Programmable Protein-DNA Crosslinking for the Direct Capture and Quantification of 5-Formylcytosine

Mario Gieß, Álvaro Muñoz-López, Benjamin Buchmuller, Grzegorz Kubik and Daniel Summerer*

Faculty of Chemistry and Chemical Biology, TU Dortmund University, Otto-Hahn-Str. 6, 44227 Dortmund (Germany)

Supporting Information Placeholder

ABSTRACT: 5-Formylcytosine (5fC) is an epigenetic nucleobase of mammalian genomes that occurs as intermediate of active DNA demethylation. 5fC uniquely interacts and reacts with key nuclear proteins, indicating functions in genome regulation. Transcription-activator-like effectors (TALEs) are repeat-based DNA binding proteins that can serve as probes for the direct, programmable recognition and analysis of epigenetic nucleobases. However, no TALE repeats for the selective recognition of 5fC are available, and the typically low genomic levels of 5fC represent a particular sensitivity challenge. We here advance TALE-based nucleobase targeting from recognition to covalent crosslinking. We report TALE repeats bearing the ketone-amino acid p-acetylphenylalanine (pAcF) that universally bind all mammalian cytosine nucleobases, but selectively form diaminooxy-linker-mediated dioxime crosslinks to 5fC. We identify repeat-linker combinations enabling single CpG resolution, and demonstrate the direct quantification of 5fC levels in a human genome background by covalent enrichment. This strategy provides a new avenue to expand the application scope of programmable probes with selectivity beyond A, G, T and C for epigenetic studies.

Methylation of the cytosine 5-position in DNA serves as regulatory element of mammalian transcription with important roles in development, X-chromosome inactivation, genomic imprinting and diseases.[1] The resulting nucleobase 5mC can be oxidized by ten-eleven translocation (TET) dioxygenases to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC, Fig. 1a) which are intermediates of an active demethylation process and may have inherent regulatory functions.[2-6] Among these nucleobases, 5fC stands out by the presence of an electrophilic aldehyde group that is available for addition and condensation reactions.[7-8] Indeed, besides exhibiting unique protein interactions[9-11] and influencing DNA flexibility and nucleosome stability,[12] 5fC can form imine-crosslinks with lysine-containing nuclear proteins and is involved in nucleosome organization.[13-15]

Analysis of 5fC in user-defined DNA sequences is key to further unravel its biological functions.[1] Here, the reactivity of 5fC has been exploited for altering its behavior in bisulfite conversion or its Watson-Crick pairing during sequencing,[16-18] or for the conjugation of click handles, haptens, fluorophores or oligonucleotides[16, 19-22] for subsequent analysis by e.g. affinity enrichment.[7]

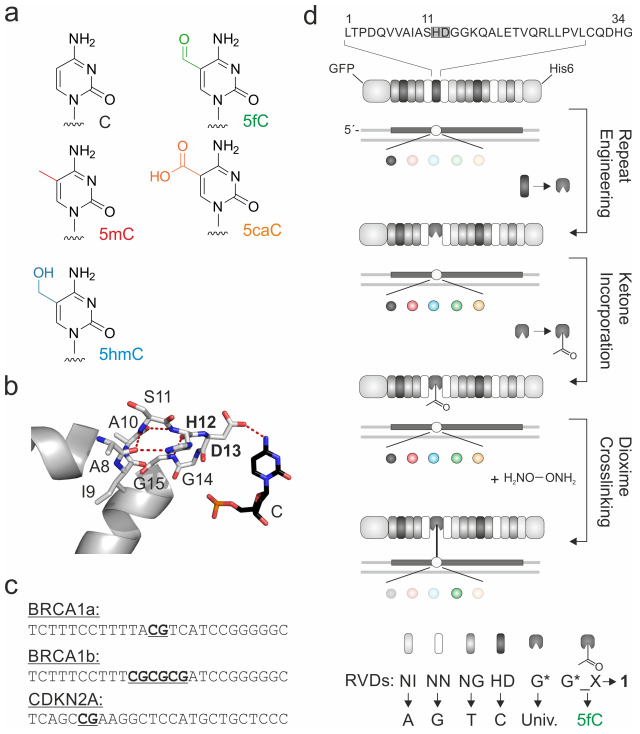

Figure 1. Concept of programmable protein-DNA crosslinking at 5fC. a.) Structures of C, 5mC, 5hmC, 5fC, and 5caC. b.) TALE repeat loop with RVD HD binding to C. Hydrogen bonds in red (pdb 3V6T[23]). c.) Target DNA sequences used in this study. d.) Concept. Sequence of one repeat on top (RVD in grey box). RVD selectivities on the bottom. TALE target sequence is shown as dark grey bar, cytosines as spheres (color-code as in Fig. 1a). TALE-bound nucleobases are saturated, non-bound nucleobases pale.

The design of programmable probes that provide selectivity for 5fC directly in its canonical sequence context promises a particular simple, yet highly resolved analysis at user-defined sites.[7-8, 20] TALE proteins[24] have emerged as a promising scaffold for engineering such probes. TALEs are repeat-based proteins and recognize DNA in a one repeat versus one nucleobase pairing via a repeat variable di-residue (RVD) positioned in a loop (Fig. 1b).[23, 25] This recognition occurs with one DNA strand via the major groove.[23, 25] The selectivity of TALE repeats is defined by the RVD, with the RVDs NI, NN, NG and HD binding A, G(A), T, and C, respectively.[24] TALE repeats with selectivity for

5mC and oxidized derivatives have been engineered,[27-31] and enabled detection at single, user-defined positions by affinity enrichment.[32] However, 5fC exhibits low genomic levels,[17] and no TALE repeats for its selective, sensitive detection have been reported.

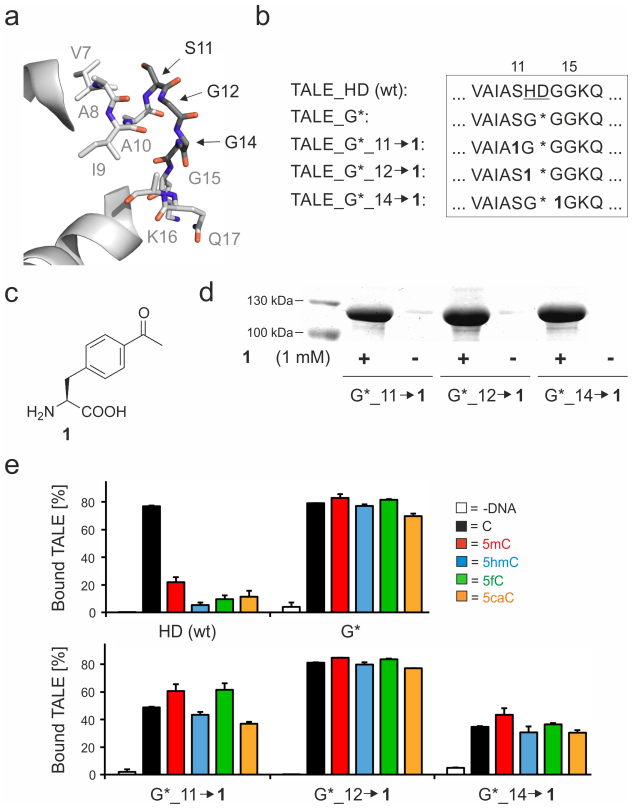

Figure 2. Design and nucleobase recognition selectivity of ketone-bearing TALE repeats. a.) Modelled TALE repeat G*[26] with incorporation positions for **1** marked with arrows. b.) Repeat designs used in this study (RVD underlined, * = deletion). c.) Structure of *p*AcF (**1**). d.) SDS PAGE analysis of TALE expressions with in-frame amber codon in presence or absence of **1** purified via a C-terminal His6 tag. e.) EMSA interaction analysis (20 mM Tris-HCl, pH = 8.0, 50 mM NaCl, 5 mM Mg_2Cl, 5 % glycerol) of TALEs with DNA oligonucleotides containing indicated nucleobase in BRCA1_a sequence opposite single indicated TALE RVD.

We here design repeats that bear a noncanonical ketone amino acid and universally bind all five cytosine nucleobases. This allows guiding a conjugation handle to user-defined DNA sequences and install crosslinks with 5fC via diaminooxy-linkers. Crosslinking overcomes the limited 5fC selectivity and sensitivity of available TALE repeats and enables the quantification of 5fC levels at single CpG by affinity enrichment in a human genome background.

We designed a TALE targeting a promoter sequence of the tumor suppressor BRCA1 (BRCA1a, Fig. 1c) and assembled[33] its gene into vector pGFP-ENTRY[32] for expression and purification of TALEs with N-terminal GFP domain, shortened, AvrBs3-type TALE N-terminal region (NTR) and a C-terminal His6 tag in *E. coli* (Fig. 1d and SI Fig. 1, 2). Instead of the C-selective RVD HD, we positioned RVD G* (* = deletion at position 13) opposite the C of the first CpG (Fig. 2a). This repeat bears a cavity opposite the pyrimidine 5-position and universally binds to all five cytosines.[28]

We then targeted loop positions 11, 12 and 14 (Fig. 2a, b) for the incorporation of the ketone-bearing ncAA *p*-acetyl-L-phenylalanine (**1**, Fig. 2c) via an orthogonal *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase/tRNA pair[34] in response to the amber codon in *E. coli* (Fig. 2d). We evaluated these repeats via electromobility shift assays (EMSA) using TALE_G*_11→**1**, TALE_G*_12→**1** and TALE_G*_14→**1** and oligonucleotide duplexes containing the BRCA1_a sequence with a single C, 5mC, 5hmC, 5fC or 5caC opposite the engineered repeat. In contrast to the corresponding C-selective TALE_HD, we observed universal binding of all five DNA targets for all three TALEs (Fig. 2e and SI Fig. 3 and 5). Compared to TALE_G*, TALE_G*_12→**1** and TALE_G*_11→**1** thereby showed similar or slightly reduced affinity, whereas binding of TALE_G*_14→**1** was reduced by ~50%. This indicates that the bulky aromatic side chain of **1** does not lead to nucleobase-selective clash at any of the positions,[28] but may somewhat interfere with overall TALE folding or complex formation when present at position 14.

Programmable TALE-DNA crosslinking for 5fC analysis requires high yields for maximal sensitivity, but should also provide high positional resolution by avoiding undesired crosslinking to adjacent off-target positions. This requires a suited position and orientation of **1** in combination with an optimal linker length to balance accessibility of target and off-target nucleobases, and entropic cost of crosslink formation. Among available condensation reactions with 5fC, we decided for oxime formations because of their simplicity, fast kinetics, and proven applicability in genomic tagging reactions. We synthesized a set of linkers with one to six ethyleneglycole moieties via Mitsunobu reaction with N-hydroxyphthalimide and subsequent hydrazinolysis (Fig. 3a, L1-5).[35] We incubated an excess of linkers with above BRCA1a oligonucleotide duplex bearing a single 5fC under conditions for *para*-phenylenediamine (*p*-PDA)-catalyzed oxime formation compatible with DNA hybridization (Fig. 3b and SI Fig. 8 for MS analyses). Briefly, we incubated 100 pmol DNA with 2 µmol *p*-PDA and 10 µmol L1-5 in 20 µL 20 mM phosphate buffer at pH = 6.0 and room temperature.[26] We then added 100 pmol TALE_G*_11→**1** (BRCA1a) in 30 µL binding buffer, and analyzed crosslink formation after incubation by SDS PAGE. Crosslink formation indicated that the TALE was still able to bind DNA, despite the bulky, linker-functionalized 5fC and acidic pH value. Crosslinking required a linker with a minimal length of two ethylenglycole moieties (L2) and occurred with maximal yields with linkers containing at least three moieties (L3-5, Fig. 3c). No crosslinking was observed between TALEs or between the TALE and an oligonucleotide duplex bearing a 5fC in the off-target sequence CDKN2A (Fig. 1c and 3c). Moreover, no crosslink was formed between BRCA1a oligonucleotides or with the TALE not bearing **1** (SI Fig. 6-9). Finally, experiments with linker L4 and DNA containing C, 5mC, 5hmC, 5fC or 5caC opposite the reactive repeat of TALE_G*_11→**1** indicated complete 5fC selectivity of crosslinking (Fig. 3d and SI Fig. 10).

Quantification of crosslinking for all TALE repeat - linker combinations revealed high crosslink yields for linkers L3-L5 with all repeat designs. However, influence of the position and orientation of **1** became apparent for linker L2,

with crosslinking only observed for TALE_G*_11→**1** and somewhat for TALE_G*_14→**1**, whereas L1 again did not afford any crosslinks (Fig. 3e and SI Fig. 11).

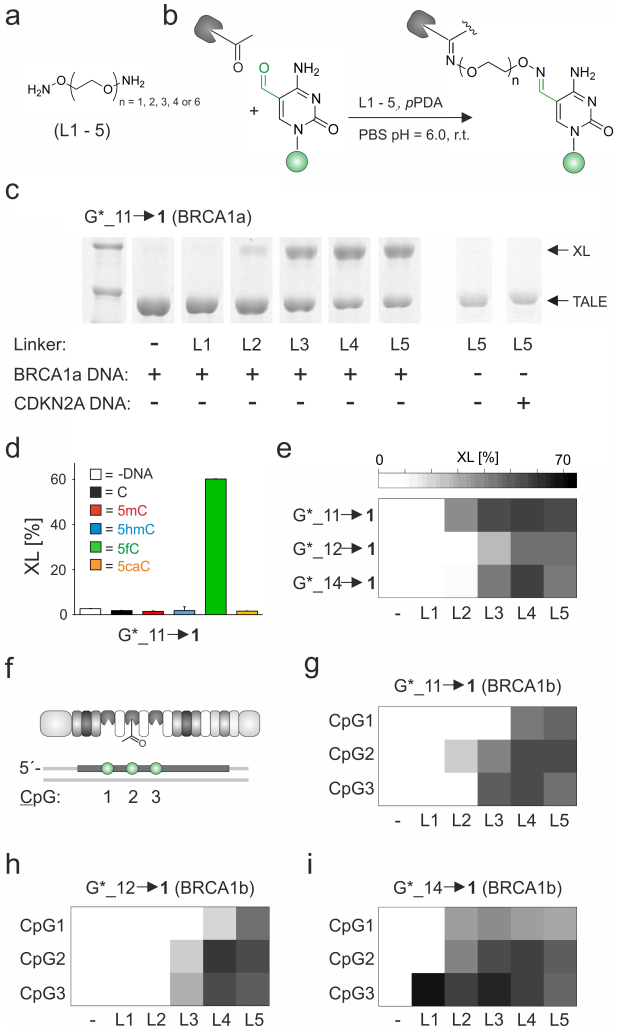

Figure 3. Programmable TALE-DNA crosslinking at 5fC. a.) Linkers used in this study. b.) Dioxime crosslinking between ketone-bearing TALE repeat and 5fC. c.) SDS PAGE analysis of basic crosslinking experiments with TALE_G*_11→**1** (BRCA1a). XL = Crosslink product. d.) Quantification of SDS PAGE analysis of crosslinking reactions with TALE_G*_11→**1** (BRCA1a) and DNA bearing single nucleobase as indicated opposite the engineered repeat. e.) Quantification of crosslinking reactions with indicated TALE-linker combinations by SDS-PAGE analysis using ^{32}P-labeled DNA. f.) Cartoon of strategy for studying positional resolution of crosslinking. g. - i.) Analysis of positional resolution of crosslinking conducted as in Fig. 3e. CpG1-3 indicate DNA with 5fC either at the first, second or third CpG with reactive repeat of TALEs opposite C of CpG2 and universal G* repeats opposite Cs of CpG1 and 3.

To assess potential off-target crosslinking, we constructed TALEs targeting a clustered triple-CpG sequence in the BRCA1 sequence context (BRCA1b, Fig. 1c). These beared one of the reactive repeats opposite the central CpG2 and universal G* repeats opposite the C positions of the two adjacent CpG1 and CpG3 (Fig. 3f). Crosslinking experiments with DNAs bearing a single 5fC in any of the three CpGs revealed some degree of off-target crosslinking for L3 - L5 in all cases (Fig. 3g-i and SI Fig. 12). However, though also L2 exhibited this promiscuity in several cases, its combination with repeat G*11→**1** showed selectivity for the central CpG2 (Fig. 3g). Interestingly, linker L1 in combination with repeat G*14→**1** crosslinked to the adjacent CpG1 with excellent yield and selectivity (Fig. 3i).

We next aimed to achieve detection of 5fC in the background of mammalian genomes. We previously reported an affinity enrichment assay enabling detection of epigenetic nucleobases at single, user-defined positions in the human genome that may be advanced to crosslinking-based analysis.[32] In this assay, magnetic bead-immobilized TALEs are incubated with fragmented genomic DNA (gDNA), and the TALE-bound DNA is eluted and quantified by qPCR after stringent washing. To obtain samples with precisely defined 5fC levels at single positions, we employed 5fC-free, fragmented human gDNA (Encode NA18507) and spiked in PCR targets of the BRCA1 locus at a concentration of a single genome copy. We introduced 5fC or 5hmC (as the structurally most similar off-target nucleobase) at the BRCA1a target position via synthetic primers (SI Fig. 13-17). For enrichments, we reacted 1 µg DNA with 10 µmol linker L4 and 2 µmol *p*-PDA in 20 µL phosphate buffer as above, and subsequently incubated the DNA with 10 pmol TALE_G*_11→**1** and 0.5 eq. *p*-PDA in 30 µL binding buffer. We quenched the reaction with 18.5 µL Tris buffer (pH = 9.0), purified it via gel filtration and incubated it with Ni-NTA magnetic beads in enrichment buffer. We heated the washed beads in water and used the supernatant for qPCRs targeting a sequence adjacent to the BRCA1 TALE binding site (Fig. 4a).

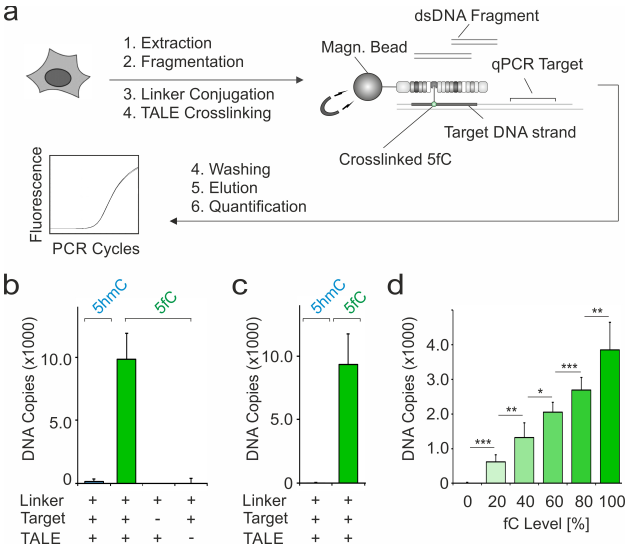

Figure 4. Covalent enrichment of 5fC-containing DNA sequences from mammalian genomic DNA backgrounds. a.) Assay workflow. b.) qPCR quantification of enrichments with TALE_G*_11→**1** and DNA with 100 % 5fC or 5hmC levels in the BRCA1 target CpG, including basic controls. A background control without linker was subtracted from other enrichments. Enrichments were conducted in triplicates, each quantified with triplicate qPCRs. c.) qPCR quantification of enrichments with TALE_G*_11→**1** performed as in 4b, but with background of mESC gDNA. d.) Enrichments as in Fig. 4b with different 5fC levels. * = $p < 0.005$. ** = $p < 0.0005$. *** = $p < 0.00005$ from student-t-test. R^2 from linear regression = 0.99.

We observed effective target enrichment with 5fC-containing gDNA, but not with 5hmC-containing DNA.

Moreover, no 5fC-target was enriched in absence of linker L4, confirming covalent enrichment (Fig. 4b). We also found that selective target enrichment was not affected by the presence of 5fC in the background gDNA by analogous experiments with non-amplified mESC gDNA extracted from J1 cells (Fig. 4c). We next asked, if this assay would allow for quantification of 5fC levels. We repeated the above experiment with mixed 5fC and 5hmC targets to obtain defined 5fC levels and observed a linear correlation between levels and number of enriched target copies (Fig. 4d). This shows the selective quantification of 5fC levels at a single, user-defined CpG in a human genome background by TALE crosslinking.

In conclusion, we advanced the direct, programmable targeting of epigenetic DNA nucleobases with TALEs from recognition to chemoselective crosslinking. By genetically encoding the ketone-bearing ncAA **1**, we created TALE repeats that universally recognize all five mammalian cytosine nucleobases, but readily react in linker-mediated, 5fC-selective oxime condensations for covalent capture of target DNA. This overcomes the lack of 5fC selectivity and sensitivity of previous, solely recognition-based TALE repeats. Successful transfer to a simple affinity enrichment assay enables the selective and sensitive quantification of 5fC at a single CpG site in a human genome background. Besides selective condensation reactions at aldehydes, conjugation chemistries are available for primary alcohols and carboxylic acids that are present in 5hmC and 5caC. Suited functional groups are available in canonical and noncanonical amino acids, providing the opportunity for an extension of the approach. Taken together, our strategy provides a new avenue to advance the sensitivity and selectivity of probes for programmable DNA targeting beyond A, G, T and C, and thus expand their application scope for epigenetics studies.

ASSOCIATED CONTENT

Supporting Information. Synthesis procedures, analytical data, DNA/protein sequences and biochemical procedures and data. This material is available free of charge via the Internet at http://pubs.acs.org."

AUTHOR INFORMATION

Corresponding Author

*daniel.summerer@tu-dortmund.de

Funding Sources

No competing financial interests. This work was funded by the DFG (Priority Program SPP1623) and the ERC (Grant EPICODE).

ACKNOWLEDGMENT

We thank A. J. Bogdanove and D. F. Voytas for TALE assembly plasmids obtained via Addgene. We thank Peter G. Schultz for plasmid pEVOL-pAcF. Structures in Fig. 1-2 have been prepared with PyMOL, Version 1.7.4 Schrödinger, LLC.

REFERENCES

1. Wu, X. J.; Zhang, Y., TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* **2017,** *18* (9), 517-534.
2. 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* (5929), 930-5.
3. Kriaucionis, S.; Heintz, N., The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **2009,** *324* (5929), 929-30.
4. Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y., Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **2011,** *333* (6047), 1300-3.
5. He, Y. F.; Li, B. Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; Sun, Y.; Li, X.; Dai, Q.; Song, C. X.; Zhang, K.; He, C.; Xu, G. L., Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **2011,** *333* (6047), 1303-7.
6. Pfaffeneder, T.; Hackner, B.; Truss, M.; Munzel, M.; Muller, M.; Deiml, C. A.; Hagemeier, C.; Carell, T., The Discovery of 5-Formylcytosine in Embryonic Stem Cell DNA. *Angew. Chem. Int. Ed. Engl.* **2011,** *50* (31), 7008-7012.
7. Dietzsch, J.; Feineis, D.; Hobartner, C., Chemoselective labeling and site-specific mapping of 5-formylcytosine as a cellular nucleic acid modification. *FEBS Lett.* **2018,** *592* (12), 2032-2047.
8. Berney, M.; McGouran, J. F., Methods for detection of cytosine and thymine modifications in DNA. *Nat. Rev. Chem.* **2018,** *2* (11), 332-348.
9. Spruijt, C. G.; Gnerlich, F.; Smits, A. H.; Pfaffeneder, T.; Jansen, P. W.; Bauer, C.; Munzel, M.; Wagner, M.; Muller, M.; Khan, F.; Eberl, H. C.; Mensinga, A.; Brinkman, A. B.; Lephikov, K.; Muller, U.; Walter, J.; Boelens, R.; van Ingen, H.; Leonhardt, H.; Carell, T.; Vermeulen, M., Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **2013,** *152* (5), 1146-59.
10. Du, Q.; Luu, P. L.; Stirzaker, C.; Clark, S. J., Methyl-CpG-binding domain proteins: readers of the epigenome. *Epigenomics* **2015,** *7* (6), 1051-73.
11. Kellinger, M. W.; Song, C. X.; Chong, J.; Lu, X. Y.; He, C.; Wang, D., 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **2012,** *19* (8), 831-3.
12. Ngo, T. T.; Yoo, J.; Dai, Q.; Zhang, Q.; He, C.; Aksimentiev, A.; Ha, T., Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat. Commun.* **2016,** *7*, 10813.
13. Li, F. C.; Zhang, Y. Q.; Bai, J.; Greenberg, M. M.; Xi, Z.; Zhou, C. Z., 5-Formylcytosine Yields DNA-Protein Cross-Links in Nucleosome Core Particles. *J. Am. Chem. Soc.* **2017,** *139* (31), 10617-10620.
14. Ji, S. F.; Shao, H. Z.; Han, Q. Y.; Seiler, C. L.; Tretyakova, N. Y., Reversible DNA-Protein Cross-Linking at Epigenetic DNA Marks. *Angew. Chem. Int. Ed. Engl.* **2017,** *56* (45), 14130-14134.
15. Raiber, E. A.; Portella, G.; Cuesta, S. M.; Hardisty, R.; Murat, P.; Li, Z.; Iurlaro, M.; Dean, W.; Spindel, J.; Beraldi, D.; Liu, Z.; Dawson, M. A.; Reik, W.; Balasubramanian, S., 5-Formylcytosine organizes nucleosomes and forms Schiff base interactions with histones in mouse embryonic stem cells. *Nat. Chem.* **2018,** *10* (12), 1258-1266.
16. Song, C. X.; Szulwach, K. E.; Dai, Q.; Fu, Y.; Mao, S. Q.; Lin, L.; Street, C.; Li, Y.; Poidevin, M.; Wu, H.; Gao, J.; Liu, P.; Li, L.; Xu, G. L.; Jin, P.; He, C., Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **2013,** *153* (3), 678-91.
17. Booth, M. J.; Marsico, G.; Bachman, M.; Beraldi, D.; Balasubramanian, S., Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. *Nat. Chem.* **2014,** *6* (5), 435-40.
18. Xia, B.; Han, D.; Lu, X.; Sun, Z.; Zhou, A.; Yin, Q.; Zeng, H.; Liu, M.; Jiang, X.; Xie, W.; He, C.; Yi, C., Bisulfite-free, base-resolution analysis of 5-formylcytosine at the genome scale. *Nat. Methods* **2015,** *12* (11), 1047-50.
19. Wang, S. R.; Song, Y. Y.; Wei, L.; Liu, C. X.; Fu, B. S.; Wang, J. Q.; Yang, X. R.; Liu, Y. N.; Liu, S. M.; Tian, T.; Zhou, X., Cucurbit[7]uril-Driven Host-Guest Chemistry for Reversible Intervention of 5-Formylcytosine-Targeted Biochemical Reactions. *J. Am. Chem. Soc.* **2017,** *139* (46), 16903-16912.
20. Su, M.; Kirchner, A.; Stazzoni, S.; Muller, M.; Wagner, M.; Schroder, A.; Carell, T., 5-Formylcytosine Could Be a Semipermanent Base in Specific Genome Sites. *Angew. Chem. Int. Ed. Engl.* **2016,** *55* (39), 11797-800.
21. Guo, P.; Yan, S.; Hu, J.; Xing, X.; Wang, C.; Xu, X.; Qiu, X.; Ma, W.; Lu, C.; Weng, X.; Zhou, X., Selective detection of 5-formyl-2'-deoxycytidine in DNA using a fluorogenic hydroxylamine reagent. *Org. Lett.* **2013,** *15* (13), 3266-9.
22. Samanta, B.; Seikowski, J.; Hobartner, C., Fluorogenic Labeling of 5-Formylpyrimidine Nucleotides in DNA and RNA. *Angew. Chem. Int. Ed. Engl.* **2016,** *55* (5), 1912-6.

23. Deng, D.; Yan, C.; Pan, X.; Mahfouz, M.; Wang, J.; Zhu, J. K.; Shi, Y.; Yan, N., Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* **2012,** *335* (6069), 720-3.
24. Boch, J.; Bonas, U., Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu. Rev. Phytopath.* **2010,** *48*, 419-36.
25. Mak, A. N. S.; Bradley, P.; Cernadas, R. A.; Bogdanove, A. J.; Stoddard, B. L., The Crystal Structure of TAL Effector PthXo1 Bound to Its DNA Target. *Science* **2012,** *335* (6069), 716-719.
26. Giess, M.; Witte, A.; Jasper, J.; Koch, O.; Summerer, D., Complete, Programmable Decoding of Oxidized 5-Methylcytosine Nucleobases in DNA by Chemoselective Blockage of Universal Transcription-Activator-Like Effector Repeats. *J. Am. Chem. Soc.* **2018,** *140* (18), 5904-5908.
27. Kubik, G.; Batke, S.; Summerer, D., Programmable sensors of 5-hydroxymethylcytosine. *J. Am. Chem. Soc.* **2015,** *137* (1), 2-5.
28. Maurer, S.; Giess, M.; Koch, O.; Summerer, D., Interrogating Key Positions of Size-Reduced TALE Repeats Reveals a Programmable Sensor of 5-Carboxylcytosine. *ACS Chem. Biol.* **2016,** *11* (12), 3294-3299.
29. Tsuji, S.; Futaki, S.; Imanishi, M., Sequence-specific recognition of methylated DNA by an engineered transcription activator-like effector protein. *Chem. Commun. (Camb)* **2016,** *52* (99), 14238-14241.
30. Zhang, Y.; Liu, L.; Guo, S.; Song, J.; Zhu, C.; Yue, Z.; Wei, W.; Yi, C., Deciphering TAL effectors for 5-methylcytosine and 5-hydroxymethylcytosine recognition. *Nat. Commun.* **2017,** *8* (1), 901.
31. Maurer, S.; Buchmuller, B.; Ehrt, C.; Jasper, J.; Koch, O.; Summerer, D., Overcoming conservation in TALE–DNA interactions: a minimal repeat scaffold enables selective recognition of an oxidized 5-methylcytosine *Chem. Sci.* **2018,** *9*, 7247-7252.
32. Rathi, P.; Maurer, S.; Kubik, G.; Summerer, D., Isolation of Human Genomic DNA Sequences with Expanded Nucleobase Selectivity. *J. Am. Chem. Soc.* **2016,** *138* (31), 9910-8.
33. Cermak, T.; Doyle, E. L.; Christian, M.; Wang, L.; Zhang, Y.; Schmidt, C.; Baller, J. A.; Somia, N. V.; Bogdanove, A. J.; Voytas, D. F., Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting (vol 39, pg e82, 2011). *Nucleic Acids Res.* **2011,** *39* (17), 7879-7879.
34. Wang, L.; Zhang, Z. W.; Brock, A.; Schultz, P. G., Addition of the keto functional group to the genetic code of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **2003,** *100* (1), 56-61.
35. Holder, P. G.; Francis, M. B., Integration of a self-assembling protein scaffold with water-soluble single-walled carbon nanotubes. *Angew. Chem. Int. Ed. Engl.* **2007,** *46* (23), 4370-3.

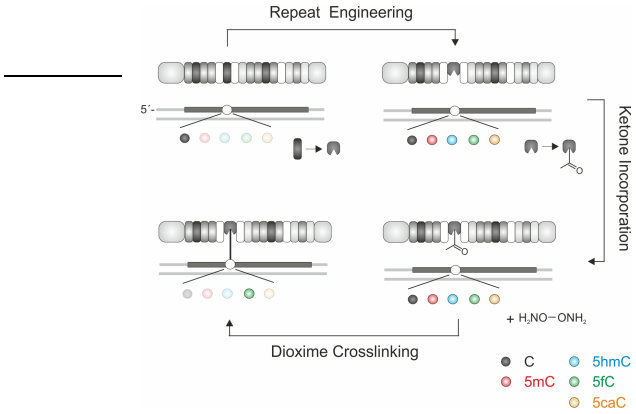

Table of Contents Artwork