Published in final edited form as:

Chem Sci. 2014 February 1; 5(2): 567–574. doi:10.1039/C3SC51849C.

A Chemical Probe Targets DNA 5-Formylcytosine Sites and Inhibits TDG Excision, Polymerases Bypass, and Gene Expression

Liang Xu^a, Ying-Chu Chen^b, Satoshi Nakajima^{c,d}, Jenny Chong^a, Lanfeng Wang^a, Li Lan^{c,d}, Chao Zhang^b, and Dong Wang^a

Chao Zhang: zhang.chao@usc.edu; Dong Wang: dongwang@ucsd.edu

^aSkaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, California, 92093, United States. Fax: (858) 822 1953

^bDepartment of Chemistry, University of Southern California, Los Angeles, California 90089, United States

^cDepartment of Microbiology and Molecular Genetics; University of Pittsburgh School of Medicine; Pittsburgh, Pennsylvania 15213, United States

^dUniversity of Pittsburgh Cancer Institute; University of Pittsburgh School of Medicine; 5117 Centre Avenue, Pittsburgh, Pennsylvania 15213, United States

Abstract

Dynamic regulation and faithful maintenance of proper DNA methylation patterns are essential for many cellular functions. 5-Formylcytosine (5fC), a newly discovered oxidized form of methylcytosine (mC) is involved in active DNA demethylation process. The latest progresses suggest exciting novel functional roles of this residue. Chemical tools are desired to further elucidate the functional roles of 5fC and to modulate dynamics of DNA demethylation and downstream biological processes. Here we designed and constructed a chemical probe, consisting of an aldehyde targeting group and an intercalation group. This molecule can selectively react with 5fC and subsequently inhibit base excision by thymine DNA glycosylase (TDG) and cause significant pausing for both DNA and RNA polymerase elongation. Further investigation using a GFP reporter system in living cells revealed that the ligand modification in 5fC sites at 5'-UTR of the GFP gene greatly inhibited the GFP expression level. These results altogether confirmed our successful design and established a new approach for generating functional ligands that target the formylcytosine sites and modulate 5fC-related biological processes.

Introduction

DNA methylation is an important epigenetic mechanism in controlling many cellular functions, such as gene expression, X-chromosome inactivation, maintenance of genomic

stability, cellular identity maintenance and differentiation, and embryo development.¹⁻⁷ High-resolution DNA methylation maps reveal that stem cells, progenitors, and differentiated cells exhibit distinct DNA methylation signatures.⁸⁻¹⁰ Therefore, it is crucial to maintain a proper pattern and dynamics of DNA methylation and demethylation processes during stem cell self-renewal and differentiation. Aberrant patterns of DNA methylation cause cancer and many human diseases.¹¹⁻¹³

Dynamic regulation and faithful maintenance of correct DNA methylation patterns are essential for many cellular functions and are the harmonious outcome of both methylation and demethylation processes.¹³⁻¹⁷ The mechanisms of DNA methylation process are well studied. A family of enzymes termed DNMTs is involved in DNA methylation. In contrast, the mechanisms of DNA demethylation were poorly understood and have become an exciting area to study recently. DNA demethylation can follow two pathways: passive and active demethylation pathways. Passive DNA demethylation simply dilutes global DNA methylation level by DNA replication. On the other hand, active DNA demethylation is much more complicated and involves more than one enzyme. Instead of directly removing the methyl group, 5-methylcytosine (5mC) is oxidized stepwise into 5-hydroxymethyl-,^[18,19] 5-formyl-^[20,21] and 5-carboxyl-cytosine^[20,22] by Ten-eleven translocation (TET) enzymes.^[23,24] These oxidized species of 5mC are proposed to be intermediates in DNA demethylation processes,^[16] removed via excision by thymine DNA glycosylase (TDG) and subsequently repaired through base excision repair (BER) mechanism (Figure 1).^[22,25,26] Intriguingly, these oxidized cytosines are at very high levels in embryonic stem cells, neuron progenitor cells and adult brain tissue, but are at relatively low levels in other tissues and cell lines.^[24] These characteristic patterns of oxidized cytosines may provide novel fingerprints of cellular identity in addition to DNA methylation pattern.^[27,28]

The roles of these oxidized species of 5mC in epigenetic and transcription regulation have been an area of intensive study recently.^[3,29] Our recent research^[30] on RNA polymerase II (pol II) transcription rates through these oxidized bases indicates a potential functional interplay between these DNA modifications and pol II transcription dynamics.^[31] Consistent with our *in vitro* studies, genome-wide mapping of 5fC distribution also reveals a high correlation of 5fC sites and high peaks of pol II binding sites (presumed pausing pol II) in the gene bodies.^[32] Intriguingly, a similar relationship is also found between bacteria RNAP transcription and DNA-J base (β -D-glucosyl-hydroxymethyluracil), another 5-position modification.^[31,33,34] A recent mass-spectrometry based proteomics study has identified quite a few protein complexes that selectively recognize these oxidized cytosines in a cell-type specific manner.^[35] Chromatin remodelling complex, DNA repair proteins, transcription and splicing machineries are among the top hits. These specialized protein complexes that specifically recognize 5fC or 5caC suggest novel functional roles of these oxidized cytosines.

To further understand the functional roles of oxidized methylcytosine residues in genomic DNA, it is important to design chemical tools that can target these modified cytosines and modulate dynamics of DNA demethylation and downstream biological processes. The DNA demethylation process may also provide novel attractive targets for therapeutic treatment.

Here we report the rational design of a chemical ligand, 9-(hydrazinylmethyl)acridine (HMA), which selectively and covalently target 5fC sites in DNA. We demonstrate that this adduct strongly inhibit TDG excision and greatly stall DNA pol I elongation and RNA pol II transcription *in vitro*. Furthermore, we reveal that the ligand-modified 5fC sites at 5'-UTR of a GFP gene abolish GFP expression in living cells. This rationally designed compound, which target 5fC sites and intervene 5fC related biological processes, paves the way for developing additional chemical probes that can manipulate DNA demethylation and related processes *in vivo* in the future.

Materials and methods

Materials

Oligonucleotides without 5-formylcytosine were purchased from IDT. The 29mer template with 5fC (denoted as F-C DNA template) was a gift from Dr. Chuan He (University of Chicago).³⁶ DNA oligos were purified by HPLC and PAGE. Chemical materials for ligand synthesis were purchased from TCI and Sigma. Recombinant *Methanobacterium thermoautotrophicum* Thymine DNA Glycosylase (TDG) was obtained from Trevigen. Klenow fragment (exo-) was bought from NEB (New England Biology). RNA pol II was purified as previously described.³⁷

Synthesis of HMA

To a solution of hydrazine monohydrate (2 mL, 41 mmol) in EtOH (1.5 mL), 9-bromoethylacridine (100 mg, 0.37 mmol, dissolved in 4 mL 1,4-dioxane) was added dropwise at 0 °C under N₂. The reaction mixture was stirred at room temperature for 1 h. Saturated aqueous NaHCO₃ was added to the reaction mixture at 0 °C and extracted with 10 mL dichloromethane three times. The organic phase was washed with brine, dried over Na₂SO₄(s) and concentrated in vacuo to give the desired product (73 mg, 89%). C₁₄H₁₃N₃; green solid; ¹H NMR (CDCl₃, 400 MHz) δ 8.37 (2 H, d, *J* = 8.8 Hz), 8.34 (2 H, d, *J* = 8.8 Hz), 7.74–7.83 (2 H, m), 7.55–7.62 (2 H, m), 4.91 (2 H, s), 3.29 (3 H, br); ¹³C NMR (CDCl₃, 100 MHz) δ 148.76 (2 ×), 140.00, 130.46 (2 ×), 129.84 (2 ×), 126.32 (2 ×), 125.50 (2 ×), 124.02 (2 ×), 50.66; ESI: [M+H]⁺, calculated: 224.12; found: 224.02 (Figure S1).

TLC analysis

5 mM nucleosides and 10 mM HMA were incubated at 25 °C for 1 h in 60 mM sodium phosphate buffer (pH 6.0). The products were analyzed on silica thin layer chromatography (TLC) plate (Whatman Ltd) with CH₂Cl₂ (NH₃ saturated)/CH₃CH₂OH (volume ratio: 15/85) as mobile phase. Untreated nucleosides were used as controls. The plate was then visualized at 245 nm and 365 nm UV light, respectively.

HMA ligand coupling test

Different concentrations of HMA (1, 10, 100 and 1000 μM) and 40 nM ³²P-labeled F-C DNA template were incubated at 37 °C for 1 h in the presence of 20 mM sodium phosphate buffer under different pH values (5, 6 and 7). The reacted mixtures were subsequently analyzed by 15% denatured PAGE. The control DNA strand (without 5fC site) was treated

using the same procedure except the incubation time was prolonged for 12 h. DNA with mC, hmC and caC were also tested in the presence of 1 mM HMA at pH=7.

For HMA ligand reaction with F-C DNA template in whole cell lysate, 100 nM F-C DNA template and 1 mM HMA were incubated with H295R cell lysate at 37 °C for 1, 3, and 6hr, respectively. As a control, C-template was also incubated with HMA under the same condition. For preparing cell lysate, H295R cells were harvested and resuspended in 1× PBS buffer (pH=7.4) and then lysed by sonication, followed by incubation on ice for 30 min. To prevent nuclease activity, 20 mM EDTA was added into the lysate.

Preparation of HMA-modified DNA template (L-C DNA)

1 μ M F-C DNA template and 1 mM HMA were incubated in pH=6 buffer at room temperature for 2 hr to achieve almost quantitative modification. The HMA-modified DNA template (L-C DNA) product was further purified by PAGE.

Fluorescent spectra

To detect the interaction of acridine with DNA duplex structure, fluorescent spectra of 1 μ M L-C DNA in 10 mM Tris-HCl were recorded by Fluorolog fluorescence spectrometer (Horiba Jobin Yvon, FL-1000). 1 μ M 20 nt or 15 nt non-template strand was then added into L-C DNA to form a duplex structure. The fluorescent spectra of these DNA duplex samples were subsequently detected after annealing. The excitation wavelength was 350 nm and the emission was recorded from 400 to 470 nm.

Thymine DNA glycosylase assay

2 μ M 32 P-labeled F-C DNA or L-C DNA was annealed with 2.4 μ M complementary DNA to form double-stranded DNA (dsDNA) in TDG reaction buffer (10 mM HEPES-KOH pH 7.4, 100 mM KCl, and 10 mM EDTA). For the TDG reaction, 100 nM dsDNA (either F-C or L-C DNA) and 2 U/ μ L TDG were incubated at 60 °C in the reaction buffer for 3 hr. To complete the cleavage of abasic site, equal volume of cleavage solution (100 mM NaOH and 3.3 mM EDTA) was added into the reaction solution. The mixture was then heated at 85 °C for 10 min. The cleavage product was analyzed by 15% denatured PAGE.

DNA pol I replication assay

0.4 μ M 32 P-labeled DNA primer was annealed with 0.8 μ M F-C or L-C DNA template in the presence or absence of 1 μ M non-template strands (20 mM Tris-HCl (pH 7.8), 10 mM MgCl_2). An aliquot of 0.5 U/ μ L DNA polymerase I (Klenow fragment exo-) was subsequently mixed with 40 nM of annealed scaffold in polymerase buffer (50 mM NaCl, 10 mM Tris-HCl (pH 7.9), 10 mM MgCl_2 and 1 mM DTT) at room temperature for 10 min before the addition of dT/C/GTP mixture (1 μ M each). The reaction was then quenched by stop buffer (20 mM EDTA and 8 M urea) after 5 min incubation. The elongation products were analyzed by 15% denatured PAGE.

As a control, 0.5 U/ μ L DNA polymerase I (Klenow fragment exo-) was mixed with the control DNA scaffold (no 5fC site) in the presence of different concentrations of HMA at room temperature for 10 min before the addition of dT/C/GTP (1 μ M each). The reaction

was quenched by addition of stop buffer (20 mM EDTA and 8 M urea) after 5 min incubation. The elongation products were analyzed by 15% denatured PAGE.

RNA pol II transcription elongation assay

0.5 μ M RNA primer, 1 μ M F-C or L-C template, and 1.5 μ M non-template strand NTS19 were annealed in RNA pol II elongation buffer (20 mM Tris-HCl (pH 7.5), 40 mM KCl, 5 mM MgCl₂). 200 nM RNA pol II were then mixed with 50 nM annealed scaffold at room temperature for 10 min before the addition of 2 mM T/U/GTP mixture. The transcription reaction was quenched at designated time period by stop buffer. Products were analyzed by 15% denatured PAGE.

Construction of GFP expression plasmid containing site-specific 5fC sites

The GFP expression plasmid containing site-specific 5fC sites (denoted as pFC plasmid) was constructed from pEGFP plasmid based on a similar construction method previously reported.^[38] Briefly, the pEGFP plasmid was double digested by Sal I and BamH I restriction enzymes to generate a linearized vector. Purified synthetic oligonucleotides containing 5fC sites were phosphorylated and annealed to form short dsDNA insertion fragments with 4-nt sticky ends that complement the vector (Sequence see Supplementary Figure S10). The insertion fragments were subsequently ligated into the linearized vector by T4 ligase at 16 °C for 18 hr. The circular pFC plasmids were purified from 0.8% agarose gel. The control plasmids with no 5fC sites in insertion fragments (denoted as pC plasmid) were prepared in the same manner. These purified plasmids were ready for transfection. For ligand treatment, the plasmids (pFC or pC plasmids) were incubated with 1 mM HMA ligand at 37 °C for 2 hr (pH=6) before transfection.

GFP expression assay

HeLa cell line was used in this study and cultured in DMEM with 10% FBS at 37 °C. 1×10^5 HeLa cells were seeded in 6 well dishes 24 hr. Cells were transfected with or without 100 ng GFP plasmid by Lipofectamine 2000 (Life technologies) and incubated for 48 – 72 hr. For imaging pictures, the Olympus FV1000 confocal microscopy system was employed (Cat. F10PRDMYR-1, Olympus) and FV1000 software was used for acquisition of images. For FACS (Fluorescence-activated cell sorting) analysis, cells were harvested by trypsinization, washed with PBS, and applied on the Accuri C6 (Becton Dickinson).

Results and discussion

Targeting 5fC by a functional group

We chose 5fC as the main target for modulating DNA demethylation and related biological processes for the following reasons. First, 5fC excision by TDG is a crucial step in generating an abasic site that is subsequently repaired by BER pathway during active DNA demethylation process. Indeed, RNAi knockdown of TDG activity increases levels of 5fC in the genome.^[32,39,40] Second, the unique reactivity of the aldehyde functional group in 5fC among natural nucleobases makes it an attractive target for selective chemical modification. This chemically reactive property has been recently utilized for selectively pulling down (and enrichment) in genome-wide 5fC mapping and *in vitro* 5fC detection.^[21,32,40–42] Thus,

we reasoned that a small molecule that selectively targets 5fC site and inhibits TDG excision would modulate downstream BER/DNA demethylation and 5fC related pathway. To achieve this goal, as a proof-of-principle, we designed a bifunctional small molecule, 9-(hydrazinylmethyl)acridine (HMA) (Figure 2a). HMA consists of two main functional groups: hydrazine moiety which can selectively react with the 5fC to generate a hydrazone adduct, and acridine moiety which can intercalate into base pairs of double-stranded DNA and interfere with TDG activity, DNA replication and transcription activity. We therefore termed this family of molecules “linked-intercalator”.

Reactivity and selectivity of HMA with 5-formylcytosine

To test the reactivity and selectivity of HMA with 5fC, HMA was incubated with individual nucleosides and subsequently analyzed by TLC. TLC results clearly indicated that only 5fC can react with HMA to form a covalent adduct, whereas no detectable product was observed from incubation of HMA with all canonical nucleosides (A, G, C, T) under the same conditions (Figure S2). To further test whether HMA can react selectively with 5fC in the context of DNA strand, we then incubated HMA with either 5fC-containing DNA strand (denoted as F-C DNA template), or unmodified DNA strand (denoted as C DNA template, as a control). The products were analyzed by denatured PAGE after 1 hr reaction at 37 °C. As clearly seen in Figure 2b, a new product band was emerging for DNA template containing a site-specific 5fC, whereas no new product band is observed for unmodified DNA template even with prolonged incubation (12 hr) (Figure 2c). In addition, we found that HMA does not react with DNA templates containing a site-specific 5mC, 5hmC and 5caC, respectively (Figure S3). Taken together, our results revealed that HMA can selectively react with the 5fC base in the DNA strand, but not the other bases. The identity of this single-HMA DNA adduct with 5fC is further confirmed by MALDI-TOF analysis (Figure S4). The efficiency of product formation is also pH dependent. We observed a full conversion at high concentration of HMA at pH 6 or 7, but not at pH 5. This is probably because the hydrazone hydrolysis is highly accelerated at lower pH.^[43] To further test whether HMA can react with F-C DNA in the content of cell lysate at the physiological condition, we incubated HMA ligand with F-C DNA in cell lysate and we found that a substantial portion of 5fC sites can be modified by the HMA ligand within a few hours incubation (almost 100% conversion of product after 6hr incubation) (Figure 2d). This rate, although slower than that in pure buffer, is compatible with the time range of most of drug/inhibitor treatment for *in vivo* studies. Collectively, we showed that HMA can selectively modify 5fC in DNA under physiological relevant condition.

To further mechanistically investigate the impacts of HMA-modifications of 5fC on 5fC-related processes, the HMA ligand-modified cytosine (denoted as L-C DNA template) was purified by PAGE gel and prepared for the following functional investigations.

Inhibition of TDG activity by HMA modification

To test whether the formation of HMA-5fC DNA adducts can interfere with TDG activity, we performed *in vitro* TDG excision assay with L-C and F-C template, respectively using purified *Methanobacterium thermoautotrophicum* TDG (as a model enzyme) (Figure 3a). As shown in Figure 3b, only a small portion of cleavage product (~8%) was observed from L-C

template, whereas a significantly higher percentage of cleavage product was observed from F-C template under the same condition (Figure 3b). Therefore, 5fC excision by TDG is greatly inhibited upon covalent attachment of HMA ligand. In addition, based on the crystal structures of human TDG–DNA complex,^[26,44–46] the binding site of human TDG for recognition of the flipped modified base is too small to accommodate HMA-ligand modified 5fC base (Figure S5). Since the TDG excision at abasic (AP) site is the first and crucial step to trigger subsequent BER pathway and DNA demethylation,^[47] strong inhibition of TDG activity by HMA ligand suggests that this ligand-modified cytosine could in principle prevent downstream DNA demethylation process. It would be interesting to further test this concept *in vivo* in the future.

HMA modification stalls DNA pol I replication

Uncleaved HMA-5fC adduct on DNA can also interfere with multiple DNA translocation events through 5fC sites. Hence, we investigated the impact of HMA-5fC adduct on DNA replication and transcription, using Klenow fragment (DNA polymerase I/exo- or Pol I/exo-) and RNA polymerase II as model enzymes.

To test the impact of HMA-DNA adduct on DNA replication, we first tested the primer extension by DNA pol I on single-stranded template DNA. Interestingly, we found that DNA pol I ran through this ligand-modified template smoothly and there was no detectable difference comparing with DNA pol I elongation on the F-C DNA template (Figure S6). This indicates that DNA pol I can efficiently bypass HMA-adduct even in the presence of bulky acridine group. Since acridine is a well-studied DNA intercalation group, which can intercalate into DNA duplex structure,^[48,49] we then test the primer extension by Klenow fragment on double-stranded template DNA. To this end, we systematically designed six different scaffolds with a variety of non-template strand (NTS) length ranging from 15nt to 20nt (Figure 4a). The covalent attachment of HMA restraining the acridine moiety interacts with long NTS, but not short NTS as revealed by fluorescence behaviors of HMA (Figure S7).

Intriguingly, we observed distinct DNA pol I replication results obtained from these six scaffolds. For scaffold with NTS-20, we observed a very strong pausing site at the 13th position for L-C template in comparison with F-C template. We observed the similar strong pausing bands for scaffolds of L-C template with NTS-19 and NTS-18. In sharp contrast, the pausing band completely disappeared for scaffold with short NTS-15 (Figure 4b). We found that the intensities of the pausing band at the 13th position are in an inverse correlation with the length of NTS among six scaffolds (Figures 4b and 4c). In addition, we noticed that this replication stall pattern is well-correlated with the capability of acridine intercalation in the downstream DNA duplex scaffolds. Notably, the acridine moiety of HMA covalently linked with 5fC can only efficiently intercalate into the nearby base pairs of DNA duplex in the scaffold with long NTS strands (NTS-18, NTS-19 or NTS-20). On the other hand, the intercalation of acridine group is greatly diminished for scaffolds containing short NTS strands (or no NTS) due to the restraint of linker length of HMA.

Our results clearly indicate that DNA pol I replication can be site-specifically paused by HMA-5fC adduct. Interestingly, this pausing position is not exactly located at the 5fC

position (the 11th position from 3' end) but two bases downstream (the 13th position from 3' end). This indicates that the bulky group of HMA *per se* does not serve as a translocation barrier or prevent nucleotide addition opposite to HMA-5fC adduct. Rather, the strong pausing at two-base downstream is likely due to the specific linked-intercalation of HMA group with downstream DNA duplex. Here we termed this interaction mode of HMA with DNA as “linked-intercalation”, since one part of HMA is covalently linked to DNA while the acridine group of HMA is intercalated with DNA. This linked-intercalation with the downstream DNA serves as a strong roadblock against DNA pol I. It is important to point out that the HMA ligand must be linked with the DNA strand (linked-intercalation mode) to be a strong roadblock against DNA polymerase, as we did not observe any replication stall with high concentrations of free HMA ligand (simple intercalation mode) (Figure S8). Collectively, DNA pol I can be stalled by HMA via its linked-intercalation mode and neither the bulky group of HMA *per se* (non-intercalation mode) nor free HMA ligand (simple intercalation) can block DNA pol I replication.

HMA modification stalls RNA pol II transcription

To investigate the effect of HMA-DNA adduct on transcription, we performed RNA polymerase II transcription assay using the scaffold shown in Figure 5a. With a short incubation, we observed a transient pausing at 5fC site and a downstream strong pausing on L-C template (Figure S9). These downstream strong pausing bands persist (indicating pol II stall) even after 2 hr incubation for the HMA ligand-modified L-C template (Figure 5b), whereas the majority of RNA pol II fully bypassed through F-C template (Figure 5b).

Analysis of elongation products through both F-C and L-C templates at different time points explicitly indicated that transcription elongation on L-C template is greatly stalled (Figure 5c). Therefore, RNA pol II transcription can be greatly interfered by treatment of HMA. Similar to DNA replication results, the positions of transcription stall are also located 2–4 base pairs downstream (the 13th–15th position from 3' end), suggesting that RNA pol II is stalled by HMA via its linked-intercalation mode. Very recently, genome-wide mapping of 5fC revealed high abundant levels of 5fC sites in gene bodies, promoters, and enhancers.^[32,39,40] Interestingly, 5fC is enriched at poised and active enhancers, but exhibits a preference to poised enhancers, in which 5fC is accumulated with increased binding of the transcriptional co-activator p300 in TDG-deficient cells.^[40] These latest progresses implied a great potential role of chemical ligand modification in modulating 5fC-related transcription process in future.

HMA modification inhibits the GFP expression in living cells

In order to understand the effects of HMA ligand modification on gene expression in living cells, we constructed a GFP reporter plasmid containing site-specific 5fC sites at 5'-UTR as shown in Figure 6a. Detailed sequence and construction procedure of the plasmid can be found *Material and Methods*, and Supplementary Information (Figure S10). This reporter system allows us to evaluate how HMA-ligand modifications would influence the gene expression in living cells by measuring GFP expression using microscopy and flow cytometry. If HMA-ligand modifications block polymerases bypass or TDG excision/BER

processes *in vivo*, we expect to observe a strong inhibition of GFP 5 gene expression in our reporter system.

As shown in Figure 6, we observed high levels of GFP expression for cells with either pC or pFC plasmid transfections (positive controls), indicating the success of GFP expression from our new constructed plasmids (in frame) and transfection procedure. In order to test the effects of ligand modification for GFP expression, we treated both the pC and pFC plasmid with HMA ligand. In this situation, pFC was covalently modified by the ligand in the 5fC sites, while pC could not. Strikingly, the GFP expression level for cells with pFC plasmid transfection was greatly diminished from 67% to around 1% upon HMA-ligand treatment, whereas HMA ligand treatment had no significant effect on GFP expression levels for cells with pC plasmid transfection (as control).

These *in vivo* results reveal that HMA ligand modification of 5fC sites strongly inhibit gene expression in living cells, which are consistent with our *in vitro* results. Moreover, the HMA-modified 5fC appears to be quite stable once it is formed and cannot be efficiently corrected by mammalian DNA repair enzymes (such as TDG and BER) *in vivo*. Taken together, as a proof of principle, our results initiate a promising and important step forward for the future investigations on dissecting the potential functional roles of 5fC in gene expression and cell differentiation.

Conclusions

In summary, we designed and developed a small-molecule ligand, HMA, which selectively target 5fC sites and modulates 5fC related biological process. We demonstrate that this linked-intercalation HMA ligand greatly inhibit the base excision by TDG, and block DNA and RNA polymerase elongation *in vitro*. We further reveal that HMA-ligand modification of 5fC sites in the 5'-UTR of a GFP gene greatly inhibits the gene expression in living cells. Taken together, our studies suggest that this ligand can be a useful chemical tool for *in vivo* studies on 5fC related biological processes, and our plasmid expression studies initiated a very promising step into the application *in vivo*. Our strategy of combining a 5fC targeting group with a bulky DNA intercalator to inhibit DNA demethylation process and interrupt gene expression established a novel framework for evolving more versatile ligands to modulate DNA demethylation pathways and pave the way to construct additional elaborate tools for the functional studies of 5-formylcytosine and other oxidized methylcytosines. Future studies will focus on the utilization of HMA or analogous ligands to manipulate processes mediated by endogenous 5fC in living cells.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

D.W. acknowledges the NIH (GM085136) and (GM102362), Kimmel Scholars award from the Sidney Kimmel Foundation for Cancer Research, start-up funds from Skaggs School of Pharmacy and Pharmaceutical Sciences, UCSD and Academic Senate Research Award from UCSD. C.Z. acknowledges start-up funds from the University of Southern California. L.L. acknowledges the Competitive Medical Research Fund (CMRF) of the University of

Pittsburgh Medical Center (UPMC) to SN and grants from the National Institutes of Health (NIH) AG045545-01 to LL and PLO. Support for the UPCI Imaging Facility and UPCI Cytometry Facility were provided by the Cancer Center Support Grant from the National Institutes of Health (P30 CA047904). We acknowledge Drs. Jia Yang and Marion Sewer for whole cell lysate. We thank Dr. Peter Geiduschek for critical readings of our manuscript.

References

1. Baylin SB, Jones PA. *Nat Rev Cancer*. 2011; 11:726. [PubMed: 21941284]
2. Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A. *Genes Dev*. 2009; 23:781. [PubMed: 19339683]
3. Dawson MA, Kouzarides T. *Cell*. 2012; 150:12. [PubMed: 22770212]
4. Bock C, Beerman I, Lien WH, Smith ZD, Gu H, Boyle P, Gnirke A, Fuchs E, Rossi DJ, Meissner A. *Mol Cell*. 2012; 47:633. [PubMed: 22841485]
5. Liang G, Zhang Y. *Cell Res*. 2013; 23:49. [PubMed: 23247625]
6. Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A, Regev A, Meissner A. *Nature*. 2012; 484:339. [PubMed: 22456710]
7. Smith ZD, Meissner A. *Nat Rev Genet*. 2013; 14:204. [PubMed: 23400093]
8. Cantone I, Fisher AG. *Nat Struct Mol Biol*. 2013; 20:282. [PubMed: 23463313]
9. Messerschmidt DM. *Epigenetics*. 2012; 7:969. [PubMed: 22869105]
10. Fu Y, He C. *Curr Opin Chem Biol*. 2012; 16:516. [PubMed: 23092881]
11. Jones PA, Baylin SB. *Cell*. 2007; 128:683. [PubMed: 17320506]
12. Feinberg P, Ohlsson R, Henikoff S. *Nat Rev Genet*. 2006; 7:21. [PubMed: 16369569]
13. Bergman Y, Cedar H. *Nat Struct Mol Biol*. 2013; 20:274. [PubMed: 23463312]
14. Ooi SK, Bestor TH. *Cell*. 2008; 133:1145. [PubMed: 18585349]
15. Song CX, He C. *Genome Biol*. 2012; 13:173. [PubMed: 23092522]
16. Bhutani N, Burns DM, Blau HM. *Cell*. 2011; 146:866. [PubMed: 21925312]
17. Lee ST, Xiao Y, Muench MO, Xiao J, Fomin ME, Wiencke JK, Zheng S, Dou X, de Smith A, Chokkalingam A, Buffler P, Ma X, Wiemels JL. *Nucleic Acids Res*. 2012; 40:11339. [PubMed: 23074194]
18. Branco MR, Ficiz G, Reik W. *Nat Rev Genet*. 2011; 13:7. [PubMed: 22083101]
19. Munzel M, Globisch D, Carell T. *Angew Chem Int Ed*. 2011; 50:6460.
20. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y. *Science*. 2011; 333:1300. [PubMed: 21778364]
21. Pfaffeneder T, Hackner B, Truss M, Munzel M, Muller M, Deiml CA, Hagemeyer C, Carell T. *Angew Chem Int Ed*. 2011; 50:7008.
22. He YF, Li BZ, 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 GL. *Science*. 2011; 333:1303. [PubMed: 21817016]
23. Wu H, Zhang Y. *Genes Dev*. 2011; 25:2436. [PubMed: 22156206]
24. Carell T, Brandmayr C, Hienzsch A, Muller M, Pearson D, Reiter V, Thoma I, Thumbs P, Wagner M. *Angew Chem Int Ed*. 2012; 51:7110.
25. Maiti, Drohat AC. *J Biol Chem*. 2011; 286:35334. [PubMed: 21862836]
26. Zhang L, Lu X, Lu J, Liang H, Dai Q, Xu GL, Luo C, Jiang H, He C. *Nat Chem Biol*. 2012; 8:328. [PubMed: 22327402]
27. Song CX, Yi C, He C. *Nat Biotechnol*. 2012; 30:1107. [PubMed: 23138310]
28. Kriukiene E, Liutkeviciute Z, Klimasauskas S. *Chem Soc Rev*. 2012; 41:6916. [PubMed: 22842880]
29. Jones PA. *Nat Rev Genet*. 2012; 13:484. [PubMed: 22641018]
30. Kellinger MW, Song CX, Chong J, Lu XY, He C, Wang D. *Nat Struct Mol Biol*. 2012; 19:831. [PubMed: 22820989]
31. Huang Y, Rao A. *Nat Struct Mol Biol*. 2012; 19:1061. [PubMed: 23132381]
32. Raiber EA, Beraldi D, Ficiz G, Burgess H, Branco MR, Murat P, Oxley D, Booth MJ, Reik W, Balasubramanian S. *Genome Biol*. 2012; 13:R69. [PubMed: 22902005]

33. van Luenen HG, Farris C, Jan S, Genest PA, Tripathi P, Velds A, Kerkhoven RM, Nieuwland M, Haydock A, Ramasamy G, Vainio S, Heidebrecht T, Perrakis A, Pagie L, van Steensel B, Myler PJ, Borst P. *Cell*. 2012; 150:909. [PubMed: 22939620]
34. Borst P, Sabatini R. *Annu Rev Microbiol*. 2008; 62:235. [PubMed: 18729733]
35. Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PW, Bauer C, Munzel M, Wagner M, Muller M, Khan F, Eberl HC, Mensinga A, Brinkman AB, Lephikov K, Muller U, Walter J, Boelens R, van Ingen H, Leonhardt H, Carell T, Vermeulen M. *Cell*. 2013; 152:1146. [PubMed: 23434322]
36. Dai Q, He C. *Org Lett*. 2011; 13:3446. [PubMed: 21648398]
37. Wang D, Bushnell DA, Westover KD, Kaplan CD, Kornberg RD. *Cell*. 2006; 127:941. [PubMed: 17129781]
38. Enou M, Ho TV, Long DT, Walter JC, Schärer OD. *Methods Mol Biol*. 2012; 920:203. [PubMed: 22941606]
39. Shen L, Wu H, Diep D, Yamaguchi S, D'Alessio AC, Fung HL, Zhang K, Zhang Y. *Cell*. 2013; 153:692. [PubMed: 23602152]
40. Song CX, Szulwach KE, Dai Q, Fu Y, Mao SQ, Lin L, Street C, Li Y, Poidevin M, Wu H, Gao J, Liu P, Li L, Xu GL, Jin P, He C. *Cell*. 2013; 153:678. [PubMed: 23602153]
41. Hu J, Xing X, Xu X, Wu F, Guo P, Yan S, Xu Z, Xu J, Weng X, Zhou X. *Chem Eur J*. 2013; 19:5836. [PubMed: 23512548]
42. Guo P, Yan S, Hu J, Xing X, Wang C, Xu X, Qiu X, Ma W, Lu C, Weng X, Zhou X. *Org Lett*. 2013; 15:3266. [PubMed: 23767758]
43. Dirksen, Dawson PE. *Bioconj Chem*. 2008; 19:2543–2548. [PubMed: 19053314]
44. Maiti A, Morgan MT, Pozharski E, Drohat AC. *Proc Natl Acad Sci USA*. 2008; 105:8890. [PubMed: 18587051]
45. Hashimoto H, Hong S, Bhagwat AS, Zhang X, Cheng X. *Nucleic Acids Res*. 2012; 40:10203. [PubMed: 22962365]
46. Hashimoto H, Zhang X, Cheng X. *DNA Repair*. 2013; 12:535. [PubMed: 23680598]
47. Dalton SR, Bellacosa A. *Epigenomics*. 2012; 4:459. [PubMed: 22920184]
48. Ferguson LR, Denny WA. *Mutat Res*. 2007; 623:14. [PubMed: 17498749]
49. Demeunynck M, Charmantray F, Martelli A. *Curr Pharm Des*. 2001; 7:1703. [PubMed: 11562307]

Figure 1.

Scheme of major DNA methylation and demethylation pathways. Chemical structures of 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC, with the reactive aldehyde group highlighted in red) and 5-carboxylcytosine (5caC) are depicted.

Figure 2. Selective reaction of HMA with 5-formylcytosine-containing DNA strand (F-C). (a) HMA ligand reacts with 5fC through formation of a hydrazone between the aldehyde and hydrazine groups (highlighted in red). (b) Gel analysis of HMA ligand-modified F-C DNA. F-C DNA were incubated with 1, 10, 100 or 1000 μM of HMA at 37 °C for 1 h before being subjected to PAGE gel analysis. The HMA-modified F-C DNA refers to L-C DNA. (c) HMA ligand does not react with unmodified DNA. Unmodified DNA (containing C instead of fC) were incubated with 1, 10, 100 or 1000 μM of HMA at 37 °C for 12 h. (d) HMA ligand-modified F-C DNA in whole cell lysate. Samples with F-C DNA or C-DNA were incubated with 1mM HMA at 37 °C for 0, 1, 3, and 6 hr, respectively.

Figure 3.

TDG cleavage is greatly inhibited by HMA ligand modification. (a) 5fC can be excised by TDG to generate an abasic (AP) site, which is subject to downstream BER repair pathway. Ligand-modified 5-fC can inhibit this process. (b) Cleavage results of F-C and L-C DNA after incubating with TDG. HMA modification inhibits DNA cleavage as revealed in gel analysis. Percentage cleavage is quantified and shown in column format on the right. The error bars represent standard deviations of three repeats.

Figure 4. DNA polymerase I stalls at L-C DNA template. (a) Scaffolds used in DNA replication tests. The numbers below the template refer to the corresponding nucleotide positions from the 3' end. (b) DNA polymerase I replication is stalled at the L-C scaffolds with long non-template strand (NTS). The numbers on the right indicate the corresponding nucleotide positions of the template DNA from the 3' end. Only dG/T/CTP was added for elongation. (c) Stalling percentage of DNA polymerase I replication at L-C (orange) and F-C (blue) DNA templates, respectively. The stalling percentage was calculated from the intensity of pausing band over the extended bands. The error bars represent standard deviations of three repeats.

Figure 5. RNA pol II stalls at L-C DNA template. (a) Scaffold used in pol II transcription assay. The numbers below the template refer to the corresponding nucleotide positions from the 3' end. (b) RNA pol II greatly stalled at ligand-modified template (L-C). The numbers on the right indicate the corresponding nucleotide positions of the template DNA from the 3' end. Only G/U/CTP was added for the RNA transcription assay. (c) Different bypass percentages through F-C and L-C strands at different time points.

Figure 6.

GFP expression is greatly inhibited by HMA modification *in vivo*. (a) Scheme of the pC or site-specific 5fC-containing pFC plasmid for GFP reporter assay. X refers to either cytosine residue or 5-fC residue at the template strand. (b) Fluorescence microscopy images for GFP expression. The first row are GFP fluorescent images in the dark field; the middle row are the DIC (differential interference contrast) images of cells, showing the total amounts of cells; the third row are merging of the upper two. (c) FACS analysis of GFP expression. The GFP-positive cells in the dashed triangle box are quantified. (d) Quantitative analysis of GFP expression level in a column format. The error bars represent standard deviations of four repeats of FACS measurements from two independent transfections. pC+L and pFC+L refer to the ligand-treated plasmids.