

Carboxymethylcytosine is a natural base modification and a handle for bacteriophage DNA hypermodification

Received: 10 September 2025

Accepted: 20 November 2025

Published online: 07 December 2025


 Check for updates

Qiaoyu Yang^{1,2,3,10}, Lin Zhang^{4,10}, Yantao Liang^{5,10}, Haoyu Ma⁶, Lifu Song⁷, Lin Luo⁵, Jason Tan⁸, Yiling Hu¹, Kailiang Ma¹, Yiwei Chen¹, Yang Tong¹, Chuyuan Zhang¹, Suwen Zhao⁶ , Min Wang⁵ , Liang Zhang⁴ , Yifeng Wei⁸  & Yan Zhang^{1,2,3,9} 

Bacteriophages possess a wide array of DNA modifications, with many acting as molecular camouflage to evade host immune defenses. Sequence databases contain numerous bacteriophage enzymes of unknown function, with some potentially involved in yet to be identified DNA modifications. Here we report the discovery of a DNA cytosine C5-carboxymethyltransferase (CmoX) in *Synechococcus* phage S-B43, which catalyzes the formation of a 5-carboxymethylcytosine (5cxmC), previously reported as an unnatural DNA modification formed by an engineered cytosine methyltransferase. The carboxy-S-adenosyl-L-methionine (Cx-SAM) cofactor required by CmoX is provided by a phage-encoded Cx-SAM synthase (CmoA), a homolog of the bacterial CmoA involved in tRNA modification. A crystal structure of CmoX in complex with Cx-SAM revealed the basis for its substrate selectivity, involving a key Arg residue interacting with the substrate carboxy group. In addition, we characterize a phage-encoded ATP-dependent amide ligase, CmoY that catalyzes the formation of 5cxmC-glycine amide. CmoA is present in many bacteriophage genomes, typically alongside CmoX and homologs of CmoY, suggesting that 5cxmC modification is a widespread naturally occurring DNA modification serving as a handle for further hypermodifications in bacteriophages. Our study underscores the ability of bacteriophages to repurpose RNA modification enzymes to expand their repertoire of DNA modifications.

DNA modifications play important roles in all domains of life, including eukarya, bacteria, archaea, and viruses¹. In vertebrate epigenetics, 5-methylcytosine (5mC) modifications within CpG motifs function in maintaining cell identity, regulating gene expression and chromatin structure, as well as ensuring genome stability by suppressing transposable elements^{2,3}. In bacterial restriction-modification (RM) systems, 5mC, N6-methyladenine (6mA) and N4-methylcytosine (4mC) added by modification methyltransferases serve to distinguish self and foreign DNA, protecting bacterial genomes from their own restriction

enzymes¹. In comparison, tailed bacteriophages in the class Caudoviricetes possess far more diverse and complex DNA modifications that aid in evading bacterial RM systems, likely arising from an evolutionary arms race between phages and their hosts^{4,5}. The study of phage DNA modifications has broadened our understanding of DNA structure and of nucleotide biosynthetic pathways, with promising applications in various DNA-based technologies. The sheer abundance and variety of bacteriophages in nature suggests that many phage DNA modifications and DNA modifying enzymes remain to be discovered.

A full list of affiliations appears at the end of the paper.  e-mail: zhaosw@shanghaitech.edu.cn; mingwang@ouc.edu.cn; liangzhang2014@sjtu.edu.cn; weiyf@a-star.edu.sg; yan.zhang@tju.edu.cn

Phage DNA modifications generally occur either at the nucleotide level prior to incorporation by DNA polymerase, or at the nucleic acid level after the DNA has already been synthesized. An example of nucleotide level modification is 2-amino-dA (Z), which involves conversion of dGMP to dZMP by a phage-encoded succinyl-2-amino-dAMP synthetase (PurZ/PurZO)⁶ and bacterial adenylosuccinate lyase (PurB), followed by phosphorylation to dZTP and incorporation by phage DNA polymerase^{7–9}. An example of nucleic acid level modifications is the substitution of guanine by 7-deazaguanine in certain phage genomes, catalyzed by DNA-guanine transglycosylase (DpdA)¹⁰. This DNA modification system is homologous to the 7-deazaguanine tRNA modification system¹⁰, involving synthesis of the 7-deazaguanine base, and incorporation into tRNA via tRNA-guanine transglycosylase (Tgt), a distant homolog of DpdA^{11,12}. In certain phages, the 7-deazaguanine base undergoes further modifications, enhancing structural diversity¹³.

Compared to DNA, RNA exhibits a far greater variety of modifications, with broad effects on folding, structural stability, intermolecular interactions, splicing, gene expression, and immunogenicity¹⁴. Structural diversity is highest in tRNA, with over 100 distinct modifications identified¹⁵, mostly at the 5' wobble position of the anticodon loop¹⁶. We decided to examine whether any of these RNA modifications may have been repurposed by phages. We noticed a *cmoA* gene, involved in 5-carboxymethoxyuridine (cmo5U) modification of Gram-negative bacteria tRNA¹⁷, is present in certain phages. The carboxymethyl moiety originates from the unusual cofactor Cx-SAM, synthesized by the SAM carboxylase CmoA using prephenate as a co-substrate and involving a unique sulfonium ylide intermediate¹⁸. In *E. coli*, the O₂-dependent TrhO or prephenate-dependent TrhP catalyzes the formation of 5-hydroxyuridine (ho5U) in tRNA¹⁷. The methyltransferase homolog CmoB then catalyzes carboxymethyl transfer from Cx-SAM to ho5U, yielding mature cmo5U-modified tRNA¹⁸. In some cases, the carboxylate group is further methylated by the SAM-dependent methyltransferase CmoM, forming 5-methoxycarbonylmethoxyuridine¹⁹.

Here, we report the discovery of an unprecedented DNA modification system, involving synthesis of Cx-SAM by the phage CmoA, and DNA cytosine carboxymethylation by a homolog of DNA cytosine-5-methyltransferase (C5-MTase), which we name CmoX. The carboxymethyl group serves as a handle for subsequent hypermodification through ATP-dependent ligation to glycine catalyzed by an ATP-dependent ligase (CmoY). We present the crystal structure of CmoX, and examine its distribution across diverse bacteriophages, supporting the existence of a widespread modification-hypermodification system.

Results

Identification of CmoA homologs in bacteriophages

While searching for homologs of RNA modification enzymes in phage, we observed the presence of CmoA homologs in several phages, including *Synechococcus* phage S-B43²⁰ (Fig. 1a), which lacks a homolog of *E. coli* CmoB. Instead, the CmoA genome neighborhood contains a homolog of C5-MTase (which we designate CmoX) (Fig. 1a), suggesting that this enzyme may in fact catalyze DNA cytosine-C5-carboxymethylation (Fig. 1b). This reaction has previously been observed only in an engineered variant of the methyltransferase M.Mpel and the 5cxmC product was called unnatural base²¹. In addition, the genome neighborhoods also often contain a homolog of asparagine synthetase (which we designate CmoY) (Fig. 1a), suggesting further hypermodification by ATP-dependent ligation to an unknown amine (which we later identified as glycine, Fig. 1b). Sequence comparisons of these phage enzymes with their related *E. coli* homologs are summarized in Supplementary Table 1. To investigate this biochemical hypothesis, the enzymes were recombinantly produced in *E. coli* and purified (Supplementary Fig. 1).

SpCmoA is a highly active Cx-SAM synthase

SpCmoA and *EcCmoA* as a positive control were separately incubated with SAM and prephenate for 1 h, followed by precipitation of the proteins and analysis of the reaction products by HPLC and LC-MS (Fig. 2). The substrate SAM elutes with retention time of 26.57 min and positive ion *m/z* of 399.1. Incubation with either *SpCmoA* or *EcCmoA* led to a decrease in the SAM peak, and the appearance of a product peak with retention time of 23.48 min and a positive ion *m/z* of 443.0, corresponding to Cx-SAM, demonstrating that *SpCmoA* is indeed a Cx-SAM synthase. A significantly higher product yield was observed for *SpCmoA*, and subsequent assays determined *k_{cat}* values of 2.33 s⁻¹ for *SpCmoA* and 0.23 s⁻¹ for *EcCmoA* (Supplementary Fig. 2). The *SpCmoA* reaction mixture was thus used as a source of Cx-SAM for subsequent biochemical assays.

SpCmoX is a DNA cytosine C5-carboxymethyltransferase

SEC analysis indicated that recombinant *SpCmoX* exists as a monomer in solution (Supplementary Fig. 3). To assay for carboxymethyltransferase activity, *SpCmoX* was incubated with a GC-rich 110-bp dsDNA fragment, and the *SpCmoA* reaction mixture as a source of Cx-SAM, followed by purification and digestion of the dsDNA product to nucleosides. LC-MS analysis revealed the presence of a product peak with a positive ion *m/z* of 286.1 (Fig. 3a, b) and a MS/MS fragmentation pattern consistent with 5-carboxymethyl-2'-deoxycytidine (5cxmdC) (Supplementary Fig. 4). To investigate the effect of carboxymethylation on restriction enzyme digestion, HhaI (cleavage sequence GCG°C) was incubated with purified carboxymethylated dsDNA, or unmodified dsDNA as a control. Analysis by agarose gel electrophoresis revealed that carboxymethylation confers resistance to HhaI digestion (Fig. 3c).

We next attempted *in vivo* reconstitution of the CmoA-CmoX system in *E. coli* BL21(DE3) to determine whether it could modify plasmid DNA. Since *E. coli* natively expresses CmoA to produce the carboxymethyl donor Cx-SAM, we expressed only *SpcmoX*, from a modified pET28a (+)-HT vector. To investigate the requirement for CmoA, a Δ *cmoA* strain was constructed using a CRISPR-aided homologous recombination method, confirmed by colony PCR. The plasmid HT-*SpcmoX* was transformed into the wild-type and Δ *cmoA* strains, followed by growth and induction of *SpcmoX* with IPTG. The plasmids were then extracted and digested to nucleosides, followed by LC-MS analysis (Fig. 3d,e). A 5cxmdC peak with retention time of 3.30 min and positive ion *m/z* of 286.1 was only observed for wild-type but not the Δ *cmoA* strain, consistent with the ability of the bacterial CmoA to supply Cx-SAM for the carboxymethylation reaction.

Nanopore sequencing identifies a GC motif required for modification

To investigate the presence of modified bases in the genomic DNA of *Synechococcus* phage S-B43, the DNA was extracted, followed by sequencing by Oxford Nanopore technology. Analysis of the sequencing data using Tombo 1.5.1 revealed a cytosine with a high likelihood of modification (Fig. 4a), although its chemical structure remained undetermined. Sequences surrounding modified cytosine were analyzed using the MEME suit and visualized using WebLogo (Fig. 4b), revealing that the modification occurs within a highly conserved GC motif.

To further verify the sequence specificity of *SpCmoX*, a series of palindromic dsDNA substrates containing GC motifs were incubated with *SpCmoX*, then digested and analyzed by LC-MS to detect the presence of 5cxmdC. Controls included no *SpCmoX* and dsDNA substrates that do not contain any GC motif. Enzyme activity was detected only in assays with dsDNA substrates containing GC motifs, confirming that this motif is essential for recognition by *SpCmoX* (Fig. 4c, d).

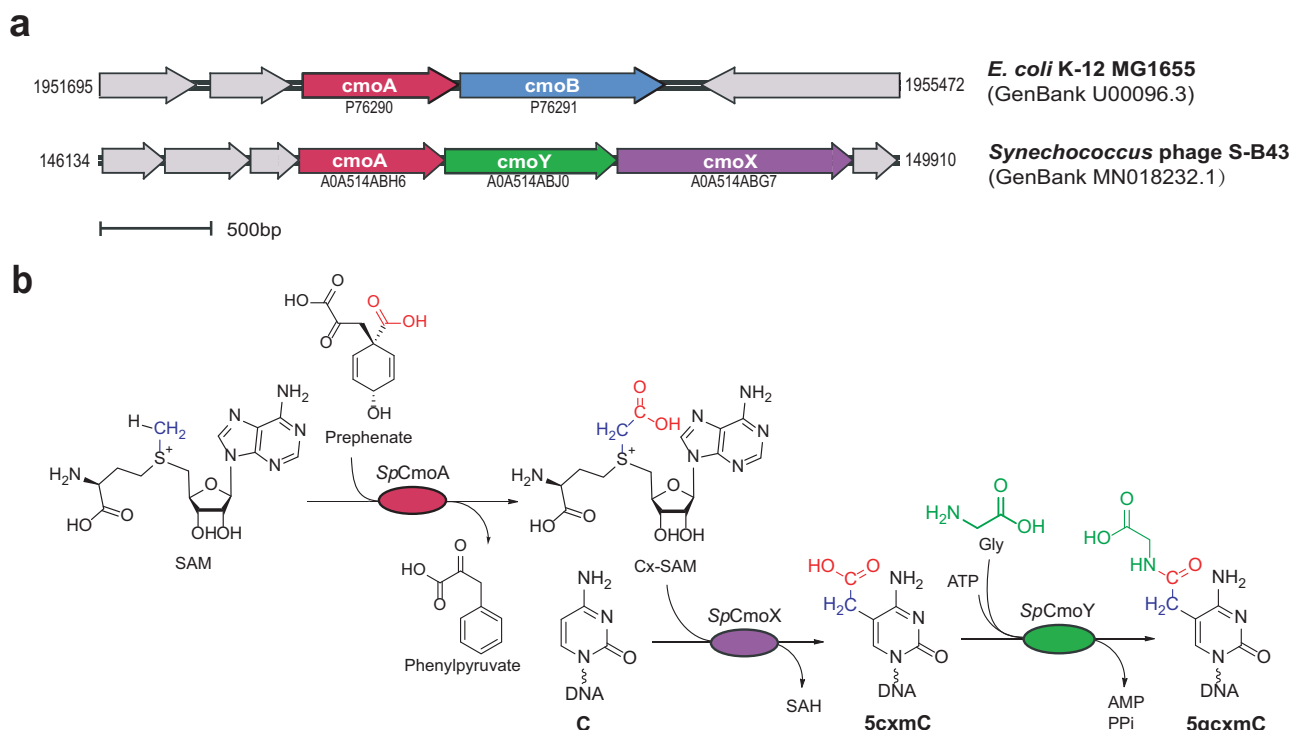


Fig. 1 | A putative 5-glycylcarboxymethylcytosine (5gcxmC) synthesis pathway. a Gene clusters containing *cmoA* in *E. coli* and *Synechococcus* phage S-B43. UniProt accession numbers for coded proteins are indicated. **b** Proposed pathway for 5gcxmC synthesis.

Crystal structure of *SpCmoX* in complex with Cx-SAM

To gain insights into the catalytic mechanism and substrate selectivity of CmoX, the crystal structure of *SpCmoX* in complex with the substrate Cx-SAM was determined to 1.9 Å resolution (Fig. 5, Supplementary Table 4). Like other C5-MTases, *SpCmoX* adopts a β/α -barrel fold (Fig. 5a). Key interactions between Cx-SAM and the active site residues of *SpCmoX* are detailed in Fig. 5b. The Cx-SAM adenine ring is stabilized by hydrogen bonding with Asp-69 and stacking interactions with aromatic residues Phe-25 and Phe-48, while the 2'-OH and 3'-OH of ribose form hydrogen bond interactions with Glu-47. A model of the *SpCmoX* ternary complex was constructed using an alignment with the crystal structure of HhaIM in complex with S-adenosyl-L-homocysteine (SAH) and a dsDNA (PDB: 1MHT, Fig. 5c). The Cys-81 nucleophile in HhaIM forms a covalent bond with cytosine C6, and aligns with Cys-98 in *SpCmoX*. Notably, Asn-304 in HhaIM is replaced with Arg-350 in *SpCmoX*, and is in direct interaction with the Cx-SAM carboxymethyl group. The interaction of the Cx-SAM carboxymethyl group with a specific Arg residue was previously proposed to contribute to substrate recognition by CmoB from *Vibrio vulnificus*²². We also tested whether the *SpCmoX* mutant R350N could accept SAM or Cx-SAM as substrates. The results showed that *SpCmoX* (R350N) could not accept either, while the wild-type could only accept Cx-SAM (Supplementary Fig. 5). The crystal structure of *SpCmoX* is consistent with a catalytic mechanism for cytosine-5-carboxymethylation that is analogous to related SAM-dependent methyltransferases⁴. In this pathway, nucleophilic attack of Cys-98 on the C6 position of cytosine forms a covalent enzyme-DNA carbanion intermediate, followed by transfer of the Cx-SAM carboxymethyl group onto the cytosine C5 carbanion, forming SAH. Finally, β -elimination of Cys-98 is facilitated by the deprotonation of C5, releasing the carboxymethylated DNA product (Fig. 5d).

SpCmoY ligates DNA 5cxmC to glycine

SpCmoA and *SpCmoX* are located adjacent to a gene encoding a homolog of asparagine synthetase, an ATP-dependent amide ligase, which we designate *SpCmoY*. The genomic proximity of *SpCmoY* with

CmoX suggests that *SpCmoY* may catalyze ATP-dependent ligation of the carboxymethyl group to an unidentified amine substrate. To test this hypothesis, *E. coli* BL21(DE3) strains harboring either the plasmid HT-*SpcmoX* (expressing *SpCmoX*, Fig. 6a) or the dual-expression plasmid HT-*SpcmoX-SpcmoY* (co-expressing *SpCmoX* and *SpCmoY*, Fig. 6b) were constructed. The genomic DNAs from both strains were extracted and digested to nucleosides, followed by LC-MS analysis (Fig. 6c, d). Expression of *SpCmoX* resulted in a 5cxmC peak (retention time of 3.42 min and positive ion m/z of 286.1), consistent with previous observations above. Strikingly, co-expression of *SpCmoX* and *SpCmoY* abolished the 5cxmC peak and generated a new peak (retention time of 3.36 min and m/z 343.1) consistent with glycine ligation to form 5-glycylcarboxymethyl-2'-deoxycytidine (5gcxmC) (Fig. 6e, f), as corroborated by MS/MS fragmentation (Supplementary Fig. 6).

To further validate this modification in vitro, 5cxmC-modified genomic DNA from the HT-*SpcmoX* transformant strain was incubated with purified *SpCmoY* in the presence of ATP and glycine. Workup and LC-MS analysis confirmed the formation of 5gcxmC (m/z 343.2 \rightarrow 227.1), which was absent in negative controls omitting ATP, enzyme, or glycine (Fig. 6g). The reaction was repeated with a synthetic 110-bp PCR-amplified dsDNA substrate modified first with *SpCmoX* and then *SpCmoY*, further confirming the formation of 5gcxmC (Supplementary Fig. 7). In analogy to related ATP-dependent ligases, we propose a catalytic mechanism, involving adenylation of the carboxyl group of 5cxmC, followed by nucleophilic attack by glycine, forming the amide bond in the modified base 5gcxmC (Fig. 6h).

Analysis of the *Synechococcus* phage S-B43 genomic DNA

We initially investigated the 5gcxmC modification in the recombinant system above due to its experimental simplicity compared to the native phage system. To investigate the native phage system, genomic DNA of *Synechococcus* phage S-B43 was purified and digested to nucleosides. LC-MS analysis confirmed the presence of 5gcxmC (Fig. 7), but not 5cxmC (Fig. 7a), demonstrating the relevance of the

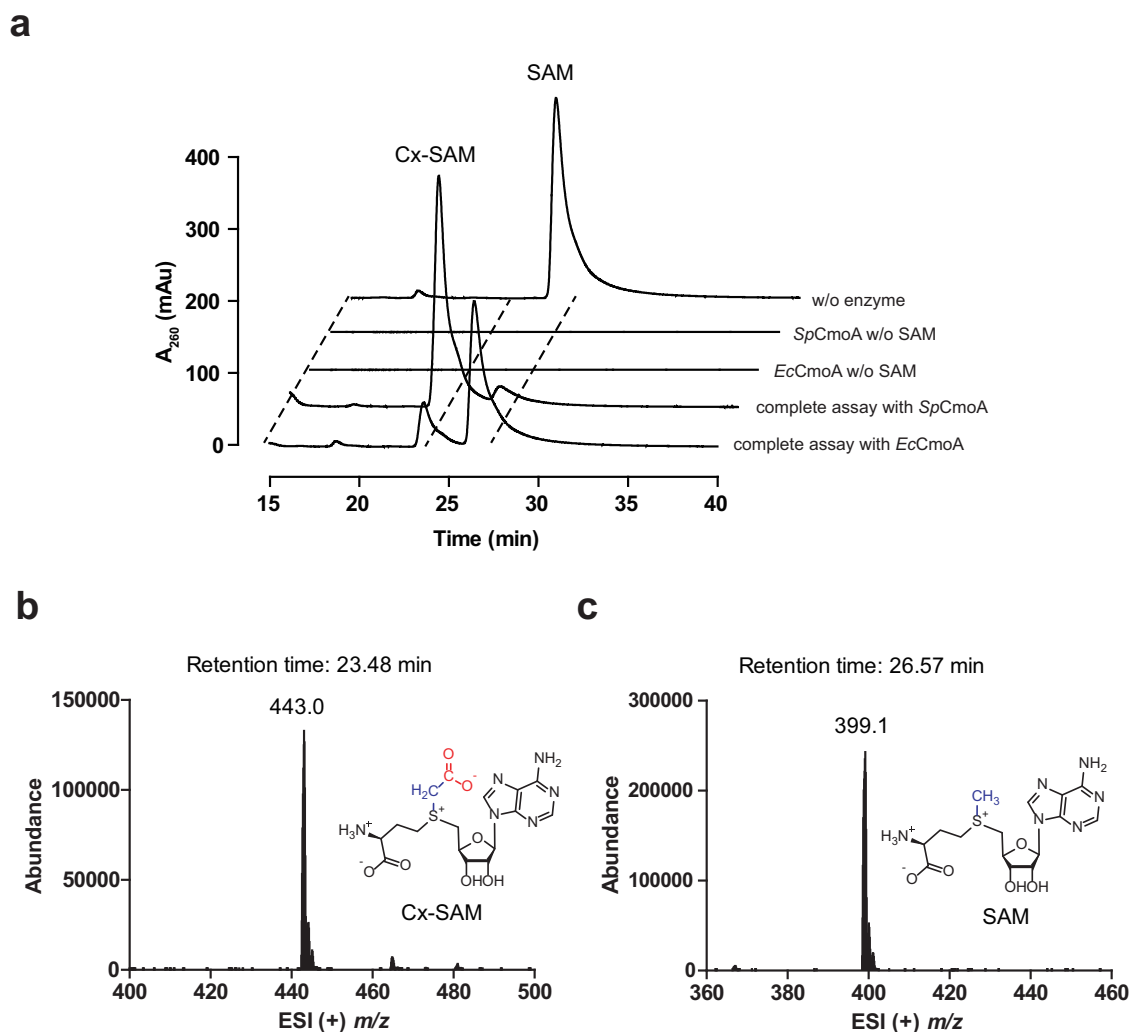


Fig. 2 | The activity assay of SpCmoA. a HPLC elution profiles of assays of SpCmoA and EcCmoA, monitoring the absorbance at 260 nm. ESI (+) m/z spectra corresponding to EIC peaks for **(b)** Cx-SAM ($m/z=443.0$) and **(c)** SAM ($m/z=399.1$).

5gcxmC modification in the native phage. An estimation based on the integrative peaks suggested that 10% of C bases and 75% of the GC sequences in the phage genome are modified.

Bioinformatics investigation of CmoX and CmoY homologs in phages

The identification of Cx-SAM-binding residues in the crystal structure enabled the search for candidate DNA cytosine carboxymethyltransferases in phage-derived genome sequences. Using SpCmoX as a query, homologous sequences were retrieved from IMGVR v4, UniProtKB, and a custom database prepared by extracting phage sequences from latest metagenomic data in MGnify, using jackhmmer (E-value 10^{-5} , 3 iterations). This search yielded 5,615 candidate CmoX sequences containing both the 'SPPC' motif (S95 P96 P97 C98) and the conserved residue R350, which is involved in carboxymethyl group binding. A Sequence Similarity Network (SSN) was constructed for these CmoX sequences to visualize their sequence diversity and genomic context (Fig. 8a–c). Multiple sequence alignments and weblogo analyses revealed conservation of the active site E and R residues, which are positioned to interact with the N⁴ and O² atoms of cytosine, respectively (Fig. 8d, e).

Many contigs contain a ParBc domain protein (PF02195), which is a small DNA-binding domain that frequently occurs as a fusion to various DNA methyltransferases²³, although its precise function in this

context remains unknown (Fig. 8c). CmoA and AMP-forming “Ligase” genes were identified within these contigs using both single-protein searches and Pfam family searches, and the results were merged. In total, 4257 CmoA, 2251 ligase, and 1660 ParBc sequences were identified (Fig. 8a–c). A SSN was constructed for the AMP-forming ligases to visualize their sequence diversity (Fig. 8f). The largest cluster contains A0A514ABJ0 (SpCmoY), and structural modeling indicates that members of this cluster, along with several smaller clusters, contain active site residues involved in binding the glycine carboxylate group (Supplementary Fig. 8a, b), suggesting that the substrate may be glycine or another α -amino acid. The second largest cluster contains A0A7TOQ3R0, and these proteins possess a glutamine-hydrolyzing amidase domain, homologous to that of asparagine synthetase, suggesting that their substrate is glutamine-derived ammonia. Unfortunately, attempts to obtain soluble recombinant enzyme from this cluster were unsuccessful. Several other clusters contain distinct active site residues, indicating the possibility of alternative amine substrates that remain to be identified (Supplementary Fig. 8c).

Discussion

Cx-SAM was first identified as a coenzyme involved in RNA modification¹⁵ and later shown to be utilized by bacteria in natural product biosynthesis²⁴. Our study of *Synechococcus* phage S-B43 shows that Cx-SAM also plays a role in the distinct DNA modification 5cxmC,

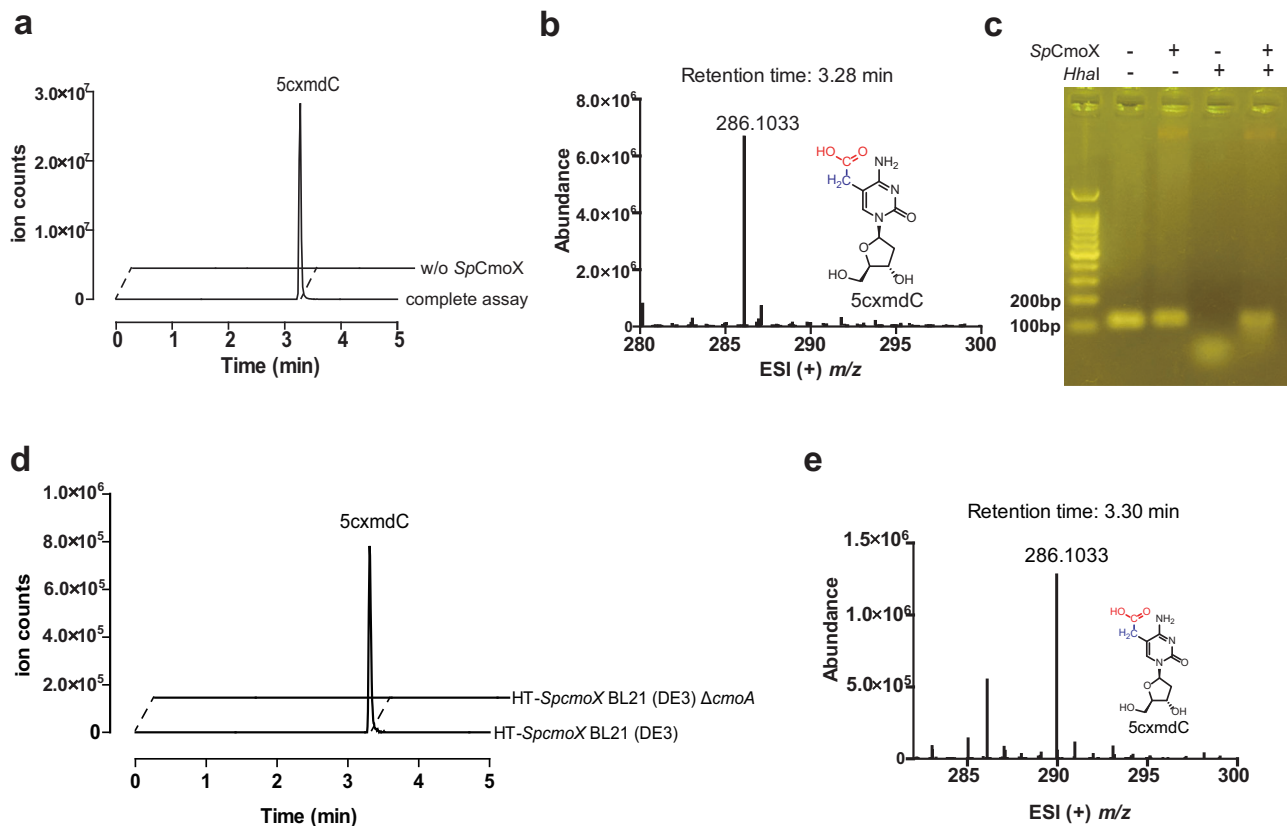


Fig. 3 | The activity assay and physiological function of SpCmoX in vitro and in vivo. **a** LC-MS extracted ion chromatogram of the digested dsDNA product, monitoring the formation of 5cxmdC ($m/z = 286.1$). **b** ESI (+) m/z spectrum corresponding to the EIC peak for 5cxmdC. **c** Agarose gel analysis of HhaI digestion of the SpCmoX-treated dsDNA. This experiment was independently repeated three

times with consistent results. **d** LC-MS extracted ion chromatograms detecting the presence of 5cxmdC ($m/z = 286.1$) from digested HT-SpcmoX plasmids extracted from BL21 (DE3) and BL21 (DE3) $\Delta cmoA$ cells. **e** The ESI (+) m/z spectrum corresponding to EIC peak for 5cxmdC.

installed by a phage-encoded cytosine C5-methyltransferase homolog, CmoX. Thus, just as bacteria have adopted carboxy-SAM for secondary metabolite synthesis in the context of interspecies competition²⁴, phages have recruited it for DNA modification in their evolutionary arms race with bacterial hosts, providing an example of how phages co-opt host RNA modification machinery to expand their DNA modification repertoire¹⁰. Notably, the phage-encoded CmoA exhibits higher activity than its bacterial counterpart, which may reflect the increased demand for Cx-SAM to achieve genome-wide modification. Further investigation is needed into the availability of prephenate, the co-substrate for CmoA and a key shikimate pathway intermediate²⁵, across different bacterial hosts. Additionally, since some phages encode SAM lyases that disrupt host SAM-dependent defenses such as BREX²⁶, it warrants investigation whether CmoA-mediated SAM depletion similarly impacts host metabolism and signaling, including the production of antiviral messengers like SAM-AMP²⁷.

Unlike typical DNA methyltransferases that install a methyl group as a steric marker, CmoX introduces a reactive carboxylate handle that enables subsequent hypermodification, in this case via ligation of glycine. The crystal structure of CmoX reveals an active-site arginine that interacts with the substrate carboxymethyl group, and strikingly mirrors a previously reported DNA methyltransferase variant M.MpeI N374K, which was engineered to accept Cx-SAM. While the engineered carboxymethyltransferase preferentially targets CG motifs²¹, CmoX exhibits complementary sequence specificity, recognizing GC motifs. Insights from the crystal structure enabled the identification of additional candidate carboxymethyltransferases in phages, although further work is needed to determine their sequence specificities.

To date, the most chemically diverse DNA modification-hypermodification systems involve the installation of a hydroxyl group in 5-hydroxymethyl-dU or -dC, serving as a handle for subsequent transformations such as glycosylation, amination, and carbamylation^{28,29}. In contrast, the carboxymethyltransferase system described here installs a carboxyl group as a complementary and versatile handle for amide bond formation. Our bioinformatics analyses suggest the presence of additional amide ligases associated with CmoA and CmoX, and further investigation is needed to uncover the full spectrum of amide modifications in natural phages. Apart from protection against restriction endonucleases, several natural and synthetic DNA modifications have been shown to either enhance or inhibit transcription by RNA polymerases, and reactive derivatives of hydroxymethyl-dC and dU have been utilized to achieve enzymatic or photochemical control of transcription^{30–34}. Further studies are warranted to determine how 5cxmdC and its derivatives affect RNA polymerase activity and to explore their potential applications in transcriptional regulation.

The strategy for DNA functionalization in phage S-B43 mirrors earlier engineering approaches using DNA, RNA, or protein methyltransferases and propargyl analogs of SAM, to enable subsequent modification via click chemistry^{35–37}. Identification of variants of CmoX and ligase in diverse phages highlights the potential for applications of such systems in DNA modification. While CmoX mediates global DNA modification to evade restriction systems, enzyme engineering could produce carboxymethyltransferases with extended recognition sequences, enabling site-specific DNA modification. Although our preliminary search did not identify candidate

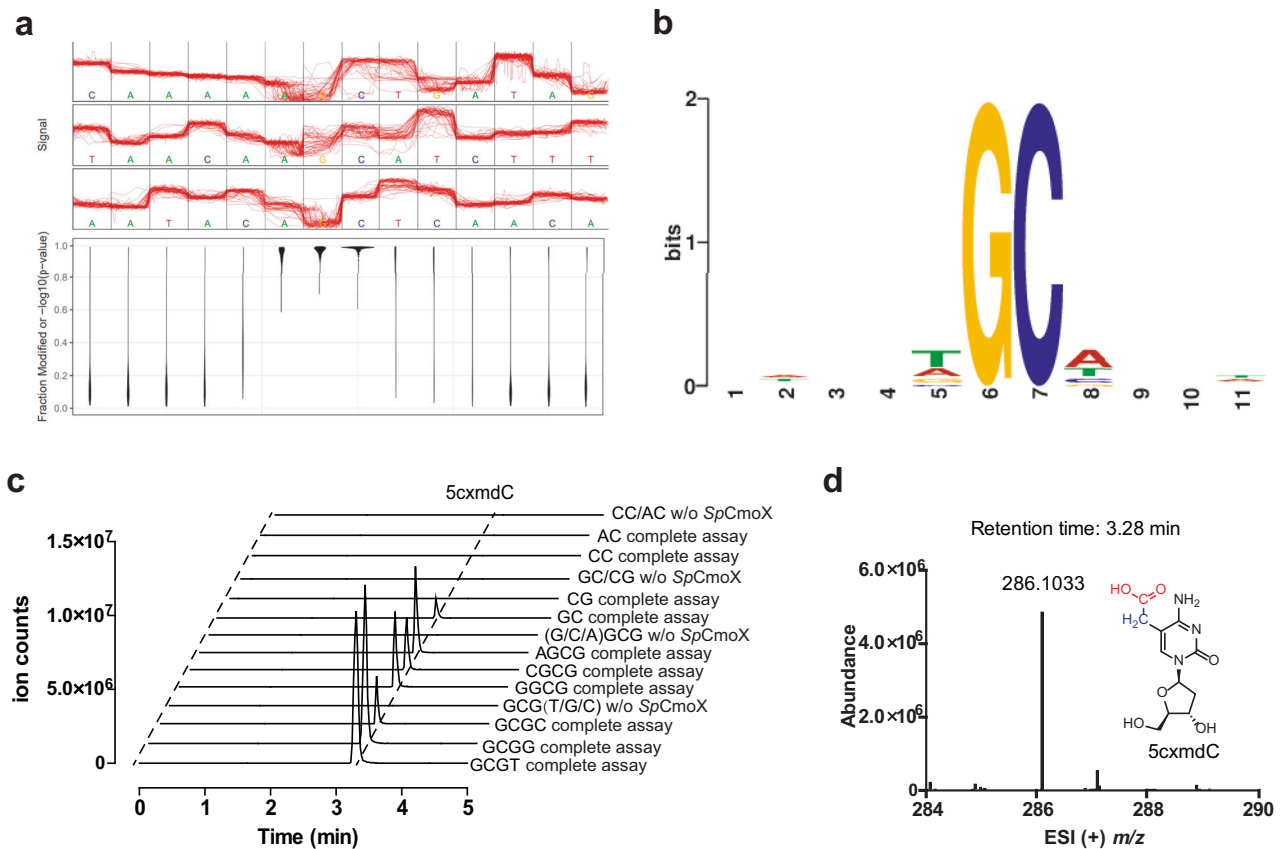


Fig. 4 | Nanopore sequencing of *Synechococcus* phage S-B43 genome and sequence specificity for SpCmoX. a Plot of the estimation results of the relative position of the modified base in the enriched motif. **b** Motif predicted by MEME from significantly modified regions of the *Synechococcus* phage S-B43 genome. To facilitate motif discovery, five positions on either side of the most modified base per region were included. **c** LC-MS extracted ion chromatograms monitoring the

formation of 5cxmdC ($m/z = 286.1$) in various assays. The substrate dsDNA was prepared by annealing 25 bp palindromic oligonucleotides (Supplementary Table 3). In the enzyme-free negative controls, GCG(T/G/C) and (G/C/A)GCG, mixed annealed oligos of the three preceding assays, were used as substrates. **d** ESI (+) m/z spectrum corresponding to EIC peak for 5cxmdC.

adenine N6- or cytosine N4-carboxymethyltransferases associated with CmoA in phages, such activities could theoretically be realized through enzyme engineering. Finally, engineering the ligase to accept a broader range of amine substrates could enable the incorporation of diverse functional groups, such as bioorthogonal handles for click chemistry, for applications in sequencing, DNA therapeutics, and chemical biology.

Methods

Materials and general methods

Lysogeny broth (LB) medium was prepared with yeast extract and tryptone (Oxoid, England). Ultrapure deionized (DI) water used in this work was prepared using Millipore Direct-Q. Ultrapure methanol and acetonitrile, used for liquid chromatography-mass spectrometry (LC-MS), were purchased from Concord Technology Co., Ltd (China). Prephenate barium salt was purchased from Sigma-Aldrich, Merck (U.S.A.). S-adenosyl-L-methionine (SAM) was acquired from Shanghai Yuanye Bio-Technology Co., Ltd (China). Oligonucleotide primers were synthesized by GENEWIZ, Inc. (Beijing, China). All protein purification chromatographic experiments were performed on liquid chromatography (FPLC) systems, including the ÄKTA pure, ÄKTA prime plus (GE Healthcare, U.S.A.) and UEV-25L (LI) system (Union-Biotech, China), equipped with appropriate columns (GE Healthcare, U.S.A.). Protein concentrations were calculated from the absorption at 280 nm, measured using MiuLab Ultra Trace Ultraviolet Spectrophotometer ND-100C (Hangzhou, China) or Eppendorf Biophotometer D30 (Hamburg, Germany).

Gene synthesis, molecular cloning and plasmid construction

Escherichia coli codon-optimized DNA fragments, encoding *Synechococcus* phage S-B43 CmoA (SpCmoA, A0A514ABH6), DNA cytosine carboxymethyltransferase (SpCmoX, A0A514ABG7) and ATP-dependent amide ligase (SpCmoY, A0A514ABJO), were synthesized and inserted into the *SspI* restriction site of the modified pET28a (+)-HT vector by General Biosystems, Inc (Anhui, China). The resulting plasmids (HT-SpcmoA, HT-SpcmoX, HT-SpcmoY) contain an N-terminal His₆-tag and a Tobacco Etch Virus (TEV) protease cleavage site, followed by the gene of interest. SpcmoA was also amplified from the HT-SpcmoA using p1F and p1R and inserted into the modified pET28a (+)-HMT vector at the *SspI* restriction site via Gibson assembly. The resulting plasmid (HMT-SpcmoA) contains in tandem: an N-terminus His₆-tag, maltose binding protein (MBP) and a TEV protease cleavage site, followed by SpcmoA. The gene for *E. coli* (strain K12) CmoA (EccmoA, P76290) was amplified by colony PCR with primers p2F and p2R and inserted into the pET28a (+)-HT vector (HT-EccmoA) for expression with an N-terminal His₆-tag via Gibson assembly. The plasmid HT-SpcmoX-SpcmoY was constructed by amplifying the HT-SpcmoX vector with primers p3F and p3R and inserting the SpcmoY gene amplified from HT-SpcmoY using primers p4F and p4R via Gibson assembly (Supplementary Table 2). All plasmids were confirmed by DNA sequencing.

Protein expression, purification and quantification

HMT-SpcmoA, HT-SpcmoX, HT-SpcmoY and HT-EccmoA plasmids were transformed into BL21 (DE3) *E. coli* cells. Transformants were selected

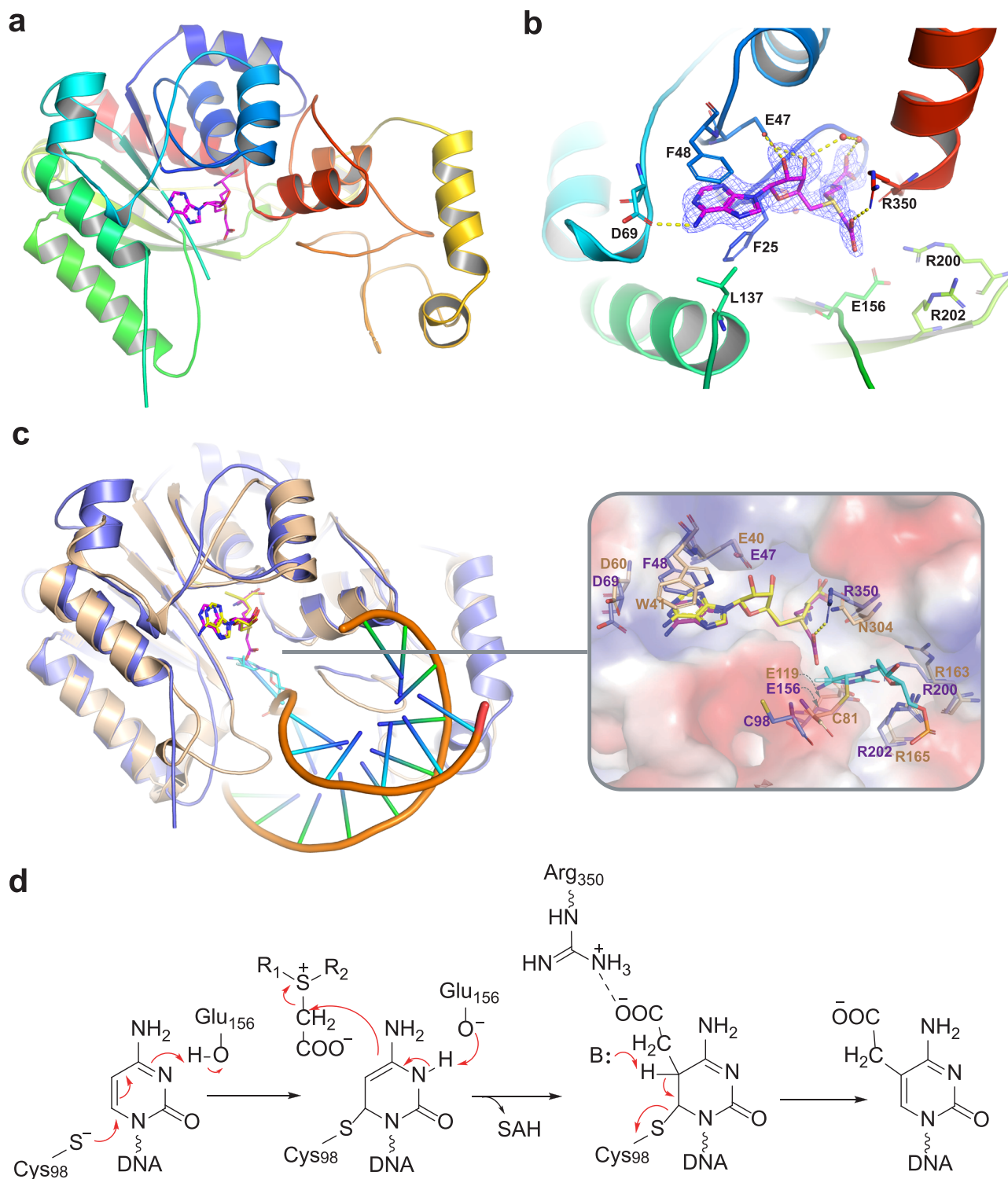


Fig. 5 | X-ray crystal structure of *SpCmoX* in complex with Cx-SAM. **a Overall structure of the *SpCmoX* monomer. **b** Interactions between Cx-SAM and active site residues of *SpCmoX*. Hydrogen bonds are represented by dashed lines. The $F_o - F_c$ omit map is shown in black mesh, contoured at 3.0σ around the Cx-SAM binding site. **c** Overlay of the overall structures and active sites of *SpCmoX* (slate blue, PDB**

entry 8YF3) and HhaIM (wheat, PDB entry 1MHT). α -based superposition of the two structures shows an RMSD of 1.027 Å. Key residues involved in substrate and cofactor binding are shown. **d** Proposed mechanism for the formation of 5cxdmC in the active site of *SpCmoX*.

on LB agar plates with 50 $\mu\text{g}/\text{mL}$ kanamycin. After an overnight incubation at 37 °C, a single colony was inoculated into 5 mL of LB medium, supplemented with 50 $\mu\text{g}/\text{mL}$ kanamycin, as starter culture. Cells were grown at 37 °C for 5 h, before being transferred into 1 L of the same medium at 37 °C with constant shaking at 220 rpm until OD_{600} reached 0.6–0.8. The temperature was then decreased to 18 °C, followed by the

addition of isopropyl β -D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.3 mM to induce protein expression. After continuous shaking for 16 h, the cells were harvested by centrifugation ($5,000 \times g$, 10 min, 4 °C). The resulting cell pellets for *SpCmoA*, *SpCmoY* and *EcCmoA* were then resuspended in lysis buffer containing Tris-HCl (50 mM, pH 8.0), KCl (150 mM), phenylmethanesulfonyl

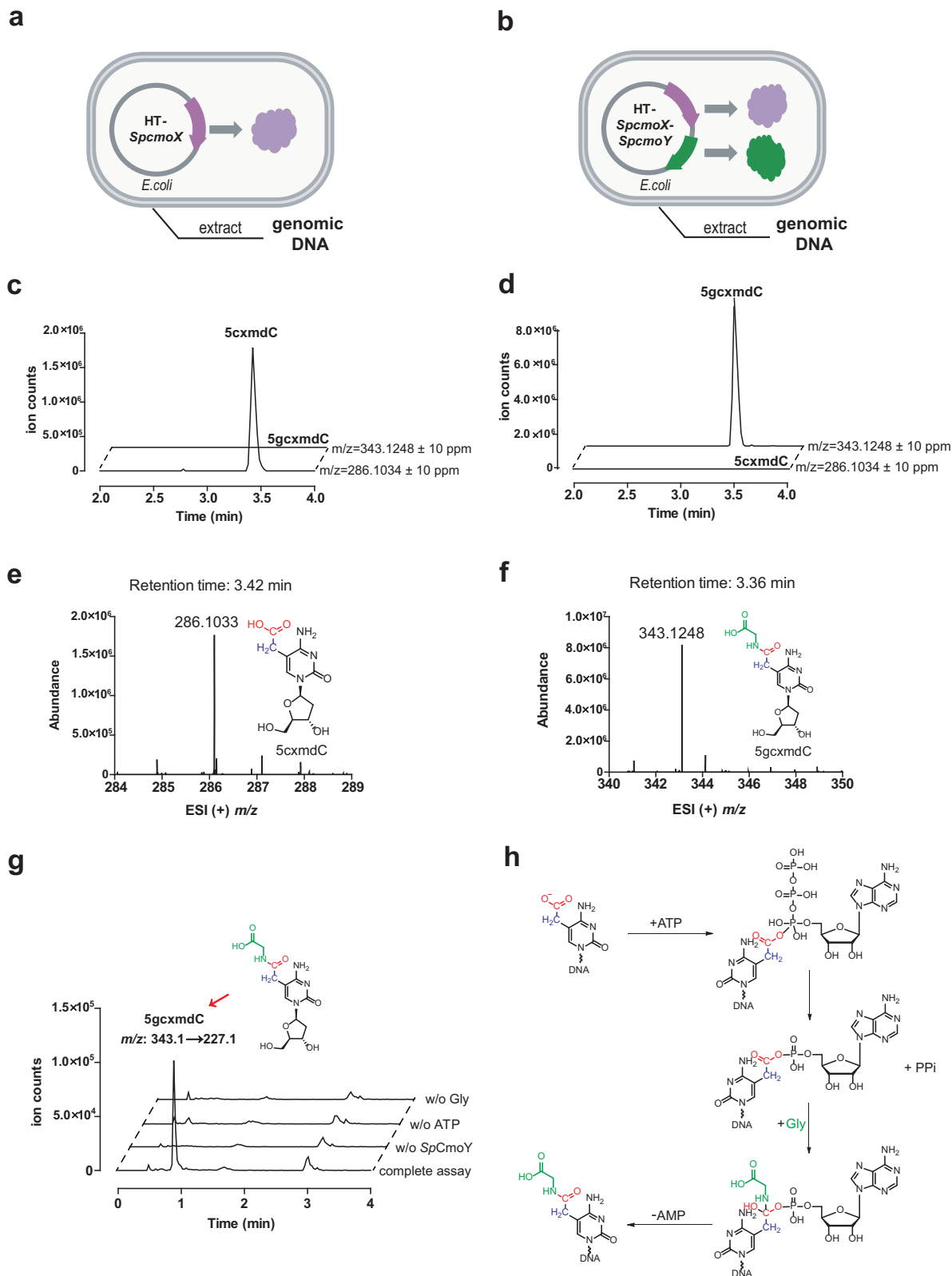


Fig. 6 | The activity assay of *SpCmoY*. **a** Extraction of the genomic DNA from *E. coli* cells overexpressing *SpCmoX*. **b** Extraction of the genomic DNA from *E. coli* cells overexpressing both *SpCmoX* and *SpCmoY*. **c** LC-MS extracted ion chromatograms monitoring the formation of 5cxmdC ($m/z = 286.1$) and 5gcxmdC ($m/z = 343.1$) in the genomic DNA of **(a)**. **d** LC-MS extracted ion chromatograms monitoring the formation of 5cxmdC ($m/z = 286.1$) and 5gcxmdC ($m/z = 343.1$) in

the genomic DNA of **(b)**. **e** ESI (+) m/z spectrum corresponding to EIC peak for 5cxmdC in **(c)**. **f** ESI (+) m/z spectrum corresponding to EIC peak for 5gcxmdC in **(d)**. **g** In vitro *SpCmoY* enzyme activity assay with the genomic DNA of **(a)** as the substrate. LC-MS/MS extracted ion chromatograms were used to monitor the formation of 5gcxmdC ($m/z = 343.1 \rightarrow 227.1$). **h** Proposed model for the *SpCmoY*-catalyzed reaction.

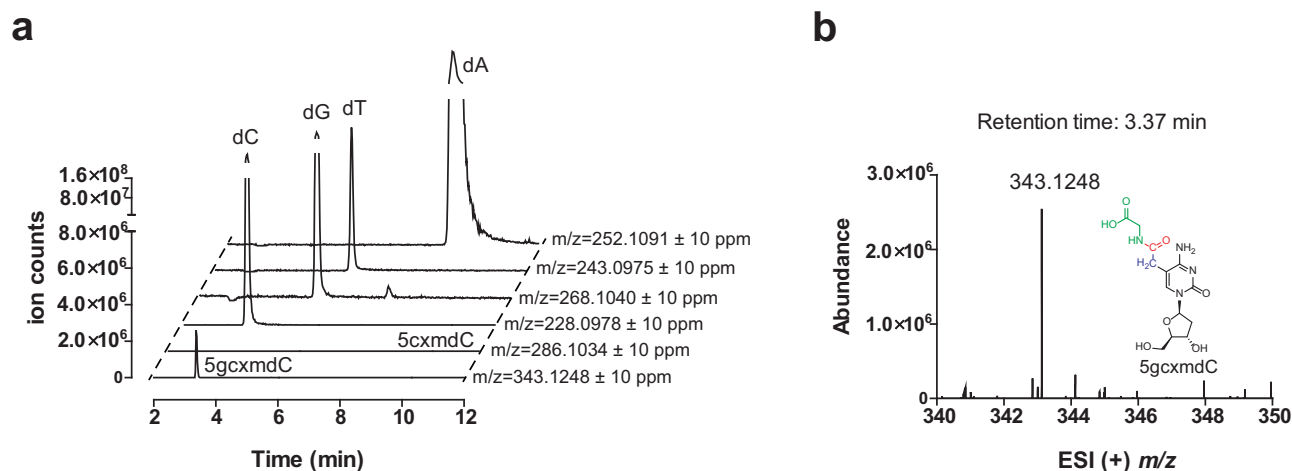


Fig. 7 | Validation of 5gcxmdC incorporation in the genome of *Synechococcus* Phage S-B43. **a LC-MS extracted ion chromatogram of the digested *Synechococcus* Phage S-B43 genome, monitoring the formation of 5gcxmdC ($m/z = 343.1$). **b** ESI (+) m/z spectrum corresponding to the EIC peak for 5gcxmdC.**

fluoride (PMSF, 1 mM), lysozyme (0.2 mg/mL), 0.03% Triton X-100, DNase I (0.02 mg/mL) and frozen in a -80°C freezer. The cell pellet for *SpCmoX* was resuspended in a different lysis buffer containing HEPES (20 mM, pH 6.8), NaCl (250 mM), PMSF (1 mM), lysozyme (0.2 mg/mL), 0.03% Triton X-100 and DNase I (0.02 mg/mL) before frozen. The cell suspensions were thawed and incubated at room temperature (RT) for 50 min to allow lysis. Nucleic acids were precipitated by addition of 6% streptomycin sulfate stock solution (final concentration, 1%), followed by centrifugation ($12,000 \times g$, 10 min, 4°C) to remove precipitated nucleic acids and cell debris. For all protein samples except *SpCmoX*, the supernatant was filtered and loaded onto a 5 mL TALON Co^{2+} column (Takara Bio USA, Inc.), pre-equilibrated with a solution of Tris-HCl (20 mM, pH 7.5) and KCl (200 mM), containing β -mercaptoethanol (BME, 5 mM) (Buffer A). The column was then washed with 10 column volumes (CV) of Buffer A and the protein was eluted with 5 CV of Buffer A containing 150 mM imidazole. The eluted protein (30 mL) was dialyzed against 2 L of Buffer A at 4°C for 4 h to remove the imidazole. The purification of *SpCmoX* employed Buffer B, containing HEPES (20 mM, pH 6.8) and 250 mM NaCl. The purified proteins were examined on an 8–16% SDS polyacrylamide gradient gel, with a commercial protein marker (Genstar, Beijing) and visualized by Coomassie staining. The protein samples were then aliquoted and glycerol was added to a final concentration of 10%. Lastly, the samples were flash frozen in liquid N_2 before being stored at -80°C . Additional chromatographic steps were performed to further purify *SpCmoX* for protein crystallization. The eluate from the TALON Co^{2+} column was diluted with Buffer C (20 mM HEPES, pH 6.8) to a NaCl concentration of 100 mM before loaded onto a 5 mL Heparin Sepharose high performance column. The column was eluted with 60 CV linear salt gradient from 100–675 mM NaCl in Buffer C. *SpCmoX* was collected, concentrated and buffer-exchanged with Buffer B containing 1 mM tris (2-carboxyethyl) phosphine (TCEP). The concentrations of the proteins were calculated based on their absorptions at 280 nm, using extinction coefficient values [*SpCmoA* ($\epsilon_{280} = 106,230 \text{ M}^{-1} \text{ cm}^{-1}$), *SpCmoX* ($\epsilon_{280} = 34,840 \text{ M}^{-1} \text{ cm}^{-1}$), *SpCmoY* ($\epsilon_{280} = 31,860 \text{ M}^{-1} \text{ cm}^{-1}$), *EcCmoA* ($\epsilon_{280} = 19,940 \text{ M}^{-1} \text{ cm}^{-1}$)].

Analysis of the oligomeric state of *SpCmoX*

The oligomeric state of *SpCmoX* was investigated using size exclusion chromatography (SEC). Briefly, a solution of *SpCmoX* (5 mL, 6.5 mg/mL) was injected into a Superdex 200 gel filtration column (300 mL) and eluted over 170 min with Buffer B containing 1 mM dithiothreitol (DTT) at 2 mL/min. The same conditions were used to analyze a mixture of molecular weight markers (Sigma MWGF 1000-1KT) to establish

a standard curve using a second-degree polynomial for the relationship between $\log(\text{molecular weight})$ and retention volume. The observed molecular weight for *SpCmoX* was 42.9 kDa indicating that it exists as a monomer (43.1 kDa) in solution.

LC-MS analysis of *SpCmoA* activity

A 200 μL reaction mixture containing 10 μM of *SpCmoA*, 5 mM SAM and 5 mM prephenate in 20 mM sodium phosphate buffer, pH 6.8 was incubated at 37°C for 1 h. This assay was performed with *EcCmoA* as a positive control, and without SAM or *SpCmoA* as negative controls. At the end of the reaction, an equal volume of acetonitrile was added to precipitate the protein. After removing the precipitate using a 0.22 μm nylon membrane filter, the filtrate was analyzed using LC-MS (Agilent 6420 Triple Quadrupole LC-MS instrument; ESI positive ion mode detection; Merck ZIC-HILIC column, bead size of 5 mm, pore size of 200 \AA , 150×4.6 mm) with a 9:1 (v/v) 0.1 M ammonium acetate/acetonitrile and acetonitrile gradient solvent system (90–70% vol. acetonitrile over 15 min, 70–50% vol. acetonitrile over 25 min, and 50% acetonitrile over 5 min, sample injection volume: 20 μL , UV 260 nm detection, flow rate: 0.75 mL/min). To compare the activities of *SpCmoA* and *EcCmoA*, enzymatic reactions each in a volume of 400 μL containing 20 mM sodium phosphate buffer (pH 6.8), 2 μM *CmoA*, 10 mM SAM, and 10 mM prephenate were incubated at 37°C . At 2-min intervals, 40 μL of the reaction mixture was sampled and immediately quenched with an equal volume of acetonitrile. The samples were then centrifuged to remove the enzyme before injection into the LC-MS system for peak integrative analysis referring to a standard curve established with known concentrations of chemically synthesized coxoy-SAM. Each assay was performed in triplicate.

Chemical and enzymatic synthesis of Cx-SAM

Chemical synthesis of Cx-SAM was performed as previously described³⁸. Briefly, 5.0 mg of SAH was dissolved in 0.8 mL of 150 mM ammonium bicarbonate (NH_4HCO_3). Subsequently, 166.7 mg of 2-iodoacetic acid was added. The reaction mixture was incubated at 37°C for 24 h with constant agitation. Following incubation, 20 mL of methanol was added, and the solution was maintained at 4°C overnight. The resulting precipitate was collected by centrifugation at $2000 \times g$ for 30 min at 4°C , followed by two washes with cold methanol and dissolved in nuclease-free water. Cx-SAM was purified using an LC-MS system (Agilent 6420 Triple Quadrupole LC-MS) equipped with an electrospray ionization (ESI) source in positive ion mode. Separation was performed on a Merck ZIC-HILIC column (5 μm

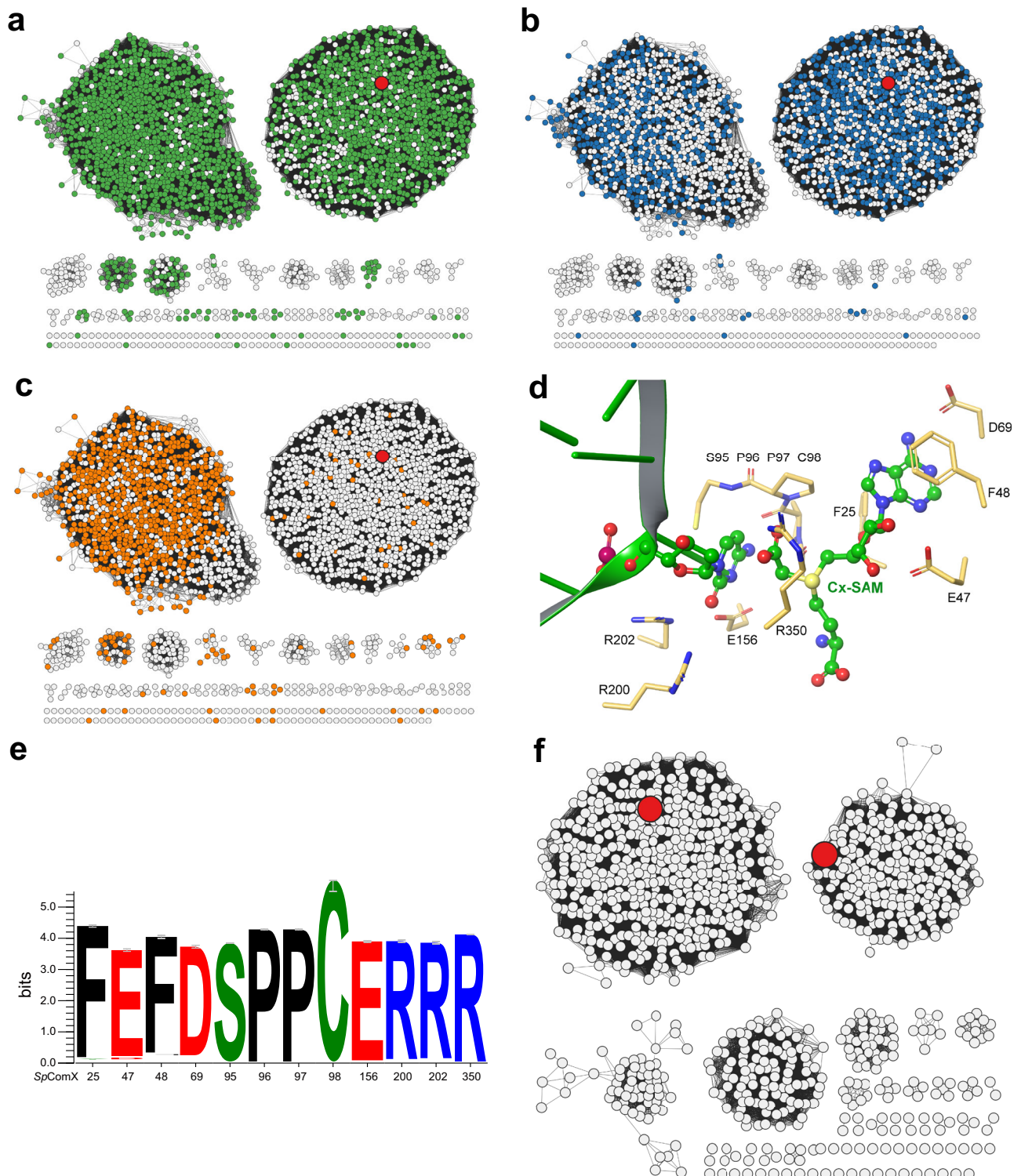


Fig. 8 | Bioinformatic support for a two-step cytosine modification pathway in phages. a–c The SSN of 5,615 CmoXs displayed at the e -value threshold of 10^{-70} (~50% sequence identity). The node coloring indicates genomic context gene associations: CmoA, AMP-forming ligase and ParBc-domain protein (PF02195), respectively. The large red node in each SSN represents *SpCmoX*. **d** Predicted structure model of *SpCmoX* bound to Cx-SAM, showing the ligand binding pocket.

The structure was generated using AlphaFold 3 and used to visualize the predicted binding mode of Cx-SAM. **e** Weblogo of active site residues of those CmoXs containing the SPPC motif and the conserved R350 residue. **f** The SSN of 2,251 AMP-forming ligases displayed at the e -value threshold of 10^{-56} (~30% sequence identity). A0A514ABJ0 (*SpCmoY*) and A0A7TQ3R0 are represented by large red nodes.

bead size, 200 Å pore size, 150 × 4.6 mm) using a gradient solvent system of 0.1M ammonium acetate/acetonitrile (9:1, v/v) and acetonitrile, as previously described. Cx-SAM concentration was determined by absorbance at 260 nm ($\epsilon = 15,400 \text{ L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$). We note that the chemical synthesis yields a mixture of stereoisomers at the

sulfur center³⁹. For non-quantitative assays of *SpCmoX* activity, Cx-SAM was prepared by a *SpCmoA*-catalyzed reaction. A reaction mixture (200 μL) containing HEPES buffer (20 mM, pH 6.8), NaCl (250 mM), *SpCmoA* (10 μM), SAM (10 mM) and prephenate (10 mM) was incubated at 37 °C for 1 h.

LC-MS analysis of *SpCmoX* activity

A reaction mixture (200 μ L) containing HEPES buffer (20 mM, pH 6.8), NaCl (250 mM), DTT (1 mM), Cx-SAM reaction mixture (40 μ L), *SpCmoX* (10 μ M) and double-stranded DNA (dsDNA) fragment (10 μ g, 5'-TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAAGCGCGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCGCTCCTTCGCTT-3'), amplified from modified pET28a (+)-HT vector using p5F and p5R. After overnight incubation at 18 °C, the dsDNA fragment was purified from the crude reaction mixture using a DNA Fragment Purification Kit (Genstar, Beijing). 5 μ g of the purified dsDNA fragment was then enzymatically digested with the Nucleoside Digestion Mix (New England Biolabs, M0649S) in a 100 μ L reaction mixture. The products were subjected to LC-MS analysis (Thermo Fisher Q Exactive HF/UltiMate 3000 RSLCnano; ESI positive ion mode detection; Synchronis aQ C18 column, particle size of 3 μ m, 150 \times 4.6 mm) with a 10 mM pH 4.6 aqueous ammonium acetate and methanol gradient solvent system (20-32% vol. methanol over 12 min, sample injection volume: 18 μ L, flow rate: 0.5 mL/min). The eluate was detected by absorbance at 260 nm and subjected to MS-MS analyses using the same machine and elution protocol.

Restriction enzyme digestion of DNA with 5cxmC

The *SpCmoX*-treated dsDNA fragment was incubated with *HhaI* (New England Biolabs) and CutSmart buffer (New England Biolabs) at 37 °C according to the manufacturer's instruction. The reaction mixture was then separated on a 2% agarose gel with gel electrophoresis conditions at 150 V for 30 min. Controls for the same gel assay included samples omitting treatment of *SpCmoX*, *HhaI* or both.

Assay of *SpCmoX* activity on annealed double-stranded palindromic oligo substrates

Palindromic oligos listed in Supplementary Table 3 were designed by us and synthesized by GENEWIZ. A 100 μ L reaction mixture containing Tris-HCl (10 mM, pH 8.0), NaCl (50 mM), palindromic oligo (100 μ M) was incubated at 95 °C for 5 min, before being cooled to RT. The annealed dsDNA was stored at -20 °C till use. A reaction mixture (100 μ L) containing HEPES (20 mM, pH 6.8), NaCl (250 mM), Cx-SAM reaction mixture (10 μ L), DTT (1 mM), *SpCmoX* (10 μ M) and annealed double-stranded palindromic oligo (25 μ L) was incubated at 18 °C overnight. The assay was performed without *SpCmoX* as negative control. At the end of the reaction, the mixture was heated at 95 °C for 5 min to precipitate the protein. The precipitate was removed by centrifugation and the supernatant was enzymatically digested with the Nucleoside Digestion Mix (New England Biolabs, M0649S) in a 100 μ L reaction mixture. LC-MS analysis was performed as described earlier.

Construction of the Δ *cmoA* *E. coli* strain

The BL21 (DE3) Δ *cmoA* *E. coli* strain was constructed using a CRISPR-aided homologous recombination strategy. Briefly, primer pairs p6F/p6R and p7F/p7R (Supplementary Table 2) were used to amplify the 500-bp region upstream and the 504-bp region downstream of *EccmoA* to produce fragments 1 and 2, respectively. Fragments 1 and 2 were then assembled using PCR with primers p6F and p7R to produce fragment 3. With the pRed_Cas9_RecA plasmid as template, fragments 4 and 5 were produced by PCR amplification using primer pairs p8F/p8R and p9F/p9R (Supplementary Table 2). Fragments 3, 4 and 5 were assembled using Gibson assembly to produce the recombinant pRed_Cas9_RecA_Δ*cmoA* plasmid and confirmed by sequencing, which contained a guide RNA sequence (5'-CGCCATTGAAAACGCATCGA-3') to introduce a Cas9-catalyzed cleavage near the homologous recombination site, thereby increasing the rate of the desired homologous recombination⁴⁰. pRed_Cas9_RecA_Δ*cmoA* plasmid was transformed into BL21 (DE3) cells and plated on LB agar, supplemented with 50 μ g/mL kanamycin. After overnight incubation at 30 °C, a colony was

inoculated into 5 mL of LB medium, supplemented with 50 μ g/mL kanamycin. Cells were grown at 30 °C for 2 h before the addition of 2 g/L L-arabinose to induce Cas9 and gRNA expression. After 24 h, the cells were harvested and streaked onto LB agar, supplemented with 50 μ g/mL kanamycin. The agar plate was then incubated at 37 °C, a non-permissive temperature resulting in the ejection of the plasmid. Surviving colonies were selected as candidates for *EccmoA* knockout via homologous recombination and were further verified by colony PCR using primers p10F and p10R and confirmed by sequencing (Supplementary Table 2).

SpCmoX activity in a recombinant *E. coli* system

HT-*SpCmoX* was transformed into BL21 (DE3) WT and Δ *cmoA* cells, followed by induction of *SpCmoX* expression. The plasmids were extracted and purified using the TIANprep Mini Plasmid Kit (TIANGEN). 10 μ g each was enzymatically digested with the Nucleoside Digestion Mix (New England Biolabs, M0649S) in a 100 μ L reaction volume and subjected to the LC-MS analysis as described in the LC-MS analysis of *SpCmoX* activity section of the Methods.

Crystallization, data collection and structure determination of *SpCmoX*

Initial screening of *SpCmoX* crystals was done by an automated liquid handling robotic system (Gryphon, Art Robbins), utilizing the sitting-drop vapor diffusion method in a 96-well format, at 291 K using various sparse matrix crystal screening kits (Hampton Research and Molecular Dimensions). The diffraction quality crystals co-crystallized with Cx-SAM were obtained by mixing *SpCmoX* (8.0 mg/mL) with BICINE (100 mM, pH 9.0), 20% (w/v) PEG 6000 and 1 mM Cx-SAM. The crystals were cryoprotected by the addition of 25% glycerol before being flash cooled in liquid nitrogen. Diffraction data were collected on a BL10U1 at the Shanghai Synchrotron Radiation Facility (SSRF) and processed to 1.9 Å resolution using HKL3000. Molecular replacement was performed by PHENIX (Version 1.14-3247)-Phaser⁴¹, using a homology model of *SpCmoX* created by AlphaFold. The structure was manually built according to the modified experimental electron density using Coot (Version 0.8.9.2)⁴² and further refined by PHENIX-Refine in iterative cycles. Supplementary Table 4 shows the statistics for data collection and final refinement. Protein structures were visualized and all structural figures were generated using PyMol (Version 3.1).

Site-directed mutagenesis of *SpCmoX* and activity assays of the mutant enzyme

A point mutation in the active site of *SpCmoX* (R350N) was introduced by PCR using primers p11F and p11R with HT-*SpCmoX* as the template and confirmed by sequencing. The expression, purification, and activity assays of the mutant enzyme were performed as described for the wild-type enzyme.

SpCmoY activity in a recombinant *E. coli* system

HT-*SpCmoX*-*SpCmoY* was transformed into BL21 (DE3) cells. The transformant was induced to express both proteins. Subsequently, genomic DNA was extracted and purified using the Bacteria Genomic DNA Kit from Beijing Zoman Biotechnology Co., Ltd (China) for analysis. The HT-*SpCmoX* transformant was used as a control. 10 μ g each was enzymatically digested with the Nucleoside Digestion Mix (New England Biolabs, M0649S) in a 50 μ L reaction volume and subjected to the LC-MS analysis as described in the LC-MS analysis of *SpCmoX* activity section of the Methods.

LC-MS analysis of *SpCmoY* activity

Genomic DNA containing 5cxmC was extracted from the HT-*SpCmoX* transformant cells and used as the substrate for *SpCmoY* enzyme activity assays. A reaction mixture (200 μ L) was prepared, containing HEPES (20 mM, pH 7.5), NaCl (250 mM), MgCl₂ (5 mM), ATP (5 mM),

Glycine (5 mM), 10 μ M SpCmoY, and 15 μ g of genomic DNA. The reaction was incubated overnight at 18 °C. Following incubation, the genomic DNA was purified from the crude reaction mixture using the Bacteria Genomic DNA Kit (Beijing Zoman Biotechnology Co., Ltd, China). 5 μ g of the purified genomic DNA was then enzymatically digested with the Nucleoside Digestion Mix (New England Biolabs, M0649S) in a 50 μ L reaction mixture and subjected to quantitative mass spectrometry (LC–MS/MS). The nucleosides were separated by an SB-Aq C18 column (2.1 \times 100 mm, 1.8 μ m; Agilent Technologies) and detected by ABCIEX QTRAP 6500 systems in positive ion mode. A gradient solvent system consisting of 5 mM aqueous ammonium acetate (pH 5.3) and acetonitrile was employed, with the acetonitrile concentration increasing from 0 to 50% over 7 min. The sample injection volume was 10 μ L, and the flow rate was maintained at 0.4 mL/min. Negative controls included assays in the absence of SpCmoY or one of the substrates, ATP or Glycine. The SpCmoX-treated, 5xcmC-modified, 110-bp dsDNA fragment obtained by PCR amplification from pET28a (+)-HT vector was used as another substrate in the SpCmoY assay.

Synechococcus phage S-B43 genomic DNA extraction and analysis

Synechococcus sp. Strain MW02 (NCBI accession number KP113680) was cultivated in artificial seawater-based f/2 medium at 25 °C, under 25 μ mol quanta $m^{-2}s^{-1}$ and a 12-h light: 12-h dark cycle. Phage expansion was performed by adding a solution of *Synechococcus* phage S-B43 (GenBank: MN018232.1) at a ratio of 1:9 (v/v) to the exponentially growing host *Synechococcus*. The phage-host suspension was then incubated at 25 °C, under constant irradiance of 25 μ mol quanta $m^{-2}s^{-1}$ for 7 days until host cell lysis was observed based on the color and turbidity of the lysate²⁰. The resulting lysate (200 mL) was centrifuged (7000 \times g, 10 min, 4 °C) to remove most of the cell debris. 8.3% (w/v) PEG 6000 and 2% (w/v) NaCl were then added to the supernatant followed by incubation at 4 °C in the dark overnight. The phage and cell debris was then pelleted by centrifugation (15,800 \times g, 15 min, 4 °C), and resuspended in 20 mL of SM buffer, containing Tris-HCl (50 mM, pH 7.5), NaCl (100 mM), magnesium sulfate (8.0 mM) and gelatin (0.1 g/L). The cell debris was then removed by differential centrifugation (10,000 \times g, 10 min, 4 °C). The genomic DNA of *Synechococcus* phage S-B43 was extracted using a Lambda phage genomic DNA Kit from Beijing Zoman Biotechnology Co., Ltd (China). 5 μ g genomic DNA was digested using the Nucleoside Digestion Mix (New England Biolabs, M0649S) in a 50 μ L reaction volume, followed by LC–MS analysis as described in the LC–MS analysis of SpCmoX activity section of the Methods.

Nanopore sequencing and data analysis

3–4 μ g of phage DNA per sample was used as input for library preparation. After the sample was qualified, size selection of long DNA fragments was performed using the PippinHT system (Sage Science, U.S.A.). The ends of the DNA fragments were then repaired followed by A-ligation reaction with the NEBNext Ultra II End Repair/da-Tailing Kit (New England Biolabs, E7546). The adapter in the SQK-LSK109 (Oxford Nanopore Technologies, United Kingdom) was used for further ligation and the DNA library was measured by Qubit 4.0 Fluorometer (Invitrogen, U.S.A.). Approximately 700 ng of DNA library was constructed and performed on a Nanopore PromethION sequencer instrument (Oxford Nanopore Technologies, United Kingdom) at the Genome Center of Grandomics (Wuhan, China). The nanopore sequencing data was analyzed with the Tombo 1.5.1 software, using the protocol designed for de novo identification of DNA modification⁴³. Modification sites, with significant values of at least 0.9, were chosen for the motif enrich analysis using the MEME suite 5.4.1⁴⁴. The relative position of the modified base, with the highest significant value, was then estimated using Tombo 1.5.1.

Bioinformatics

Sequence similarity network (SSN) analysis was carried out using the web-based Enzyme Function Initiative Enzyme Similarity Tool (EFI-EST)⁴⁵, and the results were visualized using Cytoscape version 3.8⁴⁶.

Statistics and reproducibility

All experiments in this study except Nanopore sequencing and X-ray crystallography were repeated independently at least three times. For LC–MS traces, and agarose gel assays, representative results are shown.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The atomic coordinates and structure factors for the crystal structure of SpCmoX in complex with Cx-SAM generated in this study have been deposited in the Protein Data Bank under accession code [8YF3](#). The large raw LC–MS and Nanopore FAST5 datasets can be shared by the corresponding author upon request. Other data supporting the findings of this study, including enzyme activity measurements, kinetic plots, LC–MS quantifications, and Nanopore modification analyses, are provided in the Source Data file. Source data are provided with this paper.

References

- Clark, T. A. et al. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res.* **40**, e29–e29 (2011).
- Kumar, S., Chinnusamy, V. & Mohapatra, T. Epigenetics of modified DNA bases: 5-methylcytosine and beyond. *Front. Genet.* **9**, 2018 (2018).
- Paulsen, M. & Ferguson-Smith, A. C. DNA methylation in genomic imprinting, development, and disease. *J. Pathol.* **195**, 97–110 (2001).
- Weigle, P. & Raleigh, E. A. Biosynthesis and function of modified bases in bacteria and their viruses. *Chem. Rev.* **116**, 12655–12687 (2016).
- Sood, A. J., Viner, C. & Hoffman, M. M. DNAmod: the DNA modification database. *J. Cheminform.* **11**, 30 (2019).
- Tong, Y. et al. Alternative Z-genome biosynthesis pathway shows evolutionary progression from Archaea to phage. *Nat. Microbiol.* **8**, 1330–1338 (2023).
- Pezo, V. et al. Noncanonical DNA polymerization by aminoadenine-based siphoviruses. *Science* **372**, 520–524 (2021).
- Sleiman, D. et al. A third purine biosynthetic pathway encoded by aminoadenine-based viral DNA genomes. *Science* **372**, 516–520 (2021).
- Zhou, Y. et al. A widespread pathway for substitution of adenine by diaminopurine in phage genomes. *Science* **372**, 512–516 (2021).
- Thiaville, J. J. et al. Novel genomic island modifies DNA with 7-deazaguanine derivatives. *Proc. Natl. Acad. Sci. USA* **113**, E1452–E1459 (2016).
- Reader, J. S., Metzgar, D., Schimmel, P. & de Crécy-Lagard, V. Identification of four genes necessary for biosynthesis of the modified nucleoside queuosine. *J. Biol. Chem.* **279**, 6280–6285 (2004).
- Hutinet, G. et al. 7-Deazaguanine modifications protect phage DNA from host restriction systems. *Nat. Commun.* **10**, 5442 (2019).
- Cui, L. et al. Four additional natural 7-deazaguanine derivatives in phages and how to make them. *Nucleic Acids Res.* **51**, 9214–9226 (2023).
- Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA modifications in gene expression regulation. *Cell* **169**, 1187–1200 (2017).
- Boccalletto, P. et al. MODOMICS: a database of RNA modification pathways. 2021 update. *Nucleic Acids Res.* **50**, D231–D235 (2021).

16. Agris, P. F., Vendeix, F. A. P. & Graham, W. D. tRNA's Wobble decoding of the genome: 40 years of modification. *J. Mol. Biol.* **366**, 1–13 (2007).
17. Sakai, Y., Kimura, S. & Suzuki, T. Dual pathways of tRNA hydroxylation ensure efficient translation by expanding decoding capability. *Nat. Commun.* **10**, 2858 (2019).
18. Kim, J. et al. Structure-guided discovery of the metabolite carboxy-SAM that modulates tRNA function. *Nature* **498**, 123–126 (2013).
19. Sakai, Y., Miyauchi, K., Kimura, S. & Suzuki, T. Biogenesis and growth phase-dependent alteration of 5-methoxycarbonylmethoxyuridine in tRNA anticodons. *Nucleic Acids Res.* **44**, 509–523 (2015).
20. Wang, M. et al. Genomic analysis of *Synechococcus* phage S-B43 and its adaption to the coastal environment. *Virus Res.* **289**, 198155 (2020).
21. Wang, T. & Kohli, R. M. Discovery of an unnatural DNA modification derived from a natural secondary metabolite. *Cell Chem. Biol.* **28**, 97–104.e104 (2021).
22. Jeong, S. & Kim, J. Structural snapshots of CmoB in various states during wobble uridine modification of tRNA. *Biochem. Biophys. Res. Commun.* **534**, 604–609 (2021).
23. Letunic, I., Khedkar, S. & Bork, P. SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res.* **49**, D458–D460 (2020).
24. Serebryakova, M. et al. A Trojan-Horse peptide-carboxymethyl-cytidine antibiotic from *Bacillus amyloliquefaciens*. *J. Am. Chem. Soc.* **138**, 15690–15698 (2016).
25. Kanehisa, M., Furumichi, M., Sato, Y., Mitsuura, Y. & Ishiguro-Watanabe, M. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* **53**, D672–D677 (2024).
26. Andriianov, A. et al. Phage T3 overcomes the BREX defense through SAM cleavage and inhibition of SAM synthesis by SAM lyase. *Cell Rep.* **42**, 112972 (2023).
27. Chi, H. et al. Antiviral type III CRISPR signalling via conjugation of ATP and SAM. *Nature* **622**, 826–833 (2023).
28. Hutinet, G., Lee, Y. J., de Crécy-Lagard, V. & Weigle, P. R. Hypermodified DNA in Viruses of *E. coli* and *Salmonella*. *EcoSal* **9**, eESP-0028-2019 (2021).
29. Lee, Y.-J. et al. Pathways of thymidine hypermodification. *Nucleic Acids Res.* **50**, 3001–3017 (2021).
30. Janoušková, M. et al. 5-(Hydroxymethyl)uracil and -cytosine as potential epigenetic marks enhancing or inhibiting transcription with bacterial RNA polymerase. *Chem. Commun.* **53**, 13253–13255 (2017).
31. Vaníková, Z., Janoušková, M., Kambová, M., Krásný, L. & Hocek, M. Switching transcription with bacterial RNA polymerase through photocaging, photorelease and phosphorylation reactions in the major groove of DNA. *Chem. Sci.* **10**, 3937–3942 (2019).
32. Chakrapani, A. et al. Glucosylated 5-Hydroxymethylpyrimidines as Epigenetic DNA Bases Regulating Transcription and Restriction Cleavage. *Chem. A Eur. J.* **28**, e202200911 (2022).
33. Gracias, F. et al. Homologues of epigenetic pyrimidines: 5-alkyl-, 5-hydroxyalkyl and 5-acyluracil and -cytosine nucleotides: synthesis, enzymatic incorporation into DNA and effect on transcription with bacterial RNA polymerase. *RSC Chem. Biol.* **3**, 1069–1075 (2022).
34. Gracias, F., Pohl, R., Sýkorová, V. & Hocek, M. Bacteriophage-related epigenetic natural and non-natural pyrimidine nucleotides and their influence on transcription with T7 RNA polymerase. *Commun. Chem.* **7**, 256 (2024).
35. Fischer, T. R. et al. Chemical biology and medicinal chemistry of RNA methyltransferases. *Nucleic Acids Res.* **50**, 4216–4245 (2022).
36. Brown, T., Nguyen, T., Zhou, B. & Zheng, Y. G. Chemical probes and methods for the study of protein arginine methylation. *RSC Chem. Biol.* **4**, 647–669 (2023).
37. Dalhoff, C., Lukinavičius, G., Klimašauskas, S. & Weinhold, E. Synthesis of S-adenosyl-L-methionine analogs and their use for sequence-specific transalkylation of DNA by methyltransferases. *Nat. Protoc.* **1**, 1879–1886 (2006).
38. Kim, J. et al. Determinants of the CmoB carboxymethyl transferase utilized for selective tRNA wobble modification. *Nucleic Acids Res.* **43**, 4602–4613 (2015).
39. Schülke, K. H., Ospina, F., Hörschemeyer, K., Gergel, S. & Hammer, S. C. Substrate profiling of anion methyltransferases for promiscuous synthesis of S-adenosylmethionine analogs from haloalkanes. *ChemBioChem* **23**, e202100632 (2022).
40. Zhao, D. et al. Development of a fast and easy method for *Escherichia coli* genome editing with CRISPR/Cas9. *Microb. Cell Fact.* **15**, 205 (2016).
41. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D* **66**, 213–221 (2010).
42. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126–2132 (2004).
43. Stoiber, M. et al. *De novo* Identification of DNA Modifications Enabled by Genome-Guided Nanopore Signal Processing. *bioRxiv*, 094672 (2017).
44. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).
45. Oberg, N., Zallot, R. & Gerlt, J. A. EFI-EST, EFI-GNT, and EFI-CGFP: enzyme function initiative (EFI) web resource for genomic enzymology tools. *J. Mol. Biol.* **435**, 168018 (2023).
46. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (NSFC) Distinguished Young Scholar of China Program (32125002, Y.Z.), the New Cornerstone Science Foundation (NCI2002321, Y.Z.), and the funding support from Agency for Science, Technology and Research (C211917011, Y.W.). We thank the instrument analytical center of School of Pharmaceutical Science and Technology at Tianjin University, in particular, Zhi Li and Yan Gao for providing assistance in our LC–MS analysis. We also thank the supports from the Core Facility of Basic Medical Sciences at Shanghai Jiao Tong University School of Medicine for LC–MS/MS data collection and analysis. We are grateful to Prof. Xiang Yu and Dr. You Wu, from the School of Life Sciences and Biotechnology at Shanghai Jiao Tong University, for their helpful discussions and analytical insights regarding Nanopore sequencing data. In addition, we thank Chuxiao Wang and Zhenquan Deng, Master students at the College of Marine Life Sciences, Ocean University of China, for their valuable assistance in phage cultivation and related experimental procedures, and doctoral student Hao Yu for his support in bioinformatics analyses.

Author contributions

Q.Y., Lin.Z. and Y.L. conducted most of biochemical and molecular biology experiments. H.M. carried out the bioinformatic analyses. L.S. performed the Nanopore sequencing data analysis. L.L., Y.H., K.M., Y.C., Y.T., C.Z., and Y.W. participated in data acquisition. Q.Y., J.T., H.M., S.Z., M.W., Liang.Z., Y.W., Y. Zhang designed the experiments and wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-66999-9>.

Correspondence and requests for materials should be addressed to Suwen Zhao, Min Wang, Liang Zhang, Yifeng Wei or Yan Zhang.

Peer review information *Nature Communications* thanks Michal Hocke and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

¹New Cornerstone Science Laboratory, School of Pharmaceutical Science and Technology, Tianjin University, Tianjin, China. ²Frontiers Science Center for Synthetic Biology (Ministry of Education), Tianjin University, Tianjin, China. ³Key Laboratory of Systems Bioengineering (Ministry of Education), School of Chemical Engineering and Technology, Tianjin University, Tianjin, China. ⁴School of Chemistry and Chemical Engineering, and Zhangjiang Institute for Advanced Study, Shanghai Jiao Tong University, Shanghai, China. ⁵College of Marine Life Sciences, Ocean University of China, Qingdao, China. ⁶Human Institute and School of Life Science and Technology, ShanghaiTech University, Shanghai, China. ⁷Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, China. ⁸Singapore Institute of Food and Biotechnology Innovation (SIFBI), Agency for Science, Technology and Research (A*STAR), 31 Biopolis Way, Nanos, Singapore, Singapore. ⁹State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China. ¹⁰These authors contributed equally: Qiaoyu Yang, Lin Zhang, Yantao Liang.

✉ e-mail: zhaosw@shanghaitech.edu.cn; mingwang@ouc.edu.cn; liangzhang2014@sjtu.edu.cn; weiyf@a-star.edu.sg; yan.zhang@tju.edu.cn