ACCOUNTS
of chemical research

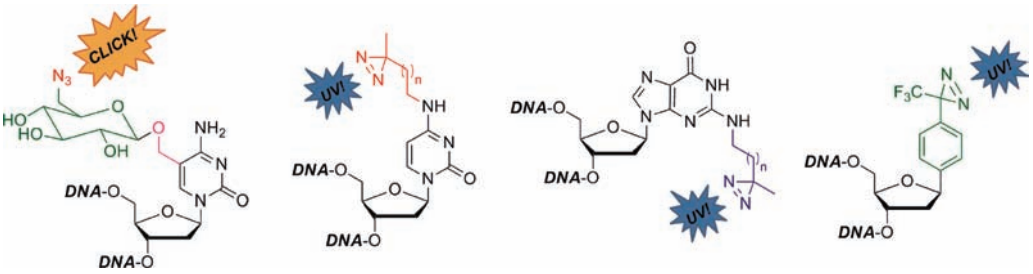
Bioorthogonal Labeling of 5-Hydroxymethylcytosine in Genomic DNA and Diazirine-Based DNA Photo-Cross-Linking Probes

CHUN-XIAO SONG AND CHUAN HE*

Department of Chemistry and Institute for Biophysical Dynamics, The University of Chicago, 929 East 57th Street, Chicago, Illinois 60637, United States

RECEIVED ON FEBRUARY 24, 2011

CONSPECTUS



DNA is not merely a combination of four genetic codes, namely A, T, C, and G. It also contains minor modifications that play crucial roles throughout biology. For example, the fifth DNA base, 5-methylcytosine (5-mC), which accounts for ~1% of all the nucleotides in mammalian genomic DNA, is a vital epigenetic mark. It impacts a broad range of biological functions, from development to cancer. Recently, an oxidized form of 5-methylcytosine, 5-hydroxymethylcytosine (5-hmC), was found to constitute the sixth base in the mammalian genome; it was believed to be another crucial epigenetic mark. Unfortunately, further study of this newly discovered DNA base modification has been hampered by inadequate detection and sequencing methods, because current techniques fail to differentiate 5-hmC from 5-mC. The immediate challenge, therefore, is to develop robust methods for ascertaining the positions of 5-hmC within the mammalian genome. In this Account, we describe our development of the first bioorthogonal, selective labeling of 5-hmC to specifically address this challenge.

We utilize β-glucosyltransferase (βGT) to transfer an azide-modified glucose onto 5-hmC in genomic DNA. The azide moiety enables further bioorthogonal click chemistry to install a biotin group, which allows for detection, affinity enrichment, and, most importantly, deep sequencing of the 5-hmC-containing DNA. With this highly effective and selective method, we revealed the first genome-wide distribution of 5-hmC in the mouse genome and began to shed further light on the biology of 5-hmC. The strategy lays the foundation for developing high-throughput, single-base-resolution sequencing methods for 5-hmC in mammalian genomes in the future.

DNA and RNA are not static inside cells. They interact with protein and other DNA and RNA in fundamental biological processes such as replication, transcription, translation, and DNA and RNA modification and repair. The ability to investigate these interactions will also be enhanced by developing and utilizing bioorthogonal probes. We have chosen the photoreactive diazirine photophore as a bioorthogonal moiety to develop nucleic acid probes. The small size and unique photo-cross-linking activity of diazirine enabled us to develop a series of novel cross-linking probes to streamline the study of protein–nucleic acid and nucleic acid–nucleic acid interactions. In the second half of this Account, we highlight a few examples of these probes.

Introduction

Bioorthogonal labeling methods are valuable tools in nucleic acid research to study their cellular behaviors, in particular recent bioorthogonal assays developed to measure DNA and RNA syntheses inside cells.[1,2] Our group has been utilizing bioorthogonal labeling methods relying on chemical or photochemical strategies to study DNA modification and interactions involving nucleic acids.[3–5]

Published on the Web 05/03/2011 www.pubs.acs.org/accounts
10.1021/ar2000502

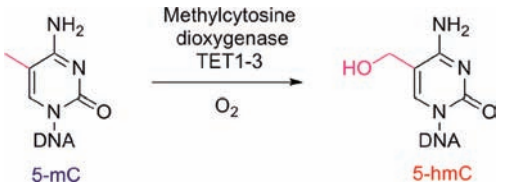


FIGURE 1. Oxidization of 5-mC to 5-hmC by TET1-3.

In eukaryotes, besides the genomic information encoded by the primary DNA sequence, there are other layers of heritable epigenetics information. One layer of such information is carried by 5-methylcytosine (5-mC), a DNA modification that constitutes ~1% of total nucleotides in mammalian genomic DNA (Figure 1). 5-mC is a vital epigenetic mark that impacts a broad range of biological functions, including gene expression, maintenance of genome integrity, parental imprinting, X-chromosome inactivation, regulation of development, aging, and cancer.[6−10] Recently, 5-hydroxymethylcytosine (5-hmC), an oxidized form of 5-mC, has been discovered in surprising abundance in mammalian genomic DNA of certain cell types (Figure 1).[11,12] Meanwhile, Tet proteins, a group of iron(II)/αKG-dependent dioxygenases similar to the AlkB family proteins and HIF prolyl-hydroxylases,[13,14] have been shown to utilize dioxygen to oxidize 5-mC to 5-hmC in mammalian genome, with Tet1 and Tet2 exhibiting roles in ES cell maintenance and normal myelopoiesis, respectively (Figure 1).[15−17] These findings suggested that 5-hmC might also be an important epigenetic mark.[18]

In order to reveal the biology of 5-hmC, the first step is to determine the genomic locations of 5-hmC. Two challenges exist to identify the locations of 5-hmC within genomic DNA: (1) distinguishing 5-hmC from 5-mC and (2) enriching 5-hmC-containing genomic DNA fragments. Current sequencing methods for 5-mC, including widely employed bisulfite sequencing and methylation-sensitive restriction digestion, failed to distinguish 5-hmC from 5-mC.[19−22] The bisulfite sequencing method, relying on the presence of a steric methyl group at the 5-position of cytosine in 5-mC to block deamination of the base as compared to normal cytosine, cannot differentiate a similarly steric 5-hydroxymethyl group on 5-hmC. All previous bisulfite sequencing data will need to be revisited, since all 5-hmC modifications were "mistaken" as 5-mC. Similarly, most restriction enzymes sensitive to the presence of 5-methyl in 5-mC are also inhibited by 5-hydroxymethyl in 5-hmC. Recently, Pacific Biosciences, a company that pioneers the use of DNA polymerase to perform single-molecule, real time (SMRT) sequencing, has applied the technology to distinguish among cytosine, 5-mC, and 5-hmC, but the method cannot enrich 5-hmC-containing DNA, and the signal differences, with the current SMRT protocol, have difficulties to achieve conclusive assignments.[23] Obviously, research in this field desperately needs a highly effective detection and sequencing method for 5-hmC. Described here is our recent work which developed a selective bioorthogonal 5-hmC labeling method to achieve detection, enrichment, and sequencing of 5-hmC in genomic DNA.

Cellular DNA and RNA are dynamic, constantly interacting with other biomolecules to sustain life processes. Photochemical cross-linking approaches have been proven to be effective in studying protein−DNA and nucleic acid−DNA interactions.[24−33] Among them, the bioorthogonal diazirine-based approach has been studied recently to probe the structure and function of biomolecule interactions.[24,34−45] Unlike other photoactive groups which may suffer from low reactivity or photodecomposition, diazirine-based probes have superior chemical stability and form highly reactive carbene species upon UV irradiation that cross-link efficiently with nearby biomolecules. The small size of the probe offers an additional advantage to minimize potential interference with biological functions. The second half of the Account will summarize our recent efforts in designing new diazirine-based DNA probes for improved protein−nucleic acid and nucleic acid−nucleic acid photo-cross-linking.

Bioorthogonal Labeling of 5-HmC

To address the challenge faced for 5-hmC detection, we borrowed a DNA modification system known to exist in nature and combined it with a bioorthogonal tagging technology.[3] 5-hmC was long known to be a component of T-even bacteriophage genomic DNA, in which it was further glucosylated to be β-glucosyl-5-hydroxymethyl-cytosine (5-gmC) by β-glucosyltransferase (βGT) using the cofactor uridine diphosphoglucose (UDP-Glc) (Figure 2a).[46−49] Our general strategy is to glucosylate 5-hmC with an azide-modified UDP-Glc. The azide-glucose installed on 5-hmC can be further functionalized using click chemistry to incorporate tags such as biotin.[50−54] The azide group is chosen because this functional group is not present inside cells; the click chemistry to label the azide group is completely bioorthogonal with no interference from biological samples. The installed biotin moiety can then serve as a detection/affinity tag for detection, enrichment, and sequencing purposes (Figure 2b).

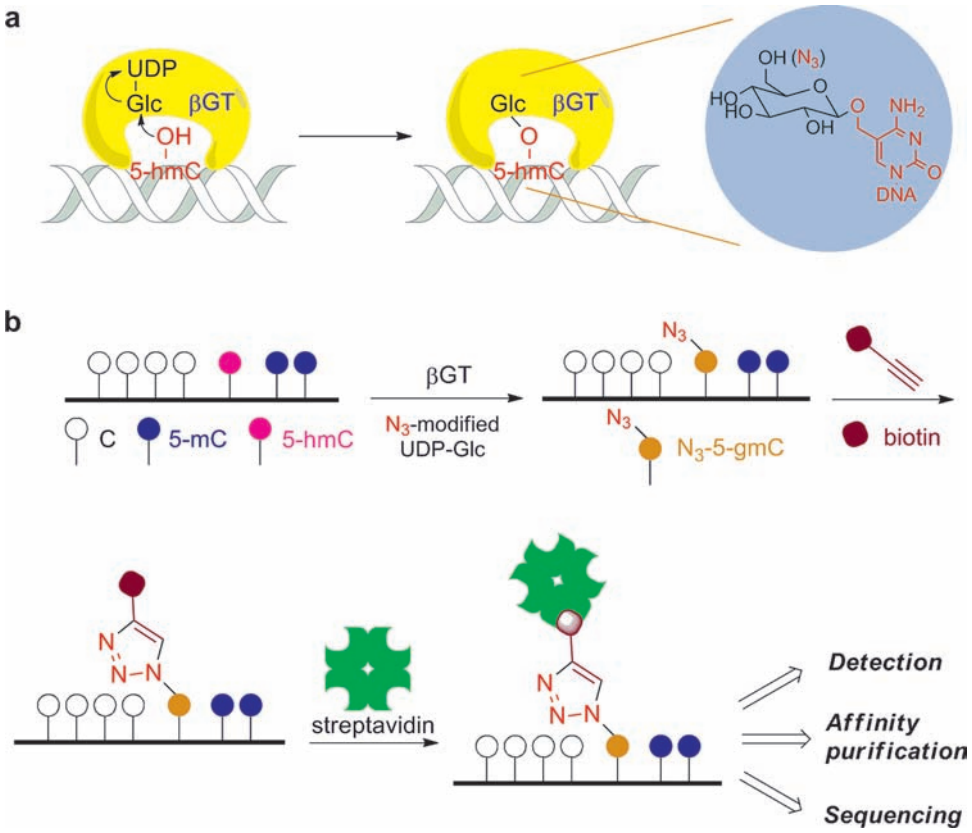


FIGURE 2. (a) The hydroxyl group of 5-hmC in duplex DNA can be glucosylated by βGT to form β-glucosyl-5-hydroxymethylcytosine (5-gmC) using UDP-Glc as a cofactor. (b) An azide group can be installed onto 5-hmC using chemically modified UDP-Glu (UDP-6-N_3-Glc), which in turn can be labeled with a biotin moiety using click chemistry for subsequent detection, affinity purification, and sequencing.

Selective Labeling of 5-hmC in Duplex DNA. To test if azide-glucose can be transferred by βGT, UDP-6-N_3-Glc was synthesized in six steps from a commercially available glucose derivative (Figure 3a).[3] We chose to attach the azide group at the 6-position of the glucose because of the synthetic accessibility of UDP-6-N_3-Glc as well as less steric hindrance for the subsequent click chemistry. To our delight, the glycosylation reaction of wild-type βGT with UDP-6-N_3-Glc went efficiently on synthetic 5-hmC-containing DNA (Figure 3a).[55,56] Quantitative conversion of 5-hmC to the glycosylated form (5-N_3-gmC) can be achieved using 1% βGT within minutes. Next, we used the dibenzocyclooctyne-PEG3-biotin (Figure 3a) as a copper-free click chemistry partner to install the biotin group.[57−59] The robust and reliable click chemistry gave high yield (~90%), and the resulting biotin-N_3-5-gmC was confirmed by MALDI-TOF, HPLC, and HRMS analysis of the corresponding HPLC hydrolysates (Figure 3b). Note that biotin-N_3-5-gmC have two isomers (50:50 ratio) due to the attachment of the biotin linker to the eight-member ring, which was confirmed by HPLC.

The similar behavior of 5-hmC and 5-mC in duplex DNA toward restriction enzyme digestion and polymerization renders it difficult to differentiate these two cytosine modifications.[20−22] With a bulky glucose group or a subsequent biotin attachment, 5-N_3-gmC or biotin-N_3-5-gmC exhibited distinguished properties compared to 5-hmC, 5-mC, and cytosine.[60] For example, we showed that, with the addition of streptavidin that binds biotin tightly, biotin-N_3-5-gmC could completely stall polymerization mainly at one base before, but also right at the modified position using Taq polymerase in a primer extension experiment (Figure 4). The two isomers of biotin-N_3-5-gmC mentioned above might explain such two-stage stops. Thus, our method has the potential to provide single-base resolution detection of 5-hmC in a genome.

Labeling and Detection of 5-hmC in Genomic DNA from Cell Lines and Animal Tissues. With a biotin group selectively attached to 5-hmC, it became handy to quantify the percentage of 5-hmC in a genome using dot-blot assay with avidin-horseradish peroxidase (HPR).[3] We have successfully applied the new labeling method on genomic DNA isolated from various sources and found significantly varied 5-hmC levels among different tissues and cell lines (Figure 5a).[3] Tissues tend to have relatively high 5-hmC

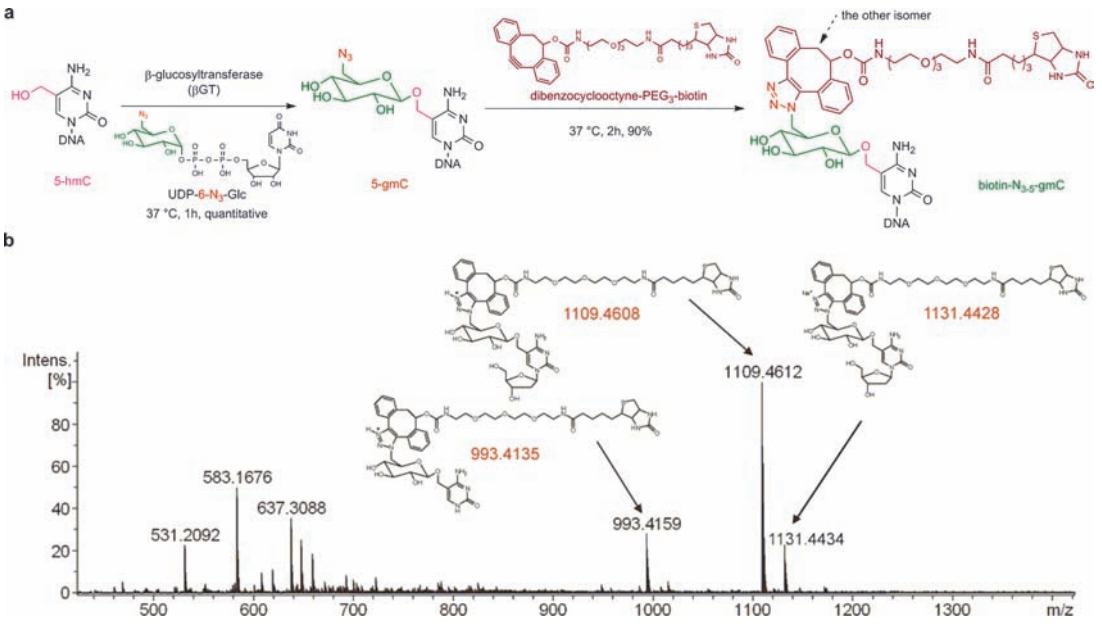


FIGURE 3. (a) Reactions of the βGT-catalyzed formation of N_3-5-gmC and the subsequent copper-free click chemistry to yield biotin-N_3-5-gmC in duplex DNA. (b) HRMS of biotin-N_3-5-gmC (structures are shown in the insets). Theoretical *m/z* values are shown in red; observed *m/z* values are shown in black.

levels (>0.1% of total nucleotide). In mouse cerebellum, we found a developmental stage-dependent increase of 5-hmC from 0.1% of total nucleotide in postnatal day 7 (P7) to 0.4% of total nucleotide in adult stage, suggesting that 5-hmC plays an important role in neuron development. Cell lines tend to have less 5-hmC compared to tissue samples. Mouse embryonic stem cells (mESC) and mouse adult neural stem cells (aNSC) were determined to have 5-hmC contents (~0.04−0.05% of total nucleotide) about one tenth of that for adult mouse cerebellum. Other cell lines, especially cancer cell lines like HeLa, have much less 5-hmC (~0.01%). We have further confirmed the existence of 5-hmC in HeLa cells by obtaining the HRMS of biotin-N_3-5-gmC from the digestion of labeled Hela genomic DNA, which is identical to the authentic standard from synthetic DNA (Figure 3b).[3] With over 40 cancer cells we have tested, the 5-hmC level is consistently low but with substantial variations (unpublished data). The vast dynamic range of 5-hmC level suggests interesting biology behind this modification.

Early studies failed to detect 5-hmC in HeLa cells due to low sensitivity of previous detection methods. Unlike antibodies which may have a sequence-bias to recognize 5-hmC in DNA,[16,61] a major advantage of our approach is the ability to selectively label all 5-hmC in the genome. Together with the amplification power of HRP, our method dramatically enhanced the sensitivity for 5-hmC detection in mammalian genomes. In fact, among the known detection methods for 5-hmC including radiolabeled TLC,[11] LC-MS,[62] and radiolabeled enzymatic quantification,[61] our method possesses the lowest detection limit for 5-hmC (Figure 5b), which is especially important for the study of samples containing low levels of 5-hmC such as cancer cell lines.

Pull-Down of 5-hmC-Containing DNA for Deep Sequencing. The biotin tag that has been selectively introduced to 5-hmC presents multiple options. Besides detection and quantification, the tag also allows high-affinity purification/enrichment of 5-hmC-containing DNA fragments. Genomic DNA with 5-hmC selectively labeled with biotin can be sheared into fragments containing several hundred base pairs. The 5-hmC-containing fragments can be enriched through specific binding of biotin−avidin/streptavidin. This interaction is highly selective and is one of the strongest noncovalent interactions known in nature,[63] which survives under harsh reaction conditions that typical antibody−antigen interactions will not tolerate. Coupled with the high specificity of the bioorthogonal covalent click chemistry, this method ensures robust and comprehensive capture of 5-hmC-containing DNA fragments. The enriched 5-hmC-containing DNA fragments are suitable for deep sequencing to reveal the genomic location/distribution of 5-hmC. Compared to the traditional noncovalent antibody-based immunoprecipitation, which may suffer from its density-dependent recognition/capture of 5-hmC, our approach provides considerable advantages.

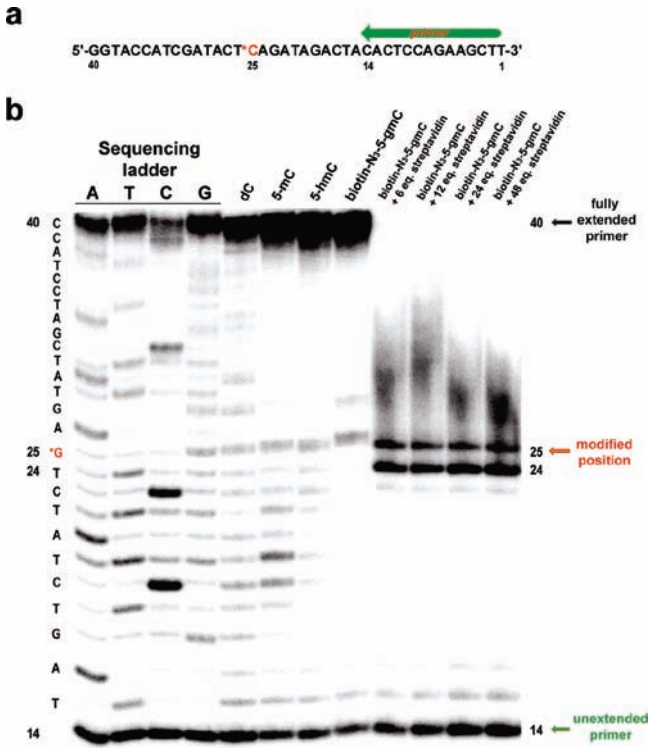
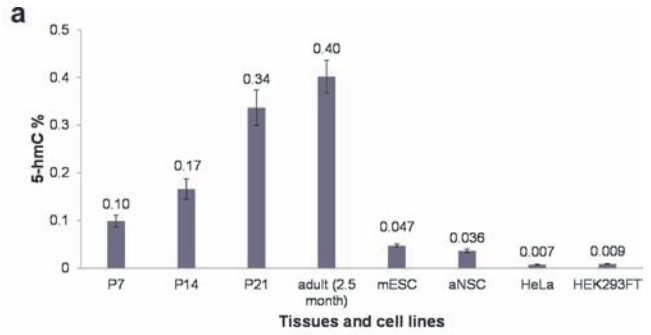


FIGURE 4. Streptavidin adduct of biotin-N_3-5-gmC hinders primer extension. (a) Sequence of 40-mer DNA used in primer extension containing the cytosine modification (*C). The green arrow represents the reverse PCR primer used for primer extension. (b) Primer extension assays for the 40-mer DNA containing different cytosine modifications, shown beside a Sanger sequencing ladder. Primer extension was not hindered with regular cytosine, 5-mC, 5-hmC, or biotin-N_3-5-gmC at the *C position. Primer extension was completely stalled when the biotin-N_3-5-gmC-containing DNA was treated with 6–48 equiv of streptavidin. The most significant stalling position was one base before the modification base with significant stalling also observed at the modified position (position 25, red arrow).



b

Comparison of current detection methods for 5-hmC

Method	Detection Limit[a]	Source
Radiolabled TLC	0.08%	ref. 11
LC-MS	HeLa not determined	ref. 62
Antibody	0.008%	MethylFlash, Epigentek
Radiolabled enzymatic quantification	0.025%[b]	ref. 61
Our method	0.004%	ref. 3

[a] 5-hmC of total nucleotide, unless noted otherwise; [b] 5-hmC of total cytosine.

FIGURE 5. (a) Quantification of 5-hmC in various tissues and cell lines, showing the percentage of 5-hmC in total nucleotides. (b) Comparison of the sensitivity of current detection and quantification methods for 5-hmC.

Indeed, we reported the first genome-wide 5-hmC distribution in mouse cerebellum using this method (Figure 6).[3] The superior enrichment ability of the method allows 5-hmC-containing DNA pieces to be affinity purified and sequenced to identify the specific genomic regions enriched for 5-hmC in this tissue. Enriched regions were nonrandomly distributed among chromosomes and clearly distinguishable from both input genomic DNA and control DNA group (labeled with regular glucose) (Figure 6). This information already provides us with extremely interesting biological information of 5-hmC. For example, through analyses of the deep sequencing results on genomic DNAs derived from mouse cerebellums at postnatal day 7 (P7) and adult stages, we showed that 5-hmC is enriched in the gene bodies that are highly expressed, suggesting that 5-hmC is associated with active or highly transcribed genes.[3] Furthermore, an age-dependent acquisition of 5-hmC in specific gene bodies, in particular those linked to neurodegenerative disorders as well as angiogenesis and hypoxia response, was observed.[3] We anticipate that this application of bioorthogonal labeling concept to 5-hmC will lead to simple and sensitive single-base resolution detection methods and high-throughput single-base sequencing methods of this new epigenetic mark in genomic DNA in the future.

Diazirine-Based DNA/RNA Photo-Cross-Linking Probes for Studying Biomolecule–Nucleic-Acid Interactions

Trapping and studying labile/transient interactions in biology is a challenge, but could be highly rewarding. While our group has been working on chemical disulfide cross-linking to study protein–DNA interactions,[64–68] we have also expanded to a different strategy of developing diazirine-based photo-cross-linking probes.[4,5] While the chemical disulfide cross-linking can provide considerable selectivity, it requires sufficient structure-based knowledge to install cross-linking sites. The method is also limited to the cysteine side chain, which carries a naturally abundant thiol group. Diazirine-based photo-cross-linking probes can readily cross-link to the nearby biomolecules without prior understanding of the interactions and does not require installation of particular

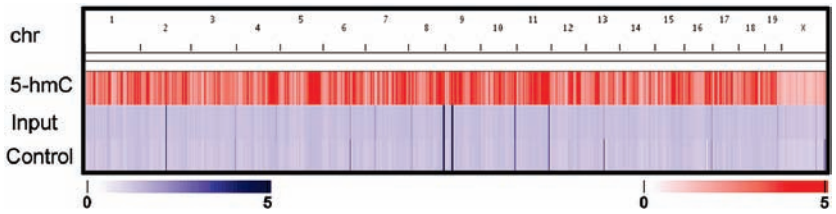


FIGURE 6. Genome-wide distribution of 5-hmC in female adult mouse cerebellum. Genome-wide heat map representation of 5-hmC reads. Heat map scale representing read density is indicated below. 5-hmC, affinity purified 5-hmC DNA; Input, genomic DNA; Control, control treated DNA with regular UDP-Glc.

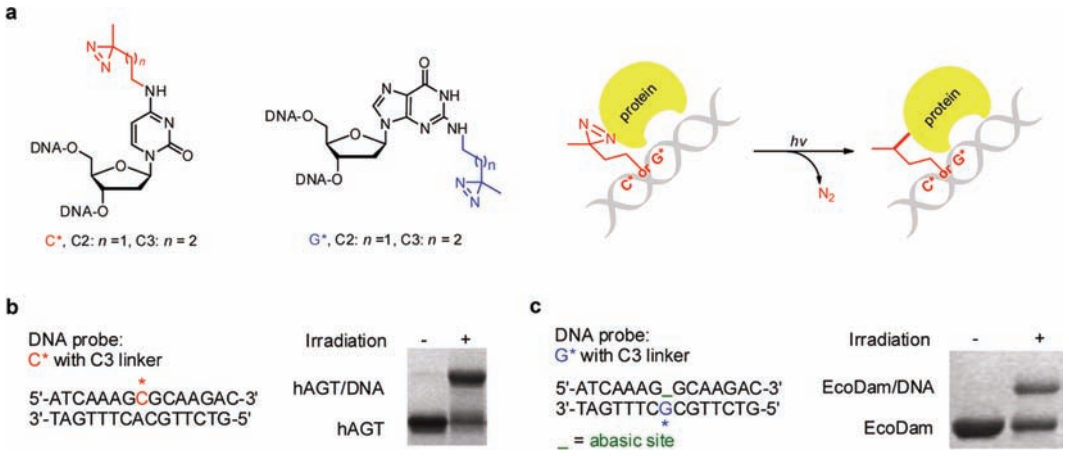


FIGURE 7. (a) Diarizine-based probes C* and G* with C2 ($n = 1$) or C3 ($n = 2$) linkers can cross-link to nearby protein upon UV irradiation. (b,c) SDS-PAGE analyses of photo-cross-linking reactions between diazirine probes and hAGT or EcoDam with the sequences of the probes showing on the side.

functional groups on the target biomolecules. Compared to other photoreactive moieties, diazirine derivatives process several benefits: (i) they have excellent chemical stability prior to photolysis and are therefore easy to handle; (ii) they photolyze rapidly at 350−360 nm, beyond the absorbance range of most biomolecules; (iii) upon UV radiation, they eliminate a molecule of N_2 to form a highly reactive carbene intermediate that can rapidly cross-link nonspecifically to nearby biomolecules through various functional groups, even inert aliphatic C−H bonds;[69] (iv) the cross-linked product is a carbon-based bond that is quite stable; (v) the relatively small size of the diazirine group reduces potential hindrance that otherwise may be introduced by other photoactive groups. Therefore, we initiated research to develop new diazirine-based nucleic acid probes to perform bioorthogonal cross-linking studies.

Diazirine-Based DNA Probes for Protein−DNA Interactions. Previously, aryltrifluoromethyldiazirine has been introduced to modified DNA bases for photo-cross-linking with proteins, but steric issues may occur due to the large phenyl group introduced.[36,39,70] Our plan was to replace the large aryltrifluoromethyldiazirine with a much smaller alkyl diazirine group to reduce potential steric hindrance (Figure 7a).[5] A convertible nucleoside method was employed to introduce the diazirine group by postsynthetic modification/deprotection, thus eliminating the trouble of preparing the phosphoramidite of the final diazirine modified nucleoside.[71−73] Namely, diazirine amines with C2 ($n = 1$) or C3 ($n = 2$) linkers were used to react with synthetic DNA incorporated with the commercially available O^4-triazolyl-dU-CE phosphoramidite or 2-F-dI-CE phosphoramidite at the modification site in order to generate the desired N^4-diazirine-modified cytosine (C*) for major groove cross-linking and N^2-diazirine-modified guanosine (G*) for minor groove cross-linking (Figure 7a). DNA or RNA probes bearing a small alkyle diazirine tether can be readily prepared using this strategy. Since the diazirine group in the two tethers point away from the DNA, these probes are ideal for protein−DNA cross-linking.

We have demonstrated that these probes could photo-cross-link efficiently to two DNA-binding proteins, human O^6-alkylguanine-DNA alkyltransferase (hAGT) and *E. coli* DNA adenine methyltransferase (EcoDam), with up to 75% cross-linking yields (Figure 7b,c).[5] No cross-linking product

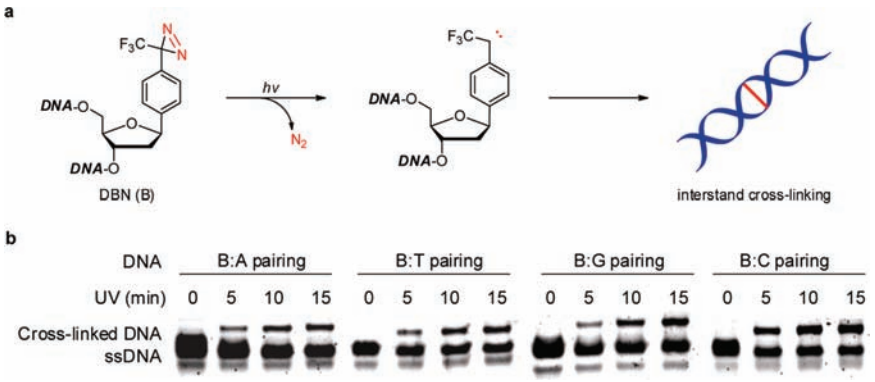


FIGURE 8. (a) Diazirine-based nucleoside analogue (DBN, B) can form a DNA interstrand cross-link upon UV irradiation. (b) Interstrand photo-cross-link for the DBN paired with A, T, G, and C, respectively. B represents the 15-mer ssDNA containing DBN; A, T, G, and C represent the complementary ssDNA strands.

formed without UV irritation, showing the excellent chemical stability of diazirine. After UV irritation, very good cross-linking yields (50−75%) could be achieved. Note that in the EcoDam case, the probe that contains a nonspecific sequence for EcoDam can be titrated away by excess of unmodified DNA. This result indicates that the diazirine probe can trap the nonspecific as well as specific interactions with the protein. With this method, we may trap specific and nonspecific interactions for future proteomic profiling of nucleic acid-binding proteins, and for potential structural characterization of protein-nucleic acid interactions.

Diazirine-Based DNA Probes for DNA−DNA Interactions. Previous work on labeling DNA with diazirine-based nucleoside analogues all resulted in extrahelical attachment of the diazirine group to the DNA helices.[5,35−44] We showed that direct link of an aryltrifluoromethyldiazirine moiety to the ribose replacing the nucleobase could generate an intrahelical attachment of diazirine to the DNA helices, which upon UV irradiation could cross-link to the opposite DNA strand (Figure 8a).[4] To this end, we designed and synthesized a diazirine-based nucleoside analogue (DBN) that can be incorporated into DNA by solid state oligonucleotide synthesis.

Oligonucleotides containing DBN (B) were synthesized and annealed to complementary oligonucleotides containing A, T, G, or C at the opposite position to give dsDNA_B:A, dsDNA_B:T, dsDNA_B:G, and dsDNA_B:C, respectively. The measured melting temperatures of these duplex DNAs showed only mild reduction of T_m (about the same as expected for a single base pairing mismatch) through introducing the DBN base analogue into duplex DNA, which suggests minimum steric hindrance to the normal dsDNA conformation. In the absence of light activation, no cross-linking product was detected. Upon UV irradiation, these four dsDNAs formed interstrand cross-linked products. The cross-linking efficiency varied with different bases opposite DBN, but the B:C pair gave the highest yield (Figure 8b).[4] This property, upon further optimization, could be used to design sequence-specific detection probes.

Conclusion and Outlook

While bioorthogonal chemistry has been implemented to reveal exciting biology of proteins, sugars, lipids, and metabolites, we show here examples of employing bioorthogonal chemistry to reveal new biology associated with nucleic acids. In the first half of this Account, we described a highly efficient and selective bioorthogonal method for labeling 5-hmC in genomic DNA. The key strategy is to transfer an azide-modified glucose to the 5-hydroxymethyl group of every 5-hmC in isolated genomic DNA using a naturally existing β-glucosyltransferase. The azide group allows for the subsequent installation of various functional tags using click chemistry for detection, enrichment, and sequencing of this potentially extremely important epigenetic mark in mammalian DNA. Using this approach, we revealed the first distribution map of 5-hmC in mammalian genomes. The covalent labeling strategy offers versatile tags and properties to be selectively installed onto 5-hmC. Future efforts will focus on developing high-throughput, single base-resolution sequencing method for 5-hmC with the strategy introduced here. In addition, in situ labeling methods may be developed using similar approaches to visualize the dynamics of 5-hmC on genomic DNA in living cells. The new technology presented here and to be developed in the future will arm biologists with essential tools to uncover new biology of 5-hmC in gene regulation and development.

We also introduced our recent efforts on designing more efficient diazirine-based photo-cross-linking probes for investigating protein−DNA and DNA−DNA interactions. Unlike

our 5-hmC labeling, photo-cross-linking can trap nonspecific interactions. Biochemical and biological validations are often required to establish physiological relevance of targets identified from photo-cross-linking. The N^4-diazirine-modified cytosine (C*) and N^2-diazirine-modified guanosine (G*) are ideal small probes for highly efficient cross-linking to trap specific and nonspecific protein–DNA interactions in the major groove and minor groove, respectively. Related probes may be used to profile DNA- or RNA-interacting proteins from cell extracts or even in living cells. For example, a DNA probe containing 5-hmC and photoactive C* or G* may be synthesized and employed to pull-down potential 5-hmC binding proteins. The DBN probe, on the other hand, is an excellent interstrand cross-linking probe for capturing nucleic acid–nucleic acid interactions. For example, we are exploring these probes to identify the genome-wide micro-RNA (miRNA) targets. Current methods using cross-linking immunoprecipitation sequencing (CLIP-Seq) have been used to identify protein-bound RNA species.[74,75] The diazirine-based nucleic acid probes can afford higher cross-linking efficiency. The DBN type probes may offer particular opportunity to trap nucleic acid–nucleic acid interaction partners in high efficiency. These probes may also be further developed to achieve sequence-specific cross-linking and thus may be used to identify the single-nucleotide polymorphism (SNP) of DNA for disease prognosis and diagnosis.

BIOGRAPHICAL INFORMATION

Chun-Xiao Song is a graduate student in the Department of Chemistry at The University of Chicago.

Chuan He received his Ph.D. from Massachusetts Institute of Technology, where he worked with Stephen J. Lippard. He did his postdoctoral work with Gregory L. Verdine at Harvard University. He is a professor of Chemistry and Institute for Biophysical Dynamics at The University of Chicago.

This work was partially supported by the National Institutes of Health (GM 071440 to C.H.). We thank Sarah Frank Reichard, M.A. for help with editing the manuscript.

FOOTNOTES

*To whom correspondence should be addressed. E-mail: chuanhe@uchicago.edu.

REFERENCES

1 Salic, A.; Mitchison, T. J. A chemical method for fast and sensitive detection of DNA synthesis in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 2415–2420.

2 Jao, C. Y.; Salic, A. Exploring RNA transcription and turnover in vivo by using click chemistry. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 15779–15784.

3 Song, C.-X.; Szulwach, K. E.; Fu, Y.; Dai, Q.; Yi, C.; Li, X.; Li, Y.; Chen, C. H.; Zhang, W.; Jian, X.; Wang, J.; Zhang, L.; Looney, T. J.; Zhang, B.; Godley, L. A.; Hicks, L. M.; Lahn, B. T.; Jin, P.; He, C. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* **2011**, *29*, 68–72. A provision patent application on 5-hmC labeling using glucose and modified glucose by β-GT was officially filed in April 2010.

4 Qiu, Z.; Lu, L.; Jian, X.; He, C. A diazirine-based nucleoside analogue for efficient DNA interstrand photocross-linking. *J. Am. Chem. Soc.* **2008**, *130*, 14398–14399.

5 Shigdel, U. K.; Zhang, J.; He, C. Diazirine-based DNA photo-cross-linking probes for the study of protein-DNA interactions. *Angew. Chem., Int. Ed.* **2008**, *47*, 90–93.

6 Gal-Yam, E. N.; Saito, Y.; Egger, G.; Jones, P. A. Cancer epigenetics: modifications, screening, and therapy. *Annu. Rev. Med.* **2008**, *59*, 267–280.

7 Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **2007**, *447*, 425–432.

8 Goll, M. G.; Bestor, T. H. Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **2005**, *74*, 481–514.

9 Weber, M.; Schubeler, D. Genomic patterns of DNA methylation: targets and function of an epigenetic mark. *Curr. Opin. Cell Biol.* **2007**, *19*, 273–280.

10 Klose, R. J.; Bird, A. P. Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* **2006**, *31*, 89–97.

11 Kriaucionis, S.; Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **2009**, *324*, 929–930.

12 Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **2009**, *324*, 930–935.

13 Yi, C.; Yang, C. G.; He, C. A non-heme iron-mediated chemical demethylation in DNA and RNA. *Acc. Chem. Res.* **2009**, *42*, 519–529.

14 Hausinger, R. P. FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 21–68.

15 Ito, S.; D'Alessio, A. C.; Taranova, O. V.; Hong, K.; Sowers, L. C.; Zhang, Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **2010**, *466*, 1129–1133.

16 Ko, M.; Huang, Y.; Jankowska, A. M.; Pape, U. J.; Tahiliani, M.; Bandukwala, H. S.; An, J.; Lamperti, E. D.; Koh, K. P.; Ganetzky, R.; Liu, X. S.; Aravind, L.; Agarwal, S.; Maciejewski, J. P.; Rao, A. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **2010**, *468*, 839–843.

17 Koh, K. P.; Yabuuchi, A.; Rao, S.; Huang, Y.; Cunniff, K.; Nardone, J.; Laiho, A.; Tahiliani, M.; Sommer, C. A.; Mostoslavsky, G.; Lahesmaa, R.; Orkin, S. H.; Rodig, S. J.; Daley, G. Q.; Rao, A. Tet1 and Tet2 Regulate 5-Hydroxymethylcytosine Production and Cell Lineage Specification in Mouse Embryonic Stem Cells. *Cell Stem Cell* **2011**, *8*, 200–213.

18 Loenarz, C.; Schofield, C. J. Oxygenase Catalyzed 5-Methylcytosine Hydroxylation. *Chem. Biol.* **2009**, *16*, 580–583.

19 Hayatsu, H.; Shiragami, M. Reaction of bisulfite with the 5-hydroxymethyl group in pyrimidines and in phage DNAs. *Biochemistry* **1979**, *18*, 632–637.

20 Nestor, C.; Ruzov, A.; Meehan, R. R.; Dunican, D. S. Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. *BioTechniques* **2010**, *48*, 317–319.

21 Huang, Y.; Pastor, W. A.; Shen, Y.; Tahiliani, M.; Liu, D. R.; Rao, A. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One* **2010**, *5*, e8888.

22 Jin, S. G.; Kadam, S.; Pfeifer, G. P. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Res.* **2010**, *38*, e125.

23 Flusberg, B. A.; Webster, D. R.; Lee, J. H.; Travers, K. J.; Olivares, E. C.; Clark, T. A.; Korlach, J.; Turner, S. W. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* **2010**, *7*, 461–465.

24 Hatanaka, Y.; Sadakane, Y. Photoaffinity labeling in drug discovery and developments: chemical gateway for entering proteomic frontier. *Curr. Top. Med. Chem.* **2002**, *2*, 271–288.

25 Meisenheimer, K. M.; Koch, T. H. Photocross-linking of nucleic acids to associated proteins. *Crit. Rev. Biochem. Mol. Biol.* **1997**, *32*, 101–140.

26 Kotzyba-Hibert, F.; Kapfer, I.; Goeldner, M. Recent Trends in Photoaffinity Labeling. *Angew. Chem., Int. Ed.* **1995**, *34*, 1296–1312.

27 Khodyreva, S. N.; Lavrik, O. I. Photoaffinity labeling technique for studying DNA replication and DNA repair. *Curr. Med. Chem.* **2005**, *12*, 641–655.

28 Brunner, J. New photolabeling and crosslinking methods. *Annu. Rev. Biochem.* **1993**, *62*, 483–514.

29 Magdalena Coman, M.; Jin, M.; Ceapa, R.; Finkelstein, J.; O'Donnell, M.; Chait, B. T.; Hingorani, M. M. Dual functions, clamp opening and primer-template recognition, define a key clamp loader subunit. *J. Mol. Biol.* **2004**, *342*, 1457–1469.

30 Zhang, C. X.; Chang, P. V.; Lippard, S. J. Identification of Nuclear Proteins that Interact with Platinum-Modified DNA by Photoaffinity Labeling. *J. Am. Chem. Soc.* **2004**, *126*, 6536–6537.

31 Lagrange, T.; Kim, T. K.; Orphanides, G.; Ebright, Y. W.; Ebright, R. H.; Reinberg, D. High-resolution mapping of nucleoprotein complexes by site-specific protein-DNA photocrosslinking: organization of the human TBP-TFIIA-TFIIB-DNA quaternary complex. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 10620–10625.

32 Musier-Forsyth, K.; Schimmel, P. Acceptor helix interactions in a class II tRNA synthetase: photoaffinity cross-linking of an RNA miniduplex substrate. *Biochemistry* **1994**, *33*, 773–779.

33 Pingoud, V.; Thole, H.; Christ, F.; Grindl, W.; Wende, W.; Pingoud, A. Photocross-linking of the homing endonuclease PI-SceI to its recognition sequence. *J. Biol. Chem.* **1999**, *274*, 10235–10243.

34 Tomohiro, T.; Hashimoto, M.; Hatanaka, Y. Cross-linking chemistry and biology: development of multifunctional photoaffinity probes. *Chem. Rec.* **2005**, *5*, 385–395.

35 Yamaguchi, T.; Saneyoshi, M. A photolabile 2′,3′-dideoxyuridylate analog bearing an aryl(trifluoromethyl)diazirine moiety: photoaffinity labeling of HIV-1 reverse transcriptase. *Nucleic Acids Res.* **1996**, *24*, 3364–3369.

36 Yamaguchi, T.; Suyama, K.; Narita, K.; Kohgo, S.; Tomikawa, A.; Saneyoshi, M. Synthesis and evaluation of oligodeoxyribonucleotides containing an aryl(trifluoromethyl)diazirine moiety as the cross-linking probe: photoaffinity labeling of mammalian DNA polymerase beta. *Nucleic Acids Res.* **1997**, *25*, 2352–2358.

37 Sergiev, P. V.; Lavrik, I. N.; Wlasoff, V. A.; Dokudovskaya, S. S.; Dontsova, O. A.; Bogdanov, A. A.; Brimacombe, R. The path of mRNA through the bacterial ribosome: a site-directed crosslinking study using new photoreactive derivatives of guanosine and uridine. *RNA* **1997**, *3*, 464–475.

38 Topin, A. N.; Gritsenko, O. M.; Brevnov, M. G.; Gromova, E. S.; Korshunova, G. A. Synthesis of a new Photo-Cross-Linking Nucleoside Analogue Containing an Aryl(Trifluoromethyl)-Diazirine Group: Application for Eco RII and Mva I Restriction-Modification Enzymes. *Nucleosides Nucleotides* **1998**, *17*, 1163–1175.

39 Tate, J. J.; Persinger, J.; Bartholomew, B. Survey of four different photoreactive moieties for DNA photoaffinity labeling of yeast RNA polymerase III transcription complexes. *Nucleic Acids Res.* **1998**, *26*, 1421–1426.

40 Persinger, J.; Sengupta, S. M.; Bartholomew, B. Spatial organization of the core region of yeast TFIIIB-DNA complexes. *Mol. Cell. Biol.* **1999**, *19*, 5218–5234.

41 Zofall, M.; Bartholomew, B. Two novel dATP analogs for DNA photoaffinity labeling. *Nucleic Acids Res.* **2000**, *28*, 4382–4390.

42 Halbfinger, E.; Gorochesky, K.; Levesque, S. A.; Beaudoin, A. R.; Sheihet, L.; Margel, S.; Fischer, B. Photoaffinity labeling on magnetic microspheres (PALMm) methodology for topographic mapping: preparation of PALMm reagents and demonstration of biochemical relevance. *Org. Biomol. Chem.* **2003**, *1*, 2821–2832.

43 Sergiev, P.; Dokudovskaya, S.; Romanova, E.; Topin, A.; Bogdanov, A.; Brimacombe, R.; Dontsova, O. The environment of 5S rRNA in the ribosome: cross-links to the GTPase-associated area of 23S rRNA. *Nucleic Acids Res.* **1998**, *26*, 2519–2525.

44 Liebmann, M.; Di Pasquale, F.; Marx, A. A new photoactive building block for investigation of DNA backbone interactions: photoaffinity labeling of human DNA polymerase beta. *ChemBioChem* **2006**, *7*, 1965–1969.

45 Suchanek, M.; Radzikowska, A.; Thiele, C. Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nat. Methods* **2005**, *2*, 261–267.

46 Wyatt, G. R.; Cohen, S. S. The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. *Biochem. J.* **1953**, *55*, 774–782.

47 Jesaitis, M. A. Differences in the Chemical Composition of the Phage Nucleic Acids. *Nature* **1956**, *178*, 637–637.

48 Josse, J.; Kornberg, A. Glucosylation of Deoxyribonucleic Acid 0.3. Alpha and Beta-Glucosyl Transferases from T4-Infected Escherichia Coli. *J. Biol. Chem.* **1962**, *237*, 1968–1976.

49 Lariviere, L.; Morera, S. Structural evidence of a passive base-flipping mechanism for beta-glucosyltransferase. *J. Biol. Chem.* **2004**, *279*, 34715–34720.

50 Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. A stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem., Int. Ed.* **2002**, *41*, 2596–2599.

51 Speers, A. E.; Cravatt, B. F. Profiling enzyme activities in vivo using click chemistry methods. *Chem. Biol.* **2004**, *11*, 535–546.

52 Best, M. D. Click chemistry and bioorthogonal reactions: unprecedented selectivity in the labeling of biological molecules. *Biochemistry* **2009**, *48*, 6571–6584.

53 Sletten, E. M.; Bertozzi, C. R. Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angew. Chem., Int. Ed.* **2009**, *48*, 6974–6998.

54 Mamidyala, S. K.; Finn, M. G. In situ click chemistry: probing the binding landscapes of biological molecules. *Chem. Soc. Rev.* **2010**, *39*, 1252–1261.

55 Tardy-Planechaud, S.; Fujimoto, J.; Lin, S. S.; Sowers, L. C. Solid phase synthesis and restriction endonuclease cleavage of oligodeoxynucleotides containing 5-(hydroxymethyl)-cytosine. *Nucleic Acids Res.* **1997**, *25*, 553–559.

56 Dai, Q.; Song, C.-X.; Pan, T.; He, C. Syntheses of Two 5-Hydroxymethyl-2′-deoxycytidine Phosphoramidites with TBDMS as the 5-Hydroxymethyl Protecting Group and Their Incorporation into DNA. *J. Org. Chem.* 2011, published online April 4, http://dx.doi.org/10.1021/jo200566d.

57 Baskin, J. M.; Prescher, J. A.; Laughlin, S. T.; Agard, N. J.; Chang, P. V.; Miller, I. A.; Lo, A.; Codelli, J. A.; Bertozzi, C. R. Copper-free click chemistry for dynamic in vivo imaging. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 16793–16797.

58 Ning, X.; Guo, J.; Wolfert, M. A.; Boons, G. J. Visualizing metabolically labeled glycoconjugates of living cells by copper-free and fast huisgen cycloadditions. *Angew. Chem., Int. Ed.* **2008**, *47*, 2253–2255.

59 Jewett, J. C.; Bertozzi, C. R. Cu-free click cycloaddition reactions in chemical biology. *Chem. Soc. Rev.* **2010**, *39*, 1272–1279.

60 Song, C.-X.; Yu, M.; Dai, Q.; He, C. Detection of 5-Hydroxymethylcytosine in a Combined Glycosylation Restriction Analysis (CGRA) Using Restriction Enzyme Taq$^{\alpha}$I. *Bioorg. Med. Chem. Lett.* 2011, published online April 9, http://dx.doi.org/10.1016/j.bmcl.2011.03.118.

61 Szwagierczak, A.; Bultmann, S.; Schmidt, C. S.; Spada, F.; Leonhardt, H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* **2010**, *38*, e181.

62 Munzel, M.; Globisch, D.; Bruckl, T.; Wagner, M.; Welzmiller, V.; Michalakis, S.; Muller, M.; Biel, M.; Carell, T. Quantification of the Sixth DNA Base Hydroxymethylcytosine in the Brain. *Angew. Chem., Int. Ed.* **2010**, *49*, 5375–5377.

63 Laitinen, O. H.; Hytonen, V. P.; Nordlund, H. R.; Kulomaa, M. S. Genetically engineered avidins and streptavidins. *Cell. Mol. Life Sci.* **2006**, *63*, 2992–3017.

64 Verdine, G. L.; Norman, D. P. Covalent trapping of protein-DNA complexes. *Annu. Rev. Biochem.* **2003**, *72*, 337–366.

65 Yang, C. G.; Yi, C.; Duguid, E. M.; Sullivan, C. T.; Jian, X.; Rice, P. A.; He, C. Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature* **2008**, *452*, 961–965.

66 Yi, C.; Yang, C. G.; He, C. A non-heme iron-mediated chemical demethylation in DNA and RNA. *Acc. Chem. Res.* **2009**, *42*, 519–529.

67 Yi, C.; Jia, G.; Hou, G.; Dai, Q.; Zhang, W.; Zheng, G.; Jian, X.; Yang, C. G.; Cui, Q.; He, C. Iron-catalysed oxidation intermediates captured in a DNA repair dioxygenase. *Nature* **2010**, *468*, 330–333.

68 Shigdel, U. K.; He, C. A New 1′-Methylenedisulfide Deoxyribose that Forms an Efficient Cross-Link to DNA Cytosine-5 Methyltransferase (DNMT). *J. Am. Chem. Soc.* **2008**, *130*, 17634–17635.

69 Church, R. F. R.; Weiss, M. J. Diazirines. II. Synthesis and properties of small functionalized diazirine molecules. Observations on the reaction of a diaziridine with the iodine-iodide ion system. *J. Org. Chem.* **1970**, *35*, 2465–2471.

70 Korshunova, G. A.; Sumbatyan, N. V.; Topin, A. N.; Mtchedlidze, M. T. Photoactivatable Reagents Based on Aryl(trifluoromethyl)diazirines: Synthesis and Application for Studying Nucleic Acid−Protein Interactions. *Mol. Biol.* **2000**, *34*, 823–839.

71 Xu, Y. Z.; Zheng, Q.; Swann, P. F. Synthesis of DNA containing modified bases by post-synthetic substitution. Synthesis of oligomers containing 4-substituted thymine: O4-alkylthymine, 5-methylcytosine, N4-dimethylamino-5-methylcytosine, and 4-thiothymine. *J. Org. Chem.* **1992**, *57*, 3839–3845.

72 Macmillan, A. M.; Verdine, G. L. Synthesis of Functionally Tethered Oligodeoxynucleotides by the Convertible Nucleoside Approach. *J. Org. Chem.* **1990**, *55*, 5931–5933.

73 Macmillan, A. M.; Verdine, G. L. Engineering Tethered DNA-Molecules by the Convertible Nucleoside Approach. *Tetrahedron* **1991**, *47*, 2603–2616.

74 Licatalosi, D. D.; Mele, A.; Fak, J. J.; Ule, J.; Kayikci, M.; Chi, S. W.; Clark, T. A.; Schweitzer, A. C.; Blume, J. E.; Wang, X.; Darnell, J. C.; Darnell, R. B. HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **2008**, *456*, 464–469.

75 Chi, S. W.; Zang, J. B.; Mele, A.; Darnell, R. B. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **2009**, *460*, 479–486.