# RNA Binding by DNMT1: Sequence and Structure Dependence and Evidence for a DNMT1-RNA-DNA Ternary Complex

Camila Sousa
Department of Biochemistry
University of Colorado at Boulder

Defense Date: April 1, 2022

Honors Thesis Advisor:
Dr. Thomas R. Cech: Biochemistry

Defense Committee:
Dr. Jeffrey Cameron: Biochemistry
Dr. Nausica Arnoult: Molecular, Cellular, and Developmental Biology

## ABSTRACT

Epigenetic modifications are a dynamic network of covalent modifications made to DNA in order to modulate gene expression. DNA methylation is one type of epigenetic marker which is essential for organismal development and cellular differentiation. DNA methylation patterns are established by the DNA methyltransferases (DNMTs) DNMT3A and DNMT3B, and maintained by DNMT1. DNMT1 recognizes newly synthesized hemimethylated DNA and retains the methylation pattern. Aberrant DNA methylation patterns are an early hallmark of various cancers, but it is unclear what misregulation event leads to these changes in methylation patterns. A number of RNAs have been shown to interact with DNMT1 in cells via crosslinking and pull-down experiments, however the mechanism of interaction between DNMT1 and RNA remains to be elucidated. Here, similar crosslinking and pull down experiments have shown that DNMT1 interacts with its own fully spliced mRNA in cellular nuclei, suggesting a possible mode of auto-regulation of DNMT1 by its mRNA. DNMT1-RNA electrophoretic mobility shift assays (EMSAs) were performed to determine RNA sequence and structure motifs which confer high affinity binding to DNMT1. Surprisingly, a GU dinucleotide repeated RNA sequence bound with the highest affinity out of all the RNAs tested. DNMT1 seemed to have a slight affinity for a quadruplex-like folded version of the GU repeated sequence, which has not been previously reported. Additionally, DNA-RNA and RNA-RNA competition EMSAs showed that DNMT1 can simultaneously bind DNA and RNA, and that multiple RNA molecules can bind DNMT1 at once. Findings from this study provide an exciting avenue of future research where DNMT1 activity may be allosterically regulated by RNA.

## INTRODUCTION

Epigenetics involves the modulation of gene expression without changes in the DNA sequence, covalent modifications are made to chromatin to regulate the levels of gene expression. The epigenetic landscape is dynamic, meaning that epigenetic patterns can change throughout the lifetime of a cell. DNA methylation is one such modification which was discovered in eukaryotes as early as 1944 along with the isolation of genetic material [1, 2], it was soon after proposed that DNA methylation may be affecting gene expression by interfering with transcription factor binding to DNA [3]. In an experiment where a mouse thymoma cell line was treated with a DNA methyltransferase inhibitor, Compere et al. showed that DNA methylation played a role in gene expression. By treating the cells with 5-azacytidine, which is incorporated into DNA instead of cytosine and cannot be methylated, they were able to induce expression of a gene which was usually methylated and show that methylated genes are generally inactive [4].

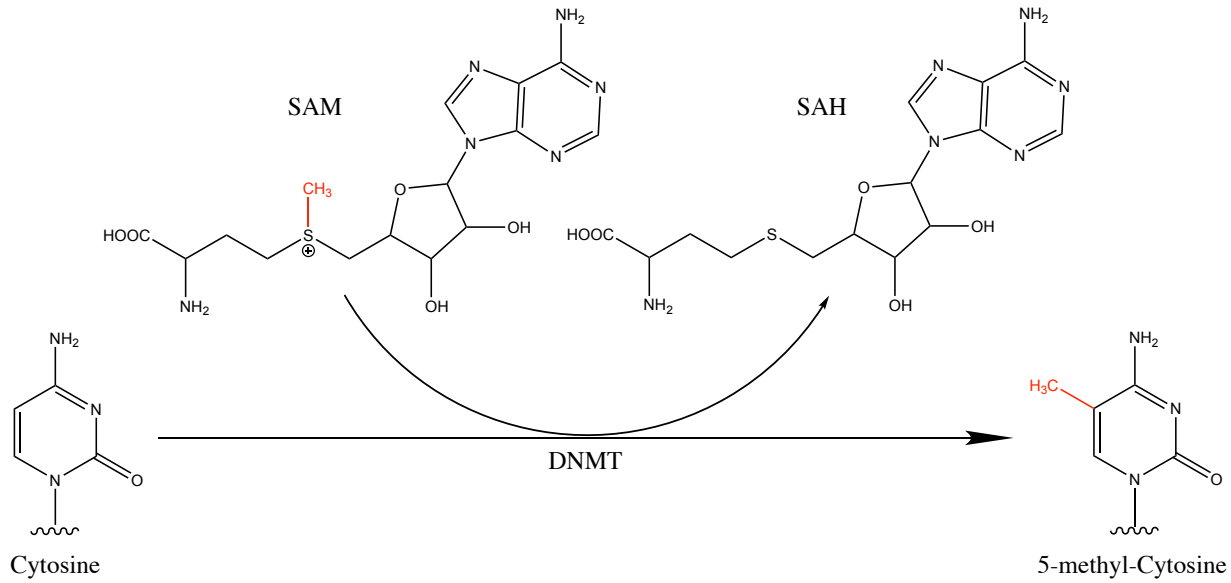
DNA methylation is found in both eukaryotes and prokaryotes, but plays different functions in different species. There are three types of DNA methylation marks in bacteria: C5-methylcytosine, N4-methylcytosine, and N6-methyladenine, these modifications are involved in bacterial virulence and bacterial immunity [5]. The primary methyl mark in eukaryotes is C5-methylcytosine (5mC), however there have been N6-methyladenine marks detected in *C. elegans*, *D. melanogaster*, as well as some protozoans [6]. In vertebrates, the methylation sites are at CpG dinucleotides, and 5mC is found almost exclusively [6, 7]. The vertebrate genome is heavily methylated compared to other eukaryotes, 60-80% of the approximately 29 million CpG sites are methylated [7]. Methylated CpG sites can be found within gene bodies, transposable

elements, and transcription enhancer sequences, which can all modulate gene expression. Transcriptional promoters of the vertebrate genome are often found in sequences with a high local density of CpG sites, known as CpG islands. Constitutively inactive genes generally have a promoter located within a heavily methylated CpG island, such as genes on the inactivated X chromosome in females [8]. However, most CpG islands are not heavily methylated, and these regions are typically resistant to methylation [7].

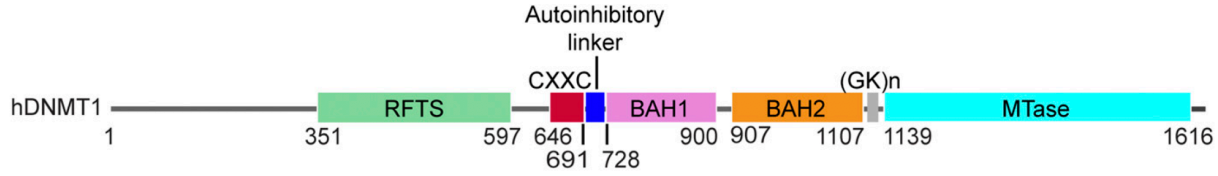
There are three canonical cytosine DNA methyltransferase (DNMT) enzymes in vertebrates which catalyze the methylation reaction. They all use S-adenosylmethionine (SAM) as the methyl donor, where the methyl group is transferred to cytosine as it is converted to 5mC (Figure 1A). There are two *de novo* DNA methyltransferases, DNMT3A and DNMT3B, which recognize unmethylated DNA and establish methylation patterns [9]. These DNMTs interact with DNMT3L, which does not have catalytic activity but is involved in targeting DNMT3A and DNMT3B to specific genomic loci during development and differentiation. Expression of these DNMTs decreases as cells differentiate [10]. There is one *maintenance* DNA methyltransferase, DNMT1, which is responsible for maintaining the established methylation pattern on newly replicated DNA [11]. DNMT1 recognizes hemimethylated DNA, when only one strand of DNA is methylated, and methylates the other strand to maintain the methylation pattern [11]. Since maintenance DNA methylation activity is only needed after DNA replication, expression of DNMT1 peaks during S phase of the cell cycle and is primarily expressed in differentiated cells [10, 12].

Overall DNMT1 structure and function is conserved among vertebrates [13], with N-terminal auto-regulatory domains and a C-terminal methyltransferase domain (Figure 1B). Human DNMT1 is 1616 amino acids long, Figure 1B shows the domains of human DNMT1 along with the residue numbering of each domain. The N terminal ~350 amino acids are unstructured. The replication-foci targeting domain (RFTS) interacts with ubiquitin-like, containing PHD and RING finger domains, 1 (UHRF1) which localizes to the replication fork and recruits DNMT1 [14, 15]. This interaction targets DNMT1 to newly synthesized DNA for methylation. The CXXC domain is a zing finger domain (where C=cysteine and X=any amino acid) which is responsible for binding unmethylated DNA [15, 16] and has been shown to be essential for catalytic activity [17]. The auto-inhibitory linker between the CXXC and Bromo-adjacent homology (BAH) 1 domains prevents unmethylated DNA from entering the catalytic pocket [16], and deletion of this auto-inhibitory linker increases DNMT1's *de novo* methylation activity [16]. Conformational changes upon binding unmethylated or hemimethylated DNA, including movement of the auto-inhibitory linker into the catalytic domain, prevent DNMT1 from having *de novo* methyltransferase activity (Figure 1C). The two BAH domains can interact with the target recognition domain (TRD) within the methyltransferase domain to further prevent binding of unmethylated DNA in the active site [18]. The glycine-lysine repeat $(GK)_n$ linker was originally thought to only provide flexibility between the regulatory and catalytic domains [15], but has been shown to interact with ubiquitin-specific protease 7 (Usp7) [19], suggesting that the linker may play a role in DNMT1 stability. The methyltransferase domain contains the catalytic site, and the mechanism of DNA methylation is highly conserved between species [15]. The target cytosine is flipped out of the DNA duplex into the catalytic site [20], shown in magenta in Figure 1D, where the catalytic cysteine 1229 (C1229) forms a covalent intermediate with C6 of cytosine, and the methyl group from SAM is transferred to C5 followed by deprotonation of C5 to form 5mC. The orientation of C1229 changes depending on if DNMT1 is in the active or auto-inhibitory conformation (Figure 1C, inset).

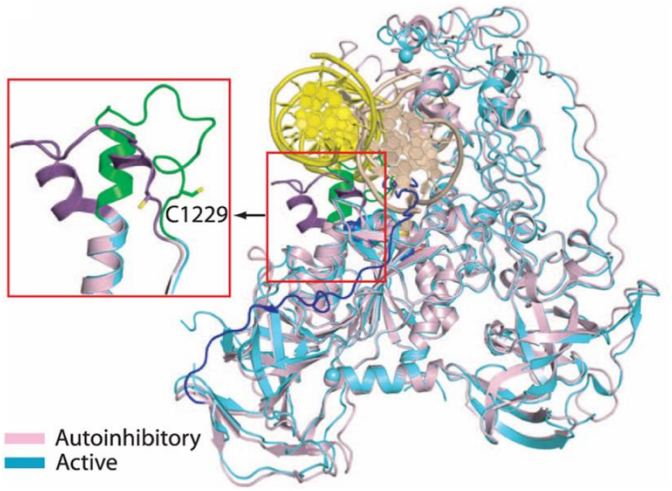
**A.**


SAM
SAH
DNMT
Cytosine
5-methyl-Cytosine

**B.**


Autoinhibitory
linker
CXXC
(GK)n
hDNMT1
RFTS
BAH1
BAH2
MTase
1
351
597
646
691
728
900
907
1107
1139
1616

**C.**


C1229
Autoinhibitory
Active

**D.**

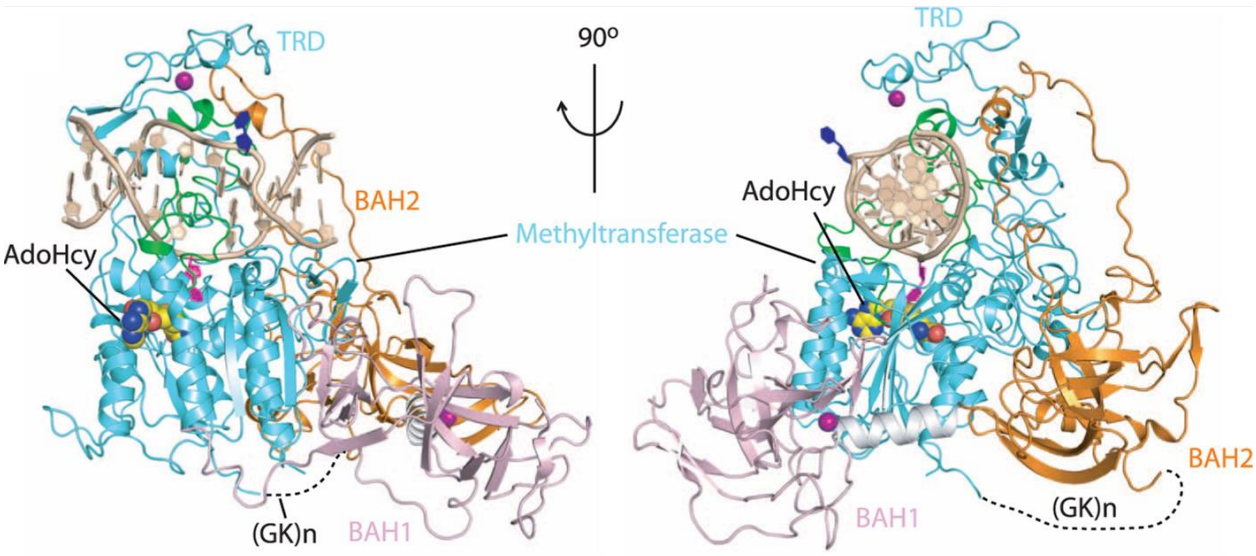



***Figure 1:*** *Reaction catalyzed by the DNMT enzymes and structure of DNMT1.* ***(A)*** *The cytosine is converted to 5-methyl-cytosine through the addition of a methyl group on carbon 5. SAM is the methyl group donor and is converted to S-adenosyl-homocysteine (SAH) during the reaction.* ***(B)*** *Figure courtesy Ren et al. 2020 [21]. Domains of human DNMT1, each domain annotated by residue numbers.* ***(C)*** *Figure courtesy Song et al. 2012 [22]. Crystal structure of the active conformation of mouse DNMT1 (731-1602) bound to hemimethylated DNA (shown in cyan and beige), superimposed over the auto-inhibited conformation of mouse DNMT1 (650-1602) bound to hemimethylated DNA (shown in pink and yellow). Inset shows the catalytic loop, containing C1229, in the active (green) and auto-inhibited (purple) conformations.* ***(D)*** *Figure courtesy Song et al. 2012 [22]. Crystal structure of mouse DNMT1 (731-1602) bound to hemimethylated DNA, in two orthogonal views. The BAH1 and BAH2 domains are shown in pink and orange, respectively. The* $(GK)_n$ *linker shown in black dashed lines. The methyltransferase domain, containing the TRD, shown in cyan. The catalytic loop and TRD loops 1 and 2 shown in green. Hemimethylated DNA shown in beige. Zinc ions shown in magenta. The target cytosine is shown flipped into the catalytic pocket in magenta, and SAH (AdoHcy) shown in space-filling representation.*

Changes in DNA methylation patterns are an important aspect of organismal development, cellular differentiation, and cancer onset and progression. In addition to DNA methylation, DNA can be actively or passively demethylated, which is mediated by various ten-eleven translocation (TET) enzymes. Active demethylation reactions involve TET-mediated oxidation of a methylated cytosine [23, 24], followed by thymine DNA glycosylase mediated base excision repair to replace the modified cytosine residue [18]. Newly fertilized mammalian zygotes undergo TET-mediated erasure of DNA methylation, followed by DNMT3A and DNMT3B mediated DNA methylation of the blastocyst after implantation [10]. Cells that will become part of the germline undergo further TET-mediated DNA demethylation [10]. DNA methylation patterns are an important part of cellular differentiation, and differentiation of embryonic stem cells is almost completely inhibited in the absence of all DNMTs [7]. Aberrant DNA methylation patterns are an early hallmark of various cancers [25-27]. In healthy cells,

most CpG dinucleotides within CpG islands are not methylated, however CpG islands containing promoters for tumor-suppressor genes are often hypermethylated in cancer cells [7, 28]. There is also a global decrease of DNA methylation levels in cancer cells, which can lead to demethylation of oncogenes [28] and genomic instability [29]. Although changes in DNA methylation patterns are known to be a hallmark of cancer, it is unclear how those changes in methylation arise.

There has been increasing focus on the regulation of DNMT1, and recently RNA has emerged as a possible regulator of DNA methyltransferase targeting and activity [30]. Various RNAs have been identified which bind to DNMT1 via crosslinking and pull down experiments [31], showing that DNMT1 can interact with a range of RNAs in different cell types. Several RNAs have been shown to change global [32] or locus specific [33, 34] genomic methylation patterns, however the mechanism of RNA binding to and regulating DNMT1 remains unclear. Hemimethylated DNA is known to bind the catalytic methyltransferase of the protein, but DNMT1 does not have any canonical RNA binding domains, and it is therefore unknown how RNA binds to DNMT1. Even though a number of different RNAs have been shown to interact with DNMT1, the specific recognition and binding mechanism remains to be fully understood.

This project is in collaboration with Linnea Jansson-Fritzberg, a Postdoctoral Fellow in the Cech lab. Jansson-Fritzberg performed formaldehyde crosslinking immunoprecipitation followed by RNA sequencing (fRIP-seq) on DNMT1. Interestingly, the experiment showed that DNMT1 was interacting with its own fully spliced mRNA (data unpublished). DNMT1 interacting with its own mRNA has not been previously shown, which prompted us to investigate the nature of this potential interaction. Even though DNMT1 has been shown to bind to a number of RNAs both in cells [30] and in vitro [32, 33], it is unknown which RNA features are being recognized by DNMT1 to allow it to bind RNA. By performing binding assays with RNAs with various sequence and secondary structure motifs, this study aims to elucidate the mechanism of interaction between DNMT1 and RNA. Additionally, since the RNA binding site on DNMT1 is unknown, competition binding assays between DNA and various RNAs were performed.

## METHODS

### Production and purification of *DNMT1* mRNA fragments

E. coli glycerol stock of *DNMT1* cDNA with UTRs was ordered from TransOmic (TCH1003-BC092517, vector pT7T3D-PacI). Bacteria were streaked onto a carbenicillin plate and grown at 37ºC overnight. Restriction enzyme cloning, using EcoRI and NotI (NEB) was used to subclone the DNMT1 cDNA into a pFastBac B vector to produce a plasmid containing the *DNMT1* cDNA without the T7 promoter.

PCR primers were designed to amplify ~200 nucleotide regions of the DNMT1 cDNA, and added to the T7 RNA polymerase promoter sequence (5′ TAATACGACTCACTATAG 3′, where G indicates the first transcribed base) on the 5’ end of the coding strand of each fragment. PCR was performed to amplify each segment using NEB Phusion polymerase, cycle parameters: 98ºC 1min, followed by 30 cycles of 98ºC 30 seconds, 61ºC 30 seconds, 72ºC 1 minute, and 72ºC 10 minutes. PCR products were run on a 1% agarose gel at 100V for 1 hour to check for correct size amplification. Each PCR product was purified using Omega Bio-tek E.Z.N.A. Cycle Pure Kit, according to manufacturer’s instructions. Pure PCR products were used as the template for in vitro transcription, described below.

In vitro transcription (IVT) with in-house purified T7 RNA polymerase was used to produce each segment of the *DNMT1* mRNA. IVT conditions varied with RNA sequence, but each reaction contained: 6-40mM $MgCl_2$, 1-4mM each rATP, rUTP, rGTP, and rCTP, 1-90mM DTT, 40mM Tris-HCl pH 7.5, 2mM spermidine, 4ng/uL inorganic pyrophosphatase (Sigma I1891), 1ng/uL DNA template, and 20uL T7 RNA polymerase per 0.5mL IVT reaction. The reactions were incubated at 25ºC overnight, then treated with Turbo DNase (Invitrogen) for 15 minutes at 37ºC to remove template DNA. Phenol chloroform extraction was used to purify the RNA, followed by ethanol precipitation (addition of 1/10X volume 3M sodium acetate and 3X volume 100% cold ethanol, incubated at -80ºC overnight). Precipitated RNA was centrifuged at maximum speed for 30 minutes at 4ºC to pellet, pellet was resuspended in formamide denaturing loading dye and run on a 6% 7M urea PAGE gel at 25W for 2.5 hours. The gel was visualized with UV shadowing and the correct band was excised and forced through a syringe. The band was eluted twice in 10mL TE pH 7.5 (10mM Tris-HCl pH 7.5, 1mM EDTA pH 8), the supernatant was 0.2um filtered and concentrated using an Amicon 10K filter unit, the concentrated solution was ethanol precipitated and pelleted (as mentioned previously), the pellet was resuspended in TE pH 7.5 and RNA concentration determined using a NanoDrop.

**RNA native gels**

RNA native gel protocol was modified from Wang et al. 2017 [35]. For figures 2D and 5C, the gels did not contain salt. Radiolabeled RNAs were refolded as described in "Kd determination by EMSA" section, and 1000 counts per minute of RNA were loaded onto 6% or 10% native polyacrylamide gels, depending on size of RNA. Gels were run at 100V in 1X TBE at 25ºC for 1 hour. Gels were vacuum dried for 1 hour at 80ºC on two layers of Whatman 3mm chromatography paper. Dried gels were exposed to Phosphor screens (Amersham Biosciences) overnight and signal acquisition was performed using a Typhoon FLA 9500 Phosphorimager. Native gels in figure 6A, in the presence of KCl or LiCl, were obtained by Linnea Jansson-Fritzberg. 10% native polyacrylamide gels were poured containing either 100mM KCl or LiCl. Unlabeled RNA was refolded as mentioned above, and was run at 80V at 25ºC for 1 hour in 0.5X TBE supplemented with either 100mM KCl or LiCl in the running buffer. The gel was visualized by post-staining with SYBR Gold Nucleic Acid Gel Stain (Invitrogen).

**Protein expression**

Protein expression protocol was modified from Davidovich et al. 2013 [36]. A pFastBac1 expression vector carrying sequences encoding full length human DNMT1 was used to generate baculovirus stocks with the Bac-to-Bac system (Life Technologies), according to the manufacturer's instructions. DNMT1 was expressed under a PreScission cleavable N-terminal MBP and 3XFLAG tag. Yield-optimized amounts of baculoviruses were used to infect 2-4 liters of Sf9 cells at a density of 2 million cells/mL in Sf-900 III SFM (Invitrogen 12658-027). Cells were incubated for 72 hours at 27ºC at 130rpm, cells were harvested and snap frozen in liquid nitrogen, then stored at -80ºC until protein purification.

**Protein purification**

Purification protocol modified from Davidovich et al. 2013 [36]. 5g frozen cell paste was dissolved in 40mL cold 1X lysis buffer (10mM Tris-HCl pH 7.5, 150mM NaCl, 0.5% Nonidet P-40, 2mM β-mercaptoethanol, 5mM imidazole) and supplemented with EDTA free Complete protease inhibitor (Roche). Lysate was centrifuged at 14,500rpm at 4ºC in Beckman JA-20.

Supernatant was agitated with 8mL amylose beads (NEB) for 1hr at 4ºC, and slurry was transferred to Econo-Pac chromatography column to remove unbound protein. The beads were washed with 4 column volumes (cv) water and 10 cv lysis buffer. Beads were washed with 16 cv wash buffer I (10mM Tris-HCl pH 7.5, 500mM NaCl, 1mM TCEP) and 16cv wash buffer II (10mM Tris-HCl pH 7.5, 150mM NaCl, 1mM TCEP), protein was slowly eluted with 3 cv elution buffer (10mM maltose, 10mM Tris-HCl pH 7.5, 150mM NaCl, 1mM TCEP). Elution was concentrated using Amicon 30K filter unit, PreScission protease was added to cleave MBP tag, and incubated at 4ºC overnight, MBP cleavage efficiency was checked with SDS-PAGE gel.

Liquid chromatography using an AKTA pure FPLC was used to further purify the protein. The sample was loaded onto a HiTrap Heparin HP affinity column (Cytiva) which was pre-equilibrated with wash buffer II, fractions were collected as protein was eluted with a linear buffer gradient change to wash buffer I. Fraction samples were run on SDS-PAGE gel to determine which fractions to pool and concentrate. The concentrated sample was loaded on a Superose 6 Increase 10/300 GL (Cytiva) column pre-equilibrated in 150mM NaCl, 20mM HEPES, 1mM TCEP. The elution fractions were checked by SDS-PAGE before pooling and concentrating. Concentrated protein was aliquoted and snap frozen, followed by storage at -80ºC.

Purified protein was checked for proper folding and catalytic activity using Abcam Colorimetric DNMT Activity Quantification Kit (ab113467). In-house purified DNMT1 had comparable activity to commercially available DNMT1 (AMSBIO 51101).

**Radio labeling of nucleic acids**

20pmol of RNA made by IVT was treated with calf intestinal alkaline phosphatase (NEB QuickCIP) and incubated at 37ºC for 10 minutes, the reaction was stopped by incubation at 80ºC for 2 minutes. The CIP treated RNA was then incubated with [$^{32}$P]γ-ATP (PerkinElmer) and T4 polynucleotide kinase (NEB) at 37ºC for 30 minutes. Synthesized RNA and DNA oligonucleotides were ordered from Integrated DNA Technologies or Dharmacon and directly incubated with [$^{32}$P]γ-ATP and polynucleotide kinase. Roche G-25 Sephadex Columns for Radiolabeled RNA and DNA Purification were used to separate unincorporated [$^{32}$P]γ-ATP from labeled nucleic acid, according to manufacturer's instructions. Labelling efficiency was determined with Beckman LS 6500 Scintillation Counter.

**Kd determination by EMSA**

Electrophoretic mobility shift assays (EMSAs) were performed with purified DNMT1 and radio labeled DNA or RNA for determination of the dissociation constant (Kd). Radio labeled RNA samples were refolded by boiling in TE pH 7.5 for 5 minutes, followed by snap cooling on ice for 2 minutes, then refold buffer was added (final concentrations: 50mM Tris-HCl pH 7.5, 100mM KCl, 2.5mM $MgCl_2$, 0.1mM $ZnCl_2$, 2.0mM β-mercaptoethanol, 0.05% Nonidet P-40, 5% glycerol) and samples were incubated at 37ºC for 30 minutes. The same protocol was followed for EMSAs using double stranded RNA, where the complementary strands of RNA were mixed in equimolar amounts prior to boiling in TE. DNMT1 was diluted in binding buffer (refold buffer supplemented with 0.1mg/mL Baker's Yeast tRNA (Sigma R5636) and 0.1mg/mL BSA (NEB)). 5uL refolded RNA, or DNA, was mixed with 5uL diluted DNMT1, mixture was incubated at 30ºC for 1 hour. 8ul of each mixture was loaded on a 1% agarose gel, gel was run at 66V for 1.5 hours at 4ºC in 1X TBE. Gels were vacuum dried for 1.5 hours at 80ºC on Hybond-N+ membrane and two layers of Whatman 3mm chromatography paper. Dried gels were exposed

to Phosphor screens (Amersham Biosciences) overnight and signal acquisition was performed using a Typhoon FLA 9500 Phosphorimager.

**ImageQuant and PRISM**

ImageQuant software was used to calculate the pixel density of manually picked bands on the EMSA gels. Rolling ball background subtraction was used, with the radius depending on the specific gel. The fraction bound values were calculated using Microsoft Excel (version 16.43). GraphPad Prism (version 9.3.1) was used for Kd and Hill coefficient determination. Data was fit to "Specific binding with Hill slope" equation, Y=Bmax*X^h/(Kd^h + X^h), and Bmax value constrained to equal 1. Error for Kd and Hill coefficient values was determined by showing the confidence intervals with 95% confidence.

**RESULTS**

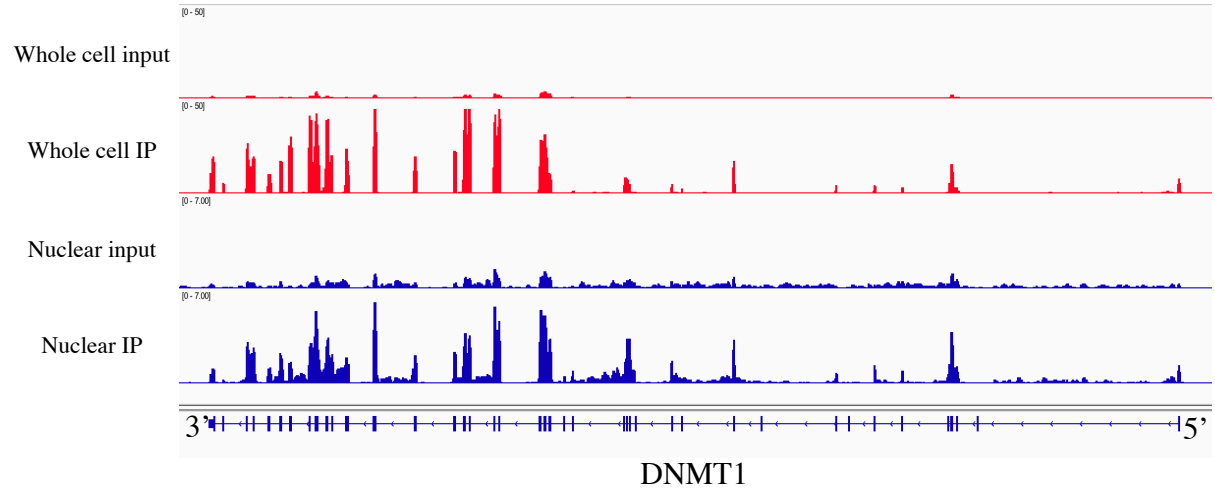
**DNMT1 binds specifically to its own messenger RNA**

It has been shown that DNMT1 can interact with various RNAs in cells [30, 31]. The Cech lab has shown that DNMT1 can interact with its own fully spliced mRNA in K562 and iPSC cells by formaldehyde crosslinking followed by RNA sequencing (Figure 2A). To verify that this is a true nuclear interaction, as opposed to an artifact of the nascent DNMT1 protein being in close proximity to the mRNA during translation, the fRIP-seq experiment was repeated with only the nuclear fraction of K562 cells. The binding pattern remained the same (Figure 2A), indicating that the DNMT1 enzyme was interacting with its own fully spliced mRNA in the nucleus. The fRIP-seq data also showed that there appeared to be a bias for DNMT1 to bind toward the 3' end of the mRNA transcript, note in figure 2A that the *DNMT1* gene is shown 3' to 5'.

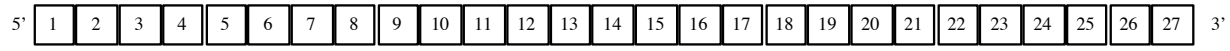
To biochemically verify this interaction and further resolve what region of the transcript is being bound by DNMT1, electrophoretic mobility shift assays (EMSAs) were performed with ~200 nucleotide fragments of the *DNMT1* mRNA. The fragments were designed to tile the entire mRNA and were produced by T7 polymerase in vitro transcription (IVT), where consecutive fragments overlap each other, and all nucleotides of the mRNA are represented by the fragments. The starting and ending residues of the fragments were determined based on the possible T7 transcription start sites (GGG, GGA, GGC) present in the native *DNMT1* sequence. The fragments were numbered simply based on their location along the mRNA, fragments with lower numbers are near the 5' end of the transcript while fragments with higher numbers are near the 3' end of the transcript (Figure 2B). EMSAs were performed with select fragments which span the mRNA, showing that DNMT1 does bind to its own mRNA in vitro, and has varying affinities for different regions of the mRNA (Figure 2C). The highest affinity *DNMT1* fragment was "RNA3", with a dissociation constant (Kd) of 262.3±5.9nM. The lowest affinity was "RNA13" which bound DNMT1 with a Kd of 481±105nM, however this value may have been impacted by the multiple folding species of "RNA13", as evident by the smeared band on the native gel (Figure 2D). The fRIP-seq accompanied by the in vitro EMSA binding studies confirmed that the interaction between DNMT1 and its own mRNA was a true interaction, not an artifact from the fRIP-seq experiment. However, the fragmentation of the mRNA for EMSA experiments was somewhat arbitrary, and secondary structures that may have present in the full-length mRNA were likely disrupted. The entire *DNMT1* mRNA is 5,246 nucleotides long, and it was not feasible to produce the full length mRNA by IVT or perform EMSAs with the full length

mRNA. This prompted us to investigate and attempt to understand, more generally, which features of RNA allow it to bind to DNMT1. The focus of the study shifted to determine which RNA sequences or structures are important to confer high binding affinity to DNMT1.

**A.**

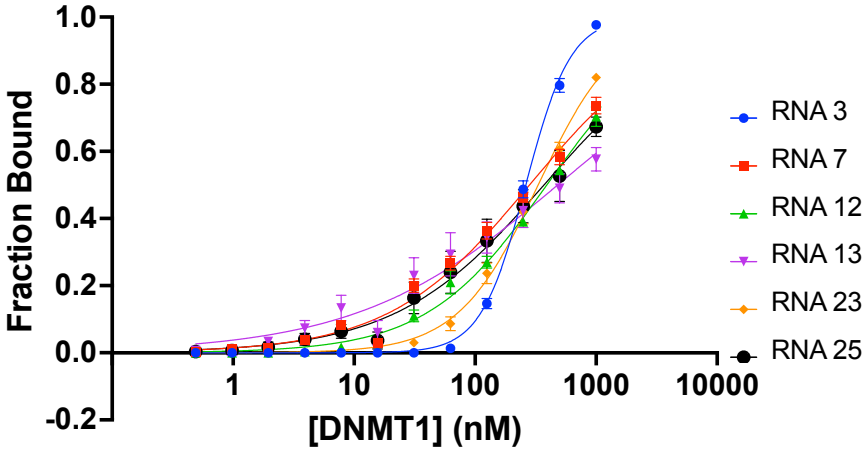



**B.**




**C.**

| RNA | Kd (nM) | Hill |
|---|---|---|
| RNA3 | 262.3±5.9 | 2.3±0.1 |
| RNA7 | 281±26 | 0.7±0.1 |
| RNA12 | 390±27 | 0.9±0.1 |
| RNA13 | 481±105 | 0.5±0.1 |
| RNA23 | 332±19 | 1.3±0.1 |
| RNA25 | 370±50 | 0.7±0.1 |



**D.**

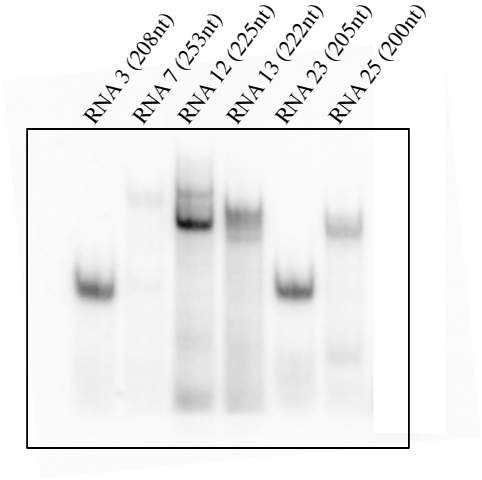



***Figure 2:*** *DNMT1 binds to its own fully spliced mRNA in cells and in vitro.* ***(A)*** *fRIP-seq trace showing the input and immunoprecipitated (IP) fractions. Note that DNMT1 gene shown 3' to 5'. This data shows results from fRIP-seq performed with whole K562 cells as well as K562 nuclei, same pattern seen when done with induced pluripotent stem cells (data not shown). fRIP-seq data obtained by Linnea Jansson-Fritzberg (unpublished).* ***(B)*** *Naming convention for ~200 nucleotide fragments of DNMT1 mRNA.* ***(C)*** *Binding curves of DNMT1 mRNA fragments to DNMT1 protein, with associated calculated Kd and Hill coefficient values.* ***(D)*** *Native gel of DNMT1 mRNA fragments.*

**DNMT1 binds RNA in a sequence and/or structure dependent manner**

When interacting with DNA, DNMT1 is known to preferentially bind double stranded DNA and methylate CpG dinucleotides [20]. We wondered if DNMT1 has a preference for single stranded RNA (ssRNA) or double stranded RNA (dsRNA), and whether or not it recognizes CpG dinucleotides in RNA. To test this, single and double stranded RNAs, all 40 nucleotides long, with and without CpG sites were designed for EMSA binding studies. The sequences were derived from the native *DNMT1* mRNA "RNA3" sequence, which contained two CpG dinucleotides. The two CpG dinucleotides were additionally mutated to CpA dinucleotides to determine whether DNMT1 recognizes CpG sites on RNA (Figure 3A). The sequences shown in Figure 3A show the ssRNA designs, which have no predicted secondary structure. Those oligomers were annealed to their complementary RNA sequence to test the affinity for dsRNA. EMSAs showed that DNMT1 does not have a strong preference for single or double stranded RNA or for the presence of CpG dinucleotides in either single stranded or double stranded RNA context (Figure 3B).

Synthetic RNA 40mers of single nucleotide tracts and repeated sequences were tested to determine if DNMT1 has any RNA sequence or structure motif preference (Figure 3C). The results show that the affinity between DNMT1 and RNA is not solely dependent on length, as all the RNAs were 40 nucleotides long, but there was a wide range of binding affinities. Not surprisingly, DNMT1 bound with moderate affinity to C and G rich RNAs, including Poly(G), Poly(C), and $(CGG)_{12}$. At 36 nucleotides long, $(CGG)_{12}$ is slightly shorter than the other oligomers tested but was used to measure the affinity of an RNA that contained both C and G residues. The A and U rich RNAs bound with low affinity, the lowest affinity RNA was Poly(A) ($Kd > 4\mu M$), which was used as a low binding control for all subsequent binding experiments.

DNMT1 has been shown to preferentially bind to G-quadruplexes [32], which is supported by DNMT1 binding to $(GGAA)_{10}$ with high affinity compared to $(GA)_{20}$, which both have the same G and A content, but only $(GGAA)_{10}$ can form G-quadruplexes. Surprisingly, DNMT1 bound with very high affinity to $(GU)_{20}$, which has not been previously reported. It was not immediately obvious why DNMT1 bound so tightly to $(GU)_{20}$, but given this result, we decided to investigate what feature of $(GU)_{20}$ allowed it to bind to DNMT1 with such high affinity.

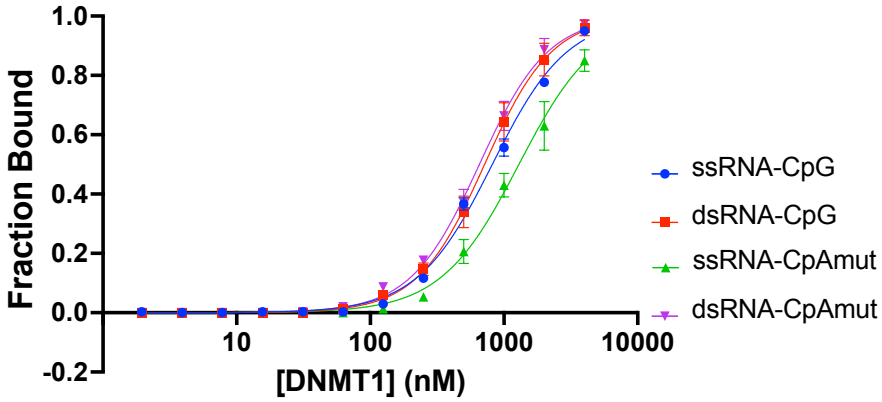
**A.**

CpG: CCAACAGCCCCCCCAAACCCCUUUCCAAACCU**CG**CA**CG**CC

CpA: CCAACAGCCCCCCCAAACCCCUUUCCAAACCU**CA**CA**CA**CC

**B.**

| RNA | Kd (nM) | Hill |
|---|---|---|
| ssRNA-CpG | 817±36 | 1.5±0.1 |
| dsRNA-CpG | 715±33 | 1.7±0.1 |
| ssRNA-CpAmut | 1293±74 | 1.5±0.1 |
| dsRNA-CpAmut | 654±27 | 1.7±0.1 |




**C.**

| RNA | Kd (nM) | Hill |
|---|---|---|
| Poly(A) | N/A | N/A |
| Poly(G) | 1030±26 | 2.4±0.1 |
| Poly(C) | 390±15 | 1.5±0.1 |
| Poly(U) | 1711±116 | 1.0±0.1 |
| $(GA)_{20}$ | 347±18 | 1.4±0.9 |
| $(GGAA)_{10}$ | 215±20 | 0.9±0.1 |
| $(CGG)_{12}$ | 1074±43 | 1.9±0.1 |
| $(GU)_{20}$ | 82.1±6.8 | 1.1±0.1 |

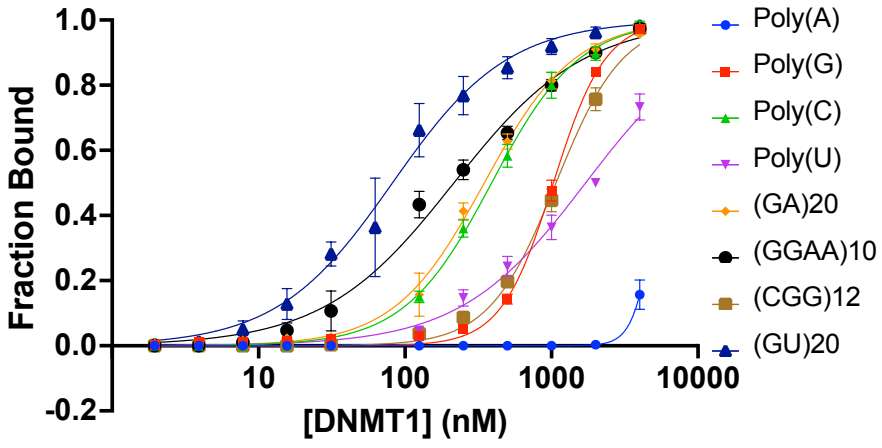



***Figure 3:*** *DNMT1 does not recognize CpG dinucleotides in RNA, but does have sequence and/or structure preference for RNA binding.* ***(A)*** *RNA oligonucleotide sequences used for EMSAs. ssRNA sequences shown, mFOLD RNA predicts no secondary structure for ssRNA. Shown sequences annealed to complement (not shown) for dsRNA EMSAs.* ***(B)*** *Binding curves of ssRNA and dsRNA, and CpG and CpA RNA oligonucleotides shown in (A), along with their calculated*

*Kd and Hill coefficients. **(C)** Binding curves of polynucleotide tracts and di-/tri-nucleotide repeat RNA 40mers, along with their calculated Kd and Hill coefficients. All RNAs are 40 nucleotides long except for (CGG)12, which is 36 nucleotides long.*

**DNMT1 binds with high affinity to GU-quadruplex RNA**

Given that $(GU)_{20}$ unexpectedly bound to DNMT1 with the highest affinity, we decided to investigate whether or not GU dinucleotide content allowed for RNA binding. The fRIP-seq data showed that there seemed to be bias for DNMT1 to bind toward the 3' end of the transcript, and we noticed by eye that the 3' untranslated region (3'UTR) appeared to be fairly GU rich. An EMSA was thus performed with the entire 3'UTR, which is 321 nucleotides long, and the Kd was determined to be 69.1±7.6nM (Figure 4). This RNA bound to DNMT1 with the highest affinity of all our constructs, but it was unknown whether or not the 3'UTR was binding with higher affinity due to it being at least 100 nucleotides longer than any other RNA previously tested. Therefore, the affinity was compared to that of a region of human telomerase RNA component (hTR34-328) which is similar in size to the *DNMT1* 3'UTR (294 nucleotides) but completely unrelated. DNMT1 bound hTR34-328 with a Kd of 82±27 (Figure 4), indicating that longer RNA could bind to DNMT1 with higher affinity, and the 3'UTR affinity to DNMT1 was not necessarily due to its high GU content.

Several fragments of the 3'UTR were also tested to compare their affinity to the full length 3'UTR. The "GU rich 50mer" and "GU rich 66mer" were two separate regions of the 3'UTR which contained many GU dinucleotides. Their binding affinity was determined to be 416±28nM and 997±59nM, respectively (Figure 4). The "conserved 56mer" was a region of the 3'UTR which is conserved between humans and chickens [37], which is not GU rich, and the Kd was determined to be 714±28nM (Figure 4). This seemed to suggest that DNMT1 binding to RNA is not determined by the dinucleotide content of the RNA.

The sequences of the *DNMT1* mRNA fragments from Figure 1C were analyzed for their GU content to investigate whether it could predict high affinity binding to DNMT1. However, it seemed that the GU content of a sequence does not determine the binding affinity to DNMT1 (Table 1). Most notably, "RNA3" which was the ~200nt *DNMT1* mRNA fragment which bound with the highest affinity had a GU content of 4.8%. When compared to "RNA13", for example, which had about twice the GU content but bound to DNMT1 with much lower affinity, it shows that GU dinucleotide content does not correlate with a tight binding interaction. Together, these results indicated that GU content alone is not enough to confer high affinity binding to DNMT1.

| RNA | Kd (nM) | Hill |
|---|---|---|
| 3'UTR full length | 69.1±7.6 | 0.9±0.1 |
| GU rich 50mer | 416±28 | 1.6±0.1 |
| GU rich 66mer | 997±59 | 1.1±0.1 |
| Conserved 56mer | 714±28 | 2.2±0.2 |
| hTR 34-328 | 82±27 | 0.9±0.1 |

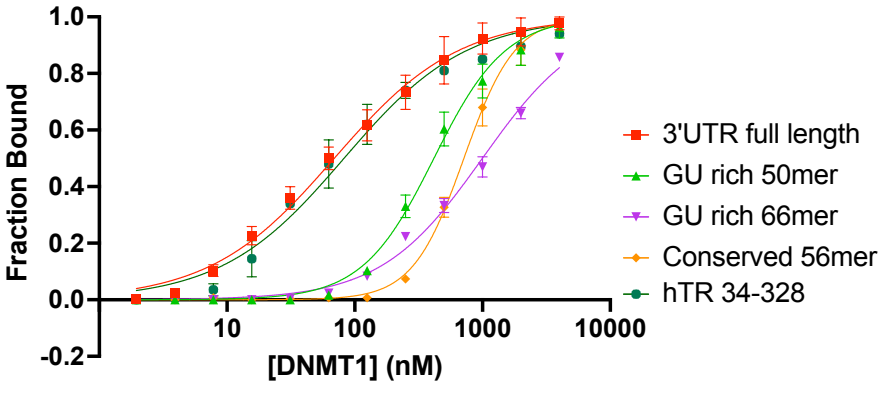


***Figure 4:*** *DNMT1 does not bind with high affinity to the DNMT1 3'UTR due to its GU content. Binding curves of full length DNMT 3'UTR and hTR 34-328, as well as fragments of the DNMT1 3'UTR, along with their calculated Kd and Hill coefficients.*

| RNA | Total length (nucleotides) | # GU dinucleotides | %GU content | Kd (nM) |
|---|---|---|---|---|
| 3'UTR full length | 321 | 45 | 14.0 | 69.1±7.6 |
| 3'UTR GU rich 50mer | 50 | 15 | 30.0 | 416±28 |
| 3'UTR GU rich 66mer | 66 | 17 | 25.6 | 997±59 |
| 3'UTR conserved 56mer | 56 | 4 | 7.1 | 714±28 |
| RNA3 | 208 | 10 | 4.8 | 262.3±5.9 |
| RNA7 | 253 | 28 | 11.1 | 281±26 |
| RNA12 | 225 | 22 | 9.8 | 390±27 |
| RNA13 | 225 | 24 | 10.7 | 481±105 |
| RNA23 | 205 | 21 | 10.2 | 332±19 |
| RNA25 | 200 | 15 | 7.5 | 370±50 |
| hTR34-328 | 294 | 23 | 7.8 | 82±27 |

***Table 1:*** *GU content is not a predictor of binding affinity to DNMT1 protein. Analysis of total GU content of DNMT1 3'UTR and fragments of the 3'UTR, as well as the DNMT1 mRNA fragments. GU dinucleotides were manually counted and %GU content calculated manually.*

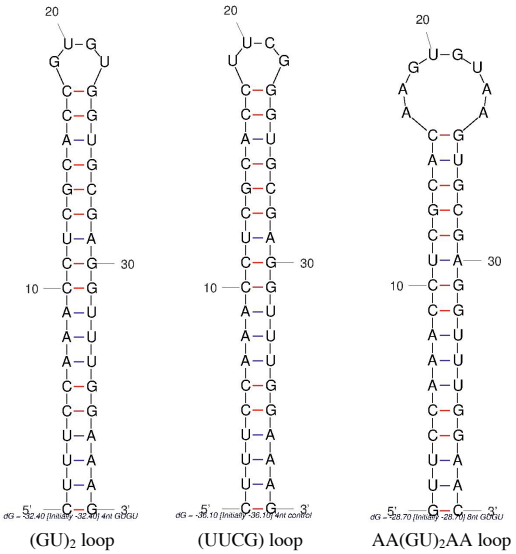
Since it was determined that the amount of GU dinucleotides within an RNA sequence was not the feature being recognized by DNMT1 for high affinity binding, we then decided to investigate possible secondary structures that could be formed by $(GU)_{20}$. It is known that G-U can wobble base pair with each other in RNA, so we predicted that $(GU)_{20}$ may be forming wobble-pair stem-loop structures. To test if stable stem-loop RNAs had comparable affinity to DNMT1 as $(GU)_{20}$, RNA oligonucleotides designed to form stem-loops with varying loop sizes and GU repeats within the loop were designed (Figure 5A). The dsRNA "stem" region of the stem-loops were the same sequences from the dsRNA shown in Figure 3A and 3B. The "$(GU)_2$ loop" contained 2 GU repeats (Figure 5A) and was used to test if repeated GU sequences in the "loop" structure were being recognized by DNMT1. The tetraloop ("(UUCG) loop") was used as a negative control as it is a stable and physiologically relevant loop structure [38] (Figure 5A). The "AA$(GU)_2$AA loop" was designed to form a larger "loop" structure with the same number of GU repeats as "$(GU)_2$ loop". An additional 8 nucleotide control loop structure was designed, where the loop sequence was AAUUCGAA, but given the low affinity of the other designed stem-loops was not used for Kd determination. The designed stem-loops were tested by EMSA and did not bind with comparable affinity to $(GU)_{20}$ (Figure 5B).

A native gel was run to compare the relative mobility of $(GU)_{20}$ to the designed stem-loops, showing that $(GU)_{20}$ runs with a lower mobility compared to stable stem-loop RNAs (Figure 5C), suggesting that it may be adopting a different structure. Previous publications have suggested that DNMT1 binds to hairpin RNAs [32, 33], so we decided to do EMSAs with the

same RNA hairpins that were used in these studies to confirm those findings. The “Di Ruscio R4” and “Di Ruscio R5” RNA are sequences published in Di Ruscio et. al. 2013, where “Di Ruscio R4” had no predicted secondary structure and “Di Ruscio R5” was predicted to form a short hairpin structure. The authors showed that “Di Ruscio R5” bound with high affinity while “Di Ruscio R4” did not [33]. However, our data showed that DNMT1 did not bind either RNA with high affinity, and bound “Di Ruscio R4” which is predicted to be unstructured with a slightly higher affinity than “Di Ruscio R5” (Figure 5B).

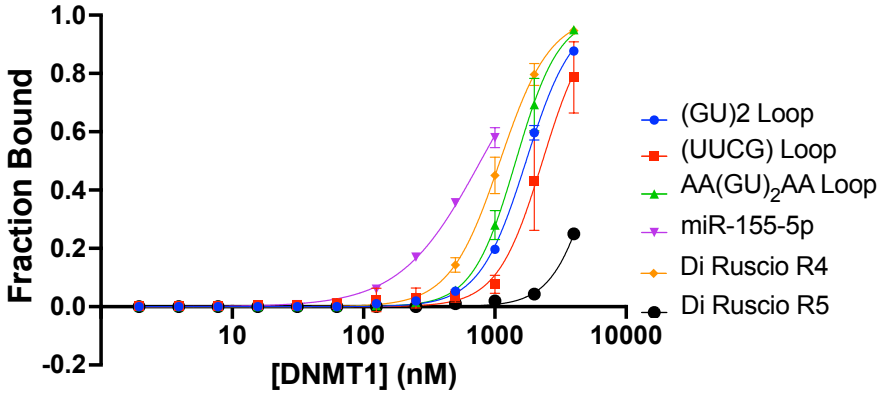
Another short hairpin RNA, “miR-155-5p”, was shown by Zhang et. al. 2015 to bind to DNMT1 in the catalytic site and inhibit DNMT1 methyltransferase activity [32]. They proposed that “miR-155-5p” formed a hairpin structure which allowed it to bind to and inhibit DNMT1. Our data, again, showed that DNMT1 did not bind to this hairpin structure with high affinity (Figure 5B). Together, these data indicate that DNMT1 does not recognize RNA hairpins or stem-loop structures for binding.

**A.**




**B.**

| RNA | Kd (nM) | Hill |
|---|---|---|
| $(GU)_2$ loop | 1731±27 | 2.4±0.1 |
| (UUCG) loop | 2309±190 | 2.5±0.5 |
| $AA(GU)_2AA$ loop | 1445±53 | 2.6±0.3 |
| miR-155-5p | 779±24 | 1.4±0.1 |
| Di Ruscio R4 | 1095±32 | 2.3±0.1 |
| Di Ruscio R5 | 6000±200 | 2.7±0.2 |



**C.**

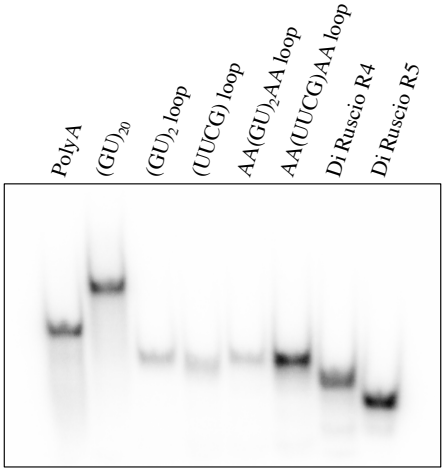
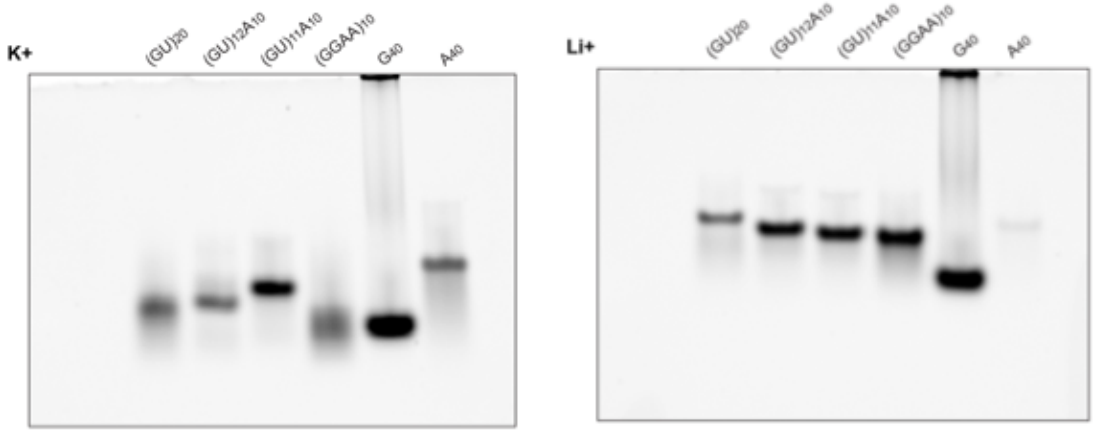



***Figure 5:*** *DNMT1 does not bind with high affinity to RNA stem-loop or hairpin structures.* ***(A)*** *Stem-loop structure designs, mFOLD RNA software used for structure prediction.* ***(B)*** *Binding curves for stem-loops designed for this study, as well as the hairpin RNAs from Di Ruscio et. al. 2013 ("Di Ruscio R4", "Di Ruscio R5") and Zhang et. al. 2015 ("miR-155-5p"), along with their calculated Kd and Hill coefficients.* ***(C)*** *Native gel comparing $(GU)_{20}$ mobility to designed stem-loops and hairpin RNAs from Di Ruscio et. al. 2013 and Zhang et. al. 2015. Note the presence of "AA(UUCG)AA loop" on the native gel, designed as a larger loop control, but the Kd of this stem-loop was not measured, which is why it does not appear in Figure 5B.*

We showed that DNMT1 does not bind to RNA stem-loops or hairpins, and that $(GU)_{20}$ was likely not folding into a stable stem-loop structure, meaning that $(GU)_{20}$ may be forming a distinct secondary structure that allowed it to bind DNMT1 with high affinity. Addition of poly(UG) (pUG) repeats is an mRNA processing step of C. elegans. It has been shown that the pUG forms a G-quadruplex-like structure called a pUG fold. This structure requires at least 12 GU repeats and, like canonical G-quadruplexes, requires potassium for proper folding and cannot fold in the presence of lithium [39]. $(GU)_{20}$ RNA fulfills the requirements to form a pUG fold, and all EMSAs presented thus far had been performed in 100mM KCl. Therefore, we wondered if DNMT1 could be recognizing the pUG-folded $(GU)_{20}$, or whether DNMT1 was recognizing the GU repeated sequence in a pUG-folded independent manner. To confirm that $(GU)_{20}$ was capable of forming a pUG-fold, a native gel was run both in lithium and potassium, and different relative mobilities were observed in the presence of the different salts, as compared to Poly(A) which had the same mobility regardless of salt conditions (Figure 6A). Additionally, to determine if the pUG fold structure was necessary for DNMT1 binding, an EMSA with DNMT1 and $(GU)_{20}$ was performed in the presence of 100mM lithium. The binding affinity of $(GU)_{20}$ in the presence of lithium was 105.8±5.2nM, compared to 82.9±6.8nM in potassium (Figure 6B). This indicates that DNMT1 has a slight preference for binding pUG folded RNA, which agrees with the fact that DNMT1 also binds to canonical G-quadruplexed RNA with moderate affinity [32]. However, given that DNMT1 binding to $(GU)_{20}$ was not completely disrupted by the presence of lithium instead of potassium, there is likely some other factor that is allowing $(GU)_{20}$ to bind to DNMT1 with high affinity.

**A.**




**B.**

| $(GU)_{20}$ EMSA Condition | Kd (nM) | Hill |
|---|---|---|
| 100mM KCl | 82.9±6.8 | 1.1±0.1 |
| 100mM LiCl | 105.8±5.2 | 1.5±0.1 |

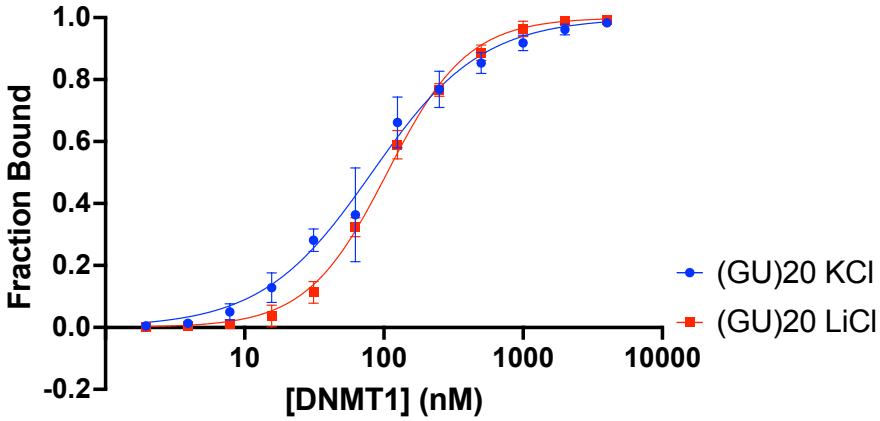



***Figure 6:*** *DNMT1 has a slightly higher affinity for pUG folded* $(GU)_{20}$ *over unfolded* $(GU)_{20}$. ***(A)*** *Native gels of quadruplex-forming RNAs in the presence of 100mM KCl or LiCl.* $(GU)_{12}$ *can form the pUG fold while* $(GU)_{11}$ *cannot. Native gel data obtained by Linnea Jansson-Fritzberg.* ***(B)*** *Binding curves of* $(GU)_{20}$ *binding to DNMT1 in the presence of 100mM KCl or 100mM LiCl, along with their calculated Kd's and Hill coefficients.*

**DNMT1 can simultaneously bind multiple nucleic acid molecules**

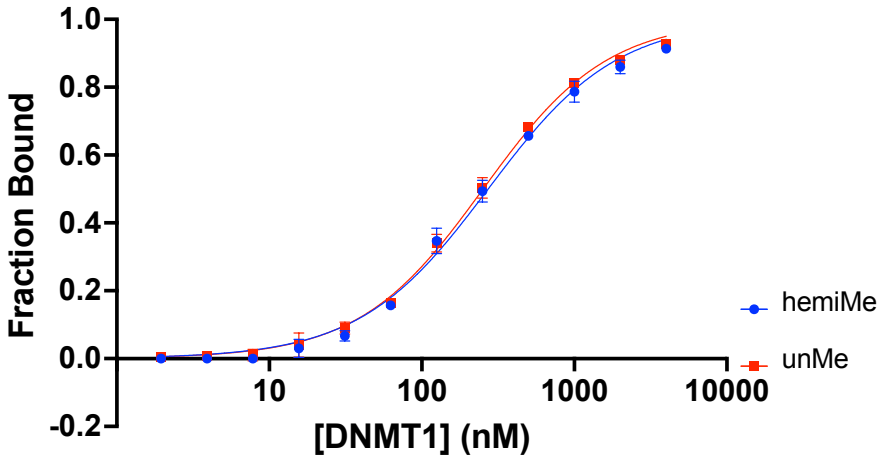
DNA is the substrate of the DNMT1-catalyzed DNA methylation reaction, and DNA is known to bind toDNMT1's catalytic domain near the C-terminus of the protein [15, 20, 40]. However, even though RNA has been shown to bind DNMT1, DNMT1 has no canonical RNA binding domains [15] and it is not known which site of the protein interacts with RNA. To determine whether or not RNA also binds in the catalytic site of DNMT1, EMSA competition assays between DNA and RNA were performed. EMSAs between DNMT1 and double stranded unmethylated $C(CGGCCG)_{12}$ and hemimethylated $C(C_{me}GGCCG)_{12}$ DNA were performed first to determine their relative Kds. Both unmethylated and hemimethylated DNA bind DNMT1 with similar affinities, 252±12nM and 272±17nM, respectively (Figure 7A), in agreement with previous reports [33, 41]. As DNMT1 is a maintenance methyltransferase, hemimethylated DNA

is its most physiologically relevant substrate and therefore it was used for all DNA-RNA competition assays. The DNA-RNA competition assays were performed with constant 0.5uM DNMT1, which is slightly above the Kd between DNMT1 and hemimethylated DNA, constant radio labeled DNA across all lanes. Unlabeled RNA was then titrated in two-fold increasing concentrations, while Poly(A) was used as a negative control (Figure 7B). "RNA3" was titrated from 30nM to 1uM, where "RNA3" was fully bound to DNMT1 at 1uM (Figure 2C). However, the "RNA3" fragment from *DNMT1* mRNA did not completely compete off DNA binding, as evident by the presence of a shifted band at 1uM "RNA3" (Figure 7B), suggesting multiple nucleic acid binding sites on DNMT1. At the time of carrying out the competition assays, $(GU)_{20}$ RNA bound to DNMT1 with the highest affinity, it was therefore also tested for DNA competition. Interestingly, it seemed that $(GU)_{20}$ RNA was able to compete off DNA binding (Figure 7B), at high $(GU)_{20}$ concentrations. It is unknown why there only appears to be competition for the same binding site at high (GU)20 concentrations. However, this may indicate that different RNAs can bind to different sites on DNMT1.

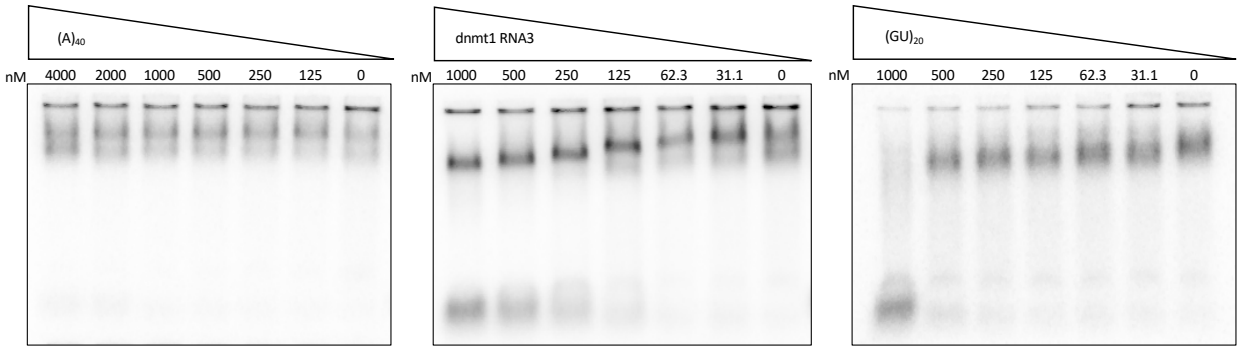
To confirm that multiple RNAs could bind to DNMT1 simultaneously, RNA-RNA competition assays were performed. The experimental set up was similar to that of the DNA-RNA competition assays, where radiolabeled $(GU)_{20}$ concentration was constant across all lanes, and 125nM DNMT1 was constant in all lanes. Unlabeled Poly(A) RNA or "RNA3" were added in increasing concentrations, Poly(A) was the negative control and did not affect $(GU)_{20}$ binding to DNMT1 (Figure 7C). Titrating in unlabeled "RNA3" appeared to match the shifting pattern as DNA-"RNA3" competition, where even high concentrations of "RNA3" did not completely compete off $(GU)_{20}$ binding (Figure 7C). This indicates that multiple RNAs are able to bind DNMT1 at the same time.

**A.**

| DNA | Kd (nM) | Hill |
|---|---|---|
| hemiMe | 272±17 | 1.03±0.06 |
| unMe | 252±12 | 1.07±0.05 |



**B.**




**C.**

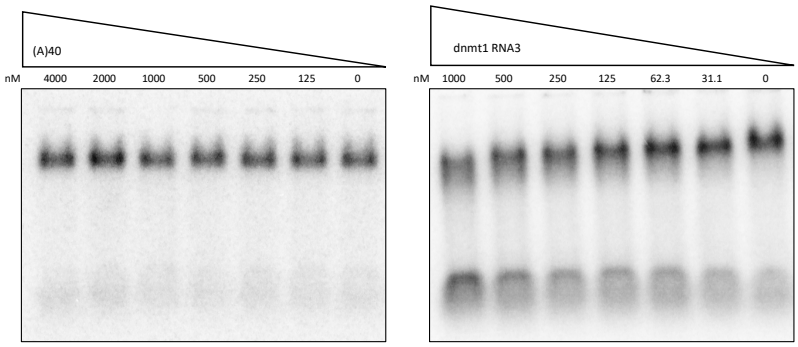



***Figure 7:*** *DNMT1 can form a ternary complex with multiple nucleic acid molecules.* ***(A)*** *Binding curves between DNMT1 and hemimethylated or unmethylated double stranded DNA, along with the calculated Kds and Hill coefficients.* ***(B)*** *Competition assay gels between DNA and RNA. Radiolabeled DNA concentration and 0.5uM DNMT1 protein concentration across all lanes, unlabeled RNA titrated in across lanes, see wedges.* ***(C)*** *Competition assay gels between different molecules of RNA. Radiolabeled $(GU)_{20}$ concentration and 125nM DNMT1 protein concentrations constant across all lanes, unlabeled competing RNA titrated in across lanes, see wedges.*

## DISCUSSION

In this study, purified DNMT1 was used in binding assays with various RNAs to determine which RNA sequence and secondary structure motifs allowed them to bind to DNMT1 with high affinity. By fRIP-seq analysis, DNMT1 was shown to interact with its own fully spliced mRNA in cellular nuclei, with a bias for binding toward the 3'end of the mRNA. Subsequent EMSAs were done to verify this binding interaction, showing that DNMT1 can bind to segments of its own mRNA in vitro. The interaction between DNMT1 and its own mRNA has not been previously shown, and may indicate a possible mechanism of autoregulation, where DNMT1 activity may be modulated upon mRNA binding, or where DNMT1 may be targeted to specific genomic loci by an mRNA scaffold. The approaches used in this study were able to identify the interaction between DNMT1 and its mRNA, however the arbitrary fragmentation of the mRNA for EMSA experiments likely disrupted secondary structure features that may be present in the full length mRNA. Therefore the focus of the study shifted to identify common RNA sequence and structure motifs that can be recognized by DNMT1 for binding.

Since DNMT1 is known to bind to and methylate CpG dinucleotides in double stranded DNA, we tested whether or not the same features were necessary for RNA binding. Results show that DNMT1 has a very slight preference for double stranded over single stranded RNA, but does not seem to recognize CpG dinucleotides in RNA. This suggests that RNA binding to DNMT1 is likely distinct from DNA binding. Various synthetic polynucleotide tract and di-/tri-nucleotide tract small RNAs were tested to determine if binding is RNA sequence or structure specific. Firstly, these experiments showed that RNA binding affinity is not solely dependent on RNA length, as all RNAs were 40 nucleotides long but displayed a wide range of binding affinities. DNMT1 bound with moderate affinity to C and G rich RNA oligomers, and bound with low or very low affinity to A and U rich RNAs.

DNMT1 has previously been shown to preferentially bind G-quadruplex forming RNAs [32], which is corroborated by the fact that DNMT1 bound to $(GGAA)_{10}$, which is able to form G-quadruplexes, with higher affinity than $(GA)_{20}$, which has the same G and A content but is unable to form any secondary structure. Most notably, DNMT1 bound to $(GU)_{20}$ with surprisingly high affinity which has not been shown prior to this study. We decided to further investigate the possible reasons that allow $(GU)_{20}$ to bind to DNMT1 with higher affinity than any other RNA tested.

The DNMT1 fRIP-seq data showed that there seems to be a bias for DNMT1 to interact with the 3' end of its mRNA transcript. We also noticed that the 3'UTR of *DNMT1* seems to be fairly GU rich. An EMSA with the full length 3'UTR showed high affinity binding. However a similar length RNA that was not GU rich, hTR34-328, also bound with high affinity indicating that the 3'UTR may be binding with high affinity to DNMT1 because it is longer than other RNAs previously tested. Additionally, when shorter regions of the 3'UTR which appeared to be GU rich were tested, they also did not bind with comparable affinity to the full length 3'UTR. This prompted the analysis of the GU content of the *DNMT1* mRNA fragments, showing that the GU content of an RNA sequence alone was not enough to confer high affinity binding to DNMT1.

To determine why $(GU)_{20}$ RNA was binding with such high affinity to DNMT1, we investigated what type of secondary structures could be formed by $(GU)_{20}$. Since G and U could wobble base pair with each other, it was possible that $(GU)_{20}$ was forming a stem-loop structure. Several 40 nucleotide long RNAs designed to form stem-loop structures, with varying loop sizes and sequences, were used in EMSAs to determine their affinity to DNMT1. We showed that DNMT1 does not bind to the tested stem-loops with comparable affinity to $(GU)_{20}$. Previous publications have proposed that DNMT1 preferentially binds stem-loop and hairpin RNA structures over unstructured RNA [32, 33] but when the same RNAs from those publications were tested here, they did not bind to DNMT1 with high affinity. This study thus contradicts previous publications stating that DNMT1 preferentially binds stem-loop structures, instead showing that stem-loop RNAs bind with moderate to low affinity.

It has also been shown that GU repeated sequences can form a G-quadruplex like structure, a pUG fold, that is dependent on the presence of potassium, and is unable to form in lithium [39]. Native polyacrylamide gels confirmed that $(GU)_{20}$ was folding differently in potassium and lithium. The EMSA between $(GU)_{20}$ shown in Figure 3C was performed in 100mM KCl, to determine whether the pUG fold was necessary for $(GU)_{20}$-DNMT1 binding, an EMSA was performed in 100mM LiCl to prevent the formation of the pUG fold. In the presence of lithium, DNMT1's affinity for $(GU)_{20}$ was slightly lower than in the presence of potassium. This indicates that DNMT1 may have a slight preference for the pUG fold structure, however the

fact that the binding interaction was not completely disrupted when performed in lithium, suggests that there may be some other feature of $(GU)_{20}$ that allows it to bind with high affinity. It is worth noting that no other RNA was tested in the presence of lithium instead of potassium, it is therefore unknown whether lithium ions somehow alter DNMT1's affinity for any RNA. It is possible that DNMT1 may be recognizing total G and U content within an RNA sequence, which could be tested by measuring the binding affinity with various scrambled G and U rich sequence, or an RNA may need to have a certain number of repeated GU dinucleotides. Conversely, DNMT1 may have affinity for both the folded and unfolded versions of $(GU)_{20}$. The results from this study merit further investigation into what feature of $(GU)_{20}$ allows it to bind to DNMT1 with high affinity.

Finally, it was shown that DNMT1 is able to bind multiple nucleic acid molecules simultaneously. The binding site of RNA on DNMT1 has not been previously described and the results from this study show that different RNAs can bind to DNMT1 in different manners. $(GU)_{20}$ can bind DNMT1 with high affinity and compete with DNA for binding, the mechanism of competition remains unknown but may result from directly binding the catalytic site of DNMT1 or allosterically preventing DNA binding. Conversely, the *DNMT1* fragment "RNA3" only partially competes with DNA binding, indicating a ternary complex between DNMT1, DNA, and RNA. Additionally, similar experiments showed that two different RNA molecules can simultaneously bind to DNMT1. A complex of DNMT1 bound to multiple nucleic acid molecules opens an exciting new avenue for future research in the field of RNA-mediated regulation of DNMT1. RNA binding to multiple sites on DNMT1 may be acting allosterically to activate or inhibit catalytic activity, and could be interacting with the N-terminal autoregulatory domains of DNMT1 to modulate methyltransferase activity.

There are several areas of potential future research for this study. To elucidate the possible effects of RNA binding on DNMT1 methyltransferase activity, DNA methylation assays in the presence of these different RNAs will be performed. Protocols have been established to measure the incorporation of [$^3$H] SAM into a hemimethylated CG DNA substrate. These experiments will help determine the biological function of DNMT1 interacting with RNA. In addition, structural studies of DNMT1 complexed with multiple nucleic acid molecules will help determine the exact RNA binding site(s) on DNMT1. Further understanding the mechanism of interaction, as well as the functional relevance, of RNA binding to DNMT1 will provide insight into potential therapeutics and treatments for diseases related to DNA methylation.

**REFERENCES**


1. Avery, O.T., C.M. Macleod, and M. McCarty, *Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii*. J Exp Med, 1944. **79**(2): p. 137-58.
2. McCarty, M. and O.T. Avery, *Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Ii. Effect of Desoxyribonuclease on the Biological Activity of the Transforming Substance*. J Exp Med, 1946. **83**(2): p. 89-96.
3. Riggs, A.D., *X inactivation, differentiation, and DNA methylation*. Cytogenet Cell Genet, 1975. **14**(1): p. 9-25.
4. Compere, S.J. and R.D. Palmiter, *DNA methylation controls the inducibility of the mouse metallothionein-I gene lymphoid cells*. Cell, 1981. **25**(1): p. 233-40.
5. Sanchez-Romero, M.A., I. Cota, and J. Casadesus, *DNA methylation in bacteria: from the methyl group to the methylome*. Curr Opin Microbiol, 2015. **25**: p. 9-16.
6. Niller, H.H., A. Demcsák, and J. Minarovits, *DNA Methylation in Eukaryotes: Regulation and Function*, in *Cellular Ecophysiology of Microbe*, T. Krell, Editor. 2017, Springer International Publishing: Cham. p. 1-62.
7. Kim, M. and J. Costello, *DNA methylation: an epigenetic mark of cellular memory*. Exp Mol Med, 2017. **49**(4): p. e322.
8. Li, E. and Y. Zhang, *DNA methylation in mammals*. Cold Spring Harb Perspect Biol, 2014. **6**(5): p. a019133.
9. Okano, M., et al., *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development*. Cell, 1999. **99**(3): p. 247-57.
10. Greenberg, M.V.C. and D. Bourc'his, *The diverse roles of DNA methylation in mammalian development and disease*. Nat Rev Mol Cell Biol, 2019. **20**(10): p. 590-607.
11. Hermann, A., R. Goyal, and A. Jeltsch, *The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites*. J Biol Chem, 2004. **279**(46): p. 48350-9.
12. Jones, P.A. and G. Liang, *Rethinking how DNA methylation patterns are maintained*. Nat Rev Genet, 2009. **10**(11): p. 805-11.
13. Alvarez-Ponce, D., et al., *Molecular evolution of DNMT1 in vertebrates: Duplications in marsupials followed by positive selection*. PLoS One, 2018. **13**(4): p. e0195162.
14. Bronner, C., et al., *Coordinated Dialogue between UHRF1 and DNMT1 to Ensure Faithful Inheritance of Methylated DNA Patterns*. Genes (Basel), 2019. **10**(1).
15. Tajima, S., et al., *Domain Structure of the Dnmt1, Dnmt3a, and Dnmt3b DNA Methyltransferases*. Adv Exp Med Biol, 2016. **945**: p. 63-86.
16. Song, J., et al., *Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation*. Science, 2011. **331**(6020): p. 1036-40.
17. Pradhan, M., et al., *CXXC domain of human DNMT1 is essential for enzymatic activity*. Biochemistry, 2008. **47**(38): p. 10000-9.
18. He, Y.F., et al., *Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA*. Science, 2011. **333**(6047): p. 1303-7.
19. Qin, W., H. Leonhardt, and F. Spada, *Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1*. J Cell Biochem, 2011. **112**(2): p. 439-44.

20. Adam, S., et al., *DNA sequence-dependent activity and base flipping mechanisms of DNMT1 regulate genome-wide DNA methylation*. Nat Commun, 2020. **11**(1): p. 3723.
21. Ren, W., et al., *Direct readout of heterochromatic H3K9me3 regulates DNMT1-mediated maintenance DNA methylation*. Proc Natl Acad Sci U S A, 2020. **117**(31): p. 18439-18447.
22. Song, J., et al., *Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation*. Science, 2012. **335**(6069): p. 709-12.
23. Tahiliani, M., et al., *Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1*. Science, 2009. **324**(5929): p. 930-5.
24. Ito, S., et al., *Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification*. Nature, 2010. **466**(7310): p. 1129-33.
25. Somasundaram, S., et al., *The DNMT1-associated lincRNA DACOR1 reprograms genome-wide DNA methylation in colon cancer*. Clin Epigenetics, 2018. **10**(1): p. 127.
26. Merry, C.R., et al., *DNMT1-associated long non-coding RNAs regulate global gene expression and DNA methylation in colon cancer*. Hum Mol Genet, 2015. **24**(21): p. 6240-53.
27. Sharma, S., T.K. Kelly, and P.A. Jones, *Epigenetics in cancer*. Carcinogenesis, 2010. **31**(1): p. 27-36.
28. Kulis, M. and M. Esteller, *DNA methylation and cancer*. Adv Genet, 2010. **70**: p. 27-56.
29. Chen, T., et al., *Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells*. Nat Genet, 2007. **39**(3): p. 391-6.
30. Zhao, Y., H. Sun, and H. Wang, *Long noncoding RNAs in DNA methylation: new players stepping into the old game*. Cell Biosci, 2016. **6**: p. 45.
31. Hendrickson, D., et al., *Widespread RNA binding by chromatin-associated proteins*. Genome Biol, 2016. **17**: p. 28.
32. Zhang, G., et al., *Small RNA-mediated DNA (cytosine-5) methyltransferase 1 inhibition leads to aberrant DNA methylation*. Nucleic Acids Res, 2015. **43**(12): p. 6112-24.
33. Di Ruscio, A., et al., *DNMT1-interacting RNAs block gene-specific DNA methylation*. Nature, 2013. **503**(7476): p. 371-6.
34. Chalei, V., et al., *The long non-coding RNA Dali is an epigenetic regulator of neural differentiation*. Elife, 2014. **3**: p. e04530.
35. Wang, X., et al., *Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines*. Mol Cell, 2017. **65**(6): p. 1056-1067 e5.
36. Davidovich, C., et al., *Promiscuous RNA binding by Polycomb repressive complex 2*. Nat Struct Mol Biol, 2013. **20**(11): p. 1250-7.
37. Detich, N., S. Ramchandani, and M. Szyf, *A conserved 3'-untranslated element mediates growth regulation of DNA methyltransferase 1 and inhibits its transforming activity*. J Biol Chem, 2001. **276**(27): p. 24881-90.
38. Hall, K.B., *Mighty tiny*. RNA, 2015. **21**(4): p. 630-1.
39. Roschdi, S., Yan, J., Nomura, Y., Escobar, C.A., Petersen, R.J., Bingman, C.A., Tonelli, M., Vivek, R., Montemayor, E.J., Wickens, M., Kennedy, S.G., *An Atypical RNA Quadruplex Marks RNAs as Vectors for Gene Silencing*. bioRxiv, 2021.
40. Takeshita, K., et al., *Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1)*. Proc Natl Acad Sci U S A, 2011. **108**(22): p. 9055-9.
41. Flynn, J., R. Azzam, and N. Reich, *DNA binding discrimination of the murine DNA cytosine-C5 methyltransferase*. J Mol Biol, 1998. **279**(1): p. 101-16.