factors/epigenome modifications and co-regulators will help future development of protein-protein interaction inhibitors targeting oncogenic pathways for personalized cancer therapeutics.

1.7.1 Targeting *de novo* DNMT3s to gene bodies via specific histone epigenetic marks

Among chromatin-binding proteins, several (e.g., PWWP, MBT, tudor, and PHD domains) are involved in protein-protein interactions with epigenetic marks deposited on histone or DNA. These epigenetic "readers" define potential druggable targets for developing selective protein-protein interaction inhibitors as a new promising strategy for chemotherapy (Cole, 2008). The proof-of-concept was established in 2010 with the BET (bromodomain and extraterminal) protein BRD4. This oncoprotein, involved in transcriptional regulation, binds lysine-acetylated residues on histone tails (Filippakopoulos et al., 2010). The discovered small molecule JQ1 (Fig. 6) binds into the acetyl-lysine recognition motif of BRD4 preventing the binding to chromatin and leads to a specific antiproliferative effect (Fig. 8).
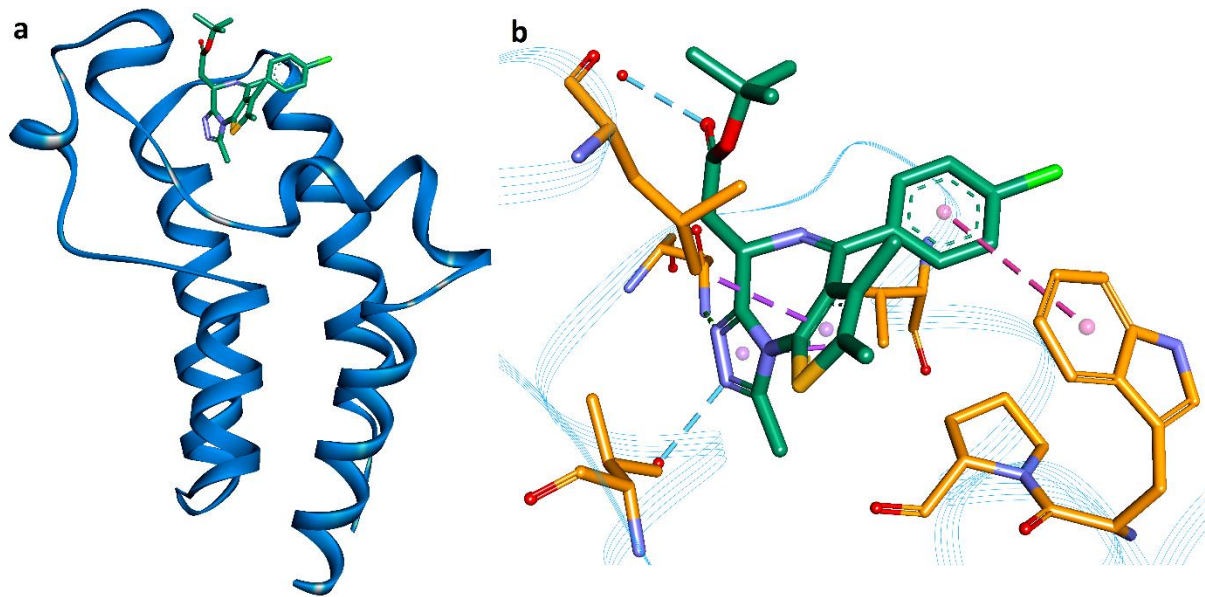
Figure 8. Human BRD4-JQ1 complex structure. *Structural overview of the BRD4 domain structure with the JQ1 inhibitor. BRD4 domain is represented as a blue solid ribbon. JQ1 inhibitor is represented as green stick model. Residues of the binding site are represented as orange stick models.*

In cancer, promoter CpG hypermethylation is well known to silence tumor suppressor genes and is the target of current treatments. Recently, however, a novel therapeutic target of DNA methylation was proposed, the gene body methylation. Indeed, gene body methylation mediates transcriptional activation of the potential oncogene ITPKA (Inositol-trisphosphate 3-kinase A) and genes up-regulated by the oncogenic transcription factor c-Myc (Wang et al., 2016; Yang et al., 2014).

Recently, regulatory mechanisms of gene body methylation were, in part, elucidated and require both *de novo* methyltransferase DNMT3B and the permissive histone epigenetic mark H3K36me3 (Baubec et al., 2015; Jin et al., 2012). More precisely, DNMT3B PWWP domain recognizes the trimethylation state of lysine 36 on histone H3 (H3K36me3) distributed within the gene body in a particular gene context (Figs. 4 and 9) (Baubec et al., 2015; Jin et al., 2012). Structure of DNMT3B PWWP domain in complex with the epigenetic mark H3K36me3 could lead to development of small-molecule inhibitors of this epigenetic complex (Fig. 9) (Rondelet et al., 2016). This will result in a DNA demethylation of gene bodies and a specific downregulation of oncogenic pathways as the small-molecule inhibitor JQ1 of BRD4 was proposed to target the c-Myc gene expression (Delmore et al., 2011; Filippakopoulos et al., 2010; Yang et al., 2014). The selective recruitment of DNMT3B to gene bodies compared to

DNMT3A still remains to be explained as their PWWP domain share a high sequence identity (53% using blastp). This could be determined by the cellular context or the difference in protein-protein interactions with the N-terminal part and transcription factors for example (Baubec et al., 2015; Rondelet et al., 2016). The N-terminal part of DNMT3s (1-281 aa) seems also important for nucleosome binding, but no structural information is available to date (Fig. 5) (Baubec et al., 2015; Jeong et al., 2009).
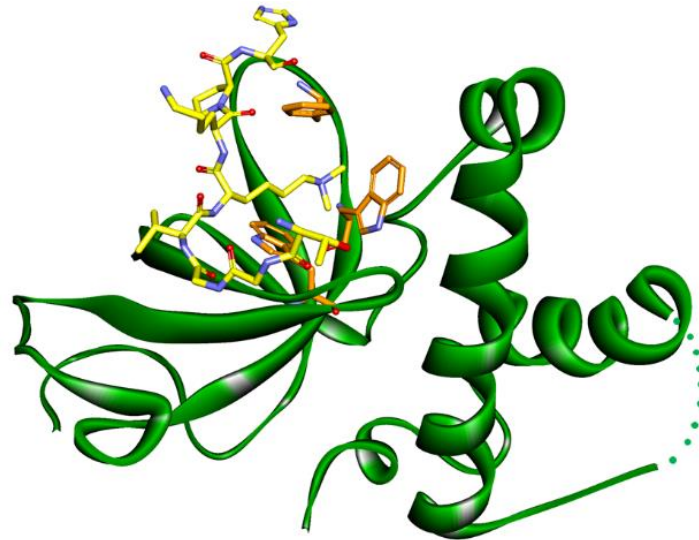
Figure 9. Structural overview of the DNMT3B PWWP domain structure in complex with the H3K36me3 peptide. *PWWP domain is represented as a green solid ribbon. Epigenetic mark H3K36me3 peptide is represented as yellow stick models. Aromatic cage residues are represented as orange stick models.*

Another recognition mechanism of DNMT3s, is the interaction with the unmethylated histone tail (H3K4me0) on nucleosome via their ADD domain (Fig. 5) (Ooi et al., 2007). This domain plays an important regulatory role for *de novo* methylation. Indeed, the recognition of H3K4me0 with DNMT3A ADD domain permits to release the autoinhibition of DNMT3A by disrupting the interaction between the ADD and catalytic domains of DNMT3A (PDB code: 4U7T) (Fig. 7) (Guo et al., 2015).

1.7.2 Targeting DNMTs to specific promoters via transcription factors and co-regulators

Transcription factors and co-regulators direct DNA methylation to specific promoters during cancerogenesis and tumor progression (Blattler and Farnham, 2013). These site-specific factors recruit, directly or indirectly, DNMTS to form a repressive complex and silence target genes. DNMTs can be also the scaffolds on DNA to recruit other repressor proteins.

The following table (Table 1) is a summary of the main partners involved in the recruitment of DNMTs to specific gene region. Details of the various recruitment mechanisms are explained in the following paragraphs.

Table 1. Partners of DNMTs associated to a particular cancer type.

DNMTs	Partners	Cancer type	References
DNMT1	UHRF1- USP7	breast cancer	(Felle et al., 2011; Qin et al., 2016)
DNMT1	DMAP1 - HDAC2	Glioma	(Rountree et al., 2000)
DNMTs	PML-RAR	acute promyelocytic leukemia	(Di Croce et al., 2002)
DNMT3A/B	EVI1	acute myeloid leukemia	(Lugthart et al., 2011)
DNMT3A	c-Myc	human astrocytoma	(Brenner et al., 2005; Hervouet et al., 2009)
DNMT3A/B	DNMT3L- NFkB	glioma cell lines	(Pacaud et al., 2014)
DNMT1	STAT3	breast cancers; T-cell lymphoma	(Lee et al., 2012)
DNMT1	NF-κB	breast cancer	(Liu et al., 2012)
DNMT3A	SETDB1	human breast adenocarcinoma cells	(Li et al., 2006)
DNMTs	EZH2	colon cancer and osteosarcoma cells	(Viré et al., 2006)
DNMTs	CBX7	embryonal carcinoma	(Mohammad et al., 2009)
DNMT3A	G9a/GLP - MPP8	epithelial tumor cells	(Chang et al., 2011)
DNMT1 and DNMT3A	SUV39H1 - HP1β	cervical carcinoma	(Fuks et al., 2003; Machado et al., 2010)
DNMT3A	CUL4B - SUV39H1/HP1	cervical carcinoma	(Yang et al., 2015)
DNMT1 or DNMT3B	NuRD	colon cancers cells	(Cai et al., 2014)
DNMT3A	ZN217/CoREST	myelodysplastic syndromes	(Thillainadesan et al., 2012)

Colocalization of DNMT1 with an essential activator for maintenance methylation, UHRF1 (E3 ubiquitin–protein ligase), is observed in late S phase (Bostick et al., 2007). UHRF1 recognizes hemimethylated DNA and can recruit DNMT1 through its SRA (Set and Ring Associated) domain and the RFTS domain of DNMT1 (Fig. 5) (Achour et al., 2008; Bostick et al., 2007). UHRF1 could also relieve the DNA binding inhibition mediated by the acidic linker (Auto) between BAH1 and CXXC domains (Fig. 5) (observed in mouse DNMT1 in complex with DNA (PDB code: 3PT6)) (Syeda et al., 2011). UHRF1 is overexpressed in various cancers and maintains the repression of several TSGs (p16^{INK4A}, p14^{ARF}, hMLH1, BRCA1, RB1, FHIT, RARb, APC and DAPK) during cell proliferation (Alhosin et al., 2011; Unoki et al., 2009).

Recently, USP7 (also known as HAUSP) was identified as an important partner of DNMT1 and UHRF1. Indeed, this partner interacts with DNMT1 and stimulates its activity, then this dimeric complex (DNMT1-USP7) is recruited by UHRF1 on chromatin to silence genes where USP7 stabilizes UHRF1 (Felle et al., 2011). If interaction between UHRF1/USP7 and DNMT1 are disrupted, expression of the tumor suppressor genes can be triggered to cause apoptosis and cell cycle arrest. In this way, the structure of the DNMT1-USP7 complex with DNMT1 was recently solved and shows molecular interactions between the C-terminal domain of USP7 and the RFTS domain of DNMT1 (PDB code: 4YOC) (Cheng et al., 2015). In another way, a team discovered recently small molecules binding the 5-methylcytosine pocket of UHRF1 to inhibit its recognition with the hemimethylated DNA (Myrianthopoulos et al., 2016).

In addition to its DNA methylation maintenance ability, DNMT1 forms different repressive transcription complexes during DNA replication at replication foci. DNMT1 can recruit both DMAP1 (via the first 120 aa - DMAP) and HDAC2 to deacetylate local chromatin. DMAP1 further recruits the TSG101 (co-repressor) (Rountree et al., 2000). Crystal structure of the DMAP1 protein is known (PDB code: 4IEJ), but the structure of the DMAP1 domain of DNMT1, despite its important role to form a repressive transcription complex, still has not been solved (Fig. 5). Recently, inhibition of protein-protein interaction between DNMT1 and DMAP1 was proposed to sensitize cancer cells to chemotherapy (Cheray et al., 2013).

The oncogenic transcription factor PML-RAR, found in acute promyelocytic leukemias, recruits DNMTs to target promoters for gene silencing (hypermethylation) (Di Croce et al., 2002). As in acute myeloid leukemia, EVI1 (ecotropic viral integration site 1) directs promoter methylation by recruitment of *de novo* DNMT3A/B (Lugthart et al., 2011). Another oncogenic transcription factor, c-Myc, induces gene repression by recruitment of DNMT3A within the promoters of CDKN1a, CCND1, TIMP2 and p21Cip1 genes in human astrocytoma (Brenner et al., 2005; Hervouet et al., 2009).

The catalytically inactive DNMT3L interacts with a site-specific transcription factor, NFkB-p65, to recruit *de novo* DNMT3A/B on the TRAF1 promoter in glioma cell lines (Pacaud et al., 2014). DNMT3L has a dual function as it interacts with the catalytic domain for the recruitment and the enzymatic activity stimulation. The structure between DNMT3L and DNMT3A was previously solved (PDB code: 2QRV) (Jia et al., 2007).

Acetylation at Lys685 of the oncogenic transcription factor STAT3 (Signal transducer and activator of transcription 3) permits the binding of DNMT1 to promoters in several cancers (e.g., breast cancers and T-cell lymphoma) (Lee et al., 2012). The mutation of Lys685 (K685R) disrupts DNMT1-STAT3 complex providing evidence of a focus region ("hot spot") for small protein-protein interaction inhibitors development. Phosphorylated RelA/p65, the subunit of NF-κB, functions as an active transcriptional repressor via a direct recruitment of DNMT1 to BRMS1 (breast cancer metastasis suppressor 1) promoter (Liu et al., 2012). Inhibition of this protein-protein interaction with protein transduction domain (PTD) peptide (PTD-RelA/p65 peptide) resulted in an abrogation of BRMS1 methylation and the transcriptional repression.

In addition to transcription factors that interact directly with DNMTs, indirect recruitments to specific promoters involve repressive histone modifications complexes. H3K9 methyltransferase SETDB1 interacts with DNMT3A to form a complex localized at promoters of the endogenous p53BP2 gene (encoding the tumor suppressor p53-binding protein 2) in HeLa cells and the RASSF1A gene (encoding the Ras association domain-containing protein 1) in human breast adenocarcinoma cells to silence their expression (Li et al., 2006). H3K27 methyltransferase EZH2 (Enhancer of Zeste homolog 2) which catalyzes trimethylation of histone H3 lysine 27 is a component of the Polycomb Repressive Complexes 2 and 3 (PRC2/3) and can recruit, via its N-terminal domain, DNMTs to silence genes (Viré et al., 2006). In colon

cancer and osteosarcoma cells, this Polycomb complex mediated *de novo* methylation and plays an important part in carcinogenesis (Schlesinger et al., 2007; Viré et al., 2006). In an identical manner, another component, the CBX7 (Chromobox protein homolog 7), targets DNMTs to gene promoters in cancer (Mohammad et al., 2009).

The Polycomb Repressive Complex 2 can also be regulated by G9a (histone-lysine N-methyltransferase 2) and GLP (G9a-like protein) which deposit the repressive epigenetic mark H3K9me1 and H3K9me2 observed in colorectal and breast cancer cells (McGarvey et al., 2006; Mozzetta et al., 2014; Tachibana et al., 2005; Wozniak et al., 2007). More interestingly, G9a/GLP interact with MPP8 (M-phase phosphoprotein 8) and DNMT3A to silence tumor suppressor E-cadherin gene in epithelial tumor cells (Kokura et al., 2010). In 2011, the mechanism of formation of this repressive complex was established (Chang et al., 2011). First, G9a or GLP dimethylate the N-terminal lysine 44 of mouse DNMT3A (K47 for human DNMT3A). Then, the chromodomain of MPP8 recognizes the methylated DNMT3A and the self-methylated GLP to form a repressive complex on the promoter of E-cadherin gene. In addition, structure between GLP and DNMT3AK44me0/2 peptide (PDB code: 3SW9 and 3SWC) were solved as well as the MPP8-DNMT3AK47me2 (PDB code: 3SVM) peptide complex (Fig. 7). As MPP8 and G9a/GLP direct *de novo* methylation for tumor progression in epithelial tumor cells, future protein-protein interaction inhibitors targeting these complexes could lead to the reexpression of the cell-adhesion molecule E-cadherin. For the repression of E-cadherin gene transcription in breast cancer cells, the transcription factor ZEB1 (δEF1) recruits DNMT1 through its Smad-binding domain (SBD).

The histone-lysine N-methyltransferase, SUV39H1 (Su(var)3-9 homolog 1), is targeted to histone H3 for the repressive H3K9 methylation on major satellite repeats at pericentric heterochromatin (Lehnertz et al., 2003). HP1β chromodomain (Heterochromatin Protein 1) binds the methylated histone H3K9me3 and can recruit DNMT1 and DNMT3A to direct DNA methylation (Fuks et al., 2003; Machado et al., 2010). In cervical carcinoma, CUL4B (Cullin 4B), coordinates the function of this repressive complex (SUV39H1/HP1/DNMT3A) to repress tumor suppressor IGFBP3 by catalysing H2AK119 mono-ubiquitination (Yang et al., 2015). This interplay between histone ubiquitination/methylation and DNA methylation to repress TSGs in cervical carcinoma represents an important target for cancer therapy.

Finally, NuRD and ZN217/CoREST chromatin remodeling complex cooperate with DNMTs to maintain silencing of TSGs in human cancers. NuRD complex (also known as Mi-2) cooperates with DNMT1 or DNMT3B to silence tumor suppressor genes in colon cancers cells whereas ZN217/CoREST complex cooperates with DNMT3A for repression of the p15^{ink4b} gene in myelodysplastic syndromes (Cai et al., 2014; Thillainadesan et al., 2012).

1.8 Conclusion

Nucleoside analogs targeting DNMTs are used as chemotherapeutic agents in the treatment of myelodysplastic syndromes and acute myeloid leukemia. However, these compounds are not selective for isoforms of DNMTs, have a poor bioavailability and a high toxicity. To address these limitations, development of nonnucleoside compounds acting directly on the catalytic domain of DNMTs has progressed despite low cellular activities compared to nucleoside analogs. Over the last two decades, and in parallel to this development, DNMT-interacting partners were described in various cancers to regulate gene expression. These novel mechanistic insights provide new epigenetic therapy strategies for cancer as exemplified by the combination treatment targeting the catalytic functions of multiprotein complexes to induce a synergistic cytotoxicity effect. Here, we reviewed these molecular machineries involving DNMTs and propose a new approach to increase therapeutic efficacy and specificity by designing protein-protein inhibitors of DNMT-interacting partners. Some structural studies of recognition systems described in the present review await further biochemical characterization for the development of a new generation of DNMT inhibitors based on this approach.

1.9 References

Achour, M., Jacq, X., Ronde, P., Alhosin, M., Charlot, C., Chataigneau, T., Jeanblanc, M., Macaluso, M., Giordano, A., Hughes, A., 2008. The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of VEGF gene expression. Oncogene 27, 2187-2197.

Alhosin, M., Sharif, T., Mousli, M., Etienne-Selloum, N., Fuhrmann, G., Schini-Kerth, V.B., Bronner, C., 2011. Down-regulation of UHRF1, associated with re-expression of tumor suppressor genes, is a common feature of natural compounds exhibiting anti-cancer properties. Journal of Experimental & Clinical Cancer Research 30, 1.

Almassalha, L.M., Bauer, G.M., Chandler, J.E., Gladstein, S., Cherkezyan, L., Stypula-Cyrus, Y., Weinberg, S., Zhang, D., Ruhoff, P.T., Roy, H.K., 2016. Label-free imaging of the native, living cellular nanoarchitecture using partial-wave spectroscopic microscopy. Proceedings of the National Academy of Sciences, 113, E6372-E6381.

Ball, M.P., Li, J.B., Gao, Y., Lee, J.-H., LeProust, E.M., Park, I.-H., Xie, B., Daley, G.Q., Church, G.M., 2009. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nature biotechnology 27, 361-368.

Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., Schubeler, D., 2015. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature 520, 243-247.

Berger, S.L., Kouzarides, T., Shiekhattar, R., Shilatifard, A., 2009. An operational definition of epigenetics. Genes & development 23, 781-783.

Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R., 2010. The NIH roadmap epigenomics mapping consortium. Nature biotechnology 28, 1045-1048.

Bestor, T.H., 2000. The DNA methyltransferases of mammals. Human molecular genetics 9, 2395-2402.

Bird, A., 2002. DNA methylation patterns and epigenetic memory. Genes & development 16, 6-21.

Blattler, A., Farnham, P.J., 2013. Cross-talk between site-specific transcription factors and DNA methylation states. Journal of Biological Chemistry 288, 34287-34294.

Bostick, M., Kim, J.K., Estève, P.-O., Clark, A., Pradhan, S., Jacobsen, S.E., 2007. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science (New York, N.Y.) 317, 1760-1764.

Bourc'his, D., Xu, G.-L., Lin, C.-S., Bollman, B., Bestor, T.H., 2001. Dnmt3L and the establishment of maternal genomic imprints. Science (New York, N.Y.) 294, 2536-2539.

Brenner, C., Deplus, R., Didelot, C., Loriot, A., Viré, E., De Smet, C., Gutierrez, A., Danovi, D., Bernard, D., Boon, T., 2005. Myc represses transcription through recruitment of DNA methyltransferase corepressor. The EMBO journal 24, 336-346.

Cai, Y., Geutjes, E., de Lint, K., Roepman, P., Bruurs, L., Yu, L., Wang, W., van Blijswijk, J., Mohammad, H., de Rink, I., 2014. The NuRD complex cooperates with DNMTs to maintain silencing of key colorectal tumor suppressor genes. Oncogene 33, 2157-2168.

Chang, Y., Sun, L., Kokura, K., Horton, J.R., Fukuda, M., Espejo, A., Izumi, V., Koomen, J.M., Bedford, M.T., Zhang, X., 2011. MPP8 mediates the interactions between DNA methyltransferase Dnmt3a and H3K9 methyltransferase GLP/G9a. Nature communications 2, 533.

Chédin, F., Lieber, M.R., Hsieh, C.-L., 2002. The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. Proceedings of the National Academy of Sciences 99, 16916-16921.

Chen, T., Li, E., 2006. Establishment and maintenance of DNA methylation patterns in mammals. Current topics in microbiology and immunology 301, 179-201.

Cheng, J., Yang, H., Fang, J., Ma, L., Gong, R., Wang, P., Li, Z., Xu, Y., 2015. Molecular mechanism for USP7-mediated DNMT1 stabilization by acetylation. Nature communications 6.

Cheng, X., Blumenthal, R.M., 2008. Mammalian DNA methyltransferases: a structural perspective. Structure 16, 341-350.

Cheray, M., Pacaud, R., Nadaradjane, A., Vallette, F.M., Cartron, P.-F., 2013. Specific inhibition of one DNMT1-including complex influences tumor initiation and progression. Clinical epigenetics 5, 1.

Chodavarapu, R.K., Feng, S., Bernatavichute, Y.V., Chen, P.-Y., Stroud, H., Yu, Y., Hetzel, J.A., Kuo, F., Kim, J., Cokus, S.J., 2010. Relationship between nucleosome positioning and DNA methylation. Nature 466, 388-392.

Christman, J.K., 2002a. 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. Oncogene 21, 5483-5495.

Christman, J.K., 2002b. 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. Oncogene 21, 5483-5495.

Chuang, J.C., Warner, S.L., Vollmer, D., Vankayalapati, H., Redkar, S., Bearss, D.J., Qiu, X., Yoo, C.B., Jones, P.A., 2010. S110, a 5-Aza-2'-Deoxycytidine–Containing Dinucleotide, Is an Effective DNA Methylation Inhibitor In vivo and Can Reduce Tumor Growth. Molecular cancer therapeutics 9, 1443-1450.

Clozel, T., Yang, S., Elstrom, R.L., Tam, W., Martin, P., Kormaksson, M., Banerjee, S., Vasanthakumar, A., Culjkovic, B., Scott, D.W., 2013. Mechanism-based epigenetic chemosensitization therapy of diffuse large B-cell lymphoma. Cancer discovery 3, 1002-1019.

Cole, P.A., 2008. Chemical probes for histone-modifying enzymes. Nature chemical biology 4, 590-597.

Consortium, I.H.G.S., 2004. Finishing the euchromatic sequence of the human genome. Nature 431, 931-945.

de Bustros, A., Nelkin, B.D., Silverman, A., Ehrlich, G., Poiesz, B., Baylin, S.B., 1988. The short arm of chromosome 11 is a" hot spot" for hypermethylation in human neoplasia. Proceedings of the National Academy of Sciences 85, 5693-5697.

Deaton, A.M., Webb, S., Kerr, A.R., Illingworth, R.S., Guy, J., Andrews, R., Bird, A., 2011. Cell type–specific DNA methylation at intragenic CpG islands in the immune system. Genome research 21, 1074-1086.

Delmore, J.E., Issa, G.C., Lemieux, M.E., Rahl, P.B., Shi, J., Jacobs, H.M., Kastritis, E., Gilpatrick, T., Paranal, R.M., Qi, J., 2011. BET bromodomain inhibition as a therapeutic strategy to target c-Myc. Cell 146, 904-917.

Deng, T., Kuang, Y., Wang, L., Li, J., Wang, Z., Fei, J., 2009. An essential role for DNA methyltransferase 3a in melanoma tumorigenesis. Biochemical and biophysical research communications 387, 611-616.

Di Croce, L., Raker, V.A., Corsaro, M., Fazi, F., Fanelli, M., Faretta, M., Fuks, F., Coco, F.L., Kouzarides, T., Nervi, C., 2002. Methyltransferase recruitment and DNA hypermethylation of target promoters by an oncogenic transcription factor. Science (New York, N.Y 295, 1079-1082.

Eden, S., Hashimshony, T., Keshet, I., Cedar, H., Thorne, A., 1998. DNA methylation models histone acetylation. Nature 394, 842-842.

Erdmann, A., Halby, L., Fahy, J., Arimondo, P.B., 2015. Targeting DNA Methylation with Small Molecules: What's Next? Miniperspective. Journal of medicinal chemistry 58, 2569-2583.

Esteller, M. 2008. Epigenetics in cancer. The New England Journal of Medicine 358, 1148-1159

Estève, P.-O., Chang, Y., Samaranayake, M., Upadhyay, A.K., Horton, J.R., Feehery, G.R., Cheng, X., Pradhan, S., 2011. A methylation and phosphorylation switch between an adjacent lysine and serine determines human DNMT1 stability. Nature structural & molecular biology 18, 42-48.

Falkenberg, K.J., Johnstone, R.W., 2014. Histone deacetylases and their inhibitors in cancer, neurological diseases and immune disorders. Nature reviews Drug discovery 13, 673-691.

Felle, M., Joppien, S., Németh, A., Diermeier, S., Thalhammer, V., Dobner, T., Kremmer, E., Kappler, R., Längst, G., 2011. The USP7/Dnmt1 complex stimulates the DNA methylation activity of Dnmt1 and regulates the stability of UHRF1. Nucleic acids research 39, 8355-8365.

Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W.B., Fedorov, O., Morse, E.M., Keates, T., Hickman, T.T., Felletar, I., 2010. Selective inhibition of BET bromodomains. Nature 468, 1067-1073.

Fuks, F., Hurd, P.J., Deplus, R., Kouzarides, T., 2003. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic acids research 31, 2305-2312.

Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., 2015. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature 517, 640-644.

Hackett, J.A., Surani, M.A., 2013. DNA methylation dynamics during the mammalian life cycle. Phil. Trans. R. Soc. B 368, 20110328.

Hansen, K.D., Timp, W., Bravo, H.C., Sabunciyan, S., Langmead, B., McDonald, O.G., Wen, B., Wu, H., Liu, Y., Diep, D., 2011. Increased methylation variation in epigenetic domains across cancer types. Nature genetics 43, 768-775.

Hervouet, E., Vallette, F.M., Cartron, P.-F., 2009. Dnmt3/transcription factor interactions as crucial players in targeted DNA methylation. Epigenetics 4, 487-499.

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., 2009. The human colon cancer methylome shows similar hypo-and hypermethylation at conserved tissue-specific CpG island shores. Nature genetics 41, 178-186.

Issa, J.-P.J., Kantarjian, H.M., 2009. Targeting DNA methylation. Clinical Cancer Research 15, 3938-3946.

Issa, J.-P.J., Roboz, G., Rizzieri, D., Jabbour, E., Stock, W., O'Connell, C., Yee, K., Tibes, R., Griffiths, E.A., Walsh, K., 2015. Safety and tolerability of guadecitabine (SGI-110) in patients with myelodysplastic syndrome and acute myeloid leukemia: a multicentre, randomised, dose-escalation phase 1 study. The Lancet Oncology 16, 1099-1110.

Jaenisch, R., Bird, A., 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nature genetics 33, 245-254.

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

Jeong, S., Liang, G., Sharma, S., Lin, J.C., Choi, S.H., Han, H., Yoo, C.B., Egger, G., Yang, A.S., Jones, P.A., 2009. Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA. Molecular and cellular biology 29, 5366-5376.

Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., Cheng, X., 2007. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature 449, 248-251.

Jin, B., Ernst, J., Tiedemann, R.L., Xu, H., Sureshchandra, S., Kellis, M., Dalton, S., Liu, C., Choi, J.-H., Robertson, K.D., 2012. Linking DNA methyltransferases to epigenetic marks and nucleosome structure genome-wide in human tumor cells. Cell reports 2, 1411-1424.

Jones, P.A., 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature Reviews Genetics 13, 484-492.

Jones, P.A., Taylor, S.M., 1980. Cellular differentiation, cytidine analogs and DNA methylation. Cell 20, 85-93.

Jones, P.A., Liang, G., 2009. Rethinking how DNA methylation patterns are maintained. Nature Reviews Genetics 10, 805-811.

Jones, P.A., Archer, T.K., Baylin, S.B., Beck, S., Berger, S., Bernstein, B.E., Carpten, J.D., Clark, S.J., Costello, J.F., Doerge, R.W., 2008. Moving AHEAD with an international human epigenome project. Nature 454, 711-715.

Juergens, R.A., Wrangle, J., Vendetti, F.P., Murphy, S.C., Zhao, M., Coleman, B., Sebree, R., Rodgers, K., Hooker, C.M., Franco, N., 2011. Combination epigenetic therapy has efficacy in patients with refractory advanced non–small cell lung cancer. Cancer discovery 1, 598-607.

Jurkowska, R.Z., Jurkowski, T.P., Jeltsch, A., 2011. Structure and function of mammalian DNA methyltransferases. Chembiochem 12, 206-222.

Keshet, I., Lieman-Hurwitz, J., Cedar, H., 1986. DNA methylation affects the formation of active chromatin. Cell 44, 535-543.

Klimasauskas, S., Kumar, S., Roberts, R.J., Cheng, X., 1994. HhaI methyltransferase flips its target base out of the DNA helix. Cell 76, 357-369.

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

Kokura, K., Sun, L., Bedford, M.T., Fang, J., 2010. Methyl-H3K9-binding protein MPP8 mediates E-cadherin gene silencing and promotes tumor cell motility and invasion. The EMBO journal 29, 3673-3687.

Lee, H., Zhang, P., Herrmann, A., Yang, C., Xin, H., Wang, Z., Hoon, D.S., Forman, S.J., Jove, R., Riggs, A.D., 2012. Acetylated STAT3 is crucial for methylation of tumor-suppressor gene promoters and inhibition by resveratrol results in demethylation. Proceedings of the National Academy of Sciences 109, 7765-7769.

Lehnertz, B., Ueda, Y., Derijck, A.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., Peters, A.H., 2003. Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Current Biology 13, 1192-1200.

Ley, T.J., Ding, L., Walter, M.J., McLellan, M.D., Lamprecht, T., Larson, D.E., Kandoth, C., Payton, J.E., Baty, J., Welch, J., 2010. DNMT3A mutations in acute myeloid leukemia. New England Journal of Medicine 363, 2424-2433.

Li, H., Rauch, T., Chen, Z.-X., Szabó, P.E., Riggs, A.D., Pfeifer, G.P., 2006. The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. Journal of Biological Chemistry 281, 19489-19500.

Li, J.-Y., Pu, M.-T., Hirasawa, R., Li, B.-Z., Huang, Y.-N., Zeng, R., Jing, N.-H., Chen, T., Li, E., Sasaki, H., 2007. Synergistic function of DNA methyltransferases Dnmt3a and Dnmt3b in the methylation of Oct4 and Nanog. Molecular and cellular biology 27, 8748-8759.

Liu, Y., Mayo, M.W., Nagji, A.S., Smith, P.W., Ramsey, C.S., Li, D., Jones, D.R., 2012. Phosphorylation of RelA/p65 promotes DNMT-1 recruitment to chromatin and represses transcription of the tumor metastasis suppressor gene BRMS1. Oncogene 31, 1143-1154.

Low, D.A., Weyand, N.J., Mahan, M.J., 2001. Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. Infection and Immunity 69, 7197-7204.

Lugthart, S., Figueroa, M.E., Bindels, E., Skrabanek, L., Valk, P.J., Li, Y., Meyer, S., Erpelinck-Verschueren, C., Greally, J., Löwenberg, B., 2011. Aberrant DNA hypermethylation signature in acute myeloid leukemia directed by EVI1. Blood 117, 234-241.

Lyko, F., Brown, R., 2005. DNA methyltransferase inhibitors and the development of epigenetic cancer therapies. Journal of the National Cancer Institute 97, 1498-1506.

Machado, M.R., Dans, P.D., Pantano, S., 2010. Isoform-specific determinants in the HP1 binding to histone 3: insights from molecular simulations. Amino acids 38, 1571-1581.

McGarvey, K.M., Fahrner, J.A., Greene, E., Martens, J., Jenuwein, T., Baylin, S.B., 2006. Silenced tumor suppressor genes reactivated by DNA demethylation do not return to a fully euchromatic chromatin state. Cancer research 66, 3541-3549.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454, 766-770.

Miranda, T.B., Cortez, C.C., Yoo, C.B., Liang, G., Abe, M., Kelly, T.K., Marquez, V.E., Jones, P.A., 2009. DZNep is a global histone methylation inhibitor that reactivates developmental genes not silenced by DNA methylation. Molecular cancer therapeutics 8, 1579-1588.

Mohammad, H.P., Cai, Y., McGarvey, K.M., Easwaran, H., Van Neste, L., Ohm, J.E., O'Hagan, H.M., Baylin, S.B., 2009. Polycomb CBX7 promotes initiation of heritable repression of genes frequently silenced with cancer-specific DNA hypermethylation. Cancer research 69, 6322-6330.

Mohn, F., Schübeler, D., 2009. Genetics and epigenetics: stability and plasticity during cellular differentiation. Trends in Genetics 25, 129-136.

Mozzetta, C., Pontis, J., Fritsch, L., Robin, P., Portoso, M., Proux, C., Margueron, R., Ait-Si-Ali, S., 2014. The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing. Molecular cell 53, 277-289.

Myrianthopoulos, V., Cartron, P.F., Liutkevičiūtė, Z., Klimašauskas, S., Matulis, D., Bronner, C., Martinet, N., Mikros, E., 2016. Tandem virtual screening targeting the SRA domain of UHRF1 identifies a novel chemical tool modulating DNA methylation. European journal of medicinal chemistry 114, 390-396.

Nguyen, C., Liang, G., Nguyen, T.T., Tsao-Wei, D., Groshen, S., Lübbert, M., Zhou, J.-H., Benedict, W.F., Jones, P.A., 2001. Susceptibility of nonpromoter CpG islands to de novo methylation in normal and neoplastic cells. Journal of the National Cancer Institute 93, 1465-1472.

Nosho, K., Shima, K., Irahara, N., Kure, S., Baba, Y., Kirkner, G.J., Chen, L., Gokhale, S., Hazra, A., Spiegelman, D., 2009. DNMT3B expression might contribute to CpG island methylator phenotype in colorectal cancer. Clinical Cancer Research 15, 3663-3671.

Okano, M., Bell, D.W., Haber, D.A., Li, E., 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99, 247-257.

Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D., 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448, 714-717.

Pacaud, R., Sery, Q., Oliver, L., Vallette, F.M., Tost, J., Cartron, P.-F., 2014. DNMT3L interacts with transcription factors to target DNMT3L/DNMT3B to specific DNA sequences: Role of the DNMT3L/DNMT3B/p65-NFκB complex in the (de-) methylation of TRAF1. Biochimie 104, 36-49.

Pandolfi, P.P., 2001. Transcription therapy for cancer. Oncogene 20, 3116-3127.

Qin, D., Wang, W., Lei, H., Luo, H., Cai, H., Tang, C., Wu, Y., Wang, Y., Jin, J., Xiao, W., 2016. CDDO-Me reveals USP7 as a novel target in ovarian cancer cells. Oncotarget 123, 359-373.

Renbaum, P., Abrahamove, D., Fainsod, A., Wilson, G.G., Rottem, S., Razin, A., 1990. Cloning, characterization, and expression in Escherichia coli of the gene coding for the CpG DNA methylase from Spiroplasma sp. strain MQ1 (M SssI). Nucleic acids research 18, 1145-1152.

Rhee, I., Bachman, K.E., Park, B.H., Jair, K.-W., Yen, R.-W.C., Schuebel, K.E., Cui, H., Feinberg, A.P., Lengauer, C., Kinzler, K.W., 2002. DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. Nature 416, 552-556.

Rideout, W., Coetzee, G.A., Olumi, A.F., Jones, P.A., 1990. 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. Science (New York, N.Y.) 249, 1288-1290.

Riggs, A.D., 1975. X inactivation, differentiation, and DNA methylation. Cytogenetic and Genome Research 14, 9-25.

Roll, J.D., Rivenbark, A.G., Jones, W.D., Coleman, W.B., 2008. DNMT3b overexpression contributes to a hypermethylator phenotype in human breast cancer cell lines. Molecular cancer 7, 1.

Rondelet, G., Dal Maso, T., Willems, L., Wouters, J., 2016. Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. Journal of Structural Biology 194, 357-367.

Rountree, M.R., Bachman, K.E., Baylin, S.B., 2000. DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. Nature genetics 25, 269-277.

Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., Eden, E., Yakhini, Z., Ben-Shushan, E., Reubinoff, B.E., 2007. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nature genetics 39, 232-236.

Shell, S.S., Prestwich, E.G., Baek, S.-H., Shah, R.R., Sassetti, C.M., Dedon, P.C., Fortune, S.M., 2013. DNA methylation impacts gene expression and ensures hypoxic survival of Mycobacterium tuberculosis. PLoS Pathog 9, e1003419.

Shirohzu, H., Kubota, T., Kumazawa, A., Sado, T., Chijiwa, T., Inagaki, K., Suetake, I., Tajima, S., Wakui, K., Miki, Y., 2002. Three novel DNMT3B mutations in Japanese patients with ICF syndrome. American journal of medical genetics 112, 31-37.

Shlush, L.I., Zandi, S., Mitchell, A., Chen, W.C., Brandwein, J.M., Gupta, V., Kennedy, J.A., Schimmer, A.D., Schuh, A.C., Yee, K.W., 2014. Identification of pre-leukaemic haematopoietic stem cells in acute leukemia. Nature 506, 328-333.

Song, J., Rechkoblit, O., Bestor, T.H., Patel, D.J., 2011. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science (New York, N.Y.) 331, 1036-1040.

Song, J., Teplova, M., Ishibe-Murakami, S., Patel, D.J., 2012. Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science (New York, N.Y.) 335, 709-712.

Stewart, K.R., Veselovska, L., Kim, J., Huang, J., Saadeh, H., Tomizawa, S.-i., Smallwood, S.A., Chen, T., Kelsey, G., 2015. Dynamic changes in histone modifications precede de novo DNA methylation in oocytes. Genes & development 29, 2449-2462.

Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., Tajima, S., 2004. DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. Journal of Biological Chemistry 279, 27816-27823.

Syeda, F., Fagan, R.L., Wean, M., Avvakumov, G.V., Walker, J.R., Xue, S., Dhe-Paganon, S., Brenner, C., 2011. The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1. Journal of Biological Chemistry 286, 15344-15351.

Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T., Shinkai, Y., 2005. Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. Genes & development 19, 815-826.

Tatton-Brown, K., Seal, S., Ruark, E., Harmer, J., Ramsay, E., del Vecchio Duarte, S., Zachariou, A., Hanks, S., O'Brien, E., Aksglaede, L., 2014. Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. Nature genetics 46, 385-388.

Thillainadesan, G., Chitilian, J.M., Isovic, M., Ablack, J.N.G., Mymryk, J.S., Tini, M., Torchia, J., 2012. TGF-β-dependent active demethylation and expression of the p15 ink4b tumor suppressor are impaired by the ZNF217/CoREST complex. Molecular cell 46, 636-649.

Tomizawa, S., Nowacka-Woszuk, J., Kelsey, G., 2012. DNA methylation establishment during oocyte growth: mechanisms and significance. Int J Dev Biol 56, 867-875.

Unoki, M., Brunet, J., Mousli, M., 2009. Drug discovery targeting epigenetic codes: the great potential of UHRF1, which links DNA methylation and histone modifications, as a drug target in cancers and toxoplasmosis. Biochemical pharmacology 78, 1279-1288.

Viré, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.-M., 2006. The Polycomb group protein EZH2 directly controls DNA methylation. Nature 439, 871-874.

Wang, S.-C., Oelze, B., Schumacher, A., 2008. Age-specific epigenetic drift in late-onset Alzheimer's disease. PloS one 3, e2698.

Wang, Y.-W., Ma, X., Zhang, Y.-A., Wang, M.-J., Yatabe, Y., Lam, S., Girard, L., Chen, J.-Y., Gazdar, A.F., 2016. ITPKA gene body methylation regulates gene expression and serves as an early diagnostic marker in lung and other cancers. Journal of Thoracic Oncology 11, 1469-1481.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., Schubeler, D., 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nature genetics 39, 457-466.

White, A.W., Westwell, A.D., Brahemi, G., 2008. Protein–protein interactions as targets for small-molecule therapeutics in cancer. Expert reviews in molecular medicine 10, e8.

Wijermans, P., Rüter, B., Baer, M., Slack, J.L., Saba, H., Lübbert, M., 2008. Efficacy of decitabine in the treatment of patients with chronic myelomonocytic leukemia (CMML). Leukemia research 32, 587-591.

Wolffe, A.P., Matzke, M.A., 1999. Epigenetics: regulation through repression. Science (New York, N.Y 286, 481-486.

Wozniak, R., Klimecki, W., Lau, S., Feinstein, Y., Futscher, B.W., 2007. 5-Aza-2'-deoxycytidine-mediated reductions in G9A histone methyltransferase and histone H3 K9 di-methylation levels are linked to tumor suppressor gene reactivation. Oncogene 26, 77-90.

Wutz, A., 2011. Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. Nature Reviews Genetics 12, 542-553.

Yan, C., Higgins, P.J., 2013. Drugging the undruggable: transcription therapy for cancer. Biochimica et Biophysica Acta (BBA)-Reviews on Cancer 1835, 76-85.

Yang, X., Han, H., De Carvalho, D.D., Lay, F.D., Jones, P.A., Liang, G., 2014. Gene body methylation can alter gene expression and is a therapeutic target in cancer. Cancer cell 26, 577-590.

Yang, Y., Liu, R., Qiu, R., Zheng, Y., Huang, W., Hu, H., Ji, Q., He, H., Shang, Y., Gong, Y., 2015. CRL4B promotes tumorigenesis by coordinating with SUV39H1/HP1/DNMT3A in DNA methylation-based epigenetic silencing. Oncogene 34, 104-118.

Zhao, L., Cao, D., Chen, T., Wang, Y., Miao, Z., Xu, Y., Chen, W., Wang, X., Li, Y., Du, Z., 2013. Fragment-based drug discovery of 2-thiazolidinones as inhibitors of the histone reader BRD4 bromodomain. Journal of medicinal chemistry 56, 3833-3851.

Zhou, L., Cheng, X., Connolly, B.A., Dickman, M.J., Hurd, P.J., Hornby, D.P., 2002. Zebularine: a novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. Journal of molecular biology 321, 591-599.

RESEARCH AIMS & METHODS

Research Aims and Methods

The motivations and objectives of the presented thesis are detailed and integrated in the epigenetic cancer therapy research. In addition, a brief description through the different chapters will be provide.

Background: Epigenetic modifications control chromatin state and, therefore, regulate gene expression. These events occur in normal cells, but, in cancer cells, their deregulations conduct to the repression of tumor suppressor genes and the activation of oncogenic pathways. Among these epigenetic modifications, DNA methylation contributes to transcriptional regulation and is catalyzed by DNA (cytosine-5) methyltransferases (DNMTs). These DNMTs are targeted to specific regions of the genome to repress tumor suppressor genes by hypermethylation of their promoters. Alternatively, they mediate the transcription of oncoprotein-regulated genes by gene body methylation. As DNA methylation is a reversible process, treatment with demethylating agents could revert the epigenome cancer cell to stop cancer progression. Current treatments against DNMTs use nucleoside analogs for the treatment of myelodysplastic syndromes and acute myeloid leukemia.

Problem definition: Nucleoside analogs are not selective for isoforms of DNMTs, have a poor bioavailability and a high toxicity. To address these limitations, development of non-nucleoside compounds acting directly in the catalytic domain of DNMTs has progressed. However, the mode of action of these compounds remains unclear and their cellular activities are still low when compared with nucleoside analogs. Furthermore, the efficacy of these compounds appears to be compromised by their lack of specificity. Therefore, the mode of action of current non-nucleoside DNMT inhibitors has to be investigated.

Aims & Methods: The first aim of this research was to rationalize the catalytic inhibition of DNMTs by maleimide derivatives (RG108-1 and RG119-1) of RG108 (N-phthaloyl-L-tryptophan) (Fig.1). This was achieved by enzymatic and cellular assays, differential scanning fluorimetry assays, crystallographic studies and molecular docking. The related work is discussed in chapter 2, entitled "Rationalization of the inhibition of DNA (cytosine-5) methyltransferases by maleimide derivatives of RG108 as non-nucleoside inhibitors".

Figure 1. Chemical structures of the non-nucleoside DNA methylation inhibitor RG108 and its maleimide derivatives.

To enhance selectivity and specificity against DNMTs, a new strategy is proposed and is based on the specific gene body methylation associated with DNMT3B. DNMT3B recognizes, *via* its PWWP domain, the permissive histone epigenetic mark H3K36me3 for gene body methylation in cancer cells. The disruption of this protein-protein interaction could downregulate expression of oncogenes and oncoprotein-regulated genes.

Chapter 3 details the complex structure between DNMT3B PWWP domain and H3K36me3 peptide (see introduction Fig. 9). Molecular modeling was also performed to propose a model of DNMT3A in a nucleosomal context. The related work is discussed in chapter 3, entitled "Structural basis for recognition of histone H3K36me3 nucleosome by human *de novo* DNA methyltransferases 3A and 3B".

Based on the results of chapter 3, the third aim of this work was to develop and identify potential protein-protein interaction inhibitors of the PWWP DNMT3B-H3K36me3 complex. To this end, a similarity-based virtual screening was performed using the bis-tris molecule, identified from a complex with DNMT3B PWWP domain (PDB code: 3QKJ), and H3K36me3 epigenetic mark as both chemical references. The hits identified were co-crystallized with the DNMT3B PWWP domain to generate *in fine* a pharmacophore model. Other virtual screenings

were performed to identify potential protein-protein interaction inhibitors targeting the DNMT3B PWWP domain. The related work is discussed in chapter 4, entitled "Targeting PWWP domain of DNA methyltransferase 3B for epigenetic cancer therapy: Identification and structural characterization of new potential protein-protein interaction inhibitors".

CHAPTER 2

Chapter 2. Rationalization of the inhibition of DNA (cytosine-5) methyltransferases by maleimide derivatives of RG108 as non-nucleoside inhibitors
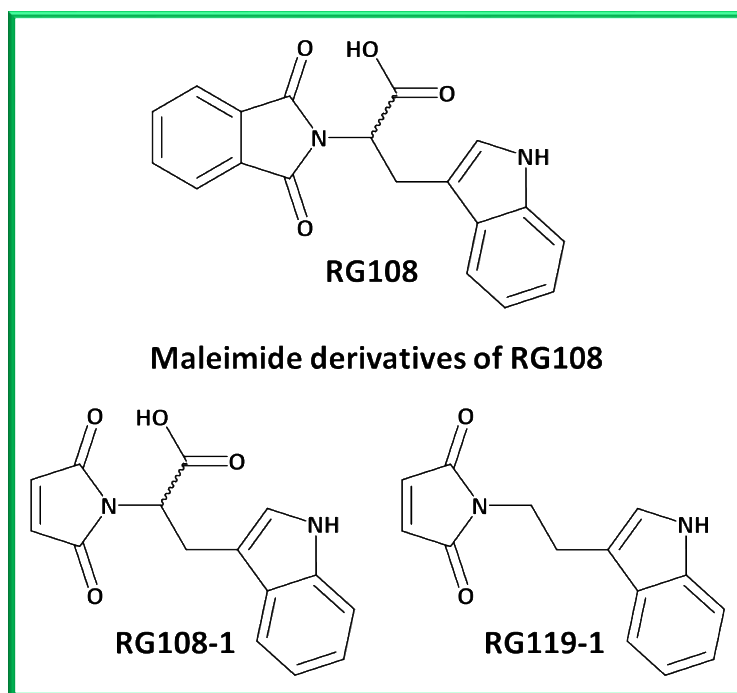
Grégoire Rondelet*[a], Laurence Fleury[b], Céline Faux[b], Véronique Masson[b], Jean Dubois[a], Paola B. Arimondo[b,c], Luc Willems[d], and Johan Wouters[a]

[a]Department of Chemistry, NAmur MEdicine & Drug Innovation Center (NAMEDIC-NARILIS), University of Namur, 61 rue de Bruxelles, B-5000 Namur, Belgium

[b]Unité de Service et de Recherche CNRS-Pierre Fabre USR n°3388, CNRS FRE n°3600, ETaC, CRDPF, 31100 Toulouse, France

[c]Churchill College, Cambridge, CD3 0DS, UK

[d]Molecular and Cellular Epigenetics (GIGA) and Molecular Biology (Gembloux Agro-Bio Tech), University of Liège (ULg), 4000 Liège, Belgium

*Corresponding author: gregoire.rondelet@unamur.be

† **Submitted**

Personal contribution: Participated in research design, conducted experiments (enzyme production (M.*Hha*I), enzyme crystallization, structure determination, docking simulations and cell viability assays), performed data analysis and literature search, generated figures and tables, and wrote the manuscript.

Abstract

DNA methyltransferases (DNMTs) are important drug targets for epigenetic therapy of cancer. Identification of novel non-nucleoside inhibitors are needed to decrease side effects and increase selectivity against DNMTs. Among these inhibitors, maleimide derivatives of RG108 seem promising compounds to improve inhibitory potency and selectivity for DNMTs. RG119-1 was found to be 10-fold more potent (IC50 values in the low micromolar range) than RG108-1 against DNMT3A and M.*Sss*I. While RG119-1 was active against DNMT1, RG108-1 showed no inhibition. This was reflected in the cytotoxicity activity of RG119-1 on the mesothelioma cells line with TC50 values in the low micromolar range. In order to understand the difference of activity against DNMTs isoforms, crystallographic assays were performed and conducted to the first apo structure of M.*Hha*I. This structure reveals that the *S*-adenosyl-L-methionine (SAM) cofactor does not affect the cofactor binding site conformation. Finally, a

new docking protocol was proposed based on differential scanning fluorimetry results, crystallographic studies and the mechanism of methylation of DNMTs. Covalent docking studies were also performed with RG108-1 and RG119-1 on DNMTs, and revealed essential residues for the stabilization of such compounds.

Keywords: DNA methylation, DNA methyltransferases, Covalent inhibitors, Michael acceptor, DNMT3A, DNMT1, M.*Hha*I, M.*Sss*I, Apoenzyme crystal structure

2.1 Introduction

In mammalian cells, DNA methylation is an important epigenetic modification involved in physiological events like embryonic development, genomic imprinting, X-chromosome inactivation, maintenance of chromosomal stability and gene silencing (Wutz, 2011). In cancers, hypermethylation of CpG islands of tumor suppressor genes (TSGs) promoters leads to their inactivation and, consequently, induces tumorigenesis and tumor maintenance (Esteller, 2008; Jones, 2012). This covalent epigenetic modification, which occurs at the C5 position of cytosine (5-methylcytosine, 5-mC), in CpG dinucleotides is catalyzed by C5-DNA methyltransferases (DNMTs) in the presence of the *S*-adenosyl-L-methionine (AdoMet or SAM) cofactor as methyl group donor (Bestor, 2000). Based on crystallographic and enzymatic studies, mostly conducted with bacterial C5-DNA methyltransferase M.*Hha*I, a mechanism of DNA cytosine-C5 methylation involving the nucleophilic addition of a cysteine to the cytosine prior to methylation was proposed (Fig. S1) (Bestor and Verdine, 1994; Goll and Bestor, 2005).

As DNA methylation is a reversible process, DNMTs are targeted for epigenetic therapy to re-express tumor suppressor genes leading to proliferation arrest and apoptosis in cancer cells (Goll and Bestor, 2005; Gros et al., 2012; Mai and Altucci, 2009). In 2004 and 2006, two nucleoside analogs of cytosine, namely 5-azacytidine (azacitidine, Vidaza®) and 5-aza-deoxycytidine (decitabine, Dacogen®) (see Fig. 6 in chapter 1), have been approved by FDA for myelodysplastic syndromes, acute myeloid leukemia and chronic myelomonocytic leukemia (Christman, 2002; Issa and Kantarjian, 2009; Jones and Taylor, 1980; Wijermans et al., 2008). These drugs are incorporated into DNA instead of deoxycytidine, the β-elimination reaction (releasing normally the 5-methylcytosine (Fig. S1)) cannot occur and thus irreversibly trap DNMT. This results in a degradation of DNMTs and a methylation marks depletion during DNA replication (Stresemann and Lyko, 2008). However, these nucleoside analogs are also associated with significant toxicity, low specificity, strong side effects, chemical instability and poor bioavailability (Lyko and Brown, 2005).

During the last decade, an attractive and novel alternative for epigenetic cancer therapy tends to develop more specific and selective DNMTis such as non-nucleoside compounds. These compounds act directly on DNMTs without incorporation into DNA expecting lower cytotoxicity and better selectivity. With this approach, there is an increasing number of non-

nucleoside analogs with various chemical structures and different expected modes of action targeting, *inter alia*, cofactor and/or substrate pockets (Erdmann et al., 2015a). Despite the fact that *in vitro* inhibitory activity increased for the new generation, development of non-nucleoside compounds is currently slowed by the lack of structural information of DNMTs in complex with these compounds, their low selectivity for DNMTs, the lack of information about their mechanism of inhibition and the low inhibitory activity in cells. Therefore, there is a need to understand the mode of action of current non-nucleoside DNMT inhibitors.

To this end, we focused our research on maleimide derivatives of the most studied compound, the RG108 (N-phthaloyl-L-tryptophan) (Fig. 1) (Brueckner et al., 2005). Since its discovery, several studies reported a weak inhibitory activity of RG108 and its demethylating activity remains controversial. However, this compound present interesting features and pharmacomodulations of RG108 were performed in different studies to increase its potency. In 2010, novel DNMTis derived from RG108 and characterized by the replacement of phthalimide group with a maleimide group were identified (Suzuki et al., 2010). Among these, RG108-1 (Fig. 1) was found to be the most potent DNMT1 inhibitor (60% inhibition at 10 µM), having a better inhibitory effect than RG108. Recently, a systematic structure-activity relationship (SAR) study on the RG108 and its analogs was performed to understand the role of the different moieties and to increase its activity, leading to the thionitropyridine analog, which presents an IC50 value of 20 µM against DNMT1 (Fig. 1) (Asgatay et al., 2014). In the same study, authors showed that the carboxylate anion of RG108 does not influence its inhibitory activity.

Based on these results, we decided to study the RG108-1 and RG119-1 (a tryptamine derivative lacking carboxylate functionality) (Fig. 1) (Suzuki et al., 2010). These compounds are expected to target directly DNMTs by covalent bonding with the catalytic cysteine residue without incorporation into DNA unlike nucleoside inhibitors. To decrease the toxicity of such compounds caused by the off target effects, specific non-covalent interactions with the scaffold of the compound are important. This will stabilize the compound inside the binding site and preferentially orientate the warhead relative to the cysteine, the electrophile moiety, to increase the rate, the selectivity and the potency of covalent bond formation (London et al., 2014). The challenge is to develop a covalent non-nucleoside inhibitor combining a scaffold for non-covalent interaction and an electrophilic warhead capable to covalently link the sulfur

atom of the key cysteine residue on the catalytic loop (Fig. S1). Selective inhibition of DNMTs will be based on the structural differences between isoforms of DNMTs (Erdmann et al., 2015b). Designing compounds with both specific non-covalent interactions and covalent binding will minimize off-target reactivity and ameliorate safety concerns (Zhu et al., 2014). The objective of the present study is to rationalize the inhibition of DNMTs by RG108-1 and RG119-1 as the binding mode of such inhibitors remains to be explored. Thereto, both compounds were analyzed in enzymatic and cellular assays. Crystallographic studies were also conducted leading to the first apo structure of M.*Hha*I. Finally, a new docking protocol (including covalent docking) was proposed based on differential scanning fluorimetry (DSF) assays, crystallographic studies and the mechanism of methylation of DNMTs.
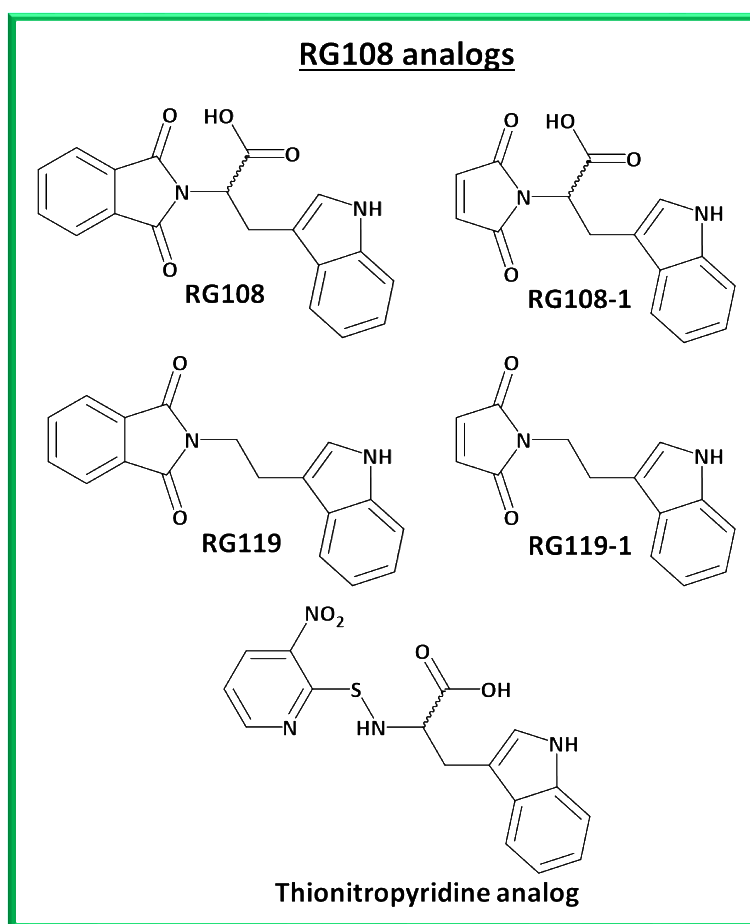

Figure 1. Chemical structures of the non-nucleoside DNA methylation inhibitor RG108 and its analogs. *RG108-1 and RG119-1 are the compounds studied in this work. The compounds are named according to the corresponding articles.*

2.2 Results and discussion

The reference RG108 compound (Fig. 1) was discovered by virtual screening using the National Cancer Institute (NCI) database and a homology model of human DNMT1 (Brueckner et al., 2005; Siedlecki et al., 2006). Authors reported an IC50 value of 0.6 µM against M.*Sss*I DNA methyltransferase and the reactivation of tumor suppressor genes, SFRP1 and p16^{Ink4a}, in colon cancer cells (Brueckner et al., 2005; Siedlecki et al., 2006). Nevertheless, other studies reported a weak activity of this compound toward DNMT1 (IC50 value of 390 µM) and murine DNMT3A (IC50 value >500 µM), reconsidering its selectivity and potency toward DNMTs (Asgatay et al., 2014; Ceccaldi et al., 2011; Suzuki et al., 2010). In 2010, novel DNMTis derived from RG108 and characterized by a maleimide scaffold were identified and presented a better inhibitory effect than RG108 (Fig. 1) (Suzuki et al., 2010).

So, we conducted this study on maleimide derivatives as their potential use as therapeutic agents recently emerged to treat cancer (Cashman et al., 2010), Alzheimer (Halim et al., 2007; Sivaprakasam et al., 2006), bacterial infection (López et al., 2003) and inflammatory diseases (Mahle et al., 2010; Matuszak et al., 2009). Covalent inhibitors can form a covalent bond with their target protein assisted by non-covalent interactions to improve selectivity, potency and be used for biological studies. We can cite as potential covalent non-nucleoside inhibitors of DNMTs, natural compounds parthenolide (IC50 of 3.5 µM against DNMT1), nanaomycin A (IC50 of 0.5 µM against DNMT3B) and curcumin (IC50 of 30nM against M.*Sss*I) (see Fig. 6 in chapter 1) (Kuck et al., 2010; Liu et al., 2009a; Liu et al., 2009b; Merfort, 2011). Such compounds are useful pharmacological tools for understanding the mechanisms of DNA methylation and links with other epigenetic modulations as exemplified by the recent probe develop for activity-based protein profiling technology based on the RG108-1 (Fig.1) (London et al., 2014; Zhu et al., 2015). These compounds are the basis for the design of new inhibitors of DNMTs. We evaluated also the activity of RG119-1 (Fig. 2), a tryptamine derivative, on DNMTs and compared it to RG108-1 as the presence of the carboxylate moiety does not seem necessary to improve the activity of such compound (Asgatay et al., 2014).

For the different studies (*in vitro* and crystallographic assays), our sample of RG108-1 was purchased in racemic form. Indeed, in the first report of RG108 study, the inhibitory activity was determined with the racemate (NSC401077) (Siedlecki et al., 2006). Furthermore, a study

of how the chirality of RG108 affects the activity was previously reported and showed no significant difference of activity between the enantiomers (Asgatay et al., 2014). Finally, our docking studies show that the binding poses of the two enantiomers are equivalent and form the same interactions with residues in the binding site (Fig. S2). RG119-1 was synthesized based on a previously described procedure and its structure was confirmed by X-ray diffraction (Dubois et al., 2016).

2.2.1 Inhibition of human DNMT1, human DNMT3A and bacterial M.*Sss*I by RG108-1 and RG119-1

Enzymatic assays were developed for human DNMT1, human DNMT3A and bacterial M.*Sss*I (Ceccaldi et al., 2011; Gros et al., 2013; Rilova et al., 2014). Dose-dependent inhibitions were observed and IC50 values were determined for RG108-1 and RG119-1 against those DNMTs (Table 1). RG119-1 showed a better inhibitory effect than RG108-1 against all DNMTs. For human DNMT3A and bacterial M.*Sss*I, RG119-1 has similar activities with IC50 values of 3.1 μM and 3.9 μM, respectively, and is 10-fold more potent than RG108-1 (Table 1). However, the inhibitory activity of RG119-1 against DNMT1 (IC50 value of 45 μM) is much better than RG108-1 (IC50 value in the millimolar range). The ability for RG108-1 to inhibit preferentially human DNMT3A can be exploited to develop a selective inhibitor.

Table 1. Inhibition of human DNMT1, human DNMT3A and bacterial M.SssI.

	DNMT1		DNMT3A		M.*Sss*I	
	IC50	95% CI	IC50	95% CI	IC50	95% CI
RG108-1 (μM)	**>1000**		**27**	17-42	**37**	16-87
RG119-1 (μM)	**45**	12-174	**3.1**	2.0-4.8	**3.9**	2.5-6.0
IC50, 50% inhibitory concentration in DNMTs assay; 95% CI, 95% confidence interval.						

2.2.2 Cytotoxic effect of RG108-1 and RG119-1 in mesothelioma cell lines (M14K and H28)

In this study, we evaluated also the cytotoxic effects of RG108-1 and RG119-1 on mesothelioma cell lines M14K and H28 (Fig. S3). M14K cell line is more sensitive compared to the more chemoresistant H28. These cell lines were chosen as hypermethylation of tumor suppressor genes was reported in malignant pleural mesothelioma (Batra et al., 2006). Results from cytotoxic assays show that RG108-1 has a weak activity, while RG119-1 is the most active compound (Table 1). Interestingly, their cytotoxicity activities are related to their inhibitory activities toward DNMT1 (Table 1 and Table 2). In mesothelioma cells, this enzyme represents more than 90% of total DNMTs expressed compared to the other isoforms (Amatori et al., 2009). RG119-1 has an IC50 value of 45 µM against DNMT1 and presents TC50 values in the low micromolar range (TC50$_{H28}$= 6.2 µM; TC50$_{M14K}$= 5.4 µM).

Table 2. Cytotoxic effect of RG108-1 and RG119-1 in M14K and H28 cell lines.

	H28		M14K	
	TC50	95% CI	TC50	95% CI
RG108-1 (µM)	>1000		>1000	
RG119-1 (µM)	6.2	3.9-9.6	5.4	3.2-9.2
TC50, 50% inhibitory concentration in cytotoxicity; 95% CI, 95% confidence interval.				

2.2.3 Apo structure of M.*Hha*I

In order to understand the mechanism of molecular recognition of RG108-1 and RG119-1 by DNMTs, crystallographic studies were performed using the bacterial C5-DNA methyltransferase from *haemophilus haemolyticus*, M.*Hha*I. This enzyme is the smallest DNMT (327 aa), is largely studied and serves as a model to investigate the inhibition mechanism of human DNMTs due to its high structural homology and sequence identity for the conserved motifs with human DNMTs (Figs. S4 and S5).

In 1993, the first crystal structure of M.*Hha*I with SAM was determined (PDB code: 1HMY) (Cheng et al., 1993). Since then, different ternary complexes of M.*Hha*I complexed with SAM cofactor or SAH product and DNA, containing methylated, mismatched target base or

nucleoside inhibitor were determined (Klimasauskas et al., 1994; Sheikhnejad et al., 1999; Zhou et al., 2002). However, no complex of DNMTs with any non-nucleoside inhibitors has been reported in the literature to date.

Prior to crystallization trials, differential scanning fluorimetry (DSF) assays were conducted to study the stabilizing effect of RG108-1 and RG119-1 with human DNMT1 alone, in presence of DNA and/or SAM. DSF is a highly sensitive method used to measure the thermal stability of a protein given by its melting temperature (Tm). If a compound stabilizes a protein, the free energy of unfolding (ΔGu) increases and so does the Tm (Niesen et al., 2007). The stability of the protein-ligand complex is therefore reflected by the difference in Tm (ΔTm) in the presence and absence of the ligand. In this study, RG108-1 does not appear to stabilize DNMT1 in any form (Table 3), an observation that can be correlated with the low *in vitro* inhibitory activity toward DNMT1 (Table 1). In contrast, RG119-1 induces a pronounced thermal shift (ΔTm$_{DNMT1}$ = ~2.8°C) in presence of DNA. We can interpret this result in the following way: as DNA binding stabilizes the closed form of the catalytic loop, which is the active conformation of DNMTs, interactions engaged by RG119-1 should be more favourable (Song et al., 2012). This hypothesis is reinforced by the fact that without DNA, we observe a slight stabilization of the DNMT1 enzyme alone and no stabilization in presence of cofactor SAM (Table 3). In presence of DNA and SAM, RG119-1 caused a smaller thermal shift (ΔTm$_{DNMT1}$ = ~1.3°C) than DNMT1-DNA system. It is difficult to interpret this difference, but it could be caused by a competition of RG119-1 with the SAM cofactor as previously reported for RG108 (Asgatay et al., 2014).

Table 3. Results of DSF (ΔTm [°C]) against DNMT1 alone, **in presence of DNA and/or SAM [a].**

	DNMT1	DNMT1-SAM	DNMT1-DNA	DNMT1-DNA-SAM
RG108-1	0.0	-0.3	-0.2	-0.8
RG119-1	0.3	-0.4	2.8	1.3

[a]*Compound concentration, 200 µM; protein concentration, 2.5 µM; SAM cofactor concentration, 5 µM ; DNA duplex concentration, 5 µM. Experiments were carried out in duplicate; average values are reported. Heat map shows relative ΔTm; red is for the largest ΔTm and blue indicates small ΔTm.*

Based on these results, we tried to obtain a crystal structure of the ternary complex of M.*Hha*I with DNA and RG108-1 or RG119-1 by co-crystallization or crystal-soaking using crystal

ternary complex (M.*Hha*I-SAH-DNA) but without success. Crystallization trials were also conducted in the absence of DNA in the hope to obtain a crystal structure complex of M.*Hha*I with RG108-1 or RG119-1 with no more success. However, we obtained the first crystal structure of the apoenzyme of M.*Hha*I despite a strong and stable association between the cofactor and M.*Hha*I, even after several purification steps (Kumar et al., 1992) and several dialysis (complex of M.*Hha*I-SAM obtained without addition of exogenous SAM).

Apo crystals of M.*Hha*I were obtained using two different methodologies. After a dialysis to decrease the concentration of endogenous SAM, the purified M.*Hha*I was crystallized by hanging-drop vapor-diffusion method in the presence of RG108-1 or RG119-1. For the second method, cocrystals of M.*Hha*I with exogenous SAM were soaked with RG108-1 or RG119-1. In both cases, we obtained apo crystals of M.*Hha*I crystallizing in the monoclinic space group P 1 21 1 (Table 4) as the previous binary complex of M.*Hha*I, with two molecules of M.*Hha*I in the asymmetric unit (O'Gara et al., 1999).

This crystal structure (Fig. 2) displays a vacant SAM-binding pocket with slight backbone displacement for key residues Glu40, Trp41, Asp60 and Ile 61 (Fig. 3a). These residues, except for Trp41, are conserved among DNMTs and mediate interactions with the cofactor (Fig. S6). In addition to these movements of the backbone, the most significant side chain conformational changes compared to the binary complex (M.*Hha*I-SAM, PDB code: 2HMY) (O'Gara et al., 1999) are found for residues Phe18 and Trp41 (Fig. 3b). These two residues interact normally with the adenine ring of the SAM, the Phe18 by π-π T-shaped interaction and the Trp41 by parallel π-π stacking interaction (Fig. S6). These observations indicate that the cofactor does not affect the cofactor binding site conformation of M.*Hha*I. More interestingly, RG119-1 seems to compete against SAM binding by displacing the cofactor from the binary complex as observed for the RG108 on the murine catalytic domain of DNMT3A (Asgatay et al., 2014). Unfortunately, RG119-1 was not observed in the electron density map probably reflecting high mobility and a low occupancy of the inhibitor in the active site. This also could be connected with our DSF results. Indeed, RG119-1 better stabilizes DNMT1 in presence of DNA than with DNMT1 alone, and it would probably be difficult to get a complex with the inhibitor in the absence of DNA.
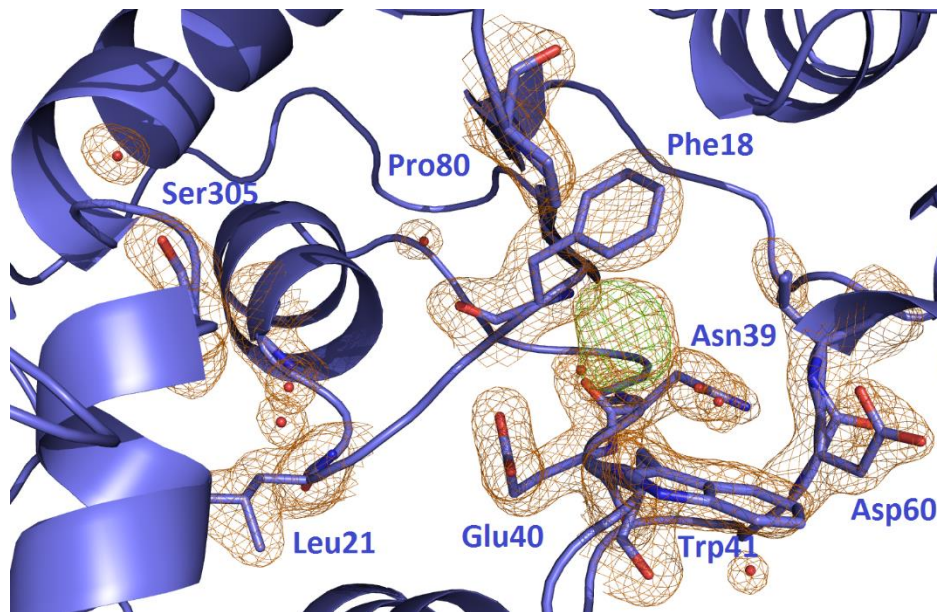
Figure 2. Structure of the M.HhaI apoenzyme. *Solid Ribbon representation of cofactor binding site with 2Fo-Fc electron density map contoured at 1.0 σ for residues (orange mesh) normally involved in cofactor interaction and a Fo-Fc omit map contoured at 2.0 σ (green mesh) near the cofactor-binding site.*
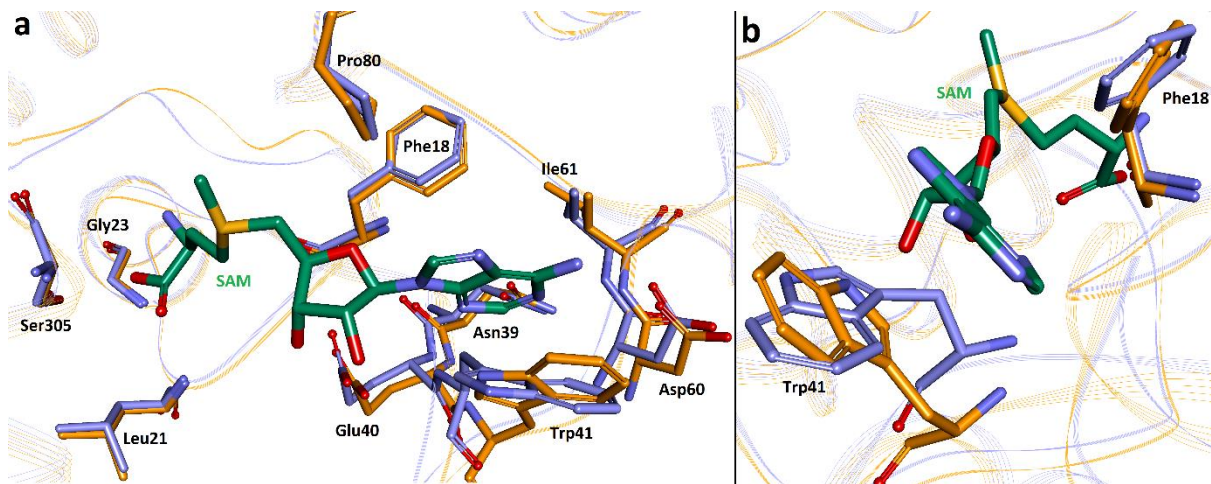
Figure 3. Superposition of binary complex M.HhaI-SAM cofactor (orange) (O'Gara et al., 1999) (PDB code: 2HMY) with the apoenzyme (blue). *Amino acid residues of the binding site are rendered as stick models while rest of the protein backbone is displayed as line ribbon. SAM cofactor is rendered as green stick model. (**a**) Slight backbone displacement for residues Glu40, Trp41, Asp60, and Ile 61 and (**b**) conformational side-chain rearrangements for residues Phe18 and Trp41 when cofactor binds to M.HhaI.*

Table 4. Data collection and refinement statistics.

	Apoenzyme M.*Hha*I
Data collection	
Space group	P 1 21 1 (No. 4)
Cell dimensions	
a, *b*, *c* (Å)	53.0, 70.0, 87.8
α, β, γ (°)	90.0, 101.3, 90.0
Resolution (Å)	41.73 - 1.90 (1.97 - 1.90) *
R_{sym} or R_{merge} (%)	5.8 (42.9)
$I/\sigma I$	12.42 (2.77)
Completeness (%)	99.61 (99.74)
Redundancy	3.68 (3.57)
Refinement	
Resolution (Å)	41.73 - 1.9
No. reflections	49517
R_{work} / R_{free} (%)	17.84/21.94
No. atoms	
Protein	5122
Ligand/ion	25 (sulfate)
Water	246
B-factors	
Protein	34.27
Ligand/ion	40.63
Water	39.54
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.08

*Values in parentheses are for highest-resolution shell.

2.2.4 Molecular docking of RG108-1 and RG119-1 into DNMTs

To investigate the mode of interaction of RG108-1 and RG119-1 with DNMTs, we performed non-covalent docking followed by covalent docking. Firstly, we studied by non-covalent docking the conformation of RG108-1 and RG119-1 inside the active site, and the orientation of the maleimide moiety relative to the catalytic cysteine to perform the covalent binding. Secondly, we examined by covalent docking the conserved interactions and the ones newly formed to maintain and stabilize this complex during the reaction process.

Based on the results of the DSF experiments with DNMT1, crystallographic studies and the mechanism of methylation of DNMTs, we propose here a new docking protocol. Several structures of DNMTs are available in the PDB (Protein Data Bank) (Berman et al., 2000) either

in complex with the SAM cofactor, in complex with a DNA probe and SAH, or alone. Crystallographic structures of mouse DNMT1 (PDB code: 4DA4, 2.6 Å) (Song et al., 2012) and M.*Hha*I (PDB code: 1M0E, 2.5 Å) (Zhou et al., 2002) both in ternary complex with SAH and DNA were used for both non-covalent and covalent docking. Indeed, the conformation of DNMTs approaches the active conformation as the DNA stabilizes the closed form of the catalytic loop which covers the active site. This effect is especially pronounced for the bacterial M.*Hha*I (Fig. S7). When DNA binds to M.*Hha*I, the catalytic loop undergoes a major conformational change (2.5 nm) to close the catalytic site (Sankpal and Rao, 2002). This ensures that the nucleophilic catalytic Cys81 is close to the target cytosine and allows a direct attack at the C6 position.

Docking simulations reported in the literature with DNMTs were conducted with the active conformation of the catalytic domain taken from the ternary complex by removing the DNA duplex and cofactor (Rilova et al., 2014; Suzuki et al., 2010), with the binary complex by removing the cofactor (Valente et al., 2014) and homology models (Kuck et al., 2010; Yoo and Medina-Franco, 2011). In our study, we decided to keep the DNA double-helical structure during docking simulations as this partner permits to close the loop for M.*Hha*I and to change the conformation of side-chain residues by steric effects into the active form for human DNMTs (Fig. S5). Furthermore, based on our DSF experiments, we conclude that RG119-1 stabilizes more the DNMT1 in presence of the DNA. However, to study the substrate pocket in the active form, we extracted the flipped substrate as the base flipping is a reversible process. DNMTs, as explained in the methylation mechanism (Fig. S1), flip the deoxycytidine out of the DNA helix into the catalytic pocket. The kinetic scheme proposed for M.*Hha*I includes first the binding of DNA, followed by the binding of the cofactor, the loop closing and the base flipping occurring almost simultaneously (Gerasimaitė et al., 2011; Kurkcuoglu et al., 2012). All these elements led us to keep the DNA during the docking simulations.

The docking protocol was first validated by docking the SAH inside the DNMTs. The proposed binding mode for the cofactor within the active site superimposed well with M.*Hha*I and DNMT1 with a RMSD of 0.89 Å and 1.34 Å, respectively (<2 Å) (Fig. S8). Binding of RG108-1 and RG119-1 was then explored inside the cofactor-binding site and the catalytic cavity, where the cytosine binds (Fig. S5), in order to model competition with the cofactor and DNA substrate, respectively. The SAM binding pocket is highly conserved among DNMTs with residues

forming hydrogen bond interactions and surrounding the adenine ring of cofactor with a hydrophobic pocket (Fig. S6). However, Trp41 in M.*Hha*I is replaced by Met1172 in DNMT1 and Val661 in DNMT3A.

Our non-covalent docking simulations of RG108-1 and RG119-1 on bacterial M.*Hha*I and mouse DNMT1 (Fig. 4) suggest that the indole ring of both ligands, like the adenine ring of the cofactor, occupies a hydrophobic pocket surrounded by residues Phe18, Trp41, Ile61 and Leu100 for M.*Hha*I, and residues Phe1148, Met1172 and Leu1250 for DNMT1 (Fig. 4c and d). Furthermore, this moiety superimposes well with the adenine ring of SAH and forms similar interactions with the cofactor site (Fig. S9). The indole ring of RG108-1 and RG119-1 is involved in π-π T-shaped interaction with the phenylalanine residue and parallel π-π stacking interaction with Trp41 for M.*Hha*I and π-sulfur interaction with Met1172 for DNMT1. Such aromatic interactions seem important as derivatives lacking the indole moiety of RG108 are inactive and its bioisosteric replacement by a benzothiophene, containing a sulfur atom in place of the nitrogen atom, increases its inhibitory potency by two (Asgatay et al., 2014). Moreover, the donor NH group from the indole ring does not interact with any surrounding residue.

For M.*Hha*I, maleimide ring of both compounds is directed to the substrate pocket and mediates, via a carbonyl group, a moderate hydrogen interaction (D_{A-B}= 3.0 Å) with Gln82 residue. This residue seems required for the preferential orientation of the warhead of RG108-1 and RG119-1 directed to the catalytic cysteine. Interestingly, for DNMT1, only RG119-1 forms a hydrogen bond with Gln1230 via the carbonyl group (D_{A-B}= 3.1 Å). For RG108-1, the maleimide ring (warhead) seems not stabilized by hydrogen bonding as reflected by the multiple conformations of the top docking solutions for this moiety.

In addition, the carboxylate makes a salt bridge with Lys1247 (D = 2.8 Å) and a hydrogen bond with Gln1230 (D_{A-B}= 3.1 Å) which could prevent the RG108-1 to react with the catalytic cysteine by blocking the compound. This could explain the relatively low inhibitory activity of RG108-1 against DNMT1. In the case of human DNMT3A and M.*Hha*I, the lysine residue (Lys1247) is replaced by a threonine residue (Thr727) and an arginine residue (Arg97), respectively. However, it is unlikely that Arg97 from M.*Hha*I interacts with the carboxylate of RG108-1 as it is engaged in a bidentate interaction with a phosphate group of the DNA

backbone and is apart from the binding site. Additionally, a π-sulfur interaction is formed also between the maleimide group of both compounds and the catalytic cysteine (Cys81 residue for M.*Hha*I and Cys1229 for DNMT1).
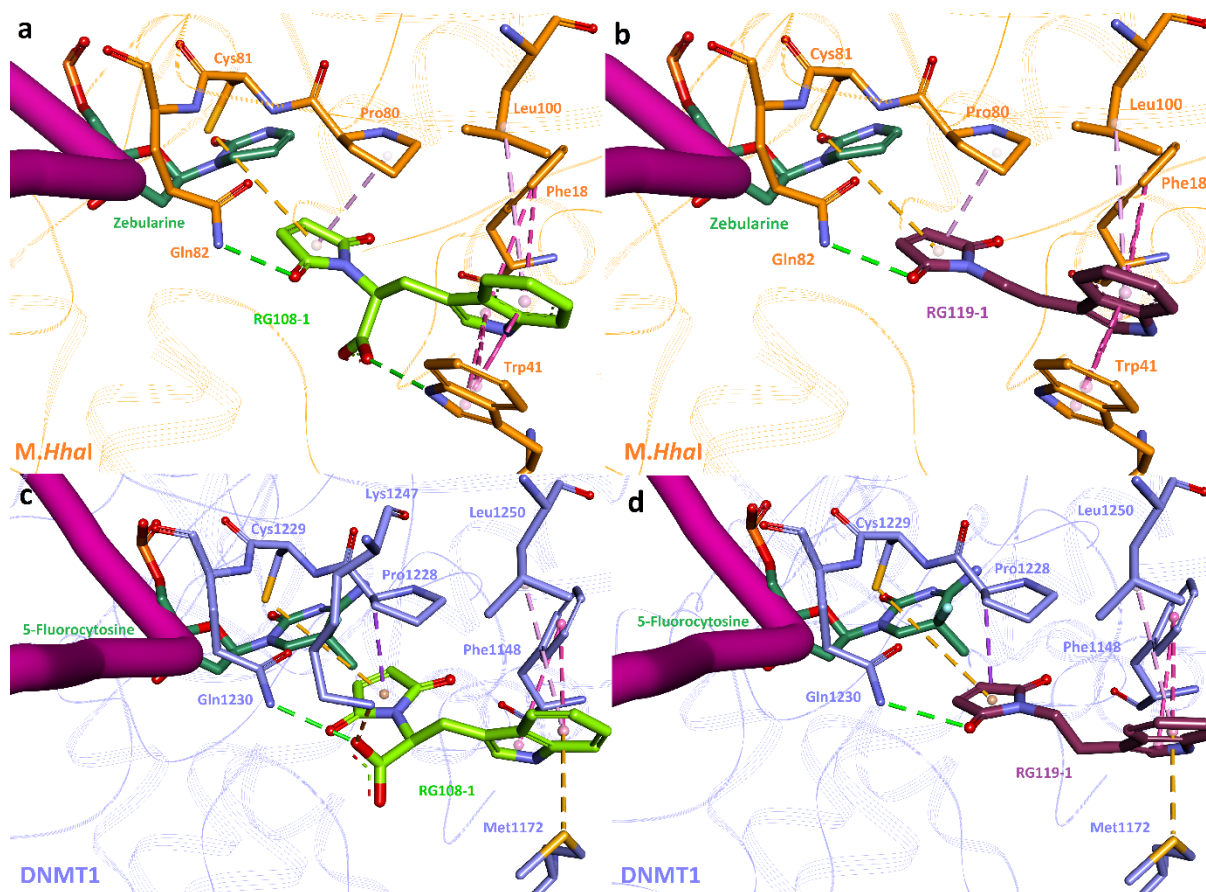

Figure 4. Non-covalent docking of RG108-1 (a,c) and RG119-1 (b,d) into bacterial M.HhaI (top) and mouse DNMT1 (bottom). *Ligands and amino acid residues of the binding site are rendered as stick models while rest of the protein backbone is displayed as line ribbon. DNA backbone is displayed as purple tube model. Zebularine and 5-Fluorocytosine, removed prior to docking simulations, were added for final visualization. Molecular interactions are shown in dotted lines with the following color codes: green for conventional hydrogen bond, yellow for π-sulfur interaction, orange for salt bridge, purple for π-sigma interaction, light pink for π-alkyl interaction and pink for π-π stacking interaction. (**a** and **b**) Details of the binding mode of RG108-1 (**a**; green; docking pose) and RG119-1 (**b**; purple; docking pose) to M.HhaI (orange; PDB code: 1M0E). RG108-1 and RG119-1 both interact with M.HhaI via hydrogen bond involving Gln82, van der Waals interaction involving Pro80 and Leu100, π-π T-shaped interaction involving Phe18, parallel π-π stacking interaction involving Trp41 and π-sulfur interaction involving Cys81. RG108-1 makes an additional interaction with Trp41 by hydrogen bonding. (**c** and **d**) Details of the binding mode of RG108-1 (**c**; green; docking pose) and RG119-1 (**d**; purple; docking pose)*

to DNMT1 (blue; PDB code: 4DA4). RG108-1 and RG119-1 both interact with DNMT1 via van der Waals interaction involving Pro1128 and Leu1250 and π-sulfur interaction involving Cys1229. Carboxylate of RG108-1 interacts with Gln1230 via a hydrogen bond and Lys1247 via a salt bridge. RG119-1 interacts with Gln1230 via the carbonyl group.

Based on this non-covalent docking study, we propose that RG108-1 and RG119-1 preferentially occupy the cofactor pocket and could act as inhibitors competitive to the SAM. This hypothesis was verified with RG108 on the murine catalytic domain of DNMT3A by conducting a WaterLOGSY (water-ligand observed via gradient spectroscopy) (Asgatay et al., 2014). The same hypothesis was proposed from previous reported docking studies but with different binding mode as these simulations were performed without DNA duplex (Suzuki et al., 2010; Yoo and Medina-Franco, 2011). In these studies, the carboxylate group of RG108-1 interacts with arginine residues normally involved in DNA binding. Interestingly, the maleimide moiety of each compound points into the cytosine cavity close to the catalytic cysteine.

Concerning the catalytic domain of mouse DNMT3A in complex with C-terminal DNMT3L (PDB code: 2QRV) (Jia et al., 2007), no preferential docking pose was obtained despite a good conformation of the active site loop stabilized by the DNMT3L. After analysis of structural alignment between ternary complexes of DNMT1 and M.*Hha*I with DNMT3A, it appears that the residue Asn707 in DNMT3A possesses a different conformation (flip of 90°) compared to Gln1230 for DNMT1 and Gln82 for M.*Hha*I. In our results, this residue mediates a moderate hydrogen bonding with a carbonyl group of the maleimide ring.

These non-covalent interactions could assist covalent bond formation, and we further performed a covalent docking simulation to investigate this complex formation. The role of the α,β-unsaturated ketone seems important for the formation of a Michael-type adduct with the sulfhydryl group of the conserved catalytic cysteine in DNMTs, as the replacement of the maleimide group of RG108-1 by a succinimide group conducts to an inactive compound (Suzuki et al., 2010). The covalent binding was then modelled using GOLD Suite software (v5.2.2, CCDC, Cambridge, UK) by forcing the ligand link atom to occupy the same steric volume as the DNMT sulfur atom. Recently, It was suggested that deprotonation of the catalytic cysteine residue occurs after DNA binding (Aranda et al., 2016). The unprotonated

cysteine facilitates the nucleophilic attack onto the 6-position of cytosine and the transition state is stabilized by hydrogen bond interactions with protonated Glu119 (Fig. S1).

For M.*Hha*I, aromatic interactions are maintained after covalent binding and Pro80 interacts at present with the indole ring (Fig. 5a and b) as observed for DNMT1 (Fig. 5c and d), but not for the Met1172 residue. However, interaction of maleimide group with Gln82 is lost, but the protonation state of Glu119 permits to form a moderate hydrogen bond (D_{A-B} = 3.2 Å) with the opposite carbonyl moiety. The same observation is made for Glu1269 of DNMT1 (D_{A-B} = 2.9 Å for RG108-1 and D_{A-B} = 2.6 Å for RG119-1) (Fig 5c and d). The carboxylate group of RG108-1 does not interact with any residue of M.*Hha*I and DNMT1 (Fig 5a and c). The contribution of the negative charge was recently experienced by replacing the carboxylate anion with a neutral isosteric moiety (Asgatay et al., 2014). Replacement of the carboxylate group did not change the inhibitory activity of this compound leading to reassess the role of this moiety for RG108 and RG108-1. Interestingly, for DNMT1, indole group of RG119-1 maintains π-π T-shaped interaction with Phe1148 and makes an additional interaction with Glu1171 (as observed for M.*Hha*I) with a strong hydrogen bonding interaction (D_{A-B} = 2.5 Å) (Fig 5d), while indole ring of RG108-1 is only surrounded by Pro1128 for DNMT1 (Fig 5c).
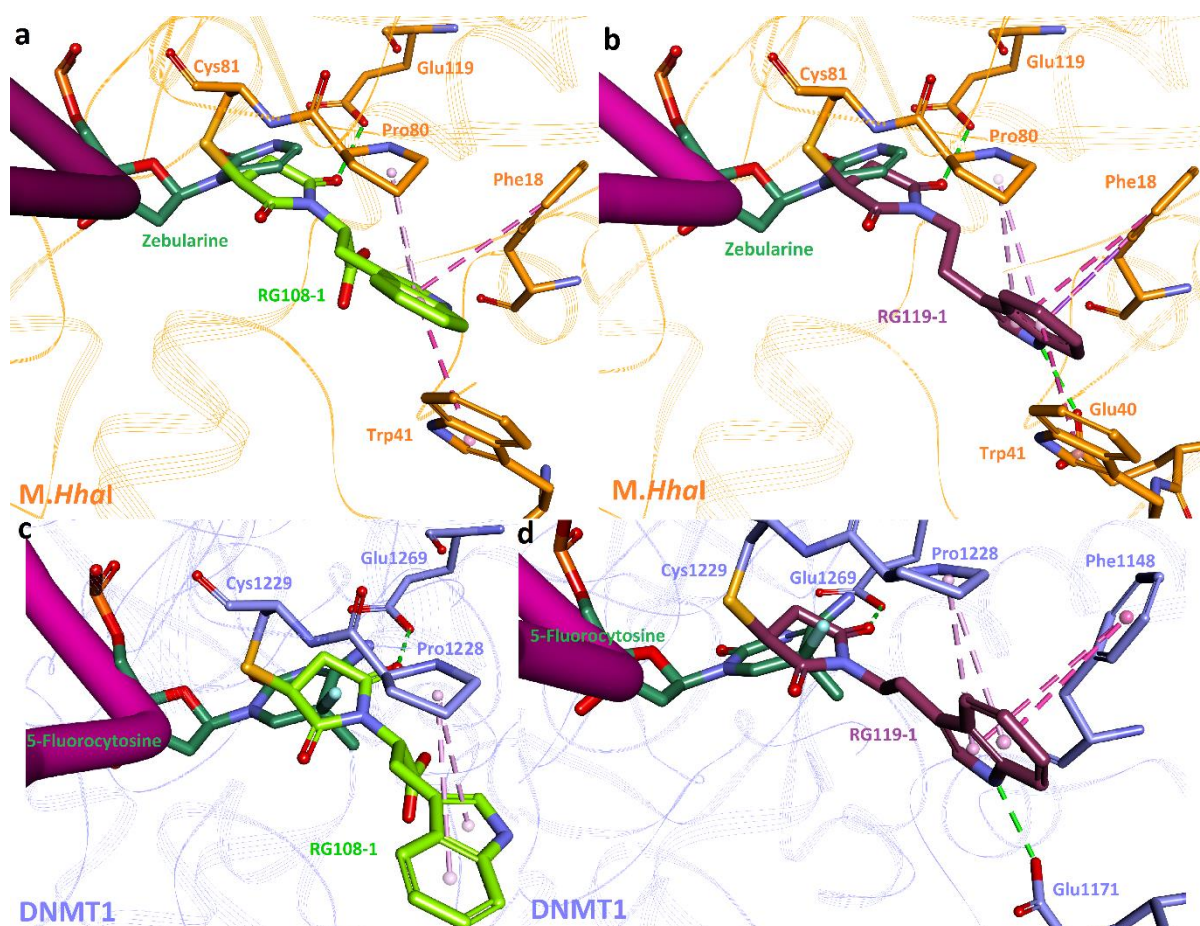
Figure 5. Covalent docking of RG108-1 and RG119-1 into bacterial M.HhaI and mouse DNMT1.
*Ligands and amino acid residues of the binding site are rendered as stick models, while the rest of the protein backbone is displayed as line ribbon. DNA backbone is displayed as purple tube model. Zebularine and 5-Fluorocytosine, removed prior to docking simulations, were added for final visualization. Molecular interactions are shown in dotted lines with the following color codes: green for conventional hydrogen bond, yellow for π-sulfur interaction, purple for π-sigma interaction, light pink for π-alkyl interaction and pink for π-π stacking interaction. (**a** and **b**) Details of the binding mode of RG108-1 (**a**; green; docking pose) and RG119-1 (**b**; purple; docking pose) to M.HhaI (orange; PDB code: 1M0E). RG108-1 and RG119-1 both interact with M.HhaI via hydrogen bond involving Glu119, van der Waals interaction involving Pro80, π-π T-shaped interaction involving Phe18 and parallel π-π stacking interaction involving Trp41. RG119-1 makes an additional interaction with Glu40 by hydrogen bonding. (**c** and **d**) Details of the binding mode of RG108-1 (**c**; green; docking pose) and RG119-1 (**d**; purple; docking pose) to DNMT1 (blue; PDB code: 4DA4). RG108-1 and RG119-1 both interact with DNMT1 via hydrogen bond involving Glu1269, and van der Waals interaction involving Pro1128. Indole group of RG119-1 makes additional π-π T-shaped interaction with Phe1148 and with Glu1171 by hydrogen bonding.*

2.3 Conclusion

DNA methyltransferases (DNMTs) are highly studied because of their role in epigenetic regulation leading to the silencing of tumor suppressor genes during oncogenesis. Accordingly, identification of novel non-nucleoside DNMT inhibitors acting directly on DNMTs without incorporation into DNA are necessary to decrease side effects of current treatments and increase selectivity against DNMTs for personal cancer therapy.

In this study, we selected two maleimide derivatives of RG108 (RG108-1 and RG119-1) to rationalize their mode of action against DNMTs. From enzymatic assays, *in vitro* inhibitory activities of RG119-1 are slightly better than those of RG108-1 on DNMT3A and M.*Sss*I (by a factor of 10), but much better against DNMT1 (IC50 value of 45 μM). Therefore, RG108-1 seems to be selective for DNMT3A (IC50 value of 27 μM), since no activity against DNMT1 was observed. Interestingly, this weak activity of RG108-1 toward DNMT1 is reflected in the cytotoxicity on mesothelioma cells while RG119-1 demonstrates potent inhibition of cell proliferation. This difference in activity can be explained by the high expression of DNMT1 in mesothelioma cells *versus* other isoforms.

The weak activity of RG108-1 towards DNMT1 is still difficult to explain, but could be related to the additional interactions of the carboxylate with Lys1247 (salt bridge) and Gln1230 (hydrogen bond) that could block RG108-1 in this conformation which does not allow covalent reaction with the catalytic cysteine of DNMT1. Furthermore, our non-covalent docking shows that the maleimide ring of RG108-1 makes no significant interactions. In addition, if a covalent complex is formed, the carboxylate group of RG108-1 does not interact with any residue of M.*Hha*I and DNMT1. Therefore, this moiety does not appear to participate in the stabilization of the covalent complex. Docking analysis suggests that RG108-1 and RG119-1 inhibit DNMTs activity by competition with both cytosine substrate and SAM cofactor. In the non-covalent docking analysis, the maleimide moiety of each inhibitor points into the cytosine cavity and could form a covalent bond with the catalytic cysteine. The covalent docking reveals a certain conservation of the non-covalent interactions after reaction and highlights essential residues for the stabilization of the inhibitors.

Crystallization trials did not allow to observe a complex of M.*Hha*I with RG108-1 and RG119-1 but the presence of the inhibitors conducted to the first apo structure of M.*Hha*I. This reveals that the SAM cofactor does not affect the cofactor binding site conformation of DNMTs and that it is probably displaced by our inhibitors, suggesting a competitive mode of action. Covalent inhibitors can improve selectivity, potency and be used for biological studies. Indeed, such compounds are useful pharmacological tools for understanding the mechanisms of DNA methylation and links with other epigenetic modulations and the basis for the design of inhibitors of DNMTs.

2.4 Materials and methods

General

All commercially available reagents and solvents were purchased from Sigma-Aldrich, VWR France, radioactive [methyl-^{3}H] SAM from Perkin-Elmer. 10 mM stock solutions of chemicals were prepared in DMSO and aliquoted.

2.4.1 Chemistry

The racemic mixture of RG108-1 was purchased from Enamine (Kiev, Ukraine). Synthesis of RG119-1 (1-[2-(1H-Indol-3-yl)-ethyl]-pyrrole-2,5-dione) was adapted from previous work (Casimir et al., 2002) and the structure was confirmed by X-ray diffraction (Dubois et al., 2016).

2.4.2 Enzyme production

Full length human DNMT1 (182 kDa, histidine-tagged) was produced and purified according to Lee et al. (2005). Catalytic human DNMT3Acat (623-908 a.a.) was produced and purified according to Gros et al. (2013). The *Sss*I methyltransferase (M.*Sss*I) was obtained from New England Biolabs (France). Overexpression of M.*Hha*I in *E. coli* strain ER1727 containing the pUHE25HhaIM plasmid (provided by Dr. S. Kumar, New England Biolabs) and purification were performed according to previous work (Kumar et al., 1992; Rondelet et al., 2016).

2.4.3 DNMT inhibition assays

DNMT1 inhibition assay was developed and described in Gros et al. (2013), DNMT3Acat inhibition assay was described in Rilova et al. (2014) and M.*Sss*I inhibition assay was described in Ceccaldi et al. (2011). The concentration at which 50 % of inhibition is observed (IC50) was determined by analysis of a concentration range of the tested compound in duplicate. The negative and positive controls were defined as points without enzyme and points without any compound, respectively. The percentages of inhibition (%I) were calculated as following: %I = 100-100x(cpm_i-cpm_{neg})/(cpm_{pos}-cpm_{neg}), where cpm_i is the inhibitor signal, cpm_{neg} the

negative control signal and cpm_{pos} the positive control signal. The non-linear regression fittings with sigmoidal dose-response (variable slope) were performed with GraphPad Prism 4.03 (GraphPad Software). SGI-1027 was used as a reference inhibitor ($IC50_{DNMT3A}$ = 0.9 µM and $IC50_{DNMT1}$ = 10 µM) (Erdmann et al., 2015b).

2.4.4 Differential Scanning Fluorimetry assay

Experiments were conducted using a CFX384™ Real-Time System (C1000 Thermal cycler, Bio-Rad CFX Manager 2.0 Software, Bio-Rad, USA). The samples were heated at 0.5°C/s, from 10 to 80°C. The fluorescence intensity was plotted as a function of the temperature. $T_{1/2}$ was given by the inflection point of the fluorescence curve. $\Delta T_{1/2}$ were calculated by subtracting the $T_{1/2}$ in the absence of the compound to the $T_{1/2}$ in the presence of the compound in the same condition (*i.e.,* in the presence of other partners DNA and/or SAM).

The protein was scanned to assess suitability of the method and the lowest concentration of DNMT1 protein needed to generate a strong signal was determined to be 2.5 µM. Compound concentration was 200 µM. The DNA duplex used in the enzymatic assays was chosen and added at 5 µM. The SAM cofactor was added or not in a final concentration of 5 µM. The Sypro orange dye (Invitrogen) was diluted to 1/400th in each sample. Each experiment was repeated for at least two times in duplicate.

2.4.5 Cell viability assays

M14K and H28 mesothelioma cell lines were maintained in Dulbecco's Modified Eagle Medium-High glucose with L-glutamine (DMEM-HG, Lonza, Westburg, Leusden, Netherlands) supplemented with 10% of fetal bovine serum (FBS, Gibco) and penicillin-streptomycine 10% (v/v) (Vandermeers et al., 2013). Cells were grown at 37 °C in a humidified atmosphere containing 5% CO_2. Cells were seeded and incubated in 96-well plates (3.10^3 cells per well) for 4 hours prior to treatment with the drugs. Drugs were suspended in DMSO and prepared to achieve a concentration range of 0 to 10 mM for RG108-1 and 0 to 1 mM for RG119-1. The final concentration of DMSO in the wells was 1% (v/v). After drugs treatment, the cells were maintained in culture for 72 h. Then, cell viability was determined using the CellTiter 96® AQueous Non-Radioactive Cell Proliferation Assay Kit (Promega). For that, 20 µL of a reagent mix (MTS tetrazolium/phenazine methosulfate) was added to 100 µL media per well. After 1

h incubation at 37°C, the absorbance of the formazan product in the 96-well plates was measured at a wavelength of 490 nm on a Wallac 1420 Victor2 Microplate Reader (PerkinElmer Life Sciences). Data are normalized to the control condition. The non-linear regression fittings with sigmoidal dose-response (variable slope) were performed with GraphPad Prism 5.0 (GraphPad Software). Experiments were repeated two times in triplicate.

2.4.6 Crystallization and structure determination of the apoenzyme form of M.*Hha*I

Crystallization conditions for the binary complex (M.*Hha*I-SAM) were previously reported (Kumar et al., 1992) and crystallization assays to obtain the ternary complex (M.*Hha*I-DNA-Inhibitors) were based and adapted from previously published data (O'Gara et al., 1996; O'Gara et al., 1998). Apo crystals of M.*Hha*I were obtained following two different methodologies. In the first place, dialysis is performed at 4 °C overnight to decrease the concentration of endogenous SAM after the final step of purification. Co-crystallization assays of M.*Hha*I (10 mg/mL) and RG108-1 or RG119-1 were carried out with a molar ratio of 1:3 using the hanging-drop vapor-diffusion method at room temperature. Drops consisted of 1 µL of mixture plus 1 µL of reservoir solution equilibrated against a reservoir volume of 700 µL. Crystals were grown against a reservoir solution consisting of 50 mM ammonium sulfate, 100 mM citrate buffer at pH 6.0-6.6, and 22-32% (w/v) PEG-4000. For the second method, cocrystals of M.*Hha*I with exogenous SAM (Sigma Aldrich) obtained with a molar ratio of 1:2 were soaked with a concentration range of 0.5-5 mM of RG108-1 or RG119-1 for 10-30 min. Single crystals were flash-cooled in liquid nitrogen and data sets were collected at SOLEIL Synchrotron, Gif-sur-Yvette, France on beamline PROXIMA 2 (PX2-A) using ADSC Q315r detector at a wavelength of 0.9801 Å. The select crystal diffracts up to 1.9 Å resolution and was used for structure determination. The data were processed using XDS/XSCALE (Kabsch, 2010). A summary of the data-collection and refinement statistics is presented in Table 4. Initial phases were calculated by molecular replacement using the program PHASER in PHENIX (Adams et al., 2010; McCoy et al., 2007) with experimental reflection data set (Fo) and the search model (Fc) solved at 2.6 Å resolution (PDB code: 2HMY) (O'Gara et al., 1999). The complex was built using the Coot program (Emsley et al., 2010) and refined with the program PHENIX (Adams et al., 2010). Ramachandran plot for the final model shows 97.5% residues in

favoured regions, 2.2% in allowed regions and 0.3% in disallowed regions. Figures were drawn using both Pymol (DeLano, 2002) and Discovery Studio (BIOVIA, 2015).

2.4.7 Molecular docking

Ligand preparation

Non-covalent: RG108-1 (R and S) and RG119-1 were prepared by using the "Prepare Ligands" tool in Discovery Studio protocols. The pH was set to 7.4 to calculate the degree of ionization of compounds and the other parameters were turned to "False". A minimization procedure was applied to the molecules with MMFF (Merck molecular force field) force field (Halgren, 1996) and the conjugate gradient algorithm with a convergence criteria of 0.01 $kcal.mol^{-1}.Å^{-1}$.

Covalent: Based on the non-covalent docking and the identification of the bond-forming group, ligands were changed for covalent docking. Structure of the ligands contains the succinimide group and the link atom with a free valence (here the sulfur atom of cysteine) which was also specify for the protein. The covalent link between the succinimide group of ligands and the sulfur atom of the catalytic cysteine (Cys81 for M.*Hha*I and Cys1229 for mouse DNMT1) was specified in GOLD.

Protein preparation

Ternary complex of mouse DNMT (DNMT1-SAH-hemimethylated CpG duplex, PDB code: 4DA4, 2.6 Å) and ternary complex DNA of methyltransferase *Hha*I (M.*Hha*I-SAH-DNA, PDB code: 1M0E, 2.5 Å) were used for both non-covalent and covalent docking. Binary complex of human DNMT3A (DNMT3A-SAH, PDB code: 2QRV, 2.9 Å) was used for non-covalent docking. The binding site was defined as a sphere of 15 Å radius centered on the sulfur atom of SAH including both the cofactor and cytosine binding pockets. Cofactor product SAH found in all structures studied, 5-fluorocytosine from mouse DNMT1 (5-FC) and Zebularine from bacterial M.*Hha*I complex were extracted prior to docking simulations. DNA double-helical structure was kept during docking simulations. Rotamers of asparagine and glutamine residues were checked and corrected (e.g., Gln1230) using NQ-Flipper (Weichenberger and Sippl, 2006). Proteins were prepared for docking by using the "Prepare Protein" tool in Discovery Studio

protocols. This procedure permits to adjust the pH to 7.4, add missing hydrogens, fix the missing side chains and assign atom types (corrected connectivity and bond orders) by applying CHARMm (Chemistry at HARvard Macromolecular mechanics) force field (version c36b2) (Brooks et al., 1983).

Docking Simulations

RG108-1 and RG119-1 were docked within the defined binding site of each protein using GOLD 5.3.0 docking software tool (Jones et al., 1997). Number of genetic algorithm (GA) iterations was changed to 100 and the top 20 docking poses were generated for every ligand. Final poses were scored with ChemPLP fitness function (Korb et al., 2009). In each cluster of 20 conformations generated, conformational analysis was performed to select the most representative conformation with the highest score. Images were generated using Discovery Studio (BIOVIA, 2015).

Data deposition

Coordinates and diffraction data for the apo structure of M.*Hha*I have been deposited in the Protein Data Bank with accession code 5LOD.

Footnotes

2.5 References

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D: Biological Crystallography 66, 213-221.

Amatori, S., Papalini, F., Lazzarini, R., Donati, B., Bagaloni, I., Rippo, M., Procopio, A., Pelicci, P., Catalano, A., Fanelli, M., 2009. Decitabine, differently from DNMT1 silencing, exerts its antiproliferative activity through p21 upregulation in malignant pleural mesothelioma (MPM) cells. Lung Cancer 66, 184-190.

Aranda, J., Zinovjev, K., Świderek, K., Roca, M., Tuñón, I., 2016. Unraveling the Reaction Mechanism of Enzymatic C5-Cytosine Methylation of DNA. A Combined Molecular Dynamics and QM/MM Study of Wild Type and Gln119 Variant. ACS Catalysis 6, 3262-3276.

Asgatay, S., Champion, C., Marloie, G., Drujon, T., Senamaud-Beaufort, C., Ceccaldi, A., Erdmann, A., Rajavelu, A., Schambel, P., Jeltsch, A., 2014. Synthesis and evaluation of analogs of N-phthaloyl-l-tryptophan (RG108) as inhibitors of DNA methyltransferase 1. Journal of medicinal chemistry 57, 421-434.

Batra, S., Shi, Y., Kuchenbecker, K.M., He, B., Reguart, N., Mikami, I., You, L., Xu, Z., Lin, Y.-C., Clément, G., 2006. Wnt inhibitory factor-1, a Wnt antagonist, is silenced by promoter hypermethylation in malignant pleural mesothelioma. Biochemical and biophysical research communications 342, 1228-1232.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein data bank. Nucleic acids research 28, 235-242.

Bestor, T.H., 2000. The DNA methyltransferases of mammals. Human molecular genetics 9, 2395-2402.

Bestor, T.H., Verdine, G.L., 1994. DNA methyltransferases. Current opinion in cell biology 6, 380-389.

BIOVIA, D.S., 2015. Discovery studio modeling environment. Dassault Systèmes: San Diego, CA, USA.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M., 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. Journal of computational chemistry 4, 187-217.

Brueckner, B., Boy, R.G., Siedlecki, P., Musch, T., Kliem, H.C., Zielenkiewicz, P., Suhai, S., Wiessler, M., Lyko, F., 2005. Epigenetic reactivation of tumor suppressor genes by a novel small-molecule inhibitor of human DNA methyltransferases. Cancer research 65, 6305-6311.

Cashman, J.R., MacDonald, M., Ghirmai, S., Okolotowicz, K.J., Sergienko, E., Brown, B., Garcia, X., Zhai, D., Dahl, R., Reed, J.C., 2010. Inhibition of Bfl-1 with N-aryl maleimides. Bioorganic & medicinal chemistry letters 20, 6560-6564.

Casimir, J.R., Guichard, G., Briand, J.-P., 2002. Methyl 2-((succinimidooxy) carbonyl) benzoate (MSB): a new, efficient reagent for N-phthaloylation of amino acid and peptide derivatives. The Journal of organic chemistry 67, 3764-3768.

Ceccaldi, A., Rajavelu, A., Champion, C., Rampon, C., Jurkowska, R., Jankevicius, G., Sénamaud-Beaufort, C., Ponger, L., Gagey, N., Dali Ali, H., 2011. C5-DNA Methyltransferase Inhibitors: From Screening to Effects on Zebrafish Embryo Development. Chembiochem 12, 1337-1345.

Cheng, X., Kumar, S., Posfai, J., Pflugrath, J.W., Roberts, R.J., 1993. Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. Cell 74, 299-307.

Christman, J.K., 2002. 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. Oncogene 21, 5483-5495.

DeLano, W.L., 2002. The PyMOL molecular graphics system.

Dubois, J., Colaço, M., Rondelet, G., Wouters, J., 2016. Synthesis and Crystallographic Characterization of a Maleimide Derivative of Tryptamine. Crystals 6, 153.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta Crystallographica Section D: Biological Crystallography 66, 486-501.

Erdmann, A., Halby, L., Fahy, J., Arimondo, P.B., 2015a. Targeting DNA Methylation with Small Molecules: What's Next? Miniperspective. Journal of medicinal chemistry 58, 2569-2583.

Erdmann, A., Menon, Y., Gros, C., Molinier, N., Novosad, N., Samson, A., Gregoire, J.-M., Long, C., Ausseil, F., Halby, L., 2015b. Design and synthesis of new non nucleoside inhibitors of DNMT3A. Bioorganic & medicinal chemistry 23, 5946-5953.

Esteller, M. 2008. Epigenetics in cancer. The New England Journal of Medicine 358, 1148-1159

Gerasimaitė, R., Merkienė, E., Klimašauskas, S., 2011. Direct observation of cytosine flipping and covalent catalysis in a DNA methyltransferase. Nucleic acids research 39, 3771-3780.

Goll, M.G., Bestor, T.H., 2005. Eukaryotic cytosine methyltransferases. Annu. Rev. Biochem. 74, 481-514.

Gros, C., Chauvigné, L., Poulet, A., Menon, Y., Ausseil, F., Dufau, I., Arimondo, P.B., 2013. Development of a universal radioactive DNA methyltransferase inhibition test for high-throughput screening and mechanistic studies. Nucleic acids research 41, e185.

Gros, C., Fahy, J., Halby, L., Dufau, I., Erdmann, A., Gregoire, J.-M., Ausseil, F., Vispé, S., Arimondo, P.B., 2012. DNA methylation inhibitors in cancer: recent and future approaches. Biochimie 94, 2280-2296.

Halgren, T.A., 1996. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. Journal of computational chemistry 17, 490-519.

Halim, D., Caron, K., Keillor, J.W., 2007. Synthesis and evaluation of peptidic maleimides as transglutaminase inhibitors. Bioorganic & medicinal chemistry letters 17, 305-308.

Issa, J.-P.J., Kantarjian, H.M., 2009. Targeting DNA methylation. Clinical Cancer Research 15, 3938-3946.

Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., Cheng, X., 2007. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature 449, 248-251.

Jones, G., Willett, P., Glen, R.C., Leach, A.R., Taylor, R., 1997. Development and validation of a genetic algorithm for flexible docking. Journal of molecular biology 267, 727-748.

Jones, P.A., 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature Reviews Genetics 13, 484-492.

Jones, P.A., Taylor, S.M., 1980. Cellular differentiation, cytidine analogs and DNA methylation. Cell 20, 85-93.

Kabsch, W., 2010. Xds. Acta Crystallogr D Biol Crystallogr 66, 125-132.

Klimasauskas, S., Kumar, S., Roberts, R.J., Cheng, X., 1994. HhaI methyltransferase flips its target base out of the DNA helix. Cell 76, 357-369.

Korb, O., Stutzle, T., Exner, T.E., 2009. Empirical scoring functions for advanced protein– ligand docking with PLANTS. Journal of chemical information and modeling 49, 84-96.

Kuck, D., Caulfield, T., Lyko, F., Medina-Franco, J.L., 2010. Nanaomycin A selectively inhibits DNMT3B and reactivates silenced tumor suppressor genes in human cancer cells. Molecular cancer therapeutics 9, 3015-3023.

Kumar, S., Cheng, X., Pflugrath, J.W., Roberts, R.J., 1992. Purification, crystallization, and preliminary X-ray diffraction analysis of an M.HhaI-AdoMet complex. Biochemistry 31, 8648-8653.

Kurkcuoglu, Z., Bakan, A., Kocaman, D., Bahar, I., Doruker, P., 2012. Coupling between catalytic loop motions and enzyme global dynamics. PLoS Comput Biol 8, e1002705.

Lee, B.H., Yegnasubramanian, S., Lin, X., Nelson, W.G., 2005. Procainamide is a specific inhibitor of DNA methyltransferase 1. Journal of Biological Chemistry 280, 40749-40756.

Liu, Z., Xie, Z., Jones, W., Pavlovicz, R.E., Liu, S., Yu, J., Li, P.-k., Lin, J., Fuchs, J.R., Marcucci, G., 2009a. Curcumin is a potent DNA hypomethylation agent. Bioorganic & medicinal chemistry letters 19, 706-709.

Liu, Z., Liu, S., Xie, Z., Pavlovicz, R.E., Wu, J., Chen, P., Aimiuwu, J., Pang, J., Bhasin, D., Neviani, P., 2009b. Modulation of DNA methylation by a sesquiterpene lactone parthenolide. Journal of Pharmacology and Experimental Therapeutics 329, 505-514.

London, N., Miller, R.M., Krishnan, S., Uchida, K., Irwin, J.J., Eidam, O., Gibold, L., Cimermančič, P., Bonnet, R., Shoichet, B.K., 2014. Covalent docking of large libraries for the discovery of chemical probes. Nature chemical biology 10, 1066-1072.

López, S.N., Sortino, M., Escalante, A., de Campos, F., Correˆa, R., Cechinel Filho, V., Nunes, R.J., Zacchino, S.A., 2003. Antifungal Properties of Novel N-and α, β-Substituted Succinimides against Dermatophytes. Arzneimittelforschung 53, 280-288.

Lyko, F., Brown, R., 2005. DNA methyltransferase inhibitors and the development of epigenetic cancer therapies. Journal of the National Cancer Institute 97, 1498-1506.

Mahle, F., da Rosa Guimarães, T., Meira, A.V., Corrêa, R., Cruz, R.C.B., Cruz, A.B., Nunes, R.J., Cechinel-Filho, V., de Campos-Buzzi, F., 2010. Synthesis and biological evaluation of N-antipyrine-4-substituted amino-3-chloromaleimide derivatives. European journal of medicinal chemistry 45, 4761-4768.

Mai, A., Altucci, L., 2009. Epi-drugs to fight cancer: from chemistry to cancer treatment, the road ahead. The international journal of biochemistry & cell biology 41, 199-213.

Matuszak, N., Muccioli, G.G., Labar, G., Lambert, D.M., 2009. Synthesis and in vitro evaluation of N-substituted maleimide derivatives as selective monoglyceride lipase inhibitors. Journal of medicinal chemistry 52, 7410-7420.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. Journal of applied crystallography 40, 658-674.

Merfort, I., 2011. Perspectives on sesquiterpene lactones in inflammation and cancer. Current drug targets 12, 1560-1573.

Niesen, F.H., Berglund, H., Vedadi, M., 2007. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nature protocols 2, 2212-2221.

O'Gara, M., Klimasauskas, S., Roberts, R.J., Cheng, X., 1996. Enzymatic C5-cytosine methylation of DNA: mechanistic implications of new crystal structures for HhaL methyltransferase-DNA-AdoHcy complexes. Journal of molecular biology 261, 634-645.

O'Gara, M., Horton, J.R., Roberts, R.J., Cheng, X., 1998. Structures of HhaI methyltransferase complexed with substrates containing mismatches at the target base. Nature structural biology 5, 872-877.

O'Gara, M., Zhang, X., Roberts, R.J., Cheng, X., 1999. Structure of a binary complex of HhaI methyltransferase with S-adenosyl-L-methionine formed in the presence of a short non-specific DNA oligonucleotide. Journal of molecular biology 287, 201-209.

Rilova, E., Erdmann, A., Gros, C., Masson, V., Aussagues, Y., Poughon-Cassabois, V., Rajavelu, A., Jeltsch, A., Menon, Y., Novosad, N., 2014. Design, Synthesis and Biological Evaluation of 4-Amino-N-(4-aminophenyl) benzamide Analogs of Quinoline-Based SGI-1027 as Inhibitors of DNA Methylation. ChemMedChem 9, 590-601.

Rondelet, G., Dal Maso, T., Willems, L., Wouters, J., 2016. Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. Journal of Structural Biology 194, 357-367.

Sankpal, U.T., Rao, D.N., 2002. Structure, Function, and Mechanism of Hha I DNA Methyltransferases. Critical reviews in biochemistry and molecular biology 37, 167-197.

Sheikhnejad, G., Brank, A., Christman, J.K., Goddard, A., Alvarez, E., Ford, H., Jr., Marquez, V.E., Marasco, C.J., Sufrin, J.R., O'Gara, M., Cheng, X., 1999. Mechanism of inhibition of DNA (cytosine C5)-methyltransferases by oligodeoxyribonucleotides containing 5,6-dihydro-5-azacytosine. Journal of molecular biology 285, 2021-2034.

Siedlecki, P., Boy, R.G., Musch, T., Brueckner, B., Suhai, S., Lyko, F., Zielenkiewicz, P., 2006. Discovery of two novel, small-molecule inhibitors of DNA methylation. Journal of medicinal chemistry 49, 678-683.

Sivaprakasam, P., Xie, A., Doerksen, R.J., 2006. Probing the physicochemical and structural requirements for glycogen synthase kinase-3α inhibition: 2D-QSAR for 3-anilino-4-phenylmaleimides. Bioorganic & medicinal chemistry 14, 8210-8218.

Song, J., Teplova, M., Ishibe-Murakami, S., Patel, D.J., 2012. Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science (New York, N.Y 335, 709-712.

Stresemann, C., Lyko, F., 2008. Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. International journal of cancer 123, 8-13.

Suzuki, T., Tanaka, R., Hamada, S., Nakagawa, H., Miyata, N., 2010. Design, synthesis, inhibitory activity, and binding mode study of novel DNA methyltransferase 1 inhibitors. Bioorganic & medicinal chemistry letters 20, 1124-1127.

Valente, S., Liu, Y., Schnekenburger, M., Zwergel, C., Cosconati, S., Gros, C., Tardugno, M., Labella, D., Florean, C., Minden, S., 2014. Selective non-nucleoside inhibitors of human DNA methyltransferases active in cancer including in cancer stem cells. Journal of medicinal chemistry 57, 701-713.

Vandermeers, F., Sriramareddy, S.N., Costa, C., Hubaux, R., Cosse, J.-P., Willems, L., 2013. The role of epigenetics in malignant pleural mesothelioma. Lung cancer 81, 311-318.

Weichenberger, C.X., Sippl, M.J., 2006. NQ-Flipper: validation and correction of asparagine/glutamine amide rotamers in protein crystal structures. Bioinformatics 22, 1397-1398.

Wijermans, P., Rüter, B., Baer, M., Slack, J.L., Saba, H., Lübbert, M., 2008. Efficacy of decitabine in the treatment of patients with chronic myelomonocytic leukemia (CMML). Leukemia research 32, 587-591.

Wutz, A., 2011. Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. Nature Reviews Genetics 12, 542-553.

Yoo, J., Medina-Franco, J.L., 2011. Homology modeling, docking and structure-based pharmacophore of inhibitors of DNA methyltransferase. Journal of computer-aided molecular design 25, 555-567.

Zhou, L., Cheng, X., Connolly, B.A., Dickman, M.J., Hurd, P.J., Hornby, D.P., 2002. Zebularine: a novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. Journal of molecular biology 321, 591-599.

Zhu, B., Ge, J., Yao, S.Q., 2015. Developing new chemical tools for DNA methyltransferase 1 (DNMT 1): A small-molecule activity-based probe and novel tetrazole-containing inhibitors. Bioorganic & medicinal chemistry 23, 2917-2927.

Zhu, K., Borrelli, K.W., Greenwood, J.R., Day, T., Abel, R., Farid, R.S., Harder, E., 2014. Docking covalent inhibitors: a parameter free approach to pose prediction and scoring. Journal of chemical information and modeling 54, 1932-1940.

2.6 Supplementary Data
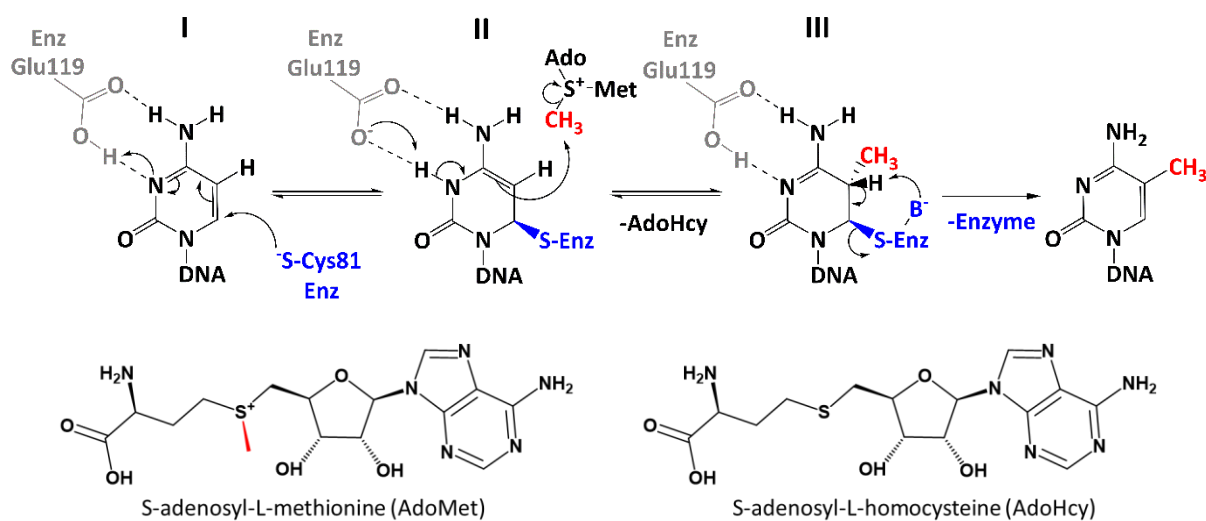
Supplementary Figure S1: Proposed mechanism of cytosine methylation catalyzed by DNMTs based on M.HhaI studies (Bestor and Verdine, 1994; Goll and Bestor, 2005). *The catalytic mechanism involves first the binding of DNMT to the target DNA. Then, the cytosine flips out of the double helix (base-flipping) to be localized in the active site. Glu119 acts as a hydrogen donor to the N3 atom of the cytosine allowing the nucleophilic addition of the unprotonated Cys81 which is situated on the catalytic loop (**I**). This leads to the formation of a covalent intermediate. The 5-position of the cytosine is activated and then accepts the methyl group from SAM (or AdoMet) to form the 5-methyl covalent adduct and SAH (or AdoHcy) (**II**). Subsequently, the covalent complex between the DNA methyltransferase and the methylated cytosine is released by β-elimination (**III**).*
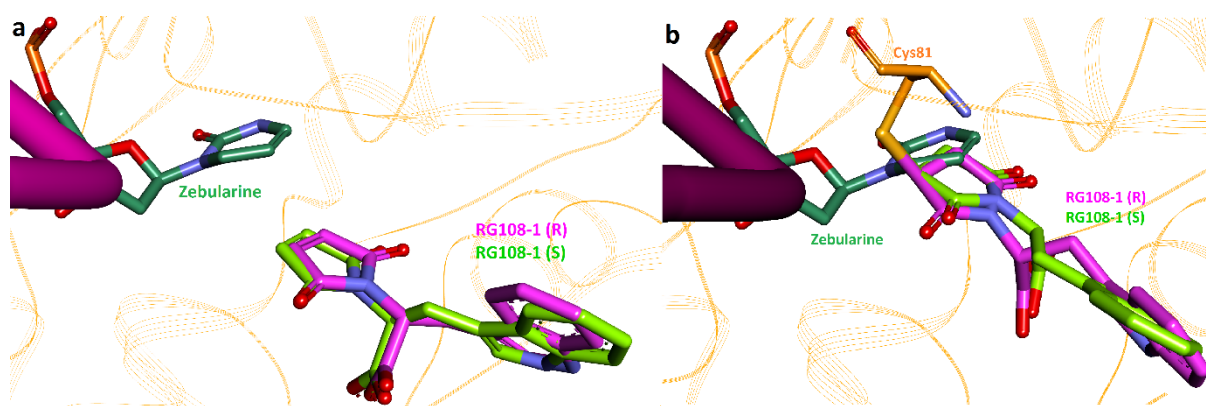
Supplementary Figure S2: Solutions of the docking simulations of RG108-1 with M.HhaI. *Zebularine, removed prior to docking simulations, was added for final visualization. (**a**) Superimposition of non-covalent docking poses of RG108-1 (S) and RG108-1 (R) (green, pink sticks, respectively). (**b**) Superimposition of covalent docking poses of RG108-1 (S) and RG108-1 (R) (green, pink sticks, respectively). The carboxylate anions of both enantiomers overlap well and binding poses of the two enantiomers are the same. They formed same interactions with residues in the binding site for non-covalent docking simulation.*
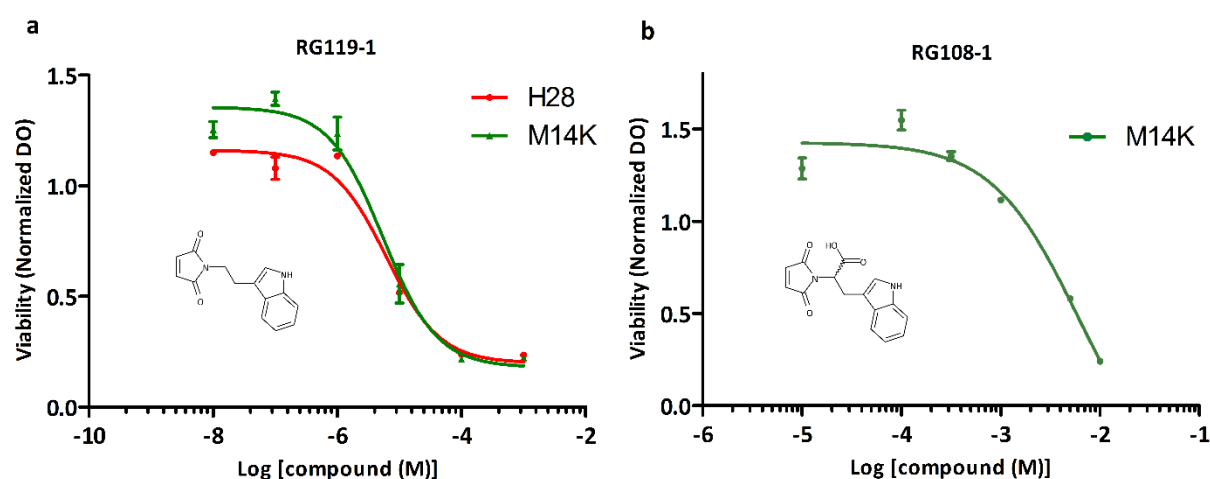
Supplementary Figure S3: Determination of the half maximal toxic concentration (TC50) of RG119-1 and RG108-1. (a) *Evaluation of the TC50 of RG119-1. H28 and M14K cells were incubated 72 h with the indicated concentrations of RG119-1 (n=6). Data are normalized to the control condition. RG119-1 has a TC50 of 5.4 µM and 6.2 µM against M14K and H28, respectively. (**b**) Evaluation of the TC50 of RG108-1. H28 and M14K cells were incubated 72 h with the indicated concentrations of RG108-1 (n=6). Data are normalized to the control condition. RG108-1 has no marked inhibitory effect on H28 and M14K cells.*
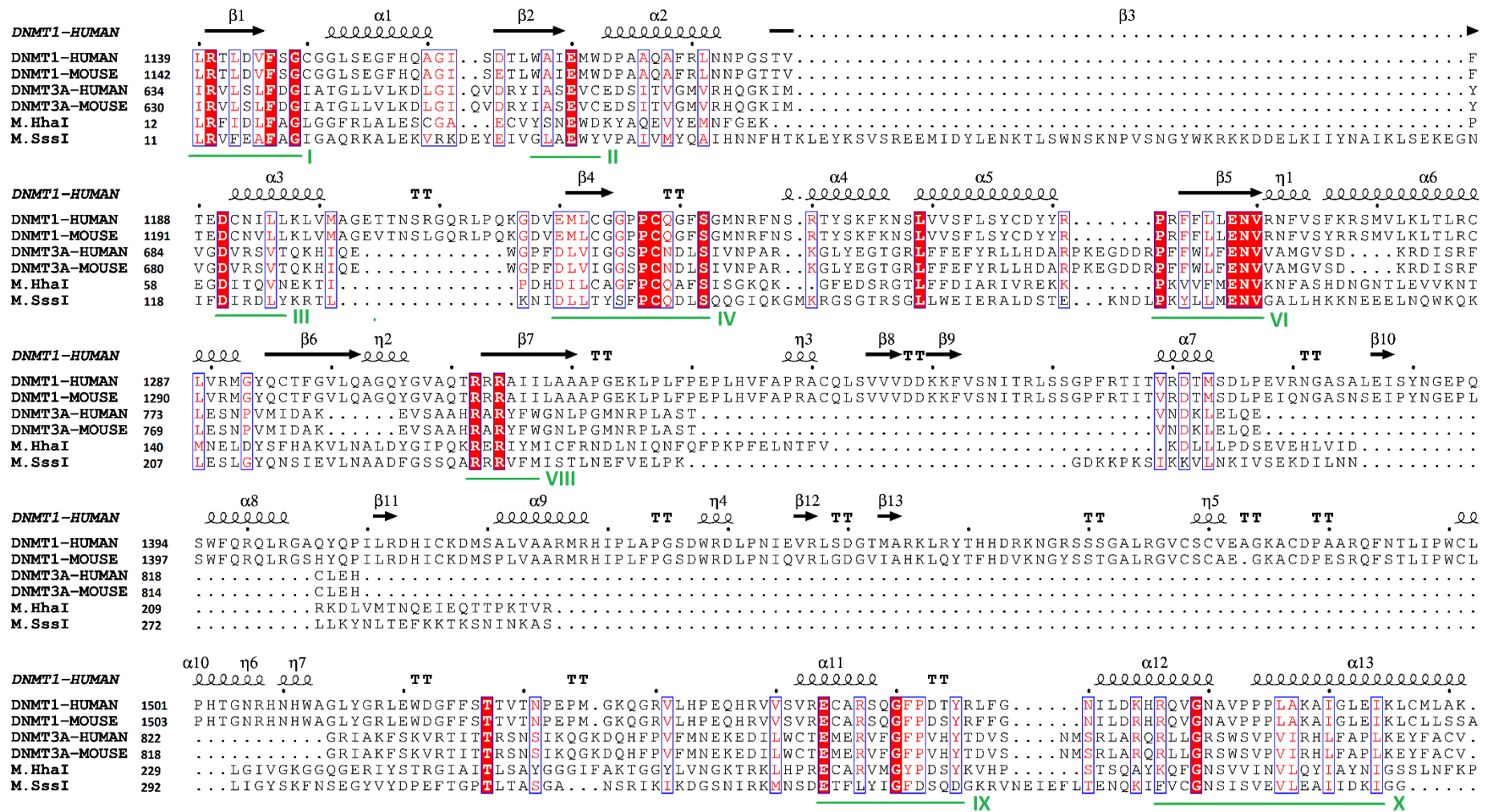
Supplementary Figure S4: Structure-based sequence alignment of DNMTs. *Alignment of catalytic domain of human DNMT1 (uniprotkb:P26358), mouse DNMT1 (uniprotkb:P13864), human DNMT3A (uniprotkb:Q9Y6K1), mouse DNMT3A (uniprotkb:O88508), bacterial M.HhaI (uniprotkb:P05102) and bacterial M.SssI (uniprotkb:P15840). Identical residues are indicated with a red background while similar residues are boxed in blue. Conserved motifs are underlined in green and numbered according to (Gowher et al., 2006). This multiple structure and sequence alignments was performed with PROMALS3D (Pei et al., 2008) and formatted with Espript (Gouet et al., 1999).*
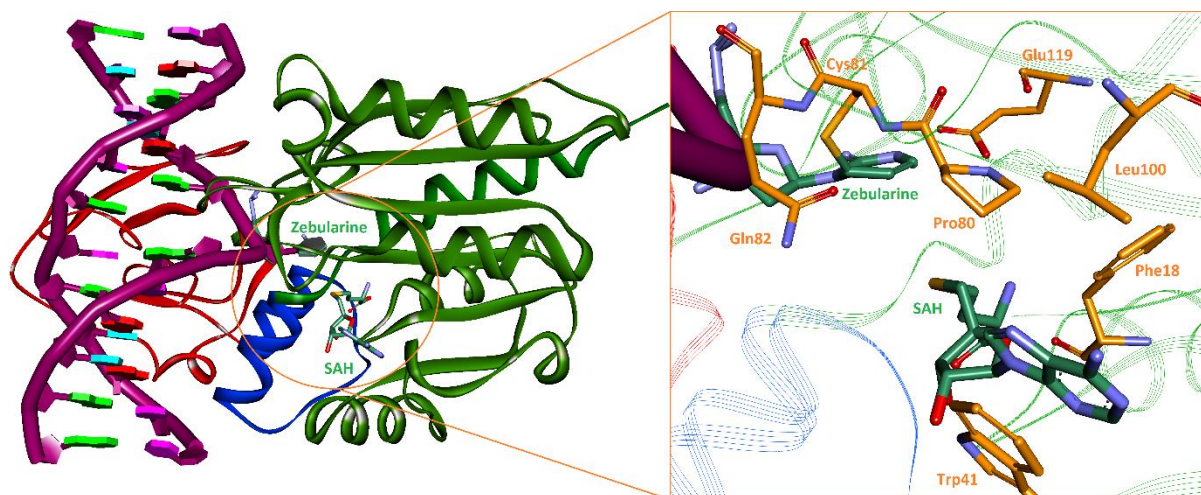
Supplementary Figure S5: Conformation of M.HhaI, and by extension DNMT1, used for non-covalent and covalent docking simulations. *Zebularine and cofactor are represented as sticks models in green. M.HhaI is displayed in solid ribbon: green for the large catalytic domain, red for the small regulation and blue for the hinge region connecting the two domains. Three-dimensional structure of catalytic domain of DNMTs is schematically divided into three parts: a pocket to bind the adenosyl group of SAM, another to bind the amino acid of SAM, and a pocket to bind the flipped out cytosine of the DNA double helix (Rilova et al., 2014). In the ternary complex, the catalytic loop is in a conformation allowing the cysteine to carry out the nucleophilic attack on cytosine. SAH and Zebularine (5-fluorocytosine for DNMT1) were extracted prior to docking simulations, and DNA double helical structure was kept during docking simulations. For docking simulation, we chose the eukaryotic mouse DNMT1 in the active form bound with hemimethylated CpG duplex (PDB code: 4DA4) (Song et al., 2012). Indeed, the published crystal structure of human DNMT1 (PDB code: 3PTA) (Song et al., 2011) bound with an unmethylated CpG duplex shows an autoinhibitory mechanism. In this conformation, the catalytic loop does not cover the active site, putting the catalytic cysteine far away from its active conformation (8 Å). This crystal structure does not possess the active conformation for the catalytic mechanism of DNA methylation. Moreover, the catalytic domain of mouse and human DNMT1 share a high sequence identity (92%, BLAST). For bacterial M.SssI, since no structural information is available in the PDB, we conducted docking studies on M.HhaI since the catalytic pocket is well conserved among the bacterial DNMTs.*
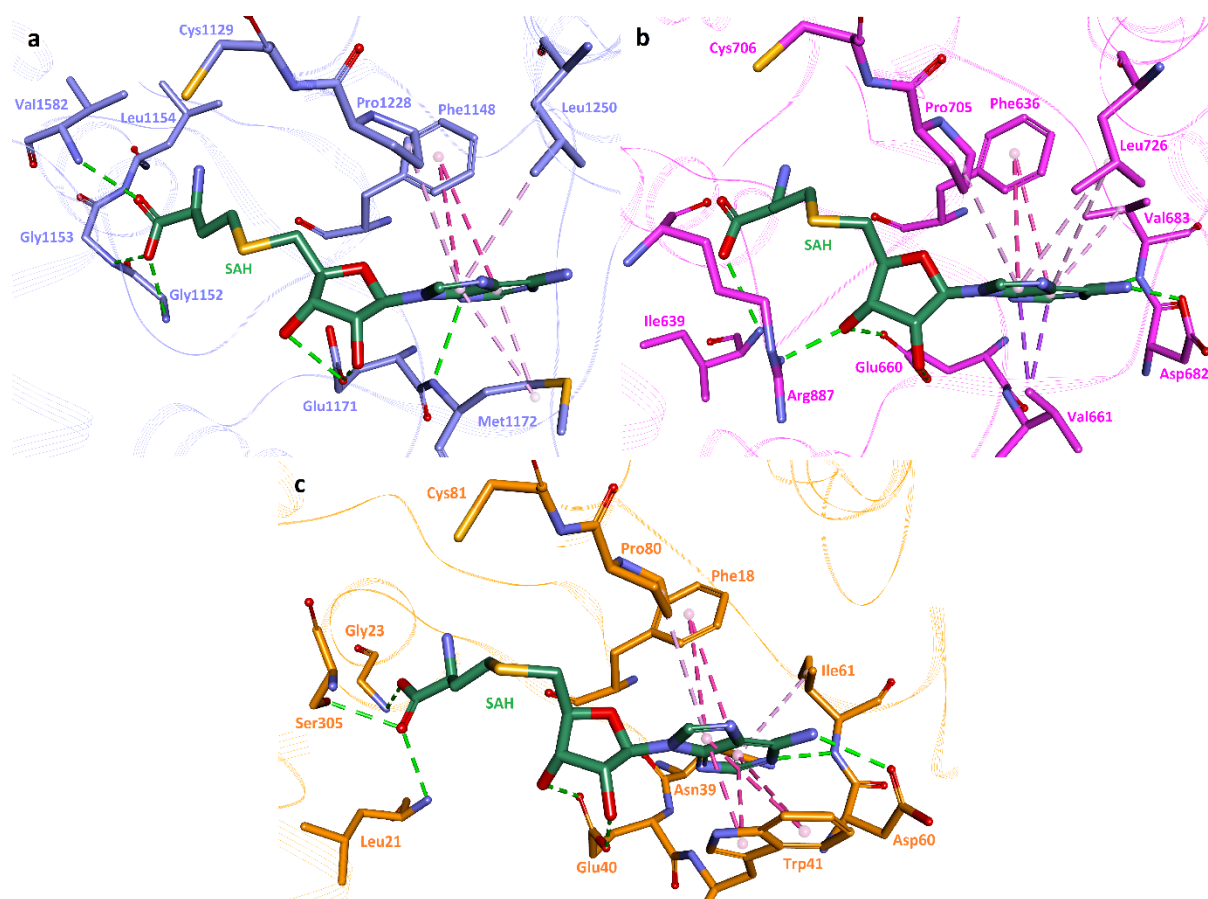
Supplementary Figure S6: Cofactor binding site of mouse DNMT1 (a; blue; PDB code: 4DA4) (Song et al., 2012), human DNMT3A (b; pink; PDB code: 2QRV) (O'Gara et al., 1999) and M.HhaI (c; orange; PDB code: 1M0E) (Zhou et al., 2002) in complex with SAH cofactor (green). *Key interactions between SAH cofactor and DNMTs are mediated through conserved residues. Adenine ring of the cofactor is stabilized in a conserved hydrophobic pocket. Trp41 in M.HhaI is replaced by Met1172 in DNMT1 and Val661 in DNMT3A. Hydroxyl groups of ribose form hydrogen bonds with a conserved glutamate. Different residues are engaged in interaction with the carboxylate group of SAH methionine.*
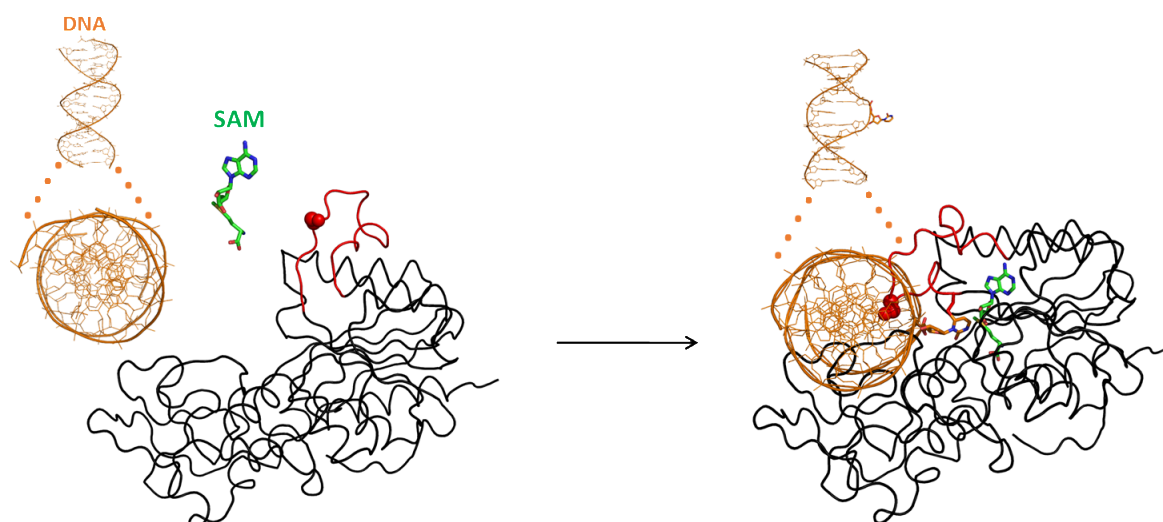
Supplementary Figure S7: Mechanism of base flipping and catalytic loop closure adapted from Gerasimaitė et al. (2011). *During the formation of the ternary complex (M.HhaI-SAM-DNA), the cytosine flips out of the DNA duplex and the catalytic loop (residues 81-100) undergoes a conformational change (2.5 nm) to close the catalytic site (Sankpal and Rao, 2002). This active conformation permits the nucleophilic attack of Cys81 (red balls) at the C6 position of cytosine.*
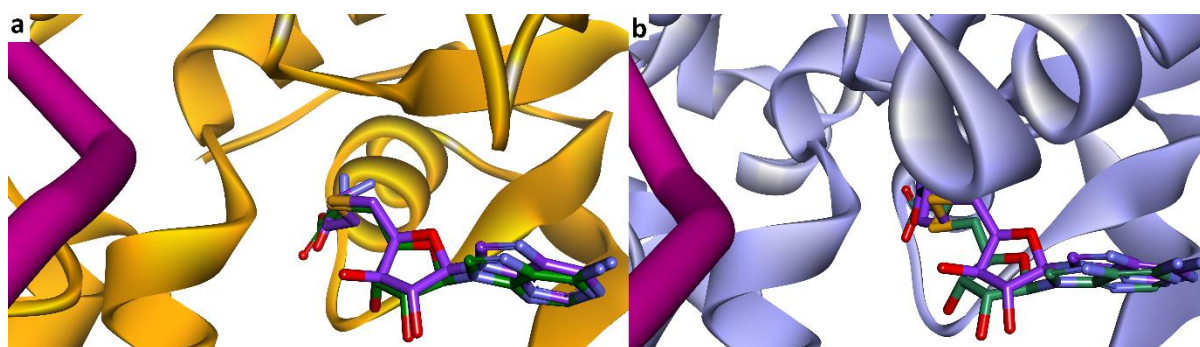
Supplementary Figure S8*: Docking validation of the new proposed protocol. (**a**)* *Superimposition of the crystal structure of M.HhaI (orange, solid ribbon) (PDB code: 1M0E) (O'Gara et al., 1999) in complex with the cofactor SAH (green) and the top docking solution of SAH in M.HhaI model (purple) (RMSD$_{SAM\ crystal\text{-}SAM\ model}$ of 0.89 Å). (**B**) Superimposition of the crystal structure of DNMT1 (blue, solid ribbon) (PDB code: 4DA4) (Song et al., 2012) in complex with the cofactor SAH (green) and the top docking solution of SAH in DNMT1 model (purple) (RMSD$_{SAM\ crystal\text{-}SAM\ model}$ of 1.34 Å).*
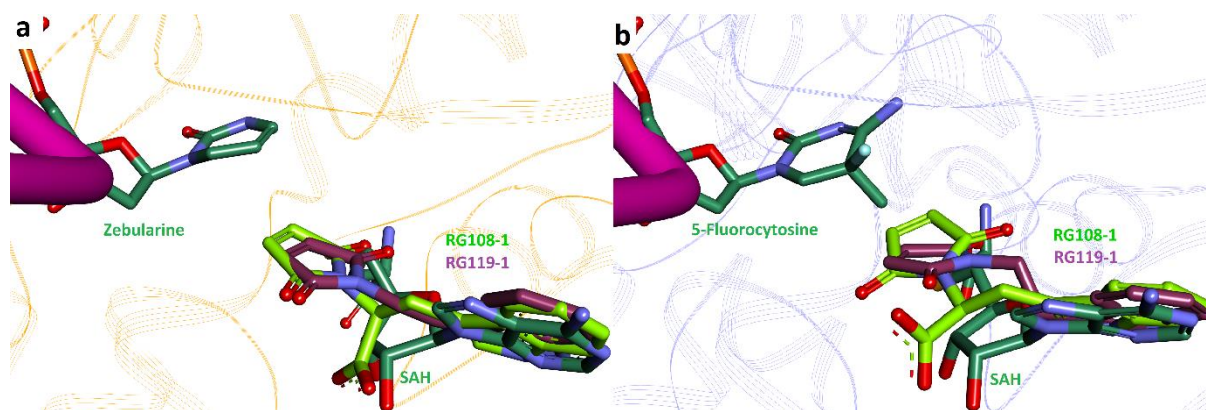
Supplementary Figure S9: Solutions of the non-covalent docking simulations. *Zebularine and 5-Fluorocytosine, removed prior to docking simulations, were added for final visualization. (**a**) Superimposition of non-covalent docking poses of RG108-1 and RG119-1 (green, purple sticks, respectively) with cofactor SAH inside M.HhaI. (**b**) Superimposition of non-covalent docking poses of RG108-1 and RG119-1 (green, purple sticks, respectively) with cofactor SAH inside DNMT1. Indole ring of RG108-1 and RG119-1 superimposed well with adenine ring of SAH interacting with the same aromatic interactions. Maleimide ring is directed to the cytosine substrate pocket and is stabilized by the same interactions for M.HhaI and DNMT1.*

2.7 Supplemental References

Bestor, T.H., Verdine, G.L., 1994. DNA methyltransferases. Current opinion in cell biology 6, 380-389.

Gerasimaitė, R., Merkienė, E., Klimašauskas, S., 2011. Direct observation of cytosine flipping and covalent catalysis in a DNA methyltransferase. Nucleic acids research 39, 3771-3780.

Goll, M.G., Bestor, T.H., 2005. Eukaryotic cytosine methyltransferases. Annu. Rev. Biochem. 74, 481-514.

Gouet, P., Courcelle, E., Stuart, D.I., 1999. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics 15, 305-308.

Gowher, H., Loutchanwoot, P., Vorobjeva, O., Handa, V., Jurkowska, R.Z., Jurkowski, T.P., Jeltsch, A., 2006. Mutational analysis of the catalytic domain of the murine Dnmt3a DNA-(cytosine C5)-methyltransferase. Journal of molecular biology 357, 928-941.

O'Gara, M., Zhang, X., Roberts, R.J., Cheng, X., 1999. Structure of a binary complex of HhaI methyltransferase with S-adenosyl-L-methionine formed in the presence of a short non-specific DNA oligonucleotide. Journal of molecular biology 287, 201-209.

Pei, J., Kim, B.-H., Grishin, N.V., 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic acids research 36, 2295-2300.

Rilova, E., Erdmann, A., Gros, C., Masson, V., Aussagues, Y., Poughon-Cassabois, V., Rajavelu, A., Jeltsch, A., Menon, Y., Novosad, N., 2014. Design, Synthesis and Biological Evaluation of 4-Amino-N-(4-aminophenyl) benzamide Analogs of Quinoline-Based SGI-1027 as Inhibitors of DNA Methylation. ChemMedChem 9, 590-601.

Sankpal, U.T., Rao, D.N., 2002. Structure, Function, and Mechanism of Hha I DNA Methyltransferases. Critical reviews in biochemistry and molecular biology 37, 167-197.

Song, J., Rechkoblit, O., Bestor, T.H., Patel, D.J., 2011. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science (New York, N.Y.) 331, 1036-1040.

Song, J., Teplova, M., Ishibe-Murakami, S., Patel, D.J., 2012. Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science (New York, N.Y.) 335, 709-712.

Zhou, L., Cheng, X., Connolly, B.A., Dickman, M.J., Hurd, P.J., Hornby, D.P., 2002. Zebularine: a novel DNA methylation inhibitor that forms a covalent complex with DNA methyltransferases. Journal of molecular biology 321, 591-599.

CHAPTER 3

Chapter 3. Structural basis for recognition of histone H3K36me3 nucleosome by human *de novo* DNA methyltransferases 3A and 3B

Grégoire Rondelet[a], Thomas Dal Maso[a], Luc Willems[b], and Johan Wouters[a]

[a]Department of Chemistry, University of Namur, 61 rue de Bruxelles, B-5000 Namur, Belgium

[b]Molecular and Cellular Epigenetics (GIGA) and Molecular Biology (Gembloux Agro-Bio Tech), University of Liège (ULg), 4000 Liège, Belgium

*Corresponding author: gregoire.rondelet@unamur.be

Personal contribution: Participated in research design, conducted experiments (protein production, protein crystallization, structure determination and docking simulations), performed data analysis and literature search, generated figures and tables, and wrote the manuscript.

Abstract

DNA methylation is an important epigenetic modification involved in chromatin organization and gene expression. The function of DNA methylation depends on cell context and is correlated with histone modification patterns. In particular, trimethylation of Lys36 on histone H3 tail (H3K36me3) is associated with DNA methylation and elongation phase of transcription. PWWP domains of the *de novo* DNA methyltransferases DNMT3A and DNMT3B read this epigenetic mark to guide DNA methylation. Here we report the first crystal structure of the DNMT3B PWWP domain-H3K36me3 complex. Based on this structure, we propose a model of the DNMT3A PWWP domain-H3K36me3 complex and build a model of DNMT3A (PWWP-ADD-CD domains) in a nucleosomal context. The trimethylated side chain of Lys36 (H3K36me3) is inserted into an aromatic cage similar to the "Royal" superfamily domains known to bind methylated histones. A key interaction between trimethylated Lys36 and a conserved water molecule stabilized by Ser270 explains the lack of affinity of mutated DNMT3B (S270P) for the H3K36me3 epigenetic mark in the ICF (Immunodeficiency, Centromeric instability and Facial abnormalities) syndrome. The model of the DNMT3A-DNMT3L heterotetramer in complex with a dinucleosome highlights the mechanism for

recognition of nucleosome by DNMT3s and explains the periodicity of *de novo* DNA methylation.

Keywords: DNA methylation, Methyltransferases, M.*Hha*I, DNMT3A, DNMT3B, PWWP, H3K36me3, Structure, Nucleosome

3.1 Introduction

DNA methylation occurs at CpG dinucleotides in mammalian cells and is catalyzed by DNA methyltransferases (DNMTs). DNMT3A and DNMT3B are both involved in *de novo* methylation establishing the methylation pattern in genome during embryogenesis while DNMT1 maintains the pattern during chromosome replication. DNA methylation is essential for cell differentiation and development but is also involved in pathologies like cancer (Bird, 2002). The function of DNA methylation depends on cell context and is correlated with histone modification patterns (Jones, 2012). In particular, trimethylation of Lys36 on histone H3 tail (H3K36me3) is associated with gene body methylation in embryonic stem (ES) cells and elongation phase of transcription (Baubec et al., 2015; Hahn et al., 2011; Lee and Shilatifard, 2007). Both DNMT3A and DNMT3B PWWP domains read H3 tails containing H3K36me3 to guide DNA methylation (Baubec et al., 2015; Dhayalan et al., 2010). A point mutation in the DNMT3B PWWP domain (S270P) leads to loss of recognition with H3K36me3-modified nucleosome (Baubec et al., 2015; Ge et al., 2004) and a decrease in DNA methylation at pericentromeric satellite repeat II as observed for ICF syndrome (Chen et al., 2004; Shirohzu et al., 2002).

DNMT3A and DNMT3B possess a C-terminal catalytic domain (CD) and an N-terminal part with a regulatory function mediated by the ADD (ATRX–DNMT3–DNMT3L) domain and a nucleosome recognition PWWP fold. Similar to the chromodomain, MBT and Tudor domains, the PWWP domain is a member of the "Royal" superfamily domains which recognize methylated histone tails through a conserved aromatic cage (Qin and Min, 2014). The H3K36me3-binding ability was established for the PWWP domains of BRPF1 (Vezzoli et al., 2010), DNMT3A (Dhayalan et al., 2010), DNMT3B (Baubec et al., 2015), PSIP1, MSH6, ZMYND11 + H3.3K36me3 (Qin and Min, 2014), LEDGF (Pradeepa et al., 2012), and Tudor domains of PHF1 and PHF19 (Ballaré et al., 2012; Musselman et al., 2012). The PWWP domain contains an anti-parallel β-barrel-like fold formed by five β-strands (β1–β5) (Fig. 1a), where a short 3_{10} helix is found between β4 and β5 (η2), an insertion motif of different lengths between β2 and β3 (η1) and a C-terminal helix bundle of 1-6 α-helices (Qiu et al., 2002). PWWP domains of DNMT3A and DNMT3B are characterized by a short motif insertion (η1) and five α-helices following the β-barrel. The conserved Pro-Trp-Trp-Pro (PWWP) motif becomes SWWP and is found in the β2 strand. DNMT3A and DNMT3B PWWP domains would

synergistically bind both histone and DNA through their conserved aromatic cage for the recognition of H3K36me3 epigenetic mark, and a positively charged surface that interacts with DNA (Qin and Min, 2014; Qiu et al., 2002). Recently, DNMT3B was shown to be involved in the selective targeting of transcribed genes in mouse stem cells (Baubec et al., 2015). This association occurs through the binding of DNMT3B PWWP domain to trimethylated lysine 36 on histone H3.

Here we report the first crystal structure of the DNMT3B PWWP domain in complex with histone H3K36me3. Based on this structure, we propose a model for the DNMT3A PWWP domain-H3K36me3 complex and predict a structure of DNMT3A in a nucleosomal context to propose a mechanism of nucleosome recognition and *de novo* DNA methylation.

3.2 Results and discussion

3.2.1 Structure of the DNMT3B PWWP domain in complex with histone H3K36me3

We determined the crystal structure of the DNMT3B PWWP domain in complex with the epigenetic mark H3K36me3 (H3$_{32–38}$K36me3) to gain structural insight into this recognition (Fig. 1a-d).

Compared to the open form (Qiu et al., 2002), the loop between the β1 and β2 strand undergoes an induced fit that closes the aromatic cage in order to enhance recognition of this epigenetic mark (see Fig. 6 in chapter 4). The histone peptide occupies a surface groove formed by the β1 strand, the loop between β1 and β2, and the β4 strand (Fig. 1a and b). The trimethylated side chain of Lys36 is well resolved in the density map and is inserted into the aromatic cage formed by the three aromatic residues Phe236, Trp239 (β2 strand) and Trp263 (β3 strand) (Fig. 1c). The trimethyl-ammonium group forms interaction with this aromatic cage by van der Waals and π-cation interactions with aromatic side chains of Phe236, Trp239 and Trp263, and electrostatic–cation interaction with the carboxylate group of Asp266 (Fig. 1d). These interactions are common to different chromatin-binding proteins in complex with the H3K36me3 peptide (Qin and Min, 2014). Another important common feature of "Royal" superfamily domains for binding preferentially H3K36me3 involves hydrophobic contacts surrounding the trimethylated lysine 36.

In our complex, Pro38 is involved in a hydrophobic contact with the side chain of Ile233 as observed for the PHF1 Tudor domain (Cai et al., 2013; Musselman et al., 2012) and Val35 for the PWWP domain of BRPF1 (Vezzoli et al., 2010; Wu et al., 2011). For complex between epigenetic marks H3K4me3, H3K9me3, and H3K27me3, and other chromatin-binding proteins, we observe polar contacts in this surface area (Cheng et al., 2013; Pena et al., 2006; Sanulli et al., 2015). Besides that, an additional key interaction involved the CO main chain group of trimethylated Lys36 with a conserved water molecule stabilized by the oxygen atom of the lateral chain of Ser270. The loss of this last strong hydrogen bond (D_{A-B} = 2.41 Å in chain C and 2.59 Å in chain D) in the ICF syndrome (S270P) explains the lack of methyltransferase recognition and activity on nucleosomal substrates (Baubec et al., 2015; Dhayalan et al., 2010; Shirohzu et al., 2002). The rest of the modified peptide is involved in intermolecular hydrogen-bonds with backbone amino and carbonyl groups: CO of Thr32 and conserved waters stabilized by CO of Asp266, CO of Gly34 and NH of Phe269, and NH of trimethylated Lys36 and CO of Phe269. The previously discussed intermolecular interactions between DNMT3B PWWP domain and H3K36me3 peptide are presented in Table S2. Some of these observed interactions can be related to previous mutational studies. Mutations among conserved residues of DNMT3A and DNMT3B PWWP domains were performed in the β-barrel (for DNMT3B: $W_{239}P_{240}$-ST; D_{266}-A) and led to the loss of ability of both DNMT3A and DNMT3B to bind to H3K36me3 modified nucleosomes (Baubec et al., 2015; Chen et al., 2004; Dhayalan et al., 2010; Ge et al., 2004). As observed for other H3K36me3 binding domains, van der Waals and electrostatic interactions contribute to a specific association of DNMT3B PWWP domain with H3K36me3 peptide.
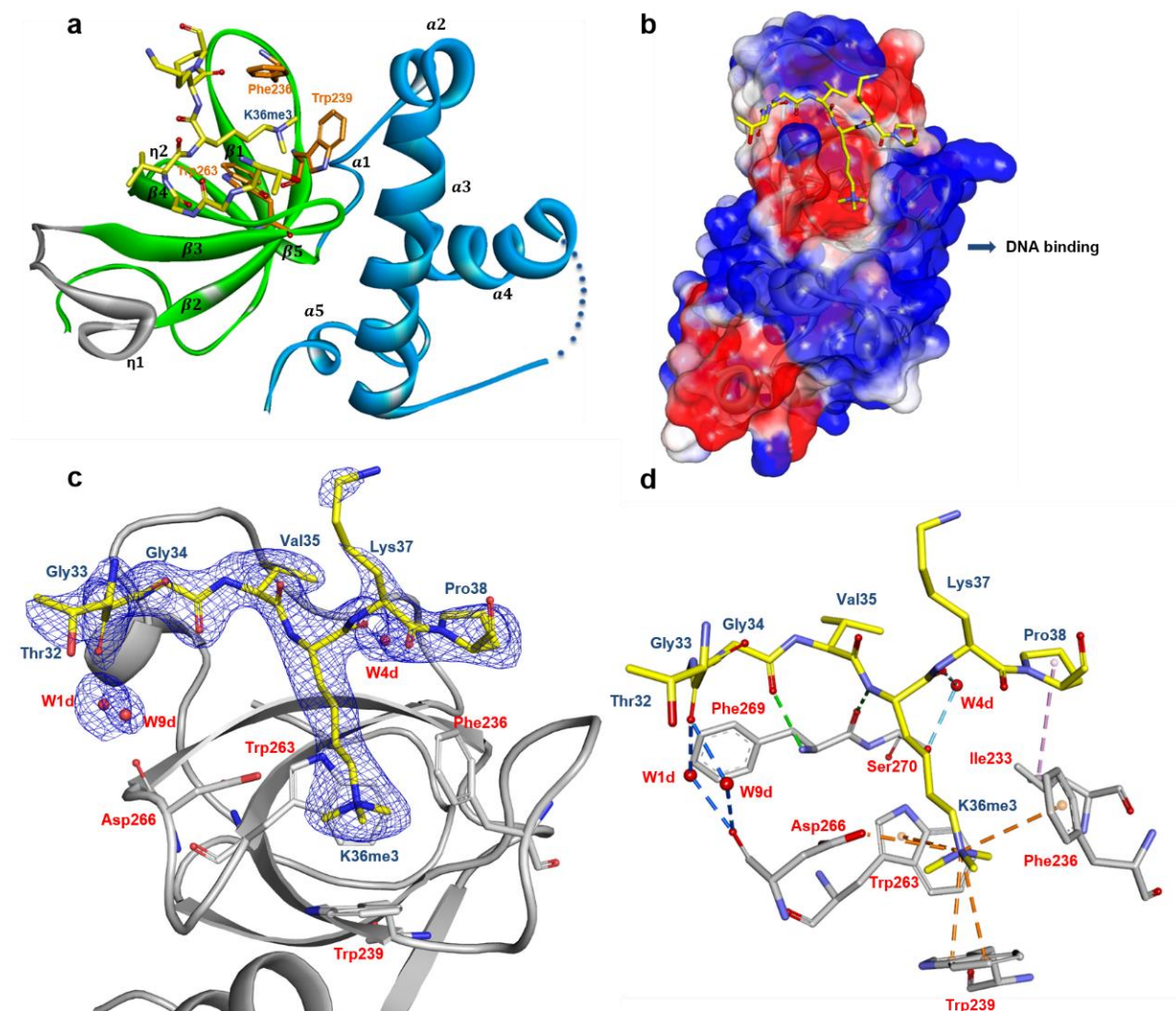
Figure 1. Structure of the DNMT3B PWWP domain in complex with H3K36me3 peptide. (a) *Solid ribbon representation of the crystal structure of the complex between the DNMT3B PWWP domain and the H3K36me3 peptide. The anti-parallel β-barrel-like is colored in green, the insertion motif η1 in grey and the C-terminal helix bundle in blue. (b) DNMT3B PWWP domain surface colored by electrostatic potential of the residues (red: negatively charged area and blue: positively charged area) with the peptide represented in stick. (c) 2Fo-Fc electron density map contoured at 1.0 σ for the H3K36me3 peptide (blue mesh) with the stick representation of the residues of the aromatic cage: Phe236, Trp263, Asp266 and Trp239. (d) Selected intermolecular interactions of the peptide with the DNMT3B PWWP domain (Table S2).*

One additional observation is that the trimethylated lysine can be engaged differently as observed for the PHF1 and PHF19 Tudor domains (Fig. 2a). We observed a flip of 180 °C as these domains lack the α-helix bundle and interact with another surface of the nucleosome-core. H3K36me3 peptide adopts the same binding mode with both DNMT3B PWWP and

BRPF1 (Fig. 2b) (Vezzoli et al., 2010). However, the β-β-α insertion motif of the BRPF1 interacts with peptide residues preceding the trimethylated Lys36 compared to the DNMT3B PWWP exhibiting a short α-helix motif (η1 insertion motif).
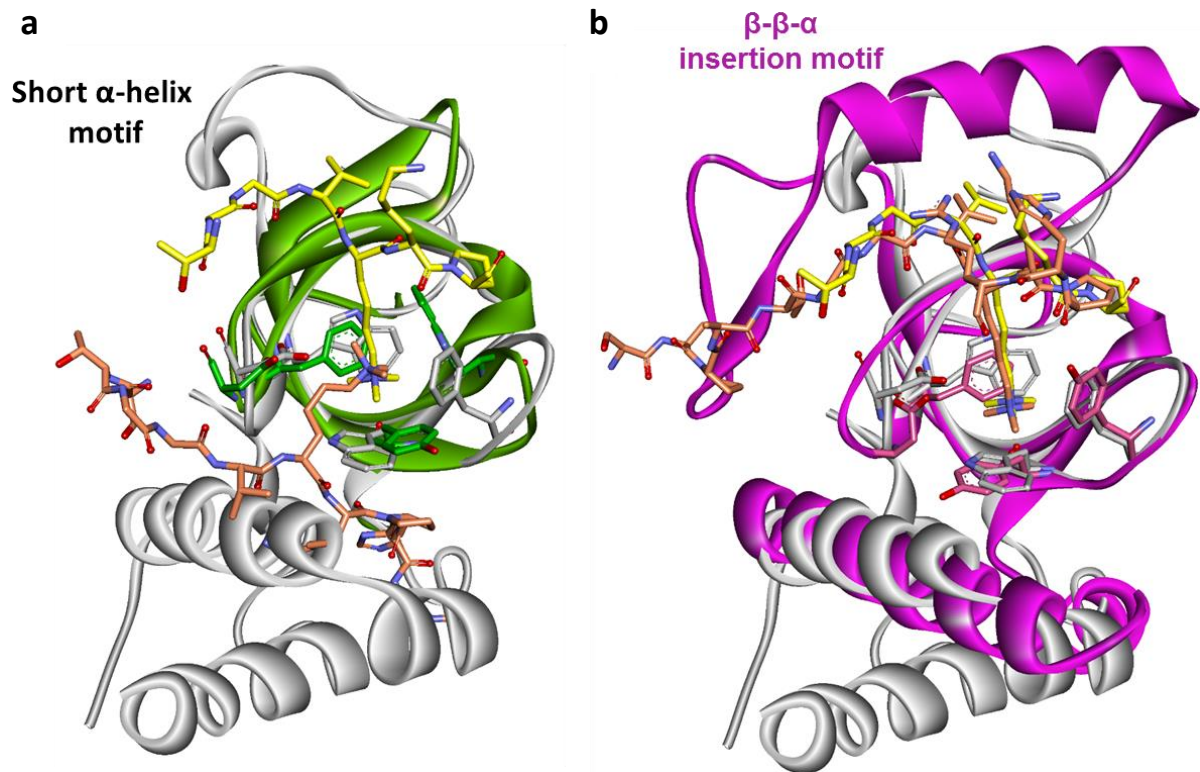
Figure 2. Different orientations and interactions of H3K36me3 peptide observed among the "Royal" superfamily. (a) Superposition of the DNMT3B PWWP domain (gray) – H3K36me3 (yellow) complex with the PHF1 (green) – H3K36me3 (orange) complex (Musselman et al., 2012) (PDB code: 4HCZ). The absence of a helix bundle for the PHF1 allows the H3K36me3 to be engaged differently than with DNMT3B PWWP. (b) Superposition of the DNMT3B PWWP domain (gray) – H3K36me3 (yellow) complex with the BRPF1 (purple) – H3K36me3 (orange) complex (Vezzoli et al., 2010) (PDB code: 2X4W).

Experiments to confirm binding of the H3K36me3 peptide to the DNMT3B PWWP domain in vitro were attempted, so far without success. In particular, due to unanticipated technical problems inherent to the system under study (precipitation and non-specific binding), we were not able to confirm the interaction using different biophysical approaches (ITC, DSF, and Bio-layer interferometry).

Table 1. Data Collection and Refinement Statistics.

	DNMT3B PWWP-H3K36me3	M.*Hha*I-DNA-SAH
Data collection		
Space group	P 32 2 1 (No. 154)	H 3 2 (No. 155)
Cell dimensions		
a, b, c (Å)	73.44, 73.44, 158.20	95.33, 95.33, 314.69
α, β, γ (°)	90, 90, 120	90, 90, 120
Resolution (Å)	50.0 - 2.24 (2.32 - 2.24) *	47.7-1.59 (1.65 - 1.59) *
R_{sym} or R_{merge} (%)	6.2 (63.3)	4.5 (69.4)
$I / \sigma I$	18.50 (3.19)	24.96 (1.71)
Completeness (%)	99.74 (99.96)	98.50 (90.27)
Redundancy	6.9 (7.0)	6.6 (3.2)
Refinement		
Resolution (Å)	36.72 - 2.24	36.55 - 1.59
No. reflections	24436	72699
R_{work} / R_{free} (%)	22.4/25.9	16.5/18.8
No. atoms		
Protein	1997	2606
Peptide/DNA	100 (peptide)	484 (DNA)
Ligand/ion	6 (citrate)	26 (SAH); 40 (sulfate)
Water	163	634
B-factors		
Protein	45.90	16.81
Peptide/DNA	52.60 (peptide)	23.20 (DNA)
Ligand/ion	55.14 (citrate)	27.74 (SO_4)
Water	49.30	30.70
R.m.s. deviations		
Bond lengths (Å)	0.009	0.006
Bond angles (°)	1.20	1.16

*Values in parentheses are for highest-resolution shell.

3.2.2 Modeling of the DNMT3A PWWP domain in complex with histone H3K36me3

As we were not able to crystallize the complex between the DNMT3A PWWP domain and H3K36me3 peptide, we performed a protein-peptide docking based on our structural study. DNMT3A and DNMT3B PWWP domains are small domains of ~150 residues very similar in structural organization and amino acid sequence (Fig. S1). They share 53% of identity in sequence and 67% of similarity using blastp.

Based on our crystal structure (DNMT3B PWWP-H3K36me3) and the high structural similarity between the two domains, we decided to dock this short peptide ($H3_{32-38}K36me3$) inside the DNMT3A PWWP domain (Wu et al., 2011) (PDB code: 3LLR). We used for this the Rosetta FlexPepDock protocol introduced for the refinement of coarse starting structure of peptide-protein complex into high-resolution models (Raveh et al., 2010). The proposed model provides a binding mode of the H3K36me3 peptide similar to the one observed in the crystallographic complex with the DNMT3B PWWP domain (Fig. 3a and b). Residues interacting with the H3K36me3 peptide are conserved among DNMT3A and DNMT3B PWWP domains (Fig. 4 and Tables S1 and S3). The trimethyl-ammonium group is stabilized into the cage formed by Phe303, Trp306, Trp330, and Asp333 (Phe236, Trp239, Trp263 and Asp266 for DNMT3B). The backbone of the peptide is involved in intermolecular hydrogen-bonding interactions with Phe336 ($Phe269_{3B}$) and Ser337 ($Ser270_{3B}$). A hydrophobic contact is observed between Pro38 and the isopropyl-group of Leu300 ($Ile233_{3B}$).
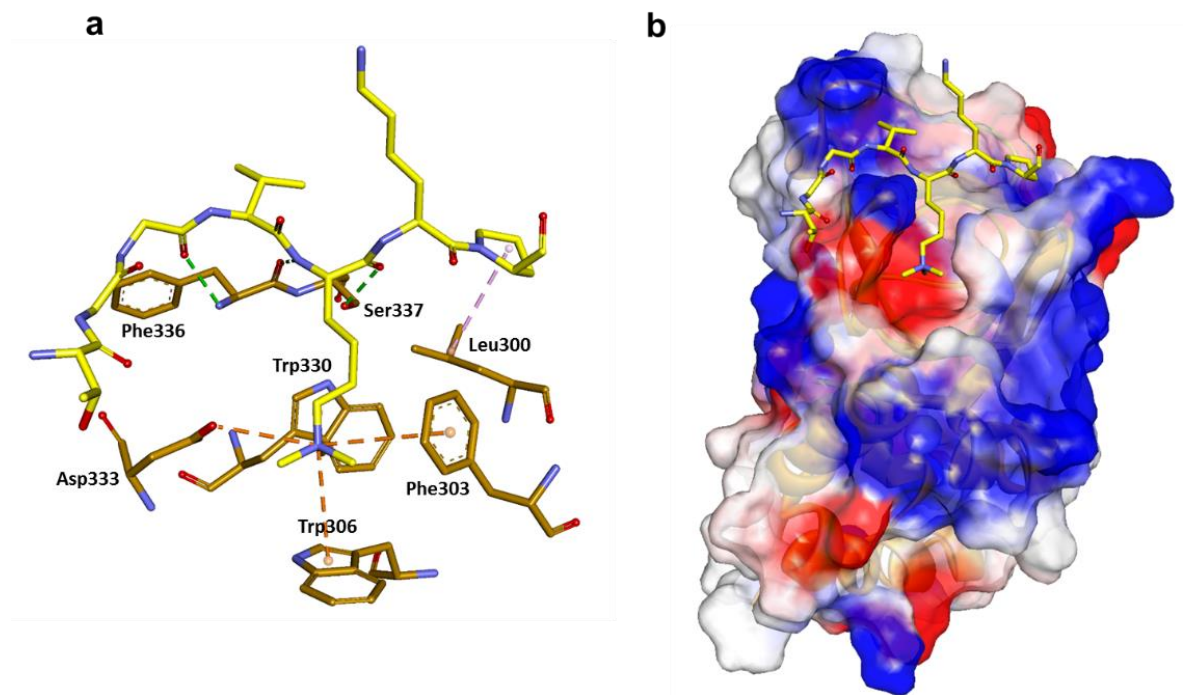
Figure 3. Molecular modeling of interactions between the DNMT3A PWWP domain and H3K36me3 peptide (H3$_{32-38}$K36me3). *(**a**) Intermolecular interactions of the top scoring model of the DNMT3A PWWP domain in complex with the H3K36me3 peptide. (**b**) DNMT3A PWWP domain surface colored by electrostatic potential of the residues (red: negatively charged area and blue: positively charged area) with the peptide represented in stick.*
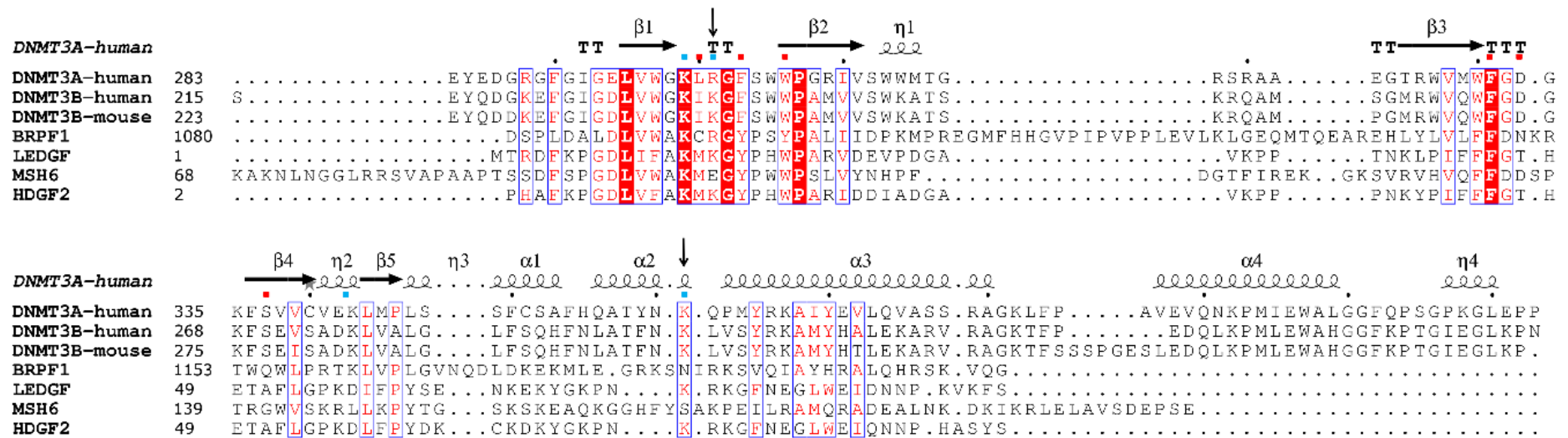
Figure 4. Structure-based sequence alignment of homologous PWWP domains reveals conserved residues involved in nucleosomal DNA binding. *This multiple structure and sequence alignments of homologous PWWP domains was performed with PROMALS3D (Pei et al., 2008) and formatted with Espript (Gouet et al., 1999). In the family of PWWP domains, a hydrophobic interface interacts with the histone tail and an adjacent more basic surface interacts with the negative phosphate backbone of the DNA in order to increase selectivity and affinity with the nucleosome. Residues involved in the H3K36me3 peptide binding: ■ DNA-binding residues: ■. For LEDGF protein, residues Lys16 (↓) and Lys73 (↓) have a large contribution to the intermolecular energy (van Nuland et al., 2013); they correspond to Arg301 and Lys361 in DNMT3A, and Lys234 and Lys294 in DNMT3B.*

3.2.3 DNMT3A/3B PWWP domains in complex with the histone H3K36me3 nucleosome core particle

In addition to their histone tail binding abilities, PWWP domains of DNMT3A and DNMT3B interact non-specifically with DNA through a basic surface adjacent to the histone binding site (Fig. 1b) (Purdy et al., 2010; Qiu et al., 2002). Individually, different PWWP domains bind free DNA in the μM range and H3K36me3 peptide in the μM–mM range which is quite low (Lukasik et al., 2006, Qiu et al., 2002; Yu et al., 2012). However, the cooperative binding of these partners can increase the binding affinity up to 10^4-fold and enhance specificity of PWWP domains for H3K36me3-NCP (nucleosome core particle) (Musselman et al., 2012; Wang et al., 2014). This seems to be a general property of PWWP modules and involves conserved interactions at the same interface (Eidahl et al., 2013; Qin and Min, 2014; van Nuland et al., 2013). Indeed, multiple structure and sequence alignments of homologous PWWP domains (Fig. 4) show that these domains present the same aromatic cage to bind the histone tail and the basic surface have conserved residues to interact with the nucleosomal DNA.

The question remains how the DNMT3A and DNMT3B PWWP domains bind the H3K36me3 epigenetic mark which is near the nucleosome core. To investigate this mechanism of recognition, we superimposed the DNMT3A and DNMT3B PWWP domains with the previously studied LEDGF PWWP domain (provided by Prof. Mamuka Kvaratskhelia), sharing high structural similarity, bound to H3K36me3-NCP (Eidahl et al., 2013). We observe that the basic surface is positioned on the DNA wrapped around histone core and is likely to interact with the phosphate backbone of DNA through the same conserved residues as those of LEDGF PWWP domain (Figs. 4 and 5A, B). DNMT3A and DNMT3B PWWP domains are situated between two DNA duplexes from where the epigenetic mark H3K36me3 emerges to interact with them (Figs. 5A and 6C). The selectivity of DNMT3A and DNMT3B for the H3K36me3-NCP is then mediated by interactions of the PWWP domain through the aromatic cage for H3K36me3 recognition and the basic surface which could bind two DNA duplexes of the nucleosome. Experiments showing this direct interaction between the PWWP domain and H3K36me3 nucleosomes were performed in previous studies for both DNMT3A and DNMT3B by GST pull-down assays (Baubec et al., 2015; Dhayalan et al., 2010). These interactions are

disrupted by different point mutations (as mentioned earlier in the description of the complex between the DNMT3B PWWP domain and histone H3K36me3).
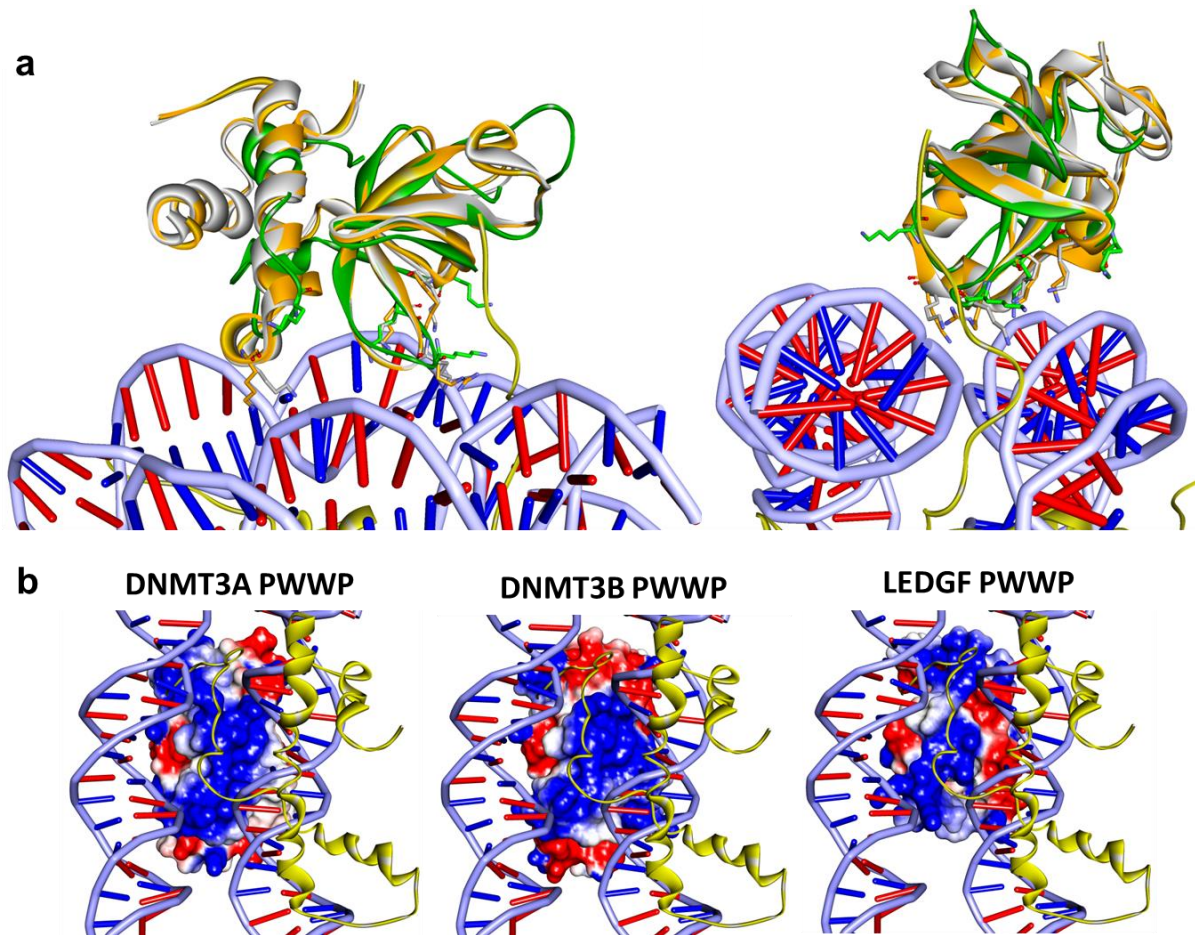
***Figure 5. PWWP domains in complex with H3K36me3-NCP.** (**a**) HADDOCK model of the complex between the LEDGF PWWP (green) domain and H3K36me3-NCP (Eidahl et al., 2013) superimposed with DNMT3A (orange) (RMSD of 1.9 Å) and DNMT3B (gray) (RMSD of 1.9 Å) PWWP domains. (**b**) PWWP domains surface colored by electrostatic potential of the residues (red: negatively charged area and blue: positively charged area). The positively charged surface is positioned between two DNA strands and contains the following residues for the DNMT3A PWWP domain: Lys299, Arg301, Lys343 and Lys 361. These residues correspond to Lys232, Lys 234, Lys276, and Lys 294 in the DNMT3B PWWP domain. They are situated on β1, on the loop between β1 and β2, between β4 and β5 (n2), on the loop between α2 and α3, and on helix α3.*

3.2.4 DNMT3A (PWWP-ADD-CD domains) in a nucleosomal context

DNA methylation functions in specific cellular and genomic contexts. For example, DNA methylation is dependant of a number of post-translational modifications on histones to

recruit and activate the *de novo* DNA methylation complex. DNMT3L, which lacks the PWWP domain and the catalytic domain, recognizes specifically the unmethylated histone H3 tail (H3K4me0) through the ADD domain (Table S1) (Eustermann et al., 2011; Ooi et al., 2007; Otani et al., 2009). This protein is expressed especially in oocytes to stabilize the conformation of the active loop and stimulates the activity of DNMT3A and DNMT3B for the establishment of the genomic imprinting (Jia et al., 2007; Smallwood et al., 2011; Suetake et al., 2004). The epigenetic mark H3K4me0 is also bound by DNMT3A via the same ADD domain to allosterically activates the DNMT3A (Guo et al., 2014; Li et al., 2011; Otani et al., 2009). Without this recognition, the ADD domain interacts specifically with the catalytic domain and prevents interaction with the DNA substrate inhibiting the activity. Binding of this epigenetic mark H3K4me0 to the ADD domain disrupts this autoinhibitory structure and an important conformational change permits to the catalytic domain to recognize the DNA. The ADD domain of DNMT3B binds the H3K4me0 (1–19) with the same affinity as DNMT3A suggesting a similar mechanism (Table S1) (Zhang et al., 2010). H3K36me3 epigenetic mark, as explained before, is read by both DNMT3A and DNMT3B PWWP domains to recruit these enzymes on the nucleosomal DNA in the pericentromeric heterochromatin regions and gene bodies (Bachman et al., 2001; Baubec et al., 2015; Chen et al., 2004; Dhayalan et al., 2010). The combination of histone modifications permissive unmethylated H3K4me0 and trimethylated H3K36me3 seems therefore necessary for the DNMT3A and DNMT3B to access and methylate DNA (Stewart et al., 2015; Tomizawa et al., 2012).

Different structural and biophysical data are available for separated domains of DNMT3s with individual parts of nucleosome. However, no structural information is available for the recognition of DNMT3s with the integral nucleosome. In this regard, we propose here a model of the DNMT3A (PWWP-ADD-CD) in a nucleosomal context. The crystal structure of the active complex of DNMT3A (ADD-CD)-DNMT3L-H3K4me0 peptide has recently been solved without the PWWP domain (Guo et al., 2014) (PDB code: 4U7T). Based on this structure, we predict the DNMT3A (PWWP-ADD-CD) structure including the missing PWWP domain by using three protein–protein docking servers emerging from the CAPRI experiment (Janin et al., 2003).

This reconstructed DNMT3A structure (Fig. 6a) shows that the β-strands, β2 and β3, and the insertion motif η1 (α-helix motif) of the PWWP domain are positioned at the interface with the ADD domain (Figs. 6A and S3). The PWWP and the ADD domains are associated by

92

strong electrostatic interactions and diverse hydrogen bonds. In addition, due to the high structural and sequence similarity of the catalytic domain of DNMT3A and the M.*Hha*I, a bacterial DNMT we solved in complex with a short DNA duplex, we superimposed these domains (root-mean-square deviation of 1.2 Å). The short oligonucleotide is in the continuity of the nucleosomal DNA and can be connected easily to the border of the nucleosome to form a contiguous DNA segment (Fig. 6b and c).

This approach yields a complete structural model of the H3K36me3-nucleosome-DNMT3A complex which can explain DNMT3A recruitment to a genomic site (Fig. 6c). The PWWP domain specifically binds the H3K36me3-NCP followed by an activation of the catalytic domain through the binding of the H3K4me0 with the ADD domain to methylate the nearby cytosine (Fig. 8a). We also introduced in this study, the DNMT3L from the active complex of DNMT3A (ADD-CD)-DNMT3L-H3K4me0 peptide (Guo et al., 2014) and the dimerization of DNMT3A through the catalytic domain (Chodavarapu et al., 2010; Jia et al., 2007; Yazdi et al., 2015). This gives a heterotetramer (DNMT3L-DNMT3A-DNMT3A-DNMT3L) in complex with a dinucleosome exhibiting a bent linker DNA (Figs. 7 and 8B). The DNMT3A-DNMT3A dimer is responsible for methylation of DNA in a periodic pattern of ~10-bp between two CpG sites corresponding to a helical turn as previously demonstrated (Jia et al., 2007; Lister et al., 2009; Smallwood et al., 2011). Moreover, this model highlights, as described in the literature, that *de novo* DNA methylation occurs on the linker DNA region between nucleosomes and at the border of the nucleosome (Baubec et al., 2015; Felle et al., 2011; Morselli et al., 2015; Takeshima et al., 2008). Indeed, DNMT3A and DNMT3B prefer to target the nucleosome-bound DNA through a synergetic effect to methylate the linker DNA region between two nucleosomes which is more accessible (Felle et al., 2011, Morselli et al., 2015; Zhang et al., 2010). Furthermore, the nucleosomes are more regularly connected in the pericentromeric heterochromatin regions with a linker DNA of ~30-bp which corresponds to the linker DNA size of our model (Chodavarapu et al., 2010). The four ADD domains present in the heterotetrameric complex (DNMT3L-DNMT3A-DNMT3A-DNMT3L) could read the H3K4me0 state for the DNA methylation establishment. The question may be asked whether DNMT3L could interact with the second copy of H3 tail of the histone octamer from adjacent nucleosomes. This could lead to a synergistic effect of binding proteins on nucleosome. This model needs further testing to answer these questions, nevertheless taken together our

observations emphasize the functional importance of the combination of histone tail modification status for the formation of a stable *de novo* methylation complex to methylate DNA in a periodic pattern.
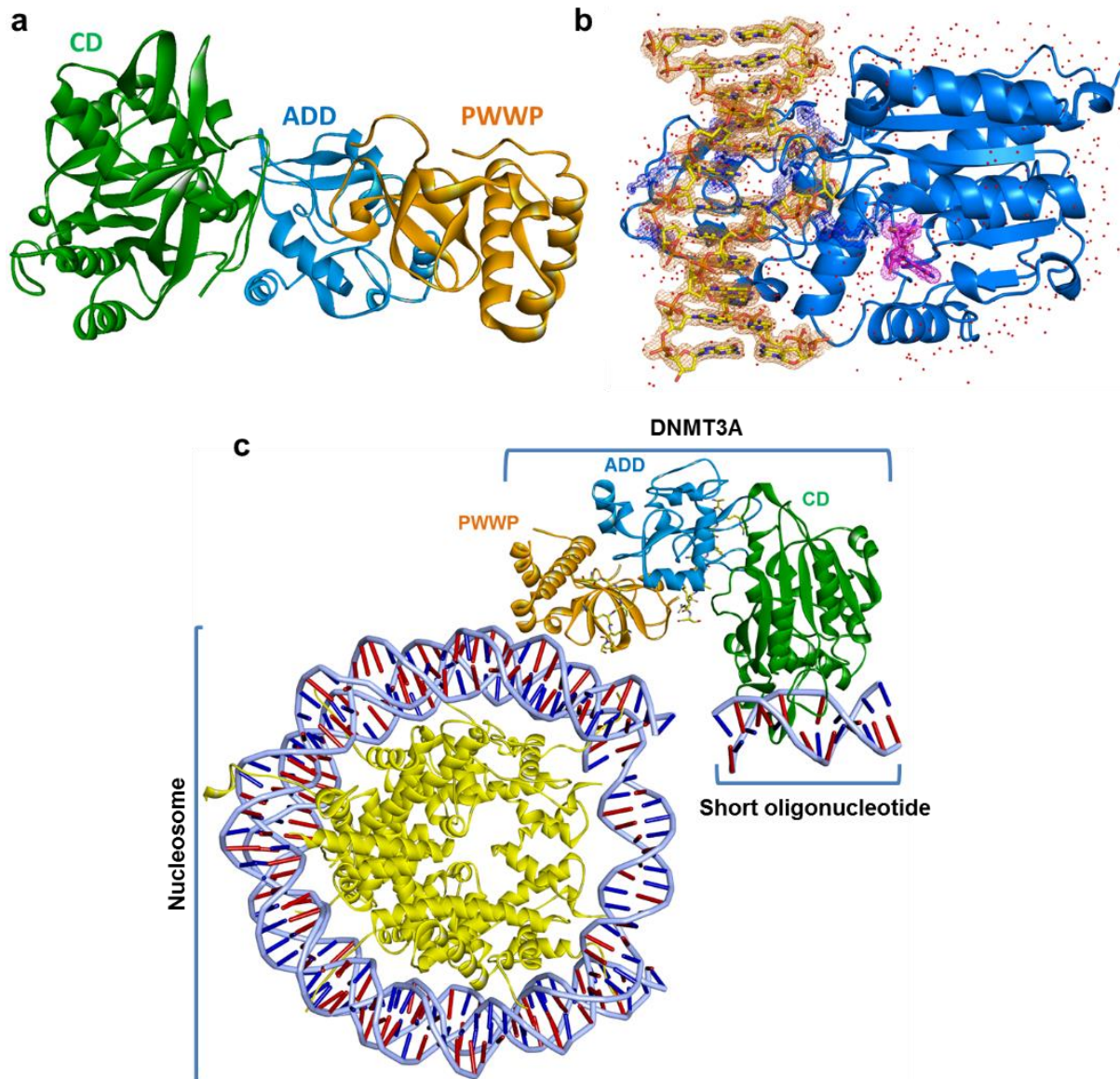
***Figure 6. Construction of the complete DNMT3A (PWWP-ADD-CD domains) in a nucleosomal context. (a)** Cartoon view of the complete reconstructed DNMT3A (PWWP-ADD-CD domains). **(b)** The bacterial M.HhaI (a DNA (cytosine-5) methyltransferase from Haemophilus haemolyticus) in complex with DNA and SAH (S-adenosyl-l-homocysteine). 2Fo-Fc electron density map contoured at 1.0 σ for short oligonucleotide (orange mesh), SAH (pink mesh) and selected residues interacting with DNA (blue mesh). **(c)** Structural model of the DNMT3A in complex with H3K36me3 nucleosome core particle.*
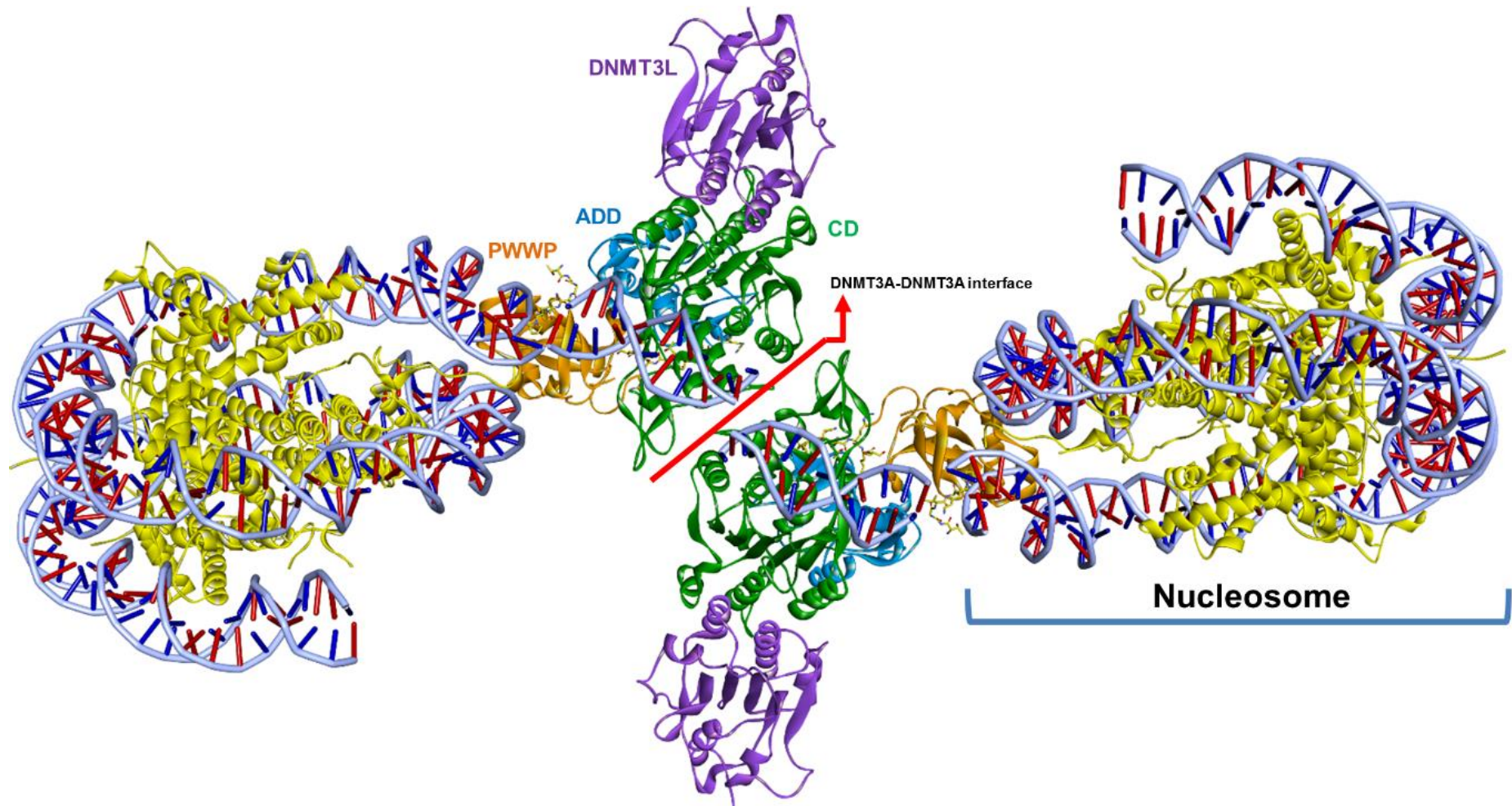
Figure 7. Structural model of DNMT3A (PWWP-ADD-CD domains)-DNMT3L heterotetramer in complex with H3K36me3-modified dinucleosome. *Two active sites of distinct DNMT3A (green solid ribbon) are located on the linker DNA region between and at the border of the nucleosomes. The length of the linker DNA is ~30-bp and corresponds to the average length in the pericentromeric heterochromatin regions (Chodavarapu et al., 2010). The center-to-center distance between the core particles is 20 nm. The model is available online: http://dx.doi.org/10.1016/j.jsb.2016.03.013*
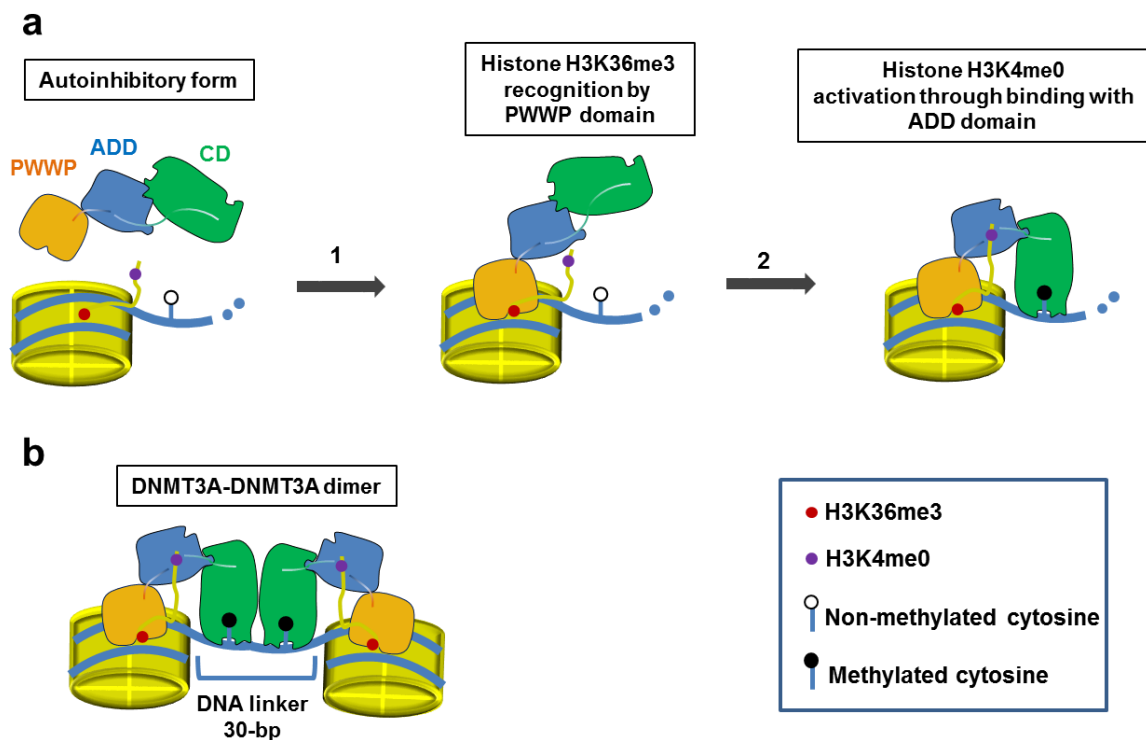
Figure 8. Proposed mechanism for recognition of nucleosome and DNA methylation by DNMT3A (for clarity, the DNMT3L was removed). (a) *Sequential recognition mechanism of DNMT3A for nucleosome and methylation activation with H3 tail: **1**-Nucleosome recognition by PWWP domain through interaction with methylated histone H3 tail (H3K36me3). **2**- Histone H3 tail (H3K4me0) allosterically activates DNMT3A through binding with ADD domain. ADD domain of DNMT3A interacts with the catalytic domain and inhibits its activity by preventing it to interact with the DNA. **(b)** Periodicity in de novo methylation patterns of ~10-bp between two CpG sites.*

It is still difficult to understand the difference of methylation activity between DNMT3A and DNMT3B. Indeed, as for DNMT3A, DNMT3B methylates the linker DNA with a periodicity of 10-bp and a periodicity corresponding to the nucleosome repeat length of ~180 bp (nucleosomal DNA (147-bp) + linker DNA (~30-bp)) (Baubec et al., 2015; Cokus et al., 2008; Morselli et al., 2015). DNMT3B binding to nucleosome is also guided by the presence of two epigenetic marks, the H3K4me0 and the H3K36me3 peptides (Baubec et al., 2015; Morselli et al., 2015). DNMT3A and DNMT3B act in a same nucleosome context in terms of organization of chromatin but methylate different regions of the genome. DNMT3B is however more enriched within gene bodies (Baubec et al., 2015; Jin et al., 2012). DNMT3A and DNMT3B are also distributed differently across the genome and this depends on the

cellular context (Jin et al., 2012). The mechanism of DNMT3s targeting to specific site of methylation throughout the genome is actually more complex and involves a series of events including the timing of expression (Lees-Murdock et al., 2005). The differences of recruitment of DNMT3s across the genome may be influenced by other specific histone tail modifications. Their distinct biological functions can be explained by the N-terminal sequence which shares very low sequence identity (10.9% using the EMBOSS needle program) between the two DNMT3s (Fig. S1). In our model we do not present this N-terminal part of DNMT3A (amino acids 1–281) as no structural information is available for neither DNMT3A nor DNMT3B. This N-terminal part of both DNMT3A/3B seems important for anchoring to nucleosomes (Baubec et al., 2015; Jeong et al., 2009). Indeed, based on multiple sequence alignments including this sequence, we found relate domains involved in DNA binding including the PDS5 homolog B protein (uniprotkb:Q6TRW4). This protein is a regulator which stabilizes cohesion complex association with chromatin suggesting that the N-terminal part of the DNMT3A could interact with DNA through its 21 arginines and 22 lysines (Fig. S1). Furthermore, this N-terminal part of both DNMT3A/3B could be important to mediate protein–protein interactions to discriminate the recruitment of these two enzymes on the nucleosomal-DNA (Geiman et al., 2004; Li et al., 2006; Rigbolt et al., 2011; Sarraf and Stancheva, 2004; Velasco et al., 2010).

3.3 Conclusion

Here, we report the crystal structure of the *de novo* DNA methyltransferase DNMT3B PWWP domain in complex with the epigenetic mark H3$_{32-38}$K36me3. This structure emphasizes that a conserved water molecule mediates strong hydrogen bonding with the epigenetic mark explaining the loss of affinity in the ICF syndrome with the mutated DNMT3B PWWP domain (S270P). Based on this structure and the docked structure DNMT3A PWWP-H3K36me3, a structural model highlights interactions between PWWP domains of DNMT3s and a nucleosome core particle. PWWP domains, situated between two DNA duplexes, engage the same binding surface and residues to interact with the H3K36me3 nucleosome. Finally, the complete model structure of the DNMT3A-DNMT3L heterotetramer in complex with a dinucleosome provides structural information about DNA methylation on the linker DNA region of nucleosomes and the *de novo* DNA methylation patterns in agreement with the findings of the literature.

3.4 Materials and methods

3.4.1 Cloning, expression and purification of DNMT3B PWWP domain

The plasmid coding for the N-terminal human PWWP domain of DNMT3B was obtained from Addgene (plasmid 32044, C. Arrowsmith) as a bacterial stab culture. The plasmid containing the pET28-MHL vector codes for an N-terminal His6-tagged fusion protein with integrated TEV protease site. Plasmids were recovered from kanamycin-resistant colonies and purified following the Addgene protocol. Plasmid was sent for sequencing to Beckman Coulter Genomics (Hope end, Takeley CM22 6TA, Essex United Kingdom). Plasmid was transformed into *Escherichia coli* strain Rosetta™ 2 (DE3) competent cells (Novagen®). Cells were grown in TB medium containing 50 μg/ml kanamycin and 34 μg/ml chloramphenicol at 37 °C. When the OD 600 nm was about 1.5, the temperature was reduced to 18 °C and expression of PWWP domain was induced with 1 mM IPTG overnight. The cells were recovered by centrifugation at 7000 rpm for 30 min at 4 °C. The harvested cells were resuspended in ice-cold lysis buffer (50 mM HEPES, pH 7.4, 500 mM NaCl, 2 mM β-mercaptoethanol, 5% glycerol, 0,1% CHAPS) with protease inhibitor cocktail (Complete Mini EDTA-Free, 1 tablet for 10 mL of solution, Roche). Cells were lysed on ice by sonication using a cell disrupter. Crude extract were centrifuged at 9000 rpm for 30 min at 4 °C to recover the lysate. The supernatant was filtered through a 0.2 μm filter (Whatman® FP 30/0.2, GE Healthcare) and loaded onto a 5 ml HisTrap™ FF crude Ni-Chelating column (GE Healthcare) equilibrated with 20 mM HEPES, pH 7.4, 500 mM NaCl, 5% glycerol, 50 mM imidazole. The charged column was washed with the equilibration buffer until a stable baseline was attained and the elution of PWWP domain of DNMT3B was performed with a linear gradient from 50 mM to 1 M imidazole over a total volume of 50 ml. After pooling the appropriate fractions, the solution was dialyzed overnight against 20 mM HEPES, pH 7.4 and 250 mM NaCl in presence of TEV protease (50 μg/mg of protein) to cleave the His-tag. After this, the solution was reloaded onto a 5 ml HiTrap™ IMAC FF Ni-Chelating column (GE Healthcare) and the column was washed with 10 column volumes of 20 mM HEPES, pH 7.4, 500 mM NaCl, 5% glycerol and 50 mM imidazole to elute the cleaved protein. Then, the protein was dialyzed against equilibration buffer (20 mM PIPES, pH 6.5). The sample was filtered through a 0.2 μm filter (Whatman® FP 30/0.2, GE Healthcare) and was loaded into an AKTA FPLC system (GE-Amersham Biosciences) with a cation exchange

column (1 ml, Resource™ S, GE Healthcare) to purify to homogeneity. The charged column was washed with the equilibration buffer until a stable baseline was attained and the elution of the protein was performed with a linear gradient of NaCl up to 1 M concentration over a total volume of 20 ml. The solution was concentrated in a stirred cell using a 5 kDa MWCO membrane (Vivaspin 15R; Sartorius Stedim Biotech GmbH, Göttingen, Germany) up to 22 mg/mL prior to crystallization. All the different stages of expression and purification were followed by SDS–PAGE.

3.4.2 H3K36me3 peptide production and purification

H3K36me3 peptide (SAPATGGV{K(Me3)}KPHRYR) 28–42 was purchased from GenScript (Piscataway, NJ, USA) with a purity up to 95% using reverse-phase HPLC.

3.4.3 Crystallization and structure determination of DNMT3B PWWP domain in complex with histone H3K36me3

Crystallization trials were performed using the Hampton Research Crystal Screen kits 1–2 and the sitting-drop vapor-diffusion method in 96-well plates at room temperature. Drops consisted of 1 µl of DNMT3B PWWP domain at 22 mg/ml – peptide mixture (1:5 M ratio) plus 1 µl of reservoir solution equilibrated against a reservoir volume of 50 µl. Crystals of complex between DNMT3B PWWP domain and H3K36me3 peptide were grown against a reservoir consisting of 1.6 M Sodium citrate tribasic pH 6.5 (Hampton Research Crystal Screen 2 No. 28). A single-crystal (crystal size: 0.45 mm × 0.4 mm) was cryoprotected in 20% glycerol and flash-cooled in liquid nitrogen. Data set was collected at SOLEIL Synchrotron, Gif-sur-Yvette, France on beamline PROXIMA 2 (PX2-A) using ADSC Q315 detector at a wavelength of 0.9801 Å. The data were processed using XDS/XSCALE (Kabsch, 2010). Initial phases were calculated by molecular replacement using the program PHASER in PHENIX (Adams et al., 2010; McCoy et al., 2007) with experimental reflection data set (Fo) and the search model (Fc) DNMT3B PWWP domain in complex with bis–tris ligand solved at 2.0 Å resolution (Wu et al., 2011) (PDB code: 3QKJ). The complex was built using the Coot program (Emsley et al., 2010) and refined with the program PHENIX (Adams et al., 2010). Single group occupancy of the peptide was set to 0.5 on both monomers and refined. After refinement, occupancy of 0.8 for each peptide was retained and gave good agreement with B values of surrounding residues. A

summary of the data-collection and refinement statistics is presented in Table 1. Ramachandran plot for the final model shows 93.92% residues in favored regions and 6.08% in allowed regions. Figures were drawn using both Pymol (DeLano, 2002) and Discovery Studio (Visualizer, 2013).

3.4.4 Modeling of the DNMT3A PWWP domain in complex with histone H3K36me3

First, we validated the method by docking the H3K36me3 peptide into the DNMT3B PWWP domain (PDB code: 5CIU). For the validation procedure, the assessment of model quality was evaluated by the backbone RMSD (RMSBB) between the crystallographic structure and the peptide model. Flexpepdock retrieved the same binding mode and backbone positioning for eight models among the ten top scoring models, and the RMSBB of the best peptide model with the native peptide was 1.0 Å (<2 Å) (Fig. S2a). To model the H3K36me3 peptide (H3$_{32-38}$K36me3)-DNMT3A PWWP domain complex, we created an initial starting structure of the H3K36me3 peptide within the aromatic cage of DNMT3A PWWP domain by superposition with the DNMT3B PWWP domain–H3K36me3 structure, expecting the peptide to be in the correct binding site. We merged the estimated conformation for the H3K36me3 peptide with the DNMT3A PWWP domain receptor into an input PDB file for FlexPepDock refinement on the server. We also provided a Rosetta atom-pair constraints file to fix the trimethyl-ammonium group of Lys36 inside the aromatic cage during the simulation. The new structure was then refined in 600 independent FlexPepDock simulations. 300 simulations were performed in high-resolution mode (full-atom mode) and, in order to increase samples of the conformational space, 300 simulations including a low-resolution pre-optimization step (centroid-based optimization) prior to the high-resolution refinement were performed. After a prepacking to remove internal clashes, FlexPepDock applied a Monte Carlo minimization to iteratively (ten iterative cycles) optimize the rigid body orientation, the backbone and side chain flexibility of the peptide. During this minimization, side chain flexibility of protein receptor was also considered. The total 600 models were ranked based on the Rosetta full-atom energy score (Rohl et al., 2004). Among the ten top scoring models, nine were found to have the same binding mode and backbone positioning (Fig. S2b).

3.4.5 Construction of the DNMT3A (PWWP-ADD-CD) structure

We selected as input structures the opened form and catalytically active structure for CD-ADD of DNMT3A with H3K4me0 epigenetic mark (Guo et al., 2014) (PDB code: 4U7T) and the DNMT3A PWWP domain (Wu et al., 2011) (PDB code: 3LLR). The docking servers Cluspro 2.0 (Comeau et al., 2004a; Comeau et al., 2004b; Kozakov et al., 2006; Kozakov et al., 2013) and Zdock 3.0.2 (Pierce et al., 2014) served to identify the interface interaction and the conformation of the complex. These servers perform both an automated rigid-body docking by using the fast Fourier transform correlation method and explore all the different binding modes by combination of translation and rotation of the ligand. Cluspro clusters the low-energy docked conformations based on the RMSD and classify the clusters according to their size. Zdock uses an energy-based scoring function to evaluate each pose. For the calculations, known binding sites of the DNMT3A as the DNMT3L-DNMT3A (catalytic domain) interface and the DNA binding site of the PWWP domain were not considered. After calculations, we get the same conformation for the lowest scoring function value for Zdock and Cluspro and identified a common binding interface between the ADD domain and the PWWP domain (Fig. S3). Moreover, the obtained conformation was retrieved seven times among the ten top scoring models of Zdock. The resulting models from Zdock and Cluspro were then subjected to rigid-body minimization and side-chain conformation optimization using the "docking-local-refine" on the RosettaDock server (Chaudhury et al., 2011; Lyskov and Gray, 2008; Lyskov et al., 2013). The lowest RosettaDock binding score model (ROSETTADOCK binding score was −522.4 for Zdock and −534.8 for Cluspro) shown in orange in Fig. 6a was selected for structural analysis with the nucleosome (Figs. 6c and 7).

3.4.6 Expression and purification of M.*Hha*I

LB medium containing 100 µg/ml ampicillin was inoculated with Escherichia coli K-12 strain ER1727 containing the pUHE25*Hha*I plasmid. When the OD600 nm reached 0.6, expression of M.*Hha*I was induced with 1 mM IPTG. After 3 h of induction at 37 °C, the cells were collected by centrifugation at 4000 rpm for 30 min at 4 °C. The cell paste was resuspended in ice-cold lysis buffer (10 mM Hepes (pH 7.0), 5 mM EDTA, 10% (v/v) glycerol,

0.1% (v/v) β-mercaptoethanol). Cells were lysed on ice by sonication using a cell disrupter. After centrifugation at 4000 rpm for 1 h at 4 °C, the supernatant was discarded and the pellet was resuspended in a high-salt buffer (10 mM Tris, 5 mM EDTA (pH 7.5), 10% (v/v) glycerol, 0.1% (v/v) β-mercaptoethanol, 500 mM NaCl) to re-solubilize the protein. The high-salt suspension was centrifuged at 14000 rpm for 1 h at 4 °C, the supernatant was kept and the pellet undergoing the same processing with high-salt buffer. Precipitation of nucleic acids from the supernatant was performed by progressive addition of a half volume of protamine sulfate solution (10 mg/ml, solubilized in high-salt buffer). After incubation at room temperature for 5 min, the solution was centrifuged at 14000 rpm for 40 min at 4 °C and the supernatant dialyzed against 50 mM Tris (pH 6.7), 5 mM EDTA, 10% (v/v) glycerol, 0.1% (v/v) β-mercaptoethanol, 100 mM NaCl. All the following steps of purification were performed at 4 °C. The sample was filtered through a 0.2 μm filter (Whatman® FP 30/0.2, GE Healthcare) and loaded onto a cation exchange column (1 ml, HiTrap™ SP FF, GE Healthcare). The charged column was washed with the equilibration buffer (50 mM Tris (pH 6.7), 5 mM EDTA, 10% (v/v) glycerol, 0.1% (v/v) β-mercaptoethanol, 100 mM NaCl) until a stable baseline was attained and the elution of the M.*Hha*I was realized with a linear gradient from 100 mM to 500 mM NaCl over a total volume of 50 ml. After pooling the appropriate fractions, the solution was concentrated in a stirred cell using a 10 kDa MWCO membrane (Vivaspin 15R; Sartorius Stedim Biotech GmbH, Göttingen, Germany) until ~5 ml and dialyzed against 50 mM Tris (pH 7.4), 5 mM EDTA, 10% (v/v) glycerol, 0.1% (v/v) β-mercaptoethanol, 100 mM NaCl. The sample was filtered through a 0.2 μm filter (Whatman® FP 30/0.2, GE Healthcare) and was injected into an AKTA FPLC system (GE-Amersham Biosciences) with an anion exchange column (1 ml, Mono-Q HR 5/5, GE Healthcare) equilibrated with 50 mM Tris (pH 7.4), 5 mM EDTA, 10% (v/v) glycerol, 0.1% (v/v) β-mercaptoethanol, 100 mM NaCl. The charged column was washed with the equilibration buffer until a stable baseline was attained. The collected flow-through fractions containing the purified M.*Hha*I were pooled and concentrated up to 10 mg/mL concentration prior to crystallization. All the different stages of expression and purification were followed by SDS–PAGE.

3.4.7 Crystallization and structure determination of M.*Hha*I in complex with SAH and a short DNA duplex

The crystallization conditions were based and adapted from previously published data (O'Gara et al., 1996; O'Gara et al., 1998). The oligonucleotides used to form the 12 bp duplex were purchased from Eurogentec (Seraing, Belgium).

5'-T-C-C-A-T-G-C-G-C-T-G-A-C- -3'

3'- -G-G-T-A-C-G-X-G-A-C-T-G-T-5'
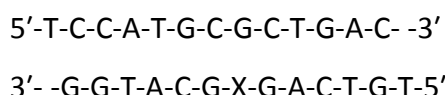
These two synthetic complementary strands of oligonucleotides were designed with a single 5' thymidine (T)-overhangs at both ends and an abasic residue was incorporated with a dSpacer (X) at the target base pair (GC → GX). The two strands were hybridized in TE buffer at 90 °C for 5 min, followed by room temperature incubation for 1 h. Crystallization trials were performed using the hanging-drop vapor-diffusion method at room temperature with purified M.*Hha*I concentrated at 10 mg/ml, SAH (dissolved in water) and DNA with following molar ratio: 1:3:1.3. Drops consisted of 2 µl of M.*Hha*I-SAH-DNA mixture plus 2 µl of reservoir solution equilibrated against a reservoir volume of 700 µl. Crystals of complex between M.*Hha*I, SAH and DNA were grown against a reservoir consisting of 50 mM Citrate pH 5.6 and 1.8 M ammonium sulfate. A single-crystal (crystal size: 0.3 mm × 0.2 mm) was cryoprotected in 20% glycerol and flash-cooled in liquid nitrogen. Data set was collected at SOLEIL Synchrotron, Gif sur Yvette, France on beamline PROXIMA 2 (PX2-A) using an ADSC Q315r detector at a wavelength of 0.9801 Å. The data were processed using XDS/XSCALE (Kabsch, 2010). A summary of the data-collection and refinement statistics is presented in Table 1. Initial phases were calculated by molecular replacement using the program PHASER in PHENIX (Adams et al., 2010; McCoy et al., 2007) with experimental reflection data set (Fo) and the search model (Fc) solved at 2.4 Å resolution (O'Gara et al., 1998) (PDB code: 9MHT). The complex was built using the Coot program (Emsley et al., 2010) and refined with the program PHENIX (Adams et al., 2010). Ramachandran plot for the final model shows 97.85% residues in favored regions, 1.85% in allowed regions and 0.31% in disallowed regions. Figures were drawn using both Pymol (DeLano, 2002) and Discovery Studio (Visualizer, 2013).

Data deposition

Coordinates and diffraction data for the structures of DNMT3B PWWP-H3K36me3 and M.*Hha*I-DNA-SAH have been deposited in the Protein Data Bank with accession codes 5CIU and 5CIY, respectively.

Footnote

Acknowledgements

3.5 References

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D: Biological Crystallography 66, 213-221.

Ballaré, C., Lange, M., Lapinaite, A., Martin, G.M., Morey, L., Pascual, G., Liefke, R., Simon, B., Shi, Y., Gozani, O., 2012. Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. Nature Structural & Molecular Biology 19, 1257-1265.

Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., Schubeler, D., 2015. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature 520, 243-247.

Bird, A., 2002. DNA methylation patterns and epigenetic memory. Genes & development 16, 6-21.

Cai, L., Rothbart, S.B., Lu, R., Xu, B., Chen, W.-Y., Tripathy, A., Rockowitz, S., Zheng, D., Patel, D.J., Allis, C.D., 2013. An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. Molecular cell 49, 571-582.

Chaudhury, S., Berrondo, M., Weitzner, B.D., Muthu, P., Bergman, H., Gray, J.J., 2011. Benchmarking and analysis of protein docking performance in Rosetta v3. 2. PloS one 6, e22477.

Chen, T., Tsujimoto, N., Li, E., 2004. The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. Molecular and cellular biology 24, 9048-9058.

Cheng, J., Yang, Y., Fang, J., Xiao, J., Zhu, T., Chen, F., Wang, P., Li, Z., Yang, H., Xu, Y., 2013. Structural insight into coordinated recognition of trimethylated histone H3 lysine 9 (H3K9me3) by the plant homeodomain (PHD) and tandem tudor domain (TTD) of UHRF1 (ubiquitin-like, containing PHD and RING finger domains, 1) protein. Journal of Biological Chemistry 288, 1329-1339.

Chodavarapu, R.K., Feng, S., Bernatavichute, Y.V., Chen, P.-Y., Stroud, H., Yu, Y., Hetzel, J.A., Kuo, F., Kim, J., Cokus, S.J., 2010. Relationship between nucleosome positioning and DNA methylation. Nature 466, 388-392.

Comeau, S.R., Gatchell, D.W., Vajda, S., Camacho, C.J., 2004a. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. Bioinformatics 20, 45-50.

Comeau, S.R., Gatchell, D.W., Vajda, S., Camacho, C.J., 2004b. ClusPro: a fully automated algorithm for protein–protein docking. Nucleic acids research 32, W96-W99.

DeLano, W.L., 2002. The PyMOL molecular graphics system.

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

Eidahl, J.O., Crowe, B.L., North, J.A., McKee, C.J., Shkriabai, N., Feng, L., Plumb, M., Graham, R.L., Gorelick, R.J., Hess, S., 2013. Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. Nucleic acids research 41, 3924-3936.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta Crystallographica Section D: Biological Crystallography 66, 486-501.

Eustermann, S., Yang, J.-C., Law, M.J., Amos, R., Chapman, L.M., Jelinska, C., Garrick, D., Clynes, D., Gibbons, R.J., Rhodes, D., 2011. Combinatorial readout of histone H3 modifications specifies localization of ATRX to heterochromatin. Nature Structural & Molecular Biology 18, 777-782.

Felle, M., Hoffmeister, H., Rothammer, J., Fuchs, A., Exler, J.H., Längst, G., 2011a. Nucleosomes protect DNA from DNA methylation in vivo and in vitro. Nucleic acids research 39, 6956-6969.

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

Geiman, T.M., Sankpal, U.T., Robertson, A.K., Chen, Y., Mazumdar, M., Heale, J.T., Schmiesing, J.A., Kim, W., Yokomori, K., Zhao, Y., 2004. Isolation and characterization of a novel DNA methyltransferase complex linking DNMT3B with components of the mitotic chromosome condensation machinery. Nucleic acids research 32, 2716-2729.

Gouet, P., Courcelle, E., Stuart, D.I., 1999. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics 15, 305-308.

Grigoryev, S.A., 2012. Nucleosome spacing and chromatin higher-order folding. Nucleus 3, 493-499.

Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., 2014. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature 517, 640-644.

Hahn, M.A., Wu, X., Li, A.X., Hahn, T., Pfeifer, G.P., 2011. Relationship between gene body DNA methylation and intragenic H3K9me3 and H3K36me3 chromatin marks. PloS one 6, e18844.

Janin, J., Henrick, K., Moult, J., Eyck, L.T., Sternberg, M.J., Vajda, S., Vakser, I., Wodak, S.J., 2003. CAPRI: a critical assessment of predicted interactions. Proteins: Structure, Function, and Bioinformatics 52, 2-9.

Jeong, S., Liang, G., Sharma, S., Lin, J.C., Choi, S.H., Han, H., Yoo, C.B., Egger, G., Yang, A.S., Jones, P.A., 2009. Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA. Molecular and cellular biology 29, 5366-5376.

Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., Cheng, X., 2007. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature 449, 248-251.

Jin, B., Ernst, J., Tiedemann, R.L., Xu, H., Sureshchandra, S., Kellis, M., Dalton, S., Liu, C., Choi, J.-H., Robertson, K.D., 2012. Linking DNA methyltransferases to epigenetic marks and nucleosome structure genome-wide in human tumor cells. Cell reports 2, 1411-1424.

Jones, P.A., 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature Reviews Genetics 13, 484-492.

Kabsch, W., 2010. Xds. Acta Crystallogr D Biol Crystallogr 66, 125-132.

Kozakov, D., Brenke, R., Comeau, S.R., Vajda, S., 2006. PIPER: an FFT-based protein docking program with pairwise potentials. Proteins: Structure, Function, and Bioinformatics 65, 392-406.

Kozakov, D., Beglov, D., Bohnuud, T., Mottarella, S.E., Xia, B., Hall, D.R., Vajda, S., 2013. How good is automated protein docking? Proteins: Structure, Function, and Bioinformatics 81, 2159-2166.

Lee, J.-S., Shilatifard, A., 2007. A site to remember: H3K36 methylation a mark for histone deacetylation. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 618, 130-134.

Lees-Murdock, D.J., Shovlin, T.C., Gardiner, T., De Felici, M., Walsh, C.P., 2005. DNA methyltransferase expression in the mouse germ line during periods of de novo methylation. Developmental dynamics 232, 992-1002.

Li, B.-Z., Huang, Z., Cui, Q.-Y., Song, X.-H., Du, L., Jeltsch, A., Chen, P., Li, G., Li, E., Xu, G.-L., 2011. Histone tails regulate DNA methylation by allosterically activating de novo methyltransferase. Cell research 21, 1172-1181.

Li, H., Rauch, T., Chen, Z.-X., Szabó, P.E., Riggs, A.D., Pfeifer, G.P., 2006. The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. Journal of Biological Chemistry 281, 19489-19500.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462, 315-322.

Lukasik, S.M., Cierpicki, T., Borloz, M., Grembecka, J., Everett, A., Bushweller, J.H., 2006. High resolution structure of the HDGF PWWP domain: a potential DNA binding domain. Protein science 15, 314-323.

Lyskov, S., Gray, J.J., 2008. The RosettaDock server for local protein–protein docking. Nucleic acids research 36, W233-W238.

Lyskov, S., Chou, F.-C., Conchúir, S.Ó., Der, B.S., Drew, K., Kuroda, D., Xu, J., Weitzner, B.D., Renfrew, P.D., Sripakdeevong, P., 2013. Serverification of molecular modeling applications: the Rosetta Online Server that Includes Everyone (ROSIE). PloS one 8, e63906.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. Journal of applied crystallography 40, 658-674.

Morselli, M., Pastor, W.A., Montanini, B., Nee, K., Ferrari, R., Fu, K., Bonora, G., Rubbi, L., Clark, A.T., Ottonello, S., 2015. In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. Elife 4, e06205.

Musselman, C.A., Avvakumov, N., Watanabe, R., Abraham, C.G., Lalonde, M.-E., Hong, Z., Allen, C., Roy, S., Nuñez, J.K., Nickoloff, J., 2012. Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. Nature Structural & Molecular Biology 19, 1266-1272.

O'Gara, M., Klimasauskas, S., Roberts, R.J., Cheng, X., 1996. Enzymatic C5-cytosine methylation of DNA: mechanistic implications of new crystal structures for HhaL methyltransferase-DNA-AdoHcy complexes. Journal of molecular biology 261, 634-645.

O'Gara, M., Horton, J.R., Roberts, R.J., Cheng, X., 1998. Structures of HhaI methyltransferase complexed with substrates containing mismatches at the target base. Nature Structural & Molecular Biology 5, 872-877.

Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D., 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448, 714-717.

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

Pei, J., Kim, B.-H., Grishin, N.V., 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic acids research 36, 2295-2300.

Pena, P.V., Davrazou, F., Shi, X., Walter, K.L., Verkhusha, V.V., Gozani, O., Zhao, R., Kutateladze, T.G., 2006. Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. Nature 442, 100-103.

Pierce, B.G., Wiehe, K., Hwang, H., Kim, B.-H., Vreven, T., Weng, Z., 2014. ZDOCK server: interactive docking prediction of protein–protein complexes and symmetric multimers. Bioinformatics 30, 1771-1773.

Pradeepa, M.M., Sutherland, H.G., Ule, J., Grimes, G.R., Bickmore, W.A., 2012. Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. PLoS genetics 8, e1002717.

Purdy, M.M., Holz-Schietinger, C., Reich, N.O., 2010. Identification of a second DNA binding site in human DNA methyltransferase 3A by substrate inhibition and domain deletion. Archives of biochemistry and biophysics 498, 13-22.

Qin, S., Min, J., 2014. Structure and function of the nucleosome-binding PWWP domain. Trends in biochemical sciences 39, 536-547.

Qiu, C., Sawada, K., Zhang, X., Cheng, X., 2002. The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. Nature Structural & Molecular Biology 9, 217-224.

Raveh, B., London, N., Schueler-Furman, O., 2010. Sub-angstrom modeling of complexes between flexible peptides and globular proteins. Proteins: Structure, Function, and Bioinformatics 78, 2029-2040.

Rigbolt, K.T., Prokhorova, T.A., Akimov, V., Henningsen, J., Johansen, P.T., Kratchmarova, I., Kassem, M., Mann, M., Olsen, J.V., Blagoev, B., 2011. System-wide temporal

characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. Sci. Signal. 4, rs3-rs3.

Rohl, C.A., Strauss, C.E., Misura, K.M., Baker, D., 2004. Protein structure prediction using Rosetta. Methods in enzymology 383, 66-93.

Sanulli, S., Justin, N., Teissandier, A., Ancelin, K., Portoso, M., Caron, M., Michaud, A., Lombard, B., Da Rocha, S.T., Offer, J., 2015. Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation. Molecular cell 57, 769-783.

Sarraf, S.A., Stancheva, I., 2004. Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. Molecular cell 15, 595-605.

Shirohzu, H., Kubota, T., Kumazawa, A., Sado, T., Chijiwa, T., Inagaki, K., Suetake, I., Tajima, S., Wakui, K., Miki, Y., 2002. Three novel DNMT3B mutations in Japanese patients with ICF syndrome. American journal of medical genetics 112, 31-37.

Smallwood, S.A., Tomizawa, S.-i., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., Sato, S., Hata, K., Andrews, S.R., Kelsey, G., 2011. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nature genetics 43, 811-814.

Stewart, K.R., Veselovska, L., Kim, J., Huang, J., Saadeh, H., Tomizawa, S.-i., Smallwood, S.A., Chen, T., Kelsey, G., 2015. Dynamic changes in histone modifications precede de novo DNA methylation in oocytes. Genes & development 29, 2449–2462.

Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., Tajima, S., 2004. DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. Journal of Biological Chemistry 279, 27816-27823.

Takeshima, H., Suetake, I., Tajima, S., 2008. Mouse Dnmt3a preferentially methylates linker DNA and is inhibited by histone H1. Journal of molecular biology 383, 810-821.

Tomizawa, S., Nowacka-Woszuk, J., Kelsey, G., 2012. DNA methylation establishment during oocyte growth: mechanisms and significance. Int J Dev Biol 56, 867-875.

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

Velasco, G., Hubé, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., Galvani, A., Viegas-Péquignot, E., Francastel, C., 2010. Dnmt3b recruitment through E2F6

transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. Proceedings of the National Academy of Sciences 107, 9281-9286.

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

Visualizer, D.S., 2013. Release 4.0. Accelrys Software Inc., San Diego, CA, USA.

Wang, J., Qin, S., Li, F., Li, S., Zhang, W., Peng, J., Zhang, Z., Gong, Q., Wu, J., Shi, Y., 2014. Crystal structure of human BS69 Bromo-ZnF-PWWP reveals its role in H3K36me3 nucleosome binding. Cell research 24, 890-893.

Wen, H., Li, Y., Xi, Y., Jiang, S., Stratton, S., Peng, D., Tanaka, K., Ren, Y., Xia, Z., Wu, J., 2014. ZMYND11 links histone H3. 3 K36 trimethylation to transcription elongation and tumor suppression. Nature 508, 263.

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

Yazdi, P.G., Pedersen, B.A., Taylor, J.F., Khattab, O.S., Chen, Y.-H., Chen, Y., Jacobsen, S.E., Wang, P.H., 2015. Nucleosome Organization in Human Embryonic Stem Cells. PloS one 10, e0136314.

Yu, Q., Wen, Z., Chen, Z., Yan, W., Weiwei, W., Jiahai, Z., Zhiyong, Z., Guohong, L., Yunyu, S., Xiaoming, T., 2012. Solution structure of the Pdp1 PWWP domain reveals its unique binding sites for methylated H4K20 and DNA. Biochemical Journal 442, 527-538.

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

3.6 Supplementary Data

Supplementary Table S1: Residues of DNMT3A and DNMT3B involved in the interaction between domains (CD-ADD-PWWP), DNMT3L and epigenetic marks H3K4me0 and H3K36me3. The residues involved in interactions, identified from the literature and in our study, are highlighted in green. These residues between DNMT3A and DNMT3B are highly conserved (highlight in red for identical residues and orange for similar residues).

	DNMT3A residues	DNMT3B residues
DNMT3A-DNMT3L	Arg729	Arg670
(Guo et al., 2014; Jia et al., 2007)	Phe732	Phe673
	Glu733	Glu674
	Tyr735	Tyr676
	Arg771	Arg712
	Glu774	Glu715
DNMT3A-DNMT3A interface	His873	His814
(Guo et al., 2014; Jia et al., 2007)	Asp876	Asp817
	Arg885	Arg826
DNMT3A ADD-CD active form	Asp530	Asp476
(Guo et al., 2014)	Arg556	Arg503
	Arg899	Arg846
	Glu907	Asp854
DNMT3A ADD-H3K4me0	Asp529	Asp475
(Guo et al., 2014)	Asp531	Asp477
	Tyr536	Tyr482
DNMT3B PWWP-H3K36me3	Leu300	Ile233
(our study)	Phe303	Phe236
	Trp306	Trp236
	Trp330	Trp239
	Asp333	Asp266
	Ser337	Ser270

Supplementary Table S2: Intermolecular interactions between DNMT3B PWWP domain and H3K36me3 peptide.

Types	Donor atom	Acceptor atom	Distance (Å)
Electrostatic	M3L36:NZ	ASP266:OD2	4.04
Electrostatic	M3L36:NZ	PHE236 (centroid)	4.28
Electrostatic	M3L36:NZ	TRP239 (centroid of the pyrrole ring)	4.17
Electrostatic	M3L36:NZ	TRP239 (centroid of the benzene ring)	4.30
Electrostatic	M3L36:NZ	TRP263 (centroid of the pyrrole ring)	4.64
Electrostatic	M3L36:NZ	TRP263 (centroid of the benzene ring)	4.98
Hydrogen Bond	W4d:O	M3L36:O	2.59
Hydrogen Bond	PHE269:N	GLY34:O	2.96
Hydrogen Bond	M3L36:N	PHE269:O	2.99
Hydrogen Bond	W1d:O	THR32:O	3.28
Hydrogen Bond	W9d:O	THR32:O	2.66
Hydrophobic	PRO38	ILE233	5.12

Supplementary Table S3: Intermolecular interactions between DNMT3A PWWP domain and H3K36me3 peptide.

Types	H3K36me3	DNMT3A PWWP
Electrostatic	M3L36	ASP253
Hydrogen Bond	GLY34	PHE256
Hydrogen Bond	M3L36	SER257
Hydrogen Bond	M3L36	PHE256
Electrostatic	M3L36	PHE223
Electrostatic	M3L36	TRP226
Electrostatic	M3L36	TRP250
Hydrophobic	PRO38	LEU220

Supplementary Figure S1: Sequence alignment of DNMT3s. *Alignment of Human DNMT3A (uniprotkb:Q9Y6K1), mouse DNMT3A (uniprotkb:O88508), human DNMT3B (uniprotkb:Q9UBC3) and mouse DNMT3B (uniprotkb:O88509). Identical residues are indicated with a red background while similar residues are boxed in blue. This multiple structure and sequence alignments was performed with PROMALS3D (Pei et al., 2008) and formatted with Espript (Gouet et al., 1999).*

***Supplementary Figure S2: Modeling of H3K36me3 peptide-DNMT3A PWWP interactions. (a)** Validation of the procedure: superposition of the native H3K36me3 peptide (yellow) in complex with the DNMT3B PWWP domain (PDB code: 5CIU) with the H3K36me3 peptide model (green) after docking. **(b)** Starting model (purple) and the ten top scoring models (yellow) after docking with the DNMT3A PWWP domain.*
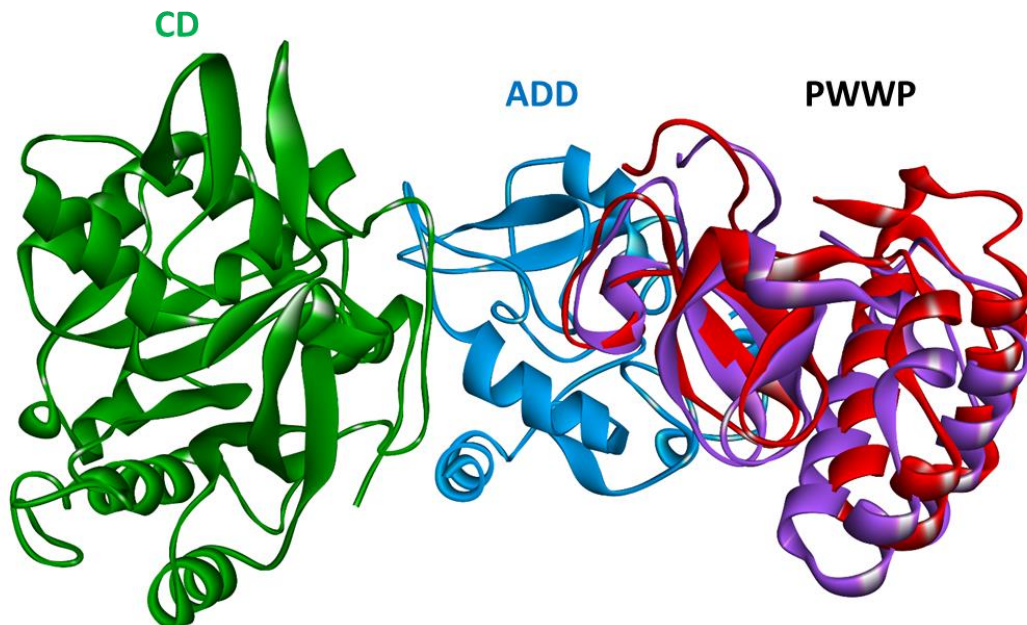
***Supplementary Figure S3: Complete reconstructed DNMT3A (PWWP-ADD-CD domains).** Best solutions of protein-protein docking of Zdock (PWWP in red) and Cluspro (PWWP in purple).*

3.7 Supplemental References

Gouet, P., Courcelle, E., Stuart, D.I., 1999. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics 15, 305-308.

Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., 2014. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature 517, 640-644.

Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., Cheng, X., 2007. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature 449, 248-251.

Pei, J., Kim, B.-H., Grishin, N.V., 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic acids research 36, 2295-2300.

CHAPTER 4

Chapter 4. Targeting PWWP domain of DNA methyltransferase 3B for epigenetic cancer therapy: Identification and structural characterization of new potential protein-protein interaction inhibitors

Grégoire Rondelet*, Thomas Dal Maso, Antonin Maniquet, Quentin Thémans, and Johan Wouters

Department of Chemistry, University of Namur, 61 rue de Bruxelles, B-5000 Namur, Belgium

*Corresponding author: gregoire.rondelet@unamur.be

† In preparation

Personal contribution: Participated in research design, conducted experiments (molecular similarity searching, docking simulations, pharmacophore generation, virtual screening, protein production, protein crystallization, structure determination and docking simulations), performed data analysis and literature search, generated figures and tables, and wrote the manuscript.

Abstract

PWWP domain of *de novo* DNA methyltransferase DNMT3B interacts with both DNA and the permissive epigenetic histone mark H3K36me3 to guide DNA methylation in intragenic regions. In cancer, gene body methylation mediates transcriptional activation of oncoprotein-regulated genes and oncogenes. Therefore, DNMT3B PWWP domain represents a novel target for epigenetic therapy cancer. To identify protein-protein interaction inhibitors of the DNMT3B PWWP-H3K36me3 complex, we performed a similarity-based virtual screening based on the recently solved crystal structures of DNMT3B PWWP domain in complex with H3K36me3 and a bis-tris molecule. The top hits identified were co-crystallized with the DNMT3B PWWP domain in order to gain structural insight into their binding mode. Pharmacophore model was constructed using these complexes and a pharmacophore virtual screening was then performed against ZINC database. Finally, molecular docking simulations were carried out with new hits to identify and explore chemical diversity for development of protein-protein interaction inhibitors of DNMT3B PWWP domain.

4.1 Introduction

Epigenetic modifications control gene regulation through modulation of the chromatin structure without changes in DNA sequence. For example, histone proteins can be modified at their tails to change nucleosomes organization state, and so the gene expression pattern. Another important epigenetic modification regulating gene expression is the DNA methylation. DNA methylation occurs on cytosine bases and is catalyzed by DNA methyltransferases (DNMTs).

In cancer, promoter hypermethylation of tumor suppressor genes conducts to their repression preventing normal cell function and apoptosis induction, if necessary. Inhibition of DNMTs is promising since DNA methylation is a reversible process. For a decade, two nucleoside analogs of cytosine (5-azacytidine (Vidaza®) and decitabine (Dacogen®)) are used as therapeutic treatment for myelodysplastic syndromes and acute myeloid leukemia. They act by covalently trapping DNMTs at their catalytic domain (Christman, 2002; Issa and Kantarjian, 2009; Jones and Taylor, 1980). These treatments lead to demethylation and reactivation of tumor suppressor genes, and, consequently, proliferation arrest and apoptosis of cancer cells (Goll and Bestor, 2005; Mai and Altucci, 2009). Current studies develop non-nucleoside analogs targeting directly the catalytic domain of DNMTs without incorporation into DNA, in contrast to nucleoside analogs, in order to decrease side effects (Erdmann et al., 2015). However, their selectivity towards DNMTs must be further improved.

In the family of the DNA methyltransferases, DNMT3A and DNMT3B have a C-terminal catalytic domain and a N-terminal regulatory part containing PWWP and ADD (ATRX–DNMT3–DNMT3L) domains which interact with nucleosomes (Cheng and Blumenthal, 2008; Goll and Bestor, 2005; Hermann et al., 2004; Jeong et al., 2009). The unique recognition PWWP domain interacts with the extremity of histone 3 tail containing the trimethylated lysine 36 (H3K36me3) epigenetic mark (Baubec et al., 2015; Dhayalan et al., 2010; Rondelet et al., 2016). This result is related with the importance of PWWP domain in the methylation activity of DNMTs on nucleosomal DNA (Chen et al., 2004). Furthermore, the catalytic domain binds to CpG DNA sites and interacts with DNMT3L and other proteins to be recruited on specific genomic regions (Gopalakrishnan et al., 2009; Jia et al., 2007; Jurkowska et al., 2008; Ooi et al., 2007). Indeed, depending on cellular differentiation status, DNMT1, DNMT3A and

DNMT3B have different distribution across histone modification patterns (Jin et al., 2012; Jones, 2012).

Recently, gene body methylation was shown to be associated with pluripotent human embryonic carcinoma and transcriptional activation of oncogenes as ITPKA and genes up-regulated by the oncogenic transcription factor c-Myc (Jin et al., 2012; Yang et al., 2014; Wang et al., 2016). Interestingly, the permissive epigenetic mark H3K36me3 is necessary for gene body methylation and permits the specific recruitment of DNMT3B by association with its PWWP domain (Baubec et al., 2015; Jin et al., 2012; Yang et al., 2014). Therefore, DNMT3B PWWP domain represents a novel target for epigenetic cancer therapy. Research of small-molecule inhibitors targeting this epigenetic complex (DNMT3B PWWP-H3K36me3) will downregulate oncogenic pathways and arrest cancer progression (Yang et al., 2014).

Nowadays, number of detrimental protein-protein interactions (PPIs) have been identified as potential treatment targets for cancer therapy (White et al., 2008). For a while, research in this field has slowly emerged. Indeed, interaction regions in these systems were thought relatively large and represented a high challenge for the development of molecules inhibiting PPIs (Lyne, 2002). But since 2004, targeting PPIs in drug research for cancer therapy was reconsidered with the discovery of small-molecules inhibiting the p53-MDM2 (a p53 negative regulator) interaction (Vassilev et al., 2004). In addition, a study showed that, for most of the PPIs, only certain residues were responsible for the binding affinity in a focus region, called "hot spot" (Lipinski et al., 2012). These hot spots have the size of small organic molecules and are defined as druggable binding sites. Identification of small molecules that disrupt specific PPIs is therefore important for developing anticancer agents. PPIs are attractive drug targets and the proof-of-concept was established in 2010 for epigenetic readers (e.g., PWWP domains) with BET bromodomains (Filippakopoulos et al., 2010).
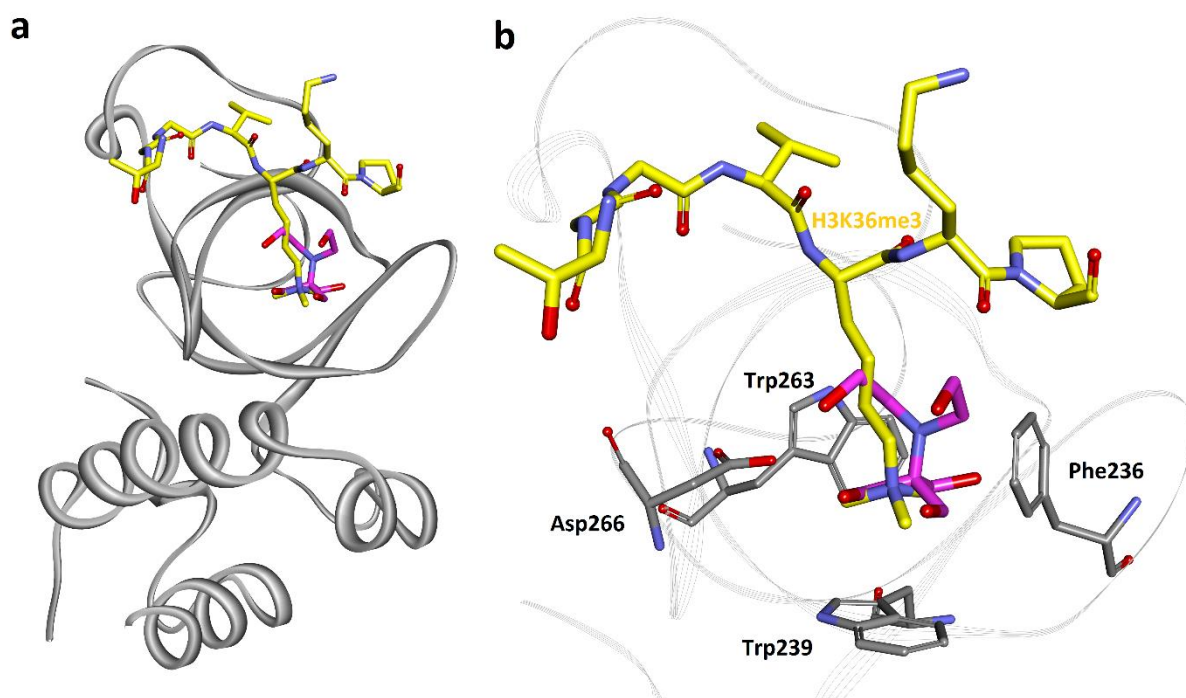
Figure 1. Superimposition of complex structures of DNMT3B PWWP-H3K36me3 (PDB code: 5CIU) with DNMT3B PWWP-bis-tris (PDB code: 3QKJ). Epigenetic mark H3K36me3 and bis-tris compound are shown as yellow and pink stick representation, respectively. (a) Structural overview of the DNMT3B PWWP domain structure (solid Ribbon representation) with the superposition of H3K36me3 peptide and bis-tris compound. (b) Zoom on the aromatic binding site of DNMT3B PWWP domain. Bis-tris molecule is bound at the same position as the trimethyl-ammonium group of epigenetic mark H3K36me3.

The aim of the present work is the identification and characterization of ligands for the PWWP regulatory domain of human DNA methyltransferase 3B. In 2011, a team solved for the first time the structure of the complex of DNMT3B PWWP domain with a ligand, the bis-tris molecule, at a resolution of 2.0 Å (PDB code: 3QKJ) (Wu et al., 2011). The bis-tris binds into the aromatic cage of DNMT3B PWWP at the same position as the trimethyl-ammonium group of lysine of histone 3 (H3K36me3) (PDB code: 5CIU) (Fig. 1) (For the related result, see Chapter 3) (Rondelet et al., 2016). To identify novel potential PPIs, we performed a similarity-based virtual screening using bis-tris molecule and H3K36me3 epigenetic mark as both chemical references. Several crystallographic complexes of DNMT3B PWWP domain with the best hits were obtained and used for building a pharmacophore model. This pharmacophore model was screened against ZINC database. Finally, new identified compounds which fit the pharmacophore model were subject to docking simulations in order to discover more potent

small orthosteric inhibitors showing chemical diversity. The flowchart for identification of new ligands of DNMT3B PWWP domain is shown in Figure 2. These compounds could inhibit the protein-protein interaction between DNMT3B PWWP domain and the H3K36me3 epigenetic mark.

4.2 Results and discussion

Structural studies of DNMT3B PWWP domain revealed a small contact surface between binding site and H3K36me3 epigenetic mark (see chap. 2) (Rondelet et al., 2016). The key residue of the peptide is the trimethylated Lys36 which is inserted into the concave aromatic surface formed by three aromatic residues (Phe236, Trp239 and Trp263) (Fig. 1) and forms van der Waals and π-cation interactions with these residues and electrostatic–cation interaction with the carboxylate group of Asp266. DNMT3B PWWP domain binds this short peptide and could therefore bind small molecules as shown for bromodomains (Cierpicki and Grembecka, 2015; Filippakopoulos and Knapp, 2014). The key hot spot is localized in the conserved aromatic cage where the bis-tris molecule and trimethyl-ammonium group of lysine of histone 3 (H3K36me3) are bound. To identify new potential protein-protein inhibitors of this epigenetic complex, we identified, firstly, lead compounds by similarity-based virtual screening using the bis-tris compound and H3K36me3 epigenetic mark as both chemical references.
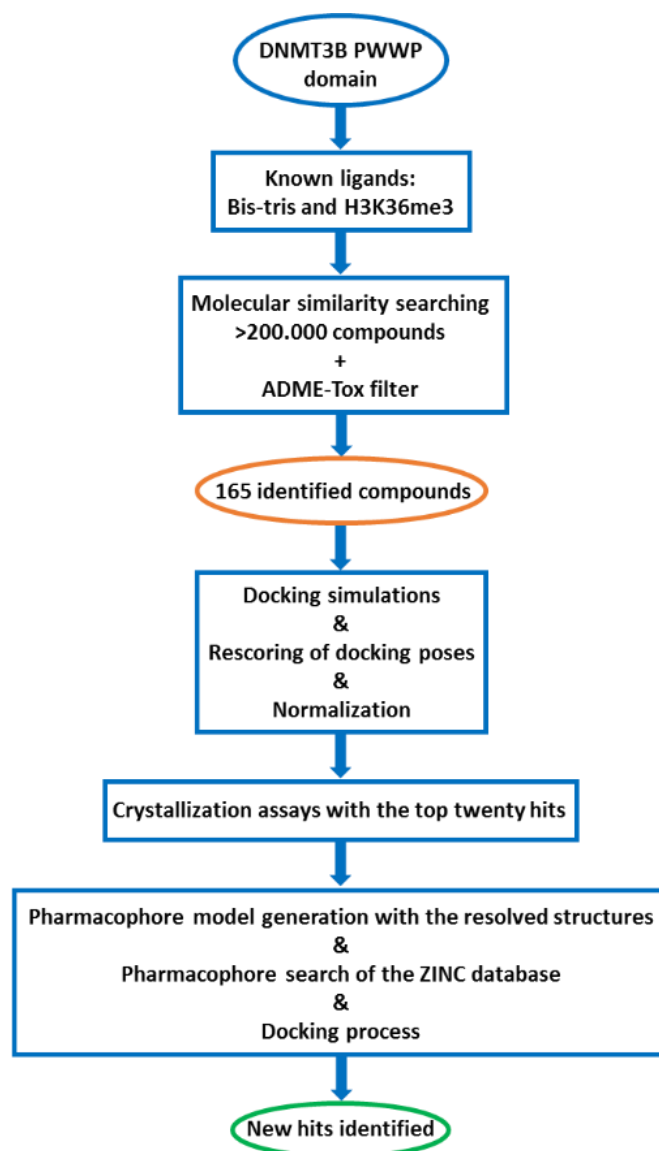
Figure 2. Overall workflow for the identification of potential protein-protein interaction inhibitors of DNMT3B PWWP domain.

4.2.1 Molecular similarity searching - libraries of drug candidates

The similar property principle (SPP) in drug discovery postulates that structurally similar compounds are assumed to exhibit similar physicochemical properties and/or biological activities (Zhang et al., 2010). In addition, molecular similarity is one of the most used concepts in computer-aided drug design (Goll and Bestor, 2005). There are many types of structural similarity search procedures presented in the literature (Gopalakrishnan et al., 2009; Ooi et al., 2007; Otani et al., 2009), but the method of choice for computing the similarity between a query structure and each molecule of the database is based on the 2D

fingerprints. This similarity measure includes binary fingerprints for the characterization of the molecular structures being compared, a weighting of representation features. So, molecular similarity searching of the commercial Sigma-Aldrich library (~200.000 compounds) was carried out using 2D fingerprints and the Tanimoto coefficient (threshold = 0.6) (Fig. S1) to identify new compounds structurally similar to the bis-tris compound and trimethyl-ammonium group of lysine of histone. Molecules in the database were ranked according to their computed similarity to both query molecules. We identified for bis-tris and trimethyl-ammonium group, 163 and 133 hits respectively.

4.2.2 ADMET prediction

After identification of similar compounds using the 2D method and prior to in silico screening stage of these compounds into the receptor-binding site of DNMT3B PWWP domain, an ADME-Tox (Absorption, Distribution, Metabolism, Excretion and Toxicity) filter was applied to the selected library with the FAF-Drugs3 tool (Lagorce et al., 2008a). These properties are considered and evaluated in early stage of drug discovery process to discard compounds with poor ADME properties and containing toxic groups (Lyne, 2002; Selick et al., 2002). Indeed, compounds with inadequate properties increase development costs, prolong development timelines and poorly water-soluble drugs need to be administered to patients at a higher dose and/or intravenously (Kerns and Di, 2010).

This evaluation was performed based on Lipinski rule which takes into account these properties: the number of H-bond donors (HBD) and acceptors (HBA), the molecular weight (MW) and the logP (Lipinski et al., 2012). Compounds properties are evaluated to classify those with a lead-like or drug-like profile by applying "rule of 5" (Lipinski et al., 2012; Lyne, 2002; Oprea et al., 2001; Teague et al., 1999). Since then, different additional rules were proposed by combination with the original "rule of 5" (Veber et al., 2002). In the present study, both Drug-Like Soft and Lead-Like Soft filters (Table S1) available in FAF-Drugs2 were used. The Drug-Like Soft filter was established by analyzing the descriptors values of 916 FDA oral drugs. The Lead-Like Soft filter thresholds were combined partially with the Drug-Like Soft filter and properties (e.g., molecular weight, logP, HBD and HBA) to allow addition of moieties in order to increase the affinity without decreasing the proper ADMET properties. After application of these filters, we found 83 drug-like compounds and 29 lead-like

compounds similar to bis-tris molecule, and 39 drug-like compounds and 14 lead-like compounds similar to trimethyl-ammonium group of lysine of histone.

4.2.3 Docking screen against DNMT3B PWWP domain

In order to validate our molecular modeling procedure with GOLD Suite software (v5.2.2, CCDC, Cambridge, UK), docking of the bis-tris into the binding site of DNMT3B PWWP domain (PDB code: 3QKJ) was performed. Docking results show that bis-tris exhibits similar interactions as those observed experimentally in the crystal structure (Fig. S2). In addition, the crystallographic structures were aligned with the best pose prediction to predict the root-mean-square deviation of heavy atom positions. Results was 0.98 Å RMSD for DNMT3B PWWP domain (3QKJ) which is much less than 2.0 Å, the usually accepted threshold for a docking accuracy success (Fig. S2) (Gohlke et al., 2000). We will use this method for the docking screening study.

The virtual screening allowed the identification of new compounds similar to bis-tris (from 101 to 292 MW, bis-tris 209) and trimethyl lysine group (from 101 to 442 MW). As the bis-tris and trimethyl lysine compounds are the starting lead-like compounds, we want to measure the drug-likeness of the new identified compounds, their potential to become a drug. This approach is dependent on the interpretation of the ranking scores. If the compounds having a different molecular weight are classified according to their dock scores, the selection of high MW compounds will be favored as the contribution of the internal energy to the total dock energy score dominates. Moreover, the energy score includes additionally the energy of interaction, sum of the van der Waals energy and electrostatic energy, between the ligand and the binding site, so the bigger the molecule is the more interactions it may form. In order to eliminate molecular size bias in the final ranking, one of the strategies of energy score normalization is based on the number of heavy atoms (non-hydrogen atoms) in the ligands (Pan et al., 2003). So, the score of the best pose for each ligand is divided by the number of heavy atoms, N, and leads to selected compounds with molecular weight distributions lower (≤ 300 daltons) than the original database with similarly good interactions. These compounds are potentially the best candidates selected as lead-like compounds. Indeed, this size-normalization procedure is motivated by a search for molecules with low molecular weight having an enhanced absorption rate and being more suitable for structural optimization. As

the aim of our study is the identification of drug-like compounds, the normalization was performed by dividing energy score of rescoring (Table 1) for each ligand by the cube root of the number of non-hydrogen atoms, $N^{1/3}$ (based on empirical considerations) (Oprea et al., 2001; Pan et al., 2003). Rescoring of docked ligands has a favourable impact on the overall rank ordering of ligands in virtual screening. We explored the chemical space defined by protein-protein interaction complex of DNMT3B PWWP domain with H3K36me3 with screening of small analogs of both bis-tris molecule and H3K36me3 epigenetic mark (Table 1 and Fig. 3).

Among all the analogs, 6-dipropylamino-1-hexanol showed the best chemscore (29.87) and binding energy (-30.96 kcal.mol[-1]) values, even after normalization (Table 1). This approach gives multiple hits presenting a better affinity for DNMT3B PWWP domain than the chemical references as exhibited by the lowest binding energy of bis-tris (-14.88 kcal.mol[-1]).
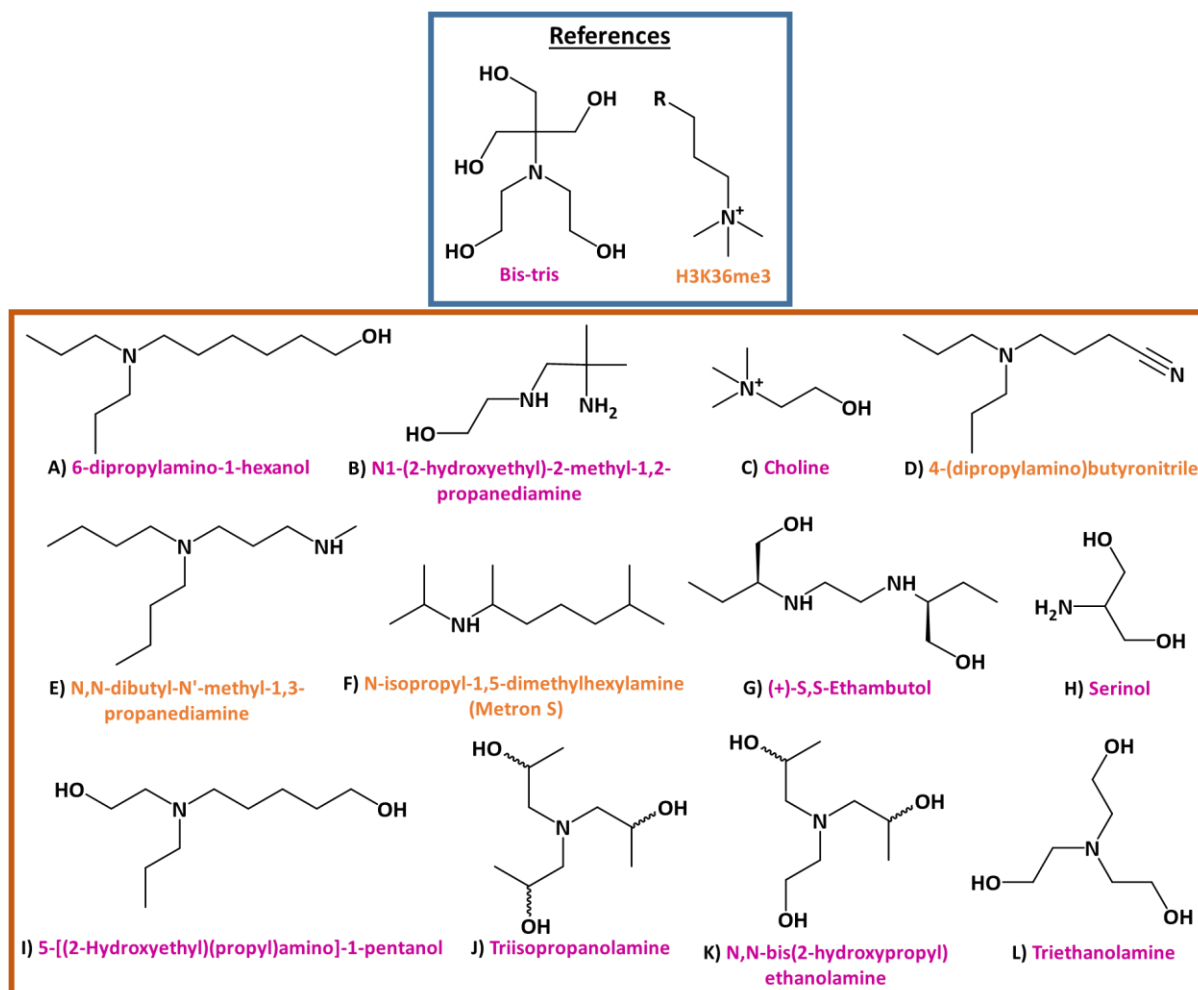
Figure 3. Identified compounds after similarity-based virtual screening. *bis-tris compound (pink) and H3K36me3 epigenetic mark (orange) defined as both chemical references (blue top box). Structure of analogs (orange bottom box) identified after virtual screening against DNMT3B PWWP domain. Bis-tris and H3K36me3 epigenetic mark analogs are represented in pink and orange, respectively.*

Table 1. Physicochemical descriptors and normalized scores of virtual screening hits. *Bis-tris and H3K36me3 epigenetic mark analogs are represented in pink and orange, respectively.*

ID	Name	MW (g.mol^{-1})	logP	logD	tPSA (Å^2)	Heavy Atoms (N)	Solub. pH 7.4 (mg/ml)	Rescoring (ChemScore)	ΔG (kcal/mol)	Normalized Score (ChemScore/N$^{1/3}$)
A	6-dipropylamino-1-hexanol	202.36	2.80	-0.42	24.67	14	22.45	29.87	-30.96	12.39
B	N1-(2-hydroxyethyl)-2-methyl-1,2-propanediamine	134.22	-1.21	-3.82	64.48	9	191.26	25.35	-25.51	12.19
C	Choline	104.17	-3.70	-4.66	20.00	7	500.00	21.76	-22.00	11.38
D	4-(dipropylamino)butyronitrile	169.29	2.10	-0.28	28.23	12	29.40	25.45	-26.05	11.12
E	N,N-dibutyl-N'-methyl-1,3-propanediamine	202.38	2.69	-1.74	21.05	14	24.06	26.51	-27.89	11.00
F	N-isopropyl-1,5-dimethylhexylamine (Metron S)	172.33	3.62	0.49	16.61	12	10.55	24.91	-25.21	10.88
G	(+)-S,S-Ethambutol	205.32	-0.08	-2.25	69.10	14	128.51	25.83	-27.68	10.72
H	Serinol	92.12	-2.02	-3.76	68.10	6	248.77	18.53	-18.57	10.20
I	5-[(2-Hydroxyethyl)(propyl)amino]-1-pentanol	190.30	0.86	-1.85	44.90	13	72.31	23.8	-24.38	10.12
J	Triisopropanolamine	192.28	-0.49	-2.50	65.13	13	138.59	21.81	-22.15	9.28
K	N,N-bis(2-hydroxypropyl)ethanolamine	178.25	-0.92	-2.64	65.13	12	183.76	19.57	-19.88	8.55
L	Triethanolamine	149.19	-1.11	-2.96	64.00	10	149.00	17.76	-18.22	8.24
	Bis-tris	210.25	-3.27	-4.04	105.59	14	891.45	13.39	-14.88	5.56

4.2.4 Structural insights into binding of bis-tris and H3K36me3 analogs to DNMT3B PWWP domain

Crystallographic complexes were obtained to gain structural insight into the binding mode of new identified compounds to DNTM3B PWWP domain. These complexes (Figs. 4 and 5) revealed the canonical anti-parallel β-barrel-like fold that we retrieved for other PWWP domains (Wen et al., 2014). However, the different complexes highlighted a conformational flexibility of the binding site and this effect is really marked compared to the apo-structure (PDB code: 3FLG) (Fig. 6). Upon binding of ligands, the closure of the binding site is initiated by conformational change of the Phe236 side-chain (displacement of the phenyl ring centroid of ~11 Å) accompanied by backbone rearrangements. This change reveals a flexibility of the loop situated between the β1 and β2 strand. Flexibility of the Phe236 is also observed when comparing different complexes. The complex formed with the triisopropylamine (Fig. 5j), one of the most crowded amine compound, showed the larger aromatic cavity with Phe236-ring centroid displacement of 1.65 Å compared to the 6-dipropylamino-1-hexanol complex (the smallest cavity) (Fig. 5a). This conformational adaptation reveals the importance to this induced-fit mechanism to enhance recognition of analogs.

In all complexes, the analogs have similar binding modes to DNMT3B PWWP domain. All structures were refined between 1.6 and 2.6 Å and the statistics are summarised in Table S2. Complexes with bis-tris analogs show that these compounds mimic scaffold of H3K36me3 epigenetic mark as exemplified by choline. This compound (Fig. 5c) engages van der Waals and electrostatic interactions as observed for H3K36me3 peptide and makes an additional hydrogen bond interaction with Asp266. Interestingly, the supplementation of choline in mice induced a reduction of hypermethylation in the hippocampus and prefrontal cortex after alcohol exposure (Otero et al., 2012) and a deficiency of choline alters brain development (Niculescu et al., 2006). Choline is a methyl group donor like folate, methionine, and vitamin B12 and its deficiency is associated with carcinoma in rats (Newberne and Rogers, 1986). These results show that choline impacts the epigenome depending on the cellular context and our solved structure may explain partially the link with the regulation of DNA methylation. Furthermore, we identified the FDA-approved drug ethambutol (Fig. 5g) (EMB) used in tritherapy (rifampicin, isoniazid, ethambutol) against *Mycobacterium tuberculosis* (*Mtb*).

Ethambutol inhibits arabinosyl transferase and prevents the cell-wall biosynthesis of *Mtb.* However, its anti-cancer activity has to be demonstrated. The antihistamine agent, Metron S (Fig. 5f), was also identified.
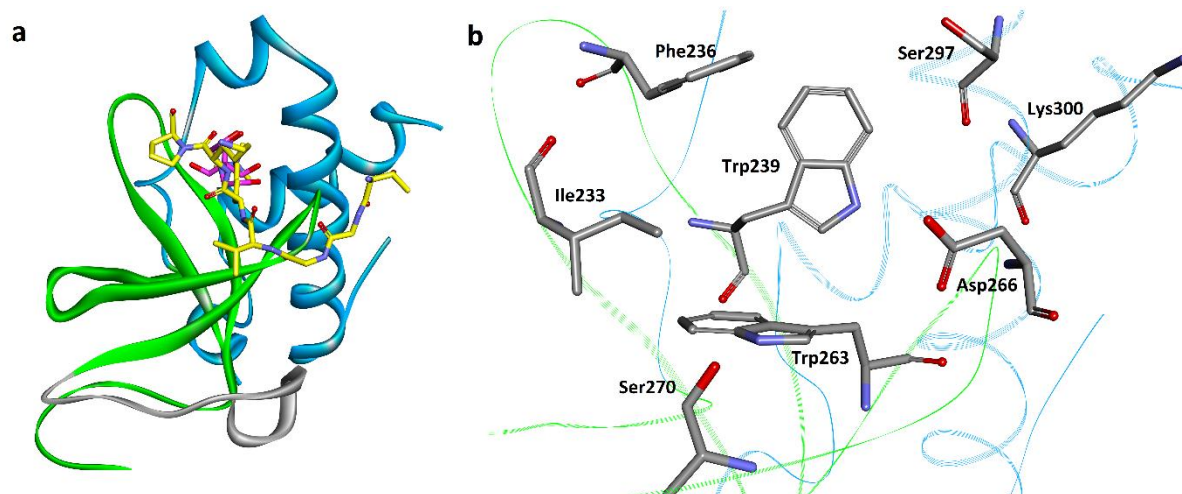
Figure 4. Structure of the DNMT3B PWWP domain. (a) Structural overview of the DNMT3B PWWP domain structure with the superposition of H3K36me3 peptide (yellow) and bis-tris compound (pink). The PWWP domain contains an anti-parallel β-barrel-like fold (green) with a short 3_{10} helix (gray) and a helix bundle of α-helices (blue). (b) Zoom on the aromatic cage of DNMT3B PWWP domain. The main residues involved in the interaction with bis-tris and H3K36me3 analogs (Fig. 5) are labeled and shown in stick representation.

Figure 5. *Molecular binding mode of identified ligands in the binding site of DNMT3B PWWP domain. (A-L) Detailed view of crystallographic complexes obtained at 1.6-2.6 Å resolution. These small molecules bind at PPI interfaces. In the top figure, binding site of PWWP domain binding site is depicted as a transparent white surface and a 2Fo-Fc electron density map was contoured at 1.0 σ for each*

ligand. The ligands are well defined in the electron density (blue mesh). The bis-tris analogs are shown in pink stick model and H3K36me3 analogs in 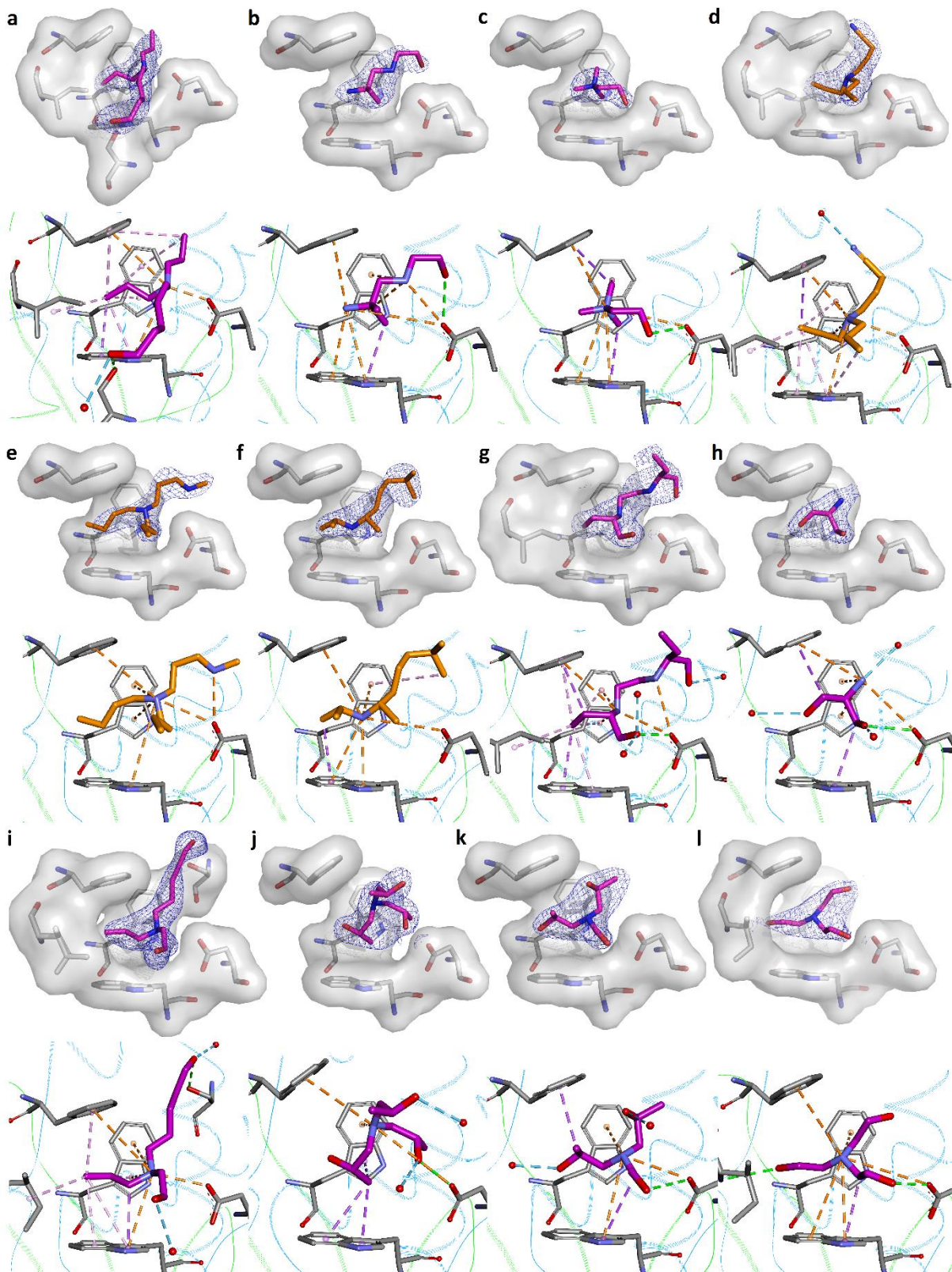orange stick model. For the bottom figure, molecular interactions are shown in dotted lines with the following color codes: blue for water hydrogen bond, orange for π-cation interaction, green for conventional hydrogen bond, purple for π-sigma interaction, light pink for π-alkyl interaction and pink for π-π T-shaped interaction. DNMT3B PWWP domain is represented as line ribbon with interacting residues in grey stick representation.
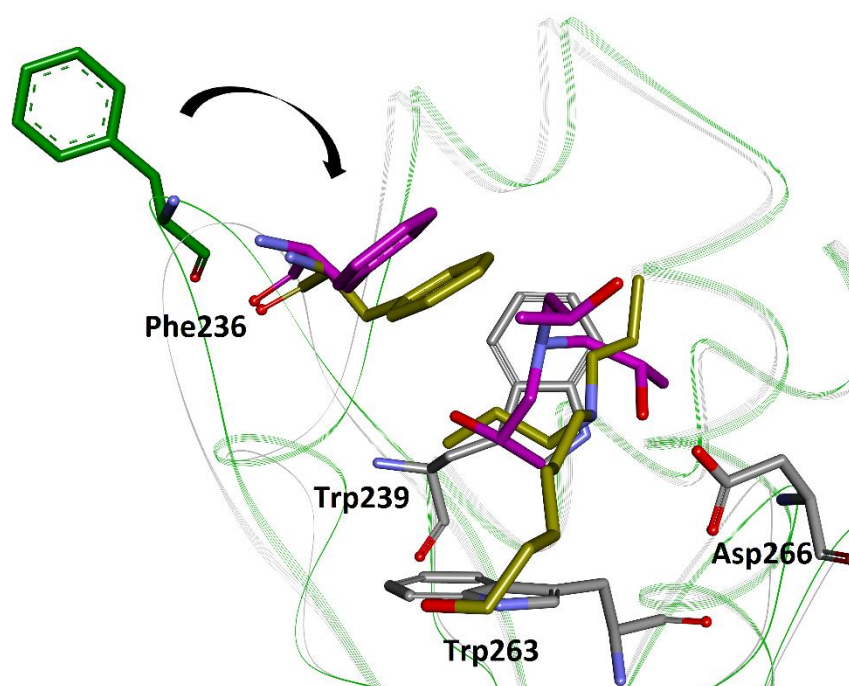
Figure 6. Superimposition of the apo structure of PWWP domain (green) and 6-dipropylamino-1-hexanol (kaki), triisopropanolamine (pink) ligand complexes.

4.2.5 Pharmacophore model generation and pharmacophore-based virtual screening

In order to increase the chemical diversity and identify new ligands with high binding affinity towards DNTM3B PWWP domain, pharmacophore generation and pharmacophore-based search were performed. The pharmacophore model (Fig. 7b) was created with LigandScout (Wolber and Langer, 2005) using the solved structures of complexes with both analogs of bis-tris and H3K36me3 peptide. Pharmacophore identified from each complex were merged into one model (Fig. 7a). The final model contains two positive-ionic centers, two lipophilic parts, four hydrogen bond acceptors and five hydrogen bond donors (Fig. 7b). The resulting pharmacophore model was screened against ZINC database which contains over 35 million

compounds. ZINCPharmer software (http://zincpharmer.csb.pitt.edu) was used for the pharmacophore research. The hits identified were subjected to ADME-Tox prediction using FAF-Drugs3 (Lagorce et al., 2015) prior to docking simulations.

Figure 7. Pharmacophore model generation of DNMT3B PWWP ligands. *Pharmacophore sites are colored red for H-bond acceptors, green for H-bond donors, blue for positively io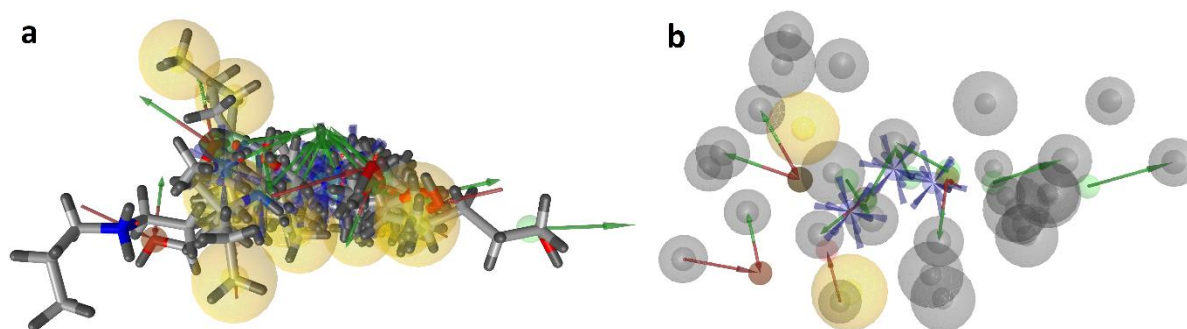nizable groups and yellow for hydrophobic groups. (**a**) Aligned ligands from crystallographic complexes with pharmacophore features. (**b**) Pharmacophore model generation based on common chemical features. Excluded volumes (ligand-inaccessible) of residues forming the binding site are represented as gray spheres.*

4.2.6 Binding mode analysis of new hits

Compounds that fit the pharmacophore model (61 compounds) were further docked into the binding site of DNMT3B PWWP domain to gain insights into their binding mode. These compounds share some structural chemical similarities with small difference in physical properties as the logP and the solubility (Fig. 8 and Table 2).

Table 2. Physicochemical descriptors and normalized scores of virtual screening hits. *logD at pH 7.4 were calculated for both compounds using MarvinSketch v.15.3.2 (ChemAxon Kft., Budapest, Hungary, www.chemaxon.com).*

| ID	Name (ZINC database)	MW (g.mol⁻¹)	logP	logD	tPSA (Å²)	Heavy Atoms (N)	Solub. pH 7.4 (mg/ml)	ΔG (kcal/mol)	Normalized Score (ChemScore/ N ^(1/3))
A	ZINC69436570	316.45	1.65	-3.29	70	23	30.36	-31.53	10.30
B	ZINC72429462	322.48	0.41	-3.03	80	23	29.78	-30.02	10.25
C	ZINC77263678	397.46	-0.68	-0.36	120	29	45.58	-33.40	10.01
D	ZINC38689942	362.37	2.10	1.17	103	27	11.46	-31.53	9.84
E	ZINC73562089	350.51	2.15	-0.68	61	25	26.08	-28.14	9.35
F	ZINC86715539	370.50	1.74	-1.77	84	27	21.01	-27.99	9.22
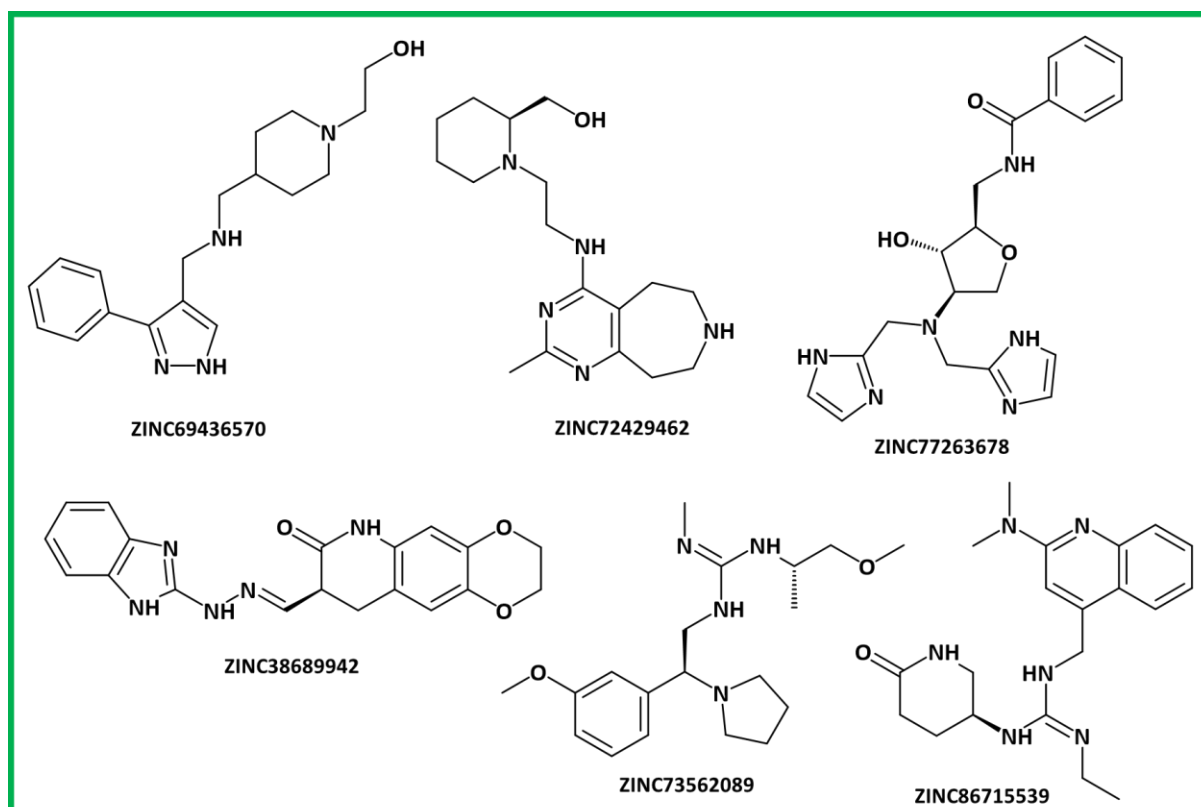
Figure 8. Identified hits obtained after pharmacophore model generation and pharmacophore-based virtual screening. *The six best ranked compounds after simulations docking are represented.*

A detailed binding mode analysis of the three best compounds (Fig. 9) presenting novel chemical features is described in detail (ZINC IDs: ZINC69436570 (2-[4-[[(3-phenyl-1H-pyrazol-4-yl)methylamino]methyl]-1-piperidyl]ethanol), ZINC72429462 ([[(2S)-1-[2-[[(2-methyl-6,7,8,9-tetrahydro-5H-pyrimido[4,5-d]azepin-4-yl)amino]ethyl]-2-piperidyl]meth) and ZINC77263678 (N-[[(2R,3S,4R)-4-[bis(1H-imidazol-2-ylmethyl)amino]-3-hydroxy-tetrahydrofuran-2-yl]methyl]benzamide).

Their simulated modes of binding are quite similar as both phenyl group of ZINC69436570 and ZINC77263678 forming T-shaped π–π interactions with all aromatic residues Phe236, Trp239 and Trp263 from the binding site (Fig. 9a and c). The methoxy group carried by the piperidyl moety of ZINC69436570 and ZINC72429462 compounds interacts by hydrogen bond with the carbonyl group of the backbone with an additional interaction with the Asp266 for ZINC69436570 compound (Fig. 9a and b). Additionally, pyrazol group of ZINC69436570 engaged hydrogen bonding network with residues Trp263, Phe269 and Ser270. Two salt bridges are also observed between both positively charged amine groups and the residue

Asp266. This compound exhibited a binding energy of -31.53 kcal.mol^{-1}. For the ZINC72429462 (Fig. 9b), the methyl 2-Pyrimidine interacts by parallel π-π stacking interaction with residue Trp239 and T-shaped π–π interaction with residue Phe236. The positively charged amine of the azepin group forms a pi-cation interaction with residue Trp263. In addition to the previously described interactions, ZINC77263678 (Fig. 9c) forms multiple hydrogen bonds with residues Trp263, Phe296 and Ser270. This compound has the lowest binding energy (-33.40 kcal.mol^{-1}). *In fine*, these complexes form intermolecular interfaces of high shape complementarity.
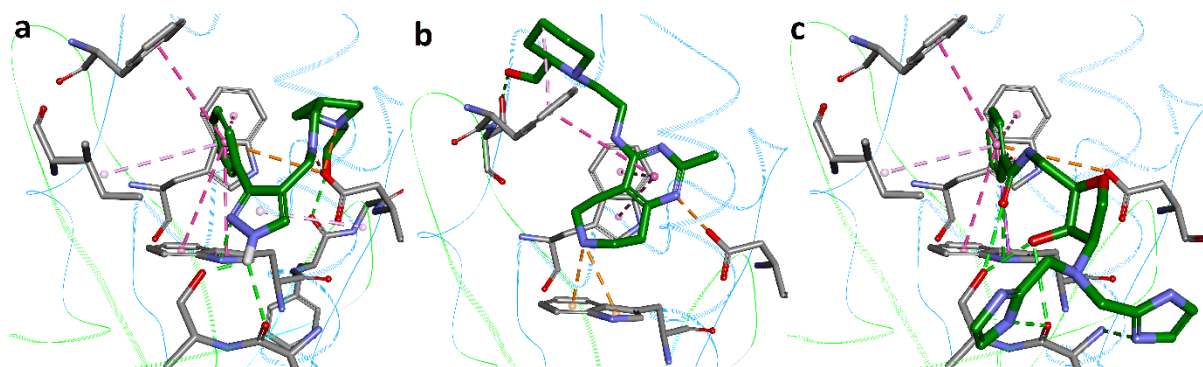
Figure 9. Binding modes of identified potential PPIs to DNMT3B PWWP domain from the pharmacophore-based virtual screening. a) ZINC69436570; b) ZINC72429462; c) ZINC77263678. Compounds are shown in green stick representation and DNMT3B PWWP domain is represented as line ribbon with interacting residues in grey stick representation.

4.3 Conclusion

The research of small drug-like molecules modulating protein-protein interactions of epigenetic "readers" is quite recent. Here we developed a workflow to identify new ligands for the DNMT3B PWWP domain. The similarity-based virtual screening approach gives multiple hits after virtual screening and allowed us to obtain complexes with DNMT3B PWWP domain. All the complexes structures led to the construction of a pharmacophore model for the discovery of novel and potential inhibitors by virtual screening. This conducts to the identification of the structural requirements for binding of ligands to DNMT3B PWWP domain and a first step for the development of protein-protein interaction inhibitors. The small molecules identified may modulate protein-protein interactions between DNMT3B PWWP domain and H3K36me3 epigenetic mark and downregulate expression of oncogenes (e.g., ITPKA) and oncoprotein-regulated genes (e.g., c-Myc).

4.4 Materials and methods

4.4.1 ADMET prediction

The compounds were exported in standard structure-data file (SDFile; SDF) after similarity research with Tanimoto coefficient (threshold = 0.6) (Fig. S1). Prior to in silico screening, ADME-Tox (absorption, distribution, metabolism, excretion and Toxicity) properties of compounds identified from similarity research and the pharmacophore screening were predicted and studied using FAF-Drugs2 (Lagorce et al., 2008b) and FAF-Drugs3 (Free ADME-Tox Filtering Tool) (Lagorce et al., 2015). Pre-defined filters "Drug-Like Soft" and "Lead-Like Soft" were used (Table S1) to select the best candidates for further evaluation as PPI inhibitors.

4.4.2 Molecular docking

Ligand preparation

SDF file containing chemical description of molecules was imported into Discovery Studio 3.5 and converted into 3D representations. Ligands were prepared by using the "Prepare Ligands" tool in Discovery Studio protocols. The pH was set to 7.4 to calculate the degree of ionization of compounds and the other parameters were turned to "False". A minimization procedure was applied to the molecules with MMFF (Merck molecular force field) force field (Halgren, 1996) and the conjugate gradient algorithm with a convergence criteria of 0.01 kcal.mol^{-1}.Å^{-1}.

Protein preparation

Crystallographic structure of DNMT3B PWWP domain (PDB code: 3QKJ) in complex with bis-tris was used. The three dimensional structure was visualized with Discovery Studio 4.0. (Visualizer, 2013). Conserved water molecule HOH-362 from chain A was kept in the binding pocket and the bis-tris ligand was extracted. In order to explore all the binding poses into DNMT3B PWWP domain, the binding site was defined as a sphere of 13 Å radius centered on the bis-tris molecule (Fig. S2). Protein was prepared for docking by using the "Prepare Protein" tool in Discovery Studio protocols. This procedure permits to adjust the pH to 7.4 in

order to generate protonation states of proper residues, add missing hydrogens, fixe the missing side chains and assign atom types (corrected connectivity and bond orders) by applying CHARMm (Chemistry at HARvard Macromolecular mechanics) force field (version c36b2) (Brooks et al., 1983).

Docking Simulations

Number of genetic algorithm (GA) iterations was changed to 100 and 20 docking poses were generated for every ligand using GOLD Suite software (v5.2.2, CCDC, Cambridge, UK) and scored with ChemPLP (Korb et al., 2009) scoring function. ChemPLP is one of the best scoring functions at both pose prediction and virtual screening performance of GOLD (top ranked success rate = 81%). It contains the Chemscore hydrogen bonding term and different linear potentials to model van der Waals and repulsive terms. Rescoring of docking poses was performed using ChemScore scoring function. The option "receptor depth scaling" was checked to increase score for hydrogen bonds located deep into the binding site. Prior to this stage, a local optimization was carried out to minimize the docked ligand into the binding site. In each cluster of twenty random conformations generated, the top ranked binding poses were selected for visual inspection and analysed to select the most representative conformation with the highest score. As a final step, normalization of docking scores (ChemScore) was performed by dividing energy score for each ligand by the cube root of the number of non-hydrogen atoms, $N^{1/3}$.

4.4.3 Over$x$pression and purification of human DNMT3B PWWP domain

Overexpression of human DNMT3B PWWP domain was done in *E. coli* strain Rosetta™ 2 (DE3) competent cells (Novagen®) containing the plasmid obtained from Addgene (plasmid 32044, C. Arrowsmith). The protein was produced and purified as previously described (Rondelet et al., 2016; Wu et al., 2011).

4.4.4 Crystallization

Purified PWWP domain of DNMT3B was used for crystallization trials. Co-crystallization trials were performed using the Hampton Research Crystal Screen kits 1–2 and the sitting-drop vapor-diffusion method in 96-well plates at room temperature. Drops consisted of 1 μl

of DNMT3B PWWP domain at 22 mg/ml – ligand (1:2 M ratio) plus 1 μl of reservoir solution equilibrated against a reservoir volume of 50 μl. Complex crystals of DNMT3B PWWP domain and 5-[(2-Hydroxyethyl)(propyl)amino]-1-pentanol were grown against a reservoir consisting of 0.1 M HEPES sodium pH 7.5, 2% v/v Polyethylene glycol 400, 2.0 M Ammonium sulfate (Hampton Research Crystal Screen 1 No. 39).

To obtain crystals for soaking, conditions were adapted from previous research (Wu et al., 2011) using sitting drop vapour-diffusion method at room temperature. Sitting drop was prepared by mixing 2 μl of the protein solution (22 mg/mL) with 2 μl of the reservoir solution containing 0.1 M Mes (pH 5.7 to 6.7), 0.2 M Li_2SO_4, and 23 to 33% PEG 3350 in a 24-well plate.

4.4.5 Crystal soaking

For crystal soaking trials, DNMT3B PWWP domain crystals were introduced into a 2 μl fresh drop of soaking solution containing 0.1 M Mes (pH 5.7 to 6.7), 0.2 M Li_2SO_4, 23 to 33% PEG 3350 and 25mM ligand for 10-30 min before being flash-cooled in liquid nitrogen.

4.4.6 Structure determination

Single crystals and data sets were collected at SOLEIL Synchrotron, Gif-sur-Yvette, France on beamline PROXIMA 1 (PX1) and PROXIMA 2 (PX2-A) using PILATUS 6M detector at a wavelength of 0.97857 Å for PX1 and an ADSC Q315r detector at a wavelength of 0.9801 Å for PX2-A. The data were processed using XDS/XSCALE (Kabsch, 2010). A summary of the data-collection and refinement statistics is presented in Table S2. Initial phases were calculated by molecular replacement using the program PHASER in PHENIX (Adams et al., 2010 and McCoy et al., 2007) with experimental reflection data set (Fo) and the search model (Fc) solved at 2.0 Å resolution (PDB code: 3QKJ) (Wu et al., 2011). The complex was built using the Coot program (Emsley et al., 2010) and refined with the program PHENIX (Adams et al., 2010). Figures were drawn using both Pymol (DeLano, 2002) and Discovery Studio (BIOVIA, 2015).

4.4.7 Generation of pharmacophore model and Pharmacophore-based virtual screening

The 3D pharmacophore model was built using LigandScout v.4.09.2 (Wolber and Langer, 2005). Crystal structures of protein-ligand complexes solved in this study were used to generate different pharmacophore models. Features pharmacophore identified were merged into a single pharmacophore model. The pharmacophore model was used for virtual screening against the ZINC database using the free pharmacophore search software ZINCPharmer (Koes and Camacho, 2012). Some unrepresentative features (not located in the same spatial position) were deleted to increase identified hits. The final hits from screening were filtered using FAF-Drugs3.

4.4.8 Structure-Based Virtual Screening

Structure-based virtual screening was carried out using compounds identified from pharmacophore virtual screening. As previously described, GOLD Suite software (v5.2.2, CCDC, Cambridge, UK) and poses were scored with ChemPLP (Korb et al., 2009) scoring function. Rescoring of docking poses was performed using ChemScore scoring function. The top ranked binding poses were selected for visual inspection and analysis to select the most representative conformation with the highest score. Finally, normalization of docking scores (ChemScore) was performed by dividing energy score for each ligand by the cube root of the number of non-hydrogen atoms, $N^{1/3}$.

4.5 References

Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., Schubeler, D., 2015. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature 520, 243-247.

BIOVIA, D.S., 2015. Discovery studio modeling environment. Dassault Systèmes: San Diego, CA, USA.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M., 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. Journal of computational chemistry 4, 187-217.

Chen, T., Tsujimoto, N., Li, E., 2004. The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. Molecular and cellular biology 24, 9048-9058.

Cheng, X., Blumenthal, R.M., 2008. Mammalian DNA methyltransferases: a structural perspective. Structure 16, 341-350.

Christman, J.K., 2002. 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. Oncogene 21, 5483-5495.

Cierpicki, T., Grembecka, J., 2015. Targeting protein–protein interactions in hematologic malignancies: still a challenge or a great opportunity for future therapies? Immunological reviews 263, 279-301.

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

Erdmann, A., Halby, L., Fahy, J., Arimondo, P.B., 2015. Targeting DNA Methylation with Small Molecules: What's Next? Miniperspective. Journal of medicinal chemistry 58, 2569-2583.

Filippakopoulos, P., Knapp, S., 2014. Targeting bromodomains: epigenetic readers of lysine acetylation. Nature reviews Drug discovery 13, 337-356.

Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W.B., Fedorov, O., Morse, E.M., Keates, T., Hickman, T.T., Felletar, I., 2010. Selective inhibition of BET bromodomains. Nature 468, 1067-1073.

Gohlke, H., Hendlich, M., Klebe, G., 2000. Knowledge-based scoring function to predict protein-ligand interactions. Journal of molecular biology 295, 337-356.

Goll, M.G., Bestor, T.H., 2005. Eukaryotic cytosine methyltransferases. Annu. Rev. Biochem. 74, 481-514.

Gopalakrishnan, S., Sullivan, B.A., Trazzi, S., Della Valle, G., Robertson, K.D., 2009. DNMT3B interacts with constitutive centromere protein CENP-C to modulate DNA methylation and the histone code at centromeric regions. Human molecular genetics 18, 3178-3193.

Halgren, T.A., 1996. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. Journal of computational chemistry 17, 490-519.

Hermann, A., Gowher, H., Jeltsch, A., 2004. Biochemistry and biology of mammalian DNA methyltransferases. Cellular and Molecular Life Sciences CMLS 61, 2571-2587.

Issa, J.-P.J., Kantarjian, H.M., 2009. Targeting DNA methylation. Clinical Cancer Research 15, 3938-3946.

Jeong, S., Liang, G., Sharma, S., Lin, J.C., Choi, S.H., Han, H., Yoo, C.B., Egger, G., Yang, A.S., Jones, P.A., 2009. Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA. Molecular and cellular biology 29, 5366-5376.

Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., Cheng, X., 2007. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature 449, 248-251.

Jin, B., Ernst, J., Tiedemann, R.L., Xu, H., Sureshchandra, S., Kellis, M., Dalton, S., Liu, C., Choi, J.-H., Robertson, K.D., 2012. Linking DNA methyltransferases to epigenetic marks and nucleosome structure genome-wide in human tumor cells. Cell reports 2, 1411-1424.

Jones, P.A., 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature Reviews Genetics 13, 484-492.

Jones, P.A., Taylor, S.M., 1980. Cellular differentiation, cytidine analogs and DNA methylation. Cell 20, 85-93.

Jurkowska, R.Z., Anspach, N., Urbanke, C., Jia, D., Reinhardt, R., Nellen, W., Cheng, X., Jeltsch, A., 2008. Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L. Nucleic acids research 36, 6656-6663.

Kerns, E., Di, L., 2010. Drug-like properties: concepts, structure design and methods: from ADME to toxicity optimization. 1st ed. Elsevier/Academic Press: New York, 2008, 526.

Koes, D.R., Camacho, C.J., 2012. ZINCPharmer: pharmacophore search of the ZINC database. Nucleic acids research 40, W409-W414.

Korb, O., Stutzle, T., Exner, T.E., 2009. Empirical scoring functions for advanced protein– ligand docking with PLANTS. Journal of chemical information and modeling 49, 84-96.

Lagorce, D., Sperandio, O., Galons, H., Miteva, M.A., Villoutreix, B.O., 2008a. FAF-Drugs2: free ADME/tox filtering tool to assist drug discovery and chemical biology projects. BMC bioinformatics 9, 396.

Lagorce, D., Sperandio, O., Galons, H., Miteva, M.A., Villoutreix, B.O., 2008b. FAF-Drugs2: free ADME/tox filtering tool to assist drug discovery and chemical biology projects. BMC bioinformatics 9, 1.

Lagorce, D., Sperandio, O., Baell, J.B., Miteva, M.A., Villoutreix, B.O., 2015. FAF-Drugs3: a web server for compound property calculation and chemical library design. Nucleic acids research 43, W200-W207.

Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J., 2012. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced drug delivery reviews 64, 4-17.

Lyne, P.D., 2002. Structure-based virtual screening: an overview. Drug Discovery Today 7, 1047-1055.

Mai, A., Altucci, L., 2009. Epi-drugs to fight cancer: from chemistry to cancer treatment, the road ahead. The international journal of biochemistry & cell biology 41, 199-213.

Newberne, P.M., Rogers, A.E., 1986. Labile methyl groups and the promotion of cancer. Annual review of nutrition 6, 407-432.

Niculescu, M.D., Craciunescu, C.N., Zeisel, S.H., 2006. Dietary choline deficiency alters global and gene-specific DNA methylation in the developing hippocampus of mouse fetal brains. The FASEB Journal 20, 43-49.

Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D., 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448, 714-717.

Oprea, T.I., Davis, A.M., Teague, S.J., Leeson, P.D., 2001. Is there a difference between leads and drugs? A historical perspective. Journal of Chemical Information and Computer Sciences 41, 1308-1315.

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

Otero, N.K., Thomas, J.D., Saski, C.A., Xia, X., Kelly, S.J., 2012. Choline supplementation and DNA methylation in the hippocampus and prefrontal cortex of rats exposed to alcohol during development. Alcoholism: Clinical and Experimental Research 36, 1701-1709.

Pan, Y., Huang, N., Cho, S., MacKerell, A.D., 2003. Consideration of molecular weight during compound selection in virtual target-based database screening. Journal of Chemical Information and Computer Sciences 43, 267-272.

Rondelet, G., Dal Maso, T., Willems, L., Wouters, J., 2016. Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. Journal of Structural Biology 194, 357-367.

Selick, H.E., Beresford, A.P., Tarbit, M.H., 2002. The emerging importance of predictive ADME simulation in drug discovery. Drug Discovery Today 7, 109-116.

Teague, S.J., Davis, A.M., Leeson, P.D., Oprea, T., 1999. The design of leadlike combinatorial libraries. Angewandte Chemie International Edition 38, 3743-3748.

Vassilev, L.T., Vu, B.T., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., Kong, N., Kammlott, U., Lukacs, C., Klein, C., 2004. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. Science (New York, N.Y.) 303, 844-848.

Veber, D.F., Johnson, S.R., Cheng, H.-Y., Smith, B.R., Ward, K.W., Kopple, K.D., 2002. Molecular properties that influence the oral bioavailability of drug candidates. Journal of medicinal chemistry 45, 2615-2623.

Visualizer, D.S., 2013. Release 4.0. Accelrys Software Inc., San Diego, CA, USA.

Wang, Y.-W., Ma, X., Zhang, Y.-A., Wang, M.-J., Yatabe, Y., Lam, S., Girard, L., Chen, J.-Y., Gazdar, A.F., 2016. ITPKA gene body methylation regulates gene expression and serves as an early diagnostic marker in lung and other cancers. Journal of Thoracic Oncology 11, 1469-1481.

Wen, H., Li, Y., Xi, Y., Jiang, S., Stratton, S., Peng, D., Tanaka, K., Ren, Y., Xia, Z., Wu, J., 2014. ZMYND11 links histone H3. 3 K36 trimethylation to transcription elongation and tumor suppression. Nature 508, 263.

White, A.W., Westwell, A.D., Brahemi, G., 2008. Protein–protein interactions as targets for small-molecule therapeutics in cancer. Expert reviews in molecular medicine 10, e8.

Wolber, G., Langer, T., 2005. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. Journal of chemical information and modeling 45, 160-169.

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

Yang, X., Han, H., De Carvalho, D.D., Lay, F.D., Jones, P.A., Liang, G., 2014. Gene body methylation can alter gene expression and is a therapeutic target in cancer. Cancer cell 26, 577-590.

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

4.6 Supplementary Data

a
b

$$T(a, b) = \frac{N_c}{N_a + N_b - N_c}$$
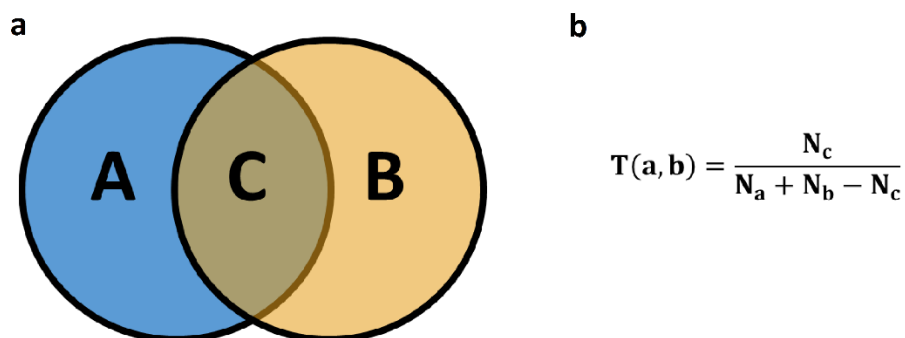
Supplementary Figure S1: *Tanimoto coefficient. (**a**) Intersection C between two sets (A and B) defined by Tanimoto coefficient (**b**) to measure molecular similarity. Tanimoto coefficient, also called the Jaccard coefficient (Otani et al., 2009) is used to determine the degree of similarity between two structures made of binary bits, each bit representing the absence (0) or the presence (1) from a list of structure fragments and/or particular features (Jurkowska et al., 2008). Consider two molecules A and B defined each by a 2D structure fingerprint, with C the intersection of the fingerprints sets with the same bits as 1, then the Tanimoto coefficient is defined by equation (**b**). In this equation, N_a and N_b represent the number of present features (digit 1) in each structure A and B, and N_c represents the common features to fingerprints A and B. So, T quantifies the fraction of features common to A and B to the total number of features of A or B. The Tanimoto coefficient gives values between 0 (dissimilar) and 1 (similar). So the higher the Tanimoto value, the closer the target structures are to the query structure. A Tanimoto coefficient higher than 0.85 indicates that two molecules may have similar activities (Chen et al., 2004).*
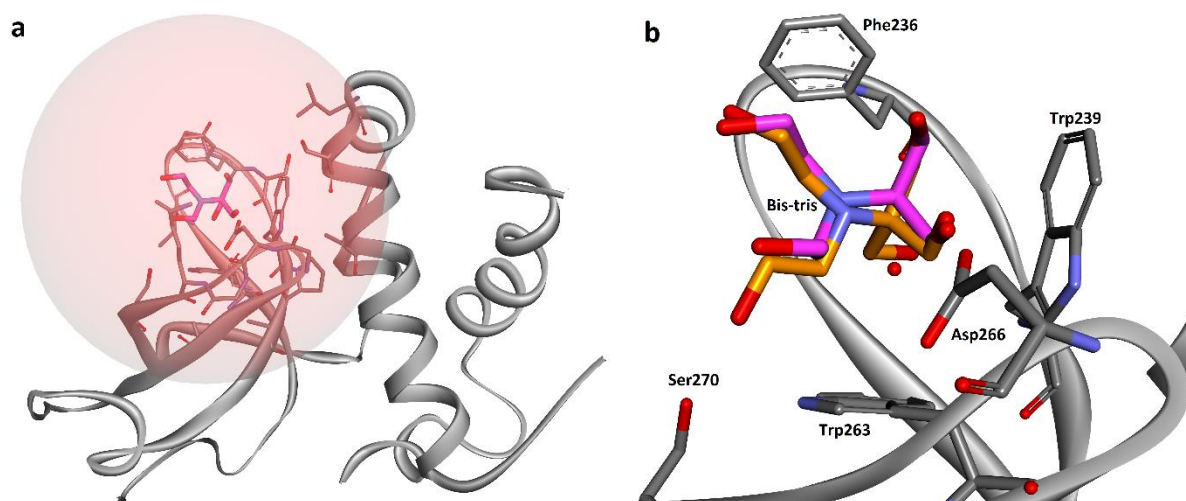
Supplementary Figure S2: *PWWP domain of human DNA (cytosine-5)-methyltransferase 3B in complex with a bis-tris molecule. (a) Binding site defined as a sphere of 13 Å radius centered on the bis-tris molecule In order to explore all the binding poses into the DNMT3B PWWP domain. (b) Superimposition of the co-crystallized pose (pink) of bis-tris (PDB code: 3QKJ) and the top ranked pose (orange) based on chemPLP function score.*

Supplementary Table S1: *Physico-chemical property filters developed in FAF-Drugs3. Orally administered drugs should have good oral bioavailability in order to be effective. The pioneering research by Lipinski et al. led to the well-known "Rule of Five" for selecting drug-like molecules (Lipinski, 2000). According to the "Rule of Five", a drug-like molecule should have no more than one of the following violations: No more than 5 hydrogen bond donors No more than 10 hydrogen bond acceptors Molecular weight no more than 500 LogP no more than 5 (Polar surface area <150 $Å^2$).*

	Rule of 3	Rule of 5	Drug-Like Soft	Lead-Like Soft
MW	≤ 300	(≤ 500)	100 - 600	150 - 400
logP	-3 to 3	(≤ 5)	-3 to 6	-3 to 4
HBA	≤ 3	(≤ 10)	≤ 12	≤ 7
HBD	≤ 3	(≤ 5)	≤ 5	≤ 4
HBonds	-	-	-	-
tPSA ($Å^2$)	≤ 60	-	≤ 180	≤ 160
Rotatable Bonds	≤ 3	-	≤ 11	≤ 9
Rigid Bonds	-	-	≤ 30	≤ 30
Rings	-	-	≤ 6	≤ 4
Max Size System Ring	-	-	≤ 18	≤ 18
Carbons	-	-	3 - 35	3 - 35
HeteroAtoms	-	-	1 - 15	1 - 15
H/C Ratio	-	-	0.1 to 1.1	0.1 to 1.1
Charges	-	-	≤ 3	≤ 3
Total Charge	-	-	-2 to 2	-2 to 2
RO5 Violations	-	2	-	-
Stereo Centers	-	-	-	≤ 2

Supplementary Table S2: Data Collection and Refinement Statistics for the twelve ligands (Table1)-DNMT3B PWWP complexes.

	A	B	C
Data collection			
Space group	P 32 2 1	P 32 2 1	P 32 2 1
Cell dimensions			
a, b, c (Å)	73.55 73.55 155.30	74.68 74.68 157.58	74.27 74.27 158.2
α, β, γ (°)	90 90 120	90 90 120	90 90 120
Resolution (Å)	40.37 - 2.07	49.97 - 2.54	40.78 - 2.40
	(2.21 - 2.07) *	(2.72 - 2.54)	(2.55 - 2.40)
R_{sym} or R_{merge} (%)	5.2 (65.8)	5.3 (64.3)	4.7 (89.7)
$I/\sigma I$	21.37 (2.90)	23.12 (3.46)	27.78 (1.85)
Completeness (%)	99.66 (97.19)	99.63 (99.04)	99.92 (99.20)
Redundancy	7.3 (7.3)	9.8 (9.7)	9.3 (7.2)
Refinement			
Resolution (Å)	40.17 - 2.07	40.77 - 2.57	40.78 - 2.40
	(2.15 - 2.07)	(2.66 - 2.57)	(2.49 - 2.40)
No. reflections	30107	16831	20415
R_{work} / R_{free} (%)	20.71/24.35	19.79/23.29	20.66/24.40
No. atoms			
Protein	2087	2104	2060
Ligand/ion	78	43	59
Water	183	45	70
Overall B-factor	38.50	56.50	57.20
Ligand	36.01	63.22	58.02
Occupancy (ligand)	0.9	0.8	0.9
R.m.s. deviations			
Bond lengths (Å)	0.008	0.011	0.009
Bond angles (°)	1.10	1.21	1.23
Clashscore	7.1	7.1	8.7
Ramachandran plot			
Favored regions (%)	96	96	96
Allowed regions (%)	3.2	3.2	3.2
Outlier regions (%)	0.8	0.8	0.8

*Values in parentheses are for highest-resolution shell

	D	E	F
Data collection			
Space group	P 32 2 1	P 32 2 1	P 32 2 1
Cell dimensions			
a, b, c (Å)	74.26 74.26 159.1	74.61 74.61 156.51	74.63 74.63 156.67
α, β, γ (°)	90 90 120	90 90 120	90 90 120
Resolution (Å)	40.92 - 2.40	40.59 - 2.45	40.62 - 2.03
	(2.54 - 2.40)	(2.59 - 2.45)	(2.15 - 2.03)
R_{sym} or R_{merge} (%)	4.1 (88.6)	8.2 (65.2)	4.2 (87.1)
I /σI	29.24 (2.48)	19.64 (1.93)	26.62 (1.98)
Completeness (%)	99.91 (99.25)	99.72 (97.14)	99.87 (98.79)
Redundancy	9.7 (9.8)	9.7 (9.3)	9.8 (9.9)
Refinement			
Resolution (Å)	40.92 - 2.40	40.59 - 2.45	40.62 - 2.03
	(2.48 - 2.40)	(2.53 - 2.45)	(2.10 - 2.03)
No. reflections	20558	19248	33473
R_{work} / R_{free} (%)	19.44/24.35	19.96/23.93	20.13/24.05
No. atoms			
Protein	2128	2070	2142
Ligand/ion	69	58	64
Water	73	61	104
Overall B-factor	51.40	45.90	50.70
Ligand	39.83	47.93	63.12
Occupancy (ligand)	0.9	0.8	0.9
R.m.s. deviations			
Bond lengths (Å)	0.009	0.008	0.008
Bond angles (°)	1.97	1.25	1.17
Clashscore	7.7	9.1	5.8
Ramachandran plot			
Favored regions (%)	96	96	97
Allowed regions (%)	3.2	3.6	2.2
Outlier regions (%)	0.8	0.4	0.8

	G	H	I
Data collection			
Space group	P 32 2 1	P 32 2 1	P 32 2 1
Cell dimensions			
a, b, c (Å)	74.20 74.20 155.80	74.63 74.63 156.57	73.50 73.50 159.96
α, β, γ (°)	90 90 120	90 90 120	90 90 120
Resolution (Å)	40.36 - 2.30	40.60 - 2.71	49.88 - 1.70
	(2.43 - 2.30)	(2.83 - 2.71)	(1.81 - 1.70)
R_{sym} or R_{merge} (%)	7.8 (87.3)	9.1 (75.3)	3.4 (65.6)
$I / \sigma I$	26.96 (3.65)	16.97 (3.54)	25.73 (2.28)
Completeness (%)	99.95 (99.96)	99.79 (97.91)	99.96 (99.98)
Redundancy	9.6 (9.4)	9.8 (10.1)	6.6 (6.5)
Refinement			
Resolution (Å)	40.39 - 2.30	40.6 - 2.71	40.87 - 1.70
	(2.38 - 2.30)	(2.79 - 2.71)	(1.76 - 1.70)
No. reflections	22766	18134	55899
R_{work} / R_{free} (%)	19.61/23.79	20.06/24.73	19.35/21.32
No. atoms			
Protein	2104	2072	2223
Ligand/ion	43	52	51
Water	171	48	340
Overall B-factor	35.90	38.00	41.50
Ligand	48.13	42.23	49.87
Occupancy (ligand)	0.9	1.0	0.9
R.m.s. deviations			
Bond lengths (Å)	0.008	0.008	0.007
Bond angles (°)	1.13	1.20	1.09
Clashscore	8.3	9.9	5.8
Ramachandran plot			
Favored regions (%)	95	96	97
Allowed regions (%)	4.2	3.2	2.3
Outlier regions (%)	0.8	0.8	0.7

	J	K	L
Data collection			
Space group	P 32 2 1	P 32 2 1	P 32 2 1
Cell dimensions			
a, b, c (Å)	74.26 74.26 158.19	74.93 74.93 157.47	74.87 74.87 157.63
α, β, γ (°)	90 90 120	90 90 120	90 90 120
Resolution (Å)	40.77 - 2.27	40.81 - 2.30	40.82 - 2.50
	(2.41 - 2.27)	(2.44 - 2.30)	(2.65 - 2.50)
R_{sym} or R_{merge} (%)	3.5 (66.1)	5.0 (89.3)	6.0 (57.8)
I /σI	37.57 (2.80)	24.72 (1.66)	24.21 (1.85)
Completeness (%)	99.91 (99.06)	99.49 (94.91)	99.92 (99.61)
Redundancy	12.6 (10.9)	9.3 (7.2)	9.8 (9.9)
Refinement			
Resolution (Å)	40.78 - 2.27	40.81 - 2.30	33.68 - 2.50
	(2.35 - 2.27)	(2.38 - 2.30)	(2.59 - 2.50)
No. reflections	24069	23390	18364
R_{work} / R_{free} (%)	22.68/26.81	19.51/24.77	19.59/22.80
No. atoms			
Protein	2102	2128	2118
Ligand/ion	33	49	45
Water	115	60	30
Overall B-factor	55.00	60.50	63.70
Ligand	64.42	70.49	75.11
Occupancy (ligand)	1.0	1.0	0.9
R.m.s. deviations			
Bond lengths (Å)	0.012	0.014	0.014
Bond angles (°)	1.33	1.39	1.28
Clashscore	17.3	7.0	6.3
Ramachandran plot			
Favored regions (%)	92	97	96
Allowed regions (%)	6.4	2.2	3.6
Outlier regions (%)	1.6	0.8	0.4

4.7 Supplemental References

Chen, T., Tsujimoto, N., Li, E., 2004. The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. Molecular and cellular biology 24, 9048-9058.

Jurkowska, R.Z., Anspach, N., Urbanke, C., Jia, D., Reinhardt, R., Nellen, W., Cheng, X., Jeltsch, A., 2008. Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L. Nucleic acids research 36, 6656-6663.

Lipinski, C.A., 2000. Drug-like properties and the causes of poor solubility and poor permeability. Journal of pharmacological and toxicological methods 44, 235-249.

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

GENERAL CONCLUSIONS
&
PERSPECTIVES

General conclusions

DNA methylation is an important epigenetic modification involved in gene expression regulation in normal cells. This modification occurs on cytosine residues of CpG dinucleotides and is catalyzed by DNA methyltransferases (DNMTs). In cancer, DNA methylation deregulation leads to tumorigenesis and tumor maintenance. Thereby, gene body methylation promotes activation of genes up-regulated by oncoproteins, and hypermethylation status of tumor suppressor genes promoters leads to their inactivation. As DNA methylation is a reversible process, epi-drugs targeting DNMTs are promising anticancer agents to reverse the deregulated methylome to stop cancer proliferation. To develop more specialized and personalized treatments for cancer, selective inhibition of DNMTs is required. Indeed, this methylome deregulation depends on DNMT isoforms overexpression in a cell type-specific manner or, *inter alia*, post-translational modification of histones to recruit DNMTs to particular genomic regions.

The objectives of this thesis were to: **i)** investigate the mode of action of current non-nucleoside targeting the C-terminal catalytic domain of DNMTs; **ii)** understand mechanism of H3K36me3 nucleosome recognition by DNMT3s; and **iii)** develop protein-protein interaction inhibitors of the PWWP DNMT3B-H3K36me3 complex to downregulate expression of oncogenes and oncoprotein-regulated genes.

In the first part, we focused our research on maleimide derivatives of RG108 (Fig. 1) (RG108-1 and RG119-1) as covalent inhibitors. These compounds show differences in selectivity and potency towards DNMTs. Enzymatic assays revealed that RG108-1 is selective for DNMT3A with an IC50 value of 27 µM (IC50$_{DNMT1}$ value in the millimolar range). RG119-1 exhibited better activity against both DNMT3A and DNMT1 with IC50 values of 3.1 µM and 45 µM, respectively. This difference of activity towards DNMT1 was reflected in the cytotoxic assays on mesothelioma cell lines. Indeed, these cells lines overexpress preferentially DNMT1 and RG119-1 shows a potent inhibition in the low micromolar range (TC50$_{H28}$= 6.2 µM; TC50$_{M14K}$= 5.4 µM) compared to RG108-1 (TC50 values in the millimolar range).
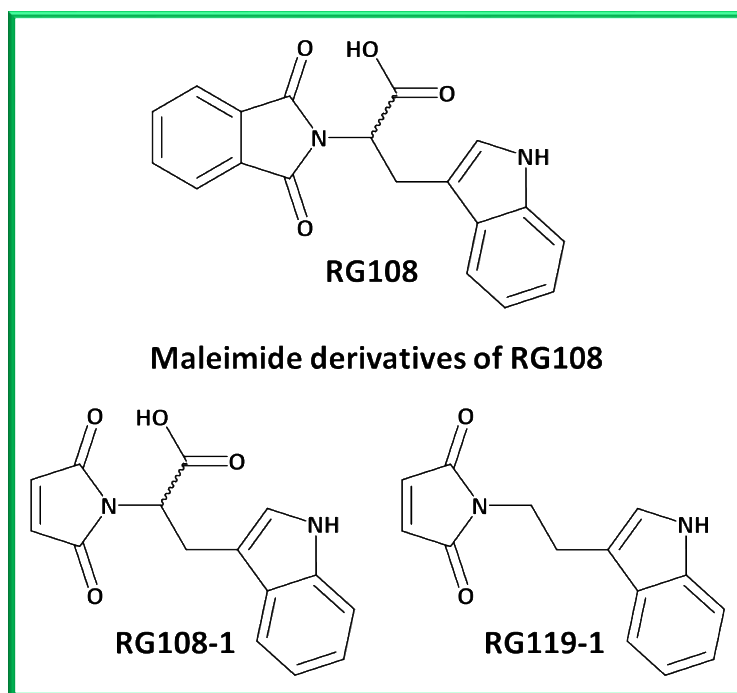
Figure 1. Chemical structures of the non-nucleoside DNA methylation inhibitor RG108 and its maleimide derivatives.

In order to understand the difference of selectivity between RG108-1 and RG119-1, differential scanning fluorimetry assays and molecular docking were performed on DNMT1. The stability of DNMT1 is more pronounced in presence of DNA and RG119-1 with a thermal shift of ~2.8°C and, as expected, RG108-1 does not stabilize DNMT1 in any form. This ternary system was retained for non-covalent and covalent docking simulations into DNMT1 (Fig. 2). From non-covalent docking studies, indole ring of RG119-1 occupies a hydrophobic pocket and is involved in π-π T-shaped interactions with Phe1148 residue and π-sulfur interactions with Met1172 residue (Fig. 2). The maleimide ring is preferentially oriented to the substrate cytosine pocket containing the catalytic cysteine via a moderate hydrogen bond interaction with Gln1230 residue. Based on this conformation, a covalent docking was performed by fixing a covalent link between ligands (succinimide group group) and the catalytic cysteine. The indole group of RG119-1 maintains the aromatic interaction with Phe1148 and makes an additional strong hydrogen-bonding interaction with Glu1171 (D_{A-B} = 2.5 Å). Another strong hydrogen-bonding between maleimide ring and the protonated Glu1269 (D_{A-B} = 2.6 Å) is newly formed.
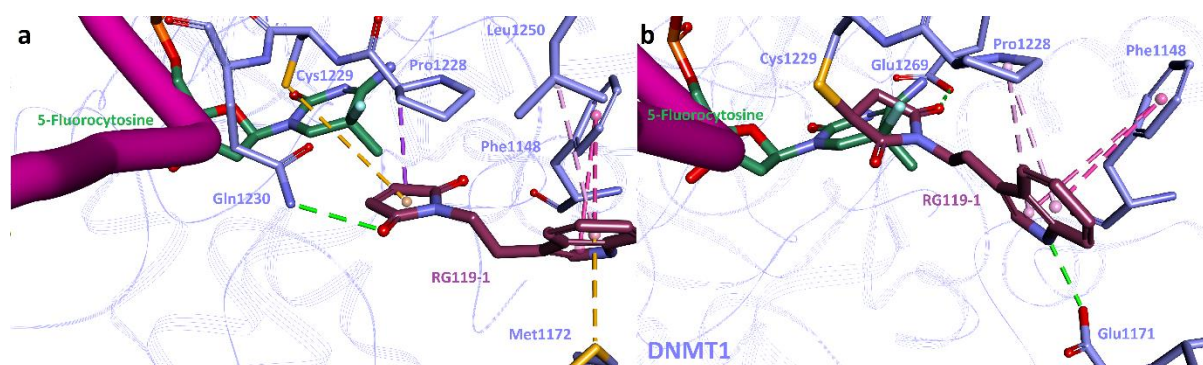
Figure 2. Binding mode of RG119-1 into mouse DNMT1. *(a) Non-covalent docking of RG119-1 into mouse DNMT1. (b) Covalent docking of RG108-1 and RG119-1 into bacterial M.HhaI and mouse DNMT1. Ligands and amino acid residues of the binding site are rendered as stick models, while the rest of the protein backbone is displayed as line ribbon. DNA backbone is displayed as purple tube model. 5-Fluorocytosine, removed prior to docking simulations, was added for final visualization.*

In this same study, crystallization assays have been conducted to the crystal structure of the M.*Hha*I apoenzyme (Fig. 3). This structure reveals that the SAM cofactor does not affect the cofactor binding site conformation of M.*Hha*I and that RG119-1 is able to displace this strong association between the cofactor and M.*Hha*I suggesting a competition mechanism of RG119-1 as suggested by our docking results. For RG108-1, the difference of selectivity compared to RG119-1 could arise from the fact that the maleimide ring is not stabilized inside the catalytic domain, and the compound is blocked by non-covalent interactions which prevent covalent reaction with the catalytic cysteine. Even if a covalent bond is formed, the compound seems not strongly stabilized inside the binding pocket.
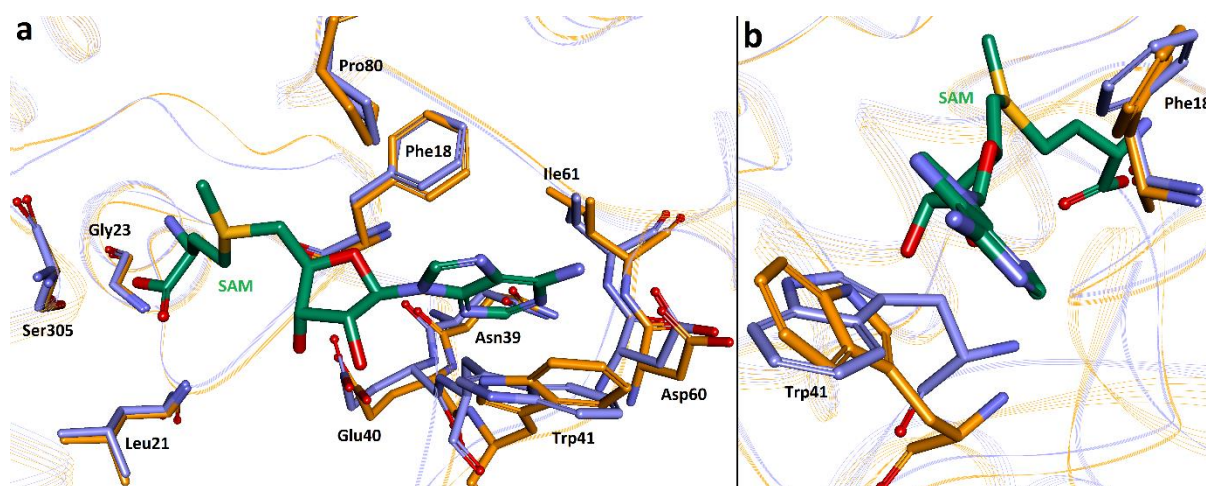
Figure 3. Superposition of binary complex M.HhaI-SAM cofactor (orange) (PDB code: 2HMY) with the apoenzyme (blue). *Amino acid residues of the binding site are rendered as stick models while rest of the protein backbone is displayed as line ribbon. SAM cofactor is rendered as green stick model. **(a)** Slight backbone displacement for residues Glu40, Trp41, Asp60, and Ile 61 (from 0.8 Å to 1.5 Å) and **(b)** conformational side-chain rearrangements for residues Phe18 and Trp41 when cofactor binds to M.HhaI.*

For a decade, efforts to discover non-nucleoside inhibitors that act directly on DNMTs is currently slowed by the lack of structural information of DNMTs, the low selectivity for DNMTs, the lack of information about their mechanism of inhibition and the low inhibitory activity in cells (Erdmann et al., 2015). These non-nucleoside compounds are proposed as novel alternative for epigenetic cancer therapy compared to the approved nucleoside analogs Vidaza® and Dacogen® which present a significant toxicity (Christman, 2002; Issa and Kantarjian, 2009; Jones and Taylor, 1980; Wijermans et al., 2008). In the present work, we showed that our series of compounds present IC50 in the low micromolar, a selectivity towards DNMTs isoform and, from crystallography and modeling, we were able in part to rationalize the mode of action of these compounds. However, a series of experiments has to be addressed to go further. In order to obtain crystal structures of complexes with these inhibitors, different isoforms of DNMTs can be considered. Conditions to achieve this can be based on the newly crystal structure of DNMT1 (Zhang et al., 2015). To investigate more thoroughly the mechanism of action of these compounds, determination of the different catalytic constant have to be addressed to determine the nature of the covalent bond. Finally, study of their effect on specific tumor suppressor genes reactivation has also to be addressed.

In the following part, we proposed a new therapeutic strategy against specific cancers based on the inhibition of protein-protein interactions involved during gene body methylation. This feature of DNA methylation mediates activation of oncogenes (e.g., ITPKA) and genes up-regulated (e.g., c-Myc) and is related with a specific distribution of histone H3K36me3 within the genome in a particular cell phase. This permissive epigenetic mark, H3K36me3, recruits DNMT3B to gene bodies via its PWWP domain.

Chapter 3 deals with the structural basis for recognition of the histone mark H3K36me3 by human *de novo* DNA methyltransferases 3A and 3B in a nucleosomal context. For this, structure of DNMT3B PWWP domain in complex with the epigenetic mark H3K36me3 ($H3_{32-38}K36me3$) was solved (Fig. 4a and b). This complex revealed an induced fit of the protein which enhances the recognition of the peptide by π-cation interactions in addition to van der Waals and electrostatic interactions which contribute to a specific association of DNMT3B PWWP domain with H3K36me3 peptide. A key strong hydrogen-bonding interaction for the recognition of this epigenetic mark involved the CO main chain group of trimethylated Lys36 with a conserved water molecule stabilized by the oxygen atom of the lateral chain of Ser270. This is consistent with the observation that the mutation (S270P) in the ICF (Immunodeficiency, Centromeric instability and Facial abnormalities) syndrome conducts to the lack of methyltransferase recognition and activity on nucleosomal substrates. The proposed model of DNMT3A in complex with the H3K36me3 peptide ($H3_{32-38}K36me3$) provides a similar binding mode (Fig. 4c).
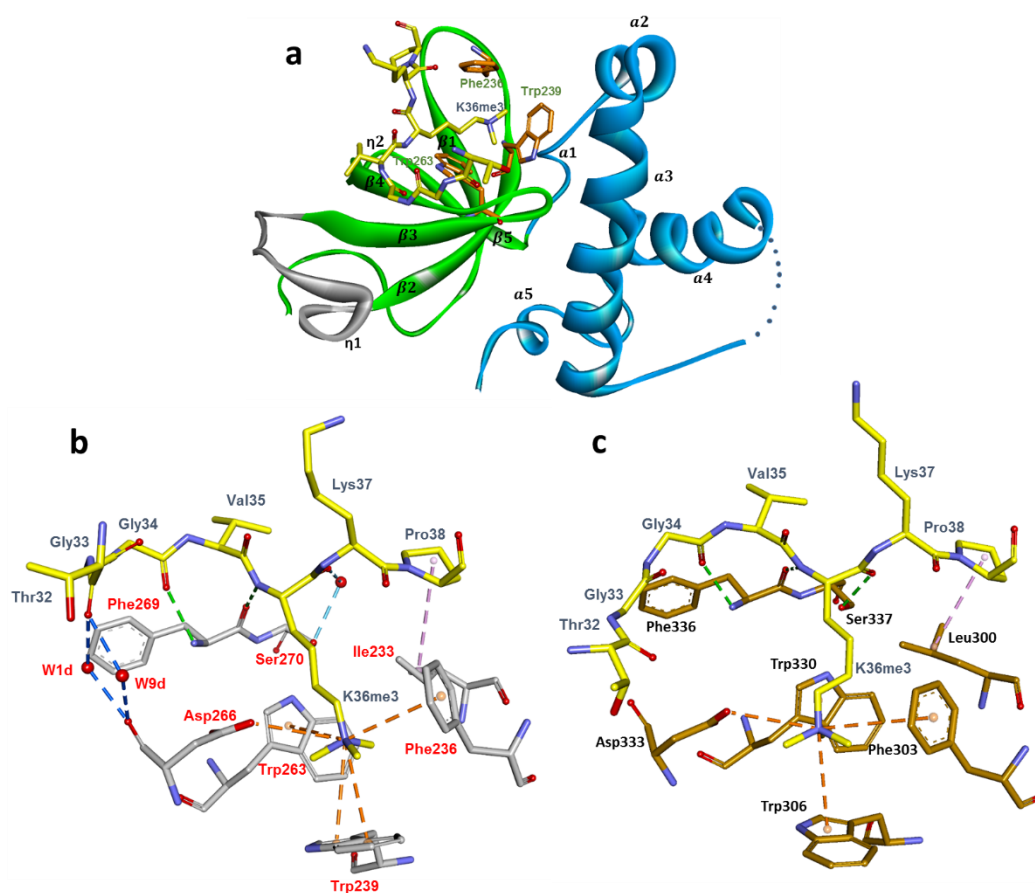
Figure 4. Structures of DNMT3s PWWP domain in complex with H3K36me3 peptide. *(**a**) Solid ribbon representation of the crystal structure of the complex between the DNMT3B PWWP domain and the H3K36me3 peptide (H3$_{32-38}$K36me3). (**b**) Crystal structure of DNMT3B PWWP domain (grey stick representation) in complex with H3K36me3 peptide (yellow stick representation). Selected intermolecular interactions are represented. (**c**) Molecular modeling of interactions between the DNMT3A PWWP domain (orange stick representation) and H3K36me3 peptide (yellow stick representation).*

From the model generated for the DNMT3A and DNMT3B PWWP domains in complex with the nucleosome core (Fig. 5), we conclude that the selectivity of DNMT3A and DNMT3B for the H3K36me3-NCP (nucleosome core particle) is mediated by interactions of the PWWP domain through the aromatic cage for H3K36me3 recognition and the basic surface which could bind two DNA duplexes of the nucleosome.
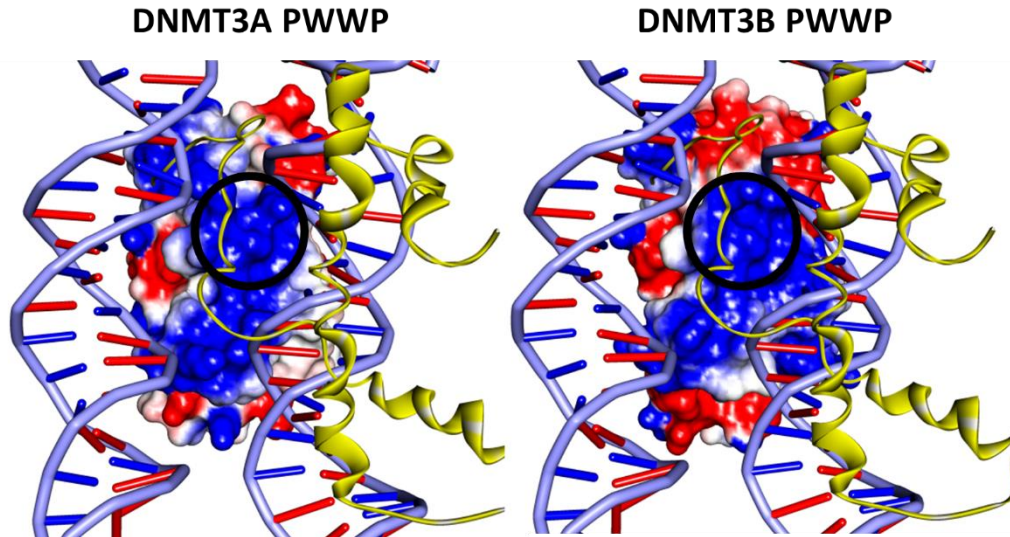
DNMT3A PWWP   **DNMT3B PWWP**

Figure 5. ***PWWP domains in complex with H3K36me3-NCP.*** *PWWP domains surface colored by electrostatic potential of the residues (red: negatively charged area and blue: positively charged area). The black circle represents the zone of interaction between the aromatic cage of DNMT3s PWWP domains and the epigenetic mark H3K36me3.*

Finally, we reconstructed the complete structure of DNMT3A (PWWP-ADD-CD domains) and determined the structural model of DNMT3A-DNMT3L heterotetramer in complex with H3K36me3-modified dinucleosome. This model provides structural information about DNA methylation on the linker DNA region of nucleosomes and the *de novo* DNA methylation patterns in agreement with the findings of the literature (Fig. 6).
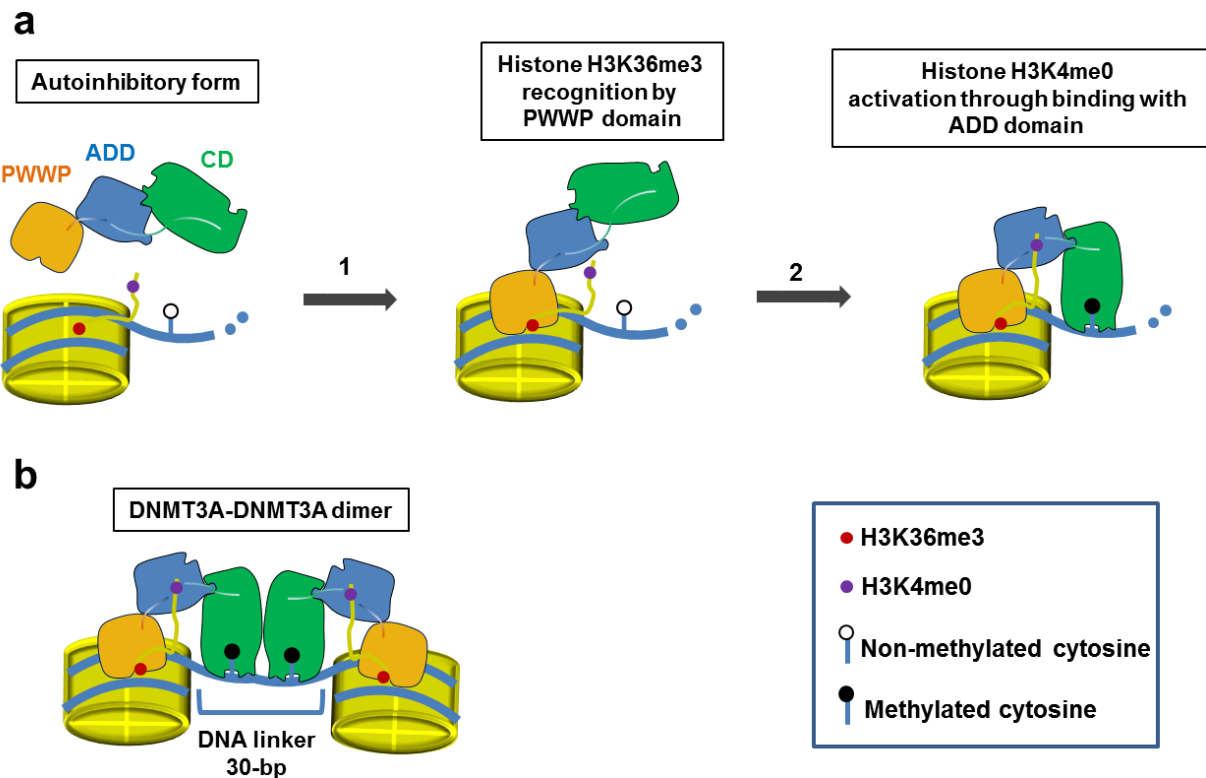
Figure 6. Proposed mechanism for recognition of nucleosome and DNA methylation by DNMT3A. (a) *Sequential recognition mechanism of DNMT3A for nucleosome and methylation activation with H3 tail: **1**-Nucleosome recognition by PWWP domain through interaction with methylated histone H3 tail (H3K36me3). **2**-Histone H3 tail (H3K4me0) allosterically activates DNMT3A through binding with ADD domain. ADD domain of DNMT3A interacts with the catalytic domain and inhibits its activity by preventing it to interact with DNA. (**b**) Periodicity in de novo methylation patterns of ~10-bp between two CpG sites. (The model is available online: http://dx.doi.org/10.1016/j.jsb.2016.03.013)*

One of the most difficult tasks for the future will be to understand the difference of methylation activity between DNMT3A and DNMT3B. Structure determination of these enzyme with nucleosome could answer this question. However, such crystallographic complexes remain challenging to obtain. In the PDB (Protein Data Bank), only two protein in complex with a nucleosome core particle were deposited since 2013 (PDB code: 5E5A (chromatin-tethering domain of Human cytomegalovirus IE1 protein) and 4LD9 (BAH domain of Sir3)). The model proposed in this thesis could help to set up a procedure for obtaining such complexes with H3K36me3-modified dinucleosome.

In the last chapter, we explored potent protein-protein interaction inhibitors of our previous solved crystal structure complex between DNMT3B PWWP domain and the tissue-specific cancer post-translational modification of histone 3, H3K36me3.

Ten years ago, protein-protein interactions emerged as a new class of therapeutic targets. Nowadays, identification of residues responsible of the binding affinity (hot spot) in a focus region in protein-protein interactions have permitted to develop small-inhibitors for dozens of protein-protein interactions (Arkin et al., 2014). In our case, the hot spot identified in the aromatic cage of DNMT3B PWWP domain is a novel druggable site for the development of small molecules (Fig. 7). These potential inhibitors are designed in order to downregulate c-Myc-regulated genes and oncogenes promoted by gene body methylation. For this, we performed first a similarity-based virtual screening using the bis-tris molecule, identified from a complex with DNMT3B PWWP domain (PDB code: 3QKJ), and H3K36me3 epigenetic mark as both chemical references.
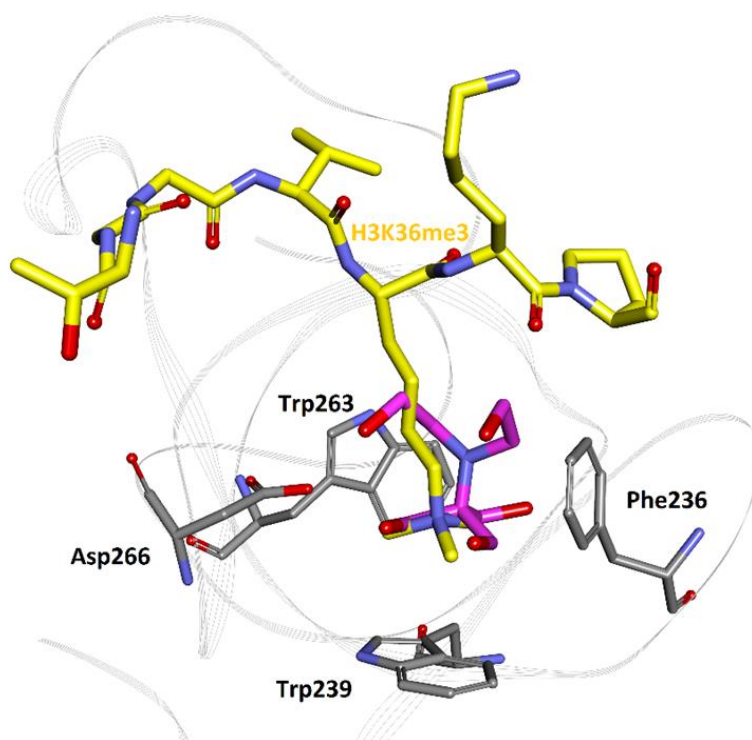

Figure 7. Superimposition of complex structures of PWWP-H3K36me3 (PDB code: 5CIU) and PWWP-bis-tris (PDB code: 3QKJ) of DNMT3B. *Bis-tris molecule (pink) is bound at the same position as the trimethyl-ammonium group of epigenetic mark H3K36me3.*

Among the twenty top ranked ligands, twelve led to a crystal complex with the DNMT3B PWWP domain. These crystallographic complexes gave structural insight into the binding mode including the conformational flexibility of the binding site. Interestingly, the identified choline showed, in a previous study, a reduction of hypermethylation in hippocampus and prefrontal cortex after alcohol exposure and is possibly linked with its capacity to inhibit the H3K36me3-DNMT3B PWWP interactions. Furthermore, we identified the FDA-approved drug ethambutol used in tritherapy against *Mycobacterium tuberculosis* and the antihistamine agent Metron S. Based twelve crystallographic complexes, we constructed a pharmacophore model (Fig. 8) and performed a pharmacophore-based virtual screening against ZINC database.
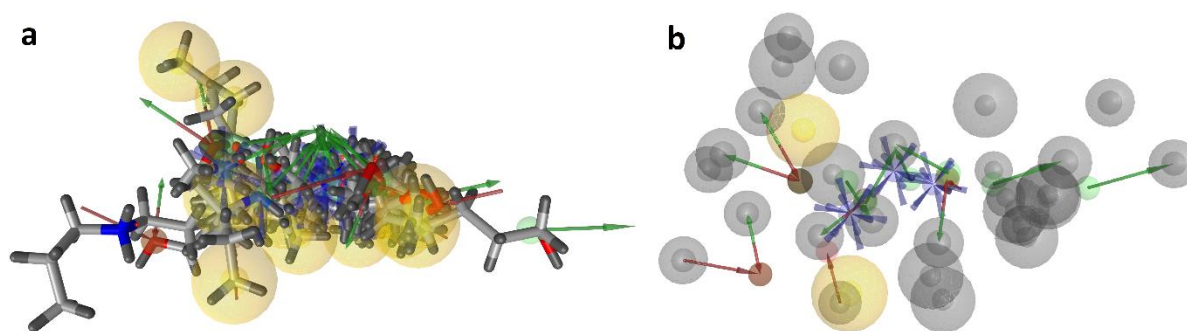

Figure 8. Pharmacophore model generation of DNMT3B PWWP ligands. Pharmacophore sites are colored red for H-bond acceptors, green for H-bond donors, blue for positively ionizable groups and yellow for hydrophobic groups. (a) Aligned ligands from crystallographic complexes with pharmacophore features. (b) Pharmacophore model generation based on common chemical features. Excluded volumes (ligand-inaccessible) of residues forming the binding site are represented as gray spheres.

Finally, molecular docking simulations were carried out into DNMT3B PWWP domain with compounds (Fig. 9) that fit the pharmacophore model to gain insights into the binding mode to identify and explore chemical diversity for development of protein-protein interaction inhibitors with high binding affinity towards DNMT3B PWWP domain. These complexes show intermolecular interfaces of high shape complementarity.
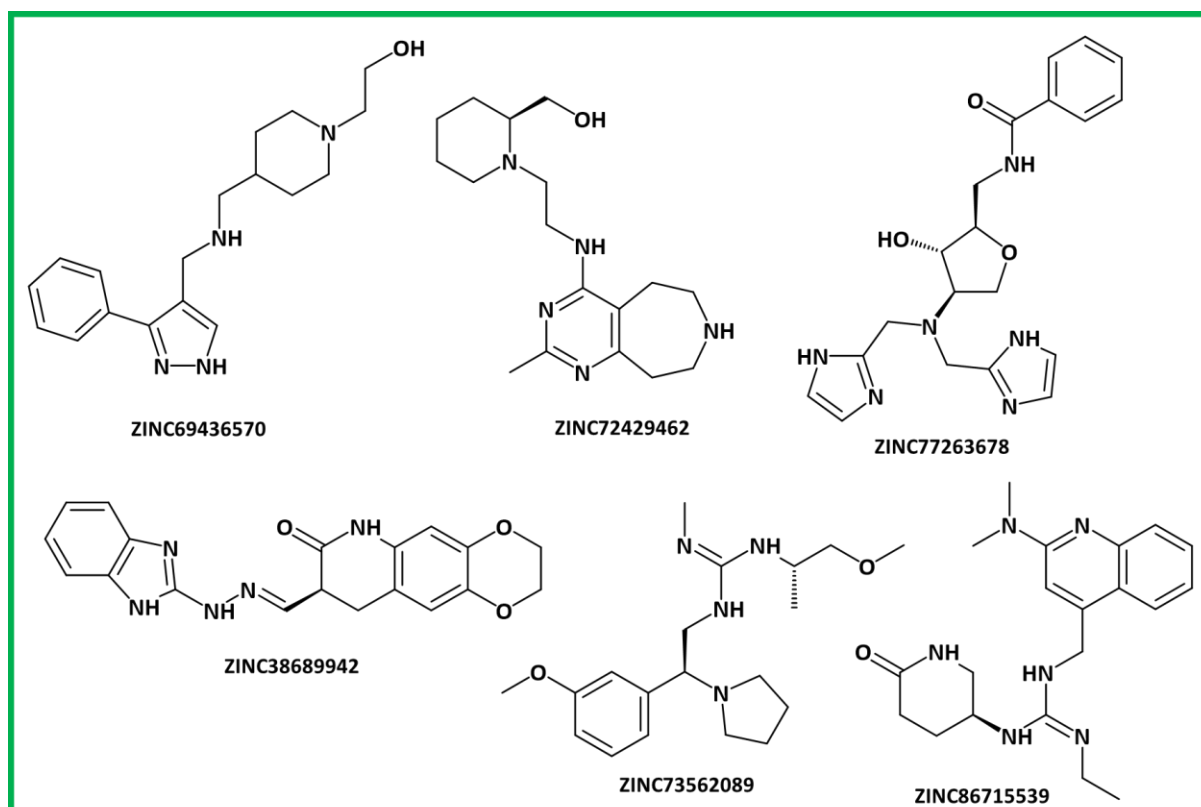
Figure 9. Identified hits obtained after pharmacophore model generation and pharmacophore-based virtual screening. *The six best ranked compounds after simulations docking are represented.*

One of the direct perspectives of chapter 4 is to obtain structures of complexes between the newly identified hits (Fig. 9) and the DNMT3B PWWP domain to get insight in the binding recognition. Determination of the binding constant and thermodynamics constants (isothermal titration calorimetry experiments) will provide important information on the affinity of these compounds for DNMT3B PWWP domain. To determine if molecules are able to inhibit the protein-protein interaction between DNMT3B PWWP domain and H3K36me3, HTRF (Homogenous Time Resolved fluorescence) experiment based on FRET (Fluorescence Resonance Energy Transfer) technology can be set up as for the BRD4-acetylated lysine system (Degorce et al., 2009; Duffy et al., 2015). Nowadays, several inhibitors of bromodomains used in cancer and inflammation are in clinical trials, such as I-BET762 (Glaxo Smith-Klein), CPI-0610 (Constellation) and OTX15 (OncEthix) which are based on the JQ1 scaffold (see chap. 1 section 7.1) (Fig. 10) (Filippakopoulos and Knapp, 2014; Arkin et al., 2014). The same strategy to design protein-protein inhibitors of the DNMT3B PWWP-H3K36me3 complex could lead to drugs able to downregulate oncogenes.
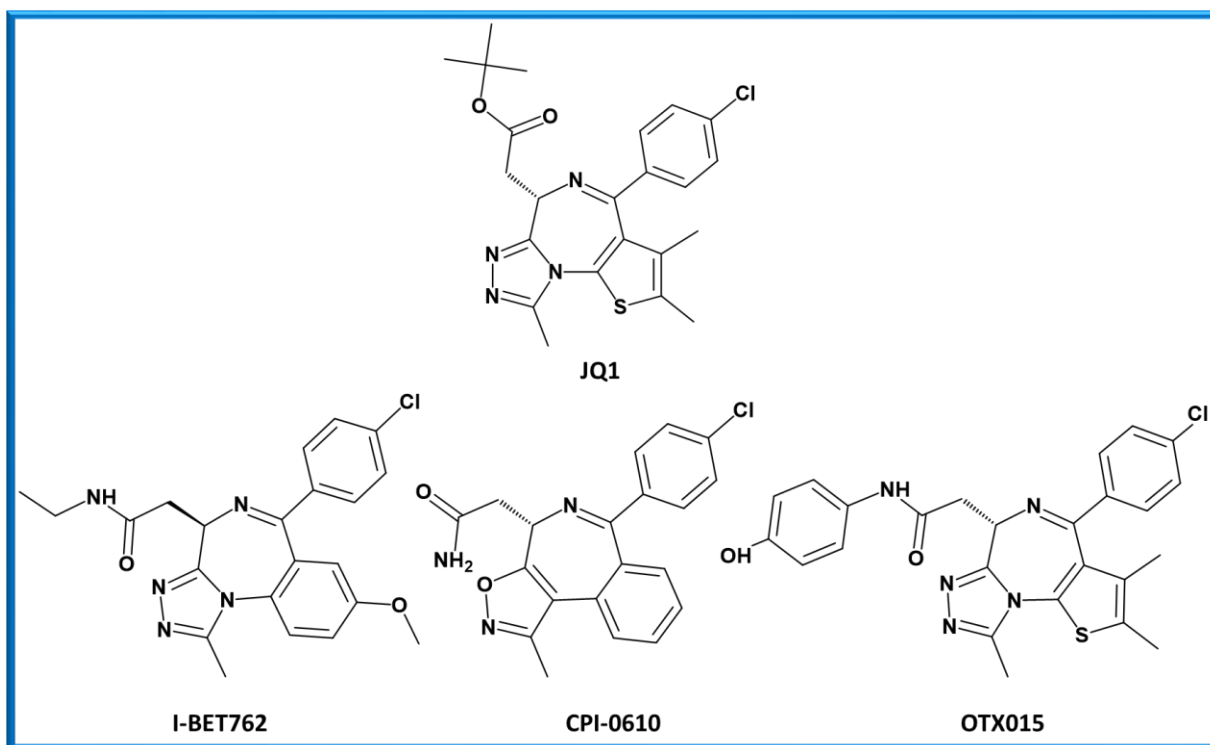
Figure 10. Small-molecule inhibitors of BRD4 bromodomain. Clinical compounds (I-BET762, CPI-0610 and OTX15) are shown below JQ1.

General Perspectives

In addition to the direct perspectives of each chapter provided in the above concluding section, some perspectives are proposed here. These are based on the combinatorial readout of histone H3 modifications by *de novo* methyltransferases (DNMT3A/3B) to enhance inhibition of these regulators for cancer therapy.

To be targeted to a specific genomic region in a particular cellular context, the combination of permissive epigenetic marks H3K4me0 and H3K36me3 are needed to activate and permit methylation of the genome. To understand the structural mechanism of these recognition events, the structure of DNMT3A ADD domain in complex with H3K4me0 and DNMT3B PWWP domain in complex with H3K36me3 were solved (see introduction Figs. 6 and 7).

However, structures of the full DNMT3A and DNMT3B enzymes have still to be addressed and, especially in this combinatorial readout system. These structural information will be useful in determining conformation and molecular mechanisms underlying these events. Indeed, specific protein-protein inhibitors can be developed by targeting either ADD domain or PWWP domain. Dual inhibition therapy could, therefore, produce a significant synergistic inhibitory effect towards aberrant DNMT3s activity. Moreover, physiological conformation of DNMT3s will highlight the relative position between hotspots in ADD and PWWP domains. An inhibitor targeting both domains could be developed, for example, by cross linking of identified protein-protein inhibitors of each domain.

Finally, small-molecule inhibitors of epigenetic regulators in combination with personal epigenome analysis could lead to a more efficiency and personalized cancer treatment.

References

Arkin, M.R., Tang, Y., Wells, J.A., 2014. Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. Chemistry & biology 21, 1102-1114.

Christman, J.K., 2002. 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. Oncogene 21, 5483-5495.

Degorce, F., Card, A., Soh, S., Trinquet, E., Knapik, G.P., Xie, B., 2009. HTRF: a technology tailored for drug discovery–a review of theoretical aspects and recent applications. Current Chemical Genomics and Translational Medicine 3.

Duffy, B.C., Liu, S., Martin, G.S., Wang, R., Hsia, M.M., Zhao, H., Guo, C., Ellis, M., Quinn, J.F., Kharenko, O.A., 2015. Discovery of a new chemical series of BRD4 (1) inhibitors using protein-ligand docking and structure-guided design. Bioorganic & medicinal chemistry letters 25, 2818-2823.

Erdmann, A., Halby, L., Fahy, J., Arimondo, P.B., 2015. Targeting DNA Methylation with Small Molecules: What's Next? Miniperspective. Journal of medicinal chemistry 58, 2569-2583.

Filippakopoulos, P., Knapp, S., 2014. Targeting bromodomains: epigenetic readers of lysine acetylation. Nature reviews Drug discovery 13, 337-356.

Issa, J.-P.J., Kantarjian, H.M., 2009. Targeting DNA methylation. Clinical Cancer Research 15, 3938-3946.

Jones, P.A., Taylor, S.M., 1980. Cellular differentiation, cytidine analogs and DNA methylation. Cell 20, 85-93.

Wijermans, P., Rüter, B., Baer, M., Slack, J.L., Saba, H., Lübbert, M., 2008. Efficacy of decitabine in the treatment of patients with chronic myelomonocytic leukemia (CMML). Leukemia research 32, 587-591.

Zhang, Z.-M., Liu, S., Lin, K., Luo, Y., Perry, J.J., Wang, Y., Song, J., 2015. Crystal Structure of Human DNA Methyltransferase 1. Journal of molecular biology 427, 2520-2531.

CURRICULUM VITÆ

PERSONAL INFORMATION   **Grégoire RONDELET**

in   Professional profile: www.linkedin.com/in/grégoire-rondelet-8a58a57a

R^G   Academic profile: www.researchgate.net/profile/Gregoire_Rondelet

Date of birth: February 7, 1989 | Nationality: Belgian

WORK EXPERIENCE

10.2012–12.2016   **Researcher in Chemistry (Ph.D.)**

Employer: National Fund for Scientific Research *(FNRS)* / Télévie campaign - Thesis performed at University of Namur *(Belgium)*

Sector: Health, Sciences and Technology

Research project: Structural study of non-nucleoside inhibitors of DNA (cytosine-5) methyltransferases

Laboratory: Laboratory of Structural Biological Chemistry *(Doctoral Advisor: Prof. Johan Wouters)*

02.2012–05.2012   **Researcher in Biochemistry (Internship)**

Employer: University Health Network, Toronto *(Canada)*

Sector: Health, Sciences and Technology

Research project: Structural study of N-domains of human CEACAMs binding Opa proteins

Laboratory: Biochemistry *(Promoter: Prof. Gil Privé)*

EDUCATION AND TRAINING

2012–2016   **Doctorate of Chemical Sciences**

University of Namur *(Belgium)*

2010–2012   **Master Degree in Chemical Sciences**

University of Namur *(Belgium)*

2007–2010   **Bachelor Degree in Chemical Sciences**

University of Namur *(Belgium)*

PERSONAL SKILLS

Mother tongue   French

Other languages   English (C1), Dutch (A2)

Communication skills   Scientific communication *(oral presentations and writing articles)*

Organisational/ managerial skills	▪ Training and coaching *(4 trainees and 3 master thesis students)* ▪ Main organizer of a scientific event "Industrial Researches in Structural Biological Chemistry" ▪ Responsibility for the management of protein purification equipment
Technical skills	▪ Synthesis of molecules of interest, characterization *(NMR, IR, UV, fluorescence, mass spectrometry, crystallography)* ▪ Biological evaluation *(MTS, ELISA, ITC, fluorescence)* ▪ Cloning, cell culture, expression of proteins of interest, protein purification *(HPLC, FPLC - affinity, ion exchange, size-exclusion chromatography)* ▪ Protein crystallography *(crystallization, synchrotron data collection, structural resolution)* ▪ Biomolecular modeling *(homology modeling, docking, structure-activity relationship, generation of pharmacophores)*
Computer skills	▪ Programming languages *(Python, Perl, HTML)* ▪ MS Office *(Word, Excel, Outlook, Powerpoint)* ▪ Specific softwares *(Discovery Studio, LigandScout, ChemDraw, KaleidaGraph, Origin, Gaussian, Mercury, Pymol, GOLD, Phenix, Coot, XDS, GraphPad Prism, ChemBioDraw, GIMP)* ▪ Bibliographic search *(Scifinder, ISI Web of Knowledge, PubMed, Patentscope, Espacenet, Google Scholar)*

COMMUNICATIONS AND PRIZE

Awards	**Best Communication Prize** (French Medicinal Chemistry Society (SCT) - 22nd Young Research Fellows Meeting, Paris, 2015)
Scientific Communications	**Oral Presentations** (5), **Posters** (10), **International Articles**: First author (FA); Co-author (CA) ▪ (FA) "Targeting PWWP domain of DNA methyltransferase 3B for epigenetic cancer therapy". In preparation. ▪ (FA) "Rationalization of the inhibition of DNA (cytosine-5) methyltransferases by maleimide derivatives of RG108 as non-nucleoside inhibitors". *Future Medicinal Chemistry.* Accepted. ▪ (FA) "Human DNA (cytosine-5) methyltransferases: a functional and structural perspective for epigenetic cancer therapy". *Biochimie*. June 2017. doi.org/10.1016/j.biochi.2017.06.003 ▪ (CA) "Synthesis and Crystallographic Characterization of a Maleimide Derivative of Tryptamine". *Crystals*. November 2016. doi.org/10.3390/cryst6110153 ▪ (FA) "Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B". *Journal of Structural Biology*. June 2016. doi.org/10.1016/j.jsb.2016.03.013 ▪ (CA) "Synthesis, evaluation and structure-activity relationship of new 3-carboxamide coumarins as FXIIa inhibitors". *European Journal of Medicinal Chemistry.* March 2016. doi.org/10.1016/j.ejmech.2016.01.023