

**BIOCHEMICAL INVESTIGATION OF THE *DE NOVO* DNA
METHYLTRANSFERASES DNMT3A AND DNMT3B**

by

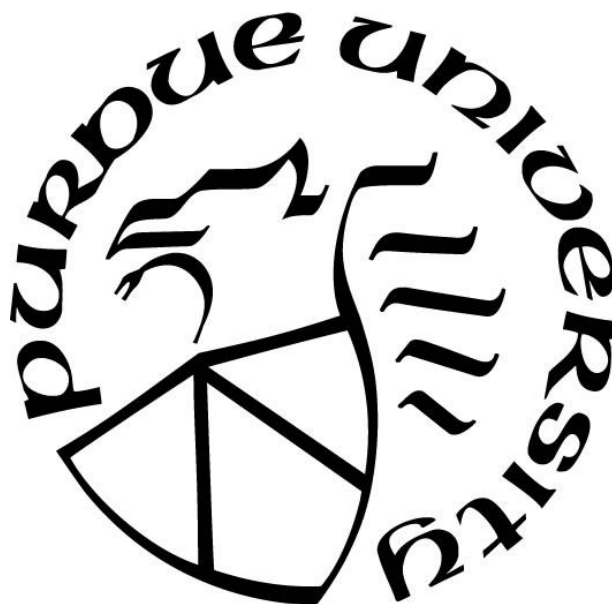
Allison Briggs Norvil

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy



Department of Biochemistry

West Lafayette, Indiana

August 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Humaira Gowher, Chair

Department of Biochemistry

Dr. Andrew Mesecar

Department of Biological Sciences

Dr. Frederick Gimble

Department of Biochemistry

Dr. Chongli Yuan

Department of Chemical Engineering

Approved by:

Dr. Andrew Mesecar

This dissertation is dedicated to my family and friends. I could not have done this without you.

ACKNOWLEDGMENTS

I would like to acknowledge all of those who have helped me develop as an independent scientist during my years as a graduate student:

My family and friends for their love and support, and understanding the commitments attached to research,

Dr. Humaira Gowher for being a wonderful mentor and friend,

My Thesis Committee Members, Drs. Andrew Mesecar, Frederick Gimble, and Chongli Yuan for feedback on my thesis work, mentorship, and helping me grow as an independent scientist,

Current and past lab members Drs. Christopher Petell, Qianyi Yang, Ming He, Debapriya Saha, Mohd Saleem Dar, Lama AlAbdi, and graduate students Isaiah Mensah and Sarah McGovern for their support, insight, and helpful discussions,

The undergraduate researchers that I was privileged enough to mentor who helped me grow and significantly contributed to my work, Amie Michie, Nicole Forstoffer, Thomas Audley, Sophia Tu, Alycia Perez-Johnson, Michelle Ramirez, and Mark Gutay,

I would also like to thank Dr. Phillip San Miguel for performing and analyzing the bisulfite-sequencing for the publication,

Our collaborators Drs. Andrew Mesecar, Jeremy Day, Mohamed Seleem, Abdelrahman Mayhoub, Tasneem Khanam,

And my colleagues who supported me and whose scientific discussions helped me grow as a scientist,

Lastly, I would like to thank several funding sources:

Department of Biochemistry, College of Agriculture, Purdue University for providing the optimal environment and financial support through the Bird Stair Fellowship that went to completing this work, and the Beach Travel Grant that helped fund attending conferences,

National Institute of Health for the research project grant (1R01GM118654-01A1) that supported me and the funded the projects in these publications.

TABLE OF CONTENTS

LIST OF TABLES	9
LIST OF FIGURES	10
ABSTRACT	12
CHAPTER 1. INTRODUCTION	13
1.1 The genome	13
1.2 Regulation of Gene Transcription	13
1.3 DNA Methylation	14
1.3.1 Prokaryotic DNA Methylation	15
1.3.2 Eukaryotic DNA Methylation	16
1.4 DNA Methyltransferases	17
1.4.1 DNMT1 Maintenance Methyltransferase	18
1.4.2 The DNMT3 <i>De Novo</i> Methyltransferases	19
1.5 Structural Comparison of the DNMT3 Enzymes	20
1.6 Kinetics mechanisms of the DNMT3 Enzymes	21
1.6.1 DNMT3A is a Distributive Enzyme that Functions in a Cooperative Manner	22
1.6.2 Sequence Specificity of DNMT3 Enzymes	22
1.7 Disease Causing Mutations to the DNMTs	23
1.7.1 ICF Syndrome	23
1.7.2 DNMT Mutations in Cancer	24
1.8 Methods to Detect DNMT Activity	25
1.8.1 Bisulfite Sequencing	26
1.8.2 Biotin-Avidin Elisa Plate Assay	26
1.9 Summary	27
1.9.1 The Kinetics Mechanism and Catalytic Properties of DNMT3B	27
1.9.2 Significance	28
CHAPTER 2. DNMT3B METHYLATES DNA BY A NONCOOPERATIVE MECHANISM, AND ITS ACTIVITY IS UNAFFECTED BY MANIPULATIONS AT THE PREDICTED DIMER INTERFACE	29
2.1 Declaration of collaborative work	29

2.2	Abstract	29
2.3	Introduction.....	30
2.4	Materials and Methods.....	32
2.4.1	Protein Purification.....	32
2.4.2	Enzyme Assays	33
2.4.3	Cooperativity Assays	34
2.4.4	Stimulation of DNMT3B-C by the Inactive Mutant Glu703Ala.....	35
2.4.5	Processivity Assays	35
2.4.6	DNA Binding Assays	35
2.4.7	Data Analysis.....	36
2.5	Results.....	36
2.5.1	Investigating the Cooperativity of DNMT3B-C.....	38
2.5.2	DNMT3B-C is Not Stimulated by the Catalytically Inactive Mutant	41
2.5.3	DNMT3B-C and Full Length DNMT3B Methylate DNA in a Processive Manner..	42
2.5.4	Processivity of the Initial Phase of the Reaction	46
2.5.5	Effect of Pre-incubation with DNA on Catalytic Activity	47
2.5.6	Order of Addition for Steady State Kinetics Analysis.....	48
2.5.7	Activity and Processivity of DNMT3B-C are Not Affected by pH	49
2.5.8	Influence of pH on the Dissociation Constant.....	52
2.5.9	The Processivity of DNMT3B-C is Not Affected by Mutation of the Arg829His Residue	53
2.6	Discussion.....	57
CHAPTER 3. THE ACUTE MYELOID LEUKEMIA VARIANT DNMT3A ARG882HIS IS A DNMT3B-LIKE ENZYME.....		61
3.1	Declaration of collaborative work	61
3.2	Abstract.....	61
3.3	Introduction.....	62
3.4	Materials and Methods.....	64
3.4.1	Protein Purification.....	64
3.4.2	DNA Methylation assays using radiolabeled AdoMet	65
3.4.3	Processivity Assays	65

3.4.4	Cooperativity Assays	65
3.4.5	Flanking Sequence Preference	66
3.4.6	Bisulfite Sequencing	67
3.4.7	Rescue Experiments in mESCs	68
3.4.8	Comparative Flanking Sequence Preference Analysis	68
3.4.9	Individual Flanking Sequence Preference Analysis	71
3.4.10	Equations	73
3.4.11	Consensus Sequence Analysis of Major and Minor Satellite Repeats	73
3.4.12	Line and Bar Graphs	74
3.5	Results	74
3.5.1	The cooperative kinetics mechanism is absent in DNMT3A Arg882 variants	74
3.5.2	Loss of cooperativity modulates flanking sequences preference of DNMT3A	76
3.5.3	Flanking sequence of DNMT3B has similarity to that of DNMT3A Arg882His	79
3.5.4	The DNMT3A Arg882His variant acquires DNMT3B-like substrate preference	84
3.5.5	Temporal change in flanking sequence preference by WT and variant DNMT3 enzymes	88
3.5.6	DNMT3A-C Arg882His and DNMT3B-C preferably methylate the same CpG site in the <i>Meis1</i> enhancer	93
3.5.7	Mouse DNMT3A Arg878His retains its activity for minor satellite DNA in mESCs ..	94
3.6	Discussion	99
CHAPTER 4. EFFECT OF DISEASE-ASSOCIATED GERMLINE MUTATIONS ON STRUCTURE FUNCTION RELATIONSHIP OF DNA METHYLTRANSFERASES.....		102
4.1	Declaration of collaborative work	102
4.2	Abstract	102
4.3	Introduction	102
4.4	Structural and functional alterations of DNMT1 by disease associated mutations	104
4.4.1	HSAN1E	107
4.4.2	ADCA-DN	108
4.5	Structural and functional alterations of DNMT3 by disease associated mutations	109
4.5.1	Tatton-Brown-Rahman syndrome	110

4.5.2 Hereditary tumors and Microcephalic dwarfism	116
4.6 Conclusion and Perspectives.....	118
CHAPTER 5. PERSPECTIVES	121
5.1 Outlook on Characterizing the Biochemical Properties of the DNMT3 Enzymes	121
5.1.1 Conclusions.....	121
5.1.2 TRD and Dimer Interface in Substrate Specificity.....	122
5.1.3 Post Translational Modifications to the DNMT3 Enzymes	125
5.2 Adaption of the Biotin-Avidin ELISA Plate Assay to Measure RNA Methyltransferase Activity	126
APPENDIX A. VEZF1	128
APPENDIX B. UPPP AND UPPS	156
REFERENCES	185
VITA	214
PUBLICATIONS.....	221

LIST OF TABLES

Table 2.1 Oligomers used for DNA methylation analysis.....	34
Table 3.1 Table containing the sequences of each 30-bp substrate used to determine substrate specificity at the N+2 and N+3 nucleotide positions.	67
Table 3.2 Data analysis for Figures 3.4B and 3.6A.	69
Table 3.3 Data for Figure 3.9A, DNMT3A-C WT at 10 minutes on the 509-bp substrate.	72

LIST OF FIGURES

Figure 2.1 Expression of DNMT3A-C, DNMT3B-C and their respective mutants.....	37
Figure 2.2 Allosteric activation of the DNMT3 enzymes.....	38
Figure 2.3 Catalytic mechanism of DNMT3B-C.....	40
Figure 2.4 DNMT3B-C Glu703Ala is catalytically inactive.	41
Figure 2.5. DNMT3B-C is not cooperatively stimulated by addition of inactive DNMT3B-C Glu703Ala.....	42
Figure 2.6 DNMT3B-C exhibits processivity on DNA substrate with 2 CpG sites.	45
Figure 2.7 Variable processivity conditions of M.SssI and DNMT3B-C.....	46
Figure 2.8 Pre-incubation with DNA decreases the methylation activity of DNMT3B-C.....	48
Figure 2.9. pH sensitivity of the DNMT3B-C processive catalysis.	51
Figure 2.10 Effect of pH on DNA binding by DNMT3B-C.....	53
Figure 2.11. Loss of catalytic activity of the DNMT3A-C Arg878His variant.....	55
Figure 2.12 The Arg829His mutation of DNMT3B-C does not reduce its activity and processivity..	56
Figure 2.13 Illustration of the histidine bases that are not conserved in DNMT3B.	59
Figure 2.14 . Model showing the different kinetics mechanisms of DNMT3A-C and DNMT3B-C.....	60
Figure 3.1 Coomassie-stained SDS-PAGE gel showing purified recombinant His-tagged DNMT3 enzymes.....	75
Figure 3.2 Relative activity and kinetics mechanism of DNMT3A-C WT and Arg882 variants.	75
Figure 3.3 Bisulfite sequencing of the 509-bp substrate.....	77
Figure 3.4 Effect of cooperativity on flanking sequence preference by DNMT3A-C..	78
Figure 3.5 Methyltransferase activity of wild type and variant DNMT3 enzymes on multiple substrates.....	81
Figure 3.6 Comparative analysis of the flanking sequence preferences of DNMT3A-C WT, DNMT3A-C Arg882His and DNMT3B-C WT.....	83
Figure 3.7 Flanking sequence preference at steady state kinetics.	85
Figure 3.8 Enrichment for each relative site preference..	86

Figure 3.9 Trinucleotide sequence flanking CpG preferred by WT and variant DNMT3 enzymes.	89
Figure 3.10 Methylation activity using short DNA substrates with varying N +2 and N +3 nucleotide sequences with T at the N +1 position.	92
Figure 3.11 Relative activity and site preference of DNMT3 WT and mutant enzymes on <i>Meis1</i> enhancer substrate.	94
Figure 3.12 Rescue of DNA methylation at the major and minor satellite repeats in <i>DNMT3A/3B</i> DKO mESCs.	96
Figure 3.13 Model showing the different kinetics mechanisms of DNMT3A and DNMT3B and their influence their flanking sequence preferences.	99
Figure 4.1 HSN1E and ADCA-DN mutations in DNMT1	106
Figure 4.2 TBRS mutations in the PWWP and ADD domain of DNMT3A	111
Figure 4.3 TBRS mutations in the catalytic domain of DNMT3A	114
Figure 4.4 PCC/PGL and MD mutations in PWWP domain of DNMT3A	117
Figure 5.1 DNMT3A target recognition domain interacting with DNA.	123
Figure 5.2 DNMT3B target recognition domain interacting with DNA..	124

ABSTRACT

DNA methylation is an epigenetic modification that is nearly ubiquitous. Eukaryotic DNA methylation contributes to the regulation of gene expression and maintaining genome integrity. In mammals, DNA methylation occurs primarily on the C5 carbon of cytosine in a CpG dinucleotide context and is catalyzed by the DNA methyltransferases, DNMT1, DNMT3A and DNMT3B. While *dnmt3a* and *dnmt3b* genes are highly homologous, the enzymes have distinct functions. Some previous reports suggested differences in the enzymatic behavior of DNMT3A and 3B, which could affect their biological roles. The goal of my thesis work was to characterize kinetics mechanisms of DNMT3A and 3B, and to identify the similarities and differences in their catalytic properties that contribute to their distinct biological functions. Given the sequence similarity between the enzymes, we asked whether DNMT3B was kinetically similar to DNMT3A. In a series of experiments designed to distinguish between various kinetics mechanisms, we reported that unlike DNMT3A, DNMT3B methylated tandem CpG on DNA in a processive manner. We also reported that the disruption of the R-D interface, critical for the cooperativity of DNMT3A, had no effect on DNMT3B activity, supporting the non-cooperative mechanism of this enzyme.

DNMT3A is frequently mutated in numerous cancers. Acute Myeloid Leukemia (AML) is a malignancy of hematopoietic stem cells in which numerous patients exhibit a high frequency of the heterozygous somatic mutation Arg882His in DNMT3A. Through thorough consensus motif building, we discovered a strong similarity in CpG flanking sequence preference between DNMT3A Arg882His variant and DNMT3B enzyme. Moreover, we found that the variant enzyme has the same kinetics mechanism as DNMT3B, indicating a gain-of-function effect caused by the mutation. This change is significant because the variant enzyme can aberrantly methylate DNMT3B targets in AML cells and effect global gene expression. In particular, given that DNMT3B has been shown to have oncogenic properties, this suggests that the Arg882His variant can acquire similar oncogenic properties and drive AML development.

Taken together, my thesis work provides novel insights into the relationship between the biochemical properties and the biological functions of DNMT3A and 3B.

CHAPTER 1. INTRODUCTION

1.1 The genome

The genetic material of all organisms is made either from DNA or RNA. Genomic DNA in prokaryotes and eukaryotes is organized into genes and intergenic regions that may influence nearby gene expression (1, 2). Through the process of transcription and translation, the information in genes is used for the synthesis of all proteins, which maintains the viability of the cell. Besides the sequence that codes for a protein, genes contain elements that control transcription. At the 5' end of the coding region there is a proximal regulatory sequence known as the promoter, where transcription machinery can bind (3). In higher organisms, the coding region consists of exons, coding information, and introns, non-coding information, and the 3' end of the gene contains a termination sequence that signals for the dissociation of the transcription machinery (4). Many genes also have distal regulatory elements such as enhancers, which serve as the binding site for activator proteins to increase gene transcription (5). The complex organization of the gene allows for precise regulation of its expression, which is necessary for the development of complex organisms.

1.2 Regulation of Gene Transcription

The transcription of DNA into the messenger RNA molecule is the first step in gene expression, which is initiated by a family of proteins known as transcription factors (TFs) (6). These proteins contain at least one DNA-binding domain that recognizes specific DNA sequences (consensus motifs) and assemble on the gene promoter in a sequence dependent manner (7, 8). Once bound, TFs recruit RNA polymerase (RNAP) and its associated transcriptional activators to the promoter through protein-protein interactions that form the transcriptional initiation complex (9). There are three transcriptional stages of the RNAP complex, paused, elongating, and terminating (10, 11). RNAP pausing plays an important role in regulating transcription by allowing for the recruitment of additional protein complexes such as elongation regulators (12). After the paused RNAP complex is released, RNAP performs elongation through uninterrupted synthesis of the nascent RNA molecule (13, 14). Transcription is disrupted at the termination site and RNAP dissociates from the DNA and releases the nascent messenger RNA transcript for processing (15).

Further regulation of gene transcription can occur prior to the initiation stage through the accessibility of regulatory elements. The binding of transcriptional repressors to regulatory elements blocks transcription by preventing RNAP association to the gene (16, 17). While these mechanisms provide direct control over gene expression through TF binding, cell specialization and the need to respond to environmental signals necessitate additional layers of regulating gene transcription.

In order to fine tune and ensure proper gene transcription, epigenetic mechanisms regulate gene expression without changing the DNA sequence. Epigenetic mechanisms in eukaryotes include chemical modifications to the DNA or nearby histones, expression of non-coding RNA, and chromatin conformation. The deposition of a methyl group on DNA is known as DNA methylation and plays an important role in regulating gene expression in bacteria (18). In eukaryotes, 147-bp of DNA wrap around histone octamers to make nucleosomes, which are spaced by 20-30-bp linker DNA. This pattern creates beads on string structure, which condense into a fiber-like nucleoprotein complex called chromatin (19, 20). The histone octamer consists of two copies of each of the core histone proteins (H2A, H2B, H3, and H4), and the wrapped nucleosomes are stabilized by linker histone H1 (21). The N-terminal residues of the histone proteins can be chemically modified to alter the charge state of the chromatin (22-25). These modifications regulate the structure of chromatin making the DNA either accessible (euchromatin) or inaccessible (heterochromatin) to the transcriptional machinery (26, 27). Collectively, these epigenetic mechanisms allow for dynamic control over gene expression that results in changes in an organism or cell's biological function. DNA methylation is largely associated with gene repression when deposited at the regulatory elements (28). Bulk DNA methylation in the genome is found in the pericentromeric and centromeric regions of the heterochromatin where it is critical for the suppression of retroviral particles and transposons (28). Furthermore, the conservation of DNA methylation from bacteria to higher eukaryotes highlights its importance in regulating gene expression during growth and development.

1.3 DNA Methylation

DNA methylation is the covalent addition of a methyl group (CH_3) to the DNA molecule that provides stable epigenetic control of the genetic program (29). The methyl group attached to the DNA protrudes from the major groove of the double helix and may be recognized by some

DNA-binding proteins and regulate gene expression (30). Enzymatic DNA methylation takes place at various positions on adenine and cytosine bases. In bacteria DNA methylation includes N⁶-methyl-adenine (6mA), C⁵-methyl-cytosine (5mC), and N⁴-methyl-cytosine (4mC) (30-32). The roles of 6mA and 5mC in bacterial physiology are well characterized, however very little is known about the function of 4mC. In eukaryotes, DNA methylation occurs almost exclusively as 5mC (33) and is predominantly found in a CpG dinucleotide context. About 70-80% of cytosine residues in CpG dinucleotides are methylated in mammals (34). Methylation of CpG sites is critical for cellular development and differentiation. Non-CpG methylation such as CpA, CpT, and CpC has also been identified in mammals on a genome-wide level in neuronal cells, however its function is not yet elucidated (34). The conservation of DNA methylation in prokaryotes and eukaryotes allows it to play a vital role in growth, development, and environmental response (35, 36).

1.3.1 Prokaryotic DNA Methylation

In prokaryotes, the 6mA and 5mC modifications have two main roles in bacterial physiology: defense against viral infection and regulating gene expression (37). To protect itself from phage DNA incorporation into its genome, bacteria evolved a restriction modification system to differentiate between self and exogenous DNA. The DNA methyltransferases target the host genome for methylation, and the resultant methylation mark blocks the cognate endonuclease activity (38). When phage DNA enters the bacteria, it is unmodified and targeted by the endonucleases for degradation to prevent it from becoming incorporated into the host's genome. This restriction modification system is often considered a primitive immune system in bacteria, and has been adopted for genetic engineering (39-41).

DNA methylation in bacteria, particularly at the 6mA site, also regulates gene transcription. When 6mA is located in the gene promoter, it can modulate the binding of RNA polymerase and transcription factors to regulate the expression of the nearby gene clusters (18, 42). Studies looking at the regulatory mechanisms behind 6mA sites have classified them as either clock-like or switch-like controls (43). During clock-like control, the methylation state of the DNA is used as a signal to couple gene expression to a cell cycle specific stage. For switch-like control, the 6mA pattern is used to regulate gene expression in a positive or negative manner by regulating the interaction of DNA-binding proteins to the DNA (18). In this case, DNA-binding proteins determine the 6mA

landscape by blocking the deposition of 6mA, which creates distinct regions of methylation. Comparatively, less is known about the role of 5mC in regulating bacterial gene expression, however it seems to be necessary for virulence in some pathogenic bacteria (44). Depletion of 5mC in these bacteria lead to increased expression of stress response genes and decreased motility, demonstrating the importance of 5mC for bacterial gene expression (44, 45).

1.3.2 Eukaryotic DNA Methylation

Eukaryotic DNA methylation occurs primarily as 5mC in the CpG dinucleotide context, whereas non-CpG methylation occurs at a much lower rate (46). 5mC levels vary across different taxa, as does the genomic context (46). For example, eukaryotes like fungi have very low levels of 5mC, which is limited to the repetitive elements to keep them in a repressed state (47, 48). Conversely, mammals have much higher levels of genomic methylation, with >70% of CpG sites being methylated (36, 46). When deposited at gene regulatory elements like enhancers and promoters, 5mC is typically associated in transcriptional repression, in part by preventing the association of activating TFs (49, 50). This system is similar to what is observed in bacteria with the switch-like mechanism to regulate gene expression. The highest levels of 5mC are present at the repetitive regions of the genome, which are made up of retroviral sequences and transposable elements (51). In these regions, 5mC prevents the transcription of these transposable elements and contributes to heterochromatin formation by targeting the binding of methyl-binding domain (MBD) proteins (52, 53). The histone modifiers such as methyltransferases and deacetylases in complex with MBD proteins can then work in tandem to condense the chromatin that prevents transcription (54).

Gene bodies also contain 5mC deposition, however its functional role in gene expression is less well known (36). Gene body methylation is typically found in actively transcribing genes and some studies have shown that it prevents spurious transcription outside the transcription start site (55, 56). Further studies have suggested that gene body 5mC influences splicing when it occurs at exon/intron borders by altering the activity of RNAP (57, 58). However, alternative splicing events are controlled by numerous mechanisms, so the direct role of 5mC in this still needs to be fully elucidated (59).

(49). CpG sites are underrepresented in the genome, and it is speculated that this is due to the spontaneous deamination of 5mC to thymine, which increases mutations due to mis-matched

base repair (60). Regions of high CpG density (>60%) do occur in the mammalian genome, known as CpG islands (CGIs), but typically remain unmethylated (61, 62). Over two-thirds of all mammalian promoters contain CGIs, which encompasses almost all ubiquitously expressed housekeeping genes and a few development genes (63, 64). Aberrant CGI hyper-methylation and dysregulation of gene expression is commonly found in many types of cancer, indicating the importance of their unmethylated state in maintaining proper gene expression (49).

1.4 DNA Methyltransferases

DNA methyltransferases are a family of enzymes that catalyze the transfer of the methyl moiety from the donor S-adenosylmethionine (AdoMet) to the target site on the DNA (35). These enzymes follow multi-substrate enzymes kinetics, in which the two substrates, AdoMet and DNA, are converted into two products, S-adenosylhomocysteine (AdoHcy) and methylated DNA, known as a bi bi reaction (65). Multi-substrate kinetics follows either ping-pong or sequential mechanism, which is dictated by the binding order and release of the substrates and products (66). DNA methyltransferases follow a sequential mechanism because the enzyme needs to bind both substrates before the first product can be formed. Sequential substrate binding can happen in an ordered or random manner, which is determined by the preferential binding order of the substrates to the enzyme. Both mechanisms are used by DNA methyltransferases (67) although, DNA methyltransferases with high sequence similarity utilize the same mechanism. For example, the bacterial cytosine C5 methyltransferase M.HhaI is highly homologous to the catalytic domain of the mammalian *de novo* C5 methyltransferase DNMT3 family, and they follow an ordered bi-bi reaction mechanism (68). These intrinsic properties help specialize methyltransferases for their specific function.

The DNA methyltransferases contain ten conserved amino acids motifs that determine their catalytic mechanism (69, 70). These motifs are involved in AdoMet binding (I, II, III), catalysis (IV – VI), and DNA binding (VII – X) (30, 70). Additionally, there is a highly variable region between motifs VIII and IV that defines sequence specificity around the target cytosine known as the target recognition domain (TRD). While these motifs are important for catalytic activity, there are distinct amino acids that play a direct role in catalysis in motifs IV, VI, and VIII. Once the enzyme is bound to the DNA and AdoMet, the target base is flipped into the catalytic site and stabilized by Arg residues in motif VIII (71). For methylation at the C5 position of cytosine base,

the Cys residue of motif IV performs the initial nucleophilic attack on C6 of cytosine (30). The Glu of motif VI interacts with N3 of the cytosine and the amine group of C4 positions the nucleotide in the active site. This stabilizes C5 to perform nucleophilic attack on the methyl moiety of the AdoMet bound to motifs I – III. Finally, C5 is deprotonated and there is a beta elimination of the Cys residue.

Mammalian DNA methyltransferases methylate DNA at C5 position of cytosine largely in CpG dinucleotide context. Two families of mammalian DNA methyltransferases have been discovered: Dnmt1 and Dnmt3. DNMT3 family has three members, DNMT3A, DNMT3B and DNMT3L. (72).

1.4.1 DNMT1 Maintenance Methyltransferase

DNMT1 is the mammalian maintenance methyltransferase that is highly expressed in dividing cells and found at the DNA replication foci (73). It is a multi-modular enzyme comprising an N-terminal domain that contains a replication foci-targeting sequence (RFTS), a DNA binding CXXC domain, two bromo-adjacent homology (BAH) domains, and a C-terminal catalytic domain with an intrinsic preference for hemi-methylated DNA (74, 75). The RFTS and CXXC domains further contribute to specificity through auto-inhibition. The CXXC domain binds to unmethylated CpG sites and pushes an autoinhibitory linker between the DNA and the catalytic site to prevent interaction (76). The autoinhibition by the RFTS domain is relieved through the interaction with UHRF1 (ubiquitin-like, containing PHD and RING finger domains 1), which binds hemi-methylated DNA sites (77). While the direct mode of transfer is unknown, it has been suggested that UHRF1 passes the hemi-methylated DNA to the DNMT1 active site once RFTS has been displaced. Despite being conserved in all mammalian DNMT1 homologs, little is known about the role of the BAH domains (78). Some BAH domains have been reported to bind to histone tails in a modification dependent manner, so it has been suggested that these domains may play a role in DNMT1 localization (79). The BAH domains are required for DNMT1 to localize to the replication fork, but the mechanism by which it accomplishes this remains unknown (78).

Post replication, CpG sites are hemimethylated and DNMT1 recognizes hemimethylated DNA at 100 fold higher efficiency than unmethylated sites (80). This feature allows the enzyme to copy the methylation from parent to the daughter strand. DNMT1 is a highly processive enzyme that can methylate up to 30 CpG sites before dissociating from the DNA (80). DNMT1 diffuses

along the DNA through electrostatic interactions and methylates each target site it encounters (80, 81). Studies have shown that this mechanism prevents DNMT1 from swapping DNA strands, allowing it to keep its orientation during catalysis (80, 82). While DNMT1 has a high preference for hemi-methylated DNA, it is also capable of performing *de novo* DNA methylation at CCGG sites, and potentially plays a role in methylating CGIs during development (80, 83, 84). The mechanism for this reaction remains unknown, but it has been suggested that DNMT1 does this by cooperating with the *de novo* DNMT3 family of methyltransferases (85).

1.4.2 The DNMT3 *De Novo* Methyltransferases

The *de novo* DNA methyltransferases responsible for establishing methylation patterns in mammals are the DNMT3 enzymes (86-88). This family of enzymes consists of two catalytically active members, DNMT3A and DNMT3B, and a catalytically inactive member, DNMT3L (87). These enzymes are primarily expressed during early development and during stem cell differentiation to reestablish methylation patterns after a genome-wide loss of methylation (89). DNMT3A and DNMT3B are highly homologous with similar domain organizations that include an N-terminal regulatory region and a C-terminal catalytic domain (90). These enzymes have a high preference for unmethylated CpG sites and several common target sites (86). However, murine knockouts for DNMT3A or DNMT3B showed that they cannot fully compensate for the loss of one another. DNMT3B knock out mice die at 9 days post implantation, while DNMT3A knock out mice survive to term, but died at 6 weeks old (86). Genome-wide DNA methylation analysis of the knockout mice revealed specific targets for these enzymes. DNMT3A showed a preference to methylate the major satellite repeats, while DNMT3B preferably methylated the minor satellite repeats. These data suggest that despite their high homology, DNMT3A and DNMT3B have unique biochemical properties that differentiate their substrate specificity.

DNMT3L is inactive as a DNA methyltransferase, but still plays an important role in establishing methylation patterns and organism development (91-93). DNMT3L is specifically expressed during male and female germ cell development and not in somatic cells (94, 95). Knockout DNMT3L mice are viable, however, males are sterile and oocytes from female knockouts show hypomethylation at the maternal imprints (96-99). Global DNA hypomethylation at the repetitive regions was also observed in *DNMT3L* ^{-/-} murine embryonic stem cells (mESCs). Co-immunoprecipitation experiments showed that DNMT3L interacts with DNMT3A and

DNMT3B, and biochemical studies found that these interactions stimulate the catalytic activity of DNMT3A and DNMT3B (93, 97, 100). Therefore, despite being catalytically inactive, DNMT3L is critical for establishing DNA methylation patterns in germ cells and during early development.

1.5 Structural Comparison of the DNMT3 Enzymes

DNMT3A and DNMT3B are highly homologous with roughly 40% sequence identity and have similar domain organizations. Their N-terminal regulatory region consists of two domains, a Pro-Trp-Trp-Pro (PWWP) domain followed by a Cys-rich Zn-binding region called the ARTX-DNMT3-DNMT3L (ADD) domain (90). The C-terminal region consists of the methyltransferase domain (MTase). The PWWP domain targets DNMT3A and DNMT3B activity by interacting with the DNA and histone H3 methylated at the Lysine 36 residue (H3K36me_{2/3}) (101-103). The PWWP domain of DNMT3A targets both H3K36me₂ and H3K36me₃, with a preference for H3K36me₂, to be recruited to genic/intergenic regions and gene bodies, respectively (104, 105). The PWWP domain of DNMT3B binds to H3K36me₃ to target methylation at gene bodies of actively transcribing genes by DNMT3B (56). The ADD domain of DNMT3A binds to the histone H3 at lysine 4 when it is unmethylated (H3K4me₀) to recruit DNMT3A to gene regulatory elements. In the absence of an unmodified H3K4 tail, the ADD domain interacts with the DNA binding region of the MTase domain and autoinhibits catalytic activity (106). This prevents DNA methylation from being deposited at regions adjacent to methylated H3K4 sites, which are typically associated with active gene transcription. This mechanism was only shown for DNMT3A, but has been speculated for DNMT3B as well (106, 107).

The MTase domains of DNMT3A and DNMT3B contain the conserved motifs also found in bacterial DNA methyltransferases. Motifs I – III are involved in AdoMet binding, while motifs IV and VI are required for the catalysis. The region between motifs VIII and IX are used for target recognition (TRD) and binds DNA. The MTase domain of DNMT3A and DNMT3B have >80% sequence similarity and can function independently of their N-terminal region (108). The highest variability between the MTase domains of DNMT3A and DNMT3B occurs in the TRD. The TRD binds to the major groove of the DNA for a direct readout of the target sequence to dictate substrate specificity (65). Conversely, DNMT3L lacks motifs IX and X and the TRD, making it unable to bind DNA and catalytically inactive (107).

Co-crystallization of DNMT3A and DNMT3L revealed an elongated, butterfly shaped hetero-tetrameric structure (109). Two DNMT3A monomers form the center of the complex that are flanked by two DNMT3L monomers on either side, forming a 3L – 3A – 3A – 3L structure. The DNMT3A – 3L interaction is mediated by two Phe residues, called the F-F interface. In the absence of DNMT3L, DNMT3A forms homo-tetramers, which interact to form long oligomers (55, 110). The DNMT3A-3A interaction occurs primarily through two pairs of Arg and Asp residues at the RD interface. Additionally, two His residues interact through base stacking, and a hydrophobic pocket is formed with two pairs of Trp and Asn residues to further stabilize the RD interface (111). The importance of the DNMT3A RD interface is highlighted by the 40 – 100 % loss in catalytic activity of the variant enzymes with mutations in the interface residues. Co-crystallization of DNMT3B with DNMT3L revealed a similar hetero-tetrameric structure mediated by conserved residues, however almost no oligomerization was observed (55, 112). These data role of variant residues in DNMT3A and DNMT3B in oligomerization process.

1.6 Kinetics mechanisms of the DNMT3 Enzymes

DNA methyltransferases use various kinetics mechanisms to methylate multiple target sites on DNA. In a distributive mechanism, the enzyme locates its target sites in a random diffusion process and after methylating the target base dissociates from methylated DNA. (113). Processive enzymes can bind locate target sites by sliding along the DNA and after methylating the target base stay bound to the DNA sliding across to find the next target. (114-116). A perfectly processive enzyme would methylate all sites on a DNA molecule before dissociating whereas the distributive enzyme dissociates and randomly binds to the same or a different DNA molecule after every catalytic event. (116). Typically, processive enzymes are considered advantageous for multiple turnover kinetics, although some distributive enzymes mitigate this by oligomerizing on DNA to methylate multiple targets in a cooperative manner. Cooperativity occurs when the binding of one enzyme to the substrate alters the enzyme's structure to allow for the second enzyme to bind efficiently, increasing catalytic activity (117)(118).

1.6.1 DNMT3A is a Distributive Enzyme that Functions in a Cooperative Manner

DNMT3A oligomerization along the DNA has been observed both *in vitro* and *in vivo*, and this mechanism was suggested to support a cooperative mechanism for methylating multiple CpG sites (109, 119). Mutations in TRD of DNMT3A that disrupt oligomerization but not DNA binding cause >2-fold loss of catalytic activity, supporting a cooperative mechanism (120). Biochemical studies were performed using long and short DNA substrate which could accommodate multiple or single enzyme binding respectively (118). The methylation rate on a long DNA substrate had an exponential relationship with the concentration of DNMT3A, while the short substrate had a linear relationship. This data supports a cooperative mechanism in which as higher enzyme concentrations stimulates the catalytic activity. Further studies also confirmed that at lower concentrations, DNMT3A methylated multiple CpG sites using a distributive catalytic mechanism.

1.6.2 Sequence Specificity of DNMT3 Enzymes

While DNMT3A and DNMT3B preferably methylate CpG sites over non-CpG sites, data from the Human Epigenome Project suggested that these enzymes have a flanking sequence preference that extends beyond the target site (121, 122). Regions with highest levels of DNA methylation were analyzed for nucleotide sequences flanking the target CpG up to ± 4 base-pairs. Sites with highest methylation had 5'-CTTGCGCAAG-3' flanking sequence, while the lowest methylated sites had 5'-TGTTTCGCGTGG-3' flanking sequence. Multiple studies showed that DNMT3A prefers CpG sites flanked by pyrimidines (5'-YNCGY-3'), particularly at the N+1 position (5'-CGY-3'). Conversely, DNMT3B had a preference for purines at the N+1 (5'-CGR-3') compared to DNMT3A.

Co-crystallization of DNMT3A with DNA demonstrated a mechanism by which the flanking sequence could influence catalytic activity (123). DNMT3A makes direct contacts with the DNA from the N-2 position to the N+6 position. The TRD mediates interactions on the 3' end of the CpG site by adopting an ordered structure that is stabilized by an Arg residue, forming a hydrogen bond network with the surrounding amino acids (112). Crosstalk between the residues of the TRD and RD interface create a hydrogen bond network that stabilizes protein-DNA interactions (124). These data suggest a mechanism by which mutations in the R-D interface result in a reduced DNA binding. While other regions of DNMT3A also interact with the DNA, such as the catalytic loop,

these interactions are only involved in CpG recognition. The TRD and RD interface play major roles in dictating the flanking sequence preference.

1.7 Disease Causing Mutations in the DNMTs

Aberrant DNA methylation patterns are a hallmark of several diseases and all types of cancer, which can be caused by either mis-regulation of DNMT expression or loss of function (49, 125). Advances in whole genome association studies have helped identify mutations of the DNMT enzymes in various diseases that alter biological functions and catalytic activity of these enzymes. Although mutations have been found across DNMT genes, most mutations tend to cluster in their functional domains. Among the DNMTs, disease-causing mutations were first discovered in DNMT3B in patients with immunodeficiency, centromeric instability, and facial anomalies (ICF) syndrome. Advances in whole-genome sequencing studies have identified numerous somatic mutations in DNMT1 and DNMT3A that cause colorectal carcinoma and acute myeloid leukemia (AML), respectively. Many of these mutations have been well characterized and their impact on DNMT catalytic activity is partially understood.

1.7.1 ICF Syndrome

ICF syndrome is a rare, recessive disease that induces chromosome breakage and DNA rearrangements (126). Symptoms of the disease include mild facial dysmorphism, growth retardation and immunoglobulin deficiency, and patients often die in infancy or childhood (127, 128). ICF patients show DNA hypomethylation at the satellite sequences centromere-adjacent in heterochromatin, which is a specific target of DNMT3B. This disease is typically caused by mono- or bi-allelic missense mutations in DNMT3B that can occur throughout the gene (129). In a total of 50 ICF patients identified worldwide, about 15 distinct DNMT3B variants have been reported, with the majority of these substitutions occurring as missense mutations (126). In the PWWP domain of DNMT3B, only one substitution was identified, Ser282Pro, which impairs H3K36me3 binding and DNMT3B localization (130, 131).

Other ICF DNMT3B variants are found in the MTase domain near the conserved catalytic motifs and can be grouped based on their effect on enzyme function (112, 126). Three mutations are located near motifs I and II (AdoMet binding) and three other mutations are located in motifs

IV and VI (catalysis). Eight mutations are located in motifs involved in DNA binding, four near the DNA binding regions of motifs VIII and X, and the TRD, and four at the RD interface. These DNMT3B ICF variants have reduced catalytic activity, and some perturbed the enzyme's ability to interact with and be stimulated by DNMT3L, providing a mechanism for the hypomethylation patterns observed in ICF patients (129, 132).

1.7.2 DNMT Mutations in Cancer

Somatic DNMT1 mutations have been identified in patients with colorectal cancer (133). Two different mutations were detected through exome sequencing: a single nucleotide deletion that results in a stop codon causing truncation of the MTase domain and one amino acid substitution in the MTase domain. The biochemical effects of the amino acid substitution, Tyr-to-Cys, on DNMT1 structure and catalytic activity are not known. However, both mutations lead to genome-wide alterations in DNA methylation in colorectal cancer patients, (133, 134). Interestingly, none of the patients with hereditary neurological disorders harboring DNMT1 germline mutations developed colorectal cancer (135, 136). These observations demonstrate the different effects of germline and somatic DNMT1 mutations on growth and development.

Besides being a *de novo* MTase and establishing DNA methylation patterns during development, DNMT3A also plays an important role in somatic stem cell differentiation. In hematopoietic stem cells (HSCs), DNMT3A is recruited to the regulatory elements of HSC specific genes for DNA methylation. Knocking out DNMT3A in HSCs leads to a decrease in cell differentiation and increased self-renewal (137). These cells display significant genome-wide hypomethylation with focal areas of hypermethylation (125, 138). The phenotype of these cells is similar to what is observed in human hematological malignancies harboring DNMT3A loss of function mutations.

Genome studies identified somatic DNMT3A mutations in 22% of adult patients with acute myeloid leukemia (AML), which are associated with poor prognosis (139, 140). Over 50 unique mutations have been mapped to all three domains of DNMT3A, and include a range of nonsense, frameshift, and missense mutations. The PWWP and ADD domains harbor the majority of the nonsense and frameshift mutations, while the MTase domain mostly acquires missense mutations. Similar to the DNMT3B mutations found in patients with ICF, most of the DNMT3A MTase mutations mapped to the conserved motifs, RD interface, or the TRD.

In AML, about 65% of DNMT3A heterozygous missense mutations affect Arg882, with the majority being an Arg-to-His substitution. This variant is typically found in older patients with AML that had significantly lower rates of overall and disease-free survival (141). The DNMT3A Arg882His variant typically arises during the early stages of cancer development and causes a 40-80% loss of catalytic activity (142, 143). The co-crystallization of DNMT3A with DNA showed that Arg882 is located close to the R-D interface and interacts with the phosphate backbone of the nucleotide at the N+3 position (123). Biochemical studies found that this variant disrupts intermolecular interactions at the R-D interface and decreases DNA binding (110, 111). Additionally, DNMT3A Arg882His was shown to have an altered flanking sequence preference around the target CpG site compared to wild type DNMT3A (144). Based on its prominent effect on genomic methylation levels, the Arg882His variant is suggested to have a dominant-negative effect on the wild type enzyme, however this was not confirmed by *in vitro* experiments (142, 143, 145). Other variants of Arg882 also occur in AML patients, such as Arg882Cys, Arg882Ser, and Arg882Pro, which have similarly reduced catalytic activity and altered flanking sequence preferences (146).

Over 10 other DNMT3A missense mutations have been identified in AML patients, but they occur at lower frequencies compared to Arg882His. *In vitro* studies uncovered a range of functional consequences of other DNMT3A variants, which displayed reduced stimulation by DNMT3L and decreased catalytic activity (147). *Dnmt3a*^{-/-} mouse embryonic stem cells (mESCs) expressing these variants showed widespread hypomethylation, suggesting that other DNMT3A missense mutations have a similar effect on biochemical properties as the Arg882His variant.

1.8 Methods to Detect DNMT Activity

Several tools have been developed to quantify DNA methylation and detect DNMT activity. *In vitro* DNA methylation assays allow for characterization of DNMT enzymatic properties without the influence of other cellular complexes. Additionally, DNMT assays provide a platform for screening putative DNMT inhibitors as therapeutics for cancer or other diseases (148).

1.8.1 Bisulfite Sequencing

Bisulfite sequencing is the gold standard technique used for detecting DNA methylation. It is based on the selective deamination of cytosine residues by treating the DNA with sodium bisulfite (149, 150). Sodium bisulfite solution will deaminate cytosine to uracil but will not alter 5-methylated cytosine. The modified DNA is PCR amplified and sequenced. This method is commonly used to study DNA methylation *in vivo* (151, 152), however it can also be used to study DNMT activity *in vitro* (144). During *in vitro* DNMT assays, the DNA substrate is methylated by the DNMT enzymes in the presence of the methyl group donor AdoMet, and the reaction is stopped by flash freezing with liquid nitrogen. The DNA purified from *in vitro* reaction or genomic DNA purified from cells is subjected to bisulfite conversion. The sample is then purified and amplified by PCR with primers specific for the bisulfite converted DNA. During this process, the uracil residues will be amplified as thymine, while 5-methylcytosine will be amplified as cytosine. The DNA is sequenced and compared to the reference DNA sequence using specific alignment programs such as Bowtie to determine methylated cytosine bases. Genome wide bisulfite sequencing has revealed changes methylation patterns across mammalian genome during development and in various diseases, thus contributing immensely in our understanding of the role of DNA methylation in various biological processes.

1.8.2 Biotin-Avidin Elisa Plate Assay

The biotin-avidin microplate assay is a medium-throughput method to measure methylation activity of DNMTs (153, 154). The methylation reaction is performed using ^3H -labeled AdoMet as the methyl group donor, which is transferred to a biotinylated DNA substrate by the DNMTs. The reaction is quenched by high concentrations of unlabeled AdoMet and DNA substrate is immobilized on avidin coated ELISA plate. The wells are washed with a high salt buffer to remove excess enzyme and ^3H -labeled AdoMet. The ^3H -labelled CH₃ group is detected by liquid scintillation counting and converted into the amount of methylation in nmoles. This method is extremely sensitive and can detect very low DNMT activity in an inexpensive and robust manner. This method has been used to characterize the kinetics mechanisms of the DNMTs and test the activity of DNMT variants.

1.9 Summary

1.9.1 Elucidating the Catalytic Properties of DNMT3B and the DNMT3A Arg882His variant

Our studies showed that DNMT3B has a unique kinetics mechanism compared to DNMT3A, but has similar biochemical properties as the DNMT3A Arg882His variant found in AML patients. We demonstrated that DNMT3B is a non-cooperative enzyme and methylates DNA in a processive manner. Most residues in the R-D interface that are critical for DNMT3A activity are conserved in DNMT3B, but perturbations of the R-D interface had no effect on DNMT3B. This suggests that these interface interactions are not critical for DNMT3B activity, unlike DNMT3A. Based on these data, we investigated the effect of the AML variant, DNMT3A Arg882His, on the kinetics mechanism of DNMT3A. The Arg882 residue is located in the R-D interface and plays a role in DNMT3A tetramer formation. We showed for the first time that the Arg882His substitution disrupts the cooperative mechanism of DNMT3A. This residue is conserved in DNMT3B (Arg829), and we predicted that the Arg829His substitution will have no effect on DNMT3B due to its non-cooperative mechanism. Our data show that DNMT3B wild type and the Arg829His variant have similar levels of catalytic activity and processivity. Taken together, we propose that DNMT3A and DNMT3B have distinct kinetics mechanisms, which could contribute to their target specificity observed *in vivo*.

A previous study showed that the DNMT3A Arg882His variant has an altered flanking sequence preference compared to the wild type enzyme. We hypothesized that the loss of cooperativity in the Arg882His variant modulates the flanking sequence preference. Under conditions where cooperativity is disfavored, we found that wild type DNMT3A has a similar flanking sequence preference to the Arg882His variant. Since DNMT3B is a non-cooperative enzyme, we anticipated its flanking sequence preference to be similar to the DNMT3A Arg882His variant. Our data showed similar flanking sequence preference between DNMT3B and DNMT3A Arg882His and suggested that the Arg882His variant could target DNMT3B specific sites *in vivo*. We performed rescue experiments in *DNMT3A/3B* double knockout mESCs, which showed that the corresponding Arg878His mutation in mouse DNMT3A fails to rescue DNA methylation at the major satellite repeats, a DNMT3A specific site. Conversely, the Arg878His mutation had no effect on its ability to methylate the minor satellite repeats, a DNMT3B specific target. Analysis of the consensus sequences of the major and minor satellite repeats revealed a previously

unappreciated flanking sequence bias, consistent with DNMT3A and DNMT3B specificity. Taken together, these data suggest that DNA methylation patterns are guided by the sequence preferences of these enzymes.

1.9.2 Significance

Although an extensive investigation has shown that the DNMTs play important roles in development and disease formation, there are still unsolved questions about the kinetics properties of DNMT3A and DNMT3B (155). Our studies demonstrate the fundamental differences between these homologous enzymes that contribute to their target specificity, and suggest that intrinsic specificity of DNMTs guides genome wide methylation patterns. Mutations in either DNMT3A or DNMT3B have been reported in several diseases, and aberrant DNMT expression can be found in nearly all types of cancer. Therefore, understanding the biochemical properties of the DNMT3 enzymes and their variants is paramount for our understanding of disease development. We showed that the DNMT3A Arg882His variant gains DNMT3B-like activity and can target DNMT3B specific sites, suggesting a mechanism for the global hypomethylation and focal hypermethylated regions observed in AML patients. Our data also highlight the importance of the cooperative mechanism for DNMT3A activity and suggest that disrupting the kinetics mechanisms of the DNMT3 enzymes contributes to disease pathogenesis. In the long term, the biochemical properties determined in these studies for the DNMT3 wild type and variant enzymes could be used to discover new therapeutics to rescue DNMT3 activity and restore proper DNA methylation patterns.

CHAPTER 2. DNMT3B METHYLATES DNA BY A NONCOOPERATIVE MECHANISM, AND ITS ACTIVITY IS UNAFFECTED BY MANIPULATIONS AT THE PREDICTED DIMER INTERFACE

*Formatted for dissertation from the article published in *Biochemistry*

Reprinted with permission from *Biochemistry* 2018, 57, 29, 4312–4324. Copyright 2016 American Chemical Society.

Allison B. Norvil, Christopher J. Petell, Lama AlAbdi, Lanchen Wu, Sandra Rossie, Humaira Gowher.

2.1 Declaration of collaborative work

Dr. Humaira Gowher and Allison B. Norvil designed experiments.

Allison B. Norvil, Lanchen Wu, Christopher J. Petell, and Lama AlAbdi performed experiments.

Dr. Humaira Gowher, Dr. Sandra Rossie, Allison B. Norvil, and Christopher J. Petell performed data analysis and wrote the manuscript.

2.2 Abstract

The MTase domains of the *de novo* DNA methyltransferases, DNMT3A-C and DNMT3B-C are highly homologous. However, their unique biochemical properties could potentially contribute to differences in the substrate preferences or biological functions of these enzymes. DNMT3A-C forms tetramers through interactions at the R-D interface, which also promote multimerization on DNA and cooperativity. Similar to processive enzymes, cooperativity allows DNMT3A-C to methylate multiple sites on the same DNA molecule, however it is unclear whether DNMT3B-C methylates DNA by cooperative or processive mechanism. The importance of the tetramer structure and cooperative mechanism is emphasized by the observation that Arg882His mutation in the R-D interface of DNMT3A is highly prevalent in acute myeloid leukemia (AML), and leads to a substantial loss of its activity. Under conditions that distinguish between cooperativity and processivity, we show that in contrast to DNMT3A-C, the activity of DNMT3B-C is not cooperative and confirm processivity of the DNMT3B-C and the full length DNMT3B enzyme. Whereas the mutation Arg878His (mouse homolog of Arg882His) led to the loss of cooperativity of DNMT3A-C, the activity and processivity of the analogous DNMT3B-C Arg829His variant

was comparable to WT enzyme. Additionally, buffer acidification that attenuates the R-D interface interactions of DNMT3A-C had no effect on DNMT3B-C activity. Taken together, these results demonstrate an important mechanistic difference between DNMT3B and DNMT3A and suggest that interactions at the R-D interface may play a limited role in regulating DNMT3B-C activity. These new insights have potential implications for the distinct biological roles of DNMT3A and 3B.

2.3 Introduction

In mammals, DNA methylation at the C5 position of cytosine bases is catalyzed by three known DNA methyltransferases (DNMTs): DNMT1, DNMT3A, and DNMT3B. Although mammalian DNMTs use a conserved catalytic mechanism that involves base flipping to methylate largely at CpG dinucleotides, these enzymes are structurally and functionally distinct (35, 72, 75, 156). DNMT1 is primarily responsible for the post-replication maintenance of DNA methylation by copying it from the parent to daughter strand (157). DNMT3A and DNMT3B, together with their catalytically inactive homologue, DNMT3L, establish DNA methylation *de novo* (75). DNMT3A and DNMT3B have different functions, supported by the distinct phenotypes of their respective murine knockouts, (86) their tissue specific expression patterns, and their unique roles in the development of cancer and other epigenetic disorders (49, 138, 158-163). However, the distinct biochemical properties of DNMT3A and DNMT3B that potentially contribute to their unique roles *in vivo* have not been fully examined.

A common feature of all mammalian DNMTs is the presence of a C-terminal MTase domain that contains 10 motifs conserved among prokaryotic and eukaryotic C5 DNA methyltransferases. The C-terminal MTase domains of DNMT3A and DNMT3B (DNMT3A-C and DNMT3B-C, respectively) share approximately 85% sequence similarity and are active methyltransferases in the absence of their respective N terminal regulatory domains (108). Another member of the DNMT3 family is DNMT3L, in which the key catalytic residues required for binding of AdoMet and DNA are absent; DNMT3L is therefore enzymatically inactive. DNMT3L interacts with both DNMT3A and DNMT3B and allosterically stimulates their catalytic activity, therefore acting as a positive regulator of *de novo* DNA methylation (100, 164). The crystal structure of the DNMT3A MTase domain (DNMT3A-C) alone and complexed with DNMT3L exhibits a hetero-tetrameric

structure of DNMT3A-C/3L or a homodimer of DNMT3A-C that can self-tetramerize through two interaction surfaces, the R-D dimer interface and the F-F tetramer interface (106, 109). The interactions at the dimer interface of DNMT3A-C tetramers support its multimerization on DNA forming nucleoprotein filaments. This allows the enzyme to bind and methylate multiple CpG sites on DNA in a cooperative manner, thus increasing its activity (110, 119, 165). Although controversial, some evidence also supports a processive mechanism for DNMT3A (108, 118, 166, 167). Similar to the case for a processive enzyme, cooperative binding of multiple DNMT3A molecules to DNA would allow it to methylate multiple sites on the same DNA molecule, complicating data analysis. However, the processivity of DNMTs is defined by their ability to diffuse along the DNA and methylate multiple sites before dissociation (82, 168). Therefore, it is important to use methods that differentiate processivity and cooperative activity of DNMTs. Previous studies have shown that DNMT3A-C methylates DNA in a distributive manner. Supporting its cooperative mechanism, its activity was shown to increase exponentially with a 2-fold increase in enzyme concentration, however only with long DNA substrates (118, 169). The critical role of the tetramer formation and cooperative mechanism of DNMT3A in its biological activity is apparent in acute myeloid leukemia (AML) cells that have a high frequency of mutation Arg882His present at the dimer interface of DNMT3A (82). *In vitro* studies show that the Arg882His mutant protein has only ~20% activity (142, 170) and is suggested to cause genome wide hypomethylation in AML cells (139, 143).

Multiple-sequence alignment reveals that most residues in the dimer interface are conserved between DNMT3A and DNMT3B. However, the role of the dimer interface and that of cooperativity of DNMT3B-C has not been evaluated. An earlier investigation using methylation-dependent restriction protection analysis of a 430-mer DNA showed that DNMT3B-C methylates its substrate in a processive manner (108). However, it is not clear from this study whether the activity of DNMT3B-C was influenced by cooperativity in a manner similar to that reported for DNMT3A-C. In the study presented here, we performed experiments that distinguish between cooperative and processive mechanisms. Our data show that in contrast to DNMT3A-C, DNMT3B-C methylates DNA in a non-cooperative manner and by a processive mechanism on both unmethylated and hemimethylated substrates at physiological pH and ionic strength and confirms the processivity of the full length DNMT3B enzyme. Experiments further characterizing the catalytic mechanism of DNMT3B-C show that whereas pre-incubation of the enzyme with DNA

reduces its activity, pre-incubation with AdoMet decreases its K_M by 10-fold, indicating that the cofactor AdoMet-bound form has an increased specificity for target sites. To further confirm the non-cooperative mechanism of DNMT3B-C, we tested the role of the Arg829 residue of DNMT3B-C that is analogous to dimer interface residue Arg882 of DNMT3A-C. Our data here show that whereas the reduced activity of the DNMT3A-C Arg882His variant could be partially attributed to the loss of its cooperativity, mutation of the conserved analogous arginine Arg829 had no effect on the activity or processivity of DNMT3B-C, thus supporting its non-cooperative mechanism. Similarly, disruption of interactions at the dimer interface of DNMT3A-C, with a decrease in the pH of buffer, causes a 2–3-fold loss of DNMT3AC activity (118, 169); however, it has no effect on the DNA binding or catalytic activity of DNMT3B-C. These data suggest that formation of tetramers may not be critical for the processive activity of DNMT3B-C. Collectively, these data reveal important biochemical differences between DNMT3A and DNMT3B that can potentially impact their activity and function during development and in cancer cells.

2.4 Materials and Methods

2.4.1 Protein Purification

Mouse DNMT3A-C, DNMT3B-C, DNMT3B-C Glu703Ala, DNMT3A-C Arg878His, DNMT3B-C Arg829His, and DNMT3L cloned in pET28a with a six-His tag were expressed and purified using affinity chromatography as described previously (167). Briefly, transformed BL21-DE3 cells were induced with 1 mM IPTG and grown for 2 h at 32 °C. Harvested cells were washed with STE buffer [10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, and 0.1 M NaCl] and suspended in buffer A [20 mM potassium phosphate (pH 7.5), 0.5 M NaCl, 10% (v/v) glycerol, 1 mM EDTA, 0.1 mM DTT, and 80 mM imidazole]. Cells were disrupted by sonication, followed by removal of cell debris by centrifugation. The clarified lysate was incubated with 0.75 mL of Ni-NTA agarose for 3 h at 4 °C. The protein-bound slurry was packed in a 2 mL Bio-Rad column and washed with 50 mL of buffer A. Protein (5–10 μ M) was eluted using 200 mM imidazole in buffer A at pH 7.5 and then stored in 20 mM HEPES (pH 7.5), 40 mM KCl, 0.2 mM DTT, 1 mM EDTA, and 20% glycerol at –80 °C. The purity and integrity of recombinant proteins were checked by SDS–PAGE

and Western blot analysis using a mouse monoclonal anti-His antibody (Invitrogen, MA1-21315). M.SssI methyltransferase was obtained from New England Biolabs.

2.4.2 Enzyme Assays

Most methylation assays for determining kinetics parameters of recombinant enzymes were performed using ^3H -labeled S-adenosylmethionine (AdoMet) as a cofactor and biotinylated oligonucleotides of varying sizes bound on avidin-coated high-binding ELISA plates (Corning) as described previously (153). The DNA methylation reaction was performed using either 250 nM 30-mer/32-mer DNA substrate or 150 nM 509-mer/719-mer DNA substrate in methylation buffer [20 mM HEPES (pH 7.5), 50 mM KCl, and 1 mM EDTA supplemented with 5 $\mu\text{g/mL}$ BSA]. The methylation reaction included 0.76 μM [methyl- ^3H] AdoMet (PerkinElmer Life Sciences). Storage buffer was added to compensate for the different enzyme volumes in all reaction mixtures. The substrates consisted of biotinylated oligonucleotides, including (1) a 30-bp oligonucleotide containing one CpG site, (2) a 32-bp oligonucleotide containing two CpG sites, (3) a 509-mer with 58 CpGs (amplified from the human SUHW1 gene promoter), or (4) a 719-mer DNA fragment with 46 CpGs (amplified from the mouse Aprt gene promoter). Primers used to amplify these substrates are listed in Table 2.1.

Table 2.1 Oligomers used for DNA methylation analysis. List of DNA oligomers used in this study. For the DNA binding experiments, the 30-mer oligomers used Cy5 5' end labels in place of the biotin modification (/5BiosG/).

Oligomer	Sequence	# CpG Sites
30-mer F	/5BiosG/GAAGCTGGGACTTCCGGGAGGAGAGTGCAA	1
30-mer R	TTGCACTCTCCTCCCGGAAGTCCCAGCTTC	1
32-mer F	/5BiosG/TGGGACTTCCGGGAGCTTCCGGGAGGAGAGTG	2
32-mer R	CACTCTCCTCCCGGAAGTCCCAGGAAGTCCCA	2
509-mer F	/5BiosG/AGATTAGGGAAGGGGGTGTG	58
509-mer R	AAGATCCTTTCAAGGCCTCAG	58
719-mer F	/5BiosG/CCCATGCGCCTGCGCCGGGTGCC	47
719-mer R	ATGCTCTAGACCTGCGATGTAGTTCGATC	47

For steady state kinetics analysis, the larger amounts of the substrate could not be accommodated in the standard plate assay, which has a maximal binding capacity per well for DNA substrate of 2 μ M. These methylation assays were, therefore, performed using a filter binding assay (171). Briefly, 10 μ L of reaction mix was spotted on a 1.5 cm DE81 filter that was then washed three to five times in 0.2 M ammonium bicarbonate, followed by 100% ethanol, and air-dried. Incorporated radioactivity was quantified by scintillation counting.

2.4.3 Cooperativity Assays

To examine cooperativity, the following changes were made to the plate assay described above. Increasing concentrations of enzyme were pre-incubated with DNA substrate for 10 min at room temperature prior to the addition of AdoMet to start the reaction. AdoMet was a mixture of unlabeled and 0.76 μ M 3 H-labeled AdoMet, which yielded a final concentration of 2 μ M. Methylation assays were performed using 30-mer, 509-mer, and 719-mer DNA substrates. On the basis of the structural studies of DNMT3AC, it was estimated that the 30-mer substrate is too short to allow potential multimerization of the enzyme (109, 165) and could therefore be used as a control. Additionally, cooperativity assays using 100 ng of the pUC19 plasmid as a substrate were performed using the filter binding assay described above. For these assays, unmethylated pUC19 was purified from the *dam*⁻/*dcm*⁻ *Escherichia coli* strain (C2925I, NEB).

2.4.4 Stimulation of DNMT3B-C by the Inactive Mutant Glu703Ala

For assays aiming to investigate the ability of an inactive mutant to stimulate wild-type (WT) DNMT3B-C activity, the activity of a 1:1 μM mixture of DNMT3B-C Glu703Ala mutant and WT DNMT3B-C was compared to the activity of WT DNMT3B-C (1 or 2 μM), using two different DNA substrates, 1 μM 30-mer substrate with 1 CpG or 150 nM 509-mer substrate with 56 CpGs. Reactions were initiated by addition of enzyme. The WT and/or Glu703Ala enzymes were mixed together and incubated for 5 min at room temperature before addition to the reaction mix. A mixture of unlabeled and 0.76 μM ^3H -labeled AdoMet was used at a final concentration of 2 μM . Methylation rates were determined by using linear regression to analyze data. The fold change in methylation rate compared to that for 1 μM WT enzyme was plotted for each enzyme treatment.

2.4.5 Processivity Assays

Methylation kinetics analyses were performed using a range of enzyme concentrations and short oligonucleotide 30- and 32-mer substrates with one and two CpG sites, respectively. Low enzyme concentrations relative to DNA substrate concentrations were included to ensure that the reaction occurred under multiple-turnover conditions. Each DNA substrate was used at a concentration of 250 nM, and a 1:1 ratio of labeled and unlabeled AdoMet (final concentration of 1.5 μM) was used. To compare the identical CpG molarity, a parallel reaction using 500 nM 30-mer was also performed. Reactions were started by enzyme addition. M.SssI and DNMT3A-C were used as positive and negative controls, respectively.

2.4.6 DNA Binding Assays

DNA binding of DNMT3B-C was performed using Cy5-labeled 30-mer DNA containing one CpG site (Table 2.1) in nitrocellulose filter binding assays. Binding reactions were performed in 20 mM HEPES (pH 7.5), 100 mM KCl, 1 mM EDTA buffer, in the presence of 0.2 mM S-adenosylhomocysteine (AdoHcy) and 30 nM DNA, and increasing concentrations of DNMT3B-C. Reaction mixtures were incubated at room temperature for 15 minutes prior to being spotted on

the nitrocellulose membrane in a dot blot apparatus, followed by three washes with binding buffer. The enzyme bound fraction of Cy5-DNA was quantified by fluorescence measurement (Typhoon).

2.4.7 Data Analysis

Data were analyzed using Prism software (GraphPad). For time-dependent kinetics measurements, values were fitted using linear regression of a nonlinear fit, which was weighted by $1/Y^2$. Each time point for methylation kinetics was an average and standard deviation of three to six experimental replicates. For secondary plots, a least-squares fitting method was used to plot the data, and the linear regression was not weighted. To determine the equilibrium binding constant, data were fitted to a one-site binding model with a Hill coefficient. Standard errors of the mean were calculated for three to six independent experiments, as described in the figure legends.

2.5 Results

N-Terminally His-tagged DNMT3B-C, DNMT3B-C Glu703Ala, and DNMT3L were overexpressed and purified on Ni-NTA agarose to 90–95% purity as estimated by the Coomassie-stained SDS gel (Figure 2.1A). The K_M and the turnover rate constant (K_{cat}) value for purified DNMT3B-C were determined to be $(3.4 \pm 1) \times 10^{-7} \text{ M}$ and $(3.3 \pm 0.3) \times 10^{-4} \text{ s}^{-1}$, respectively (Figure 2.1B), and catalytic efficiency $K_{cat}/K_M \sim 10^3 \text{ M}^{-1} \text{ s}^{-1}$. To confirm the catalytic robustness of the enzyme, we assessed the allosteric activation of DNMT3B-C by DNMT3L (170). Consistent with earlier observations, we observed ~6-fold activation of DNMT3B activity in the presence of DNMT3L (170). This DNMT3B-C protein was next used to determine the catalytic mechanism.

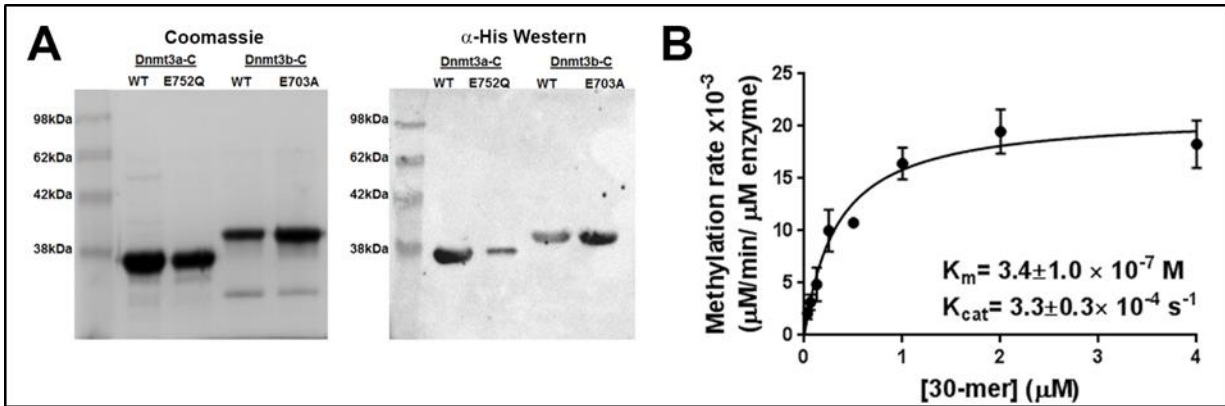


Figure 2.1 Expression of DNMT3A-C, DNMT3B-C and their respective mutants. (A) WT: wild type; Coomassie stained SDS-PAGE gel and Western Blot showing purified His-tagged WT DNMT3A-C, inactive mutant DNMT3A-C E752Q, WT DNMT3B-C, and the inactive mutant DNMT3B-C Glu703Ala. Mouse monoclonal Anti-His antibody was used to detect proteins on the western blot. (B) Steady-state kinetics analysis of DNMT3B-C activity. Methylation reactions were performed for 10 min and initial velocities were calculated as methyl groups transferred per min per μ M enzyme for varying concentrations of 30-mer substrate ranging from 0.031-4 μ M in the presence of 0.75 μ M AdoMet. The reactions were started by addition of enzyme to the substrate cocktail. The data were fit to the Michaelis-Menten equation to yield the kinetics parameters. The data are average \pm standard error of the mean ($n \geq 3$ independent experiments).

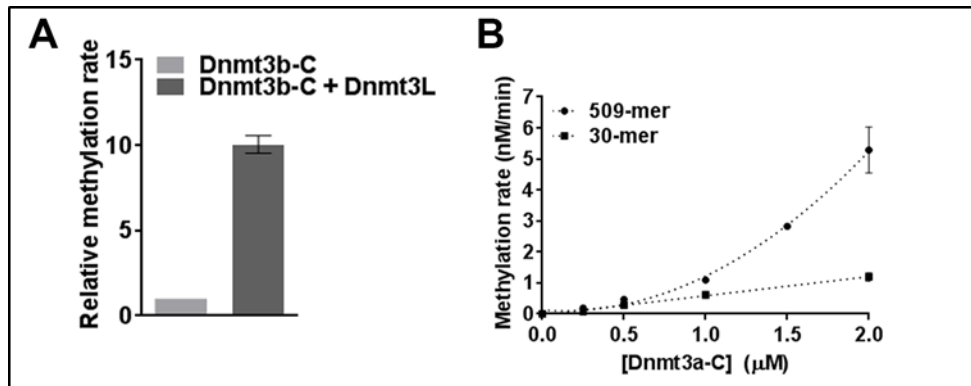


Figure 2.2 Allosteric activation of the DNMT3 enzymes. (A) DNMT3B-C is allosterically activated by DNMT3L. Analysis of the fold change in activation of DNMT3B-C by DNMT3L toward 30-mer DNA (500 nM). Time points were taken at 2, 4, 8, and 16 min in the presence of 1 μ M of either DNMT3B-C alone or with DNMT3L. Average \pm SEM of three independent experiments. B, Cooperative mechanism of DNMT3A-C. DNA methylation by DNMT3A-C at various enzyme concentrations. For 30-mer with 1 CpG and 509-mer with 58 CpGs. Data were fit to linear regression (30-mer) or second order polynomial (509-mer).

2.5.1 Investigating the Cooperativity of DNMT3B-C

DNMT3A-C monomers interact with each other through two interfaces, the DNMT3A-DNMT3L interface (F-F interface) and the DNMT3A-DNMT3A dimer interface (R-D interface), to form tetramers (109). DNMT3A-C tetramers multimerize on DNA potentially through interactions along the dimer interface, which supports cooperativity of DNMT3A (118). Many amino acids known to be critical for the dimer and tetramer interface mediated interactions of DNMT3A are conserved in DNMT3B (106, 109). However, the cooperativity of DNMT3B has not been tested. The activity of a cooperative enzyme like DNMT3A increases in a nonlinear manner at higher enzyme concentrations because of the allosteric effect of enzyme subunit interactions (118), in contrast to a non-cooperative enzyme for which the activity is expected to increase in a linear manner as a function of enzyme concentration. Previous studies showing cooperativity of DNMT3A-C used a 509-bp DNA substrate to allow the binding of multiple protein molecules at higher enzyme concentrations (118). To compare the cooperativity of DNMT3A-C and DNMT3B-C, DNA methylation analysis was performed using the same 509-mer DNA substrate. Increasing concentrations of the DNMT3A-C and DNMT3B-C enzymes ranging from 0 to 2 μ M were pre-incubated with 509-mer DNA substrate for 10 min to allow the formation of nucleoprotein complexes. The methylation reaction was initiated by the addition of radiolabeled AdoMet and was monitored by incorporation of radioactivity into DNA for an additional 10

minutes. The methylation activity was plotted against DNMT3A-C and DNMT3B-C enzyme concentrations (Figure 2.3A). The data show that for DNMT3A-C with an increase in concentration from 1 to 2 μ M the activity increased \sim 5-fold, whereas for DNMT3B-C, there was only \sim 2-fold increase in activity for every 2-fold increase in enzyme concentration. As shown previously (118), this exponential increase in methylation activity of DNMT3A-C was specific for only the long DNA substrate (509-mer) and was absent for a short 30-mer DNA substrate (Figure 2.2B). On the basis of the structural studies of DNMT3A-C, the 30-mer DNA substrate is not expected to accommodate more than one or two tetramers and thus cannot support multimerization and cooperativity (109, 118, 165). To test if DNMT3B-C cooperativity could be supported on longer DNA substrates, we repeated the methylation assays using the pUC19 plasmid as a substrate. As shown in Figure 2.3B, whereas the activity of DNMT3A-C was cooperative, the activity of DNMT3B-C increased linearly with the increase in enzyme concentration similar to the data shown in Figure 2.3A. We also noticed that compared to the 509-mer, the plasmid substrate stimulated cooperativity of DNMT3A-C at lower enzyme concentrations, suggesting cooperativity of DNMT3A-C may be influenced by substrate length. To confirm the results for DNMT3B-C, we performed a time course of DNA methylation to determine the initial rate of DNA methylation using three biotinylated substrates, a 30-mer with one CpG site at 250 nM, a 509-bp DNA substrate with 58 CpGs, and a 719-bp substrate with 46 CpG sites at concentrations of 150 nM each. Increasing concentrations of the DNMT3B-C enzyme ranging from 0.5 to 2 μ M were pre-incubated with the DNA, and the reaction was initiated via the addition of radiolabeled AdoMet and was monitored by the incorporation of radioactivity into the DNA (Figure 2.3C-E). Irrespective of the substrate length, the methylation rate increased linearly with the increase in DNMT3B concentration (Figure 2.3F), with a \sim 2fold increase in the rate of methylation for every 2-fold increase in enzyme concentration. We observed rates of methylation slightly lower than expected at the lowest enzyme concentration for all three substrates potentially because of the slow turnover under these conditions. At higher enzyme concentrations in which cooperative methylation was observed for DNMT3A-C (118) (Figure 2.3A), the absence of an exponential or nonlinear increase in the methylation rate for DNMT3B-C suggests that DNMT3B methylates DNA by a non-cooperative mechanism.

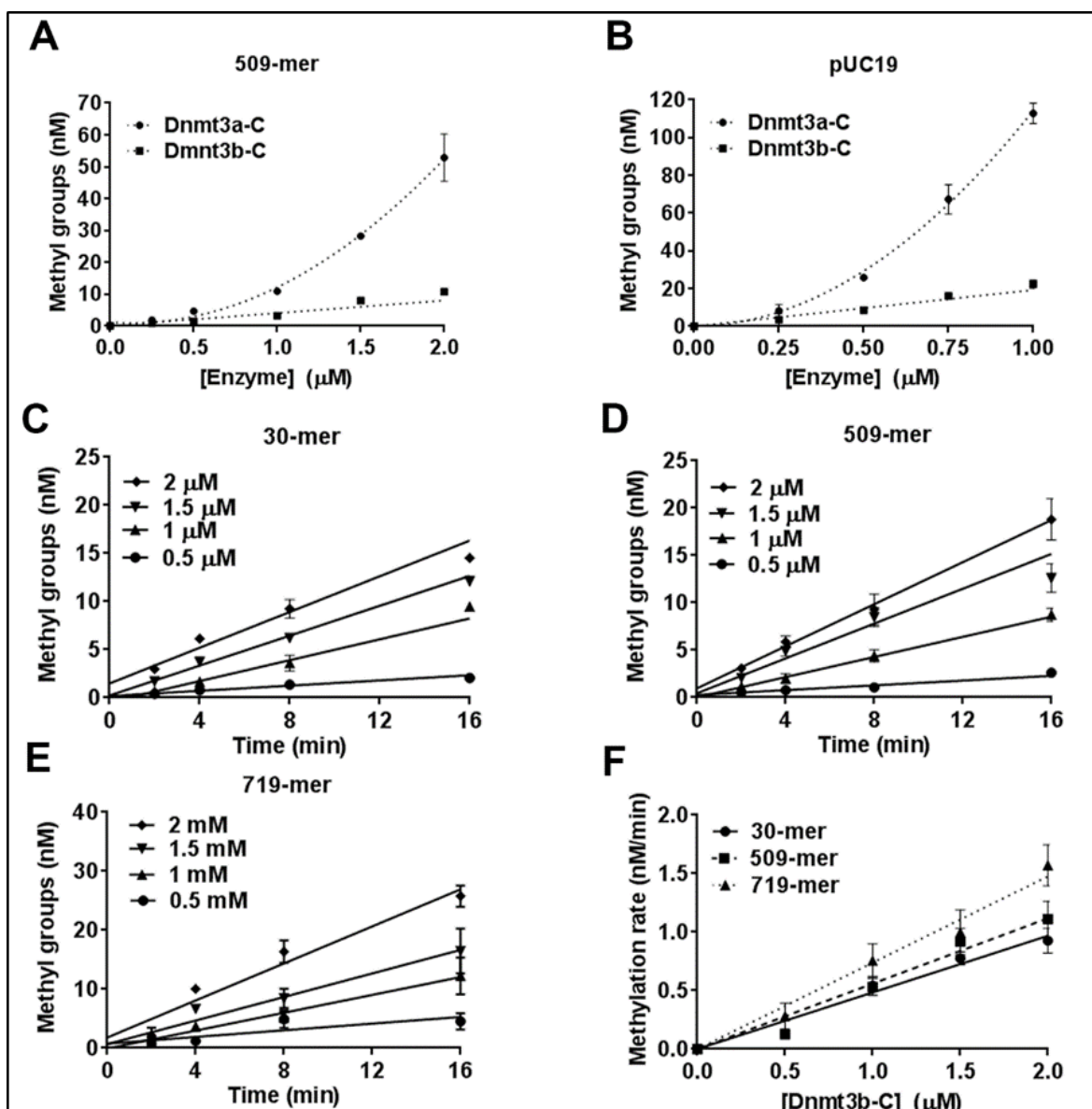


Figure 2.3 Catalytic mechanism of DNMT3B-C. (A and B) Methylation activity of DNMT3A-C and DNMT3B-C enzymes was measured for 10 min with the 509-bp or pUC19 substrate in the presence of 0.25 to 2 μM or 0.25 to 1 μM enzyme respectively. The enzymes were pre-incubated with DNA for 10 min at room temperature and the reaction was initiated by addition of AdoMet.

Total methylation activity was plotted against enzyme concentration using an average and standard deviation ($n \geq 3$ independent experiments). (C, D and E) Time course of DNA methylation with 30-mer, 509-mer, and 719-mer DNA substrates in the presence of 0.5, 1, 1.5, or 2 μM DNMT3B-C enzyme which was pre-incubated with DNA for 10 min at room temperature and the reaction was initiated by addition of AdoMet. The data were fitted by linear regression, weighted by $1/Y^2$. (F) Methylation rates for all three DNA substrates were plotted against enzyme concentrations using linear regression without weighting. Average \pm standard error of the mean are shown ($n \geq 3$ independent experiments).

2.5.2 DNMT3B-C is Not Stimulated by the Catalytically Inactive Mutant

Addition of a catalytically inactive mutant to WT DNMT3A-C was shown to stimulate its catalytic activity because it contributes to the cooperativity of DNMT3A on long DNA substrates, resulting in an allosteric effect (118). This variant carries a mutation in conserved motif IV (ENV) that is required for catalysis, but this mutation does not affect the DNA binding activity of the methyltransferase enzymes (118). The corresponding mutant DNMT3B-C Glu703Ala has very low residual activity (Figure 2.4). As an additional test of cooperative stimulation, we tested the influence of the DNMT3B-C Glu703Ala mutant on the activity of WT DNMT3B-C. Methylation kinetics reactions were performed with two biotinylated substrates at near saturating concentrations: a 30-bp substrate with one CpG site at 1 μ M and a 509-bp substrate at 150 nM using either 1 μ M WT enzyme or 1 μ M WT mixed with 1 μ M inactive mutant (Figure 2.5A,B). The fold change in the rates of DNA methylation compared to that of 1 μ M WT enzyme alone was then plotted (Figure 2.5C). A 2-fold increase in activity was observed when the enzyme concentration was increased from 1 to 2 μ M for both substrates. The 1:1 WT/inactive mutant mixture yielded a small increase in activity, however, with both short (30-mer) and long (509-mer) DNA substrates. Because the short DNA substrate cannot bind multiple enzyme units (109, 118, 165), this increase in activity is not likely due to cooperativity but may be due to interactions at the tetramer interface (F-F interface) as shown previously for DNMT3A-C (100, 118). These data provide further support for the non-cooperative mechanism of DNMT3B activity.

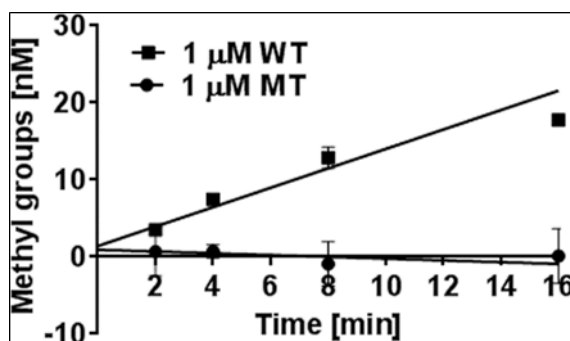


Figure 2.4 DNMT3B-C Glu703Ala is catalytically inactive. Catalytic activity of WT DNMT3B-C and DNMT3B-C Glu703Ala. Methylation assays were performed as described in Materials and Methods using 250 nM of the 30-mer substrate and data were fit to linear regression. Average \pm SEM of four independent experiments.

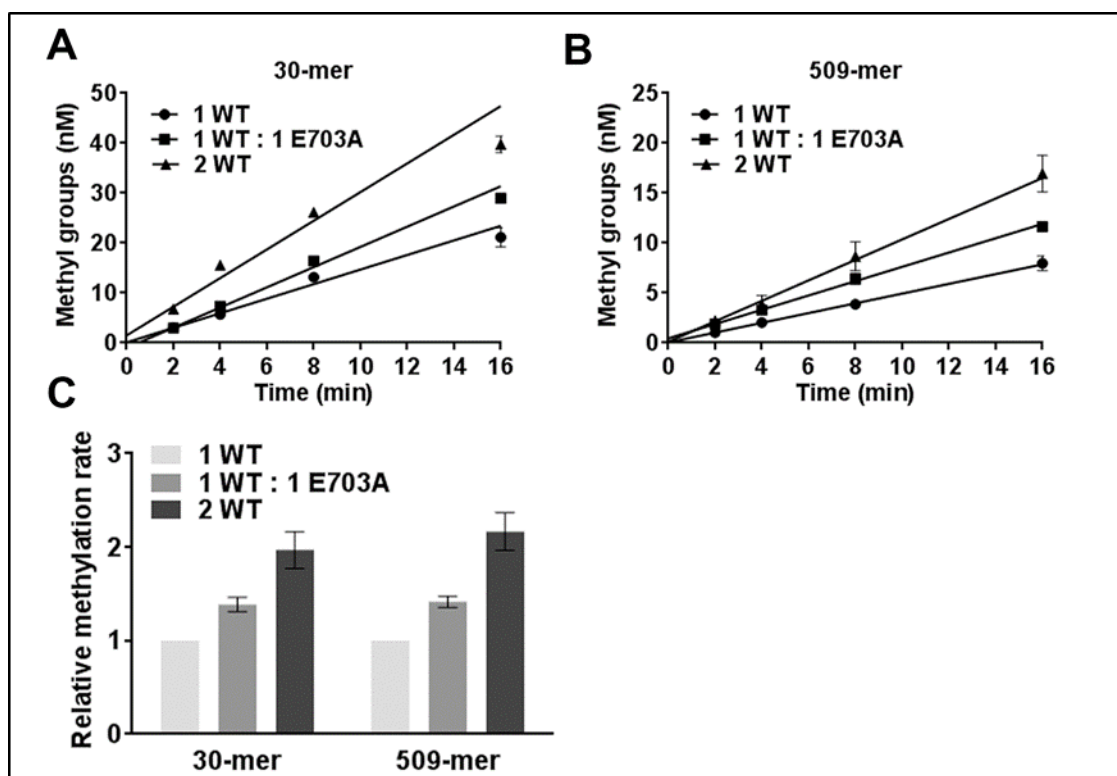


Figure 2.5. DNMT3B-C is not cooperatively stimulated by addition of inactive DNMT3B-C Glu703Ala. (A and B) Methylation kinetics of 30-mer and 509-mer substrates, respectively by WT (DNMT3B-C); and mutant (catalytically inactive DNMT3B-C Glu703Ala). Methylation was carried out using 1 μ M WT, a ratio of 1 μ M WT : 1 μ M Glu703Ala, and 2 μ M WT with either 1 μ M 30-mer or 150 nM 509-mer substrates. The reaction was initiated by addition of enzyme to the substrate mix and data were fitted by linear regression, weighted by $1/Y^2$. (C) Methylation rates were measured from the slopes in A and B. Rate of methylation for each substrate was normalized to the rate for 1 μ M enzyme and plotted in the bar graph with normalized error. Average \pm standard error of the mean is shown ($n \geq 4$ independent experiments).

2.5.3 DNMT3B-C and Full Length DNMT3B Methylate DNA in a Processive Manner

Unlike a previous method that used methylation-dependent restriction protection to test the processivity of DNMT3B (108), here we assayed the processivity using two DNA substrates of the same sequence, 30 and 32-bp in length, except that one substrate contained one CpG and the second contained two CpGs (167). In a processive reaction, the second site on the two-site DNA substrate is expected to be methylated faster, so the methylation rate for the two-site substrate would be expected to be ~ 2 -fold higher than that for the one-site substrate. This effect should be prominent at a low enzyme:DNA ratio, in which each DNA molecule is occupied by only one

enzyme molecule, in other words, under multiple turnover conditions. Therefore, methylation reactions were performed using a range of DNMT3B-C enzyme concentrations at near saturating substrate (DNA and AdoMet) concentrations. Unlike the cooperativity assays, the enzyme was not pre-incubated with DNA, and the methylation reactions were initiated by addition of enzyme to a buffer/substrate cocktail. This allowed the enzyme to bind to its preferred substrate first and catalysis to proceed at maximal efficiency. Short oligonucleotides were used as substrates to eliminate any potential cooperativity from occurring, because they cannot bind more than a few molecules of the enzyme. Assays of DNMT3A-C and DNMT3B-C (0.0625–2 μM) were performed with 0.25 μM one-site and two-site substrates (0.5 and 1 μM CpG sites). As previously reported, DNMT3A-C methylated one and two-site substrates to the same degree, indicating a distributive mechanism of DNA methylation (108, 118). In contrast, DNMT3B-C methylated the two-site substrate with a 2-fold higher efficiency (Figure 2.6A). Because this effect was evident at the lowest concentration of enzyme assayed, the data indicate that the DNMT3B enzyme methylates DNA in a processive fashion. We also demonstrated the processive activity of DNMT3B by using the full length DNMT3B enzyme (DNMT3B-FL) in the assays described above (Figure 2.6A and Figure 2.7A). Our data show that similar to the truncated DNMT3B-C, at lower enzyme concentrations DNMT3B-FL methylates DNA in a processive manner. Because of its large size and tendency to precipitate, DNMT3B-FL could be purified at a maximal concentration of 0.5–1 μM ; consequently, the assays at higher concentrations could not be performed. A similar effect was observed for the known processive methyltransferase M.SssI (Figure 2.7B). To determine if these results were influenced by the presence of a 2-fold higher CpG molarity of the two-site substrate, we performed a parallel reaction at an equal CpG molarity by comparing methylation of 0.5 μM one-site substrate and 0.25 μM two-site substrate. The ratio of DNA methylation at either 0.25 or 0.5 μM DNMT3B-C remained at 2, confirming its processive mechanism of DNA methylation (Figure 2.7C). We next tested the effect of ionic strength on the processivity of the DNMT3B-C enzyme. Figure 2.6B shows that increasing the ionic strength from 50 to 150 mM KCl had no effect on the degree of methylation of one- or two-site substrate, indicating that DNMT3B can methylate DNA processively at a physiological ionic strength ranging from 100 to 150 mM. Salt concentrations of >250 mM disrupted enzyme activity. DNMT3B has been shown to collaborate with DNMT1

for the maintenance of DNA methylation (172). We therefore tested the processivity of DNMT3B-C on one- or two-site hemimethylated substrates. Similar to the unmethylated substrate, DNMT3B methylated the two-site hemimethylated substrate with an efficiency 2-fold higher than that of the one-site substrate (Figure 2.6C).

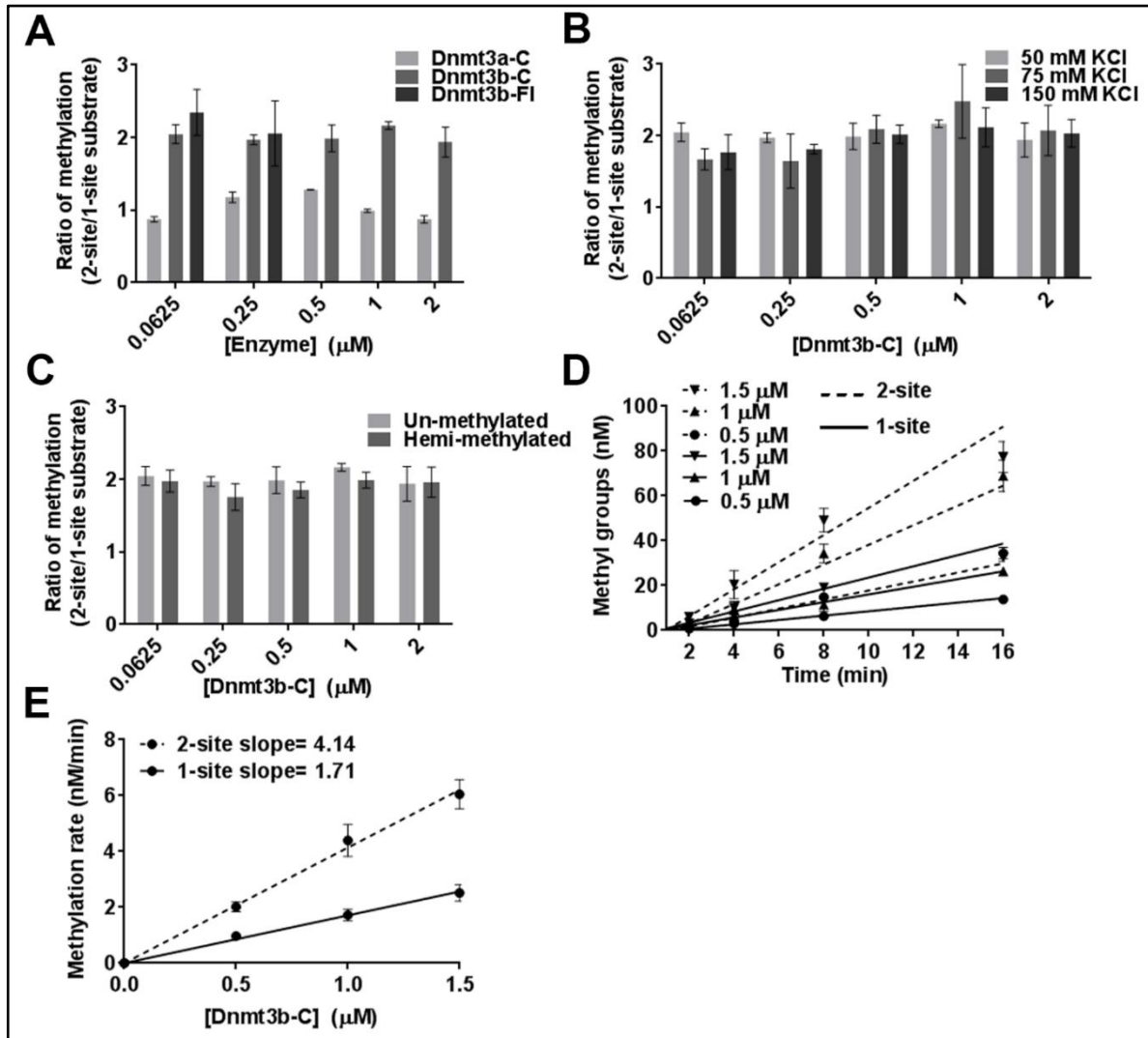


Figure 2.6 DNMT3B-C exhibits processivity on DNA substrate with 2 CpG sites. (A) Steady state DNA methylation levels were measured for 1 h at various concentrations of DNMT3B-C, DNMT3B- FL, or DNMT3A-C enzymes using 1.5 μ M 3 H-labeled AdoMet (1:1 mixture of labeled and unlabeled) and 250 nM of 1-site or 2-site substrate. For each enzyme concentration, the methylation of 2-site substrate was normalized to that of 1-site substrate to show the relative change in methylation level. (B and C) Relative DNA methylation as described in (A) at physiologically relevant salt concentrations and with 250 nM of hemimethylated 1-site and 2-site substrates. (D) Initial velocity of methylation was measured at early time points (2, 4, 8, and 16 min) for 1-site (solid line) and 2-site (dashed line) substrates. Concentration of DNMT3B-C ranged from 0.5 to 1.5 μ M and the data were fit to linear regression, weighted by $1/Y^2$. (E) DNA methylation rates measured in (D) were plotted against enzyme concentration and the slopes were determined from the linear regression without weighting. Data are an average \pm standard error of the mean ($n \geq 3$ independent experiments).

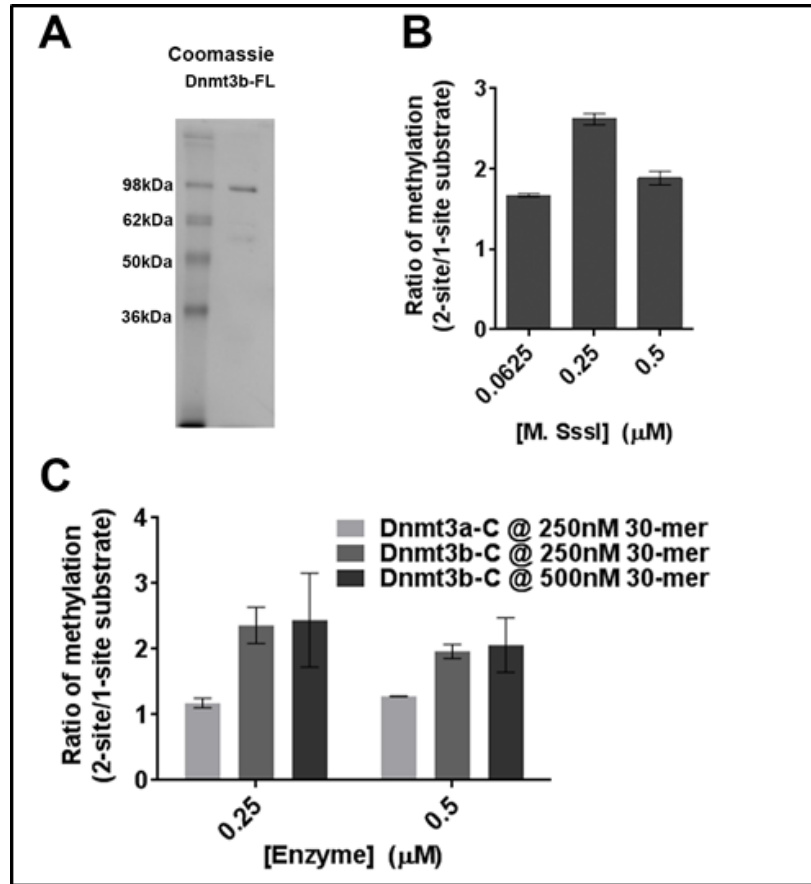


Figure 2.7 Variable processivity conditions of M.SssI and DNMT3B-C. A, WT: wild type; Coomassie stained SDS-PAGE gel showing purified His-tagged WT DNMT3B-FL. B, As a positive control for processivity, steady state DNA methylation levels were measured for 1 h at various concentrations of M.SssI using 1.5 μ M 3 H-labeled AdoMet (1:1 mixture of labeled and unlabeled) and either 1-site or 2-site substrates. For each enzyme concentration, the methylation of 2-site substrate was normalized to that of 1-site substrate to show the relative change in methylation level. C, DNMT3B-C is processive when 1 and 2-site substrates contain the same [CpG]. Steady state DNA methylation levels were measured after 500 nM of 1-site and 250 nM of 2-site substrate were incubated with DNMT3B-C and 1.5 μ M (0.75 μ M of 3 H-labeled and 0.75 μ M of unlabeled) AdoMet. The methylation level for 2-site substrate was normalized to that of 1-site substrate to show the relative change. Average \pm SEM of three experiments.

2.5.4 Processivity of the Initial Phase of the Reaction

To confirm our conclusion from the steady state experiments described above, we next examined the processivity of DNMT3B-C during the initial phase of the reaction. The initial rate of catalysis of a processive enzyme is also expected to be 2-fold higher with a two-site substrate than with a one-site substrate. DNA methylation using different enzyme concentrations was measured as a function of time, and initial rates were calculated from the slope after fitting the data

by linear regression (Figure 2.6D). Methylation rates increased linearly with enzyme concentration, indicating multiple-turnover conditions. As shown in the secondary plot, DNA methylation rates for the two-site substrate were 2-fold higher than those for the one-site substrate, strongly suggesting that DNMT3B-C operates in a processive fashion by methylating multiple CpGs on the same molecule of DNA before dissociation (Figure 2.6E).

Taken together, our data here confirm the processive mechanism of DNMT3B and validate an important difference between the catalytic mechanism of DNMT3A and DNMT3B.

2.5.5 Effect of Pre-incubation with DNA on Catalytic Activity

Processivity is expected to increase the catalytic activity of a methyltransferase toward long DNA substrates with multiple target sites. However, for experiments in which DNMT3B-C was pre-incubated with DNA, we did not observe a substantial increase in the methylation rate for long DNA substrates compared to that with the 30-mer (Figure 2.3F). Although pre-incubation with DNA may simply lead to a decrease in the active fraction of the enzyme, it is also possible that in the absence of AdoMet, nonspecific DNA binding by DNMT3B-C keeps it in a nonproductive complex. This can potentially slow the methylation reaction and influence the processivity of the enzyme. To investigate the effect of pre-incubation with DNA on the processivity of DNMT3B, the rate of DNA methylation for the 509-bp DNA substrate was compared under two conditions. The first condition, in which various concentrations of DNMT3B-C were pre-incubated with DNA substrate and the methylation reaction was initiated with AdoMet, is represented in Figure 2.3D. For the second condition, no pre-incubation was performed and methylation was initiated by addition of enzyme to the substrate cocktail. Kinetics of DNA methylation at each enzyme concentration was performed, and the initial methylation rates were determined by linear regression (Figure 2.8A). At all the tested enzyme concentrations, methylation rates were significantly slower when DNA was pre-incubated with enzyme than when it was not (Figure 2.8B), confirming a negative effect of pre-incubation with DNA on the catalytic activity of DNMT3B-C. This is not intuitive for a processive enzyme, because during multiple reactions the exchange of AdoHcy with AdoMet should not require the enzyme to be dislodged from the DNA. We speculate that in the absence of a cofactor (AdoMet or AdoHcy), the enzyme pre-incubated with DNA binds strongly to nonspecific sites, which slows catalysis suggesting that the enzyme prefers to bind AdoMet for the first turnover.

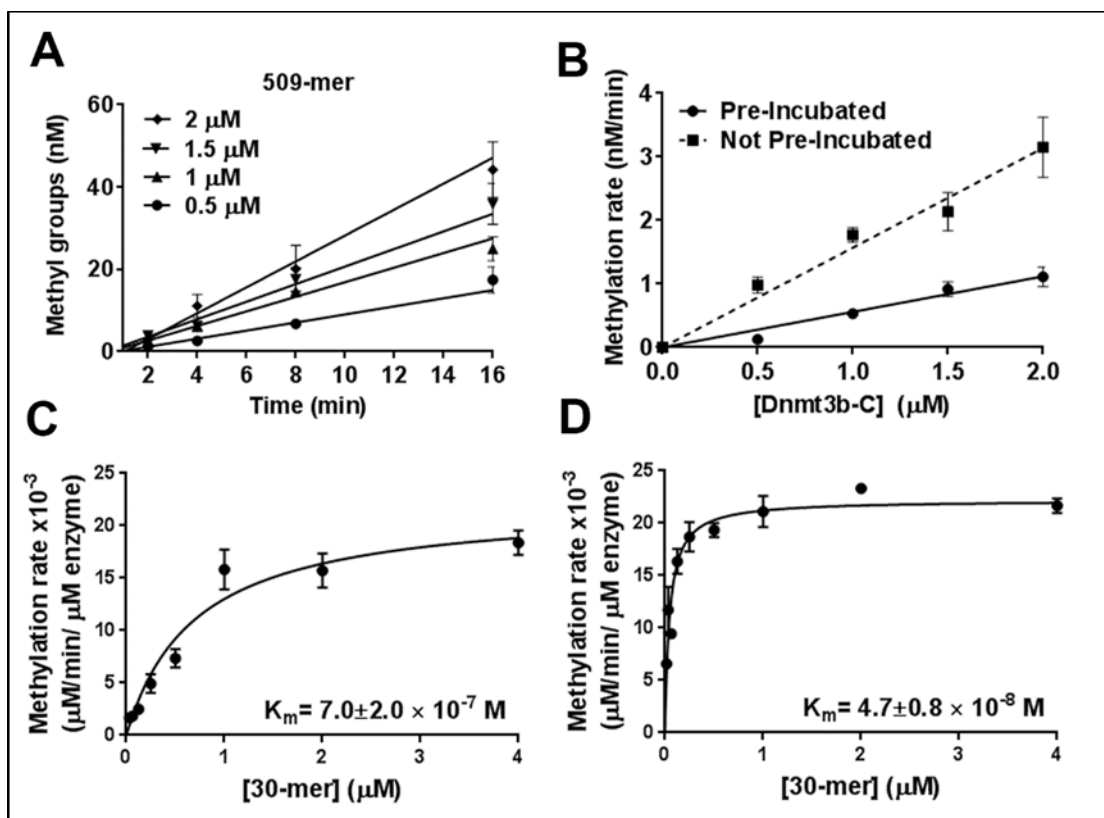


Figure 2.8 Pre-incubation with DNA decreases the methylation activity of DNMT3B-C. (A) Methylation activity of DNMT3B-C for the 509-mer was measured without pre-incubation with DNA for a range of enzyme concentrations. The reaction was initiated with addition of enzyme to substrate cocktail. The data, which is the average \pm standard error of the mean ($n \geq 3$ independent experiments), were fit to linear regression, weighted by $1/Y^2$. (B) Rates of DNA methylation calculated from (A) were plotted against enzyme concentrations and compared to the rates obtained from assays with pre-incubated enzyme from Figure 2.3D. (C and D) Steady-state kinetics analysis of methylation by DNMT3B-C. Methylation reactions were performed for 10 min and initial velocities were calculated as methyl groups transferred per min per μM enzyme for varying concentrations of 30-mer substrate ranging from 0.031–4 μM and 0.75 μM AdoMet. The enzyme was pre-incubated with DNA for 15 min and the reaction was started with AdoMet (C), or the enzyme was pre-incubated with AdoMet for 15 min and initiated with DNA (D). The data, which are an average \pm standard error of the mean ($n \geq 3$ independent experiments), were fit to the Michaelis-Menten equation and the significance was estimated using Student's paired t-test yielding a p-value= 0.014.

2.5.6 Order of Addition for Steady State Kinetics Analysis

To determine the effect of pre-incubation with the substrate and cofactor, we analyzed the steady state methylation activity of DNMT3B-C. The avidin plate assay used for methylation assays has a maximal DNA binding capacity of 2 μM (153); therefore, we used a filter binding assay for this experiment (171). Using a 30-mer substrate with one CpG site and 0.5 μM enzyme,

steady state kinetics experiments were performed, and data were fitted to the Michaelis–Menten equation. Three different experimental conditions were used, one in which the enzyme was pre-incubated with DNA or AdoMet for 15 min before the reaction was started by addition of AdoMet or DNA, respectively, or the reaction was initiated by adding enzyme to a substrate/cofactor mix. All data sets fit well to the hyperbolic curve (Figures 2.1B and 2.8C, D). However, the data set for pre-incubation with DNA (Figure 2.8D) also fitted to a sigmoidal curve, indicating a slower reaction at lower substrate concentrations that recovers at higher enzyme concentrations. The comparison using Akaike’s Information Criteria (AICc) or the extra sum of squares F-test suggests that the hyperbolic Michaelis–Menten model is a better fit with a 71.5% higher probability of being correct. The estimated V_{\max} under all the conditions showed no significant change. Under these three conditions, the K_M value was estimated to be $(3.4 \pm 1) \times 10^{-7}$ M for the reaction with non-pre-incubated enzyme (Figure 2.1B) and $(7.0 \pm 2) \times 10^{-7}$ and $(4.7 \pm 0.8) \times 10^{-8}$ M when the enzyme was pre-incubated with DNA and AdoMet, respectively (Figure 2.8C, D). A 10-fold decrease in the K_M of the enzyme upon pre-incubation with AdoMet demonstrates that the cofactor-bound DNMT3B-C enzyme exists in a conformation that favors catalysis, potentially by enhancing the specific interaction of DNMT3B-C with DNA. On the basis of these results, we propose that the binding of DNMT3B-C to DNA in the absence of AdoMet may lead to the formation of a nonspecific DNA–enzyme binary complex in a conformation that limits or restricts binding of the cofactor AdoMet, thus slowing the reaction. This may have a stronger influence on initial or pre-steady state rates, thus influencing the processivity of the enzyme at earlier time points as seen in Figure 2.8B. We propose that *in vivo* most enzyme exists in the AdoMet-bound state and that during processive DNA methylation, cofactor exchange takes place without dissociation of the enzyme from the DNA.

2.5.7 Activity and Processivity of DNMT3B-C are Not Affected by pH

Under various physiological conditions, changes in the intracellular pH can influence protein–protein and protein–DNA interactions (173). Previous studies have shown that the interaction of DNMT3A through its dimer interface can be disrupted by a change in pH from 7.5 to 6.5, which results in a decrease in its activity and cooperativity (169). Our data showing that DNMT3B methylates DNA in a non-cooperative manner suggest that the activity of DNMT3B

may not be strongly dependent on interactions involving the dimer interface. We tested this by assessing the effect of pH on the processivity of DNMT3B-C. Methylation assays were performed at pH 7.5 and 6.5 for one- and two-CpG site substrates for 60 min with varying concentrations of enzyme. As shown in Figure 2.9A, the steady state level of DNA methylation using a one-site substrate was slightly higher at pH 6.5 than at pH 7.5. As a consequence, the ratios of methylation between one- and two-site substrates were reduced to <2 (Figure 2.9B). We checked whether this difference in methylation activity was due to the impact of pH on the initial rate of DNA methylation. Methylation kinetics at various enzyme concentrations was performed with one- and two-site substrates at pH 6.5 and 7.5, and initial rates were measured (Figure 2.9C, D). The rates of DNA methylation for the one-site substrate at various enzyme concentrations were slightly higher at pH 6.5 compared to those at pH 7.5 (Figure 2.9E). We speculate that this higher methylation rate under pre-steady state conditions could be responsible for the difference observed at the steady state levels. However, the rates of methylation of the two-site substrate remained ~ 2 -fold higher than those for the one-site substrate at all enzyme concentrations tested at pH 7.5 and 6.5, indicating that the activity and processivity of DNMT3B-C are not affected by lower pH. This behavior is in contrast to that of DNMT3A-C, for which lowering the pH to 6.5 decreased its methylation activity by disruption of the dimer interface (169).

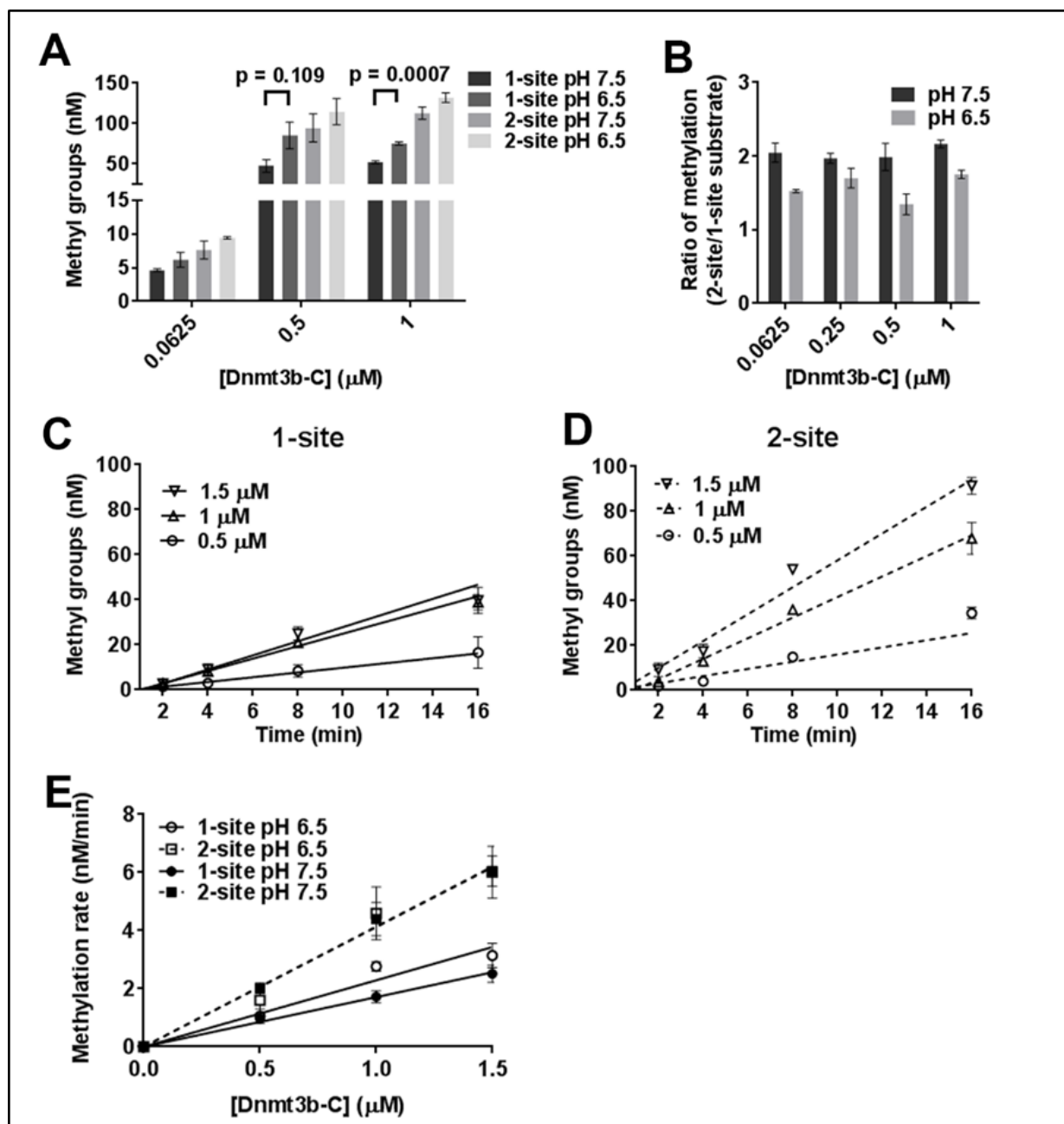


Figure 2.9. pH sensitivity of the DNMT3B-C processive catalysis. (A) Methylation at 1-site and 2-site substrates by DNMT3B-C was measured for 1 hour at pH 6.5 or 7.5. (B) Ratio of the methylation levels was calculated as the fold change at 2-site versus 1-site substrate. (C and D) Time course of DNA methylation by DNMT3B-C at pH 6.5 for 1-site and 2-site substrates. For each substrate, time course of DNA methylation reaction was done in the presence of 0.5- 2 μM DNMT3B-C. The reaction was started by the addition of enzyme and the data were fitted by linear regression which was weighted by $1/Y^2$. E The calculated methylation rates at pH 6.5 were plotted against enzyme concentration and compared to those at pH 7.5. Data are an average \pm standard error of the mean ($n \geq 4$ independent experiments).

2.5.8 Influence of pH on the Dissociation Constant

To test the effect of lower pH on DNA binding by DNMT3B-C, we performed DNA binding assays using Cy5-labeled 30-mer DNA with one CpG site. An increasing amount of DNMT3B-C ranging from 0.5 to 2 μ M was incubated with 30 nM DNA substrate, and the binding assay was performed using a dot blot assay. The binding was measured by trapping the protein-DNA complexes on a nitrocellulose membrane. The intensity of the spots was measured, and data were fitted to a one-site specific binding saturation curve. Binding curves were hyperbolic, and the data could also be fitted well with the Hill coefficient constrained to 1 (Figure 2.10A, B). The dissociation constants at pH 7.5 and 6.5 showed no significant difference, indicating that the binding of DNMT3B-C to DNA is not sensitive to lower pH and there is potentially a limited contribution of the dimer interface to the activity of DNMT3B-C.

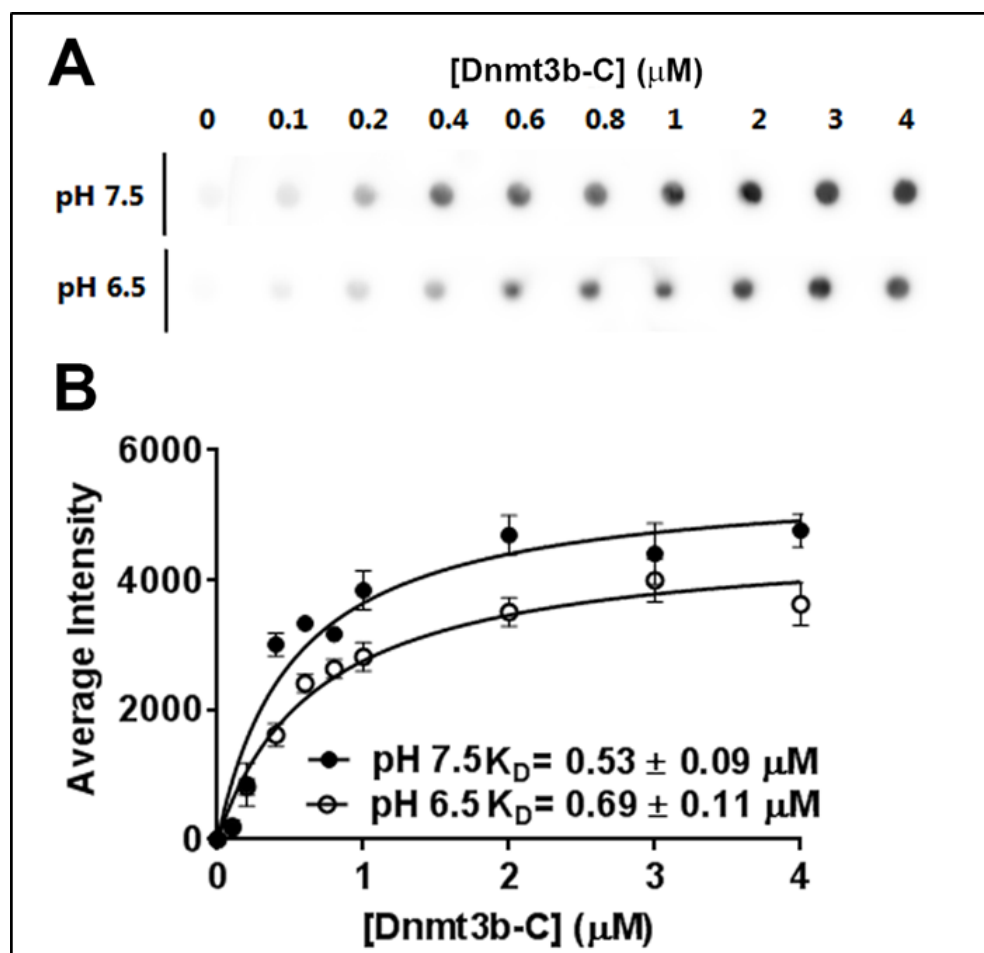


Figure 2.10 Effect of pH on DNA binding by DNMT3B-C. For DNA binding analysis DNMT3B-C enzyme was incubated with 30 nM of Cy5-labelled 30-mer DNA in the presence of 0.2 mM AdoHcy (S-Adenosylhomocysteine) in binding buffer at pH 6.5 or 7.5. (A) Representative blot for DNA binding. (B) Signal intensity on the blot was quantified using ImageQuant software and data which were an average \pm standard error of the mean ($n \geq 3$ independent experiments) were fitted to a one site binding saturation model and binding constants were determined. The binding constants are within an error of $<20\%$.

2.5.9 The Processivity of DNMT3B-C is Not Affected by Mutation of the Arg829His Residue

The Arg882His mutation, which is close to the dimer interface of DNMT3A, is highly prevalent in human AML cells. This mutation, as well as the equivalent Arg878His mutation in mouse, disrupts tetramer formation and leads to a significant decrease in enzymatic activity (139, 143, 170). Although this residue is implicated in tetramer formation, its effect on the cooperativity of DNMT3A has not been tested. We performed the cooperativity assays as described above using various concentrations of WT DNMT3AC and DNMT3A-C Arg878His enzyme (Figure 2.11A).

Methylation assays were performed using 32-mer, 509-mer, and pUC19 DNA as substrates of various lengths. Loss of 64% activity for the 32-mer substrate at two tested enzyme concentrations (0.5 and 1 μ M) confirms the previous observation that the mutation affects tetramer formation and that the tetrameric form is the most catalytically active form of DNMT3A-C (111, 142) (Figure 2.11B, C). Interestingly, on longer DNA substrates, although the activity was rescued at a lower enzyme concentration (0.5 μ M) (Figure 2.11D), at a higher enzyme concentration, the enzymatic activity did not demonstrate an exponential increase that is expected because of cooperativity (Figure 2.12A,B). These data support our observation for the WT enzyme that the length of DNA positively affects the cooperative behavior of DNMT3A. Destabilization of the tetramer structure due to the absence of Arg878 can be partially rescued by long DNA substrates, however not to the extent that it can restore cooperativity. Loss of cooperativity resulted in 78 and 63% losses of activity for 509-mer and pUC19, respectively, at higher enzyme concentrations. These data suggest that the human Arg882His mutation of DNMT3A affects its cooperative mechanism at multiple CpG sites, which would have a substantial effect on its activity *in vivo*. We next tested the effect of the analogous mutation of the conserved Arg829 in DNMT3B-C. On the basis of our data showing that the non-cooperative and processive mechanism of DNMT3B-C is not affected by acidification, we predicted that the mutation of Arg829 will not affect the processivity of DNMT3B-C. Our data in Figure 2.12C–E demonstrate that the variant enzyme Arg829His has activity and processivity comparable to those of the WT enzyme.

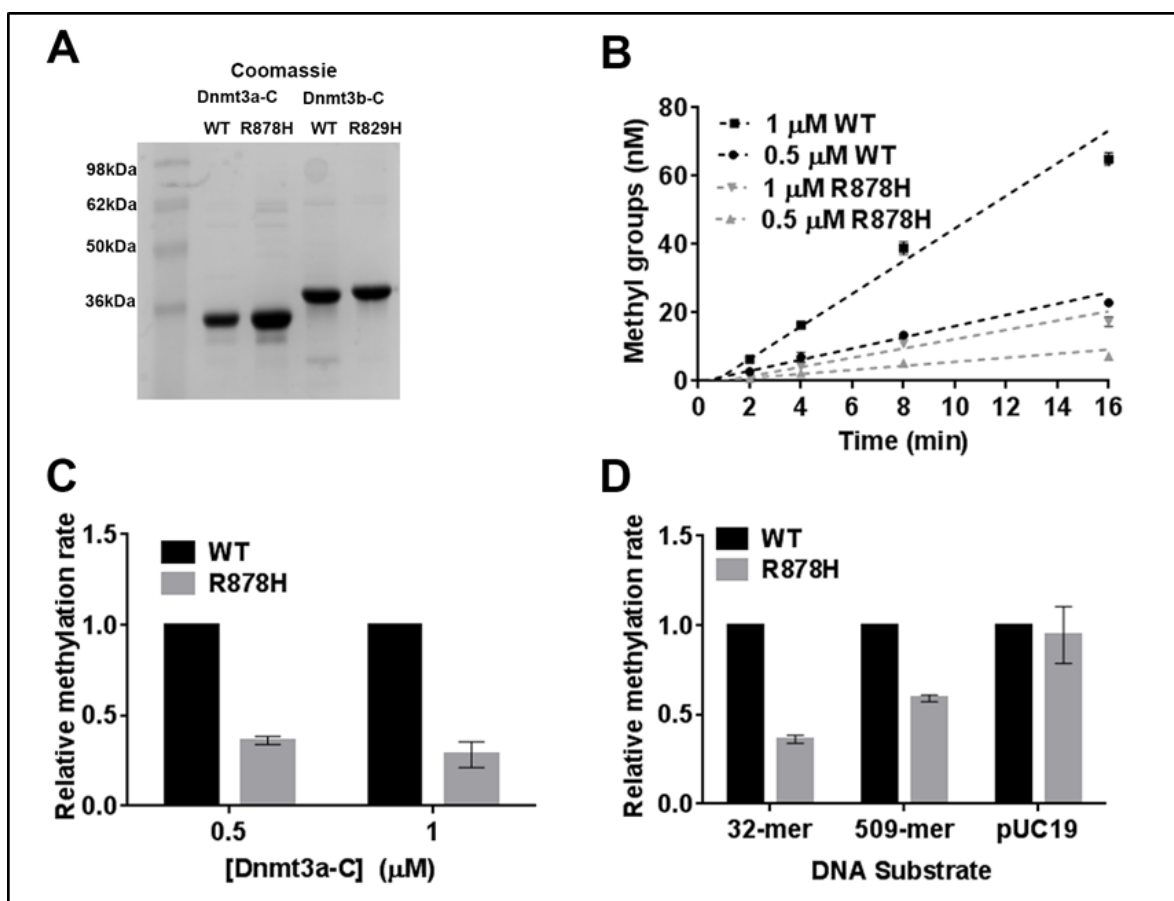


Figure 2.11. Loss of catalytic activity of the DNMT3A-C Arg878His variant. A, WT: wild type; Coomassie stained SDS-PAGE gel showing purified His-tagged WT DNMT3A-C, AML mutant DNMT3A-C Arg878His, WT DNMT3B-C, and the AML analogue mutant DNMT3B-C Arg829His. B, Initial velocity of methylation was measured at early time points (2, 4, 8, and 16 min) for 2-site substrates to compare WT and Arg878His activity, in the presence of 0.5 or 1 μ M of enzyme. The data were fit to linear regression which was weighted by $1/Y^2$. C, Methylation rates for the 32-mer of WT enzyme were set to 1 and the relative change in rate of methylation for the Arg878His mutant at different enzyme concentrations was plotted in the bar graph with normalized error. D, Methylation rates for the 0.5 μ M of WT enzyme were set to 1 and the relative change in rate of methylation for the Arg878His mutant on different DNA substrates was plotted in the bar graph with normalized error. Average \pm SD for each time point is derived from two to four independent experiments.

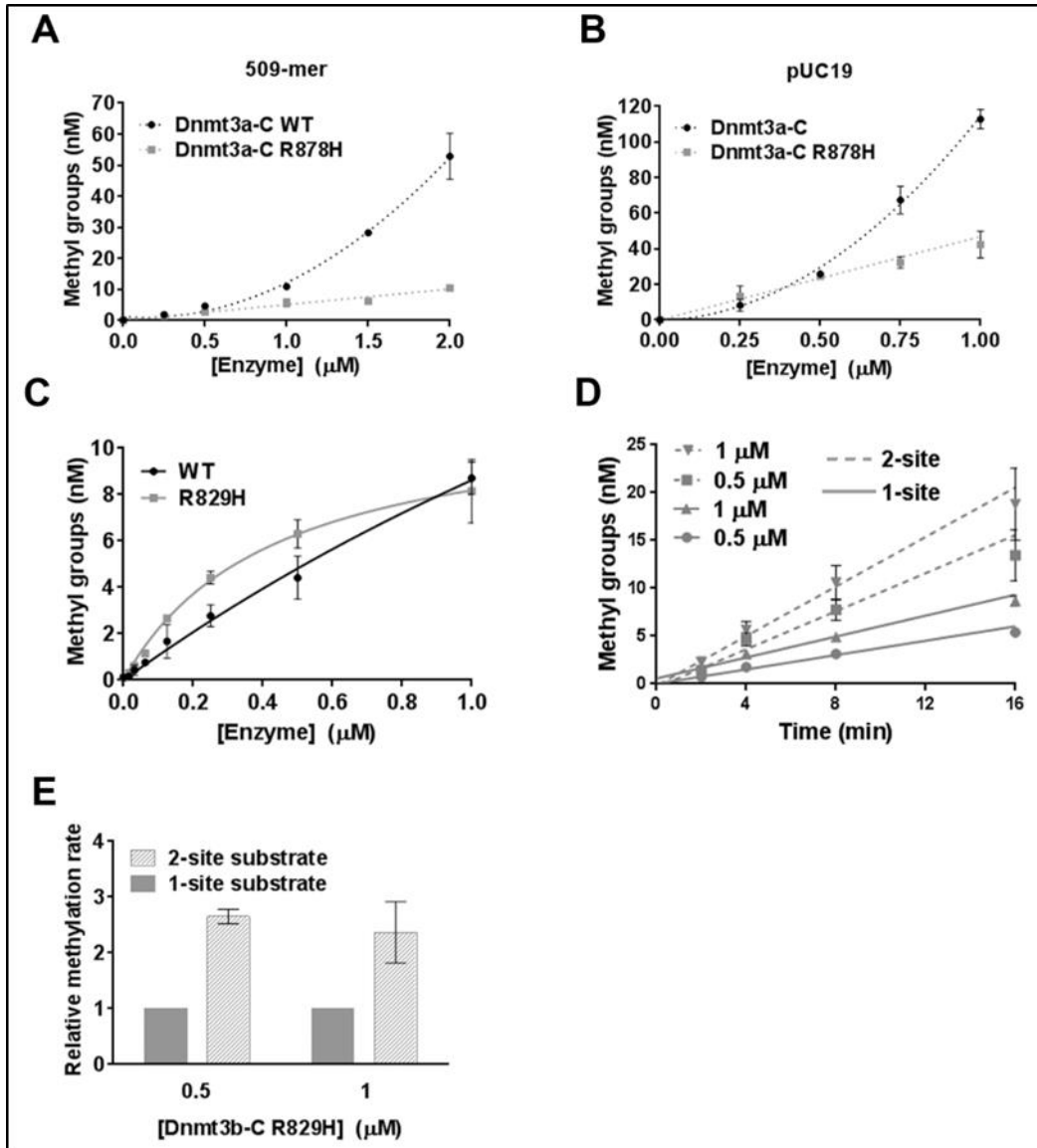


Figure 2.12 The Arg829His mutation of DNMT3B-C does not reduce its activity and processivity. (A and B) Methylation activity of the WT and the Arg878His mutant DNMT3A-C enzymes was measured for 10 minutes with the 509-bp or pUC19 substrate in the presence of 0.25 to 2 μM and 0.25 to 1 μM enzyme respectively. Enzymes were pre-incubated with DNA for 10 min at room temperature and the reaction was initiated by addition of AdoMet. Total methylation activity was plotted against enzyme concentration using an average and standard deviation ($n \geq 3$ independent experiments). (C) Methylation activity was measured for 10 minutes with the 30-mer substrate at enzyme concentrations ranging from 0.015 to 1 μM of the WT and the mutant Arg829His DNMT3B-C enzymes. (D) Initial velocity of methylation was measured for 1-site (solid line) and 2-site (dashed line) substrates, in the presence of 0.5 and 1 μM of DNMT3B-C Arg829His mutant enzyme. The data were fit to linear regression, weighted by $1/Y^2$. (E) Methylation rates for the 1-site substrate were normalized to 1, and the relative change in rate of methylation for the 2-site substrate at different enzyme concentrations were plotted with normalized error. For all the experiments average \pm standard error of the mean is derived ($n \geq 4$ independent experiments).

Taken together, these data show that DNMT3B-C methylates DNA in a non-cooperative and processive manner and suggest that DNMT3B-C may not require the interaction of DNMT3B monomers or dimers along the dimer interface for optimal activity.

2.6 Discussion

DNA methylation together with specific histone modifications plays an important role in regulating chromatin structure and thereby controlling gene expression. DNA methylation has a major role in regulation of epigenetic processes, including genomic imprinting, X-chromosome inactivation, nuclear reprogramming, and carcinogenesis (33, 174). Recent studies of the functions of the mammalian *de novo* methyltransferases DNMT3A and DNMT3B reveal both redundant and specific functions, highlighting the importance of differentiating their individual contributions to normal and diseased epigenomes (175). Whereas several biochemical studies have elucidated the structure–function relationship of DNMT3A, less is known about the biochemical properties of DNMT3B. Studies characterizing the enzymatic properties of DNMT3A-C have provided critical information that facilitates our understanding of the biochemical basis of DNMT3A function *in vivo*. This is underscored by the recent finding that ~20% of acute myeloid leukemia (AML) patients have mutations in DNMT3A of which many are present in the dimer or tetramer interface of the enzyme (111, 139, 142, 176, 177). These mutations were shown to disrupt oligomerization of DNMT3A and alter its catalytic properties *in vitro* and *in vivo*. Likewise, a high frequency of DNMT3B mutations is found in ICF syndrome, which is a rare genetic disease-causing immunodeficiency, centromeric instabilities, and facial abnormalities. Many of these mutations are present in the DNMT3B MTase domain and result in reduced catalytic activity *in vitro*, potentially linking the mutations to hypomethylation of DNA in the diseased state (86, 108, 127, 178-180). Some of these mutations are present in the potential DNA binding region of DNMT3B, and these residues are conserved in the dimer interface of DNMT3A. In the absence of a DNMT3B-C crystal structure, the detailed characterization of the catalytic mechanism of DNMT3B-C here provides important insights into the impact of these mutations on the catalytic mechanism of DNMT3B compared to that of DNMT3A and highlights its properties that are distinct from those of DNMT3A. Our findings suggest that the reduced activity of DNMT3B mutants in ICF syndrome may not be because of the dimer-interface disruption.

Processivity of DNMT3B-C has been reported previously; however, it was not clear whether the results reflected cooperative methylation of multiple sites by a multimerized enzyme or one enzyme unit processively methylating multiple sites (108). By using DNA substrates of various lengths, our assays in this study distinguish processivity from cooperativity and show that DNMT3B-C methylates DNA by a processive mechanism in a non-cooperative manner. We also show that DNMT3B can methylate both unmethylated and hemimethylated sites processively, consistent with its collaborative function with DNMT1 during the maintenance of DNA methylation. The non-cooperative mechanism is further supported by our data showing that on the longer DNA substrates, the catalytic activity of DNMT3B is not enhanced by the addition of its inactive mutant. Surprisingly, in our assays using DNA substrates of various lengths to examine the cooperativity of DNMT3B-C, the methylation rate was not higher for the long DNA, which would be expected from a processive enzyme. Because these experiments were performed by pre-incubating the enzyme with DNA, this could be explained if the enzyme prefers binding AdoMet before binding to DNA. Our data demonstrating a 10-fold decrease in the K_M of DNMT3B-C when it is pre-incubated with AdoMet support this conclusion. Our data are consistent with the model in which the binding of the enzyme to specific sites on DNA is favored by AdoMet binding, and during the processive catalysis, DNMT3B is capable of cofactor exchange without dissociating from DNA.

Perturbations in the intracellular microenvironment including pH has a regulatory effect in several DNA binding proteins and enzymes. These regulatory mechanisms are expected to be prominent during tumor progression and also during development. Acidification has been shown to disrupt cooperativity and reduce the catalytic activity of DNMT3A. Interestingly, again in contrast to that of DNMT3A, the catalytic activity of DNMT3B was not significantly affected at pH 6.5 and instead showed a slight increase in activity for a one-site substrate. We speculate that this may be because DNMT3B-C has a slightly higher K_D value at pH 6.5. This difference between DNMT3A and DNMT3B mechanisms could be explained by the differences in their amino acid sequences, especially in the DNA binding region and the dimer interface. The interaction of several DNA binding proteins involves either water mediated or direct H-bond interactions between the Arg and His residues and the DNA backbone. The protonation of His residues is pH sensitive, which can alter its interaction with DNA and affect the activity of the protein. The activity of DNMT3A decreases at lower pH because of the disruption of its dimer interface involving His821

and/or His873 (169). However, in DNMT3B, the residues at the equivalent positions are replaced by Tyr821 and Leu847, respectively, both of which are also a part of DNA binding loop L2 (106, 109) (Figure 2.13). Compared to DNMT3A, this makes the enzymatic activity of DNMT3B potentially less dependent on protein–protein interaction at the dimer interface and less sensitive to changes in pH. A high level of expression of DNMT3B has been implicated in the initiation and progression of several cancers and in some cases contributes to CpG island promoter methylation (162). Its ability to function under low-pH conditions can be exploited by cancer cells in which the intracellular environment is typically more acidic than that of normal cells.

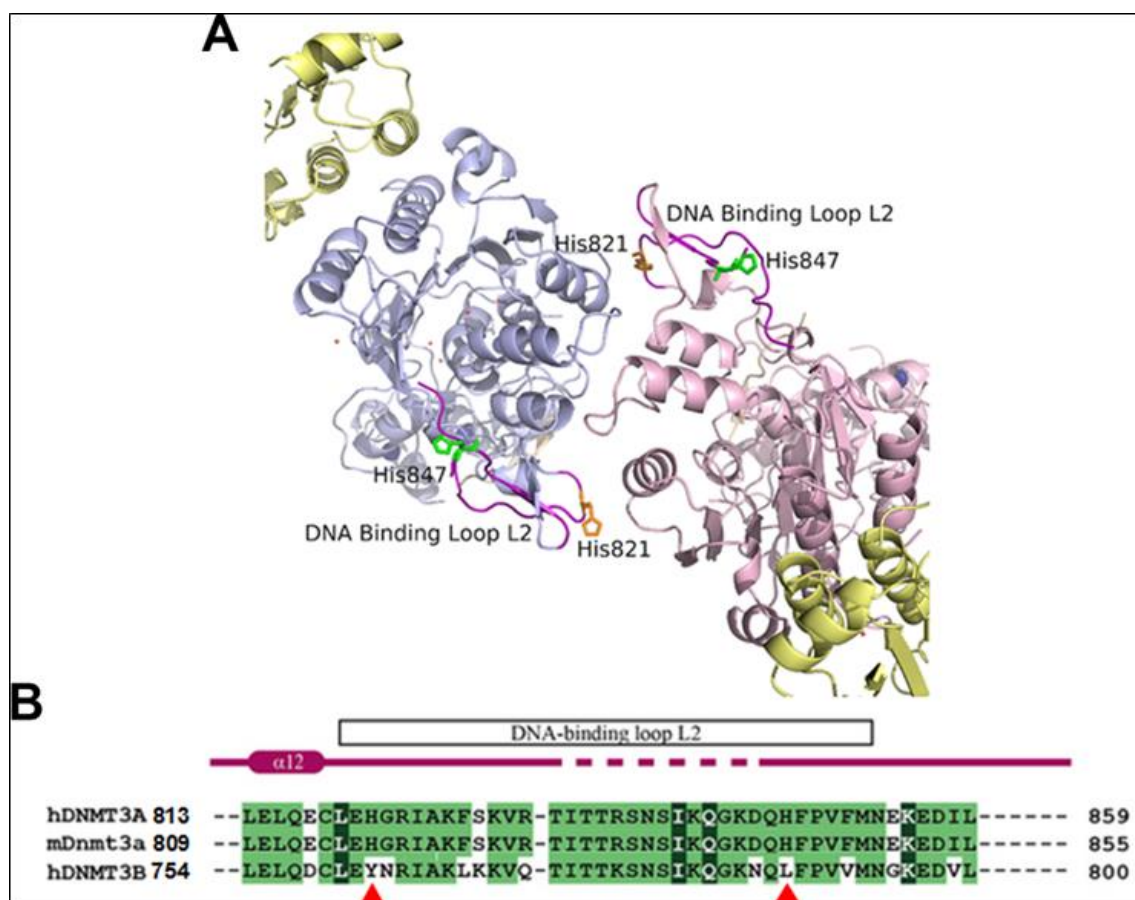


Figure 2.13 Illustration of the histidine bases that are not conserved in DNMT3B. (A) Interface of the symmetrical DNMT3A dimer in DNMT3A-DNMT3L hetero-tetramer based on the crystal structure (PDB, 2QRV). DNMT3A monomers are shown in blue and pink, while flanking DNMT3L is shown in yellow. Labeled histidine bases are located in the second DNA binding loop and are not conserved in DNMT3B. H821 is predicted to contribute to the pH sensitivity of DNMT3A (169). (B) Sequence alignment of DNMT3A and DNMT3B adapted from Guo. X, et al. 2015 (106). Red arrows mark DNMT3A H821 and H847, showing these residues are not conserved in DNMT3B.

Further supporting a non-cooperative mechanism, our data here show for the first time that the DNMT3A Arg878His mutation disrupts the cooperativity of the enzyme whereas the analogous Arg829His mutation in DNMT3B-C has no effect on its processivity. Although the SNP causing the Arg to His mutation in human DNMT3A was shown to be present at a very high frequency in AML cells (181), the analogous mutation is not reported for DNMT3B in any epigenetic diseases. Therefore, we speculate that a smaller contribution of the dimer interface would promote the processivity of DNMT3B, which may work efficiently for enzymes that do not form oligomers on DNA. These observations also emphasize that the minor difference in the amino acid sequence of DNMT3A and DNMT3B catalytic domains dictates a processive versus a cooperative mechanism.

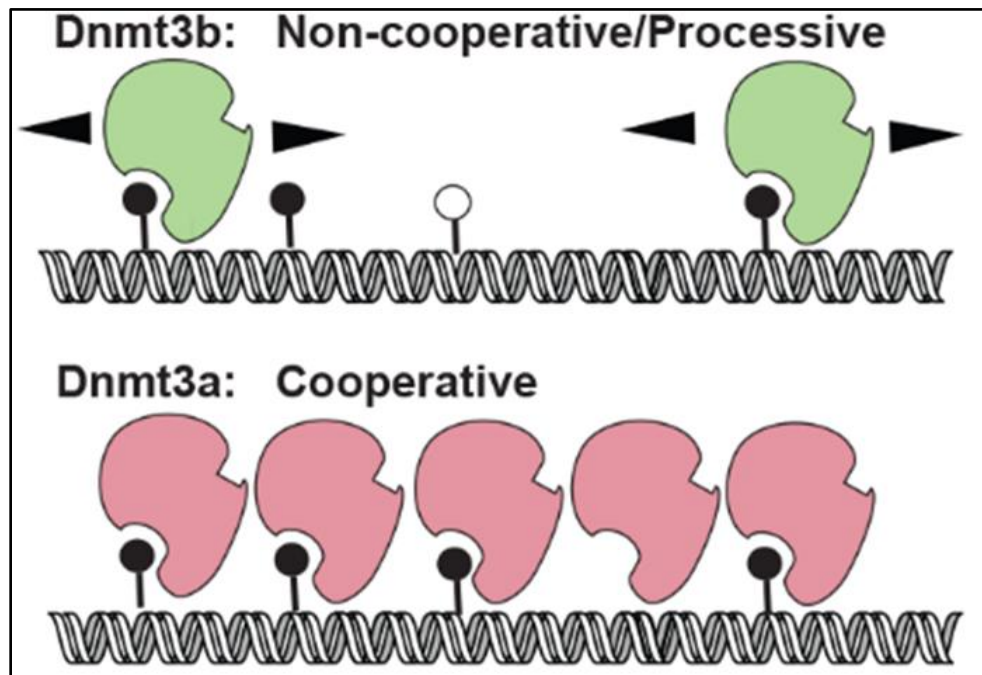


Figure 2.14 . Model showing the different kinetics mechanisms of DNMT3A-C and DNMT3B-C.

CHAPTER 3. THE ACUTE MYELOID LEUKEMIA VARIANT DNMT3A ARG882HIS IS A DNMT3B-LIKE ENZYME

*Formatted for dissertation from the article published in *Nucleic Acids Research*, Volume 48, Issue 7, 17 April 2020, Pages 3761–3775

Allison B. Norvil, Lama AlAbdi, Bigang Liu, Yu Han Tu, Nicole E. Forstoffer, Amie R. Michie, Taiping Chen, Humaira Gowher.

3.1 Declaration of collaborative work

Dr. Humaira Gowher and Allison B. Norvil conceived the project and designed experiments.

Allison B. Norvil, Lama AlAbdi, Nichole E. Forstoffer, and Amie R Michie purified recombinant proteins and performed radiolabeled methylation assays.

Allison B. Norvil and Yu Han Tu performed *in vitro* methylation assays and bisulfite conversion.

Dr. Humaira Gowher, Dr. Taiping Chen, Allison B. Norvil analyzed the data, generated figures, and wrote the manuscript.

Bigang Liu performed the rescue experiments in mESCs, and Dr. Taiping Chen supervised the tissue culture work.

3.2 Abstract

We have previously shown that the highly prevalent acute myeloid leukemia (AML) mutation, Arg882His, in DNMT3A disrupts its cooperative mechanism and leads to reduced enzymatic activity, thus explaining the genomic hypomethylation in AML cells. However, the underlying cause of the oncogenic effect of Arg882His in DNMT3A is not fully understood. Here, we discovered that DNMT3A WT enzyme under conditions that favor non-cooperative kinetics mechanism as well as DNMT3A Arg882His variant acquire CpG flanking sequence preference akin to that of DNMT3B, which is non-cooperative. We tested if DNMT3A Arg882His could preferably methylate DNMT3B-specific target sites *in vivo*. Rescue experiments in DNMT3A/3B double knockout mouse embryonic stem cells show that the corresponding Arg878His mutation in mouse DNMT3A severely impairs its ability to methylate major satellite DNA, a DNMT3A-preferred target, but has no overt effect on the ability to methylate minor satellite DNA, a DNMT3B-preferred target. We also observed a previously unappreciated CpG flanking sequence

bias in major and minor satellite repeats that is consistent with DNMT3A and DNMT3B specificity suggesting that DNA methylation patterns are guided by the sequence preference of these enzymes. We speculate that aberrant methylation of DNMT3B target sites could contribute to its oncogenic potential of DNMT3A AML variant.

3.3 Introduction

DNA methylation in mammals is critical for development and maintenance of the somatic cell state (182, 183). DNA methylation has diverse functions including regulation of gene expression and silencing of repetitive elements (184-186). In mammals, CpG methylation is established and maintained by DNA methyltransferases (DNMTs) (187). DNMT1 largely functions as a maintenance methyltransferase by copying the methylation pattern from parent to daughter strand during DNA replication (188, 189). The DNMT3 family includes two active homologs, DNMT3A and DNMT3B, and an inactive homolog DNMT3L (86, 91). These enzymes perform *de novo* DNA methylation, which predominantly takes place during early embryogenesis and stem cell differentiation (86, 190, 191). Whereas DNMT3A is expressed ubiquitously, DNMT3B is highly expressed during early embryogenesis and largely silenced in somatic cell types (89, 159, 192). Despite having a high sequence similarity (>40%), DNMT3A and DNMT3B have distinct preferences for some target sites, although many sites can be methylated redundantly (33). This is represented by a preference of DNMT3A for major satellite repeats and a preference of DNMT3B for minor satellite repeats (86, 190).

Mutations in DNMT3A and DNMT3B have been identified in several diseases. Germline transmitted mutations of DNMT3A and DNMT3B cause Tatton-Brown-Rahman syndrome and ICF (Immunodeficiency, Centromeric instability, and Facial anomalies) syndrome, respectively (86, 126, 127, 178, 193). Somatic mutations in DNMT3A are commonly found in patients with acute myeloid leukemia (AML) and other hematologic neoplasms (125, 139). Detailed studies revealed that ~20% of AML cases have heterozygous DNMT3A mutations, with the majority (~60-70%) carrying the Arg882His mutation, which cause genome-wide hypomethylation (139, 194-196). Given that genetic knockout of one copy of DNMT3A exhibits no obvious phenotype, the heterozygous DNMT3A Arg882His mutation was suggested to have a dominant negative effect (86, 139). The observation that the expression of the murine DNMT3A Arg878His variant (equivalent to Arg882His in human DNMT3A) in mouse embryonic stem cells (mESCs) causes

genome-wide loss of DNA methylation further supported the dominant negative activity of this variant (142, 143). In contrast, DNMT3B mutations have not been identified in cancer. However, aberrant overexpression of DNMT3B is highly tumorigenic, including in AML (192, 197-199). Overexpression of DNMT3B in AML leads to disease prognosis similar to that of patients with DNMT3A Arg882His mutation (198, 199).

The effect of Arg882His on the activity of DNMT3A could be predicted from the crystal structure of the DNMT3A MTase domain (DNMT3A-C), which shows that DNMT3A forms homodimers and tetramers through two interaction surfaces (106, 109). Several AML-associated DNMT3A mutations, including Arg882His, are present at or close to the protein-protein interface and disrupt tetramer formation. This leads to reduced catalytic activity *in vitro* (111, 139, 140, 142, 147, 200). Further, DNMT3A-C tetramers can oligomerize on DNA, forming nucleoprotein filaments (120, 165). This oligomerization allows the enzyme to bind and methylate multiple CpG sites in a cooperative manner, thus exponentially increasing its catalytic activity (118). Our previous biochemical studies show that in DNMT3A-C, the Arg882His substitution results in loss of cooperativity, causing a significant decrease in its catalytic activity. Interestingly, we also reported that in DNMT3B, which is a non-cooperative enzyme, the homologous mutation, Arg829His, has no effect on its catalytic activity *in vitro* (201). Further, the DNMT3A-C Arg882His variant was shown to have a bias for G at N+3 position from CpG site (N=CG dinucleotide) (144). This is supported by the interaction of Arg882 with phosphate backbone at the N+3 position observed in co-crystal structure of the DNMT3A-C/DNA complex (123). However, it is unclear whether change in flanking sequence preference is directly due to the substitution of Arg with His or caused by loss of cooperative mechanism in the variant enzyme.

Our data show that all variants of Arg882 found in AML patients have low catalytic activity and lack a cooperative mechanism. We further show that in the absence of cooperative mechanism at low enzyme concentrations, the WT DNMT3A-C enzyme prefers G at the N+3 position, thus behaving like the AML variant, DNMT3A-C Arg882His. To test the role of cooperativity, we compared the flanking sequence preference of non-cooperative enzyme DNMT3B, with that of the DNMT3A-C WT and the Arg882His variant. DNA methylation of 124 CpGs in four different substrates was rank ordered to compute consensus sequence motifs that are preferred by these enzymes. Interestingly, our data show that the Arg882His variant and DNMT3B-C have a similar preference for nucleotides at N+1, 2 and 3 positions flanking the CpG site. These data strongly

support that the gain of flanking sequence preference is due to loss of cooperative mechanism in DNMT3A-C Arg882His enzyme and suggest that the variant could methylate DNMT3B-C preferred sites. We tested this “gain of function” prediction by expressing WT mouse DNMT3A or the Arg878His (corresponding to Arg882His in human DNMT3A) in DNMT3A/3B double knockout (DKO) mESCs. Our data show that whereas the DNMT3A Arg878His variant failed to rescue methylation at the major satellite repeats (DNMT3A preferred target sites), its ability to methylate the minor satellite repeats (DNMT3B preferred target sites) was comparable to that of DNMT3A WT enzyme. Further, our analysis of the CpG flanking sequences of the satellite repeats show that whereas major satellite repeats have an A or T at the N+3 position, the minor satellite repeats are enriched in G at N+3 position. This observation provides the previously unknown explanation for the substrate specificity of DNMT3A and DNMT3B for major and minor satellite repeats, respectively. Taken together, our data provide novel mechanistic insights into DNMT3A and DNMT3B substrate specificities that could influence the oncogenic potential of these enzymes. We suggest that in leukemia, DNMT3A Arg882His substitution establishes a DNMT3B-like activity, the tumorigenic properties of which are exploited by cancer cells.

3.4 Materials and Methods

3.4.1 Protein Purification

Human DNMT3A-C WT, DNMT3A-C Arg882His, DNMT3A-C Arg882Cys, and DNMT3A-C Arg882Ser, and mouse DNMT3B-C WT, DNMT3B-C Arg829His, and DNMT3B-C Glu703Ala in pET28a(+) with a 6X His tag, were expressed and purified using affinity chromatography as described (201). Briefly, transformed BL21 (DE3) pLys cells were induced with 1 mM IPTG at OD₆₀₀ 0.3 and expressed for 2 h at 32°C. Harvested cells were washed with STE buffer (10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, 0.1 M NaCl), and suspended in Buffer A (20 mM Potassium Phosphate (pH 7.5), 0.5 M NaCl, 10% (v/v) glycerol, 1 mM EDTA, 0.1 mM DTT, 40 mM imidazole). Cells were disrupted by sonication, followed by removal of cell debris by centrifugation. Clarified lysate was incubated with 0.4 mL Ni-NTA agarose for 3 h at 4°C. The protein bound slurry was packed in a 2 ml BioRad column and washed with 150 ml Buffer A. Protein was eluted using 200 mM imidazole in Buffer A at pH 7.5, then stored in 20 mM HEPES

pH 7.5, 40 mM KCl, 1 mM EDTA, 0.2 mM DTT and 20% (v/v) glycerol at -80°C. The purity and integrity of recombinant proteins were checked by SDS-PAGE gel.

3.4.2 DNA Methylation assays using radiolabeled AdoMet

Radioactive methylation assays to determine kinetics parameters of recombinant enzymes were performed using ³H-labelled S-Adenosylmethionine (AdoMet) as a methyl group donor and biotinylated oligonucleotides of varying sizes bound on avidin-coated high-binding Elisa plates (Corning) as described(153). The DNA methylation reactions were carried out using either 250 nM 30-bp/32-bp DNA substrate in methylation buffer (20 mM HEPES pH 7.5, 50 mM KCl, and 1 mM EDTA, 5 µg/ml BSA). The methylation reaction included 0.76 µM [methyl-³H] AdoMet (PerkinElmer Life Sciences). Storage buffer was added to compensate for the different enzyme volumes in all reactions. Incorporated radioactivity was quantified by scintillation counting.

3.4.3 Processivity Assays

Methylation kinetics analyses were performed using two enzyme concentrations and short oligonucleotide 30-bp and 32-bp substrates with 1 and 2 CpG sites, respectively. Low enzyme concentrations relative to DNA substrate concentrations were used to ensure that the reaction occurred under multiple turnover conditions. Each DNA substrate was used at 250 nM, and a 1:1 ratio of labelled and unlabeled AdoMet (final concentration 1.5 µM) was used.

3.4.4 Cooperativity Assays

To examine cooperativity, increasing concentrations of enzyme were pre-incubated with DNA substrate for 10 minutes at room temperature prior to the addition of AdoMet to start the reaction. AdoMet was a mixture of unlabeled and 0.76 µM ³[H]-labeled AdoMet to yield a final concentration of 2 µM. Methylation assays were performed using 100 ng of an unmethylated pUC19 plasmid purified from dam-/dcm- *E.coli* strain (C2925I, NEB) or a 1-kb fragment containing 14 CpG sites amplified from the Meis1 enhancer were used as DNA substrates for filter binding assays. Briefly, 10 µl reaction mix was spotted on 0.5 in DE81 filter that was then washed 5 times in 0.2 M ammonium bicarbonate (NH₄HCO₃), washed 2-3 times with 100% ethanol, and air dried. Incorporated radioactivity was quantified by scintillation counting (202).

3.4.5 Flanking Sequence Preference

In vitro DNA methylation reactions were carried out using 100ng of a 509-bp DNA fragment amplified from the SUHW1 promoter, 100 ng of a 1-kb DNA fragment amplified from the Meis1 enhancer region, 100 ng of a 721-bp DNA fragment amplified from the Sirt4 enhancer region, or 100 ng of a 1092-bp region of the pUC19 plasmid. Methylation reactions were carried out in methylation buffer (20 mM HEPES pH 7.5, 50 mM KCl, 1 mM EDTA, and 0.05 mg/ml BSA using varying concentrations of each enzyme at 37°C. Samples were taken at 10, 30, or 60 minutes and reaction was stopped by freeze/thaw. DNA methylation was analyzed by bisulfite sequencing as described below.

Flanking sequence preference was also measured using short oligonucleotides and DNA methylation assays using radiolabeled AdoMet as described above. Sixteen different 30-bp substrates were used with varying combinations of the second and third nucleotide around the CpG site on either side (Table 3.1).

Table 3.1 Table containing the sequences of each 30-bp substrate used to determine substrate specificity at the N+2 and N+3 nucleotide positions. To make sure each side of the substrate is methylated equally, the flanking sequencing is the same on each side of the CpG site.

30 bp oligomers
5' - GAA GAT GGG ATT CAC GTG ACT AGA GTG TAA - 3'
5' - GAA GAT GGG ATA AAC GTT TCT AGA GTG TAA - 3'
5' - GAA GAT GGG ATA CAC GTG TCT AGA GTG TAA - 3'
5' - GAA GAT GGG ATT TAC GTA ACT AGA GTG TAA - 3'
5' - GAA GAT GGG ATT GAC GTC ACT AGA GTG TAA - 3'
5' - GAA GAT GGG ATG CAC GTG CCT AGA GTG TAA - 3'
5' - GAA GAT GGG ATC TAC GTA GCT AGA GTG TAA - 3'
5' - GAA GAT GGG ATG AAC GTT CCT AGA GTG TAA - 3'
5' - GAA GAT GGG ATG TAC GTA CCT AGA GTG TAA - 3'
5' - GAA GAT GGG ATG GAC GTC CCT AGA GTG TAA - 3'
5' - GAA GAT GGG ATT AAC GTT ACT AGA GTG TAA - 3'
5' - GAA GAT GGG ATC AAC GTT GCT AGA GTG TAA - 3'
5' - GAA GAT GGG ATC GAC GTC GCT AGA GTG TAA - 3'
5' - GAA GAT GGG ATA GAC GTC TCT AGA GTG TAA - 3'
5' - GAA GAT GGG ATA TAC GTA TCT AGA GTG TAA - 3'
5' - GAA GAT GGG ATC CAC GTG GCT AGA GTG TAA - 3'

3.4.6 Bisulfite Sequencing

Bisulfite conversion was performed using EpiTect Fast Bisulfite Conversion Kit (Qiagen, 59802) according to the manufacturer's protocol. PCR amplifications were performed with primers as described (118). The pooled samples were sequenced using NGS on Wide-Seq platform. The reads were assembled and analyzed by Bismark and Bowtie2. Methylated and unmethylated CpGs for each target were quantified, averaged, and presented as percent CpG methylation. Methylation assays were done at least 3 times using 3 different enzyme preparations and average percent methylation for each CpG was used to calculate preference. The standard error of the mean (SEM) for total percent methylation at all CpGs for each enzyme was also reported.

3.4.7 Rescue Experiments in mESCs

WT (J1) and DNMT3A/3B DKO mESCs were cultured on gelatin-coated petri dishes in mESC medium (DMEM supplemented with 15% fetal bovine serum, 0.1 mM nonessential amino acids, 0.1 mM β -mercaptoethanol, 50 U/mL penicillin, 50 μ g/mL streptomycin, and 103 U/mL leukemia inhibitory factor) (86). For the generation of stable clones expressing DNMT3A or DNMT3B proteins, mESCs were transfected with plasmid vectors using Lipofectamine 2000 (Invitrogen) and then seeded at low density on dishes coated with feeder cells, selected with 6 μ g/mL of Blasticidin S HCl (Gibco) for 7-10 days, and individual colonies were picked. Southern blot analysis of DNA methylation at the major and minor satellite repeats were performed as previously described (143, 190).

3.4.8 Comparative Flanking Sequence Preference Analysis

To compare the flanking sequence preference of different enzymes, we first calculated average percent methylation of all CpG sites in the substrate. The fractional variance (v) at each CpG site was calculated by dividing the percent methylation of each site by average methylation. The preference for a site by an enzyme (B) over enzyme (A) was calculated as relative change $(v_B - v_A) / v_A$. Positive values indicate a preference of enzyme B for a site and values greater than or equal to 1 were considered significant given the preference is more than 2-fold (Table 3.2). To determine the fractional distribution of nucleotides at preferred sites the occurrence of each nucleotide at N+1/2/3 positions was calculated as a fraction (N=CG dinucleotide). At each position, the number of times each nucleotide occurred was divided by the total number of preferred sites to compute the fractional occurrence.

Table 3.2 Data analysis for Figures 3.4B and 3.6A.

Data from Fig 2b.											
Base	Site No.	DNMT3A-C: WT 1 μM ²	Fractional Variance (A)	DNMT3A-C: WT 0.25 μM ²	Fractional Variance (B)	Relative Change ((B-A)/A)	Fold Change 2 ² ((B-A)/A)	Sites preferred by DNMT3B-C: WT compared to DNMT3A-C: WT			
71	1	9.5	0.526367864	1.5	0.109147476	0.936701584	1.91416941				
89	2	34.9	1.93370931	2.4	0.1631067961	-0.156508193	0.897193956	G	N + 1	N + 2	N + 3
94	3	30.5	1.689917879	2.6	0.1766990291	0.045607194	1.03211748	C	A		C
100	4	7.1	0.393390719	1.6	0.1087378641	1.764118693	3.396664423	C	G		C
104	5	21.3	1.180172158	2.3	0.1563106796	0.32447354	1.252207407	G	C	G	
112	6	5.5	0.30473929	1.1	0.747572816	1.45315534	2.738062427				
114	7	4.8	0.265954289	1.1	0.747572816	1.81090716	3.508628402	Fractional distribution of bases at the preferred sites			
120	8	19.8	1.097061443	1.3	0.883495146	-0.194671227	0.873771996		N + 1	N + 2	N + 3
122	9	0.2	0.509745721	1	0.67961165	0.333236598	1.259836572	A		0	0.25
127	10	10.5	0.581775007	1.2	0.815533981	0.401803051	1.321158036	C	0.5	0.25	0.25
137	11	24.4	1.351934303	1.7	0.155339806	-0.145417197	0.904117888	G	0.5	0.25	0.75
142	12	37.6	2.083308598	2.2	0.495145631	-0.282321576	0.822266765	T	0	0.25	0
176	13	10.6	0.587315722	1.1	0.747572816	0.27286362	1.208203625				
185	14	13.4	0.742455724	1.5	0.109147476	0.373034705	1.295074152				
187	15	29.6	1.640051449	1.6	0.1087378641	-0.336985043	0.791694073				
192	16	38.5	2.133175027	2.1	0.1427184466	-0.330957635	0.795008596				
198	17	45.5	2.521025032	2.3	0.1563106796	-0.379971727	0.76845265				
212	18	14.4	0.797862867	1.2	0.815533981	0.022148058	1.015470309				
228	19	29.9	1.656673593	1.8	0.122300971	-0.261592038	0.834166894				
230	20	12.6	0.698130009	1.3	0.883495146	0.265516644	1.202066444				
238	21	15.7	0.869892154	1.4	0.951456311	0.093763527	1.067150407				
244	22	25.1	1.390719303	1.4	0.951456311	-0.315853092	0.803375799				
255	23	5.7	0.315820718	0.7	0.475728155	0.506323454	1.420425784				
261	24	19.1	1.058276442	1.3	0.883495146	-0.16515656	0.891831734				
267	25	11.7	0.64826358	1.3	0.883495146	0.362864078	1.285976321				
278	26	8.7	0.482042149	1.5	0.109147476	1.114789086	2.165633464				
281	27	8.8	0.487582863	1.1	0.747572816	0.533222087	1.447157643				
284	28	8.7	0.482042149	0.8	0.54368932	0.127887513	1.092692538				
290	29	52.1	2.88671218	3.6	2.446601942	-0.152460727	0.899714559				
306	30	22.2	1.230038587	1.5	0.109147476	-0.171231304	0.8880844				
309	31	16.4	0.908677154	1.4	0.951456311	0.04078499	1.033170602				
312	32	18.8	1.041654299	1.7	0.155339806	0.109193982	1.078584631				
315	33	9.2	0.509745721	0.9	0.611650485	0.199912938	1.148629037				
318	34	8.3	0.459879292	1.1	0.747572816	0.625584864	1.542836159				
320	35	14.3	0.792322153	1.2	0.815533981	0.029295947	1.020513981				
324	36	41.6	2.304937172	2.9	1.970873786	-0.144933836	0.904420855				
334	37	20.1	1.113683586	1.7	0.155339806	0.037403999	1.026265491				
336	38	18.4	0.109491442	1.6	0.1087378641	0.066589278	1.047237946				
338	39	14.8	0.820025725	1.2	0.815533981	-0.005477565	0.99621044				
340	40	24.5	1.357475017	1.5	0.109147476	-0.24903408	0.841459605				
357	41	13.5	0.747996438	1.6	0.1087378641	0.453721683	1.369568744				
375	42	15.3	0.847729297	1.4	0.951456311	0.122358652	1.088513011				
378	43	13.5	0.747996438	1.6	0.1087378641	0.453721683	1.369568744				
380	44	10.5	0.581775007	1.4	0.951456311	0.635436893	1.553408099				
386	45	6.4	0.354605719	0.6	0.40776699	0.149916566	1.109505305				
389	46	5.6	0.310280004	0.8	0.54368932	0.752253814	1.68442222				
391	47	11.3	0.626100722	0.9	0.611650485	-0.023079732	0.984129632				
397	48	30.5	1.689917879	1.7	0.155339806	-0.316333758	0.803108182				
412	49	31.3	1.734243594	2.6	0.1766990291	0.01888241	1.013174315				
414	50	18.4	0.109491442	1.4	0.951456311	-0.066734382	0.954796788				
416	51	25.3	1.401800732	1.5	0.109147476	-0.272780038	0.827723006				
424	52	17.9	0.99178787	1.2	0.815533981	-0.177713294	0.884103211				
436	53	13.9	0.770159296	0.7	0.475728155	-0.382299015	0.767214018				
439	54	12.1	0.670426437	0.9	0.611650485	-0.087669502	0.94104166				
447	55	5.7	0.315820718	0.6	0.40776699	0.291134389	1.223602016				
449	56	5.7	0.315820718	0.8	0.54368932	0.721512519	1.648909843				
AVG		18.04821429		1.471428571							

Table 3.2 continued

Data from Fig 3a												
Base	Site No.	DNMT3A-C WT %	Fractional Variance (A)	DNMT3B-C WT %	Fractional Variance (B)	Relative Change ((B-A)/A)	Fold Change 2 ⁴ ((B-A)/A)		Sites preferred by DNMT3B-C WT compared to DNMT3A-C WT			
71	1	9.5	0.526367864	2.8	1.154639175	1.193597396	2.287223575		N + 1	N + 2	N + 3	
89	2	34.9	1.93370931	0.8	0.329896907	-0.829396846	0.562764471		G	G	G	
94	3	30.5	1.689917879	11.9	4.907216495	1.903819503	3.742025794		G	C	G	
100	4	7.1	0.393390719	1	0.412371134	0.048248252	1.034008648		G	C	T	
104	5	21.3	1.180172158	7.1	2.927835052	1.480854197	2.791139434		G	C	A	
112	6	5.5	0.30473929	1.8	0.742268041	1.435747757	2.705223436		C	G	G	
114	7	4.8	0.265954289	0.9	0.371134021	0.395480486	1.315380772		G	C	C	
120	8	19.8	1.097061443	2	0.824742268	-0.248226001	0.841931053		C	G	G	
122	9	9.2	0.509745721	2.1	0.865979381	0.698845809	1.623205669		G	C	T	
127	10	10.5	0.581775007	1.1	0.453608247	-0.220302967	0.858385156		G	C	T	
137	11	24.4	1.351934303	4.2	1.731958763	0.28109684	1.215118353		C	T	G	
142	12	37.6	2.083308598	0.9	0.371134021	-0.821853555	0.565714653		C	G	G	
176	13	10.6	0.587315722	2.9	1.195876289	1.036172785	2.050780073					
185	14	13.4	0.742455724	0.9	0.371134021	-0.500126393	0.707044835		Fractional distribution of bases at the preferred sites			
187	15	29.6	1.640051449	1.8	0.742268041	-0.547411734	0.6842466		N + 1	N + 2	N + 3	
192	16	38.5	2.133175027	3	1.237113402	-0.420060058	0.747393511		A	0	0	0.090909
198	17	45.5	2.521025032	2	0.824742268	-0.672854392	0.627264408		C	0.363636	0.545455	0.090909
212	18	14.4	0.797862867	1.6	0.659793814	-0.173048601	0.886966425		G	0.636364	0.363636	0.545455
228	19	29.9	1.656673593	1.3	0.536082474	-0.676410322	0.625720241		T	0	0.090909	0.272727
230	20	12.6	0.698130009	1.6	0.659793814	-0.054912687	0.962652702					
238	21	15.7	0.868892154	2.1	0.865979381	-0.004497997	0.996887081					
244	22	25.1	1.390719303	1.5	0.618556701	-0.555225343	0.680550751					
255	23	5.7	0.315820718	0.9	0.371134021	0.175141462	1.12907511					
261	24	19.1	1.058276442	0.8	0.329896907	-0.68826963	0.62059775					
267	25	11.7	0.64826358	1.4	0.577319588	-0.109436955	0.926949755					
278	26	8.7	0.482042149	1.6	0.659793814	0.368747143	1.29123102					
281	27	8.8	0.487582863	2.3	0.948453608	0.945215223	1.92547611					
284	28	8.7	0.482042149	1.3	0.536082474	0.112107054	1.080805599					
290	29	52.1	2.88671218	2.8	1.154639175	-0.60001583	0.659746716					
306	30	22.2	1.230385887	5.4	2.226804124	0.810353063	1.753640549					
309	31	16.4	0.908677154	0.9	0.371134021	-0.591566687	0.663621859					
312	32	18.8	1.041654299	1.1	0.453608247	-0.564530912	0.676175237					
315	33	9.2	0.509745721	1.4	0.577319588	0.132563873	1.096240145					
318	34	8.3	0.459879292	1.4	0.577319588	0.255372003	1.193643488					
320	35	14.3	0.792322153	1.2	0.494845361	-0.375449293	0.770865308					
324	36	41.6	2.304937172	2.6	1.072164948	-0.534839838	0.690235299					
334	37	20.1	1.113683586	2.7	1.113402062	-0.000252786	0.999824797					
336	38	18.4	1.019491442	2.2	0.907216495	-0.110128386	0.926505608					
338	39	14.8	0.820025725	1.3	0.536082474	-0.346261394	0.786619909					
340	40	24.5	1.357475017	1.6	0.659793814	-0.513955096	0.700299957					
357	41	13.5	0.747996438	2.5	1.030927835	0.378252332	1.299766376					
375	42	15.3	0.847729297	2	0.824742268	-0.027116001	0.981380152					
378	43	13.5	0.747996438	1.6	0.659793814	-0.117918508	0.921516238					
380	44	10.5	0.581775007	2.7	1.113402062	0.913801809	1.884003715					
386	45	6.4	0.354605719	4.4	1.81443299	4.116761782	17.34877367					
389	46	5.6	0.310280004	1.5	0.618556701	0.993543551	1.99106946					
391	47	11.3	0.626100722	3.3	1.360824742	1.17349173	2.255569485					
397	48	30.5	1.689917879	4.3	1.773195876	0.049279316	1.034747897					
412	49	31.3	1.734243594	2.4	0.989690722	-0.429324274	0.742609525					
414	50	18.4	1.019491442	2.6	1.072164948	0.051666453	1.036461447					
416	51	25.3	1.401800732	2.2	0.907216495	-0.352820644	0.783051639					
424	52	17.9	0.99178787	3.2	1.319587629	0.330513983	1.257461285					
436	53	13.9	0.770159296	1.5	0.618556701	-0.196845763	0.872455974					
439	54	12.1	0.670426437	6.2	2.556701031	2.813544469	7.030096382					
447	55	5.7	0.315820718	2	0.824742268	1.611425471	3.05535983					
449	56	5.7	0.315820718	5.2	2.144329897	5.789706224	55.31911704					
AVG		18.04821429		2.425								

3.4.9 Individual Flanking Sequence Preference Analysis

Analysis of the bisulfite sequencing data was performed to determine optimal flanking sequence preferred by each enzyme. To consider the uneven distribution of nucleotides flanking the CpG site, the occurrence of each nucleotide at the N+1/2/3 positions (p1n, p2n, and p3n) was calculated by dividing the number of times it occurred by the total number of CpG sites (s(p1n), s(p2n), s(p3n)). The expected occurrence of a trinucleotide set (O) was computed by multiplying the occurrence of three nucleotides as described by equation (1). This created an expected value, which would predict the probability at which this trinucleotide set would be methylated if there were no flanking sequence preference by the enzyme. With the data obtained from bisulfite sequencing, the fractional methylation (f) for each CpG site was calculated by dividing percent methylation at a site by sum of percent methylation at all sites as described by equation (2). Fractional methylation was sorted by nucleotides at position N+1/2/3 (f(p1n), f(p2n), and f(p3n)) and summed for each nucleotide at a position ($\Sigma f(p1n)$, $\Sigma f(p2n)$, and $\Sigma f(p3n)$). The methylation for the flanking trinucleotide sets (M) was calculated then multiplying the summed value for a nucleotide at the three positions as described by equation (3). This gives us the value for observed methylation value at a CpG with a specific flanking trinucleotide. The preference for a flanking trinucleotide by an enzyme was calculated by determining the fold change between the observed and expected values as described by equation (4) (Table 3.3).

Table 3.3 Data for Figure 3.9A, DNMT3A-C WT at 10 minutes on the 509-bp substrate.

Data from Fig 5a													
Base	N+1	N+2	N+3	DNMT3A WT %	Fractional Methylation (f)	Occurrence (O)	No. of sites (S)	Tri-nucleotide Seq	Occurrence (O = f(p1,1)*f(p2,1)*f(p3,1))	Methylated (M = f(p1,1)*f(p2,1)*f(p3,1))	Fold Change ((M-O)/O)		
71	G	C	T	9.5	0.00399406	N + 1 position (p1)	s(p1,1)	AAA	0.001912625	0.001727035	1.69133907		
89	C	T	C	34.9	0.034530523	A (p1,1)	0.142857143	AAC	0.005737875	0.003251841	1.364222573		
94	G	C	A	30.5	0.030177105	C (p1,1)	0.517857143	AAG	0.004911621	0.001937353	0.135002855		
100	C	A	C	7.1	0.007024834	G (p1,1)	0.303571429	AAT	0.002738879	0.001913086	1.220831198		
104	G	C	C	21.3	0.021074503	T (p1,1)	0.035714286	ACA	0.008294	0.013750278	0.98310352		
112	G	G	G	5.5	0.005441773			ACC	0.024882	0.025890453	0.74090418		
114	G	T	G	4.8	0.004749184	N + 1 position (p2)	s(p2,1)	ACG	0.021298992	0.015424785	-0.164236381		
120	C	G	A	19.8	0.019930383	A (p2,1)	0.107142857	ACT	0.011877079	0.015215172	0.630317401		
122	A	C	C	9.2	0.009102602	C (p2,1)	0.464285714	AGA	0.005737875	0.006666135	0.495241151		
127	C	A	G	10.5	0.010388839	G (p2,1)	0.321428571	AGC	0.017213625	0.012551682	0.312625156		
137	G	C	T	24.4	0.024141684	T (p2,1)	0.107142857	AGG	0.014734863	0.007477936	-0.369842198		
142	C	C	A	37.6	0.037201939			AGT	0.008216637	0.007384266	0.233013732		
176	C	T	G	10.6	0.010487781	N + 1 position (p3)	s(p3,1)	ATA	0.001912625	0.001870038	0.45999212		
185	C	G	C	13.4	0.013258138	A (p3,1)	0.125	ATC	0.005737875	0.003521102	0.281681141		
187	C	A	A	29.6	0.029286633	C (p3,1)	0.375	ATG	0.004911621	0.002097771	-0.384697629		
192	C	C	C	38.5	0.038092411	G (p3,1)	0.321428571	ATT	0.002738879	0.002071494	0.203946488		
198	T	C	C	45.5	0.045018304	T (p3,1)	0.178571429	CAA	0.006948	0.00417995	1.094252843		
212	G	A	T	14.4	0.014247551			CAC	0.020544	0.011636252	0.838478672		
228	C	G	A	29.9	0.029583457	Methylated (M)		CAG	0.017585664	0.009312543	-0.11739336		
230	A	G	C	12.6	0.012466607	N + 1 position (p1)	2f(p1,1)	CAT	0.009806336	0.006485705	0.726973948		
238	G	C	C	15.7	0.015533788			CCA	0.029696	0.04020342	0.542110954		
244	C	A	C	25.1	0.024834723	A (p1,1)	0.126955575 2f(p1,1)	CCC	0.089088	0.002645315	0.351770962		
255	G	A	G	5.7	0.005629656	C (p1,1)	0.597704561 2f(p1,1)	CCG	0.076259328	0.051356405	-0.350280282		
261	C	T	C	19.1	0.018897794	G (p1,1)	0.23680934 2f(p1,1)	CCT	0.042534672	0.054504222	0.271663759		
267	G	G	G	11.7	0.011576135	T (p1,1)	0.069259329 2f(p1,1)	CGA	0.020544	0.023853818	0.162736959		
278	G	C	G	8.7	0.008607896			CGC	0.061632	0.044914449	0.020730189		
281	G	C	G	8.8	0.008706837	N + 2 position (p2)	2f(p2,1)	CGG	0.052756992	0.026758732	-0.509973514		
284	A	C	T	8.7	0.008607896	A (p2,1)	0.076283764 2f(p2,1)	CGT	0.029419008	0.02642355	-0.041177647		
290	C	C	T	52.1	0.051548432	C (p2,1)	0.543880479 2f(p2,1)	CTA	0.006848	0.006691605	0.13532643		
306	G	C	G	22.2	0.021964975	G (p2,1)	0.272781241 2f(p2,1)	CTC	0.020544	0.012599763	-0.003332652		
309	C	C	G	16.4	0.016226378	T (p2,1)	0.107054517 2f(p2,1)	CTG	0.017585664	0.007506075	-0.521525469		
312	C	G	G	18.8	0.01860097			CTT	0.009806336	0.007112547	-0.061781864		
315	A	C	G	9.2	0.009102602	N + 3 position (p3)	2f(p3,1)	GAA	0.004052625	0.005380643	0.544628336		
318	C	G	C	8.3	0.00821213	A (p3,1)	0.169585436 2f(p3,1)	GAC	0.012157875	0.010507814	0.355808788		
320	C	A	C	14.3	0.01414861	C (p3,1)	0.446621154 2f(p3,1)	GAG	0.010407141	0.006260253	-0.349028267		
324	C	C	T	41.6	0.041159592	G (p3,1)	0.183536163 2f(p3,1)	GAT	0.005803359	0.006181836	0.27379821		
334	C	G	C	20.1	0.019887207	T (p3,1)	0.200257247 2f(p3,1)	GCA	0.017574	0.044431868	0.137393029		
336	C	G	C	18.4	0.018205204			GCC	0.052722	0.083660941	-0.00518449		
338	C	G	T	14.8	0.014643317			GCG	0.045138032	0.048842775	-0.520654917		
340	T	C	C	24.5	0.024240625			GCT	0.025165968	0.042128439	-0.062076894		
357	A	T	T	13.5	0.013357079			GGA	0.012157875	0.02154057	-0.142416498		
375	G	C	G	15.3	0.015138023			GGC	0.036473625	0.040558824	-0.247154429		
378	C	G	A	13.5	0.013357079			GGG	0.031221423	0.04163777	-0.638578075		
380	A	G	A	10.5	0.010388839			GGT	0.017410377	0.028610599	-0.293814019		
386	G	C	G	6.4	0.006332245			GTA	0.004052625	0.006042734	-0.162633295		
389	C	G	G	5.6	0.005540714			GTC	0.012157875	0.011377888	-0.264902119		
391	G	C	T	11.3	0.01118037			GTG	0.010407141	0.006778617	-0.647096388		
397	A	G	C	30.5	0.030177105			GTT	0.005480339	0.006693708	-0.309486202		
412	C	G	C	31.3	0.03096836			TAA	0.0004815	0.000579974	0.540693657		
414	C	G	C	18.4	0.018205204			TAC	0.0014445	0.001092036	0.352526481		
416	C	T	C	25.3	0.025032156			TAG	0.001236492	0.000560004	-0.350686591		
424	C	C	C	17.9	0.017710498			TAT	0.000885928	0.000543454	0.274959502		
436	A	C	G	13.9	0.013752845			TCA	0.003088	0.004617631	0.134495566		
439	G	C	T	12.1	0.011971501			TCC	0.006264	0.008894555	-0.004062041		
447	C	G	G	5.7	0.005639656			TCG	0.005361984	0.005179965	-0.52187563		
449	G	G	G	5.7	0.005639656			TCT	0.002990316	0.00511508	-0.064466215		
			SUM	1010.7				TGA	0.0014445	0.002238627	-0.144601157		
								TGC	0.0043335	0.00421512	-0.249072272		
								TGG	0.003709476	0.002511247	-0.639498783		
								TGT	0.002068324	0.003479791	-0.294616444		
								TTA	0.0004815	0.000627998	-0.164766453		
								TTT	0.0014445	0.001182459	-0.266774751		
								TTG	0.001236492	0.000704475	-0.647997291		
								TTT	0.000885928	0.000895651	-0.311245258		

3.4.10 Equations

v = fractional variance

f = fractional methylation

n = any nucleotide

A,C,G,T = specific nucleotides

s = fraction (number of sites out / 56)

p1 = nucleotide at position N+1

p2 = nucleotide at position N+2

p3 = nucleotide at position N+3

M = observed methylation based on fractional methylation

O = normalized occurrence of each site based on their frequency in the DNA substrate

Σ = sum

$$O = s(p1n) * s(p2n) * s(p3n) \quad (1)$$

$$f = \frac{\% \text{Methylation at each CpG}}{\text{Sum of all methylation}} \quad (2)$$

$$M = \Sigma f(p1n) * \Sigma f(p2n) * \Sigma f(p3n) \quad (3)$$

$$\text{Relative Change} = \frac{M-O}{O} \quad (4)$$

3.4.11 Consensus Sequence Analysis of Major and Minor Satellite Repeats

From the mm9 genome, three regions, chr9: 3000466 – 3028144, chr9: 3033472 – 3037264, and chr2: 98506820 – 98507474, were used for the sequences of the major satellite repeats. The minor satellite repeat sequences were obtained from chr2:98505036-98505275 and chr2:98506495-98506615. Additionally, more sequences were obtained from GenBank and used for analysis (accession no. X14462.1, X14463.1, X14464.1, X14465.1, X14466.1, X14468.1, X14469.1, X14470.1). Consensus sequences of the CpG sites and flanking regions in the major and minor satellite repeats were built using WebLogo (<https://weblogo.berkeley.edu/logo.cgi>).

3.4.12 Line and Bar Graphs

Data were analyzed using the Prism software (GraphPad). For cooperativity graphs, data were fit with nonlinear, second order polynomial regression curves. Errors were calculated as standard error of the mean (SEM) for two to four independent experiments, as described in the figure legends. P-values for the relative change in flanking sequence preference was calculated with the following formula with Excel.

$$p\text{-value} = 1 - \text{BINOM.DIST}(\text{number_s}, \text{trials}, \text{probability_s}, \text{TRUE}) \quad (5)$$

3.5 Results

3.5.1 The cooperative kinetics mechanism is absent in DNMT3A Arg882 variants

DNMT3A-C WT methylates neighboring CpGs on DNA by a cooperative kinetics mechanism, in which the DNA bound tetramer promotes successive binding events through protein-protein interaction (118). Our previous studies also show that the DNMT3A-C Arg882His variant fails to methylate DNA by a cooperative mechanism, which could be due to impaired tetramerization (111, 143, 201). It is not clear, however, whether this is due to the loss of Arg or gain of His at this position. Based on the observation that many AML patients have mutations that lead to substitution of Arg882 to Cys or Ser, we asked if these variant enzymes could methylate multiple CpGs on a single DNA molecule in a cooperative manner. His-tagged recombinant DNMT3A-C (MTase domain) WT and Arg882 variants were produced in *E. coli* and purified using Ni-NTA affinity chromatography to about 90 – 95% purity (Figure 3.1A) (108). The catalytic activity of DNMT3A-C Arg882 variants was compared to the WT enzyme by performing DNA methylation assays using a 30-bp substrate containing one CpG site. Consistent with previous reports, a 60 - 80% loss of catalytic activity was observed in all variants compared to WT enzyme (Figure 3.2A) (111, 140, 142, 144, 200, 201). We next assayed the cooperative kinetics mechanism of these enzymes. This is observed as an exponential relationship between the catalytic activity and enzyme concentrations as shown for DNMT3A-C WT enzyme. However, all Arg882 variant enzymes failed to methylate the substrate in a cooperative fashion (Figure 3.2B). From these data, we conclude that the Arg882 residue plays a key role in the cooperative mechanism of DNMT3A-C.

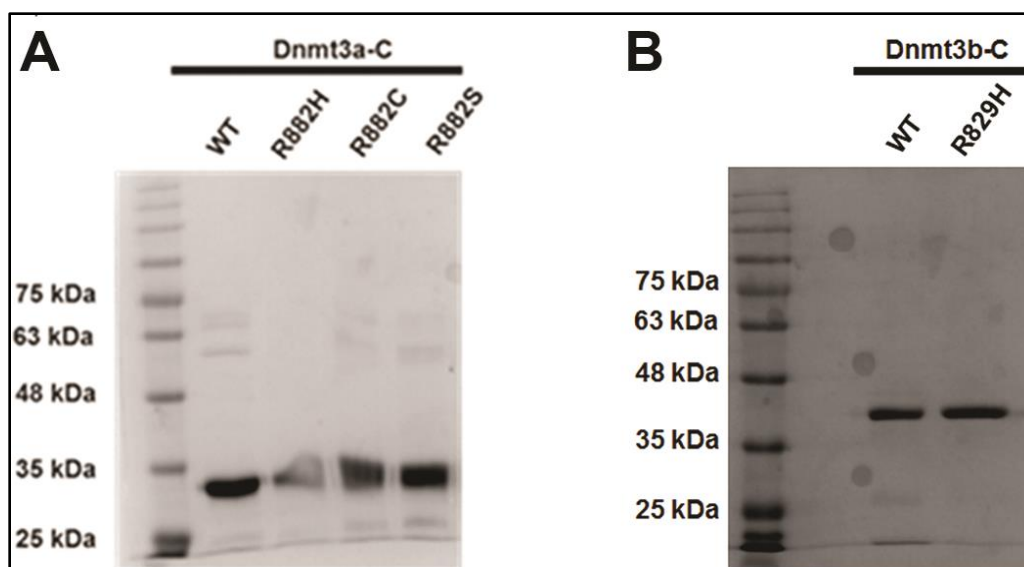


Figure 3.1 Coomassie-stained SDS-PAGE gel showing purified recombinant His-tagged DNMT3 enzymes. (A) DNMT3A-C WT, Arg882His, Arg882C, and Arg882S. (B) DNMT3B-C WT and DNMT3B-C Arg829His.

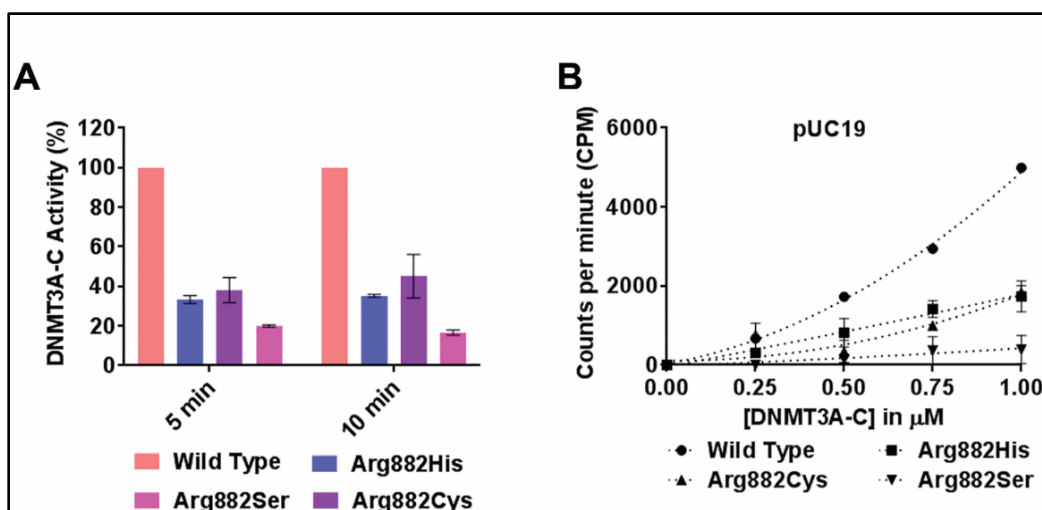


Figure 3.2 Relative activity and kinetics mechanism of DNMT3A-C WT and Arg882 variants. (A) Methylation activity of 1 μ M DNMT3A-C WT and Arg882 variants was measured for 5 and 10 minutes using 3 [H] labelled S-Adenosylmethionine (AdoMet). The transfer of radiolabeled –CH₃ group to DNA was measured as counts per minute (CPM) using the MicroBeta scintillation counter. (B) Methylation activity of DNMT3A-C WT and Arg882 variants was measured for 10 minutes using 100 ngs of pUC19 plasmid as a substrate, at concentrations of enzymes varying from 0.25 to 1 μ M. The enzymes were pre-incubated with DNA for 10 minutes at room temperature and the reaction was initiated by addition of AdoMet. Each data point is an average and standard error of the mean ($n \geq 3$ independent experiments). The data shows reduced activity and loss of cooperativity for all the variant enzymes.

3.5.2 Loss of cooperativity modulates flanking sequences preference of DNMT3A

Recent co-crystal structure of DNMT3A with DNA shows that the Arg882 residue interacts with the phosphate backbone of the nucleotide at N+3 position (N=CG dinucleotide) (123). It was also shown that Arg882His has a preference for G at the N+3 position compared to the WT enzyme by *in vitro* methylation assays where a 509-bp DNA fragment containing 56 CpG sites (SUHW1 promoter region) was used to compute relative preference of nucleotides at positions flanking the central CpG (144). Given that the Arg882 residue is also necessary for the cooperative kinetics mechanism of DNMT3A, we used the same experimental system and method of calculation to test if this preference was affected by cooperative kinetics mechanism of DNMT3A (Table 3.1). We performed *in vitro* methylation of the SUHW1 DNA substrate using the WT and the Arg882His variant enzymes (Figure 3.3A). DNA methylation was quantified by bisulfite conversion and high throughput sequencing. The preference of sites is represented by a relative change greater than 1. Our data confirmed the previously reported G preference at N+3 for Arg882His variant compared to the WT enzyme (Figure 3.4A). We next tested the relationship between loss of cooperative mechanism and flanking sequence preference. We have previously shown that at a low concentration (0.25 μ M), DNMT3A-C does not multimerize on the SUHW1 promoter DNA substrate and therefore cannot methylate multiple CpGs using cooperative mechanism, which is observed at a higher concentration (1 μ M) of the enzyme (201). DNA methylation assays were performed using 0.25 μ M and 1 μ M enzyme, and the flanking sequence preferences were compared. Strikingly, 6 out of 7 sites preferred by DNMT3A-C at 0.25 μ M have G at N+3 position, which matches the flanking sequence preference of the DNMT3A-C Arg882His variant (Figure 3.4B, Figure 3.3B). Our data therefore suggest the role of cooperativity in modulating the interaction of the Arg882 residue with DNA.

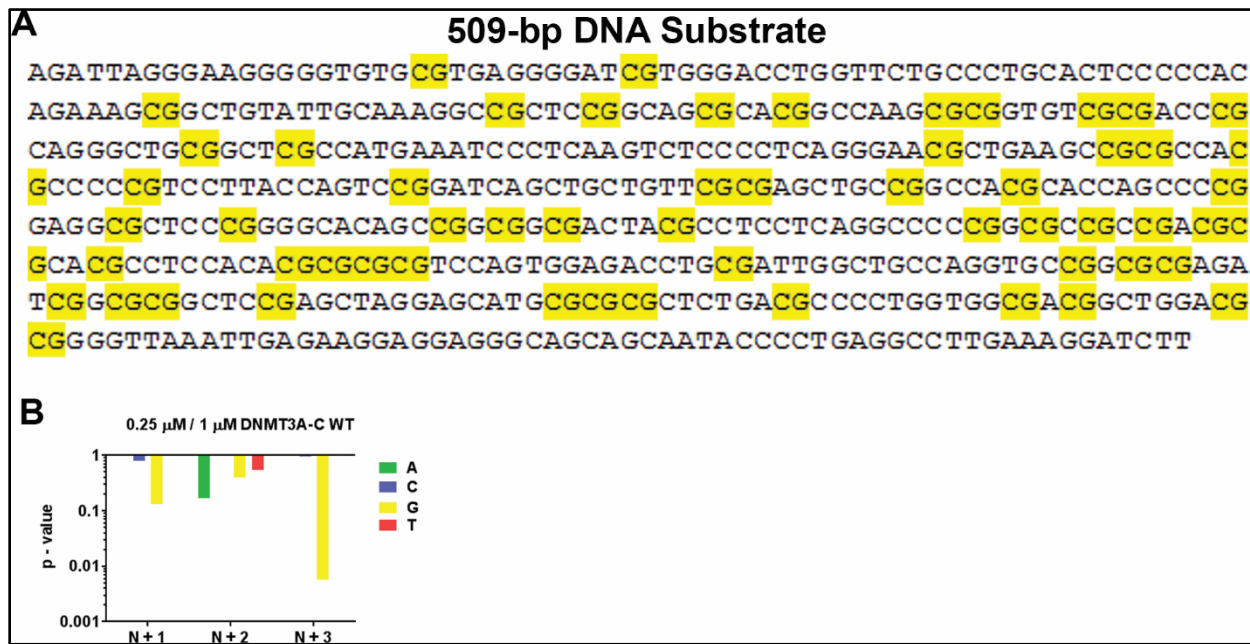


Figure 3.3 Bisulfite sequencing of the 509-bp substrate. (A) Sequence of the 509-bp (*SUHWI* promoter region) DNA substrate used for DNA methylation assays. Underlined sequence indicates primer-binding site for bisulfite sequencing. (B) p-values of enrichment for each base at N+1/2/3, corresponding with Figure 3.4B.

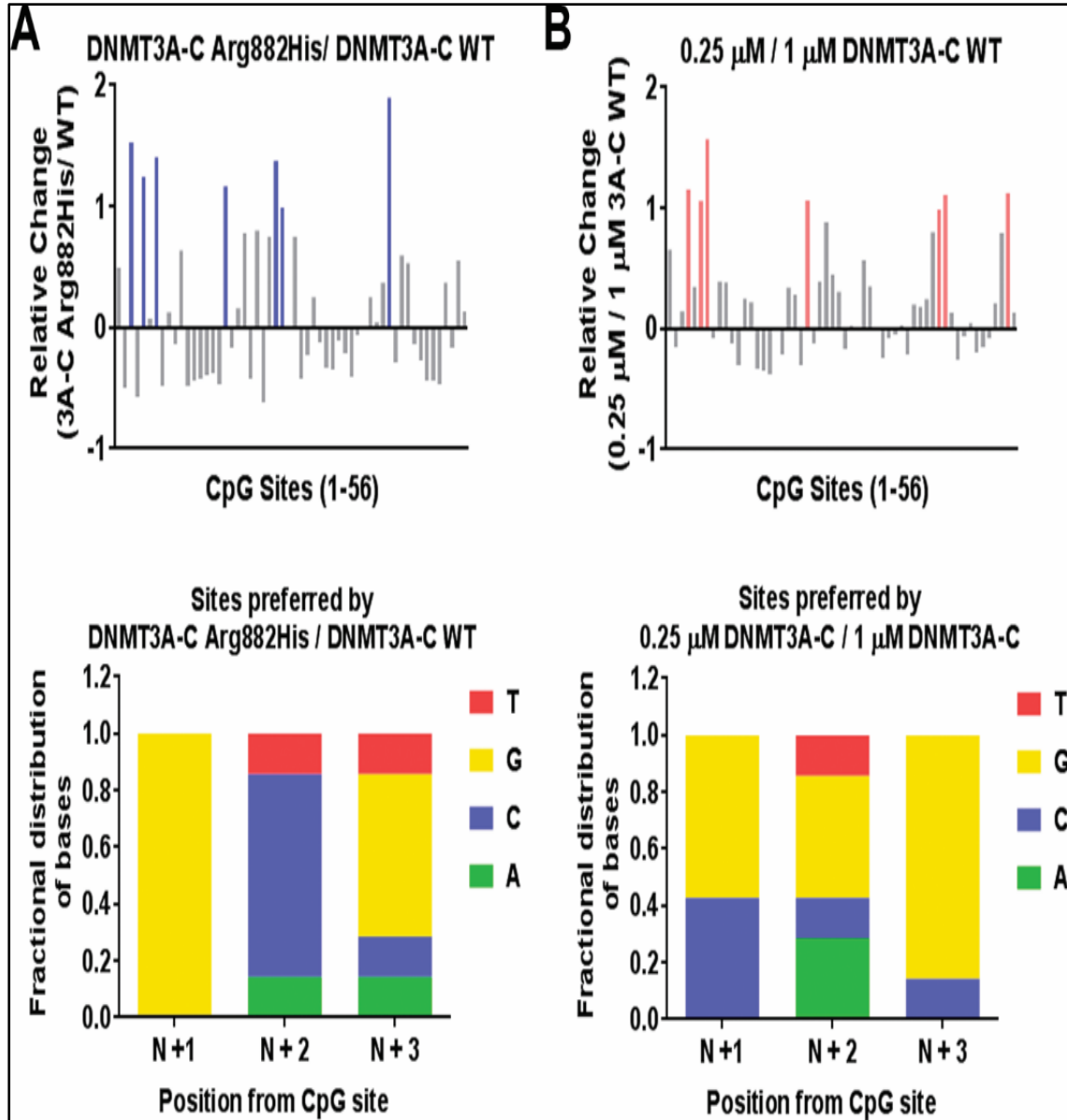


Figure 3.4 Effect of cooperativity on flanking sequence preference by DNMT3A-C. (A,B) DNA methylation of the 56 CpG sites in the *SUHWI* promoter region (509-bp) substrate was analyzed using bisulfite sequencing. Methylation reaction was carried out using 1 μ M DNMT3A-C WT, and 1 μ M DNMT3A-C Arg882His in (A), and 0.25 μ M and 1 μ M DNMT3A-C WT in (B). *Top panel* shows relative preference calculated for each CpG site (1-56) by the DNMT3A-C Arg882His compared to DNMT3A-C WT in (A), and 0.25 μ M compared to 1 μ M of DNMT3A-C WT in (B). Relative preference of 1 is equal to a 2-fold change, and is represented by blue and light pink bars, respectively. *Bottom panel* shows the fractional distribution of nucleotides at the preferred sites. Each bar represents nucleotides at positions N+1/2/3 respectively from the CpG site. Data presented are an average $n=3$ independent experiments. The data show DNMT3A-C WT at low concentrations have flanking sequence preference similar to DNMT3A-C Arg882His variant.

3.5.3 Flanking sequence of DNMT3B has similarity to that of DNMT3A Arg882His

DNMT3B is a homolog of DNMT3A that is frequently overexpressed in tumors, including AML (192, 198). Similar to DNMT3A-C Arg882His variant, DNMT3B-C functions as a non-cooperative enzyme (201). Therefore, we asked if DNMT3B-C and the DNMT3A-C Arg882His variant have a similar flanking sequence preference. Although the nucleotide preference of DNMT3B is reported for N+1 position (G), the extended flanking sequence preference has not been thoroughly evaluated (122). Using recombinant DNMT3B-C WT and the Arg829His variant (homologous to DNMT3A Arg882His), we performed *in vitro* methylation assays using 4 different DNA substrate ranging in size from 500-1000-bp. As DNMT3B and DNMT3A Arg882His have lower activity compared to DNMT3A WT, we used catalytically dead DNMT3B Glu703Ala variant to detect the background of methylation analysis by bisulfite sequencing. The methylation level was significantly higher for all the enzymes compared to the control. The SEM for the percent methylation at all CpG sites was in the same range for all the active enzymes (Figure 3.5A, B). Therefore, using the average percent methylation at each CpG site we computed the relative preference for each CpG site by DNMT3A-C WT, DNMT3A-C Arg882His, and DNMT3B-C WT. The preferred sites of DNMT3A-C Arg882His compared to DNMT3A-C WT showed pronounced increase at 17 sites, of which 9 had G at the N+3 position and 11 had G at N+1 position (Figure 3.6A - C). Simultaneously, there were 21 sites preferred by DNMT3B-C WT compared to DNMT3A-C WT, of which 10 had G at the N+3 position and 12 had G at N+1 position (Figure 3.6D - F). Therefore, we discovered a strong overlap in the fractional distribution of bases preferred by DNMT3B-C with those preferred by DNMT3A-C Arg882His over the DNMT3A-C enzyme. This was confirmed by a direct comparison of the preferred sites of DNMT3B-C WT and the DNMT3A-C Arg882His enzyme, which showed no significant preference for a specific nucleotide at examined CpG flanking positions at the 19 preferred sites (Figure 3.6G - I). These data suggest that the Arg882His substitution alters the specificity of DNMT3A to be like that of DNMT3B. Furthermore, a comparison between DNMT3B-C Arg829His variant and DNMT3B-C WT showed a very similar flanking sequence preference between the two enzymes, suggesting little or no effect of the Arg829 mutation on DNMT3B-C (Figure 3.5C, D). DNMT3B methylates DNA using a processive kinetics mechanism (201), therefore, we assayed the kinetics mechanism of the DNMT3A-C Arg882His variant. Our data confirmed that the Arg882His variant, like the

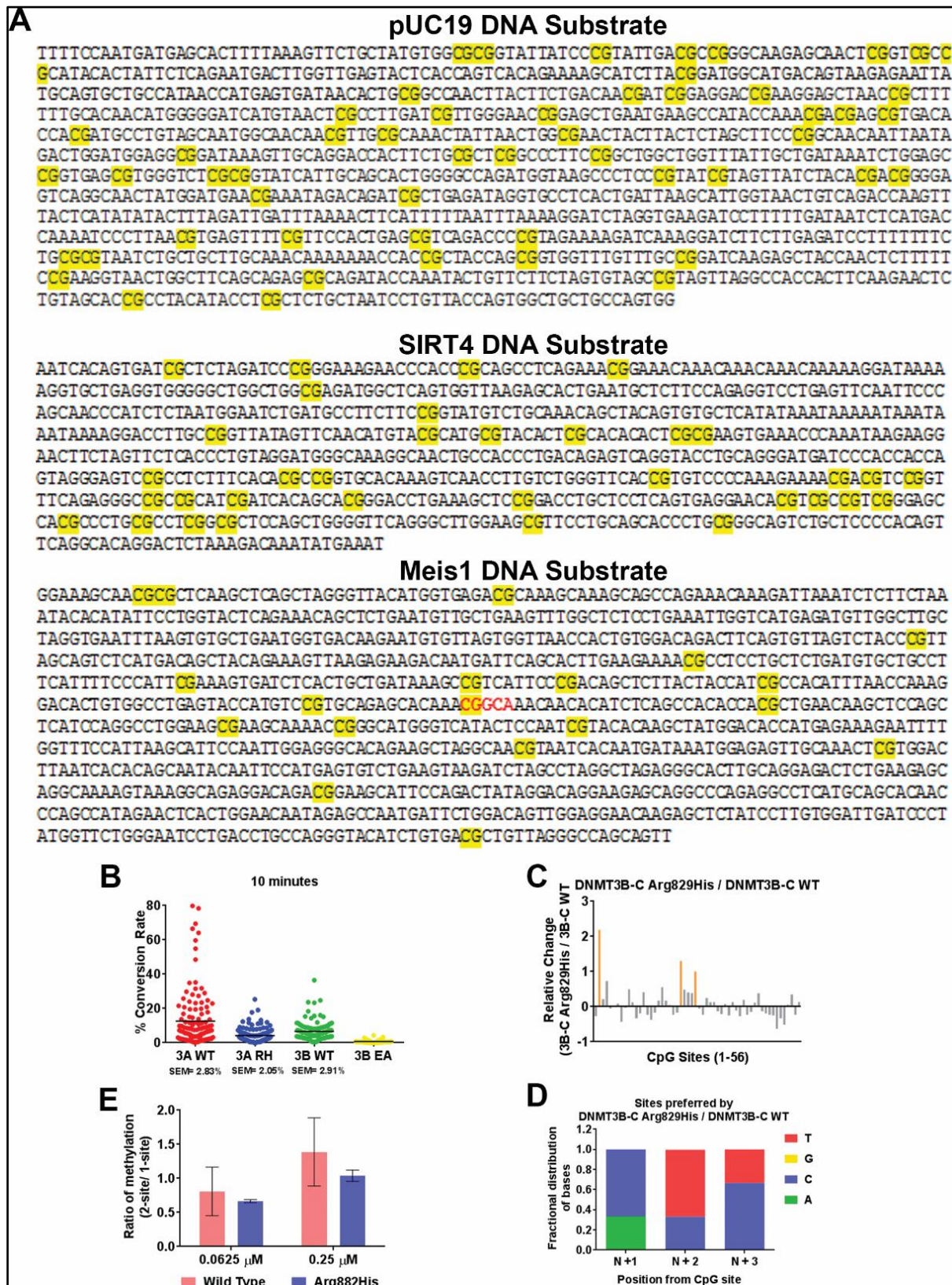
WT enzyme, is non-processive, suggesting a specific effect of Arg882His mutation on the flanking sequence preference of DNMT3A (Figure 3.5E).

Figure 3.5 Methyltransferase activity of wild type and variant DNMT3 enzymes on multiple substrates. (A) Sequences of the DNA substrates used for *in vitro* DNA methylation assays. In the *Meis1* enhancer sequence, the CpG site in red is commonly preferred by both DNMT3A-C Arg882His and DNMT3B-C WT compared to DNMT3A-C WT from Figures 6B, C. (B) Scatter plot of the average methylation of all four substrates by DNMT3A-C WT (3A WT), DNMT3A-C Arg882His (3A RH), DNMT3B-C WT (3B WT), and catalytically inactive DNMT3B-C Glu703Ala (3B EA). Each data point is calculated from the 50,000-80,000 reads and the average is represented by the black line. The average methylation by active enzymes ranged between 4.13 – 12.64%, and the average methylation background recorded using 3B EA mutant is 0.73%.

SEM was calculated for percent methylation for at least three replicates. (C) Methylation reaction was carried out using 1 μ M DNMT3B-C WT, and 1 μ M DNMT3B-C Arg829His using the *SUHWI* DNA substrate for 10 minutes. DNA methylation of the 56 CpG was analyzed using bisulfite sequencing. Relative preference calculated for each CpG site (1-56) by the DNMT3B-C Arg829His compared to DNMT3B-C WT is represented by orange bars. (D) Fractional

distribution of nucleotides at the preferred sites. Each bar represents nucleotides at positions N+1/2/3 respectively from the CpG site. (E) Methylation of a 1-site and 2-site substrate at two different enzyme concentrations. All reactions were carried out using 0.25 μ M of DNA substrate and a 1:1 ratio of radioactively labelled and unlabeled AdoMet. Samples were incubated for 60 minutes and the incorporation of radioactivity was measured as counts per minute using scintillation counter. The ratio of incorporated methylation of the 2-site substrate compared to the 1-site substrate was plotted at different concentrations for each enzyme. Data presented are an average $n=3$ independent experiments.

Figure 3.5 continued



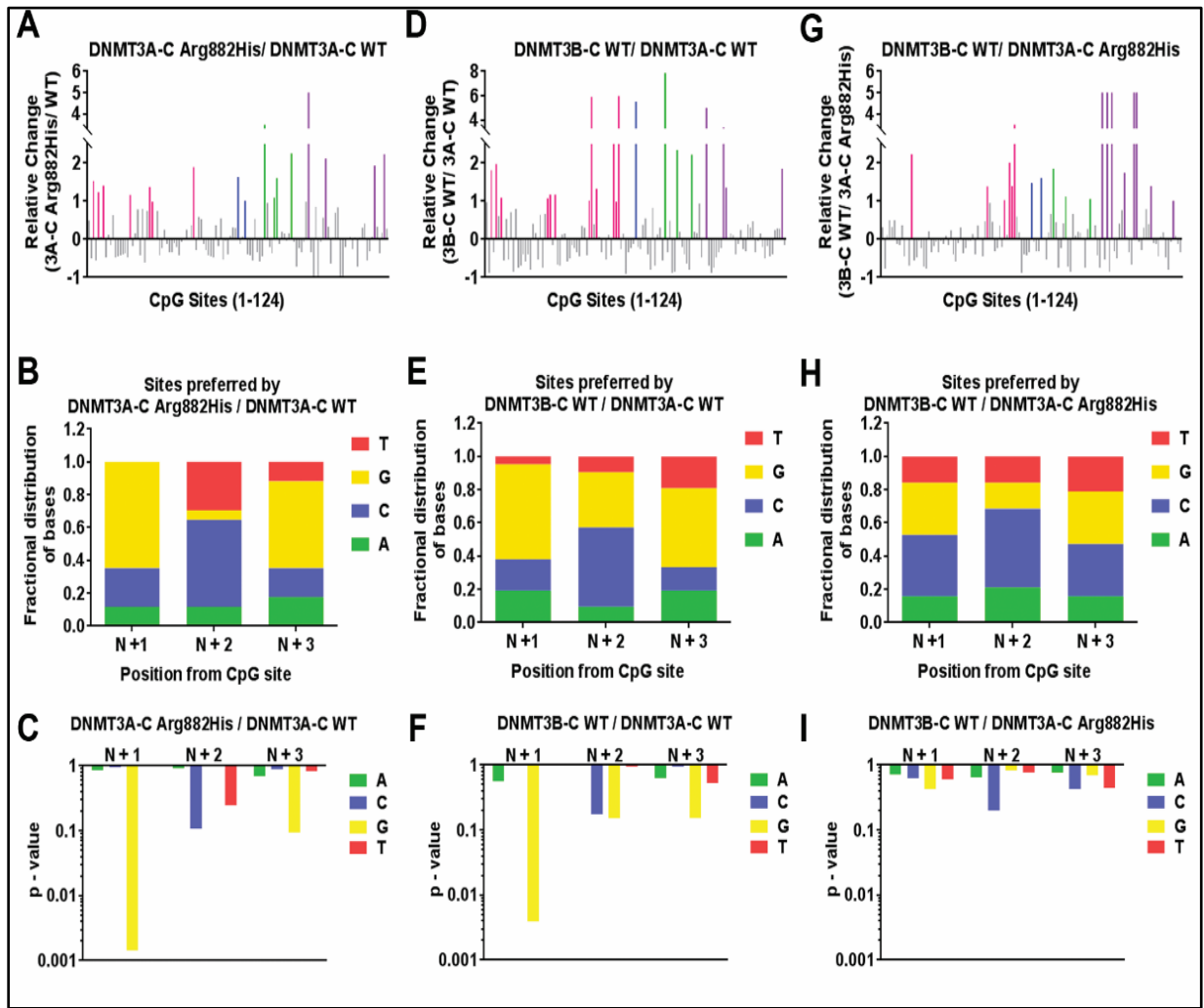


Figure 3.6 Comparative analysis of the flanking sequence preferences of DNMT3A-C WT, DNMT3A-C Arg882His and DNMT3B-C WT. DNA methylation of 124 CpG sites in 4 substrates, *SUHW1* promoter region (509-bp), *Meis1* (1000-bp) and *Sirt4* (721-bp) enhancer regions, and pUC19 fragment (1092-bp), was analyzed using bisulfite sequencing. Methylation reaction was carried out using 1 μ M enzyme for 10 minutes and relative preference was calculated for each CpG site represented on X axis from 1-124. A 2 fold or more change in substrate preference is represented by bars (colored) greater than 1. Preferred CpG sites are shown in pink for *SUHW1*, blue for *Meis1*, green for *Sirt4*, and purple for the pUC19 substrate. DNMT3A-C Arg882His compared to DNMT3A-C WT in (A), DNMT3B-C WT compared to DNMT3A-C WT in (D), and DNMT3B-C WT compared to DNMT3A-C Arg882His in (G). The distribution of nucleotides at positions N+1/2/3 respectively was plotted from 17 preferred CpG sites by DNMT3A-C Arg882His in (B), 21 preferred sites by DNMT3B-C WT in (E), and the 19 sites preferred by DNMT3B-C WT in (H). p-values of enrichment for each base at N+1/2/3, corresponding to B, E, and H, are shown in C, F, and I respectively. Data presented are an average $n=3$ independent experiments. These data show that substrate preference of DNMT3A-C Arg882His variant is similar to that of DNMT3B-C WT.

3.5.4 The DNMT3A Arg882His variant acquires DNMT3B-like substrate preference

The methylation assays to determine the flanking sequence preference described above were performed for 10 minutes, which represents initial enzyme kinetics. To evaluate the substrate preference under multiple turnover conditions, the methylation assays were carried out for 30 and 60 minutes, allowing the enzyme kinetics to enter the steady state. Compared to the flanking sequence preference at 10 minutes, DNMT3A-C Arg882His and DNMT3B-C both showed a slight decrease in relative preference for G at the N+3 as well as at N+1 position (Figure 3.7A-D, Figure 3.8A-C, E-G). This indicates that after methylating the preferred sites during the initial reaction, the enzyme methylates other sites under multiple turnover conditions. A similar comparison of site preference of DNMT3B-C WT over DNMT3A-C Arg882His variant showed no change in the preference during the steady state reaction that is decreased at the 30 and 60 minute time points (Figure 3.7E-F, Figure 3.8D, H). These data confirm the similarity between the DNMT3A-C Arg882His variant and DNMT3B-C WT both in the initial and steady state reaction conditions.

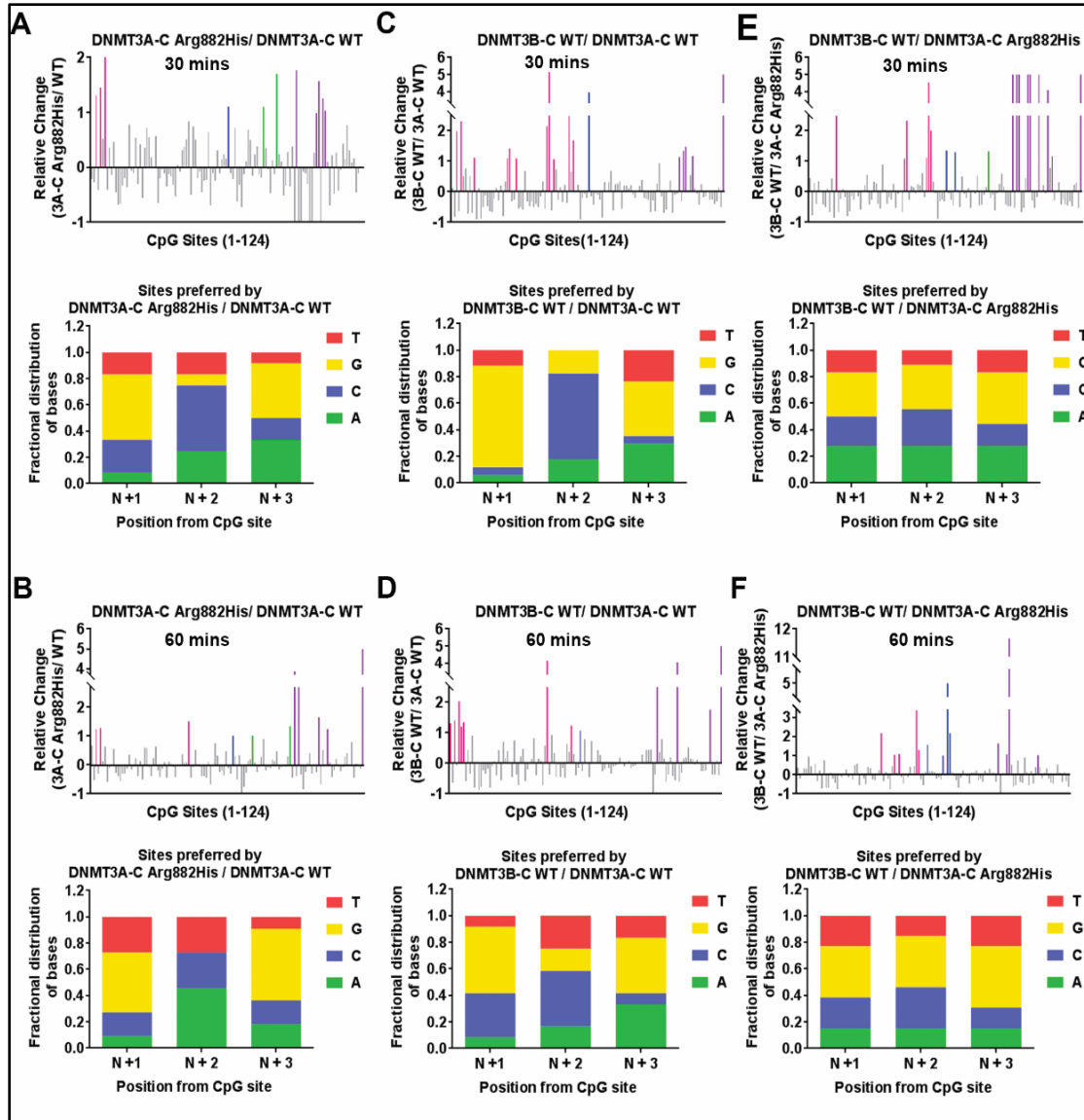
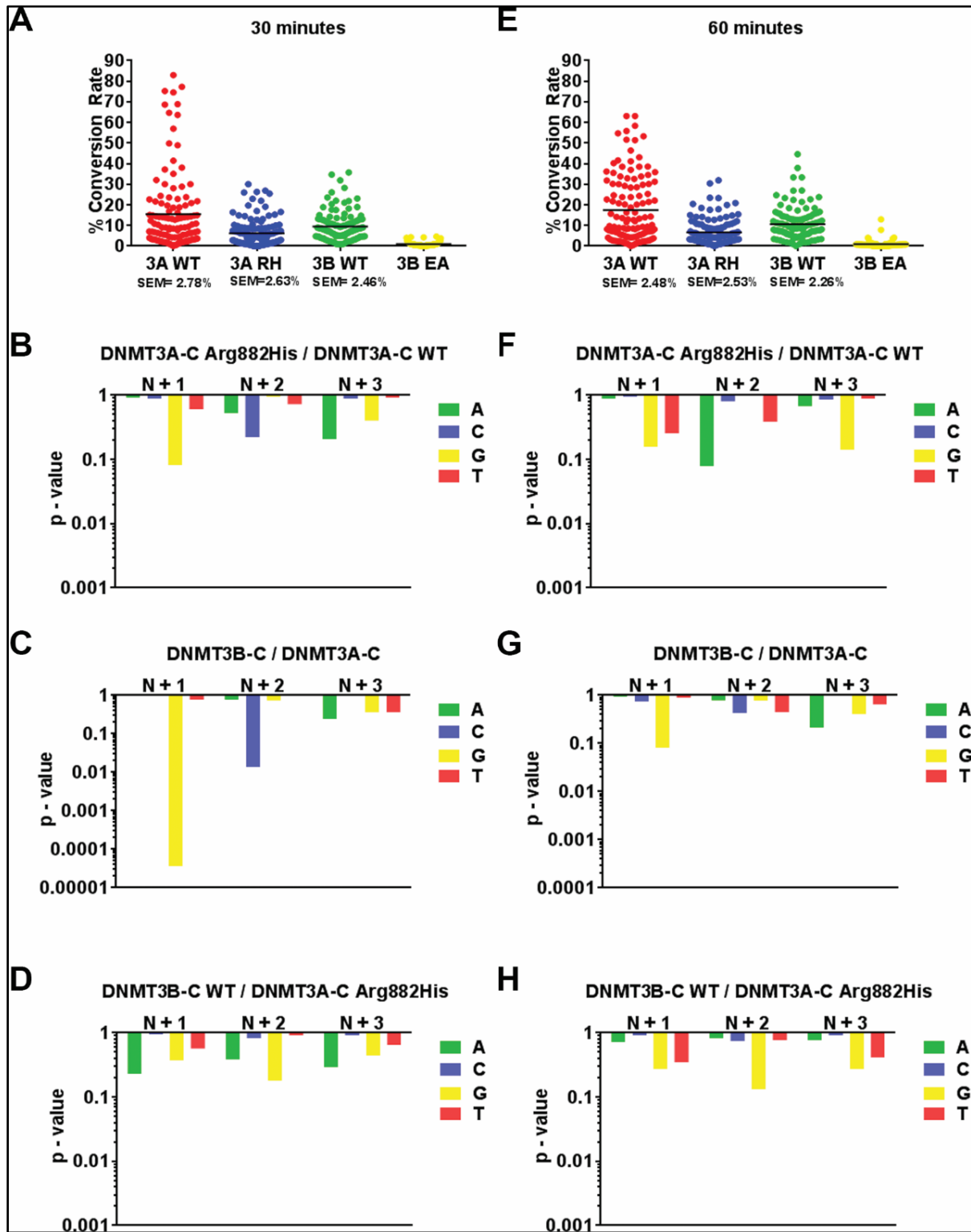


Figure 3.7 Flanking sequence preference at steady state kinetics. Methylation reactions were carried out using 1 μ M enzyme and 4 DNA substrates for 30 minutes and 60 minutes. DNA methylation of the 124 CpG sites was analyzed using bisulfite sequencing. *Top panels* show the relative preference calculated for each CpG site (1-124) at 30 min and 60 minutes by DNMT3A-C WT compared to the DNMT3A-C Arg882His variant in (A, B), DNMT3B-C WT compared to DNMT3A-C WT in (C, D), and DNMT3B-C WT compared to DNMT3A-C Arg882His in (E, F). A 2 fold or more change in substrate preference is represented by bars (colored) greater than 1. Preferred CpG sites are shown in pink for SUWH1, blue for *Meis1*, green for *Sirt4*, and purple for the pUC19 substrate. *Bottom panels* show the fractional distribution of nucleotides at the preferred sites. Each bar represents nucleotides at positions N+1/2/3 respectively from the CpG site. Data presented are an average $n=3$ independent experiments. The data show that the similarity in flanking sequence preference between DNMT3A-C Arg882His and DNMT3B-C is maintained during steady state kinetics.

Figure 3.8 Enrichment for each relative site preference. (A, E) Scatter plot of the methylation of the 509-bp substrate by each enzyme represented in Figure 4 after 30 minutes (A) or 60 minutes (E). Each dot represents the % methylation of a CpG site, calculated from the 50,000-80,000 reads the average is represented by the black line which ranges from 6.3-15.4% for 30 minutes and 6.7-17.5% for 60 minutes reactions. The background methylation using 3.6B EA was 0.93% for 30 minutes and 0.91% for 60 minute time points. (B - D) p-values of enrichment for each base at N+1/2/3 at 30 minutes, corresponding to Figure 3.7A, C, and E, respectively. (F - H) p-values of enrichment for each base at N+1/2/3 at 60 minutes, corresponding to Figure 3.7B, D, and F, respectively.

Figure 3.8 continued



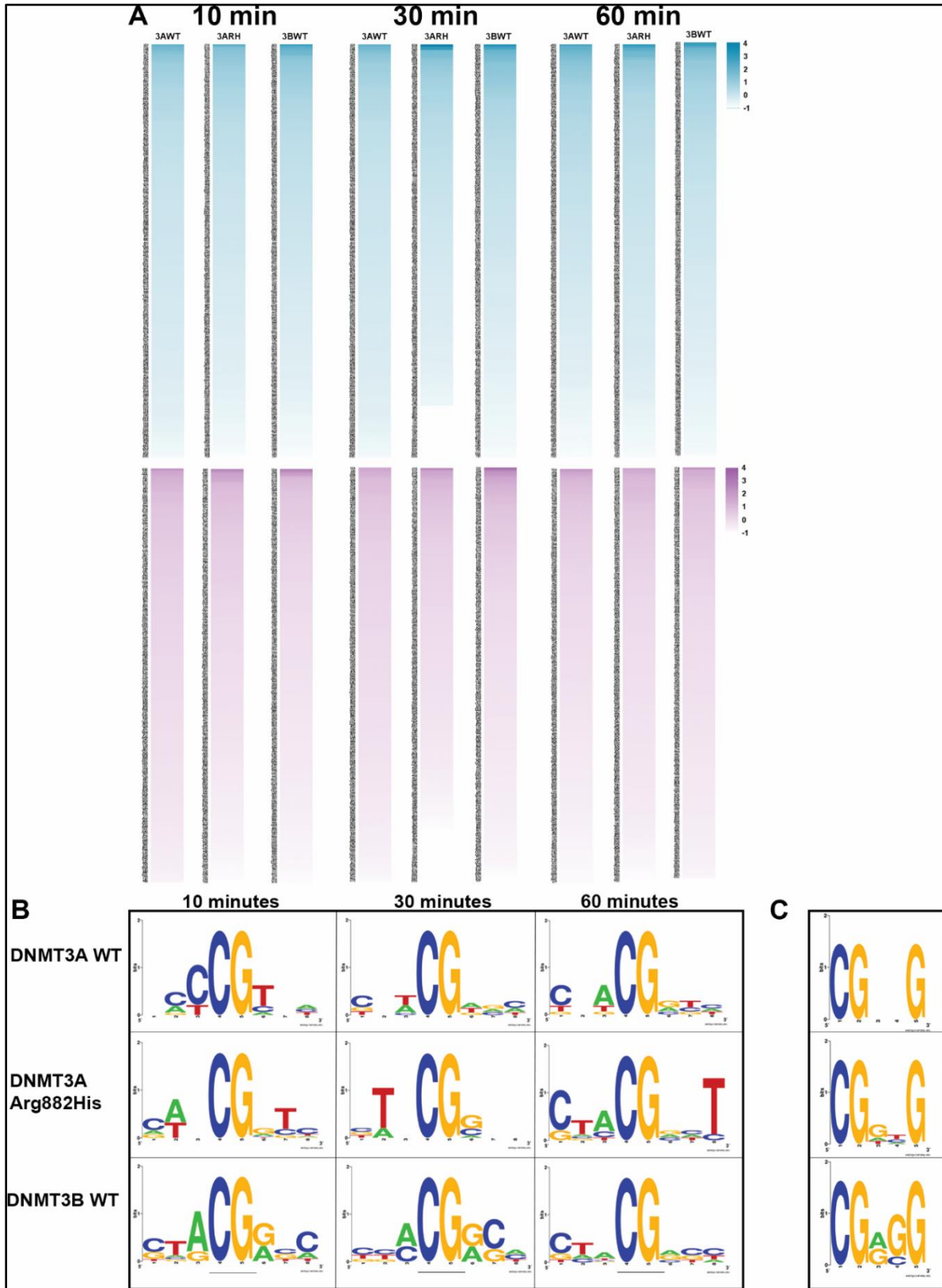
3.5.5 Temporal change in flanking sequence preference by WT and variant DNMT3 enzymes

We extended the analysis of the flanking sequence preference to determine the preferred trinucleotides at the N+1/2/3 and N-1/2/3 positions by DNMT3A-C WT, DNMT3A-C Arg882His, and DNMT3B-C WT. The bisulfite sequencing data of four DNA substrates was used for this analysis. Based on the occurrence of different sites, the methylation at a site was calculated as a ratio of observed to expected fractional methylation for the specific substrate (See Methods, Table 3.2).

The top 10 preferred sites were used in WebLogo application (203) to determine the consensus sequence logo flanking the CpG site for each enzyme. The analysis was performed for methylation data collected at the 10, 30, and 60-minute time points to monitor the temporal order of site preference by these enzymes (Figure 3.9A, B). As expected, the data again show a dramatic difference in flanking sequence preference between DNMT3A-C and DNMT3B-C. Whereas DNMT3A-C WT shows a preference for C or T at the N+1/2/3 positions, DNMT3B-C WT prefers G and A. The preferred sequence of the DNMT3A-C Arg882His variant strikingly resembles that of DNMT3B-C, particularly at N+1, where it loses preference for T and gains a preference for G, and at N+2, where it gains a preference for C. Consistent with previous observations (122), the sites with T at the N+1 position are most preferred by DNMT3A-C WT which is evident by its preference at 10 minutes (Figure 3.9B). DNMT3A-C WT enzyme has the least preference for sites with G at the N+1 position, evident by its appearance at 60 minutes. Comparatively, the preference pattern at the flanking positions is similar for DNMT3B-C and DNMT3A-C Arg882His, but different from that of DNMT3A-C WT. Both DNMT3B-C and DNMT3A-C Arg882His enzymes prefer G at the N+1 position and C at N+2, and disfavor sites with T at the N+1 position. The preference of all three enzymes for nucleotides at position N-1/2/3 showed a weak or no preference (Figure 3.9B). This is in agreement with crystal structure data showing fewer interactions between DNMT3A and the nucleotides upstream of CpG site (123).

Figure 3.9 Trinucleotide sequence flanking CpG preferred by WT and variant DNMT3 enzymes. (A) Heat-map showing the preference of different trinucleotide sets by DNMT3A-C WT (3AWT), DNMT3A-C Arg882His (3ARH), DNMT3B-C WT (3BWT) at 10, 30, and 60 minutes. Sequences with values greater or equal to 1 are considered as preferred. Upper panels in blue represent the flanking sequence preference at the N +1/2/3 positions. Lower panels in purple represent the flanking sequence preference at the N-1/2/3 positions. (B) Consensus sequence generated by WebLogo from top ten preferred sequences by each enzyme as methylation reaction proceeds from 10 to 60 minutes. (C) Consensus flanking sequence for methylated sites with a G at the N+3 position. The data show that for the DNMT3A-C Arg882His variant the nucleotide preference at N+1 and N+3 matches that of DNMT3B-C. Similarly, at methylated sites with G at the N+3 position, inner flanking sequence are similar between DNMT3A-C Arg882His and DNMT3B-C, whereas it is different for DNMT3A-C WT enzyme.

Figure 3.9 continued



We next selected CpG sites that had G at the N+3 position and generated the preferred flanking sequence logo using the WebLogo application (203). A comparison between the consensus sequences again show a striking similarity between DNMT3A-C Arg882His and DNMT3B-C WT enzymes. Whereas G at the N+1 position is the most preferred nucleotide by these enzymes, it is the least preferred by the DNMT3A-C WT enzyme. The consensus flanking sequence of DNMT3A-C WT has a T at the N+1 position. These observations uncover the importance of the nucleotide at the N+1 position, which can affect the interaction of DNMT3 enzymes with DNA (Figure 3.9C). We therefore tested if T at the N+1 position affects the preference for G at the N+3 position by DNMT3B-C and DNMT3A-C Arg882His enzymes. Methylation assays were performed using 30-bp oligonucleotide substrates, which contain a central CpG site, and varying combinations of nucleotides at N+2/3 or N-2/3 positions. The positions N+1 and N-1 were held constant with T and A, respectively (Table 3.3). Radiolabeled AdoMet was used in enzyme assays and total methylation at 10 minutes was measured. Our data show that for these substrates, the preference of DNMT3B-C and the DNMT3A-C Arg882 variants for G at the N+3 position is lost (Figure 3.10A-E). DNMT3A-C and its variants show rather strong preference for sites with A at the N+2 position, whereas DNMT3B-C shows a weak preference for sites with A or C occupying the N+2 position (Figure 3.10E). Interestingly, DNMT3B-C Arg829His shows reduced activity when compared to the WT enzyme, indicating an adverse effect of T at N+1 position on its activity (Figure 3.10F). These data confirm our previous observations that the interaction of DNMT3A-C Arg882His at the N+3 position is strongly influenced by the nucleotide at the N+1 position.

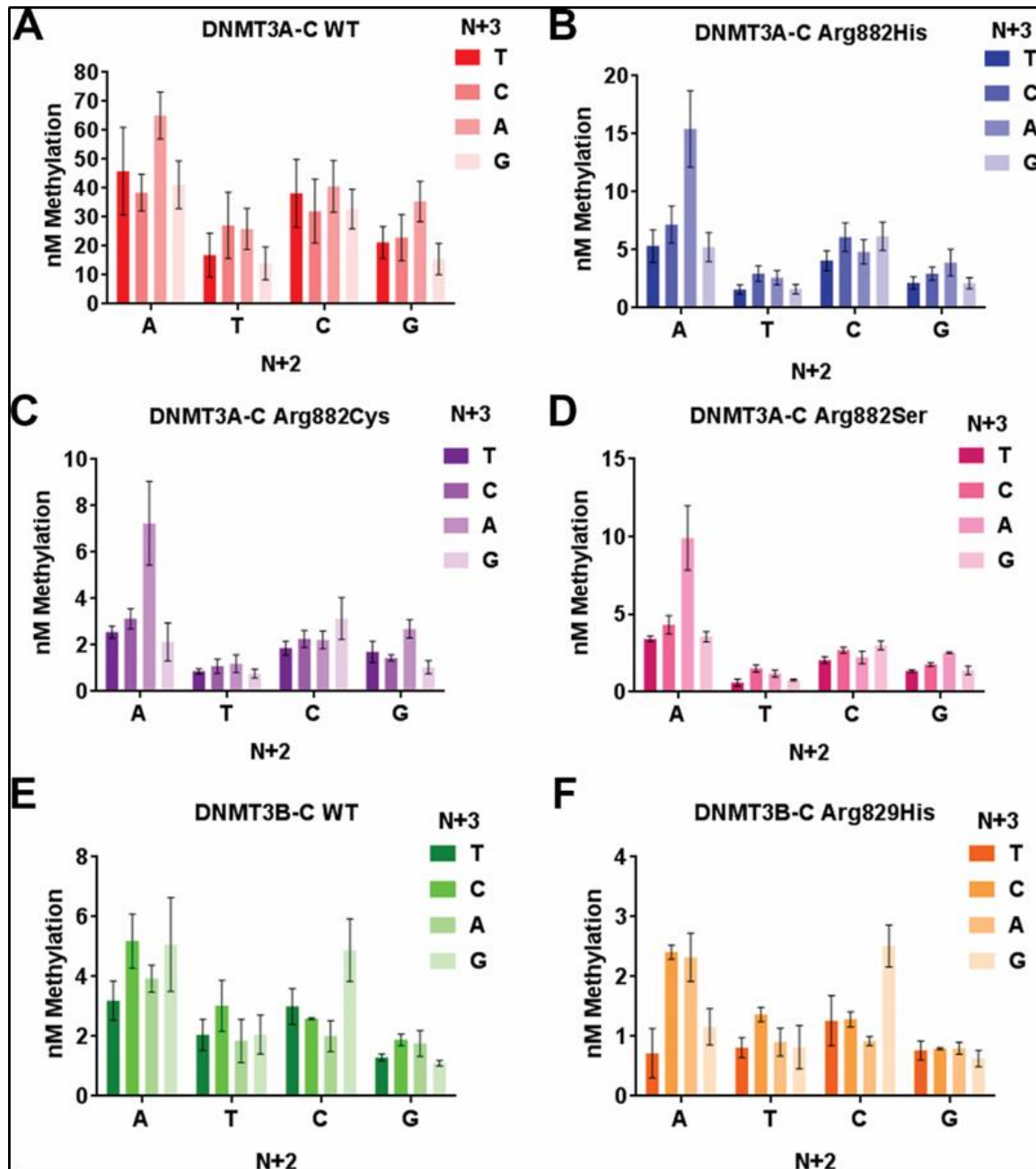


Figure 3.10 Methylation activity using short DNA substrates with varying N +2 and N +3 nucleotide sequences with T at the N +1 position. Each enzyme at 1 μ M concentration was incubated with 250 nM 30-bp substrate and a mixture of labeled and unlabeled AdoMet for 10 minutes. Total methylation activity was plotted for each dinucleotide set grouped by second nucleotide. Error bars represent SEM of three independent experiments with two different enzyme purifications.

3.5.6 DNMT3A-C Arg882His and DNMT3B-C preferably methylate the same CpG site in the *Meis1* enhancer

We next tested if the change in sequence preference of DNMT3A-C Arg882His could affect methylation of the regions that are known to be spuriously hypomethylated in AML patients (204). The *Meis1* gene is expressed during development and, by promoting the self-renewal of progenitor-like cells, it regulates leukemogenesis and hematopoiesis (205, 206). The enhancer of *Meis1* is methylated by DNMT3A during normal hematopoietic stem cell (HSC) differentiation, whereas in AML patients expressing the DNMT3A Arg882His variant, this region is largely hypomethylated (204, 207). A 1-kb region of the *Meis1* enhancer was used as a substrate for methylation reactions (Figure 3.5A). The average methylation by DNMT3A-C Arg882His and DNMT3B-C WT was significantly lower than by DNMT3A-C WT (Figure 3.11A). Examination of the flanking sequence preference showed that DNMT3B methylated with a strong bias for one site (#7), which had G at N+1 and A at the N+3 position (Figure 3.11B, Figure 3.5A). Interestingly, this site was one of the two sites preferred by DNMT3A-C Arg882His (Figure 3.11C). A comparison between DNMT3B-C and DNMT3A-C Arg882His confirmed that DNMT3B-C has a higher preference for this site (Figure 3.11D). Although there are 3 sites in this substrate with G at N+3 position, these sites have either T, C, or A at the N+1 position, which are weakly preferred by DNMT3B-C as well as by the DNMT3A-C Arg882His variant. The data again confirms the effect of the nucleotide at the N+1 position on the flanking sequence preference of DNMT3 enzymes and the effect of DNMT3A-C Arg882His on target site specificity.

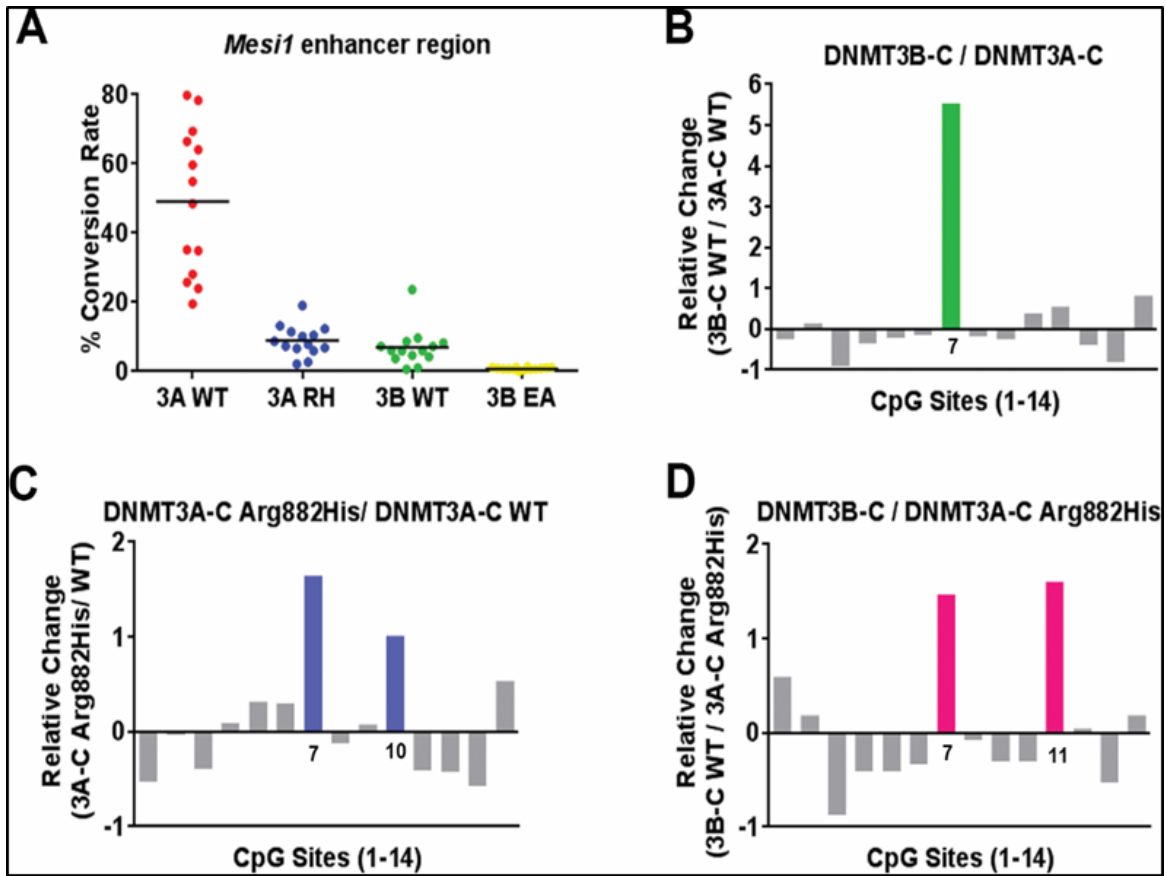


Figure 3.11 Relative activity and site preference of DNMT3 WT and mutant enzymes on *Meis1* enhancer substrate. A 1-kb *Meis1* enhancer region was used as substrate for *in vitro* methylation reactions by DNMT3A-C WT, the Arg882His variant, and DNMT3B-C WT. (A) Methylation activity of DNMT3A-C WT, DNMT3A-C Arg882His, DNMT3B-C WT, and catalytically inactive DNMT3B-C Glu703Ala on the *Meis1* substrate at 10 minutes, analyzed by bisulfite sequencing. Relative preference for each of the 14 CpG sites was calculated for DNMT3B-C compared to DNMT3A-C in (B), DNMT3A-C Arg882His compared to DNMT3A-C WT in (C), and DNMT3B-C WT compared to DNMT3A-C Arg882His in (D). Relative preference of 1 which is equal to a 2-fold change, is represented by green, blue and pink bars, respectively. Data presented are an average $n=3$ independent experiments. DNMT3A-C Arg882His and DNMT3B-C prefer to methylate same site in this substrate, however DNMT3B-C shows a very strong preference compared to the DNMT3A-C Arg882His variant.

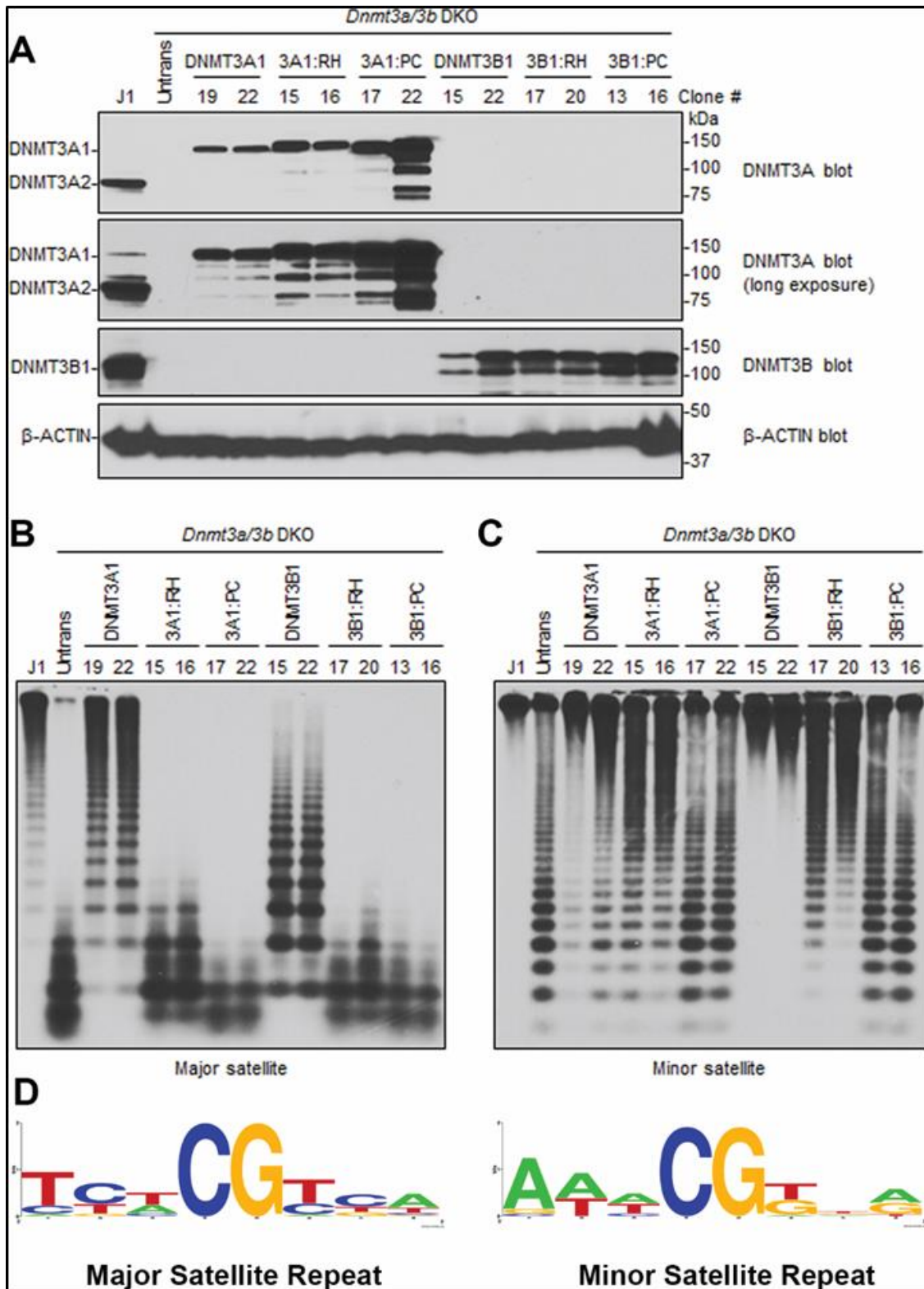
3.5.7 Mouse DNMT3A Arg878His retains its activity for minor satellite DNA in mESCs

While DNMT3A and DNMT3B redundantly methylate many genomic regions in cells, they also have preferred and specific targets (86, 138, 190). For example, in murine cells, DNMT3A preferentially methylates the major satellite repeats in pericentric regions, whereas DNMT3B preferentially methylates the minor satellite repeats in centromeric regions (86, 190).

Based on our observations, we predicted that DNMT3A Arg882His would prefer DNMT3B-specific targets. To test the idea, we carried out rescue experiments in late-passage DNMT3A/3B DKO mESCs, which show severe loss of global DNA methylation, including at the major and minor satellite repeats (190). mESCs express two major DNMT3A isoforms, DNMT3A1 (full length) and DNMT3A2 (a shorter form that lacks the N terminus of DNMT3A1), with both showing identical activity (143, 190, 208). We transfected DNMT3A/3B DKO mESCs with plasmid vectors and generated stable lines expressing mouse DNMT3A1 WT, DNMT3A1 Arg878His, DNMT3B1 WT, or DNMT3B1 Arg829His (catalytically inactive DNMT3A1 and DNMT3B1, with their PC motif in the catalytic center being mutated (101), were included as negative controls) (Figure 3.12A). The genomic DNA from these cell lines was harvested, and DNA methylation at the major and minor satellite repeats was analyzed by digestion with methylation-sensitive restriction enzymes followed by Southern blot. Consistent with our previous results (143), the ability of DNMT3A1 to rescue methylation at the major satellite repeats is severely impaired with the Arg878His substitution (Figure 3.12B). However, at the minor satellite repeats, which are largely methylated by DNMT3B, DNMT3A Arg878His rescues DNA methylation comparable to the DNMT3A WT enzyme (Figure 3.12C). These data demonstrate that the DNMT3A AML mutant specifically retains its ability to methylate DNMT3B-preferred target sites, while losing preference for sites methylated by DNMT3A.

Figure 3.12 Rescue of DNA methylation at the major and minor satellite repeats in *DNMT3A/3B* DKO mESCs. (A) *DNMT3A/3B* DKO mESCs were transfected with plasmids encoding mouse DNMT3A1 WT, DNMT3A1 Arg878His (3A1:RH), DNMT3A Pro705Val/Cys706Asp (3A1:PC), DNMT3B1 WT, DNMT3B1 Arg829His (3B1:RH), or DNMT3B1 Pro656Gly/Cys657Thr (3B1:PC), and stable clones were derived. Total cell lysates were used to analyze the expression of DNMT3A or DNMT3B proteins by western blot with anti-DNMT3A, anti-DNMT3B, and anti- β -Actin antibodies. A long exposure of the DNMT3A blot is included to show endogenous DNMT3A1 in WT (J1) mESCs. Note that stable clones showing similar expression levels to those of endogenous DNMT3A or DNMT3B were used for the experiments. (B, C) DNA methylation was analyzed by Southern blot. Genomic DNA was digested with MaeII (major satellite repeats) or HpaII (minor satellite repeats), and probed for the major (B) or minor (C) satellite repeats. J1 (WT) and untransfected DKO mESCs were used as controls. The numbers on the top indicate clone#. Complete digestion due to low or no DNA methylation is indicated by low molecular weight bands as seen in untransfected DKOs, and high molecular weight bands as seen in J1 indicate high DNA methylation and protection from digestion. Comparing the activity of DNMT3A1 clones 19/22 with 3A1:RH clones 15/16 at major and minor satellite repeats shows that 15/16 methylate minor repeats similar to 19/22 whereas at major satellite repeats the activity of 15/16 is severely impaired. (D) Consensus sequence of the nucleotides flanking the CpG site in either the major or minor satellite repeats, created using WebLogo shows high prevalence G at N+1 and N+3 positions at minor repeats.

Figure 3.12 continued



To test if the preference of DNMT3A and DNMT3B was driven by a potential sequence bias in the major or minor satellite repeats, we computed the consensus sequence logo for +3 and -3 flanking nucleotides around the CpG sites using the WebLogo application (138) (Figure 3.12D). The analysis shows that the major satellite repeats are enriched with CpG sites carrying T at N+1 and A at N+3 and are depleted in CpG sites carrying G at the N+1 and N+3 positions. However, the minor satellite repeats have a high percentage of sites with G at the N+1 and N+3 positions, which are highly preferred by DNMT3B as well as DNMT3A Arg882His. These data confirm that DNMT3A Arg878His acquires catalytic properties analogous to DNMT3B, which may allow it to target DNMT3B-specific sites in somatic cells and contribute to cancer development.

Our data also show that unlike the WT enzyme, the DNMT3B Arg829His variant was unable to rescue methylation of the major satellite repeats and only partially rescued methylation of the minor satellite repeats (Figure 3.12B, C). This is explained by the observation that substrates with T at the N+1 position are strongly disfavored by the DNMT3B-C Arg829His variant (Figure 3.10F), and, about half of the CpG sites in both the major and minor satellite repeats have T at the N+1 position.

Taken together, we propose a model in which substrate specificity and kinetics mechanism of DNMT3A and DNMT3B regulate the DNA methylation of various genomic regions. At repetitive elements where CpG content is intermediate to high, DNMT3A WT enzyme acts cooperatively on multiple CpGs where it prefers A and T at N+1 and N+3 positions, respectively. Loss of cooperativity in the DNMT3A WT or DNMT3A Arg882His variant modifies the specificity to G at both N+1 and N+3 analogous to DNMT3B (Figure 3.13). Whereas DNMT3B methylates its targets in a processive manner, DNMT3A WT and Arg882His methylate these sites distributively explaining the lower activity of these enzymes at these sites. Given that the DNA methylation at repetitive elements represents the bulk of DNA methylation in mammalian genomes, these data indicate that besides being targeted by protein-protein interaction and chromatin modifications, DNA methylation by DNMT3A and DNMT3B is largely guided by their intrinsic sequence preference.

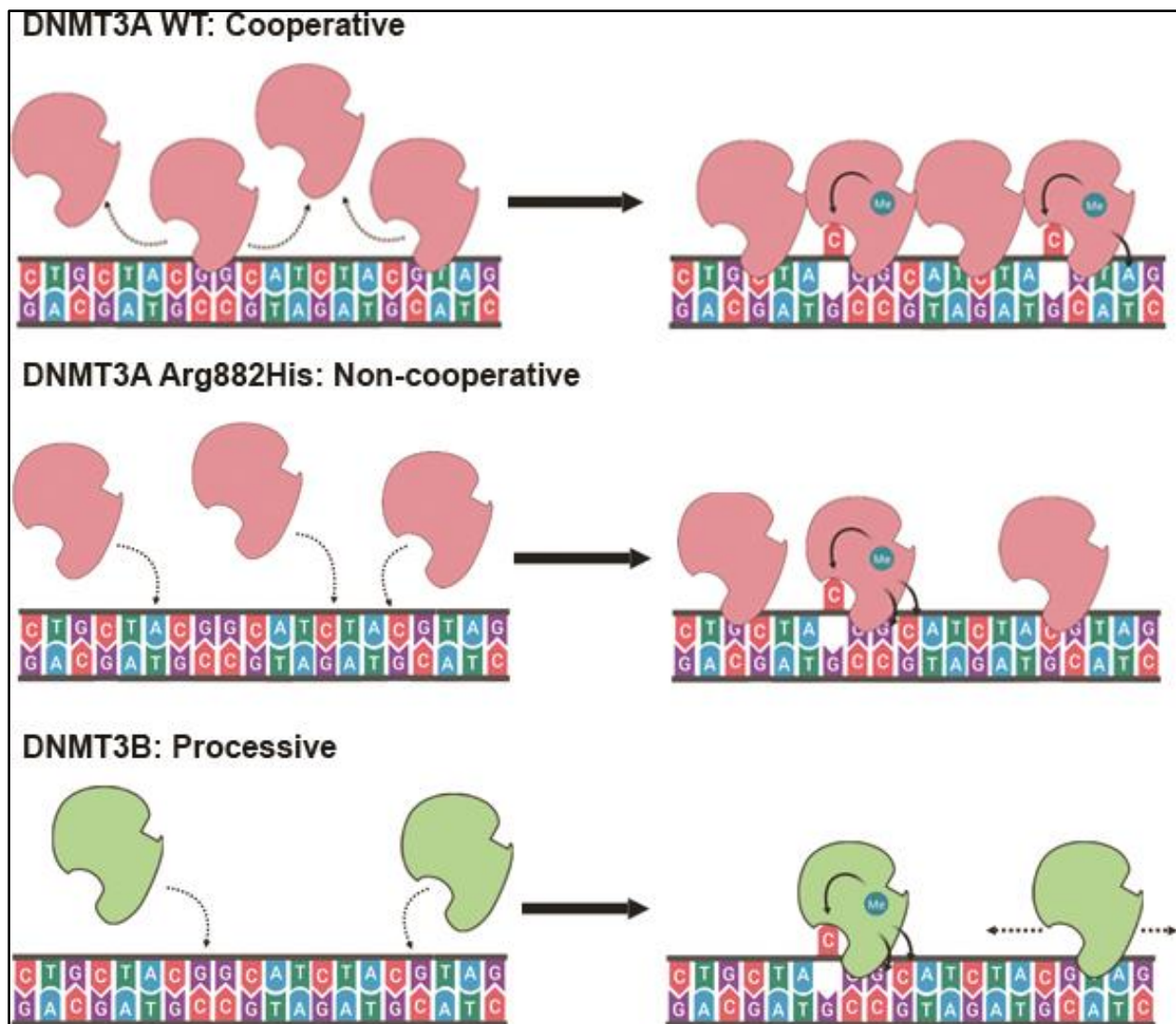


Figure 3.13 Model showing the different kinetics mechanisms of DNMT3A and DNMT3B and their influence their flanking sequence preferences. At high concentrations, DNMT3A methylates multiple CpG sites rapidly by using cooperative mechanism and has no strong preference at N+1 and N+3, and a minor preference for A at N+2. However, in absence of cooperative kinetics mechanism at low concentrations of DNMT3A enzyme, for the Arg882His variant, and DNMT3B, the flanking sequence preference for G at the N+1 and C at the N+2 position is observed.

3.6 Discussion

Despite numerous studies addressing the biological roles of DNMT3A and DNMT3B in development and diseases, the differences and similarities in their kinetics mechanisms remain poorly understood. Germline mutations in DNMT3A and DNMT3B have deleterious effects and are associated with congenital diseases (126, 193). In AML and other types of leukemia, the

majority of somatic mutations in DNMT3A affects Arg882, mostly leading to an Arg-to-His substitution (139). Because of its high prevalence (~20% in AML) and early occurrence during disease development, the Arg882His variant is considered a founder mutation (209). Therefore, the variant enzyme has been the subject of many studies. Through these studies, the Arg882His substitution was shown to alter DNA binding properties, attenuate tetramerization, disrupt cooperativity, and change the flanking sequence preference of the DNMT3A enzyme (111, 125, 142, 144, 201). In this study we show that the altered sequence preference of the AML variant can potentially cause aberrant methylation of DNMT3B target sites, thus contributing to its oncogenic potential.

We show that disruption of cooperativity alters the flanking sequence preference of DNMT3A-C WT to match the sequence preferred by DNMT3A-C Arg882His variant, demonstrating the effect of kinetics mechanism on substrate specificity. Using the DNA methylation analysis of 124 CpG sites, we systematically computed the temporal order of site preference for the DNMT3A-C WT, DNMT3A-C Arg882His, and DNMT3B-C WT enzymes. The most important finding in this study is the discovery that the altered flanking sequence preference of the DNMT3A-C Arg882His variant is nearly identical to the preferred substrate sequence of DNMT3B-C. Consequently, we predicted that the DNMT3A Arg882His variant would potentially methylate DNMT3B-specific targets. Indeed, DNA methylation of major and minor satellite repeats rescued by the WT or mutant DNMTs in DNMT3A/3B DKO mESCs provide strong evidence supporting this prediction. The data revealed that, compared to the DNMT3A WT, the mouse DNMT3A Arg878His (equivalent to human DNMT3A Arg882His) variant has little effect in rescuing DNA methylation at major satellite repeats. However, its ability to methylate the minor satellite repeats is almost fully retained. These data demonstrated that target site specificity of DNMT3A and DNMT3B *in vivo* could be strongly influenced by their respective CpG flanking sequence preference. Consistent with this, we observed that at the minor satellite repeats CpG flanking sequence showed an overrepresentation of G at N+1 and N+3 positions that is preferred by DNMT3B and the DNMT3A Arg882His variant. The major satellite repeats have high percentages of T and A at the N+1 and N+3 positions, respectively, which are preferred by DNMT3A WT enzyme and disfavored by DNMT3B WT and the Arg829His variant.

Given that under conditions favoring non-cooperative mechanism, the flanking sequence preference of DNMT3A WT is modified to match that of DNMT3A Arg878His variant, we

propose that DNMT3A methylates minor satellite repeats using a non-cooperative kinetics mechanism. This prediction also explains the rationale behind the lower activity of DNMT3A at minor satellite repeats compared to that of DNMT3B, which methylates its target sites using processive kinetics mechanism. Similarly, although DNMT3A Arg878His variant prefers DNMT3B sites, it methylates these sites using a non-cooperative distributive mechanism, explaining an incomplete rescue by the variant enzyme compared to DNMT3B. A high activity of DNMT3A at the major satellite repeats could be justified by the cooperative mechanism used to methylate multiple CpG sites. Therefore, these data also provide evidence for the role of distinct kinetics properties and substrate specificity of DNMT3A and DNMT3B in selection of their genomic target sites.

These observations also suggest that at genomic regions with sparse or dispersed CpG sites, the absence of cooperativity would favor methylation of sequences with Gs at N+1, 3 by both DNMT3A and DNMT3B at common target sites. However, the distinct tissue specific expression of DNMT3A and DNMT3B ensures that these enzymes methylate their specific targets, thus regulating their biological activity. The importance of this regulation is highlighted by aberrant expression of DNMT3B in various types of cancer, including AML (192, 197-199). The effect of DNMT3B overexpression in AML is similar to that of DNMT3A Arg882His mutation, leading to an increase in stemness, downregulation in apoptotic genes, and poor patient prognosis (142, 198). Our data here reveal a mechanism by which the DNMT3A Arg882His variant acts like the DNMT3B enzyme, thus providing a mechanistic explanation to the aforementioned observations in AML patients. We speculate that the oncogenic potential of DNMT3A Arg882His variant may not be only due to its lower activity causing DNA hypomethylation in AML cells, but also due to the gain of DNMT3B-like activity generating aberrant patterns of DNA methylation. While the DNMT3A Arg882 mutation may have other effects, these data suggest that changes of substrate specificity contribute to leukemogenesis caused by the DNMT3A Arg882His mutation. Thus, our observations provide novel insights into consequences of cancer-causing mutations on the enzymatic activity of DNMT3 enzymes.

CHAPTER 4. EFFECT OF DISEASE-ASSOCIATED GERMLINE MUTATIONS ON STRUCTURE FUNCTION RELATIONSHIP OF DNA METHYLTRANSFERASES

*Formatted for dissertation from the article published in *Genes* **2019**, 10(5), 369

Allison B Norvil, Debapriya Saha, Mohd Saleem Dar, Humaira Gowher

4.1 Declaration of collaborative work

All authors contributed equally to this review.

4.2 Abstract

Despite a large body of evidence supporting the role of aberrant DNA methylation in etiology of several human diseases, the fundamental mechanisms that regulate the activity of mammalian DNA methyltransferases (DNMTs) are not fully understood. Recent advances in whole genome association studies have helped identify mutations and genetic alterations of DNMTs in various diseases that have a potential to affect the biological function and activity of these enzymes. Several of these mutations are germline-transmitted and associated with a number of hereditary disorders, which are potentially caused by aberrant DNA methylation patterns in the regulatory compartments of the genome. These hereditary disorders usually cause neurological dysfunction, growth defects, and inherited cancers. Biochemical and biological characterization of DNMT variants can reveal the molecular mechanism of these enzymes and give insights on their specific functions. In this review, we introduce roles and regulation of DNA methylation and DNMTs. We discuss DNMT mutations that are associated with rare diseases, the characterized effects of these mutations on enzyme activity and provide insights on their potential effects based on the known crystal structure of these proteins.

4.3 Introduction

DNA methylation is a highly conserved epigenetic modification in mammals and takes place at the 5' position of cytosine, largely at the CpG dinucleotide (28, 35). The distribution of DNA methylation in mammalian genome is bimodal such that the repetitive elements and transposons are most densely methylated and the regions with highest propensity of CpG (CpG islands) are

least methylated (49). DNA methylation increases the information content of the genome through its potential to control gene expression. At regulatory elements of genes, including promoters and enhancers, DNA methylation is largely associated with repressed genes and is often tissue specific. Conversely, high DNA methylation is found in gene bodies of highly transcribing genes. All these observations indicate that interpretation of DNA methylation is dependent on the genomic context. Despite the complexity of DNA demethylation, the active loss of DNA methylation has been observed both during early development and at certain inducible genes in later adulthood (210). Although a functional demethylase that can directly remove $-CH^3$ groups from the 5'C of cytosine has not been discovered, reversal of DNA methylation can be mediated by the conversion of methyl to hydroxymethyl and higher oxidation states by TET family of methylcytosine dioxygenases. This can lead to progressive loss of the modification because these oxidized states unlike DNA methylation cannot be maintained. Besides being an intermediate of DNA demethylation process, hydroxymethylation at regulatory elements, which are in the “primed state”, alters the signal output of DNA methylation by changing its detectability (136, 211-213).

DNA methyltransferases (DNMTs) are a class of enzymes that catalyze transfer of a methyl group from S-adenosyl-L-methionine (AdoMet) to DNA. Mammalian DNMTs belong to two structurally and functionally distinct families, DNMT1 and DNMT3 (214-216). In somatic cells, 60-80% of CpG sites are methylated and DNMT1 diligently copies the methylation pattern from the parent to the daughter strand post-replication and repair (35, 217, 218). This activity, particularly, ensures the maintenance of the unmethylated regions by averting spurious *de novo* DNA methylation. The DNMT3 family includes DNMT3A and DNMT3B, which are the *de novo* methyltransferases, and one regulatory factor, DNMT3-Like protein (DNMT3L) (216). DNMT3L is catalytically inactive but interacts with both DNMT3A and 3B to enhance their enzymatic activity (100). The DNMT3 proteins are required for the establishment of genomic DNA methylation during embryogenesis after its erasure at the preimplantation stage (218). Whereas DNMT1 is ubiquitously expressed, DNMT3 enzymes show distinct tissue specific expression and methylate the regulatory elements of the transcriptionally inactive genes. Several lines of evidence including distinct phenotype of DNMT3A and DNMT3B KO mice indicate distinct biological functions of these enzymes (86). Studies of molecular and cellular phenotypes resulting from DNMT mutations have largely contributed to our understanding of the biological roles of DNA methylation (187).

Mammalian DNMTs constitute a C-terminal MTase domain, the structure of which is highly conserved from bacteria to humans (214, 219). Given mammalian DNMTs have weak sequence specificity, their target site recognition is guided by the N-terminal regulatory region, which interacts with transcription factors, chromatin binding proteins, and histone-tail modifications (220). Consequently, these interactions regulate site-specific DNA methylation leading to differential gene expression. Abnormal pattern of DNA methylation has been observed in several diseases and in all types of cancer. This could be caused by either loss of function of DNMTs or their interactions with modulators (125, 221). Recent high throughput sequencing have revealed mutations in DNMTs associated with several diseased states. Interestingly, in all reported disorders only one of the three enzymes accumulates the mutations, leading to distinct phenotypes (222). Although these mutations are distributed throughout the *DNMT* gene, most of them tend to cluster in the functional domains of these enzymes. The adverse effect of these mutations on the catalytic activity and function of DNMTs has been established in multiple reports (144, 145, 223-225). The following content will discuss recent advancements in the investigation of etiological consequences of germline - transmitted mutations in DNMT1 and DNMT3A and the effect of these mutations on catalytic and targeting mechanism of the enzymes.

4.4 Structural and functional alterations of DNMT1 by disease associated mutations

The eukaryotic DNMT1 is a multimodular protein comprising of a replication foci-targeting sequence (RFTS), a DNA binding CXXC domain, two bromo-adjacent homology (BAH) domains and a C-terminal MTase domain (Figure 4.1A) (226). DNMT1 has an intrinsic preference for hemimethylated CpG sites. This preference is further modulated by interactions of both CXXC and RFTS domains with the DNA binding region of DNMT1 leading to autoinhibition. The CXXC domain binds to unmethylated CpG dinucleotides and sandwiches a section of highly acidic amino acids (the autoinhibitory BAH1-CXXC linker) between the DNA and DNMT1 active site. The autoinhibition by RFTS is relieved by its interaction with UHRF1 (ubiquitin-like, containing PHD and RING finger domains 1), which binds to hemimethylated CpG dinucleotides. It is suggested that hemimethylated DNA from UHRF1 is transferred to the active site of DNMT1 after the inhibitory RFTS has been displaced by UHRF1/hemimethylated DNA complex (227). The multiple functional domains of the N- terminus have different roles including coordination of

methylation and replication during S-phase, partial suppression of *de novo* methylation and nuclear localization (28, 216, 228).

Exome sequencing studies revealed several mutations in DNMT1 that result in two adult onset, progressive neurological disorders. These mutations are germline dominant and include 13 amino acid substitutions in the RFTS domain that potentially disrupt the catalytic activity of the enzyme (Figure 4.1A). The first is hereditary sensory neuropathy with dementia and hearing loss (HSAN1E) and the second is autosomal dominant cerebellar ataxia, deafness and narcolepsy (ADCA-DN), caused by progressive loss of sensory neuron function (229-231).

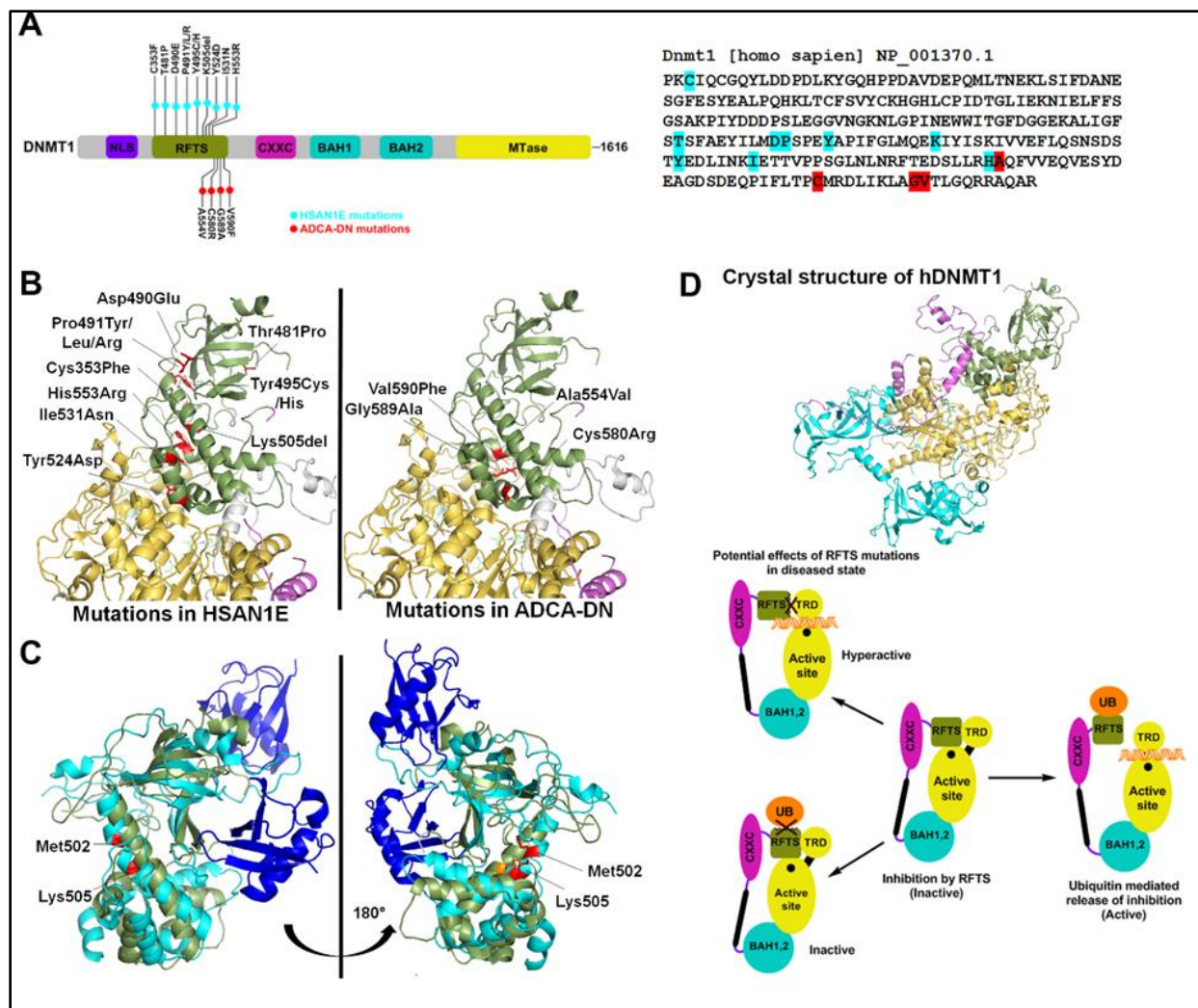


Figure 4.1 HSN1E and ADCA-DN mutations in DNMT1 (A) Left Schematic representation of the hDNMT1 gene. HSN1E mutations are listed above the gene in light blue, while ADCA-DN mutations are listed below the gene in red. Right Nucleotide sequence of the RFTS domain, with the mutations highlighted in color corresponding to the schematic. (B) Crystal structure of hDNMT1 (351 – 1600) from PDB: 4WXX. The cartoon structure of the RFTS domain is green, the CXXC domain is purple, and the MTase domain is yellow. All disease mutations are located in the RFTS domain, and are shown as stick structures in red. The positions of HSN1E and ADCA-DN mutations are shown in the left and right DNMT1 structure respectively. (C) Overlay of the hDNMT1 RFTS domain bound (light blue) and unbound (green) to two molecules of ubiquitin (dark blue) from PDB: 4WXX and 5YDR. When ubiquitin is bound, the RFTS domain bends about 30° at Met502. The HSN1E mutation Lys505del and Met502 are shown in red in RFTS domain bound to ubiquitin and in orange in the unbound RFTS domain. (D) Model showing the effect of mutations in the RFTS domain on the catalytic mechanism of DNMT1. DNMT1 is auto-inhibited by the interaction of its RFTS domain with TRD in MTase domain that prevents DNA binding. When RFTS interacts with ubiquitin, auto-inhibition is released allowing TRD to interact with the hemi-methylated DNA. However, mutations in the RFTS that alter its binding to ubiquitin will prevent enzyme activation, while mutations that alter its binding to the TRD will leave the enzyme in a hyperactive state.

4.4.1 HSN1E

To date nine heterozygous mutations of *DNMT1* gene have been identified in HSN1E patients. These mutations are located mostly in exon 20, which encodes part of the RFTS domain (Figure 4.1A). The first discovered mutations include substitution of two contiguous amino acids Asp490Glu and Pro491Arg, and a Tyr495Cys substitution. Systematic investigation revealed that these mutations lead to protein degradation, reduced DNMT1 activity and defective binding to heterochromatin in G2 phase (231). This ultimately leads to widespread DNA hypomethylation including pericentromeric satellite 2 sequences, other repetitive elements, intergenic regions, imprinted genes, and transcriptional start sites. At some CpG islands, site-specific hypermethylation was also reported (232). In cancer cells, the deletion of RFTS domain was shown to make DNMT1 hyperactive and available for euchromatic binding. Further, the binding to heterochromatin was decreased, potentially due to a missing interaction of RFTS with an unknown heterochromatin binding protein (233). Studies in mouse embryonic stem cells (ESCs) revealed that DNMT1 mutations Pro491Tyr and Tyr495Cys lead to decreased binding with UHRF1. The ESCs overexpressing the variant DNMT1 enzymes failed to properly differentiate into neuronal progenitor cells, suggesting a developmental defect as a possible mechanism for disease progression (234). In addition to the aforementioned mutations, six other mutations in RFTS domain were identified in HSN1E patients (Figure 4.1A) (229, 235-237). However, in absence of their biochemical characterization, the role of several of these mutations in disease development is not understood. Based on crystal structure analysis, we speculate that these mutations may lead to altered domain structure or interfere with protein-protein interactions (Figure 4.1B) (238). The UHRF1 interaction region spans from residues 458 – 500, suggesting that Thr481Pro mutation may also cause decreased UHRF1 binding. Biochemical studies however show no effect of this mutation on its localization to replication foci (229). Crystal structure of RFTS domain co-crystallized with ubiquitin shows interaction with two ubiquitin molecules (239). To accommodate the binding of two ubiquitin molecules, the RFTS domain undergoes a drastic conformational change, bending an alpha helix by about 30° at Met502 (Figure 4.1C) (239). Deletion of Lys505 may prevent the conformational change and affect the RFTS-ubiquitin interaction (239) (Figure 4.1C). His553, which is located in exon 21, interacts with Glu504 and Lys505 and may facilitate this conformational change, which can be affected by His553Arg mutation (237) (Figure 4.1C). Based on its position in RFTS domain, the Cys353Phe substitution may perturb zinc binding in

the RFTS domain and/or affect protein stability (Figure 4.1B). Indeed a study using recombinant expression of many of these variants showed cytosolic aggregation and early degradation of the GFP-tagged mutant proteins (229). Given that, DNMT1 protein is present in appreciable levels in neurons, cellular toxicity caused by protein aggregates, may underlie the clinical manifestations (229). This speculation is supported by the fact that adult neurons are post-mitotic performing little to no maintenance methylation and that the conditional deletion of DNMT1 gene in post-mitotic neurons of mice show no obvious phenotype (240). Together these studies support the conclusion that loss of DNMT1 targeting causes site-specific changes in DNA methylation in the HSAN1E patients.

4.4.2 ADCA-DN

ADCA-DN patients are reported to have four missense substitutions, Ala554Val, Cys580Arg, Gly589Ala, and Val590Phe. Similar to HSAN1E, all four mutations map to the RFTS domain of DNMT1, however they occur exclusively in exon 21 (230, 236, 241). All four mutations are located in the alpha helical bundle of the RFTS C-lobe, which has a hydrophobic pocket at the center. Therefore substitutions, Ala554Val and Val590Phe, present in the hydrophobic pocket and Gly589Ala, closely located to hydrophobic pocket, may destabilize the RFTS domain (Figure 4.1B). Further, these substitutions could also impair autoinhibition by weakening the interaction of the RFTS C-lobe with the DNA binding region of the MTase domain, rendering the enzyme hyperactive with the potential to be mis-targeted (Figure 4.1D). This speculation is supported by data showing that truncation of RFTS domain leads to dysregulation of DNMT1 activity (233). Methylation profiling of ADCA-DN patients showed global hypomethylation and hypermethylation specifically at around 80 CpG islands of which nearly half were associated with promoters and rest were inter- or intragenic (242). However, the effect of these changes on expression of associated genes and consequent biological function is unknown.

Besides HSAN1E and ADCA-DN disorders, previous work has established the role of aberrant DNA methylation in neurological disorders, such as Alzheimer's and Parkinson's disease (243-247). However the methylation defects are potentially caused by mis-regulation of DNMTs and/or interactions of MBD proteins, such as, MeCP2 with methylated DNA (248). Mutation of DNMT1 also cause alterations in the genome-wide DNA methylation patterns in colorectal cancer patients (133), however none of the patients with neurological disorders were shown to develop cancer

(136, 249). Conversely, mice carrying DNMT1 alleles with partial loss of function have developmental delays and high incidences of leukemia without discernible neurological defects. In summary, these studies support that germline versus somatic mutations have a spatiotemporal effect on the activity of DNMT1 during development and adulthood.

4.5 Structural and functional alterations of DNMT3 by disease associated mutations

The DNMT3 family consists of two catalytically active DNMTs, DNMT3A and DNMT3B and a catalytically inactive protein, DNMT3L. DNMT3A and DNMT3B have similar domain organization; both have a variable region at the N-terminus, followed by the Pro-Trp-Trp-Pro (PWWP) domain, a Cys-rich Zn-binding domain also called ATRX-DNMT3-DNMT3L (ADD) domain and a C-terminal methyltransferase (MTase) domain (250). The PWWP domain targets DNMT3A activity by binding to DNA and histone H3 methylated at Lys36 residue (H3K36me_{2/3}) (101-103). Co-crystal structure of the DNMT3L –ADD domain with histone H3 peptide shows that it specifically interacts with the Lys4 residue only when it is unmethylated (H3K4me₀). Methylation of histone H3K4 (H3K4me_{1/2/3}) disrupts this interaction (107). Interaction of the ADD domain with the DNA binding region of the DNMT3A MTase domain was revealed in a recent crystal structure suggesting its role in autoinhibition of the DNMT3A enzymatic activity. This autoinhibition is relieved by the interaction of the DNMT3A-ADD domain with histone H3K4me₀ (251). The dynamic role of this regulatory mechanism was shown to regulate DNA methylation at the enhancers of pluripotency genes during embryonic stem cell differentiation (252).

The MTase domain comprises ten sequence motifs, which are conserved in all cytosine DNMTs and have direct role in the catalysis (214, 219). Motifs I – III are involved in binding to the AdoMet, whereas motifs IV and VI are required for the catalysis. The region between and including motifs VIII and IX is called the target recognition domain (TRD) and is responsible for DNA binding in DNMT3A (253, 254). DNMT3A forms a hetero-tetrameric structure with DNMT3L in which two DNMT3A monomers form the center of the complex, flanked by two DNMT3L monomers on either side (255). In the heterotetramer, the DNMT3A – 3L interaction is mediated by two Phe residue and DNMT3A –3A interaction surface comprises Arg and Asp, therefore named as the RD interface (100, 255). In the absence of DNMT3L, DNMT3A forms homo-tetramers and can oligomerize on DNA (225, 254). This property facilitates DNMT3A's

cooperativity, where multiple enzyme units interact with DNA to methylate it at a faster rate (256). Mutations in RD interface disrupts DNMT3A DNA binding and activity demonstrating the critical role of protein dimerization in catalysis (225).

While there are only a handful of diseases caused by DNMT3 germline mutations, these diseases are caused by a plethora of mutations. Among all DNMTs, the disease-causing mutations were first discovered in DNMT3B in patients with immunodeficiency, centromeric instability, and facial anomalies (ICF) syndrome (86, 127, 178). The implications of these mutations on DNMT3B activity and on the etiology of ICF has been extensively investigated and reviewed (257). More recently, a high prevalence of DNMT3A somatic mutations were observed in hematological malignancies, Acute Myeloid Leukemia (AML) and Myelodysplastic syndrome (MDS). A series of germline mutations in DNMT3A were discovered in patients with growth syndromes, Tatton-Brown-Rahman syndrome (TBRS) and microcephalic dwarfism (MD). Some of these mutations were also found in hereditary tumors, pheochromocytomas (PCC) and paragangliomas (PGL).

4.5.1 Tatton-Brown-Rahman syndrome

DNMT3A related overgrowth syndrome, also known as Tatton-Brown-Rahman syndrome (TBRS) is an autosomal dominant condition characterized by overgrowth, distinctive facial appearance, and intellectual disability. It is caused by heterozygous mutations in DNMT3A that are transmitted through the germ line. In 55 TBRS patients, more than 40 distinct DNMT3A variants have been reported. Of these, most are missense mutations (30 variants), and the rest are nonsense variants, frameshift variants or whole gene deletions (258-265). The TBRS mutations are specifically localized in each of the three functional domains of DNMT3A, 11 of which overlap with the somatic DNMT3A variants found in hematological malignancies (Figure 4.2A). A recent study performed genome-wide DNA methylation analysis of 16 TBRS patients and detailed analysis of the methylation distribution in one patient with Arg771Gln substitution. Their data showed widespread DNA hypomethylation at specific genomic sites located near genes involved in morphogenesis, development, differentiation, and malignancy predisposition pathways, thus providing an important insight into developmental mechanisms that are dysregulated in the disease (266).

In the PWWP domain, several frameshift mutations, one deletion, and five missense substitutions were reported (Figure 4.2A). Analysis of the crystal structure shows that the missense

mutations cluster around the aromatic cage which interacts with H3K36me2/3 (103). Notable mutations include Arg301Trp, Gly298Trp/Arg, Tyr365Cys and Trp297del, which are near to or interact with the aromatic cage residues, and therefore may disrupt its binding to H3K36me2/3 (Figure 4.2B).

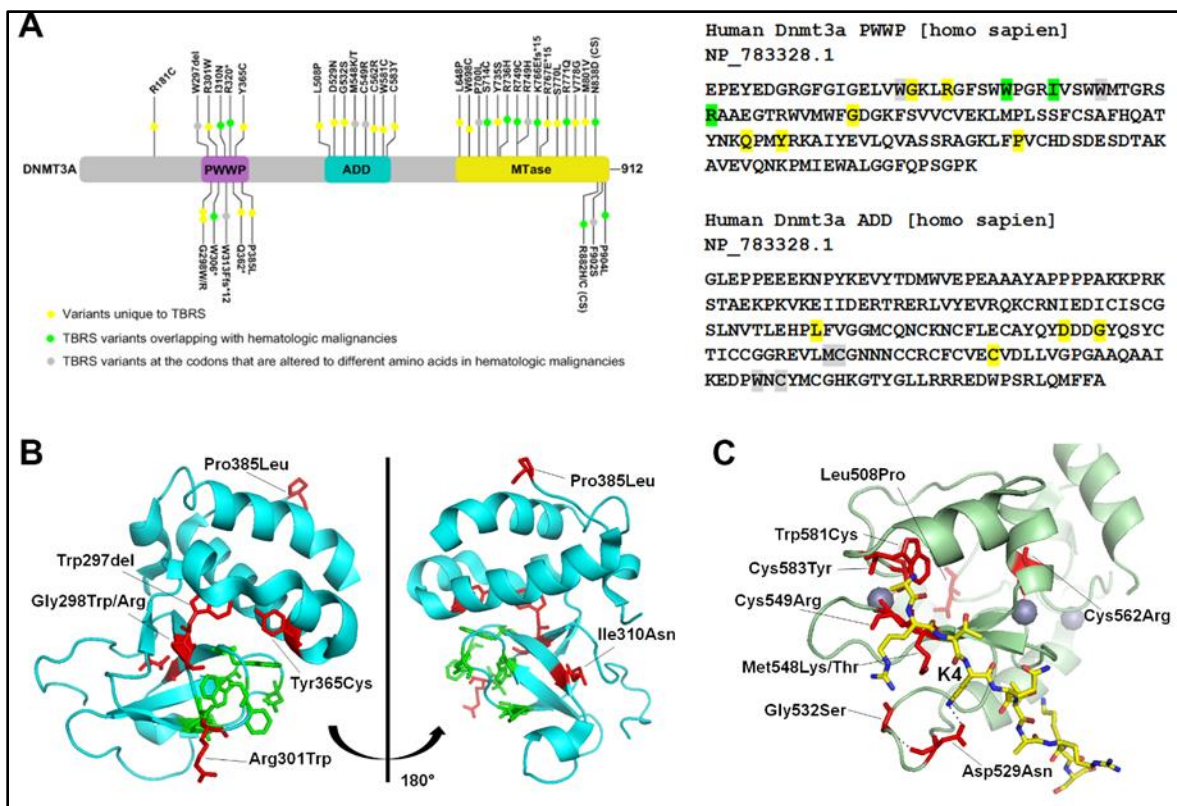


Figure 4.2 TBRS mutations in the PWWP and ADD domain of DNMT3A (A) Left Schematic representation of the hDNMT3A gene, with mutations listed in different domains. Variants unique to TBRS are highlighted in yellow, TBRS variants overlapping with hematologic malignancies are highlighted in green, and TBRS variants at the codon that are altered to different amino acids in hematologic malignancies are highlighted in grey. In the schematic, (*) is used to indicate a stop codon replacement, Ffs indicates a frame-shift mutation, and CS indicates the catalytic site. Right Nucleotide sequence of the PWWP and ADD domain, with the mutations highlighted in the color corresponding to the schematic. (B) Two orientations of the DNMT3A PWWP domain (PDB: 3LLR). The positions of TBRS mutations are shown as stick structures in red, whereas residues part of the aromatic cage that bind to H3K36me2/3 shown as stick structures in green. (C) Crystal Structure of the DNMT3A ADD domain (PDB: 4U7T), in green bound to an unmodified H3 peptide shown as a stick structure in yellow. Grey spheres represent bound zinc. The positions of TBRS mutations are shown as a stick structure in red.

In TBRS patients, mutations in the ADD domain cluster around the H3K4 binding site. Crystal structure of DNMT3A bound to H3K4me0 shows that Asp529 in the ADD domain makes direct contact with the Lys4 of the histone H3 (251). In absence of the histone H3 peptide, Asp529 also interacts with the DNA binding region, suggesting its involvement in regulation of the autoinhibited state. The TBRS variant Asp529Asn therefore could potentially be hyperactive and mis-targeted. Biochemical data showing that the variant Asp529Ala is neither autoinhibited by the ADD domain nor activated by unmethylated H3 peptide support this speculation (251). Met548 and Trp581 also make direct contact with the H3 tail, and therefore the variants Met548Thr and Trp581Cys may have reduced interaction with histone H3. Indeed, as reported, Met548Trp variant cannot be released from the autoinhibited state in the presence of unmethylated H3 peptide (251). Besides these mutations Cys549, Cys562, and Cys583 are present in the zinc finger regions. These residues mediate zinc binding and the mutations could potentially alter the DNMT3A structure or stability.

In TBRS patients, 15 substitutions and 2 frameshift mutations in the MTase domain are distributed in all sequence motifs (Figure 4.3A) (258, 259). We analyzed the crystal structure of DNMT3A to predict the potential effects of TBRS mutations on its activity (Figure 4.3B). As visualized in the structure of DNMT3A, Trp698, Pro700 and Arg736, are spatially located near Motifs I – III, and therefore the mutations at these residues could affect the AdoMet binding and catalysis (Figure 4.3C). Pro700 interacts with Arg635 in Motif I, and Arg736 contacts the backbones of Arg688 in Motif III (Figure 4.3C). The most recent co-crystal structure of DNMT3A bound to DNA shows that Ser714 interacts with phosphodiester backbone and Arg749 interacts with Asp702 of motif IV, which is involved in catalysis (Figure 4.3D) (267). Biochemical investigation of the TBRS substitutions Arg736His and Ser714Cys were recently reported (223). The data show that Arg736His variant, interestingly, has a threefold increase in catalytic turnover but is weakly stimulated by DNMT3L, and has an increased preference for non-CpG sites. However, Ser714Cys has reduced activity as well as reduced stimulation by DNMT3L (223, 267). These data suggest that besides having a direct effect on catalysis, these mutations could alter the structure disrupting the interaction of DNMT3A with DNMT3L. The regions where DNMT3A interacts with the DNA (TRD) spans from motif VIII – IX. Residues that are mutated in or near this region include Val778, Met801, Asn838, Arg882, Phe902, and Pro904. Val778 and Met801 are spatially located near motif VIII, so their mutations to Gly and Val, respectively, may alter

DNMT3A's ability to bind to DNA (Figure 4.3E). Asn838 interacts with the phosphodiester bond between nucleotides at N+2 and N+3 from the target CpG site, suggesting that the Asn838Asp variant may have a weak binding to DNA (Figure 4.3F) (254, 267). Phe902, Pro904 and Leu648 are spatially located near motif X (Figure 4.3G). *In vitro* studies show that Pro904Leu variant has higher catalytic turnover and negligible effect on DNMT3L mediated DNMT3A stimulation (223).

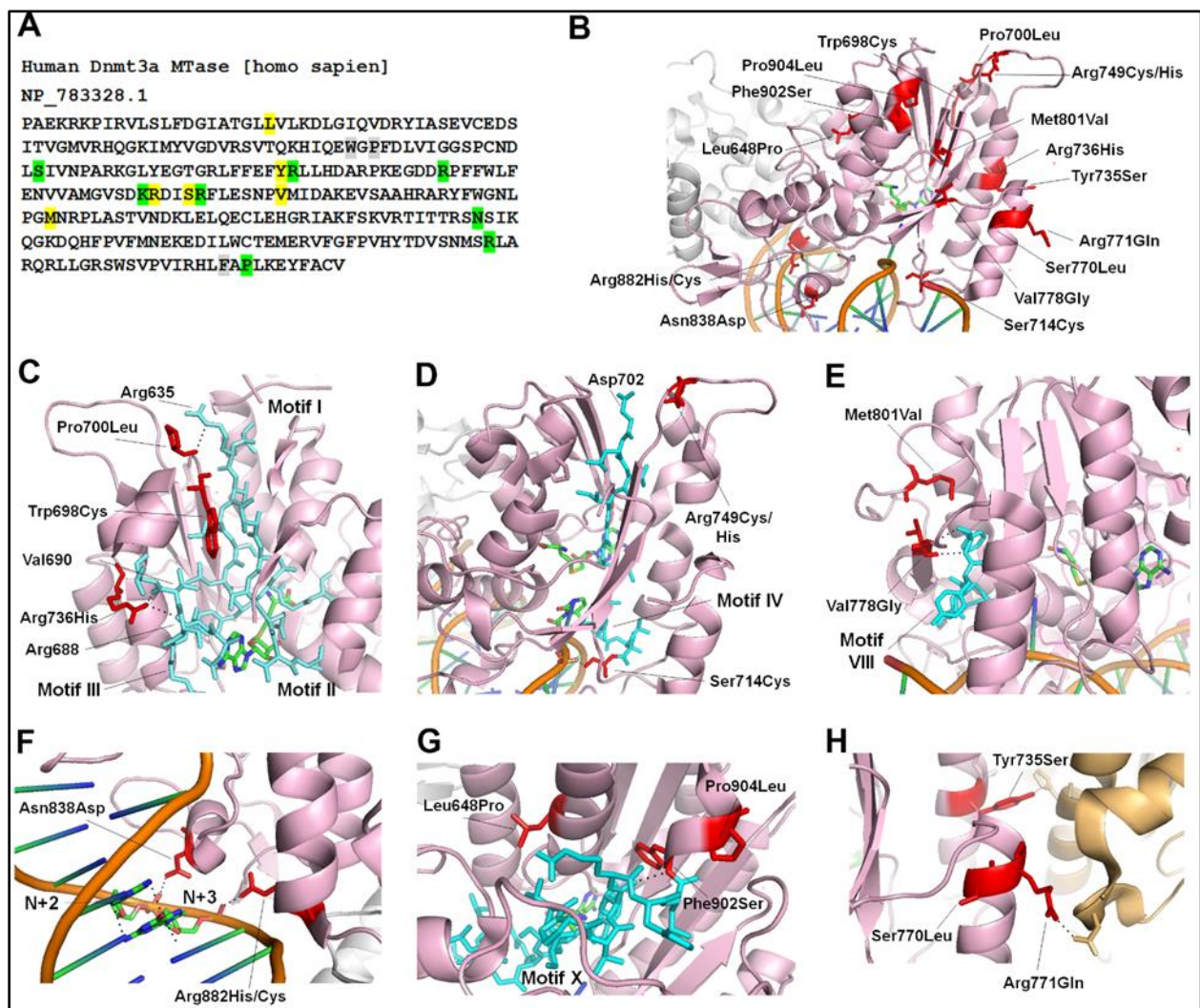


Figure 4.3 TBRS mutations in the catalytic domain of DNMT3A (A) Nucleotide sequence of the hDNMT3A MTase domain. As seen in Figure 4.2A, variants unique to TBRS are highlighted in yellow, TBRS variants overlapping with hematologic malignancies are highlighted in green, and TBRS variants at the codon that are altered to different amino acids in hematologic malignancies are highlighted in grey. (B) Crystal structure of DNMT3A bound to DNA, zoomed in to show only one of the two monomers in the tetrameric structure (PDB: 5YX2). The MTase domain is shown in a pink cartoon structure and the positions of TBRS mutations are shown as stick structures in red. The second monomer of the MTase is shown in grey. (C - G) Magnified view of the motifs I-III (C), motif IV (D - E), motif VIII (F), the TRD (G), and motif X (H) shown as stick structures in blue and the positions of TBRS mutations shown as stick structures in red. The black dotted lines represent interactions with nearby residues or with the DNA. (H) The interface between DNMT3A, pink cartoon, and DNMT3L, orange cartoon. The positions of TBRS mutations shown as stick structures in red. The black dotted lines represent interactions with nearby residues of DNMT3L shown as stick structures in orange.

The TBRS variant, Arg882His, which is also the most prominent somatic variant of DNMT3A in acute myeloid leukemia (AML) patients, has been extensively studied (Figure 4.3F) (125). Interestingly mutation of Arg882 is found in 25% AML patients and 25% of TBRS patients indicating this to be a hotspot (265). Given the majority of TBRS patients are pediatric or young adults, it is difficult to determine the risk of AML, because of its late onset. More recently, two TBRS patients were diagnosed with AML in childhood supporting potential risk and susceptibility of TBRS patients to develop AML (259).

The effect of Arg882His substitution on DNMT3A activity, reported by several studies, causes 40%-80% loss of catalytic activity (268, 269). Arg882 interacts with the phosphodiester bond of the nucleotide at N+3 from the target CpG site. Given the position of R882 close to the RD interface, the mutation disrupts intermolecular interactions, thus preventing the enzyme to form tetramers (Figure 4.3F). This in turn negatively affects DNMT3A cooperativity and decreases its DNA binding capacity (224). More recently, the variant was shown to have altered flanking sequence preference around the CpG site (144). However, it is not clear whether altered flanking sequence preference is a direct consequence of the amino acid substitution or an indirect effect of the loss of cooperativity. Based on its effect on genomic DNA methylation, Arg882His is also suggested to have a dominant-negative effect on the wild type enzyme, however this activity was not confirmed by *in vitro* experiments (144, 269, 270).

Another group of TBRS mutations includes residues Tyr735, Ser770, and Arg771, involved in intermolecular interactions with DNMT3L (Figure 4.3H). Tyr735 and Arg771 make direct contact with His and Asp residues in DNMT3L, respectively, so their mutation may alter the stability of this interface. Biochemical analysis of Arg771Gln variant interestingly shows an increase in the catalytic activity of the enzyme by 6 fold and no change in the level of stimulation by DNMT3L (223). However other substitutions of this residue resulted in a decrease in the stimulation by DNMT3L showing the role of this residue in stabilization of the DNMT3A-3L interactions.

It is interesting to note that while majority of the mutated residues are highly conserved in DNMT3B, three residues Leu648, Ser714, and Arg736 are conserved in bacterial DNA cytosine methyltransferases including M.HhaI and M.HaeIII, suggesting the effect of substitutions on their conserved structure and catalytic mechanism. However, compared to the level of overlap between DNMT3A mutations found in hematologic malignancies and growth syndromes, very few

coincide with mutations of DNMT3B found in ICF patients. Further, the effect of TBRS mutation, Arg736His, on the activity of DNMT3A is notable given that this substitution occurs normally in DNMT3B. This suggests that Arg736 in DNMT3A is important for the catalytic mechanism, which has common and distinct features from that of DNMT3B. These observations suggest that the catalytic mechanisms of DNMT3A and DNMT3B are critical for their unique biological functions. While this speculation is anticipated, effects of these substitutions on the catalytic mechanism of DNMT3A compared to DNMT3B need to be further elucidated.

4.5.2 Hereditary tumors and Microcephalic dwarfism

Pheochromocytoma/paraganglioma (PCC/PGL) is a rare neuroendocrine malignancy that may develop at various body sites, including the head, neck, and abdomen, and has a five-year survival rate of only 40% patients (271, 272). PCC/PGL is the most heritable of all tumors and carries both germline and somatic mutations in 1 of 20 known genes including metabolic genes. Recently de novo germline mutations in DNMT3A were reported in PCC/PGLs. The mutations occur in the PWWP domain, and result in substitutions Lys299Ile and Arg318Trp (Figure 4.4A). Although these residues are not the part of the aromatic cage, crystal structure analysis shows that Lys299 interacts with the backbone of Phe303, which stabilizes the interaction of PWWP domain with H3K36me_{2/3} (Figure 4.4B). Conversely, Arg318 interacts with the Val35 of the H3 tail, so the substitution Arg318Trp may effect H3 binding. However, future biochemical studies will be needed to show the effect of these mutations on the activity of the enzymes. Methylation profiling of PCC/PGL patients and of HeLa cells carrying CRISPR/Cas9- mediated knock-in of the PGG/PCL DNMT3A mutation show site-specific hypermethylation at homeobox genes, genes involved in dopaminergic neurogenesis, neural crest differentiation, and embryonic morphogenesis. Given that previously known mutations in DNMT3A causing overgrowth syndrome result in genome-wide hypomethylation, the PWWP mutation in PCC/PGL leading to hypermethylation is described as gain-of-function mutation.

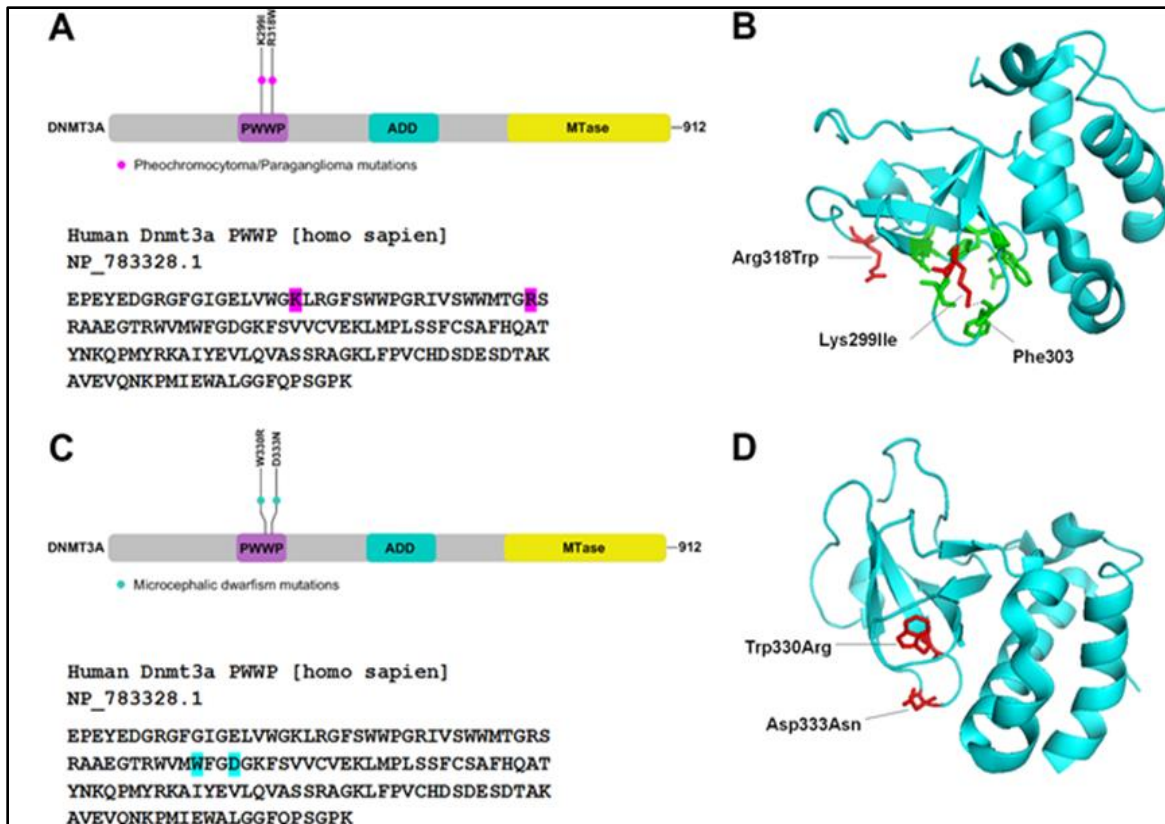


Figure 4.4 PCC/PGL and MD mutations in PWWP domain of DNMT3A (A, C) Top Schematic representation of the hDNMT3A gene, showing PCC/PGL and MD mutations in the PWWP domain. Below Nucleotide sequence of the PWWP domain, with the mutations highlighted in pink and blue respectively. (B, D) Crystal structure of DNMT3A PWWP domain shown in blue (PDB: 3LLR). The positions of PCC/PGL mutation (B) and MD mutations (D) shown as stick structures in red. The aromatic cage residues are shown in green.

Similar DNMT3A heterozygous gain-of-function mutations were recently shown to cause microcephalic dwarfism, a hypocellular disorder of extreme global growth failure, including a reduction in head size and height (273). The mutations result in substitutions Trp330Arg and Asp333Asn, both of which are located in the DNMT3A PWWP domain (Figure 4.4C). Both residues are part of the aromatic cage that interacts with H3K36me2/3, and when mutated, the interaction with chromatin is abrogated (Figure 4.4D) (273). In contrast to some TBRS mutations, which lead to PWWP domain instability, MD mutations do not affect the stability of this domain (273). Genome-wide DNA methylation analysis of patients' fibroblasts showed that similar to PCC/PGL, the majority of differentially methylated regions (DMRs) were hypermethylated compared to wild type samples, and these DMRs were associated with developmental transcription

factors and morphogen genes. In the control fibroblasts these regions were marked by trimethylation of lysine 27 on histone H3 (H3K27), established by the PRC2 complex, and the DNA remains unmethylated (273). Notably, many of the hypermethylated regions were identified previously as broad non-methylated islands or differentially methylated valleys and often associated with polycomb-regulated developmental genes (274-277). Reduced H3K27me3 at these hypermethylated sites occurs despite normal levels of EZH2. Previous observations have shown an antagonism between DNA methylation and deposition of H3K27me3, and that DNA methylation can abrogate SUZ12 binding to nucleosomes (275, 278). These observations support the speculation that PRC2 binding is impaired by DNA methylation at hypermethylated regions in MD patients. Further based on the data that MD DNMT3A variants cannot be targeted to H3K36me3 chromatin domains, they are speculated to methylate transcriptionally repressed regions spuriously, which are otherwise regulated by the PRC2 complex. Compared to PRC mediated repression, which supports plasticity and allows reactivation of genes under various developmental cues, DNA methylation is considered to silence genes stably. Whereas loss of DNMT3A increases the stemness with reduced developmental potential of stem cells, loss of EZH2 promotes premature differentiation. The DNMT3A MD variants are proposed to be a gain-of-function mutations, which might increase cellular differentiation causing premature depletion of stem/progenitor cell pools. This in turn could affect the growth of tissues and lead to reduced organism size (273).

4.6 Conclusion and Perspectives

DNA methylation defects in absence of DNMT mutations have been reported in plethora of disorders. However, with exception of DNMT3B germline transmitted mutations that cause ICF, disease-causing mutations in DNMT1 and DNMT3A were recently discovered. These mutations cause growth disorders, HSN1E, ADCA-DN, TBR1 and MD that have distinct clinical manifestations.

The ADCA-DN and HSN1E mutations were mapped to the regions that could potentially impair the interaction of DNMT1-RFTS domain either with the ubiquitin or with the DNA binding region of the MTase domain. The former will render the enzyme in an autoinhibited state, and the latter will lead to an unregulated activity of DNMT1 enzyme making it hyperactive (Figure 4.1D). Methylome mapping of the patients show aberrant DNA methylation including both widespread

hypomethylation and site-specific hypermethylation. This methylation pattern is similar to the one observed in various cancers indicating a similar loss of DNMT1 function, however none of the HSN1E and ADCA-DN patients develop cancer. It is speculated that in these patients, a gradual loss of DNA methylation over time may cause late onset and progressive neurological disabilities. However, given that post mitotic neurons are terminally differentiated and do not perform maintenance methylation, the effect of RFTS mutations on the activity of DNMT1 could be due to protein misfolding and aggregation (229). Interestingly, ADCA-DN and HSN1E have some overlapping clinical features typical of mitochondrial diseases, which is supported by biochemical evidence of mitochondrial dysfunction (232, 236, 279). Given that presence of methyl cytosine in mtDNA is still debated, it is not clear how DNMT1 mutations can cause mitochondrial dysfunction through its effect on mtDNA methylation. Taken together, these data suggest that DNMT1 mutants could exert their damaging effect through at least two mechanisms including impairment of epigenetic pathways, and cellular stress by the protein misfolding (229, 231, 232).

Compared to TBRS in which DNMT3A mutations are present in all functional domains of the protein, in MD, the two mutations map only to the PWWP domain. However, in contrast to MD-specific PWWP mutations, the TBRS-specific PWWP mutations affected protein stability resulting in loss of protein function. Based on the observation that the MD variant, DNMT3A Trp330Arg, is unable to interact with H3K36me2/3, it was proposed to be “available” to methylate sites that are normally polycomb repressed (273). This is consistent with the observation that polycomb repressed developmental genes in ESCs gain DNA methylation during differentiation and are often found hypermethylated in cancer cells (217, 276, 280). It is possible that by interacting with PRC2, DNMT3A Trp330Arg is targeted to these regions in MD patients (274, 277). In addition, once the DNA methylation is established, it could potentially block the activity of PRC2 by interfering with the binding of SUZ12 (275) and result in stable repression of the developmental genes (278). Furthermore, mutations of epigenetic modulators, EZH2 and NSD1, cause the overgrowth syndromes Weaver and Sotos syndrome respectively, which supports a critical role of epigenetic regulation in organism size.

The discovery of DNMT mutations in rare diseases has shaped our understanding of the cause and consequence of aberrant DNA methylation in various disorders. It is clear that global hypomethylation, often found in cancers as well as growth disorders, may not be a direct consequence of DNMT loss of function, rather an indirect response to a diseased state. Targeted

hypermethylation seems to be direct consequence of aberrant DNMT activity. This is also supported by inconsistency between the biochemical outcomes of DNMT mutations and their effect on genomic methylation. Future studies designed to address the direct effect of DNMT mutations on genomic DNA methylation patterns will help understand the contribution of DNMTs in pathogenesis.

CHAPTER 5. PERSPECTIVES

5.1 Outlook on Characterizing the Biochemical Properties of the DNMT3 Enzymes

5.1.1 Conclusions

Several studies have highlighted the role of the specific enzymatic properties of DNMTs in regulation of their activity in vivo. Although DNMTs are targeted to specific genomic locations through their interactions with various proteins and modified histidine tails, the intrinsic specificity of DNMTs for specific sequences can dictate their target site specificity. In our studies, we demonstrated that unique enzymatic properties of DNMT3A and DNMT3B control their choice of genomic targets. Our studies demonstrated that DNMT3B has a distinct kinetics mechanism compared to DNMT3A, and is a non-cooperative, processive enzyme (118, 120). Cooperativity of DNMT3A is disrupted in the AML Arg882His variant, while the conserved mutation in DNMT3B had no effect on its processivity. Based on these data, we predicted that the observed change in flanking sequence preference of the DNMT3A Arg882His variant was due to the loss of its kinetics mechanism. Under conditions that favor non-cooperativity, wild type DNMT3A had a similar flanking sequence preference as the Arg882His variant, suggesting the role of cooperativity in influencing flanking sequence preference. Further, we showed that the non-cooperative DNMT3B has a similar flanking sequence preference as the DNMT3A Arg882His variant. Based on these data, we predicted that the DNMT3A Arg882His variant could target DNMT3B specific sites. DNMT3A methylates major satellite repeats whereas DNMT3B methylates minor satellite repeats (86). In *DNMT3A/3B* double knock out mESCs, the DNMT3A Arg882His variant rescued methylation at the minor satellite repeats to a similar level as the wild type enzyme. From these data, we found a previously unappreciated sequence bias in the consensus motifs of the major and minor satellite repeats that matches the preferences of DNMT3A and DNMT3B, respectively. These studies provide a mechanism by which the substrate specificity and kinetics mechanisms of DNMT3A and DNMT3B regulate DNA methylation in various genomic regions. Mutations to DNMT3A and DNMT3B have been identified in germ-line diseases as well as cancer, and our findings demonstrate how these mutations could disrupt the kinetics mechanisms of these enzymes, leading to differentially methylated regions and aberrant gene expression.

In our studies, we primarily focused on the effect of the Arg882His variant on the kinetics mechanism of DNMT3A, since it is the most common disease-causing mutation in the DNMT3 enzymes. However, there are multiple other disease associated variants of the DNMT3A and DNMT3B MTase domains that remain uncharacterized. Due to their occurrence in the conserved motifs, we predicted that many of mutations will disrupt DNA binding, catalysis, or alter kinetics mechanisms. Further studies could investigate the effect of these substitution on the kinetics mechanisms of DNMT3A and DNMT3B to determine their role in disease progression.

5.1.2 The Role of the TRD and R-D Interface in Substrate Specificity

The TRD is a highly variable region located between motifs VIII and IX of the DNMT3 MTase domains that defines sequence specificity outside the target CpG site. Co-crystal structures of DNMT3A with either 5'-CGT-3' or 5'-CGA-3' DNA substrates revealed that the TRD adopts different conformations depending on the flanking sequence. When T is at the N+1 position, Arg836 and Asn838 of the TRD make base specific interactions that dictate substrate specificity (123), while Arg882 interacts with Ser837 and the DNA phosphate backbone between the N+3 and N+4 position to stabilize the TRD structure (Figure 5.1A) (124). Conversely, when DNMT3A binds to a substrate with A at the N+1 position, the Arg836 and Asn838 interactions change, and Arg882 cannot interact with Ser837 to stabilize the TRD (Figure 5.1B) (124). In DNMT3A variant with Arg882His substitution, the capacity for binding a 5'-CGA-3' substrate is retained, since Arg882 only binds to the phosphate backbone. However, in presence of the 5'-CGT-3' substrate, the His residue cannot interact with Ser837, which prevents the TRD from adopting the proper structure. These studies provide an in-depth analysis on the mechanisms that govern preference at the N+1 position, but the biochemical properties that drive flanking sequence preference at other sites remain unknown. Our studies show that cooperativity modulates the flanking sequence preference of DNMT3A, which suggests that multimerization of DNMT3A alters the conformation of the DNA binding region. Mutational analyses of the residues in the DNA binding interface could be performed to determine how cooperativity modulates flanking sequence preference. Further, these assays would identify the residues that dictate the flanking sequence preference of DNMT3A beyond the N+1 position.

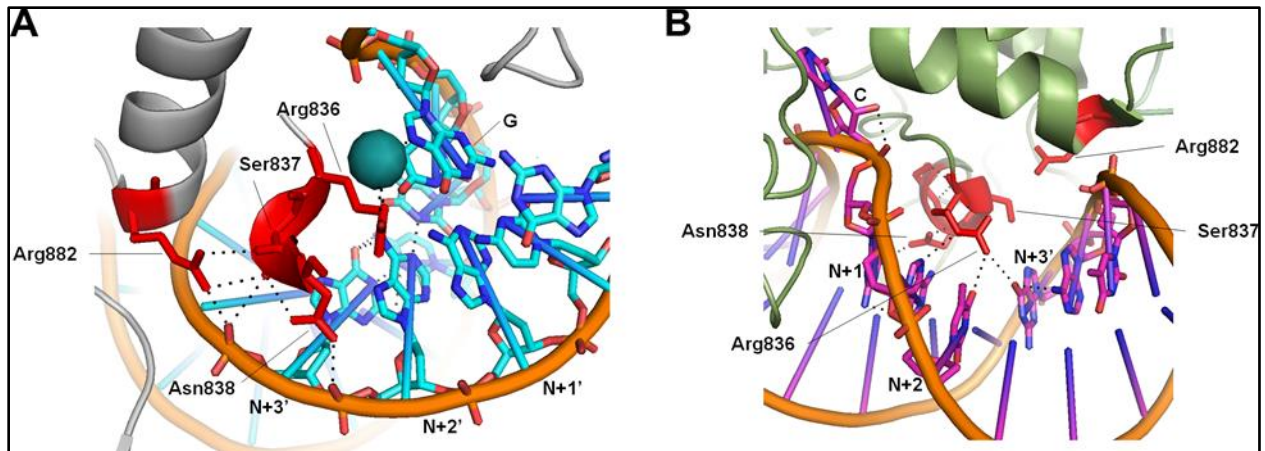


Figure 5.1 DNMT3A target recognition domain interacting with DNA. (A) Crystal structure of DNMT3A bound to DNA with a CpGpT sequence, zoomed in to show only one of the two monomers in the tetrameric structure (PDB: 5YX2). The MTase domain is shown in a grey cartoon structure and the residues that dictate flanking sequence preference are shown as stick structures in red. The black dotted lines represent interactions with nearby residues or with the DNA. (B) Crystal structure of DNMT3A bound to DNA with a CpGpA sequence, zoomed in to show only one of the two monomers in the tetrameric structure (PDB: 6W8B). The MTase domain is shown in a green cartoon structure and the residues that dictate flanking sequence preference are shown as stick structures in red. The black dotted lines represent interactions with nearby residues or with the DNA.

Co-crystallization of DNMT3B with DNA showed that the TRD domain of DNMT3B becomes ordered when bound to DNA, similar to DNMT3A (112). In the presence of a 5'-CGG-3' substrate, which is preferred by DNMT3B, Asn779 (Asn838 in DNMT3A) and Lys777 (Arg836 in DNMT3A) interact with G of the CpG target site and G at the N+1 position (Figure 5.2A). While these interactions are similar to those made by DNMT3A when T is at the N+1 position, the DNMT3B Arg829 residue (Arg882 in DNMT3A) does not support the rigid TRD structure like it does in DNMT3A. Instead, Gln772 forms a hydrogen bond network to stabilize the TRD of DNMT3B. When DNMT3B binds to a disfavored substrate, like T at N+1, nearly all the interactions between Lys777, Asn779, and the DNA are disrupted (Figure 5.2B). These data provide novel insights on the structure of DNMT3B that can influence its catalytic mechanism. Like DNMT3A, it is unknown how DNMT3B substrate specificity is regulated beyond the N+1 position. Additionally, DNMT3B mutations observed in ICF patients have been shown to decrease catalytic activity, but it is unknown how these substitutions effect other biochemical properties.

Future studies could determine if the ICF variants of DNMT3B lose processivity or have an altered flanking sequence preference.

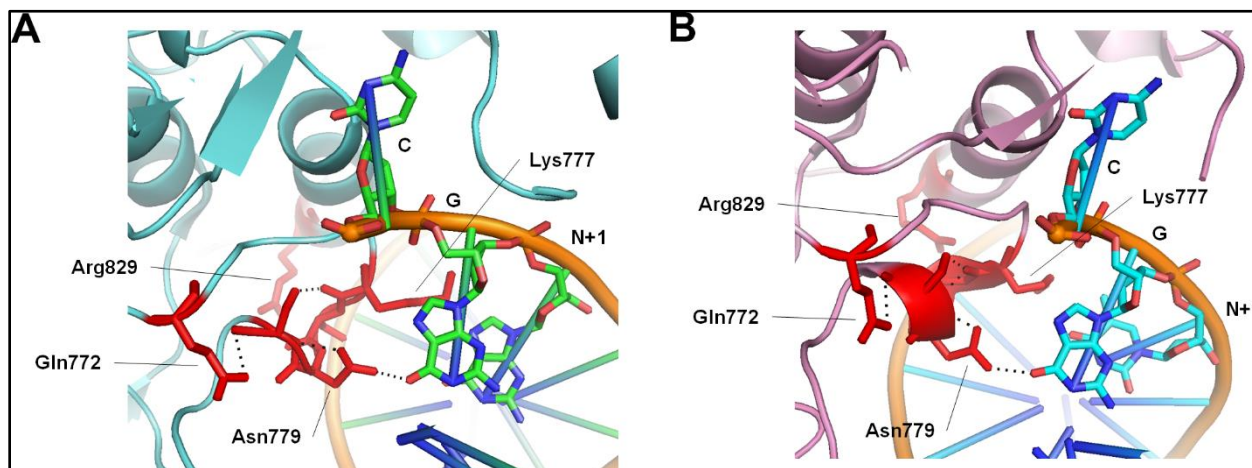


Figure 5.2 DNMT3B target recognition domain interacting with DNA. (A) Crystal structure of DNMT3B bound to DNA with a CpGpG sequence, zoomed in to show only one of the two monomers in the tetrameric structure (PDB: 6KDA). The MTase domain is shown in a blue cartoon structure and the residues that dictate flanking sequence preference are shown as stick structures in red. The black dotted lines represent interactions with nearby residues or with the DNA. (B) Crystal structure of DNMT3B bound to DNA with a CpGpT sequence, zoomed in to show only one of the two monomers in the tetrameric structure (PDB: 6KDB). The MTase domain is shown in a pink cartoon structure and the residues that dictate flanking sequence preference are shown as stick structures in red. The black dotted lines represent interactions with nearby residues or with the DNA.

Many studies have suggested that the TRD influences the kinetics mechanisms of DNMT3A and DNMT3B. Several mutations in the TRD of DNMT3A disrupt cooperativity without altering DNA binding (120). In DNMT3B, a Gln772Arg substitution in the TRD decreased its ability to perform multiple rounds of catalysis on a long DNA substrate (112). Future biochemical assays could determine the roles of the DNMT3A and 3B TRD regions in their respective kinetics mechanism. The TRD variants with disrupted kinetics mechanisms could be assayed for cooperativity or processivity using the methods described in our studies, and domain swapping between DNMT3A and DNMT3B would determine if the TRD is enough to drive kinetics mechanisms.

5.1.3 Post Translational Modifications to the DNMT3 Enzymes

The regulation of the DNMT3 enzymes by post translational modifications (PTMs) is poorly understood, however many PTMs of DNMT3A and DNMT3B have been mapped (281-285). About 70 phosphorylation sites have been mapped on DNMT3A and DNMT3B proteins, and sumoylation is known to occur on both enzymes. The most well characterized PTM is the phosphorylation of DNMT3A at Ser386 and Ser389, located near the PWWP domain. Casein kinase 2 (CK2) phosphorylates these sites, which increases heterochromatin targeting and decreases methyltransferase activity (286). It remains unknown how the phosphorylation of these sites alters localization and catalytic activity, since these residues are not located in any of the functional domains. In human ESCs, there are three different groups of phosphorylation sites in DNMT3A and DNMT3B (281, 283). The first group of phosphorylation sites are maintained as cells undergo differentiation, while the second class either appear or disappear during differentiation. In the second group, the method of differentiation determines which sites are phosphorylated. The third group includes phosphorylation sites that occur during differentiation regardless of the method used. There is only one phosphorylation site in the third group, DNMT3B Thr383, located adjacent to the putative nuclear localization signal. Further studies could investigate the changes in the biochemical properties of DNMT3A and DNMT3B due to phosphorylation, characterize their role during differentiation, and identify the kinases/phosphatases responsible for modifying these sites.

The functional role of DNMT3A and DNMT3B sumoylation has been documented, although the specific sites for these modifications remains unknown. DNMT3A is sumoylated by Small Ubiquitin Modifier 1 (SUMO-1) in the N-terminus, which disrupts its interaction with histone deacetylases (HDAC1/2) (284). This abolishes the ability of DNMT3A to repress transcription, possibly because it can no longer be recruited to those regions (252). DNMT3B is also modified by SUMO-1 in its N-terminal region (1-432) and interacts with SUMO-conjugating UBC9, however the biological significance of this modification remains unknown (285). Multiple modified forms of DNMT3B were detected through *in vitro* assays, suggesting that DNMT3B is poly-sumoylated. Further studies could map the sumoylation sites of DNMT3A and DNMT3B, and identify the biological roles of DNMT3B sumoylation.

5.2 Biotin-Avidin ELISA Assay for RNA Methyltransferase Activity

Recent discoveries and world-wide events have made it necessary to study RNA methyltransferase activity and characterize the enzymes that catalyze these reactions. The detection RNA modifications like N⁶-methyladenosine in mammals opened up a new field of epitranscriptomics to understand the role of RNA methylation in cell biology (287). N⁶-methyladenosine can modulate the binding affinities of RNA binding proteins and regulate mRNA stability. The multicomponent complex that catalyzes N⁶-methyladenosine contains two poorly characterized RNA methyltransferases, methyltransferase-like 3 (METTL3) and methyltransferase-like 14 (METTL14) (288). These enzymes play important roles in embryo development and stem cell self-renewal, making it necessary to understand their biochemical properties to target and catalyze N⁶-methylation of mRNA transcripts.

In the last five years, there have been two outbreaks of RNA viruses that caused world-wide pandemics. The 2015 pandemic was caused by the Zika virus, a member of the *flaviviridae* family that includes Dengue Fever and West Nile that is transmitted primarily through mosquitos (289). While many patients have only mild symptoms, Zika virus can spread from a pregnant woman to her baby, resulting in microcephaly and other birth defects. In rare cases, Zika virus infection in adults can cause Guillain-Barre syndrome, an immune dysfunction that leads to peripheral nerve damage. The 2020 pandemic was caused by the Coronavirus, a member of the *coronaviridae* family that results in a severe acute respiratory syndrome (290). This virus is transmitted primarily through respiratory droplets produced from coughs or sneezes and can have long term cardiovascular complications (291, 292). There are no vaccines against either Zika or Coronavirus, which has prompted researchers to investigate therapeutics against these viruses. Both Zika and Coronaviruses encode for RNA methyltransferases that are necessary for viral propagation inside the host cell (293-296). RNA methylation occurs on internal nucleotides of the single-stranded RNA genome and at the 5'-guanosine cap to facilitate genome stability and prevent host recognition. Viral knockouts of the RNA methyltransferases allow for host identification and degradation of the viral genome, making these enzymes a potential therapeutic targets.

Conventional methods to study RNA methyltransferases use diethylaminoethyl (DEAE) cellulose filter binding assays. The methyl group from ³[H] labeled AdoMet is transferred to the RNA substrate, which is bound to the DEAE filter (171). The negative charge of the phosphate backbone forms ionic interactions with the positively charged DEAE. The unused ³[H] AdoMet is

washed off and radioactivity on the bound RNA is measured by liquid scintillation counting. There are several disadvantages in this method which include high background and tedious processing of multiple samples (153). Previous studies on RNA methyltransferases used the conventional DEAE filter binding assay to characterize the enzyme's biochemical properties and screen potential inhibitors, however low processing power of the assay has limited screening (296)

The biotin-avidin ELISA assay to measure DNA methylation is a medium throughput technique with low background levels that utilizes a biotin-tagged DNA substrate and $^3\text{[H]}$ -labeled AdoMet (148). It has been widely used to characterize the biochemical properties of both bacterial and mammalian DNMTs due to its ease of use and reproducibility. We postulated that the biotin-avidin ELISA microplate assay could be adapted to measure RNA methylation by tagging the 3' end of an RNA substrate with biotin. We modified the buffers and method to ensure stability of RNA substrate during the processing of samples. This included use of RNase inhibitor in the reaction mixture and use of RNase free water for making reagents. We performed preliminary experiments to validate this technique using the recombinant Zika Virus RNA MTase, and found that the background radiation levels in the microplate assay are reduced by 90% compared to the DEAE filter assay. We determined the Michaelis-Menten constants of the Zika Virus RNA MTase (NS5) and found that it binds to RNA in the nano-molar range. Additionally, we performed product inhibition assays with AdoHcy and other potential inhibitors of NS5. The microplate assay was able to measure low levels RNA methyltransferase activity than previously reported (296). The biotin-avidin microplate assay could also be used to measure 5'-guanosine cap methylation and to screen potential small molecule inhibitors of other viral MTases.

APPENDIX A. VEZF1

*** The transcription factor Vezf1 represses the expression of the antiangiogenic factor Cited2 in endothelial cells**

*Originally published in *J. Biol. Chem.*

†AlAbdi, L., †He, M., Yang, Q. Y., **Norvil, A. B.**, and Gowher, H. (2018) The transcription factor Vezf1 represses the expression of the antiangiogenic factor Cited2 in endothelial cells. *J Biol Chem* 293, 11109-11118

†Co-first author

Declaration of collaborative work

Lama AlAbdi, Ming He, and Dr. Humaira Gowher wrote the manuscript.

Lama AlAbdi, Ming He, Qianyi Yang, Allison Norvil, and Dr. Humaira Gowher performed experiments.

Dr. Humaira Gowher performed the Microarray experiment.

Lama AlAbdi performed Vezf1 ChIP-qPCR experiments, Western blot, and RT-qPCRs.

Ming He performed the remaining ChIP-qPCRs, DNA methylation analysis, and endothelial lineage differentiation and RT-qPCR at increasing concentrations of VEGF1.

Qianyi Yang generated the Cited2 knockdown cell lines, Alkaline Phosphatase staining experiments, and endothelial lineage differentiation.

Ming He and Qianyi Yang performed tube formation assays.

Allison Norvil performed cloning, protein purifications and PCRs.

*** Characterization of Small Molecules Inhibiting the Pro-Angiogenic Activity of the Zinc Finger Transcription Factor Vezf1**

*Originally published in *Molecules*.

He, M.; Yang, Q.; Norvil, AB; Sherris, D.; Gowher, H. Characterization of Small Molecules Inhibiting the Pro-Angiogenic Activity of the Zinc Finger Transcription Factor Vezf1. (2018) *Molecules* 23, 7, 1615.

Declaration of collaborative work

Humaira Gowher and David Sherris designed the experiments

Ming He, Qianyi Yang, and Allison Norvil performed the experiments.

Humaira Gowher wrote the manuscript.



The transcription factor *Vezf1* represses the expression of the antiangiogenic factor *Cited2* in endothelial cells

Received for publication, March 13, 2018, and in revised form, May 19, 2018. Published, Papers in Press, May 24, 2018, DOI 10.1074/jbc.RA118.002911

Lama AlAbdi^{†1}, Ming He^{†1}, Qianyi Yang^{‡2}, Allison B. Norvil[‡], and Humaira Gowher^{‡§3}

From the [†]Department of Biochemistry and [‡]Purdue University Center for Cancer Research, Purdue University, West Lafayette, Indiana 47907

Edited by Joel Gottesfeld

Formation of the vasculature by angiogenesis is critical for proper development, but angiogenesis also contributes to the pathogenesis of various disorders, including cancer and cardiovascular diseases. Vascular endothelial zinc finger 1 (*Vezf1*), is a Krüppel-like zinc finger protein that plays a vital role in vascular development. However, the mechanism by which *Vezf1* regulates this process is not fully understood. Here, we show that *Vezf1*^{−/−} mouse embryonic stem cells (ESC) have significantly increased expression of a stem cell factor, Cbp/p300-interacting transactivator 2 (*Cited2*). Compared with WT ESCs, *Vezf1*^{−/−} ESCs inefficiently differentiated into endothelial cells (ECs), which exhibited defects in the tube-formation assay. These defects were due to reduced activation of EC-specific genes concomitant with lower enrichment of histone 3 acetylation at Lys²⁷ (H3K27) at their promoters. We hypothesized that overexpression of *Cited2* in *Vezf1*^{−/−} cells sequesters P300/CBP away from the promoters of proangiogenic genes and thereby contributes to defective angiogenesis in these cells. This idea was supported by the observation that shRNA-mediated depletion of *Cited2* significantly reduces the angiogenic defects in the *Vezf1*^{−/−} ECs. In contrast to previous studies that have focused on the role of *Vezf1* as a transcriptional activator of proangiogenic genes, our findings have revealed a role for *Vezf1* in modulating the expression of the antiangiogenic factor *Cited2*. *Vezf1* previously has been characterized as an insulator protein, and our results now provide insights into the mechanism, indicating that *Vezf1* can block inappropriate, nonspecific interactions of promoters with *cis*-located enhancers, preventing aberrant promoter activation.

Development of a proper vascular system is indispensable for embryogenesis. Accurate spatial and temporal control of gene

expression is required in endothelial cells (ECs),⁴ which are committed to the formation of the vasculature (1, 2). Angiogenesis involves migration, growth, and differentiation of ECs and takes place during development as well as in adulthood. Angiogenesis is regulated by an interplay between pro- and anti-angiogenic factors (3). Hif-1 α is a major pro-angiogenic factor, which interacts with p300/CBP and activates the expression of a number of pro-angiogenic genes including *VEGF*, thus initiating new vessel formation. Treatment of cultured ECs with VEGF-A₁₆₅ induces Hif-1 α expression, suggesting a bidirectional stimulatory relationship between VEGF and Hif-1 α (4). *In vivo*, Hif-1 α can be induced by a variety of factors including hypoxic conditions, (5–7), and certain cytokines and growth factors under normoxic conditions (8–12). Among many known factors, *Vezf1* (DB1/Bgp1) and *Cited2* (Mrg1/p35srj) play important roles in regulation of angiogenesis during development and in adulthood.

Cited2 (Cbp/p300-interacting transactivator with Glu/Asp-rich carboxyl-terminal domain 2) also named Mrg1/p35srj is a ubiquitously expressed essential transcriptional regulator that binds strongly to the histone acetyltransferases p300 and CBP (cAMP-responsive element-binding protein). *Cited2* plays a critical role in heart development, neurulation, and maintenance of fetal and adult hematopoietic stem cells. It is expressed throughout early embryogenesis and in embryonic stem cells (ESCs) (13–17, 19–21). Several studies have demonstrated that by competing with Hif-1 α to bind CBP/P300, *Cited2* prevents the activation of pro-angiogenic genes such as *VEGF*, and inhibits angiogenesis (22–24). For example, over-expression of *Cited2* suppresses *VEGF* promoter activity, and siRNA knockdown of *Cited2* increases *VEGF* promoter activity (25). The expression levels of Hif-1 α responsive genes including *VEGF* is increased in *Cited2*^{−/−} embryos (14). Another study showed that *Cited2* is a negative regulator of fracture healing, and its expression is inversely related to the expression of genes involved in extracellular matrix remodeling and angiogenesis, such as matrix metalloprotease, *VEGF*, and Hif-1 α (26). These studies suggest that the proper up-regulation of pro-angiogenic genes requires the levels of *Cited2* to be tightly controlled espe-

This work was supported by American Heart Association (AHA) Scientist Development Award 17SDG33700153 (to H.G.) a Graduate Student Fellowship from the King Saud University (KSU) (to L.A.) and NCI, National Institutes of Health Grant P30 CA023168. The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

✉ Author's Choice—Final version open access under the terms of the Creative Commons CC-BY license.

This article contains Fig. S1 and Table S1.

¹ Both authors contributed equally to the results of this work.

² Present address: Dept. of Anesthesiology, Washington University School of Medicine, 660 S. Euclid Ave., St. Louis, MO 63110.

³ To whom correspondence should be addressed. Tel: 301-820-2794; E-mail: hgowher@purdue.edu.

⁴ The abbreviations used are: EC, endothelial cell; VEGF, vascular endothelial growth factor; *Cited2*, Cbp/p300-interacting transactivator with Glu/Asp-rich carboxyl-terminal domain 2; CBP, cAMP-responsive element-binding protein; ESC, embryonic stem cell; *Vezf1*, vascular endothelial zinc finger 1; LIF, leukemia inhibitory factor; IN, input; IP, immunoprecipitated, qPCR, quantitative PCR; Gapdh, glyceraldehyde-3-phosphate dehydrogenase; D, days post-differentiation.

Cited2 causes angiogenesis defects in *Vezf1* null cells

cially at the onset of angiogenesis. Because its aberrant high expression in a specific developmental window can inhibit Hif-1 α activity, this regulation can potentially be critical for angiogenesis in both embryonic stages and adulthood.

Mammalian *Vezf1* is an essential transcription factor, which is expressed in the anterior-most mesoderm at E7.5 during development. Its expression is later restricted in the vascular endothelium, an observation that revealed its role in regulation of angiogenesis. *Vezf1* null mice die at approximately E9.5 (27). *Vezf1*^{-/-} ESCs grow slower and make smaller embryoid bodies, which have defects in vascularization and cease to grow a few days post-differentiation (28, 29). *Vezf1* is expressed in both adult and embryonic ECs. Blocking the activity of *Vezf1* by small molecule inhibitor Vec6 inhibits wound healing suggesting its role in postnatal angiogenesis (30). *Vezf1* contains six Cys₂/His₂-type zinc finger motifs and binds poly(dG) or poly(dC) sequences (31, 32). It carries a glutamine-rich stretch and a proline-rich region that are characteristic of transcriptional activation or repression domains (33). It is proposed to act as a transcriptional activator of pro-angiogenic genes including *endothelin 1*, *microtubule turnover protein*, *stathmin/OP18*, and *metallothionein 1 (MT1)* (34–36). However, no change in the expression of pro-angiogenic genes was seen in *Vezf1*^{-/-} embryos (28). Other studies suggested an indirect role of *Vezf1* by interacting with RhoB that promotes expression of RhoB-regulated pro-angiogenesis genes (30, 37). Therefore, the mechanism by which *Vezf1* regulates angiogenesis is unclear.

To study the specific role of *Vezf1* in endothelial development and angiogenesis, we examined the differentiation of WT and *Vezf1*^{-/-} ESCs to ECs by treatment with VEGF-A₁₆₅ and tested their angiogenic potential by a *in vitro* tube-formation assay. Our findings suggest that *Vezf1* controls activation of angiogenesis in ECs by restricting *Cited2* expression to basal levels, which allows Hif-1 α -mediated activation of the pro-angiogenic genes. We observed a strong increase in the expression of *Cited2* in *Vezf1*^{-/-} ESCs compared with WT cells. In addition, the elevated expression of *Cited2* in *Vezf1*^{-/-} ESCs affected their efficiency of differentiation into ECs and attenuated the ability of *Vezf1*^{-/-} ECs to form vascular structures in a tube-formation assay. Concomitant with this, there was reduced activation of endothelial/pro-angiogenic genes in differentiating *Vezf1*^{-/-} ECs. Based on our data showing reduced levels of histone H3K27Ac at the promoters of angiogenesis-specific genes, we propose that high levels of *Cited2* sequester the histone acetyltransferase p300 away from angiogenesis-specific gene promoters, thus reducing their activation and gene expression. Together these observations substantiate the critical role of *Vezf1* in controlling the expression of developmental regulators such as *Cited2*. Given that the expression of *Cited2* in ESCs is not completely turned off, we suggest the role of *Vezf1* in fine-tuning *Cited2* expression in ESCs. Our previous work using genome-wide ChIP-SEQ showed that binding sites for *Vezf1* are mostly present in CpG-rich regions (38). We also showed that *Vezf1* binds to the chicken β -globin insulator suggesting a role in regulating enhancer-mediated control of gene expression (32, 38). Based on these studies, we speculate that the insulator function of *Vezf1* blocks inappropriate interac-

tions of the *Cited2* promoter with nearby enhancer/s, thus modulating the magnitude and spatiotemporal regulation of its expression.

Results

Cited2 expression is high in *Vezf1*^{-/-} ESCs

To elucidate the mechanism of *Vezf1*, we had previously analyzed changes in gene expression of *Vezf1*^{-/-} ESCs compared with WT ESCs using a microarray analysis (Fig. S1) (39). We found *Cited2* among the top 20 genes that were up-regulated in *Vezf1*^{-/-} ESCs by more than 5-fold. To confirm this observation, we measured the gene expression of *Cited2* quantitatively in WT and *Vezf1*^{-/-} ESCs by qRT-PCR and protein levels by Western blotting. The data show a 4–5-fold higher transcript and protein levels of *Cited2* in *Vezf1*^{-/-} ESCs compared with WT ESCs (Fig. 1, A and B). We found these observations consistent with our previously published ChIP-SEQ of Ser² phosphorylated RNA Pol II (elongating form of RNA Pol II) in WT and *Vezf1*^{-/-} ESCs (38). Data analysis of elongating RNA Pol II showed more than a 2-fold higher enrichment in the *Cited2* gene body in *Vezf1*^{-/-} ESCs compared with the WT ESCs (Fig. 1C). Based on the known function of *Cited2* as an anti-angiogenic factor (22–24), we hypothesized that a presence of high *Cited2* level in *Vezf1*^{-/-} cells could interrupt or delay the differentiation of ECs, or reduce their angiogenic potential.

Vezf1^{-/-} ESCs are defective in EC differentiation

Cited2 expression is critical for pluripotency and differentiation of ESCs; therefore, we tested if overexpression of *Cited2* affected the pluripotency of *Vezf1*^{-/-} ESCs. We quantified the expression of the pioneer transcription factor, Oct4, in *Vezf1*^{-/-} ESCs compared with that in the WT ESCs and observed no significant difference in its expression (Fig. 2A). Additionally, both WT and *Vezf1*^{-/-} ESCs showed positive alkaline phosphate staining suggesting that high expression of *Cited2* had little if any effect on the pluripotency of *Vezf1*^{-/-} ESCs (Fig. 2B). Previous studies showed a reduced growth of *Vezf1*^{-/-} EBs and defects in their vascular structures, but reported little or no difference in the endothelial differentiation in 3D cultures (27). We performed *in vitro* differentiation of WT and *Vezf1*^{-/-} ESCs to ECs on gelatinized plates in the presence of 10 ng/ μ l of VEGF-A₁₆₅. The differentiation of ESCs to ECs was monitored by microscopy. During differentiation, the WT ESCs showed an expected loss of ~5–10% of cells, and differentiated efficiently into ECs. Comparatively, during the first 3 days of differentiation, over 80% *Vezf1*^{-/-} cells died, leading to a reduced efficiency of EC derivation from the *Vezf1*^{-/-} ESCs. This was confirmed by positive alkaline phosphatase staining of the surviving *Vezf1*^{-/-} cells at 10 days post-differentiation (Fig. 2C).

Given that the expression of the EC lineage is driven by VEGF-A in an autoregulatory loop (9, 40), we tested if increasing the concentration of VEGF-A₁₆₅ in the medium could improve EC differentiation and survival of *Vezf1*^{-/-} cells. EC differentiation was monitored by measurement of endothelial-specific gene expression including *VEGF-A*, *Flk1*, *Flt1*, *CD31*, and *Tie2* (41). WT and *Vezf1*^{-/-} ESCs were differentiated using 20, 40, and 60 ng/ μ l of VEGF-A₁₆₅. Increasing the VEGF-

Cited2 causes angiogenesis defects in *Vezf1* null cells

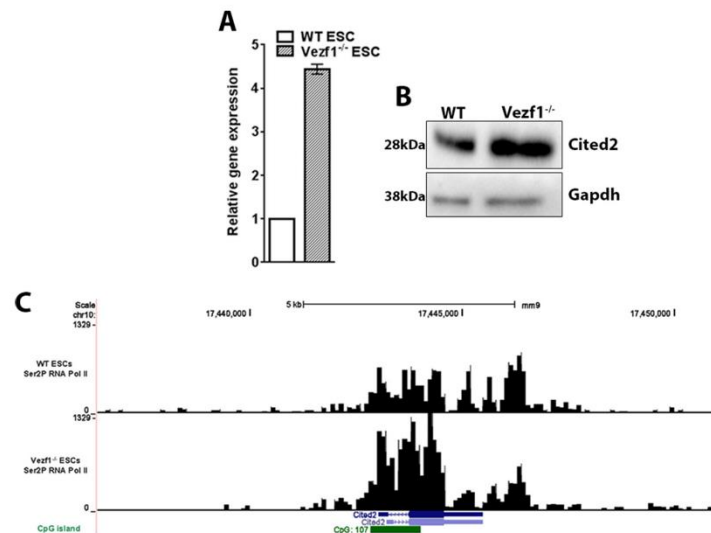


Figure 1. A, gene expression analysis of *Cited2* by RT-qPCR in WT and *Vezf1*^{-/-} ESCs showed a 4–5-fold higher expression of *Cited2* in *Vezf1*^{-/-} ESCs. Gene expression was normalized to Gapdh and represented relative to gene expression in WT ESC, set to 1. B, 50 μ g of total cell extract were loaded from the WT and *Vezf1*^{-/-} ESCs followed by Western blot analysis using anti-Cited2 antibody and anti-GAPDH antibody for loading control. C, screen shot from UCSC genome browser showing the occupancy of Ser²-phosphorylated Pol II at the *Cited2* locus in WT and *Vezf1*^{-/-} ESCs. These data were obtained from the previously published publicly available ChIP-SEQ data. WT, wildtype ESCs; *Vezf1*^{-/-}, *Vezf1*-knockout ESCs.

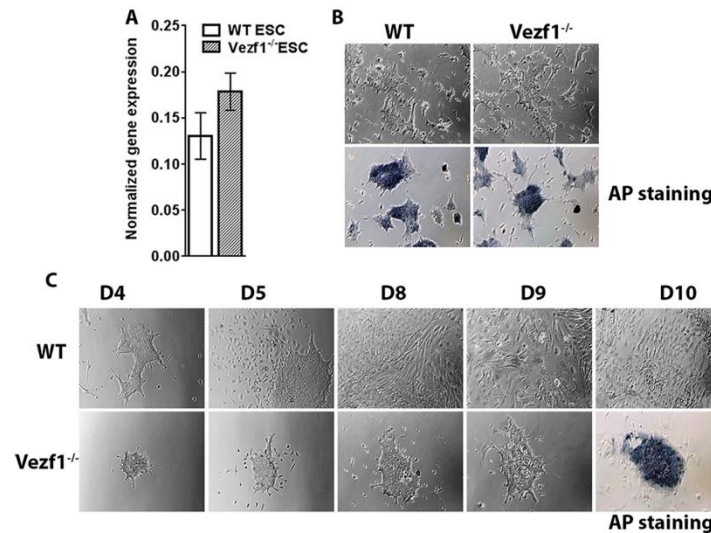


Figure 2. A, gene expression analysis by RT-qPCR of Oct4 in WT and *Vezf1*^{-/-} ESCs. B, alkaline phosphatase (AP) staining for pluripotency in WT and *Vezf1*^{-/-} ESCs; the presence of dark blue stain indicates positive for pluripotency. C, WT and *Vezf1*^{-/-} ESCs were differentiated using 10 ng/ μ l of VEGF-A₁₆₅ for 5, 6, 7, and 10 days and visualized using brightfield microscopy at $\times 10$ magnification. Unlike the WT cells, the *Vezf1*^{-/-} cells were unable to differentiate and most of the cells died. The field view is the representation of proliferation during differentiation and cell number. The panel for D10 shows alkaline phosphatase staining. A strong signal of *Vezf1*^{-/-} cells indicates presence of undifferentiated stem cells. WT, wildtype ESCs; *Vezf1*^{-/-}, *Vezf1* knockout ESCs; UD, undifferentiated; D4–D10, days post-differentiation.

A₁₆₅ to 20 ng/ μ l stimulated *Vezf1*^{-/-} cells to differentiate to ECs with a higher efficiency (Fig. 3A). In both WT and *Vezf1*^{-/-} ESCs, differentiation induced repression of the pluripotency gene, Oct4, and activation of endothelial-specific genes, VEGF-A, *Fli1*, *CD31*, and *Tie2*, albeit at lower levels in the

Vezf1^{-/-} cells (Fig. 3B). Reduced activation of *CD31* and *Tie2* in *Vezf1*^{-/-} cells suggested a partial defect in EC differentiation. WT and *Vezf1*^{-/-} ECs day (D) 6 and D8 post-differentiation treated with 20, 40, and 60 ng/ μ l of VEGF-A₁₆₅ were collected for gene expression analysis by qRT-PCR. The data

Cited2 causes angiogenesis defects in *Vezf1* null cells

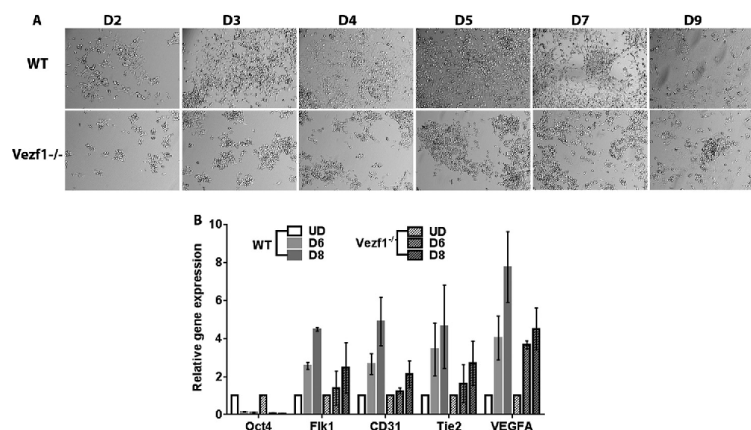


Figure 3. A and B. WT and *Vezf1*^{-/-} ESCs were differentiated using 20 ng/μl of VEGF-A₁₆₅. A, differentiating WT and *Vezf1*^{-/-} ECs were visualized using brightfield microscopy. Whereas most of the WT cells undergo distinct morphological changes, only a fraction acquire a similar morphology in *Vezf1*^{-/-} cells. B, gene expression by RT-qPCR plotted as a relative change to the expression in UD where UD was set to 1. Endothelial specific genes, *Flk1*, *VEGF-A*, *CD31*, and *Tie2* show an expected increase in expression in differentiating WT ECs. Differentiating *Vezf1*^{-/-} ECs, however, show significantly low expression of all tested endothelial-specific genes. A decrease in *Oct4* expression is observed in both WT and *Vezf1*^{-/-} ECs indicating loss of pluripotency. WT, wildtype; *Vezf1*^{-/-}, *Vezf1* knockout; UD, undifferentiated; D2–D9, days post-differentiation.

showed that at all three concentrations of VEGF-A₁₆₅, the expression of *VEGF-A* and its receptor, *Flk1*, was significantly higher in WT ECs compared with that in the *Vezf1*^{-/-} ECs, indicating that gene expression is not further rescued by higher doses of VEGF-A₁₆₅ (Fig. 4). Previous studies have shown that *Hif-1α* expression can also be activated by treatment of ECs with VEGF-A₁₆₅ thus showing that VEGF regulates the expression of its own transcription factor (4). In WT and *Vezf1*^{-/-} ECs, we checked the expression of *Hif-1α* and *Flt1*, which is the angiogenesis-specific VEGF receptor. Although no difference in the expression of *Hif-1α* was observed between WT and *Vezf1*^{-/-} ECs, similar to *Flk1* and *VEGF-A*, *Flt1* expression was also comparatively lower in *Vezf1*^{-/-} ECs at 20, 40, and 60 ng/μl of VEGF-A₁₆₅ in both D6 (Fig. 4, A–D) and D8 (Fig. 4, E–H) post-differentiation. These data show that in *Vezf1*^{-/-} ECs, reduced expression of angiogenic genes including VEGF is not due to lower *Hif-1α* expression, but potentially due to its reduced activity.

Vezf1^{-/-} ECs are defective in forming vascular networks in 3D cultures

We next tested the angiogenic potential of *Vezf1*^{-/-} ECs by an *in vitro* tube-formation assay. We differentiated WT and *Vezf1*^{-/-} ESCs to ECs for 8 days in the presence of VEGF-A₁₆₅ at 20 ng/μl. The ECs were collected and placed in Matrigel to form vascular networks or tubes in 3D culture. The mouse endothelial cell line, MSS31, was used as a positive control. Whereas WT ECs formed distinct tubes within 4–6 h in Matrigel, *Vezf1*^{-/-} ECs showed significant defects in tube formation as indicated by the shorter tube length (Fig. 5, A and B).

Taken together, these data show that *Vezf1*^{-/-} ESCs have reduced competence to differentiate into ECs and to form vascular structures in Matrigel. Given that *Flt1* receptor function is required for tubulogenesis (42), the inability of *Vezf1*^{-/-} ECs to form tubes in Matrigel could be the consequence of strongly

reduced expression of *Flt1* in these cells. Because no change was observed in *Hif-1α* gene expression, we predict that the reduced expression of EC-specific genes is due to high *Cited2* expression in *Vezf1*^{-/-} ESCs and ECs.

Induced repression of *Cited2* partially rescues EC differentiation and vascular defects in *Vezf1*^{-/-} ECs

We next tested the hypothesis that the defective vasculature formation by *Vezf1*^{-/-} ECs is due to anomalous high expression of *Cited2*. We therefore asked if depletion of *Cited2* in *Vezf1*^{-/-} ESCs could rescue their ability to differentiate and make vascular networks. *Vezf1*^{-/-} ESCs were transfected with *Cited2* shRNA, to generate stable transgenic ESCs lines, *Vezf1*^{-/-sh}. Of the nine transgenic lines, some showed more than 10-fold reduction in *Cited2* expression when compared with WT ESCs. Because *Cited2* is known to be important for pluripotency (43), we chose to use the *Vezf1*^{-/-sh} cell line (7-2), in which *Cited2* expression is reduced to levels similar to that of WT ESC (Fig. 6, A and B). The *Vezf1*^{-/-sh} ESCs were differentiated to ECs using 20, 40, and 60 ng/μl of VEGF-A₁₆₅. Compared with *Vezf1*^{-/-} ESCs, *Vezf1*^{-/-sh} ESCs showed better survival and higher efficiency of EC differentiation at 20 ng/μl of VEGF-A₁₆₅, which was similar to the WT cells (Fig. 6C). The *Vezf1*^{-/-sh}-derived ECs were collected on day 6 post-differentiation and RNA was extracted for gene-expression analysis. The data show that expression of VEGF-A was largely rescued, whereas *Flt1* and *Flk1* were partially rescued when compared with their expression in WT and *Vezf1*^{-/-} ECs (Fig. 6D). Next, we tested if the derived ECs form vascular structures by a tube-formation assay. WT, *Vezf1*^{-/-}, and *Vezf1*^{-/-sh}-derived ECs ((7-2) and (3-3)) were collected after 10 days and used to perform an *in vitro* tube-formation assay. *Vezf1*^{-/-sh} (3-3) ECs have ~7-fold reduced expression of *Cited2* compared with WT ESCs. This cell line was therefore used to test the effect of *Cited2* deficiency on EC differentiation and tube formation. The repression of

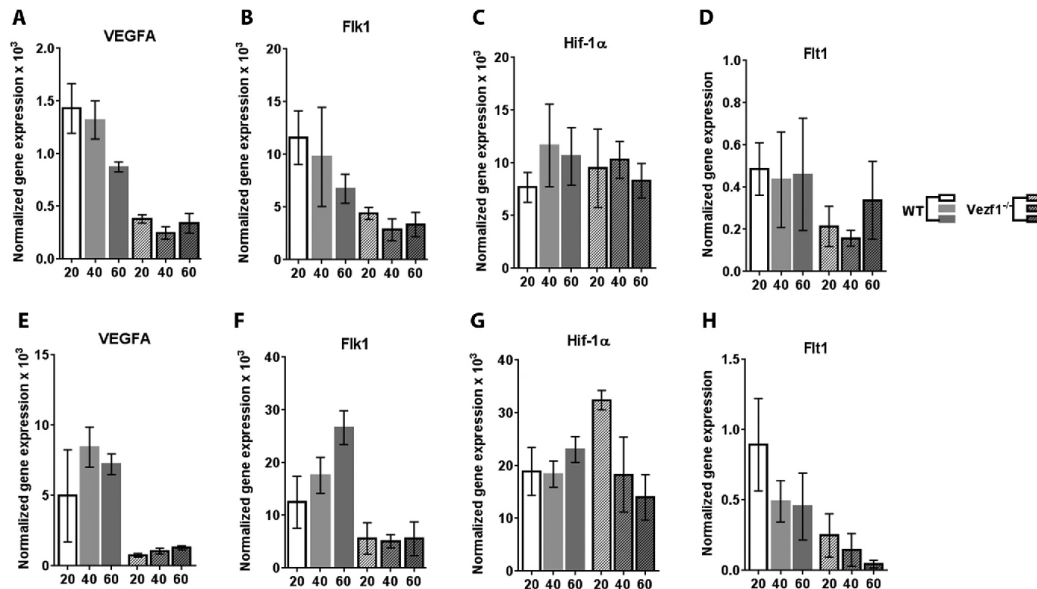


Figure 4. A–D, gene expression analysis in WT and *Vezf1*^{−/−} D6 ECs, differentiated using 20, 40, and 60 ng/μl of VEGF-A₁₆₅. E–H, gene expression analysis in WT and *Vezf1*^{−/−} D8 ECs differentiated using 20, 40 and 60 ng/μl of VEGF-A₁₆₅. The expression of all genes is increased in D8 compared with D6 post-differentiation. Higher doses of VEGF-A₁₆₅ have no further effect on expression of VEGF-A, Flk1, Hif-1α, and Fit1 in both WT or in *Vezf1*^{−/−} cells. The data represents average and S.D. of 3 to 4 replicates. WT, wildtype; *Vezf1*^{−/−}, *Vezf1* knockout; UD, 20, 40, 60 ng/μl of VEGF-A₁₆₅ used for differentiation.

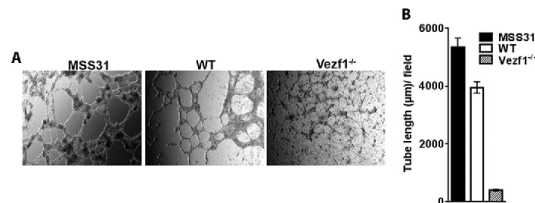


Figure 5. A, differentiated WT and *Vezf1*^{−/−} EC plated in VEGF-supplemented Matrigel were incubated at 37 °C for 5–15 h. The formation of tube structures is visualized by brightfield microscopy. Mouse endothelial cell line MSS31 is used as a positive control. The images were taken at ×10 magnification at 12 h for MSS31, and 6 h for WT and *Vezf1*^{−/−}. B, measurement of tube length using ImageJ software. Compared with MSS31 and WT ECs, *Vezf1*^{−/−} ECs were unable to make tube-like structures in Matrigel. WT, wildtype; *Vezf1*^{−/−}, *Vezf1* knockout.

Cited2 in *Vezf1*^{−/sh} ECs largely rescued the defective tube formation on Matrigel, which was more prominent in *Vezf1*^{−/sh} (7–2) compared with the (3–3) cell line (Fig. 7, A and B). This observation supports the known role of *Cited2* in pluripotency and differentiation of ESCs and emphasizes the importance of the appropriate levels of developmental transcription factors for proper differentiation. These data directly support our hypothesis that an aberrant high expression of *Cited2* prevents the activation of EC-specific gene expression potentially by sequestering p300 from the promoters of angiogenesis-specific genes.

P300 activity is regulated by *Cited2* at the VEGF-A promoter

In response to VEGF signaling, P300 acetyltransferase interacts with Hif-1α, which targets it to the HBS (HIF-1-binding

element) of the promoters of angiogenesis-specific genes where it acetylates histone H3 at Lys²⁷. To test the impact of *Cited2* expression on the activity of P300 histone acetyltransferase at *VEGF* and *Flk-1* promoters, we performed a chromatin immunoprecipitation (ChIP) assay using anti-histone H3K27Ac antibody. We observed an expected increase in the fold-enrichment of H3K27Ac at VEGF and Flk-1 promoters in WT-differentiated ECs compared with the undifferentiated ESCs. However, this increase was markedly reduced in the differentiated *Vezf1*^{−/−} ECs particularly at the *VEGF* promoter, which is a direct target of Hif-1α. As a control, we used the *Oct4* promoter where H3K27 acetylation is decreased post-differentiation in both the WT and *Vezf1*^{−/−} ESCs (Fig. 8A). Although *Flk-1* has an HBS in its promoter, some studies indicate that it is targeted by both Hif-1α and HIF-1β (40). These data support our hypothesis that reduced P300 activity at angiogenesis-specific gene promoters inhibits their complete activation during *Vezf1*^{−/−} EC differentiation.

Transcriptional regulation of *Cited2* expression by *Vezf1*

Vezf1^{−/−} ESCs show a genome-wide loss of DNA methylation at several CpG islands (CpGi's) flanking tissue-specific genes and a significant decrease in the expression of the DNA methyltransferase, Dnmt3b (39). *Cited2* is encoded by a relatively small gene, which has a large CpGi at its promoter and exon 1 (Fig. 1C). We therefore asked if the increase in *Cited2* expression in *Vezf1*^{−/−} ESCs could be due to loss of DNA methylation at its CpGi. To test the potential role of Dnmt3b in regulation of *Cited2* CpGi DNA methylation, genomic DNA from Dnmt3b^{−/−} ESCs, *Vezf1*^{−/−} ESCs, and WT ESCs was

Cited2 causes angiogenesis defects in Vezf1 null cells

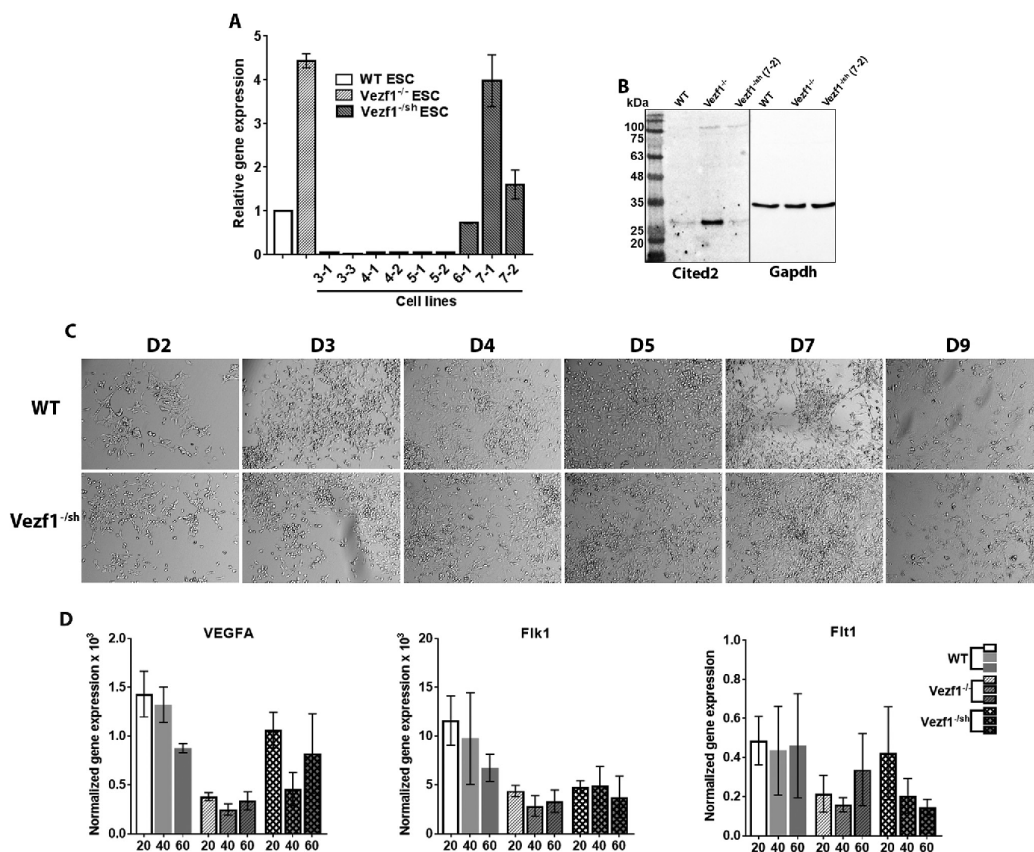


Figure 6. A, gene expression analysis of *Cited2* by RT-qPCR in *Vezf1*^{-/-} cell lines. Change in gene expression was plotted relative to that of WT ESCs, set to 1. Labels on the x axis represent various stable cell lines, of which the *Vezf1*^{-/-} cell line (7-2) has *Cited2* expression reduced to the levels similar to WT ESCs. B, Western blot analysis using 50 µg of total cell extract from the WT, *Vezf1*^{-/-}, and *Vezf1*^{-/-} (7-2) ESCs probed with anti-*Cited2* antibody and anti-GAPDH. C, differentiation of WT, *Vezf1*^{-/-}, and *Vezf1*^{-/-} (7-2) ESCs was induced using 20 ng/µl of VEGF-A₁₆₅. D2–D6 are days post-differentiation. Compared with the differentiating WT cells, *Vezf1*^{-/-} (7-2) show similar morphology and cell number indicating at least a partial rescue of their ability to differentiate into ECs. D, gene expression analysis of *Vezf1*^{-/-} (7-2) ESCs, which were differentiated to ECs at 20, 40, and 60 ng/µl of VEGF-A₁₆₅. Expression of pro-angiogenic genes *VEGFA*, *Flk1*, and *Flt1* was measured. Compared with WT and *Vezf1*^{-/-} cells, gene expression was largely rescued in *Vezf1*^{-/-} (7-2) ECs. The data represent the average and S.D. of 3 to 4 replicates. WT, wildtype ESCs; *Vezf1*^{-/-}, *Vezf1* knockout ESCs; *Vezf1*^{-/-} (7-2), stable transgenic *Vezf1*^{-/-} ESCs expressing *Cited2*-shRNA; 20, 40, 60, ng/µl of VEGF-A₁₆₅ used for differentiation.

extracted to quantify DNA methylation using bisulfite sequencing. Our data show very low CpG methylation at *Cited2* CpGi in WT ESCs that did not change in *Vezf1*^{-/-} and *Dnmt3b*^{-/-} ESCs (Fig. 8B). *Cited2* gene expression also showed no change in *Dnmt3b*^{-/-} compared with that in WT ESCs (Fig. 8C). These data confirm that expression of *Cited2* is not regulated by *Dnmt3b* or changes in DNA methylation at its CpGi and support the direct role of *Vezf1* in regulating *Cited2* expression. Therefore, we investigated the binding of *Vezf1* near the *Cited2* promoter by ChIP assay using custom-made rabbit polyclonal anti-*Vezf1* antibody, which was previously characterized and used in ChIP studies in ESCs (12, 34). We observed a high relative enrichment of *Vezf1* at the *Cited2* promoter in WT ESCs (Fig. 8D). These data suggest a direct regulation of *Cited2* expression by *Vezf1* through its binding at the promoter-associated CpGi.

Taken together our *in vitro* differentiation experiments show that the aberrant high expression of *Cited2* in *Vezf1*^{-/-} ESCs suppresses their angiogenic potential by sequestering P300/CBP away from the pioneer transcription factor Hif-1α. This abbreviates the promoter activation of the downstream pro-angiogenic genes (Fig. 8E). These findings suggest that *Vezf1* regulates angiogenesis by fine-tuning the level of anti-angiogenic factor *Cited2*.

Discussion

The transcription factor *Vezf1* is highly expressed in vascular endothelium and its role in vascular development has been observed by several earlier studies (28, 34, 36, 44, 45). For example, recent studies have shown that a small molecule inhibitor of *Vezf1*, *Vec6*, prevents wound healing and angiogenesis (30). Although the function of *Vezf1* has been suggested through its

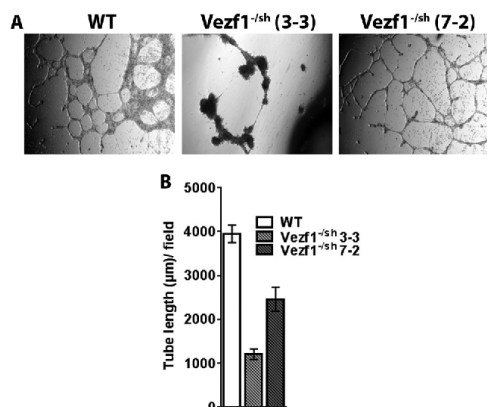


Figure 7. A, WT, *Vezf1*^{-/-sh} (7-2), and *Vezf1*^{-/-sh} (3-3) cells were differentiated for 10 days and used in a tube-formation assay. The *Vezf1*^{-/-sh} cell line (3-3) had about 7-fold lower *Cited2* expression than WT. Compared with the WT cells, tube formation was rescued in *Vezf1*^{-/-sh} (7-2) ECs, which was absent in *Vezf1*^{-/-sh} (3-3) ECs. Images were taken at 12 h for *Vezf1*^{-/-sh} (3-3) and 6 h for WT and *Vezf1*^{-/-sh} (7-2). B, tube length was measured by ImageJ software and plotted.

role as a transcriptional activator of some genes that are known to promote angiogenesis, previous studies were performed using semi-quantitative RT-PCR to measure the expression of pro-angiogenic genes in *Vezf1*^{-/-} embryos. These data showed no change in the expression of pro-angiogenic factors compared with WT embryos (28). Unlike the previous study (28), we used quantitative RT-PCR to measure gene expression changes and our data show about 2–3-fold lower expression of several pro-angiogenic genes, including *CD31*, *Tie2*, *VEGF-A*, and its receptors *Flk-1* and *Flt-1* in the *in vitro* derived *Vezf1*^{-/-} ECs. The defective EC differentiation of *Vezf1*^{-/-} ESCs was also supported by impaired morphological changes associated with EC differentiation. We further show that the expression of some of these genes can be largely rescued by down-regulating *Cited2*, which is aberrantly overexpressed in *Vezf1*^{-/-} ESCs. Our data support the previously suggested role of *Vezf1* in angiogenesis, however, through a different mechanism. In contrast to its previously predicted role as a transcriptional activator, our data show that *Vezf1* restricts the expression of the anti-angiogenic gene *Cited2* to basal levels, ensuring a balanced gene expression during angiogenesis. Our data also emphasize that small but quantifiable changes in gene expression of developmental transcription factors and regulators can have profound effect on cell differentiation.

Cited2 (Mrg1/p35srj) belongs to a family of transactivators that lack direct DNA binding but contain glutamic acid/aspartic acid (ED)-rich tail, which interacts with P300/CBP acetyltransferase. Whereas, on one hand, *Cited2* competes with Hif-1 α to interact with P300/CBP, *Cited2* promoter has HIF1-binding sites and its expression is up-regulated in hypoxia by Hif-1 α . Therefore, *Cited2* participates in a negative-feedback loop with Hif-1 α in which *Cited2* accumulates during hypoxia. During restoration of normoxia, it will inhibit Hif-1 α activity and prevent hypervascularization. An anomalous high expression of *Cited2* at the onset of angiogenesis could interfere with

Cited2 causes angiogenesis defects in *Vezf1* null cells

Hif-1 α -mediated activation of pro-angiogenic genes, consequently the expression of *Cited2* must be tightly controlled. We propose that in undifferentiated ESCs, *Vezf1* modulates the expression of *Cited2*, thus enabling Hif-1 α -mediated gene activation during angiogenesis. This mechanism is supported by our data showing that overexpression of *Cited2* causes defective tube formation by *Vezf1*^{-/-} ECs when there is no difference in Hif-1 α expression between the WT and *Vezf1*^{-/-} cells. *Cited2* is expressed throughout early embryogenesis and its expression in ESCs is critical for pluripotency, and appropriate differentiation (13, 43). Our data showing absence of rescue in a *Vezf1*^{-/-sh} (3-3) cell line, which has *Cited2* expression significantly lower than WT ESCs, supports the role of the basal expression of *Cited2* in maintenance of pluripotency and differentiation potential of ESCs. *Vezf1*-mediated regulation of *Cited2* expression is also potentially relevant during adult angiogenesis and wound healing where *Vezf1* could down-regulate or maintain low *Cited2* expression in the ECs. This prediction is supported by our observation from a published microarray analysis of *Vezf1*-silenced BVEC's, listing *Cited2* among the up-regulated genes (30). This study also showed that loss of *Vezf1* causes inhibition of wound healing and blood vessel formation (30).

Based on the previously characterized role of *Vezf1* as an insulator binding protein, it is highly plausible that *Vezf1* insulates its target promoters from interaction with nonspecific enhancers, and in the case of *Cited2* from the enhancers in the downstream gene, *β -taxilin* (*Txlnb*). The insulator function of *Vezf1* is supported by our published ChIP-SEQ studies showing that a significant number of *Vezf1*-binding sites are adjacent to insulator protein, CTCF. *Vezf1* shows widespread binding at CpG's present in the regulatory elements of genes including promoters, enhancers, and insulators (32, 38). By using *in vitro* EC differentiation as a developmental model system, it will be important to explore the regulatory potential of *Vezf1*-mediated insulator function in modulating gene expression during development.

Experimental procedures

Embryonic stem cell culture

Undifferentiated WT, *Dnmt3b*^{-/-}, and *Vezf1*^{-/-} ESCs were cultured in Dulbecco's modified Eagle's medium containing 15% ESC qualified fetal bovine serum (Millipore), supplemented with nonessential amino acids, glutamine, 1000 units/ml of leukemia inhibitory factor (LIF) (ESGRO; Chemicon International), and 50 μ M β -mercaptoethanol. Cells were cultured on 0.1% gelatin for one passage before switching to differentiation conditions.

Endothelial lineage differentiation

ESCs were plated at a density of 3×10^3 cells/cm² on gelatinized plates in ESC medium with LIF and incubated overnight to attach. The next day the media was removed and cells were washed with PBS. Differentiation was induced by adding ESC medium without LIF and VEGF-A₁₆₅ (R&D Systems) at 20–60 ng/ μ L. VEGF-A₁₆₅ was supplemented to the culture every alternate day for 10 days to drive differentiation into endothelial lineage. Cell morphology was monitored using phase-contrast

Cited2 causes angiogenesis defects in *Vezf1* null cells

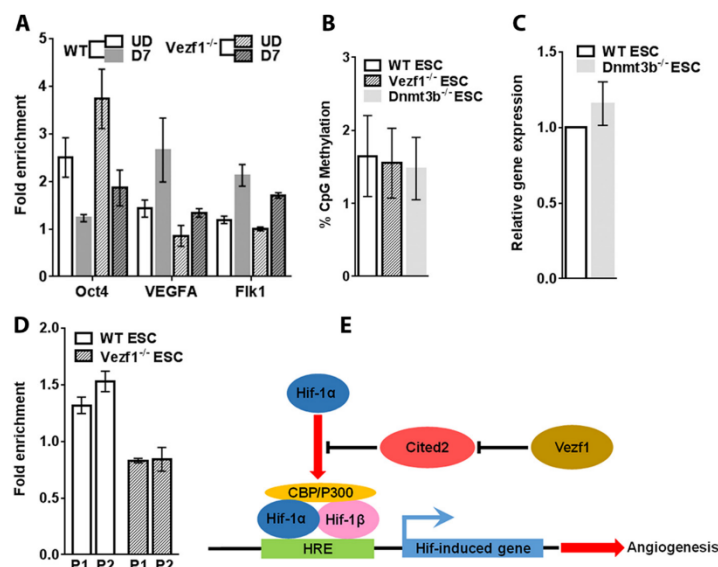


Figure 8. A, ChIP-qPCR of H3K27Ac showed an increase in the fold-enrichment of H3K27Ac at VEGF-A and Flk1 promoters in WT ECs that were reduced in *Vezf1*^{-/-} ECs. At the Oct4 promoter, post-differentiation, deacetylation of H3K27 is accompanied by the loss of gene expression. Therefore, it serves as a negative control. B, DNA methylation was analyzed at *Cited2* CpG using bisulfite sequencing in undifferentiated ESCs. The average percent methylation of 27 CpG sites in the CpG was plotted. C, gene expression analysis of *Cited2* in WT and Dnmt3b^{-/-} ESCs. There was no significant difference in *Cited2* expression levels. D, ChIP-qPCR using anti-Vezf1 antibody showed an increase in the fold-enrichment at *Cited2* promoter in WT ESCs. Fold-enrichment below 1 in *Vezf1*^{-/-} ESCs indicates absence of binding, and was used as negative control. The data in each bar graph represents average and S.D. of 3 to 4 replicates. E, model showing the effect of Cited2 on Hif-1α-mediated regulation of pro-angiogenic genes. Vezf1 regulates the expression of Cited2 at basal levels in ESCs before the induction of endothelial differentiation. This allows the interaction of Hif-1α and P300 that activate the pro-angiogenic genes. In the absence of Vezf1, high expression of Cited2 sequesters P300 away from Hif-1α, thus inhibiting the activation of pro-angiogenic genes. WT, wildtype; *Vezf1*^{-/-}, *Vezf1* knockout; *Dnmt3b*^{-/-}, Dnmt3b knockout; UD, undifferentiated; D7, days post-differentiation; P1 and P2, two primer pairs in *Cited2* promoter used in ChIP-qPCR.

microscopy and pluripotency was monitored by alkaline phosphatase staining.

Tube-formation assay

The ECs on D10 post-differentiation were collected by trypsinization and counted using Bio-Rad Cell Counter. The tube-formation assay was performed by plating 2×10^5 ECs on a 24-well plate coated with VEGF supplemented Matrigel (BD Biosciences) according to the manufacturer's protocol (46, 47). The cells were incubated at 37 °C for 3–18 h. Tubing was scored using images from phase-contrast microscopy (29). The length of the tubes was measured by ImageJ software.

Transfection and generation of stable ESC lines

The lentivirus construct pLKO.1 carrying shRNA specific for *Cited2* was purchased from Dharmacon. The recombinant lentivirus containing *Cited2* shRNA was packaged using Vira Power (Fisher Thermo Scientific) in 293FT cells using the manufacturer's protocol. For shRNA-mediated depletion of *Cited2*, WT, and *Vezf1*^{-/-} ESCs were transfected by lentivirus at a multiplicity of infection of 2 followed by selection of transgenic lines with stably integrated lentivirus construct using 3 μg/ml of puromycin.

Gene expression by quantitative RT-PCR and Western blotting

RNA from cells was purified using TRIzol (Invitrogen, 15596026) according to the manufacturer's protocol. Genomic

DNA contamination was removed by DNase (Roche Applied Science, 04716728001) treatment at 37 °C overnight. Quantitative RT-PCR was performed for equal amounts of RNA by using Verso One-Step kit (Thermo Scientific, AB-4104A). The data were analyzed and gene expression was normalized to *Gapdh* expression. The change in expression is represented either as normalized gene expression or as relative gene expression that is changed relative to expression in the undifferentiated cells, set to 1. See Table S1 for primers used. Western blot analysis was performed using commercially available antibodies anti-Cited2 (ab108345, Abcam) and anti-Gapdh (sc47724, Santa Cruz) according to the manufacturers' recommendation.

DNA-methylation assay

DNA-methylation assay was performed by bisulfite sequencing (Bis-SEQ). Bisulfite sequencing was performed using the EpiTect Fast Bisulfite Conversion Kit (Qiagen, 59802). Genomic DNA was purified from WT, *Vezf1*^{-/-}, and *Dnmt3b*^{-/-} ESCs. 1 μg of genomic DNA was used and bisulfite-converted DNA was amplified using nested primers and *Taq* polymerase (New England Biolabs, M0267L). Products from the inner PCR were used to generate a library for high throughput sequencing using Wide-SEQ. The primers used are listed in Table S1.

ChIP

ChIP was performed using cross-linked chromatin from WT and *Vezf1*^{-/-} ESCs using a previously published protocol (18). Briefly, cells were cross-linked for 5 min with 1% formaldehyde in buffer (0.1 M NaCl, 1 mM EDTA, 0.5 mM EGTA, and 50 mM HEPES, pH 8). Nuclei were isolated and chromatin was sheared to 0.5–1-kb fragments using a Covaris E210 device, according to the manufacturer's protocols. Antibodies (8 µg) were immobilized on Protein A/G magnetic beads (Life Technologies, 10002D and 10004D) by overnight incubation. The magnetic beads were washed to remove unconjugated antibody and mixed with 8 µg of sonicated chromatin. After an overnight incubation, magnetic beads were washed, and bound DNA was purified. DNA was quantified using PicoGreen (Life Technologies, P11495) in a NanoDrop 3300 fluorospectrometer. Quantitative PCR was then performed using equal amounts of IN (input) and IP (immunoprecipitated sample) DNA. Fold-enrichment was calculated as follows: $C_t(\text{IN}) - C_t(\text{IP})$ and the fold-change was calculated by using $2^{-(C_t(\text{IN}) - C_t(\text{IP}))}$. The fold-enrichment of 1 or less indicates no binding. See Table S1 for primers used. The H3K27Ac antibody used is commercially available (39133, Active Motif) and anti-*Vezf1* antibody is a previously characterized custom-made polyclonal antibody.

Author contributions—L. A., M. H., Q. Y., and H. G. formal analysis; L. A., M. H., and A. B. N. validation; L. A., M. H., Q. Y., and A. B. N. investigation; L. A. and Q. Y. methodology; L. A., M. H., and H. G. writing-review and editing; M. H. visualization; M. H. and H. G. project administration; Q. Y. and H. G. writing-original draft; H. G. conceptualization; H. G. supervision; H. G. funding acquisition.

Acknowledgments—We thank Dr. Sandra Rossie for reading the manuscript, Dr. Taiping Chen for providing *Dnmt3b*^{-/-} ESCs, and Dr. Heidi Stuhlmann for *Vezf1*^{-/-} ESCs.

References

- Carmeliet, P. (2000) Mechanisms of angiogenesis and arteriogenesis. *Nat. Med.* **6**, 389–395 [CrossRef Medline](#)
- Risau, W. (1997) Mechanisms of angiogenesis. *Nature* **386**, 671–674 [CrossRef Medline](#)
- Marcelo, K. L., Goldie, L. C., and Hirschi, K. K. (2013) Regulation of endothelial cell differentiation and specification. *Circ. Res.* **112**, 1272–1287 [CrossRef Medline](#)
- Deudero, J. J., Caramelo, C., Castellanos, M. C., Neria, F., Fernández-Sánchez, R., Calabia, O., Peñate, S., and González-Pacheco, F. R. (2008) Induction of hypoxia-inducible factor 1α gene expression by vascular endothelial growth factor. *J. Biol. Chem.* **283**, 11435–11444 [CrossRef Medline](#)
- Imanirad, P., and Dzierzak, E. (2013) Hypoxia and HIFs in regulating the development of the hematopoietic system. *Blood Cells Mol. Dis.* **51**, 256–263 [CrossRef Medline](#)
- Semenza, G. L. (1999) Regulation of mammalian O₂ homeostasis by hypoxia-inducible factor 1. *Annu. Rev. Cell Dev. Biol.* **15**, 551–578 [CrossRef Medline](#)
- Semenza, G. L. (2001) Hypoxia-inducible factor 1: oxygen homeostasis and disease pathophysiology. *Trends Mol. Med.* **7**, 345–350 [CrossRef Medline](#)
- Zelzer, E., Levy, Y., Kahana, C., Shilo, B. Z., Rubinstein, M., and Cohen, B. (1998) Insulin induces transcription of target genes through the hypoxia-inducible factor HIF-1α/ARNT. *EMBO J.* **17**, 5085–5094 [CrossRef Medline](#)
- Hellwig-Bürgel, T., Rutkowski, K., Metzen, E., Fandrey, J., and Jelkmann, W. (1999) Interleukin-1β and tumor necrosis factor-α stimulate DNA binding of hypoxia-inducible factor-1. *Blood* **94**, 1561–1567 [Medline](#)
- Görlach, A., Diebold, I., Schini-Kerth, V. B., Berchner-Pfannschmidt, U., Roth, U., Brandes, R. P., Kietzmann, T., and Busse, R. (2001) Thrombin activates the hypoxia-inducible factor-1 signaling pathway in vascular smooth muscle cells: role of the p22(phox)-containing NADPH oxidase. *Circ. Res.* **89**, 47–54 [CrossRef Medline](#)
- Spinella, F., Rosanò, L., Di Castro, V., Natali, P. G., and Bagnato, A. (2002) Endothelin-1 induces vascular endothelial growth factor by increasing hypoxia-inducible factor-1α in ovarian carcinoma cells. *J. Biol. Chem.* **277**, 27850–27855 [CrossRef Medline](#)
- Laughner, E., Taghavi, P., Chiles, K., Mahon, P. C., and Semenza, G. L. (2001) HER2 (neu) signaling increases the rate of hypoxia-inducible factor 1α (HIF-1α) synthesis: novel mechanism for HIF-1-mediated vascular endothelial growth factor expression. *Mol. Cell. Biol.* **21**, 3995–4004 [CrossRef Medline](#)
- Pacheco-Leyva, L., Matias, A. C., Oliveira, D. V., Santos, J. M., Nascimento, R., Guerreiro, E., Michell, A. C., van De Vrugt, A. M., Machado-Oliveira, G., Ferreira, G., Domian, I., and Bragança, J. (2016) CITED2 cooperates with ISL1 and promotes cardiac differentiation of mouse embryonic stem cells. *Stem Cell Rep.* **7**, 1037–1049 [CrossRef Medline](#)
- Yin, Z., Haynie, J., Yang, X., Han, B., Kiatchoosakun, S., Restivo, J., Yuan, S., Prabhakar, N. R., Herrup, K., Conlon, R. A., Hoit, B. D., Watanabe, M., and Yang, Y.-C. (2002) The essential role of Cited2, a negative regulator for HIF-1α, in heart development and neurulation. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10488–10493 [CrossRef](#)
- Bamforth, S. D., Bragança, J., Eloranta, J. J., Murdoch, J. N., Marques, F. I., Kranc, K. R., Farza, H., Henderson, D. J., Hurst, H. C., and Bhattacharya, S. (2001) Cardiac malformations, adrenal agenesis, neural crest defects and exencephaly in mice lacking Cited2, a new Tfp2 co-activator. *Nat. Genet.* **29**, 469–474 [CrossRef Medline](#)
- Bamforth, S. D., Bragança, J., Farthing, C. R., Schneider, J. E., Broadbent, C., Michell, A. C., Clarke, K., Neubauer, S., Norris, D., Brown, N. A., Anderson, R. H., and Bhattacharya, S. (2004) Cited2 controls left-right patterning and heart development through a Nodal-Pitx2c pathway. *Nat. Genet.* **36**, 1189–1196 [CrossRef Medline](#)
- Qu, X., Lam, E., Doughman, Y. Q., Chen, Y., Chou, Y. T., Lam, M., Turakhia, M., Dunwoodie, S. L., Watanabe, M., Xu, B., Duncan, S. A., and Yang, Y. C. (2007) Cited2, a coactivator of HNF4α, is essential for liver development. *EMBO J.* **26**, 4445–4456 [CrossRef Medline](#)
- Petell, C. J., Alabdi, L., He, M., San Miguel, P., Rose, R., and Gowher, H. (2016) An epigenetic switch regulates *de novo* DNA methylation at a subset of pluripotency gene enhancers during embryonic stem cell differentiation. *Nucleic Acids Res.* **44**, 7605–7617 [CrossRef Medline](#)
- Du, J., and Yang, Y. C. (2013) Cited2 in hematopoietic stem cell function. *Curr. Opin. Hematol.* **20**, 301–307 [CrossRef Medline](#)
- Du, J., and Yang, Y. C. (2012) HIF-1 and its antagonist Cited2: regulators of HSC quiescence. *Cell Cycle* **11**, 2413–2414 [CrossRef Medline](#)
- Kranc, K. R., Schepers, H., Rodrigues, N. P., Bamforth, S., Villadsen, E., Ferry, H., Bouriez-Jones, T., Sigvardsson, M., Bhattacharya, S., Jacobsen, S. E., and Enver, T. (2009) Cited2 is an essential regulator of adult hematopoietic stem cells. *Cell Stem Cell* **5**, 659–665 [CrossRef Medline](#)
- Bragança, J., Eloranta, J. J., Bamforth, S. D., Ibbitt, J. C., Hurst, H. C., and Bhattacharya, S. (2003) Physical and functional interactions among AP-2 transcription factors, p300/CREB-binding protein, and CITED2. *J. Biol. Chem.* **278**, 16021–16029 [CrossRef Medline](#)
- Fox, S. B., Bragança, J., Turley, H., Campo, L., Han, C., Gatter, K. C., Bhattacharya, S., and Harris, A. L. (2004) CITED4 inhibits hypoxia-activated transcription in cancer cells, and its cytoplasmic location in breast cancer is associated with elevated expression of tumor cell hypoxia-inducible factor 1α. *Cancer Res.* **64**, 6075–6081 [CrossRef Medline](#)
- Freedman, S. J., Sun, Z. Y., Kung, A. L., France, D. S., Wagner, G., and Eck, M. J. (2003) Structural basis for negative regulation of hypoxia-inducible factor-1α by CITED2. *Nat. Struct. Biol.* **10**, 504–512 [CrossRef Medline](#)
- Agrawal, A., Gaighate, S., Smith, H., Anderson, D. G., Albert, T. J., Shapiro, I. M., and Risbud, M. V. (2008) Cited2 modulates hypoxia-inducible factor-dependent expression of vascular endothelial growth

Cited2 causes angiogenesis defects in Vezf1 null cells

- factor in nucleus pulposus cells of the rat intervertebral disc. *Arthritis Rheum.* **58**, 3798–3808 [CrossRef Medline](#)
26. Lee, J. Y., Taub, P. J., Wang, L., Clark, A., Zhu, L. L., Maharam, E. R., Leong, D. J., Ramcharan, M., Li, Z., Liu, Z., Ma, Y. Z., Sun, L., Zaidi, M., Majeska, R. J., and Sun, H. B. (2009) Identification of CITED2 as a negative regulator of fracture healing. *Biochem. Biophys. Res. Commun.* **387**, 641–645 [CrossRef Medline](#)
 27. Xiong, J. W., Leahy, A., Lee, H. H., and Stuhlmann, H. (1999) Vezf1: a Zn finger transcription factor restricted to endothelial cells and their precursors. *Dev. Biol.* **206**, 123–141 [CrossRef Medline](#)
 28. Kuhnert, F., Campagnolo, L., Xiong, J. W., Lemons, D., Fitch, M. J., Zou, Z., Kiosses, W. B., Gardner, H., and Stuhlmann, H. (2005) Dosage-dependent requirement for mouse Vezf1 in vascular system development. *Dev. Biol.* **283**, 140–156 [CrossRef Medline](#)
 29. Zou, Z., Ocaya, P. A., Sun, H., Kuhnert, F., and Stuhlmann, H. (2010) Targeted Vezf1-null mutation impairs vascular structure formation during embryonic stem cell differentiation. *Arterioscler. Thromb. Vasc. Biol.* **30**, 1378–1388 [CrossRef](#)
 30. Gerald, D., Adini, I., Shechter, S., Perruzzi, C., Varnau, J., Hopkins, B., Kazeroonian, S., Kurschat, P., Blachon, S., Khedkar, S., Bagchi, M., Sherris, D., Prendergast, G. C., Klagsbrun, M., Stuhlmann, H., Rigby, A. C., Nagy, J. A., and Benjamin, L. E. (2013) RhoB controls coordination of adult angiogenesis and lymphangiogenesis following injury by regulating VEZF1-mediated transcription. *Nat. Commun.* **4**, 2824 [CrossRef Medline](#)
 31. Clark, S. P., Lewis, C. D., and Felsenfeld, G. (1990) Properties of BGP1, a poly(dG)-binding protein from chicken erythrocytes. *Nucleic Acids Res.* **18**, 5119–5126 [CrossRef Medline](#)
 32. Dickson, J., Gowher, H., Strogantsev, R., Gaszner, M., Hair, A., Felsenfeld, G., and West, A. G. (2010) VEZF1 elements mediate protection from DNA methylation. *PLoS Genet.* **6**, e1000804 [CrossRef Medline](#)
 33. Koyano-Nakagawa, N., Nishida, J., Baldwin, D., Arai, K., and Yokota, T. (1994) Molecular cloning of a novel human cDNA encoding a zinc finger protein that binds to the interleukin-3 promoter. *Mol. Cell. Biol.* **14**, 5099–5107 [CrossRef Medline](#)
 34. Miyashita, H., and Sato, Y. (2005) Metallothionein 1 is a downstream target of vascular endothelial zinc finger 1 (VEZF1) in endothelial cells and participates in the regulation of angiogenesis. *Endothelium* **12**, 163–170 [CrossRef Medline](#)
 35. Miyashita, H., Kanemura, M., Yamazaki, T., Abe, M., and Sato, Y. (2004) Vascular endothelial zinc finger 1 is involved in the regulation of angiogenesis: possible contribution of stathmin/OP18 as a downstream target gene. *Arterioscler. Thromb. Vasc. Biol.* **24**, 878–884 [CrossRef](#)
 36. Aitsebaomo, J., Kingsley-Kallesen, M. L., Wu, Y., Quertermous, T., and Patterson, C. (2001) Vezf1/DB1 is an endothelial cell-specific transcription factor that regulates expression of the endothelin-1 promoter. *J. Biol. Chem.* **276**, 39197–39205 [CrossRef Medline](#)
 37. Lebowitz, P. F., and Prendergast, G. C. (1998) Functional interaction between RhoB and the transcription factor DB1. *Cell Adhes. Commun.* **6**, 277–287 [CrossRef Medline](#)
 38. Gowher, H., Brick, K., Camerini-Otero, R. D., and Felsenfeld, G. (2012) Vezf1 protein binding sites genome-wide are associated with pausing of elongating RNA polymerase II. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 2370–2375 [CrossRef](#)
 39. Gowher, H., Stuhlmann, H., and Felsenfeld, G. (2008) Vezf1 regulates genomic DNA methylation through its effects on expression of DNA methyltransferase Dnmt3b. *Genes Dev.* **22**, 2075–2084 [CrossRef Medline](#)
 40. Hu, C.-J., Wang, L.-Y., Chodosh, L. A., Keith, B., and Simon, M. C. (2003) Differential roles of hypoxia-inducible factor 1 α (HIF-1 α) and HIF-2 α in hypoxic gene regulation. *Mol. Cell. Biol.* **23**, 9361–9374 [CrossRef Medline](#)
 41. Yamaguchi, T. P., Dumont, D. J., Conlon, R. A., Breitman, M. L., and Rossant, J. (1993) flk-1, an flt-related receptor tyrosine kinase is an early marker for endothelial cell precursors. *Development* **118**, 489–498 [Medline](#)
 42. Yang, S., Xin, X., Zlot, C., Ingle, G., Fuh, G., Li, B., Moffat, B., de Vos, A. M., and Gerritsen, M. E. (2001) Vascular endothelial cell growth factor-driven endothelial tube formation is mediated by vascular endothelial cell growth factor receptor-2, a kinase insert domain-containing receptor. *Arterioscler. Thromb. Vasc. Biol.* **21**, 1934–1940 [CrossRef](#)
 43. Kranc, K. R., Oliveira, D. V., Armesilla-Diaz, A., Pacheco-Leyva, I., Catarina Matias, A., Luisa Escapa, A., Subramani, C., Wheadon, H., Trindade, M., Nichols, J., Kaji, K., Enver, T., and Bragança, J. (2015) Acute loss of Cited2 impairs Nanog expression and decreases self-renewal of mouse embryonic stem cells. *Stem Cells* **33**, 699–712 [Medline](#)
 44. Bruderer, M., Alini, M., and Stoddart, M. J. (2013) Role of HOXA9 and VEZF1 in endothelial biology. *J. Vasc. Res.* **50**, 265–278 [CrossRef Medline](#)
 45. Yang, Z., and Li, J. C. (2008) Stimulation of endothelin-1 gene expression by insulin via phosphoinositide-3 kinase-glycogen synthase kinase-3 β signaling in endothelial cells. *Life Sci.* **82**, 512–518 [CrossRef Medline](#)
 46. Valster, A., Tran, N. L., Nakada, M., Berens, M. E., Chan, A. Y., and Symons, M. (2005) Cell migration and invasion assays. *Methods* **37**, 208–215 [CrossRef Medline](#)
 47. Eccles, S. A., Box, C., and Court, W. (2005) Cell migration/invasion assays and their application in cancer drug discovery. *Biotechnol. Annu. Rev.* **11**, 391–421 [CrossRef Medline](#)

**The transcription factor Vezf1 represses the expression of the antiangiogenic factor
Cited2 in endothelial cells**

Lama AlAbdi, Ming He, Qianyi Yang, Allison B. Norvil and Humaira Gowher

J. Biol. Chem. 2018, 293:11109-11118.

doi: 10.1074/jbc.RA118.002911 originally published online May 24, 2018

Access the most updated version of this article at doi: [10.1074/jbc.RA118.002911](https://doi.org/10.1074/jbc.RA118.002911)

Alerts:


- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

This article cites 47 references, 19 of which can be accessed free at
<http://www.jbc.org/content/293/28/11109.full.html#ref-list-1>

Article

Characterization of Small Molecules Inhibiting the Pro-Angiogenic Activity of the Zinc Finger Transcription Factor Vezf1

Ming He ¹, Qianyi Yang ^{1,2}, Allison B. Norvil ¹, David Sherris ³ and Humaira Gowher ^{1,4,*} 
¹ Department of Biochemistry, Purdue University, West Lafayette, IN 47907, USA; he261@purdue.edu (M.H.); qianyiyang@wustl.edu (Q.Y.); anorvil@purdue.edu (A.N.)

² Department of Anesthesiology, Washington University School of Medicine, 660 S Euclid Ave, St. Louis, MO 63110, USA

³ GenAdam Therapeutics, Inc., 37 Neillian Crescent, Jamaica Plain, MA 02130, USA; dsherris@genadamthera.com

⁴ Purdue University Center for Cancer Research, Purdue University, West Lafayette, IN 47907, USA

* Correspondence: hgowher@purdue.edu; Tel.: +1-3018202794

Received: 14 May 2018; Accepted: 24 June 2018; Published: 3 July 2018



Abstract: Discovery of inhibitors for endothelial-related transcription factors can contribute to the development of anti-angiogenic therapies that treat various diseases, including cancer. The role of transcription factor Vezf1 in vascular development and regulation of angiogenesis has been defined by several earlier studies. Through construction of a computational model for Vezf1, work here has identified a novel small molecule drug capable of inhibiting Vezf1 from binding to its cognate DNA binding site. Using structure-based design and virtual screening of the NCI Diversity Compound Library, 12 shortlisted compounds were tested for their ability to interfere with the binding of Vezf1 to DNA using electrophoretic gel mobility shift assays. We identified one compound, T4, which has an IC₅₀ of 20 μM. Using murine endothelial cells, MSS31, we tested the effect of T4 on endothelial cell viability and angiogenesis by using tube formation assay. Our data show that addition of T4 in cell culture medium does not affect cell viability at concentrations lower or equal to its IC₅₀ but strongly inhibits the network formation by MSS31 in the tube formation assays. Given its potential efficacy, this inhibitor has significant therapeutic potential in several human diseases.

Keywords: Vezf1; angiogenesis; vascular biology; endothelial cells; MSS31; tube formation; small molecule inhibitors computational modeling

1. Introduction

Vascular Endothelial Zinc Finger 1 (Vezf1), previously named Bgp1, was discovered as a protein binding to a poly (dG) sequence present in the neighborhood of chicken β-globin promoter [1,2]. Vezf1 belongs to the family of Kruppel-like zinc finger protein that contains six C2H2 class zinc finger motifs (~99% identical between homologues). It is about 65 kd protein and recognizes long strings of poly (dG).poly (dC) ('G-strings'). Biochemical studies revealed that the minimal binding site of Vezf1 is a (dG)⁷ string, but it also binds to a bipartite poly G string containing a (dG)⁶ and (dG)⁴ separated by three to four nucleotides present in the upstream hypersensitive site 5'/HS4 of chicken β-globin locus [3]. Interestingly, poly G tracts are highly prevalent in CpG islands in mammalian genomes resulting in a number of putative binding sites in these regions. Similar to some other zinc finger proteins such as CTCF, Vezf1 is conserved on among vertebrates and the proteins are near identical between mouse and human [4,5].

The expression of *Vezf1* is highly restricted in vascular endothelium during embryogenesis. Targeted inactivation of *Vezf1* gene in mice causes embryonic lethality and it acts in a dosage dependent fashion to regulate the development of blood and lymphatic vascular system [6,7]. Embryonic stem cells (ESCs) derived from *Vezf1*^{-/-} embryos differentiated into embryoid bodies were shown to have defect sprouting angiogenesis. However, the loss of *Vezf1* in embryos had no significant effect on the expression of pioneer factors that regulate vasculogenesis including VEGF-A-A (Vascular Endothelial Growth Factor) and ET-1, indicating a complex mechanism [8]. Our genome-wide gene expression analysis of the *Vezf1*^{-/-} ESCs showed a significant high expression of the antiangiogenic factor *Cited2/Mrg1* [9]. Using an in vitro endothelial cell (EC) differentiation and tube formation as model system for angiogenesis, our data showed that some of the differentiation and angiogenesis defects in *Vezf1*^{-/-} ECs could be rescued by reducing the *Cited2* expression in these cells [10]. Other studies have shown the function of *Vezf1* in regulation of adult angiogenesis by activating pro-angiogenic genes including microtubule turnover protein, stathmin/OP18, and metallothionein 1 (MT1) [7,11,12]. Haploinsufficient *Vezf1* mice exhibit reduced angiogenic response to injury. Using primary human endothelial cells (BVECs), *Vezf1* was shown to be functionally linked with Rho B which is a small GTPase known to be involved in angiogenesis during post-natal retinal development [13,14]. To understand the mechanism of *Vezf1*, our previous studies showed a major loss of DNA methylation genome-wide with a concomitant reduction in the expression of a major DNA methyltransferase *Dnmt3b* in *Vezf1*^{-/-} ESCs [15]. We further showed that *Vezf1* binding sites overlap with the RNA polymerase 2 pausing sites genome-wide and interacts with chromatin modifying proteins [9]. However the details of *Vezf1* mechanism and how it impacts gene expression in early development or in somatic cells is not known.

Formation of vascular system is the most critical and earliest step in development and is a critical requirement for tumor progression and metastasis [16,17]. Several studies have led to the identification of many transcription factors that regulate angiogenesis but most of them are not endothelial specific thus making it difficult to use them as targets for anti-cancer drug development [18,19]. *Vezf1* is unique among endothelial factors since during development its function is restricted to vascular system thus making it an attractive target for cancer therapeutics [6]. Although *Vezf1* null mutants are embryonic lethal, *Vezf1* null embryonic stem cell derived teratocarcinomas were able to differentiate into each of the three germ layer. However, these tumors showed decreased proliferative ability and delayed differentiation [8]. Currently approved angiogenesis inhibitors target the VEGF pathway by direct inhibition of VEGF-A (bevacizumab) or VEGF-A receptors (sunitinib and sorafenib), and few explore alternatives to this direct inhibition [20]. New endothelial-related transcription factors can contribute to the development of antiangiogenic therapies that treat diseases ranging from cardiovascular disease to cancer. The need for new cancer therapeutics and the necessity to understand the complex mechanisms involved in regulation of cell-type specific genes has increased the demand for comprehensive characterization of small molecule inhibitors against factors such as *Vezf1* that can potentially work in a restricted fashion and block angiogenesis and hyper-vascularization in diseased state.

In the present study we used recombinant *Vezf1* protein to characterize 12 compounds that could potentially inhibit its binding to DNA. We found T4 to block *Vezf1*-DNA binding at low μ M concentration. We tested the cellular toxicity and activity of T4 by using MSS31 endothelial cells. We show a specific inhibition of tube formation by MSS31 cells which are viable when treated with a dose equal to IC₅₀ of T4. We therefore report an effective compound that can potentially block angiogenesis without affecting cell viability.

2. Results

2.1. Computational Modelling of *Vezf1* Structure, Evaluation of Potential Binding Sites and Design of Small Molecule Inhibitors

Computational modeling and design of small molecules was done at Vasculomedics Inc. (Boston, MA, USA). A model of the zinc finger, *Vezf1* that would compare favorably to the experimentally

determined structures of other zinc fingers was determined. Using high resolution DNA bound structure of Zinc Finger 268 and 1AAY [21], the structure was further refined using Monte Carlo searching (see Figure S1). VasculoMedics, Inc. had previously tested a few lead compounds that inhibit Vezf1 DNA binding using previously designed model of Vezf1. Docking tools of FLO (McDock+) were used to dock these known ligands (see Figures 1 and 2). New compounds were designed such that they incorporate only minimal changes from the lead compounds thus using docking experiments of lead compounds as a guide. The de novo program AlleGrow was used to discover novel side chains to replace those of the leads. Each designed compound was docked and evaluated in the model. Those structures were selected which formed the most favorable interactions and had good energy profiles. After evaluation of several lead compounds for complexity in synthesis, six compounds were selected and synthesized at MedChem Partners. Next, a major virtual screening program was performed in order to find compounds at the NCI Diversity Compound Library.

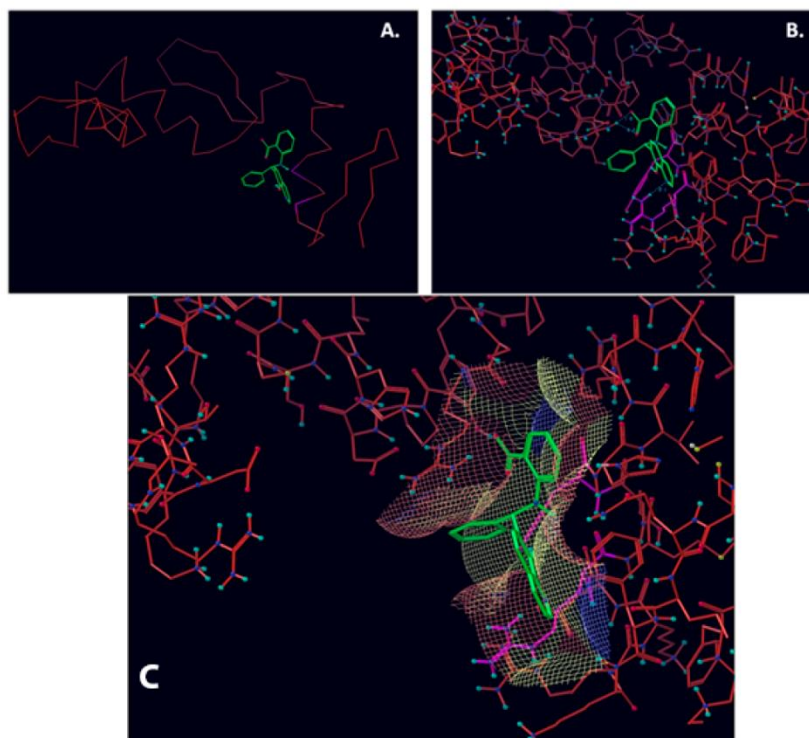


Figure 1. NSC1012 docked into Vezf1 Model binding site. (A) The alpha carbon trace and (B) complete model. (C) NSC1012 docked into Model binding site with the solvent accessible surface: yellow covers hydrophobic areas of the binding site, red, hydrogen bond acceptor, and blue hydrogen bond donor.

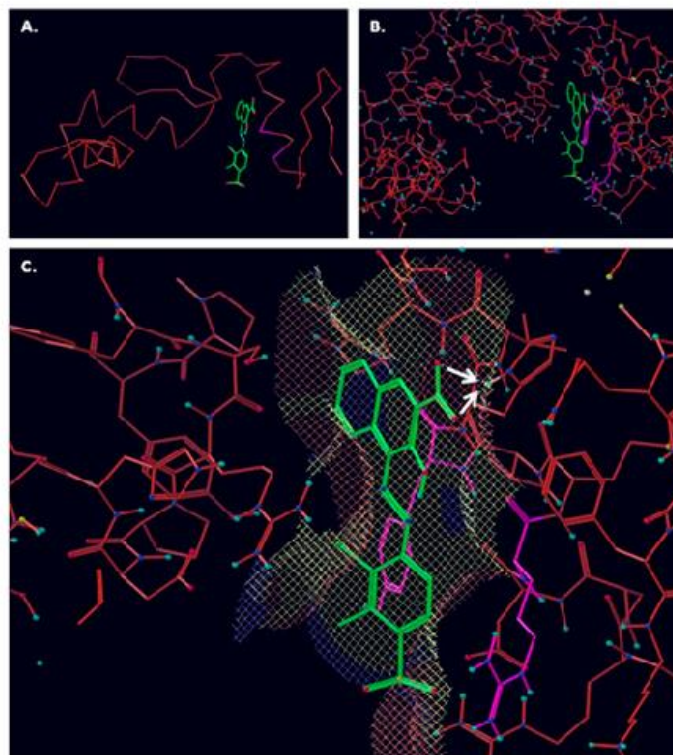


Figure 2. NSC16087 docked into Model II binding site. In (A) alpha carbon trace, in (B) complete model and in (C) NSC16087 docked into Model binding site with the accessible surface mesh. Hydrogen bonds to His 48 are shown with white arrows.

2.2. Virtual Screening

We used major virtual screening program to find compounds that could be obtained from the NCI Diversity Compound Library. The first step for virtual screening was to identify the docking protocol, which gave the best result when docking compounds of known activity. Therefore, two docking algorithms were tested by docking all compounds. The methods tested included:

- (1) Sdock+ is a very fast docking algorithm that uses a novel method for generating vast numbers of conformers in the target-binding site. The best scoring conformers are energy minimized in the binding site.
- (2) McDock+ uses a Monte Carlo algorithm for generating new conformers. Conformers generated by McDock+ are optimized with energy minimization and the FLO scoring function. The calculations are much more time consuming than Sdock+, however, often the Mcdock+ results are more reliable.
- (3) SDock+ followed by McDock+ Conformers found by SDock+ served as a starting point for McDock+.

Each method was evaluated by determining the rank order obtained for the three lead compounds as shown in Table 1. SDock+ alone was much faster with results as good as Sdock+ followed by McDock+ and better than straight McDock+. Hence, SDock+ was used to screen the NCI Diversity

Compound Library. Upon finding positive compounds selected from virtual screening in the NCI database, each compound was evaluated in the model (see Figure 3).

Table 1. Rank order of known compounds using each method. Results are shown of testing docking protocols.

	SDock+	McDock+	SDock+ Plus McDock+
M1012	2	7	3
M16087	3	3	2
M609974	12	15	17

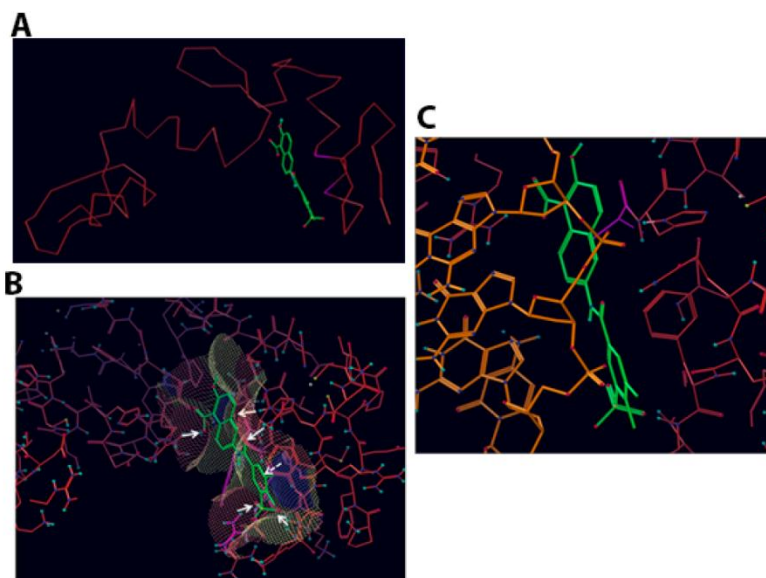


Figure 3. Compound T4 docked into Vezf1 Model binding site. (A) Only the alpha carbon trace is shown to indicate how T4 fits between two fingers. (B) Compound T4 docked into binding site with the accessible surface mesh. Arrows show important interactions with solid arrows indicating hydrogen bonds and dotted lines show stacking interactions between the compounds phenyl and phenyl ring of PHE123. (C) Compound T4 (green carbon atoms) docked into binding site (red carbon atoms). The structure 1AAY of the zinc finger 268 with DNA has been superimposed. Only DNA (orange carbon atoms) is shown.

Those forming the most favorable interactions were selected, for synthesis by MedPharma Partners. All compound synthesized are shown in Table 2.

2.3. Cloning and Purification of Recombinant Vezf1

The Vezf1 gene was cloned to encode an N-terminal His6-tag fusion protein into pQE10 (Qiagen, Hilden, Germany). XLBlue (MRE, T7) *E. coli* cells, transformed with pQE10 Vezf1, were grown at 32 °C in 500 mL of LB medium containing 75 mg/mL ampicillin. Protein expression was induced at a cell density of 0.3 A600 nm by the addition of 1 mM IPTG and the cells were grown for an additional two hours at 30 °C. All purification steps were carried out at 4 °C. Since the Vezf1 protein turned out to be susceptible to proteolysis, the purification was carried out in the presence of Protease Inhibitor Cocktail (Roche, Basel, Switzerland) in the sonication buffer. The Vezf1 concentration was estimated

from Coomassie blue-stained SDS-PAGE gels using protein standards of known concentration and Western blots were carried out using an anti-His6-tag antibody (Abcam, Cambridge, UK), according to the instructions of the supplier (see Figure 4A).

2.4. Determination of DNA Binding Constant Vezf1 to Its Specific DNA Sequence

In order to perform the biochemical testing of the potential small molecule inhibitors, we first determined the DNA binding constant of the recombinant His-Vezf1. Gel mobility shift analysis was used to analyze the interactions of His-Vezf1 recombinant protein with ^{32}P -labelled oligonucleotides containing Vezf1 binding sites characterized at the chicken beta-globin insulator element. Binding assays were carried by incubating 25 nM of radiolabeled oligonucleotides with two-fold increasing concentration of protein ranging from 50–1000 nM in the binding buffer. Protein bound DNA runs as a slower species on the gel (see Figure 4B). The band intensities are used to determine the ratio of bound to unbound nucleic acid on the gel which reflects the fraction of free and bound probe molecules as the binding reaction enters the gel. The data were fitted to the following expression which directly follows from the definition of a bimolecular binding equilibrium and was used to determine the binding constant (K_{dis}) of Vezf1:

$$v_1 = \frac{D_{\text{tot}} + P_{\text{tot}} + K_{\text{dis}} - \sqrt{(D_{\text{tot}} + P_{\text{tot}} + K_{\text{dis}})^2 - 4 \times D_{\text{tot}} \times P_{\text{tot}}}}{2 \times D_{\text{tot}}}$$

v_1 is the fractional degree of saturation of the available protein binding sites on the DNA at increasing protein concentrations. The dissociation constant of Vezf1 as shown in Figure 4C was approximately 640 nM.

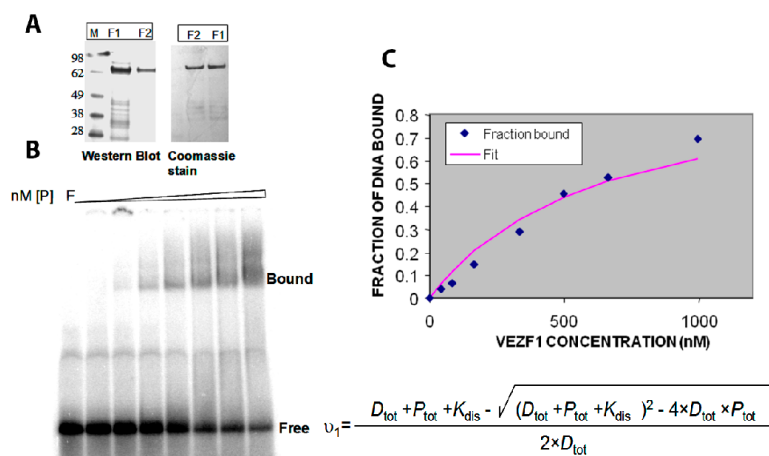


Figure 4. (A) His tagged Vezf1 was purified with affinity chromatography using Ni-NTA column. F1 and F2 represent the eluted fractions 1 and 2 which are on SDS PAGE stained with Coomassie blue. The integrity of the recombinant protein was tested by Western blot probed with anti-His antibody. (B) EMSA (electrophoretic mobility shift assay) was performed to determine the DNA binding constant of Vezf1. 25 nM of radiolabeled DNA were incubated with increasing amount of purified Vezf1 protein 50–1000 nM). (C) The band intensities of the free and bound fractions were measured using Image Quant L software in Typhoon Imager. The data were fitted into an equation of binding equilibrium to determine binding constant of 640 nM.

2.5. Effect of Small Molecule Inhibitors on DNA Binding Property of Vezf1

Next, we tested the effect of candidate inhibitors (see Table 1) on the DNA binding affinity of Vezf1. Vezf1 protein at 640 nM concentration was incubated with 500 μ M of potential inhibitors for 10 min. All the 12 compounds tested are soluble in water and/or DMSO. 25 nM labeled DNA was added to the reaction and binding assays were done as described above (see Figure 5A). The bound and the free DNA fractions were quantified and plotted (see Figure 5B). Various compounds had varying effect on the DNA binding ability of Vezf1. 3 of the 12 tested compounds, A, B and C, had the greatest degree of potency resulting in a near complete loss of DNA binding by Vezf1. The percent bound DNA is also shown in Table 1.

Next we determined the minimum concentration at which these 3 compounds were effective, and determine the IC₅₀. DNA binding assays were done by pre-incubating 1 μ M Vezf1 with varying concentrations of the inhibitors and test the ability of Vezf1 to bind DNA using the gel shift assays. Out of the 3 small molecule inhibitors, T4 (503-1-83) was most potent. It was able to block DNA binding by Vezf1 effectively at a concentration of 20 μ M. T6 (503-1-71) and NSC1012 inhibited DNA binding by Vezf1 at concentrations of 100 μ M and 500 μ M respectively (see Figure 5C).

Table 2. Summary of the compounds used and their activity.

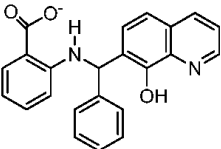
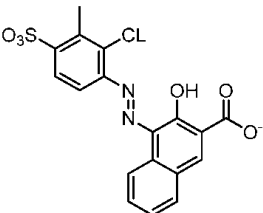
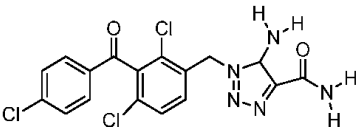
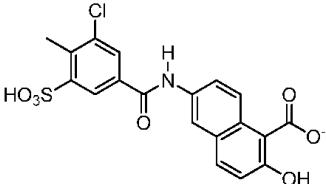
Compound Structure	Compound Number	% DNA Bound in EMSA (500 μ M)	Activity Concentration
	NSC1012 (1)	7.38%	500 μ M
	NSC16087 (2)	22.56%	500 μ M
	NSC609974 (3)	38.45%	500 μ M
	T4	5.94%	20 μ M

Table 2. Cont.

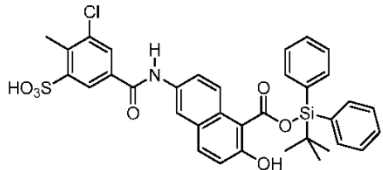
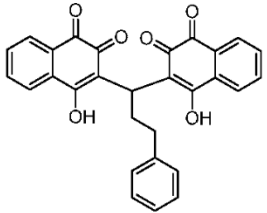
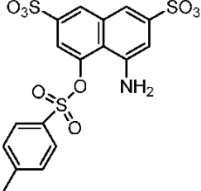
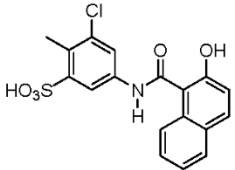
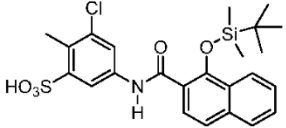
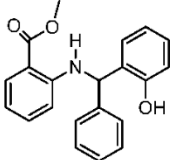
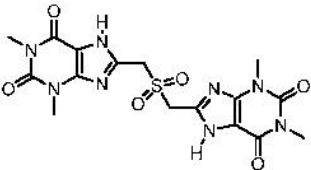
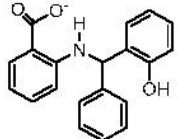
Compound Structure	Compound Number	% DNA Bound in EMSA (500 μ M)	Activity Concentration
	T6	5.92%	100 μ M
	CID 272651 (NSC117274)	27.13%	NA
	CID 23270	29.37%	NA
	T3	30.81%	NA
	T5	36.81%	NA
	T2	38.23%	

Table 2. Cont.

Compound Structure	Compound Number	% DNA Bound in EMSA (500 μ M)	Activity Concentration
	CID 267103	38.95%	
	T1	43.58%	

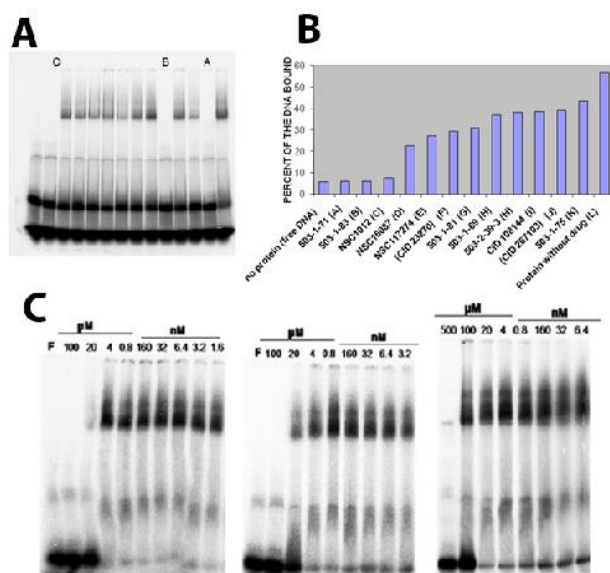


Figure 5. (A) EMSA showing the effect of candidate inhibitors on the DNA binding affinity of Vezf1. The protein was pre incubated 500 nM of potential inhibitors for 10 min before adding the DNA to the binding reaction. Compounds noted as A, B and C. Presence of compounds have varying effect on the DNA binding ability of VEZF1. 3 of the 12 tested compounds, A, B and C, severely affected DNA binding of the recombinant Vezf1. (B) The bound and free reactions were quantified and plotted as percent bound. A, B and C compounds were identified as 501-1-71, 503-1-83 and NSC1012 respectively. (C) To determine the minimum concentration at which these 3 drugs were effective, DNA binding assays were done by pre-incubating 640 nM Vezf1 with varying concentrations of the 3 selected inhibitors, followed by DNA binding assay. As shown, 503-1-83 was most effective in blocking DNA binding by Vezf1 at 20 μ M, whereas 503-1-71 showed similar affect at about 100 μ M and NSC1012 at 500 μ M.

2.6. Effect of Small Molecule Inhibitors on Cell Viability

Following the biochemical tests, we wanted to test the ability of these compounds to inhibit Vezf1 activity in cells. However, we first tested the cellular toxicity of some selected compounds. We tested four candidates: T4: most potent based on IC_{50} , followed by T6 and T5. We also tested T2 as a control which has no effect in vitro on Vezf1 DNA binding but it has a similar structure to T4 (see Figure 6A). Due its critical role in endothelial development and angiogenesis, we used the mouse endothelial cell line, MSS31, for these studies. Cultured MSS31 cells were treated with increasing concentration of the four inhibitors, T4, T5, T6 and T2 for 24 h. Cells were collected and the live cell population was counted by Bio-Rad cell counter using Trypan Blue exclusion test. Live cells are impermeable to Trypan blue dye, which only stains the dead population. The Bio-Rad cell counter identifies and counts only the unstained cells. As shown in Figure 6B, exposure to T2 and T6 causes cell death at very low concentration. T4 is less toxic and can be tolerated until 50 μ M. These data show that T4 can be tolerated by cells at concentrations close to its estimated IC_{50} .

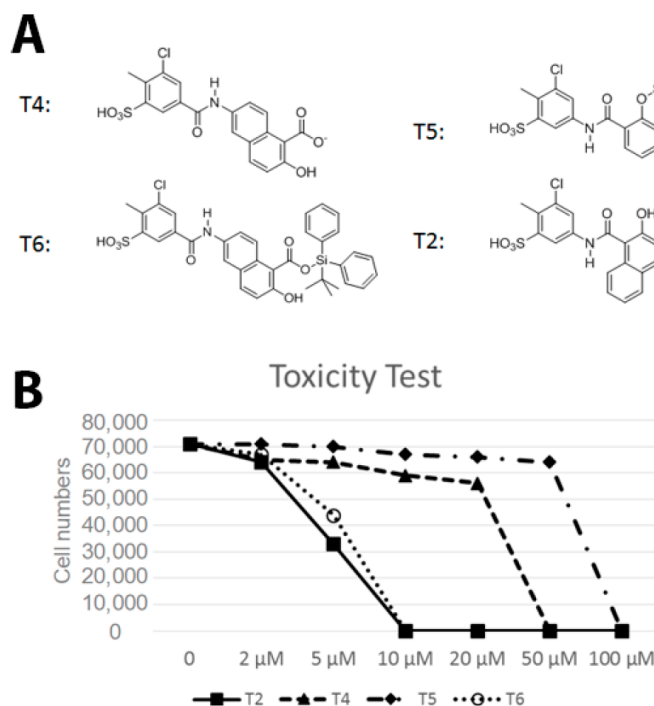


Figure 6. Effect of small molecule inhibitors on cell viability. **(A)** To determine the minimum concentration at which the small molecule inhibitors can be tolerated by cells, four inhibitors were chosen based on their effect on the DNA binding by Vezf1 in gel shift assays. **(B)** MSS31 cells were treated with inhibitors for 24 h and live cells were counted using Trypan blue exclusion test and plotted against the concentration of the inhibitors used. Whereas T5 and T6 were deleterious at very low concentrations, T4, which as a lowest IC_{50} of 20 μ M, was tolerated by cells in this range.

2.7. T4 Treated MSS31 Cells Are Incapable of Tube Formation in Matrigel™

Vezf1 has a critical function in differentiation of endothelial cells and angiogenesis. Angiogenesis can be modeled in tissue culture by plating MSS31 endothelial cells in VEGF-A containing Matrigel™. The cells arrange into distinct tube like structures within 6–10 h. To determine the ability of the small molecule inhibitors to disrupt this biological activity of Vezf1, we tested the effect of T4 and T5 on the tube formation by endothelial cells MSS31. We exposed MSS31 cells to T4 at 20 μ M, a concentration close to its estimated IC₅₀ value and T5 at 75 μ M, given at 100 μ M it is lethal for 24 h. Treated and untreated MSS31 cells were plated in Matrigel™. Within 6–10 h the untreated cells organized into networks as shown by phase contrast microscopy (see Figure 7A). The cells treated with T5 were not at all affected and formed networks similar to untreated cells, however the network formation by cells treated with T4 were strongly inhibited. The tube length was measured which shows a clear absence of tube formation by T4 treated MSS31 cells (see Figure 7B).

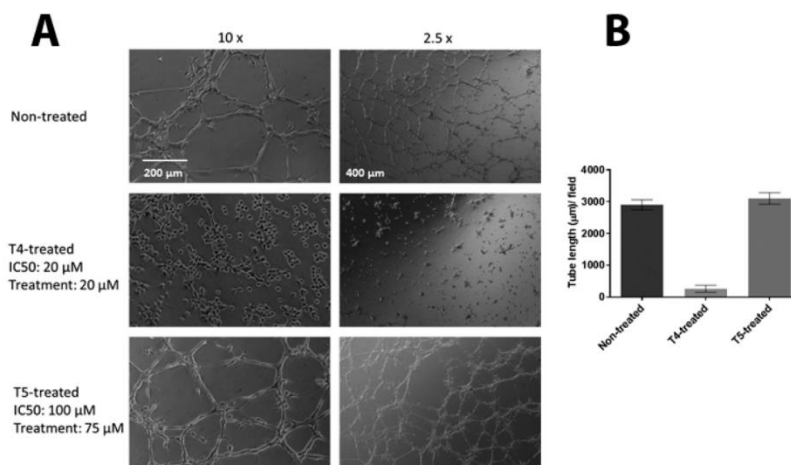


Figure 7. Effect of Vezf1 inhibitors on its biological activity. (A) MSS31 cells were treated with 20 μ M compound T4 or 75 μ M compound T5 for 24 h. Cells were placed on Matrigel™, and tube formation was evaluated after 18 h by bright field microscopy. T4 treatment clearly prevents tube formation in MSS31 cells compare to non-treated and T5-treated cells. 10 \times and 2.5 \times show two magnifications of cells in Matrigel™. (B) Tube length in A was measured using Image Quant L software and plotted. T4 treated cells show no tube formation whereas T5 treated cells, the tube length is similar to that of the untreated cells.

These data strongly support T4 to be inhibitor of Vezf1 activity in endothelial cells and has ability to block angiogenesis at non-toxic concentrations.

3. Discussion

Multiple signaling pathways that regulate proliferation and differentiation of cells converge into a small subset of transcription factors (TFs) which in turn drive oncogenic transformation by controlling the expression of a large number of genes. This makes TFs compelling targets for cancer therapy. The well-known cancer target TF include MYC, the ETS family, STAT, Fos, Jun, Myb, Sox2 and more. Indirect targeting of TF's has been used as a therapeutic strategy though inhibition of upstream signaling pathways including kinases. Direct inhibition by small molecule inhibitors against TFs works by blocking their site of activation or the interaction of TF with DNA. These small molecules could be purified from natural source or by high throughput screening of thousands of synthetic compounds

using biological assays facilitated by the repository of more than 400,000 candidate drugs at National Cancer Institute. Designing molecules that interrupt DNA binding by TFs has drawbacks since DNA is negatively charged and charged molecules have low permeability. Frequently, small molecules have low affinity and low specificity therefore; they must be used at high doses to achieve therapeutic effects [22]. An example of direct inhibition of TFs include ligands such as Tamoxifen, which binds to nuclear receptor protein, the oestrogen receptor [23]. Besides nuclear receptors, there are only few examples of the direct inhibition of TFs by small molecules. These include the inhibition of FoxM1 DNA binding by naturally occurring antibiotic thiostrepton [24]. Small molecules could be designed to target protein-protein interactions and an example is nutlin, which disrupts the interaction between the transcription factor p53 and its negative regulator MDM2 [25].

Angiogenesis is regulated by the activity of several proteins including Vezf1. Several transcription factors, like VEGF-A, VEGF-AR, and ET-1 have been studied and the mechanistic details of their activity in angiogenesis have been worked out [16]. The discovery of Vezf1 as an endothelial transcription factor and its role in angiogenesis is relatively new. Genetic knockout studies revealed the role of Vezf1 in vascularization, however the mechanism of its regulation is not fully understood [6]. Our previous studies have characterized the activity of Vezf1 at chicken β -globin insulator. Insulators are chromatin elements that block the non-specific activity of enhancers by disallowing their interaction with unassigned promoters [26]. However, the mechanism by which Vezf1 regulates gene expression and angiogenesis during development and disease is unknown. Through construction of a new computational model for Vezf1, work here has identified a potential small molecule binding site (occupied by DNA in the X-ray structure used to construct the model) [21] for use in creating novel small molecule drugs capable of inhibiting Vezf1 from binding to its cognate DNA binding site. The structure of Zif268 was used to model the structure of Vezf1 given the high sequence similarity between the zinc fingers of Vezf1 and Zif268. The Zif268 structure was determined in complex with DNA and demonstrates detailed molecular interactions between zinc fingers and DNA backbone. This motif provides a framework to design molecules that can interrupt DNA binding by zinc fingers. Using AlleGrow and docking tools of FLO to design new compounds, and use of and use of NCI Diversity Compound Library, a novel compound was discovered which inhibits binding of DNA at 20 μ M concentration. Of particular note, out of 12 designed compounds, identification of such a small molecule inhibitor may be considered a significant achievement.

Discovery of a small molecular inhibitor against Vezf1 will not only have therapeutic use, but can also be employed to understand its mechanism in development and disease. Tumor cells treated with these inhibitors can be tested for proliferation, migration and tube formation. Further this inhibitor can be used to study wound healing in animal models. This new lead compound is also a candidate for the determination of an X-ray crystal structure of the zinc finger/ligand complex. It is also ideal as a starting point for new rounds of design, synthesis and biological evaluation to find medicinal agents interfering with zinc finger/DNA interactions, a new mechanism of action for development of agents against cancer.

4. Materials and Methods

4.1. Protein Purification and DNA Binding Assays

XLBlue (MRE, T7) *E. coli* cells, transformed with pQE10 Vezf1 were induced using IPTG and two hours later harvested by centrifugation at 6 K RPM. The cells were washed with STE buffer (10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, 0.1 M NaCl) and centrifuged at 6K RPM. The cell pellet was suspended in buffer A (20 mM KPi (pH 7.5), 1 mM EDTA, 0.1 mM DTT, 0.5 M NaCl, 10% (v/v) glycerol, 20 mM imidazole) and the cells were disrupted by sonication. The insoluble cell debris was removed by centrifugation (60 min, 13,000 \times g). The supernatant was applied onto a Ni-NTA (Qiagen, Mettman Germany) column (1 mL gel bed) equilibrated with buffer A. After washing with 150 mL of buffer A, the His-Vezf1 was eluted with 5 mL of elution buffer (20 mM KPi (pH 7.5), 1 mM EDTA, 0.1 mM DTT,

0.5 M NaCl, 10% (v/v) glycerol, 200 mM imidazole). The eluate was dialyzed overnight against storage buffer (20 mM Hepes (pH 7.5), 40 mM NaCl, 1 mM EDTA, 0.2 mM DTT, 20% glycerol) aliquoted and stored at -80°C . For DNA binding assays, 50 pmoles of oligonucleotide were end-labeled by polynucleotide kinase in presence of γ [32P] ATP. Purified recombinant Vezf1 was allowed to bind to the oligonucleotide containing the Vezf1 binding site in HS4-FPIII. The reaction was carried out in 15 μL volume in binding buffer (20 mM Hepes 7.5, 10 mM MgCl_2 , 50 mM NaCl, 1 mM EDTA, 5 ng/ μL poly dA.dT) and incubated for 10 min at room temperature followed by the addition of 5% Ficoll.

The bound and the free DNA was separated on 5% acrylamide gel run in .5XTBE buffer for 2 h at 150 V. The gel was dried and imaged by Typhoon (GE Healthcare Lifesciences, Chicago, IL, USA). The band densities were quantified by the ImageQuant TL software (GE Healthcare Lifesciences). The top strand sequence for the duplex used for gel mobility shift analyses is 5'AGGCGCGCCCCGGTCCGGCGCTCCCCCGCATCCCCGAGCCGGGGCGCGCT3'. The relative band intensities were used to fit the data to an equation representing bimolecular binding equilibrium using Microsoft Excel and the binding constant was determined.

4.2. Small Molecule Inhibitors

The small molecule synthesized by MedPharma Partners (Boston, MA, USA) were suspended in DMSO at 10 mg/mL concentration. Further dilutions for use in the gel shift assays and cell culture were done in water. The diluted samples were stored in -20°C whereas the stock in DMSO was stored in 4°C .

4.3. Cell Viability and Tube Formation Assays

MSS31 mouse endothelial cells were collected by trypsinization and counted using a Bio-Rad Cell Counter (Hercules, CA, USA). The cell viability assay was performed plating 1×10^6 MSS31 cells in each well of a 6-well plate. After 24 h the cells were treated with various small molecules by directly adding into the medium at various concentrations. The cells were incubated for 24 h and viability of cells was measured. The cells were trypsinized and treated with Trypan Blue. The live cell population which is stained is counted using Bio-Rad cell counter. The tube formation or sprouting assay was performed by plating 2×10^5 cells on a 24-well plate coated with VEGF-A supplemented Matrigel™ (BD Biosciences, East Rutherford, NJ, USA) according to the manufacturers protocol. The cells were incubated at 37 degrees for 3–18 h. Sprouting was scored using images from phase contrast microscopy. The length of the tubes was measured by Image J software (National Institutes of Health, Bethesda, MA, USA). For inhibitor treatment, MSS31 cells were treated for 24 h with compounds, after which cells were trypsinized and plated on Matrigel™ for tube formation assays.

Supplementary Materials: The following are available online: Figure S1.

Author Contributions: H.G. and D.S. designed the experiments, M.H., Q.Y. and A.N. performed the experiments and H.G. wrote the manuscript.

Acknowledgments: This work was supported by Scientist Development Award from American Heart Association to H.G., ACS institutional grant from PUCCR supported by P30 CA0231681 and R43 GM093385-01 to DS.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lewis, C.D.; Clark, S.P.; Felsenfeld, G.; Gould, H. An erythrocyte-specific protein that binds to the poly(dG) region of the chicken beta-globin gene promoter. *Genes Dev.* **1988**, *2*, 863–873. [[CrossRef](#)] [[PubMed](#)]
2. Clark, S.P.; Lewis, C.D.; Felsenfeld, G. Properties of BGP1, a poly(dG)-binding protein from chicken erythrocytes. *Nucleic Acids Res.* **1990**, *18*, 5119–5126. [[CrossRef](#)] [[PubMed](#)]
3. Dickson, J.; Gowher, H.; Strogantsev, R.; Gaszner, M.; Hair, A.; Felsenfeld, G.; West, A.G. VEZF1 elements mediate protection from DNA methylation. *PLoS Genet.* **2010**, *6*, e1000804. [[CrossRef](#)] [[PubMed](#)]

4. Weth, O.; Renkawitz, R. CTCF function is modulated by neighboring DNA binding factors. *Biochem. Cell Biol.* **2011**, *89*, 459–468. [[CrossRef](#)] [[PubMed](#)]
5. Bruderer, M.; Alini, M.; Stoddart, M.J. Role of HOXA9 and VEZF1 in endothelial biology. *J. Vasc. Res.* **2013**, *50*, 265–278. [[CrossRef](#)] [[PubMed](#)]
6. Kuhnert, F.; Campagnolo, L.; Xiong, J.W.; Lemons, D.; Fitch, M.J.; Zou, Z.; Kiessses, W.B.; Gardner, H.; Stuhlmann, H. Dosage-dependent requirement for mouse *Vezf1* in vascular system development. *Dev. Biol.* **2005**, *283*, 140–156. [[CrossRef](#)] [[PubMed](#)]
7. Aitsebaomo, J.; Kingsley-Kallesen, M.L.; Wu, Y.; Quertermous, T.; Patterson, C. *Vezf1*/DB1 is an endothelial cell-specific transcription factor that regulates expression of the endothelin-1 promoter. *J. Biol. Chem.* **2001**, *276*, 39197–39205. [[CrossRef](#)] [[PubMed](#)]
8. Zou, Z.; Ocaya, P.A.; Sun, H.; Kuhnert, F.; Stuhlmann, H. Targeted *Vezf1*-null mutation impairs vascular structure formation during embryonic stem cell differentiation. *Arterioscler. Thromb. Vasc. Biol.* **2010**, *30*, 1378–1388. [[CrossRef](#)] [[PubMed](#)]
9. Gowher, H.; Brick, K.; Camerini-Otero, R.D.; Felsenfeld, G. *Vezf1* protein binding sites genome-wide are associated with pausing of elongating RNA polymerase II. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 2370–2375. [[CrossRef](#)] [[PubMed](#)]
10. Alabdi, L.; He, M.; Yang, Q.; Norvil, A.B.; Gowher, H. The transcription factor *Vezf1* represses the expression of the antiangiogenic factor *Cited2* in endothelial cells. *J. Biol. Chem.* **2018**. [[CrossRef](#)] [[PubMed](#)]
11. Miyashita, H.; Sato, Y. Metallothionein 1 is a downstream target of vascular endothelial zinc finger 1 (VEZF1) in endothelial cells and participates in the regulation of angiogenesis. *Endothel. J. Endothel. Cell Res.* **2005**, *12*, 163–170. [[CrossRef](#)] [[PubMed](#)]
12. Miyashita, H.; Kanemura, M.; Yamazaki, T.; Abe, M.; Sato, Y. Vascular endothelial zinc finger 1 is involved in the regulation of angiogenesis: Possible contribution of stathmin/OP18 as a downstream target gene. *Arterioscler. Thromb. Vasc. Biol.* **2004**, *24*, 878–884. [[CrossRef](#)] [[PubMed](#)]
13. Gerald, D.; Adini, I.; Shechter, S.; Perruzzi, C.; Varnau, J.; Hopkins, B.; Kazerounian, S.; Kurschat, P.; Blachon, S.; Khedkar, S.; et al. RhoB controls coordination of adult angiogenesis and lymphangiogenesis following injury by regulating VEZF1-mediated transcription. *Nat. Commun.* **2013**, *4*, 2824. [[CrossRef](#)] [[PubMed](#)]
14. Aitsebaomo, J.; Wennerberg, K.; Der, C.J.; Zhang, C.; Kedar, V.; Moser, M.; Kingsley-Kallesen, M.L.; Zeng, G.Q.; Patterson, C. p68RacGAP is a novel GTPase-activating protein that interacts with vascular endothelial zinc finger-1 and modulates endothelial cell capillary formation. *J. Biol. Chem.* **2004**, *279*, 17963–17972. [[CrossRef](#)] [[PubMed](#)]
15. Gowher, H.; Stuhlmann, H.; Felsenfeld, G. *Vezf1* regulates genomic DNA methylation through its effects on expression of DNA methyltransferase *Dnmt3b*. *Genes Dev.* **2008**, *22*, 2075–2084. [[CrossRef](#)] [[PubMed](#)]
16. Carmeliet, P. Mechanisms of angiogenesis and arteriogenesis. *Nat. Med.* **2000**, *6*, 389–395. [[CrossRef](#)] [[PubMed](#)]
17. Flamme, I.; Frolich, T.; Risau, W. Molecular mechanisms of vasculogenesis and embryonic angiogenesis. *J. Cell. Physiol.* **1997**, *173*, 206–210. [[CrossRef](#)]
18. Eccles, S.A.; Box, C.; Court, W. Cell migration/invasion assays and their application in cancer drug discovery. *Biotechnol. Annu. Rev.* **2005**, *11*, 391–421. [[PubMed](#)]
19. Gasparini, G.; Longo, R.; Fanelli, M.; Teicher, B.A. Combination of antiangiogenic therapy with other anticancer therapies: Results, challenges, and open questions. *J. Clin. Oncol.* **2005**, *23*, 1295–1311. [[CrossRef](#)] [[PubMed](#)]
20. Hayden, E.C. Cutting off cancer's supply lines. *Nature* **2009**, *458*, 686–687. [[CrossRef](#)] [[PubMed](#)]
21. Pavletich, N.P.; Pabo, C.O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **1991**, *252*, 809–817. [[CrossRef](#)] [[PubMed](#)]
22. Darnell, J.E., Jr. Transcription factors as targets for cancer therapy. *Nat. Rev. Cancer* **2002**, *2*, 740. [[CrossRef](#)] [[PubMed](#)]
23. Johnston, S.J.; Carroll, J.S. Transcription factors and chromatin proteins as therapeutic targets in cancer. *Biochim. Biophys. Acta* **2015**, *1855*, 183–192. [[CrossRef](#)] [[PubMed](#)]
24. Hegde, N.S.; Sanders, D.A.; Rodriguez, R.; Balasubramanian, S. The transcription factor FOXM1 is a cellular target of the natural product thiostrepton. *Nat. Chem.* **2011**, *3*, 725–731. [[CrossRef](#)] [[PubMed](#)]

25. Vassilev, L.T.; Vu, B.T.; Graves, B.; Carvajal, D.; Podlaski, F.; Filipovic, Z.; Kong, N.; Kammlott, U.; Lukacs, C.; Klein, C.; et al. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **2004**, *303*, 844–848. [[CrossRef](#)] [[PubMed](#)]
26. Ghirlando, R.; Giles, K.; Gowher, H.; Xiao, T.; Xu, Z.; Yao, H.; Felsenfeld, G. Chromatin domains, insulators, and the regulation of gene expression. *Biochim. Biophys. Acta* **2012**, *1819*, 644–651. [[CrossRef](#)] [[PubMed](#)]

Sample Availability: The compounds can be ordered from MedPharma Partners who have standardized protocol for synthesis.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

APPENDIX B. UPPP AND UPPS

***Lipophilic efficient phenylthiazoles with potent undecaprenyl pyrophosphatase inhibitory activity**

*Originally published in *European Journal of Medicinal Chemistry*.

Elsebaei, M.; Mohammad, H.; Samir, A.; Abutaleb, A.; Moustafa, M.; **Norvil, AB**; Michie, AR; Gowher, H.; Seleem, M. Lipophilic Efficient Phenylthiazoles with Potent Undecaprenyl Pyrophosphatase Inhibitory activity. (2019) *European Journal of Medicinal Chemistry* 175, 49-62.

Declaration of collaborative work

Mohamed M Esebaei, Haroon Mohammad, Amgad Samir, Nader S Abutaleb, Mahmoud M Moustafa, Hebatallah Samy, Mohamed N Seleem, and Abdelrahman S Mayhoub designed and performed antibacterial assays, synthesized compounds, and wrote the manuscript.

Allison B Norvil and Humaira Gowher designed *in vitro* inhibition assays and assisted with writing the manuscript. Amie R Michie performed *in vitro* inhibition assays.

*** From Phenylthiazoles to Phenylpyrazoles: Broadening the Antibacterial Spectrum towards Carbapenem-Resistant Bacteria.**

*Originally published in *Journal of Medicinal Chemistry*

Hammad, A.; Abutaleb, NS.; Elsebaei, M.; **Norvil , AB**; Alswah, M.; Ali, AO.; Abdel-Aleem, JA.; Alattar, A.; Bayoumi, S.; Gowher, H.; Seleem, MN.; Mayhoub, AS. From Phenylthiazoles to Phenylpyrazoles: Broadening the Antibacterial Spectrum towards Carbapenem-Resistant Bacteria. (2019) *Journal of Medicinal Chemistry* 2019 62 (17),7998-8010.

Declaration of collaborative work

Ali Hammad, Nader S Abutaleb, Mohamed M Elsebaei, Mohamed Alswah, Alsagher O Ali, Jelan A Abdel-Aleem, Abdelaziz Alattar, Sammar A Bayoumi, Mohamed N Seleem, and Abdelrahman S Mayhoub designed and performed antibacterial assays, synthesized compounds, and wrote the manuscript.

Allison B Norvil and Humaira Gowher designed *in vitro* inhibition assays and assisted with writing the manuscript. Allison B Norvil performed *in vitro* inhibition assays.



Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: <http://www.elsevier.com/locate/ejmech>

Research paper

Lipophilic efficient phenylthiazoles with potent undecaprenyl pyrophosphatase inhibitory activity



Mohamed M. Elsebaei ^{a,1}, Haroon Mohammad ^{b,1}, Amgad Samir ^a, Nader S. Abutaleb ^b, Allison B. Norvil ^c, Amie R. Michie ^c, Mahmoud M. Moustafa ^d, Hebatallah Samy ^e, Humaira Gowher ^{c,f}, Mohamed N. Seleem ^{b,g,**}, Abdelrahman S. Mayhoub ^{a,h,*}

^a Department of Pharmaceutical Organic Chemistry, College of Pharmacy, Al-Azhar University, Cairo, 11884, Egypt

^b Department of Comparative Pathobiology, College of Veterinary Medicine, Purdue University, West Lafayette, IN, 47907, USA

^c Department of Biochemistry, College of Agriculture, Purdue University, West Lafayette, IN, 47907, USA

^d Department of Pharmaceutical Chemistry, Imam Abdulrahman Bin Faisal University, Dammam, 31441, Saudi Arabia

^e University of Science and Technology, Biomedical Sciences Program, Zewail City of Science and Technology, October Gardens, 6th of October, Giza, 12578, Egypt

^f Purdue University Center for Cancer Research, Purdue University, West Lafayette, IN, 47907, USA

^g Purdue Institute of Inflammation, Immunology, and Infectious Disease, West Lafayette, IN, 47907, USA

^h University of Science and Technology, Nanoscience Program, Zewail City of Science and Technology, October Gardens, 6th of October, Giza, 12578, Egypt

ARTICLE INFO

Article history:

Received 6 February 2019

Received in revised form

23 April 2019

Accepted 24 April 2019

Available online 27 April 2019

ABSTRACT

Antibiotic resistance remains a pressing medical challenge for which novel antibacterial agents are urgently needed. The phenylthiazole scaffold represents a promising platform to develop novel antibacterial agents for drug-resistant infections. However, enhancing the physicochemical profile of this class of compounds remains a challenging endeavor to address to successfully translate these molecules into novel antibacterial agents in the clinic. We extended our understanding of the SAR of the phenylthiazoles' lipophilic moiety by exploring its ability to accommodate a hydrophilic group or a smaller sized hetero-ring with the objective of enhancing the physicochemical properties of this class of novel antimicrobials. Overall, the 2-thienyl derivative **20** and the hydroxyl-containing derivative **31** emerged as the most promising antibacterial agents inhibiting growth of drug-resistant *Staphylococcus aureus* at a concentration as low as 1 µg/mL. Remarkably, compound **20** suppressed bacterial undecaprenyl pyrophosphatase (UppP), the molecular target of the phenylthiazole compounds, in a sub nano-molar concentration range (almost 20,000 times more potent than the lead compounds **1a** and **1b**). Compound **31** possessed the most balanced antibacterial and physicochemical profile. The compound exhibited rapid bactericidal activity against *S. aureus*, and successfully cleared intracellular *S. aureus* within infected macrophages. Furthermore, insertion of the hydroxyl group enhanced the aqueous solubility of **31** by more than 50-fold relative to the first-generation lead **1c**.

© 2019 Elsevier Masson SAS. All rights reserved.

Abbreviations: MIC, minimum inhibitory concentration; UppP, undecaprenyl diphosphate phosphatase; UppS, undecaprenyl diphosphate synthase.

* Corresponding author. Department of Pharmaceutical Organic Chemistry, College of Pharmacy, Al-Azhar University, Cairo, 11884, ASM, Egypt

** Corresponding author. Department of Comparative Pathobiology, College of Veterinary Medicine, Purdue University, West Lafayette, IN, 47907, USA

E-mail addresses: mseleem@purdue.edu (M.N. Seleem), amayhoub@azhar.edu.eg (A.S. Mayhoub).

¹ These two authors contributed equally.

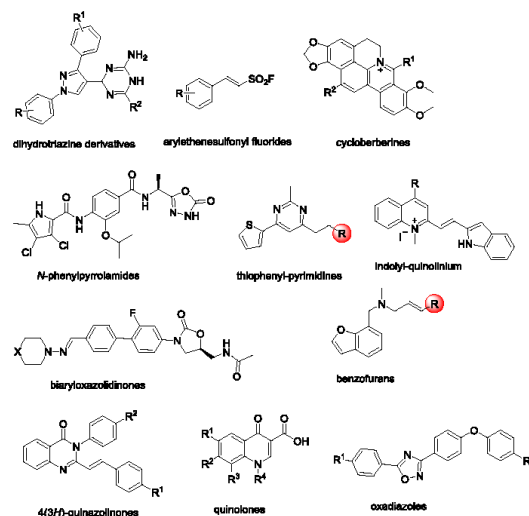
<https://doi.org/10.1016/j.ejmech.2019.04.063>

0223-5234/© 2019 Elsevier Masson SAS. All rights reserved.

1. Introduction

A recent report from the World Health Organization (WHO) indicated that the current pipeline of molecules with antibacterial activity in clinical trials is insufficient to overcome the global threat of antibiotic resistance [1]. The vast majority of new chemical entities currently tested in the antibiotic development pipeline are modifications of the existing antibiotic classes. Therefore, they are not expected to overcome the current resistance mechanisms and they only provide short-term solutions that cannot control the exponential increase in pan-resistant pathogens [1]. Consequently,

there is a dire need for novelty in the field of antibacterial discovery. Hence, the WHO established three criteria to assess innovativeness of novel chemical entities in this field: (1) agents must belong to a new chemical scaffold, (2) agent should target a new pathway that has not been targeted by existing classes of antibiotics or possess a new mechanism of action, and (3) an absence of cross-resistance to known antibiotics [1]. In this regard, several scaffolds were recently introduced as novel antibacterial agents including dihydrotriazine derivatives [2], aryloethanesulfonyl fluorides [3], cycloberberines [4], *N*-phenylpyrrolamides [5], thiophenyl-pyrimidines [6], indolyl-quinolinium derivatives [7], biaryloxazolidinones [8], benzofurans [9], 4(3*H*)-quinazolinones [10], quinolones [11], and oxadiazoles [12]. In addition, many phenylthiazoles were recently discovered by our group [13–17] as potential anti MRSA. Meeting the WHO criteria as a new, previously unexploited chemical scaffold for antibiotic development, these compounds are active against a panel of multi-drug resistance bacterial pathogens [18–21], and inhibit two enzymes involved in bacterial cell wall synthesis [21] that are not targeted by FDA-approved antibiotics currently.



In an effort to improve the antibacterial activity of the first-generation lead compound **1a**, the structure-activity-relationship (SAR) of the lipophilic moiety (colored blue in Fig. 1) was rigorously studied [13–16]. Thus far, increasing the lipophilicity of lead compound **1a** via replacement of the *n*-butyl group by more rigid acetylenyl moieties furnished compounds **1b–1d** with a notable improvement in the antibacterial activity against methicillin-resistant and vancomycin-resistant *Staphylococcus aureus* (MRSA & VRSA) (Fig. 1). In addition to ameliorating the minimum inhibitory concentration (MIC) values, subsequent generations of phenylthiazoles possessed two advantages over the glycopeptide antibiotic vancomycin (the drug of choice for treatment of invasive MRSA infections). These two features were rapid bactericidal activity against MRSA and the ability to reduce the burden of intracellular MRSA within infected macrophages [18], a significant source of recurring staphylococcal infection for afflicted patients [22]. However, these advantages came at a cost to the compounds' drug-like properties. The high increase in lipophilicity dramatically deteriorated the aqueous solubility of the phenylthiazole compounds (Fig. 1). Therefore, in this report, we investigated the effect of decreasing the lipophilicity of the compounds, by inserting a

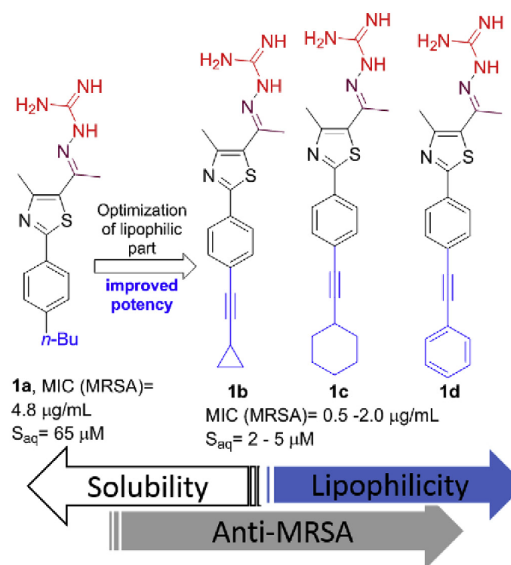


Fig. 1. Developmental progress of phenylthiazole antibacterial agents.

polar group or using smaller ring sizes, on antibacterial activity and the physicochemical profile of the new analogues, with the aim to bring this class of novel antibacterial agents closer to clinical applications.

2. Results and discussion

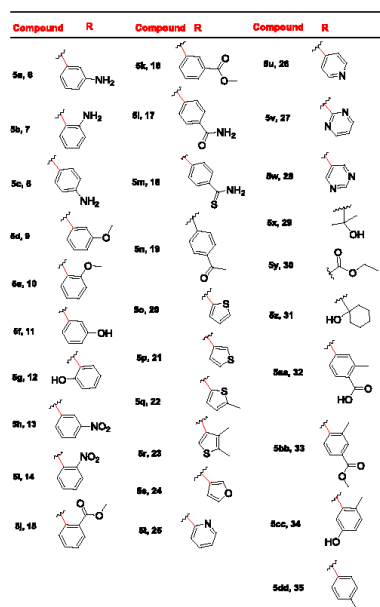
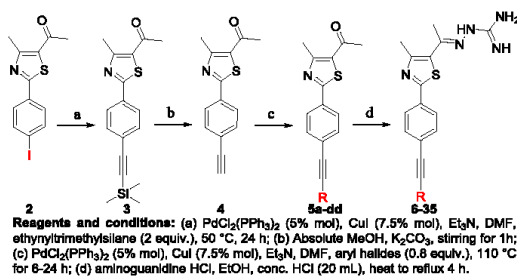
2.1. Chemistry

The starting material, *p*-iodophenylthiazole **2**, was prepared as previously reported [23]. This step was followed by two tandem Sonogashira cross-coupling conversions to obtain the final products **6–35** as depicted in Scheme 1.

2.2. Biological results and discussion

2.2.1. Antibacterial activity of new analogues against MRSA

All newly synthesized final products were initially screened against MRSA USA400, a highly virulent strain associated with many cases of community-acquired MRSA skin and soft tissue infections in North America. MRSA USA400 has also been isolated from invasive MRSA infections including necrotizing pneumonia, pulmonary abscesses, and sepsis [24–26]. The initial screening against MRSA USA400 revealed a wealth of structure-activity-relationships (SAR) as the MIC values ranged between 2 to more than 64 μ g/mL (Table 1). In general, bulkier substituents such as nitro or ester groups negated anti-MRSA activity regardless of the position of the substituent on the terminal phenyl ring. In this regard, the nitro-containing derivatives **13** (MIC = 64 μ g/mL) and **14** (MIC > 64 μ g/mL) exhibited weak activity or was inactive. The same lack of potency was also observed with all carbonyl-containing derivatives containing an ester (compounds **15**, **16** and **33**), amide (compound **17**), thioamide (compound **18**) or carboxylic acid (compound **32**). On the other hand, polar small-sized substituents were position-sensitive: hydroxyl-, methoxy- and amino-



Scheme 1..

containing derivatives at the meta position of the terminal phenyl ring (**6**, **9** and **11**) were all active against MRSA USA400 with MIC values ranging between 2 and $4\ \mu\text{g/mL}$. Moving the hydroxyl group from the meta (compound **11**) to the ortho position (compound **12**) dramatically negated the compound's antibacterial activity as the MIC value increased by a factor of 64 (Table 1).

Alternatively, decreasing the lipophilicity by replacement of the terminal phenyl ring with a heterocyclic system provided distinct SAR as high anti-MRSA potency was observed with the smaller-sized ring and the larger ones were inactive. In this vein, all derivatives with pyridine (compounds **25** (MIC = $32\ \mu\text{g/mL}$) and **26** (MIC > $64\ \mu\text{g/mL}$)) or pyrimidine (compounds **27** and **28**, MIC > $64\ \mu\text{g/mL}$) side chains were devoid of potent antibacterial activity against MRSA. On the other hand, the 2-thienyl derivatives **20** and **22** were almost as potent as the two reference antibiotics linezolid and vancomycin, inhibiting MRSA growth at a concentration of $2\ \mu\text{g/mL}$. Interestingly, shifting the sulfur atom of the terminal thiophen ring to position-3 resulted in an inactive compound (**21**, MIC > $64\ \mu\text{g/mL}$). The last tactic used to overcome the high lipophilicity issue was to insert a hydroxyl group next to the acetylene

Table 1

The minimum inhibitory concentration (MIC in $\mu\text{g/mL}$) of the tested phenylthiazole derivatives initially screened against methicillin-resistant *Staphylococcus aureus* USA 400.

Compound/Control Antibiotic	MIC ($\mu\text{g/mL}$)
6	4
7	>64
8	4
9	4
10	16
11	2
12	>64
13	64
14	>64
15	32
16	16
17	>64
18	64
19	>64
20	2
21	>64
22	2
23	16
24	4
25	32
26	>64
27	>64
28	>64
29	2
30	>64
31	4
32	>64
33	>64
34	>64
35	32
Linezolid	1
Vancomycin	1

linker by utilizing the commercially available 1-ethynylcyclohexan-1-ol and 2-methylbut-3-yn-2-ol. This resulted in two potent compounds, **29** and **31**, with MIC values of $2\ \mu\text{g/mL}$ and $4\ \mu\text{g/mL}$, respectively.

2.2.2. Antibacterial activity of new analogues against other bacterial strains

Next, we moved to examine the spectrum of antibacterial activity of six selected compounds against a panel of clinically-relevant Gram-positive and Gram-negative bacterial pathogens (Table 2 and Table 3).

In accordance with the results of the initial screening against MRSA USA400, the tested compounds exhibited potent antibacterial activity against a panel of pathogenic staphylococcal strains with MIC values ranging from 1 to $8\ \mu\text{g/mL}$. Compounds **8**, **11**, and **20** inhibited growth of *S. epidermidis* at concentrations of $1\text{--}2\ \mu\text{g/mL}$. *S. epidermidis* is an important pathogen involved in nosocomial bloodstream infections, cardiovascular infections, and infections of the eye, ear, nose and throat [25]. Additionally, *S. epidermidis* has been associated with infections on foreign bodies, such as indwelling catheters and implanted medical devices due its ability to form strong biofilms [27]. The six phenylthiazole analogues also inhibited growth of *L. monocytogenes* and *S. pneumoniae* isolates at concentrations as low as $2\ \mu\text{g/mL}$ and inhibited growth of vancomycin-resistant *E. faecalis* at concentrations as low as $4\ \mu\text{g/mL}$. The hydroxyl-containing derivatives **29** and **31** were the most potent compounds against the bacterial pathogens tested with MIC values ranging from 1 to $8\ \mu\text{g/mL}$.

The MBC values for both compounds **29** and **31** were found to be equal to or up to two-fold higher than the compounds' MIC values

Table 2
The minimum inhibitory concentration (MIC in $\mu\text{g/mL}$) and minimum bactericidal concentration (MBC in $\mu\text{g/mL}$) of the tested compounds, linezolid, and vancomycin against a panel of Gram-positive bacterial pathogens including methicillin-sensitive *Staphylococcus aureus*, methicillin-resistant *Staphylococcus aureus* (MRSA), vancomycin-resistant *Staphylococcus epidermidis*, *Enterococcus faecalis*, *E. faecium*, *Listeria monocytogenes*, and *Streptococcus pneumoniae*.

	Compounds/Control antibiotics											
	8		11		20		22		29		31	
	MIC	MBC	MIC	MBC	MIC	MBC	MIC	MBC	MIC	MBC	MIC	MBC
<i>Staphylococcus aureus</i>												
ATCC 6538	2	4	2	4	1	2	4	4	2	4	2	4
<i>Staphylococcus aureus</i>												
NRS107	4	8	4	4	2	4	4	4	4	4	4	4
MRSA	8	16	2	4	2	4	4	4	2	4	2	4
NRS119	4	16	4	4	4	4	8	16	4	16	4	16
MRSA	4	16	4	4	4	4	8	16	4	16	4	16
NRS384 (USA300)	16	32	4	4	8	16	4	8	4	16	2	8
VRS10 (VISA)	16	32	4	4	1	2	4	8	1	4	1	8
VRS11a (VISA)	16	32	4	4	1	2	4	8	1	4	1	8
VRS12 (VISA)	16	32	4	4	8	16	8	16	4	16	4	16
Methicillin-resistant <i>Staphylococcus epidermidis</i>	2	2	2	2	1	1	4	4	4	4	4	4
NRS101	>64	>64	32	64	64	>64	16	64	4	8	8	16
<i>Enterococcus faecalis</i> ATCC 51299 (VRE) ^a	4	16	<0.5	1	4	8	2	4	2	4	1	1
<i>Enterococcus faecium</i> ATCC 700221 (VRE)	32	>64	8	64	8	32	8	32	4	8	4	8
<i>Listeria monocytogenes</i> ATCC 19111	32	32	8	32	4	4	8	16	4	8	8	8
Cephalosporin-resistant <i>Streptococcus pneumoniae</i> ATCC 51916	16	32	16	32	4	8	8	32	4	8	4	8
Methicillin-resistant <i>Streptococcus pneumoniae</i> ATCC 700677												
<i>Vancomycin-resistant enterococci</i>												

^a VRE: Vancomycin-resistant enterococci.

against the tested bacterial strains, indicating the bactericidal effect of the tested compounds. This expected mode of action was further assessed via a time-kill assay against MRSA (Fig. 2). The hydroxycyclohexyl derivative **31** exhibited rapid bactericidal action, completely eradicating MRSA USA400 CFU within 6 h. In contrast, the *p*-anilino and 2-thienyl derivatives, **8** and **22**, exhibited slow bactericidal action, similar to vancomycin, as they reduced MRSA CFU by 3-log₁₀ within 12 h and completely eradicated MRSA after 24 h.

In general, the tested compounds were inactive (MIC > 64 $\mu\text{g/mL}$) or exhibited poor activity (MIC between 32 and 64 $\mu\text{g/mL}$) against all four Gram-negative bacterial pathogens examined. This result was similar to the antibiotic erythromycin, which is known to be a substrate for the TolC-AcrAB efflux pump present in *E. coli* [28]. To examine if the antibacterial activity of the phenylthiazole analogues was similarly impeded by the presence of an efflux pump, we examined the six compounds' antibacterial activity against a mutant *E. coli* strain (JW55031) deficient in TolC expression. There was a modest improvement observed in the antibacterial activity of the tested compounds. However, the MIC values for the phenylthiazole derivatives were much higher than those observed for erythromycin. Therefore, the observed inactivity of the tested compounds against Gram-negative pathogens is likely due to the inability to cross the outer membrane in these bacteria rather than being a substrate for the TolC efflux pump.

We next investigated the ability of MRSA to develop resistance to the newly synthesized phenylthiazoles via a multi-step resistance selection assay. As presented in Fig. 3, there was no observed increase in the MIC for compound **31** over 14 passages. Only a one-fold increase in MIC was observed for both compounds **8** and **22** after the sixth passage. In contrast, MRSA developed resistance rapidly to the antibiotic rifampicin (in agreement with reports from literature) [29–31] as the MIC of the antibiotic increased 29-fold after only one passage and continued to increase rapidly thereafter (>500,000-fold increase in MIC by the ninth passage). The results indicated that MRSA USA400 was unable to acquire resistance to the newly synthesized phenylthiazoles, similar to the first-generation compounds.

As highlighted in the introduction, an important feature of the phenylthiazoles that is advantageous over existing front-line therapeutics, such as vancomycin, is the ability of compounds with this new scaffold to penetrate and reduce the burden of *S. aureus* within infected macrophages. *S. aureus* is capable of invading and surviving inside mammalian macrophages which can result in chronic and persistent infections. Many antibiotics cannot access these infected macrophages at the optimum concentration to clear the infection. Therefore, treatment of intracellular *S. aureus* infections with conventional antibiotics is a great challenge [32,33].

Examination of the activity of compounds **8**, **11**, **22** and **31** on the clearance of intracellular MRSA present in murine macrophage (J774) cells. Data are presented as log₁₀ colony forming units of MRSA USA400 per mL inside infected murine macrophages after treatment with 16 $\mu\text{g/mL}$ of compound **8**, **11**, **22**, **31** or vancomycin (tested in quadruplicates) for 24 h. A one-way ANOVA with post-hoc Dennett's test was used to determine statistical difference between the values obtained for each compound/vancomycin relative to DMSO (*, $P < 0.05$).

The toxicity profile of the most promising compounds against the macrophage J774 cell line was evaluated. Compounds **1d**, **20**, and **29** were toxic to macrophage cells even at a concentration as low as 16 $\mu\text{g/mL}$ (Fig. 1S) so they were excluded from the intracellular infection experiment. Compounds **8**, **11**, **22**, and **31** were non-toxic to macrophage cells exposed to 16 $\mu\text{g/mL}$ of the analogues over 24 h (compounds **8** and **11** were non-toxic up to 32 $\mu\text{g/mL}$). Due to their superior toxicity profile, compounds **8**, **11**, **22**, and **31**

Table 3

The minimum inhibitory concentration (MIC in $\mu\text{g/mL}$) and minimum bactericidal concentration (MBC in $\mu\text{g/mL}$) of the tested compounds and control antibiotics against Gram-negative bacterial pathogens including *Acinetobacter baumannii*, *Escherichia coli*, *Klebsiella pneumoniae*, and *Pseudomonas aeruginosa*.

Compound/Antibiotic	Bacterial Strains									
	<i>Acinetobacter baumannii</i> ATCC 19606		<i>Escherichia coli</i> BW25113 ^a		<i>Escherichia coli</i> JW55031 ^b		<i>Klebsiella pneumoniae</i> BAA-1706		<i>Pseudomonas aeruginosa</i> ATCC 15442	
	MIC	MBC	MIC	MBC	MIC	MBC	MIC	MBC	MIC	MBC
8	64	>64	>64	>64	16	16	>64	>64	>64	>64
11	>64	>64	>64	>64	64	>64	>64	>64	>64	>64
20	>64	>64	>64	>64	64	>64	>64	>64	>64	>64
22	32	64	64	>64	32	64	32	64	>64	>64
29	64	64	>64	>64	16	16	>64	>64	>64	>64
31	64	64	>64	>64	32	32	>64	>64	>64	>64
Colistin	≤ 0.5	≤ 0.5	≤ 0.5	≤ 0.5	≤ 0.5	≤ 0.5	≤ 0.5	≤ 0.5	1	1
Erythromycin	8	16	32	>64	1	2	64	>64	>64	>64

^a Wild-type strain.

^b ΔtolC mutant.

were investigated for their ability to clear intracellular MRSA. A concentration of 16 $\mu\text{g/mL}$ was selected, as all four compounds were non-toxic at this concentration. Vancomycin, given its high molecular weight and complex structure, is one antibiotic unable to sufficiently accumulate inside macrophage cells and thus is unable to clear an intracellular MRSA infection. As depicted in Fig. 4, after 24 h, compound **31** completely eradicated MRSA inside the infected macrophages. Compounds **11** and **22** generated a 0.92- \log_{10} reduction and 0.21- \log_{10} reduction in intracellular MRSA, respectively. Compound **8** was unable to reduce the burden of MRSA inside infected macrophages. Vancomycin, as expected, was not able to reduce the presence of MRSA inside infected J774 cells, even after 24 h of treatment. The results collectively indicated that compounds **11**, **22**, and **31** had the ability to gain entry into macrophage cells at a concentration high enough to significantly reduce the burden of MRSA inside infected macrophages.

2.2.3. Investigating physicochemical properties

After confirming the potent *in vitro* activity of the newly synthesized derivatives against MRSA, we moved to evaluate the physicochemical profile of the compounds. Aqueous solubility is a critical attribute in drug development as it controls the pharmacokinetic profile and route of administration for a drug [34,35]. For instance, highly lipophilic compounds, in general, bind tightly with plasma proteins, possess higher affinity to liver microsomes and are distributed widely. Highly lipophilic active pharmaceutical ingredients (APIs) present an immense challenge in the formulation

stage [32]. In this regard, APIs with a log P value of four or more are hard to put in an oral dosage form [36]. Consequently, to investigate the effect of the chemical modifications reported here on the physicochemical properties of the phenylthiazole scaffold, all promising compounds were subjected to a turbidometric solubility assay. As presented in Table 4, the most effective approach to enhance the aqueous solubility of the phenylthiazole compounds was the insertion of a hydroxyl group. This modification decreased the partition coefficient from 5.2 in the case of the lead compound **1c** to 3.1 in the case of **31** and enhanced the aqueous solubility of this compound 55-fold.

2.2.4. Biochemical assay

Previously, we found that the lead phenylthiazole **1a** exerted its antibacterial effect by targeting undecaprenyl diphosphate phosphatase, UppP ($\text{IC}_{50} = 6 \mu\text{M}$), a key enzyme involved in cell wall synthesis [21]. In addition, compound **1a** also inhibited undecaprenyl diphosphate synthase UppS ($\text{IC}_{50} = 19 \mu\text{M}$), albeit to a lower extent. Similar IC_{50} values were reported with the cycloacetylenyl derivative **1b** [14]. Therefore, in order to explore whether the new series of phenylthiazoles possess the same antibacterial mode of action or not, all active compounds were assayed for their inhibitory activities against both enzymes. Unexpectedly, none of the tested compounds showed any significant inhibitory effect on UppS. However, two thiophen-containing compounds, **20** and **22**, were significantly more potent inhibitors of UppP with IC_{50} values

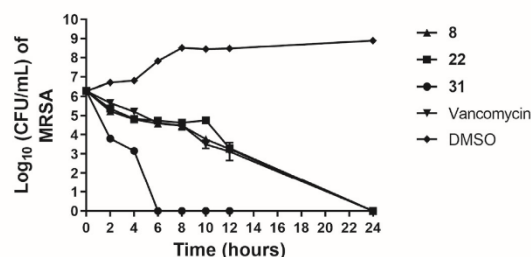


Fig. 2. Time-kill analysis of tested compounds and vancomycin (all at $4 \times \text{MIC}$) against methicillin-resistant *Staphylococcus aureus* (MRSA USA400) over a 24-h incubation period at 37°C. DMSO (solvent for the compounds) served as a negative control. The error bars represent standard deviation values obtained from triplicate samples used for each compound/antibiotic studied.

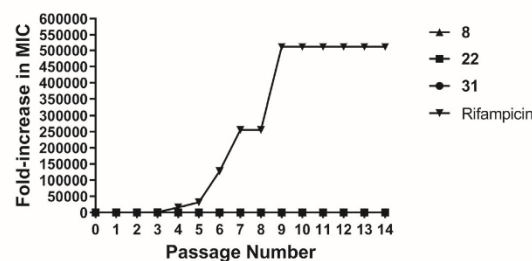


Fig. 3. Multi-step resistance selection of tested compounds and rifampicin against methicillin-resistant *S. aureus* USA400 (NRS123). Bacteria were serially passaged over 14 days, and the broth microdilution assay was used to determine the minimum inhibitory concentration of each compound against MRSA after each successive passage. A four-fold shift in MIC would be indicative of bacterial resistance having formed to the test agent.

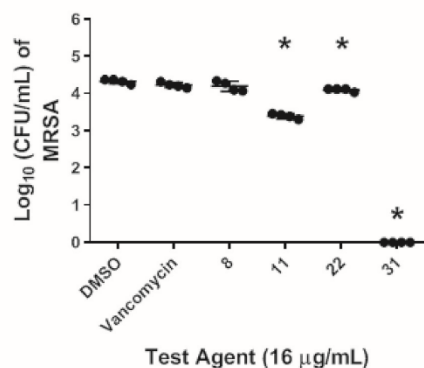


Fig. 4. Examination of clearance of intracellular MRSA USA400 (NRS123) present in infected macrophage (J774) cells.

Table 4
Evaluation of solubility in phosphate buffered saline (PBS).

Compound/Drug	Solubility limit (μM)	cLogP
1a	65	4.4
1c	2.7	5.2
1d	2.5	5.0
8	105	3.7
11	55	4.3
20	50	4.6
22	12	5.1
29	320	1.8
31	147	3.1
Tamoxifen	15.6	
Verapamil	>500	

Solubility limit corresponds to the highest concentration of test compound where no precipitate was detected (OD_{540}).

of 0.3 and 6 nM, respectively (Fig. 5). These two values are 1000 to 20,000 times more potent than the lead compound 1a.

In order to rationalize the result of the biochemical assay, a computational study has been conducted to predict the binding poses of the most potent derivatives obtained so far 20 and 22 in comparison with the lead compound 1a. In general, the thiophene derivatives 20 and 22 have shown more negative docking scores than the lead compound 1a (Table 1S), indicating a more stable

binding with the enzyme. This is attributed to the formation of more interactions with the residues of the active site (Fig. 2S). In all the cases, the pharmacophoric guanidine moiety was involved in the formation of a hydrogen bond with a hydrophilic residue in the UppP active pocket (either Ser173 or Glu21). However, in the thiophene-containing derivatives 20 and 22, the aromatic thiophene ring is also involved in an arene-H interaction with Val103 in the hydrophobic portion of the pocket. This is due to the geometrical arrangement of the CH bond of the side chain of Val103 in a direct line with a p-orbital in the thiophene ring. Additionally, the values of Van der Waals volume of the molecules are shown in Table 2S. The thiophene derivatives are found to have higher Van der Waals volume which would result in a better interaction with the hydrophobic portion of the active site of UppP, resulting in a better binding.

3. Conclusion

It is well-known in drug development that escalating the lipophilic property of small molecules is always accompanied by pharmacokinetic drawbacks and dosage-form formulation challenges. The present study addressed the problematic increase in the lipophilic property of our recently developed phenylthiazole antibacterial compounds by exploring the capacity of the lipophilic moiety of the phenylthiazole scaffold to accommodate polar groups. Inclusion of small polar groups at the meta-position of the terminal phenyl enhanced both the antibacterial potency and aqueous solubility profile of the compounds. Alternatively, replacing the terminal phenyl group with smaller sized hetero rings provided thiophene-containing analogues that, in addition to their potent antibacterial effect, suppressed undecaprenyl diphosphate phosphatase in the nanomolar range, for the first time. To the best of our knowledge, the IC_{50} reported here for compound 20 is the lowest ever against UppP in the literature. An excellent balance between antibacterial potency and improving solubility was obtained by inserting a hydroxyl group next to the acetylene bridge. In this vein, compound 31 was active against fourteen highly-resistant clinical isolates of MRSA, *S. epidermidis*, *S. pneumoniae*, *L. monocytogenes* and vancomycin-resistant enterococci. This promising antibacterial effect was accompanied with a notable enhancement in aqueous solubility. The impact of the inserted hydroxyl group was extended to the mode of inhibition in which compound 31 cleared MRSA cells in only 6 h compared with the bacteriostatic effect of compound 1c. Interestingly, the *in vitro* activity of compound 31 was superior to vancomycin due to the rapid bactericidal effect and its ability to eradicate intracellular MRSA.

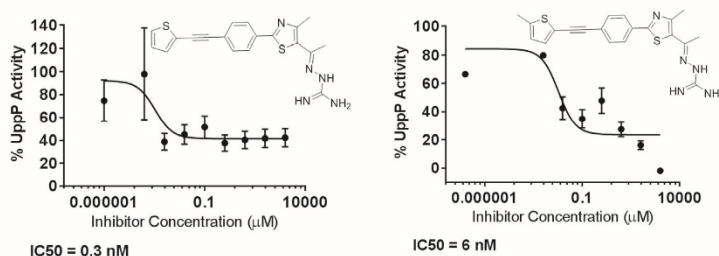


Fig. 5. Dose response curves for inhibition of undecaprenyl diphosphate phosphatase (UppP) by compounds 20 and 22. The UppP inhibition assay was performed as indicated in the methods section. The activity of UppP was measured by the release of inorganic phosphate in the presence of increasing concentration of inhibitors. Error bars represent standard deviation values obtained from duplicate samples used for each compound studied.

4. Experimental

4.1. Chemistry

4.1.1. General

¹H NMR spectra were run at 400 MHz and ¹³C spectra were determined at 100 MHz in deuterated chloroform (CDCl₃) or dimethyl sulfoxide (DMSO-*d*₆) on a Varian Mercury VX-400 NMR spectrometer. Chemical shifts are given in parts per million (ppm) on the delta (δ) scale. Chemical shifts were calibrated relative to those of the solvents. Flash chromatography was performed on 230–400 mesh silica. The progress of reactions was monitored with Merck silica gel IB2–F plates (0.25 mm thickness). Mass spectra were recorded at 70 eV. High resolution mass spectra for all ionization techniques were obtained from a Finnigan MAT XL95. Melting points were determined using capillary tubes with a Stuart SMP30 apparatus and are uncorrected. All yields reported refer to isolated yields.

4.1.2. 1-(4-Methyl-2-(4-((trimethylsilyl)ethynyl)phenyl)thiazol-5-yl)ethan-1-one (3)

4.1.2.1. Method A. To dry DMF (3 mL) in a 25-mL sealed tube, compound **2** (500 mg, 1.45 mmol), ethynyltrimethylsilane (0.4 mL, 2.91 mmol), and triethylamine (1 mL) were added. After the reaction mixture was purged with dry nitrogen gas for 20 min, dichlorobis (triphenylphosphine)palladium (II) (52 mg, 0.072 mmol) and copper (I) iodide (28 mg, 0.145 mmol) were added. The sealed tube was then placed in an oil bath and stirred at 65 °C for 2 h. After cooling down to room temperature, the reaction mixture was passed through celite, followed by chloroform (2 × 50 mL). The organic materials were then concentrated under reduced pressure. The crude materials were purified washed by silica gel flash column chromatography using hexane-ethyl acetate (9:1).

4.1.2.2. Method B. To dry DME (2 mL) in a 25-mL sealed tube compound **2** (500 mg, 1.45 mmol), ethynyltrimethylsilane (0.4 mL, 2.91 mmol), triethylamine (2 mL). After the reaction mixture was purged with dry nitrogen gas for 20 min, dichlorobis (triphenylphosphine)palladium (II) (52 mg, 0.072 mmol) and copper (I) iodide (28 mg, 0.145 mmol) were added. The sealed tube was then placed in an oil bath and stirred at 65 °C for 2 h. After cooling down to room temperature, the reaction mixture was passed through celite, followed by chloroform (2 × 50 mL). The organic materials were then concentrated under reduced pressure. The crude materials were purified washed by silica gel flash column chromatography using hexane-ethyl acetate (9:1). Yield, physical properties and spectral data of isolated purified product are listed below:

Light-brown solid (430 mg, 94%) m. p: 79–80 °C. ¹H NMR (DMSO-*d*₆) δ: 7.94 (d, *J* = 8.4 Hz, 2H), 7.55 (d, *J* = 8.4 Hz, 2H), 2.78 (s, 3H), 2.58 (s, 3H), 0.27 (s, 9H); Anal. Calc. for: C₁₇H₁₉NOSSi (313): C, 65.13; H, 6.11; N, 4.47%; Found: C, 65.12; H, 6.12; N, 4.45%.

4.1.3. 1-(2-(4-Ethynylphenyl)-4-methylthiazol-5-yl)ethan-1-one (4)

To dry methanol (10 mL) in a 25-mL round flask, compound **3** (430 mg, 1.37 mmol) and anhydrous potassium carbonate (20 mg, 0.137 mmol) were added. The reaction mixture was stirred at room temperature for 1 h until an off-white precipitate formed, and then concentrated under reduced pressure. The reaction was then quenched with very dilute hydrochloric acid and extracted with diethyl ether and then dried over anhydrous MgSO₄, the solvent evaporated under reduced pressure. Off-white solid (325 mg, 98%) m. p: 102–103 °C. ¹H NMR (DMSO-*d*₆) δ: 7.96 (d, *J* = 8.4 Hz, 2H), 7.59 (d, *J* = 8.4 Hz, 2H), 3.23 (s, 1H), 2.79 (s, 3H), 2.58 (s, 3H); ¹³C NMR

(DMSO-*d*₆) δ: 191.08, 167.79, 158.67, 133.34, 132.75, 127.38, 126.81, 124.93, 83.95, 83.28, 30.88, 18.61; Anal. Calc. for: C₁₄H₁₁NOS (241): C, 69.68; H, 4.59; N, 5.80%; Found: C, 69.69; H, 4.58; N, 5.80%.

4.1.4. 1-(2-(4-((Substitutedphenyl)ethynyl)phenyl)-4-methylthiazol-5-yl)ethan-1-one (5a-dd)

General procedure: to dry DME (3 mL) in a 25-mL sealed tube, compound **4** (200 mg, 0.82 mmol), substituted aryl halides (0.66 mmol), and triethylamine (2 mL) were added. After the reaction mixture was purged with dry nitrogen gas for 20 min, dichlorobis (triphenylphosphine)palladium (II) (30 mg, 0.041 mmol) and copper (I) iodide (16 mg, 0.082 mmol) were added. The sealed tube was then heated and stirred at 85 °C for 20 h. After cooling down to room temperature, the reaction mixture was passed through celite, followed by chloroform (2 × 50 mL). The organic materials were then concentrated under reduced pressure. The crude materials were purified washed by silica gel flash column chromatography using hexane-ethyl acetate (9:1). Yield, physical properties and spectral data of isolated purified products are listed below:

4.1.4.1. 1-(2-(4-((3-Aminophenyl)ethynyl)phenyl)-4-methylthiazol-5-yl)ethan-1-one (5a). Yellow oil (220 mg, 80%); ¹H NMR (DMSO-*d*₆) δ: 8.05 (d, *J* = 8.8 Hz, 2H), 7.67 (d, *J* = 8.4 Hz, 2H), 7.09 (t, *J* = 8.0 Hz, 1H), 6.76 (s, 1H), 6.72 (d, *J* = 7.2 Hz, 1H), 6.64 (d, *J* = 7.2 Hz, 1H), 5.31 (brs, 2H), 2.73 (s, 3H), 2.61 (s, 3H); ¹³C NMR (DMSO-*d*₆) δ: 191.21, 167.78, 158.76, 150.73, 142.37, 137.36, 133.21, 133.07, 132.94, 128.03, 127.31, 124.56, 124.32, 92.06, 88.06, 30.90, 18.62; Anal. Calc. for: C₂₀H₁₆N₂OS (332): C, 72.26; H, 4.85; N, 8.43%; Found: C, 72.28; H, 4.84; N, 8.44%.

4.1.4.2. 1-(2-(4-((2-Aminophenyl)ethynyl)phenyl)-4-methylthiazol-5-yl)ethan-1-one (5b). Orange oil (200 mg, 73%); ¹H NMR (DMSO-*d*₆) δ: 8.20 (d, *J* = 8.0 Hz, 2H), 7.99 (d, *J* = 8.4 Hz, 2H), 7.95 (d, *J* = 8.4 Hz, 1H), 7.84 (t, *J* = 7.2 Hz, 1H), 7.71 (d, *J* = 8.4 Hz, 1H), 7.57 (d, *J* = 8.0 Hz, 1H), 5.79 (brs, 2H), 2.61 (s, 3H), 2.32 (s, 3H); ¹³C NMR (DMSO-*d*₆) δ: 191.18, 168.18, 158.76, 150.73, 133.68, 132.66, 132.52, 132.50, 131.90, 130.75, 129.29, 129.17, 127.17, 92.06, 88.06, 30.95, 18.62; Anal. Calc. for: C₂₀H₁₆N₂OS (332): C, 72.26; H, 4.85; N, 8.43%; Found: C, 72.28; H, 4.84; N, 8.44%.

4.1.4.3. 1-(2-(4-((4-Aminophenyl)ethynyl)phenyl)-4-methylthiazol-5-yl)ethan-1-one (5c). Yellow oil (250 mg, 90.9%); ¹H NMR (DMSO-*d*₆) δ: 8.00 (d, *J* = 8.4 Hz, 2H), 7.58 (d, *J* = 8.4 Hz, 2H), 7.25 (d, *J* = 8.4 Hz, 2H), 7.59 (d, *J* = 8.4 Hz, 2H), 5.66 (brs, 2H), 2.71 (s, 3H), 2.58 (s, 3H); ¹³C NMR (DMSO-*d*₆) δ: 191.15, 168.20, 158.73, 150.42, 133.32, 132.79, 132.01, 131.33, 127.20, 126.91, 114.11, 107.99, 95.14, 86.85, 30.91, 18.65; Anal. Calc. for: C₂₀H₁₆N₂OS (332): C, 72.26; H, 4.85; N, 8.43%; Found: C, 72.27; H, 4.84; N, 8.43%.

4.1.4.4. 1-(2-(4-((3-Methoxyphenyl)ethynyl)phenyl)-4-methylthiazol-5-yl)ethan-1-one (5d). Green oil (240 mg, 83.6%); ¹H NMR (DMSO-*d*₆) δ: 7.95 (d, *J* = 8.4 Hz, 2H), 7.65 (d, *J* = 8.0 Hz, 2H), 7.37 (t, *J* = 8.0 Hz, 1H), 7.17 (d, *J* = 8.0 Hz, 1H), 7.13 (s, 1H), 7.02 (d, *J* = 6.4 Hz, 1H), 3.80 (s, 3H), 2.60 (s, 3H), 2.32 (s, 3H); ¹³C NMR (DMSO-*d*₆) δ: 191.11, 161.61, 160.05, 159.69, 148.77, 143.17, 136.42, 133.55, 132.58, 130.42, 126.33, 124.30, 123.72, 116.58, 91.69, 89.31, 55.73, 18.67, 16.54; Anal. Calc. for: C₂₁H₁₇NO₂S (347): C, 72.60; H, 4.93; N, 4.03%; Found: C, 72.61; H, 4.94; N, 4.02%.

4.1.4.5. 1-(2-(4-((2-Methoxyphenyl)ethynyl)phenyl)-4-methylthiazol-5-yl)ethan-1-one (5e). Yellowish oil (245 mg, 85%); ¹H NMR (DMSO-*d*₆) δ: 8.05 (d, *J* = 8.4 Hz, 2H), 7.66 (d, *J* = 8.1 Hz, 2H), 7.52 (d, *J* = 7.5 Hz, 1H), 7.43 (t, *J* = 7.6 Hz, 1H), 7.12 (d, *J* = 8.4 Hz, 1H), 6.99 (d, *J* = 8.4 Hz, 1H), 3.87 (s, 3H), 2.72 (s, 3H), 2.48 (s, 3H); ¹³C

NMR (DMSO- d_6) δ : 191.11, 161.76, 160.18, 159.93, 136.21, 133.61, 133.28, 132.41, 131.14, 126.34, 124.31, 121.02, 111.89, 111.49, 93.08, 88.68, 56.02, 18.67, 16.58; Anal. Calc. for: $C_{21}H_{17}NO_2S$ (347): C, 72.60; H, 4.93; N, 4.03%; Found: C, 72.65; H, 4.97; N, 4.01%.

4.1.4.6. 1-(2-(4-((3-Hydroxyphenyl)ethynyl)phenyl)-4-methylthiazol-5-yl)ethan-1-one (**5f**). Orange oil (220 mg, 79.7%); 1H NMR (DMSO- d_6) δ : 7.94 (d, J = 8.4 Hz, 2H), 7.64 (d, J = 8.4 Hz, 2H), 7.26 (t, J = 8.0 Hz, 1H), 7.01 (d, J = 7.6 Hz, 1H), 6.94 (s, 1H), 6.86 (d, J = 8.0 Hz, 1H), 6.00 (brs, 1H), 2.60 (s, 3H), 2.33 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 191.21, 161.83, 157.87, 149.09, 143.51, 133.41, 132.57, 130.41, 126.33, 123.91, 123.37, 122.77, 118.24, 117.02, 91.91, 88.88, 18.67, 16.69; Anal. Calc. for: $C_{20}H_{15}NO_2S$ (333): C, 72.05; H, 4.54; N, 4.20%; Found: C, 72.02; H, 4.52; N, 4.17%.

4.1.4.7. 1-(2-(4-((2-hydroxyphenyl)ethynyl)phenyl)-4-methylthiazol-5-yl)ethan-1-one (**5g**). Light brown oil (200 mg, 72%); 1H NMR (DMSO- d_6) δ : 7.95 (d, J = 8.4 Hz, 2H), 7.71 (d, J = 8.4 Hz, 2H), 7.66 (d, J = 8.4 Hz, 1H), 7.57 (t, J = 6.4 Hz, 1H), 7.29 (t, J = 7.2 Hz, 1H), 7.19 (t, J = 8.0 Hz, 1H), 5.70 (brs, 1H), 2.60 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 191.11, 167.63, 161.11, 160.09, 152.23, 147.96, 143.19, 136.91, 134.57, 133.71, 132.92, 126.36, 121.67, 115.22, 94.78, 83.38, 30.51, 18.69; Anal. Calc. for: $C_{20}H_{15}NO_2S$ (333): C, 72.05; H, 4.54; N, 4.20%; Found: C, 72.04; H, 4.53; N, 4.19%.

4.1.4.8. 1-(4-Methyl-2-(4-((3-nitrophenyl)ethynyl)phenyl)thiazol-5-yl)ethan-1-one (**5h**). Red oil (250 mg, 83%); 1H NMR (DMSO- d_6) δ : 8.39 (s, 1H), 8.28 (d, J = 7.2 Hz, 1H), 8.04 (d, J = 8.0 Hz, 2H), 7.98 (d, J = 8.4 Hz, 1H), 7.77 (d, J = 8.0 Hz, 2H), 7.75 (t, J = 8.4 Hz, 1H), 2.61 (s, 3H), 2.33 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 191.21, 167.78, 158.76, 150.73, 142.37, 137.36, 133.21, 133.07, 132.94, 128.03, 127.31, 124.56, 124.32, 91.76, 88.06, 30.90, 18.62; Anal. Calc. for: $C_{20}H_{14}N_2O_5S$ (362): C, 66.29; H, 3.89; N, 7.73%; Found: C, 66.27; H, 3.86; N, 7.75%.

4.1.4.9. 1-(4-Methyl-2-(4-((2-nitrophenyl)ethynyl)phenyl)thiazol-5-yl)ethan-1-one (**5i**). Brown oil (230 mg, 76%); 1H NMR (DMSO- d_6) δ : 8.20 (d, J = 8.0 Hz, 1H), 7.99 (d, J = 8.4 Hz, 2H), 7.91 (d, J = 7.2 Hz, 1H), 7.84 (t, J = 7.2 Hz, 1H), 7.71 (d, J = 8.4 Hz, 2H), 7.68 (t, J = 8.0 Hz, 1H), 2.61 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 191.28, 160.25, 148.69, 142.93, 136.89, 135.02, 134.72, 132.89, 130.51, 126.43, 125.36, 122.87, 117.22, 96.16, 86.92, 18.68, 16.47; Anal. Calc. for: $C_{20}H_{14}N_2O_5S$ (362): C, 66.29; H, 3.89; N, 7.73%; Found: C, 66.33; H, 3.91; N, 7.77%.

4.1.4.10. Methyl-3-((4-(5-acetyl-4-methylthiazol-2-yl)phenyl)ethynyl)benzoate (**5j**). Green oil (265 mg, 85%); 1H NMR (DMSO- d_6) δ : 8.11 (s, 1H), 8.01 (d, J = 7.2 Hz, 1H), 7.97 (d, J = 8.0 Hz, 2H), 7.87 (d, J = 7.6 Hz, 1H), 7.71 (d, J = 8.0 Hz, 2H), 7.64 (t, J = 7.6 Hz, 1H), 3.89 (s, 3H), 2.61 (s, 3H), 2.34 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 191.23, 167.77, 158.74, 150.33, 142.37, 137.36, 133.21, 133.17, 132.91, 131.08, 128.13, 127.31, 124.56, 124.32, 119.37, 92.16, 88.07, 47.57, 30.90, 18.62; Anal. Calc. for: $C_{22}H_{17}NO_3S$ (375): C, 70.38; H, 4.56; N, 3.73%; Found: C, 70.41; H, 4.54; N, 3.70%.

4.1.4.11. Methyl-2-((4-(5-acetyl-4-methylthiazol-2-yl)phenyl)ethynyl)benzoate (**5k**). Yellow oil (250 mg, 80%); 1H NMR (DMSO- d_6) δ : 8.21 (d, J = 8.0 Hz, 2H), 8.03 (d, J = 8.0 Hz, 1H), 7.97 (d, J = 6.8 Hz, 1H), 7.91 (d, J = 8.0 Hz, 2H), 7.74 (t, J = 8.0 Hz, 1H), 7.65 (t, J = 6.8 Hz, 1H), 4.03 (s, 3H), 2.62 (s, 3H), 2.33 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 191.21, 167.78, 160.18, 149.33, 148.52, 142.92, 133.67, 133.32, 132.44, 129.73, 128.63, 126.19, 124.67, 122.70, 119.37, 92.76, 88.15, 55.60, 18.67, 16.47; Anal. Calc. for: $C_{22}H_{17}NO_3S$ (375): C, 70.38; H, 4.56; N, 3.73%; Found: C, 70.36; H, 4.54; N, 3.70%.

4.1.4.12. 4-((4-(Acetyl-4-methylthiazol-2-yl)phenyl)ethynyl)benzamide (**5l**). Yellow oil (270 mg, 90.6%); 1H NMR (DMSO- d_6) δ : 8.01 (d, J = 8.4 Hz, 2H), 7.96 (d, J = 8.0 Hz, 2H), 7.72 (d, J = 8.4 Hz, 2H), 7.68 (d, J = 8.0 Hz, 2H), 4.36 (brs, 2H), 2.60 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 191.18, 167.85, 167.50, 158.74, 134.88, 133.71, 133.16, 132.70, 131.84, 129.28, 128.35, 127.31, 125.18, 91.86, 90.99, 30.93, 18.65; Anal. Calc. for: $C_{21}H_{16}N_2O_2S$ (360): C, 69.78; H, 4.47; N, 7.77%; Found: C, 69.79; H, 4.48; N, 7.78%.

4.1.4.13. 4-((4-(Acetyl-4-methylthiazol-2-yl)phenyl)ethynyl)benzothioamide (**5m**). Yellow oil (290 mg, 92.9%); 1H NMR (DMSO- d_6) δ : 8.09 (d, J = 8.4 Hz, 2H), 7.94 (d, J = 8.4 Hz, 2H), 7.80 (d, J = 8.4 Hz, 2H), 7.63 (d, J = 8.4 Hz, 2H), 7.63 (brs, 2H), 2.73 (s, 3H), 2.59 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 191.22, 167.46, 158.75, 133.28, 133.14, 133.06, 133.02, 132.75, 127.34, 127.21, 124.62, 118.86, 111.78, 92.99, 90.84, 30.94, 18.64; Anal. Calc. for: $C_{21}H_{16}N_2OS_2$ (376): C, 67.00; H, 4.28; N, 7.44%; Found: C, 67.02; H, 4.29; N, 7.46%.

4.1.4.14. 1-(4-((4-(5-Acetyl-4-methylthiazol-2-yl)phenyl)ethynyl)phenyl)ethan-1-one (**5n**). Yellow powder (500 mg, 95.6%); m. p: 125–127; 1H NMR (DMSO- d_6) δ : 8.07 (d, J = 8.0 Hz, 2H), 7.98 (d, J = 8.0 Hz, 2H), 7.70 (d, J = 8.0 Hz, 2H), 7.63 (d, J = 8.0 Hz, 2H), 2.64 (s, 3H), 2.46 (s, 3H), 2.38 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 191.22, 191.02, 168.76, 158.75, 133.28, 133.14, 133.06, 133.02, 132.75, 127.34, 127.21, 124.62, 122.86, 119.78, 92.99, 90.84, 18.84, 18.64, 16.49; Anal. Calc. for: $C_{22}H_{17}NO_2S$ (359): C, 73.51; H, 4.77; N, 3.90%; Found: C, 73.53; H, 4.76; N, 3.89%.

4.1.4.15. 1-(4-Methyl-2-(4-thiophen-2-ylethynyl)phenyl)thiazol-5-yl)ethan-1-one (**5o**). Yellow oil (250 mg, 93%); 1H NMR (DMSO- d_6) δ : 7.94 (d, J = 8.4 Hz, 2H), 7.71 (d, J = 6.4 Hz, 1H), 7.63 (d, J = 8.4 Hz, 2H), 7.47 (d, J = 4.5 Hz, 1H), 7.16 (t, J = 5.2 Hz, 1H), 2.60 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 191.21, 160.20, 148.63, 142.99, 133.63, 133.49, 132.30, 129.84, 128.38, 126.33, 123.34, 122.07, 93.07, 85.04, 18.68, 16.48; Anal. Calc. for: $C_{18}H_{13}NOS_2$ (323): C, 66.85; H, 4.05; N, 4.33%; Found: C, 66.86; H, 4.04; N, 4.32%.

4.1.4.16. 1-(4-Methyl-2-(4-thiophen-3-ylethynyl)phenyl)thiazol-5-yl)ethan-1-one (**5p**). Yellow oil (240 mg, 89.5%); 1H NMR (DMSO- d_6) δ : 8.02 (d, J = 8.0 Hz, 2H), 7.95 (s, 1H), 6.84 (d, J = 7.2 Hz, 1H), 7.65 (d, J = 8.0 Hz, 2H), 7.30 (d, J = 6.0 Hz, 1H), 2.71 (s, 3H), 2.57 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 191.21, 160.19, 148.61, 136.53, 133.40, 132.26, 130.79, 130.09, 127.55, 126.33, 123.86, 121.43, 88.81, 87.29, 18.66, 16.49; Anal. Calc. for: $C_{18}H_{13}NOS_2$ (323): C, 66.85; H, 4.05; N, 4.33%; Found: C, 66.86; H, 4.06; N, 4.34%.

4.1.4.17. 1-(4-Methyl-2-(4-((5-methylthiophen-2-yl)ethynyl)phenyl)thiazol-5-yl)ethan-1-one (**5q**). Yellow oil (250 mg, 89.6%); 1H NMR (DMSO- d_6) δ : 7.95 (d, J = 8.0 Hz, 2H), 7.93 (d, J = 8.0 Hz, 1H), 7.65 (d, J = 8.4 Hz, 2H), 7.30 (d, J = 6.0 Hz, 1H), 2.68 (s, 3H), 2.51 (s, 3H), 2.39 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 191.11, 161.77, 148.77, 144.28, 136.29, 136.19, 133.45, 133.18, 132.60, 132.41, 126.18, 124.16, 98.75, 85.55, 32.00, 18.64, 16.35; Anal. Calc. for: $C_{19}H_{15}NOS_2$ (337): C, 67.63; H, 4.48; N, 4.15%; Found: C, 67.64; H, 4.47; N, 4.14%.

4.1.4.18. 1-(2-(4-((4,5-Dimethylthiophen-3-yl)ethynyl)phenyl)-4-methylthiazol-5-yl)ethan-1-one (**5r**). Yellow oil (230 mg, 79%); 1H NMR (DMSO- d_6) δ : 8.05 (d, J = 8.4 Hz, 2H), 7.67 (d, J = 8.4 Hz, 1H), 7.58 (s, 1H), 2.73 (s, 3H), 2.59 (s, 3H), 2.35 (s, 3H), 2.20 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 191.22, 167.99, 158.75, 134.99, 133.02, 132.17, 129.35, 128.77, 127.29, 126.87, 125.82, 122.84, 90.20, 88.39, 30.93, 18.64, 13.66, 12.95; Anal. Calc. for: $C_{20}H_{17}NO_2S$ (351): C, 68.34; H, 4.88; N, 3.99%; Found: C, 68.32; H, 4.87; N, 3.98%.

- 4.1.4.19.** 1-(2-(4-(Furan-3-ylethynyl)phenyl)-4-methylthiazol-5-yl)ethan-1-one (**5s**). Green oil (235 mg, 92.5%); ^1H NMR (DMSO- d_6) δ : 8.14 (s, 1H), 7.92 (d, J = 8.1 Hz, 2H), 7.77 (d, J = 4.2 Hz, 1H), 7.59 (d, J = 7.8 Hz, 2H), 6.71 (d, J = 4.2 Hz, 1H), 2.59 (s, 3H), 2.51 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 191.19, 160.27, 147.23, 144.63, 142.89, 136.63, 133.44, 132.29, 126.32, 123.81, 112.79, 107.06, 91.01, 83.36, 18.66, 16.45; Anal. Calc. for: $\text{C}_{18}\text{H}_{13}\text{NO}_2\text{S}$ (307): C, 70.34; H, 4.26; N, 4.56%; Found: C, 70.32; H, 4.24; N, 4.55%.
- 4.1.4.20.** 1-(4-Methyl-2-(4-(pyridin-2-ylethynyl)phenyl)thiazol-5-yl)ethan-1-one (**5t**). Yellow oil (230 mg, 87%); ^1H NMR (DMSO- d_6) δ : 8.36 (d, J = 6.4 Hz, 1H), 7.97 (d, J = 8.4 Hz, 2H), 7.88 (t, J = 6.8 Hz, 1H), 7.70 (d, J = 8.4 Hz, 2H), 7.68 (d, J = 7.6 Hz, 1H), 7.45 (t, J = 6.8 Hz, 1H), 2.60 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 191.21, 167.78, 158.76, 150.73, 142.37, 137.36, 133.21, 133.07, 132.94, 128.03, 127.31, 124.56, 124.36, 91.76, 88.06, 30.90, 18.62; Anal. Calc. for: $\text{C}_{19}\text{H}_{14}\text{N}_2\text{OS}$ (318): C, 71.68; H, 4.43; N, 8.80%; Found: C, 71.67; H, 4.44; N, 8.78%.
- 4.1.4.21.** 1-(4-Methyl-2-(4-(pyridin-4-ylethynyl)phenyl)thiazol-5-yl)ethan-1-one (**5u**). Yellow oil (240 mg, 91%); ^1H NMR (DMSO- d_6) δ : 8.65 (d, J = 6.4 Hz, 2H), 8.08 (d, J = 8.0 Hz, 2H), 7.97 (d, J = 8.4 Hz, 2H), 7.70 (d, J = 8.0 Hz, 2H), 2.72 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 191.21, 160.14, 150.46, 148.79, 143.06, 133.68, 132.97, 130.43, 126.37, 125.81, 122.58, 93.60, 88.89, 18.67, 16.52; Anal. Calc. for: $\text{C}_{19}\text{H}_{14}\text{N}_2\text{OS}$ (318): C, 71.68; H, 4.43; N, 8.80%; Found: C, 71.63; H, 4.45; N, 8.77%.
- 4.1.4.22.** 1-(4-Methyl-2-(4-(pyrimidin-2-ylethynyl)phenyl)thiazol-5-yl)ethan-1-one (**5v**). Yellow oil (230 mg, 87%); ^1H NMR (DMSO- d_6) δ : 8.92 (d, J = 8.0 Hz, 2H), 8.00 (d, J = 8.0 Hz, 2H), 7.70 (d, J = 8.0 Hz, 2H), 7.44 (t, J = 6.8 Hz, 1H), 2.61 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 191.21, 167.78, 158.76, 150.73, 142.37, 132.94, 128.03, 127.31, 124.56, 124.32, 120.32, 90.46, 86.34, 18.70, 16.58; Anal. Calc. for: $\text{C}_{18}\text{H}_{13}\text{N}_3\text{OS}$ (319): C, 67.69; H, 4.10; N, 13.16%; Found: C, 67.66; H, 4.09; N, 13.15%.
- 4.1.4.23.** 1-(4-Methyl-2-(4-(pyrimidin-5-ylethynyl)phenyl)thiazol-5-yl)ethan-1-one (**5w**). Yellow oil (240 mg, 90.9%); ^1H NMR (DMSO- d_6) δ : 9.2 (s, 1H), 9.03 (s, 2H), 7.98 (d, J = 8.4 Hz, 2H), 7.71 (d, J = 8.4 Hz, 2H), 2.60 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 191.21, 161.23, 159.50, 157.44, 148.63, 142.87, 134.30, 132.82, 126.40, 122.51, 119.21, 95.69, 85.12, 18.67, 16.45; Anal. Calc. for: $\text{C}_{18}\text{H}_{13}\text{N}_3\text{OS}$ (319): C, 67.69; H, 4.10; N, 13.16%; Found: C, 67.66; H, 4.09; N, 13.13%.
- 4.1.4.24.** 1-(2-(4-(3-Hydroxy-3-methylbut-1-yn-1-yl)phenyl)-4-methylthiazol-5-yl)ethan-1-one (**5x**). Yellow oil (400 mg, 91.9%); ^1H NMR (DMSO- d_6) δ : 7.89 (d, J = 8 Hz, 2H), 7.51 (d, J = 8.4 Hz, 2H), 5.44 (brs, 1H), 2.59 (s, 3H), 2.31 (s, 3H), 1.49 (s, 6H); ^{13}C NMR (DMSO- d_6) δ : 190.20, 160.20, 148.53, 142.99, 136.49, 132.56, 126.50, 123.56, 94.41, 83.50, 65.50, 29.08, 18.64, 16.17; Anal. Calc. for: $\text{C}_{17}\text{H}_{17}\text{NO}_2\text{S}$ (299): C, 68.20; H, 5.72; N, 4.68%; Found: C, 68.18; H, 5.75; N, 4.66%.
- 4.1.4.25.** Ethyl-3-(4-(5-acetyl-4-methylthiazol-2-yl)phenyl)propionate (**5y**). Brown oil (400 mg, 87.7%); ^1H NMR (DMSO- d_6) δ : 8.09 (d, J = 8.4 Hz, 2H), 7.81 (d, J = 8.4 Hz, 2H), 4.27 (q, J = 7.2 Hz, 2H), 2.72 (s, 3H), 2.58 (s, 3H), 1.27 (t, J = 6.9 Hz, 3H); ^{13}C NMR (DMSO- d_6) δ : 190.98, 165.12, 153.46, 149.11, 144.12, 137.19, 135.71, 134.15, 126.36, 119.58, 85.57, 82.38, 62.65, 18.68, 16.32, 14.31; Anal. Calc. for: $\text{C}_{17}\text{H}_{15}\text{NO}_3\text{S}$ (313): C, 65.16; H, 4.82; N, 4.47%; Found: C, 65.18; H, 4.80; N, 4.43%.
- 4.1.4.26.** 1-(2-(4-((1-Hydroxycyclohexyl)ethynyl)phenyl)-4-methylthiazol-5-yl)ethan-1-one (**5z**). Yellow oil (450 mg, 91%); ^1H NMR (DMSO- d_6) δ : 7.98 (d, J = 8.4 Hz, 2H), 7.53 (d, J = 8.1 Hz, 2H), 2.69 (s, 3H), 2.56 (s, 3H), 2.49 (brs, 1H), 1.86–1.48 (m, 10H); ^{13}C NMR (DMSO- d_6) δ : 190.40, 168.75, 159.56, 132.18, 131.46, 131.27, 127.33, 126.60, 97.73, 80.13, 38.65, 32.55, 29.79, 25.86, 24.88, 18.43; Anal. Calc. for: $\text{C}_{20}\text{H}_{21}\text{NO}_2\text{S}$ (339): C, 70.77; H, 6.24; N, 4.13%; Found: C, 70.80; H, 6.22; N, 4.10%.
- 4.1.4.27.** 4-((4-(5-Acetyl-4-methylthiazol-2-yl)phenyl)ethynyl)-2-methylbenzoic acid (**5aa**). Brown oil (250 mg, 80%); ^1H NMR (DMSO- d_6) δ : 12.36 (brs, 1H), 7.95 (d, J = 8.4 Hz, 2H), 7.71 (d, J = 8.0 Hz, 2H), 7.57 (d, J = 5.6 Hz, 1H), 7.51 (s, 1H), 7.29 (d, J = 7.2 Hz, 1H), 2.60 (s, 3H), 2.58 (s, 3H), 2.33 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 191.11, 161.11, 160.09, 152.23, 147.96, 143.19, 136.91, 134.57, 133.71, 132.92, 126.36, 121.67, 115.22, 94.78, 83.38, 30.51, 18.69, 16.55; Anal. Calc. for: $\text{C}_{22}\text{H}_{17}\text{NO}_3\text{S}$ (375): C, 70.38; H, 4.56; N, 3.73%; Found: C, 70.40; H, 4.55; N, 3.71%.
- 4.1.4.28.** Methyl-4-((4-(5-acetyl-4-methylthiazol-2-yl)phenyl)ethynyl)-3-methyl benzoate (**5bb**). Brown oil (255 mg, 79%); ^1H NMR (DMSO- d_6) δ : 7.97 (d, J = 7.6 Hz, 2H), 7.93 (s, 1H), 7.83 (d, J = 7.6 Hz, 1H), 7.70 (d, J = 7.2 Hz, 2H), 7.27 (d, J = 6.8 Hz, 1H), 3.87 (s, 3H), 2.61 (s, 3H), 2.55 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 191.11, 167.78, 160.31, 148.18, 142.78, 140.78, 132.71, 132.42, 130.56, 128.13, 127.15, 126.36, 124.56, 124.32, 119.37, 92.76, 88.15, 50.60, 20.69, 18.68, 16.45; Anal. Calc. for: $\text{C}_{23}\text{H}_{19}\text{NO}_3\text{S}$ (389): C, 70.90; H, 4.92; N, 3.60%; Found: C, 70.92; H, 4.90; N, 3.61%.
- 4.1.4.29.** 1-(2-(4-((5-Hydroxy-2-methylphenyl)ethynyl)phenyl)-4-methylthiazol-5-yl)ethan-1-one (**5cc**). Brown oil (200 mg, 69.4%); ^1H NMR (DMSO- d_6) δ : 7.97 (d, J = 8.0 Hz, 2H), 7.74 (d, J = 8.4 Hz, 2H), 7.29 (s, 1H), 7.27 (d, J = 7.2 Hz, 1H), 7.19 (d, J = 7.2 Hz, 1H), 6.27 (s, 1H), 2.62 (s, 3H), 2.51 (s, 3H), 2.34 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 191.21, 168.63, 161.21, 160.09, 152.33, 148.96, 143.09, 136.90, 134.58, 133.70, 132.91, 126.37, 121.68, 94.68, 83.28, 30.51, 18.69, 16.54; Anal. Calc. for: $\text{C}_{21}\text{H}_{17}\text{NO}_2\text{S}$ (347): C, 72.60; H, 4.93; N, 4.03%; Found: C, 72.63; H, 4.91; N, 4.00%.
- 4.1.4.30.** 1-(4-Methyl-2-(4-(p-tolylethynyl)phenyl)thiazol-5-yl)ethan-1-one (**5dd**). Brown oil (380 mg, 78.8%); ^1H NMR (DMSO- d_6) δ : 7.99 (d, J = 8.4 Hz, 2H), 7.90 (d, J = 8.4 Hz, 2H), 7.44 (d, J = 8 Hz, 2H), 7.39 (d, J = 6.8 Hz, 2H), 2.59 (s, 3H), 2.31 (s, 3H), 2.29 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 191.22, 167.76, 158.75, 133.28, 133.14, 133.06, 133.02, 132.75, 127.34, 127.21, 124.62, 118.86, 111.78, 92.99, 90.84, 30.94, 18.64, 16.49; Anal. Calc. for: $\text{C}_{21}\text{H}_{17}\text{NOS}$ (331): C, 76.10; H, 5.17; N, 4.23%; Found: C, 76.08; H, 5.14; N, 4.21%.
- 4.1.5.** 2-(1-(4-Methyl-2-(4-Substitutedphenylethynyl)phenyl)thiazol-5-yl)ethylidene hydrazine-1-carboximidamide (6–35)
- General procedure:** Thiazole derivatives **5a–dd** (0.63 mmol) were dissolved in absolute ethanol (10 mL) and aminoguanidine hydrochloride (70 mg, 0.63 mmol) and hydrochloric acid (0.1 mL) were added. The reaction mixture was heated at reflux for 4 h. The solvent was concentrated under reduced pressure then poured in crushed ice and neutralized with sodium carbonate to pH 7–8. The precipitated solid was collected by filtration and washed with copious amounts of water. Crystallization from ethylacetate afforded the desired products as solids. Yield, physical properties and spectral data of isolated purified products are listed below:
- 4.1.5.1.** 2-(1-(2-(4-((3-Aminophenyl)ethynyl)phenyl)-4-methylthiazol-5-yl)ethylidene)hydrazine-1-carboximidamide (**6**). Yellow solid (215 mg, 92%); m. p: 260–262; ^1H NMR (DMSO- d_6) δ : 7.92 (d, J = 8.4 Hz, 2H), 7.60 (d, J = 8.4 Hz, 2H), 7.09 (t, J = 7.6 Hz, 1H), 6.76 (s, 1H), 6.71 (d, J = 7.2 Hz, 1H), 6.64 (d, J = 6.8 Hz, 1H), 5.81 (brs, 2H), 5.71 (brs, 2H), 5.27 (brs, 2H), 2.60 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 161.55, 160.28, 149.33, 148.52, 142.92, 136.56, 133.32,

132.44, 129.73, 126.19, 124.10, 122.70, 119.37, 116.64, 115.31, 92.76, 88.15, 18.62, 16.47; Anal. Calc. for: $C_{21}H_{20}N_6S$ (388): C, 64.93; H, 5.19; N, 21.63%; Found: C, 64.92; H, 5.18; N, 21.64%.

4.1.5.2. 2-(1-(2-(4-((2-Aminophenyl)ethynyl)phenyl)-4-methylthiazol-5-yl) ethylidene) hydrazine-1-carboximidamide (**7**). Yellow solid (185 mg, 79%); m. p: 257–258; 1H NMR (DMSO- d_6) δ : 8.20 (d, J = 8.0 Hz, 2H), 7.99 (d, J = 8.4 Hz, 2H), 7.95 (d, J = 8.4 Hz, 1H), 7.84 (t, J = 7.2 Hz, 1H), 7.71 (d, J = 8.4 Hz, 1H), 7.57 (d, J = 8.0 Hz, 1H), 5.79 (brs, 4H), 5.27 (brs, 2H), 2.61 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 168.18, 161.18, 158.76, 150.73, 133.68, 132.66, 132.52, 132.50, 131.90, 130.75, 129.29, 129.17, 127.17, 122.17, 92.16, 88.26, 30.95, 18.62; Anal. Calc. for: $C_{21}H_{20}N_6S$ (388): C, 64.93; H, 5.19; N, 21.63%; Found: C, 64.93; H, 5.18; N, 21.63%.

4.1.5.3. 2-(1-(2-(4-((4-Aminophenyl)ethynyl)phenyl)-4-methylthiazol-5-yl) ethylidene)hydrazine-1-carboximidamide (**8**). Yellow solid (225 mg, 96.5%); m. p: 293–294; 1H NMR (DMSO- d_6) δ : 7.85 (d, J = 8.4 Hz, 2H), 7.53 (d, J = 8.4 Hz, 2H), 7.27 (d, J = 8.0 Hz, 2H), 6.59 (d, J = 8.4 Hz, 2H), 6.63 (brs, 4H), 5.58 (brs, 2H), 2.60 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 162.88, 160.00, 158.73, 150.42, 143.12, 133.32, 132.79, 132.01, 131.33, 127.20, 126.91, 114.11, 107.99, 95.14, 86.85, 30.91, 18.65; Anal. Calc. for: $C_{21}H_{20}N_6S$ (388): C, 64.93; H, 5.19; N, 21.63%; Found: C, 64.92; H, 5.18; N, 21.64%.

4.1.5.4. 2-(1-(2-(4-((3-Methoxyphenyl)ethynyl)phenyl)-4-methylthiazol-5-yl) ethylidene) Hydrazine-1-carboximidamide (**9**). Yellow solid (205 mg, 88.6%); m. p: 295–297; 1H NMR (DMSO- d_6) δ : 7.95 (d, J = 8.4 Hz, 2H), 7.65 (d, J = 8.0 Hz, 2H), 7.37 (t, J = 8.0 Hz, 1H), 7.17 (d, J = 8.0 Hz, 1H), 7.13 (s, 1H), 7.02 (d, J = 6.4 Hz, 1H), 5.83 (brs, 4H), 3.80 (s, 3H), 2.60 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 161.61, 160.05, 159.69, 148.77, 143.17, 136.42, 133.55, 132.58, 130.42, 126.33, 124.67, 124.30, 123.72, 116.58, 116.03, 91.69, 89.31, 55.73, 18.67, 16.54; Anal. Calc. for: $C_{22}H_{21}N_5OS$ (347): C, 65.49; H, 5.25; N, 17.36%; Found: C, 65.48; H, 5.26; N, 17.35%.

4.1.5.5. 2-(1-(2-(4-((2-Methoxyphenyl)ethynyl)phenyl)-4-methylthiazol-5-yl) ethylidene) Hydrazine-1-carboximidamide (**10**). Yellow solid (220 mg, 94.8%); m. p: 288–289; 1H NMR (DMSO- d_6) δ : 8.05 (d, J = 8.4 Hz, 2H), 7.66 (d, J = 8.1 Hz, 2H), 7.52 (d, J = 7.5 Hz, 1H), 7.43 (t, J = 7.6 Hz, 1H), 7.12 (d, J = 8.4 Hz, 1H), 6.99 (d, J = 8.4 Hz, 1H), 5.85 (brs, 4H), 3.87 (s, 3H), 2.72 (s, 3H), 2.48 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 161.76, 160.18, 159.93, 148.12, 143.56, 136.21, 133.61, 133.28, 132.41, 131.14, 126.34, 124.31, 121.02, 111.89, 111.49, 93.08, 88.68, 56.02, 18.67, 16.58; Anal. Calc. for: $C_{22}H_{21}N_5OS$ (347): C, 65.49; H, 5.25; N, 17.36%; Found: C, 65.48; H, 5.26; N, 17.35%.

4.1.5.6. 2-(1-(2-(4-((3-Hydroxyphenyl)ethynyl)phenyl)-4-methylthiazol-5-yl) ethylidene) hydrazine-1-carboximidamide (**11**). Orange solid (210 mg, 90%); m. p: 295–296; 1H NMR (DMSO- d_6) δ : 9.75 (brs, 2H), 7.94 (d, J = 8.4 Hz, 2H), 7.64 (d, J = 8.4 Hz, 2H), 7.26 (t, J = 8.0 Hz, 1H), 7.24 (brs, 2H), 7.01 (d, J = 7.6 Hz, 1H), 6.94 (s, 1H), 6.86 (d, J = 8.0 Hz, 1H), 6.00 (brs, 1H), 2.60 (s, 3H), 2.33 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 161.83, 159.76, 157.87, 149.09, 143.51, 136.05, 133.41, 132.57, 130.41, 126.33, 123.91, 123.37, 122.77, 118.24, 117.02, 91.91, 88.88, 18.67, 16.69; Anal. Calc. for: $C_{21}H_{19}N_5OS$ (389): C, 64.76; H, 4.92; N, 17.98%; Found: C, 64.75; H, 4.93; N, 17.99%.

4.1.5.7. 2-(1-(2-(4-((2-hydroxyphenyl)ethynyl)phenyl)-4-methylthiazol-5-yl) ethylidene) hydrazine-1-carboximidamide (**12**). Yellow solid (205 mg, 87.9%); m. p: 266–267; 1H NMR (DMSO- d_6) δ : 7.95 (d, J = 8.4 Hz, 2H), 7.71 (d, J = 8.4 Hz, 2H), 7.66 (d, J = 8.4 Hz, 1H), 7.57 (t, J = 6.4 Hz, 1H), 7.40 (brs, 4H), 7.29 (t, J = 7.2 Hz, 1H), 7.19 (t, J = 8.0 Hz, 1H), 5.70 (brs, 1H), 2.60 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 167.63, 161.11, 160.09, 152.23, 147.96, 143.19, 136.91,

134.57, 133.99, 133.71, 132.92, 126.36, 124.35, 121.67, 115.22, 94.78, 83.38, 30.51, 18.69; Anal. Calc. for: $C_{21}H_{19}N_5OS$ (389): C, 64.76; H, 4.92; N, 17.98%; Found: C, 64.75; H, 4.93; N, 17.99%.

4.1.5.8. 2-(1-(4-Methyl-2-(4-((3-nitrophenyl)ethynyl)phenyl)thiazol-5-yl) ethylidene)hydrazine-1-carboximidamide (**13**). Red solid (200 mg, 86.9%); m. p: 296–297; 1H NMR (DMSO- d_6) δ : 8.39 (s, 1H), 8.28 (d, J = 7.2 Hz, 1H), 8.04 (d, J = 8.0 Hz, 2H), 7.98 (d, J = 8.4 Hz, 1H), 7.77 (d, J = 8.0 Hz, 2H), 7.75 (t, J = 8.4 Hz, 1H), 7.56 (brs, 2H), 5.87 (brs, 2H), 2.61 (s, 3H), 2.33 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 161.53, 159.99, 153.52, 148.42, 143.62, 142.92, 138.02, 134.06, 132.89, 130.98, 126.38, 124.13, 122.94, 91.57, 84.57, 18.62, 16.57; Anal. Calc. for: $C_{21}H_{18}N_6O_2S$ (418): C, 60.27; H, 4.34; N, 20.08%; Found: C, 60.28; H, 4.33; N, 20.07%.

4.1.5.9. 2-(1-(4-Methyl-2-(4-((2-nitrophenyl)ethynyl)phenyl)thiazol-5-yl) ethylidene) hydrazine-1-carboximidamide (**14**). Red solid (190 mg, 82%); m. p: 298–299; 1H NMR (DMSO- d_6) δ : 8.20 (d, J = 8.0 Hz, 1H), 7.99 (d, J = 8.4 Hz, 2H), 7.91 (d, J = 7.2 Hz, 1H), 7.84 (t, J = 7.2 Hz, 1H), 7.71 (d, J = 8.4 Hz, 2H), 7.68 (t, J = 8.0 Hz, 1H), 5.79 (brs, 2H), 5.72 (brs, 2H), 2.61 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 161.28, 160.25, 149.65, 148.69, 142.93, 136.89, 135.02, 134.27, 133.69, 132.89, 130.51, 126.43, 125.36, 122.87, 117.22, 96.16, 86.92, 18.68, 16.47; Anal. Calc. for: $C_{21}H_{18}N_6O_2S$ (418): C, 60.27; H, 4.34; N, 20.08%; Found: C, 60.26; H, 4.33; N, 20.07%.

4.1.5.10. Methyl-3-((4-(5-(1-(2-carbamimidoylhydrazono)ethyl)-4-methylthiazol-2-yl)phenyl)ethynyl)benzoate (**15**). Yellow solid (210 mg, 91.7%); m. p: 260–261; 1H NMR (DMSO- d_6) δ : 8.11 (s, 1H), 8.01 (d, J = 7.2 Hz, 1H), 7.97 (d, J = 8.0 Hz, 2H), 7.87 (d, J = 7.6 Hz, 1H), 7.71 (d, J = 8.0 Hz, 2H), 7.64 (t, J = 7.6 Hz, 1H), 6.12 (brs, 4H), 3.89 (s, 3H), 2.61 (s, 3H), 2.34 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 169.97, 167.77, 161.23, 158.74, 150.33, 142.37, 137.36, 133.21, 133.17, 132.91, 131.08, 128.13, 127.31, 124.56, 124.32, 119.37, 92.76, 88.07, 47.54, 30.90, 18.62; Anal. Calc. for: $C_{23}H_{21}N_5O_2S$ (431): C, 64.02; H, 4.91; N, 16.23%; Found: C, 64.03; H, 4.90; N, 16.22%.

4.1.5.11. Methyl-2-((4-(5-(1-(2-carbamimidoylhydrazono)ethyl)-4-methylthiazol-2-yl)phenyl)ethynyl)benzoate (**16**). Yellow powder (205 mg, 89.5%); m. p: 270–271; 1H NMR (DMSO- d_6) δ : 8.21 (d, J = 8.0 Hz, 2H), 8.03 (d, J = 8.0 Hz, 1H), 7.97 (d, J = 6.8 Hz, 1H), 7.91 (d, J = 8.0 Hz, 2H), 7.74 (t, J = 8.0 Hz, 1H), 7.65 (t, J = 6.8 Hz, 1H), 5.88 (brs, 4H), 4.03 (s, 3H), 2.62 (s, 3H), 2.33 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 167.78, 160.18, 149.33, 148.52, 142.92, 136.56, 133.67, 133.32, 132.44, 129.73, 128.63, 126.19, 124.67, 122.70, 119.37, 116.64, 92.76, 88.15, 55.60, 18.67, 16.47; Anal. Calc. for: $C_{23}H_{21}N_5O_2S$ (431): C, 64.02; H, 4.91; N, 16.23%; Found: C, 64.03; H, 4.90; N, 16.22%.

4.1.5.12. 4-((4-(5-(1-(2-carbamimidoylhydrazono)ethyl)-4-methylthiazol-2-yl) phenyl)ethynyl)benzamide (**17**). Yellow solid (210 mg, 90.9%); m. p: 253–255; 1H NMR (DMSO- d_6) δ : 8.01 (d, J = 8.4 Hz, 2H), 7.96 (d, J = 8.0 Hz, 2H), 7.72 (d, J = 8.4 Hz, 2H), 7.68 (d, J = 8.0 Hz, 2H), 5.83 (brs, 4H), 4.36 (brs, 2H), 2.60 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 167.85, 167.50, 158.74, 148.78, 143.32, 134.88, 133.71, 133.16, 132.70, 131.84, 129.28, 128.35, 127.31, 125.18, 91.86, 90.99, 30.93, 18.65; Anal. Calc. for: $C_{22}H_{20}N_6OS$ (416): C, 63.44; H, 4.84; N, 20.18%; Found: C, 63.43; H, 4.83; N, 20.17%.

4.1.5.13. 4-((4-(5-(1-(2-carbamimidoylhydrazono)ethyl)-4-methylthiazol-2-yl) phenyl)ethynyl)benzothioamide (**18**). Orange solid (200 mg, 87%); m. p: 260–262; 1H NMR (DMSO- d_6) δ : 7.97 (d, J = 8.0 Hz, 2H), 7.93 (d, J = 8.0 Hz, 2H), 7.91 (brs, 2H), 7.80 (d, J = 8.4 Hz, 2H), 7.63 (d, J = 8.4 Hz, 2H), 5.78 (brs, 2H), 5.70 (brs, 2H), 2.60 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 167.46, 161.91, 158.75, 143.35, 133.28, 133.14, 133.06, 133.02, 132.75, 127.34, 127.21,

124.62, 118.86, 111.78, 92.99, 90.84, 18.64, 16.46; Anal. Calc. for: $C_{22}H_{20}N_6S_2$ (432): C, 61.09; H, 4.66; N, 19.43%; Found: C, 61.08; H, 4.65; N, 19.44%.

4.1.5.14. 2-(1-(4-((4-(5-(1-(2-Carbamimidoylhydrazineylidene)ethyl)-4-methylthiazol-2-yl)phenyl)ethynyl)phenyl)ethylidene)hydrazine-1-carboximidamide (**19**). Yellow solid (240 mg, 91.6%); m. p: 302–303; 1H NMR (DMSO- d_6) δ : 11.55 (brs, 1H), 11.42 (brs, 1H), 8.07 (d, J = 8.0 Hz, 2H), 7.98 (d, J = 8.0 Hz, 2H), 7.92 (brs, 3H), 7.81 (brs, 3H), 7.70 (d, J = 8.0 Hz, 2H), 7.63 (d, J = 8.0 Hz, 2H), 2.64 (s, 3H), 2.46 (s, 3H), 2.38 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 163.96, 156.44, 156.25, 152.84, 150.91, 147.27, 137.31, 132.75, 132.39, 131.45, 127.16, 126.35, 124.16, 123.19, 120.55, 91.55, 90.63, 18.39, 18.24, 14.70; Anal. Calc. for: $C_{24}H_{25}N_9S$ (471): C, 61.13; H, 5.34; N, 26.73%; Found: C, 61.12; H, 5.33; N, 26.74%.

4.1.5.15. 2-(1-(4-Methyl-2-(4-thiophen-2-ylethynyl)phenyl)thiazol-5-yl)ethylidene)hydrazine-1-carboximidamide (**20**). Yellow solid (227 mg, 97%); m. p: 244–246; 1H NMR (DMSO- d_6) δ : 7.94 (d, J = 8.4 Hz, 2H), 7.71 (d, J = 6.4 Hz, 1H), 7.63 (d, J = 8.4 Hz, 2H), 7.47 (d, J = 4.5 Hz, 1H), 7.16 (t, J = 5.2 Hz, 1H), 5.76 (brs, 4H), 2.60 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 161.45, 160.20, 148.63, 142.99, 136.64, 133.63, 133.49, 132.30, 129.84, 128.38, 126.33, 123.34, 122.07, 93.07, 85.04, 18.68, 16.48; Anal. Calc. for: $C_{19}H_{17}N_5S_2$ (379): C, 60.13; H, 4.52; N, 18.45%; Found: C, 60.13; H, 4.52; N, 18.44%.

4.1.5.16. 2-(1-(4-Methyl-2-(4-thiophen-3-ylethynyl)phenyl)thiazol-5-yl)ethylidene)hydrazine-1-carboximidamide (**21**). Yellow solid (240 mg, 89.5%); m. p: 233–234; 1H NMR (DMSO- d_6) δ : 8.02 (d, J = 8.0 Hz, 2H), 7.95 (s, 1H), 6.84 (d, J = 7.2 Hz, 1H), 7.65 (d, J = 8.0 Hz, 2H), 7.30 (d, J = 6.0 Hz, 1H), 5.89 (brs, 4H), 2.71 (s, 3H), 2.57 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 161.75, 159.89, 148.93, 143.35, 136.20, 133.37, 132.26, 130.79, 130.09, 127.55, 126.33, 123.86, 121.43, 88.81, 87.29, 18.66, 16.49; Anal. Calc. for: $C_{19}H_{17}N_5S_2$ (379): C, 60.13; H, 4.52; N, 18.45%; Found: C, 60.14; H, 4.51; N, 18.45%.

4.1.5.17. 2-(1-(4-Methyl-2-(4-((5-methylthiophen-2-yl)ethynyl)phenyl)thiazol-5-yl)ethylidene)hydrazine-1-carboximidamide (**22**). Yellow solid (205 mg, 87.9%); m. p: 234–235; 1H NMR (DMSO- d_6) δ : 7.95 (d, J = 8.0 Hz, 2H), 7.93 (d, J = 8.0 Hz, 1H), 7.65 (d, J = 8.4 Hz, 2H), 7.30 (d, J = 6.0 Hz, 1H), 5.87 (brs, 4H), 2.68 (s, 3H), 2.51 (s, 3H), 2.39 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 167.87, 161.77, 158.47, 148.77, 144.28, 136.29, 136.19, 133.45, 133.18, 132.60, 132.41, 126.18, 124.16, 98.75, 85.55, 32.00, 18.64, 16.35; Anal. Calc. for: $C_{20}H_{19}N_5S_2$ (393): C, 61.04; H, 4.87; N, 17.80%; Found: C, 61.04; H, 4.86; N, 17.81%.

4.1.5.18. 2-(1-(2-(4-((4,5-Dimethylthiophen-3-yl)ethynyl)phenyl)-4-methylthiazol-5-yl)ethylidene)hydrazine-1-carboximidamide (**23**). Red powder (195 mg, 84%); m. p: 227–228; 1H NMR (DMSO- d_6) δ : 7.95 (d, J = 8.4 Hz, 2H), 7.59 (d, J = 8.4 Hz, 2H), 7.37 (s, 1H), 6.05 (brs, 4H), 2.60 (s, 3H), 2.59 (s, 3H), 2.35 (s, 3H), 2.20 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 167.99, 161.74, 158.75, 134.99, 134.26, 133.02, 132.17, 129.35, 128.77, 127.29, 126.87, 125.82, 122.84, 90.20, 88.39, 30.93, 18.64, 13.66, 12.95; Anal. Calc. for: $C_{21}H_{21}N_5S_2$ (407): C, 61.89; H, 5.19; N, 17.18%; Found: C, 61.90; H, 5.18; N, 17.19%.

4.1.5.19. 2-(1-(2-(4-(Furan-3-ylethynyl)phenyl)-4-methylthiazol-5-yl)ethylidene)hydrazine-1-carboximidamide (**24**). Yellow solid (212 mg, 89.8%); m. p: 222–223; 1H NMR (DMSO- d_6) δ : 8.14 (s, 1H), 7.92 (d, J = 8.1 Hz, 2H), 7.77 (d, J = 4.2 Hz, 1H), 7.59 (d, J = 7.8 Hz, 2H), 6.71 (d, J = 4.2 Hz, 1H), 5.64 (brs, 4H), 2.59 (s, 3H), 2.51 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 162.55, 160.27, 158.56, 147.23, 144.63, 142.89, 136.63, 133.44, 132.29, 126.32, 123.81, 112.79, 107.06, 91.01, 83.36, 18.66, 16.45; Anal. Calc. for: $C_{19}H_{17}N_5OS$ (363): C, 62.79; H, 4.71; N, 19.27%; Found: C, 62.78; H, 4.70; N, 19.28%.

4.1.5.20. 2-(1-(4-Methyl-2-(4-(pyridin-2-ylethynyl)phenyl)thiazol-5-yl)ethylidene)hydrazine-1-carboximidamide (**25**). Orange solid (230 mg, 97.8%); m. p: 268–269; 1H NMR (DMSO- d_6) δ : 8.63 (d, J = 6.4 Hz, 1H), 7.97 (d, J = 8.4 Hz, 2H), 7.88 (t, J = 6.8 Hz, 1H), 7.70 (d, J = 8.4 Hz, 2H), 7.68 (d, J = 7.6 Hz, 1H), 7.45 (t, J = 6.8 Hz, 1H), 5.79 (brs, 2H), 5.69 (brs, 2H), 2.60 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 167.78, 160.44, 158.76, 150.73, 148.76, 142.37, 137.36, 133.21, 133.07, 132.94, 128.03, 127.31, 124.56, 124.36, 91.76, 88.06, 18.70, 16.58; Anal. Calc. for: $C_{20}H_{18}N_6S$ (374): C, 64.15; H, 4.85; N, 22.44%; Found: C, 64.15; H, 4.84; N, 22.43%.

4.1.5.21. 2-(1-(4-Methyl-2-(4-(pyridin-4-ylethynyl)phenyl)thiazol-5-yl)ethylidene)hydrazine-1-carboximidamide (**26**). Green solid (225 mg, 95.7%); m. p: 270–271; 1H NMR (DMSO- d_6) δ : 8.65 (d, J = 6.4 Hz, 2H), 8.08 (d, J = 8.0 Hz, 2H), 7.97 (d, J = 8.4 Hz, 2H), 7.70 (d, J = 8.0 Hz, 2H), 5.81 (brs, 4H), 2.72 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 161.34, 160.14, 150.46, 148.79, 143.06, 133.68, 132.97, 130.43, 129.17, 127.32, 125.81, 122.58, 93.60, 88.89, 18.67, 16.52; Anal. Calc. for: $C_{20}H_{18}N_6S$ (374): C, 64.15; H, 4.85; N, 22.44%; Found: C, 64.15; H, 4.84; N, 22.43%.

4.1.5.22. 2-(1-(4-Methyl-2-(4-(pyrimidin-2-ylethynyl)phenyl)thiazol-5-yl)ethylidene)hydrazine-1-carboximidamide (**27**). Yellow solid (222 mg, 94%); m. p: 284–285; 1H NMR (DMSO- d_6) δ : 8.92 (d, J = 8.0 Hz, 2H), 8.00 (d, J = 8.0 Hz, 2H), 7.70 (d, J = 8.0 Hz, 2H), 7.44 (t, J = 6.8 Hz, 1H), 5.81 (brs, 2H), 5.76 (brs, 2H), 2.61 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 168.75, 163.09, 158.76, 152.45, 150.73, 142.37, 132.94, 128.03, 127.31, 124.56, 124.32, 120.32, 90.46, 86.34, 18.70, 16.58; Anal. Calc. for: $C_{19}H_{17}N_7S$ (375): C, 60.78; H, 4.56; N, 26.11%; Found: C, 60.77; H, 4.55; N, 26.12%.

4.1.5.23. 2-(1-(4-Methyl-2-(4-(pyrimidin-5-ylethynyl)phenyl)thiazol-5-yl)ethylidene)hydrazine-1-carboximidamide (**28**). Yellow solid (220 mg, 93.6%); m. p: 253–254; 1H NMR (DMSO- d_6) δ : 8.92 (d, J = 8.0 Hz, 2H), 8.00 (d, J = 8.0 Hz, 2H), 7.70 (d, J = 8.0 Hz, 2H), 7.44 (t, J = 6.8 Hz, 1H), 5.76 (brs, 4H), 2.71 (s, 3H), 2.59 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 167.75, 161.23, 158.76, 152.45, 150.73, 142.37, 132.94, 128.03, 127.31, 124.56, 124.32, 120.32, 90.46, 86.34, 18.67, 16.45; Anal. Calc. for: $C_{19}H_{17}N_7S$ (375): C, 60.78; H, 4.56; N, 26.11%; Found: C, 60.78; H, 4.55; N, 26.10%.

4.1.5.24. 2-(1-(2-(4-(3-Hydroxy-3-methylbut-1-yn-1-yl)phenyl)-4-methylthiazol-5-yl)ethylidene)hydrazine-1-carboximidamide (**29**). Yellow solid (290 mg, 93.5%); m. p: 224–226; 1H NMR (DMSO- d_6) δ : 7.89 (d, J = 8.0 Hz, 2H), 7.51 (d, J = 8.4 Hz, 2H), 5.78 (brs, 2H), 5.73 (brs, 2H), 5.44 (brs, 1H), 2.65 (s, 3H), 2.31 (s, 3H), 1.49 (s, 6H); ^{13}C NMR (DMSO- d_6) δ : 161.61, 160.20, 148.53, 142.99, 136.49, 133.46, 132.56, 126.50, 123.56, 94.41, 83.50, 65.50, 29.08, 18.64, 16.17; Anal. Calc. for: $C_{18}H_{21}N_5OS$ (355): C, 60.82; H, 5.96; N, 19.70%; Found: C, 60.81; H, 5.95; N, 19.70%.

4.1.5.25. Ethyl-3-(4-(5-(1-(2-carbamimidoylhydrazono)ethyl)-4-methylthiazol-2-yl)phenyl)propionate (**30**). Brown solid (200 mg, 85%); m. p: 177–178; 1H NMR (DMSO- d_6) δ : 8.09 (d, J = 8.4 Hz, 2H), 7.81 (d, J = 8.4 Hz, 2H), 5.79 (brs, 4H), 4.27 (q, J = 7.2 Hz, 2H), 2.72 (s, 3H), 2.58 (s, 3H), 1.27 (t, J = 6.9 Hz, 3H); ^{13}C NMR (DMSO- d_6) δ : 164.10, 156.42, 153.13, 147.47, 133.36, 132.86, 132.00, 129.35, 129.23, 126.73, 123.32, 92.14, 83.66, 48.08, 27.88, 18.64, 18.55; Anal. Calc. for: $C_{18}H_{19}N_5O_2S$ (369): C, 58.52; H, 5.18; N, 18.96%; Found: C, 58.51; H, 5.17; N, 18.95%.

4.1.5.26. 2-(1-(2-(4-((1-Hydroxycyclohexyl)ethynyl)phenyl)-4-methylthiazol-5-yl)ethylidene)hydrazine-1-carboximidamide (**31**). Yellow solid (215 mg, 92%); m. p: 279–280; 1H NMR (DMSO- d_6) δ : 7.89 (brs, 2H), 7.87 (brs, 2H), 7.52 (d, J = 8.4 Hz, 2H), 7.49 (d,

$J = 8.0$ Hz, 2H), 5.76 (brs, 1H), 2.64 (s, 3H), 2.31 (s, 3H), 1.86–1.48 (m, 10H); ^{13}C NMR (DMSO- d_6) δ : 168.75, 160.54, 159.56, 148.24, 132.18, 131.46, 131.27, 127.33, 126.60, 97.73, 80.13, 38.65, 32.55, 29.79, 25.86, 24.88, 18.43; Anal. Calc. for: $\text{C}_{21}\text{H}_{25}\text{N}_5\text{O}_5$ (395): C, 63.77; H, 6.37; N, 17.71%; Found: C, 63.78; H, 6.36; N, 17.70%.

4.1.5.27. 4-((4-(5-(1-(2-carbamimidoylhydrazono)ethyl)-4-methylthiazol-2-yl) phenyl)ethynyl)-2-methylbenzoic acid (**32**). Brown solid (195 mg, 85%); m. p: 297–298; ^1H NMR (DMSO- d_6) δ : 12.36 (brs, 1H), 7.95 (d, $J = 8.4$ Hz, 2H), 7.71 (d, $J = 8.0$ Hz, 2H), 7.57 (d, $J = 5.6$ Hz, 1H), 7.51 (s, 1H), 7.29 (d, $J = 7.2$ Hz, 1H), 5.79 (brs, 2H), 5.69 (brs, 2H), 2.60 (s, 3H), 2.58 (s, 3H), 2.33 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 167.63, 161.11, 160.09, 152.23, 147.96, 143.19, 136.91, 134.57, 133.71, 133.21, 132.92, 131.76, 129.02, 126.36, 121.67, 115.22, 94.78, 83.38, 30.51, 18.69, 16.57; Anal. Calc. for: $\text{C}_{23}\text{H}_{21}\text{N}_5\text{O}_2\text{S}$ (431): C, 64.02; H, 4.91; N, 16.93%; Found: C, 64.01; H, 4.90; N, 16.91%.

4.1.5.28. Methyl-4-((4-(5-(1-(2-carbamimidoylhydrazono)ethyl)-4-methylthiazol-2-yl)phenyl) ethynyl)-3-methylbenzoate (**33**). Yellow powder (255 mg, 79%); m. p: 228–229; ^1H NMR (DMSO- d_6) δ : 7.97 (d, $J = 7.6$ Hz, 2H), 7.93 (s, 1H), 7.83 (d, $J = 7.6$ Hz, 1H), 7.70 (d, $J = 7.2$ Hz, 2H), 7.27 (d, $J = 6.8$ Hz, 1H), 5.79 (brs, 2H), 5.69 (brs, 2H), 3.87 (s, 3H), 2.61 (s, 3H), 2.55 (s, 3H), 2.32 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 169.78, 167.78, 160.31, 158.18, 148.18, 142.78, 140.78, 132.71, 132.42, 130.56, 127.15, 126.36, 124.56, 124.32, 119.37, 92.76, 88.15, 50.60, 20.69, 18.68, 16.45; Anal. Calc. for: $\text{C}_{24}\text{H}_{23}\text{N}_5\text{O}_2\text{S}$ (445): C, 64.70; H, 5.20; N, 15.72%; Found: C, 64.69; H, 5.21; N, 15.71%.

4.1.5.29. 2-(1-(2-(4-((5-Hydroxy-2-methylphenyl)ethynyl)phenyl)-4-methylthiazol-5-yl) ethylidene)hydrazine-1-carboximidamide (**34**). Yellow solid (200 mg, 86%); m. p: 260–261; ^1H NMR (DMSO- d_6) δ : 8.65 (brs, 4H), 7.95 (d, $J = 8.0$ Hz, 2H), 7.71 (d, $J = 8.4$ Hz, 2H), 7.69 (s, 1H), 7.68 (d, $J = 8.0$ Hz, 1H), 7.59 (d, $J = 8.0$ Hz, 1H), 5.83 (s, 1H), 2.62 (s, 3H), 2.51 (s, 3H), 2.34 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 168.63, 161.21, 160.09, 152.33, 148.96, 143.19, 143.09, 136.90, 134.58, 133.70, 132.91, 127.05, 126.37, 121.68, 119.56, 94.68, 83.28, 30.51, 18.69, 16.54; Anal. Calc. for: $\text{C}_{22}\text{H}_{21}\text{N}_5\text{O}_5$ (403): C, 65.49; H, 5.25; N, 17.36%; Found: C, 65.48; H, 5.24; N, 17.37%.

4.1.5.30. 2-(1-(4-Methyl-2-(4-(*p*-tolylethynyl)phenyl)thiazol-5-yl) ethylidene) hydrazine-1-carboximidamide (**35**). Brown solid (160 mg, 68.6%); m. p: 266–268; ^1H NMR (DMSO- d_6) δ : 7.99 (d, $J = 8.4$ Hz, 2H), 7.90 (d, $J = 8.4$ Hz, 2H), 7.44 (d, $J = 8.0$ Hz, 2H), 7.39 (d, $J = 6.8$ Hz, 2H), 5.75 (brs, 2H), 5.65 (brs, 2H), 2.59 (s, 3H), 2.31 (s, 3H), 2.29 (s, 3H); ^{13}C NMR (DMSO- d_6) δ : 167.76, 161.29, 158.75, 133.28, 133.14, 133.06, 133.02, 132.75, 127.34, 127.21, 124.62, 118.86, 111.78, 92.99, 90.84, 30.94, 18.64, 16.49; Anal. Calc. for: $\text{C}_{22}\text{H}_{21}\text{N}_5\text{S}$ (387): C, 68.19; H, 5.46; N, 18.07%; Found: C, 68.18; H, 5.46; N, 18.06%.

4.2. Biological methods

4.2.1. Determination of minimum inhibitory concentration (MIC) and cinimum cactericidal concentration (MBC) values

S. aureus clinical isolates (NRS107, NRS119, NRS123, NRS384, VRS10, VRS11a, and VRS12) were obtained through the Network of Antimicrobial Resistance in *Staphylococcus aureus* (NARSA) program and BEI Resources. The MICs of the newly synthesized compounds, tested against clinical isolates of a panel of different Gram-positive and Gram-negative bacteria were determined using the broth microdilution assay in accordance with the Clinical and Laboratory Standards Institute guidelines. The MBC was determined by transferring a small aliquot (5 μL), from wells lacking bacterial growth (in the MIC plates), onto Tryptic soy agar (TSA) plates. Plates were incubated at 37 °C for at least 18 h prior to

determining the MBC. The MBC was categorized as the lowest concentration where 99.9% of bacterial growth was inhibited [37].

4.2.2. Time-kill assay against MRSA

MRSA USA400 cells in logarithmic growth phase ($\text{OD}_{600} > 0.800$) were diluted to $\sim 10^6$ CFU/mL and exposed to concentrations equivalent to $4 \times \text{MIC}$ (in triplicate) of tested compounds or vancomycin in Tryptic soy broth. After the corresponding time points, cells were collected, diluted and plated on TSA plates. Plates were incubated for 18–20 h and the viable CFU/mL was determined [38,39].

4.2.3. Resistance study against MRSA

To determine if MRSA would be capable of forming resistance to the compounds quickly, a multi-step resistance selection experiment was conducted, as described previously [40–42]. The broth microdilution assay was utilized to determine the MIC of compounds **8**, **22**, **31**, and rifampicin exposed to MRSA USA400 (NRS123) for 14 passages. A greater than four-fold increase in the initial MIC indicates development of resistance.

4.2.4. Intracellular infection of J774 cells with MRSA and treatment with alkynylphenylthiazoles

Toxicity assessment: Compounds were assayed against a murine macrophage (J774) cell line to determine their *in vitro* potential toxic effect, as described elsewhere [43]. Compounds were incubated with J774 cells at concentrations ranging from 16 to 64 $\mu\text{g/mL}$ for 24 h at 37 °C with 5% CO_2 . The toxicity data were analyzed using a two-way ANOVA with post-hoc Dunnet's test for multiple comparisons, utilizing GraphPad Prism 6.0 (GraphPad Software, La Jolla, CA).

Eradication of intracellular MRSA: Utilizing previously described methods [44,45], the ability of the compounds to reduce the burden of intracellular MRSA was evaluated. Murine macrophage cells (J774) were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% FBS at 37 °C with CO_2 (5%). J774 cells were exposed to MRSA USA400 cells at a multiplicity of infection of approximately 10:1. After 1 h of infection, J774 cells were washed with gentamicin (100 $\mu\text{g/mL}$) to kill extracellular MRSA. The compounds or vancomycin (all at 16 $\mu\text{g/mL}$) were subsequently added to each well (four replicates per test agent). After 24 h incubation at 37 °C with 5% CO_2 , the test agents were removed. J774 cells were washed and subsequently lysed using 0.1% Triton-X. The solution was serially diluted in phosphate-buffered saline and transferred to TSA plates in order to determine viable MRSA CFU inside the J774 cells. Plates were incubated at 37 °C for 18–22 h before counting viable CFU/mL. Data are presented as \log_{10} (MRSA CFU/mL) in infected J774 cells.

4.3. PBS solubility screen

The solubility screen was conducted as described previously [46]. Serial dilutions of the tested compounds were prepared in phosphate buffered saline (PBS) at $100 \times$ the final concentration. The solutions were diluted 100-fold into PBS in a 96-well plate and mixed. The absorbance of the PBS-containing plate was measured prior to addition of the test agents to determine the background absorbance. After 2 h, the presence of precipitate was detected by turbidity (absorbance at 540 nm). An absorbance value of greater than (mean + $3 \times$ standard deviation of the blank), after subtracting the pre-experiment background, is indicative of turbidity. The solubility limit is reported as the highest experimental concentration with no evidence of turbidity.

4.4. Protein purification

E. coli UppP in pET28a with a 6X His tag was expressed and purified using affinity chromatography. Briefly, transformed C41 (DE3) cells were grown exponentially in 2xYT media, induced with 1 mM IPTG at OD₆₀₀ 0.7, and expressed overnight at 22 °C. Harvested cells were washed with STE buffer (10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, 0.1 M NaCl), and resuspended in buffer A (25 mM Tris-HCl (pH 7.2), 150 mM NaCl, 5 mM 2-mercaptoethanol, 10% glycerol (v/v)). Cells were disrupted by sonication and centrifuged at 10,000 RPM for 30 min. Membrane extracts were then suspended in buffer A and the detergent DDM was added to a final concentration of 1.5%. This mixture was incubated overnight at 4 °C with end-over-end rotation. A second centrifugation yielded a soluble extract which was further used for purification. The soluble extract was incubated at 4 °C for 3 h with 0.5 mL of Ni-NTA equilibrated with buffer B (25 mM Tris-HCl (pH 7.2), 300 mM NaCl, 5 mM 2-mercaptoethanol, 20% glycerol (v/v)), supplemented with 20 mM imidazole. The protein bound slurry was packed in a 2 mL Biorad column and washed with 150 mL buffer B containing 0.1% DDM. Protein was eluted using 300 mM imidazole in buffer B at pH 7.2, then stored in buffer B without DDM at –80 °C. The purity and integrity of recombinant proteins were checked by SDS-PAGE gel.

4.5. Inhibition assays

Putative inhibitors for UppP were serially diluted from 1 mM to 1 nM in DMSO. Inhibitors were incubated with 20 nM of UppP at room temperature for 15 min in assay buffer (50 mM HEPES, 150 mM NaCl, 10 mM MgCl₂, and 0.02% DDM) before adding 35 μM FPP to start the reaction. Reaction mixtures were incubated at 37 °C for 20 min, then quenched by addition of malachite-green reagent [35]. Phosphate release was measured at 620 nm and values obtained were used to yield a dose-response curve.

4.6. Computational calculation

Molecular docking of the lead compound **1a** and the two thiophene-containing derivatives **20** and **22** against UppP was done, using MOE (Molecular Operating Environment, version 2016.0802, Chemical Computing Group ULC, Montreal, Canada). The crystal structure of UppP (code: 6CB2) was downloaded from the protein data bank. The forcefield used was OPLS-AA. The crystal structure was prepared and validated, and the ligand-protein interactions were investigated after docking. Van der Waals volume was also calculated, using Chemicalize online tool (February 2019), <https://chemicalize.com/>, developed by ChemAxon <http://www.chemaxon.com>.

Acknowledgements

This work was funded by the Academy of Scientific Research and Technology (JESOUR-D program, Grant ID# 3092). The authors would like to thank NARSA program and BEI Resources for providing the clinical isolates of *S. aureus* used in this study. The authors would like also to thank Mrs. Zainab A. Alrashed, a 5th year PharmD student at Imam Abdulrahman Bin Faisal University for designing the graphical abstract of the manuscript.

Appendix A. Supplementary data

Supporting data related to this article (scanned copies for all NMR spectra, Figures 1S and 2S and Table 1S & 2S) can be found online.

Supplementary data to this article can be found online at

<https://doi.org/10.1016/j.ejmech.2019.04.063>.

References

- [1] T.J. Beveridge, Structures of gram-negative cell walls and their derived membrane vesicles, *J. Bacteriol.* 181 (1999) 4725–4733.
- [2] T.-Y. Zhang, C.-J. Zheng, J. Wu, L.-P. Sun, H.-R. Piao, Synthesis of novel dihydrotriazine derivatives bearing 1,3-diaryl pyrazole moieties as potential antibacterial agents, *Bioorg. Med. Chem. Lett.* 29 (9) (1 May 2019) 1079–1084 (Ahead of Print).
- [3] G.-F. Zha, S.-M. Wang, K.P. Rakesh, S.N.A. Bukhari, H.M. Manukumar, H.K. Vivek, N. Mallesha, H.-L. Qin, Discovery of novel aryloxythienosulfonyl fluorides as potential candidates against methicillin-resistant of *Staphylococcus aureus* (MRSA) for overcoming multidrug resistance of bacterial infections, *Eur. J. Med. Chem.* 162 (2019) 364–377.
- [4] Y.-S. Yang, X. Lu, Q.-X. Zeng, J. Pang, T.-Y. Fan, X.-F. You, S. Tang, Y.-X. Wang, D.-Q. Song, Synthesis and biological evaluation of 7-substituted cycloberberine derivatives as potent antibacterial agents against MRSA, *Eur. J. Med. Chem.* 168 (2019) 283–292.
- [5] D.B. Tiz, Z. Skok, M. Durcik, T. Tomasic, L.P. Masic, J. Ilas, A. Zega, D. Kikelj, G. Draskovits, T. Revesz, A. Nyerges, C. Pal, C.D. Cruz, P. Tammela, D. Zigon, N. Zidar, An optimised series of substituted N-phenylpyrrolamides as DNA gyrase B inhibitors, *Eur. J. Med. Chem.* 167 (2019) 269–290.
- [6] Z. Fang, S. Zheng, W. Yuan, K.-F. Chan, K.-Y. Wong, Q. Guo, W. Wu, H.-K. Lui, Y.-C. Leung, Y. Lu, T.-H. Chan, N. Sun, Design, synthesis and antibacterial evaluation of 2,4-disubstituted-6-thiophenyl-pyrimidines, *Eur. J. Med. Chem.* 161 (2019) 141–153.
- [7] S. Cai, W. Yuan, Y. Li, X. Huang, Q. Guo, Z. Tang, Z. Fang, H. Lin, W.L. Wong, K.Y. Wong, Y.J. Lu, N. Sun, Antibacterial activity of indolyl-quinolinium derivatives and study their mode of action, *Bioorg. Med. Chem.* 27 (2019) 1274–1282.
- [8] Y. Wu, X. Ding, L. Ding, Y. Zhang, L. Cui, L. Sun, W. Li, D. Wang, Y. Zhao, Synthesis and antibacterial activity evaluation of novel biarylloxazolidinone analogues containing a hydrazone moiety as promising antibacterial agents, *Eur. J. Med. Chem.* 158 (2018) 247–258.
- [9] Y. Wang, F. Chen, H. Di, Y. Xu, Q. Xiao, X. Wang, H. Wei, Y. Lu, L. Zhang, J. Zhu, C. Sheng, L. Lan, J. Li, Discovery of potent benzofuran-derived diaphylotheone desaturase (CrtN) inhibitors with enhanced oral bioavailability for the treatment of methicillin-resistant *Staphylococcus aureus* (MRSA) infections, *J. Med. Chem.* 59 (2016) 3215–3230.
- [10] R. Bouley, D. Ding, Z. Peng, M. Bastian, E. Lastochkin, W. Song, M.A. Suckow, V.A. Schroeder, W.R. Wolter, S. Mobashery, M. Chang, Structure-activity relationship for the 4(3H)-Quinazolinone antibacterials, *J. Med. Chem.* 59 (2016) 5011–5021.
- [11] L. Zhang, K.V. Kumar, S. Rasheed, S.-L. Zhang, R.-X. Geng, C.-H. Zhou, Design, synthesis, and antibacterial evaluation of novel azolylthioether quinolones as MRSA DNA intercalators [Erratum to document cited in CA163:170094], *MedChemComm* 6 (2015) 1405–1406.
- [12] P.I. O'Daniel, Z. Peng, H. Pi, S.A. Testero, D. Ding, E. Spink, E. Leemans, M.A. Boudreau, T. Yamaguchi, V.A. Schroeder, W.R. Wolter, L.I. Llarull, W. Song, E. Lastochkin, M. Kumarasiri, N.T. Antunes, M. Espahbodi, K. Lichtenwalter, M.A. Suckow, S. Vakulenko, S. Mobashery, M. Chang, Discovery of a new class of non-β-lactam inhibitors of penicillin-binding proteins with gram-positive antibacterial activity, *J. Am. Chem. Soc.* 136 (2014) 3664–3672.
- [13] M. ElAwamy, H. Mohammad, A. Hussien, N.S. Abutaleb, M. Hagrass, R.A.T. Serya, A.T. Taher, K.A. Abouzid, M.N. Seleem, A.S. Mayhoub, Alkoxypenthythiazoles with broad-spectrum activity against multidrug-resistant gram-positive bacterial pathogens, *Eur. J. Med. Chem.* 152 (2018) 318–328.
- [14] M.M. Elsebaei, H. Mohammad, M. Abouf, N.S. Abutaleb, Y.A. Hegazy, A. Ghiaty, L. Chen, J. Zhang, S.R. Malwal, E. Oldfield, M.N. Seleem, A.S. Mayhoub, Alkynyl-containing phenylthiazoles; systemically active antibacterial agents effective against methicillin-resistant *Staphylococcus aureus* (MRSA), *Eur. J. Med. Chem.* 148 (2018) 195–209.
- [15] M. Hagrass, N.S. Abutaleb, A.O. Ali, J.A. Abdel-Aleem, M.M. Elsebaei, M.N. Seleem, A.S. Mayhoub, Naphthylthiazoles: targeting multidrug-resistant and intracellular *Staphylococcus aureus* with biofilm disruption activity, *ACS Infect. Dis.* 4 (2018) 1679–1691.
- [16] M. Hagrass, Y.A. Hegazy, A.H. Elkabbany, H. Mohammad, A. Ghiaty, T.M. Abdelghany, M.N. Seleem, A.S. Mayhoub, Biphenylthiazole antibiotics with an oxadiazole linker: an approach to improve physicochemical properties and oral bioavailability, *Eur. J. Med. Chem.* 143 (2018) 1448–1456.
- [17] A. Koth, N.S. Abutaleb, M.A. Seleem, M. Hagrass, H. Mohammad, A. Bayoumi, A. Ghiaty, M.N. Seleem, A.S. Mayhoub, Phenylthiazoles with tert-Butyl side chain: metabolically stable with anti-biofilm activity, *Eur. J. Med. Chem.* 151 (2018) 110–120.
- [18] I. Eid, M.M. Elsebaei, H. Mohammad, M. Hagrass, C.E. Peters, Y.A. Hegazy, B. Cooper, J. Pogliano, K. Pogliano, H.S. Abulkhair, M.N. Seleem, A.S. Mayhoub, Arylthiazole antibiotics targeting intracellular methicillin-resistant *Staphylococcus aureus* (MRSA) that interfere with bacterial cell wall synthesis, *Eur. J. Med. Chem.* 139 (2017) 665–673.
- [19] H. Mohammad, P.V. Reddy, D. Monteleone, A.S. Mayhoub, M. Cushman, G.K. Hammac, M.N. Seleem, Antibacterial characterization of novel synthetic Thiazole compounds against methicillin-resistant *Staphylococcus*

- pseudintermedius, *PLoS One* 10 (2015), e0130385.
- [20] M.A. Seleem, A.M. Disouky, H. Mohammad, T.M. Abdelghany, A.S. Mancy, S.A. Bayoumi, A. Elshafeey, A. El-Morsy, M.N. Seleem, A.S. Mayhoub, Second-generation phenylthiazole antibiotics with enhanced pharmacokinetic properties, *J. Med. Chem.* 59 (2016) 4900–4912.
 - [21] M. Hagrass, Y.A. Hegazy, A.H. Elkabbany, H. Mohammad, A. Ghiaty, T.M. Abdelghany, M.N. Seleem, A.S. Mayhoub, Biphenylthiazole antibiotics with an oxadiazole linker: an approach to improve physicochemical properties and oral bioavailability, *Eur. J. Med. Chem.* 143 (1 January 2018) 1448–1456.
 - [22] M.F. Mohamed, A. Abdelkhalek, M.N. Seleem, Evaluation of short synthetic antimicrobial peptides for treatment of drug-resistant and intracellular *Staphylococcus aureus*, *Sci. Rep.* 6 (2016) 29707.
 - [23] H. Mohammad, A.S. Mayhoub, A. Ghafoor, M. Soofi, R.A. Alajlouni, M. Cushman, M.N. Seleem, Discovery and characterization of potent thiazoles versus methicillin- and vancomycin-resistant *Staphylococcus aureus*, *J. Med. Chem.* 57 (2014) 1609–1615.
 - [24] R.J. Gordon, F.D. Lowy, Pathogenesis of methicillin-resistant *Staphylococcus aureus* infection, *Clin. Infect. Dis.* 46 (2008) S350–S359.
 - [25] C. Vuong, M. Otto, *Staphylococcus epidermidis* infections, *Microb. Infect.* 4 (2002) 481–489.
 - [26] C. Hunt, M. Dionne, M. Delorme, D. Murdock, A. Erdreich, D. Wolsey, A. Groom, J. Cheek, J. Jacobson, B. Cunningham, L. Shireley, K. Belani, S. Kurachek, P. Ackerman, S. Cameron, P. Schlievert, J. Pfeiffer, S. Johnson, D. Boxrud, J. Bartkus, J. Besser, K. Smith, K. LeDell, C. O'Boyle, R. Lynfield, K. White, M. Osterholm, K. Moore, R. Danila, E. officers, Four pediatric deaths from community-acquired methicillin-resistant *Staphylococcus aureus* - Minnesota and North Dakota, 1997–1999, *JAMA, J. Am. Med. Assoc.* 282 (1999) 1123–1125.
 - [27] E. Tacconelli, M. Tumbarello, M. Pittiruti, F. Leone, M. Lucia, R. Cauda, L. Ortona, Central venous catheter-related sepsis in a cohort of 366 hospitalised patients, *Eur. J. Clin. Microbiol. Infect. Dis.* 16 (1997) 203–209.
 - [28] F. Husain, M. Humbard, R. Misra, Interaction between the TolC and AcrA proteins of a multidrug efflux system of *Escherichia coli*, *J. Bacteriol.* 186 (2004) 8533–8536.
 - [29] F.J. Schmitz, A.C. Fluit, D. Hafner, A. Beeck, M. Perdikouli, M. Boos, S. Scheuring, J. Verhoef, K. Kohrer, C. Von Eiff, Development of resistance to ciprofloxacin, rifampin, and mupirocin in methicillin-susceptible and -resistant *Staphylococcus aureus* isolates, *Antimicrob. Agents Chemother.* 44 (2000) 3229–3231.
 - [30] R.H.K. Eng, S.M. Smith, M. Tillem, C. Cherubin, Rifampin resistance - development during the therapy of methicillin-resistant *staphylococcus-aureus* infection, *Arch. Intern. Med.* 145 (1985) 146–148.
 - [31] M.J.J. van Rensburg, A.C. Whitelaw, B.G. Elisha, Genetic basis of rifampicin resistance in methicillin-resistant *Staphylococcus aureus* suggests clonal expansion in hospitals in Cape Town, South Africa, *BMC Microbiol.* 12 (2012).
 - [32] B. Testa, P. Crivori, M. Reist, P.-A. Carrupt, The influence of lipophilicity on the pharmacokinetic behavior of drugs: concepts and examples, *Perspect. Drug Discov. Des.* 19 (2000) 179–211.
 - [33] M.F. Mohamed, A. Abdelkhalek, M.N. Seleem, Evaluation of short synthetic antimicrobial peptides for treatment of drug-resistant and intracellular *Staphylococcus aureus*, *Sci Rep-Uk* 6 (2016).
 - [34] K.T. Savjani, A.K. Gajjar, J.K. Savjani, Drug Solubility: Importance and Enhancement Techniques, *ISRN pharmaceutics*, 2012, 2012.
 - [35] J. Desai, Y.D. Wang, K.D. Wang, S.R.D. Malwal, E. Oldfield, Isoprenoid biosynthesis inhibitors targeting bacterial cell growth, *ChemMedChem* 11 (2016) 2205–2215.
 - [36] J.B. Taylor, D.J. Triggle, *Comprehensive Medicinal Chemistry II*, Elsevier, 2007.
 - [37] H. Mohammad, P.V.N. Reddy, D. Monteleone, A.S. Mayhoub, M. Cushman, M.N. Seleem, Synthesis and antibacterial evaluation of a novel series of synthetic phenylthiazole compounds against methicillin-resistant *Staphylococcus aureus* (MRSA), *Eur. J. Med. Chem.* 94 (2015) 306–316.
 - [38] H. Mohammad, A. AbdelKhalek, N.S. Abutaleb, M.N. Seleem, Repurposing niclosamide for intestinal decolonization of vancomycin-resistant enterococci, *Int. J. Antimicrob. Agents* 51 (2018) 897–904.
 - [39] M.F. Mohamed, G.K. Hammac, L. Guptill, M.N. Seleem, Antibacterial activity of novel cationic peptides against clinical isolates of multi-drug resistant *Staphylococcus pseudintermedius* from infected dogs, *PLoS One* 9 (2014), e116259.
 - [40] H. Mohammad, W. Younis, H.G. Ezzat, C.E. Peters, A. AbdelKhalek, B. Cooper, K. Pogliano, J. Pogliano, A.S. Mayhoub, M.N. Seleem, Bacteriological profiling of diphenylureas as a novel class of antibiotics against methicillin-resistant *Staphylococcus aureus*, *PLoS One* 12 (2017), e0182821.
 - [41] A. AbdelKhalek, N.S. Abutaleb, K.A. Elmagarmid, M.N. Seleem, Repurposing aurano-fin as an intestinal decolonizing agent for vancomycin-resistant enterococci, *Sci Rep-Uk* 8 (2018).
 - [42] A. AbdelKhalek, N.S. Abutaleb, H. Mohammad, M.N. Seleem, Repurposing eb-selen for decolonization of vancomycin-resistant enterococci (VRE), *PLoS One* 13 (2018), e0199710.
 - [43] A. AbdelKhalek, C.R. Ashby Jr., B.A. Patel, T.T. Talele, M.N. Seleem, In vitro antibacterial activity of rhodanine derivatives against pathogenic clinical isolates, *PLoS One* 11 (2016), e0164227.
 - [44] M.F.N. Abushahba, H. Mohammad, S. Thangamani, A.A.A. Hussein, M.N. Seleem, Impact of different cell penetrating peptides on the efficacy of antisense therapeutics for targeting intracellular pathogens, *Sci Rep-Uk* 6 (2016).
 - [45] Y.H. Pei, M.F. Mohamed, M.N. Seleem, Y. Yeo, Particle engineering for intracellular delivery of vancomycin to methicillin-resistant *Staphylococcus aureus* (MRSA)-infected macrophages, *J. Control. Release* 267 (2017) 133–143.
 - [46] M.A. Seleem, A.M. Disouky, H. Mohammad, T.M. Abdelghany, A.S. Mancy, S.A. Bayoumi, A. Elshafeey, A. El-Morsy, M.N. Seleem, A.S. Mayhoub, Second-generation phenylthiazole antibiotics with enhanced pharmacokinetic properties, *J. Med. Chem.* 59 (2016) 4900–4912.

From Phenylthiazoles to Phenylpyrazoles: Broadening the Antibacterial Spectrum toward Carbapenem-Resistant Bacteria

Ali Hammad,^{†,Δ} Nader S. Abutaleb,^{‡,Δ,⊙} Mohamed M. Elsebaei,[†] Allison B. Norvil,[§] Mohamed Alswah,[†] Alsagher O. Ali,^{‡,||} Jelan A. Abdel-Aleem,^{‡,⊥} Abdelaziz Alattar,[#] Sammar A. Bayoumi,[¶] Humaira Gowher,^{§,∇} Mohamed N. Seleem,^{*,‡,⊙,Ⓛ} and Abdelrahman S. Mayhoub^{*,†,◆,Ⓛ}

[†]Department of Pharmaceutical Organic Chemistry, College of Pharmacy, Al-Azhar University, Cairo 11884, Egypt

[‡]Department of Comparative Pathobiology, College of Veterinary Medicine, [§]Department of Biochemistry, College of Agriculture, and [∇]Purdue University Center for Cancer Research, Purdue University, West Lafayette, Indiana 47907, United States

^{||}Division of Infectious Diseases, Animal Medicine Department, Faculty of Veterinary Medicine, South Valley University, Qena, 83523, Egypt

[⊥]Department of Industrial Pharmacy, Faculty of Pharmacy, Assiut University, Assiut, 71515, Egypt

[#]Department of Analytical Chemistry, College of Pharmacy, Al-Azhar University, Cairo 11884, Egypt

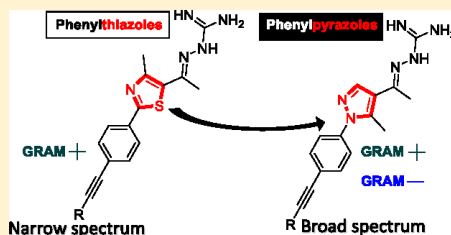
[¶]Department of Pharmaceutics, College of Pharmacy, Heliopolis University, Cairo, 11777, Egypt

[⊙]Purdue Institute for Inflammation, Immunology, and Infectious Diseases, West Lafayette, Indiana 47907, United States

[◆]University of Science and Technology, Nanoscience Program, Zewail City of Science and Technology, October Gardens, 6th of October, Giza 12578, Egypt

S Supporting Information

ABSTRACT: The narrow antibacterial spectrum of phenylthiazole antibiotics was expanded by replacing central thiazole with a pyrazole ring while maintaining its other pharmacophoric features. The most promising derivative, compound 23, was more potent than vancomycin against multidrug-resistant Gram-positive clinical isolates, including vancomycin- and linezolid-resistant methicillin-resistant *Staphylococcus aureus* (MRSA), with a minimum inhibitory concentration (MIC) value as low as 0.5 $\mu\text{g/mL}$. Moreover, compound 23 was superior to imipenem and meropenem against highly pathogenic carbapenem-resistant strains, such as *Acinetobacter baumannii*, *Klebsiella pneumoniae*, and *Escherichia coli*. In addition to the notable biofilm inhibition activity, compound 23 outperformed both vancomycin and kanamycin in reducing the intracellular burden of both Gram-positive and Gram-negative pathogenic bacteria. Compound 23 cleared 90% of intracellular MRSA and 98% of *Salmonella enteritidis* at 2X the MIC. Moreover, preliminary pharmacokinetic investigations indicated that this class of novel antibacterial compounds is highly metabolically stable with a biological half-life of 10.5 h, suggesting a once-daily dosing regimen.



■ INTRODUCTION

According to a report released from the World of Health Organization (WHO) in September 2017, the world is running out of antibiotics¹ for both Gram-positive and Gram-negative bacteria. The list of the most critical priority pathogens for antibiotic development is dominated by Gram-negative bacteria.¹ Gram-negative pathogens are characterized by an impassable outer membrane that prevents most of the chemical compounds, leaving porin channels as the only gate for the antibacterial agents to cross through.² As porins are lined with highly charged residues such as aspartate, glutamate, and arginine,³ the physicochemical properties of antibacterial compounds are a crucial factor that medicinal chemists have

to consider. In general, drugs targeting Gram-negative pathogens are more polar compared to others having limited anti-Gram-negative activity.⁴

This work is built on the discovery of a lead compound, phenylthiazole, which bears an alkyl side chain on one side and a guanidine head on the other side; it possesses antimicrobial activity against Gram-positive pathogens.⁵ This initial discovery was followed by intensive structural optimization to improve its antimicrobial activity and metabolic stability.^{6–15} The structural optimization focused on the nitrogenous part,

Received: May 2, 2019

Published: August 1, 2019

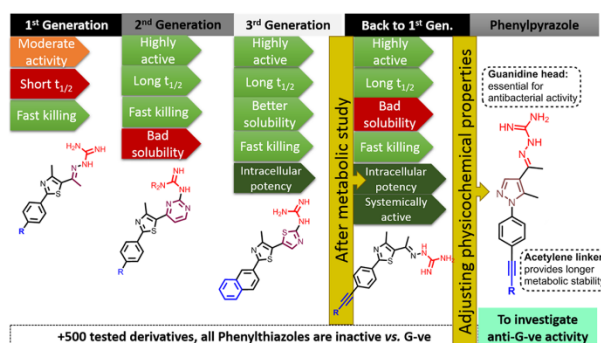


Figure 1. Summary of the developmental progress of phenylthiazoles antibiotics and the objective of the current study.

the lipophilic tail, and the connection with the phenylthiazole scaffold.^{6–15} Our structural optimization efforts with first generation compounds using different heterorings as a linker between the guanidine head and the main scaffold overcame its metabolic instability and improved its antibacterial activity (Figure 1).^{6,9,13,15} Using an alkynyl lipophilic moiety blocked the metabolic soft spot,¹⁴ increased biological half-life, and enhanced the systemic efficiency.⁸

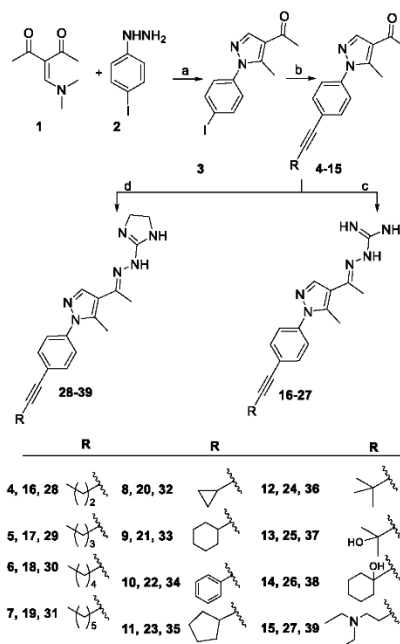
Although the synthesized phenylthiazoles had several advantages, such as being able to clear intracellular pathogens, penetrate, and reduce the burden of the bacterial biofilm and induce the fast lethal mode of action, they exhibited a narrow spectrum of activity in that they inhibit Gram-positive pathogens only (Table S1). Because recent mechanistic studies suggested that phenylthiazoles exert their antibacterial effects via the suppression of two vital proteins involved in bacterial cell wall synthesis, which are undecaprenyl pyrophosphate phosphatase (UppP) and undecaprenyl pyrophosphate synthase (UppS),^{8,12} that are essential in both Gram-positive and Gram-negative bacterial strains, we hypothesized that the observed narrow spectrum of activity can be related to the poor permeability of the compounds into Gram-negative pathogens.

The overall objective of this study is to design compounds capable of targeting Gram-negative pathogens. Because the structure–activity relationships (SAR) of our phenylthiazoles established that both the guanidine structure (colored red in Figure 1) and lipophilic part (colored blue in Figure 1) are essential for the antimicrobial activity, we turned our attention to the central thiazole ring and replaced it with a more polarazole moiety; that is, the pyrazole ring (Figure 1). This structural modification would keep the essential structural elements in their positions while decreasing the lipophilicity by a factor of 25%, as measured by log *P* values (Figure 1). This includes the design and synthesis of phenylpyrazoles bearing the most promising side chains, in terms of pharmacodynamic and pharmacokinetic properties; that is, alkynyl derivatives, and investigating their microbiological profiles versus a large panel of highly and extremely drug-resistant Gram-positive and Gram-negative clinical isolates. Additionally, the suitability of the most promising compounds for systemic administration was studied via addressing the key pharmacokinetic parameters.

RESULTS AND DISCUSSION

Chemistry. The reaction of 3-[(dimethylamino)methylene]pentane-2,4-dione (**1**), in situ generated from acetylacetone and dimethylformamide–dimethylacetal, with *p*-iodophenylhydrazine (**2**), afforded the starting building block *N*-phenylpyrazole **3** (Scheme 1). The latter key starting material was then allowed to react with terminal acetylene

Scheme 1⁴⁴



^aReagents and conditions: (a) Absolute EtOH, heat to reflux, 12 h; (b) proper terminal acetylene, PdCl₂(PPh₃)₂ (5% mol), CuI (7.5% mol), Et₃N, DME, heat at 65 °C for 16 h; (c) aminoguanidine HCl, EtOH, Conc. HCl, heat to reflux 12 h; (d) 2-hydrazinyl-4,5-dihydro-1H-imidazole HBr, EtOH, conc. HCl, heat to reflux 16 h.

synthons, under standard Sonogashira conditions, to yield compounds 4–15. Finally, condensation of acetyl-containing intermediates 4–15 with aminoguanidine or its cyclic 2-hydrazinylimidazole salts provided the final products 16–39 (Scheme 1).

Biological Results and Discussion. The highly contagious methicillin-resistant *Staphylococcus aureus* (MRSA) USA300 clinical isolate was chosen for initial evaluation of activity against Gram-positive bacteria. The TolC mutant *Escherichia coli* (Δ tolC, encoding for efflux pump) was selected for the initial anti-Gram-negative activity. Linear acetylenes (compounds 16–19) provided an almost flat SAR with narrow difference between pentyl 16 [minimum inhibitory concentration (MIC) value of 4 μ g/mL] and other longer derivatives, such as compounds 17–19 (MIC value of 2 μ g/mL), against MRSA USA300 (Table 1). On the other hand, the activity of phenylpyrazole with a longer side chain (compound 19) was one-fold less potent than other analogs against TolC mutant *E. coli*.

Table 1. MIC (in μ g/mL) of Compounds Initially Screened against MRSA USA300 and TolC Mutant *E. coli*

compounds/control drugs	<i>S. aureus</i> NRS384 (MRSA USA300)	<i>E. coli</i> JWS5031 (TolC mutant strain)
16	4	4
17	2	4
18	2	4
19	2	8
20	8	8
21	2	2
22	1	2
23	1	2
24	8	8
25	128	128
26	2	4
27	128	128
28	8	16
29	2	4
30	2	2
31	4	8
32	16	32
33	4	4
34	4	4
35	4	4
36	16	16
37	32	16
38	64	32
39	>128	>128
vancomycin	1	NT
gentamycin	NT	0.5

To explore the effect of cyclic side chains on the antibacterial activity, we started with the cyclopropylacetylenyl derivative 20, which was one-fold less potent than the corresponding open-chain with the same number of carbon units (compound 16). Expanding the size of the terminal alicyclic moiety by replacing the cyclopropyl group with a cyclopentyl group yielded compound 23, an equi-potent compound to vancomycin, with an MIC value of 1 μ g/mL against MRSA USA300. Parallel to the improvement in its antistaphylococcal activity, the anti-Gram-negative potency was also ameliorated. In this regard, the MIC value of compound

23 against *E. coli* (TolC mutant strain) was one time lower than the corresponding opened-chain analog (compound 18). Further expanding the ring size produced the cyclohexyl compound 21 and its aromatized analog 22 with almost similar antibacterial activities on both MRSA USA300 and *E. coli*. Moving from the cyclic to branched side chains was associated with four- to eight-fold drop in antibacterial activity as observed with compound 24 (Table 1).

Further branching, as in the case of compound 27, nullified the antibacterial activities against both tested microorganisms. The attempt to reduce the lipophilicity of the terminal side chain by introducing a polar hydroxyl group ended up with variable impact as the hydroxycyclohexyl derivative 26 maintained the antibacterial effect, observed with the cyclohexyl 21, on both tested species. On the other hand, the replacement of one methyl group from compound 24 with a hydroxyl led to an inactive compound 25 (Table 1). Finally, the aminoimidazoline analogs 28–39 showed close SAR.

Next, the spectrum of the antibacterial activity of eight selected compounds (16, 18, 21, 22, 23, 26, 30, and 35) was examined against a large panel of clinically relevant Gram-positive and Gram-negative bacterial pathogens. Multidrug-resistant (MDR) staphylococcal isolates continue to be a source of healthcare-associated infections, including bloodstream infections, osteomyelitis, sepsis, and necrotizing pneumonia.^{16,17} In the United States alone, nearly half of the fatalities caused by drug-resistant pathogens are attributed to MRSA.¹⁸ In addition, it has been reported that some clinical isolates of MRSA that are resistant to nearly all antibiotic classes, such as β -lactams, fluoroquinolones, macrolides, tetracyclines, and lincosamides, had been isolated.^{19,20} The problem was further exacerbated by the emergence of clinical isolates exhibiting resistance to the last resort antibiotics such as vancomycin and linezolid.^{21,22} We tested eight selected compounds (16, 18, 21, 22, 23, 26, 30, and 35) against methicillin-sensitive *S. aureus*, methicillin-resistant *S. aureus* (MRSA USA300, USA400, USA500 and USA700), vancomycin-resistant *S. aureus* (VRSA), and linezolid-resistant MRSA. Remarkably, all the tested phenylpyrazoles exhibited potent antibacterial activity against all tested staphylococcal strains, inhibiting their growth at concentrations mostly ranging between 1 and 2 μ g/mL (Table 2). In particular, compound 22 was as effective as vancomycin and linezolid, the drugs of choice for the treatment of staphylococcal infections. Moreover, the cyclopentyl 23 was superior to vancomycin and linezolid against MRSA USA500 (Table 2), the highly contagious community-acquired strain.²³ The minimum bactericidal concentration (MBC) values for the compounds were found to be equal to or up to two-fold higher than the compounds' MIC values against the tested bacterial strains, indicating the compounds are bactericidal (Table S2). The compounds' bactericidal mode of action was then confirmed by a standard time-kill assay (Figure S1).

The antibacterial evaluation was then broadened to include eight additional Gram-positive strains including *Staphylococcus epidermidis*, vancomycin-resistant enterococci (VRE), *Listeria monocytogenes*, *Clostridium difficile*, and cephalosporin-resistant *Streptococcus pneumoniae*. The eight tested phenylpyrazoles exhibited potent antibacterial activity against the tested Gram-positive pathogens, inhibiting growth of the tested strains at concentrations ranging from 0.5 to 4 μ g/mL (Table 3). The antibacterial activity of the cyclopentyl-containing derivative 23 outperformed both frontline therapeutics (vancomycin and

Table 2. Antimicrobial Activity of Phenylpyrazoles (MICs, in $\mu\text{g/mL}$) against Clinically Relevant Drug-Resistant Staphylococcal Clinical Isolates^a

bacterial strains	compounds/control antibiotics									
	16	18	21	22	23	26	30	35	Lin	van
<i>S. aureus</i> ATCC 6538	2	2	2	1	0.5	2	2	1	1	1
<i>S. aureus</i> NRS107	4	2	2	2	1	4	4	2	1	2
linezolid-resistant MRSA NRS119	2	2	2	1	1	2	2	4	32	1
MRSA NRS123 (USA400)	2	2	2	2	2	2	4	4	1	1
MRSA NRS384 (USA300)	4	2	2	1	1	2	2	4	1	1
MRSA NRS385 (USA500)	2	2	2	1	0.5	2	2	4	2	1
MRSA NRS386 (USA700)	4	2	2	2	1	2	2	4	2	1
VRSA10	2	2	2	1	0.5	2	2	8	1	>64
VRSA12	2	2	2	2	1	2	2	4	1	64

^aVRSA: vancomycin-resistant *S. aureus*. Lin: linezolid. Van.: vancomycin.**Table 3. Antimicrobial Activity of Phenylpyrazoles (MICs, in $\mu\text{g/mL}$) against Other Clinically Important Gram-Positive Bacterial Pathogens**

bacterial strains	compounds/control antibiotics									
	16	18	21	22	23	26	30	35	Lin	Van
<i>S. epidermidis</i> NRS 101	4	2	2	2	1	4	2	2	≤0.5	1
<i>E. faecalis</i> ATCC 51299 (VRE) ^a	4	2	2	1	0.5	2	4	8	1	32
<i>E. faecium</i> ATCC 700221 (VRE)	4	2	1	1	1	2	2	4	1	>64
<i>L. monocytogenes</i> ATCC 19111	4	2	2	2	0.5	2	2	4	1	2
cephalosporin-resistant <i>S. pneumoniae</i> ATCC 51916	4	2	2	1	1	4	4	8	0.5	2
methicillin-resistant <i>S. pneumoniae</i> ATCC 700677	4	2	2	2	0.5	2	2	4	1	1
<i>C. difficile</i> ATCC BAA 1870	8	4	4	4	1	4	4	4	NT ^b	1

^aVRE: vancomycin-resistant Enterococci. ^bNT: not tested.**Table 4. Antimicrobial Activity of Phenylpyrazoles (MICs in $\mu\text{g/mL}$) against Gram-Negative Bacterial Pathogens^a**

bacterial strains	compounds/control antibiotics								
	16	18	21	22	23	26	30	35	Gen
<i>A. baumannii</i> ATCC 19606	8	4	4	4	4	4	8	16	16
<i>A. baumannii</i> ATCC BAA 1141	8	4	4	4	4	8	>128	16	4
<i>A. baumannii</i> ATCC BAA 1747	8	4	4	4	4	16	>128	16	0.5
<i>A. baumannii</i> NR 17786	16	8	8	8	8	16	>128	16	64
<i>A. baumannii</i> NR 17785	8	8	8	4	4	8	>128	8	>64
<i>E. coli</i> BW25113 ¹	8	8	4	8	8	8	>128	64	0.5
<i>E. coli</i> ATCC 25922	8	8	4	4	4	8	>128	32	1
<i>E. coli</i> ATCC 35150	8	8	4	8	4	16	>128	16	1
<i>E. coli</i> ATCC 1411	16	16	8	8	8	32	>128	64	0.5
<i>E. cloacae</i> subsp. <i>cloacae</i> ATCC BAA-1143	16	8	8	8	8	16	>128	32	1
<i>K. pneumoniae</i> BAA-1706	8	8	8	8	4	8	>128	32	8
<i>K. pneumoniae</i> BAA-1144	16	16	16	8	8	16	>128	64	16
<i>Salmonella enterica</i> NR-170	16	8	8	8	8	16	>128	32	1
<i>S. enteritidis</i> ATCC 13076	16	8	16	8	4	32	>128	32	1
<i>S. enteritidis</i> ATCC 14028	16	16	32	16	8	32	>128	32	1
<i>Salmonella typhimurium</i> ATCC 700720	16	8	8	8	8	64	>128	64	1
<i>S. flexneri</i> ATCC 9199	16	8	8	8	8	16	>128	32	1
<i>P. aeruginosa</i> ATCC 9027	32	16	16	16	16	32	>128	32	0.5
<i>P. aeruginosa</i> ATCC 25619	16	32	64	32	16	32	>128	64	0.5
<i>P. aeruginosa</i> ATCC 15442	16	32	32	32	32	64	>128	64	2

^aGen: gentamicin.

linezolid), especially against VRE (*Enterococcus faecalis* and *Enterococcus faecium*, the two clinically important strains with very limited treatment options).^{24–26} The high potency of derivative **23** was also observed against MDR-*S. pneumoniae*, a notorious microorganism that is responsible for the vast mortality of lobar pneumonia.^{29,30} Moreover, compound **23** was as effective as linezolid and vancomycin against *S.*

epidermidis. *S. epidermidis* possesses the ability to form strong adherent biofilms on indwelling catheters and implanted medical devices causing a wide range of nosocomial infections, cardiovascular infections, and eye, ear, nose, and throat infections.^{31,32} Additionally, compound **23** retained the same activity against another clinically important pathogen, *L. monocytogenes*. The latter is a foodborne pathogen that infects

Table 5. Antimicrobial Activity (MICs, in $\mu\text{g/mL}$) of Phenylpyrazoles against Carbapenem-Resistant Gram-Negative Bacterial Pathogens Including *A. baumannii*, *E. coli*, *K. pneumoniae*, and *P. aeruginosa*^a

bacterial strains	compounds/control antibiotics										
	16	18	21	22	23	26	35	Imp	Mer	Tig	Col
<i>E. coli</i> ATCC BAA-2452	32	16	8	8	4	64	16	32	32	0.25	0.06
<i>E. coli</i> ATCC BAA-2469	32	8	4	8	4	64	32	32	32	0.5	0.06
<i>K. pneumoniae</i> ATCC BAA-1705	16	16	8	16	8	64	64	32	32	1	0.125
<i>K. pneumoniae</i> ATCC BAA-2146	32	32	8	16	16	>64	64	32	64	4	0.25
<i>A. baumannii</i> ATCC BAA 1605	16	8	4	4	8	32	16	32	32	1	0.25
<i>P. aeruginosa</i> ATCC 1744	32	32	32	16	16	64	32	16	16	8	1
MIC ₉₀	32	32	32	16	16	>64	64	32	64	8	1

^aImp: imipenem. Mer: meropenem. Tig: tigecycline. Col: colistin. MIC₉₀: the concentration which inhibited 90% of the tested strains.

humans and animals and can lead to serious, often fatal central nervous system infections (listeriosis). Its infections are mostly found among pregnant women, newborns, and immunocompromised patients.^{33,34} It is usually ranked as the third or fourth most common cause of bacterial meningitis in North America and Western Europe.³⁵ Again, the MBC values for all compounds were found to be less than 3-fold higher than their corresponding MIC values against the tested bacterial strains, indicating the compounds are bactericidal (Table S3).

After confirming the activity against multiple Gram-positive pathogens, we moved next to expand our understanding of the effect of the newly developed phenylpyrazoles against Gram-negative clinical isolates. Table 1 provided some primitive information about the anti-Gram-negative activity as tested against TolC mutant *E. coli*. To precisely evaluate the activity of phenylpyrazoles against Gram-negative pathogens, 20 clinical isolates belonging to different families of MDR-Gram-negative species were tested and the results are summarized in Table 4.

With the exception of the heptyl derivative 30, phenylpyrazoles exhibited reasonable antibacterial activity against Gram-negative pathogens, inhibiting the growth of the tested strains at concentrations as low as 4 $\mu\text{g/mL}$. The limited activity of compound 30 could be due to the effect of efflux pumps present in these Gram-negative bacterial cells as it showed potent activity against the TolC mutant *E. coli*, which lacks the efflux pump. Notably, compounds 22 and 23 exhibited the highest activity of all the synthesized derivatives against the highly pathogenic Gram-negative bacteria. Most importantly, compounds 22 and 23 retained their activity against the bacterial strains that showed high resistance to gentamicin, one of the most potent antibiotics used for the treatment of Gram-negative bacterial infections, such as *Klebsiella pneumoniae* ATCC BAA 1144, *Acinetobacter baumannii* ATCC 19606, *A. baumannii* NR 17785, and *A. baumannii* NR 17786. It is worth mentioning that the corresponding phenylthiazoles structures to 22 and 23 (compounds 1S and 1IS; Table S1) did not have any anti-Gram-negative activity (Table S1).

The MBC values for all compounds were found to be equal to or one-fold higher than the compounds' MIC values against the tested bacterial strains, indicating the compounds might have bactericidal activity (Table S3).

To test whether the anti-Gram-negative activity can be maintained against the biggest threat pathogens, such as carbapenem-resistant *Enterobacteriaceae* (CRE), we screened a panel of CRE in addition to MDR *Acinetobacter* and *Pseudomonas* clinical isolates. The selected strains of Gram-negative bacteria have become resistant to nearly all available

antibiotics. According to the Centers for Disease Control and Prevention (CDC), CRE is an urgent public health threat that requires immediate and aggressive action. Each year, infections caused by carbapenem-resistant *Klebsiella* and carbapenem-resistant *E. coli* result in approximately 600 deaths. Additionally, nearly 63% of *Acinetobacter* are MDR strains, including resistance to carbapenem, leading to approximately 500 deaths annually in USA. Moreover, MDR *Pseudomonas* (including carbapenem resistant) infections in USA alone lead to approximately 400 deaths per year.^{35–40} Consequently, there is an unmet urgent need for new drugs for the treatment of infections caused by carbapenem-resistant pathogens. In this context, we assessed the activity of phenylpyrazoles against carbapenem-resistant pathogens. The strains listed in Table 5 are classified as MDR strains. The results of the assessment of our newly synthesized phenylpyrazoles against these pathogens can be considered highly promising. For instance, *A. baumannii* ATCC BAA-1605, isolated from a military personnel returning from Afghanistan, is resistant to most antibiotics used for the treatment, including ceftazidime, gentamicin, ticarcillin, piperacillin, aztreonam, cefepime, ciprofloxacin, imipenem, and meropenem.⁴¹ *K. pneumoniae* ATCC 1705 is resistant to as many as 36 representative antibiotics in a variety of drug classes, including carbapenems, β -lactams, cephalosporins, quinolones, tetracyclines, glycolylcyclines, and aminoglycosides. The same resistance pattern was also reported in the cases of *K. pneumoniae* ATCC 2146 and *Pseudomonas aeruginosa* ATCC 1744.^{42–45} *E. coli* ATCC 2469 and ATCC 2452 are reported to be New Delhi metallo-beta-lactamase (NDM-1) positive and carbapenems-resistant.^{46–48}

In this regard, compounds 22 and 23 exhibited the highest activity against the tested carbapenem-resistant strains. Both compounds inhibited 90% of the tested strains at a concentration of 16 $\mu\text{g/mL}$ (one-fold lower than MIC₉₀ of tigecycline). In conclusion, phenylpyrazoles could serve as excellent starting compounds for developing effective therapeutics for the treatment of CRE, carbapenemase-producing *Acinetobacter* and *Pseudomonas*. Moreover, the MBC values for all compounds were found to be equal to or one-fold higher than the compounds' MIC values against the tested bacterial strains, indicating the compounds have bactericidal activity against these carbapenem-resistant bacteria (Table S3).

Apart from the direct inhibitory activity on bacterial cells, modern antibiotics should possess additional attributes to be efficient in clearing bacterial pathogens. For instance, the ability of a microorganism to build a biofilm mass creates a hardly surmountable shield for traditional antibiotics.^{49,50} For instance, up to a 1000 fold higher than MIC concentration of antibiotics is needed to clear a biofilm. Therefore, many

frontline antibiotics, such as teicoplanin and vancomycin, were excluded from the clinically useful lists for treating the bacterial biofilm.⁵¹ In this regard, all active compounds from the panel of newly synthesized phenylpyrazoles were tested for their ability to inhibit biofilm formation against MRSA USA300. The strain MRSA NRS384 (MRSA300) was selected for this assay based on its biofilm consistency after screening a panel of MDR-biofilm-producing staphylococcal strains (Figure S2). As seen in Figure 2, all active compounds were superior to

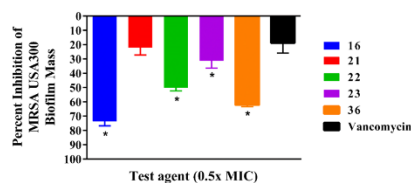


Figure 2. Investigation of the antibiofilm activity of phenylpyrazoles against MRSA USA300. Data are presented as percent inhibition of MRSA USA300 biofilm formation. Dimethyl sulfoxide (DMSO) (the solvent for tested compounds) served as a negative control and vancomycin served as the control antibiotic. The values represent an average of quadruplicates of each compound/drug. Error bars represent standard deviation values. An asterisk (*) denotes statistical significance ($P < 0.05$) between results for the tested compounds and the control antibiotic (vancomycin) analyzed via unpaired Student's *t*-test.

vancomycin in their ability to inhibit biofilm formation. At $0.5 \times \text{MIC}$, vancomycin inhibited 18% of MRSA biofilm formation. Phenylpyrazoles with linear, aromatic, and branched side chains (compounds 16, 22, and 36) exhibited the highest biofilm inhibition activity. At a nontoxic, very low concentration ($0.5 \times \text{MIC}$), compounds 16, 22, and 36 significantly inhibited 74, 50, and 62% of MRSA300 biofilm formation, respectively. This is many fold superior over the cornerstone therapeutic vancomycin. On the other hand, the most promising antibacterial candidate in this series in terms of MIC values (compound 23) was relatively less potent in terms of biofilm-mass reduction, as it significantly inhibited MRSA300 in a biofilm by only 31% at $0.5 \times \text{MIC}$ concentration, which is almost two times significantly higher than the reference drug vancomycin. Compound 21 was nonsignificantly higher than vancomycin in inhibiting MRSA300 biofilm formation, where it inhibited approximately 22% of its biofilm formation at the tested concentration ($0.5 \times \text{MIC}$).

Next, we moved to investigate the ability of our newly synthesized compounds to clear microorganisms harbored intracellularly inside infected macrophages. For the effective management of intracellular bacterial infections, there is a critical need for new types of antimicrobials that will treat persistent and MDR-intracellular bacterial infections.

First, we assessed the ability of compounds 23 and 35 to clear intracellular MRSA inside the infected macrophages. Compound 23 was capable of reducing the intracellular MRSA by $1.01 \log_{10}$ -reduction, which is equivalent to 90% reduction of intracellular MRSA, at $2 \times \text{MIC}$, while its imidazoline analog 35 reduced the intracellular MRSA400 by $1.79 \log_{10}$ -reduction. This is equivalent to approximately 98.4% reduction of intracellular MRSA at the same tested concentration ($2 \times \text{MIC}$) (Figure 3). On the other hand, vancomycin, given its

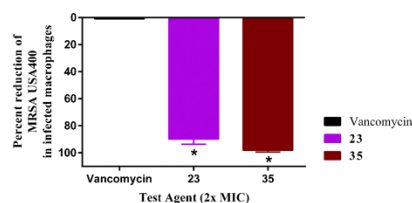


Figure 3. Activity of compounds 23 and 35 on the clearance of intracellular MRSA USA400 present in murine macrophage (J774) cells. Data are presented as percent reduction of MRSA USA400 colony-forming units inside infected murine macrophage cells after the treatment with $2 \times \text{MIC}$ of compounds 23 and 35 (tested in triplicate) for 24 h. Data were analyzed via an unpaired Student's *t*-test ($P < 0.05$). Asterisks (*) represent a significant difference between the treatment of J774 cells with compounds 23 and 35 in comparison to vancomycin.

high molecular weight and complex structure, was not able to sufficiently accumulate inside macrophage cells and clear intracellular MRSA infections, as expected.^{52,53} In agreement with the previous reports, vancomycin did not reduce the presence of MRSA inside infected J774 cells after 24 h of treatment. These results collectively indicated that this class of newly discovered phenylpyrazoles could pass into the infected macrophage cells, even at a lower concentration ($2 \times \text{MIC}$), and significantly reduce the burden of MRSA inside them.

After confirming the ability of phenylpyrazoles to clear the intracellular Gram-positive bacterial cells, we moved to determine whether the potential therapeutic application of phenylpyrazoles could be expanded beyond merely inhibiting Gram-negative bacteria. Their ability to clear the intracellular Gram-negative pathogen *S. enteritidis* harbored inside infected macrophages was evaluated (Figure 4). In this case, two of the

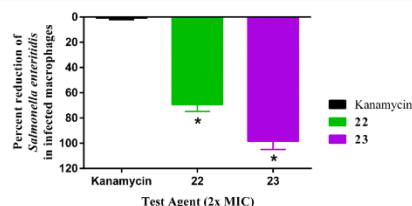


Figure 4. Examination of the activity of compounds 22 and 23 on the clearance of intracellular *Salmonella enteritidis* ATCC 14028 present in murine macrophage (J774) cells. Data are presented as percent reduction of *S. enteritidis* colony forming units inside infected murine macrophage cells after the treatment with $2 \times \text{MIC}$ of compounds 22 and 23 (tested in triplicates) for 24 h. Data were analyzed via an unpaired Student's *t*-test ($P < 0.05$). Asterisks (*) represent a significant difference between the treatment of J774 cells with compounds 23 and 35 in comparison to vancomycin.

most promising derivatives against Gram-negative bacteria were chosen: compounds, 22 and 23. The phenylpyrazole 23 was capable of significantly reducing the intracellular *S. enteritidis* by $1.76 \log_{10}$, which is equivalent to 98% reduction, at $2 \times \text{MIC}$. Additionally, compound 22 significantly reduced the burden of intracellular *S. enteritidis* by $0.57 \log_{10}$, which is equivalent to 70% reduction, at $2 \times \text{MIC}$. These results collectively indicated that phenylpyrazoles could gain entry and

accumulate inside infected macrophages and significantly reduce the burden of intracellular *Salmonella*. On the other hand, kanamycin, an efficient anti-Gram negative antibiotic, at $2 \times \text{MIC}$ ($16 \mu\text{g/mL}$), as reported earlier,⁵⁴ was unable to sufficiently accumulate inside infected macrophages and thus was unable to clear the intracellular *Salmonella* (Figure 4).

Finally, to investigate whether the pathogenic microorganisms are able to develop rapid resistance to this newly introduced class of antibacterial agents, a multi-step resistance study was conducted for the three most promising compounds 22, 23, and 35 against MRSA USA123 (MRSA400). As shown in Figure 5, the MIC values of 23 remained stable along the

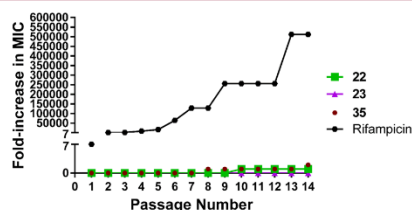


Figure 5. Multi-step resistance study of compounds 22, 23, 35 and rifampicin against methicillin-resistant *S. aureus* USA400. Bacteria were serially passaged over a 14 days period, and the broth microdilution assay was used to determine the MIC of each compound/control antibiotic against MRSA USA400 after each successive passage. A four-fold increase in MIC would be indicative of bacterial resistance to the test agent.

14-passages without any increase. On the other hand, the MIC values for 22 increased by one-fold after the 10th passage and remained stable after that. The MIC values of the imidazoline 35 increased by 2-folds after 14-passages (one-fold increase in the eighth passage and another one-fold increase on the 14th passage). In contrast, MRSA USA400 developed resistance

rapidly to the antibiotic rifampicin, in agreement with previous reports.^{55,56} The MIC of the antibiotic increased 7-folds after only one passage and continued to increase rapidly ($>500\,000$ -fold increase in MIC by the 13th passage). The results indicate that MRSA was unable to develop a rapid resistance to any of the tested phenylpyrazoles and can rapidly develop resistance to rifampicin (Figure 5).

Pharmacokinetic Evaluation. The metabolic stability and rate of clearance of the phenylpyrazoles compounds 23 and 35 were initially investigated using in vitro assays (Table 6). The most promising derivative in this series, compound 23, showed a high degree of metabolic stability with an intrinsic half-life ($t_{1/2}$) of more than 3 h. This value was increased in the absence of NADPH, the cofactor of CYP-450, indicating that phenylpyrazole 23 is mainly metabolized by this class of metabolic oxidases. The imidazoline analog compound 35 showed similar high metabolic stability in the presence of NADPH, but the half-life value was significantly increased when measured in the absence of NADPH, suggesting compound 35 is an exclusive substrate for CYP-450.

Furthermore, unlike some reported phenylthiazoles that were found to be substrates for the efflux system, the Caco-2 assay for compounds 23 and 35 indicated that our newly discovered phenylpyrazoles are not substrates for P-glycoprotein (P-gp) efflux pumps. Unfortunately, both tested phenylpyrazoles possess very limited permeation properties (Table 7), which means that they are not suitable for the oral route. Finally, the metabolic stability of this novel class of antibacterial compounds was further confirmed by measuring the biological half-life in rats, in which compound 23 showed a biological half-life of 10.5 h (Figure 6). This value roughly suggests a once-daily dosing regimen after parenteral administration.

Mechanistic study. The potent broad-spectrum activity of the synthesized compounds led us to investigate the compound's mechanism of action. Previously phenylthiazoles

Table 6. In Vitro Preliminary Pharmacokinetic Parameters; Half-Life and Intrinsic Clearance ($t_{1/2}$ and Cl_{int}) of Compounds 23 and 35

compound	incubation time (min)	% compound remaining			half-life (min)			Cl_{int} ($\mu\text{L}/\text{min}/\text{mg}$)
		1st	2nd	mean	1st	2nd	mean	
23	0	100.0	100.0	100	188.8	198.1	194	35.8
23	15	92.1	93.6	93				
23	30	95.0	88.9	92				
23	45	80.2	82.1	81				
23	60	81.4	82.1	82				
23 (no NADPH)	0	100.0	100.0	100	418.4	591.7	505	<115.5
23 (no NADPH)	15	98.9	99.1	99				
23 (no NADPH)	30	98.0	99.0	99				
23 (no NADPH)	45	{85.3}	98.9	99				
23 (no NADPH)	60	90.7	91.6	91				
35	0	100.0	100.0	100	160.9	144.8	153	42.3
35	15	103.6	93.7	99				
35	30	99.5	84.3	92				
35	45	83.7	75.2	79				
35	60	80.5	78.0	79				
35 (no NADPH)	0	100.0	100.0	100	1303.4	1129.7	1217	<115.5
35 (no NADPH)	15	97.4	99.1	98				
35 (no NADPH)	30	96.5	96.6	97				
35 (no NADPH)	45	96.5	{100.0}	97				
35 (no NADPH)	60	96.5	96.6	97				

Table 7. Evaluation of apparent Permeability of Tested Compounds, Ranitidine, and Warfarin, via the Caco-2 Permeability assay

compound tested	mean $A \rightarrow B P_{app}$ ^a ($\times 10^{-6}$ cm/s)	mean $B \rightarrow A P_{app}$ ^b ($\times 10^{-6}$ cm/s)	efflux ratio ^c	substrate of P-gp ^d
lead compound (1)	0.0 ^e	1.2	>2	yes
23	0.01	0.01	1.0	no
35	0.1	0.15	1.5	no
ranitidine	0.2	1.7	8.5	yes
warfarin	27.6	11.1	0.4	no

^aMean $A \rightarrow B P_{app}$ = mean apparent permeability of the test compound from the apical to basolateral surface. ^bMean $B \rightarrow A P_{app}$ = mean apparent permeability of the test compound from basolateral to the apical surface. ^cEfflux ratio = $P_{app}(B \rightarrow A)/P_{app}(A \rightarrow B)$. ^dP-gp = P-glycoprotein. ^eCompound not detected in the receiver compartment (the peak below the limit of detection); permeability may be underestimated.

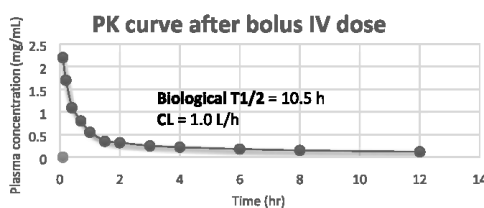


Figure 6. PK curve of compound 23 after 5 mg/kg IV bolus dose in rats (3 animals).

exhibited their antimicrobial activity by inhibiting UppP and UppS, two essential enzymes involved in bacterial cell wall synthesis.^{8,12} To investigate whether the newly developed phenylpyrazole-based antibacterial compounds possess the same mode of inhibition, the inhibitory activity of the most promising derivatives were all assessed against both UppP and UppS. Surprisingly, the newly developed phenylpyrazole-based antibacterial compounds did not inhibit these two enzymes, indicating a likely new mechanism of action (Figure S6).

CONCLUSIONS

The replacement of the central ring of phenylthiazole antibacterial compounds has an immense impact on their physicochemical properties, in which they were able to cross and accumulate inside the Gram-negative bacterial cells. Such chemical modification broadened the antibacterial spectrum of this novel class of antibiotics to include many MDR-Gram-negative pathogens. Studying the SAR at the phenyl position-4, by designing 24 congeners, revealed eight derivatives with promising antimicrobial profiles. Among them, congeners with cyclopentyl (compound 23) and phenyl (compound 22) side chains showed highly comparable results with reference drugs in terms of MIC values, vancomycin in the case of Gram-positive bacteria, and gentamicin in the case of Gram-negative bacteria. In addition to outperforming vancomycin against several MDR-Gram-positive strains, compounds 22 and 23 showed an efficient antibacterial effect against several MDR-Gram-negative clinical isolates such as *A. baumannii*, *E. coli*, *K. pneumoniae*, *Enterobacter cloacae*, *S. enteritidis*, and *Shigella flexneri*. Remarkably, compounds 22 and 23 were superior to imipenem and meropenem against the urgent-threat carbape-

nem-resistant bacteria, making them excellent starting compounds for developing an effective treatment for these superbugs. Moreover, the antimicrobial performance of this set of novel antibacterial agents was not only limited to lower microgram/mL MIC values but also extended to include the ability to clear intracellular-hidden pathogens and inhibit staphylococcal biofilm formation. Finally, the overall evaluation of the in vitro and in vivo PK profiles of phenylpyrazole antibiotics, represented by the most potent derivative compound 23, revealed that they are metabolically stable with a long biological half-life, but like many other clinically used antibiotics, they have to be administered parenterally. Further experiments to identify the cellular target of the newly developed phenylpyrazoles are underway.

EXPERIMENTAL SECTION

General. All biologically tested compounds are with purity of 95% or more. ¹H NMR spectra were run at 400 MHz and ¹³C NMR spectra were determined at 100 MHz in deuterated chloroform (CDCl₃), or dimethyl sulfoxide (DMSO-*d*₆) on a Varian Mercury VX-400 NMR spectrometer. Chemical shifts are given in parts per million (ppm) on the delta (δ) scale. Chemical shifts were calibrated relative to those of the solvents. Flash chromatography was performed on 230–400 mesh silica. The progress of reactions was monitored with Merck silica gel IB2-F plates (0.25 mm thickness). The infrared spectra were recorded in potassium bromide disks on pye Unicam SP 3300 and Shimadzu FT IR 8101 PC infrared spectrophotometer. Mass spectra were recorded at 70 eV. High-resolution mass spectra for all ionization techniques were obtained from a FinniganMAT XL95. Melting points were determined using capillary tubes with a Stuart SMP30 apparatus and are uncorrected. All yields reported refer to isolated yields. (4-Iodophenyl)hydrazine (2)⁵⁷ and 1-(1-(4-iodophenyl)-5-methyl-1H-pyrazol-4-yl)ethanone (3)⁵⁸ were prepared as reported.

1-(5-Methyl-1-(4-(substituted alkynyl)phenyl)-1H-pyrazol-4-yl)ethanone 4–15. *General Procedure.* To DME (5 mL) in a 75 mL sealed tube, compound 3 (340 mg, 1.45 mmol), appropriate alkynes (2.91 mmol), triethylamine (1 mL) were dissolved. After the reaction mixture was purged with dry nitrogen gas for 20 min, dichloro-bis(triphenylphosphine)palladium(II) (52 mg, 0.072 mmol) and copper(I) iodide (28 mg, 0.145 mmol) were added. The sealed tube was then placed in an oil bath and stirred at 65 °C for 16 h. After cooling to room temperature, the reaction mixture was passed through Celite, followed by ethyl acetate (2 × 50 mL) and dried over anhydrous magnesium sulphate. The organic materials were then concentrated under reduced pressure. The crude materials were purified via silica gel flash column chromatography using hexane–ethyl acetate (7:3). Yields, physical properties, and spectral data of isolated purified products are listed below.

1-(5-Methyl-1-(4-(pent-1-yn-1-yl)phenyl)-1H-pyrazol-4-yl)ethanone (4). It is obtained as yellow oil (320 mg, 80%). ¹H NMR (CDCl₃): δ 7.99 (s, 1H), 7.52 (d, *J* = 8.4 Hz, 2H), 7.35 (d, *J* = 8.4 Hz, 2H), 2.57 (s, 3H), 2.43 (s, 3H), 2.40 (t, *J* = 6.9 Hz, 2H), 1.68 (m, *J* = 7.2 Hz, 2H), 1.08 (t, *J* = 7.2 Hz, 3H); ¹³C NMR (CDCl₃): δ 193.33, 142.88, 141.98, 137.38, 132.32, 125.15, 124.82, 121.20, 92.27, 79.64, 28.62, 22.03, 21.36, 13.48, 12.34; MS (*m/z*): 266 (*M*⁺, 19.7%).

1-(1-(4-(Hex-1-yn-1-yl)phenyl)-5-methyl-1H-pyrazol-4-yl)ethanone (5). It is obtained as brown oil (420 mg, 97%). ¹H NMR (CDCl₃): δ 7.98 (s, 1H), 7.51 (d, *J* = 8.4 Hz, 2H), 7.31 (d, *J* = 8.4 Hz, 2H), 2.56 (s, 3H), 2.47 (s, 3H), 2.42 (t, *J* = 6.9 Hz, 2H), 1.62–1.44 (m, 4H), 0.95 (t, *J* = 6.9 Hz, 3H); ¹³C NMR (CDCl₃): δ 193.27, 142.83, 141.94, 137.34, 132.28, 125.11, 124.79, 121.16, 92.40, 79.45, 30.61, 28.58, 21.93, 19.04, 13.52, 12.30; MS (*m/z*): 280 (*M*⁺, 14.0%).

1-(1-(4-(Hept-1-yn-1-yl)phenyl)-5-methyl-1H-pyrazol-4-yl)ethanone (6). It is obtained as brown oil (420 mg, 93%). ¹H NMR (CDCl₃): δ 7.81 (s, 1H), 7.34 (d, *J* = 7.5 Hz, 2H), 7.20 (d, *J* = 6.9 Hz, 2H), 2.39 (s, 3H), 2.28 (s, 3H), 2.27 (t, *J* = 6.3 Hz, 2H), 1.47–1.19

(m, 6H), 0.79 (t, $J = 6.9$ Hz, 3H); ^{13}C NMR (CDCl_3): δ 192.53, 142.24, 141.48, 136.97, 131.77, 124.60, 124.22, 120.71, 91.91, 79.16, 30.62, 28.06, 27.84, 21.71, 18.88, 13.48, 11.81; MS (m/z): 294 (M^+ , 9.23%).

1-(5-Methyl-1-(4-(oct-1-yn-1-yl)phenyl)-1H-pyrazol-4-yl)-ethanone (7). It is obtained as gray oil (350 mg, 74%). ^1H NMR (CDCl_3): δ 7.99 (s, 1H), 7.51 (d, $J = 8.1$ Hz, 2H), 7.34 (d, $J = 8.7$ Hz, 2H), 2.58 (s, 3H), 2.56 (s, 3H), 2.47 (t, $J = 7.2$ Hz, 2H), 1.64–1.34 (m, 8H), 0.92 (t, $J = 6.9$ Hz, 3H); ^{13}C NMR (CDCl_3): δ 193.30, 142.00, 137.32, 132.29, 125.14, 124.83, 92.50, 79.48, 31.27, 28.60, 28.53, 22.47, 19.38, 13.97, 12.33; MS (m/z): 308 (M^+ , 100%).

1-(1-(4-(Cyclopropylethynyl)phenyl)-5-methyl-1H-pyrazol-4-yl)-ethanone (8). It is obtained as brown oil (340 mg, 85%). ^1H NMR (CDCl_3): δ 7.98 (s, 1H), 7.48 (d, $J = 8.4$ Hz, 2H), 7.32 (d, $J = 8.4$ Hz, 2H), 2.55 (s, 3H), 2.46 (s, 3H), 1.45 (p, $J = 7.5$ Hz, 1H), 0.90–0.80 (m, 4H); ^{13}C NMR (CDCl_3): δ 193.29, 142.83, 141.96, 137.29, 132.31, 125.10, 124.64, 121.17, 95.44, 74.67, 28.59, 12.31, 8.62, 0.10; MS (m/z): 264 (M^+ , 100%).

1-(1-(4-(Cyclohexylethynyl)phenyl)-5-methyl-1H-pyrazol-4-yl)-ethanone (9). It is obtained as yellow oil (370 mg, 78%). ^1H NMR (CDCl_3): δ 7.99 (s, 1H), 7.51 (d, $J = 8.1$ Hz, 2H), 7.33 (d, $J = 8.1$ Hz, 2H), 2.62 (m, 1H), 2.55 (s, 3H), 2.47 (s, 3H), 1.89–1.35 (m, 10H); ^{13}C NMR (CDCl_3): δ 193.27, 142.90, 142.00, 137.22, 132.32, 125.17, 124.89, 121.15, 96.43, 79.46, 32.50, 29.63, 28.60, 25.81, 24.80, 12.30; MS (m/z): 306 (M^+ , 100%).

1-(5-Methyl-1-(4-(phenylethynyl)phenyl)-1H-pyrazol-4-yl)-ethanone (10). It is obtained as brown solid (340 mg, 74%), mp 132–133 °C; ^1H NMR (CDCl_3): δ 8.02 (s, 1H), 7.67 (d, $J = 8.4$ Hz, 2H), 7.64 (d, $J = 8.4$ Hz, 2H), 7.57 (t, $J = 7.2$ Hz, 1H), 7.44 (d, $J = 8.4$ Hz, 2H), 7.37 (d, $J = 6$ Hz, 2H), 2.61 (s, 3H), 2.49 (s, 3H); ^{13}C NMR (CDCl_3): δ 193.33, 142.91, 142.10, 138.00, 132.38, 131.63, 128.61, 128.38, 125.25, 123.91, 122.71, 121.30, 91.08, 88.10, 28.65, 12.40; MS (m/z): 300 (M^+ , 37.47%).

1-(1-(4-(Cyclopentylethynyl)phenyl)-5-methyl-1H-pyrazol-4-yl)-ethanone (11). It is obtained as brown oil (350 mg, 77%). ^1H NMR ($\text{DMSO}-d_6$): δ 8.25 (s, 1H), 7.54 (d, $J = 6.8$ Hz, 2H), 7.51 (d, $J = 6.8$ Hz, 2H), 2.93 (p, $J = 6.8$ Hz, 1H), 2.51 (s, 3H), 2.44 (s, 3H), 2.01–1.63 (m, 8H); ^{13}C NMR (CDCl_3): δ 193.51, 142.94, 142.77, 137.87, 132.52, 125.80, 124.04, 121.35, 96.80, 79.64, 33.88, 30.49, 29.20, 25.11, 12.40; MS (m/z): 292 (M^+ , 25.36%).

1-(1-(4-(3,3-Dimethylbut-1-yn-1-yl)phenyl)-5-methyl-1H-pyrazol-4-yl)ethanone (12). It is obtained as yellowish oil (350 mg, 71%). ^1H NMR (CDCl_3): δ 7.98 (s, 1H), 7.51 (d, $J = 8.4$ Hz, 2H), 7.33 (d, $J = 8.4$ Hz, 2H), 2.55 (s, 3H), 2.47 (s, 3H), 1.32 (s, 9H); ^{13}C NMR (CDCl_3): δ 193.29, 142.86, 141.94, 137.29, 132.31, 125.07, 124.80, 121.18, 100.42, 77.99, 30.85, 28.60, 27.94, 12.30; MS (m/z): 280 (M^+ , 32.29%).

1-(1-(4-(3-Hydroxy-3-methylbut-1-yn-1-yl)phenyl)-5-methyl-1H-pyrazol-4-yl)ethanone (13). It is obtained as yellow oil (335 mg, 78%). ^1H NMR (CDCl_3): δ 8.00 (s, 1H), 7.53 (d, $J = 8.7$ Hz, 2H), 7.37 (d, $J = 8.4$ Hz, 2H), 2.56 (s, 3H), 2.47 (s, 3H), 1.62 (s, 6H), 1.50 (br s, 1H); ^{13}C NMR (CDCl_3): δ 193.42, 142.09, 142.19, 137.94, 132.44, 125.18, 123.48, 121.48, 95.70, 80.80, 65.44, 31.36, 28.62, 12.36; MS (m/z): 282 (M^+ , 29.17%).

1-(1-(4-(1-Hydroxycyclohexylethynyl)phenyl)-5-methyl-1H-pyrazol-4-yl)ethanone (14). It is obtained as brown oil (320 mg, 65%). ^1H NMR (CDCl_3): δ 8.00 (s, 1H), 7.54 (d, $J = 8.4$ Hz, 2H), 7.37 (d, $J = 8.1$ Hz, 2H), 2.60 (br s, 1H), 2.56 (s, 3H), 2.47 (s, 3H), 2.00–1.24 (m, 10H); ^{13}C NMR (CDCl_3): δ 193.41, 142.09, 137.85, 134.19, 132.44, 128.38, 125.80, 123.63, 94.83, 82.96, 68.91, 39.87, 28.61, 25.11, 23.28, 12.34; MS (m/z): 322 (M^+ , 100%).

1-(1-(4-(3-(Diethylamino)prop-1-yn-1-yl)phenyl)-5-methyl-1H-pyrazol-4-yl)ethanone (15). It is obtained as reddish oil (280 mg, 59%). ^1H NMR (CDCl_3): δ 7.86 (s, 1H), 7.42 (d, $J = 7.8$ Hz, 2H), 7.24 (d, $J = 8.1$ Hz, 2H), 3.53 (s, 2H), 3.35 (m, 4H), 2.43 (s, 3H), 2.32 (s, 3H), 0.98 (m, 6H); ^{13}C NMR (CDCl_3): δ 192.79, 142.43, 141.65, 137.41, 132.08, 124.77, 123.50, 120.86, 85.33, 83.53, 46.85, 41.00, 28.20, 12.14, 11.93; MS (m/z): 311 (M^+ , 8.77%).

2-(1-(5-Methyl-1-(4-(substituted)phenyl)-1H-pyrazol-4-yl)-ethylidene)hydrazine-Carboximidamide (16–27). General Procedure. Pyrazole derivatives 4–15 (0.63 mmol) were dissolved in

absolute ethanol (10 mL), aminoguanidine hydrochloride (70 mg, 0.63 mmol), and hydrochloric acid (0.1 mL). The reaction mixture was heated at reflux for 12 h. The solvent was concentrated under reduced pressure, poured onto crushed ice, and neutralized with sodium carbonate (to pH 7–8). The solid precipitate was collected by filtration and washed with a copious amount of water. Crystallization from absolute methanol afforded the desired products, as solids. Yields, physical properties, and spectral data of isolated purified products are listed below.

2-(1-(5-Methyl-1-(4-(pent-1-yn-1-yl)phenyl)-1H-pyrazol-4-yl)-ethylidene)hydrazine-Carboximidamide (16). It is obtained as yellowish solid (85 mg, 59%) mp 173–174 °C. ^1H NMR ($\text{DMSO}-d_6$): δ 7.63 (s, 1H), 7.29 (d, $J = 8.6$ Hz, 2H), 7.23 (d, $J = 8.4$ Hz, 2H), 5.54 (br s, 4H), 2.25 (s, 3H), 2.20 (t, $J = 6.4$ Hz, 2H), 1.98 (s, 3H), 1.35 (m, 2H), 0.77 (t, $J = 7.3$ Hz, 3H); ^{13}C NMR ($\text{DMSO}-d_6$): δ 158.87, 145.56, 140.16, 138.92, 136.45, 132.57, 125.41, 123.03, 121.99, 92.25, 80.44, 22.08, 21.07, 16.24, 13.80, 12.39; HRMS (EI) m/z : 322.1901 M^+ , calcd for $\text{C}_{18}\text{H}_{22}\text{N}_6$ 322.1906. Calcd for $\text{C}_{18}\text{H}_{22}\text{N}_6$ (322): C, 67.06; H, 6.88; N, 26.07%. Found: C, 66.93; H, 7.06; N, 26.19%.

2-(1-(1-(4-(Hex-1-yn-1-yl)phenyl)-5-methyl-1H-pyrazol-4-yl)-ethylidene)hydrazine-Carboximidamide (17). It is obtained as brown solid (80 mg, 67%) mp 182–183 °C. ^1H NMR ($\text{DMSO}-d_6$): δ 11.30 (br s, 1H), 8.06 (s, 1H), 7.68 (br s, 3H), 7.54 (d, $J = 8.4$ Hz, 2H), 7.42 (d, $J = 8.1$ Hz, 2H), 2.47 (s, 3H), 2.43 (t, $J = 6.9$ Hz, 2H), 2.34 (s, 3H), 1.55–1.41 (m, 4H), 0.92 (t, $J = 7.2$ Hz, 3H); ^{13}C NMR ($\text{DMSO}-d_6$): δ 156.01, 148.60, 140.32, 137.98, 137.92, 132.03, 125.13, 123.04, 119.07, 92.08, 79.68, 30.14, 21.36, 18.30, 16.81, 13.40, 12.88; HRMS (EI) m/z : 336.2074 M^+ , calcd for $\text{C}_{19}\text{H}_{24}\text{N}_6$ 336.2062. Anal. Calcd for $\text{C}_{19}\text{H}_{24}\text{N}_6$ (336): C, 67.83; H, 7.19; N, 24.98%. Found: C, 67.62; H, 7.05; N, 24.69%.

2-(1-(1-(4-(Hept-1-yn-1-yl)phenyl)-5-methyl-1H-pyrazol-4-yl)-ethylidene)hydrazine-Carboximidamide (18). It is obtained as yellow solid (120 mg, 67%) mp 198–200 °C. ^1H NMR ($\text{DMSO}-d_6$): δ 11.20 (br s, 1H), 8.06 (s, 1H), 7.62 (br s, 3H), 7.55 (d, $J = 8.4$ Hz, 2H), 7.46 (d, $J = 8.4$ Hz, 2H), 2.49 (s, 3H), 2.46 (t, 2H), 2.34 (s, 3H), 1.56 (p, $J = 7.2$ Hz, 2H), 1.36 (p, $J = 6.6$ Hz, 4H), 0.91 (t, $J = 7.2$ Hz, 3H); ^{13}C NMR ($\text{DMSO}-d_6$): δ 155.91, 148.68, 140.22, 137.92, 134.19, 131.94, 127.63, 125.09, 118.98, 92.09, 79.61, 30.39, 27.65, 21.48, 18.51, 16.65, 13.69, 12.76; HRMS (EI) m/z : 350.2233 M^+ , calcd for $\text{C}_{19}\text{H}_{24}\text{N}_6$ 350.2219. Anal. Calcd for $\text{C}_{20}\text{H}_{26}\text{N}_6$ (350): C, 68.54; H, 7.48; N, 23.98%. Found: C, 68.40; H, 7.67; N, 24.26%.

2-(1-(5-Methyl-1-(4-(oct-1-yn-1-yl)phenyl)-1H-pyrazol-4-yl)-ethylidene)hydrazine-Carboximidamide (19). It is obtained as yellow solid (93 mg, 58%) mp 280–281 °C. ^1H NMR ($\text{DMSO}-d_6$): δ 7.82 (s, 1H), 7.52 (d, $J = 8.7$ Hz, 2H), 7.48 (d, $J = 8.1$ Hz, 2H), 5.58 (br s, 2H), 5.29 (br s, 2H), 2.49 (s, 3H), 2.46 (t, $J = 6.9$ Hz, 2H), 2.20 (s, 3H), 1.58 (p, $J = 6.9$ Hz, 2H), 1.45 (p, $J = 6.3$ Hz, 2H), 1.40 (m, 4H), 0.90 (t, $J = 6.9$ Hz, 3H); ^{13}C NMR ($\text{DMSO}-d_6$): δ 158.59, 144.44, 139.42, 138.46, 135.52, 131.80, 124.78, 122.43, 121.80, 91.76, 79.73, 30.60, 27.95, 27.81, 21.84, 18.54, 15.46, 13.72, 13.06; HRMS (EI) m/z : 364.2382 M^+ , calcd for $\text{C}_{20}\text{H}_{28}\text{N}_6$ 364.2375. Anal. Calcd for $\text{C}_{20}\text{H}_{28}\text{N}_6$ (364): C, 69.20; H, 7.74; N, 23.06%. Found: C, 69.15; H, 7.42; N, 23.18%.

2-(1-(1-(4-(Cyclopropylethynyl)phenyl)-5-methyl-1H-pyrazol-4-yl)ethylidene)hydrazine-Carboximidamide (20). It is obtained as brown solid (90 mg, 57%) mp 203–204 °C. ^1H NMR ($\text{DMSO}-d_6$): δ 11.44 (br s, 1H), 7.97 (s, 1H), 7.87 (d, $J = 8.7$ Hz, 2H), 7.69 (br s, 3H), 7.47 (d, $J = 8.1$ Hz, 2H), 2.60 (s, 3H), 2.43 (s, 3H), 1.64 (p, $J = 6.8$ Hz, 1H), 1.49 (q, $J = 8.4$ Hz, 2H), 1.39 (q, $J = 5.2$ Hz, 2H); ^{13}C NMR (CDCl_3): δ 168.77, 158.87, 144.05, 142.05, 139.98, 132.01, 131.54, 127.96, 125.21, 96.58, 78.16, 18.45, 18.07, 13.41, 8.53; HRMS (EI) m/z : 320.1755 M^+ , calcd for $\text{C}_{18}\text{H}_{20}\text{N}_6$ 320.1749. Anal. Calcd for $\text{C}_{18}\text{H}_{20}\text{N}_6$ (320): C, 67.48; H, 6.29; N, 26.23%. Found: C, 67.34; H, 6.57; N, 26.46%.

2-(1-(1-(4-(Cyclohexylethynyl)phenyl)-5-methyl-1H-pyrazol-4-yl)ethylidene)hydrazine-Carboximidamide (21). It is obtained as faint yellow solid (100 mg, 71%) mp 245–246 °C. ^1H NMR ($\text{DMSO}-d_6$): δ 11.26 (br s, 1H), 8.01 (s, 1H), 7.78 (d, $J = 8.4$ Hz, 2H), 7.55 (br s, 1H), 7.54 (br s, 2H), 7.50 (d, $J = 8.4$ Hz, 2H), 2.52 (m, 1H),

2.49 (s, 3H), 2.34 (s, 3H), 1.34–1.23 (m, 10H); ^{13}C NMR (DMSO- d_6): δ 156.44, 149.16, 139.32, 137.19, 134.83, 129.77, 127.84, 125.65, 119.61, 96.33, 80.21, 32.64, 31.66, 29.23, 25.94, 17.32, 13.46; HRMS (EI) m/z : 362.2198 M^+ , calcd for $\text{C}_{21}\text{H}_{26}\text{N}_6$, 362.2219. Anal. Calcd for $\text{C}_{21}\text{H}_{26}\text{N}_6$ (362): C, 69.58; H, 7.23; N, 23.19%. Found: C, 69.43; H, 7.49; N, 23.40%.

2-(1-(5-Methyl-1-(4-(phenylethynyl)phenyl)-1H-pyrazol-4-yl)ethylidene)hydrazine-carboximidamide (22). It is obtained as brown solid (90 mg, 64%) mp 258–260 °C. ^1H NMR (DMSO- d_6): δ 8.01 (s, 1H), 7.72 (d, J = 8.4 Hz, 2H), 7.64 (d, J = 8.4 Hz, 2H), 7.58 (t, J = 7.2 Hz, 1H), 7.44 (d, J = 8.4 Hz, 2H), 7.37 (d, J = 6 Hz, 2H), 6.96 (br s, 4H), 2.52 (s, 3H), 2.31 (s, 3H); ^{13}C NMR (DMSO- d_6): δ 156.86, 147.19, 140.12, 138.82, 137.14, 132.12, 131.31, 128.86, 128.64, 125.05, 121.93, 121.67, 120.11, 90.32, 88.37, 16.28, 12.93; HRMS (EI) m/z : 356.1751 M^+ , calcd for $\text{C}_{21}\text{H}_{20}\text{N}_6$, 356.1749. Anal. Calcd for $\text{C}_{21}\text{H}_{20}\text{N}_6$ (356): C, 70.77; H, 5.66; N, 23.58%. Found: C, 70.59; H, 5.59; N, 23.81%.

2-(1-(1-(4-(Cyclopentylethynyl)phenyl)-5-methyl-1H-pyrazol-4-yl)ethylidene)hydrazine-carboximidamide (23). It is obtained as white solid (65 mg, 45%) mp 190–192 °C. ^1H NMR (DMSO- d_6): δ 11.16 (br s, 1H), 8.09 (s, 1H), 7.63 (br s, 3H), 7.55 (d, J = 8 Hz, 2H), 7.47 (d, J = 8 Hz, 2H), 2.92 (m, 1H), 2.49 (s, 3H), 2.34 (s, 3H), 2.29–1.62 (m, 8H); ^{13}C NMR (DMSO- d_6): δ 156.36, 149.19, 140.86, 138.47, 132.54, 131.63, 125.65, 123.57, 119.57, 96.64, 79.69, 33.49, 30.49, 25.10, 17.22, 13.40; HRMS (EI) m/z : 348.2072 M^+ , calcd for $\text{C}_{20}\text{H}_{24}\text{N}_6$, 348.2062. Anal. Calcd for $\text{C}_{20}\text{H}_{24}\text{N}_6$ (348): C, 68.94; H, 6.94; N, 24.12%. Found: C, 68.70; H, 7.19; N, 24.38%.

2-(1-(1-(4-(3,3-Dimethylbut-1-yn-1-yl)phenyl)-5-methyl-1H-pyrazol-4-yl)ethylidene)hydrazine-carboximidamide (24). It is obtained as white solid (38 mg, 30%) mp 190–191 °C. ^1H NMR (DMSO- d_6): δ 11.4 (s, 1H), 7.89 (s, 1H), 7.86 (d, J = 8.8 Hz, 2H), 7.69 (br s, 3H), 7.48 (d, J = 8.8 Hz, 2H), 2.61 (s, 3H), 2.43 (s, 3H), 1.30 (s, 9H); ^{13}C NMR (DMSO- d_6): δ 167.97, 163.96, 158.55, 152.41, 148.33, 131.91, 131.60, 125.87, 120.38, 101.02, 78.47, 40.33, 30.51, 27.57, 18.36, 17.99; HRMS (EI) m/z : 336.2045 M^+ , calcd for $\text{C}_{19}\text{H}_{24}\text{N}_6$, 336.2062. Anal. Calcd for $\text{C}_{19}\text{H}_{24}\text{N}_6$ (336): C, 67.83; H, 7.19; N, 24.98%. Found: C, 67.64; H, 7.35; N, 25.14%.

2-(1-(1-(4-(3-Hydroxy-3-methylbut-1-yn-1-yl)phenyl)-5-methyl-1H-pyrazol-4-yl)ethylidene)hydrazine-carboximidamide (25). It is obtained as white solid (85 mg, 65%) mp 270 °C. ^1H NMR (DMSO- d_6): δ 7.86 (s, 1H), 7.54 (d, J = 8.8 Hz, 2H), 7.50 (d, J = 8.8 Hz, 2H), 5.66 (br s, 2H), 5.40 (br s, 3H), 2.51 (s, 3H), 2.20 (s, 3H), 1.49 (s, 6H); ^{13}C NMR (DMSO- d_6): δ 159.23, 145.04, 140.15, 139.27, 136.20, 132.45, 125.39, 122.37, 122.22, 97.61, 80.14, 64.13, 32.00, 16.13, 13.75; HRMS (EI) m/z : 338.1858 M^+ , calcd for $\text{C}_{18}\text{H}_{22}\text{N}_6\text{O}$, 338.1855. Anal. Calcd for $\text{C}_{18}\text{H}_{22}\text{N}_6\text{O}$ (338): C, 63.89; H, 6.55; N, 24.83%. Found: C, 63.96; H, 6.23; N, 24.87%.

2-(1-(1-(4-(1-Hydroxycyclohexyl)ethyl)phenyl)-5-methyl-1H-pyrazol-4-yl)ethylidene)hydrazine-carboximidamide (26). It is obtained as yellow solid (77 mg, 66%) mp 250–252 °C. ^1H NMR (DMSO- d_6): δ 7.85 (s, 1H), 7.55 (d, J = 8.7 Hz, 2H), 7.51 (d, J = 8.7 Hz, 2H), 5.71 (br s, 1H), 5.60 (br s, 2H), 5.51 (br s, 2H), 2.50 (s, 3H), 2.20 (s, 3H), 1.89–1.22 (m, 10H). ^{13}C NMR (DMSO- d_6): δ 158.56, 144.54, 139.52, 138.73, 135.64, 131.85, 124.83, 121.82, 120.05, 96.01, 81.87, 66.94, 24.80, 22.69, 22.40, 15.54, 13.09; HRMS (EI) m/z : 378.2171 M^+ , calcd for $\text{C}_{21}\text{H}_{26}\text{N}_6\text{O}$, 378.2168. Anal. Calcd for $\text{C}_{21}\text{H}_{26}\text{N}_6\text{O}$ (378): C, 66.64; H, 6.92; N, 22.21%. Found: C, 66.83; H, 7.19; N, 22.39%.

2-(1-(1-(4-(3-(Diethylamino)prop-1-yn-1-yl)phenyl)-5-methyl-1H-pyrazol-4-yl)ethylidene)hydrazine-carboximidamide (27). It is obtained as light brown solid (75 mg, 62%) mp 155–156 °C. ^1H NMR (DMSO- d_6): δ 7.84 (s, 1H), 7.56 (d, J = 8.4 Hz, 2H), 7.48 (d, J = 8.7 Hz, 2H), 5.68 (br s, 2H), 5.44 (br s, 2H), 3.61 (s, 2H), 2.57 (q, J = 7.2 Hz, 4H), 2.49 (s, 3H), 2.20 (s, 3H), 1.04 (t, J = 6.9 Hz, 6H); ^{13}C NMR (DMSO- d_6): δ 158.74, 144.34, 139.57, 138.77, 135.61, 132.05, 124.84, 121.88, 121.74, 86.46, 83.73, 46.62, 40.91, 15.57, 13.18, 12.47; HRMS (EI) m/z : 365.2342 M^+ , calcd for $\text{C}_{20}\text{H}_{27}\text{N}_7$, 365.2328. Anal. Calcd for $\text{C}_{20}\text{H}_{27}\text{N}_7$ (365): C, 65.73; H, 7.45; N, 26.83%. Found: C, 65.62; H, 7.69; N, 26.89%.

2-(4-(Substituted alkynyl)phenyl)-5-(1-(2-(4,5-dihydro-1H-imidazole-2-yl)hydrazono)ethyl)-4-methyl Pyrazole (28–39).

General Procedure. Pyrazole derivatives 4–16 (0.63 mmol) were dissolved in absolute ethanol (10 mL), and then, 2-hydrazinyl-4,5-dihydro-1H-imidazole hydrobromide (11.5 mg, 0.63 mmol) and hydrochloric acid (0.1 mL) were added. The reaction mixture was heated at reflux for 16 h. The solvent was concentrated under reduced pressure, poured onto crushed ice, and neutralized with sodium carbonate (to pH 7–8). The solid precipitate was collected by filtration and washed with a copious amount of water. Crystallization from absolute methanol afforded the desired products as solids. Yields, physical properties, and spectral data of isolated purified products are listed below.

4-(1-(2-(4,5-Dihydro-1H-imidazol-2-yl)hydrazineylidene)ethyl)-5-methyl-1-(4-(pent-1-yn-1-yl)phenyl)-1H-pyrazole (28). It is obtained as yellow solid (63 mg, 40%) mp 200–202 °C. ^1H NMR (DMSO- d_6): δ 7.83 (s, 1H), 7.54 (d, J = 6.3 Hz, 2H), 7.46 (d, J = 6.6 Hz, 2H), 6.22 (br s, 1H), 6.03 (br s, 1H), 3.35 (s, 4H), 2.52 (s, 3H), 2.43 (t, J = 7.2 Hz, 2H), 2.18 (s, 3H), 1.62 (m, J = 6.9 Hz, 2H), 1.04 (t, J = 7.2 Hz, 3H); ^{13}C NMR (DMSO- d_6): δ 163.56, 145.53, 139.48, 138.49, 135.80, 131.85, 124.77, 122.40, 121.62, 91.60, 79.87, 42.28, 21.48, 20.50, 15.33, 13.17, 13.06; HRMS (EI) m/z : 348.2069 M^+ , calcd for $\text{C}_{20}\text{H}_{24}\text{N}_6$, 348.2062. Anal. Calcd for $\text{C}_{20}\text{H}_{24}\text{N}_6$ (348): C, 68.94; H, 6.94; N, 24.12%. Found: C, 68.16; H, 6.87; N, 25.49%.

4-(1-(2-(4,5-dihydro-1H-imidazol-2-yl)hydrazineylidene)ethyl)-1-(4-(hex-1-yn-1-yl)phenyl)-5-methyl-1H-pyrazole (29). It is obtained as brown solid (74 mg, 50%) mp 195–197 °C. ^1H NMR (DMSO- d_6): δ 11.81 (br s, 1H), 8.24 (br s, 1H), 8.04 (s, 1H), 7.52 (d, J = 6.9 Hz, 2H), 7.42 (d, J = 6.9 Hz, 2H), 3.70 (s, 4H), 2.49 (s, 3H), 2.42 (t, J = 7.2 Hz, 2H), 2.34 (s, 3H), 1.51–1.40 (m, 4H), 0.89 (t, J = 5.7 Hz, 3H); ^{13}C NMR (DMSO- d_6): δ 158.44, 150.21, 140.39, 138.14, 137.99, 132.04, 125.17, 123.11, 118.79, 92.12, 79.70, 42.89, 30.17, 21.39, 18.33, 16.89, 13.37; HRMS (EI) m/z : 362.2209 M^+ , calcd for $\text{C}_{21}\text{H}_{26}\text{N}_6$, 362.2219. Anal. Calcd for $\text{C}_{21}\text{H}_{26}\text{N}_6$ (362): C, 69.23; H, 7.23; N, 23.19%. Found: C, 69.79; H, 7.50; N, 23.06%.

4-(1-(2-(4,5-Dihydro-1H-imidazol-2-yl)hydrazono)ethyl)-1-(4-(hept-1-yn-1-yl)phenyl)-5-methyl-1H-pyrazole (30). It is obtained as yellow solid (96 mg, 51%) mp 189–191 °C. ^1H NMR (DMSO- d_6): δ 7.83 (s, 1H), 7.53 (d, J = 8.4 Hz, 2H), 7.47 (d, J = 8.7 Hz, 2H), 6.23 (br s, 1H), 6.04 (br s, 1H), 3.37 (s, 4H), 2.51 (s, 3H), 2.48 (t, J = 6.3 Hz, 2H), 2.18 (s, 3H), 1.59–1.33 (m, 6H), 0.92 (t, J = 7.2 Hz, 3H); ^{13}C NMR (DMSO- d_6): δ 163.62, 145.82, 139.47, 138.47, 135.80, 131.82, 124.76, 122.41, 121.61, 91.77, 79.71, 42.28, 30.39, 27.67, 21.47, 18.50, 15.32, 13.68, 13.05; HRMS (EI) m/z : 376.2382 M^+ , calcd for $\text{C}_{22}\text{H}_{28}\text{N}_6$, 376.2375. Anal. Calcd for $\text{C}_{22}\text{H}_{28}\text{N}_6$ (376): C, 70.18; H, 7.50; N, 22.32%. Found: C, 69.94; H, 7.65; N, 22.43%.

4-(1-(2-(4,5-Dihydro-1H-imidazol-2-yl)hydrazono)ethyl)-5-methyl-1-(4-(oct-1-yn-1-yl)phenyl)-1H-pyrazole (31). It is obtained as yellow solid (74 mg, 49%) mp 296–298 °C. ^1H NMR (DMSO- d_6): δ 7.83 (s, 1H), 7.52 (d, J = 8.7 Hz, 2H), 7.46 (d, J = 8.7 Hz, 2H), 6.22 (br s, 1H), 6.06 (br s, 1H), 3.34 (s, 4H), 2.49 (s, 3H), 2.46 (t, J = 6.9 Hz, 2H), 2.16 (s, 3H), 1.57–1.31 (m, 8H), 0.89 (t, J = 6.9 Hz, 3H); ^{13}C NMR (DMSO- d_6): δ 163.59, 145.49, 139.49, 138.42, 135.79, 131.83, 125.26, 124.77, 122.35, 91.77, 79.71, 42.27, 30.59, 27.92, 27.81, 21.86, 18.51, 15.34, 13.76, 13.09; HRMS (EI) m/z : 390.2547 M^+ , calcd for $\text{C}_{23}\text{H}_{30}\text{N}_6$, 390.2532. Anal. Calcd for $\text{C}_{23}\text{H}_{30}\text{N}_6$ (390): C, 70.74; H, 7.74; N, 21.52%. Found: C, 70.59; H, 8.09; N, 21.30%.

1-(4-(Cyclopropylethynyl)phenyl)-4-(1-(2-(4,5-dihydro-1H-imidazol-2-yl)hydrazono)ethyl)-5-methyl-1H-pyrazole (32). It is obtained as yellow-white solid (78 mg, 46%) mp 210–212 °C. ^1H NMR (DMSO- d_6): δ 11.71 (s, 1H), 8.14 (s, 1H), 7.87 (d, J = 7.8 Hz, 2H), 7.49 (d, J = 8.7 Hz, 2H), 5.41 (br s, 1H), 3.71 (s, 4H), 2.62 (s, 3H), 2.43 (s, 3H), 1.44 (p, J = 6.8 Hz, 1H), 0.98 (q, J = 8.4 Hz, 2H), 0.90 (q, J = 5.2 Hz, 2H); ^{13}C NMR (DMSO- d_6): δ 164.07, 158.42, 148.47, 139.59, 132.04, 131.04, 125.95, 120.61, 120.57, 96.62, 85.02, 42.88, 18.44, 17.88, 13.41, 8.50; HRMS (EI) m/z : 346.1901 M^+ , calcd for $\text{C}_{20}\text{H}_{22}\text{N}_6$, 346.1906. Anal. Calcd for $\text{C}_{20}\text{H}_{22}\text{N}_6$ (346): C, 69.34; H, 6.40; N, 24.26%. Found: C, 69.17; H, 6.12; N, 24.39%.

1-(4-(Cyclohexylethynyl)phenyl)-4-(1-(2-(4,5-dihydro-1H-imidazol-2-yl)hydrazono)ethyl)-5-methyl-1H-pyrazole (33). It is ob-

tained as yellow-white oil (87 mg, 58%) mp 215–216 °C. ¹H NMR (DMSO-*d*₆): δ 11.69 (br s, 1H), 8.25 (br s, 1H), 8.19 (s, 1H), 7.54 (d, *J* = 8.4 Hz, 2H), 7.46 (d, *J* = 8.4 Hz, 2H), 3.71 (s, 4H), 2.66 (m, 1H), 2.49 (s, 3H), 2.33 (s, 3H), 1.84–1.34 (m, 10H); ¹³C NMR (DMSO-*d*₆): δ 158.83, 150.79, 140.92, 138.63, 138.44, 132.59, 125.70, 123.54, 119.28, 96.37, 80.20, 43.36, 32.54, 29.23, 25.80, 24.69, 17.26, 13.47; HRMS (EI) *m/z*: 388.2363 M⁺, calcd for C₂₃H₂₈N₆, 388.2375. Anal. Calcd for C₂₃H₂₈N₆ (388): C, 71.10; H, 7.26; N, 23.63%. Found: C, 70.98; H, 7.45; N, 23.68%.

4-(1-(2-(4,5-Dihydro-1H-imidazol-2-yl)hydrazono)ethyl)-5-methyl-1-(4-(phenylethynyl)phenyl)-1H-pyrazole (34). It is obtained as gray solid (70 mg, 42%) mp 224–226 °C. ¹H NMR (DMSO-*d*₆): δ 7.97 (s, 1H), 7.72 (d, *J* = 8.1 Hz, 2H), 7.69 (d, *J* = 8.1 Hz, 2H), 7.57 (t, *J* = 7.2 Hz, 1H), 7.54 (d, *J* = 8.4 Hz, 2H), 7.49 (d, *J* = 6 Hz, 2H), 6.24 (br s, 2H), 3.54 (s, 4H), 2.54 (s, 3H), 2.25 (s, 3H); ¹³C NMR (DMSO-*d*₆): δ 160.99, 147.77, 140.05, 138.89, 136.98, 132.10, 131.29, 128.83, 128.63, 124.99, 121.93, 121.56, 120.33, 90.28, 88.39, 42.35, 16.09, 13.06; HRMS (EI) *m/z*: 382.1896 M⁺, calcd for C₂₃H₂₂N₆, 382.1906. Anal. Calcd for C₂₃H₂₂N₆ (382): C, 72.23; H, 5.80; N, 21.97%. Found: C, 71.98; H, 6.05; N, 21.93%.

1-(4-(Cyclopentylethynyl)phenyl)-4-(1-(2-(4,5-dihydro-1H-imidazol-2-yl)hydrazono)ethyl)-5-methyl-1H-pyrazole (35). It is obtained as brown solid (70 mg, 46%) mp 190–192 °C. ¹H NMR (DMSO-*d*₆): δ 11.79 (br s, 1H), 8.26 (br s, 1H), 8.07 (s, 1H), 7.55 (d, *J* = 9.2 Hz, 2H), 7.46 (d, *J* = 6.4 Hz, 2H), 3.71 (s, 4H), 2.66 (p, *J* = 6 Hz, 1H), 2.44 (s, 3H), 2.34 (s, 3H), 1.98–1.61 (m, 8H); ¹³C NMR (DMSO-*d*₆): δ 158.89, 150.73, 142.96, 138.61, 138.42, 132.55, 125.70, 123.61, 119.28, 96.66, 79.69, 43.35, 33.90, 29.22, 25.10, 17.32, 13.49; HRMS (EI) *m/z*: 374.2215 M⁺, calcd for C₂₇H₂₆N₆, 374.2219. Anal. Calcd for C₂₇H₂₆N₆ (374): C, 70.56; H, 7.00; N, 22.44%. Found: C, 70.43; H, 7.27; N, 22.60%.

4-(1-(2-(4,5-Dihydro-1H-imidazol-2-yl)hydrazono)ethyl)-1-(4-(3,3-dimethylbut-1-yn-1-yl)phenyl)-5-methyl-1H-pyrazole (36). It is obtained as pale yellow solid (38 mg, 30%) mp 190–192 °C. ¹H NMR (DMSO-*d*₆): δ 7.86 (s, 1H), 7.54 (d, *J* = 8.8 Hz, 2H), 7.51 (d, *J* = 8.8 Hz, 2H), 6.26 (br s, 1H), 6.10 (br s, 1H), 3.39 (s, 4H), 2.52 (s, 3H), 2.18 (s, 3H), 0.93 (s, 9H); ¹³C NMR (DMSO-*d*₆): δ 164.24, 146.11, 141.20, 139.84, 137.90, 136.45, 130.32, 128.00, 125.09, 94.24, 83.75, 42.87, 34.67, 30.86, 15.97, 13.75; HRMS (EI) *m/z*: 362.2222 M⁺, calcd for C₂₁H₂₆N₆, 362.2219. Anal. Calcd for C₂₁H₂₆N₆ (362): C, 69.58; H, 7.23; N, 23.19%. Found: C, 69.47; H, 7.58; N, 23.24%.

4-(4-(1-(2-(4,5-Dihydro-1H-imidazol-2-yl)hydrazono)ethyl)-5-methyl-1H-pyrazol-1-yl)phenyl)-2-methylbut-3-yn-2-ol (37). It is obtained as brown solid (67 mg, 52%) mp 193–195 °C. ¹H NMR (DMSO-*d*₆): δ 7.86 (s, 1H), 7.55 (d, *J* = 8.7 Hz, 2H), 7.50 (d, *J* = 8.7 Hz, 2H), 6.32 (br s, 1H), 6.09 (br s, 1H), 5.43 (br s, 1H), 3.37 (s, 4H), 2.53 (s, 3H), 2.18 (s, 3H), 1.49 (s, 6H); ¹³C NMR (DMSO-*d*₆): δ 163.76, 145.55, 139.73, 139.12, 135.95, 132.17, 124.86, 121.84, 121.15, 93.07, 82.62, 69.77, 42.40, 28.61, 15.47, 13.25; HRMS (EI) *m/z*: 364.2031 M⁺, calcd for C₂₀H₂₄N₆O, 364.2012. Anal. Calcd for C₂₀H₂₄N₆O (364): C, 65.91; H, 6.64; N, 23.06%. Found: C, 65.79; H, 6.81; N, 23.24%.

1-(4-(1-(2-(4,5-Dihydro-1H-imidazol-2-yl)hydrazono)ethyl)-5-methyl-1H-pyrazol-1-yl)phenylethynyl)cyclohexanol (38). It is obtained as yellow solid (62 mg, 50%) mp 297–298 °C. ¹H NMR (DMSO-*d*₆): δ 11.28 (s, 1H), 8.07 (s, 1H), 7.62 (d, *J* = 8.4 Hz, 2H), 7.55 (d, *J* = 8.1 Hz, 2H), 5.46 (br s, 1H), 4.92 (br s, 1H), 3.71 (s, 4H), 2.50 (s, 3H), 2.32 (s, 3H), 1.85–1.22 (m, 10H); ¹³C NMR (DMSO-*d*₆): δ 158.23, 150.39, 140.43, 138.19, 138.13, 132.04, 125.21, 122.43, 118.78, 96.28, 81.79, 66.98, 42.84, 24.82, 22.74, 22.45, 16.70, 12.97; HRMS (EI) *m/z*: 404.2320 M⁺, calcd for C₂₃H₂₈N₆O, 404.2325. Anal. Calcd for C₂₃H₂₈N₆O (404): C, 68.29; H, 6.98; N, 20.78%. Found: C, 68.15; H, 7.20; N, 20.95%.

3-(4-(1-(2-(4,5-Dihydro-1H-imidazol-2-yl)hydrazono)ethyl)-5-methyl-1H-pyrazol-1-yl)phenyl)-N,N-diethylprop-2-yn-1-amine (39). It is obtained as brown solid (64 mg, 43%) mp 122–124 °C. ¹H NMR (DMSO-*d*₆): δ 8.24 (br s, 1H), 7.86 (s, 1H), 7.65 (d, *J* = 7.8 Hz, 2H), 7.49 (d, *J* = 8.4 Hz, 2H), 6.25 (br s, 1H), 3.62 (s, 2H), 3.38 (s, 4H), 2.53 (s, 3H), 2.46 (q, *J* = 7.2 Hz, 4H), 2.19 (s, 3H), 1.02 (t, *J* = 7.2 Hz, 6H); HRMS (EI) *m/z*: 391.2472 M⁺, calcd for C₂₂H₂₈N₇,

391.2484. Anal. Calcd for C₂₂H₂₈N₇ (391): C, 67.49; H, 7.47; N, 25.04%. Found: C, 67.43; H, 7.86; N, 25.17%.

Microbiological Assays. MRSA Biofilm Inhibition Assessment. Phenylpyrazole compounds were tested for their ability to inhibit the formation of the staphylococcal biofilm using the microtiter plate biofilm formation assay following the procedure reported elsewhere.^{55,59,60}

In Vitro Cytotoxicity Analysis Against Caco-2 Cells, Vero Cells, and J774 Cells. Phenylpyrazoles were assayed for their *in vitro* cytotoxicity profile against Caco-2 cells, Vero cells, and murine macrophage (J774) cells as reported earlier.⁸

Clearance of Intracellular MRSA USA400 and *S. enteritidis* Present in Murine Macrophage (J774) Cells. The ability of phenylpyrazoles to reduce the burden of intracellular MRSA or *S. enteritidis* was evaluated using previously described methods.⁵

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jmedchem.9b00720.

Strains used in the manuscript; their corresponding sites of isolation and the protocol used for MIC and MBC determination; details about *in vitro* and *in vivo* PK measurements; toxicity data; biochemical assays; ¹H and ¹³C NMR spectra of all new described compounds; and high-performance liquid chromatography method and chromatograms (PDF)

Molecular formula strings (CSV)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: mseleem@purdue.edu (M.N.S.).

*E-mail: amayhoub@azhar.edu.eg (A.S.M.).

ORCID

Nader S. Abutaleb: 0000-0003-1730-4150

Mohamed N. Seleem: 0000-0003-0939-0458

Abdelrahman S. Mayhoub: 0000-0002-3987-3680

Author Contributions

△A.H. and N.S.A. contributed equally

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was funded by Science & Technology Development Funds (STDF-Egypt) and the Egyptian Ministry of Higher Education and Scientific Research (MHESR). The work is derived from the Subject Data funded in whole or part by NAS and USAID and that any opinions, findings, conclusions, or recommendations expressed in such article are those of the authors alone and do not necessarily reflect the views of USAID or NAS. The authors would like to thank ATCC and BEI Resources for providing the clinical isolates used in this study.

■ ABBREVIATIONS

CFUs, colony forming units; Cl_{int}, intrinsic clearance; CRE, carbapenem-resistant *Enterobacteriaceae*; DME, dimethyl ethylene glycol; MBC, minimum bactericidal concentration; MIC, minimum inhibitory concentration; NDM-1, New Delhi metallo-beta-lactamase; KPC, *K. pneumoniae* carbapenemase; P_{app}, apparent permeability; P-gp, P-glycoprotein; t_{1/2}, half-life;

UppP, undecaprenyl pyrophosphate phosphatase; UppS, undecaprenyl pyrophosphate synthase

REFERENCES

- (1) *Antibacterial Agents in Clinical Development: An Analysis of the Antibacterial Clinical Development Pipeline, Including Tuberculosis*; World Health Organization: Geneva, 2017.
- (2) Nikaido, H. Prevention of Drug Access to Bacterial Targets: Permeability Barriers and Active Efflux. *Science* **1994**, *264*, 382–388.
- (3) Yamashita, E.; Zhalnina, M. V.; Zakharov, S. D.; Sharma, O.; Cramer, W. A. Crystal Structures of the OmpF Porin: Function in a Colicin Translocon. *EMBO J.* **2008**, *27*, 2171–2180.
- (4) O'Shea, R.; Moser, H. E. Physicochemical Properties of Antibacterial Compounds: Implications for Drug Discovery. *J. Med. Chem.* **2008**, *51*, 2871–2878.
- (5) Mohammad, H.; Mayhoub, A. S.; Ghafoor, A.; Soofi, M.; Alajlouni, R. A.; Cushman, M.; Seleem, M. N. Discovery and Characterization of Potent Thiazoles Versus Methicillin- and Vancomycin-Resistant *Staphylococcus Aureus*. *J. Med. Chem.* **2014**, *57*, 1609–1615.
- (6) Eid, I.; Elsebaei, M. M.; Mohammad, H.; Hagra, M.; Peters, C. E.; Hegazy, Y. A.; Cooper, B.; Pogliano, J.; Pogliano, K.; Abulkhair, H. S.; Seleem, M. N.; Mayhoub, A. S. Arylthiazole Antibiotics Targeting Intracellular Methicillin-Resistant *Staphylococcus Aureus* (MRSA) That Interfere with Bacterial Cell Wall Synthesis. *Eur. J. Med. Chem.* **2017**, *139*, 665–673.
- (7) Eissa, I. H.; Mohammad, H.; Qassem, O. A.; Younis, W.; Abdelghany, T. M.; Elshafeey, A.; Abd Rabo Moustafa, M. M.; Seleem, M. N.; Mayhoub, A. S. Diphenylurea Derivatives for Combating Methicillin- and Vancomycin-Resistant *Staphylococcus Aureus*. *Eur. J. Med. Chem.* **2017**, *130*, 73–85.
- (8) Elsebaei, M. M.; Mohammad, H.; Abouf, M.; Abutaleb, N. S.; Hegazy, Y. A.; Ghiaty, A.; Chen, L.; Zhang, J.; Malwal, S. R.; Oldfield, E.; Seleem, M. N.; Mayhoub, A. S. Alkynyl-Containing Phenylthiazoles: Systemically Active Antibacterial Agents Effective against Methicillin-Resistant *Staphylococcus Aureus* (MRSA). *Eur. J. Med. Chem.* **2018**, *148*, 195–209.
- (9) Hagra, M.; Hegazy, Y. A.; Elkabbany, A. H.; Mohammad, H.; Ghiaty, A.; Abdelghany, T. M.; Seleem, M. N.; Mayhoub, A. S. Biphenylthiazole Antibiotics with an Oxadiazole Linker: An Approach to Improve Physicochemical Properties and Oral Bioavailability. *Eur. J. Med. Chem.* **2018**, *143*, 1448–1456.
- (10) Mohammad, H.; Mayhoub, A. S.; Cushman, M.; Seleem, M. N. Anti-Biofilm Activity and Synergism of Novel Thiazole Compounds with Glycopeptide Antibiotics against Multidrug-Resistant *Staphylococci*. *J. Antibiot.* **2015**, *68*, 259–266.
- (11) Mohammad, H.; Reddy, P. V. N.; Monteleone, D.; Mayhoub, A. S.; Cushman, M.; Seleem, M. N. Synthesis and Antibacterial Evaluation of a Novel Series of Synthetic Phenylthiazole Compounds against Methicillin-Resistant *Staphylococcus Aureus* (MRSA). *Eur. J. Med. Chem.* **2015**, *94*, 306–316.
- (12) Mohammad, H.; Younis, W.; Chen, L.; Peters, C. E.; Pogliano, J.; Pogliano, K.; Cooper, B.; Zhang, J.; Mayhoub, A.; Oldfield, E.; Cushman, M.; Seleem, M. N. Phenylthiazole Antibacterial Agents Targeting Cell Wall Synthesis Exhibit Potent Activity in Vitro and in Vivo against Vancomycin-Resistant Enterococci. *J. Med. Chem.* **2017**, *60*, 2425–2438.
- (13) Seleem, M. A.; Disouky, A. M.; Mohammad, H.; Abdelghany, T. M.; Mancy, A. S.; Bayoumi, S. A.; Elshafeey, A.; El-Morsy, A.; Seleem, M. N.; Mayhoub, A. S. Second-Generation Phenylthiazole Antibiotics with Enhanced Pharmacokinetic Properties. *J. Med. Chem.* **2016**, *59*, 4900–4912.
- (14) Yahia, E.; Mohammad, H.; Abdelghany, T. M.; Fayed, E.; Seleem, M. N.; Mayhoub, A. S. Phenylthiazole Antibiotics: A Metabolism-Guided Approach to Overcome Short Duration of Action. *Eur. J. Med. Chem.* **2017**, *126*, 604–613.
- (15) Hagra, M.; Abutaleb, N. S.; Ali, A. O.; Abdel-Aleem, J. A.; Elsebaei, M. M.; Seleem, M. N.; Mayhoub, A. S. Naphthylthiazoles: Targeting Multidrug-Resistant and Intracellular *Staphylococcus aureus* with Biofilm Disruption Activity. *ACS Infect. Dis.* **2018**, *4*, 1679–1691.
- (16) Denis, O. Route of transmission of *Staphylococcus aureus*. *Lancet Infect. Dis.* **2017**, *17*, 124–125.
- (17) David, M. Z.; Daum, R. S. Community-Associated Methicillin-Resistant *Staphylococcus Aureus*: Epidemiology and Clinical Consequences of an Emerging Epidemic. *Clin. Microbiol. Rev.* **2010**, *23*, 616–687.
- (18) Centers for Disease Control and Prevention. *Antibiotic Resistance Threats in the United States, 2013*, US Department of Health and Human Services, Meeting the Challenges of Drug-Resistant Diseases in Developing Countries Committee on Foreign Affairs Subcommittee on Africa, Global Health, Human Rights, and International Organizations, United States House of Representatives, April 23, 2013.
- (19) Han, L. L.; McDougal, L. K.; Gorwitz, R. J.; Mayer, K. H.; Patel, J. B.; Sennott, J. M.; Fontana, J. L. High Frequencies of Clindamycin and Tetracycline Resistance in Methicillin-Resistant *Staphylococcus Aureus* Pulsed-Field Type USA300 Isolates Collected at a Boston Ambulatory Health Center. *J. Clin. Microbiol.* **2007**, *45*, 1350–1352.
- (20) Frazee, B. W.; Lynn, J.; Charlebois, E. D.; Lambert, L.; Lowery, D.; Perdreau-Remington, F. High Prevalence of Methicillin-Resistant *Staphylococcus Aureus* in Emergency Department Skin and Soft Tissue Infections. *Ann. Emerg. Med.* **2005**, *45*, 311–320.
- (21) Locke, J. B.; Morales, G.; Hilgers, M.; Kedar, G. C.; Rahawi, S.; Picazo, J. J.; Shaw, K. J.; Stein, J. L. Elevated Linezolid Resistance in Clinical Cfr-Positive *Staphylococcus Aureus* Isolates Is Associated with Co-Occurring Mutations in Ribosomal Protein L3. *Antimicrob. Agents Chemother.* **2010**, *54*, 5352–5355.
- (22) Hiramatsu, K. Vancomycin-resistant *Staphylococcus aureus*: a new model of antibiotic resistance. *Lancet Infect. Dis.* **2001**, *1*, 147–155.
- (23) Shen, H.; Akoda, E.; Zhang, K. Methicillin-Resistant *Staphylococcus Aureus* Carriage among Students at a Historically Black University: A Case Study. *Int. J. Microbiol.* **2013**, *2013*, 979734.
- (24) Arias, C. A.; Murray, B. E. Emergence and Management of Drug-Resistant Enterococcal Infections. *Expert Rev. Anti-Infect. Ther.* **2008**, *6*, 637–655.
- (25) AbdelKhalek, A.; Abutaleb, N. S.; Mohammad, H.; Seleem, M. N. Repurposing Ebselen for Decolonization of Vancomycin-Resistant Enterococci (Vre). *PLoS One* **2018**, *13*, e0199710.
- (26) AbdelKhalek, A.; Abutaleb, N. S.; Elmagarmid, K. A.; Seleem, M. N. Repurposing Auranofin as an Intestinal Decolonizing Agent for Vancomycin-Resistant Enterococci. *Sci. Rep.* **2018**, *8*, 8353.
- (27) Mohammad, H.; AbdelKhalek, A.; Abutaleb, N. S.; Seleem, M. N. Repurposing Niclosamide for Intestinal Decolonization of Vancomycin-Resistant Enterococci. *Int. J. Antimicrob. Agents* **2018**, *51*, 897–904.
- (28) Gonzales, R. D.; Schreckenberger, P. C.; Graham, M. B.; Kelkar, S.; DenBesten, K.; Quinn, J. P. Infections Due to Vancomycin-Resistant Enterococcus Faecium Resistant to Linezolid. *Lancet* **2001**, *357*, 1179.
- (29) Moellering, R. C., Jr. The Continuing Challenge of Lower Respiratory Tract Infections. *Clin. Infect. Dis.* **2004**, *38*, S319–S321.
- (30) Feldman, C. Clinical Relevance of Antimicrobial Resistance in the Management of Pneumococcal Community-Acquired Pneumonia. *J. Lab. Clin. Med.* **2004**, *143*, 269–283.
- (31) Vuong, C.; Otto, M. *Staphylococcus Epidermidis* Infections. *Microbes Infect.* **2002**, *4*, 481–489.
- (32) Otto, M. *Staphylococcus epidermidis*—the “accidental” pathogen. *Nat. Rev. Microbiol.* **2009**, *7*, 555.
- (33) Schlech, W. F., III Epidemiology and Clinical Manifestations of *Listeria Monocytogenes* Infection. *Gram-Positive Pathogens*, 2nd ed.; American Society of Microbiology: Washington, DC 2003; 2904, 2006; pp 601–608.
- (34) Vázquez-Boland, J. A.; Kuhn, M.; Berche, P.; Chakraborty, T.; Domínguez-Bernal, G.; Goebel, W.; González-Zorn, B.; Wehland, J.

- Kreft, J. *Listeria* Pathogenesis and Molecular Virulence Determinants. *Clin. Microbiol. Rev.* **2001**, *14*, 584–640.
- (35) Hussein, A. S.; Shafraan, S. D. Acute Bacterial Meningitis in Adults: A 12-Year Review. *Medicine* **2000**, *79*, 360–368.
- (36) Weiner, L. M.; Webb, A. K.; Limbago, B.; Dudeck, M. A.; Patel, J.; Kallen, A. J.; Edwards, J. R.; Sievert, D. M. Antimicrobial-Resistant Pathogens Associated With Healthcare-Associated Infections: Summary of Data Reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2011–2014. *Infect. Control Hosp. Epidemiol.* **2016**, *37*, 1288–1301.
- (37) Fischbach, M. A.; Walsh, C. T. Antibiotics for emerging pathogens. *Science* **2009**, *325*, 1089–1093.
- (38) Meletis, G.; Exindari, M.; Vavatsi, N.; Sofianou, D.; Diza, E. Mechanisms responsible for the emergence of carbapenem resistance in *Pseudomonas aeruginosa*. *Hippokratia* **2012**, *16*, 303.
- (39) Evans, B. A.; Hamouda, A.; Amys, S. G. B. The Rise of Carbapenem-Resistant *Acinetobacter Baumannii*. *Curr. Pharm. Des.* **2013**, *19*, 223–238.
- (40) Poirel, L.; Nordmann, P. Carbapenem Resistance in *Acinetobacter Baumannii*: Mechanisms and Epidemiology. *Clin. Microbiol. Infect.* **2006**, *12*, 826–836.
- (41) Tien, H. C.; Battad, A.; Bryce, E. A.; Fuller, J.; Mulvey, M.; Bernard, K.; Brisebois, R.; Doucet, J. J.; Rizoli, S. B.; Fowler, R.; Simor, A. Multi-Drug Resistant *Acinetobacter* Infections in Critically Injured Canadian Forces Soldiers. *BMC Infect. Dis.* **2007**, *7*, 95.
- (42) Leski, T.; Vora, G. J.; Taitt, C. R. Multidrug Resistance Determinants from Ndm-1-Producing *Klebsiella pneumoniae* in the USA. *Int. J. Antimicrob. Agents* **2012**, *40*, 282–284.
- (43) Center of Disease Control and Prevention. *Detection of Enterobacteriaceae Isolates Carrying Metallo-Beta-Lactamase-United States, 2010. MMWR. Morbidity and Mortality Weekly Report*, 2010; Vol. 59, p 750.
- (44) Yigit, H.; Queenan, A. M.; Anderson, G. J.; Domenech-Sanchez, A.; Biddle, J. W.; Steward, C. D.; Alberti, S.; Bush, K.; Tenover, F. C. Novel Carbapenem-Hydrolyzing β -Lactamase, KPC-1, from a Carbapenem-Resistant Strain of *Klebsiella pneumoniae*. *Antimicrob. Agents Chemother.* **2008**, *52*, 809.
- (45) Lledo, W.; Hernandez, M.; Lopez, E.; Molinari, O.; Soto, R.; Hernandez, E.; Santiago, N.; Flores, M.; Vazquez, G.; Robledo, I. *Guidance for Control of Infections with Carbapenem-Resistant or Carbapenemase-Producing Enterobacteriaceae in Acute Care Facilities. Morbidity and Mortality Weekly Report*, 2009; Vol. 58, pp 256–258.
- (46) Hatrongjit, R.; Kerdin, A.; Akeda, Y.; Hamada, S. Detection of Plasmid-Mediated Colistin-Resistant and Carbapenem-Resistant Genes by Multiplex PCR. *MethodsX* **2018**, *5*, 532–536.
- (47) Gupta, N.; Limbago, B. M.; Patel, J. B.; Kallen, A. J. Carbapenem-Resistant Enterobacteriaceae: Epidemiology and Prevention. *Clin. Infect. Dis.* **2011**, *53*, 60–67.
- (48) Charrier, C.; Salisbury, A.-M.; Savage, V. J.; Moyo, E.; Forward, H.; Ooi, N.; Cheung, J.; Metzger, R.; McGarry, D.; Walker, R.; Cooper, I. R.; Ratcliffe, A. J.; Stokes, N. R. In vitro biological evaluation of novel broad-spectrum isothiazolone inhibitors of bacterial type II topoisomerases. *J. Antimicrob. Chemother.* **2016**, *71*, 2831–2839.
- (49) Ahmed, M. N.; Porse, A.; Sommer, M. O. A.; Høiby, N.; Ciofu, O. Evolution of Antibiotic Resistance in Biofilm and Planktonic *Pseudomonas aeruginosa* Populations Exposed to Subinhibitory Levels of Ciprofloxacin. *Antimicrob. Agents Chemother.* **2018**, *62*, e00320–18.
- (50) Dunne, W. M., Jr.; Mason, E. O., Jr.; Kaplan, S. L. Diffusion of Rifampin and Vancomycin through a *Staphylococcus Epidermidis* Biofilm. *Antimicrob. Agents Chemother.* **1993**, *37*, 2522–2526.
- (51) Claessens, J.; Roriz, M.; Merckx, R.; Baatsen, P.; Van Mellaert, L.; Van Eldere, J. Inefficacy of Vancomycin and Teicoplanin in Eradicating and Killing *Staphylococcus Epidermidis* Biofilms in Vitro. *Int. J. Antimicrob. Agents* **2015**, *45*, 368–375.
- (52) Mohamed, M. F.; Abdelkhalek, A.; Seleem, M. N. Evaluation of Short Synthetic Antimicrobial Peptides for Treatment of Drug-Resistant and Intracellular *Staphylococcus Aureus*. *Sci. Rep.* **2016**, *6*, 29707.
- (53) Mohamed, M. F.; Hamed, M. I.; Panitch, A.; Seleem, M. N. Targeting Methicillin-Resistant *Staphylococcus Aureus* with Short Salt-Resistant Synthetic Peptides. *Antimicrob. Agents Chemother.* **2014**, *58*, 4113–4122.
- (54) Brezden, A.; Mohamed, M. F.; Nepal, M.; Harwood, J. S.; Kuriakose, J.; Seleem, M. N.; Chmielewski, J. Dual Targeting of Intracellular Pathogenic Bacteria with a Cleavable Conjugate of Kanamycin and an Antibacterial Cell-Penetrating Peptide. *J. Am. Chem. Soc.* **2016**, *138*, 10945–10949.
- (55) Kotb, A.; Abutaleb, N. S.; Seleem, M. A.; Hagra, M.; Mohammad, H.; Bayoumi, A.; Ghiaty, A.; Seleem, M. N.; Mayhoub, A. S. Phenylthiazoles with Tert-Butyl Side Chain: Metabolically Stable with Anti-Biofilm Activity. *Eur. J. Med. Chem.* **2018**, *151*, 110–120.
- (56) Schmitz, F.-J.; Fluit, A. C.; Hafner, D.; Beeck, A.; Perdikouli, M.; Boos, M.; Scheuring, S.; Verhoef, J.; Kohrer, K.; Von Eiff, C. Development of Resistance to Ciprofloxacin, Rifampin, and Mupirocin in Methicillin-Susceptible and -Resistant *Staphylococcus aureus* Isolates. *Antimicrob. Agents Chemother.* **2000**, *44*, 3229–3231.
- (57) Brown, F. C.; Bradsher, C. K.; Moser, B. F.; Forrester, S. Structure and Antimicrobial Activity of the 3-Aminorhodanines. *J. Org. Chem.* **1959**, *24*, 1056–1060.
- (58) Wang, S.; Li, Y.; Bi, X.; Liu, Q. Synthesis of 1,3,4-Trisubstituted Pyrazoles from α -(1, 3-Dithian-2-yl) Enamine Ketones Via [4+1] Cyclization. *Synlett* **2015**, *26*, 1895–1899.
- (59) Hagra, M.; Abutaleb, N. S.; Ali, A. O.; Abdel-Aleem, J. A.; Elsebaei, M. M.; Seleem, M. N.; Mayhoub, A. S. Naphthylthiazoles: Targeting Multidrug-Resistant and Intracellular *Staphylococcus aureus* with Biofilm Disruption Activity. *ACS Infect. Dis.* **2018**, *4*, 1679–1691.
- (60) Kotb, A.; Abutaleb, N. S.; Hagra, M.; Bayoumi, A.; Moustafa, M. M.; Ghiaty, A.; Seleem, M. N.; Mayhoub, A. S. Tert-Butylphenylthiazoles with an Oxadiazole Linker: A Novel Orally Bioavailable Class of Antibiotics Exhibiting Antibiofilm Activity. *RSC Adv.* **2019**, *9*, 6770–6778.

REFERENCES

1. Nelson CE, Hersh BM, Carroll SB. The regulatory content of intergenic DNA shapes genome architecture. *Genome biology*. 2004;5(4):R25. doi: 10.1186/gb-2004-5-4-r25. PubMed PMID: 15059258; PMCID: PMC395784.
2. Dar D, Sorek R. Bacterial Noncoding RNAs Excised from within Protein-Coding Transcripts. *mBio*. 2018;9(5). doi: 10.1128/mBio.01730-18. PubMed PMID: 30254125; PMCID: PMC6156199.
3. Juven-Gershon T, Kadonaga JT. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Developmental biology*. 2010;339(2):225-9. doi: 10.1016/j.ydbio.2009.08.009. PubMed PMID: 19682982; PMCID: 2830304.
4. Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. *Annual review of genomics and human genetics*. 2006;7:29-59. doi: 10.1146/annurev.genom.7.080505.115623. PubMed PMID: 16719718.
5. Narlikar L, Ovcharenko I. Identifying regulatory elements in eukaryotic genomes. *Briefings in functional genomics & proteomics*. 2009;8(4):215-30. doi: 10.1093/bfpg/elp014. PubMed PMID: 19498043; PMCID: 2764519.
6. Palstra RJ, Grosveld F. Transcription factor binding at enhancers: shaping a genomic regulatory landscape in flux. *Frontiers in genetics*. 2012;3:195. doi: 10.3389/fgene.2012.00195. PubMed PMID: 23060900; PMCID: 3460357.
7. Fuda NJ, Ardehali MB, Lis JT. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature*. 2009;461(7261):186-92. doi: 10.1038/nature08449. PubMed PMID: 19741698; PMCID: PMC2833331.
8. Inukai S, Kock KH, Bulyk ML. Transcription factor-DNA binding: beyond binding site motifs. *Current opinion in genetics & development*. 2017;43:110-9. doi: 10.1016/j.gde.2017.02.007. PubMed PMID: 28359978; PMCID: PMC5447501.
9. Boehm AK, Saunders A, Werner J, Lis JT. Transcription factor and polymerase recruitment, modification, and movement on dhsp70 in vivo in the minutes following heat shock. *Molecular and cellular biology*. 2003;23(21):7628-37. doi: 10.1128/mcb.23.21.7628-7637.2003. PubMed PMID: 14560008; PMCID: 207607.
10. Adelman K, Lis JT. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature reviews Genetics*. 2012;13(10):720-31. doi: 10.1038/nrg3293. PubMed PMID: 22986266; PMCID: PMC3552498.
11. Mayer A, Landry HM, Churchman LS. Pause & go: from the discovery of RNA polymerase pausing to its functional implications. *Curr Opin Cell Biol*. 2017;46:72-80. doi: 10.1016/j.ceb.2017.03.002. PubMed PMID: 28363125; PMCID: PMC5505790.

12. Landick R. The regulatory roles and mechanism of transcriptional pausing. *Biochemical Society transactions*. 2006;34(Pt 6):1062-6. doi: 10.1042/BST0341062. PubMed PMID: 17073751.
13. Chen FX, Smith ER, Shilatifard A. Born to run: control of transcription elongation by RNA polymerase II. *Nature reviews Molecular cell biology*. 2018;19(7):464-78. doi: 10.1038/s41580-018-0010-5. PubMed PMID: 29740129.
14. Vassilyev DG, Vassilyeva MN, Perederina A, Tahirov TH, Artsimovitch I. Structural basis for transcription elongation by bacterial RNA polymerase. *Nature*. 2007;448(7150):157-62. doi: 10.1038/nature05932. PubMed PMID: 17581590.
15. Porrua O, Boudvillain M, Libri D. Transcription Termination: Variations on Common Themes. *Trends in genetics : TIG*. 2016;32(8):508-22. doi: 10.1016/j.tig.2016.05.007. PubMed PMID: 27371117.
16. Rojo F. Repression of transcription initiation in bacteria. *Journal of bacteriology*. 1999;181(10):2987-91. PubMed PMID: 10321997; PMCID: 93751.
17. Ip YT. Transcriptional regulation. Converting an activator into a repressor. *Current biology : CB*. 1995;5(1):1-3. doi: 10.1016/s0960-9822(95)00001-7. PubMed PMID: 7697337.
18. Casadesus J, Low D. Epigenetic gene regulation in the bacterial world. *Microbiology and molecular biology reviews : MMBR*. 2006;70(3):830-56. doi: 10.1128/MMBR.00016-06. PubMed PMID: 16959970; PMCID: 1594586.
19. Ramakrishnan V. Histone structure and the organization of the nucleosome. *Annual review of biophysics and biomolecular structure*. 1997;26:83-112. doi: 10.1146/annurev.biophys.26.1.83. PubMed PMID: 9241414.
20. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*. 1997;389(6648):251-60. doi: 10.1038/38444. PubMed PMID: 9305837.
21. Martire S, Banaszynski LA. The roles of histone variants in fine-tuning chromatin organization and function. *Nature reviews Molecular cell biology*. 2020. doi: 10.1038/s41580-020-0262-8. PubMed PMID: 32665685.
22. Felsenfeld G, Boyes J, Chung J, Clark D, Studitsky V. Chromatin structure and gene expression. *Proceedings of the National Academy of Sciences of the United States of America*. 1996;93(18):9384-8. doi: 10.1073/pnas.93.18.9384. PubMed PMID: 8790338; PMCID: 38436.
23. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell research*. 2011;21(3):381-95. doi: 10.1038/cr.2011.22. PubMed PMID: 21321607; PMCID: 3193420.

24. Geiman TM, Robertson KD. Chromatin remodeling, histone modifications, and DNA methylation-how does it all fit together? *Journal of cellular biochemistry*. 2002;87(2):117-25. doi: 10.1002/jcb.10286. PubMed PMID: 12244565.
25. Strahl BD, Allis CD. The language of covalent histone modifications. *Nature*. 2000;403(6765):41-5. doi: 10.1038/47412. PubMed PMID: 10638745.
26. Babu A, Verma RS. Chromosome structure: euchromatin and heterochromatin. *International review of cytology*. 1987;108:1-60. doi: 10.1016/s0074-7696(08)61435-7. PubMed PMID: 2822591.
27. Nair SS, Kumar R. Chromatin remodeling in cancer: a gateway to regulate gene transcription. *Molecular oncology*. 2012;6(6):611-9. doi: 10.1016/j.molonc.2012.09.005. PubMed PMID: 23127546; PMCID: 3538127.
28. Edwards JR, Yarychivska O, Boulard M, Bestor TH. DNA methylation and DNA methyltransferases. *Epigenetics & chromatin*. 2017;10:23. Epub 2017/05/16. doi: 10.1186/s13072-017-0130-8. PubMed PMID: 28503201; PMCID: PMC5422929.
29. Moore LD, Le T, Fan G. DNA methylation and its basic function. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*. 2013;38(1):23-38. doi: 10.1038/npp.2012.112. PubMed PMID: 22781841; PMCID: 3521964.
30. Jeltsch A. Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *Chembiochem : a European journal of chemical biology*. 2002;3(4):274-93. doi: 10.1002/1439-7633(20020402)3:4<274::AID-CBIC274>3.0.CO;2-S. PubMed PMID: 11933228.
31. Cheng X. DNA modification by methyltransferases. *Current opinion in structural biology*. 1995;5(1):4-10. doi: 10.1016/0959-440x(95)80003-j. PubMed PMID: 7773746.
32. Ziller MJ, Muller F, Liao J, Zhang Y, Gu H, Bock C, Boyle P, Epstein CB, Bernstein BE, Lengauer T, Gnirke A, Meissner A. Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS genetics*. 2011;7(12):e1002389. doi: 10.1371/journal.pgen.1002389. PubMed PMID: 22174693; PMCID: 3234221.
33. Schubeler D. Function and information content of DNA methylation. *Nature*. 2015;517(7534):321-6. doi: 10.1038/nature14192. PubMed PMID: 25592537.
34. Jang HS, Shin WJ, Lee JE, Do JT. CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function. *Genes*. 2017;8(6). doi: 10.3390/genes8060148. PubMed PMID: 28545252; PMCID: 5485512.
35. Goll MG, Bestor TH. Eukaryotic cytosine methyltransferases. *Annual review of biochemistry*. 2005;74:481-514. doi: 10.1146/annurev.biochem.74.010904.153721. PubMed PMID: 15952895.

36. Feng S, Jacobsen SE, Reik W. Epigenetic reprogramming in plant and animal development. *Science*. 2010;330(6004):622-7. doi: 10.1126/science.1190614. PubMed PMID: 21030646; PMCID: 2989926.
37. Sanchez-Romero MA, Cota I, Casadesus J. DNA methylation in bacteria: from the methyl group to the methylome. *Current opinion in microbiology*. 2015;25:9-16. doi: 10.1016/j.mib.2015.03.004. PubMed PMID: 25818841.
38. Blow MJ, Clark TA, Daum CG, Deutschbauer AM, Fomenkov A, Fries R, Froula J, Kang DD, Malmstrom RR, Morgan RD, Posfai J, Singh K, Visel A, Wetmore K, Zhao Z, Rubin EM, Korlach J, Pennacchio LA, Roberts RJ. The Epigenomic Landscape of Prokaryotes. *PLoS genetics*. 2016;12(2):e1005854. doi: 10.1371/journal.pgen.1005854. PubMed PMID: 26870957; PMCID: 4752239.
39. Bickle TA. Restricting restriction. *Molecular microbiology*. 2004;51(1):3-5. doi: 10.1046/j.1365-2958.2003.03846.x. PubMed PMID: 14651606.
40. Loenen WA, Dryden DT, Raleigh EA, Wilson GG, Murray NE. Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic acids research*. 2014;42(1):3-19. doi: 10.1093/nar/gkt990. PubMed PMID: 24141096; PMCID: 3874209.
41. Loenen WA, Raleigh EA. The other face of restriction: modification-dependent enzymes. *Nucleic acids research*. 2014;42(1):56-69. doi: 10.1093/nar/gkt747. PubMed PMID: 23990325; PMCID: 3874153.
42. Wion D, Casadesus J. N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nature reviews Microbiology*. 2006;4(3):183-92. doi: 10.1038/nrmicro1350. PubMed PMID: 16489347; PMCID: 2755769.
43. Low DA, Casadesus J. Clocks and switches: bacterial gene regulation by DNA adenine methylation. *Current opinion in microbiology*. 2008;11(2):106-12. doi: 10.1016/j.mib.2008.02.012. PubMed PMID: 18396448.
44. Kumar R, Mukhopadhyay AK, Ghosh P, Rao DN. Comparative transcriptomics of *H. pylori* strains AM5, SS1 and their hpyAVIBM deletion mutants: possible roles of cytosine methylation. *PloS one*. 2012;7(8):e42303. doi: 10.1371/journal.pone.0042303. PubMed PMID: 22879937; PMCID: 3411764.
45. Kahramanoglou C, Prieto AI, Khedkar S, Haase B, Gupta A, Benes V, Fraser GM, Luscombe NM, Seshasayee AS. Genomics of DNA cytosine methylation in *Escherichia coli* reveals its role in stationary phase transcription. *Nature communications*. 2012;3:886. doi: 10.1038/ncomms1878. PubMed PMID: 22673913.
46. Schmitz RJ, Lewis ZA, Goll MG. DNA Methylation: Shared and Divergent Features across Eukaryotes. *Trends in genetics : TIG*. 2019;35(11):818-27. doi: 10.1016/j.tig.2019.07.007. PubMed PMID: 31399242; PMCID: 6825889.

47. Bartels A, Han Q, Nair P, Stacey L, Gaynier H, Mosley M, Huang QQ, Pearson JK, Hsieh TF, An YC, Xiao W. Dynamic DNA Methylation in Plant Growth and Development. *International journal of molecular sciences*. 2018;19(7). doi: 10.3390/ijms19072144. PubMed PMID: 30041459; PMCID: 6073778.
48. Selker EU, Tountas NA, Cross SH, Margolin BS, Murphy JG, Bird AP, Freitag M. The methylated component of the *Neurospora crassa* genome. *Nature*. 2003;422(6934):893-7. doi: 10.1038/nature01564. PubMed PMID: 12712205.
49. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature reviews Genetics*. 2012;13(7):484-92. doi: 10.1038/nrg3230. PubMed PMID: 22641018.
50. Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, Das PK, Kivioja T, Dave K, Zhong F, Nitta KR, Taipale M, Popov A, Ginno PA, Domcke S, Yan J, Schubeler D, Vinson C, Taipale J. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*. 2017;356(6337). doi: 10.1126/science.aaj2239. PubMed PMID: 28473536.
51. Criscione SW, Zhang Y, Thompson W, Sedivy JM, Neretti N. Transcriptional landscape of repetitive elements in normal and cancer human cells. *BMC genomics*. 2014;15:583. doi: 10.1186/1471-2164-15-583. PubMed PMID: 25012247; PMCID: 4122776.
52. Meehan RR, Lewis JD, McKay S, Kleiner EL, Bird AP. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell*. 1989;58(3):499-507. doi: 10.1016/0092-8674(89)90430-3. PubMed PMID: 2758464.
53. Baubec T, Ivanek R, Lienert F, Schubeler D. Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell*. 2013;153(2):480-92. doi: 10.1016/j.cell.2013.03.011. PubMed PMID: 23582333.
54. Myant K, Termanis A, Sundaram AY, Boe T, Li C, Merusi C, Burrage J, de Las Heras JJ, Stancheva I. LSH and G9a/GLP complex are required for developmentally programmed DNA methylation. *Genome research*. 2011;21(1):83-94. doi: 10.1101/gr.108498.110. PubMed PMID: 21149390; PMCID: 3012929.
55. Neri F, Rapelli S, Krepelova A, Incarnato D, Parlato C, Basile G, Maldotti M, Anselmi F, Oliviero S. Intragenic DNA methylation prevents spurious transcription initiation. *Nature*. 2017;543(7643):72-7. doi: 10.1038/nature21373. PubMed PMID: 28225755.
56. Baubec T, Colombo DF, Wirbelauer C, Schmidt J, Burger L, Krebs AR, Akalin A, Schubeler D. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature*. 2015;520(7546):243-7. doi: 10.1038/nature14176. PubMed PMID: 25607372.

57. Gelfman S, Cohen N, Yearim A, Ast G. DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure. *Genome research*. 2013;23(5):789-99. doi: 10.1101/gr.143503.112. PubMed PMID: 23502848; PMCID: 3638135.
58. Shayevitch R, Askayo D, Keydar I, Ast G. The importance of DNA methylation of exons on alternative splicing. *Rna*. 2018;24(10):1351-62. doi: 10.1261/rna.064865.117. PubMed PMID: 30002084; PMCID: 6140467.
59. Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian development and disease. *Nature reviews Molecular cell biology*. 2019;20(10):590-607. doi: 10.1038/s41580-019-0159-6. PubMed PMID: 31399642.
60. Holliday R, Grigg GW. DNA methylation and mutation. *Mutation research*. 1993;285(1):61-7. doi: 10.1016/0027-5107(93)90052-h. PubMed PMID: 7678134.
61. Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes. *Journal of molecular biology*. 1987;196(2):261-82. doi: 10.1016/0022-2836(87)90689-9. PubMed PMID: 3656447.
62. Bird A, Taggart M, Frommer M, Miller OJ, Macleod D. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*. 1985;40(1):91-9. doi: 10.1016/0092-8674(85)90312-5. PubMed PMID: 2981636.
63. Larsen F, Gundersen G, Lopez R, Prydz H. CpG islands as gene markers in the human genome. *Genomics*. 1992;13(4):1095-107. doi: 10.1016/0888-7543(92)90024-m. PubMed PMID: 1505946.
64. Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS, Adli M, Kasif S, Ptaszek LM, Cowan CA, Lander ES, Koseki H, Bernstein BE. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS genetics*. 2008;4(10):e1000242. doi: 10.1371/journal.pgen.1000242. PubMed PMID: 18974828; PMCID: 2567431.
65. DNA Methylation: Enzymology. *eLS*. p. 1-10.
66. Ulusu NN. Evolution of Enzyme Kinetic Mechanisms. *Journal of molecular evolution*. 2015;80(5-6):251-7. doi: 10.1007/s00239-015-9681-0. PubMed PMID: 25987355; PMCID: 4472936.
67. Prasad Y, Kumar R, Chaudhary AK, Dhanaraju R, Majumdar S, Rao DN. Kinetic and catalytic properties of M.HpyAXVII, a phase-variable DNA methyltransferase from *Helicobacter pylori*. *The Journal of biological chemistry*. 2019;294(3):1019-34. doi: 10.1074/jbc.RA118.003769. PubMed PMID: 30478171; PMCID: 6341398.
68. Bheemanaik S, Reddy YV, Rao DN. Structure, function and mechanism of exocyclic DNA methyltransferases. *The Biochemical journal*. 2006;399(2):177-90. doi: 10.1042/BJ20060854. PubMed PMID: 16987108; PMCID: 1609917.

69. Posfai J, Bhagwat AS, Posfai G, Roberts RJ. Predictive motifs derived from cytosine methyltransferases. *Nucleic acids research*. 1989;17(7):2421-35. doi: 10.1093/nar/17.7.2421. PubMed PMID: 2717398; PMCID: 317633.
70. Kumar S, Cheng X, Klimasauskas S, Mi S, Posfai J, Roberts RJ, Wilson GG. The DNA (cytosine-5) methyltransferases. *Nucleic acids research*. 1994;22(1):1-10. doi: 10.1093/nar/22.1.1. PubMed PMID: 8127644; PMCID: 307737.
71. Lukashevich OV, Cherepanova NA, Jurkovska RZ, Jeltsch A, Gromova ES. Conserved motif VIII of murine DNA methyltransferase Dnmt3a is essential for methylation activity. *BMC biochemistry*. 2016;17:7. doi: 10.1186/s12858-016-0064-y. PubMed PMID: 27001594; PMCID: 4802922.
72. Hermann A, Gowher H, Jeltsch A. Biochemistry and biology of mammalian DNA methyltransferases. *Cellular and molecular life sciences : CMLS*. 2004;61(19-20):2571-87. doi: 10.1007/s00018-004-4201-1. PubMed PMID: 15526163.
73. Jones PA, Liang G. Rethinking how DNA methylation patterns are maintained. *Nature reviews Genetics*. 2009;10(11):805-11. doi: 10.1038/nrg2651. PubMed PMID: 19789556; PMCID: 2848124.
74. Svedružić ŽM. Chapter 6 - Dnmt1: Structure and Function. In: Cheng X, Blumenthal RM, editors. *Progress in molecular biology and translational science*: Academic Press; 2011. p. 221-54.
75. Jurkowska RZ, Jurkowski TP, Jeltsch A. Structure and function of mammalian DNA methyltransferases. *Chembiochem : a European journal of chemical biology*. 2011;12(2):206-22. doi: 10.1002/cbic.201000195. PubMed PMID: 21243710.
76. Song J, Rechko O, Bestor TH, Patel DJ. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science*. 2011;331(6020):1036-40. doi: 10.1126/science.1195380. PubMed PMID: 21163962; PMCID: 4689315.
77. Syeda F, Fagan RL, Wean M, Avvakumov GV, Walker JR, Xue S, Dhe-Paganon S, Brenner C. The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1. *The Journal of biological chemistry*. 2011;286(17):15344-51. doi: 10.1074/jbc.M110.209882. PubMed PMID: 21389349; PMCID: 3083197.
78. Yarychkivska O, Shahabuddin Z, Comfort N, Boulard M, Bestor TH. BAH domains and a histone-like motif in DNA methyltransferase 1 (DNMT1) regulate de novo and maintenance methylation in vivo. *The Journal of biological chemistry*. 2018;293(50):19466-75. doi: 10.1074/jbc.RA118.004612. PubMed PMID: 30341171; PMCID: 6302165.
79. Yang N, Xu RM. Structure and function of the BAH domain in chromatin biology. *Critical reviews in biochemistry and molecular biology*. 2013;48(3):211-21. doi: 10.3109/10409238.2012.742035. PubMed PMID: 23181513.

80. Hermann A, Goyal R, Jeltsch A. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *The Journal of biological chemistry*. 2004;279(46):48350-9. doi: 10.1074/jbc.M403427200. PubMed PMID: 15339928.
81. Goyal R, Reinhardt R, Jeltsch A. Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase. *Nucleic acids research*. 2006;34(4):1182-8. doi: 10.1093/nar/gkl002. PubMed PMID: 16500889; PMCID: 1383621.
82. Vilkaitis G, Suetake I, Klimasauskas S, Tajima S. Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase. *The Journal of biological chemistry*. 2005;280(1):64-72. doi: 10.1074/jbc.M411126200. PubMed PMID: 15509558.
83. Feltus FA, Lee EK, Costello JF, Plass C, Vertino PM. Predicting aberrant CpG island methylation. *Proceedings of the National Academy of Sciences of the United States of America*. 2003;100(21):12253-8. doi: 10.1073/pnas.2037852100. PubMed PMID: 14519846; PMCID: 218745.
84. Jair KW, Bachman KE, Suzuki H, Ting AH, Rhee I, Yen RW, Baylin SB, Schuebel KE. De novo CpG island methylation in human cancer cells. *Cancer research*. 2006;66(2):682-92. doi: 10.1158/0008-5472.CAN-05-1980. PubMed PMID: 16423997.
85. Fatemi M, Hermann A, Gowher H, Jeltsch A. Dnmt3a and Dnmt1 functionally cooperate during de novo methylation of DNA. *European journal of biochemistry*. 2002;269(20):4981-4. doi: 10.1046/j.1432-1033.2002.03198.x. PubMed PMID: 12383256.
86. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*. 1999;99(3):247-57. doi: 10.1016/s0092-8674(00)81656-6. PubMed PMID: 10555141.
87. Chédin F. Chapter 7 - The DNMT3 Family of Mammalian De Novo DNA Methyltransferases. In: Cheng X, Blumenthal RM, editors. *Progress in molecular biology and translational science*: Academic Press; 2011. p. 255-85.
88. Okano M, Takebayashi S, Okumura K, Li E. Assignment of cytosine-5 DNA methyltransferases Dnmt3a and Dnmt3b to mouse chromosome bands 12A2-A3 and 2H1 by in situ hybridization. *Cytogenetics and cell genetics*. 1999;86(3-4):333-4. doi: 10.1159/000015331. PubMed PMID: 10575238.
89. Watanabe D, Suetake I, Tada T, Tajima S. Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. *Mechanisms of development*. 2002;118(1-2):187-90. PubMed PMID: 12351185.
90. Cheng X, Blumenthal RM. Mammalian DNA methyltransferases: a structural perspective. *Structure*. 2008;16(3):341-50. doi: 10.1016/j.str.2008.01.004. PubMed PMID: 18334209; PMCID: 2597194.

91. Aapola U, Kawasaki K, Scott HS, Ollila J, Vihinen M, Heino M, Shintani A, Kawasaki K, Minoshima S, Krohn K, Antonarakis SE, Shimizu N, Kudoh J, Peterson P. Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. *Genomics*. 2000;65(3):293-8. doi: 10.1006/geno.2000.6168. PubMed PMID: 10857753.
92. Aapola U, Lyle R, Krohn K, Antonarakis SE, Peterson P. Isolation and initial characterization of the mouse Dnmt3l gene. *Cytogenetics and cell genetics*. 2001;92(1-2):122-6. doi: 10.1159/000056881. PubMed PMID: 11306809.
93. Chedin F, Lieber MR, Hsieh CL. The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proceedings of the National Academy of Sciences of the United States of America*. 2002;99(26):16916-21. doi: 10.1073/pnas.262443999. PubMed PMID: 12481029; PMCID: 139244.
94. Lees-Murdock DJ, Shovlin TC, Gardiner T, De Felici M, Walsh CP. DNA methyltransferase expression in the mouse germ line during periods of de novo methylation. *Developmental dynamics : an official publication of the American Association of Anatomists*. 2005;232(4):992-1002. doi: 10.1002/dvdy.20288. PubMed PMID: 15739230.
95. La Salle S, Mertineit C, Taketo T, Moens PB, Bestor TH, Trasler JM. Windows for sex-specific methylation marked by DNA methyltransferase expression profiles in mouse germ cells. *Developmental biology*. 2004;268(2):403-15. doi: 10.1016/j.ydbio.2003.12.031. PubMed PMID: 15063176.
96. Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH. Dnmt3L and the establishment of maternal genomic imprints. *Science*. 2001;294(5551):2536-9. doi: 10.1126/science.1065848. PubMed PMID: 11719692.
97. Hata K, Okano M, Lei H, Li E. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development*. 2002;129(8):1983-93. PubMed PMID: 11934864.
98. Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, Sasaki H. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature*. 2004;429(6994):900-3. doi: 10.1038/nature02633. PubMed PMID: 15215868.
99. Webster KE, O'Bryan MK, Fletcher S, Crewther PE, Aapola U, Craig J, Harrison DK, Aung H, Phutikanit N, Lyle R, Meachem SJ, Antonarakis SE, de Kretser DM, Hedger MP, Peterson P, Carroll BJ, Scott HS. Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. *Proceedings of the National Academy of Sciences of the United States of America*. 2005;102(11):4068-73. doi: 10.1073/pnas.0500702102. PubMed PMID: 15753313; PMCID: 552976.

100. Gowher H, Liebert K, Hermann A, Xu G, Jeltsch A. Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. *The Journal of biological chemistry*. 2005;280(14):13341-8. doi: 10.1074/jbc.M413412200. PubMed PMID: 15671018.
101. Chen T, Tsujimoto N, Li E. 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*. 2004;24(20):9048-58. doi: 10.1128/MCB.24.20.9048-9058.2004. PubMed PMID: 15456878; PMCID: 517890.
102. Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, Jeltsch A. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *The Journal of biological chemistry*. 2010;285(34):26114-20. doi: 10.1074/jbc.M109.089433. PubMed PMID: 20547484; PMCID: 2924014.
103. Wu H, Zeng H, Lam R, Tempel W, Amaya MF, Xu C, Dombrowski L, Qiu W, Wang Y, Min J. Structural and histone binding ability characterizations of human PWWP domains. *PloS one*. 2011;6(6):e18919.
104. Weinberg DN, Papillon-Cavanagh S, Chen H, Yue Y, Chen X, Rajagopalan KN, Horth C, McGuire JT, Xu X, Nikbakht H, Lemiesz AE, Marchione DM, Marunde MR, Meiners MJ, Cheek MA, Keogh MC, Bareke E, Djedid A, Harutyunyan AS, Jabado N, Garcia BA, Li H, Allis CD, Majewski J, Lu C. The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature*. 2019;573(7773):281-6. doi: 10.1038/s41586-019-1534-3. PubMed PMID: 31485078; PMCID: 6742567.
105. Dukatz M, Holzer K, Choudalakis M, Emperle M, Lungu C, Bashtrykov P, Jeltsch A. H3K36me2/3 Binding and DNA Binding of the DNA Methyltransferase DNMT3A PWWP Domain Both Contribute to its Chromatin Interaction. *Journal of molecular biology*. 2019;431(24):5063-74. doi: 10.1016/j.jmb.2019.09.006. PubMed PMID: 31634469.
106. Guo X, Wang L, Li J, Ding Z, Xiao J, Yin X, He S, Shi P, Dong L, Li G, Tian C, Wang J, Cong Y, Xu Y. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature*. 2015;517(7536):640-4. doi: 10.1038/nature13899. PubMed PMID: 25383530.
107. Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD, Cheng X, Bestor TH. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*. 2007;448(7154):714-7. doi: 10.1038/nature05987. PubMed PMID: 17687327; PMCID: 2650820.
108. Gowher H, Jeltsch A. Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases. *The Journal of biological chemistry*. 2002;277(23):20409-14. doi: 10.1074/jbc.M202148200. PubMed PMID: 11919202.
109. Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature*. 2007;449(7159):248-51. doi: 10.1038/nature06146. PubMed PMID: 17713477; PMCID: 2712830.

110. Holz-Schietinger C, Matje DM, Harrison MF, Reich NO. Oligomerization of DNMT3A controls the mechanism of de novo DNA methylation. *The Journal of biological chemistry*. 2011;286(48):41479-88. doi: 10.1074/jbc.M111.284687. PubMed PMID: 21979949; PMCID: 3308859.
111. Holz-Schietinger C, Matje DM, Reich NO. Mutations in DNA methyltransferase (DNMT3A) observed in acute myeloid leukemia patients disrupt processive methylation. *The Journal of biological chemistry*. 2012;287(37):30941-51. doi: 10.1074/jbc.M112.366625. PubMed PMID: 22722925; PMCID: 3438927.
112. Lin CC, Chen YP, Yang WZ, Shen JCK, Yuan HS. Structural insights into CpG-specific DNA methylation by human DNA methyltransferase 3B. *Nucleic acids research*. 2020;48(7):3949-61. doi: 10.1093/nar/gkaa111. PubMed PMID: 32083663; PMCID: 7144912.
113. Barel I, Reich NO, Brown FLH. Integrated rate laws for processive and distributive enzymatic turnover. *The Journal of chemical physics*. 2019;150(24):244120. doi: 10.1063/1.5097576. PubMed PMID: 31255081.
114. Ptashne M. Gene regulation by proteins acting nearby and at a distance. *Nature*. 1986;322(6081):697-701. doi: 10.1038/322697a0. PubMed PMID: 3018583.
115. von Hippel PH, Berg OG. Facilitated target location in biological systems. *The Journal of biological chemistry*. 1989;264(2):675-8. PubMed PMID: 2642903.
116. van Dongen SF, Elemans JA, Rowan AE, Nolte RJ. Processive catalysis. *Angewandte Chemie*. 2014;53(43):11420-8. doi: 10.1002/anie.201404848. PubMed PMID: 25244684.
117. Porter CM, Miller BG. Cooperativity in monomeric enzymes with single ligand-binding sites. *Bioorganic chemistry*. 2012;43:44-50. doi: 10.1016/j.bioorg.2011.11.001. PubMed PMID: 22137502; PMCID: 3307832.
118. Emperle M, Rajavelu A, Reinhardt R, Jurkowska RZ, Jeltsch A. Cooperative DNA binding and protein/DNA fiber formation increases the activity of the Dnmt3a DNA methyltransferase. *The Journal of biological chemistry*. 2014;289(43):29602-13. doi: 10.1074/jbc.M114.572032. PubMed PMID: 25147181; PMCID: 4207976.
119. Jurkowska RZ, Rajavelu A, Anspach N, Urbanke C, Jankevicius G, Ragozin S, Nellen W, Jeltsch A. Oligomerization and binding of the Dnmt3a DNA methyltransferase to parallel DNA molecules: heterochromatic localization and role of Dnmt3L. *The Journal of biological chemistry*. 2011;286(27):24200-7. doi: 10.1074/jbc.M111.254987. PubMed PMID: 21566127; PMCID: 3129201.
120. Rajavelu A, Jurkowska RZ, Fritz J, Jeltsch A. Function and disruption of DNA methyltransferase 3a cooperative DNA binding and nucleoprotein filament formation. *Nucleic acids research*. 2012;40(2):569-80. doi: 10.1093/nar/gkr753. PubMed PMID: 21926161; PMCID: 3258144.

121. Novik KL, Nimmrich I, Genc B, Maier S, Piepenbrock C, Olek A, Beck S. Epigenomics: genome-wide study of methylation phenomena. *Current issues in molecular biology*. 2002;4(4):111-28. PubMed PMID: 12432963.
122. Handa V, Jeltsch A. Profound flanking sequence preference of Dnmt3a and Dnmt3b mammalian DNA methyltransferases shape the human epigenome. *Journal of molecular biology*. 2005;348(5):1103-12. doi: 10.1016/j.jmb.2005.02.044. PubMed PMID: 15854647.
123. Zhang ZM, Lu R, Wang P, Yu Y, Chen D, Gao L, Liu S, Ji D, Rothbart SB, Wang Y, Wang GG, Song J. Structural basis for DNMT3A-mediated de novo DNA methylation. *Nature*. 2018;554(7692):387-91. doi: 10.1038/nature25477. PubMed PMID: 29414941; PMCID: 5814352.
124. Anteneh H, Fang J, Song J. Structural basis for impairment of DNA methylation by the DNMT3A R882H mutation. *Nature communications*. 2020;11(1):2294. doi: 10.1038/s41467-020-16213-9. PubMed PMID: 32385248; PMCID: 7210271.
125. Brunetti L, Gundry MC, Goodell MA. DNMT3A in Leukemia. *Cold Spring Harbor perspectives in medicine*. 2017;7(2). doi: 10.1101/cshperspect.a030320. PubMed PMID: 28003281; PMCID: 5287058.
126. Ehrlich M. The ICF syndrome, a DNA methyltransferase 3B deficiency and immunodeficiency disease. *Clinical immunology*. 2003;109(1):17-28. doi: 10.1016/s1521-6616(03)00201-8. PubMed PMID: 14585272.
127. Hansen RS, Wijmenga C, Luo P, Stanek AM, Canfield TK, Weemaes CM, Gartler SM. The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proceedings of the National Academy of Sciences of the United States of America*. 1999;96(25):14412-7. Epub 1999/12/10. PubMed PMID: 10588719; PMCID: 24450.
128. Jin B, Tao Q, Peng J, Soo HM, Wu W, Ying J, Fields CR, Delmas AL, Liu X, Qiu J, Robertson KD. DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function. *Human molecular genetics*. 2008;17(5):690-709. doi: 10.1093/hmg/ddm341. PubMed PMID: 18029387.
129. Moarefi AH, Chedin F. ICF syndrome mutations cause a broad spectrum of biochemical defects in DNMT3B-mediated de novo DNA methylation. *Journal of molecular biology*. 2011;409(5):758-72. doi: 10.1016/j.jmb.2011.04.050. PubMed PMID: 21549127.
130. Ge YZ, Pu MT, Gowher H, Wu HP, Ding JP, Jeltsch A, Xu GL. Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. *The Journal of biological chemistry*. 2004;279(24):25447-54. doi: 10.1074/jbc.M312296200. PubMed PMID: 14998998.

131. Shirohzu H, Kubota T, Kumazawa A, Sado T, Chijiwa T, Inagaki K, Suetake I, Tajima S, Wakui K, Miki Y, Hayashi M, Fukushima Y, Sasaki H. Three novel DNMT3B mutations in Japanese patients with ICF syndrome. *American journal of medical genetics*. 2002;112(1):31-7. doi: 10.1002/ajmg.10658. PubMed PMID: 12239717.
132. Xie ZH, Huang YN, Chen ZX, Riggs AD, Ding JP, Gowher H, Jeltsch A, Sasaki H, Hata K, Xu GL. Mutations in DNA methyltransferase DNMT3B in ICF syndrome affect its regulation by DNMT3L. *Human molecular genetics*. 2006;15(9):1375-85. doi: 10.1093/hmg/ddl059. PubMed PMID: 16543361.
133. Kanai Y, Ushijima S, Nakanishi Y, Sakamoto M, Hirohashi S. Mutation of the DNA methyltransferase (DNMT) 1 gene in human colorectal cancers. *Cancer letters*. 2003;192(1):75-82. doi: 10.1016/s0304-3835(02)00689-4. PubMed PMID: 12637155.
134. Zhang W, Xu J. DNA methyltransferases and their roles in tumorigenesis. *Biomark Res*. 2017;5:1. doi: 10.1186/s40364-017-0081-z. PubMed PMID: 28127428; PMCID: PMC5251331.
135. Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, Menzies A, Teague JW, Futreal PA, Stratton MR. The Catalogue of Somatic Mutations in Cancer (COSMIC). *Current protocols in human genetics*. 2008;Chapter 10:Unit 10 1. doi: 10.1002/0471142905.hg1011s57. PubMed PMID: 18428421; PMCID: 2705836.
136. Wu H, Zhang Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell*. 2014;156(1-2):45-68. doi: 10.1016/j.cell.2013.12.019. PubMed PMID: 24439369; PMCID: 3938284.
137. Jeong M, Park HJ, Celik H, Ostrander EL, Reyes JM, Guzman A, Rodriguez B, Lei Y, Lee Y, Ding L, Guryanova OA, Li W, Goodell MA, Challen GA. Loss of Dnmt3a immortalizes Hematopoietic Stem Cells In Vivo. *Cell reports*. 2018;23(1):1-10. doi: 10.1016/j.celrep.2018.03.025. PubMed PMID: 29617651; PMCID: PMC5908249.
138. Challen GA, Sun D, Mayle A, Jeong M, Luo M, Rodriguez B, Mallaney C, Celik H, Yang L, Xia Z, Cullen S, Berg J, Zheng Y, Darlington GJ, Li W, Goodell MA. Dnmt3a and Dnmt3b have overlapping and distinct functions in hematopoietic stem cells. *Cell stem cell*. 2014;15(3):350-64. doi: 10.1016/j.stem.2014.06.018. PubMed PMID: 25130491; PMCID: PMC4163922.
139. Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, Kandoth C, Payton JE, Baty J, Welch J, Harris CC, Lichti CF, Townsend RR, Fulton RS, Dooling DJ, Koboldt DC, Schmidt H, Zhang Q, Osborne JR, Lin L, O'Laughlin M, McMichael JF, Delehaunty KD, McGrath SD, Fulton LA, Magrini VJ, Vickery TL, Hundal J, Cook LL, Conyers JJ, Swift GW, Reed JP, Alldredge PA, Wylie T, Walker J, Kalicki J, Watson MA, Heath S, Shannon WD, Varghese N, Nagarajan R, Westervelt P, Tomasson MH, Link DC, Graubert TA, DiPersio JF, Mardis ER, Wilson RK. DNMT3A mutations in acute myeloid leukemia. *The New England journal of medicine*. 2010;363(25):2424-33. doi: 10.1056/NEJMoa1005143. PubMed PMID: 21067377; PMCID: 3201818.

140. Yan XJ, Xu J, Gu ZH, Pan CM, Lu G, Shen Y, Shi JY, Zhu YM, Tang L, Zhang XW, Liang WX, Mi JQ, Song HD, Li KQ, Chen Z, Chen SJ. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nature genetics*. 2011;43(4):309-15. doi: 10.1038/ng.788. PubMed PMID: 21399634.
141. Yuan XQ, Chen P, Du YX, Zhu KW, Zhang DY, Yan H, Liu H, Liu YL, Cao S, Zhou G, Zeng H, Chen SP, Zhao XL, Yang J, Zeng WJ, Chen XP. Influence of DNMT3A R882 mutations on AML prognosis determined by the allele ratio in Chinese patients. *Journal of translational medicine*. 2019;17(1):220. doi: 10.1186/s12967-019-1959-3. PubMed PMID: 31291961; PMCID: 6621981.
142. Russler-Germain DA, Spencer DH, Young MA, Lamprecht TL, Miller CA, Fulton R, Meyer MR, Erdmann-Gilmore P, Townsend RR, Wilson RK, Ley TJ. The R882H DNMT3A mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its ability to form active tetramers. *Cancer cell*. 2014;25(4):442-54. doi: 10.1016/j.ccr.2014.02.010. PubMed PMID: 24656771; PMCID: 4018976.
143. Kim SJ, Zhao H, Hardikar S, Singh AK, Goodell MA, Chen T. A DNMT3A mutation common in AML exhibits dominant-negative effects in murine ES cells. *Blood*. 2013;122(25):4086-9. doi: 10.1182/blood-2013-02-483487. PubMed PMID: 24167195; PMCID: 3952368.
144. Emperle M, Rajavelu A, Kunert S, Arimondo PB, Reinhardt R, Jurkowska RZ, Jeltsch A. The DNMT3A R882H mutant displays altered flanking sequence preferences. *Nucleic acids research*. 2018;46(6):3130-9. Epub 2018/03/09. doi: 10.1093/nar/gky168. PubMed PMID: 29518238; PMCID: 5887309.
145. Emperle M, Dukatz M, Kunert S, Holzer K, Rajavelu A, Jurkowska RZ, Jeltsch A. The DNMT3A R882H mutation does not cause dominant negative effects in purified mixed DNMT3A/R882H complexes. *Scientific reports*. 2018;8(1):13242. doi: 10.1038/s41598-018-31635-8. PubMed PMID: 30185810; PMCID: 6125428.
146. Emperle M, Adam S, Kunert S, Dukatz M, Baude A, Plass C, Rathert P, Bashtrykov P, Jeltsch A. Mutations of R882 change flanking sequence preferences of the DNA methyltransferase DNMT3A and cellular methylation patterns. *Nucleic acids research*. 2019;47(21):11355-67. doi: 10.1093/nar/gkz911. PubMed PMID: 31620784; PMCID: 6868496.
147. Sandoval JE, Huang YH, Muise A, Goodell MA, Reich NO. Mutations in the DNMT3A DNA methyltransferase in acute myeloid leukemia patients cause both loss and gain of function and differential regulation by protein partners. *The Journal of biological chemistry*. 2019;294(13):4898-910. doi: 10.1074/jbc.RA118.006795. PubMed PMID: 30705090; PMCID: 6442042.
148. Jurkowska RZ, Ceccaldi A, Zhang Y, Arimondo PB, Jeltsch A. DNA Methyltransferase Assays. In: Tollefsbol TO, editor. *Epigenetics Protocols*. Totowa, NJ: Humana Press; 2011. p. 157-77.

149. Grunau C, Clark SJ, Rosenthal A. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic acids research*. 2001;29(13):E65-5. doi: 10.1093/nar/29.13.e65. PubMed PMID: 11433041; PMCID: 55789.
150. Zhang Y, Rohde C, Tierling S, Stamerjohanns H, Reinhardt R, Walter J, Jeltsch A. DNA methylation analysis by bisulfite conversion, cloning, and sequencing of individual clones. *Methods in molecular biology*. 2009;507:177-87. doi: 10.1007/978-1-59745-522-0_14. PubMed PMID: 18987815.
151. Gu H, Bock C, Mikkelsen TS, Jäger N, Smith ZD, Tomazou E, Gnirke A, Lander ES, Meissner A. Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution. *Nat Methods*. 2010;7(2):133-6. Epub 2010/01/10. doi: 10.1038/nmeth.1414. PubMed PMID: 20062050; PMCID: PMC2860480.
152. Clark SJ, Smallwood SA, Lee HJ, Krueger F, Reik W, Kelsey G. Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). *Nat Protoc*. 2017;12(3):534-47. Epub 2017/02/09. doi: 10.1038/nprot.2016.187. PubMed PMID: 28182018.
153. Roth M, Jeltsch A. Biotin-avidin microplate assay for the quantitative analysis of enzymatic methylation of DNA by DNA methyltransferases. *Biological chemistry*. 2000;381(3):269-72. doi: 10.1515/BC.2000.035. PubMed PMID: 10782999.
154. Liebert K, Jeltsch A. Detection and quantitation of the activity of DNA methyltransferases using a biotin/avidin microplate assay. *Methods in molecular biology*. 2008;418:149-56. doi: 10.1007/978-1-59745-579-4_13. PubMed PMID: 18287657.
155. Jeltsch A, Gowher H. Editorial-Role of DNA Methyltransferases in the Epigenome. *Genes*. 2019;10(8). doi: 10.3390/genes10080574. PubMed PMID: 31366147; PMCID: 6723798.
156. Cedar H, Bergman Y. Programming of DNA methylation patterns. *Annual review of biochemistry*. 2012;81:97-117. doi: 10.1146/annurev-biochem-052610-091920. PubMed PMID: 22404632.
157. Nishiyama A, Yamaguchi L, Nakanishi M. Regulation of maintenance DNA methylation via histone ubiquitylation. *Journal of biochemistry*. 2016;159(1):9-15. doi: 10.1093/jb/mvv113. PubMed PMID: 26590302; PMCID: 4882649.
158. Schubeler D. ESCI award lecture: regulation, function and biomarker potential of DNA methylation. *European journal of clinical investigation*. 2015;45(3):288-93. doi: 10.1111/eci.12403. PubMed PMID: 25608229.
159. Chen BF, Chan WY. The de novo DNA methyltransferase DNMT3A in development and cancer. *Epigenetics*. 2014;9(5):669-77. doi: 10.4161/epi.28324. PubMed PMID: 24589714; PMCID: 4063825.

160. Yang L, Rau R, Goodell MA. DNMT3A in haematological malignancies. *Nature reviews Cancer*. 2015;15(3):152-65. doi: 10.1038/nrc3895. PubMed PMID: 25693834; PMCID: 5814392.
161. Jeong M, Goodell MA. New answers to old questions from genome-wide maps of DNA methylation in hematopoietic cells. *Experimental hematology*. 2014;42(8):609-17. doi: 10.1016/j.exphem.2014.04.008. PubMed PMID: 24993071; PMCID: 4137036.
162. Walton EL, Francastel C, Velasco G. Dnmt3b Prefers Germ Line Genes and Centromeric Regions: Lessons from the ICF Syndrome and Cancer and Implications for Diseases. *Biology*. 2014;3(3):578-605. doi: 10.3390/biology3030578. PubMed PMID: 25198254; PMCID: 4192629.
163. Duan F, Cui S, Song C, Dai L, Zhao X, Zhang X. Systematic evaluation of cancer risk associated with DNMT3B polymorphisms. *Journal of cancer research and clinical oncology*. 2015;141(7):1205-20. doi: 10.1007/s00432-014-1894-x. PubMed PMID: 25515408.
164. Karetka MS, Botello ZM, Ennis JJ, Chou C, Chedin F. Reconstitution and mechanism of the stimulation of de novo methylation by human DNMT3L. *The Journal of biological chemistry*. 2006;281(36):25893-902. doi: 10.1074/jbc.M603140200. PubMed PMID: 16829525.
165. Jurkowska RZ, Anspach N, Urbanke C, Jia D, Reinhardt R, Nellen W, Cheng X, Jeltsch A. Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L. *Nucleic acids research*. 2008;36(21):6656-63. doi: 10.1093/nar/gkn747. PubMed PMID: 18945701; PMCID: 2588524.
166. Holz-Schietinger C, Reich NO. The inherent processivity of the human de novo methyltransferase 3A (DNMT3A) is enhanced by DNMT3L. *The Journal of biological chemistry*. 2010;285(38):29091-100. doi: 10.1074/jbc.M110.142513. PubMed PMID: 20630873; PMCID: 2937940.
167. Gowher H, Jeltsch A. Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG [correction of non-CpA] sites. *Journal of molecular biology*. 2001;309(5):1201-8. doi: 10.1006/jmbi.2001.4710. PubMed PMID: 11399089.
168. Renbaum P, Razin A. Mode of action of the Spiroplasma CpG methylase M.SssI. *FEBS letters*. 1992;313(3):243-7. doi: 10.1016/0014-5793(92)81201-v. PubMed PMID: 1446743.
169. Holz-Schietinger C, Reich NO. De novo DNA methyltransferase DNMT3A: Regulation of oligomeric state and mechanism of action in response to pH changes. *Biochimica et biophysica acta*. 2015;1850(6):1131-9. doi: 10.1016/j.bbagen.2015.02.003. PubMed PMID: 25681155.

170. Gowher H, Loutchanwoot P, Vorobjeva O, Handa V, Jurkowska RZ, Jurkowski TP, Jeltsch A. Mutational analysis of the catalytic domain of the murine Dnmt3a DNA-(cytosine C5)-methyltransferase. *Journal of molecular biology*. 2006;357(3):928-41. doi: 10.1016/j.jmb.2006.01.035. PubMed PMID: 16472822.
171. Rubin RA, Modrich P. EcoRI methylase. Physical and catalytic properties of the homogeneous enzyme. *The Journal of biological chemistry*. 1977;252(20):7265-72. PubMed PMID: 332688.
172. Rhee I, Bachman KE, Park BH, Jair KW, Yen RW, Schuebel KE, Cui H, Feinberg AP, Lengauer C, Kinzler KW, Baylin SB, Vogelstein B. DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature*. 2002;416(6880):552-6. doi: 10.1038/416552a. PubMed PMID: 11932749.
173. Smallbone K, Gavaghan DJ, Gatenby RA, Maini PK. The role of acidity in solid tumour growth and invasion. *Journal of theoretical biology*. 2005;235(4):476-84. doi: 10.1016/j.jtbi.2005.02.001. PubMed PMID: 15935166.
174. Du J, Johnson LM, Jacobsen SE, Patel DJ. DNA methylation pathways and their crosstalk with histone methylation. *Nature reviews Molecular cell biology*. 2015;16(9):519-32. doi: 10.1038/nrm4043. PubMed PMID: 26296162; PMCID: 4672940.
175. Tiedemann RL, Putiri EL, Lee JH, Hlady RA, Kashiwagi K, Ordog T, Zhang Z, Liu C, Choi JH, Robertson KD. Acute depletion redefines the division of labor among DNA methyltransferases in methylating the human genome. *Cell reports*. 2014;9(4):1554-66. doi: 10.1016/j.celrep.2014.10.013. PubMed PMID: 25453758; PMCID: 4254561.
176. Shih AH, Abdel-Wahab O, Patel JP, Levine RL. The role of mutations in epigenetic regulators in myeloid malignancies. *Nature reviews Cancer*. 2012;12(9):599-612. doi: 10.1038/nrc3343. PubMed PMID: 22898539.
177. Walter MJ, Ding L, Shen D, Shao J, Grillot M, McLellan M, Fulton R, Schmidt H, Kalicki-Veizer J, O'Laughlin M, Kandoth C, Baty J, Westervelt P, DiPersio JF, Mardis ER, Wilson RK, Ley TJ, Graubert TA. Recurrent DNMT3A mutations in patients with myelodysplastic syndromes. *Leukemia*. 2011;25(7):1153-8. doi: 10.1038/leu.2011.44. PubMed PMID: 21415852; PMCID: 3202965.
178. Xu GL, Bestor TH, Bourc'his D, Hsieh CL, Tommerup N, Bugge M, Hulten M, Qu X, Russo JJ, Viegas-Pequignot E. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature*. 1999;402(6758):187-91. doi: 10.1038/46052. PubMed PMID: 10647011.
179. Ehrlich M, Buchanan KL, Tsien F, Jiang G, Sun B, Uicker W, Weemaes CM, Smeets D, Sperling K, Belohradsky BH, Tommerup N, Misek DE, Rouillard JM, Kuick R, Hanash SM. DNA methyltransferase 3B mutations linked to the ICF syndrome cause dysregulation of lymphogenesis genes. *Human molecular genetics*. 2001;10(25):2917-31. doi: 10.1093/hmg/10.25.2917. PubMed PMID: 11741835.

180. Hamidi T, Singh AK, Chen T. Genetic alterations of DNA methylation machinery in human diseases. *Epigenomics*. 2015;7(2):247-65. doi: 10.2217/epi.14.80. PubMed PMID: 25942534.
181. Miller CA, Wilson RK, Ley TJ. Genomic landscapes and clonality of de novo AML. *The New England journal of medicine*. 2013;369(15):1473. doi: 10.1056/NEJMc1308782. PubMed PMID: 24106950; PMCID: 4374653.
182. Reik W, Dean W, Walter J. Epigenetic reprogramming in mammalian development. *Science*. 2001;293(5532):1089-93. doi: 10.1126/science.1063443. PubMed PMID: 11498579.
183. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. *Nature reviews Genetics*. 2013;14(3):204-20. doi: 10.1038/nrg3354. PubMed PMID: 23400093.
184. Jaenisch R, Jahner D. Methylation, Expression and Chromosomal Position of Genes in Mammals. *Biochimica et biophysica acta*. 1984;782(1):1-9. doi: Doi 10.1016/0167-4781(84)90099-X. PubMed PMID: WOS:A1984ST80900001.
185. Surani MA. Imprinting and the initiation of gene silencing in the germ line. *Cell*. 1998;93(3):309-12. PubMed PMID: 9590162.
186. Ng HH, Bird A. DNA methylation and chromatin modification. *Current opinion in genetics & development*. 1999;9(2):158-63. PubMed PMID: 10322130.
187. Lyko F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nature reviews Genetics*. 2018;19(2):81-92. doi: 10.1038/nrg.2017.80. PubMed PMID: 29033456.
188. Bestor T, Laudano A, Mattaliano R, Ingram V. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *Journal of molecular biology*. 1988;203(4):971-83. doi: 10.1016/0022-2836(88)90122-2. PubMed PMID: 3210246.
189. Yoder JA, Soman NS, Verdine GL, Bestor TH. DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. *Journal of molecular biology*. 1997;270(3):385-95. doi: 10.1006/jmbi.1997.1125. PubMed PMID: 9237905.
190. Chen T, Ueda Y, Dodge JE, Wang Z, Li E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Molecular and cellular biology*. 2003;23(16):5594-605. doi: 10.1128/mcb.23.16.5594-5605.2003. PubMed PMID: 12897133; PMCID: 166327.
191. Gujar H, Weisenberger DJ, Liang G. The Roles of Human DNA Methyltransferases and Their Isoforms in Shaping the Epigenome. *Genes*. 2019;10(2). doi: 10.3390/genes10020172. PubMed PMID: 30813436; PMCID: 6409524.

192. Linhart HG, Lin H, Yamada Y, Moran E, Steine EJ, Gokhale S, Lo G, Cantu E, Ehrich M, He T, Meissner A, Jaenisch R. Dnmt3b promotes tumorigenesis in vivo by gene-specific de novo methylation and transcriptional silencing. *Genes & development*. 2007;21(23):3110-22. doi: 10.1101/gad.1594007. PubMed PMID: 18056424; PMCID: 2081977.
193. Norvil AB, Saha D, Saleem Dar M, Gowher H. Effect of Disease-Associated Germline Mutations on Structure Function Relationship of DNA Methyltransferases. *Genes*. 2019;10(5). doi: 10.3390/genes10050369. PubMed PMID: 31091831; PMCID: 6562416.
194. Cancer Genome Atlas Research N, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson A, Hoadley K, Triche TJ, Jr., Laird PW, Baty JD, Fulton LL, Fulton R, Heath SE, Kalicki-Veizer J, Kandoth C, Kline JM, Koboldt DC, Kanchi KL, Kulkarni S, Lamprecht TL, Larson DE, Lin L, Lu C, McLellan MD, McMichael JF, Payton J, Schmidt H, Spencer DH, Tomasson MH, Wallis JW, Wartman LD, Watson MA, Welch J, Wendl MC, Ally A, Balasundaram M, Birol I, Butterfield Y, Chiu R, Chu A, Chuah E, Chun HJ, Corbett R, Dhallan N, Guin R, He A, Hirst C, Hirst M, Holt RA, Jones S, Karsan A, Lee D, Li HI, Marra MA, Mayo M, Moore RA, Mungall K, Parker J, Pleasance E, Plettner P, Schein J, Stoll D, Swanson L, Tam A, Thiessen N, Varhol R, Wye N, Zhao Y, Gabriel S, Getz G, Sougnez C, Zou L, Leiserson MD, Vandin F, Wu HT, Applebaum F, Baylin SB, Akbani R, Broom BM, Chen K, Motter TC, Nguyen K, Weinstein JN, Zhang N, Ferguson ML, Adams C, Black A, Bowen J, Gastier-Foster J, Grossman T, Lichtenberg T, Wise L, Davidsen T, Demchok JA, Shaw KR, Sheth M, Sofia HJ, Yang L, Downing JR, Eley G. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *The New England journal of medicine*. 2013;368(22):2059-74. doi: 10.1056/NEJMoa1301689. PubMed PMID: 23634996; PMCID: 3767041.
195. Qu Y, Lennartsson A, Gaidzik VI, Deneberg S, Karimi M, Bengtzen S, Hoglund M, Bullinger L, Dohner K, Lehmann S. Differential methylation in CN-AML preferentially targets non-CGI regions and is dictated by DNMT3A mutational status and associated with predominant hypomethylation of HOX genes. *Epigenetics*. 2014;9(8):1108-19. doi: 10.4161/epi.29315. PubMed PMID: 24866170; PMCID: 4164496.
196. Yang L, Rodriguez B, Mayle A, Park HJ, Lin X, Luo M, Jeong M, Curry CV, Kim SB, Ruau D, Zhang X, Zhou T, Zhou M, Rebel VI, Challen GA, Gottgens B, Lee JS, Rau R, Li W, Goodell MA. DNMT3A Loss Drives Enhancer Hypomethylation in FLT3-ITD-Associated Leukemias. *Cancer cell*. 2016;29(6):922-34. doi: 10.1016/j.ccell.2016.05.003. PubMed PMID: 27300438; PMCID: 4908977.
197. Gagliardi M, Strazzullo M, Matarazzo MR. DNMT3B Functions: Novel Insights From Human Disease. *Frontiers in cell and developmental biology*. 2018;6:140. doi: 10.3389/fcell.2018.00140. PubMed PMID: 30406101; PMCID: 6204409.

198. Niederwieser C, Kohlschmidt J, Volinia S, Whitman SP, Metzeler KH, Eisfeld AK, Maharry K, Yan P, Frankhouser D, Becker H, Schwind S, Carroll AJ, Nicolet D, Mendler JH, Curfman JP, Wu YZ, Baer MR, Powell BL, Kolitz JE, Moore JO, Carter TH, Bundschuh R, Larson RA, Stone RM, Mrozek K, Marcucci G, Bloomfield CD. Prognostic and biologic significance of DNMT3B expression in older patients with cytogenetically normal primary acute myeloid leukemia. *Leukemia*. 2015;29(3):567-75. doi: 10.1038/leu.2014.267. PubMed PMID: 25204569; PMCID: 4351165.
199. Hayette S, Thomas X, Jallades L, Chabane K, Charlot C, Tigaud I, Gazzo S, Morisset S, Cornillet-Lefebvre P, Plesa A, Huet S, Renneville A, Salles G, Nicolini FE, Magaud JP, Michallet M. High DNA methyltransferase DNMT3B levels: a poor prognostic marker in acute myeloid leukemia. *PloS one*. 2012;7(12):e51527. doi: 10.1371/journal.pone.0051527. PubMed PMID: 23251566; PMCID: 3519733.
200. Yamashita Y, Yuan J, Suetake I, Suzuki H, Ishikawa Y, Choi YL, Ueno T, Soda M, Hamada T, Haruta H, Takada S, Miyazaki Y, Kiyoi H, Ito E, Naoe T, Tomonaga M, Toyota M, Tajima S, Iwama A, Mano H. Array-based genomic resequencing of human leukemia. *Oncogene*. 2010;29(25):3723-31. doi: 10.1038/onc.2010.117. PubMed PMID: 20400977.
201. Norvil AB, Petell CJ, Alabdi L, Wu L, Rossie S, Gowher H. Dnmt3b Methylates DNA by a Noncooperative Mechanism, and Its Activity Is Unaffected by Manipulations at the Predicted Dimer Interface. *Biochemistry*. 2018;57(29):4312-24. Epub 2016/10/22. doi: 10.1021/acs.biochem.6b00964. PubMed PMID: 27768276; PMCID: 5992102.
202. Hemeon I, Gutierrez JA, Ho MC, Schramm VL. Characterizing DNA methyltransferases with an ultrasensitive luciferase-linked continuous assay. *Analytical chemistry*. 2011;83(12):4996-5004. doi: 10.1021/ac200816m. PubMed PMID: 21545095; PMCID: 3115440.
203. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. 2004;14(6):1188-90. doi: 10.1101/gr.849004. PubMed PMID: 15173120; PMCID: PMC419797.
204. Lu R, Wang P, Parton T, Zhou Y, Chrysovergis K, Rockowitz S, Chen WY, Abdel-Wahab O, Wade PA, Zheng D, Wang GG. Epigenetic Perturbations by Arg882-Mutated DNMT3A Potentiate Aberrant Stem Cell Gene-Expression Program and Acute Leukemia Development. *Cancer cell*. 2016;30(1):92-107. doi: 10.1016/j.ccell.2016.05.008. PubMed PMID: 27344947; PMCID: 4945461.
205. Mercader N, Leonardo E, Azpiazu N, Serrano A, Morata G, Martinez C, Torres M. Conserved regulation of proximodistal limb axis development by Meis1/Hth. *Nature*. 1999;402(6760):425-9. doi: 10.1038/46580. PubMed PMID: 10586884.
206. Cai M, Langer EM, Gill JG, Satpathy AT, Albring JC, Kc W, Murphy TL, Murphy KM. Dual actions of Meis1 inhibit erythroid progenitor development and sustain general hematopoietic cell proliferation. *Blood*. 2012;120(2):335-46. doi: 10.1182/blood-2012-01-403139. PubMed PMID: 22665933; PMCID: 3628121.

207. Ferreira HJ, Heyn H, Vizoso M, Moutinho C, Vidal E, Gomez A, Martinez-Cardus A, Simo-Riudalbas L, Moran S, Jost E, Esteller M. DNMT3A mutations mediate the epigenetic reactivation of the leukemogenic factor MEIS1 in acute myeloid leukemia. *Oncogene*. 2016;35(23):3079-82. doi: 10.1038/onc.2015.359. PubMed PMID: 26434589; PMCID: 4705435.
208. Chen T, Ueda Y, Xie S, Li E. A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation. *The Journal of biological chemistry*. 2002;277(41):38746-54. doi: 10.1074/jbc.M205312200. PubMed PMID: 12138111.
209. Mayle A, Yang L, Rodriguez B, Zhou T, Chang E, Curry CV, Challen GA, Li W, Wheeler D, Rebel VI, Goodell MA. Dnmt3a loss predisposes murine hematopoietic stem cells to malignant transformation. *Blood*. 2015;125(4):629-38. doi: 10.1182/blood-2014-08-594648. PubMed PMID: 25416277; PMCID: 4304108.
210. Wu X, Zhang Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nature Reviews Genetics*. 2017;18:517. doi: 10.1038/nrg.2017.33.
211. Li J, Wu X, Zhou Y, Lee M, Guo L, Han W, Mo W, Cao WM, Sun D, Xie R, Huang Y. Decoding the dynamic DNA methylation and hydroxymethylation landscapes in endodermal lineage intermediates during pancreatic differentiation of hESC. *Nucleic acids research*. 2018;46(6):2883-900. Epub 2018/02/03. doi: 10.1093/nar/gky063. PubMed PMID: 29394393; PMCID: 5888657.
212. Mahe EA, Madigou T, Serandour AA, Bizot M, Avner S, Chalmel F, Palierne G, Metivier R, Salbert G. Cytosine modifications modulate the chromatin architecture of transcriptional enhancers. *Genome research*. 2017;27(6):947-58. Epub 2017/04/12. doi: 10.1101/gr.211466.116. PubMed PMID: 28396520; PMCID: 5453328.
213. Hackett JA, Dietmann S, Murakami K, Down TA, Leitch HG, Surani MA. Synergistic mechanisms of DNA demethylation during transition to ground-state pluripotency. *Stem cell reports*. 2013;1(6):518-31. Epub 2013/12/29. doi: 10.1016/j.stemcr.2013.11.010. PubMed PMID: 24371807; PMCID: 3871394.
214. Cheng X. Structure and function of DNA methyltransferases. *Annual review of biophysics and biomolecular structure*. 1995;24:293-318. doi: 10.1146/annurev.bb.24.060195.001453. PubMed PMID: 7663118.
215. Gowher H, Jeltsch A. Mammalian DNA methyltransferases: new discoveries and open questions. *Biochemical Society transactions*. 2018;46(5):1191-202. doi: 10.1042/BST20170574. PubMed PMID: 30154093; PMCID: 6581191.
216. Bestor TH. The DNA methyltransferases of mammals. *Human molecular genetics*. 2000;9(16):2395-402. Epub 2000/09/27. PubMed PMID: 11005794.

217. Mortusewicz O, Schermelleh L, Walter J, Cardoso MC, Leonhardt H. Recruitment of DNA methyltransferase I to DNA repair sites. *Proceedings of the National Academy of Sciences*. 2005;102(25):8905-9.
218. Chen T, Li E. Establishment and maintenance of DNA methylation patterns in mammals. *DNA Methylation: Basic Mechanisms*: Springer; 2006. p. 179-201.
219. Jurkowska RZ, Jeltsch A. Enzymology of Mammalian DNA Methyltransferases. *Advances in experimental medicine and biology*. 2016;945:87-122. Epub 2016/11/09. doi: 10.1007/978-3-319-43624-1_5. PubMed PMID: 27826836.
220. Tajima S, Suetake I, Takeshita K, Nakagawa A, Kimura H. Domain Structure of the Dnmt1, Dnmt3a, and Dnmt3b DNA Methyltransferases. *Advances in experimental medicine and biology*. 2016;945:63-86. Epub 2016/11/09. doi: 10.1007/978-3-319-43624-1_4. PubMed PMID: 27826835.
221. Jones PA, Issa JP, Baylin S. Targeting the cancer epigenome for therapy. *Nature reviews Genetics*. 2016;17(10):630-41. doi: 10.1038/nrg.2016.93. PubMed PMID: 27629931.
222. Velasco G, Francastel C. Genetics meets DNA methylation in rare diseases. *Clinical genetics*. 2019;95(2):210-20. Epub 2018/11/21. doi: 10.1111/cge.13480. PubMed PMID: 30456829.
223. Sandoval JE, Huang Y-H, Muise A, Goodell MA, Reich NO. Mutations in the DNMT3A DNA methyltransferase in AML patients cause both loss and gain of function and differential regulation by protein partners. *Journal of Biological Chemistry*. 2019:jbc.RA118. 006795.
224. Norvil AB, Petell CJ, Alabdi L, Wu L, Rossie S, Gowher H. Dnmt3b methylates DNA by a noncooperative mechanism, and its activity is unaffected by manipulations at the predicted dimer interface. *Biochemistry*. 2016;57(29):4312-24.
225. Holz-Schietinger C, Matje DM, Reich NO. Mutations in DNA methyltransferase (DNMT3A) observed in acute myeloid leukemia patients disrupt processive methylation. *Journal of Biological Chemistry*. 2012;287(37):30941-51.
226. Song J, Rechko O, Bestor TH, Patel DJ. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science*. 2011;331(6020):1036-40.
227. Bostick M, Kim JK, Estève P-O, Clark A, Pradhan S, Jacobsen SE. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science*. 2007;317(5845):1760-4.
228. Svedruzic ZM. Dnmt1 structure and function. *Progress in molecular biology and translational science*. 2011;101:221-54. Epub 2011/04/22. doi: 10.1016/b978-0-12-387685-0.00006-8. PubMed PMID: 21507353.

229. Baets J, Duan X, Wu Y, Smith G, Seeley WW, Mademan I, McGrath NM, Beadell NC, Khoury J, Botuyan M-V. Defects of mutant DNMT1 are linked to a spectrum of neurological disorders. *Brain*. 2015;138(4):845-61.
230. Winkelmann J, Lin L, Schormair B, Kornum BR, Faraco J, Plazzi G, Melberg A, Cornelio F, Urban AE, Pizza F, Poli F, Grubert F, Wieland T, Graf E, Hallmayer J, Strom TM, Mignot E. Mutations in DNMT1 cause autosomal dominant cerebellar ataxia, deafness and narcolepsy. *Human molecular genetics*. 2012;21(10):2205-10. doi: 10.1093/hmg/dds035. PubMed PMID: 22328086; PMCID: 3465691.
231. Klein CJ, Botuyan MV, Wu Y, Ward CJ, Nicholson GA, Hammans S, Hojo K, Yamanishi H, Karpf AR, Wallace DC, Simon M, Lander C, Boardman LA, Cunningham JM, Smith GE, Litchy WJ, Boes B, Atkinson EJ, Middha S, PJ BD, Parisi JE, Mer G, Smith DI, Dyck PJ. Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss. *Nature genetics*. 2011;43(6):595-600. Epub 2011/05/03. doi: 10.1038/ng.830. PubMed PMID: 21532572; PMCID: PMC3102765.
232. Sun Z, Wu Y, Ordog T, Baheti S, Nie J, Duan X, Hojo K, Kocher JP, Dyck PJ, Klein CJ. Aberrant signature methylome by DNMT1 hot spot mutation in hereditary sensory and autonomic neuropathy 1E. *Epigenetics*. 2014;9(8):1184-93. Epub 2014/07/18. doi: 10.4161/epi.29676. PubMed PMID: 25033457; PMCID: PMC4164503.
233. Wu BK, Mei SC, Brenner C. RFTS-deleted DNMT1 enhances tumorigenicity with focal hypermethylation and global hypomethylation. *Cell Cycle*. 2014;13(20):3222-31. Epub 2014/12/09. doi: 10.4161/15384101.2014.950886. PubMed PMID: 25485502; PMCID: 4615144.
234. Smets M, Link S, Wolf P, Schneider K, Solis V, Ryan J, Meilinger D, Qin W, Leonhardt H. DNMT1 mutations found in HSANIE patients affect interaction with UHRF1 and neuronal differentiation. *Human molecular genetics*. 2017;26(8):1522-34.
235. Klein CJ, Bird T, Ertekin-Taner N, Lincoln S, Hjorth R, Wu Y, Kwok J, Mer G, Dyck PJ, Nicholson GA. DNMT1 mutation hot spot causes varied phenotypes of HSAN1 with dementia and hearing loss. *Neurology*. 2013;80(9):824-8.
236. Moghadam KK, Pizza F, La Morgia C, Franceschini C, Tonon C, Lodi R, Barboni P, Seri M, Ferrari S, Liguori R. Narcolepsy is a common phenotype in HSAN IE and ADCA-DN. *Brain*. 2014;137(6):1643-55.
237. Yuan J, Higuchi Y, Nagado T, Nozuma S, Nakamura T, Matsuura E, Hashiguchi A, Sakiyama Y, Yoshimura A, Takashima H. Novel mutation in the replication focus targeting sequence domain of DNMT1 causes hereditary sensory and autonomic neuropathy IE. *Journal of the peripheral nervous system : JPNS*. 2013;18(1):89-93. Epub 2013/03/26. doi: 10.1111/jns5.12012. PubMed PMID: 23521649.
238. Zhang Z-M, Liu S, Lin K, Luo Y, Perry JJ, Wang Y, Song J. Crystal structure of human DNA methyltransferase 1. *Journal of molecular biology*. 2015;427(15):2520-31.

239. Li T, Wang L, Du Y, Xie S, Yang X, Lian F, Zhou Z, Qian C. Structural and mechanistic insights into UHRF1-mediated DNMT1 activation in the maintenance DNA methylation. *Nucleic acids research*. 2018;46(6):3218-31.
240. Fan G, Beard C, Chen RZ, Csankovszki G, Sun Y, Siniaia M, Biniszkiewicz D, Bates B, Lee PP, Kuhn R, Trumpp A, Poon C, Wilson CB, Jaenisch R. DNA hypomethylation perturbs the function and survival of CNS neurons in postnatal animals. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 2001;21(3):788-97. Epub 2001/02/07. PubMed PMID: 11157065.
241. Pedroso JL, Povoas Barsottini OG, Lin L, Melberg A, Oliveira AS, Mignot E. A novel de novo exon 21 DNMT1 mutation causes cerebellar ataxia, deafness, and narcolepsy in a Brazilian patient. *Sleep*. 2013;36(8):1257-9.
242. Kernohan KD, Cigana Schenkel L, Huang L, Smith A, Pare G, Ainsworth P, Boycott KM, Warman-Chardon J, Sadikovic B. Identification of a methylation profile for DNMT1-associated autosomal dominant cerebellar ataxia, deafness, and narcolepsy. *Clinical epigenetics*. 2016;8:91. Epub 2016/09/08. doi: 10.1186/s13148-016-0254-x. PubMed PMID: 27602171; PMCID: 5011850.
243. Mastroeni D, Grover A, Delvaux E, Whiteside C, Coleman PD, Rogers J. Epigenetic changes in Alzheimer's disease: decrements in DNA methylation. *Neurobiology of aging*. 2010;31(12):2025-37.
244. Yokoyama AS, Rutledge JC, Medici V. DNA methylation alterations in Alzheimer's disease. *Environmental epigenetics*. 2017;3(2):dvx008.
245. Jowaed A, Schmitt I, Kaut O, Wüllner U. Methylation regulates alpha-synuclein expression and is decreased in Parkinson's disease patients' brains. *Journal of Neuroscience*. 2010;30(18):6355-9.
246. Desplats P, Spencer B, Coffee E, Patel P, Michael S, Patrick C, Adame A, Rockenstein E, Masliah E. α -Synuclein Sequesters Dnmt1 from the Nucleus A Novel Mechanism For Epigenetic Alterations In Lewy Body Diseases. *Journal of Biological Chemistry*. 2011;286(11):9031-7.
247. Wüllner U, Kaut O, deBoni L, Piston D, Schmitt I. DNA methylation in Parkinson's disease. *Journal of neurochemistry*. 2016;139:108-20.
248. Weissman J, Naidu S, Bjornsson HT. Abnormalities of the DNA methylation mark and its machinery: an emerging cause of neurologic dysfunction. *Seminars in neurology*. 2014;34(3):249-57. Epub 2014/09/06. doi: 10.1055/s-0034-1386763. PubMed PMID: 25192503; PMCID: PMC4512289.
249. Forbes S, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, Menzies A, Teague J, Futreal P, Stratton M. The catalogue of somatic mutations in cancer (COSMIC). *Current protocols in human genetics*. 2008;57(1):10.1. 1-1. 26.

250. Cheng X, Blumenthal RM. Mammalian DNA methyltransferases: a structural perspective. *Structure*. 2008;16(3):341-50.
251. Guo X, Wang L, Li J, Ding Z, Xiao J, Yin X, He S, Shi P, Dong L, Li G. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature*. 2015;517(7536):640.
252. Petell CJ, Alabdi L, He M, San Miguel P, Rose R, Gowher H. An epigenetic switch regulates de novo DNA methylation at a subset of pluripotency gene enhancers during embryonic stem cell differentiation. *Nucleic acids research*. 2016;44(16):7605-17. doi: 10.1093/nar/gkw426. PubMed PMID: 27179026; PMCID: 5027477.
253. Gowher H, Loutchanwoot P, Vorobjeva O, Handa V, Jurkowska RZ, Jurkowski TP, Jeltsch A. Mutational analysis of the catalytic domain of the murine Dnmt3a DNA-(cytosine C5)-methyltransferase. *Journal of molecular biology*. 2006;357(3):928-41.
254. Rajavelu A, Jurkowska RZ, Fritz J, Jeltsch A. Function and disruption of DNA methyltransferase 3a cooperative DNA binding and nucleoprotein filament formation. *Nucleic acids research*. 2011;40(2):569-80.
255. Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature*. 2007;449(7159):248.
256. Emperle M, Rajavelu A, Reinhardt R, Jurkowska RZ, Jeltsch A. Cooperative DNA binding and protein/DNA fiber formation increases the activity of the Dnmt3a DNA methyltransferase. *Journal of Biological Chemistry*. 2014;289(43):29602-13.
257. Ehrlich M. The ICF syndrome, a DNA methyltransferase 3B deficiency and immunodeficiency disease. *Clinical immunology*. 2003;109(1):17-28.
258. Tatton-Brown K, Seal S, Ruark E, Harmer J, Ramsay E, del Vecchio Duarte S, Zachariou A, Hanks S, O'Brien E, Aksglaede L. Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. *Nature genetics*. 2014;46(4):385.
259. Tatton-Brown K, Zachariou A, Loveday C, Renwick A, Mahamdallie S, Aksglaede L, Baralle D, Barge-Schaapveld D, Blyth M, Bouma M. The Tatton-Brown-Rahman Syndrome: A clinical study of 55 individuals with de novo constitutive DNMT3A variants. *Wellcome open research*. 2018;3.
260. Okamoto N, Toribe Y, Shimojima K, Yamamoto T. Tatton–Brown–Rahman syndrome due to 2p23 microdeletion. *American Journal of Medical Genetics Part A*. 2016;170(5):1339-42.
261. Hollink IH, van den Ouweland AM, Beverloo HB, Arentsen-Peters ST, Zwaan CM, Wagner A. Acute myeloid leukaemia in a case with Tatton-Brown-Rahman syndrome: the peculiar DNMT3A R882 mutation. *Journal of medical genetics*. 2017;54(12):805-8.

262. Kosaki R, Terashima H, Kubota M, Kosaki K. Acute myeloid leukemia-associated DNMT3A p. Arg882His mutation in a patient with Tatton-Brown–Rahman overgrowth syndrome as a constitutional mutation. *American Journal of Medical Genetics Part A*. 2017;173(1):250-3.
263. Tlemsani C, Luscan A, Leulliot N, Bieth E, Afenjar A, Baujat G, Doco-Fenzy M, Goldenberg A, Lacombe D, Lambert L. SETD2 and DNMT3A screen in the Sotos-like syndrome French cohort. *Journal of medical genetics*. 2016;53(11):743-51.
264. Xin B, Cruz Marino T, Szekely J, Leblanc J, Cechner K, Sency V, Wensel C, Barabas M, Therriault V, Wang H. Novel DNMT3A germline mutations are associated with inherited Tatton-Brown–Rahman syndrome. *Clinical genetics*. 2017;91(4):623-8.
265. Shen W, Heeley JM, Carlston CM, Acuna-Hidalgo R, Nillesen WM, Dent KM, Douglas GV, Levine KL, Bayrak-Toydemir P, Marcelis CL. The spectrum of DNMT3A variants in Tatton–Brown–Rahman syndrome overlaps with that in hematologic malignancies. *American Journal of Medical Genetics Part A*. 2017;173(11):3022-8.
266. Jeffries AR, Maroofian R, Salter CG, Chioza BA, Cross HE, Patton MA, Temple IK, Mackay D, Rezwan FI, Aksglaede L. Growth disrupting mutations in epigenetic regulatory molecules are associated with abnormalities of epigenetic aging. *bioRxiv*. 2018:477356.
267. Zhang Z-M, Lu R, Wang P, Yu Y, Chen D, Gao L, Liu S, Ji D, Rothbart SB, Wang Y. Structural basis for DNMT3A-mediated de novo DNA methylation. *Nature*. 2018;554(7692):387.
268. Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, Kandoth C, Payton JE, Baty J, Welch J. DNMT3A mutations in acute myeloid leukemia. *New England Journal of Medicine*. 2010;363(25):2424-33.
269. Russler-Germain DA, Spencer DH, Young MA, Lamprecht TL, Miller CA, Fulton R, Meyer MR, Erdmann-Gilmore P, Townsend RR, Wilson RK. The R882H DNMT3A mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its ability to form active tetramers. *Cancer cell*. 2014;25(4):442-54.
270. Kim SJ, Zhao H, Hardikar S, Singh AK, Goodell MA, Chen T. A DNMT3A mutation common in AML exhibits dominant-negative effects in murine ES cells. *Blood*. 2013;122(25):4086-9.
271. Remacha L, Curras-Freixes M, Torres-Ruiz R, Schiavi F, Torres-Perez R, Calsina B, Leton R, Comino-Mendez I, Roldan-Romero JM, Montero-Conde C, Santos M, Perez LI, Pita G, Alonso MR, Honrado E, Pedrinaci S, Crespo-Facorro B, Percesepe A, Falcioni M, Rodriguez-Perales S, Korpershoek E, Ramon-Maiques S, Opocher G, Rodriguez-Antona C, Robledo M, Cascon A. Gain-of-function mutations in DNMT3A in patients with paraganglioma. *Genetics in medicine : official journal of the American College of Medical Genetics*. 2018;20(12):1644-51. Epub 2018/05/10. doi: 10.1038/s41436-018-0003-y. PubMed PMID: 29740169.

272. Averbuch SD, Steakley CS, Young RC, Gelmann EP, Goldstein DS, Stull R, Keiser HR. Malignant pheochromocytoma: effective treatment with a combination of cyclophosphamide, vincristine, and dacarbazine. *Annals of internal medicine*. 1988;109(4):267-73. Epub 1988/08/15. PubMed PMID: 3395037.
273. Heyn P, Logan CV, Fluteau A, Challis RC, Auchynnikava T, Martin CA, Marsh JA, Taglini F, Kilanowski F, Parry DA, Cormier-Daire V, Fong CT, Gibson K, Hwa V, Ibanez L, Robertson SP, Sebastiani G, Rappsilber J, Allshire RC, Reijns MAM, Dauber A, Sproul D, Jackson AP. Gain-of-function DNMT3A mutations cause microcephalic dwarfism and hypermethylation of Polycomb-regulated regions. *Nature genetics*. 2019;51(1):96-105. doi: 10.1038/s41588-018-0274-x. PubMed PMID: 30478443; PMCID: 6520989.
274. Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, Whitaker JW, Tian S, Hawkins RD, Leung D. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell*. 2013;153(5):1134-48.
275. Li Y, Zheng H, Wang Q, Zhou C, Wei L, Liu X, Zhang W, Zhang Y, Du Z, Wang X. Genome-wide analyses reveal a role of Polycomb in promoting hypomethylation of DNA methylation valleys. *Genome biology*. 2018;19(1):18.
276. Jeong M, Sun D, Luo M, Huang Y, Challen GA, Rodriguez B, Zhang X, Chavez L, Wang H, Hannah R. Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nature genetics*. 2014;46(1):17.
277. Long HK, Sims D, Heger A, Blackledge NP, Kutter C, Wright ML, Grützner F, Odom DT, Patient R, Ponting CP. Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. *Elife*. 2013;2:e00348.
278. Wu H, Coskun V, Tao J, Xie W, Ge W, Yoshikawa K, Li E, Zhang Y, Sun YE. Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science*. 2010;329(5990):444-8.
279. Maresca A, Zaffagnini M, Caporali L, Carelli V, Zanna C. DNA methyltransferase 1 mutations and mitochondrial pathology: is mtDNA methylated? *Frontiers in genetics*. 2015;6:90. Epub 2015/03/31. doi: 10.3389/fgene.2015.00090. PubMed PMID: 25815005; PMCID: PMC4357308.
280. Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, Eden E, Yakhini Z, Ben-Shushan E, Reubinoff BE, Bergman Y, Simon I, Cedar H. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat Genet*. 2007;39(2):232-6. Epub 2007/01/04. doi: 10.1038/ng1950. PubMed PMID: 17200670.
281. Rigbolt KT, Vanselow JT, Blagoev B. GProX, a user-friendly platform for bioinformatics analysis and visualization of quantitative proteomics data. *Molecular & cellular proteomics : MCP*. 2011;10(8):O110 007450. doi: 10.1074/mcp.O110.007450. PubMed PMID: 21602510; PMCID: PMC3149098.

282. Jeltsch A, Jurkowska RZ. Allosteric control of mammalian DNA methyltransferases - a new regulatory paradigm. *Nucleic acids research*. 2016;44(18):8556-75. doi: 10.1093/nar/gkw723. PubMed PMID: 27521372; PMCID: PMC5062992.
283. Van Hoof D, Munoz J, Braam SR, Pinkse MW, Linding R, Heck AJ, Mummery CL, Krijgsveld J. Phosphorylation dynamics during early differentiation of human embryonic stem cells. *Cell stem cell*. 2009;5(2):214-26. doi: 10.1016/j.stem.2009.05.021. PubMed PMID: 19664995.
284. Ling Y, Sankpal UT, Robertson AK, McNally JG, Karpova T, Robertson KD. Modification of de novo DNA methyltransferase 3a (Dnmt3a) by SUMO-1 modulates its interaction with histone deacetylases (HDACs) and its capacity to repress transcription. *Nucleic acids research*. 2004;32(2):598-610. doi: 10.1093/nar/gkh195. PubMed PMID: 14752048; PMCID: PMC373322.
285. Kang ES, Park CW, Chung JH. Dnmt3b, de novo DNA methyltransferase, interacts with SUMO-1 and Ubc9 through its N-terminal region and is subject to modification by SUMO-1. *Biochemical and biophysical research communications*. 2001;289(4):862-8. doi: 10.1006/bbrc.2001.6057. PubMed PMID: 11735126.
286. Deplus R, Blanchon L, Rajavelu A, Boukaba A, Defrance M, Luciani J, Rothe F, Dedeurwaerder S, Denis H, Brinkman AB, Simmer F, Muller F, Bertin B, Berdasco M, Putmans P, Calonne E, Litchfield DW, de Launoit Y, Jurkowski TP, Stunnenberg HG, Bock C, Sotiriou C, Fraga MF, Esteller M, Jeltsch A, Fuks F. Regulation of DNA methylation patterns by CK2-mediated phosphorylation of Dnmt3a. *Cell reports*. 2014;8(3):743-53. doi: 10.1016/j.celrep.2014.06.048. PubMed PMID: 25066127.
287. Blanco S, Frye M. Role of RNA methyltransferases in tissue renewal and pathology. *Curr Opin Cell Biol*. 2014;31:1-7. doi: 10.1016/j.ceb.2014.06.006. PubMed PMID: 25014650; PMCID: 4238901.
288. Liu J, Yue Y, Han D, Wang X, Fu Y, Zhang L, Jia G, Yu M, Lu Z, Deng X, Dai Q, Chen W, He C. A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nature chemical biology*. 2014;10(2):93-5. doi: 10.1038/nchembio.1432. PubMed PMID: 24316715; PMCID: 3911877.
289. Musso D, Nhan TX, de Pina JJ, Marchi J, Texier G. The Use of Simple Laboratory Parameters in the Differential Diagnosis of Acute-Phase Zika and Dengue Viruses. *Intervirology*. 2019;62(2):51-6. doi: 10.1159/000499567. PubMed PMID: 31307046.
290. Munster VJ, Koopmans M, van Doremalen N, van Riel D, de Wit E. A Novel Coronavirus Emerging in China - Key Questions for Impact Assessment. *The New England journal of medicine*. 2020;382(8):692-4. doi: 10.1056/NEJMp2000929. PubMed PMID: 31978293.

291. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KSM, Lau EHY, Wong JY, Xing X, Xiang N, Wu Y, Li C, Chen Q, Li D, Liu T, Zhao J, Liu M, Tu W, Chen C, Jin L, Yang R, Wang Q, Zhou S, Wang R, Liu H, Luo Y, Liu Y, Shao G, Li H, Tao Z, Yang Y, Deng Z, Liu B, Ma Z, Zhang Y, Shi G, Lam TTY, Wu JT, Gao GF, Cowling BJ, Yang B, Leung GM, Feng Z. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *The New England journal of medicine*. 2020;382(13):1199-207. doi: 10.1056/NEJMoa2001316. PubMed PMID: 31995857; PMCID: 7121484.
292. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, Liu L, Shan H, Lei CL, Hui DSC, Du B, Li LJ, Zeng G, Yuen KY, Chen RC, Tang CL, Wang T, Chen PY, Xiang J, Li SY, Wang JL, Liang ZJ, Peng YX, Wei L, Liu Y, Hu YH, Peng P, Wang JM, Liu JY, Chen Z, Li G, Zheng ZJ, Qiu SQ, Luo J, Ye CJ, Zhu SY, Zhong NS, China Medical Treatment Expert Group for C. Clinical Characteristics of Coronavirus Disease 2019 in China. *The New England journal of medicine*. 2020;382(18):1708-20. doi: 10.1056/NEJMoa2002032. PubMed PMID: 32109013; PMCID: 7092819.
293. Li G, Fan Y, Lai Y, Han T, Li Z, Zhou P, Pan P, Wang W, Hu D, Liu X, Zhang Q, Wu J. Coronavirus infections and immune responses. *Journal of medical virology*. 2020;92(4):424-32. doi: 10.1002/jmv.25685. PubMed PMID: 31981224; PMCID: 7166547.
294. Nugent EK, Nugent AK, Nugent R, Nugent K. Zika Virus: Epidemiology, Pathogenesis and Human Disease. *The American journal of the medical sciences*. 2017;353(5):466-73. doi: 10.1016/j.amjms.2016.12.018. PubMed PMID: 28502333.
295. Wang Y, Sun Y, Wu A, Xu S, Pan R, Zeng C, Jin X, Ge X, Shi Z, Ahola T, Chen Y, Guo D. Coronavirus nsp10/nsp16 Methyltransferase Can Be Targeted by nsp10-Derived Peptide In Vitro and In Vivo To Reduce Replication and Pathogenesis. *Journal of virology*. 2015;89(16):8416-27. doi: 10.1128/JVI.00948-15. PubMed PMID: 26041293; PMCID: 4524257.
296. Coutard B, Barral K, Lichiere J, Selisko B, Martin B, Aouadi W, Lombardia MO, Debart F, Vasseur JJ, Guillemot JC, Canard B, Decroly E. Zika Virus Methyltransferase: Structure and Functions for Drug Design Perspectives. *Journal of virology*. 2017;91(5). doi: 10.1128/JVI.02202-16. PubMed PMID: 28031359; PMCID: 5309936.

VITA

Allison B. Norvil

Department of Biochemistry

Purdue University

175 South University St, West Lafayette IN, 47907

Education

Purdue University	Ph.D.	Biochemistry	2015-2020	West Lafayette, IN
Purdue University	B.S.	Biochemistry	2011-2015	West Lafayette, IN

Professional Experience

2015 - Current	Graduate Student, Department of Biochemistry, Purdue University, West Lafayette, IN
2013-2015	Undergraduate Researcher, Department of Biochemistry, Purdue University, West Lafayette, IN

Publications

Norvil, Allison B; AlAbdi, Lama; Liu, Bigang; Tu, Yu Han; Forstoffer, Nicole E; Michie, Amie R; Chen, Taiping; Gowher, Humaira. The Acute Myeloid Leukemia variant DNMT3A Arg882His is a DNMT3B-like enzyme. *Nucleic Acids Research*, Volume 48, Issue 7, 17 April 2020, Pages 3761–3775.

Hagras, M.; Abutaleb, NS; Elhosseiny, N; Omara, M; Elsebaei, M; Alhashimi, M; **Norvil AB**; Gutay, MI; Gowher, H; Attia, A; Seleem, MN; Mayhoub, AS. Development of biphenylthiazoles exhibiting improved pharmacokinetics and potent activity against intracellular *Staphylococcus aureus*. Under Review ACS Infectious Disease.

Norvil AB, Saha D, Dar MS, Gowher H. Effect of germline disease-associated mutations on structure function relationship of DNA methyltransferases. (2019) *Genes* 10(5), 369.

Hammad, A.; Abutaleb, NS.; Elsebaei, M.; **Norvil, AB**; Alswah, M.; Ali, AO.; Abdel-Aleem, JA.; Alattar, A.; Bayoumi, S.; Gowher, H.; Seleem, MN.; Mayhoub, AS. From Phenylthiazoles to Phenylpyrazoles: Broadening the Antibacterial Spectrum towards Carbapenem-Resistant Bacteria. (2019) *Journal of Medicinal Chemistry* **2019** 62 (17),7998-8010.

Elsebaei, M.; Mohammad, H.; Samir, A.; Abutaleb, A.; Moustafa, M.; **Norvil, AB**; Michie, AR; Gowher, H.; Seleem, M. Lipophilic Efficient Phenylthiazoles with Potent Undecaprenyl Pyrophosphatase Inhibitory activity. (2019) *European Journal of Medicinal Chemistry* **175**, 49-62.

He, M.; Yang, Q.; **Norvil, AB**; Sherris, D.; Gowher, H. Characterization of Small Molecules Inhibiting the Pro-Angiogenic Activity of the Zinc Finger Transcription Factor Vezf1. (2018) *Molecules* **23**, 7, 1615.

He, M; AlAbdi, L; Yang, Q; **Norvil, AB**; Gowher, H. The transcription factor Vezf1 represses the expression of the antiangiogenic factor Cited2 in endothelial cells. (2018) *JBC* **293**, 11109-11118.

Norvil, A. B., Petell, C. J., AlAbdi, L., Wu, L. C., Rossie, S., and Gowher, H. (2018) DNMT3B Methylates DNA by a Non-cooperative Mechanism, and Its Activity Is Unaffected by Manipulations at the Predicted Dinner Interface. *Biochemistry-Us* **57**, 4312-4324

Savell, K. E., Gallus, N. V. N., Simon, R. C., Brown, J. A., Revanna, J. S., Osborn, M. K., Song, E. Y., O'Malley, J. J., Stackhouse, C. T., **Norvil, A. B.**, Gowher, H., Sweatt, J. D., and Day, J. J. Extra-coding RNAs regulate neuronal DNA methylation dynamics. (2016) *Nature Communications* **7**, 12091.

Research Summary

As a graduate student under the supervision of Dr. Humaira Gowher at Purdue University's Department of Biochemistry, I collaborated with many members of Gowher lab and other labs, and contributed multiple publications. My research involved the use of *in vitro* assays and recombinant proteins purified from bacterial expression systems to determine the kinetics mechanisms of the highly homologous *de novo* DNA methyltransferases DNMT3A and 3B. Despite being highly homologous, these enzymes have distinct genomic targets, suggesting that they have biochemical properties that distinguish their substrate preference. In our study published in *Biochemistry*, we determined that DNMT3B has a unique kinetics mechanism compared to DNMT3A. Further, we found that the DNMT3A mutant found in patients with AML, Arg882His, disrupts the cooperative kinetics mechanism of DNMT3A, while the analogous mutant in DNMT3B, Arg829His, has no effect on its kinetics mechanism. These data suggest that the unique biochemical properties of DNMT3A and 3B may contribute to their substrate specificity.

In our second study published in *Nucleic Acids Research*, we investigated if the kinetics mechanisms of DNMT3A and 3B contribute to substrate specificity. Other studies had

shown that Arg882 directly interacts with the DNA, and the AML Arg882His variant alters the flanking sequence preference compared to the wild type enzyme. We generated consensus motifs for DNMT3A WT, DNMT3A Arg882His, and DNMT3B WT, and found that the Arg882His variant has a flanking sequence preference similar to that of DNMT3B.

Additionally, we showed that under conditions favoring non-cooperativity, DNMT3A wild type has a similar flanking sequence preference as the Arg882His variant and DNMT3B, indicating that the kinetics mechanism of DNMT3A influences its substrate preference. We further showed that this mechanism is conserved *in vivo* embryonic stem cells, and there is a bias in the flanking sequence of DNMT3A and 3B specific sites in genome. This study demonstrated the interplay between kinetics mechanism and substrate specificity.

Lastly, I worked on numerous drug development studies aimed to find novel therapeutics against gram-positive bacteria and Zika Virus enzymes. For these studies, we collaborated with Dr. Saleem and Dr. Mesecar, respectively, at Purdue university to develop novel assays and test potential inhibitors for bacterial enzymes UppP and UppS, and Zika Virus RNA Methyltransferase. These studies have led to two publications, one in *European Journal of Medicinal Chemistry* and the other in *Journal of Medicinal Chemistry*, and two other manuscripts are currently being written.

During my time as a graduate student, I have attended numerous nation and international conferences such as the Keystone Meeting on RNA and DNA Methylation and the Gordon Conference on Epigenetics. I have also had the privilege of mentoring 14 undergraduate and junior graduate students. My experiences in research and training acquired here have helped prepare me for a future in the pharmaceutical industry.

Awards

2018	Graduate Student Beach Travel Award, Department of Biochemistry, Purdue University
2018	Featured in Graduate AG Research Spotlight Series, College of Agriculture, Purdue University
2017	Henry Moses Award, Department of Biochemistry, Purdue University
2017	Bird Stair Research Fellowship, Department of Biochemistry, Purdue University
2015	DowAgro Science Fellowship, Purdue University

Conferences Attended

2020*	Biochemical Horizons Symposium. Purdue University, West Lafayette, IN
-------	---

2019*	Gordon Research Conference - Epigenetic Mechanisms Across Scales: From Molecules to Environment and Evolution, Holderness, NH
2019 [†]	Hitchhiker's Guide to the Biomolecular Galaxy, West Lafayette, IN
2018*	Purdue University Department of Biochemistry Annual Research Retreat, West Lafayette, IN
2018*	Keystone Symposia – DNA and RNA Methylation, Vancouver, Canada
2017*	Biochemical Horizons Symposium, Purdue University, West Lafayette, IN
2017 [†]	Purdue University Department of Biochemistry Annual Research Retreat, West Lafayette, IN
2017	American Society for Biochemistry and Molecular Biology- Experimental Biology. Chicago, IL
2017*	Hitchhiker's Guide to the Biomolecular Galaxy, West Lafayette, IN
2016*	Purdue Cancer Center Retreat, West Lafayette, IN
2016*	Purdue University Department of Biochemistry Annual Research Retreat, West Lafayette, IN
2016*	Midwest Chromatin and Epigenetics Meeting, Van Andel Research Institute, Grand Rapids, MI
2016 [†]	Hitchhiker's Guide to the Biomolecular Galaxy, West Lafayette, IN
2015*	Purdue University Department of Biochemistry Annual Research Retreat, West Lafayette, IN

* Poster Presentation

[†] Oral Presentation

Presentations

2020*	Biochemical Horizons Symposium. <i>The Acute Myeloid Leukemia variant DNMT3A Arg882His is a DNMT3B-like enzyme.</i> Allison
-------	--

	Norvil , Lama AlAbdi, Bigang Liu, Yu Han Tu, Nicole E. Forstoffer, Amie R Michie, Taiping Chen, and Humaira Gowher
2019*	Gordon Research Conference - Epigenetic Mechanisms Across Scales: From Molecules to Environment and Evolution. <i>The Acute Myeloid Leukemia variant DNMT3A Arg882His is a DNMT3B-like enzyme.</i> Allison Norvil , Lama AlAbdi, Bigang Liu, Nicole E. Forstoffer, Amie R. Michie, Taiping Chen, and Humaira Gowher
2019†	Hitchhiker's Guide to the Biomolecular Galaxy. <i>The AML variant DNMT3A Arg882His is a DNMT3B-like enzyme.</i> Allison Norvil
2018*	Purdue University Department of Biochemistry Annual Research Retreat. <i>A Tale of Two Enzymes: Dissecting how the unique catalytic mechanisms of DNMT3A and 3B influence their biological functions.</i> Allison Norvil , Amie Michie, Nicole Forstoffer, Alycia Perez-Johnson, Lama AlAbdi, and Humaira Gowher
2018*	Keystone Symposia – DNA and RNA Methylation. <i>Determining the catalytic mechanisms of DNA and RNA methyltransferases.</i> Allison Norvil , Christopher J Petell, Lama AlAbdi, Emma Lendy, Arun Ghosh, Andrew Mesecar, and Humaira Gowher
2017*	Biochemical Horizons Symposium. <i>DNMT3B methylates DNA by a non-cooperative mechanism and its activity is unaffected by manipulations at the predicted dimer interface.</i> Allison Norvil , Christopher J Petell, Lama AlAbdi, Lanchen Wu, Sandra Rossie, and Humaira Gowher
2017†	Biochemical Horizons Symposium. <i>DNMT3B methylates DNA by a non-cooperative mechanism and its activity is unaffected by manipulations at the predicted dimer interface.</i> Allison Norvil
2017*	Hitchhiker's Guide to the Biomolecular Galaxy. <i>DNMT3B methylates DNA by a non-cooperative mechanism and its activity is unaffected by manipulations at the predicted dimer interface.</i> Allison Norvil , Christopher J Petell, Lama AlAbdi, Lanchen Wu, Sandra Rossie, and Humaira Gowher

- 2016* Purdue Cancer Center Retreat. *DNMT3B methylates DNA by a non-cooperative mechanism*. **Allison Norvil**, Christopher J. Petell, Lama AlAbdi, Lanchen Wu, Sandra Rossie, and Humaira Gowher
- 2016* Purdue University Department of Biochemistry Annual Research Retreat. *DNMT3B methylates DNA by a non-cooperative mechanism*. **Allison Norvil**, Christopher J. Petell, Lama AlAbdi, Lanchen Wu, Sandra Rossie, and Humaira Gowher
- 2016* Midwest Chromatin and Epigenetics Meeting. *Investigating the catalytic mechanism of DNMT3B*. **Allison Norvil**, Christopher J. Petell, and Humaira Gowher
- 2016[†] Hitchhiker's Guide to the Biomolecular Galaxy. *Investigating the catalytic mechanism of DNMT3B*. **Allison Norvil**
- 2015* Purdue University Department of Biochemistry Annual Research Retreat. *Dominant effect of Lsd1 inhibition on DNMT3A activity at the enhancers of pluripotency genes*. **Lama AlAbdi**, Christopher J. Petell, Ming He, Phillip San Miguel, Richard Rose, and Humaira Gowher

* Poster Presentation

[†] Oral Presentation

Mentoring/Teaching experience

- | | |
|-----------------|--|
| 2014-2020 | Mentoring junior graduate students, rotation students and undergraduates, Department of Biochemistry, Purdue University |
| 2015, 2017-2019 | Teaching assistant to Dr. Humaira Gowher, Metabolism (BCHM 462), Fall semester, Department of Biochemistry, Purdue University |
| 2016 | Teaching assistant to Dr. Jeremy Lohman, Molecules (BCHM 361), Spring semester, Department of Biochemistry, Purdue University |
| 2016 | Teaching assistant to Dr. Orla Hart, Introduction to Biochemistry (BCHM 309), Fall semester, Department of Biochemistry, Purdue University |
| 2017 | Instructor, Introduction to Biochemistry (BCHM 309), Spring semester, Department of Biochemistry, Purdue University |
| 2017 | Teaching assistant to Dr. Sandra Rossie, Molecules (BCHM 561), Summer semester, Department of Biochemistry, Purdue University |

2020	Teaching assistant to Dr. Orla Hart, Introduction to Biochemistry (BCHM 307), Spring semester, Department of Biochemistry, Purdue University
2020	Teaching assistant to Dr. Orla Hart, Introduction to Biochemistry (BCHM 307), Summer semester, Department of Biochemistry, Purdue University

PUBLICATIONS

***Extra-coding RNAs regulate neuronal DNA methylation dynamics**

*Originally published in *Nature Communications*.

Savell, K., Gallus, N., Simon, R., Brown, J., Revanna, J., Osborn, M., Song, E., O'Malley, J., Stackhouse, C., **Norvil, A.**, Gowher, H., Sweatt, J., Day, J. Extra-coding RNAs regulate neuronal DNA methylation dynamics. *Nat Commun* **7**, 12091 (2016).

Declaration of collaborative work

Katherine Savell, Nancy Gallus, J. David Sweatt and Jeremy Day designed the experiments.

Katherine Savell and Nancy Gallus executed the experiments with assistance from Esther Song, Rhiana Simon, Mary Katherine Osborn., Jordan Brown, John O'Malley, Jasmin Revanna, Allison Norvil, Humaira Gowher, and Jeremy Day.

Next-generation sequencing data sets were analyzed by Jeremy Day with assistance from Christian Stackhouse.

Jeremy Day and Katherine Savell wrote the manuscript with assistance from all authors.

ARTICLE

Received 27 Jan 2016 | Accepted 28 May 2016 | Published 7 Jul 2016

DOI: 10.1038/ncomms12091

OPEN

Extra-coding RNAs regulate neuronal DNA methylation dynamics

Katherine E. Savell¹, Nancy V.N. Gallus¹, Rhiana C. Simon¹, Jordan A. Brown¹, Jasmin S. Revanna¹, Mary Katherine Osborn¹, Esther Y. Song¹, John J. O'Malley¹, Christian T. Stackhouse¹, Allison Norvil², Humaira Gowher², J David Sweatt¹ & Jeremy J. Day¹

Epigenetic mechanisms such as DNA methylation are essential regulators of the function and information storage capacity of neurons. DNA methylation is highly dynamic in the developing and adult brain, and is actively regulated by neuronal activity and behavioural experiences. However, it is presently unclear how methylation status at individual genes is targeted for modification. Here, we report that extra-coding RNAs (ecRNAs) interact with DNA methyltransferases and regulate neuronal DNA methylation. Expression of ecRNA species is associated with gene promoter hypomethylation, is altered by neuronal activity, and is overrepresented at genes involved in neuronal function. Knockdown of the *Fos* ecRNA locus results in gene hypermethylation and mRNA silencing, and hippocampal expression of *Fos* ecRNA is required for long-term fear memory formation in rats. These results suggest that ecRNAs are fundamental regulators of DNA methylation patterns in neuronal systems, and reveal a promising avenue for therapeutic targeting in neuropsychiatric disease states.

¹Department of Neurobiology, McKnight Brain Institute, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA. ²Department of Biochemistry, Purdue University, West Lafayette, Indiana 47907, USA. Correspondence and requests for materials should be addressed to J.J.D. (email: jjday@uab.edu) or to J.D.S. (email: dsweatt@uab.edu).

Methylation of cytosine bases in DNA is a critical regulator of the function and plasticity within the central nervous system. In the developing brain, neurons exhibit unique DNA methylation patterns that are correlated with synaptogenesis and regulate the expression of neuronal genes¹. In the adult brain, active DNA methylation is required for memory formation and maintenance^{2–5}, and neuronal activity and behavioural experiences lead to site-specific reorganization of DNA methylation dynamics^{4,6–9}. In addition, maintenance of DNA methylation patterns at specific genes is disrupted in numerous cognitive and neurodegenerative disorders^{10–14}.

Despite the well-appreciated role of DNA methylation in neuronal function and physiology, the mechanisms by which individual genes or sequences of DNA are targeted for active methylation or demethylation are presently unclear. Emerging evidence from other systems suggests that site-specific regulation of DNA methylation can occur via non-coding RNA species^{15–18}. Most remarkably, extra-coding RNAs (ecRNAs), previously characterized as non-polyadenylated, sense-strand RNA overlapping protein-coding genes, have been shown to bind to DNA methyltransferases (DNMTs) and direct gene-specific methylation patterns¹⁵. These RNAs were previously found to precede synthesis of its messenger RNA (mRNA) counterpart in S phase and to be regulated by both RNA polymerases II and III (ref. 15). However, the extent and nature of this regulation in post-mitotic cells such as neurons has not been explored.

Here, we report that ecRNA species, which we define as sense-strand, non-polyadenylated RNAs that overlap with the gene boundaries, are generated from genes that are critical for neuronal responses to stimulation, and that neuronal activity modifies ecRNA production. Genome-wide, ecRNA levels are correlated with mRNA expression and promoter DNA methylation. We show that ecRNAs directly bind DNA methyltransferases to inhibit their activity, and that blockade of ecRNA production at the *Fos* gene locus results in hypermethylation and gene silencing. Finally, our data reveal that ecRNA expression from the *Fos* gene in the hippocampus is required for the formation of long-term fear memories, suggesting a critical role for ecRNA-directed DNA methylation in neuronal function within the adult brain.

Results

Genome-wide classification of neuronal ecRNAs. To characterize the prevalence and expression of ecRNAs in neuronal systems, we established a pipeline for detection of non-polyadenylated (PolyA[−]) RNA transcripts using whole-genome, directional RNA sequencing (Fig. 1a). Total RNA was extracted from neuronal cultures prepared from embryonic rat cortex, and underwent selection for polyadenylation signals. PolyA⁺ and PolyA[−] libraries underwent identical sequencing and were aligned to the rat genome (Rn5 build) using Tophat. This pipeline resulted in considerable enrichment of known non-polyadenylated transcripts (for example, genes encoding canonical histone proteins) in the PolyA[−] sequencing tracks, as well as highly reproducible PolyA⁺ and PolyA[−] sequencing between biological replicates (Supplementary Fig. 1). Comparison between PolyA⁺ and PolyA[−] RNA-seq data sets revealed the presence of non-polyadenylated, ecRNA-like transcripts overlapping protein-coding genes that also generate PolyA⁺ transcripts (Fig. 1b). Although a considerable fraction of PolyA[−] transcripts mapped to exonic regions (possibly indicative of spliced mRNA that had not yet been polyadenylated), we also observed frequent read coverage from intronic regions and regions outside the typical gene boundaries

(transcription start and end sites) specific to the PolyA[−] library (Fig. 1b, Supplementary Fig. 1c). To characterize the genome-wide expression of ecRNA-like transcripts, we computed an ecRNA index of sense-strand PolyA[−] reads arising from upstream, intronic and downstream loci surrounding genes (Fig. 1c). This index revealed that ecRNA-like expression from known coding genes was ubiquitous across the genome, and was highly correlated with mRNA expression from the overlapping gene (Fig. 1d,e).

To determine whether ecRNA production was associated with DNA methylation states genome-wide, we next performed whole-genome sequencing on methylated DNA fragments enriched with a recombinant methyl-binding domain (MBD) protein (Supplementary Fig. 2). As demonstrated previously¹⁹, we observed hypomethylation of the vast majority of gene promoters and CpG islands within the genome, and hypermethylation of gene bodies across the entire genome (Fig. 1g, Supplementary Fig. 3a,b). To examine whether ecRNA levels reflected DNA methylation patterns, we divided all genes into distinct quartiles according to their ecRNA index (Fig. 1f–g). This analysis revealed genes with higher ecRNA expression generated significantly more mRNA and possessed less DNA methylation at promoters (defined as 5 kb region surrounding the transcription start site (TSS); Fig. 1h–j). As this effect could potentially be driven by the presence of CpG islands at highly expressed genes, we next examined the relationship between ecRNA levels and CpG island status. As expected, promoters containing CpG islands were hypomethylated, whereas promoters lacking CpG islands were hypermethylated (Supplementary Fig. 3c). However, the relationship between ecRNA level and promoter DNA methylation was preserved regardless of promoter CpG status. Additionally, we found that DNA methylation at gene-overlapping CpG islands was strongly related to ecRNA expression (Supplementary Fig. 3d), indicative of potential ecRNA regulation of DNA methylation states at both promoter and intragenic sites.

Given that DNA methylation status in neurons is both activity and experience-dependent, we next assessed the degree to which ecRNAs were altered by neuronal activity. Neurons were incubated with potassium chloride (KCl) or tetrodotoxin (TTX) (Fig. 2), which increase and decrease neuronal activity, respectively. We again performed genome-wide RNA sequencing to examine both PolyA⁺ (mRNA) and PolyA[−] (including extra-coding) RNA species. Following neuronal depolarization with KCl, we observed that PolyA⁺ expression from a small set of genes (Fig. 2a; Supplementary Data 1) were modified by neuronal activity, including a number of immediate early genes (*Fos*, *Arc*, *Egr1* and *Nr4a2*) that have well-established roles in neuronal and behavioural function. To determine whether non-polyadenylated transcripts arising from these activity-regulated genes were also modified by neuronal depolarization, PolyA[−] transcripts arising from the 5', intronic, and 3' extra-genic regions were plotted and compared with PolyA⁺ transcripts (Fig. 2b; Supplementary Data 1). Depolarization-induced ecRNA transcription from each of these regions was significantly correlated with PolyA⁺ transcription, suggesting that ecRNAs arising from these genes are also activity-dependent (see Fig. 2c,d for representative examples). Several of these results were independently verified using quantitative PCR with reverse primers (RT-qPCR) with custom mRNA or ecRNA-specific primers (Fig. 2e,f), confirming significant activity-dependent regulation of ecRNA transcripts.

To examine the interplay between ecRNA levels and DNA methylation status at different functional groups within the genome, we performed unbiased hierarchical clustering using a percentile-normalized quantification of ecRNA index, promoter

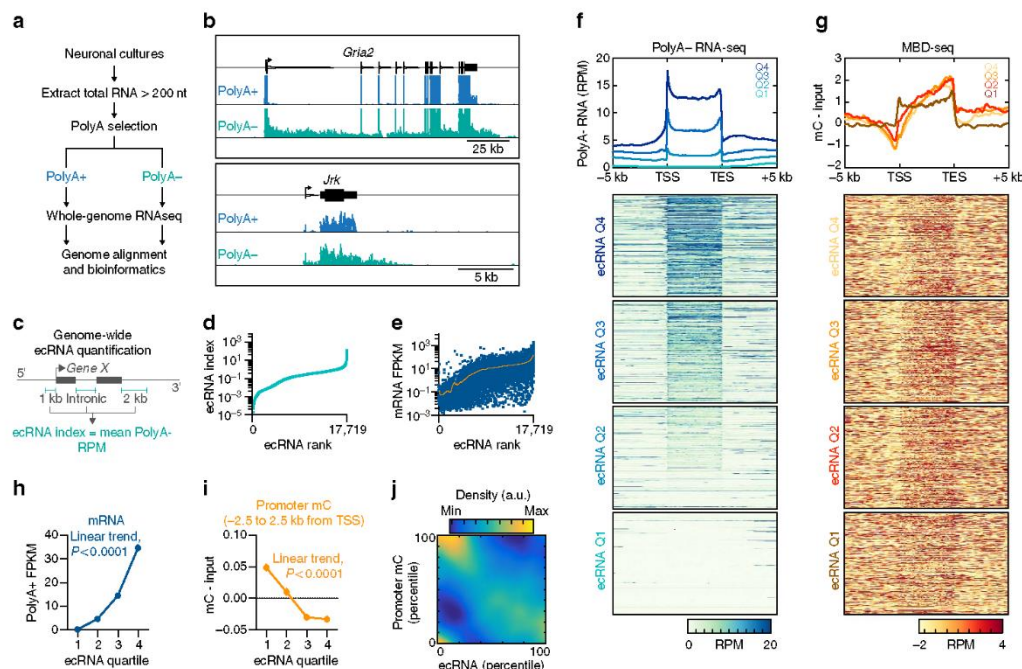


Figure 1 | Genome-wide identification and quantification of ecRNAs from neuronal systems. (a) RNA-seq workflow identifies both polyadenylated and non-polyadenylated transcripts from the same neuronal tissue. (b) Comparison of PolyA+ and PolyA- sequencing from representative gene loci reveals PolyA- transcripts arising from intronic and post-TESs. (c) Genome wide, extra-coding transcripts were characterized by averaging PolyA- reads that mapped to 5' (pre-TSS), intronic or 3' (post-TES) of a given gene. (d) Rank plot of ecRNA index at 17,719 rat genes. (e) mRNA expression (PolyA+ RNA-seq) ranked by ecRNA index reveals correlation between ecRNA and mRNA expression. (f) Division of ecRNA into discrete quartiles reveals general profile and expression of PolyA- RNA transcripts. Data are aligned to transcription start sites (TSS) and TESs. Heatmap shows PolyA- transcription from all genes. (g) MBD-seq reveals metagenomic DNA methylation profiles, including hypomethylation at TSS and hypermethylation at TES. ecRNA transcription is associated with hypomethylated promoters across the genome. (h) Genome wide, ecRNA levels are positively correlated with mRNA transcription ((h) one-way ANOVA, $F_{(3,17715)} = 612.5$, $P < 0.0001$) and negatively correlated with promoter DNA methylation ((i) one-way ANOVA, $F_{(3,17715)} = 73.27$, $P < 0.0001$). (j) Percentile-percentile density scatterplot of ecRNA and promoter DNA methylation. Data in h and i are presented as mean \pm s.e.m.

DNA methylation and gene expression (mRNA fragments per kilobase of exon per million mapped reads (FPKM)). Hierarchical clustering identified at least eight distinct subgroups of genes (Fig. 3). As highlighted above, gene clusters with the highest ecRNA levels tended to have the lowest promoter methylation, whereas clusters that did not generate ecRNAs were characterized by promoter hypermethylation. Gene ontology analysis of the largest cluster demonstrated an enrichment of genes involved in neuronal processes, including gene ontology terms for neuronal differentiation, neuronal development and dendritic spine morphogenesis (Fig. 3e, Supplementary Data 2). This subclass of genes was marked by a high ecRNA index and low promoter DNA methylation, and also overlapped with genes regulated by neuronal depolarization and genes involved in intellectual/learning disorders and neurodegenerative disease states (Fig. 3b). In contrast, the second-largest cluster was marked by low ecRNA levels and high promoter DNA methylation, and was enriched for gene ontology-terms related to immune function (Fig. 3d).

Although the vast majority of genes followed the general inverse relationship between ecRNA levels and DNA methylation, one small cluster of genes was marked by promoter hypomethylation despite the relative absence of ecRNA expression.

This category generated the lowest mRNA levels, and was heavily enriched with genes that code for olfactory receptors (Fig. 3c), which possess unique regulatory mechanisms that help to specify expression of a single olfactory receptor gene in olfactory sensory neurons²⁰. Similarly, another small cluster was marked by promoter hypermethylation despite high ecRNA quantification and high mRNA expression (Fig. 3f). This cluster was enriched for genes involved in ribosomal RNA processing, which involves robust expression of genes that are critical for function in all cells. Overall, this analysis suggests that ecRNAs may be important for regulation of DNA methylation states at certain functional classes of genes but not others.

Biogenesis and activity-dependence of Fos ecRNA. To further dissect the biogenesis, structure and function of ecRNA in neuronal systems, we focused on the ecRNA arising from the activity-regulated *Fos* gene locus (Fig. 4a). Our RNA-seq and RT-qPCR results indicated that a non-polyadenylated transcript arising from the 3' (post-transcription end site (TES)) *Fos* gene locus was induced by neuronal depolarization with KCl. Therefore, we next examined the time course of *Fos* ecRNA and mRNA induction following neuronal stimulation using RT-qPCR

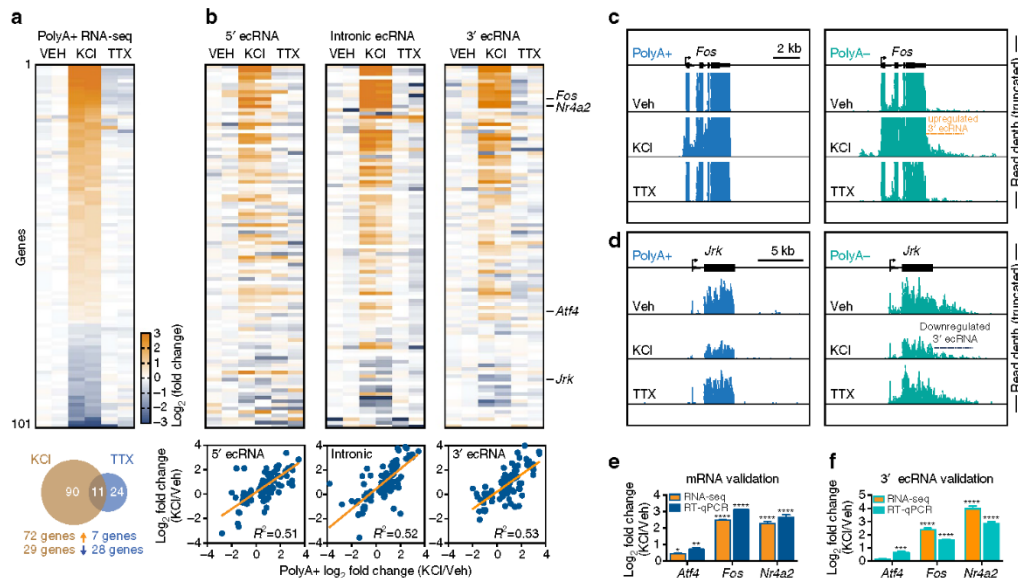


Figure 2 | Regulation of mRNA and ecRNA by neuronal activity. (a) PolyA+ RNA-seq following 1 h neuronal depolarization (25 mM KCl) or inactivation (1 μ M TTX) reveals altered mRNA expression at a small subset of genes. Top, heatmap of KCl-altered transcripts (each column = 1 biological replicate; 2 replicates per treatment). Bottom, Venn diagram of overlap between transcripts altered by KCl and TTX. (b) Corresponding heatmaps from PolyA- RNA-seq reveal relationship between activity-related mRNA and ecRNA changes. PolyA- RNA transcription from 5', intronic and 3' sites all correlated significantly with mRNA changes following neuronal depolarization with KCl (linear regression, $P < 0.0001$ for each comparison). (c, d) Representative examples of activity-induced increases (*Fos* gene, (c)) and decreases (*Jrk* gene, (d)) in 3' ecRNA levels. RNA-seq reads shown individually; read depth truncated to highlight ecRNA changes. (e, f) Validation of RNA-seq results with RT-qPCR ($n = 6$ biological replicates) confirms changes in mRNA and ecRNA levels. Data are expressed as mean \pm s.e.m. Individual comparisons made with Student's t -test versus vehicle, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ and **** $P < 0.0001$.

with ecRNA-specific primers (Fig. 4a; Supplementary Data 3). Although KCl increased both *Fos* mRNA and ecRNA, these transcripts had unique induction profiles (Fig. 4b, c). Critically, *Fos* ecRNA (but not mRNA) was significantly elevated as soon as 30 min following KCl stimulation, with maximal ~ 3 -fold induction at 45 min after stimulation (Fig. 4c). By contrast, *Fos* mRNA levels were not significantly elevated until 45 min following depolarization, and attained maximal levels until 1–4 h after stimulation (Fig. 4b). To examine whether other mechanistically distinct forms of stimulation could induce ecRNAs, neuronal cultures were treated for 1 h with the specific glutamate receptor agonists α -Amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) or N-methyl-D-aspartate (NMDA; Fig. 4b, c). Similar to depolarization with KCl, acute AMPA or NMDA treatment significantly elevated both *Fos* mRNA and 3' ecRNA, although mRNA induction was consistently higher than ecRNA induction. Finally, to determine whether *Fos* ecRNA levels were also sensitive to neuronal inactivation, neuronal cultures were incubated for 1 h with TTX. Although *Fos* mRNA was substantially reduced by acute TTX treatment, *Fos* ecRNA levels were not altered by neuronal inactivation. Together, these results suggest ecRNA levels in neurons can be activity-dependent and are potentially separate from mRNAs arising from the same gene.

The distinct ecRNA and mRNA synthesis patterns following neuronal activity indicates that ecRNAs are unlikely to be a simple byproduct of conventional mRNA production by RNA polymerases, and also suggests that these transcripts may be controlled by different mechanisms. To examine this possibility,

we treated neuronal cultures with selective inhibitors of RNA polymerase II (RNAPII) and RNA polymerase III (RNAPIII) to characterize ecRNA and mRNA biogenesis (Supplementary Fig. 4). Previous studies have demonstrated that *Fos* mRNA, along with many other immediate early genes, is transcribed by RNAPII, which can remain poised at transcription start sites^{21–23}. In contrast, RNAPIII transcribes of non-coding elements like transfer and ribosomal RNAs. As expected, we found that basal levels of *Fos* mRNA are sensitive to inhibition of transcription by RNAPII. 4 h treatment with 5, 6-dichloro-1- β -D-ribofuranosyl-1H-benzimidazole (DRB, a pTEFb inhibitor which blocks RNAPII-dependent transcription), dose-dependently decreased *Fos* mRNA (Supplementary Fig. 4a), whereas treatment with an RNAPIII inhibitor (ML-60218) did not alter *Fos* mRNA expression (Supplementary Fig. 4b). In contrast, we found that basal expression of *Fos* ecRNA was sensitive to treatment with either inhibitor, indicating that basal ecRNA transcription may be induced by either RNAPII or RNAPIII. To probe this relationship at another mRNA/ecRNA pair, we repeated this experiment at a second activity-responsive gene (*Atf4*). At this locus, we found a complete dissociation between the effects of RNAPII and RNAPIII inhibition. DRB treatment dose-dependently reduced *Atf4* mRNA, but had no effect on *Atf4* ecRNA. Conversely, RNAPIII inhibition decreased *Atf4* ecRNA but did not decrease *Atf4* mRNA (Supplementary Fig. 4).

Together, these results further suggest that ecRNA is not simply a non-polyadenylated byproduct of normal gene transcription. However, steady-state transcript levels are the result of complex dynamics between transcript synthesis and

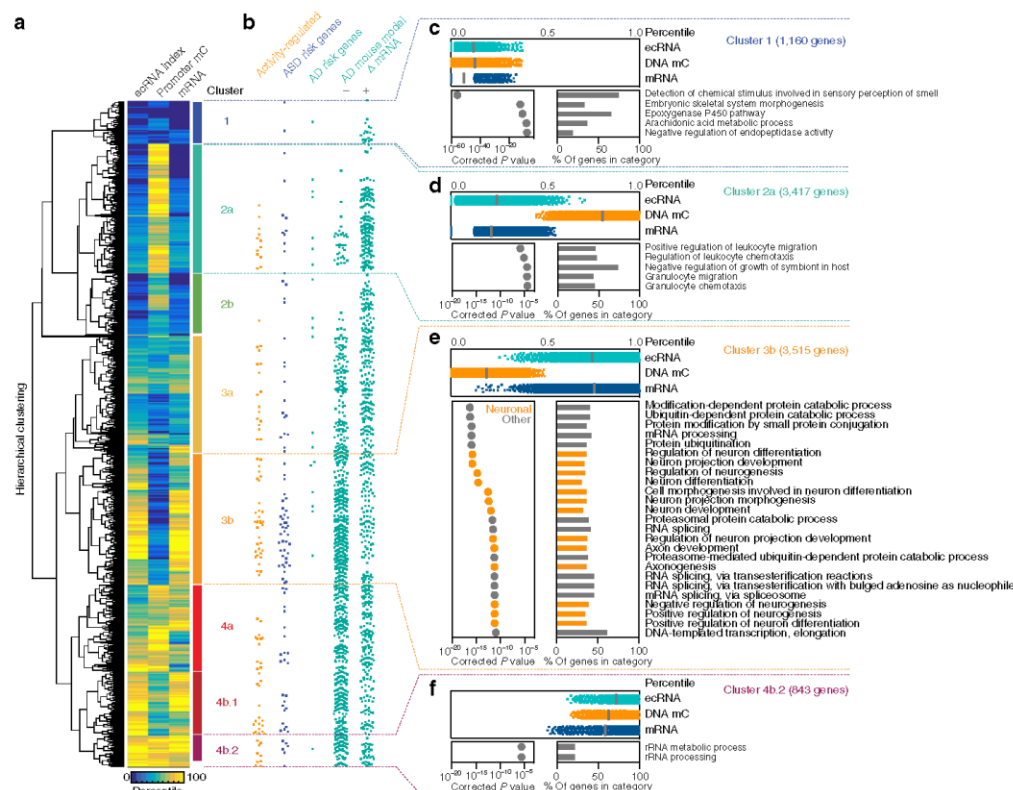


Figure 3 | High ecRNA expression is inversely related to promoter DNA methylation at genes involved in neuronal function and brain disease.

(a) Hierarchical clustering of percentile-normalized ecRNA, DNA mC and mRNA highlights major gene clusters. Clustering was performed using percentile scores instead of raw data to equalize data range. (b) Overlap between distinct gene clusters and activity-responsive (Veh versus KCl genes from Fig. 2a), genes harbouring mutations associated with autism spectrum-disorders⁴³ (ASD risk genes), genes implicated in Alzheimer's disease risk⁴⁴ (AD risk genes) and genes altered in an animal model of Alzheimer's disease³⁴ (AD mouse model Δ mRNA genes; downregulated (–) and upregulated (+)). Each circle represents a single gene. (c–f) Distribution of ecRNA, DNA mC and mRNA scores (top panels) and gene ontology analysis (bottom panels) of selected clusters. Gene ontology groups are ranked by corrected P value. Cluster 3b (e) is enriched for genes involved in neurogenesis, neuronal projection and neuronal development. This cluster represents 19.8% of genes investigated but contains 34% of activity-regulated genes (odds ratio = 1.69; $P = 0.0079$), 40% of the genes implicated in ASD (odds ratio = 2.006; $P = 0.0003$) and 39% of genes that are downregulated in a mouse AD model (odds ratio = 1.966; $P < 0.0001$).

degradation. To further examine activity-dependent synthesis of *Fos* mRNA and ecRNA, we pre-treated neuronal cultures with RNAPII or RNAPIII-dependent transcriptional inhibitors before KCl-induced depolarization (Fig. 4d,e). Here, we found that while KCl-induced increases in *Fos* mRNA were prevented by pretreatment with DRB, RNAPII-dependent transcription inhibition was unable to block increases in *Fos* ecRNA. In contrast, ML-60218 significantly attenuated *Fos* ecRNA increases following KCl (but did not alter KCl-induced *Fos* mRNA), suggesting that at least part of the *Fos* ecRNA induction in response to neuronal depolarization is driven by RNAPIII transcription. In agreement with this, RNAPIII binding at the *Fos* promoter was significantly increased following KCl stimulation, as detected using chromatin immunoprecipitation (ChIP) for the RNAPIII subunit RPC32 (Supplementary Fig. 4c). Thus, these findings reveal a critical dissociation between RNAPII and RNAPIII driven transcription at the *Fos* locus, and further confirm the independence of activity-dependent *Fos* ecRNA and *Fos* mRNA synthesis.

Given the genome-wide relationship between DNA methylation and ecRNA expression, we hypothesized that *Fos* ecRNA increases following KCl stimulation might be associated with subsequent *Fos* hypomethylation. To test this, we examined DNA methylation levels at three sites relative to the *Fos* gene (distal upstream enhancer, proximal promoter and gene body) using methylated DNA immunoprecipitation (MeDIP) 24 h after KCl-induced neuronal depolarization. As shown in Fig. 4f, we observed decreases in DNA methylation at each of these locations following KCl treatment, which is consistent with the genome-wide relationship between ecRNA and DNA methylation.

ecRNAs directly bind to DNMT3a and prevent methylation. Next, we sought to examine how ecRNAs might regulate DNA methylation status. As previous studies suggest that DNA methyltransferases bind to non-coding RNA species^{15–17}, we hypothesized that *Fos* ecRNA controls DNA methylation at this

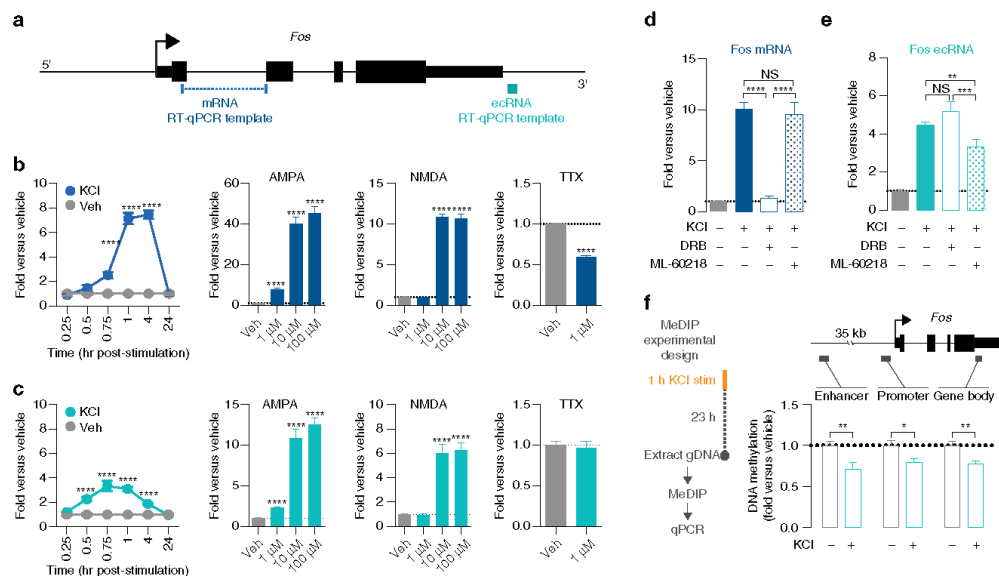


Figure 4 | *Fos* ecRNA is differentially responsive to neuronal activation and undergoes unique biogenesis. (a) RT-qPCR template locations used to distinguish between *Fos* mRNA and ecRNA. (b) Modulation of *Fos* mRNA by KCl, AMPA, NMDA and TTX reveals overall timecourse and activity-dependence of *Fos* gene transcription. *Fos* mRNA is upregulated as soon as 45 min after KCl treatment, and peaks at 4 h following stimulation. AMPA and NMDA treatment (1 h) produced dose-dependent increases in *Fos* mRNA, whereas neuronal silencing with TTX decreased *Fos* mRNA. (c) In contrast, *Fos* ecRNA is induced within 30 min of neuronal depolarization with KCl, and peaks at 45 min following KCl treatment. Neuronal stimulation with KCl, AMPA and NMDA induced much lower levels of ecRNA transcription (as compared with mRNA), and neuronal inactivation with TTX did not alter *Fos* ecRNA transcription. (*n* = 3–6 per group for KCl experiments, 3 per group for AMPA and NMDA experiments, and 6 per group for TTX experiments). (d,e) 4-h pretreatment with the RNAPII-dependent transcription inhibitor 5, 6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) had no effect on *Fos* ecRNA but blocked mRNA induction after 1-hour KCl treatment, whereas pretreatment with the RNAPIII inhibitor ML-60218 had no effect on *Fos* mRNA but decreased ecRNA induction after KCl treatment (*n* = 6–12 per group; mRNA one-way ANOVA, $F_{(3,32)} = 78.86$, $P < 0.0001$; ecRNA one-way ANOVA, $F_{(3,32)} = 66.55$, $P < 0.0001$; Tukey's *post hoc* test for individual comparisons). (f) Left, MeDIP experimental design. Right, DNA methylation decreases 24 h after KCl treatment in the enhancer, promoter and gene body of the *Fos* locus (*n* = 8, unpaired Student's *t*-test; $t_{16} > 2.733$ and $P < 0.015$ for each comparison). All data are expressed as mean \pm s.e.m. Individual comparisons, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ and **** $P < 0.0001$.

gene by interacting with DNA methyltransferases. To examine this possibility, we first verified nuclear expression of the major DNA methyltransferases (DNMT1 and DNMT3a) in our neuronal cultures using IP-grade DNMT antibodies and MAP2 co-staining to identify neurons (Fig. 5a). This staining revealed mostly nuclear localization of DNMTs. Likewise, PolyA+ RNA-seq revealed that *Dnmt1* and *Dnmt3a* are the most abundant DNMT transcripts in cortical neuronal cultures (*Dnmt1* FPKM mean = 17.18; *Dnmt3a* FPKM mean = 14.57; *Dnmt3b* FPKM mean = 0.79). To determine whether *Fos* ecRNA might interact with DNMTs, we performed an RNA immunoprecipitation (RIP) with antibodies for DNMT1 and DNMT3a. As part of the RNA-IP protocol, we gently permeabilized the cellular membrane, which allowed for nuclear isolation and collection. In non-IP (input control) samples, we found that ecRNAs were significantly enriched as compared with total lysate fractions, indicative of nuclear localization (Fig. 5b). In IP samples, RT-qPCR was used to detect the presence of ecRNAs bound to DNMTs, and revealed significant enrichment of *Fos* ecRNA on DNMT1 and DNMT3a, as compared with a normal immunoglobulin-G (IgG) control (Fig. 5c). In contrast, *Fos* mRNA was not significantly enriched in DNMT-RIP samples, demonstrating selectivity of DNMT binding to extra-coding transcripts arising from the *Fos* locus.

These results are the first indication that ecRNAs bind to the *de novo* DNA methyltransferase DNMT3a. To examine this

interaction in more detail, we performed a series of *in vitro* DNMT binding assays with synthetic, fluorescently labelled 25-nucleotide RNA and double-stranded DNA probes based on sequences from the *Fos* ecRNA locus (Fig. 5d). Interaction between synthetic RNAs and DNMTs was detected using an electrophoretic separation mobility shift assay (EMSA; Fig. 5e–j). For both RNAs examined (ecRNA-1 and ecRNA-2), we observed slight binding to DNMT1 (Fig. 5e) and robust binding to recombinant DNMT3a either in complex with DNMT3l (which prevents DNMT3a aggregation and allows greater electrophoretic mobility; Fig. 5f), or the isolated catalytic domain of DNMT3a (Fig. 5g). RNA did not bind other proteins with similar size or charge, and binding was maintained with the addition of the non-specific competitor poly-dI-dC (see Methods section and Supplementary Fig. 5c). To directly compare DNMT3a binding to RNA and DNA with the same primary sequence, we conducted complete binding curves to derive the affinity of DNMT-nucleic acid binding (Fig. 5h). Strikingly, ecRNA sequences bound the catalytic domain of DNMT3a with similar affinity as double-stranded DNA, with K_d values in the low nanomolar range (Fig. 5i). Moreover, a competition assay (Fig. 5j,k) revealed that over 20-fold excess of dsDNA was required to reduce ecRNA-DNMT binding by half, suggesting robust ecRNA-DNMT3a interactions even in the presence of double-stranded DNA.

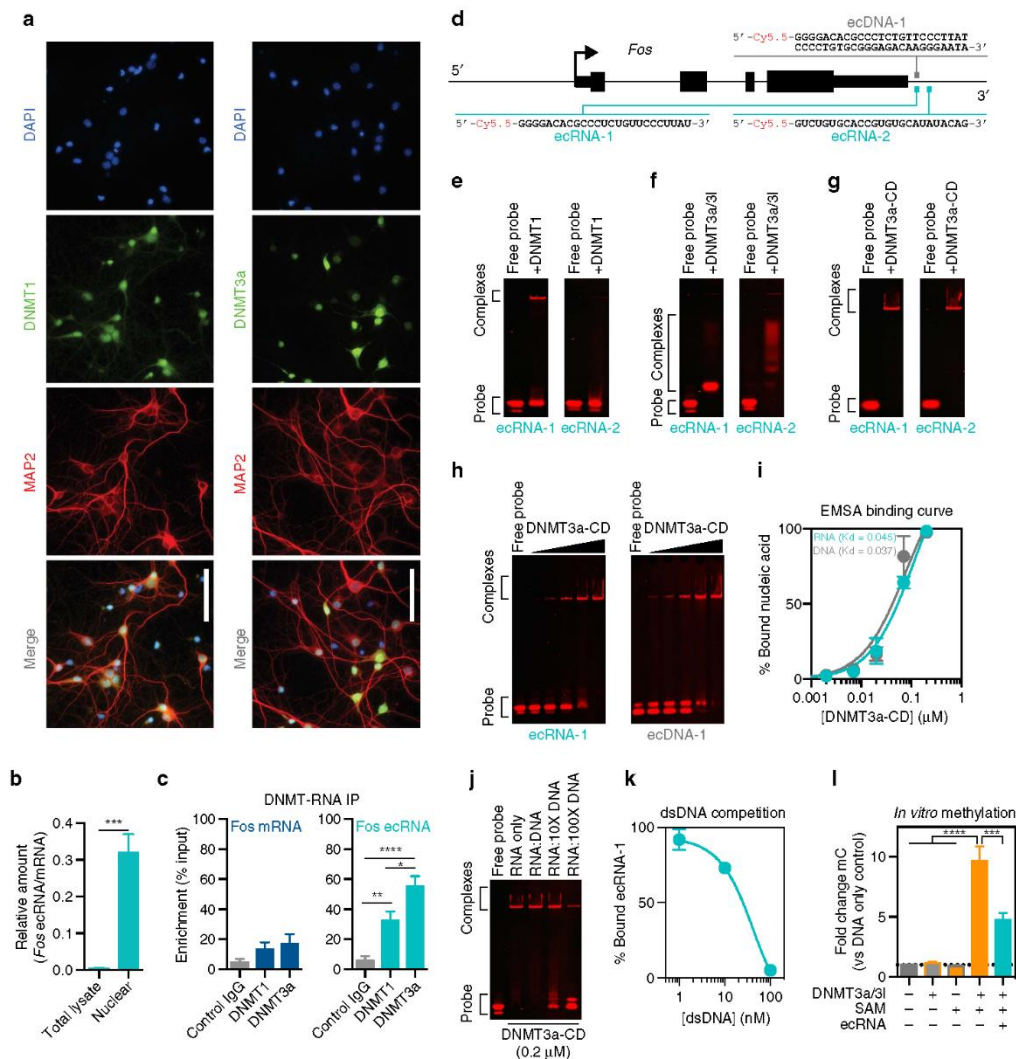


Figure 5 | *Fos* ecRNA interacts with DNA methyltransferases and blocks DNA methylation. (a) Immunostaining reveals nuclear localization of DNMT1 and DNMT3a in neuronal cultures. Cell nuclei are stained with 4,6-diamidino-2-phenylindole (DAPI), and neurons are marked by MAP2 (microtubule-associated protein 2). Scale bar, 50 μm. (b) *Fos* ecRNA/mRNA comparison in total neuronal lysate and nuclear fraction (separated during RNA-IP; $n = 4$ per group; unpaired Student's t -test, $t_6 = 6.301$, $P = 0.007$). (c) *Fos* ecRNA, but not mRNA, immunoprecipitates with anti-DNMT1 or DNMT3a antibodies but not control IgG ($n = 4$ –6 per group; ecRNA one-way ANOVA, $F_{(2,15)} = 20.53$, $P < 0.0001$; Tukey's *post hoc* test for individual comparisons). (d) Locations of synthetic RNA and DNA oligonucleotides used in mobility shift assays. (e) Electrophoretic mobility shift assay reveals only slight binding of ecRNA probes (1 nM) to recombinant DNMT1 (0.2 μM). RNA/DNMT complexes are evident as low-mobility band on native PAGE gel following incubation with DNMT protein. (f) Synthetic ecRNA probes (1 nM) bind recombinant DNMT3a/DNMT3l protein (0.2 μM). (g) Complete binding of synthetic ecRNA probes (1 nM) to truncated recombinant DNMT3a protein (0.2 μM) containing only catalytic domain (DNMT3a-CD). (h) Incubation of ecRNA-1 or ecDNA-1 probes (1 nM) with escalating concentrations of DNMT3a-CD (0.002–0.2 μM). (i) DNMT3a-CD binds equally to RNA and double-stranded DNA with the same primary sequence. Binding affinity (K_d values; derived from non-linear, one-site regression analysis of complete concentration curve) for RNA and dsDNA was not significantly different ($n = 2$ replicates; comparison of K_d , $F_{(1,22)} = 0.47$, $P = 0.49$). (j) Competition assay between ecRNA-1 (1 nM) and unlabelled ecDNA-1 probes (1–100 nM) shows intact RNA/DNMT3a-CD complexes even with 10-fold higher concentrations of DNA. (k) Quantification of dsDNA competition assay. (l) Co-incubation of DNMT3a/3l protein and the methyl donor SAM results in cytosine methylation at dsDNA (3.42 nM) from the *Fos* promoter. Methylation was significantly inhibited by addition of ecRNA (5.16 μM). For (l) $n = 4$ per group, one-way ANOVA, $F_{(4,15)} = 40.25$, $P < 0.0001$; Tukey's *post hoc* test for individual comparisons. All data are expressed as mean \pm s.e.m. Individual comparisons, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ and **** $P < 0.0001$.

Next, we used an *in vitro* DNA methylation assay to test whether the presence of ecRNA could block the ability of DNMTs to methylate double-stranded DNA. Although either DNMT3a/3l protein or the methyl donor S-adenosyl methionine (SAM) alone was insufficient to methylate an isolated dsDNA fragment from the *Fos* promoter region, co-incubation of DNMT and SAM induced robust methylation of the target DNA (Fig. 5l). Consistent with the ability of ecRNA to bind DNMT3a directly at the catalytic domain, and with the genome-wide relationship between ecRNA and DNA methylation, we found that addition of ecRNA to the *in vitro* methylation reaction was sufficient to impair methylation of the *Fos* promoter target DNA.

The ability of ecRNAs to form stem-loop structures has been shown to mediate interactions between ecRNAs and DNMT1 (ref. 15). To determine whether ecRNA-DNMT3a interactions were also dependent on stem-loop structure, we designed 'mutant' synthetic RNAs that possessed identical base composition as synthetic ecRNA-1 and ecRNA-2, but lacked a predicted secondary structure (Supplementary Fig. 5). DNMT3a-CD interactions were not impaired in mutant RNAs lacking predicted stem-loop secondary structure (Supplementary Fig. 5a,b). These results suggest that stem-loop structure is not required for ecRNA-DNMT3a interactions.

Fos ecRNA reduction alters methylation and gene expression.

Overall, our results suggest a model in which ecRNA species generated from a protein-coding gene can bind to DNA methyltransferases to inhibit DNA methylation at the promoter of that gene. This model is intriguing in that it indicates that protein-coding transcription could be dynamically tuned by manipulating ecRNA status at a given gene. To test this prediction, we designed chemically modified anti-sense oligonucleotides (ASOs) that selectively target *Fos* 3' ecRNA (at locations at least 100bp from the mRNA TES) or *Fos* mRNA (in exon 3) (Fig. 6a; Supplementary Data 3). As our model predicted, ASOs that targeted the *Fos* 3' ecRNA significantly reduced both the ecRNA and the mRNA, while also significantly decreasing *Fos* protein levels (Fig. 6c,d). In contrast, ASOs targeting *Fos* mRNA specifically decreased mRNA levels, but did not alter *Fos* ecRNA (Fig. 6b). To examine how ASO-mediated ecRNA degradation affected DNA methylation status at the *Fos* gene, we assayed DNA methylation levels using MeDIP (at the *Fos* enhancer, promoter and gene body; Fig. 6e) and bisulfite sequencing (at the *Fos* promoter; Fig. 6f). ASO-mediated *Fos* ecRNA degradation increased DNA methylation levels as assayed both by MeDIP and bisulfite sequencing. Overall, bisulfite analysis revealed that individual CpGs in the *Fos* promoter were 1.7 times more likely to be methylated following ASO treatment than in vehicle-treated controls, and this effect was largely driven by significant ASO-induced hypermethylation at two distinct CpG sites (Fig. 6f). In contrast, treatment with a *Fos* mRNA ASO that resulted in comparable degradation of *Fos* mRNA levels without altering ecRNA expression did not significantly affect *Fos* promoter methylation status as measured by MeDIP (Supplementary Fig. 6). Finally, to examine the potential stability of ASO-mediated ecRNA knockdown, we treated neuronal cultures with *Fos* ecRNA ASO for up to 7 days (Supplementary Fig. 6). This experiment demonstrated enduring ecRNA knockdown for up to 1 week after ASO treatment. Furthermore, consistent with a mechanism by which DNA methylation accumulates at the *Fos* locus with continued ecRNA knockdown, we observed a progressive decline in *Fos* mRNA expression with prolonged treatment. Together, these results suggest that loss of ecRNA at a gene locus induces promoter hypermethylation and transcriptional silencing of the associated gene.

Fos ecRNA is required for long-term memory formation. Early transcriptional induction from the *Fos* gene locus after neuronal activity is a common feature across neuronal subtypes and functionally defined regions of the central nervous system (CNS). *Fos* protein forms a heterodimer with Jun family members to generate an AP-1 transcription factor, which regulates gene expression networks in response to stimulation^{24–26}. Although baseline levels of *Fos* mRNA are relatively low in adult animals^{27,28}, *Fos* mRNA transcription is upregulated following multiple types of memory formation^{4,9,29}. Further, mice with conditional deletion of the *Fos* gene in the adult CNS exhibit impaired hippocampal synaptic long-term potentiation and impaired performance on hippocampus-dependent long-term memory tasks, including contextual fear conditioning³⁰.

To examine the potential role for *Fos* ecRNA in long-term memory formation, we trained rats in a standard contextual fear conditioning task in which three 1 mA footshocks were paired with a novel context. One hour after training, the CA1 of the hippocampus was removed. RT-qPCR using RNA extracted from the CA1 revealed a significant increase in both *Fos* mRNA and *Fos* ecRNA as compared with an experimentally naïve group that did not undergo fear conditioning (Fig. 7a). This result suggests that ecRNA species can be induced in the adult brain in an experience-dependent manner. To examine the function of *Fos* ecRNA in memory formation, we stereotactically infused either vehicle or *Fos* ecRNA ASOs into the dorsal CA1 subregion of the hippocampus (Fig. 7b), followed by either RT-qPCR to verify *Fos* ecRNA knockdown (Fig. 7c), or by behavioural testing (Fig. 7d,e). Consistent with *in vitro* results, *Fos* ecRNA ASO treatment produced a significant decrease in *Fos* ecRNA levels in the CA1 of the hippocampus up to 7 days after infusion (Fig. 7c). Seven days after ASO treatment, rats underwent contextual fear conditioning followed by short-term (1h post-training) and long-term (24h post-training) tests for memory formation. *Fos* ecRNA ASOs did not alter baseline (pre-shock) freezing during initial the initial training session or short-term contextual fear memory. In contrast, *Fos* ecRNA knockdown resulted in a significant deficit in freezing responses observed on re-exposure to the conditioning chamber after 24 h (Fig. 7d), indicative of a selective impairment in long-term contextual fear memory formation. Importantly, ASO treatment did not alter performance in an open field test (Fig. 7e), suggesting that deficits in contextual fear memory were not due to altered locomotor activity or differences in anxiety-like behaviours.

Discussion

In this study, we provide novel genome-wide evidence for an association between ecRNA species and DNA methylation states in neuronal systems. We find that in the absence of ecRNA production, genes tend to possess methylated promoters and are silent. In contrast, genes that generate ecRNAs possess demethylated promoters and are actively transcribed. Using unbiased mathematical clustering, we observed that this relationship is preserved at actively transcribed neuronal genes, including genes modulated by neuronal activity and genes implicated in developmental and neurodegenerative disorders. This genome-wide characterization provides a key initial step in understanding how ecRNAs might regulate epigenetic states within neurons, and also provides insight into the biological context of this phenomenon across gene classes.

These results are consistent with previous observations that the epigenome is highly dynamic in the adult brain, and reveal one of the first potential mechanisms for activity- and experience-dependent alterations in cytosine methylation status at specific gene loci. In addition to providing a genome-wide

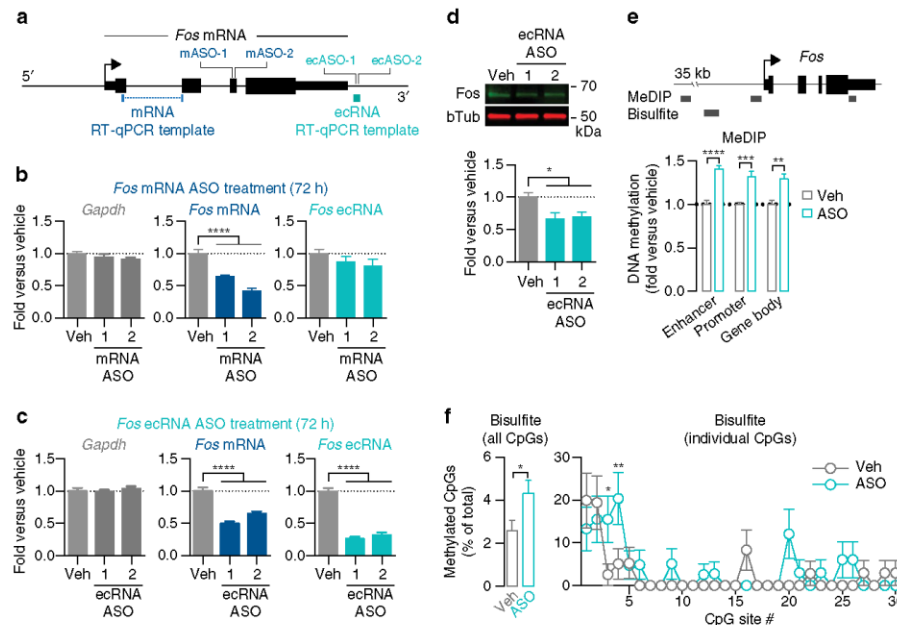


Figure 6 | *Fos* ecRNA controls *Fos* gene methylation. (a) Anti-sense oligonucleotide (ASO) target locations for *Fos* mRNA and ecRNA. (b) *Fos* mRNA ASOs decreased mRNA expression with no significant effect on ecRNA ($n = 8$ per group; one-way ANOVA for *Fos* mRNA, $F_{(2,21)} = 42.96$, $P < 0.0001$; Tukey's post hoc test for individual comparisons.). (c) *Fos* ecRNA ASOs decreased both mRNA and ecRNA ($n = 8$ per group; one-way ANOVAs for *Fos* mRNA and ecRNA, $F_{(2,21)} = 36.43$ and 95.48 , respectively, $P < 0.0001$; Tukey's post hoc test for individual comparisons.). (d) *Fos* ecRNA knockdown reduces *Fos* protein quantified by immunoblotting ($n = 6$ per group; one-way ANOVA, $F_{(2,15)} = 5.438$, $P = 0.0168$). (e) *Fos* ecRNA knockdown (72 h ASO treatment) resulted in increased enhancer, promoter and gene body methylation as measured by MeDIP ($n = 4$ per group; two-way ANOVA main effect of ASO, $F_{(1,18)} = 71.60$, $P < 0.0001$, Sidak's post hoc test for multiple comparisons). (f) Bisulfite sequencing following *Fos* ecRNA knockdown confirmed significant promoter hypermethylation (left, average of all CpG sites; 41–45 individual clones/group, χ^2 test, $z = 2.213$, $P = 0.0269$), driven largely by specific hypermethylation at two distinct CpG sites (right; methylation status between treatment groups compared with a two-way ANOVA with Sidak's post hoc tests adjusted for multiple comparisons at individual CpG sites). Data are expressed as mean \pm s.e.m. Individual comparisons, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ and $****P < 0.0001$.

characterization of activity-regulated transcriptional dynamics of both polyadenylated and non-polyadenylated RNA species in neurons, our results provide a detailed analysis of ecRNAs arising from the *Fos* gene locus. We show that *Fos* ecRNAs interact directly with DNMTs, that DNA methylation states at *Fos* are altered following ecRNA increases, that *Fos* ecRNAs undergo distinct synthesis and regulation from *Fos* mRNA, and that knockdown of *Fos* ecRNA increases DNA methylation and represses *Fos* mRNA. Furthermore, we reveal that *Fos* ecRNA is induced by the memory-forming experiences, and is necessary for expression of contextual fear memory. These findings highlight a potential role for ecRNAs in cognitive function, and further illustrate the dynamic link between active non-coding RNA transcription and epigenetic regulation in neurons.

One challenge in genome-scale identification and quantification of ecRNA species is that non-polyadenylated RNA fractions could also contain nascent unspliced pre-mRNA that has not yet been polyadenylated. Although previous reports have revealed widespread co-transcriptional splicing in neurons³¹, it is likely that nascent mRNA transcripts contributed in some way to the intronic ecRNA quantification strategy employed here. Though this issue is difficult to resolve using standard whole-genome sequencing approaches, more detailed single-molecule sequencing could be employed to capture longer transcripts and thus provide more complete disambiguation of ecRNA and pre-mRNA transcripts.

Critically, this analysis could also facilitate identification of ecRNA transcript start and end sites, as well as enriched sequence motifs that might confer ecRNAs with targeted DNA or protein recognition properties. Nevertheless, our characterization of ecRNA from the 3' end of the *Fos* gene locus is in agreement with previous observations that ecRNAs are regulated distinctly from overlapping mRNA and contribute to gene-specific DNA methylation patterns¹⁵. Based on our genome-wide gene expression and DNA methylation data, we hypothesize that ecRNA transcripts arising from other genes also possess differential biosynthesis and regulation. However, although our results suggest that activity-dependent induction of the *Fos* ecRNA is controlled by RNAPIII, it remains unclear if this is the case at other ecRNA transcription sites. Future studies will be required to explore differential polymerase contributions to transcription of functionally defined non-coding RNA subclasses.

Overall, our results add to an emerging appreciation for non-coding RNA species in epigenetic and transcriptional modulation³², including the ability to act as decoy molecules or scaffolds for transcriptional regulators, target specific modifiers to a unique DNA locus, and directly bind to DNA itself. We propose that ecRNAs directly interact with the DNMT catalytic domain to block the ability of DNMTs to target and silence overlapping genes. One prediction of this model is that ecRNAs would need to remain co-localized with the parent gene following synthesis to

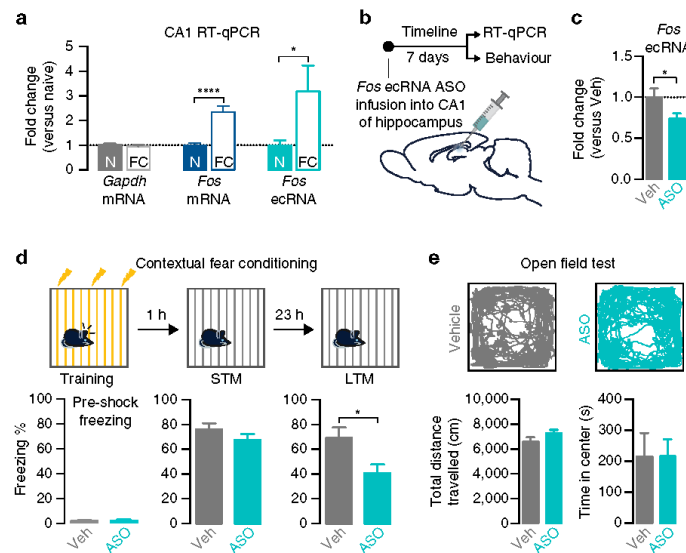


Figure 7 | Fos ecRNA is induced by behavioural experience and modulates memory formation. (a) Contextual fear conditioning is associated with increased Fos mRNA and Fos ecRNA in the CA1 subregion of the hippocampus, as compared with experimentally naïve controls (N = naïve, FC = fear conditioned; $n = 16$ per group; Mann–Whitney U test, $U = 15$ and $P < 0.0001$ for Fos mRNA, $U = 73$ and $P = 0.038$ for Fos ecRNA). Area CA1 was subdissected from total hippocampus 1 h following contextual fear conditioning. (b) Experimental timeline for Fos ecRNA ASO experiments *in vivo*. (c) Fos ecRNA ASO treatment decreased Fos ecRNA expression in the CA1 of the hippocampus ($n = 8$ –9 per group, unpaired Student's t -test, $t_{15} = 2.173$, $P = 0.0462$). (d) Top, contextual fear conditioning design. Bottom, Fos ecRNA ASO treatment impaired long-term memory but did not alter baseline freezing or short-term memory ($n = 8$ –9 per group; STM unpaired Student's t -test, $t_{15} = 1.245$, $P = 0.23$; LTM unpaired Student's t -test, $t_{15} = 2.640$, $P = 0.0186$). (e) Open field test. Top, representative traces showing animal location during 30 min test session. Fos ecRNA knockdown did not affect total distance travelled (bottom left; unpaired Student's t -test, $t_{14} = 1.517$, $P = 0.1514$) or time spent in the center of an open field test (bottom right; unpaired Student's t -test, $t_{14} = 0.005$, $P = 0.99$; $n = 8$ per group). All data are expressed as mean \pm s.e.m. Individual comparisons, * $P < 0.05$ and **** $P < 0.0001$.

provide locus-specific DNMT regulation. Although how this occurs is presently unknown, there are at least two distinct possibilities. One is that ecRNAs could possess extremely short half-lives (as has been demonstrated for enhancer RNAs (ref. 22)), in effect restricting function to a local chromatin environment surrounding the overlapping gene. However, a second and potentially more intriguing possibility is that ecRNAs could form a triplex structure with DNA (ref. 33) that serves the dual purpose of stabilizing the ecRNA and providing locus-specific anchor for interference of DNMT activity¹⁵. Regardless of the specific mechanism, characterizing the nature of this process will be important for understanding how individual cytosine nucleotides can be targeted for developmental or activity-dependent epigenetic reprogramming^{1,8}.

Our results also suggest the possibility that differential regulation of ecRNA species may serve as a unique means for endogenous control of the neuronal epigenome. Given that many neuropsychiatric and neurological diseases are associated with chronic alterations in specific gene products^{34–36}, we speculate that ecRNA targeting may provide an attractive therapeutic approach that could deliver gene-specific epigenetic reorganization to alleviate longstanding epigenetic pathologies. Indeed, co-opting endogenous mechanisms for epigenetic control represents an entirely new layer of therapeutic possibility, and could include the de-repression of methylated genes or the silencing of actively transcribed genes³⁷. Moreover, because cytosine methylation can be self-perpetuating^{38,39}, ecRNA-targeted alterations in gene methylation could conceivably persist beyond the initial triggering stimulus to generate enduring transcriptional regulation.

Methods

Cultured neuron experiments. Primary rat cortical neuronal cultures were generated from embryonic day 18 rat cortical tissue as described previously⁴. Cell culture wells were coated overnight at 37 °C with poly-L-lysine ($50 \mu\text{g ml}^{-1}$) and rinsed three times with dH_2O . Dissected cortices were incubated with papain for 20 min at 37 °C. After rinsing in Hank's Balanced Salt Solution, a single-cell suspension of the tissue was re-suspended in Neurobasal media (Invitrogen) by trituration through a series of large to small fire-polished Pasteur pipets. Primary neuronal cells passed through a $70\text{-}\mu\text{m}$ cell strainer were plated on poly-lysine coated culture wells. Cells were grown in Neurobasal media plus B-27 and L-glutamine supplement (complete Neurobasal media) for 8–11 days *in vitro* in a humidified CO_2 (5%) incubator at 37 °C.

At 8–11 days *in vitro*, neuronal cultures were treated as described. For KCl stimulation experiments, $25 \mu\text{l}$ of 1 M KCl (Sigma) or vehicle (neurobasal media alone) was added to cell culture wells to achieve a final concentration of 25 mM KCl. Cells were incubated with KCl for described time points before RNA extraction. For TTX inactivation experiments, cells were treated with $1 \mu\text{M}$ TTX (Tocris Bioscience) in neurobasal media for the described time points before RNA extraction. For AMPA and NMDA experiments, neuronal cultures were treated with the described concentrations of S-AMPA or NMDA (Sigma) for 1 h. S-AMPA and NMDA were diluted in sterile water and added to cultures at a volume of $10 \mu\text{l}$. $10 \mu\text{l}$ sterile water was added as a vehicle control. For experiments involving RNAP inhibitors, cultures were treated for 4 h (Fig. S4) or 4 h followed by a 1 h, 25 mM KCl stimulation (Fig. 4d,e). The RNAPII-dependent transcriptional inhibitor DRB and the RNAPIII inhibitor ML-60218 (Sigma) were dissolved to a 20-mM stock solution in 100% cell culture grade DMSO (Sigma) and diluted to neurobasal media to described experimental concentrations. Vehicle-treated cells received equal concentrations of DMSO in neurobasal media. At a minimum, all cell culture experiments were performed in triplicate.

RNA-seq. RNA-seq experiments using two biological replicates per condition were carried out at the Hudson Alpha Genome Services Laboratory. RNA was extracted, DNase-treated and purified (RNeasy, Qiagen). Two distinct RNA libraries were generated. Polyadenylated (PolyA+) RNA was captured with the NEBNext Poly(A) mRNA Magnetic Isolation Module. The remaining non-polyadenylated

(PolyA[−]) underwent ribosomal RNA depletion (NEBNext rRNA depletion kit). 2 µg of total RNA underwent quality control (Bioanalyzer; all RIN values > 9.5), and was prepared for directional RNA sequencing at Hudson Alpha using NEBNext reagents (New England Biolabs) according to manufacturer's recommendations with minor modifications (including the use of custom library adaptors and indexes). PolyA⁺ and PolyA[−] RNA libraries underwent sequencing (50 bp paired-end directional reads; ~25 M reads/sample) on an Illumina sequencing platform (HiSeq2000).

RNA-seq data analysis. Raw paired-end sequenced reads were quality controlled, filtered for read quality (PASTX toolkit, Galaxy) and aligned to the rat Rn5 genome sequence in Galaxy using Tophat v1.4.0 (with custom settings -p 8 -r 175). Overall, we obtained ~130 million total paired-end mapped PolyA⁺ reads and ~118 million mapped PolyA[−] reads from six independent biological replicates. Genome-aligned sequenced reads were examined in Seqmonk software release v0.28.0 (Babraham Institute), using Ensembl release v70 gene and feature annotations with modifications to update 3'UTR annotation (as described below). For all sequencing analysis, we filtered gene annotations to exclude Ensembl predicted/model genes and short genes (<1 kb), leaving 17,719 genes for characterization. For independent PolyA⁺ samples, transcript expression levels were determined by computing the FPKM. Gene expression differences between groups (Veh versus KCl) were calculated in Seqmonk using the Intensity Difference function, correcting for multiple testing. Statistical significance was assessed using Student's *t*-tests and a false-discovery rate of 0.05. Comparisons and analysis in Figs 1 and 3, and S1 was performed with data sets from unstimulated cells (vehicle group).

To characterize the genome-wide expression of ecRNA-like transcripts, we computed FPKM values at every gene in three distinct locations: (1) 5', pre-transcription start site (−1,000 bp upstream to TSS), (2) introns and (3) post-TES (TES to 2,000 bp past TES). For this analysis, we included only PolyA[−] reads from the 'sense' strand of a given gene, to exclude anti-sense transcripts. For genes containing multiple introns, composite (mean) FPKM values were computed to quantify intronic transcription across the gene. Intronic genes did not receive a value for the 'intron' quantification. Pre-TSS, intronic and post-TES FPKM values were averaged to obtain an 'ecRNA index', which provided an overall quantification for ecRNA-like transcription from every gene.

De Novo 3'UTR annotation. Recent reports have revealed robust 3'UTR extension and lengthening of neuronal genes^{40–42}, which are frequently not annotated in available rat genome assemblies. Given that 3'UTR extensions could impair accurate quantification of 3' ecRNAs across the genome, we performed a *de novo* annotation of 3'UTR end sites throughout the genome. Our 3'UTR identification pipeline (based on⁴⁰) utilized a merged library consisting of ~130 million total paired-end mapped PolyA⁺ reads from vehicle, KCl and TTX treated cells. We first identified contiguously transcribed regions of the genome using Seqmonk software, with a fivefold (21 × read depth) coverage cutoff. To identify only meaningful 3'UTR extensions, we limited our annotation to regions at least 200 bp in length, and merged contiguous regions separated by <150 bp. Given that our libraries were directional, we separated forward and reverse strands for this analysis and considered transcripts arising from each strand separately. Genome-wide, this analysis identified 52,651 regions of contiguous transcription. To identify 3'UTR extensions, we filtered this list for regions that overlapped existing transcription end annotations for our entire gene set, which identified 5,801 potential candidate 3'UTR extensions. Following additional filtering to remove regions that overlapped the start site of adjacent genes, regions from internally transcribed genes and potential anti-sense transcripts, our list of 3'UTR extensions was reduced to 4,926 genes. Visual inspection and overall global alignments indicated precise 3'UTR annotation following this pipeline, and correctly identified many previously observed 3'UTR extensions, including the 18.7 kbp 3'UTR extension at *Grin2b*, the 16.3 kbp extension at *Ntrk3*, as well as extensions at *Neddl4*, *Dnajc15* and *Hmbox1* (ref. 40). Overall, this analysis added ~8.3 Mbp to the annotated transcriptome, with an average extension of 1,703 bp at the 4,926 genes with new 3'UTR termination annotations. All 3'ecRNA characterizations were performed with these updated annotations.

Even with corrected 3'UTR annotations, it is still possible that ecRNA estimates could be skewed due to another type of assembly fault (for example, mislocated exons or exons expressed only in neurons), the presence of overlapping sense-strand genes, or the inclusion of inactive genes that bias genome-wide correlations. To address each of these concerns, we performed an additional analysis designed to estimate the normalized level of ecRNA, compared with PolyA⁺ RNA from the same loci. This analysis revealed the same relationship between ecRNA levels and promoter DNA methylation (Supplementary Fig. 7).

MBD-seq and capture verification. Genome-wide DNA methylation patterns were quantified using MBD protein capture (MBD-capture; MethylMiner Kit, Invitrogen), followed by next-generation sequencing. DNA from unstimulated neuronal cultures was extracted, RNase treated and purified (DNeasy, Qiagen). 2 µg of genomic DNA was sonicated to 200–400 bp (Bioruptor Pico, Diagenode). Methylated DNA was collected with recombinant MBD2 protein/biotin complex,

which was purified using streptavidin-coated magnetic beads (Invitrogen). DNA sequencing was performed at Hudson Alpha using NEBNext reagents (New England Biolabs) according to manufacturer's recommendations with minor modifications (including the use of custom library adaptors and indexes). DNA libraries were quantified with the Kapa Library Quant Kit (Kapa Biosystems), and underwent sequencing (25 M total 50 bp single-end reads) on an Illumina sequencing platform (HiSeq2000). We sequenced two biological replicates as well as an input (non-IP) control for normalization. To ensure that MBD-capture resulted in the specific enrichment of methylated DNA, we performed control reactions in which gDNA was spiked with synthetic methylated and non-methylated DNA fragments (1 pg each, Methyl Miner kit, Invitrogen) before immunoprecipitation with recombinant MBD2. Methylated and non-methylated DNA capture was quantified via RT-qPCR with primers specific for each synthetic sequence, using both the captured (MBD2-bound) and unbound fractions. Our results demonstrated robust enrichment (~600-fold) of methylated DNA fragments in the captured sample, and equally robust depletion of methylated DNA in the unbound fraction (Supplementary Fig. 2b).

MBD-seq data analyses. Raw single-end sequenced reads were quality controlled, filtered for read quality (PASTX toolkit, Galaxy) and aligned to the rat genome (Rn5 assembly) in Galaxy using Bowtie. Overall, we obtained ~44 M mapped single-end reads from MBD-IP samples, and ~20 M mapped reads from an input control sample. Genome-aligned sequenced reads were examined using Seqmonk (Babraham Institute). For each sample, methylation levels were examined relative to gene elements using the built-in analysis pipelines. Promoter methylation was quantified relative to the TSS by computing the reads per million for a 5-kb window surrounding the TSS (−2.5 to 2.5 kb). To remove potential confounds arising from DNA sequence mappability, all data were corrected by subtracting reads per million values from the input (non-IP) control. This pipeline resulted in highly reproducible methylation quantification at both gene bodies (Supplementary Fig. 2c) and gene promoters (Supplementary Fig. 2d). Additional analysis based on CpG identification (Supplementary Fig. 3) was performed using annotated CpG island boundaries.

Hierarchical clustering and gene ontology. Hierarchical clustering analysis and dendrogram generation (Fig. 3) was performed in Matlab using the clustergram toolbox. Clustering was performed using percentile scores instead of raw data to equalize data ranges between columns. Each row was sorted based on Euclidean distance metrics and average linkage to generate a hierarchical tree. To integrate this information with existing data sets, gene lists corresponding to autism spectrum risk genes, Alzheimer's disease risk genes, or genes altered in a mouse model of Alzheimer's disease (Fig. 3b) were obtained from published reports^{34,43,44}.

Gene ontology analysis of gene clusters identified in using this approach was performed using the ClueGO plugin in Cytoscape⁴⁵. Enrichment analysis was conducted using a reference set of all 17,719 genes. Significantly enriched biological process terms (hierarchy levels 8–15) containing at least 10% of genes in each category were identified using a Benjamini-Hochberg false-discovery rate and $\alpha = 0.001$. For comparison to previously published data sets, human and mouse gene names were mapped onto rat gene orthologues according to gene symbols. Only genes with identical rat, mouse and human symbols are shown here, resulting in the exclusion of some genes that were found in original lists but did not map to rat gene symbols.

RNA quantification with RT-qPCR. Total RNA was extracted using the RNeasy Mini kit (Qiagen) following the manufacturer's instructions. All samples were treated with DNaseI during column purification to remove contaminating DNA. RNA was reverse transcribed using the iScript RT-PCR kit (Bio-Rad). PCR amplification was performed in triplicate (cell culture experiments) or with six replicates (*in vivo* experiments) using a CFX96 real-time PCR system (Bio-Rad) at 95 °C for 3 min, followed by 40 cycles of 95 °C for 10 s and 58 °C for 30 s, followed by real-time melt analysis to verify product specificity. *Gapdh* (cell culture experiments) or *Actb* (*in vivo* experiments) were used as internal controls for normalization using the $\Delta\Delta Ct$ method⁴⁶.

Chromatin immunoprecipitation. RNAPIII binding at specific genomic loci was assayed using available ChIP protocols with minor modifications^{47,48}. Briefly, cultured cells were treated as described and immediately fixed in 1% paraformaldehyde, washed in phosphate-buffered saline (PBS), lysed and then sonicated (Bioruptor Pico, Diagenode) to shear DNA to 200–500 bp fragments. Sheared, cross-linked DNA was incubated with 5 µl RNAPIII antibody (RPC32 subunit, #sc-21754, Santa Cruz Biotechnology) and 25 µl protein-A coated magnetic beads (Invitrogen) overnight, washed sequentially in low salt, high salt, LiCl and tris-EDTA (TE) buffers, and then incubated for 2 h at 65 °C in TE buffer containing 1% SDS and proteinase K solution (Qiagen) to reverse crosslinks. Following magnetic removal of protein-A coated beads, extracted DNA was then purified (Qiagen DNA Mini Spin Column), and RNAPIII binding levels at the *Fos* gene promoter were assayed via qPCR as described above. Ct values for IP samples were normalized to unprocessed (input) DNA, which was not incubated with RNAPIII antibody.

DNMT/MAP2 immunostaining. To verify expression of DNMT isoforms in neuronal cultures, we performed immunolabeling for DNMT1 and DNMT3a. After removal of neuronal culture media, cells were washed with PBS and incubated at room temperature for 20 min in freshly prepared 4% paraformaldehyde in PBS. After fixation, cells were washed three times with PBS and neuronal membranes were permeabilized with PBS containing 0.25% Triton X-100 for 15 min at room temperature. Cells were then washed three times in PBS, blocked for 1 h (10% Thermo Blocker bovine serum albumin (BSA) #37525, 0.05% Tween-20, and 300 mM glycine in PBS) and co-incubated with DNMT (1:1,000 in PBS with 10% Thermo Blocker BSA #37525, Abcam anti-DNMT1 (ab87656) or anti-DNMT3a (ab2850)) and MAP2 antibodies (1:250, Anti-MAP2 Alexa Fluor 555 Conjugate, Invitrogen) at 4 °C overnight. Cells were washed three times in PBS and incubated for 1 h at room temperature with a fluorescent secondary antibody (Alexa 488 goat anti-rabbit, Invitrogen; 1:250 in PBS with 10% Thermo Blocker BSA #37525), washed three times with PBS, and mounted onto microscope slides with Prolong Gold anti-fade medium (Invitrogen) containing 4,6-diamidino-2-phenylindole stain as a marker for cell nuclei. Immunostaining experiments were performed in duplicate.

RNA immunoprecipitation. RIP with control and DNMT antibodies was performed based on a previously published protocol¹⁵, with minor modifications. Day 1: primary neuronal cultures (~250,000 neurons per culture well) were cross-linked with 1% formaldehyde in PBS, freshly supplemented with protease/phosphatase inhibitor cocktail (1/100th volume of Halt inhibitor cocktail (Pierce)), RNase inhibitor (40 U RNasin (Promega) per ml of buffer) and 2 μ M vanadyl complex. After incubation for 10 min at room temperature on a rocking platform, 2 M glycine was added to each well to a final concentration of 0.2 M to stop crosslinking, and rested on ice for 10 min. Cells were washed three times with ice-cold PBS (freshly supplemented with Halt cocktail, RNase inhibitor and vanadyl complex). Cells were lysed in RIP lysis buffer (50 mM HEPES, 10 mM NaCl, 1 mM EDTA, 0.5% NP-40, freshly supplemented with Halt cocktail, RNase inhibitor and vanadyl complex). Cells were detached with a cell scraper and placed on ice for 10 min to complete lysis. 1.2×10^7 cells were collected and dounce homogenized (10 strokes with pestle A, 40 strokes with pestle B) and centrifuged for 10 min at 2,500g at 4 °C to pellet cell nuclei. After resuspension in RIP resuspension buffer (50 mM HEPES, 10 mM MgCl₂, freshly supplemented with 1X Halt cocktail, RNase inhibitor, and vanadyl complex), samples were sheared by sonication (three times for 5 s at 4 °C with BioRuptor Pico, with 30 s off time between pulses). DNase treatment was performed at 37 °C for 30 min, and EDTA was added to a concentration of 20 mM to stop the reaction. After adding RIP dilution buffer (50 mM HEPES, 150 mM NaCl, 1 mM EDTA, 0.5% NP-40, 0.1% Triton X-100, 0.1% sodium deoxycholate, freshly supplemented with Halt cocktail, RNase inhibitor and vanadyl complex), 100 μ l of the sample was removed as the input control.

Immunoprecipitation for RIP was performed as follows. Samples were pre-cleared for 1 h with 25 μ l fully re-suspended magnetic beads (Protein-A coated Dynabeads Beads, Invitrogen) at 4 °C on a rocking platform. Beads were removed in a magnetic field, and the remaining sample was divided equally for three IP reactions. Antibodies for IP reactions were: 5 μ g IgG control (Rabbit IgG, Abcam ab46540), 5 μ g DNMT1 antibody (Abcam ab87656) and 5 μ g DNMT3a antibody (Abcam ab2850). Samples were incubated overnight on a rocking platform at 4 °C.

Day 2: 25 μ l Protein-A coated Dynabeads were added and incubation was carried for 1 h at 4 °C on a rocking platform. Dynabeads were separated in a magnetic field and supernatant (unbound fraction) was removed. Immuno-RNA complexes were washed six times with RIP dilution buffer. After the final wash, beads were re-suspended in RIP reversal buffer (100 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 1% SDS, freshly supplemented with RNase inhibitor and vanadyl complex) and 20 μ l Proteinase K was added to each sample. Incubation was performed for 1 h at 42 °C, then 1 h for 65 °C. Supernatant (containing RNA) was removed and RNA was extracted with miRNeasy kit (Qiagen). DNase treatment was performed on column and RNA was eluted in 30 μ l buffer EB (Qiagen). RT-qPCR was used to determine differences in RNA levels (compared with IgG control).

RNA electrophoretic mobility shift assay (REMSA). Mobility shift assays were conducted with custom synthetic Cy5.5-labelled 25-base RNA oligonucleotides (Sigma) and specified concentrations of full-length recombinant human DNMT1 protein (Sigma, SRP0126), recombinant human DNMT3a and DNMT3l protein (Sigma, SRP0396), or truncated murine DNMT3a protein containing only the catalytic methyltransferase domain (DNMT3a-CD, purified as previously described⁴⁹). RNA oligonucleotides (1 nM) were incubated with DNMT protein (0–0.2 μ M), IgG protein (0.2 μ M), or 1 nM–100 nM poly-dI-dC (Sigma) in REMSA buffer (20 mM HEPES, 40 mM KCl, 1 mM EDTA, 0.2 mM DTT, 0.1 mg ml⁻¹ BSA, 0.1% Tween-20, and 20% Glycerol) for 1 h at 37 °C, and REMSA was performed using native polyacrylamide gel electrophoresis (Bio-Rad Mini-Protein TGX Precast 4–20% gel). Electrophoretic mobility of Cy5.5-labelled RNA was assayed using fluorescence imaging on the Odyssey Infrared Imaging System (Li-Cor Biosciences). RNA-DNMT3a complex formation was quantified as the Cy5.5 signal intensity appearing at the higher band (corresponding to DNMT-bound RNA with lower electrophoretic mobility) divided by the total signal intensity (bound RNA plus free probe). Binding affinity (K_d) was calculated using non-linear regression (one-site-specific binding) in Prism 6.0 software (Graphpad), constraining B_{max} equal to the theoretical maximum of 100% binding. Synthetic RNA probes showed

no binding affinity for isomolar concentrations of proteins with similar mass (rabbit IgG, 150 kDa) or charge (BSA, pI = 6.17) as full-length human DNMT3a (101 kDa, pI = 6.57).

In vitro methylation assay. *In vitro* methylation assays were performed with 10 ng of a 189-bp gDNA template from the *Fos* promoter (containing 15 CpG sites and amplified from *Fos* gDNA MeDIP primers). DNA was incubated for 1 h at 37 °C with 0.06 μ M DNMT3a-3l (Sigma, SRP0396) and 0.5 mM S-(5'-Adenosyl)-L-methionine (SAM, Sigma) in REMSA buffer alongside DNA, protein and SAM only controls. For the methylation reaction with the added synthetic ecRNA, 5.16 μ M ecRNA (ecRNA-1 oligonucleotide sequence) was added to account for the size differences between the RNA (25 nt, single-stranded) and DNA (189 bp, double-stranded) template and to provide ~15x coverage of RNA for each DNA molecule. Purified DNA (0.2 ng) from each reaction was mixed with 500 ng sonicated rat genomic DNA to reduce non-specific binding and denatured at 95 °C for 10 min. Each reaction was diluted with 1x bind/wash buffer (0.5% NP-40, 1.1% TritonX100, 1.5 mM EDTA, 50 mM Tris-HCl, and 150 mM NaCl), and an input control was removed. To quantify DNA methylation, MeDIP was carried out with 10 μ l magnetic beads (Protein-A coated Dynabeads Beads, Invitrogen) and 1 μ l 5 mC antibody (Epigentek, #A-1014) for 1 h at 4 °C with rotation. Beads were washed 3 \times with 1 \times bind/wash buffer for 5 min/wash and re-suspended in buffer EB with 1% SDS. Proteinase K (Qiagen) was added and reactions were incubated at 50 °C for 1 h then 95 °C for 10 min to reverse binding. Beads were removed and IP reactions and inputs were purified using a PCR Purification Kit (Qiagen). Purified DNA was subjected to qPCR, and methylation levels were computed as a fraction of input DNA.

RNA secondary structure prediction. Secondary structure of synthetic RNA oligonucleotides (Supplementary Fig. 5) was predicted using RNAfold software⁵⁰ (via web server hosted at the University of Vienna), using minimum free energy and partition function algorithms.

Anti-sense oligonucleotide (ASO) design and treatment. To manipulate *Fos* mRNA or ecRNA levels, we designed four 20 bp ASOs that targeted distinct transcripts from the *Fos* gene locus (Fig. 6a; Supplementary Data 3). ASOs targeting exon 3 of *Fos* mRNA or *Fos* 3' ecRNA were synthesized with two chemical modifications: an all phosphorothioate backbone and five 2' O-methyl RNA bases on each end of the oligonucleotide (Integrated DNA Technologies). These modifications have been shown to enhance ASO stability and improve affinity for the target RNA molecule⁵¹, and have successfully been employed for targeted manipulations in animal models of degenerative and developmental disorders^{52,53}. Primary neuronal cultures were treated with vehicle (15 μ l buffer EB, Qiagen) or ASO (15 μ l in buffer EB, for a final concentration of 1.5 μ M) and incubated for 72 h. Following ASO treatment, RNA was extracted (Qiagen RNeasy kit) and *Fos* mRNA and ecRNA levels were determined using RT-qPCR with custom primers (Supplementary Data 3). Gene expression was determined by normalizing to *Gapdh* ($2^{-\Delta\Delta C_t}$).

Methylated DNA immunoprecipitation. MeDIP was performed using a 5-methylcytosine antibody (4 μ l per sample, mouse monoclonal, Epigentek #A-1014) as described previously^{42,7}, with minor modifications. Genomic DNA was extracted (DNeasy Blood and Tissue Kit, Qiagen), treated with RNase A and quantified (Quant-IT HS dsDNA kit, Invitrogen) using the manufacturer's recommended protocols. 300 ng of DNA per sample was removed and sonicated (Bioruptor Pico, Diagenode) to 200– to 800-bp fragments for methylation analysis. Fragmented DNA was incubated for 1 h with 4 μ l 5 mC antibody and methylated DNA was collected with protein-A – coated Dynabeads (Invitrogen), washed (1x wash buffer, MethylMiner kit, Invitrogen), extracted for 2 h at 60 °C with proteinase K in TE buffer with 1% SDS, and purified (Qiagen DNeasy kit). To ensure adequate enrichment of methylated DNA fragments in each sample, we spiked samples with synthetic methylated and unmethylated control DNA fragments (1 pg each, Methyl Miner kit, Invitrogen) before MeDIP. Methylation at selected DNA regions was assayed via qPCR on a CFX96 real-time PCR system (Bio-Rad) with Sso Advanced chemistry. Ct values for IP samples were normalized to *Gapdh*, which did not change across samples. There were no differences in genomic DNA from input samples. There was no difference in methylated DNA enrichment between groups. The mean enrichment for synthetic methylated DNA spike-in controls over non-methylated fragments was >150-fold for each sample.

Bisulfite sequencing. DNA methylation at the *Fos* gene promoter following *Fos* ecRNA ASO treatment was assayed using bisulfite sequencing. Genomic DNA was extracted (DNeasy Blood and Tissue Kit, Qiagen), treated with RNase A and quantified (Quant-IT HS dsDNA kit, Invitrogen) using the manufacturer's recommended protocols. 450 ng of DNA per sample underwent bisulfite conversion (EZ DNA Methylation Lightning kit, Zymo) and PCR amplification using bisulfite-compatible primer sets targeting the *Fos* promoter locus (Fig. 6e). PCR products were cloned using the TOPO-TA cloning system (Invitrogen) and sequenced using Sanger sequencing (UAB Hefflin Genomics Core). 10–15 individual clones were sequenced per conversion, and the experiment was performed in quadruplicate (resulting in 41–45 clones/group).

Western blotting. To determine whether ecRNA-mediated knockdown of *Fos* mRNA also resulted in a loss of Fos protein, we performed immunoblotting after 72 h incubation with *Fos* ecRNA ASOs. For each sample, 250,000 cells were washed with ice-cold Tris-buffered saline, lysed in 100 μ l RIPA lysis buffer (50 mM Tris-HCl, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS and 1X Halt protease and phosphatase inhibitor (Pierce)) for 30 min, and sonicated to reduce sample viscosity (Bioruptor Pico (Diagenode); 3 \times 5 s, 30 s off time between pulses). Protein sample (20 μ l) was boiled at 95 °C for 5 min with 4 \times Laemmli buffer (Bio-Rad), separated on a 15% polyacrylamide gel, and transferred to a polyvinylidene difluoride membrane. Fos protein was detected with a rabbit polyclonal anti-Fos antibody (1:200; Santa Cruz #D2513), and imaged on an Odyssey Infrared Imaging System (Li-Cor Biosciences) using a goat anti-rabbit secondary (1:10,000; IR dye 800, Li-Cor Biosciences #827-08365). As a loading control, β -Tubulin was detected using a mouse anti- β -Tubulin antibody (1:1,000; Millipore #05-661) and imaged using a goat anti-mouse secondary antibody (1:10,000; IR dye 680, Li-Cor Biosciences #926-68,170). Protein levels were quantified in Odyssey Application Software v3.0 (Li-Cor Biosciences), and Fos intensity values were normalized to β -Tubulin for analysis.

Animals. Male Sprague-Dawley rats, ~90–120-day-old and weighing 250–350 g, were co-housed in plastic cages in an AAALAC-approved animal care facility on a 12-h light/dark cycle with food and water available *ad libitum*. All experiments were performed with naïve animals in the light phase of the light/dark cycle/MBD-seq. All procedures were performed in accordance with the University of Alabama at Birmingham Institutional Animal Care and Use Committee. All animals were randomly assigned to respective groups.

Stereotaxic surgery. Naïve adult Sprague-Dawley rats were anaesthetized with 4% isoflurane and secured in a stereotaxic apparatus (Kopf Instruments). During surgical procedures, an anaesthetic plane was maintained with 1–2.5% isoflurane. Under aseptic conditions, guide holes were drilled using stereotaxic coordinates (anteroposterior, -3.3 mm from bregma, ± 2.0 mm lateral from midline) to target the dorsal CA1 region of the hippocampus. All infusions were made using a gastight 30-gauge stainless steel injection needle (Hamilton Syringes) that extended into the infusion site (2.5 mm ventral to bregma). Bilateral microinfusions of 1 μ l solution were made using a syringe pump (Harvard Apparatus) at a rate of 0.25 μ l min⁻¹. Injection needles remained in place for 5 min following infusion to allow for diffusion. Rats were infused bilaterally with either 1.5 μ M of ASO in sterile saline or a vehicle control (saline alone) into the CA1 of the hippocampus. After infusions, guide holes were covered with sterile bone wax and surgical incision sites were closed with nylon sutures. Animals received buprenorphine for pain management.

Contextual fear conditioning. To examine the role of *Fos* ecRNA in the adult brain in the context of learning and memory, we performed *in vivo* experiments with *Fos* ecRNA ASO-1, which demonstrated the most robust reduction of *Fos* ecRNA in neuronal cultures. *Fos* ecRNA ASO was delivered directly to the CA1 subregion of rat hippocampus via stereotaxic microinfusion. Knockdown was verified at 7 day following infusions with RT-qPCR on RNA extracted from dorsal CA1 tissue punches. Following recovery from surgery, animals were habituated to handling for 2 days before fear conditioning.

For both ASO and non-ASO experiments, contextual fear conditioning was conducted as previously described²³. Animals were placed in a standard behavioural chamber (Med Associates) for a 7-min behavioural conditioning session. This session consisted of three electric shocks (1 s, 1 mA each) delivered to a metal floor grid every 2 min from the start of the behavioural session. After the final shock, animals remained in the chamber for 1 min. This training paradigm has previously been shown to induce robust and long-lasting contextual fear memory. To examine induction of *Fos* mRNA and ecRNA following fear conditioning, animals were killed at 1 h following behavioural training. Brains were rapidly removed and the CA1 of the hippocampus was dissected away from other hippocampal subregions for downstream analysis. To examine the effect of *Fos* ecRNA ASO treatment, we tested memory performance at 1 h (short-term memory) and 24 h (long-term memory) after the completion of training. For memory tests, animals were returned to the fear conditioning chamber and allowed to explore the environment for 7 min. During this time, we monitored freezing behaviour as a readout of fear-related memory, which was captured using high-speed video recording. Freezing behaviour was quantified as the time spent immobile, and was manually scored by three experimenters that were blind to treatment group.

Open field test. A standard open field arena (43 \times 43 cm; Med Associates) was used to assess locomotor activity and anxiety-like behaviour. Seven days following vehicle or *Fos* ecRNA ASO infusions into the CA1 of the hippocampus, rats were placed in the open field arena for a 30-min period. Activity was tracked using automated video tracking software (CinePlex Studio, Plexon Inc.). Distance travelled (in cm) was used to quantify total locomotor activity, and time spent in the center of the open field (defined as 18 cm square in the middle of the field) was used to quantify anxiety-like behaviour.

Statistical analyses. Transcriptional and epigenetic differences from PCR experiments were compared with one- or two-way ANOVA with Tukey or Sidak *post hoc* tests, or *t*-tests where appropriate. Bisulfite sequencing data was analysed using a χ^2 test with Yates' correction to compare the difference in methylated CpGs between vehicle and ASO treatment, or with a two-way ANOVA followed by Sidak's *post hoc* test (with correction for multiple comparisons at individual CpG sites). Genome-wide correlations between ecRNA quartile and mRNA or promoter DNA methylation were performed with a one-way ANOVA followed by a *post hoc* test for a linear trend between column mean and left-to-right column order. Statistical significance was designated at $\alpha = 0.05$ for all analyses. Statistical and graphical analyses were performed with Graphpad software (Prism). Where necessary, statistical assumptions (for example, normality for parametric tests) were formally tested. Graphical analyses were performed in Matlab (v. R2014b) and DeepTools⁵⁴ (via Galaxy server).

Data availability. Sequencing data that support the findings of this study have been deposited in Gene Expression Omnibus (GEO) with the accession number GSE64988 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE64988>). All relevant data that support the findings of this study are available on request from the corresponding author (J.J.D.).

References

1. Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).
2. Miller, C. A. *et al.* Cortical DNA methylation maintains remote memory. *Nat. Neurosci.* **13**, 664–666 (2010).
3. Miller, C. A. & Sweatt, J. D. Covalent modification of DNA regulates memory formation. *Neuron* **53**, 857–869 (2007).
4. Day, J. J. *et al.* DNA methylation regulates associative reward learning. *Nat. Neurosci.* **16**, 1445–1452 (2013).
5. Peng, J. *et al.* Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat. Neurosci.* **13**, 423–430 (2010).
6. Roth, T. L., Lubin, P. D., Funk, A. J. & Sweatt, J. D. Lasting epigenetic influence of early-life adversity on the BDNF gene. *Biol. Psychiatry* **65**, 760–769 (2009).
7. Weaver, I. C. *et al.* Epigenetic programming by maternal behavior. *Nat. Neurosci.* **7**, 847–854 (2004).
8. Guo, J. U. *et al.* Neuronal activity modifies the DNA methylation landscape in the adult brain. *Nat. Neurosci.* **14**, 1345–1351 (2011).
9. Halder, R. *et al.* DNA methylation changes in plasticity genes accompany the formation and maintenance of memory. *Nat. Neurosci.* **19**, 102–110 (2016).
10. De Jager, P. L. *et al.* Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat. Neurosci.* **17**, 1156–1163 (2014).
11. Lunnon, K. *et al.* Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat. Neurosci.* **17**, 1164–1170 (2014).
12. LaPlant, Q. *et al.* Dnmt3a regulates emotional behavior and spine plasticity in the nucleus accumbens. *Nat. Neurosci.* **13**, 1137–1143 (2010).
13. Iwamoto, K. *et al.* DNA methylation status of SOX10 correlates with its downregulation and oligodendrocyte dysfunction in schizophrenia. *J. Neurosci.* **25**, 5376–5381 (2005).
14. Masliah, E., Dumaop, W., Galasko, D. & Desplats, P. Distinctive patterns of DNA methylation associated with Parkinson disease: identification of concordant epigenetic changes in brain and peripheral blood leukocytes. *Epigenetics* **8**, 1030–1038 (2013).
15. Di Ruscio, A. *et al.* DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* **503**, 371–376 (2013).
16. Watanabe, T. *et al.* Role for piRNAs and noncoding RNA in *de novo* DNA methylation of the imprinted mouse *Rasgr1* locus. *Science* **332**, 848–852 (2011).
17. Mohammad, F. *et al.* Long noncoding RNA-mediated maintenance of DNA methylation and transcriptional gene silencing. *Development* **139**, 2792–2803 (2012).
18. Holz-Schietinger, C. & Reich, N. O. RNA modulation of the human DNA methyltransferase 3A. *Nucleic Acids Res.* **40**, 8550–8557 (2012).
19. Mellen, M., Ayata, P., Dewell, S., Kriakou, S. & Heintz, N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151**, 1417–1430 (2012).
20. Dalton, R. P., Lyons, D. B. & Lomvardas, S. Co-opting the unfolded protein response to elicit olfactory receptor feedback. *Cell* **155**, 321–332 (2013).
21. Wu, J. Q. & Snyder, M. RNA polymerase II stalling: loading at the start prepares genes for a sprint. *Genome Biol.* **9**, 220 (2008).
22. Schaukowitz, K. *et al.* Enhancer RNA facilitates NELF release from immediate early genes. *Mol. Cell* **56**, 29–42 (2014).
23. Saha, R. N. *et al.* Rapid activity-induced transcription of Arc and other IEGs relies on poised RNA polymerase II. *Nature Neurosci.* **14**, 848–856 (2011).
24. O'Shea, E. K., Rutkowski, R. & Kim, P. S. Mechanism of specificity in the Fos-Jun oncoprotein heterodimer. *Cell* **68**, 699–708 (1992).

25. Chiu, R. *et al.* The c-Fos protein interacts with c-Jun/AP-1 to stimulate transcription of AP-1 responsive genes. *Cell* **54**, 541–552 (1988).
26. Malik, A. N. *et al.* Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. *Nat. Neurosci.* **17**, 1330–1339 (2014).
27. Zovkic, I. B., Paulukaitis, B. S., Day, J. J., Etikala, D. M. & Sweatt, J. D. Histone H2A.Z subunit exchange controls consolidation of recent and remote memory. *Nature* **515**, 582–586 (2014).
28. Graff, J. *et al.* Epigenetic priming of memory updating during reconsolidation to attenuate remote fear memories. *Cell* **156**, 261–276 (2014).
29. Alberini, C. M. Transcription factors in long-term memory and synaptic plasticity. *Physiol. Rev.* **89**, 121–145 (2009).
30. Fleischmann, A. *et al.* Impaired long-term memory and NR2A-type NMDA receptor-dependent synaptic plasticity in mice lacking c-Fos in the CNS. *J. Neurosci.* **23**, 9116–9122 (2003).
31. Ameur, A. *et al.* Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat. Struct. Mol. Biol.* **18**, 1435–1440 (2011).
32. Geisler, S. & Collier, J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat. Rev. Mol. Cell Biol.* **14**, 699–712 (2013).
33. Martianov, I., Ramadass, A., Serra Barros, A., Chow, N. & Akoulitchev, A. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* **445**, 666–670 (2007).
34. Gjonneska, E. *et al.* Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature* **518**, 365–369 (2015).
35. Kohen, R., Dobra, A., Tracy, J. H. & Haugen, E. Transcriptome profiling of human hippocampus dentate gyrus granule cells in mental illness. *Transl. Psychiatry* **4**, e366 (2014).
36. Hwang, Y. *et al.* Gene expression profiling by mRNA sequencing reveals increased expression of immune/inflammation-related genes in the hippocampus of individuals with schizophrenia. *Transl. Psychiatry* **3**, e321 (2013).
37. Day, J. J., Kennedy, A. J. & Sweatt, J. D. DNA methylation and its implications and accessibility for neuropsychiatric therapeutics. *Annu. Rev. Pharmacol. Toxicol.* **55**, 591–611 (2015).
38. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
39. Holliday, R. & Pugh, J. E. DNA modification mechanisms and gene activity during development. *Science* **187**, 226–232 (1975).
40. Miura, P., Shenker, S., Andreu-Agullo, C., Westholm, J. O. & Lai, E. C. Widespread and extensive lengthening of 3' UTRs in the mammalian brain. *Genome Res.* **23**, 812–825 (2013).
41. Hilgers, V. *et al.* Neural-specific elongation of 3' UTRs during Drosophila development. *Proc. Natl Acad. Sci. USA* **108**, 15864–15869 (2011).
42. Pal, S. *et al.* Alternative transcription exceeds alternative splicing in generating the transcriptome diversity of cerebellar development. *Genome Res.* **21**, 1260–1272 (2011).
43. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
44. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
45. Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).
46. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408 (2001).
47. Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).
48. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of *in vivo* protein-DNA interactions. *Science* **316**, 1497–1502 (2007).
49. Gowher, H. & Jeltsch, A. Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases. *J. Biol. Chem.* **277**, 20409–20414 (2002).
50. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31**, 3429–3431 (2003).
51. Southwell, A. L., Skotte, N. H., Bennett, C. F. & Hayden, M. R. Antisense oligonucleotide therapeutics for inherited neurodegenerative diseases. *Trends Mol. Med.* **18**, 634–643 (2012).
52. Kordasiewicz, H. B. *et al.* Sustained therapeutic reversal of Huntington's disease by transient repression of huntingtin synthesis. *Neuron* **74**, 1031–1044 (2012).
53. Meng, L. *et al.* Towards a therapy for Angelman syndrome by targeting a long non-coding RNA. *Nature* **518**, 409–412 (2015).
54. Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).

Acknowledgements

The authors' work is supported by NIH grants DA034681 & DA039650 (J.J.D.), MH091122 & MH57014 (J.D.S.), DARPA grant HR0011-12-1-0015 (J.D.S.), startup funds from UAB (J.J.D.) and the Evelyn F. McKnight Brain Research Foundation. The views, opinions and/or findings contained in this article/presentation are those of the author(s)/presenter(s) and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

Author contributions

K.E.S., N.V.N.G., J.D.S. and J.J.D. designed the experiments. K.E.S. and N.V.N.G. executed the experiments with assistance from E.Y.S., R.S., M.K.O., J.B., J.J.M., J.S.R., A.N., H.G. and J.J.D. Next-generation sequencing data sets were analysed by J.J.D. with assistance from C.T.S. J.J.D. and K.E.S. wrote the manuscript with assistance from all authors.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions>.

How to cite this article: Savell, K. E. *et al.* Extra-coding RNAs regulate neuronal DNA methylation dynamics. *Nat. Commun.* **7**:12091 doi: 10.1038/ncomms12091 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016