Published in final edited form as:

J Am Chem Soc. 2017 August 09; 139(31): 10617–10620. doi:10.1021/jacs.7b05495.

5-Formylcytosine Yields DNA-Protein Crosslinks in Nucleosome Core Particles

Fengchao Li^{†,‡}, Yingqian Zhang^{†,‡}, Jing Bai[†], Marc M. Greenberg[§], Zhen Xi^{†,‡}, and Chuanzheng Zhou^{†,‡}

[†]State Key Laboratory of Elemento-Organic Chemistry and Department of Chemical Biology, College of Chemistry, Nankai University, Tianjin 300071, China

[‡]Collaborative Innovation Center of Chemical Science and Engineering (Tianjin), Tianjin 300071, China

[§]Department of Chemistry, Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, United States

Abstract

In situ generation of 5-formylcytosine (5fC) in nucleosome core particles (NCPs) reveals that 5fC leads to essential DNA-protein crosslinks (DPCs). Mechanistic studies using chemical models and mutated histones demonstrate that DPCs form reversibly between the formyl function of 5fC and primary amines on histones. These results suggest that DPC formation from 5fC in chromatin occurs in addition to its role in DNA demethylation.

5-Formylcytosine (5fC) has been identified as a naturally occurring nucleobase in genomic DNA.^{1–4} It was initially thought to be merely an intermediate in the ten-eleven translocation (TET) enzyme-mediated DNA demethylation pathway. TETs catalyze the stepwise oxidation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5fC and 5-carboxylcytosine (5caC).⁵ Cleavage of 5fC with thymine DNA glycosylase (TDG) followed by base-excision repair (BER) results in restoration of normal cytosine.⁶ More recently, 5fC was found to be a stable DNA modification that exists for weeks in mammals.⁷ This modification can alter the structure of the DNA double helix⁸ and is possibly involved in transcription regulation and chromatin remodeling.^{9–11}

Eukaryotic nuclear DNA is assembled in nucleosome core particles (NCPs), the monomeric unit of chromatin. A NCP consists of ~145 bp of DNA tightly wrapped ~1.65 turns around an octameric core of histone proteins containing two copies each of H2A, H2B, H3, and H4. Histones are lysine rich, particularly in their flexible N-terminal tails that protrude from the core. Modifications on these lysines play very important roles in dynamically regulating

Correspondence to: Chuanzheng Zhou.

[#]Author Contributions: These authors contributed equally.

Supporting Information: General experimental methods, UPLC profiles, NMR and MS spectra, and gel pictures. This material is available free of charge via the Internet at <http://pubs.acs.org>.

nucleosome structure and function.¹² In addition, these same lysine residues catalyze DNA strand scission at abasic sites (APs) within NCPs via lysine-AP crosslink intermediates.¹³⁻¹⁶

The formyl group of 5fC is highly reactive toward nucleophilic primary amines, hydrazides, and aminoxy derivatives. 5fC labeling and detection methods have been developed based on this reactivity pattern.^{4,17-20} The high activity of 5fC raises the possibility of DNA-protein crosslink (DPC) formation by 5fC with ϵ -amino and/or α -amino groups on proteins, which may expand the role of 5fC in genomic DNA. Herein, we describe a method for generating 5fC *in situ* in NCPs and demonstrate that it yields DPCs.

The viability of DPC formation was explored by reacting 5fC nucleosides and amino acid analogues. Treatment of **1a** with excess ethylamine (**2a**) in aqueous solution yielded **3a** (Figure 1A) as the sole product. The C=N bond in **3a** was unambiguously characterized by NMR (Figure S8). This result encouraged us to explore the reaction of nucleoside **1b** with 100 eq. of protected lysine **2b** in HEPES buffer (pH 7.5). 5fC-lysine conjugate **3b** was gradually formed over time and reached steady state (4% relative to **1b**) after 3 h (Figure 1B). If water was evaporated to make the solution more concentrated, **3b** reached an ~80% steady state yield (Figure 1C). At this point, the same amount of water that has been evaporated was added to dilute the solution, and the initial steady state (4% product) was reestablished. Thus, formation of 5fC-lysine conjugate **3b** is reversible and the equilibrium is pH dependent. At pH 6.0 almost no conjugate was observed after 12 h incubation, whereas at pH 9.5, 56% of **3b** was present at equilibrium (Figure S2).

Under the same conditions, glycineamide **2c**, an N-terminal peptide analogue reacted with **1b** at pH 7.5 to yield 14% of conjugate **3c** at equilibrium (Figure 1B). To probe the reactivity of 5fC toward other amino acids that contain nucleophilic side chains, **1b** was treated with 100 eq. of **2d-h** in HEPES buffer (pH 7.5). No reaction was detected, suggesting that 5fC reacts specifically with primary amines in proteins, including the ϵ -amine of lysine and N-terminal α -amine.

The above results encouraged us to introduce 5fC site-specifically to NCP and monitor DPC formation. The formyl group of 5fC is compatible with standard DNA synthesis protocols.²¹ However, the possibility of DPC formation during NCP reconstitution led us to protect the formyl moiety. Inspired by the 1,3-dioxane protecting group strategy,²² we introduced 4-(2-nitrophenyl)-1,3-dioxane as a photolabile protecting group for the formyl function of 5fC. Thus, phosphoramidite **8** was synthesized starting from known **1a** in 4 steps (Figure 2A) and efficiently introduced into oligonucleotides through standard solid-phase DNA synthesis. The aldehyde protecting group was stable during DNA synthesis and purification, and quantitatively removed upon photolysis at 350 nm for 5 min (Figure S3). Oligonucleotides containing **9** were ligated to produce the 145 nt DNA whose sequence was based upon the strong positioning “601 DNA”.²³ Ligation was carried out at pH 7.9 to prevent loss of the aldehyde protecting group.

NCPs containing a single **9** at position 3, 45, 74, or 89 were prepared by reconstituting 5'-FAM labeled DNAs with histone octamer via the salt dialysis method (Figure 2B).¹⁴ Position 89 is located in an extremely bent region that is readily accessed by DNA-damaging

molecules.^{23,24} NCP containing **9** at position 89 was irradiated at 350 nm for 5 min and then incubated at 37 °C for 44 h. After quenching by NaBH₄, which stabilizes DPCs by reducing the hydrolysable C=N bond to a stable C-N bond, the reaction was analyzed using 10% SDS PAGE (Figure 3A), which revealed a new slower migrating DNA band. This product was converted to a product that co-migrates with the original 145 bp dsDNA after treatment with proteinase K, suggesting it is a DPC. DPC formation was attributed to the formyl moiety of 5fC because of the absence of DPC without photo irradiation (Figure 3A).

Kinetic studies revealed that DPC formation in NCP via 5fC₈₉ increased over time and reached a maximum of ~20% after 68 h incubation. At this stage, addition of 10 mM of hydroxylamine resulted in rapid DPC decomposition (Figure 3B). Adding hydroxylamine exhibited no adverse effect on NCP integrity (Figure S4). DPC loss is consistent with reversible DPC formation via Schiff base formation. Hydroxylamine addition pushes the equilibrium towards noncrosslinked material by trapping the free aldehyde as a stable oxime (Figure 3C).

To reveal which histone is responsible for DPC formation with 5fC₈₉ in the NCP, the stabilized product was subjected to in-gel tryptic digestion. UPLC-MS/MS analysis of tryptic peptide fragments identified the protein as histone H4 (Figure S5). This result is consistent with the crystal structure of a NCP containing the 601 DNA sequence in which position 89 is in close proximity to the lysine rich N-terminal of H4.²³ Deleting the N-terminal tail (1-20 amino acids, Del1-20) of H4 significantly reduced DPC formation (Table 1 and Figure S6), confirming that it is responsible for DPC formation with 5fC₈₉. Mutating all lysines in the tail to arginines (H4-K5, 8, 12, 16, 20R) only slightly reduced the DPC formation (from 20% to 15%), indicating that the N terminal α-amine group compensates for the absence of ε-amine group of the lysines. To confirm the role of the N terminal amine group in DPC formation, H4 with the N-terminal amine capped by thiazolidine (H4-CapN) was prepared (Scheme 1 & Figure S7).²⁵ The yield of DPC at equilibrium formed in the NCP containing this modified H4 was reduced to ~7%. Furthermore, the DPC yield was reduced to negligible levels when the thiazolidine cap was combined with the poly-lysine mutations (H4-CapN-K5, 8, 12, 16, 20R mutant). These results strongly indicate the ε-amine of lysines and N-terminal amine react with 5fC₈₉ to form DPCs, and that the latter is more reactive and/or forms more stable product.

We found that neither 5fC₈₉-A or a 5fC₈₉-T mismatch significantly affects DPC formation. However, DPC formation is strongly location dependent (Table 1). Positions 74 and 3 are near the dyad axis where the interaction between DNA and the N terminal tails of histones is weakest due to the distance (Figure 2B). DPC formation at 5fC₇₄ and 5fC₃ was negligible. Whereas 5fC₄₅, which is located near the tails of histones H2A and H2B, led to ~15% of DPC. The variable reactivity of 5fCs in NCPs probably owe to their different exposure and proximity to the histone tails, which is consistent with previous observations that slight changes in rotational positioning can affect the DNA damage to some extent in NCPs.²⁶⁻²⁸

In summary, 5fC has been introduced site-specifically into reconstituted NCPs using a protected precursor. In situ generation of 5fC via photolysis reveals that the modified nucleotide leads to DPC formation. The efficiency of DPC formation is location dependent

and is up to ~20%. Mechanistic studies using chemical models, as well as mutated histones suggest that DPCs are formed via C=N bond between the formyl function of 5fC and primary amines on histones, and exist in dynamic equilibrium. For 5fC₈₉, the equilibrium is dominated by the N-terminal amine of histone H4.

Though through iterative or distributive mechanism for oxidation of 5mC by TETs is still in dispute,^{29,30} the long lifetime of 5fC in mammals⁷ strongly suggest that DPC formation via 5fC in NCPs could occur in eukaryotic cells and have important biological consequences, such as chromatin remodeling, transcription, and DNA demethylation interference. This study enriches our understanding of 5fC and therefore encourages us to further examine its biological function *in vitro* and *in vivo*.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NSFC (21572109, 21332004) and Natural Science Foundation of Tianjin City (15JCYBJC53300). C. Z. is grateful for the sponsorship from the National Thousand Young Talents Program. M. M. G. is grateful to the National Institute of General Medical Science (GM-063028) for financial support.

References

1. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y. Science. 2011; 333:1300–1303. [PubMed: 21778364]
2. Booth MJ, Marsico G, Bachman M, Beraldi D, Balasubramanian S. Nat Chem. 2014; 6:435–440. [PubMed: 24755596]
3. Wagner M, Steinbacher J, Kraus TFJ, Michalak S, Hackner B, Pfaffeneder T, Perera A, Mueller M, Giese A, Kretschmar HA, Carell T. Angew Chem Int Ed Engl. 2015; 54:12511–12514. [PubMed: 26137924]
4. Xia B, Han D, Lu X, Sun Z, Zhou A, Yin Q, Zeng H, Liu M, Jiang X, Xie W, He C, Yi C. Nat Methods. 2015; 12:1047–1050. [PubMed: 26344045]
5. Lu X, Zhao BS, He C. Chem Rev. 2015; 115:2225–2239. [PubMed: 25675246]
6. Kohli RM, Zhang Y. Nature. 2013; 502:472–479. [PubMed: 24153300]
7. Bachman M, Uribe-Lewis S, Yang X, Burgess HE, Iurlaro M, Reik W, Murrell A, Balasubramanian S. Nat Chem Biol. 2015; 11:555–557. [PubMed: 26098680]
8. Raiber EA, Murat P, Chirgadze DY, Beraldi D, Luisi BF, Balasubramanian S. Nat Struct Mol Biol. 2015; 22:44–49. [PubMed: 25504322]
9. Kellinger MW, Song CX, Chong J, Lu XY, He C, Wang D. Nat Struct Mol Biol. 2012; 19:831–833. [PubMed: 22820989]
10. Iurlaro M, Ficiz G, Oxley D, Raiber EA, Bachman M, Booth MJ, Andrews S, Balasubramanian S, Reik W. Genome Biol. 2013; 14:R119. [PubMed: 24156278]
11. Nakano S, Suzuki T, Kawarada L, Iwata H, Asano K, Suzuki T. Nat Chem Biol. 2016; 12:546–551. [PubMed: 27214402]
12. Bowman GD, Poirier MG. Chem Rev. 2015; 115:2274–2295. [PubMed: 25424540]
13. Zhou CZ, Szczepanski JT, Greenberg MM. J Am Chem Soc. 2012; 134:16734–16741. [PubMed: 23020793]
14. Zhou CZ, Szczepanski JT, Greenberg MM. J Am Chem Soc. 2013; 135:5274–5277. [PubMed: 23531104]
15. Zhou CZ, Greenberg MM. J Am Chem Soc. 2012; 134:8090–8093. [PubMed: 22551239]

16. Weng LW, Greenberg MM. *J Am Chem Soc.* 2015; 137:11022–11031. [PubMed: 26290445]
17. Hu J, Xing X, Xu X, Wu F, Guo P, Yan S, Xu Z, Xu J, Weng X, Zhou X. *Chem Eur J.* 2013; 19:5836–5840. [PubMed: 23512548]
18. 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; 15:3266–3269. [PubMed: 23767758]
19. Samanta B, Seikowski J, Höbartner C. *Angew Chem Int Ed.* 2015; 55:1912–1916.
20. Su M, Kirchner A, Stazzoni S, Mueller M, Wagner M, Schroeder A, Carell T. *Angew Chem Int Ed Engl.* 2016; 55:11797–11800. [PubMed: 27561097]
21. Dai Q, He C. *Org Lett.* 2011; 13:3446–3449. [PubMed: 21648398]
22. Schroeder AS, Steinbacher J, Steigenberger B, Gnerlich FA, Schiesser S, Pfaffeneder T, Carell T. *Angew Chem Int Ed Engl.* 2014; 53:315–318. [PubMed: 24281791]
23. Vasudevan D, Chua EYD, Davey CA. *J Mol Biol.* 2010; 403:1–10. [PubMed: 20800598]
24. Kuduvalli PN, Townsend CA, Tullius TD. *Biochemistry.* 1995; 34:3899–3906. [PubMed: 7696253]
25. Zhang L, Tam JP. *Anal Biochem.* 1996; 233:87–93. [PubMed: 8789151]
26. Song Q, Cannistraro VJ, Taylor JS. *Nucleic Acids Res.* 2014; 42:13122–13133. [PubMed: 25389265]
27. Wang K, Taylor JS. *Nucleic Acids Res.* 2017; 45:7031–7041. [PubMed: 28525579]
28. Szczepanski JT, Zhou CZ, Greenberg MM. *Biochemistry.* 2013; 52:2157–2164. [PubMed: 23480734]
29. Crawford DJ, Liu MY, Nabel CS, Cao XJ, Garcia BA, Kohli RM. *J Am Chem Soc.* 2016; 138:730–733. [PubMed: 26734843]
30. Tamanaha E, Guan S, Marks K, Saleh L. *J Am Chem Soc.* 2016; 138:9345–9348. [PubMed: 27362828]

Figure 1. Reversible Schiff bases formation between 5fC nucleosides and amino acids. (A) Reaction of 5fC nucleosides with amino acids. (B) Kinetics of formation of Schiff bases **3b/c** in HEPES buffer (pH 7.5). (C) UPLC showing the dependence of reversible **3b** formation on H_2O .

Figure 2.

In situ generation of 5fC in NCP. (A) Synthesis of protected 5fC building block for oligonucleotide assembly. (B) Generation of 5fC upon photolysis in NCPs.

Figure 3. DPC formation by 5fC in NCP. (A) 10% SDS PAGE analysis of the DPC formation in NCP containing 5fC₈₉. (B) Kinetics of DPC formation in NCP containing 5fC₈₉. (C) Proposed reversible DPC formation based upon hydroxylamine trapping.

Scheme 1.

Preparation of N terminal modified histone H4-CapN.

Table 1

Efficiency of DPC formation by 5fC within NCPs.

5fC position and base pair	histone H4 mutants in NCP	DPC% at equilibrium ^a
5fC ₈₉ -G	WT ^b	20.1 ± 0.7
	Del 1-20 ^c	4.2 ± 0.2
	K5, 8, 12, 16, 20R ^d	14.7 ± 4.1
	CapN ^e	6.7 ± 0.2
	CapN-K5, 8, 12, 16, 20R ^f	3.3 ± 0.2
5fC ₈₉ -A	WT	16.6 ± 4.6
5fC ₈₉ -T	WT	21.4 ± 0.7
5fC ₇₄ -G	WT	3.2 ± 0.1
5fC ₃ -G	WT	3.6 ± 0.1
5fC ₄₅ -G	WT	14.8 ± 0.3

^aYields are averages ± standard deviations of at least three experiments.^bWild type histone H4.^cHistone H4 without the N-terminal 1-20 amino acids.^dH4 with Lys 5, 8, 12, 16, 20 to Arg mutations.^eH4 with capped N terminal by thiazolidine (see Scheme 1).^fH4 with both capped N terminal and Lys 5, 8, 12, 16, 20 to Arg mutations.