Published in final edited form as:

Tetrahedron. 2012 July 1; 68(26): 5145–5151. doi:10.1016/j.tet.2012.04.031.

Synthesis of DNA oligos containing 2'-deoxy-2'-fluoro-D-arabinofuranosyl-5-carboxylcytosine as hTDG inhibitor

Qing Dai*, Xingyu Lu, Liang Zhang, and Chuan He*

Department of Chemistry and Institute for Biophysical Dynamics, the University of Chicago, 929 East 57th Street, Chicago, Illinois 60637, USA

Abstract

As an important step of the active demethylation of 5-methylcytosine (5mC), human thymine DNA glycosylase (hTDG) efficiently excises 5-carboxylcytosine (5caC) from double-stranded DNA (dsDNA). Here, we present synthesis of DNA oligos containing a 2'-deoxy-2'-fluoro-D-arabinofuranosyl-5-carboxylcytidine (F-5caC) modification that act as hTDG inhibitors. The glycosylase activity assay showed that F-5caC oligos were resistant to excision by the hTDG catalytic domain (hTDG^{cat}, residues 111–308) and they could inhibit the excision of DNA oligos containing 5caC. The electrophoretic mobility shift assay confirmed that DNA oligos containing F-5caC could bind well with unmodified hTDG^{cat} to form a stable complex, which makes it possible to obtain the crystal structure of the complex to reveal details on how hTDG^{cat} recognizes the DNA substrate.

Keywords

human thymine DNA glycosylase (hTDG); inhibitor; 2'-deoxy-2'-fluoro-D-arabinofuranosyl-5-carboxylcytidine; dsDNA

1. Introduction

As one of the best characterized epigenetic modifications, 5-methylcytosine (5mC) plays important roles in many biological processes including regulation of gene expression, genomic imprinting, X chromosome inactivation, and suppression of transposable elements propagation.¹ Therefore, studies on the formation of 5mC and its demethylation pathways are very important. In contrast to the clear methylation pathway, the active demethylation pathway has been less understood, particularly in mammals.² Recent studies have shown that 5mC can be converted to 5-hydroxymethylcytosine (5hmC, the “sixth” base) in mammalian genome through the oxidation of TET (ten eleven translocation) dioxygenases³ and that 5hmC is a widespread DNA modification in various tissues and cells.⁴ More recently, together with our collaborators, we discovered that 5hmC can be further oxidized into 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) by TET proteins.⁵ These findings strongly supported the hypothesis of the active 5mC demethylation pathway (Fig. 1).⁶ The TET protein-catalyzed conversion of 5mC to 5hmC may represent the first step of a

multiple-step pathway of DNA demethylation. Further oxidation of 5hmC in DNA results in the formation of 5fC and 5caC in DNA. It has been proposed that 5caC may be converted back to cytosine by decarboxylation to complete the 5mC active demethylation pathway.⁷ Despite of extensive efforts, the decarboxylase responsible for removing the 5-carboxyl group has yet to be identified. On the other hand, it has been shown that hTDG can efficiently excise the 5caC base by cleaving the bond between C-1 of the sugar and N-1 of the nucleobase to generate an abasic site in DNA;⁸ subsequent incorporation of an unmethylated cytosine via the DNA base-excision-repair pathway (BER)⁹ establishes the active demethylation cycle (Fig. 1). We have recently published the crystal structures of hTDG^{cat} in complex with a duplex DNA containing 5caC.¹⁰ To avoid the excision of 5caC during the crystallization, an inactive hTDG^{cat} mutant with the active site residue Asn140 mutated to alanine was used. In the meantime, the crystal structure of the DNA complex with unmodified hTDG could be very helpful for elucidating how hTDG interacts with the DNA substrate. To achieve this goal we synthesized dsDNA oligos containing F-5caC. Herein, we described the syntheses, hTDG inhibitory effects and the binding properties of F-5caC-containing DNA oligos with hTDG^{cat}.

2. Results and Discussion

It is well-known that fluorine substitution at the 2'-C position in the 2'-deoxyribose ring can slow down or even inhibit the *N*-glycosidic bond cleavage catalyzed by alkyl-*N*-purine DNA glycosylase.¹¹ Presumably, the strong electron-withdrawing fluorine at 2'-C position destabilizes the obligate oxocarbenium ion intermediate or transition state by increasing its charge density, thereby inhibiting the deglycosylation reaction.¹² Since it is believed that the 2'-F in the arabino configuration will not change the sugar pucker in DNA,¹³ we chose to introduce the 2'-fluoro substituent to the 5caC in the arabino position to synthesize the 2'-deoxy-2'-fluoro-D-arabinofuranosyl-5-carboxylcytosine nucleosides and the phosphoramidites.

We recently reported a highly efficient and convenient method for synthesis of 5caC free nucleoside and its phosphoramidite.¹⁴ Based on this procedure, it would be ideal if 2'-F could be introduced to 5caC directly to afford the desired F-5caC. A search for the literature revealed that several methods for introducing fluoro to 2'-C of nucleosides have been reported: (1) the S_N2 fluorination via a sulfonate intermediate.¹⁵ This method is usually suitable for purine instead of pyrimidine nucleosides since O² of the pyrimidine base may attack C-2' to form a 2,2'-anhydropyrimidine byproducts; (2) direct fluorination of arabinonucleosides with diethylaminosulfur trifluoride (DAST)¹⁶ and (3) ring opening reaction of 2,2'-anhydropyrimidine.¹⁷ Unfortunately, both method 2 and 3 introduces fluoro into C-2' from *a* face.¹⁸ Introduction of 2'-F into the arabino position of nucleosides usually involves fluorination of the sugar without a base followed by the glycosylation reaction.¹⁹ Therefore, we chose to synthesize 2'-deoxy-2'-fluoro-D-arabinofuranosyl-5-iodocytidine derivative²⁰ by glycosylation reaction, then substitute 5-iodo with 5-carboxyl group through palladium-catalyzed carbonyl reaction to prepare F-5caC (Scheme 1).

The glycosylation acceptor, glycosyl bromide **2** was prepared by first fluorination of commercially available ribose derivative **1** with DAST followed by bromination with hydrogen bromide in acetic acid.²¹ The glycosylation donor, the silylated 5-iodocytosine was prepared by first treating the commercially available cytosine with iodine and iodic acid to furnish 5-iodocytosine in 85% yield followed by treatment with hexamethyldisilazane.²² *N*-Glycosylation of the bromide **2** with silylated 5-iodocytosine in 1,2-dichloroethane and acetonitrile gave a mixture of β and α anomers at C-1' with the β -anomer **3a** as the main product (the ratio of $\beta/\alpha = 3:1$ as indicated by ¹H NMR spectrum).²³ Unfortunately, the two anomers migrated the same on TLC and could not be separated by

column chromatography at this stage. It has been reported that $^3J_{1',2'}$ in β anomer with a 2'-arabino substituent is in the range of 3–4 Hz whereas the α anomer shows very small or no coupling between the 1' and 2' protons.²⁴ In our case, analysis of 6.32–6.37 ppm region reveals a doublet of doublet (dd) for the H-1' resonance of the β anomer with $J_{H1',F} = 21.0$ Hz and $^3J_{1',2'} = 3.0$ Hz, while 6.21–6.25 ppm region reveals a doublet (d) for the H-1' resonance of the α anomer with $J_{H1',F} = 14.0$ Hz and no observed coupling between H-1' and H-2'. Treatment of the mixture of **3a** and **3b** with saturated ammonia in MeOH readily removed both benzoyl groups to give a mixture of 5-iodo free nucleosides **4a** and **4b** quantitatively. Using $(\text{PhCN})_2\text{PdCl}_2$ as catalyst and methanol as solvent, **4a** and **4b** was quantitatively converted to the corresponding 5-methoxycarbonyl derivatives **5a** and **5b** in the presence of triethylamine under carbon monoxide atmosphere (75% isolated yield). This new procedure avoided using DMF as well as the more expensive catalyst $\text{Pd}_2(\text{dpa})_3$ but was better than the one we previously reported to synthesize 5-caC;¹⁴ we found the procedure can also be extended to prepare 5-methoxycarbonyl cytidine. Protection of the exocyclic amino group of **5a** and **5b** was realized by treating the mixture of **5a** and **5b** with acetic anhydride in DMF to furnish the mixture of **6a** and **6b**. On TLC, the two anomers of **3a/3b**, **4a/4b** and **5a/5b** could not be separated at all; in comparison, the mixture **6a** and **6b** gave two distinguishable spots that migrate very close on TLC and attempts to purify them by silica gel chromatography always resulted in many fractions containing both **6a** and **6b**. Fortunately, we found that after they were converted to the corresponding 5'-tritylated derivatives **7a** and **7b**, they migrated significantly away from each other, allowing complete separation of the two anomers readily by silica gel column chromatography. Phosphitylation of the 3'-OH of **7a** and **7b** under the standard conditions provided the phosphoramidites **8a** and **8b**, respectively.

Both phosphoramidites **8a** and **8b** were incorporated into a 20mer DNA separately with the sequence of 5'-GAG CGT GAC XGG AGC TGA AA-3', in which X = 1- β -2'-F-5-caC or 1- α -2'-F-5-caC to give resin-bound oligos **β 20** and **α 20**, respectively. The oligos were deprotected by treatment with 0.1 M K_2CO_3 in water and MeOH (v/v = 1:1) overnight at 40 °C. After cooled to rt, AcOH was added to adjust pH to neutral. Oligos **β 20** and **α 20** were then purified by PAGE gel electrophoresis and their masses were verified by MALDI-TOF analysis. For both **β 20** and **α 20**, $m/z = 6278$ of $[\text{MH}]^+$ was observed (6278 calculated). As controls, we also prepared another two DNA oligos **5hmC20** and **5caC20** which have the same sequence except X = 5hmC or 5caC by following the methods that we reported recently.^{14,25} Their masses were also verified by MALDI-TOF analysis. For **5hmC20**, $m/z = 6246$ was observed (6246 calculated); for **5caC20**, $m/z = 6260$ was observed (6260 calculated).

The four synthesized DNA oligos were then 5'-end ^{32}P -labeled and annealed to a complimentary 20mer DNA primer (5'-TTT CAG CTC CGG TCA CGC TC-3') to give the corresponding dsDNA. To test whether they can be excised by hTDG^{cat}, we treated the double-stranded **β 20** and **α 20** with hTDG^{cat} separately followed by NaOH treatment. After the modified base is excised by hTDG^{cat}, it generates an abasic site in DNA, which degrades upon treatment with aqueous NaOH and results in specific strand scission to generate a 5'- ^{32}P -end labeled 9mer. The 9mer product can be detected on the basis of its increased mobility on a polyacrylamide gel. However, if the modified base cannot be excised by hTDG^{cat}, DNA oligos will keep intact as a 20mer. To our delight, we found that both **β 20** and **α 20** stayed intact after treatment with hTDG^{cat} (lane 1 and 2, Fig. 2), suggesting that introduction of 2'-F in the arabino position indeed prevented the excision of base. As a positive control, more than half of the 20mer **5caC20** was converted to 9mer under the same conditions (lane 3 and 6, Fig. 2), suggesting that hTDG did possess the activity to excise the 5caC base from **5caC20**. We then tested whether the addition of **β 20** or **α 20** can inhibit the

excision activity of hTDG^{cat} on **5caC20**. Notable inhibition was observed when 10 equiv. of **α20** was added (lane 4, Fig. 2) and the activity of hTDG^{cat} was sufficiently $\square$ inhibited when 100 equiv. of **α20** was added (lane 5, Fig. 2). We found that **β20** had higher hTDG^{cat} inhibitory activity and 10 equivalent of **β20** showed higher inhibition activity of hTDG^{cat} than the same amount of **α20** (compare lane 7 to lane 4, Fig. 2). As a negative control, addition of 100 equivalent of **5hmC20** did not show any inhibitory effect (lane 9, Fig. 2).

The observation that addition of either **β20** or **α20** oligo inhibited the excision of **5caC20** substrate by hTDG^{cat} may result from the competitive binding of **β20** or **α20** to hTDG^{cat}. To test this hypothesis, we conducted an electrophoretic mobility shift assay (EMSA) to determine the apparent dissociation constant (K_d^{app}). As shown in Fig. 3, hTDG^{cat} binds with both **β20** and **α20** oligos with $K_d^{app} > 10 \mu\text{M}$ for **α20** and $K_d^{app} = 400 \pm 10 \text{ nM}$ for **β20**, indicating that **β20** binds more tightly to hTDG^{cat}. This is consistent with the observation that **β20** showed better inhibitory activity than **α20**. Since hTDG^{cat} binds well with **β20** but does not excise the oligo, it makes this DNA oligo good tools to form the complex with hTDG^{cat} as well as make it possible to obtain its crystal structure to reveal the details on how hTDG^{cat} binds with the DNA substrate. Indeed, after screening different lengths of DNA oligos containing β -F-5caC modification, we obtained the crystal structure of the complex of hTDG^{cat} with a 23mer dsDNA oligo.¹⁰

Since DNA oligos containing F-5caC showed good hTDG^{cat} inhibitory activity, we wondered whether the corresponding free nucleoside or its methyl ester could also inhibit hTDG^{cat}. Since the anomers **3a/3b**, **4a/4b** and **5a/5b** could not be separated by silica gel chromatography, we tried to improve the selectivity of the β anomer by changing the reaction conditions so that purification of the β anomer by recrystallization might be possible. We found that the ratio **3a/3b** was improved to 6:1 from 3:1 in the absence of NaI and CH₃CN in the reaction mixture although the combined yield of **3a/3b** was decreased from 90 % to 75%. Subsequent efforts to obtain pure **3a** by crystallization failed, however, when the mixture of **3a** and **3b** (**3a/3b** = 6:1) was converted to the corresponding mixture of **4a** and **4b**; pure **4a** could be readily obtained by recrystallization from acetone and hexane. Pure **4a** was then converted to **5a** in 92% yield. Compound **5a** could either be hydrolyzed to the corresponding acid **9** quantitatively or be converted to the corresponding phosphoramidite **8a** through **6a** and **7a**. In contrast to the inhibitory effects of DNA oligos **β20** and **α20** to hTDG^{cat}, both nucleosides **5a** and **9** did not show noticeable hTDG^{cat} inhibitory activity even when their concentrations were 1000 times as much as the 5caC DNA concentration. This suggests that hTDG^{cat} could not bind these nucleosides well and the binding requires a stretch of DNA. Indeed, the crystal structure shows that two hTDG molecules bind to one dsDNA with 5caC recognized by one protein while the other hTDG binds nonspecifically ten base pairs away.^{10,26} The hTDG interacts with the backbone of the 5caC-containing strand via electrostatic complementarity, and distorts the dsDNA backbone $\sim 45^\circ$ bending towards the active site.

3. Conclusion

In summary, we have successfully synthesized DNA oligos **β20** and **α20** which contain F-5caC modification. In contrast to the corresponding **5caC20** oligo which can be efficiently excised by hTDG^{cat}, both **β20** and **α20** were kept intact when treated with hTDG^{cat} due to the presence of the strong electron-withdrawing 2'-F substituent in arabino position. Besides, we demonstrated that both **β20** and **α20** could inhibit the excision of **5caC20** oligo by hTDG^{cat} and **β20** exhibited better inhibitory activity than **α20**. Furthermore, we have shown that both **β20** and **α20** could bind with hTDG^{cat} and **β20** has stronger binding affinity than **α20**. These properties of **β20** allowed it to form a stable complex with hTDG^{cat} and the crystal structure of such a complex revealed the details on how DNA substrate interacts with

hTDG^{cat}. Free nucleosides with 2'-F substitutions did not exhibit inhibition effects of hTDG^{cat}; however, these molecules could be further modified as potential small molecule inhibitors as demonstrated previously.²⁷

4. Experimental section

4.1. Synthesis of phosphoramidite 8a and 8b

4.1.1 General experimental procedures—Starting materials and reagents were obtained from commercial suppliers and used without further purifications. All reactions using air sensitive or moisture sensitive reagents were run under an argon atmosphere. ¹H and ¹³C NMR spectra were recorded on Bruker 500 MHz NMR spectrometer. ¹H chemical shifts are reported in δ (ppm) relative to tetramethylsilane, ¹³C chemical shifts δ (ppm) relative to the solvent used. Merck silica gel (9385 grade, 230–400 mesh, 60A, Aldrich) was used for column chromatography. Silica gel on glass with fluorescent indicator (Sigma) was used for TLC. High-resolution mass spectra (HRMS) were obtained from Mass Spectrometry & Proteomics Facility of University of Notre Dame.

4.1.2. Mixture of 1-(2-Deoxy-2-fluoro-3,5-di-O-benzoyl- β -D-arabinofuranosyl)-5-iodocytosine (3a) and 1-(2-deoxy-2-fluoro-3,5-di-O-benzoyl- α -D-arabinofuranosyl)-5-iodocytosine (3b)—To a round bottom flask was added 5-iodocytosine (1.38 g, 5.82 mmol, 1.5 equiv.), (Me₃Si)₂NH (20 mL) and (NH₄)₂SO₄ (100 mg). The mixture was stirred and heated to 127°C with an oil bath. After 1 h, the reaction mixture turned clear. The excess (Me₃Si)₂NH was removed by evaporation in vacuum. To the residue was added NaI (582 mg) and 1-bromo-2-deoxy-2-fluoro-3,5-di-O-benzoyl-D-arabinofuranose (1.64 g, 3.88 mmol, 1.0 equiv.) dissolved in ClCH₂CH₂Cl (30 mL) and CH₃CN (8 mL), and the mixture was stirred at rt for 2 days. Then CH₂Cl₂ (100 mL) was added to dilute the reaction mixture and the well-stirred solution was treated dropwise with MeOH (4 mL). The suspension was filtered through a Celite pad, and the filtrate was washed with water, brine and dried over Na₂SO₄ and evaporated to dryness. The residue was purified by silica gel chromatography, eluting with 1–4% MeOH in CH₂Cl₂, to give a mixture of **3a** and **3b** in a 3:1 ratio (2.02 g, 90% combined yield) as white foam. ¹H NMR (500 MHz) (CDCl₃) δ : 9.32 (br., 0.5H), 8.88 (br., 1.5 H), 7.45–8.10 (m, 11 H), 6.35 (dd, J =21.0 and 3.0 Hz, 0.75 H), 6.22 (d, J =14.0 Hz, 0.25 H), 5.68 (d, J =14 Hz, 0.25 H), 5.60 (dd, J =17.0 and 2.5 Hz, 0.75 H), 5.49 (d, J =47.5 Hz, 0.25 H), 5.41 (dd, J =50.0 and 2.5 Hz, 0.75 H), 5.04 (m, 0.25 H), 4.81 (d, J =4 Hz, 1.5 H), 4.55–4.68 (m, 0.5 H), 4.54 (m, 0.75 H). A similar procedure as above but without adding NaI and CH₃CN gave a mixture of **3a** and **3b** (**3a/3b** = 6 : 1) in 75% yield as indicated by ¹H NMR spectrum. MS calculated for C₂₃H₁₉FIN₃O₆, [MH⁺] 580 (calcd.), 580 (found).

4.1.3. Mixture of 1-(2-Deoxy-2-fluoro- β -D-arabinofuranosyl)-5-iodocytosine (4a) and 1-(2-deoxy-2-fluoro- α -D-arabinofuranosyl)-5-iodocytosine (4b)—To a round bottom flask containing the mixture of **3a** and **3b** (3:1 ratio, 1.80 g, 3.11 mmol) was added saturated ammonia in methanol (50 mL) and the mixture was stirred at rt for 2 days. Removal of all the volatile under reduced pressure. The residue was purified by silica gel chromatography, eluting with 1–10% MeOH in CH₂Cl₂, to give a mixture of **4a** and **4b** in a 3:1 ratio (1.06 g, 92% combined yield) as white solid. ¹H NMR (500 MHz) (DMSO-*d*₆) δ : 8.04 (s, 0.75 H), 7.95 (br. 1 H), 7.93 (s, 0.25 H), 6.74 (br., 1 H), 6.04 (dd, J =16.5 and 3.5 Hz, 0.75 H), 5.82 (m, 0.25 H), 5.85 (m, 1H), 5.18 (m, 1 H), 4.92–5.07 (m, 1 H), 3.46–4.28 (m, 4 H). MS calculated for C₉H₁₁FIN₃O₄, [MH⁺] 372 (calcd.), 372 (found).

4.1.4. 1-(2-Deoxy-2-fluoro- β -D-arabinofuranosyl)-5-iodocytosine (4a)—After treatment of the mixture of **3a** and **3b** (6:1 ratio, 450 mg, 0.78 mmol) with saturated

ammonia in methanol (20 mL) at rt for 2 days. The volatile were removed under reduced pressure to give a white solid. It was recrystallized by acetone and hexane to give pure **4a** (187 mg, 65%) as white solid. ¹H NMR (500 MHz) (CD₃OD) δ: 8.24 (s, 1 H), 6.15 (dd, *J*=17.5 and 3.5 Hz, 1 H), 4.99–5.10 (m, 1 H), 4.29 (m, 1 H), 3.94 (m, 1 H), 3.86 (m, 1 H), 3.75 (m, 1 H). ¹³C NMR (125.8 MHz) (CD₃OD) δ: 165.6, 156.5, 149.1, 96.8, 95.3, 86.1, 86.0, 85.59, 85.57, 74.6, 74.4, 61.4, 56.2. MS calculated for C₉H₁₁FIN₃O₄, [MH⁺] 372 (calcd.), 372 (found).

4.1.5. Mixture of 1-(2-Deoxy-2-fluoro-β-D-arabinofuranosyl)-5-methoxycarbonylcytosine (5a) and 1-(2-deoxy-2-fluoro-α-D-arabinofuranosyl)-5-methoxycarbonylcytosine (5b)—

To a round bottom flask containing the mixture of **4a** and **4b** (3:1 ratio, 420 mg, 1.13 mmol) was added methanol (30 mL), (PhCN)₂PdCl₂ (44 mg, 0.1 equiv.), Et₃N (0.7 mL) and the mixture was stirred overnight at 50 °C on an oil bath under a CO atmosphere. After cooled to rt, silica gel (5 g) was added and the volatile was removed under reduced pressure. The residue was purified by silica gel chromatography, eluting with 1–4% MeOH in CH₂Cl₂, to give a mixture of **5a** and **5b** in a 3:1 ratio (238 mg, 88% combined yield) as white foam. ¹H NMR (500 MHz) (CD₃OD) δ: 8.79 (br. 0.25 H), 8.56 (br. 0.75 H), 6.22 (dd, *J*=16.5 and 3.5 Hz, 0.25 H), 6.05 (d, *J*=14.5 Hz, 0.75 H), 5.05–5.16 (m, 1 H), 4.31–4.50 (m, 1 H), 3.66–4.03 (m, 6 H). HRMS calculated for C₁₁H₁₄FN₃O₆, [MH⁺] 304.0945 (calcd.), 304.0939 (found).

4.1.6. 1-(2-Deoxy-2-fluoro-β-D-arabinofuranosyl)-5-methoxycarbonylcytosine (5a)—

Pure **5a** was prepared from pure **4a** by the same procedure as above. ¹H NMR (500 MHz) (DMSO-*d*₆) δ: 8.62 (s, 1 H), 8.13 (br.; 1 H), 7.73 (br.; 1 H), 6.11 (dd, *J*=16.5 and 4.0 Hz, 1 H), 5.91 (d, *J*=5.0 Hz), 5.18 (t, *J*=5.5 Hz, 1 H), 5.11 (dd, *J*=4.0, 2.5 Hz, 0.5 H), 5.1 (dd, *J*=4.0 and 3.0 Hz, 0.5 H), 4.22 (m, 1 H), 3.86 (m, 1 H), 3.76 (s, 3 H), 3.62 (m, 2 H). 166.2, 164.4, 154.2, 150.1, 97.3, 96.2, 95.8, 85.9, 85.8, 74.4, 74.2, 61.3, 53.4. HRMS calculated for C₁₁H₁₅FN₃O₆, [MH⁺] 304.0945 (calcd.), 304.0961 (found).

4.1.7. The mixture of N⁴-Acetyl-1-(2-deoxy-2-fluoro-β-D-arabinofuranosyl)-5-methoxycarbonylcytosine (6a) and N⁴-acetyl-1-(2-deoxy-2-fluoro-α-D-arabinofuranosyl)-5-methoxycarbonylcytosine (6b)—

To a round bottom flask containing the mixture of **5a** and **5b** (3:1, 220 mg, 0.73 mmol) was added DMF (10 mL) and Ac₂O (0.2 mL) and the mixture was stirred at rt for 2 days. After removal of the volatile in high vacuum, the residue was purified by silica gel chromatography, eluting with 1–8% MeOH in CH₂Cl₂ to give a mixture of **6a** and **6b** (3:1, 228 mg, 91% combined yield). ¹H NMR (500 MHz) (CD₃OD) δ: 9.05 (s, 0.75 H), 8.78 (s, 0.25 H), 6.26 (m, 0.75 H), 6.23 (m, 0.25 H), 5.08–5.15 (m, 1 H), 4.34–4.38 (m, 1 H), 4.07 (m, 0.75 H), 4.03 (m, 0.25 H), 3.91 (s, 2.25 H), 3.87 (s, 0.75 H), 3.77–3.84 (m, 2 H), 2.56 (s, 3 H). ¹³C NMR (125.8 MHz) (CD₃OD) δ: 172.2, 166.2, 160.7, 153.7, 151.1, 99.0, 97.2, 95.6, 86.4, 86.3, 85.8, 73.8, 73.6, 60.9, 53.9, 27.5. HRMS calculated for C₁₃H₁₆FN₃O₇, [MH⁺] 346.1051 (calcd.), 346.1066 (found).

4.1.8. N⁴-Acetyl-1-(2-deoxy-2-fluoro-β-D-arabinofuranosyl)-5-methoxycarbonylcytosine (6a)—

Using pure **5a** as starting material, **6a** was obtained by following the same procedure as white solid. ¹H NMR (500 MHz) (DMSO-*d*₆) δ: 10.70 (br., 1 H), 8.91 (s, 1 H), 6.15 (dd, *J*=14.0 and 4.0 Hz, 1 H), 5.97 (br., 1 H), 5.29 (br., 1 H), 5.24 (m, 1 H), 4.25 (dd, *J*=16.5 and 2.5 Hz, 1 H), 3.94 (m, 1 H), 3.78 (s, 3 H), 3.68 (m, 1 H), 3.32 (s, 3 H). ¹³C NMR (125.8 MHz) (DMSO-*d*₆) δ: 172.2, 166.2, 160.7, 153.7, 151.1, 99.0, 97.2, 95.6, 86.4, 86.3, 85.8, 73.8, 73.6, 60.9, 53.9, 27.5. HRMS calculated for C₁₃H₁₆FN₃O₇, [MH⁺] 346.1051 (calcd.), 346.1045 (found).

4.1.9. N⁴-Acetyl-1-(2-deoxy-2-fluoro-5-O-DMTr-β-D-arabinofuranosyl)-5-methoxycarbonylcytosine (7a) and N⁴-Acetyl-1-(2-deoxy-2-fluoro-5-O-DMTr-α-D-arabinofuranosyl)-5-methoxycarbonylcytosine (7b)—To a pyridine (8 mL)

solution of the mixture of **6a** and **6b** (3:1 ratio, 180 mg, 0.52 mmol) was added 4,4'-dimethoxytrityl chloride (212 mg, 0.63 mmol, 1.2 equiv.) and the mixture was stirred overnight at rt. MeOH (0.5 mL) was added to quench the reaction. After 15 min stirring, the reaction mixture was diluted with CH₂Cl₂, washed with 5% aqueous NaHCO₃, and brine. The organic layer was dried over anhydrous Na₂SO₄, filtered, and concentrated. The residue was purified by silica gel chromatography eluting with 0–3% MeOH in CH₂Cl₂ containing 0.2% Et₃N to give 213 mg **7a** (63%) and 71 mg **7b** (21%). **7a**, ¹H NMR (500 MHz) (CD₃CN) δ: 10.74 (br., 1 H), 8.82 (s, 1 H), 7.51 (m, 2 H), 7.26–7.39 (m, 7 H), 6.89 (m, 4 H), 6.18 (dd, *J*=18.0 and 3.5 Hz, 1 H), 5.11–5.23 (m, 1 H), 4.41–4.45 (m, 1 H), 4.14 (m, 1 H), 4.04 (m, 1 H), 3.78 (s, 6 H), 3.43 (dd, *J*=10.5 and 3.5 Hz, 1 H), 3.34 (s, 3 H), 3.28 (m, 1H), 2.51 (s, 3 H). ¹³C NMR (125.8 MHz) (CD₃CN) δ: 167.6, 161.9, 156.5, 155.4, 149.3, 146.5, 141.6, 132.54, 132.50, 126.6, 124.6, 124.5, 123.6, 109.7, 93.1, 92.5, 91.0, 82.9, 82.8, 82.7, 80.80, 80.77, 71.3, 71.1, 58.8, 51.6, 48.6, 22.8. MS calculated for C₃₄H₃₄FN₃O₉, [MH⁺] 648.2357 (calcd.), 648.2355 (found). With **7b**, ¹H NMR (500 MHz) (CD₃CN) δ: 10.72 (br., 1 H), 8.65 (s, 1 H), 7.50 (m, 2 H), 7.24–7.38 (m, 7 H), 6.89 (m, 4 H), 5.98 (d, *J*=15.0 Hz, 1 H), 5.12 (d, *J*=48.5 Hz, 1 H), 4.66 (m, 1 H), 4.35 (d, *J*=14.5 Hz, 1 H), 4.04 (m, 1 H), 3.88 (s, 3 H), 3.78 (s, 3 H), 3.31 (m, 1 H), 3.19 (m, 1H), 2.47 (s, 3 H). ¹³C NMR (125.8 MHz) (CD₃CN) δ: 167.5, 162.0, 156.6, 155.4, 150.0, 146.4, 146.3, 141.6, 132.5, 132.4, 126.6, 124.6, 124.5, 123.6, 109.8, 96.3, 94.8, 92.8, 89.2, 88.9, 85.7, 82.9, 71.0, 70.8, 60.0, 59.9, 51.6, 49.0, 22.7. MS calculated for C₃₄H₃₄FN₃O₉, [MH⁺] 648.2357 (calcd.), 648.2352 (found).

4.1.10. N⁴-Acetyl-1-(2-deoxy-2-fluoro-5-O-DMTr-β-D-arabinofuranosyl)-5-methoxycarbonylcytosine-3-O-(2-cyanoethyl)-N,N-diisopropyl phosphoramidite (8a)—To a solution of N⁴-acetyl-1-(2-deoxy-2-fluoro-5-O-DMTr-β-D-

arabinofuranosyl)-5-methoxycarbonylcytosine (**7a**) (100 mg, 155 μmol) and *N,N*-diisopropylethyl amine (108 μL, 620 μmol) in CH₂Cl₂ (4 mL) was added 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (52 μL, 234 μmol, 1.5 equiv.). After stirring at room temperature for 1 h, the solution was diluted with CH₂Cl₂, washed with 5% aqueous NaHCO₃, and brine. The organic layer was dried over anhydrous Na₂SO₄, filtered, and concentrated under reduced pressure. The residue was purified silica gel column chromatography with 4–10% MeOH in CH₂Cl₂ containing 0.2% Et₃N to afford phosphoramidite **8a** (112 mg, 85%) as a white foam. ¹H NMR (500 MHz) (CD₃CN) δ: 10.74 (br., 1 H), 8.871 (s, 0.5 H), 8.870 (s, 0.5 H), 7.50 (m, 2 H), 7.26–7.39 (m, 7 H), 6.87–6.91 (m, 4 H), 6.24 (d, *J*=3.0 Hz, 0.5 H), 6.20 (d, *J*=3.5 Hz, 0.5 H), 5.26–5.41 (m, 1 H), 4.58–4.72 (m, 1 H), 4.21–4.28 (m, 1 H), 3.78 (s, 3 H), 3.76 (s, 3 H), 3.50–3.65 (m, 4 H), 3.27 (s, 1.5 H), 3.32 (s, 1.5 H), 2.65 (t, *J*=6.0 Hz, 1 H), 2.52 (t, *J*=6.0 Hz, 1 H), 2.51 (s, 1.5 H), 2.50 (s, 1.5 H), 1.29 (m, 2H), 1.05–1.19 (m, 12 H). ¹³C NMR (125.8 MHz) (CD₃CN) δ: 167.6, 161.9, 156.5, 155.4, 141.6, 141.5, 132.44, 132.39, 132.3, 132.2, 126.65, 126.64, 126.62, 124.65, 124.57, 124.53, 123.61, 123.58, 109.77, 109.75, 109.73, 93.2, 92.12, 92.09, 91.8, 90.6, 90.3, 82.89, 82.86, 82.7, 82.6, 82.5, 82.3, 80.0, 79.8, 73.4, 73.3, 73.2, 73.0, 72.6, 72.4, 72.3, 72.2, 58.4, 58.1, 55.5, 55.43, 55.39, 55.3, 51.60, 51.58, 48.6, 48.5, 39.9, 39.85, 39.80, 39.76, 22.8, 20.58, 20.56, 20.52, 20.50, 20.46, 20.44, 20.40, 20.38, 18.45, 17.00, 16.62, 16.56, 16.50. ³¹P NMR (202.4 MHz) (CD₃CN) δ: 150.7, 150.2. HRMS calculated for C₄₃H₅₁FN₅O₁₀P, [MH⁺] 848.3436 (calcd.), 848.3430 (found).

4.1.11. N⁴-Acetyl-1-(2-deoxy-2-fluoro-5-O-DMTr-α-D-arabinofuranosyl)-5-methoxycarbonylcytosine-3-O-(2-cyanoethyl)-N,N-diisopropyl phosphoramidite (8b)—Using **7b** as starting material, phosphoramidite **8b** was prepared

in the similar way as **8a** in 85% yield as white foam. ^1H NMR (500 MHz) (CD_3CN) δ : 10.74 (br., 0.5 H), 10.68 (br., 0.5 H), 8.60 (s, 0.5 H), 8.57 (s, 0.5 H), 7.50 (m, 2 H), 7.26–7.38 (m, 7 H), 6.89–6.92 (m, 4 H), 6.01 (d, $J=12.5$ Hz, 0.5 H), 5.99 (d, $J=14.0$ Hz, 0.5 H), 5.33 (d, $J=48.0$ Hz, 0.5 H), 5.20 (d, $J=48.0$ Hz, 0.5 H), 4.81 (t, 0.5 H), 4.67 (t, 0.5 H), 4.582–4.63 (m, 1 H), 4.27 (m, 1 H), 3.90 (s, 1.5 H), 3.88 (s, 3 H), 3.79 (s, 6 H), 3.56–3.77 (m, 2 H), 3.36–3.48 (m, 2 H), 3.23 (m, 1 H), 2.86 (t, 1 H), 2.57 (m, 2 H), 2.51 (m, 3 H), 0.98–1.30 (m, 12 H). ^{13}C NMR (125.8 MHz) (CD_3CN) δ : 171.84, 171.82, 166.4, 166.35, 160.92, 160.90, 159.8, 154.02, 153.94, 150.2, 150.1, 145.9, 136.8, 136.75, 136.70, 136.6, 130.99, 130.96, 129.0, 128.94, 128.91, 128.90, 128.0, 119.24, 119.19, 114.1, 99.9, 99.1, 98.4, 97.6, 97.3, 97.1, 93.4, 93.1, 89.95, 89.90, 89.43, 89.41, 87.41, 87.38, 77.3, 77.2, 77.1, 77.0, 76.9, 64.2, 62.8, 55.93, 55.92, 53.5, 53.4, 44.2, 44.11, 44.08, 44.01, 27.1, 27.0, 24.82, 24.75, 24.71, 24.66, 24.58, 22.79, 22.77, 21.36, 21.34, 20.70, 20.64. ^{31}P NMR (202.4 MHz) (CD_3CN) δ : 151.2, 150.7. HRMS calculated for $\text{C}_{43}\text{H}_{51}\text{FN}_5\text{O}_{10}\text{P}$, $[\text{MH}^+]$ 848.3436 (calcd.), 848.3431 (found).

4.2. DNA synthesis, deprotection, purification and characterization

Oligonucleotide synthesis was performed on an Expedite Nucleic Acid Synthesis System using standard DNA synthesis conditions (scale: 1 μM) except that the modified modifications **III** and **V** used double coupling to ensure good yields. Phosphoramidites for dA, dC, dG, dT and CPG carriers were obtained from *Glen Research*. The terminal DMTr protecting group was removed from the oligonucleotides by using the DMTr off mode. Then the resin-containing oligos were transferred to a 2 mL vial. In general, two mild conditions were used for deprotection and cleavage of the **ODNs** from the resin with the oligos containing 5-fC and 5-hmC modifications. The resin was then treated with 1 mL 0.1 M K_2CO_3 in $\text{MeOH}/\text{H}_2\text{O}$ (v/v, 1:1) overnight at 40 °C. After deprotection, filtration to get rid of the resin and the resin was rinsed with 0.5 mL the mixture of acetonitrile and water (v/v, 1:1). AcOH was added to the filtrate to pH neutral and then was concentrated by speedvac. The dried **ODN** samples were then dissolved in deionized water (1mL) and were purified native PAGE with a 20% polyacrylamide gel. The purified oligonucleotides were concentrated by speedvac and further characterized by Maldi-TOF ms.

4.3. hTDG^{cat} glycosylase activity test

The glycosylase reactions were performed with 10 nM ^{32}P DNA substrates, 100 nM hTDG^{cat} at 37 °C for 30 min in 25 mM HEPES pH=7.4, 0.5 mM EDTA, 0.5 mg/mL BSA, 0.5 mM DTT. Reactions were quenched with 0.1 M NaOH and 0.01 M EDTA (v/v, 1:1), and incubated at 100 °C for 5min. The mixture was subject to denaturing Urea-PAGE with a 20% polyacrylamide gel. The gel was imaged and quantified using Fuji screen.

4.4. Electrophoretic mobility shift test

4 nM ^{32}P -Labeled DNA substrate and the enzyme were incubated at room temperature for 30 min in 10 mM HEPES pH = 8.0, 1 mM EDTA, 50 mM KCl, 0.05% Triton 100, 5% glycerol, 10 $\mu\text{g}/\text{mL}$ Salmon DNA. The mixture was subject to native PAGE with 8% polyacrylamide gel. The gel was imaged and quantified using Fuji screen.

Acknowledgments

This study was supported by National Institutes of Health (GM071440 to C.H.). We thank S. F. Reichard at the University of Chicago for editing the manuscript.

References

1. (a) Feng S, Jacobsen SE, Reik W. *Science*. 2010; 330:622. [PubMed: 21030646] (b) Ooi SK, O'Donnell AH, Bestor TH. *J. Cell Sci.* 2009; 122:2787. [PubMed: 19657014]
2. Bhutani N, Burns DM, Blau HM. *Cell*. 2011; 146:866. [PubMed: 21925312]
3. (a) Kriaucionis S, Heintz N. *Science*. 2009; 324:929. [PubMed: 19372393] (b) Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A. *Science*. 2009; 324:930. [PubMed: 19372391] (c) Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. *Nature*. 2010; 466:1129. [PubMed: 20639862]
4. (a) Münzel M, Globisch D, Brückl T, Wagner M, Welzmler V, Michalakakis S, Müller M, Biel M, Carell T. *Angew. Chem. Int. Ed.* 2010; 49:5375. (b) Szwagierczak A, Bultmann S, Schmidt CX, Spada F, Leonhardt H. *Nucleic Acids Res.* 2010; 38:181. (c) Song C-X, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen C-H, Zhang W, Jian X, Wang J, Zhang L, Looney TJ, Zhang B, Godley LA, Hicks LM, Lahn BT, Jin P, He C. *Nat. Biotechnol.* 2011; 29:68. [PubMed: 21151123] (d) Song C-X, Clark TA, Lu X-Y, Kislyuk A, Dai Q, Turner SW, He C, Korlach J. *Nature Methods*. 2011; 9:75. [PubMed: 22101853]
5. (a) Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y. *Science*. 2011; 333:1300. [PubMed: 21778364] (b) He YF, Li B-Z, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, Dai Q, Song CX, Zhang K, He C, Xu G-L. *Science*. 2011; 333:1303. [PubMed: 21817016]
6. Wu SC, Zhang Y. *Nat. Rev. Mol. Cell Biol.* 2010; 11:607. [PubMed: 20683471]
7. Radzicka A, Wolfenden R. *Science*. 1995; 267:90. [PubMed: 7809611]
8. Maiti A, Drohat AC. *J. Biol. Chem.* 2011; 286:35334. [PubMed: 21862836]
9. Krokan HE, Standal R, Slupphaug G. *Biochem. J.* 1997; 325:1. [PubMed: 9224623] (b) Stivers JT, Jiang YL. *Chem. Rev.* 2003; 103:2729. [PubMed: 12848584]
10. Zhang L, Lu X, Lu J, Liang H, Dai Q, Xu G, Luo C, Jiang H, He C. *Nature Chem. Biol.* 2012; 8:328. [PubMed: 22327402]
11. (a) Marquez VE, Tseng CK-H, Mitsuya H, Aoki S, Kelley JA, Ford H Jr, Roth JS, Broder S, Johns DG, Friscoll JS. *J. Med. Chem.* 1990; 33:978. [PubMed: 2106581] (b) York JL. *J. Org. Chem.* 1981; 46:2171. (c) Ikeda H, Fernandez R, Eilk A Jr, Marquez VE. *Nucleic Acids Res.* 1998; 26:2237. [PubMed: 9547286] (d) Sinnot ML. *Chem. Rev.* 1990; 90:1171.
12. Scharer OD, Verdine GL. *J. Am. Chem. Soc.* 1995; 117:10781.
13. Lee S, Bowman BR, Ueno Y, Wang S, Verdine GL. *J. Am. Chem. Soc.* 2008; 130:11570. [PubMed: 18686953]
14. Dai Q, He C. *Org. Lett.* 2011; 13:3446. [PubMed: 21648398]
15. (a) Takamatsu S, Maruyama T, Katayama S, Hirose N, Izawa K. *Tetrahedron Lett.* 2001; 42:2321. (b) Kawasaki AM, Casper MD, Freier SM, Lesnik EA, Zounes MC, Cummins LL, Gonzalez C, Cook PD. *J. Med. Chem.* 1993; 36:831. [PubMed: 8464037]
16. (a) Alauddin MM, Fissekis JD, Conti PS. *J. labeled Comp. Radiopharmaceuticals*. 2003; 46:805. (b) Hayakawa H, Takai F, Tanaka H, Miyasaka T, Yamaguchi K. *Chem. Pharm. Bull.* 1990; 38:1136. (c) Clark JL, Hollecker L, Mason JC, Stuyver LJ, Tharnish PM, Lostia S, McBrayer TR, Schinazi RF, Watanabe KA, Otto MJ, Furman PA, Stec WJ, Patterson SE, Pankiewicz K, W. J. *Med. Chem.* 2005; 48:5504. [PubMed: 16107149]
17. Codington JF, Doerr IL, Fox JJ. *J. Org. Chem.* 1964; 29:558.
18. (a) Guschlbauer W, Mengel R. *Angew. Chem. Int. Ed.* 1978; 7:525. (b) Mercer JR, Knaus EE, Wiebe LI. *J. Med. Chem.* 1987; 30:670. [PubMed: 2951522]
19. (a) Wilds CJ, Damha MJ. *Nucleic Acids Res.* 2000; 28:3628. (b) Ziemkowski P, Felczak K, Pozna ski J, Kulikowski T, Zieli ski Z, Cie la J, Rode W. *Biochem Biophys Res. Commun.* 2007; 362:37. [PubMed: 17692822] (c) Wu CY, Chan PC, Chang WT, Liu RS, Alauddin MM, Wang HE. *Appl. Radiat. Isot.* 2009; 67:1362. [PubMed: 19324560]
20. Watanabe KA, Reichman U, Hirota K, Lopez C, Fox JJ. *J. Med. Chem.* 1979; 22:21. [PubMed: 218006]
21. Howell HG, Brodfuehrer PR, Brundidge SP, Benigni DA, Sapino C Jr. *J. Org. Chem.* 1998; 53:85.

22. Watanabe KA, Su T-L, Klein RS, Chu CK, Matsuda A, Chun MW, Lopez C, Fox JJ. *J. Med. Chem.* 1983; 26:152. [PubMed: 6298422]
23. Chun MW. *Arch. Pharm. Res.* 1983; 6:79.
24. Barchi JJ Jr, Marquez VE, Driscoll JS, Ford H Jr, Mitsuya H, Shirasaka T, Aoki S, Kelley JA. *J. Med. Chem.* 1991; 34:1647. [PubMed: 2033591]
25. (a) Dai Q, Song C-X, Pan T, He C. *J. Org. Chem.* 2011; 76:4182–4188. [PubMed: 21462947] (b) Dai Q, He C. *Current Protocols in Nucleic Acid Chemistry*. 2011; Chapter 4(Unit 4.47)
26. Maiti A, Morgan MT, Pozharski E, Drohat AC. *Proc. Natl. Acad. Sci. U. S. A.* 2008; 105:8890. [PubMed: 18587051]
27. Huang H, Stivers JT, Greenberg MM. *J. Am. Chem. Soc.* 2009; 131:1344. [PubMed: 19173657]

Fig. 1.
The active pathway of 5mC demethylation

Fig 2.

Cleavage assay to detect the excision of dsDNA oligos by hTDG.

Lane 1–2, α20 and β20 cannot be excised by hTDG^{cat}. Lane 3–8, α20 and β20 can inhibit hTDG^{cat} activity. Lane 9, hmC20 can not inhibit hTDG^{cat} activity.

Fig. 3.

Electrophoretic mobility shift assay for hTDG^{cat} with **α20** and **β20** oligos.

Electrophoretic mobility shift experiments were performed with 4 nM ³²P-labeled DNA substrate and the enzyme concentrations shown. Fitting the data provides $K_d^{app} > 10 \mu\text{M}$ for **α20**, $K_d^{app} = 400 \pm 10 \text{ nM}$ for **β20**.

Scheme 1.

Synthesis of 1'-β-2'-deoxy-2'-fluoro-D-arabinofuranosyl-5-carboxylcytosine phosphoramidite (**8a**) and 1'-α-2'-deoxy-2'-fluoro-D-arabinofuranosyl-5-carboxylcytosine (**8b**)