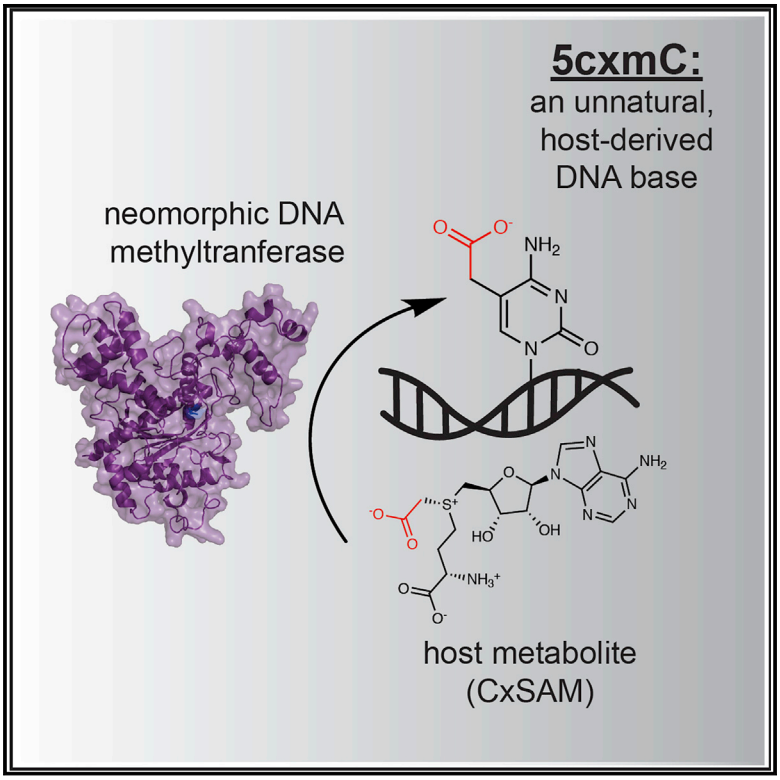
Brief Communication

Cell Chemical Biology

Discovery of an Unnatural DNA Modification Derived from a Natural Secondary Metabolite

Graphical Abstract



Authors

Tong Wang, Rahul M. Kohli

Correspondence

rkohli@pennmedicine.upenn.edu

In Brief

In Wang and Kohli, 5-carboxymethylcytosine is revealed as an unnatural genomic DNA modification that can be generated *in vivo* from a unique combination of a sparse natural host metabolite and a neomorphic cytosine DNA methyltransferase.

Highlights

- Mutagenesis unexpectedly identifies a neomorphic activity in a DNA methyltransferase
- Bacterial genetics uncovers the modification 5-carboxymethylcytosine (5cxmC)
- Structure-guided biochemistry reveals enzymatic basis for DNA carboxymethylation
- 5cxmC provides a model for how gain-of-function genomic bases may arise in nature

Wang & Kohli, 2021, Cell Chemical Biology *28*, 97–104
January 21, 2021
https://doi.org/10.1016/j.chembiol.2020.09.006

CellPress

Cell Chemical Biology

CellPress

Brief Communication

Discovery of an Unnatural DNA Modification Derived from a Natural Secondary Metabolite

Tong Wang[1] and Rahul M. Kohli[2,3,4,*]

[1]Graduate Group in Biochemistry and Molecular Biophysics, University of Pennsylvania, Philadelphia, PA, USA

[2]Department of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[3]Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA, USA

[4]Lead Contact

*Correspondence: rkohli@pennmedicine.upenn.edu

https://doi.org/10.1016/j.chembiol.2020.09.006

SUMMARY

Despite widespread interest for understanding how modified bases have evolved their contemporary functions, limited experimental evidence exists for measuring how close an organism is to accidentally creating a new, modified base within the framework of its existing genome. Here, we describe the biochemical and structural basis for how a single-point mutation in *E. coli*'s naturally occurring cytosine methyltransferase can surprisingly endow a neomorphic ability to create the unnatural DNA base, 5-carboxymethylcytosine (5cxmC), *in vivo*. Mass spectrometry, bacterial genetics, and structure-guided biochemistry reveal this base to be exclusively derived from the natural but sparse secondary metabolite carboxy-*S*-adenosyl-L-methionine (CxSAM). Our discovery of a new, unnatural DNA modification reveals insights into the substrate selectivity of DNA methyltransferase enzymes, offers a promising new biotechnological tool for the characterization of the mammalian epigenome, and provides an unexpected model for how neomorphic bases could arise in nature from repurposed host metabolites.

INTRODUCTION

Manipulation of genomes to accommodate non-canonical nucleobases creates opportunities for increasing DNA coding capacity and offers insights into the requirements for sustaining life (Krueger and Kool, 2009). Significant progress has been made in the propagation of unnatural nucleobases within the *E. coli* genome (Malyshev et al., 2014). Similarly, non-native but naturally occurring nucleobases, such as those derived from bacteriophages, have been introduced into *E. coli* (Mehta et al., 2016a, 2016b). Both strategies rely on the manipulation of nucleoside triphosphate (dNTP) pools. Although dNTP manipulation is one way to imbue the genome with chemical diversity, marks covalently introduced after DNA replication provide an alternative route to encoding additional complexity in DNA, a strategy akin to nature's existing epigenetic bases. To our knowledge, there have been no previous reports describing an organism with an unnatural DNA modification derived from its own native metabolome, a finding that may show the feasibility for an organism to accidentally accrue a gain-of-function ability to create a new DNA base *in vivo*.

Within the natural realm, an array of different DNA modifications have been described, but the vast majority of this diversity is confined to bacteriophage genomes and their prokaryotic hosts. Modifications to all canonical nucleobases have been described in phages (Weigele and Raleigh, 2016). In prokaryotes, the predominant modifications are found at the *N*6 position of adenine and *N*4 or C5 positions of cytosine. Methylation of these bases primarily serves rudimentary functions to distinguish self from non-self in the arms race against bacteriophages (Nabel et al., 2012; Wilson and Murray, 1991), although emerging models suggest that some modifications may affect genome regulation (Sánchez-Romero and Casadesús, 2020).

5-Methylcytosine (5mC) is a genomic DNA modification that extends from prokaryotes to higher organisms. Phylogenetic evidence shows that DNA cytosine methyltransferases (MTases), the enzymes that make 5mC, are conserved from prokaryotic restriction-modification systems to eukaryotic gene regulatory machinery (Iyer et al., 2011). In mammals, 5mC generation is largely confined to cytosine-guanine (CpG) dinucleotides (Portela and Esteller, 2010). Adding further complexity to this model, 5mC is a substrate for the ten-eleven translocation (TET) family enzymes, which iteratively oxidize 5mC to create 5-hydroxymethyl-, 5-formyl-, and 5-carboxylcytosine (He et al., 2011; Ito et al., 2011; Tahiliani et al., 2009). While predominantly implicated as intermediates toward 5mC erasure, the potential independent epigenetic identities of each oxidized 5mC base is the subject of numerous provocative hypotheses (Bilyard et al., 2020). Across phylogeny, there is therefore compelling evidence for a functional role for diverse DNA modifications, providing the motivation for understanding how new DNA modifications can arise.

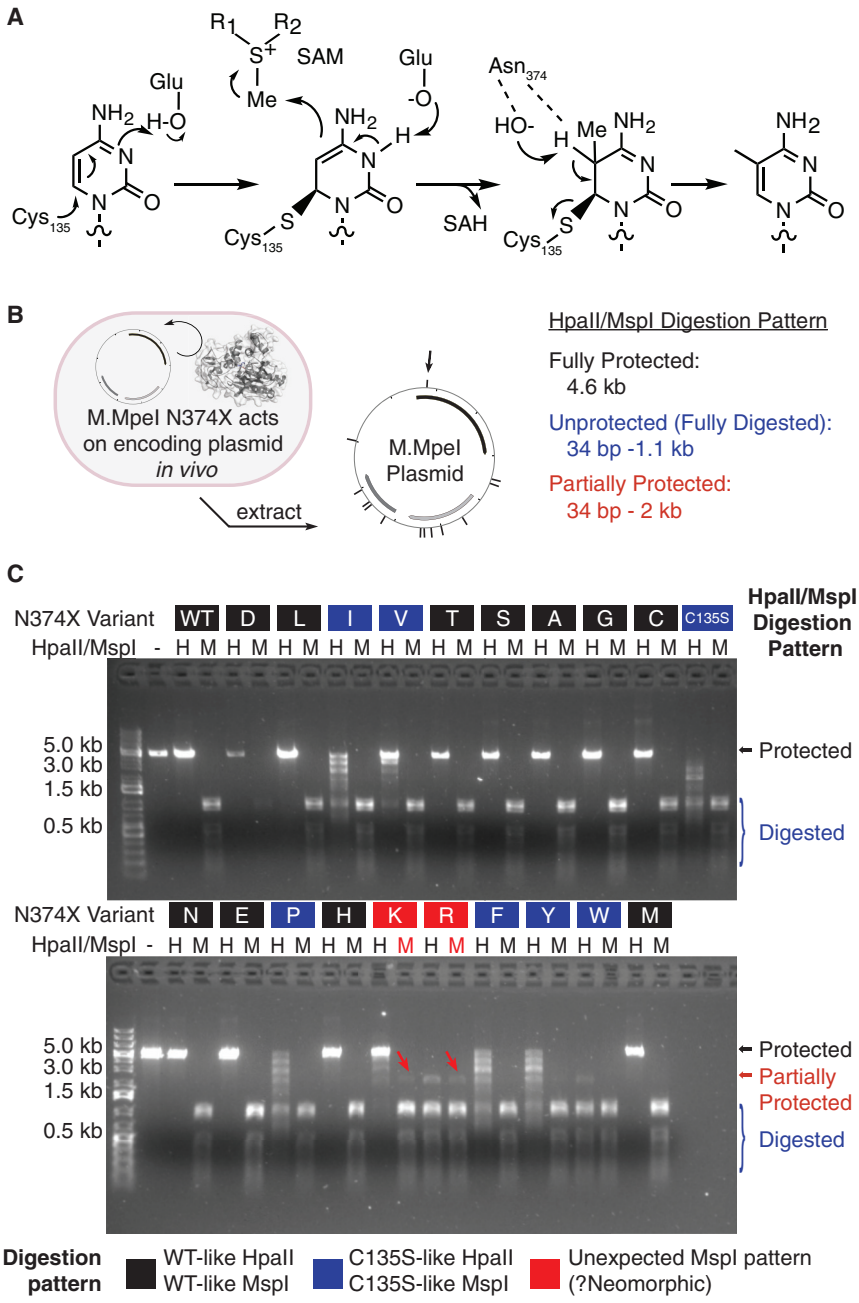


Figure 1. Saturation Mutagenesis of M.MpeI N374X Shows Activity Inconsistent with DNA Cytosine Methylation

(A) Mechanism of DNA cytosine methyltransferases.

(B) Experimental design. Individual M.MpeI N374X constructs were transformed and expressed in *E. coli*. *In vivo* modification was detected by restriction digestion of purified plasmids. The C**C**GG sites recognized by HpaII or MspI are visualized by hash marks and include CpGs that could be modified. Partial resistance to cleavage at the site marked with a black arrow above the plasmid can result in the generation of a ~2-kb fragment highlighted with the red arrows on the gel.

(C) HpaII ("H," modification-sensitive) and MspI ("M," methylation-insensitive) digestion. Each N374X mutant digestion pattern is WT-like, C135S catalytic mutant-like, or potentially neomorphic. The red arrows highlight MspI digestion bands inconsistent with methylation.

RESULTS

Saturation Mutagenesis of a CpG MTase Reveals a New DNA Modification

Given our interest in developing tools to understand mammalian modifications in the CpG context, we began by examining the crystal structure of a recently described bacterial CpG methyltransferase, M.MpeI (Wojciechowski et al., 2013). M.MpeI uses a canonical cytosine DNA MTase mechanism to make 5mC from *S*-adenosyl-L-methionine (SAM) and cytosine (Figure 1A). We focused on Asn374 of M.MpeI to assimilate two competing observations from the literature. The Asn side chain, which is heavily conserved across MTases, has been proposed to act as part of a network of hydrogen bonds with active site water molecules to promote the final elimination step for 5mC generation (Jurkowski and Jeltsch, 2011; Zhang and Bruice, 2006). Despite this model, however, mutation of this Asn to Ala is tolerated in homologous MTases and permits transfer of bulky SAM analogs *in vitro* (Lukinavicius et al., 2012). Thus, we pursued saturation mutagenesis of N374 as an unbiased way to understand its core role in MTase catalysis.

We performed an *in vivo* activity screen that relies upon the linkage of the M.MpeI mutant genotype with a cytosine-methylating phenotype. We separately transformed each of the 20 N374X variants, along with a C135S catalytic mutant, into *E. coli*. After inducing expression, the plasmids were recovered and assessed for the ability of each MTase to modify its own encoding plasmid *in vivo* (Figure 1B). Extracted plasmids were digested with one of two C**C**GG-recognizing restriction enzymes,

Here, we describe how our efforts to understand the structural basis of 5mC generation by the CpG MTase M.MpeI unexpectedly led to the identification of the previously undescribed DNA modification, 5-carboxymethylcytosine (5cxmC), which originates from the naturally occurring, trace secondary metabolite carboxy-*S*-adenosyl-L-methionine (CxSAM). Additional biochemistry and structure-guided rationalization led to the accurate prediction that *E. coli*'s native MTase, Dcm, only requires a single-point mutation to generate 5cxmC as a metabolite-derived, unnatural nucleobase in genomic DNA *in vivo*, offering a model for how new modifications can arise in nature.

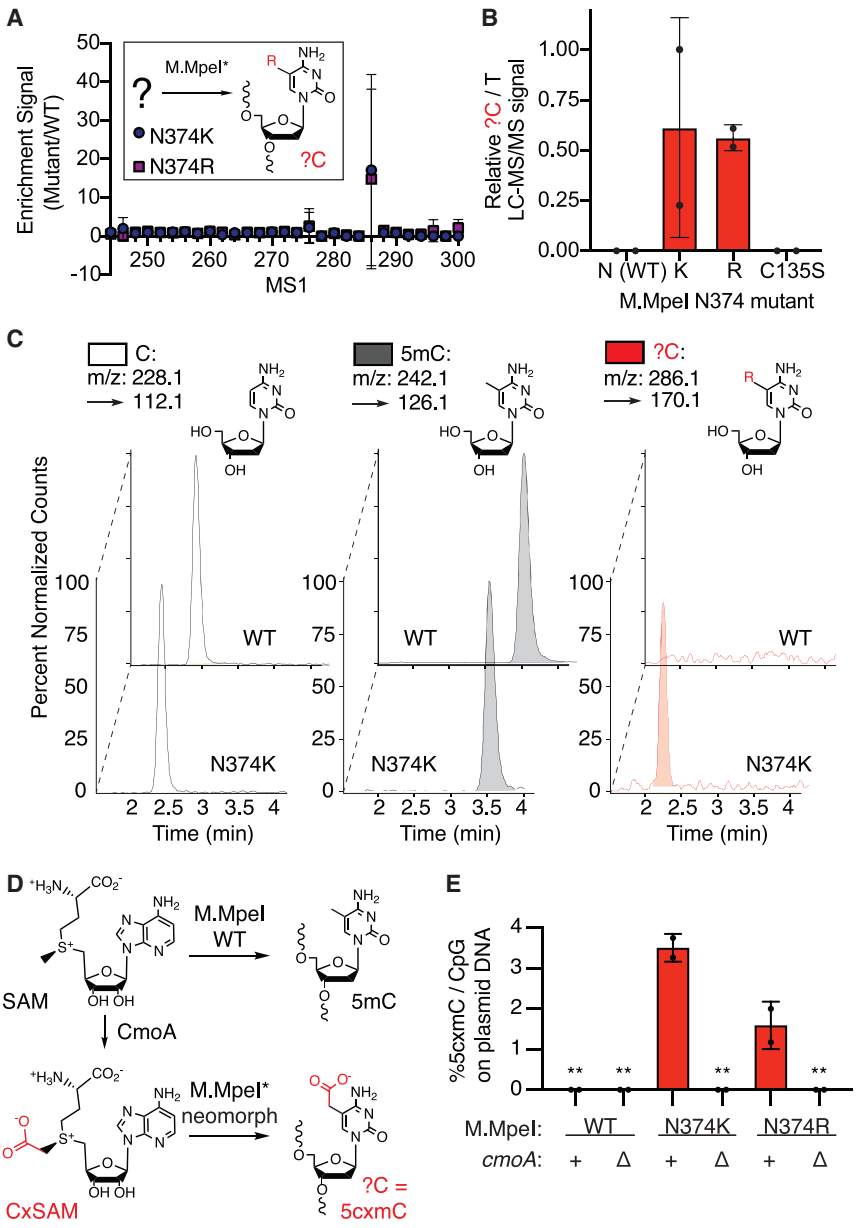


Figure 2. M.MpeI Mutants Create 5cxmC *In Vivo*

(A–C) Nucleoside LC-MS/MS of plasmid DNA modified *in vivo* by M.MpeI mutants. (A) Scanning mode identifies a candidate mass. Listed are the parent ion (MS1) *m/z* $[M + H]^+$ associated with the fragment ion (MS2) $[M + H - dR]^+$ resulting from loss of the deoxyribose sugar. (B) Quantitative detection normalized to input (T), expressed relative to maximum signal observed. (C) New modification shows a distinct retention time and detectable *m/z* in N374K but not WT M.MpeI plasmids. Peaks are normalized for maximal detection. Panels (A–C) show the same samples (n = 2 biological replicates). Error bars represent SD.

(D) CxSAM synthesized by CmoA *in vivo* provides a putative substrate for carboxymethylation.

(E) 5cxmC is derived exclusively from CxSAM as assessed by *cmoA* knockout (Δ). Shown is the % 5cxmC relative to total CpGs. Graphs show mean ± SD (n = 2 biological replicates). **Below limit of detection.

HpaII and MspI. HpaII is methylation sensitive and blocked by any covalent modification at the 5-position of the underlined cytosine. Its isoschizomer MspI is methylation insensitive and was intended to serve as a positive control for methylation (Figure 1C).

In our *in vivo* screen, both the HpaII and MspI digestion patterns were similar to wild-type (WT) M.MpeI for most of our variants, suggesting that conversion to C**5mC**GG was achieved across all plasmid sites. This finding suggests that, contrary to its postulated mechanistic role, N374 is generally tolerant to variation, although our saturating *in vivo* overexpression conditions and indirect assay may overestimate tolerance to mutagenesis. Nonetheless, partial protection from HpaII digestion, suggesting impaired catalysis, was observed with hydrophobic β-branched (Ile/Val), constrained (Pro), or bulky aromatic (Phe/Tyr/Trp) mutations at position N374. Most notably and unexpectedly, in both positively charged variants, N374K and N374R, there emerged a faint ~2-kb band resistant to MspI digestion, inconsistent with cytosine methylation (Figure 1C, red arrows). Upon reexamination of the plasmid map, we considered the possibility that if the CpG dinucleotides in the plasmid were to be randomly and partially modified to something other than 5mC *in vivo*, a C**C**GG protection event at a position between two larger fragments normally generated by MspI digestion (Figure 1B, black arrow) could account for this new ~2-kb band (red arrows, Figure 1C, red).

Identification of 5cxmC

While MspI cleaves at C**5mC**GG, it is blocked by bulkier modifications, such as the oxidized 5mCs (Liu et al., 2016). To explore the possibility that we were detecting an unexpected DNA modification, we degraded each plasmid to its individual nucleosides and performed liquid chromatography-tandem mass spectrometry (LC-MS/MS), scanning for nucleoside masses larger than 5mC (Figure 2A). In the N374K and N374R mutants but not WT or C135S, we identified a peak with a unique retention time of 2.2 min and *m/z* 286.1 $[M + H]^+ \rightarrow 170.1$ $[M + H - dR]^+$ (Figures 2B and 2C).

Upon literature review, we identified carboxy-*S*-adenosyl-L-methionine (CxSAM) as a candidate MTase substrate worth further investigation. CxSAM is a sparse secondary metabolite in *E. coli* generated from SAM and prephenate by the non-essential enzyme CxSAM synthase (CmoA) and has recently been shown to be involved in tRNA modifications of uridine in *E. coli* (Kim et al., 2013). Although CxSAM is 400-fold less prevalent than SAM *in vivo* (Kim et al., 2015), we noted that the reaction

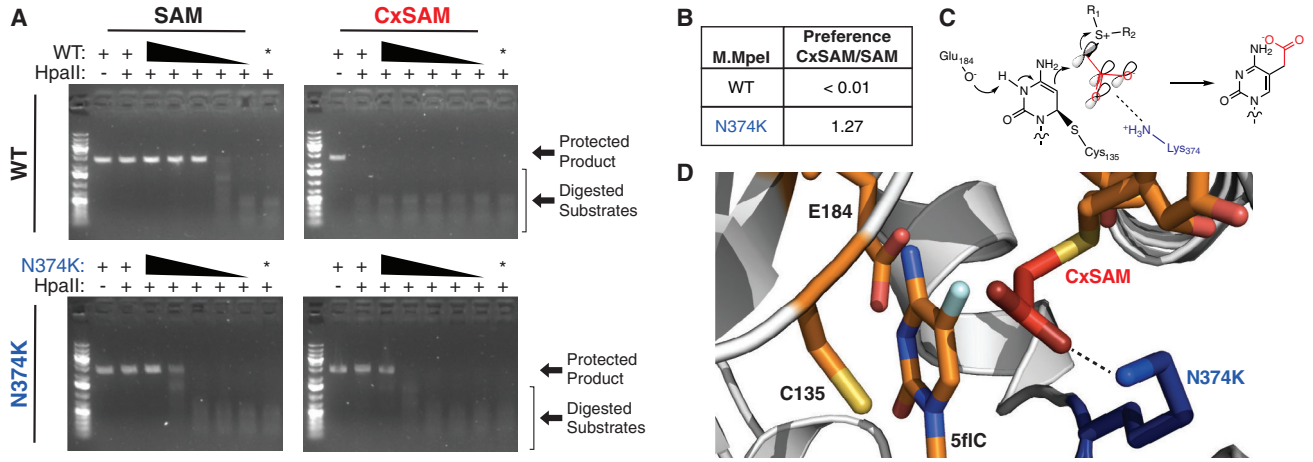


Figure 3. M.MpeI N374K Creates 5cxmC *In Vitro*

(A) Unmodified pUC19 plasmid DNA was incubated with an excess of SAM or CxSAM and serial dilutions of M.MpeI WT or N374K. The negative control lane (*) contains the highest concentration of enzyme with no SAM or CxSAM substrate. Digestion with HpaII fragments only unmodified DNA. M.MpeI N374K transfers both CxSAM and SAM *in vitro* while M.MpeI WT only transfers SAM.

(B) M.MpeI N374K prefers CxSAM over SAM in a quantitative oligonucleotide assay shown in Figure S3 (n = 3 independent replicates).

(C) Mechanism of DNA carboxymethylation visualizing a π system, which is favorable for CxSAM electrophilicity. Adjacent to catalytic residues E184 and C135, N374K (blue) could form a gain-of-function salt bridge (dashed line) with the carboxylate (red) of CxSAM.

(D) Structural model of M.MpeI with same elements as in (C). The model was obtained by the introduction of an N374K mutation and manually overlaying CxSAM from PDB: 4QNV in place of SAH in the 5-fluorocytosine (5flC)-bound M.MpeI structure (PDB: 4DKJ).

of CxSAM with a target cytosine would yield 5-carboxymethylcytosine (5cxmC) (Figure 2D), a bulky modification that could resist MspI digestion. The nucleoside LC-MS/MS signal could also be consistent with this possibility: it has a more polar retention time relative to cytosine and 5mC, and the parent and fragmentation mass patterns fit with the addition of a carboxymethyl group.

To rigorously assess if CxSAM was in fact the substrate for our mutant MTase *in vivo*, we generated a *cmoA* knockout strain (Figures S1A–S1C). While *in vivo* plasmid carboxymethylation by M.MpeI N374K and N374R can be detected in the *cmoA*$^+$ *E. coli* strain by LC-MS/MS, these signals are lost in the *ΔcmoA* strain (Figures 2D, S1D, and S2E). We also expressed M.MpeI WT, C135S, N374K, and a double mutant C135S N374K and showed that only the combination of M.MpeI N374K in *cmoA*$^+$ *E. coli* could create an MspI-resistant modification (Figure S1F). Thus, 5cxmC is generated via a similar mechanism to 5mC and is solely derived from the activity of mutant M.MpeI utilizing endogenous CxSAM as a substrate.

Validation of DNA Carboxymethyltransferase Activity

Having established the identity and origin of the 5cxmC base, we aimed to reproduce this activity *in vitro*. We purified both the WT and N374K M.MpeI variants (Figure S2A) and synthesized CxSAM as a diastereomeric mixture (Figures S2B–S2D). We then incubated the enzyme with a pUC19 plasmid DNA substrate and either SAM or CxSAM. The plasmids were subsequently digested with modification-sensitive HpaII (C**C**GG, 13 sites) (Figure 3A). Consistent with our *in vivo* analysis, WT M.MpeI is capable of completely protecting the pUC19 plasmid with SAM but not CxSAM. M.MpeI N374K, by contrast, transfers SAM less efficiently than the WT enzyme, but gains the neomorphic ability to transfer CxSAM.

For a more quantitative comparison of *in vitro* activity, we devised an oligonucleotide-based assay, whereby modification of a CpG on a fluorophore-labeled strand can be tracked by monitoring its resistance to HpaII digestion (Figure S3A). Consistent with the previous pUC19-based assay, we found that for WT M.MpeI, only SAM and not CxSAM was a substrate for cytosine modification. For the N374K variant, CxSAM was 1.3-fold preferred over SAM (Figures 3B, S3B, and S3C). While our *in vitro* studies show that N374K M.MpeI has a modest preference for CxSAM over SAM, our *in vivo* experiments suggest that the oligonucleotide assay may underestimate the extent of this preference, possibly due to our inability to separate CxSAM diastereomers or other factors that enhance CxSAM selectivity *in vivo*.

Structural Model Accurately Predicts Neomorphic Carboxymethylation by a Mutant *E. coli* MTase

Previous work with synthetic SAM analogs has suggested that transfer can be promoted by the presence of a conjugated π system at the β-carbon relative to the electrophilic carbon (Figure 3C) (Dalhoff et al., 2006). This mechanism alone, however, cannot explain why our mutant and WT M.MpeI behave differently toward the same CxSAM substrate. We thus turned to the crystal structure of M.MpeI with the *S*-adenosyl-L-homocysteine (SAH) product bound and manually overlaid CxSAM in place of SAH (Figure 3D). In the model, we observed that a mutant Lys374 is poised to form a potential salt bridge with the carboxylate anion of CxSAM, offering a likely explanation for this enzyme's ability to accept this substrate.

Given this structural model for cytosine carboxymethylation, we wondered if this neomorphic activity was also accessible for homologous MTases. We specifically chose to focus on *E. coli*'s naturally occurring DNA cytosine methyltransferase

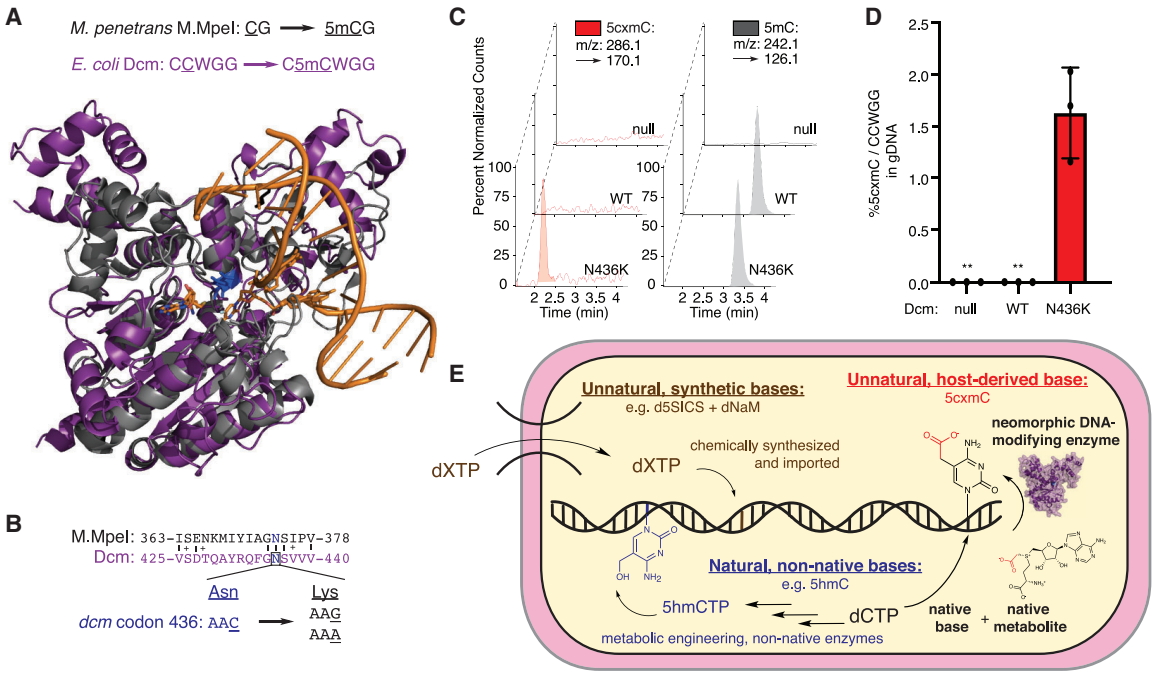


Figure 4. *E. coli*'s Native Methyltransferase Dcm Is a Single Point Mutation Away from Accruing a Gain-of-Function Ability to Carboxymethylate Genomic DNA *In Vivo*, Redefining the Boundary between Unnatural and Natural Nucleobases

(A) Alignment of Dcm (purple, Swiss-Model: P0AED9) with M.MpeI (gray, PDB: 4DKJ).

(B and C) (B) Asn436 codon. (C) Qualitative nucleoside LC-MS/MS showing that only Dcm N436K can carboxymethylate *E. coli* genomic DNA *in vivo*. Peaks are normalized for maximal detection.

(D) Quantitative 5cxmC LC-MS/MS signal in Dcm mutants. Shown is the %5cxmC relative to total CCWGGs (W = A/T) with each Dcm variant (null = no plasmid). Error bars represent mean ± SD (n = 3 biological replicates). **Below limit of detection.

(E). Multiple routes to expanded genomic diversity are shown. Brown: unnatural dXTPs can be imported and replicated *in vivo* (Malyshev et al., 2014). Blue: non-native nucleoside triphosphates (e.g., 5hmCTPs) can be incorporated into *E. coli* genomes using bacteriophage biosynthetic machinery (Mehta et al., 2016a, 2016b). Red: 5-carboxymethylcytosine (5cxmC) in DNA, synthesized by a neomorphic DNA-modifying enzyme, is a new, unnatural DNA base, derived from the native substrates cytosine and CxSAM, without dNTP manipulation.

(Dcm) because this enzyme allowed us to address the question: What does it take for an organism to make a new DNA modification, *in vivo*? While M.MpeI is native to *Mycoplasma penetrans* and generates 5mC in the CpG context, Dcm generates 5mC in C**C**WGG (W = A/G) contexts. Structural alignment showed that Asn436 of Dcm and Asn374 of M.MpeI are both positioned adjacent to C5 of the target cytosine (Figure 4A and 1A), and a simple transversion mutation in *dcm* could create an N436K mutation (Figure 4B).

Encouraged by our biochemical understanding of M.MpeI-mediated DNA carboxymethylation, we moved to *dam*$^-$/*dcm*$^-$ *E. coli* and introduced either Dcm WT or N436K on a plasmid. After induction of MTase expression, we extracted the genomic DNA (gDNA) and performed nucleoside LC-MS/MS (Figure 4C). In this setting, both the WT and N436K enzymes could methylate cytosine. However, only the N436K mutant enzyme could create 5cxmC in the native *E. coli* genome. Quantification of 5cxmC showed that >1.5% of the C**C**WGG sites were carboxymethylated (Figure 4D). Thus, as with M.MpeI, despite the excess of SAM over CxSAM, a substantial amount of 5cxmC can be detected *in vivo* (Figure S4A). In addition, tracking *E. coli* growth curves for the Dcm WT, N436K, or C177A catalytic mutants suggests that the presence of 5cxmC does not lead to any apparent deficiency in growth under these conditions (Figure S4B).

To further compare with the M.MpeI variants, we next purified the Dcm WT and N436K mutant enzymes and assayed for their ability to modify pUC19 plasmid DNA with either SAM or CxSAM using the modification-sensitive restriction enzyme EcoRII (C**C**WGG, seven sites). Consistent with our *in vivo* observations, WT Dcm can protect the pUC19 plasmid with SAM but not CxSAM. By contrast, Dcm N436K has comparable ability to transfer both CxSAM and SAM (Figures S4C and S4D). Given the extensive conservation of the active site Asn in homologous MTases (Jurkowski and Jeltsch, 2011), the consistent findings from M.MpeI to Dcm indicate that this residue may have neomorphic potential across the cytosine MTase family. Our results highlight that a single-point mutation in the native *dcm* coding sequence is sufficient to create an unnatural DNA modification in *E. coli*.

DISCUSSION

Here, we describe the discovery of a new genomic DNA base derived exclusively from the combination of a neomorphic methyltransferase and a native, but sparse, metabolite. Although not our original intent, the realization that our findings occupy a distinct space relative to similar, yet methodologically divergent, synthetic biology efforts has afforded us unique insights into

both technology development and the chemical determinants of genomic composition and evolution (Figure 4E).

Non-canonical nucleobases can originate from a variety of sources. While previous efforts have shown that synthetic and non-native sources of dNTPs can be used to create new bases *in vivo*, this study instead identifies the metabolome as an underappreciated source of genomic diversification. Outside of SAM, metabolites have been well documented to potentiate (Blaschke et al., 2013; Chen et al., 2013) or inhibit (Dang et al., 2009; Xu et al., 2016) the production of naturally occurring modified nucleobases, but only very rarely are they considered as substrates that can directly be used to modify genomic DNA. An interesting exception is provided by ascorbic acid (vitamin C), which was recently shown to be an unexpected co-substrate for generating the natural, modified base 5-glycerylmethylcytosine in the algae *Chlamydomonas reinhardtii* (Xue et al., 2019). In the case of CxSAM, while no role in DNA modification was previously known, the metabolite has been shown to act as a direct substrate for modified uridines in tRNA (Kim et al., 2013, 2015) and small-molecule cofactor modifications (Serebryakova et al., 2016). These precedents helped us to reveal that CxSAM can also be utilized by neomorphic, mutant DNA MTases. Our study thus highlights the possibility that rare metabolites similar to CxSAM, with other biological functions, could be redirected toward the creation of new, modified nucleic acids, in agreement with previous phylogenetic predictions that invoked modified SAM substrates as possible sources for undiscovered genomic diversity (Iyer et al., 2013).

Because CxSAM is a derivative of the parent metabolite SAM, substrate competition is a notable aspect of this study. This competition recapitulates a challenge faced in other synthetic biology efforts, outside of nucleic acids. For example, genetic code expansion can be limited by poor unnatural amino acid bioavailability and competition with endogenous elongation factor Tu (Chin, 2014). The experiments here establish that competition from a ubiquitous substrate (SAM) can be overcome. For technology development, this is an important observation because the synthetic SAM analog field continues to expand but is currently predominantly limited to *in vitro* settings. Given this study and others (Wang et al., 2013), it is now more feasible to consider whether SAM analogs with useful chemical handles can covalently modify DNA or proteins *in vivo*, despite their inevitable competition with native SAM.

As unnatural cytosines derived from SAM analogs have already been applied for *in vitro* sequencing applications (Kriukienė et al., 2013), the discovery of this DNA-modifying enzyme is also notable for its biotechnological potential. Here, we show that our enzyme can completely carboxymethylate a 2.6-kb plasmid substrate (Figure 3A). Given its proficiency, we speculate that our CpG DNA carboxymethyltransferase, which would selectively react with unmodified CpGs, can be leveraged to distinguish unmethylated, methylated, and oxidized CpGs when coupled to either third-generation or enzymatic methods for epigenetic sequencing (Figure S4E) (Schutsky et al., 2018).

Finally, we note that 5cxmC, but not 5mC, showed a gain-of-function ability to resist digestion by the modification-sensitive endonuclease MspI (Figures 1C, S1D, and S1F). Given the growing body of evidence that suggests that restriction-modification systems have the capacity to coevolve, it is feasible that selection focused on 5cxmC could be harnessed to improve the stability and abundance of 5cxmC modifications *in vivo* (Iyer et al., 2011; Sánchez-Romero and Casadesús, 2020). However, we caution that the levels of 5cxmC remain relatively low (~1.5% of possible sites in gDNA for Dcm N436K, Figure 4), and the neomorphic CxMTases not only can create 5cxmC from CxSAM but also have a diminished ability to create 5mC from SAM. Both properties have unknown and potentially independent effects on physiology that will require future investigation (Figure S4).

When considering the potential advantages of modified DNA bases, such as 5cxmC, it is particularly surprising that the answer to the question, "How close is an organism to accidentally procuring a new DNA modification?" may be only a point mutation. Interpreting this finding with knowledge of the extensive structural conservation within the cytosine MTase family and phylogenetic conservation of the enzyme CmoA illustrates how the boundary separating natural and unnatural DNA modifications is significantly more blurred than what might have been previously imagined.

SIGNIFICANCE

Efforts to manipulate and create synthetic organisms tolerant of unnatural or non-native nucleobases have collectively shown that life is remarkably accepting of new genomic chemistries. While these efforts have helped establish that exotic chemistries can be accommodated, here we describe and characterize a biochemical route for an unnatural DNA base to accidentally appear *in vivo*. We show that a single mutation in the active site of cytosine DNA methyltransferases confers neomorphic activity on these enzymes, allowing them to accept a sparse natural metabolite as a substrate and to thereby create 5-carboxymethylcytosine in genomic DNA. In addition to contributing a promising biotechnology for future exploration, our discovery of a host-derived, unnatural DNA modification provides evidence for the metabolome as a potentially rich and underappreciated source for accessing genomic diversity and offers a model for how new genomic DNA bases could arise in nature.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODELS AND SUBJECT DETAILS
  - *E. Coli* Strains
- METHOD DETAILS
  - Cloning
  - *In Vivo* Methyltransferase Assays
  - Nucleoside LC-MS/MS
  - Protein Purification

- CxSAM Synthesis
- Restriction Digest Based Assays
- pUC19 Assay
- Oligonucleotide Assay
- Protein Structures
- Growth Curves

- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.chembiol.2020.09.006.

ACKNOWLEDGMENTS

The authors thank Michael Cory, Juan Serrano, all members of the Kohli Laboratory, and Walraj Gosal (Cambridge Epigenetix, CEGX) for insightful discussions and feedback. In addition, we thank E. James Petersson for drawing our attention to the CxSAM metabolite, Mark Goulian and Kiran Gajula for assistance with bacteriology, Jamie DeNizio and Clementina Mesaros for providing mass spectrometry training and guidance, Uday Ghanty and Charles Ross III for CxSAM characterization support.

AUTHOR CONTRIBUTIONS

T.W. and R.M.K. designed the research and wrote the paper. T.W. performed all experiments.

DECLARATION OF INTERESTS

T.W. is supported by a Cambridge Epigenetix (CEGX) Training Fellowship. This work was supported by the NIH (R01-GM118501 and R01-HG010646 to R.M.K.). R.M.K. is supported by an Investigator in the Pathogenesis of Infectious Disease Award (Burroughs Wellcome Fund). Patent filing on the DNA carboxymethyltransferase is pending.

Received: April 28, 2020
Revised: August 19, 2020
Accepted: September 23, 2020
Published: October 13, 2020

REFERENCES

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. *2*, https://doi.org/10.1038/msb4100050.

Bilyard, M.K., Becker, S., and Balasubramanian, S. (2020). Natural, modified DNA bases. Curr. Opin. Chem. Biol. *57*, 1–7.

Blaschke, K., Ebata, K.T., Karimi, M.M., Zepeda-Martinez, J.A., Goyal, P., Mahapatra, S., Tam, A., Laird, D.J., Hirst, M., Rao, A., et al. (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. Nature *500*, 222–226.

Chen, J., Guo, L., Zhang, L., Wu, H., Yang, J., Liu, H., Wang, X., Hu, X., Gu, T., Zhou, Z., et al. (2013). Vitamin C modulates TET1 function during somatic cell reprogramming. Nat. Genet. *45*, 1504–1509.

Chin, J.W. (2014). Expanding and reprogramming the genetic code of cells and animals. Annu. Rev. Biochem. *83*, 379–408.

Dalhoff, C., Lukinavicius, G., Klimasauskas, S., and Weinhold, E. (2006). Direct transfer of extended groups from synthetic cofactors by DNA methyltransferases. Nat. Chem. Biol. *2*, 31–32.

Dang, L., White, D.W., Gross, S., Bennett, B.D., Bittinger, M.A., Driggers, E.M., Fantin, V.R., Jang, H.G., Jin, S., Keenan, M.C., et al. (2009). Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature *462*, 739–744.

DeNizio, J.E., Liu, M.Y., Leddin, E.M., Cisneros, G.A., and Kohli, R.M. (2019). Selectivity and promiscuity in TET-mediated oxidation of 5-methylcytosine in DNA and RNA. Biochemistry *58*, 411–421.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. PLoS One *3*, e3647.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303–1307.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300–1303.

Iyer, L.M., Abhiman, S., and Aravind, L. (2011). Natural history of eukaryotic DNA methylation systems. Prog. Mol. Biol. Transl. Sci. *101*, 25–104.

Iyer, L.M., Zhang, D., Burroughs, A.M., and Aravind, L. (2013). Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA. Nucleic Acids Res. *41*, 7635–7655.

Jurkowski, T.P., and Jeltsch, A. (2011). On the evolutionary origin of eukaryotic DNA methyltransferases and Dnmt2. PLoS One *6*, e28104.

Kim, J., Xiao, H., Bonanno, J.B., Kalyanaraman, C., Brown, S., Tang, X., Al-Obaidi, N.F., Patskovsky, Y., Babbitt, P.C., Jacobson, M.P., et al. (2013). Structure-guided discovery of the metabolite carboxy-SAM that modulates tRNA function. Nature *498*, 123–126.

Kim, J., Xiao, H., Koh, J., Wang, Y., Bonanno, J.B., Thomas, K., Babbitt, P.C., Brown, S., Lee, Y., and Almo, S.C. (2015). Determinants of the CmoB carboxymethyl transferase utilized for selective tRNA wobble modification. Nucleic Acids Res. *43*, 4602–4613.

Kriukienė, E., Labrie, V., Khare, T., Urbanavičiūtė, G., Lapinaitė, A., Koncevičius, K., Li, D., Wang, T., Pai, S., Ptak, C., et al. (2013). DNA unmethylome profiling by covalent capture of CpG sites. Nat. Commun. *4*, 2190.

Krueger, A.T., and Kool, E.T. (2009). Redesigning the architecture of the base pair: toward biochemical and biological function of new genetic sets. Chem. Biol. *16*, 242–248.

Kubiak, J.M., Culyba, M.J., Liu, M.Y., Mo, C.Y., Goulian, M., and Kohli, R.M. (2017). A small-molecule inducible synthetic circuit for control of the SOS gene network without DNA damage. ACS Synth. Biol. *6*, 2067–2076.

Liu, M.Y., DeNizio, J.E., and Kohli, R.M. (2016). Quantification of oxidized 5-methylcytosine bases and TET enzyme activity. Methods Enzymol. *573*, 365–385.

Lukinavicius, G., Lapinaite, A., Urbanaviciute, G., Gerasimaite, R., and Klimasauskas, S. (2012). Engineering the DNA cytosine-5 methyltransferase reaction for sequence-specific labeling of DNA. Nucleic Acids Res. *40*, 11594–11602.

Malyshev, D.A., Dhami, K., Lavergne, T., Chen, T., Dai, N., Foster, J.M., Corrêa, J., Ivan, R., and Romesberg, F.E. (2014). A semi-synthetic organism with an expanded genetic alphabet. Nature *509*, 385–388.

Mehta, A.P., Li, H., Reed, S.A., Supekova, L., Javahishvili, T., and Schultz, P.G. (2016a). Replacement of 2′-deoxycytidine by 2′-deoxycytidine analogues in the *E. coli* genome. J. Am. Chem. Soc. *138*, 14230–14233.

Mehta, A.P., Li, H., Reed, S.A., Supekova, L., Javahishvili, T., and Schultz, P.G. (2016b). Replacement of thymidine by a modified base in the *Escherichia coli* genome. J. Am. Chem. Soc. *138*, 7272–7275.

Miller, J.H. (1992). A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for *Escherichia coli* and Related Bacteria (Cold Spring Harbor Laboratory Press).

Mo, C.Y., Manning, S.A., Roggiani, M., Culyba, M.J., Samuels, A.N., Sniegowski, P.D., Goulian, M., and Kohli, R.M. (2016). Systematically altering bacterial SOS activity under stress reveals therapeutic strategies for potentiating antibiotics. mSphere *1*, e00163-16.

Nabel, C.S., DeNizio, J.E., Carroll, M., and Kohli, R.M. (2017). DNA methyltransferases demonstrate reduced activity against arabinosylcytosine: implications for epigenetic instability in AML. Biochemistry *56*, 2166–2169.

Nabel, C.S., Manning, S.A., and Kohli, R.M. (2012). The curious chemical biology of cytosine: deamination, methylation, and oxidation as modulators of genomic potential. ACS Chem. Biol. *7*, 20–30.

Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. Nat. Biotechnol. *28*, 1057–1068.

Sánchez-Romero, M.A., and Casadesús, J. (2020). The bacterial epigenome. Nat. Rev. Microbiol. *18*, 7–20.

Schutsky, E.K., DeNizio, J.E., Hu, P., Liu, M.Y., Nabel, C.S., Fabyanic, E.B., Hwang, Y., Bushman, F.D., Wu, H., and Kohli, R.M. (2018). Nondestructive, base-resolution sequencing of 5-hydroxymethylcytosine using a DNA deaminase. Nat. Biotechnol. *36*, 1083–1090.

Serebryakova, M., Tsibulskaya, D., Mokina, O., Kulikovsky, A., Nautiyal, M., Van Aerschot, A., Severinov, K., and Dubiley, S. (2016). A Trojan-horse peptide-carboxymethyl-cytidine antibiotic from *Bacillus amyloliquefaciens*. J. Am. Chem. Soc. *138*, 15690–15698.

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

Wang, R., Islam, K., Liu, Y., Zheng, W., Tang, H., Lailler, N., Blum, G., Deng, H., and Luo, M. (2013). Profiling genome-wide chromatin methylation with engineered posttranslation apparatus within living cells. J. Am. Chem. Soc. *135*, 1048–1056.

Weigele, P., and Raleigh, E.A. (2016). Biosynthesis and function of modified bases in bacteria and their viruses. Chem. Rev. *116*, 12655–12687.

Wilson, G.G., and Murray, N.E. (1991). Restriction and modification systems. Annu. Rev. Genet. *25*, 585–627.

Wojciechowski, M., Czapinska, H., and Bochtler, M. (2013). CpG underrepresentation and the bacterial CpG-specific DNA methyltransferase M.MpeI. Proc. Natl. Acad. Sci. U S A *110*, 105.

Xu, Q., Wang, K., Wang, L., Zhu, Y., Zhou, G., Xie, D., and Yang, Q. (2016). IDH1/2 mutants inhibit TET-promoted oxidation of RNA 5mC to 5hmC. PLoS One *11*, e0161261.

Xue, J.H., Chen, G.D., Hao, F., Chen, H., Fang, Z., Chen, F.F., Pang, B., Yang, Q.L., Wei, X., Fan, Q.Q., et al. (2019). A vitamin-C-derived DNA modification catalysed by an algal TET homologue. Nature *569*, 581–585.

Zhang, X., and Bruice, T.C. (2006). The mechanism of M.HhaI DNA C5 cytosine methyltransferase enzyme: a quantum mechanics/molecular mechanics approach. Proc. Natl. Acad. Sci. U S A *103*, 6148–6153.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
ER1821 *E. coli*	NEB	N/A
dcm⁻/dam⁻ E. coli	NEB	C2925I
ΔcmoA E. coli	KEIO, Baba et al. 2006	JW1859
ER1821 *ΔcmoA*	This study	N/A
Chemicals, Peptides, and Recombinant Proteins		
HpaII	NEB	R0171
MspI	NEB	R0106
HindIII-HF	NEB	R3104
NdeI	NEB	R0111S
EcoRII	Thermo Fisher	ER1921
Nucleoside Digestion Mix	NEB	M0649
M.MpeI WT His	This study	N/A
M.MpeI N374K His	This study	N/A
Dcm WT His	This study	N/A
Dcm N374K His	This study	N/A
Proteinase K	Zymo	D3001
RNAse A	Thermo	EN0531
SAM	NEB	B9003S
SAH	Sigma	A9384
Iodoacetic acid	Sigma	I4386
CxSAM	This study	N/A
Thymidine-D_3 ITSD	Toronto Research Chemicals	T412002
Critical Commercial Assays		
Oligo Clean and Concentrator kit	Zymo	D4060
Plasmid extraction kit	Qiagen	27104
gDNA extraction kit	Qiagen	69504
Oligonucleotides		
27-Top-C-FAM: GTATCTAGTTCAATCCGGTTCATAGCA-FAM	IDT	N/A
27-Bottom-mC: TGCTATGAAC/5mC/GGATTGAACTAGATAC	IDT	N/A
27-Bottom-C: TGCTATGAACCGGATTGAACTAGATAC	IDT	N/A
27-cxmC-standard	This study	N/A
Recombinant DNA		
pUC19	Miniprepped from NEB C2925I	N/A
Software and Algorithms		
Prism 8	Graphpad	www.graphpad.com
Fiji	This study	https://fiji.sc/

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Rahul Kohli (rkohli@pennmedicine.upenn.edu).

Materials Availability

Plasmids generated in this study can be obtained with a Materials Transfer Agreement.

Data and Code Availability

This study did not generate any datasets/code.

EXPERIMENTAL MODELS AND SUBJECT DETAILS

E. Coli Strains

ER1821 *E. coli* (New England Biolabs (NEB), F$^-$ *glnV44 e14$^-$(McrA$^-$) rfbD1? relA1? endA1 spoT1? thi-1 Δ(mcrC-mrr)114*::*IS10*) were used in M.MpeI experiments, including cloning. This strain is void of all methylation-specific restriction factors which recognize CpG methylation as foreign. ER1821 *ΔcmoA* was created with P1vir phage transduction using the *ΔcmoA* strain (JW1859) from the KEIO collection and kanamycin selection (Figures S1A–S1C) (Baba et al., 2006; Miller, 1992). For Dcm experiments, *dcm$^-$/dam$^-$ E. coli* were used (NEB C2925I, *ara-14 leuB6 fhuA31 lacY1 tsx78 glnV44 galK2 galT22 mcrA dcm-6 hisG4 rfbD1 R(zgb210*::*Tn10)* Tet^S *endA1 rspL136 (*Str^R*) dam13*::*Tn9 (*Cam^R*) xylA-5 mtl-1 thi-1 mcrB1 hsdR2*).

METHOD DETAILS

Cloning

The WT M.MpeI sequence was obtained from the protein FASTA file from the deposited (PDB: 4DKJ) crystal structure (Wojciechowski et al., 2013) and ordered as an *E. coli* codon-optimized GeneBlock from Integrated DNA Technologies (IDT). The gene was PCR amplified with primers containing BsaI-HF and HindIII-HF overhangs using Phusion Polymerase (NEB) and ligated into digested pMG81 plasmid, a medium copy number vector with an anhydrotetracycline-inducible promoter (Kubiak et al., 2017). The WT *dcm* gene was obtained by directly amplifying from ER1821 gDNA with Phusion Polymerase and primers introducing a C-terminal His tag and appropriate BsaI overhangs. This gene was then assembled using Golden-Gate cloning (Engler et al., 2008) into a compatible pMG81 plasmid. All mutations were obtained by performing Q5 Site Directed Mutagenesis (NEB) and confirmed by Sanger sequencing.

In Vivo Methyltransferase Assays

Individual pMG81-MMpeI or pMG81-Dcm plasmids were transformed into chemically-competent *E. coli* and plated separately. Single colonies were grown in cultures (3 mL LB, 100 μg/mL carbenicillin). Cultures were allowed to grow at 37°C until log phase (OD ∼ 0.4-0.7) before induction with 20 ng/mL anhydrotetracycline (ATc) and then grown overnight. Plasmid extractions (Qiagen) or gDNA extractions (Qiagen DNeasy) were then performed, eluted in 10 mM Tris-Cl, pH 8.0, and quantified by nanodrop.

Nucleoside LC-MS/MS

Plasmid or gDNA was digested with Nucleoside Digestion Mix (NEB) in a 10 μL total volume for 4 hours at 37°C, and the mixture was diluted 10-fold into 0.1% formic acid with the addition of 770 fmol thymidine-D_3 (T-D_3) internal standard (ITSD) into a volume of 20 μL. LC-MS/MS was performed as previously described (DeNizio et al., 2019). Briefly, 5 μL was injected onto an Agilent 1200 Series HPLC instrument equipped with a 5 μm, 2.1 mm × 250 mm Supelcosil LC18-S analytical column (Sigma) equilibrated to 45°C in buffer A (0.1% formic acid). The nucleosides were separated in a gradient of 0 to 10% buffer B (0.1% formic acid and 30% (v/v) acetonitrile) over 8 min at a flow rate of 0.5 mL/min. MS/MS was performed by positive ion mode ESI on an Agilent 6460 triple-quadrupole mass spectrometer, with a gas temperature of 225°C, a gas flow of 12 L/min, a nebulizer at 35 psi, a sheath gas temperature of 300°C, a sheath gas flow of 11 L/min, a capillary voltage of 3500 V, a fragmentor voltage of 70 V, and a delta EMV of 1000 V. Energies were 10 V for all bases except for 5cxmC (25V). When scanning in multiple reaction monitoring (MRM) mode, mass transitions were collected in increments of two mass units from 242.1 → 126.1 to 300.1 → 184.1. For targeted MRM, mass transitions were (C: 228.1 → 112.1, T: 243.1 → 127.1, T-D_3: 246.1 → 130.1, 5mC: 242.1 → 126.1, 5cxmC: 286.1 → 170.1). The amount of total input DNA injected was first obtained using T and the T-D_3 ITSD using the equations below, where *A* signifies area measured by the MS/MS instrument. This number was then used to calculate a relative quantity of 5cxmC nucleoside in the experiments that lack a chemical standard for 5cxmC. This approach allows for accurate comparisons across conditions and is used in Figures 2A–2C.

$$fmol\ T = 192.5\ fmol\ T \frac{A_T}{A_{T-D_3}}$$

$$Relative\ 5cxmC\ (arbitrary\ units) = \frac{A_{5cxmC}}{fmol\ T}$$

A standard for 5cxmC was synthesized using a chemoenzymatic approach. Excess M.MpeI N374K was reacted with CxSAM and hemimethylated substrate (see Oligonucleotide Assay methods). The substrate was verified for complete carboxymethylation by MspI digestion and visualized for fluorescein (FAM) fluorescence (excitation at 488 nm, emission at 520 nm) after separation on a

20% TBE Acrylamide Denaturing PAGE. The fully carboxymethylated standard was purified using an oligonucleotide spin column (Zymo), and quantified using the FAM fluorophore. This purified hemi-carboxymethylated oligonucleotide was then digested with Nucleoside Digestion Mix (New England Biolabs) in a 10 μL total volume for 4 hours at 37°C and diluted 10-fold into 0.1% formic acid before LC-MS/MS injection. An LC-MS/MS standard curve with serial dilutions of digested nucleoside (Figures S1D and S1E) was used to convert the integrated area of an experimental sample to fmol 5cxmC detected. The lowest concentration on the LC-MS/MS standard curve was determined to be the limit of detection (0.26 fmol Figures 2 and 4).

With knowledge of the amount of T and 5cxmC injected, it was possible to calculate the total amount of 5cxmC relative to either total CpG sites (M.MpeI) or CCWGG sites (Dcm, W = A/T). For M.MpeI experiments, the amount of T injected was converted to total amount of CpGs injected by dividing by the molar ratio of Ts to CpGs in the pMG81-MMpeI plasmid (5.07). For Dcm samples, the complete genome of K-12 MG1655, the parent strain of the *dam*$^-$/*dcm*$^-$ *E. coli* strain (GenBank: U00096.3) was analyzed. The molar ratio (101) comparing total instances of T (2,284,124) to CCWGG (22,716) was used to calculate the total amount of 5cxmC relative to total CCWGG sites.

Protein Purification

All variants were purified using a C-terminal His tag. Single colonies from transformation of pMG81-MMpeI, pMG81-M.MpeI-N374K, pMG81-Dcm, or pMG81-Dcm-N436K into ER1821 *E. coli* were started in overnight cultures (10 mL LB, 100 μg/mL carbenicillin), inoculated into large scale cultures (1 L LB, 100 μg/mL ampicillin) in the morning, and allowed to grow at 37°C until log phase (OD ~ 0.4 - 0.7) before switching the temperature to 16°C. After 20 minutes, 20 ng/mL anhydrotetracycline (ATc) was used to induce protein overexpression and cultures were left at 16°C overnight. Cells were harvested by centrifugation (8000g, 30 min, 4°C) before resuspending in 25 mL Buffer A (50 mM Tris Cl, pH 7.5 at 25°C, 150 mM NaCl, 25 mM Imidazole, 10% Glycerol (v/v)) + 1 EDTA-free Protease Inhibitor Tablet (Sigma) + 10 μL RNase A (Thermo Fisher). Resuspended cells were frozen overnight at −80°C.

Cells were lysed using a sonicator and centrifuged at 4°C for 30 min at 27,000g. Soluble lysate was passed through a gravity column containing 4 mL His Cobalt Resin, pre-equilibrated in Buffer A. After loading, the columns were washed with 25 column volumes (CV) of Buffer A for WT M.MpeI and WT Dcm, or 25 CV of Buffer B (50 mM Tris Cl, pH 7.5 at 25°C, 1 M NaCl, 25 mM Imidazole, 10% Glycerol (v/v)) followed by 5 CV of Buffer A for M.MpeI N374K and Dcm N436K. Protein was eluted with sequential fractions of Buffer C (50 mM Tris Cl, pH 7.5 at 25°C, 150 mM NaCl, 150 mM Imidazole, 10% Glycerol (v/v)). Samples were dialyzed (8,000 MWCO, Thermo Fisher) overnight at 4°C in 2 L of prechilled Dialysis Buffer (20 mM Tris HCl pH 7.5 at 25°C, 0.2 mM EDTA, 2 mM DTT, 150 mM NaCl, 10% Glycerol (v/v)). Cold 40% (v/v) glycerol was added to the protein to dilute the dialyzed protein 2-fold before flash freezing with liquid nitrogen and long-term storage at −80°C. All proteins were quantified by comparison to a BSA standard curve after running SDS-PAGE and visualizing with Coomassie Blue.

CxSAM Synthesis

Reactions were performed as previously described (Kim et al., 2015). 2-iodoacetic acid (667 mg) was added to a solution of *S*-adenosyl-L-homocysteine (20 mg) in 3.3 mL of 150 mM aqueous ammonium bicarbonate. The reaction was incubated at 37°C for 24 hrs. 80 mL of methanol was then added and incubated at 4°C overnight. Precipitates were collected by centrifugation at 4°C (2000g, 30 min) and washed twice with cold methanol to yield CxSAM. CxSAM was dissolved in nuclease free water and stored at -20°C. We attempted purification using a Zorbax 300SB-C18 column (9.4 x 250 mm, 5 μm particle size) (Agilent Technologies) with Buffer A = 10 mM Na_2HPO_4, pH 5.9 and Buffer B = 100% acetonitrile as described previously (Serebryakova et al., 2016). The absence of other products under these conditions suggested that no further purifications were necessary (Figure S2). CxSAM was quantified using absorbance measured at 260 nm (15,400 L mol^{-1} cm^{-1}). High resolution mass spectrometry (HRMS) was obtained to 443.1360 (mDa = -0.2, PPM = -0.5, Theoretical Mass: 443.1343).

Restriction Digest Based Assays

Restriction digestions were performed at 37°C for 1 hr in 1x NEB CutSmart Buffer in the specified volume. EcoRII restriction digests were performed at 37°C for 1 hr in 1x Thermo Scientific Buffer O.

pUC19 Assay

3-fold serial dilutions of M.MpeI (0.78 μM – 3.2 nM) were incubated with 160 μM SAM or CxSAM substrate and pUC19 plasmid DNA (100 ng) for 4 hrs at 37°C in M.MpeI reaction buffer (10 mM Tris Cl, 50 mM NaCl, 1 mM DTT, 1 mM EDTA, pH 7.9 at 25°C) in a 5 μL volume. 2.5 μL of the reaction was then incubated with the appropriate restriction enzyme, and the plasmid DNA was simultaneously linearized with HindIII-HF (NEB) in a final digestion volume of 25 μL. HpaII (NEB) recognizes C<u>C</u>GGs (13 sites). Samples were treated with 1 μL Proteinase K at 37°C for 10 min, separated on 1% TAE Agarose gel, and visualized with SYBR Safe DNA Gel Stain (ThermoFisher).

A similar assay was performed with the Dcm proteins except that serial 3-fold dilutions encompassed 0.26 μM – 3.2 nM and activity was instead assessed with EcoRII (Thermo), which recognizes C<u>C</u>WGGs (7 sites). The plasmid was also linearized with NdeI (NEB).

Oligonucleotide Assay

A FAM labelled top strand oligonucleotide with a single unmethylated C<u>C</u>GG and unlabeled complementary bottom strand oligonucleotides with an unmethylated or methylated C<u>C</u>GG were obtained from IDT as above. 200 nM of the duplexed, hemimethylated

oligo was reacted with serial dilutions of M.MpeI and 40 μM SAM or CxSAM substrate at 37°C in M.MpeI reaction buffer and a final volume of 5 μL for 30 min before heat inactivation at 95°C for 5 min. 25x excess of unmethylated bottom strand was reannealed before restriction digestion with HpaII in a final volume of 50 μL (Nabel et al., 2017). Undigested and digested strands were resolved on a 20% TBE Acrylamide Denaturing PAGE. *In vitro* carboxymethylation was also confirmed by purifying the reaction mixture before the strand exchange step with an Oligo Clean & Concentrator column (Zymo) and analyzing by electrospray ionization mass spectrometry (ESI-MS, Novatia, Figure S3).

Protein Structures

The structure of M.MpeI bound to SAH and a 5-fluorocytosine containing DNA substrate was used in modeling (PDB 4DKJ). The mutant N374K residue was manually created in PyMOL. Subsequently, CxSAM (PDB 4QNV) was manually overlaid on top of SAH with no energy minimization calculations to determine bond angles.

Growth Curves

dam$^-$/*dcm*$^-$ *E. coli* were transformed with pMG81-Dcm expression plasmids and plated on amplicillin. Overnight cultures were started from single colonies. The next morning, cultures were quantified by OD_{600} before dilution (LB, 100 μg/mL carbenicillin). Protein overexpression was induced with or without ATc (20 ng/mL). Cultures were monitored for 20 hours at 37°C with cycled agitation (3-mm orbital shaking, 450 rpm) in a 96-well plate (Mo et al., 2016). Optical density at 595 nm (OD_{595}) measurements were obtained every 5 minutes on a plate reader (Tecan).

QUANTIFICATION AND STATISTICAL ANALYSIS

All replicates are explicitly specified within figure legends.

FAM-labelled gels were quantified on a Typhoon imager and using Fiji software. Curves in Figure S3 were fit using GraphPad Prism 8 using non-linear least squares fit to extract EC_{50} values and 95% confidence intervals.

Cell Chemical Biology, Volume *28*

Supplemental Information

Discovery of an Unnatural DNA Modification Derived from a Natural Secondary Metabolite

Tong Wang and Rahul M. Kohli

Supplemental Figures:

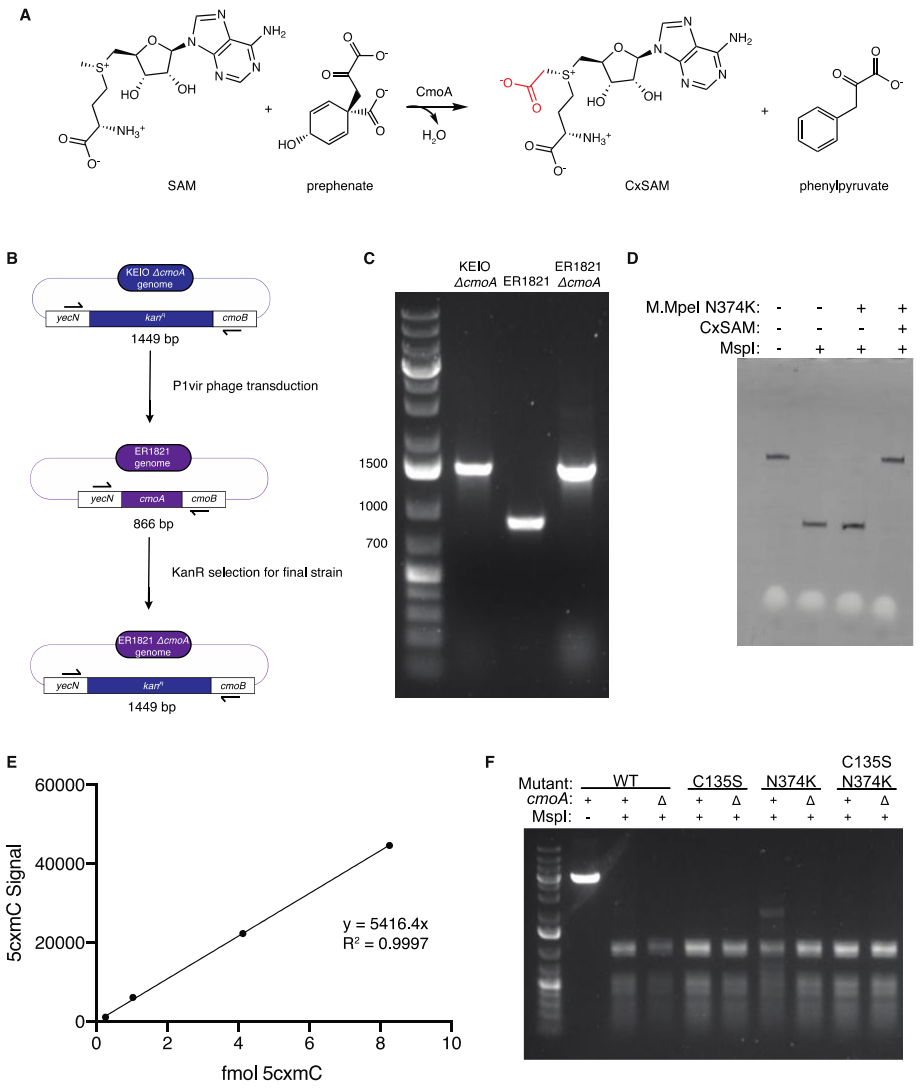


Figure S1, related to Figure 2: Validation of CmoA knockout and LC-MS/MS standard. (A) CmoA catalyzes the reaction between SAM and prephenate to yield water, CxSAM, and phenylpyruvate (Kim et al., 2013). (B) Transduction with bacteriophage P1vir was used to move a kanamycin resistance cassette from KEIO *ΔcmoA* to ER1821 cells (Baba et al., 2006) (C) Colony PCR was used to validate ER1821 *ΔcmoA* cells with primers flanking the endogenous *cmoA* locus. (D) MspI assay (see methods) confirming >98% carboxymethylation of 5cxmC oligonucleotide standard. When interpreted with Figure 1C, this experiment additionally shows that 5cxmC completely blocks MspI digestion, suggesting that this modification has a new gain-of-function ability to protect genomic DNA from restriction digestion in a way that 5mC cannot. (E) Representative LC-MS/MS standard curve for 5cxmC standard. (F) *In vivo* modification assay as described in Figure 1B. Upon overexpression, M.MpeI modifies its encoding plasmid using endogenous SAM or CxSAM. The modification can be assessed after plasmid isolation and restriction digestion. An MspI-resistant modification appears only in the presence of both M.MpeI N374K and an intact *cmoA* gene.

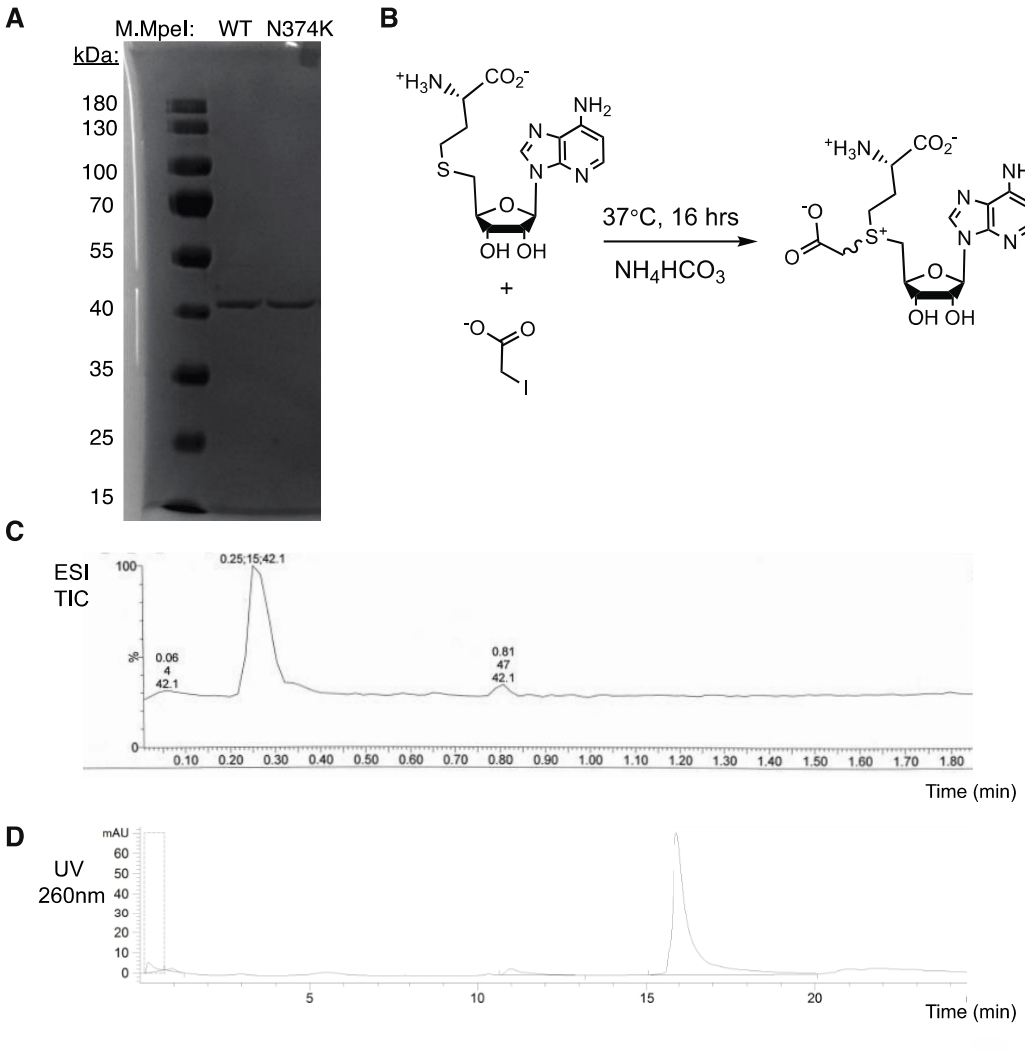


Figure S2, related to Figure 3: Validation of *in vitro* reagents. (A) SDS-PAGE of M.MpeI enzymatic preps used in this study. (B) Chemical synthesis of CxSAM. (C) LC-MS ESI$^+$ Total Ion Current (TIC) signal with observed mass of 443.5. In addition to the trace shown, HRMS was also obtained, identifying a mass of 443.1360 (mDa = -0.2, PPM = -0.5, Theoretical Mass: 443.1343). (D) HPLC purification of CxSAM showing single UV 260 nm peak.

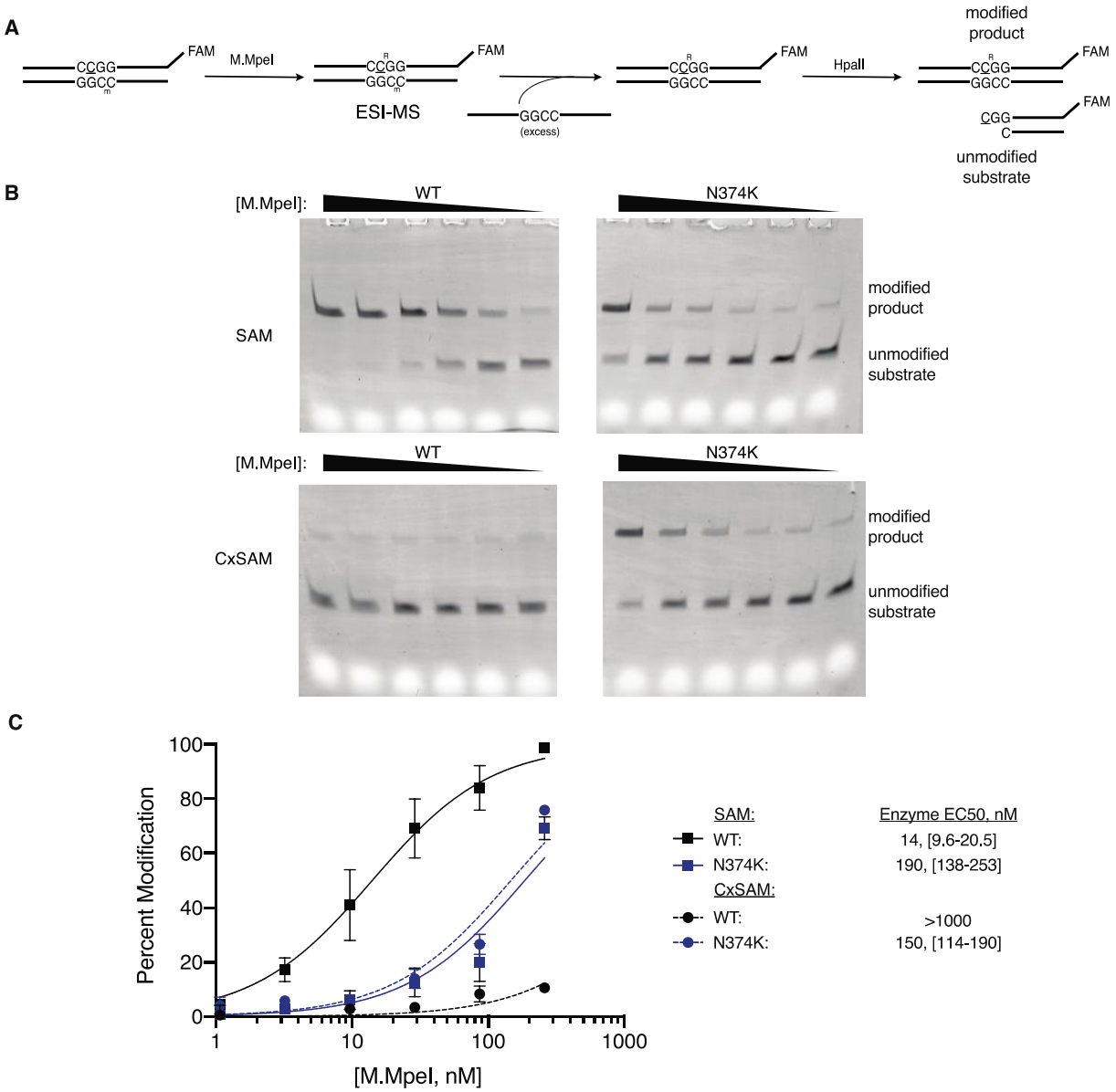


Figure S3, related to Figure 3: Quantitative oligonucleotide assay. Assay design was previously validated with homologous methyltransferases (Nabel et al., 2017). a) M.MpeI N374K was incubated with excess SAM or CxSAM and a hemimethylated CpG substrate containing a fluorophore label as shown. ESI-MS was obtained to confirm carboxymethylation of the hemimethylated substrate (expected top strand: 8877.9, observed top strand: 8876.7). HpaII digest was used to visualize total modification of the top strand after bottom strand exchange. b) Representative oligonucleotide assay gels. c) Enzyme dilution curve showing quantitative, relative activities of M.MpeI WT and N374K towards SAM and CxSAM. Points represent mean ± s.e. (n = 3 independent replicates). EC_{50} values were calculated, and 95% Confidence Intervals are reported in brackets.

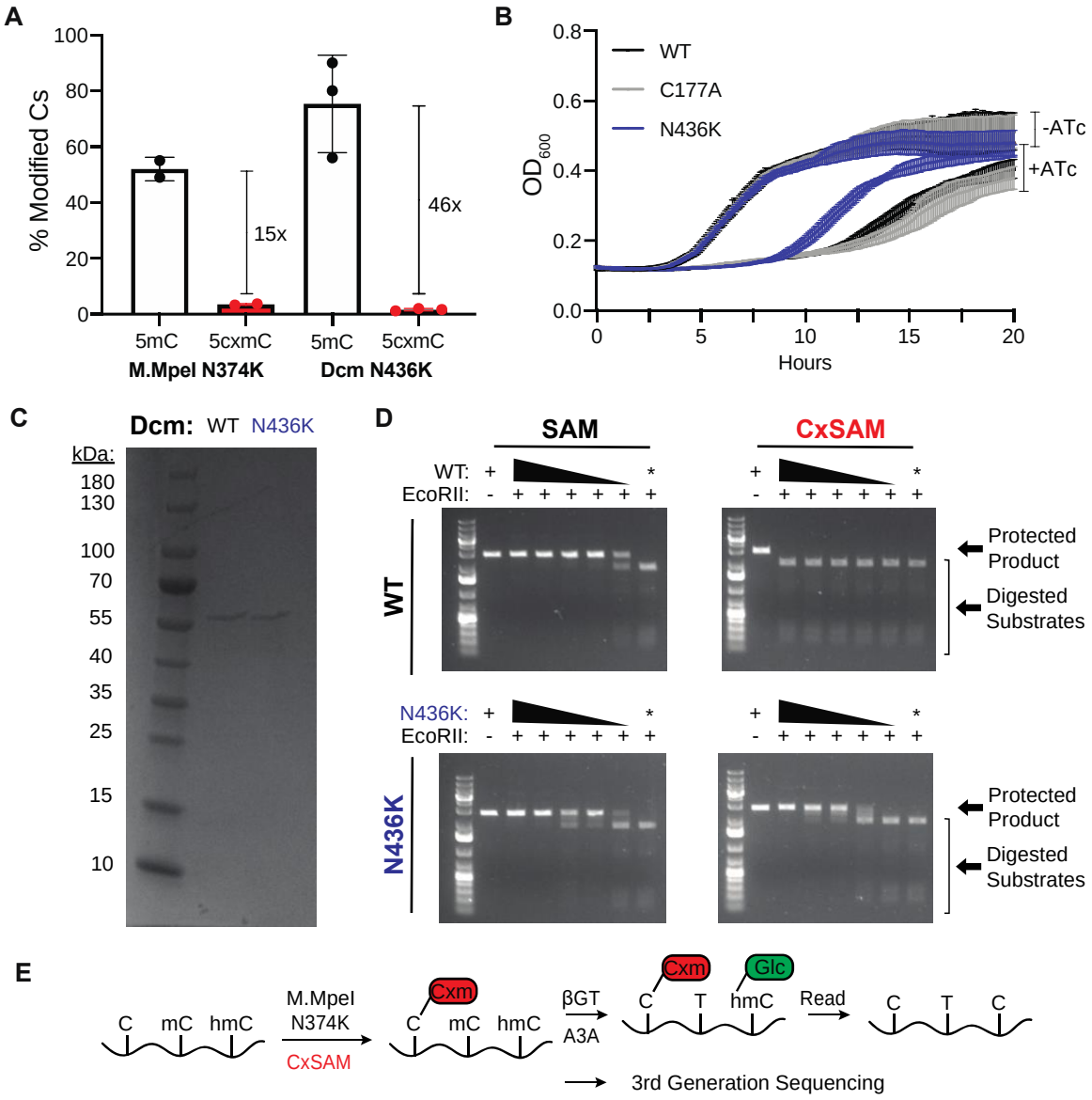


Figure S4, related to Figure 4: Generalizability and applications of DNA carboxymethyltransferases. (A) LC-MS/MS detection of 5mC and 5cxmC nucleosides catalyzed by M.MpeI N374K and Dcm N436K *in vivo* overexpression. M.MpeI experiments show percent of CpG modification on plasmid DNA. Dcm experiments are shown as a percent of CCWGG modification on gDNA. The relative 5mC vs 5cxmC modification levels are also labelled. These are the same experimental data as Figure 2E and Figure 4D, respectively. (B) Growth curves for *dam*⁻/*dcm*⁻ *E. coli* upon transformation with Dcm expression plasmids. Overnight cultures were diluted into a plate reader and monitored for 20 hrs using OD_{600}. Protein overexpression was induced with or without anhydrotetracycline (ATc). Points represent mean ± s.d. (n = 3 biological replicates). (C) SDS-PAGE of Dcm enzymatic preps used in this study. (D) Unmodified pUC19 plasmid DNA was incubated with an excess of SAM or CxSAM and 3-fold serial dilutions of Dcm WT or the N436K mutant to yield methylated or carboxymethylated DNA.

The negative control lane (*) contains the highest concentration of enzyme with no SAM or CxSAM substrate. Digestion with EcoRII fragments only unmodified DNA, allowing for visualization of digested substrate vs protected product. Dcm N436K transfers both CxSAM and SAM at comparable efficiency *in vitro* while Dcm WT only transfers SAM. (E) CpG carboxymethyltransferase can create 5-carboxymethylcytosine, which could permit differentiation of C, mC, and hmC by either enzymatic epigenetic sequencing methods or third generation sequencing (Schutsky et al., 2018).