# Supplementary Information

# A DNA Aptamer for Binding and Inhibition of DNA methyltransferase 1

**Linlin Wang,[1,] † Ju Yong Lee,[1,] † Linfeng Gao,[3] Jiekai Yin,[3] Yaokai Duan,[1] Luis A. Jimenez,[4] Gary Brent Adkins,[1] Wendan Ren[2], Linhui Li,[2] Jian Fang,[2] Yinsheng Wang,[1,3] Jikui Song,[2,3,] * and Wenwan Zhong[1,3,] ***

[1]Department of Chemistry, University of California-Riverside, Riverside, CA 92521, U.S.A. [2]Department of Biochemistry, University of California-Riverside, Riverside, CA 92521, U.S.A. [3]Environmental Toxicology Program, University of California-Riverside, Riverside, CA 92521, U.S.A. [4]Program in Biomedical Sciences; University of California-Riverside, Riverside, CA 92521, U.S

## Table of Contents

**Table S1.** Random ssDNA libraries used in SELEX on DNMT1.

| Name | Sequences |
|---|---|
| **ssDNA library used for SELEX on DNMT1** | 5’-AGC AGC ACA GAC GTC AGA TG-$(N)_{20}$ – CCTATGCGTGCTACCGTCAA-3’ |
| **Forward primer for SELEX on DNMT1** | 5’-FAM-AGCAGCACAGACGTCAGATG-3’ |
| **Reverse primer for SELEX on DNMT1** | 5’-Biotin-TTGACGGTAGCACGCATAGG-3’ |

**Table S2.** Sequence analysis of the selected round 10th and 11th pool. Four aptamers (Apt. #1-4) with highest abundances were discovered, and the strands with small sequence difference than these major ones were listed as well, with the difference in their sequences indicated in colors, coded by the aptamer groups.

| | Name | Random region sequence | Length | Seq._num |
|---|---|---|---|---|
| Round 11th | Apt#1 | TGCGAGTGACATCTCAACGG | 20 | 1269 |
| | Apt#1-1 | TGCGAGTGACATCGCCACGG | 20 | 567 |
| | Apt#2 | TGGGGGTGAGTCCACTTCTG | 20 | 805 |
| | Apt#2-1 | TGGGGGCGAGTCCACTTCTG | 20 | 563 |
| | Apt#2-2 | TGGGGGAGAGTCCACTTCTG | 20 | 455 |
| | Apt#2-3 | TGGGGAAGGGTCCACTTCTG | 20 | 293 |
| | Apt#3 | TAGCGAGTGTACTCTGAGGG | 20 | 531 |
| | Apt#3-1 | TCGCGAGTGGACTCTGAGGG | 20 | 526 |
| | Apt#3-2 | TCGCGAGTGGACTCTGAGGA | 20 | 440 |
| | Apt#4 | CCCATGCGTAGGGTCTGATG | 20 | 492 |
| | Apt#4-1 | CCAACGCGTATGGTCTGATG | 20 | 398 |
| | Apt#4-2 | CCGAACCGTACGGTCTGATG | 20 | 384 |
| | Apt#5 | TCGACGCGTACGAGCTGACTG | 21 | 508 |
| | Apt#6 | GCGAACCGTACGCTCTGCCGG | 21 | 476 |
| | Apt#7 | TGGCATGCACTCTACGACGG | 20 | 374 |

| | Name | Random region sequence | Length | Seq._num |
|---|---|---|---|---|
| Round 10th | Apt#1 | TGCGAGTGACATCTCAACGG | 20 | 946 |
| | Apt#1-1 | TGCGAGTGACATCGCCACGG | 20 | 476 |
| | Apt#2 | TGGGGGTGAGTCCACTTCTG | 20 | 387 |
| | Apt#2-1 | TGGGGGCGAGTCCACTTCTG | 20 | 300 |
| | Apt#2-2 | TGGGGGAGAGTCCACTTCTG | 20 | 299 |
| | Apt#3 | TAGCGAGTGTACTCTGAGGG | 20 | 552 |
| | Apt#3-1 | TCGCGAGTGGACTCTGAGGG | 20 | 385 |
| | Apt#3-2 | TCGCGAGTGGACTCTGAGGA | 20 | 349 |
| | Apt#3-3 | TAGCGGGTGTACTCTGTCGG | 20 | 371 |
| | Apt#3-4 | TCGCGAGTGGACTCTGCGTA | 20 | 343 |
| | Apt#3-5 | TCGCGAGTGGACTGTGACGG | 20 | 318 |
| | Apt#4 | CCCATGCGTAGGGTCTGATG | 20 | 307 |
| | Apt#4-1 | CCAACGCGTATGGTCTGATG | 20 | 243 |
| | Apt#4-2 | CCGAACCGTACGGTCTGATG | 20 | 268 |
| | Apt#5 | GCGAACCGTACGCTCTGCCGG | 21 | 391 |
| | Apt#6 | TCGACGCGTACGAGCTGACTG | 21 | 380 |

**Table S3.** The modified aptamer sequences based on the full-length aptamers shown in Figure S5. Secondary structures can be found in Figure S7.

| Label | full sequences (random region highlighted as red) | description |
|---|---|---|
| **Modified ssDNA #5** | 5’-CGT CAG ATG TGC GAG TGA CAT CTC AAC GG-3' | Modified from Apt. #1 |
| **Modified ssDNA #6** | 5’-CGT AGA TGT GCT GAC ATC TAC GG-3' | Modified from Apt. #1 |
| **Modified ssDNA #7** | 5’-CAG ATG TGC GAG TGA CAT CTC-3’ | Modified from Apt. #1 |
| **Apt. #8** | 5’-CGT AGA TGT GCG AGT GAC ATC TAC GG-3’ | Modified from Apt. #1 |
| **Apt. #9** | 5’-AGA AGT GGG GGT GAG TCC ACT TCT G-3’ | Modified from Apt. #2 |
| **Apt. #10** | 5’-CCC ATG CGTAGG GTC TGA TGC CTA TGC GTG-3’ | Modified from Apt. #4 |
| **Apt. #11** | 5’-AGA TGT GCG AGT GAC ATC TCA ACG GCC TAT-3’ | Modified from Apt. #1 |
| **Modified ssDNA #12** | 5’-TGC GAG TGA CAT CTC AAC GG-3’ | Modified from Apt. #1 |

**Table S4.** Sequences of the ssDNAs mutated from Apt. #9. Secondary structures shown in Figure S8.

| Name | Sequence |
|---|---|
| **Stems1** | AGGT GGG GGT GAG TCC ACCT G |
| **Stems2** | AT GGG GGT GAG TCC AT G |
| **Steml** | ATAGA AGT GGG GGT GAG TCC ACT TCT ATG |
| **Loops1** | AGA AGT GGG GAG TCC ACT TCT G |
| **Loops2** | AGA AGT GGG G TCC ACT TCT G |
| **Loopl** | AGA AGT GGG GGT TTTTGAG TCC ACT TCT G |
| **Bulge1** | AGA ACT GGG GGT GAG TCC ACT TCT G |
| **Bulge2** | AGA TCT GGG GGT GAG TCC ACT TCT G |
| **Apt. #9** | AGA AGT GGG GGT GAG TCC ACT TCT G |

## I. Concept of AF4-SELEX tested using IgE and its known aptamer

**Figure S1.** Illustration of AF4 channel, and the focusing step employed for on-column ssDNA-target incubation. The sample focusing step compresses the injected volume into a narrow stripe across the breadth axis of the channel by two opposite flows which enter from the channel inlet and a position further down, respectively, and meet at a location determined by the ratio of their flow rates, enhancing the aptamer and target concentrations for better binding.

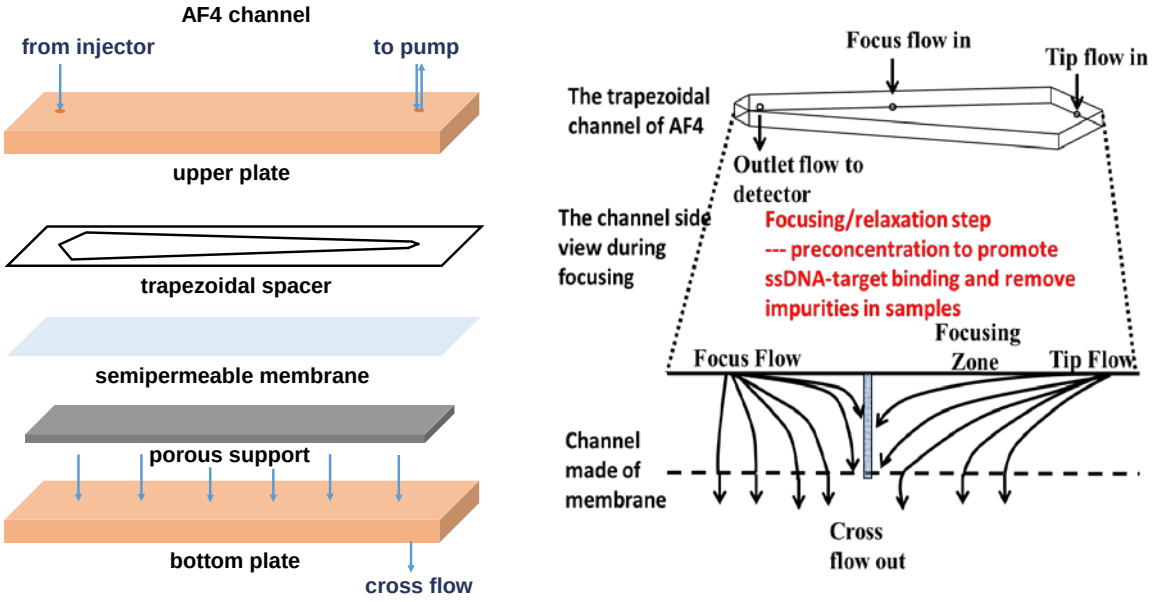



**Figure S2.** Enrichment of the ssDNA eluted by AF4 with the target protein using $TiO_2$ microfibers. G-25 represents the illustra MicroSpin G25 column from GE Healthcare for removing the phosphate salt from the enriched ssDNAs.

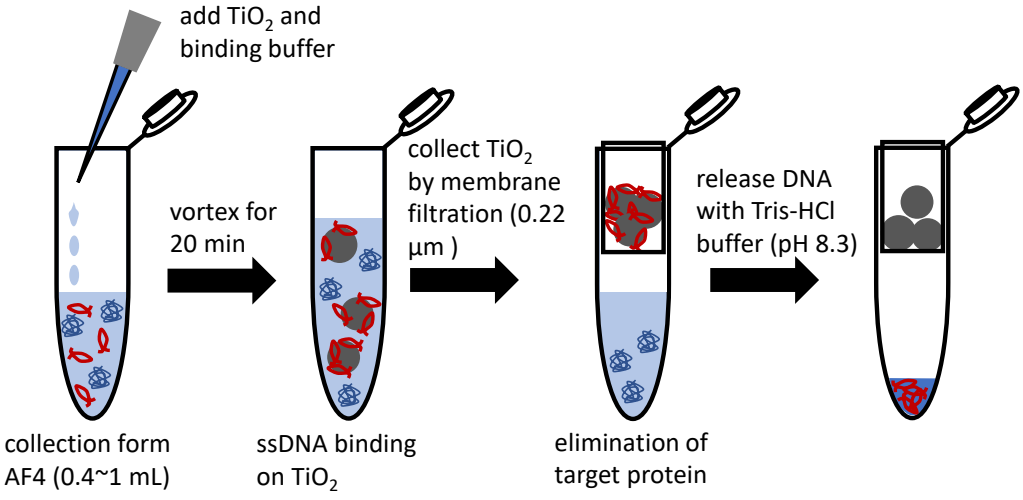
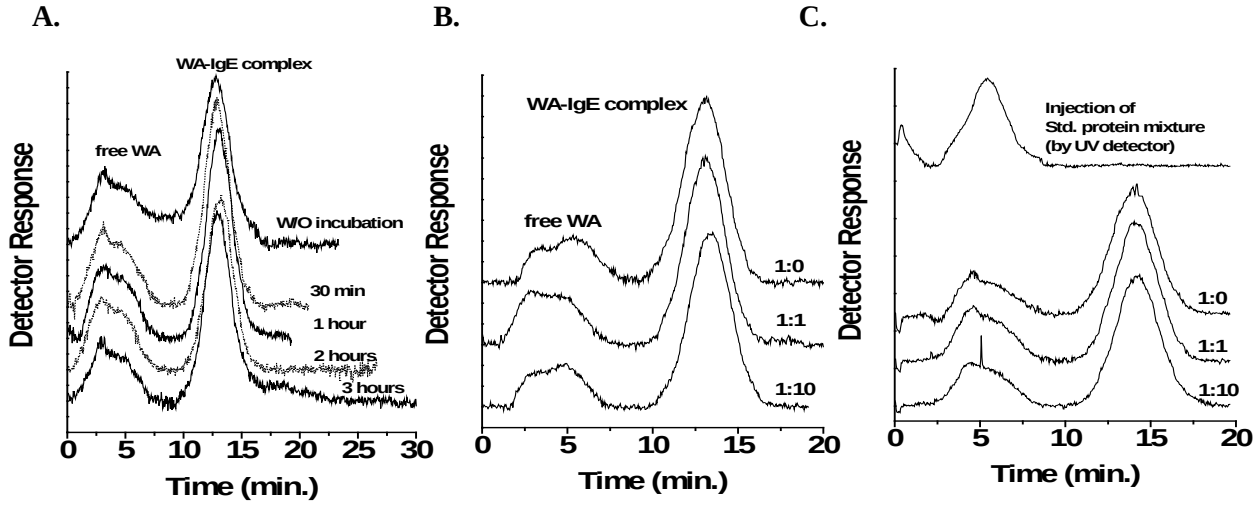


**Figure S3.** Evaluation of on-column complex formation in AF4, and impact of interfering molecules to complex integrity. One hundred pmol of the fluorescein-labeled Weigand aptamer (WA) (37 nt, Mw ~ 22.7 kDa) and 50 pmol IgE (Mw = 150 kDa) were injected; and elution was monitored by the fluorescence detector. The WA-IgE complex peak did not change, when the sample was: **A)** pre-incubated for 0, 30 min, 1 hr, 2 hrs, and 3 hrs before injection to AF4; **B)** mixed with the free, non-binding ssDNA at a molar ratio of 1:0, 1:1, and 1:10 (WA : non-binding ssDNA); or **C)** mixed with the protein mixture added at a mass ratio of 1:0, 1:1, and 1:10 (IgE : protein mixture).




Feasibility of AF4-SELEX was proved using IgE as the standard protein target. Co-injection of IgE and its aptamer WA yielded a peak for the IgE-aptamer complex well resolved from the unbound aptamer. The complex peak area was not changed if the protein and the aptamer were sequentially injected without preincubation for up to 1 hr (Figure S3A). This may be attributed to the enhanced aptamer and target concentrations in the confined sample zone during sample focusing in AF4 that helps speed up complex formation and increase the bound ratio. More effective strands thus can be collected during the first few selection cycles to increase the possibility of aptamer identification in later cycles. Complex formation was also not altered by co-injection of interfering molecules like random ssDNAs or non-target proteins (Figure. S3B & C). This feature makes it possible to include the interfering or competing molecules during SELEX to enhance selection stringency. WA is the Anti-IgE aptamer: 5'-FAM-AGC AGC ACA GAC GTC AGA TGA TG-G GGC ACG TTT ATC CGT CCC TCC TAG TGG CGT GCC CCT CCT ATG CGT GCT ACC GTC AA-3', reported in by Wiegand, T.W. et al. in 1996.[1]

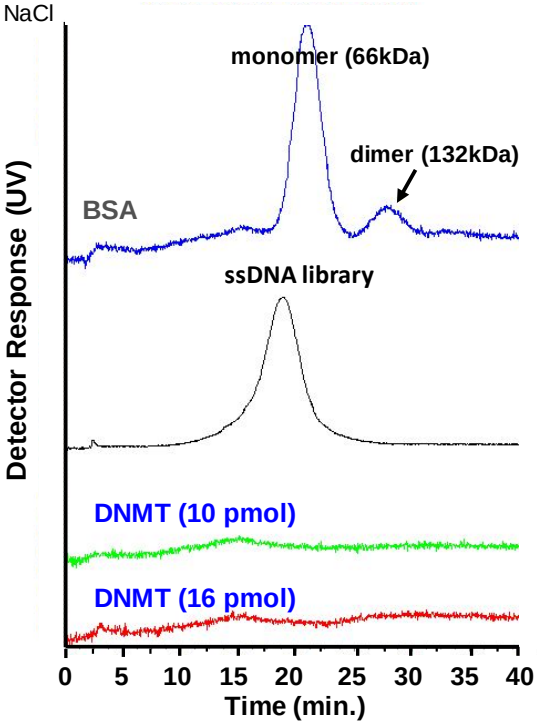
**References:**

(1) T. W. Wiegand, P. B. Williams, S. C. Dreskin, M. H. Jouvin, J. P. Kinet, D. Tasset D, *J Immunol.* **1996**, *157*, 221-230.

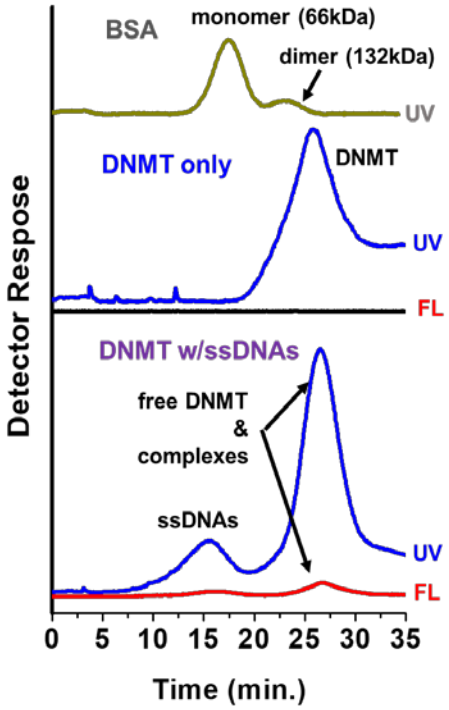
## II. Application of AF4-SELEX for discovery of DNMT1-binding aptamers

**Figure S4.** Optimization of DNMT1 separation conditions in AF4. **A)** Using the separation solution for IgE, elution of DNMT1 was not observed probably due to the denaturation of DNMT1. **B)** Only with the proper solution condition, DNMT1 could be eluted. The peak shape was improved after incubation with ssDNAs, indicating binding.

**A. AF4 separation solution:** 1x PBS with 250 mM NaCl



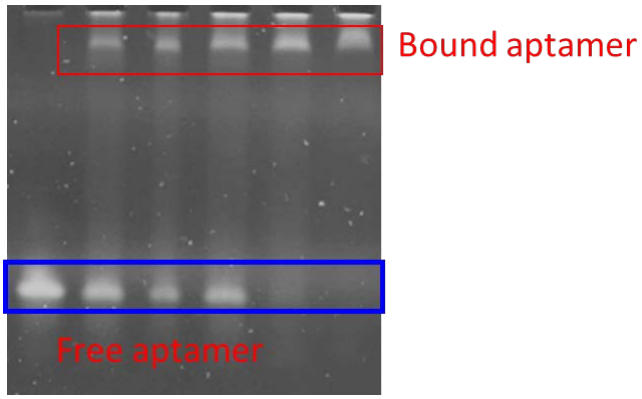

**B. AF4 Separation solution:** 20 mM Tris-HCl, with 500 mM NaCl, 5 mM DTT, 1 mM $MgCl_2$ , cooled



### III. Characterization of the discovered aptamers for DNMT1

**Figure S5.** EMSA results for evaluation of the binding of **A)** Apt. #1 and **B)** Apt. #2 to DNMT1. A fixed amount of Apt. #1 or Apt. #2 (0.4 μM) was incubated with varied amounts of DNMT1 and the mixture was injected to the gel for EMSA. The complex band formed by binding between Apt. #1 and DNMT1 showed up at a much higher position than that of the free aptamer, with the addition of 0.5 μM DNMT1, and the band intensity increased gradually with DNMT1 concentration increasing to 0.8, 1.2, and 2.0 μM bands.

**A.** The EMSA analysis of Apt. #1 binding with DNMT1.




**B.** The EMSA result of Apt. #2 binding with DNMT1, with the fitting plot shown in Figure 3A.

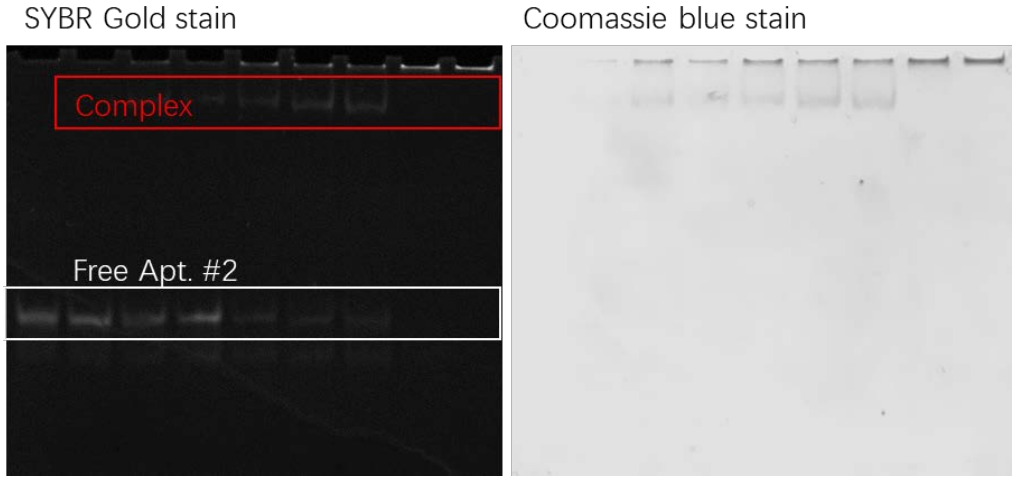


**Figure S6.** Gel images for analysis of DNMT1 binding of Apt. #1 and #2, as well as the ssDNAs shortened from the full-length aptamers as shown in Table S2. Apt. #8, 9, 10, and 11 showed significant binding to DNMT1, and Apt. #9 yielded the strongest complex band (include in the red rectangle). The gel on the left used 10 pmol aptamer (or modified ssDNA) to mix with 6.5 pmol DNMT and the mixture was incubated in 15 μL of Tris buffer (20 mM, pH 7.5, 100 mM NaCl, 5 mM DTT, 1 mM $MgCl_2$) for 30 min at 4 °C, while the one on the right employed 4 pmol aptamer or modified ssDNA.

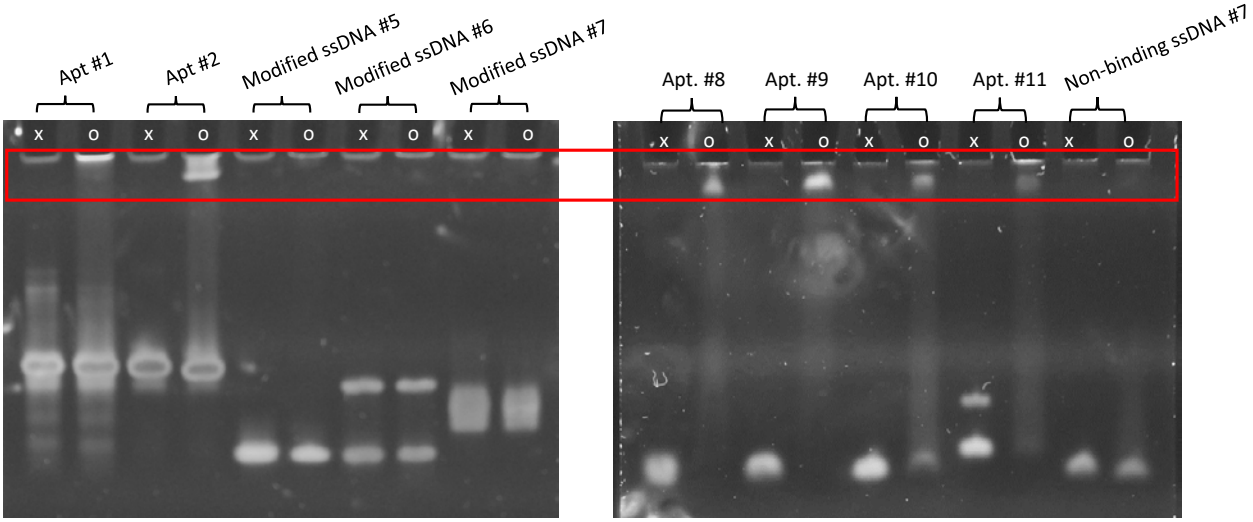
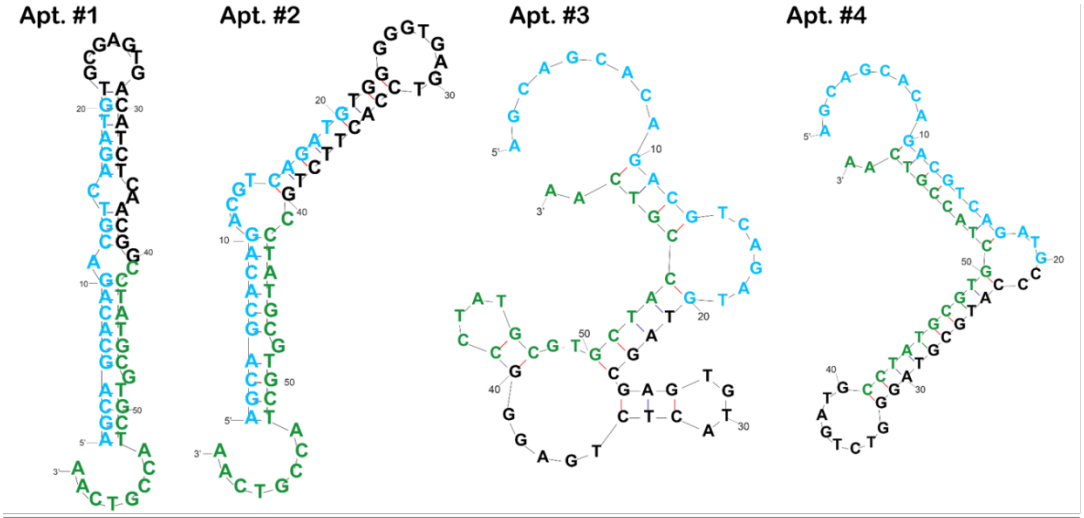


**Figure S7** Predicted structures of **A)** the full length aptamers (Groups of 20-nucleotide sequences were highlighted with blue, black, and green sequentially to facilitate sequence reading) and **B)** the truncation versions from the four full-length aptamers. The DNA folding method was employed for predicting the secondary structures using the Mfold software. ∆ G in kcal/mol.

**A.**




| Name | full sequences (random region highlighted as red) | ΔG (kcal/mol) |
|---|---|---|
| Apt. #1 | 5'-AGC AGC ACA GAC GTC AGA TGT GCG AGT GAC ATC TCA ACG GCC TAT GCG TGC TAC CGT CAA-3' | -18.18 |
| Apt. #2 | 5'-AGC AGC ACA GAC GTC AGA TGT GGG GGT GAG TCC ACT TCT GCC TAT GCG TGC TAC CGT CAA-3' | -29.86 |
| Apt. #3 | 5'-AGC AGC ACA GAC GTC AGA TGT AGC GAG TGT ACT CTG AGG GCC TAT GCG TGC TAC CGT CAA-3' | -13.87 |
| Apt. #4 | 5'-AGC AGC ACA GAC GTC AGA TGC CCA TGC GTA GGG TCT GAT GCC TAT GCG TGC TAC CGT CAA-3' | -16.13 |

**B.**

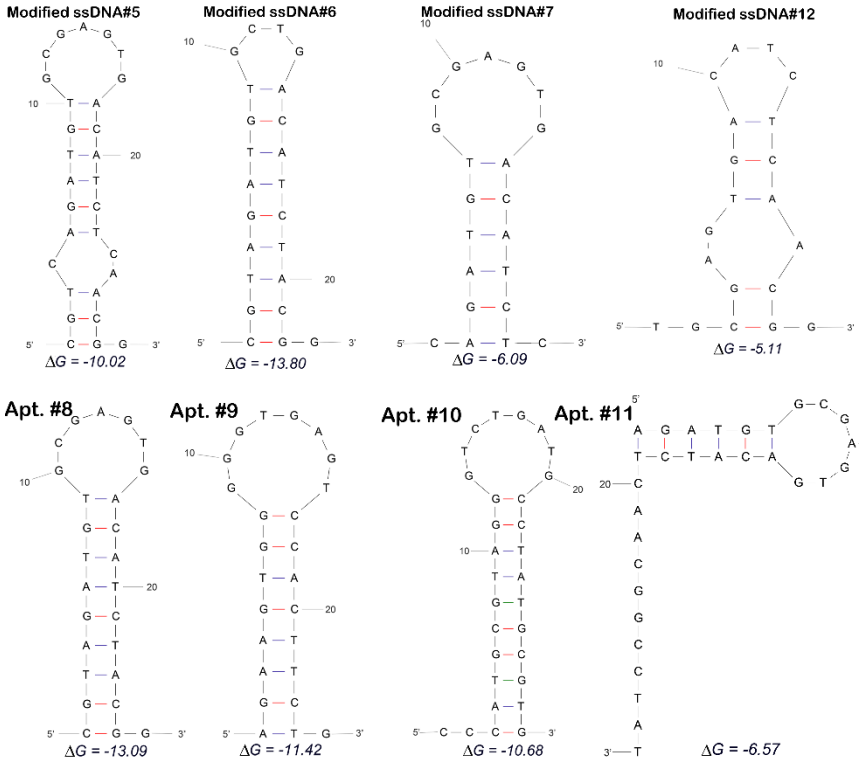
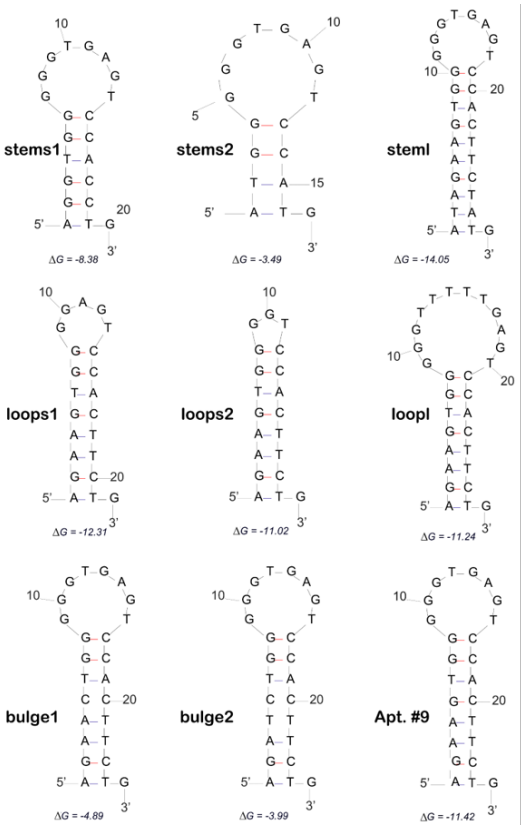
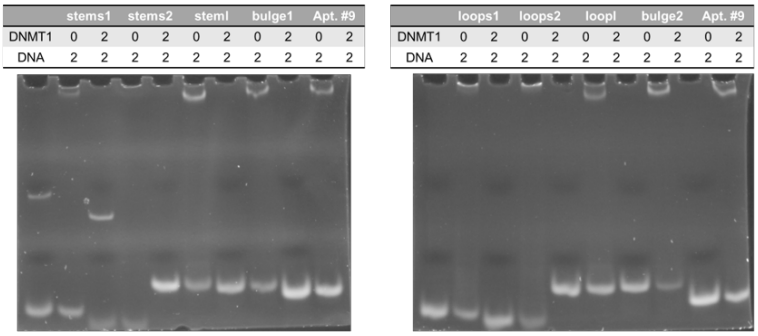


**Figure S8.** Predicted structures of the ssDNA sequences derived from Apt. #9, gel image for EMSA that tested the binding of these strands with DNMT1 (0.4 μM), and the bar plot comparing the band intensity of the DNA-DNMT1 complex calculated by ImageJ. Error bar is the standard deviation calculated from 2 repeats.



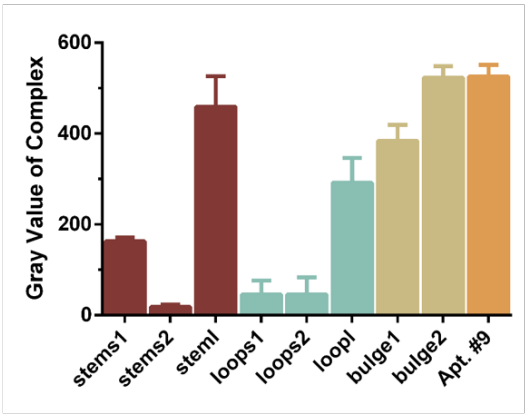
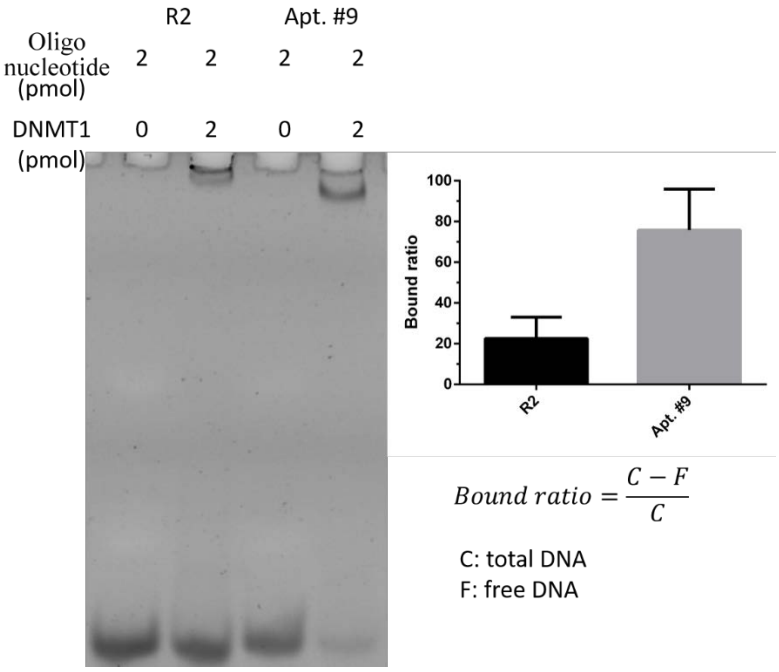



**Figure S9.** Comparison of DNMT1 binding by the RNA sequence, R2, reported by Di Ruscio *et al* in 2013 (Nature, 503, 371-376), and Apt. #9. R2 = 5'- CCC GGG ACG CGG GUC CGG GAC AG. In our EMSA experiment method, judging by the free DNA (or RNA) band intensity, a lower bound ratio was found for R2 compared to Apt. #9 when the same incubation conditions were employed.

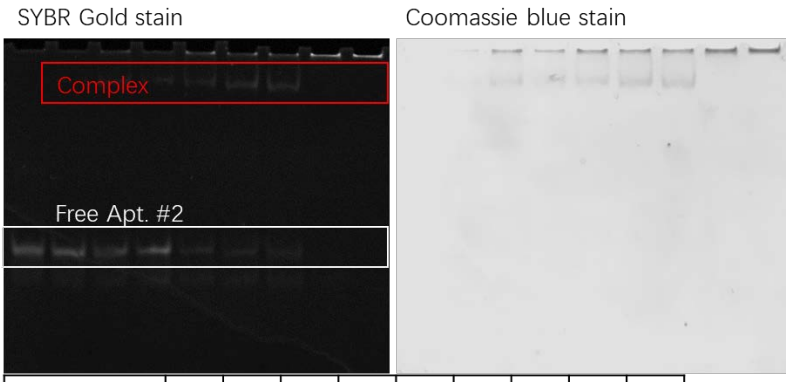



**References:**

Di Ruscio, A., Ebralidze, A.K., Benoukraf, T., Amabile, G., Goff, L.A., Terragni, J., Figueroa, M.E., De Figueiredo Pontes, L.L., Alberich-Jorda, M., Zhang, P. *et al.* (2013) DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature*, **503**, 371-376.

**Figure S10.** Evaluation of DNMT1 binding by **A)** Apt. #9 with the presence of the hemi-methylated DNA (hemiDNA). The binding curve and $K_d$ value obtained from this gel were reported in **B)**, along with those reported in Fig. 3A for Apt. #9 without the presence of hemiDNA. Bound ratio = 1 – (intensity of free aptamer band with addition of DNMT1/intensity of free aptamer band without DNMT1).

**A.**




| DNMT1 (pmol) | 0 | 1 | 2 | 3 | 4 | 6 | 8 | 10 | 14 |
|---|---|---|---|---|---|---|---|---|---|
| Apt. #2 (pmol) | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |

**B.**

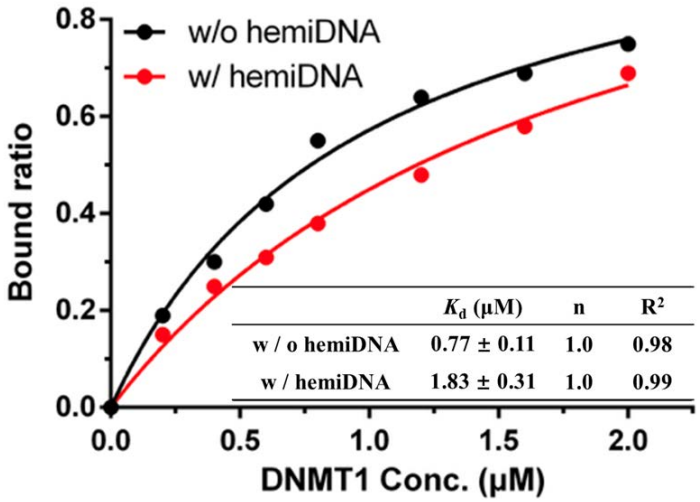


**Figure S11**. Evaluation of DNMT1 binding by hemiDNA **A)** without or **B)** with the presence of Apt. #9. The binding curves and $K_d$ values were reported in Figure 3C. Bound ratio was calculated by dividing the intensity of the hemiDNA-DNMT1 complex band with the sum intensity of all protein bands detected by Coomassie staining.

**A.** **B.**

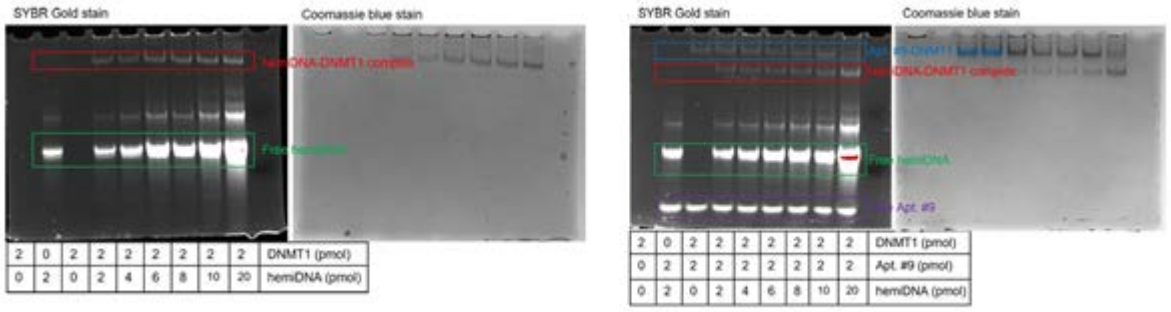



**Figure S12.** The **A)** Gel image and **B)** bar plot that demonstrate the capability of Apt. #9 in displacing hemiDNA off its complex with DNMT1. The sample volume injected was 5 μL.

**A.** **B.**

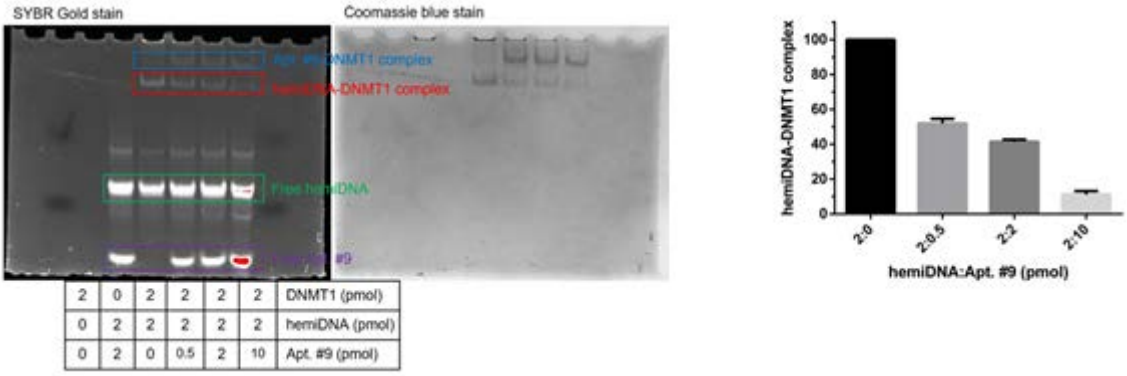
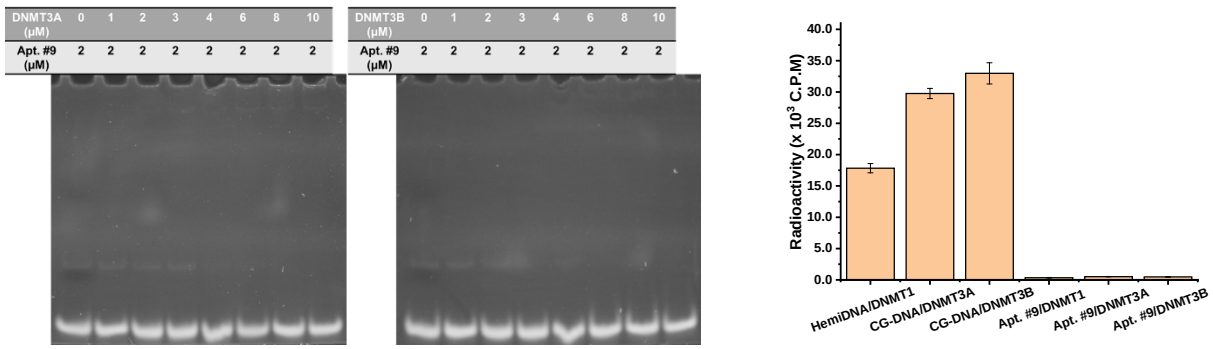



**Figure S13. A)** Gel images for EMSA that tested binding between Apt. #9 and DNMT3A or 3B. **B)** The methylation activity of DNMT1, DNMT3A, and DNMT3B on Apt. #9.

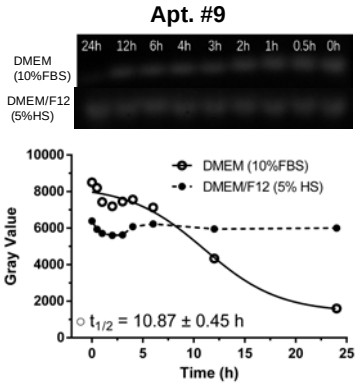
**A.** **B.**



## IV. *In vitro* activity of Apt. #9

**Figure S14.** The stability of **A)** Apt. #9 and **B)** ssDNA12, in the DMEM medium (high glucose) containing 10% FBS (used for HeLa and HEK 293T cells) or DMEM/F12 medium containing 5% horse serum (used for MCF-10A cells).

**A.**




**B.**

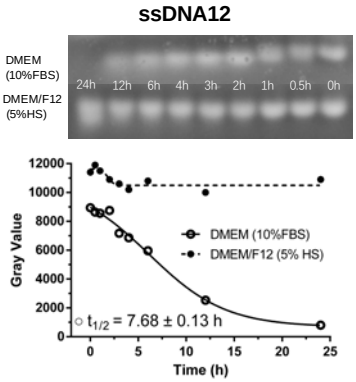
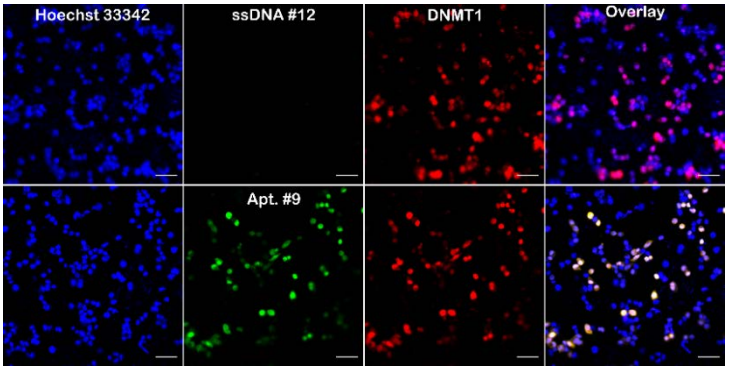
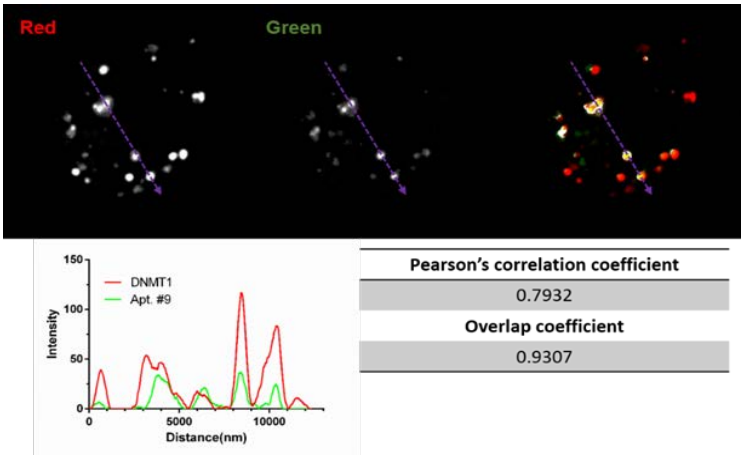



**Figure S15. A)** Live cell imaging for transfected HEK293T cells. Green - Alexa Fluor 488 -labeled Apt. #9; Red - mRuby (red)-tagged DNMT1; Blue - Hoechst 33342 stain for cell nuclei. Scale bar – 20 μm. The non – binding ssDNA #12 was also labeled with Alexa Fluor 488. **B)** Pearson's correlation coefficient of Apt. #9 with DNMT1 obtained from Figure 4A with the colocalization analysis plugin of ImageJ. To further illustrate the overlapping feature, a line (dashed line in purple) was drawn across the image, and the intensity of the green or red fluorescence was plotted to show their colocalization on the line.

**A.**




**B.**




| Pearson's correlation coefficient |
|---|
| 0.7932 |
| **Overlap coefficient** |
| 0.9307 |

**Figure S16.** Uptake of Apt. #9 by living cells. From top to bottom are incubation of 3 μM Apt. #9 with HeLa, MCF-10A, and HEK293 cells stained with Hoechst 33342. Scale bar is 10μm.

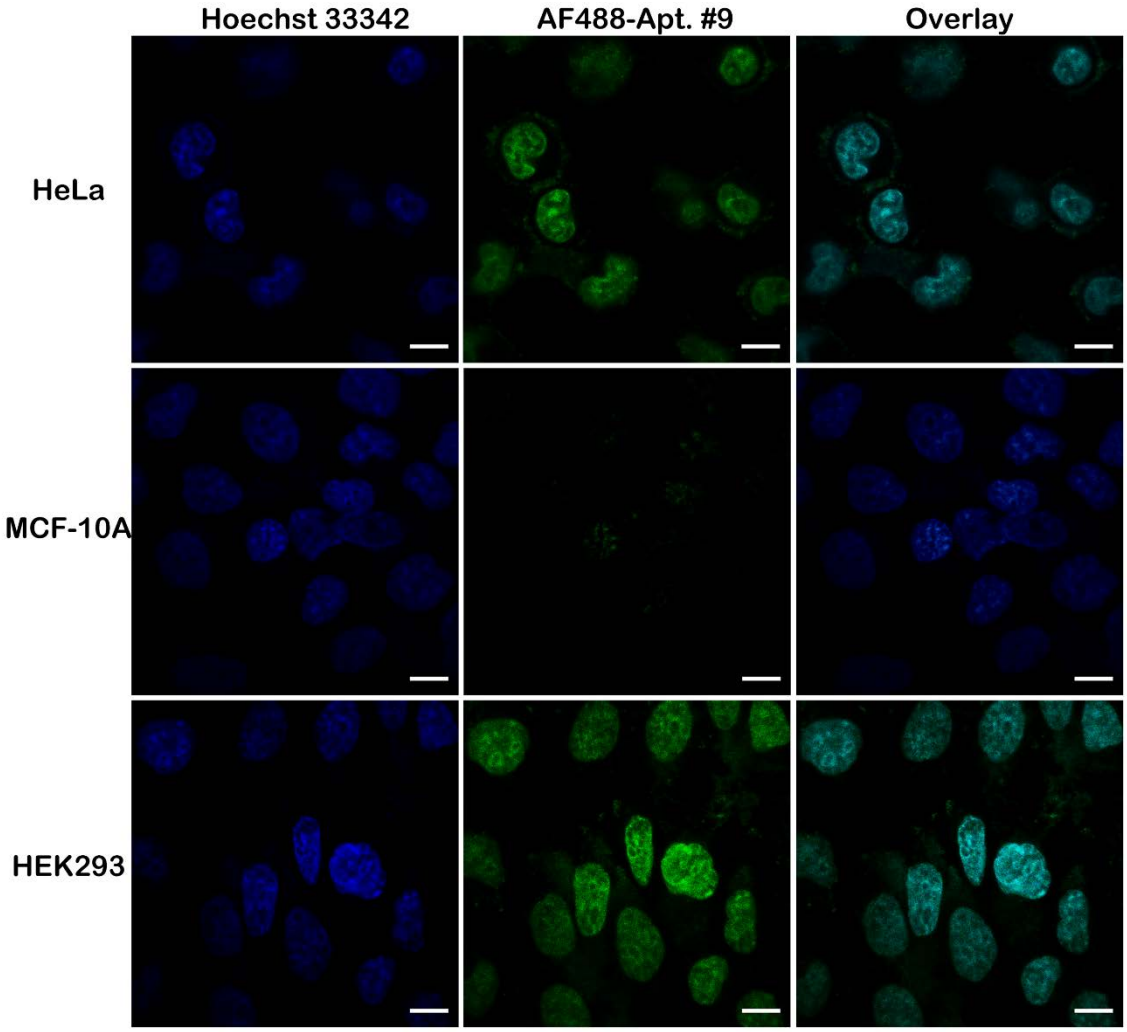


**Figure S17.** Results of restriction enzyme digestion of genomic DNA extracted from cells treated with Apt. #9. The two enzymes employed were HpaII that cleaves only the unmethylated DNA at CCGG sites; and MspI that cleaves the CCGG sites regardless of methylation situation. The HeLa cells were treated with 3 μM Apt. #9 (replenished every 12 hrs) for 12 and 24 hrs, and the genomic DNA was extracted as indicated in Methods. Then 1 μg of DNA was digested by 1 μL of each enzyme (10 units for HpaII, and 20 units for MspI due to their different catalytic efficiency) in CutSmart buffer, incubation 15 min at 37 $^{0}$C.

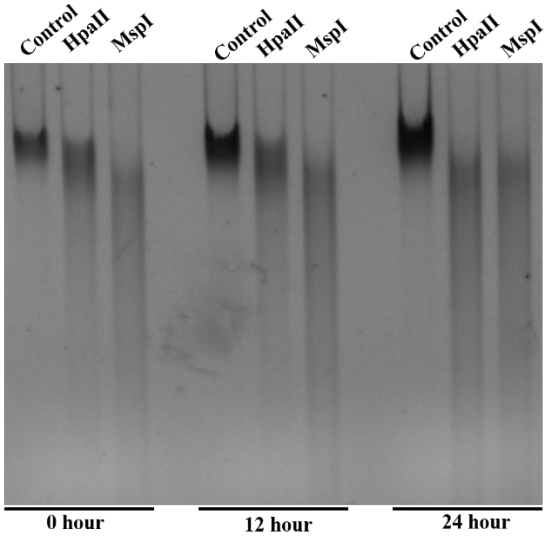
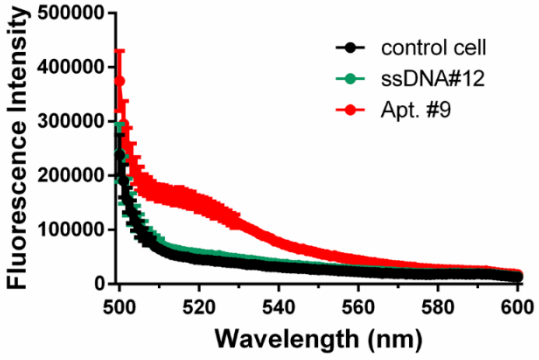



**Figure S18.** Measurement of intracellular aptamer concentration. Procedure: HeLa cells were incubated with 3 μM Apt. #9 or ssDNA#12 for 3 hours, lysed the cells in pure water with 15-min sonication, and measured the fluorescence of the cell lysate using a fluorometer. The aptamer amount taken up by the cells was calculated with the calibration curve prepared from the standard solutions of the Alexa-488-labeled Apt. #9, which was found to be 1.44 x $10^{-12}$ moles. Approximately a total of $10^{5}$ cells were used, thus each cell averagely had 1.44 x $10^{-17}$ moles of the aptamer molecules. The volume of a HeLa cell was reported to be around 2600 μm$^{3}$ (Zhao et al., NMR Biomed, 21, 159-164), thus the average intracellular aptamer concentration should equal to 1.44 x $10^{-17}$ moles/ (2600 x $10^{-18}$ x $10^{3}$ L) =5.5 μM.




**Reference:**

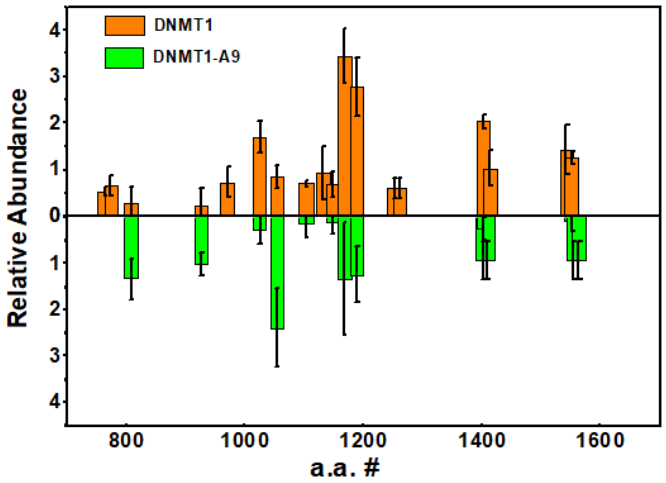
Zhao, L., Kroenke, C.D., Song, J., Piwnica-Worms, D., Ackerman, J.J. and Neil, J.J. (2008) Intracellular water-specific MR of microbead-adherent cells: the HeLa cell intracellular water exchange lifetime. NMR Biomed, 21, 159-164.

**V. Limited proteolysis for exploration of the binding interface between aptamer and DNMT1**

**Method:** DNMT1 in TBS (20 mM Tris, pH 7.5, 250 mM NaCl, 5 mM DTT) was mixed with Apt .#9 at a 1:10 molar ratio and incubated for 1 hr on ice. Later on, protease K was added into the mixture, in 1:30 mass ratio (protease K : DNMT1). The temperature was then changed to 37 $^0$C for 3 min. Afterwards, urea and DTT were added to a final concentration of 4 M and 5 mM respectively. The mixture was incubated at 95 $^0$C for 5 min. After cooled to room temperature, IAA was added into the mixture at 10 mM and incubated for 20 mins in dark. ABC of 50 mM was used to dilute the mixture for 10 times, and Glu-C in 1:50 mass ratio (Glu-C : DNMT1) was added. After 4 hr incubation at 37 $^0$C, the mixture was lyophilized and desalted. The resulted peptides were injected into a CapLC system (Waters Corporation, Milford, MA, USA) coupled with a Finnigan LTQ mass spectrometer (Fisher Scientific, Maltham, MA, USA). Separation was performed in a reverse phase separation column (75 μm i.d. × 10 cm) packed with 5 μm C18 silica beads (Dr. Maisch HPLC GmbH, Germany). Formic acid (0.1%) in water (solution A) and acetonitrile (solution B) were used as mobile phases. The linear gradient started at 2% B and increased to 50 % B during a period of 60 min. The full MS scan was collected within a range of 300-2000 m/z, and MS/MS was carried in a data-dependent mode. The energy for collision induced dissociation (CID) was set to be 35%. The MS/MS spectra were searched against the sequence of human DNMT1 (P26358), using MSGF Plus. The reversed order of the sequence was used as the decoy, and the cutoff of false discovery rate (FDR) was set as 0.01. The relative abundance (RA) of the cutting site i was used for quantification: $\%RA = \frac{SC_i}{\sum SC} \times 100\%$, where SC represents the spectra counting of each identified peptide. Skyline (version 4.1) was used to extract and quantify the peak area of peptides identified by MSGF Plus. The mass accuracy of parent ion was set as 0.7 Da.

**Figure S19. A**) The mirrored bar plot of RA value of identified cutting sties with significant changes (p-value < 0.05). The error bars are calculated for 6 repeats. **B**) Two peptides in the area showing decreased RA values are chosen as examples, and their peak areas in MS1 are integrated by Skyline.

**A.**




**B.**