Nature Communications

https://doi.org/10.1038/s41467-026-70671-1

Article in Press

A single viral enzyme drives tRNA-dependent hypermodification of DNA at adenine

Received: 11 August 2023

Accepted: 3 March 2026

Cite this article as: Silva, R.M., Slyvka, A., Lee, Y.-J. *et al.* A single viral enzyme drives tRNA-dependent hypermodification of DNA at adenine. *Nat Commun* (2026). https://doi.org/10.1038/s41467-026-70671-1

Rebekah M. B. Silva, Anton Slyvka, Yan-Jiun Lee, Chudi Guan, Sean R. Lund, Elisabeth A. Raleigh, Krzysztof Skowronek, Michael S. Kuska, Matthias Bochtler & Peter R. Weigele

A single viral enzyme drives tRNA-dependent hypermodification of DNA at adenine

Authors: Rebekah M. B. Silva[1†], Anton Slyvka[2†], Yan-Jiun Lee[1], Chudi Guan[1], Sean R. Lund[1], Elisabeth A. Raleigh[1], Krzysztof Skowronek[2], Michael S. Kuska[1], Matthias Bochtler[2,3*], and Peter R. Weigele[1*]

Affiliations:

[1]Research Department, New England Biolabs, Inc.; Ipswich, MA, USA

[2]International Institute of Molecular and Cell Biology in Warsaw; Warsaw, Poland

[3]Institute of Biochemistry and Biophysics, Polish Academy of Sciences; Warsaw, Poland

*Corresponding author. Email: mbochtler@iimcb.gov.pl; *Corresponding author. Email: weigele@neb.com

†These authors contributed equally to this work.

Abstract

Nucleic acid modifying enzymes drive diverse defense and counter-defense measures in the evolutionary arms race between viruses and their cellular hosts. Abundant and widespread bacterial viruses (bacteriophage or phage) encode for biosynthetic pathways that install elaborate DNA hypermodifications which protect their genomic DNA from host endonucleases. Here, we establish the molecular basis for the multistep biosynthesis of 6-aminocarboxymethyl-2′-deoxyadenosine (6-*N*cmdA), a nucleobase hypermodification found in the virion DNA of bacteriophage Mu that leads to restriction evasion in the context of phage-host conflicts. In the first step, we show that Mu-encoded Mom enzyme catalyzes the formation of 6-*N*cmdA by transferring glycine from charged $tRNA^{Gly}$ to the *N*6 position of adenine within double-stranded DNA. We uncover a second step where the glycyl-dA intermediate undergoes an on-base rearrangement to form 6-*N*cmdA. Examination of the proposed reaction pathways by quantum chemical calculations confirms the instability of acyl exocyclic groups at *N*6-adenine and reveals an energetically favorable orientation of 6-*N*cmdA that restores canonical base pairing. An X-ray structure confirms Mom is a member of the GNAT superfamily and suggests binding sites for both tRNA and DNA. Guided by the Mom structure and patterns of sequence conservation across metagenomic space, we show residues R111 and S124 are essential for catalysis. This work demonstrates that the Mom enzyme defines a new category of acetyltransferases utilizing charged tRNA to modify DNA.

Introduction

Nucleic acid modifying enzymes drive diverse defense and counter-defense measures in the evolutionary arms race between viruses and their cellular hosts[1,2]. Abundant and widespread bacterial viruses (bacteriophage or phage) encode for biosynthetic pathways that install elaborate DNA hypermodifications which protect their genomic DNA from host endonucleases[3–10]. The biosynthetic complexity and chemical diversity found within naturally modified DNA from phage, as collectively encoded by the most numerically abundant genetic entity in our biosphere, has only begun to be appreciated. Phage-encoded biosynthetic pathways install structurally complex nucleobase modifications in DNA, termed DNA hypermodifications after those found in tRNA (named by Ross H. Hall in 1971), by harvesting building blocks from host cellular metabolites and appending functional groups pre- and/or post-replicatively, all without affecting the coding capacity of the bases[6,7,11]. Hypermodifications are found at key positions, or 'hotspots', on the nucleobases that do not interfere with Watson-Crick base pairing. These positions are tolerant of bulky, diverse adducts, and study of DNA hypermodifications from phage are expanding our understanding of DNA plasticity. Accordingly, the repertoire of phage enzymes that synthesize hypermodifications exhibit exquisite site-specific precision when conjugating functional groups to nucleobases, and like their bacterial counterparts involved in genome defense, have potential to be engineered for therapeutic and bio-control applications[12–14].

One hypermodification of historical significance was found in the genome of phage Mu, a cornerstone of bacterial genetics[15], over 40 years ago and subsequently mapped by restriction sensitivity to a loose SASNY consensus sequence in the Mu genome (accounting for 15-25% of total dA, IUPAC code given for consensus sequence)[16–22]. In 1983, McCloskey and coworkers

determined the structure of an unusual aminocarboxymethyl (carbamoylmethyl or glycinamide) group at the *N*6-position of adenine (abbreviated herein as 6-*N*cmdA, Fig. 1a); the structure of 6-*N*cmdA was further supported by MS/MS results in subsequent work[21,23–25]. The bulky aminocarboxymethyl group does not interfere with Watson-Crick base pairing, though there is a decrease in melting temperature associated with the hypermodification[21]. These foundational studies on Mu also identified a gene, *mom* (modification of Mu), necessary for the installation of 6-*N*cmdA during late stages of Mu infection[16–20,23,26–28]. A more recent study predicted that Mom belongs to the Gcn5-related N-Acetyltransferase (GNAT) superfamily, and another established that *mom* is the only Mu gene required for 6-*N*cmdA to form *in vivo* in *E. coli* (referred to colloquially as 'momylation')[23,28–31]. However, it has remained undetermined for over 40 years whether the gene product of *mom* is an active acyltransferase enzyme, and if so, what co-substrate is used and what functional group is transferred to dA[21,23,28]. Here, we demonstrate that the phage Mu enzyme Mom is an active aminoacyltransferase that co-opts charged $tRNA^{Gly}$ ($Gly\text{-}tRNA^{Gly}$) in order to form 6-*N*cmdA. Moreover, the mechanism uncovered herein reveals that the Mom-catalyzed hypermodification is not formed via simple group transfer, helping to explain how this reaction evaded elucidation for so long.

Results

The adenine hypermodification to 6-*N*cmdA forms post-replicatively on dsDNA

Since phage enzymes can install hypermodifications before or after replication, we first sought to determine if Mom can modify the DNA polymer directly. We assayed lysates for phage enzyme activity on a biotinylated DNA substrate[32,33]. We took clarified lysates derived from *E. coli* cultures heterologously expressing wild-type (WT) Mom, added biotinylated DNA substrates exogenously, incubated, recovered, and analyzed 2'-deoxynucleosides by LC/MS (Fig. 1a, *ex vivo* assay in Fig. 1b, Supplementary Fig. 1, and Supplementary Table 1). After incubation, a nucleoside identical in mass (308 u) and retention time to the native Mu 2'-deoxynucleoside was observed. This nucleoside product replaced about 20% of dA as compared to DNA incubated in control lysates. Furthermore, PacBio SMRT sequencing and sequence context sensitivity to restriction enzymes show Mu gDNA is specifically modified at dA in the context of an SAS sequence motif accounting for 20% of dA in Mu (Supplementary Fig. 2-5 and Supplementary Table 2)[20]. Incorporation of a nucleobase modified at the mononucleotide level would be expected to produce a polymer with 6-*N*cmdA found at random positions. Together, these results establish that momylation involves direct and site-specific modification of double-stranded DNA (dsDNA) that can be achieved *in vitro* at levels comparable to the native phage.

Adenine hypermodification to 6-*N*cmdA requires components of transcription/translation

Following the demonstration that Mom acts post-replicatively in lysates, we aimed to reconstitute the momylation reaction *in vitro* using experimental approaches amenable to systematic dissection of reaction components and potential group donors. Mom was either synthesized in an *in vitro* transcription-translation (IVTT) system from plasmid template or purified following recombinant expression in *E. coli* (Supplementary Fig. 6-8). Synthesis of Mom by the IVTT system resulted in active momylation of substrate DNA. By contrast, reconstitution of 6-*N*cmdA was not achieved with purified Mom until incubation of purified Mom with the IVTT system. Strikingly, installation of 6-*N*cmdA on DNA substrates by purified Mom is readily observable in reactions containing transcription and translation machinery without templates encoding for Mom. Together, these data

show that Mom, along with purified components of the transcription/translation system, are sufficient to reconstitute Mom activity.

Modification of dA by Mom using glycine is a tRNA-dependent process

The IVTT system that supports momylation contains a complex mixture of purified enzymes and cofactors, including tRNA, amino acids, and tRNA synthetases[34,35]. Some well-established enzymes of the GNAT superfamily use charged tRNA to aminoacylate well-known substrates[36–38]. We noticed that the aminocarboxymethyl group introduced by Mom and a glycyl group are isomeric, together suggesting that $tRNA^{Gly}$ and glycine may be involved in momylation. We tested this $tRNA^{Gly}$ hypothesis more directly in an enzyme coupled assay containing purified Mom (Δ10, a construct where the first 10 amino acids have been truncated), $tRNA^{Gly/GCC}$ transcribed *in vitro* (Supplementary Table 3, Supplementary Fig. 9-10, Supplementary Table 4), glycine, ATP, purified *E. coli* glycyl-tRNA synthetase (GlyRS), and linear dsDNA substrates. As can be seen in Fig. 1c and 1d, modification of DNA by Mom is a tRNA-dependent process: 6-*N*cmdA can only be detected when $tRNA^{Gly/GCC}$, GlyRS, glycine, and ATP are included in the reactions. The mass of 6-*N*cmdA formed during Mom assays using glycine labeled with ^{15}N and ^{13}C confirms that Mom transfers an activated glycine to dA (Supplementary Fig. 11).

Gly-$tRNA^{Gly}$ is the co-substrate used by Mom to hypermodify dA

Reconstitution of 6-*N*cmdA enabled us to probe the activated forms of glycine further in order to ascertain the co-substrate used by Mom at the transfer step. Two forms of activated glycine are generated *in situ* by GlyRS – glycyladenylate as an intermediate and Gly-$tRNA^{Gly}$ (charged tRNA) as a product[39–43]– in the coupled reaction where 6-*N*cmdA is produced during Mom activity assays. Our strategy involved generating a tRNA analog with a 2'-deoxyadenosine (from 2'-dATP) enzymatically installed by a CCA-adding enzyme at the 3' end of $tRNA^{Gly/GCC}$-CC following *in vitro* transcription (Fig. 2a, Supplementary Fig. 9, Supplementary Fig. 12-16, and see Methods). All tRNA species were verified by intact mass spectrometry[44–46]. We additionally verified that GlyRS can charge $tRNA^{Gly/GCC}$-CC-2'-dA[47–51], and we note that the charging (percent aminoacylated) for $tRNA^{Gly/GCC}$-CC-2'-dA is comparable to what we observe for the native sequence $tRNA^{Gly/GCC}$-CCA (Supplementary Fig. 17). The activity assays for GlyRS confirm that glycyladenylate *and* Gly-$tRNA^{Gly}$ were generated by GlyRS for both $tRNA^{Gly/GCC}$-CCA and $tRNA^{Gly/GCC}$-CC-2'-dA and that both activated forms of glycine would be available in an activity assay for Mom.

We next tested if $tRNA^{Gly/GCC}$-CC-2′-dA can support momylation. As can be seen in Fig. 2b (n=3 using a pUC19 plasmid substrate), momylation is not observed when $tRNA^{Gly/GCC}$-CC-2′-dA is included in a Mom assay, even after overnight incubation (Supplementary Fig. 18). In controls with $tRNA^{Gly/GCC}$-CCA (n=3 using a pUC19 plasmid substrate), momylation is first detectable after 5 minutes, and the majority of dA's in an SAS sequence motif (~25% of all dA's for pUC19) are typically modified around 60 minutes under the conditions tested (Fig. 2b and Supplementary Fig. 19). Though both forms of activated glycine were available to Mom in these activity assays, $tRNA^{Gly/GCC}$-CCA exclusively supports momylation, demonstrating how critical the terminal riboadenosine is for hypermodification of dA. Our activity assays for Mom and GlyRS are in agreement and confirm Gly-$tRNA^{Gly}$-CCA is the activated form of glycine used by Mom to hypermodify DNA.

Mom is selective for Gly-tRNA^{Gly} and dsDNA

Given the sensitivity of Mom for the terminal riboadenosine of tRNA, we further explored several features related to the co-substrate and substrate preferences of Mom. We noted that *E. coli* encodes for three different isoacceptors of tRNA^{Gly} – tRNA$^{Gly/GCC}$, tRNA$^{Gly/CCC}$, and tRNA$^{Gly/TCC}$ (Supplementary Table 4). While tRNA$^{Gly/GCC}$ is reported to be the most frequently-used and most abundant isoacceptor in *E. coli*[52–54], there is precedent for a particular tRNA isoacceptor to be used in a non-canonical role outside of ribosome-dependent protein synthesis[55,56]. We therefore evaluated if tRNA$^{Gly/CCC}$ and tRNA$^{Gly/TCC}$ (Supplementary Fig. 9) could support the momylation reaction. As can be seen in Fig. 2c, we see that all isoacceptors of tRNA^{Gly} can in fact support momylation. This finding is consistent with the strategy of the viral lifecycle, where available resources are scavenged rapidly in the process of hijacking the host cellular machinery.

We next investigated how selective Mom is for Gly-tRNA^{Gly} by utilizing an editing-deficient variant of AlaRS (C666A) to mischarge tRNA$^{Ala/TGC}$ (Supplementary Fig. 9, Supplementary Fig. 20)[34,57–59]. In activity assays for AlaRS, we verified that C666A AlaRS can form Ala-tRNA$^{Ala/TGC}$ and Gly-tRNA$^{Ala/TGC}$, while WT AlaRS only forms Ala-tRNA$^{Ala/TGC}$ (Supplementary Fig. 21). Under all conditions tested for WT Mom, 6-*N*cmdA could not be detected in assays where mischarged Gly-tRNA$^{Ala/TGC}$ or Ala-tRNA$^{Ala/TGC}$ were formed *in situ* (Fig. 2c). As has been shown for other GNATs[60,61], features at the 3′ end of tRNA and the amino acid appear to be sufficient to direct co-substrate specificity. Our data here suggest similar drivers of co-substrate recognition for Mom.

We also find that WT Mom exhibits a strong substrate preference between dsDNA and ssDNA (Fig. 2d). In addition to momylating biotinylated DNA and plasmid DNA, WT Mom and a Δ10 construct hypermodify an 11 bp dsDNA containing a single adenine in the SASNY context with a very high conversion rate (>95%). Mom exhibits activity at A/C MM (93% conversion) and to a lesser extent at an A/A mismatch (MM, 72% conversion per strand). Activity on an A/G MM (10%) is diminished compared to the A/T control. Mom activity on single-stranded DNA from phage M13 or on ssDNA 11-mers could not be detected above background (Supplementary Fig. 22). Taken together, we conclude that Mom is active on dsDNA (even on very short oligomers) but not on ssDNA.

Overall, we establish that Mom exhibits strong preferences to use the terminal riboadenosine of tRNA, Gly-tRNA^{Gly}, and target dsDNA. Specificity for co-substrates, particularly tRNA^{Gly}, has been attributed to evolutionary pressures on bacterial toxins belonging to the GNAT superfamily[62], though a general understanding of the factors that tune co-substrate recognition across GNATs has been elusive possibly due to characteristically high structural homology but very low sequence homology[36,37,63,64]. Substrate specificity and scope vary even more widely for GNATs. Here we add dsDNA to the multitude of targets modified by this enzyme superfamily.

6-*N*cmdA forms following an on-base isomerization of the exocyclic glycyl group

Having established that glycine is the group transferred during momylation, we sought to explain the apparent discrepancy between the structures of a modification formed by simple group transfer of the glycyl group to *N*6 versus the mature hypermodification. As highlighted above, the aminocarboxymethyl group of 6-*N*cmdA and a glycyl group are structural isomers. We therefore hypothesized that following transfer to the *N*6 position, the glycyl group necessarily undergoes

isomerization. We utilized LC/MS-MS and stable isotope labeling to investigate our hypothesis, focusing on a daughter ion of 6-*N*cmdA produced during fragmentation by bond breakage, shown as a dashed red line in top reaction of Fig. 3a, at the exocyclic C-C bond (*m/z* 148, termed $[6mA^{\bullet}]^+$ for clarity, see also Supplementary Fig. 23-25)[23]. For the coupled Mom activity assay including glycine-1-^{13}C,^{15}N and unlabeled dsDNA supported conversion of dA to 6-*N*cmdA (*m/z* = 311) (second reaction in Fig. 3a, Supplementary Fig. 11, Supplementary Fig. 26). Similarly, reactions containing unlabeled glycine and dsDNA labeled with $^{15}N_5$-dA (where all 5 purine nitrogens in dA are labeled with ^{15}N, see Methods for substrate preparation) supported conversion of dA to 6-*N*cmdA (*m/z* = 314, third reaction in Fig. 3a, Supplementary Fig. 27 and 28). Fragmentation patterns for 6-*N*cmdA labeled with glycine-1-^{13}C,^{15}N (*m/z* = 311) produced a daughter ion with an *m/z* = 149 that we assigned to $[6m\text{-}^{15}N\text{-}A^{\bullet}]^+$, one Dalton heavier than the product produced with unlabeled glycine. Given that the labeled nitrogen originated from the primary amine of glycine, the increase in mass here (relative to the unlabeled reaction) can be explained by a rearrangement that must have occurred during the modification reaction. Conversely, reactions utilizing a DNA substrate labeled with $^{15}N_5$-dA (*m/z* = 314) produced a daughter ion corresponding to $[6m\text{-}^{15}N_4\text{-}A^{\bullet}]^+$ (*m/z* = 152) that is 4 mass units (rather than 5) heavier than $[6mA^{\bullet}]^+$, as is observed in reactions without labeled components (compare, first and third reactions in Fig. 3a).

Because the nitrogen at *N*6 is a poor nucleophile due to resonance delocalization of the lone pair into the purine ring[65], we sought to assess the possibility that the more nucleophilic nitrogen at *N*1[66,67] carries out the initial nucleophilic attack, followed by migration of *N*1-glycyl to *N*6. Migration of a methyl group from *N*1 to *N*6 of adenine has been observed *in vitro* [68] via a Dimroth rearrangement[69], wherein the purine ring opens between *N*1 and *N*6 followed by rotation about the *C*5-*C*6 bond and ring closure, effectively swapping the *N*1 and *N*6 nitrogens together with their associated substituents. To our knowledge, a biologically catalyzed Dimroth rearrangement has not been reported, and it is unclear whether an aminoacyl group instead of an alkyl group could migrate through such a mechanism. We probed the possibility of glycyl addition at *N*1 by using a DNA substrate where dA is labeled specifically at *N*6 with ^{15}N (fourth reaction, Fig. 3a, Supplementary Fig. 23, Supplementary Fig. 29). If *N*1 were the site of glycyl addition followed by a Dimroth rearrangement, the labeled nitrogen originating at *N*6 should be retained in the purine ring at *N*1. However, we observe that the $[6mA^{\bullet}]^+$ daughter ion here is equal in mass (*m/z* = 148) to mass of the $[6mA^{\bullet}]^+$ daughter ion in the top reaction where unlabeled activity assay components were used. Because we find that the single, labeled nitrogen from *N*6 exchanged to the distal position and not to the *N*1 position, we conclude that *N*6 is the nucleophile that makes the attack on Gly-tRNA^{Gly}. Other GNAT superfamily members are known to acylate exocyclic amines. For example, TmcA from *E. coli* acetylates tRNA^{Met} at the *N*4 position of the wobble base (C34) to form ac4C, a tRNA modification found in all domains of life[36].

Energetics of the on-base rearrangement leading to 6-*N*cmdA

Our labeling experiments indicate an exchange between the *N*6 proximal nitrogen and the distal primary amine occurs during momylation (chemically, a transaminoacylation followed by an isomerization). This "flipping" of the transferred group places the glycyl C_α one atom closer to the nucleobase and apparently displaces the carbonyl oxygen one atom further away from dA. To account for the nitrogen exchange, we propose a mechanism for isomerization that proceeds via a five-membered ring intermediate through ring closure upon attack of *C*6 by the exocyclic primary amine followed by ring opening through breakage of the initial *C*6-*N*6 bond to yield the native hypermodification found in phage Mu (Fig. 3b).

We have not seen an analogous multistep and controlled process – transfer and then rearrangement of the transfer group to one product during hypermodification of a nucleobase – driven by a single enzyme described in the literature. Established acylation chemistry characteristic of GNAT enzymes is notably distinct from the chemistry proposed for isomerization of an exocyclic glycyl group on dA. These factors motivated quantum-chemical exploration of the rearrangement in water in the absence of enzyme using the ORCA package (Supplementary Fig. 30-37)[70]. We calculated re-arrangement energies for educt and product dihedral angles using the efficient R2SCAN functional (with dispersion correction, in implicit water). We separately considered the neutral (Supplementary Fig. 30-33) and cationic forms (Supplementary Fig. 34-37) allowing for the glycyl-*N*6 amide bond to occur in the trans- or cis-conformation. A dihedral scan shows that the *N*1-*C*6-*N*6-C angle tends to be close to 0 or 180° (Supplementary Fig. 30 and 34), presumably to conjugate the *N*6 lone electron pair to the aromatic system of the adenine nucleobase. In the neutral form, the glycyl *N*6 trans-amide bond is favored (Supplementary Fig. 30 and 31). In the cationic form, the cis-amide, which is poised for the rearrangement reaction, is preferred (Supplementary Fig. 34 and 35). For both neutral and cationic forms, the glycyl is preferably oriented towards the Watson-Crick edge, indicating that without the rearrangement reaction, the glycyl group on *N*6 would interfere with Watson-Crick base pairing. In the product state, the preference for the Watson-Crick edge over the Hoogsteen edge is mitigated for the neutral form (Supplementary Fig. 30 and 31) or reversed for the cationic form (Supplementary Fig. 34 and 35). Thus, the rearrangement alleviates the impairment of Watson-Crick base pairing from glycyl transfer (Supplementary Fig. 38). Overall, the rearrangement reaction is favorable for both the neutral and cationic forms (Supplementary Fig. 30, 31, 34, 35). In the latter case, deprotonation of the *N*1 after the rearrangement adds to the favorability. For both the neutral and cationic forms, the free energy of the spirocyclic intermediate is also relatively low, and not much higher than the energy required to push the glycyl group of the educt "out of plane" (Supplementary Fig. 30, 31, 34, 35).

To determine whether the rearrangement reaction could occur spontaneously in water, we analyzed plausible reaction pathways (Supplementary Fig. 32, 33, 36, 37). For the neutral form, we assumed nucleophilic attack of the glycyl Nα on the *C*6 (Supplementary Fig. 32b) and proton transfer from glycyl Nα to adenine *N*1 (Supplementary Fig. 32c) to form the spirocyclic reaction intermediate. For the resolution of the intermediate, we tested proton transfer from the adenine *N*1 to the adenine *N*6 (Supplementary Fig. 33b), and departure of the *N*6 leaving group (Supplementary Fig. 33c). For the cationic form, we assumed proton transfer from Nα to *N*1 (Supplementary Fig. 36b) and Nα to *C*6 bond formation to form the intermediate (Supplementary Fig. 36c). For the resolution of the intermediate, we postulated proton transfer from *N*1 to *N*6 with *C*6-*N*6 bond scission (Supplementary Fig. 37b), and then proton transfer from Nα to *N*1 (Supplementary Fig. 37c). For both the neutral and cationic forms, formation of the spirocyclic intermediate was possible via relatively low-lying saddle points/transition states. In the case of the cationic form, these were sufficiently low (~17 kcal/mol) for a reaction rate above 10/s at room temperature according to Eyring's equation. By contrast, the barriers for the resolution of the spirocyclic intermediate came out high (Supplementary Fig. 33 and 37). This could indicate either that the optimal reaction path for a spontaneous resolution of the spirocyclic intermediate remains to be identified or that Mom facilitates the rearrangement reaction by general acid-base chemistry or through a concerted protonation/deprotonation step (Supplementary Fig. 39).

The crystal structure of S124A Mom reveals putative active site and distinct features

In order to gain a more detailed picture of momylation, we aimed to solve a crystal structure of Mom. Crystallization screens identified conditions leading to the growth of orthorhombic crystals of Mom(Δ10) S124A (a catalytically dead variant) in space group P2(1)2(1)2(1) that contained two Mom protomers in the asymmetric unit and that diffracted to ~2.0 Å resolution (Fig. 4a). The structure was solved by molecular replacement using an AlphaFold2 generated search model of the Mom protomer and refined with manual edits in Coot (Supplementary Table 5)[71–73].

As anticipated[23,28,73], Mu Mom protomers have the GNAT fold, but with an atypical insertion spanning residues ~175-209 that visually resembles a helix-turn-helix (HTH) motif (Fig. 4a-c)[63,74,75]. An insertion can also be seen with superpositions and sequence alignments of Mu Mom with the predicted structures of active homologs mined from metagenome databases (Fig. 4b-c, Supplementary Fig. 40 and Supplementary Fig. 41). Importantly, deletion of the HTH-like insertion leads to complete loss of the activity (Supplementary Fig. 5), highlighting its crucial role in momylation.

Previous work has established that the truncation of 10 N-terminal residues (Δ10) of Mom is well tolerated[23], which we also confirm. Notably, however, the truncation of 20 amino acids (Δ20) reduced the activity (Supplementary Fig. 5a, Supplementary Fig. 22a, Supplementary Fig. 42, Methods, and Supplementary Methods). This is surprising because the 11-19 region is unstructured, and our electron map does not contain densities that can be attributed to this stretch of amino acids.

Superposition of 51 diverse GNAT crystal structures solved with CoA or a CoA analogs and Mom suggested residues H58, Y59, R111, S124 D149, R151, and Y159 potentially encompass key active site residues (Fig. 4d, Supplementary Fig. 43, Supplementary Table 6, and Supplementary Table 7). This finding was in line with previous work that showed that alanine mutations of the residues R111, S124, and D149 greatly reduced Mom activity[23]. As our own *in vivo* and *in vitro* data reproduced these observations for R111A and S124A, we decided to carry out a more comprehensive structure guided mutagenesis screen informed by our modeling (Supplementary Fig. 42). In doing so, we identified another essential residue, R151, as the R151A variant lacked any detectable activity *in vivo*. In addition, the activity of R111K, H58A, Y59A, W48A, Y159A, and W181A variants were drastically reduced.

Superposition of Mom with the subset of those GNAT structures solved with the substrate or product (Fig. 4e) further provide insights to the role of other residues, such as R111, critical for momylation. Superposition of Mom with two GNATS where tRNA is a co-substrate or substrate – $FemX_{Wv}$ (from *Weissella viridescens*, PDB ID: 4II9) and TacT (from *Salmonella typhimurium*, PDB ID: 7F36), respectively – reveals two possible binding modes of tRNA each that could position the CCA stem in proximity to residues critical for activity (Fig. 4f, Supplementary Fig. 43, and Supplementary Table 6)[76]. During momylation, a FemX-like strategy could be used for binding the co-substrate (tRNA), while a TacT-like strategy could be used for binding the substrate (DNA), aided by the positively charged cleft (shown in blue in Fig. 4g-h).

Consistent with size exclusion chromatography results (Supplementary Fig. 8), the structure shows that Mom forms tight dimers with an interface area of ~850 $Å^2$, and a predicted free energy of binding of 4.7 kcal/mol according to the PISA server[77]. Based on the structure, we have tested the dimer surface by mutating the residues Y28 and W48 to alanine, which we hypothesized were

potentially involved in dimerization. While mutation did not have a significant effect for Y28A, W48A severely reduced the activity of Mom *in vivo* (Supplementary Fig. 5 and Supplementary Fig. 42). To test the role of W48 on Mom dimerization and activity further, we purified the W48A variant and compared its migration on size-exclusion column side-by-side with S124A variant. As predicted, Mom W48A migrated as a monomer (Supplementary Fig. 44a-c) and only had residual activity *in vitro* (Supplementary Fig. 44d).

The crystal structure of Mom alone combined with structure guided mutagenesis provided key insights in the understanding of residues important for momylation. Because structures of Mom bound to DNA, tRNA, or both have not been solved to date, we utilized AlphaFold3 (AF3)[78] modeling of Mom with $tRNA^{Gly}$ and dsDNA (Supplementary Fig. 42e-f, Supplementary Results, and Supplementary Methods). To verify that AF3 models were reliable, we also tested several other GNAT members that are known to utilize tRNA as an acyl donor or to modify tRNA (Supplementary Fig. 45, Supplementary Fig. 46, and Supplementary Table 10).

The AF3-generated model of Mom in a ternary complex supports our conclusions on relative orientation of tRNA and DNA based on the extensive structural comparison. Moreover, the accumulation of positive charge in the 11-19 amino acid unstructured region could contribute to tRNA binding, explaining the reduced activity of Δ20 variant (Supplementary Fig. 42g). Satisfyingly, the AF3 model also identifies the positively charged cleft formed between HTH-like insertion and the GNAT part of Mom as DNA binding region, suggesting that HTH-like insertion could interact with DNA through the major groove (Supplementary Fig. 42e-f). The AF3-generated model additionally suggests that substrate dA is flipped-out from the DNA double helix and positioned in close proximity to the 3'-terminal ribose of the CCA stem of tRNA and, importantly, both are surrounded by the essential residues, e.g. R111, D149 and R151 (Supplementary Fig. 42f).

Discussion

Presented here is the mechanistic explanation behind a genetically simple but elegant strategy used by the phage Mu-encoded Mom enzyme to protect DNA from endonucleases. It has been known for decades that Mom catalyzes carbamoylmethyl addition to the *N*6 group of adenine. However, the source of the carbamoylmethyl group remained unknown. The enigma deepened when it was demonstrated that Mom belonged to the GNAT family of acyltransferases, where substrates are typically acylated or aminoacylated, seemingly at odds with the chemically established structure for adenine hypermodified by Mom. Here, we confirm that the carbamoylmethyl modification is indeed the product of momylation and establish that the source of the modification is activated glycine from charged $tRNA^{Gly}$. Similar to other GNATs, we further establish that Mom is selective for glycine carried by $tRNA^{Gly}$. Because all $tRNA^{Gly}$ isoacceptors support momlyation, we speculate that Mom could be recognizing a small portion of charged tRNA molecule, most likely at the 3´-acceptor end (5'-tRNA-CCA-Gly)[60,79]. At present, we cannot distinguish whether acyl transfer is direct, or whether glycyl is first transferred from a tRNA molecule to one (or more) carrier residues in the active site of Mom before the transfer to adenine. A direct transfer of the glycyl group to *N*6 would require spatial proximity between the tRNA acceptor site and the binding site for the substrate adenine. An intermediate transesterification step facilitated by Mom, as has been reported for some GNATs[38], would relax this requirement and remains to be explored.

Based on the results of isotope labeling experiments and supported by our quantum chemical calculations, we propose a rearrangement reaction that converts *N*6-glycyladenine to *N*6-carbamyolmethyladenine. The product of momylation is different from tricyclic compounds previously obtained from aminoacylated adenosines under acidic, abiotic conditions[80–83]. Overall, the rearrangement after Mom-catalyzed glycyl transfer is favorable in both neutral and cationic forms. Critically, the rearrangement alleviates steric clashing at the base pairing interface resulting from aminoacylation of adenine.

Our data also shed light on how Mom achieves *N*6 acylation of glycine. As an aromatic amine, the *N*6 on adenine is not very nucleophilic. Propensity for alkylation damage, a measure of nucleophilicity, is concentrated on the *N*1 and *N*3 nitrogens, rather than the *N*6 nitrogen of adenine[66,84]. Hence, Mom would need to use an activated form of glycine, rather than free glycine, to facilitate the glycyl transfer reaction. Therefore, the harvest of glycyl from $tRNA^{Gly}$ represents an efficient solution this problem, exploiting a readily available resource in the cell. The strategy to use tRNA activated amino acids for aminoacylation of peptides, proteins, lipids, and natural products, has been adopted by some other GNAT acyltransferases (Fig. 5)[85]. In case of Mom, potential adverse effects of co-opting host translation machinery during infection may be mitigated by the expression of Mom in the late stages of the lytic infection cycle[86]. Notably, there is precedent in the literature for the acylation of aromatic amines by a GNAT family member, TmcA, which acetylates the *N*4 of cytosine in tRNA[87]. However, this enzyme uses acetyl-CoA, i.e. a thioester, as the source of the activated carboxylic acid. The use of charged tRNA, and hence of an ester as the source of carboxylic acid for acylation of an aromatic amine, as in the primary step of the Mom catalyzed reaction, defines not only a new category of GNAT enzymes but also a new type of functionality.

Within the landscape of tRNA hypermodifications, a carbamoylmethyl (ncm^5) group is also found on the *C*5 carbon of the wobble uridine U_{34} in tRNA of nearly all eukaryotes. It is widely thought that this carbamoylmethyl group represents an amidated version of the carboxymethyl group introduced in a radical reaction catalyzed by the multi-subunit Elongator complex[88,89]. If so, then carbamoylmethyl modifications found in tRNA versus DNA differ not only in the anchoring group (*N* versus *C*) but are also introduced by chemically and evolutionarily unrelated mechanisms. As such, the discoveries presented herein broadly highlight the diverse biochemical strategies used to install critical hypermodifications onto nucleic acids, provide a foundation for further investigation of the DNA hypermodification reactions made by Mom-like GNAT enzymes, and underscore the versatility of the GNAT fold in enzyme evolution.

Methods

Phage Mu preparation, genomic DNA isolation, and SMRT Sequencing

Two *E. coli* strains carrying phage Mu lysogen were obtained from Prof. Stan Hattman and were used to produce Mu phages. Of the two strains, one contains the Mu lysogen (denoted as ER3153; genotype: *sup mcrA thi hsdR mcrB zzz::Mu(Cts62)*) which produces the Mu phage with the native 6-*N*cmdA DNA modification phenotype, and the other has a mutant Mu lysogen which produces Mu phages without 6-*N*cmdA DNA modification but canonical 2´-deoxyadenosine phenotype (denoted as ER3154; genotype: *sup mcrA thi hsdR mcrB zzz::Mu(Cts62 mom)*).

A heat-shock procedure was performed to induce the production of Mu phage from the lysogen. LB medium was inoculated with the phage lysogen strain from an overnight culture and was incubated at 37 °C with agitation until an OD_{600} reading of ~1. The culture was then incubated at 45 °C for 30 minutes and then again at 37 °C with agitation. Phage were harvested when the culture appeared translucent (cells lysed), a process that normally takes about 1.5 – 3 h after heat-shock. The phage lysate solution was then centrifuged at 10000 rcf x g at 4 °C for 10 min, and the supernatant was collected. The phage was then concentrated first by precipitation in a solution of 10% (w/v) poly(ethylene glycol) average MW 8 kDa and 1 M NaCl, followed by pelleting by centrifugation, and finally resuspension in a buffer composed of 50 mM Tris-HCl 7.5, 10 mM $MgCl_2$ and 75 mM NaCl. The phage solution was stored at 4 °C in the dark. A previously published protocol was followed to extract and purify phage genomic DNA[90].

Genomic DNA isolated from phage Mu was used to prepare the single-molecular real-time (SMRT) sequencing libraries. SMRT sequencing reads were analyzed using SMRT Analysis software and pipelines from SMRT Link webtool (version 10.2.0) hosted by Pacific Biosciences. SMRT sequencing raw CCS reads were processed to generate HiFi reads in which the output sequences were used for the *de novo* assembly of the genome using MEGAHIT assembler software (version 1.2.9) [91]. The longest assembled output sequence was used as the reference for subsequent SMRT Analyses steps. Base modification detection and motif analysis were performed with the same SMRT Link version. To detect the base modifications, the Base Modification Analysis application was used to first map the raw reads to the assembled genome and the mapped reads were analyzed further to extract the inter-pulse duration (IPD) kinetics information collected during sequencing course and processed for all pulses aligned to each position in the reference sequence using an in silico kinetic reference and a t-test based kinetic score detection of modified base positions. The per-base IPD ratio mapped outputs were further used to extract the modification motifs. The top three identified modification motif sequences were extracted separately and the sequences for each motif string were used to generate sequence logo by using WebLogo 3 webtool (version 3.7). See Supplementary Fig. 2 and Supplementary Table 2.

DNA Substrates

Biotinylated DNA substrates were prepared by digesting λ DNA (NEB) with MseI (NEB) as described[32,33,90]. Digested λ DNA was backfilled with 50 µM Biotin-16-(5-aminoallyl)-dUTP (Jena Bioscience) using Klenow Fragment (3´→5´ exo-) (NEB). Empty plasmids of pUC19 (NEB) and pET28a were used as supplied. Double-stranded DNA oligomers were obtained by heating a 1:1 mix of complementary oligonucleotides at 90 °C for 5 min and annealing by gradual cooling to the room temperature for 30 min. Duplexes were stored at -20 °C until use. WM_{Mu} 34-mer was resuspended in Duplex Buffer (IDT), complementary strands were mixed in equimoloar ratios, and annealed by heating at 95 °C for 2 minutes and cooling to room temperature over a linear gradient. The 3 kbp substrates were amplified with a standard dNTP mix, $^{15}N_5$-dATP (Cambridge

Isotope Laboratories), dGTP, dTTP, and dCTP, or 6-^{15}N-dATP, dGTP, dTTP, and dCTP. 6-^{15}N-dA (Toronto Research Chemicals) was chemically phosphorylated (see Supplemental Methods). See Supplementary Table 1 for sequences.

Restriction digest protection assay

To assay for installation of 6-*N*cmdA by Mom, Mu gDNA or DNA substrates of Mom were subject to restriction digests with 6-*N*cmdA-sensitive restriction enzymes HgaI (NEB) or AasI FastDigest (Thermo) according to the manufacturer's instructions[23]. The digestion products were resolved on an 0.8% agarose gel and imaged using GelRed (Biotium) staining.

Constructs, cloning, and strains

Escherichia coli NEB® 5-alpha (NEB) or TOP10 strains (Invitrogen) was used for cloning. *E. coli* T7 Express (NEB) or BL21-CodonPlus (DE3)-RIL strains (Agilent) was used for heterologous protein expression. Plasmids were stored at -20 °C until use. See Supplementary Table 8 and Supplementary Table 9 for cloning primers sequences and a list of expression plasmids.

Mom. Mom was cloned into the pRY plasmid (derived from or pET28c) or pET28a. Expression cells harboring the pRY Mu Mom plasmid were used in downstream *in vivo* and *ex vivo* Mom activity assays (see below). Expression cells harboring pET28a plasmids expression WT Mom or Mom variants were used for *in vivo* Mom activity assays or for purifying enzyme for crystallography.

Glycyl-tRNA synthetase (GlyRS). The *E. coli* GlyRS expression construct, pET21a-GlyRS-His_6, was a gift from Sebastian Maerkl & Takuya Ueda (Addgene plasmid #124110; http://n2t.net/addgene:124110; RRID:Addgene_124110).

Alanyl-tRNA synthetase (AlaRS). C666A AlaRS[57–59] was generated on a pEQ expression plasmid using mutagenesis primers given in Supplementary Table 8.

tRNA nucleotidyltransferase cca2 (CCA2). CCA2 from *Schizosaccharomyces pombe* (*S. pombe*) was ordered from Twist on a pET28b plasmid with an N-terminal His_6 tag and TEV protease cleavage site.

E. coli tRNA genes. The pUC19 plasmid containing the gene encoding the $tRNA^{Gly/GCC}$ isoacceptor from *E. coli* was kindly gifted by Dr. Haruichi Asahara. The genes encoding for $tRNA^{Gly/CCC}$, $tRNA^{Gly/TCC}$, and $tRNA^{Ala/TGC}$ were ordered from GenScript or Twist on a pUC19 plasmid. Sequences of the tRNA isoacceptors from *E. coli* were accessed at the GtRNAdb Database[52,53].

***In vivo* Mom activity assays**

To test the *in vivo* activity of Mom, variants, and mined homologs (see below), expression plasmids from cells harvested after recombinant expression (as described in the "Protein Expression" section below and in Supplementary Methods) were isolated by miniprepping as described[32,92,93]. As a negative control, expression strains bearing the pET28a vector with no insert ("Empty") were used. Plasmids were assayed for the presence of hypermodification as described below.

UHPLC-MS of nucleosides and oligonucleotides and UHPLC-MS/MS analysis of nucleosides

Nucleoside samples were prepared from dsDNA substrates from *in vivo*, *ex vivo*, and *in vitro* Mom activity assays by using the Nucleoside Digest Mix Kit (NEB) according to the manufacturer's instructions and as previously described[32,33,90]. Nucleosides were subjected to ultra-high-performance liquid chromatography (UHPLC), UHPLC-mass spectrometry (UHPLC-MS)[32,33,90], or LC-MS/MS. Agilent ChemStation software and Adobe Illustrator were used for data processing

and visualization. Nucleosides were quantified as previously described[94–96]. tRNAs were verified by intact mass spectrometry previously described[44–46].

Recombinant protein expression of Mom

Mom was heterologously expressed in T7 Express or BL21-CodonPlus (DE3)-RIL strains. A single colony from freshly transformed T7 Express cells was used to inoculate LB/Kan (40 µg/mL) media (+ glucose, 0.1%). The starter culture was grown overnight at 37 °C at 240 rpm. A 1:100 to 1:1000 dilution of the overnight culture was made into fresh expression LB media (- glucose). Cells were grown to an OD_{600} of 0.6 at 240 rpm and expression was induced with 100 µM IPTG. Cells were grown for 18 hours at 18 °C after which they were harvested by centrifugation and stored at -20 or -80 °C until use. Mom was also heterologously expressed in the BL21-CodonPlus (DE3)-RIL strain as was GlyRS in LB media under the selective pressure of kanamycin-chloramphenicol and ampicillin-chloramphenicol, respectively. Cells were used to inoculate starter cultures that were grown in the presence of 1% glucose up to a late log phase at 37 °C in a rotary shaker at 140 rpm. A 1:200 dilution of starter culture was used to inoculate expression media. After the OD_{600} reached 0.6-0.7, the expression was induced with 250 µM IPTG. Cells were incubated with 140 rpm shaking for 18-20 h at 25 °C. Cells were then collected by centrifugation, washed with PBS (pH 7.4) and stored at -20 °C.

***Ex vivo* Mom activity assays**

Clarified lysates were prepared from cells where WT Mom or variants had been heterologously expressed[32,33,90]. *Ex vivo* assay reactions were prepared on ice and incubated with biotinylated DNA substrates at 37 °C for > 15 min as previously described[32,33,90]. Where indicated, reactions were supplemented with 1 mM $^{13}C_2$-acetyl-CoA (Sigma Aldrich) (comparable to the intracellular concentration of acetyl-CoA, BNID 114622)[97] that had been stored in 1 M aliquots at -80 °C in water until use. Streptavidin Magnetic Beads (NEB) were used to recover biotinylated DNA substrates from enzymatic reactions. Nucleoside digests were analyzed by UHPLC-MS as detailed above.

Protein purification

All protein purification steps were carried out at the temperature range 0 – 12 °C. All buffers were filtered (0.45 µm) and degassed. Equipment used for protein purification was ÄKTA Purifier FPLC system (GE Healthcare) and Minipuls 3 peristaltic pump (Gilson). Buffer composition for purification of Mom and variants is described in Supplementary Methods. N-ter 6xHis-Tag full-length Mom WT and R111A and 6xHis-SUMO-Δ10 Mom WT and S124A were purified using our original protocol that involved Ni-NTA, Heparin, and size-exclusion chromatography. *E. coli* GlyRS was also purified according to the protocol from Yoshihiro Shimizu & Takuya Ueda with additional steps[34]. Detailed protocols can be found in the Supplementary Methods. Aliquots of *E. coli* WT GlyRS and WT AlaRS were also kindly gifted by Corinna Tuckey, Dr. Ying Zhou, Dr. Bradley J. Landgraf, and Dr. Emily Chen. The expression and purification scheme for WT CC2A from *S. pombe* was kindly shared by Dr. Andew P. Sikkema and can be found in the Supplementary Methods. A contaminating RNase was removed using an immobilized aptamer[98]. C666A AlaRS was expressed and purified[57–59] (see also Supplementary Methods.).

Mom activity assays with the PURExpress® system

For cell free protein synthesis (also referred as *in vitro* transcription and translation or IVTT), the PURExpress® system (NEB) was used with 1000 ng of pRY WT Mom plasmid as the template

in a reaction incubated at 25 °C for 24 hours according to the manufacturer's instructions[34,35]. Nucleosides from recovered plasmids were prepared and analyzed by UHPLC-MS. The PURExpress® system was also utilized for activity assays (without plasmid template), which included PURExpress® kit components, purified WT Mom (2 μM), $tRNA^{Gly/GCC}$ (made by *in vitro* transcription, see below), and 1000 ng of biotinylated DNA substrates where indicated. Reactions were incubated at 25 °C for 24 hours or 37 °C for ≥ 2 hours and biotinylated substrates were recovered, digested, and analyzed by UHPLC-MS.

***In vitro* transcription (IVT) of tRNA**

Templates for IVT were amplified by PCR using Q5® High-Fidelity DNA Polymerase or Phusion™ DNA Polymerase (NEB). Unmethylated primers were used for amplification by Q5. Standard IVT reaction conditions using the HiScribe® T7 Quick High Yield RNA Synthesis Kit (NEB) were used for $tRNA^{Gly}$ (all isoacceptors) and $tRNA^{Ala/TGC}$ (see Supplementary Table 4 for sequences). $tRNA^{Gly/GCC}$ was also transcribed using purified T7 polymerase (kindly gifted by Prof. Janusz Bujnicki's lab) in previously described conditions[48,52,53,99]. In some cases, repression of nontemplated additions at the 3' end by T7 RNA polymerase was needed. For full-length $tRNA^{Gly/GCC}$-CCA (76 bases), 10% (v/v) DMSO was added to the IVT reaction mixture. For $tRNA^{Gly/GCC}$-CC (75 bases), reverse primers including a 2'-*O*-methylated base second from the 5' end were used during PCR amplification by Phusion. 10% (v/v) DMSO was also added the IVT reaction mixture containing 2'-*O*-methylated PCR products. In all cases, GMP was added to a final concentration of 20 mM to the IVT reaction mixture to favor the 5'-monophosphate product. tRNAs were purified using the Monarch® RNA Cleanup Kit (NEB), characterized by urea-PAGE, and where indicated by intact mass spectrometry.

***In vitro* coupled Mom activity assay with nucleoside analysis**

Components for coupled Mom activity assays including GlyRS were mixed on ice in the following order to a final volume of 30 μL and incubated at 37 °C for the indicated durations: activity buffer 1 (AB1, See Supplementary Table 3 for activity buffer compositions); Pyrophosphatase, Inorganic (0.025 units/mL, NEB); GlyRS (2-4 nM); glycine (2 mM); RNase Inhibitor, Murine (20 units, NEB); Mom (2 μM); DNA substrate (typically 1 μg but could range from 250 ng – 10 μg); $tRNA^{Gly}$ (~3 μg [4 μM]); and RNase-free water[48,61]. Where specified, glycine-1-^{13}C,^{15}N (2 mM, Sigma) was used in place of glycine. Reaction components were omitted where specified in the text. After incubation, all reactions treated with 1-3 μL RNase A (NEB) for > 30 min. For time courses, samples were also treated with 1-3 μL of ProtK (NEB). Biotinylated substrates were prepared for UHPLC-MS analysis as described above. The 3 kbp substrates, oligomer substrates, and plasmid substrates were purified using the Monarch® PCR and DNA Cleanup Kit (NEB) with the DNA Cleanup and Concentration and the Oligonucleotide Cleanup protocols, respectively. The 3 kbp substrates, oligomer substrates, and plasmid substrates were prepared for UHPLC-MS analysis as described above. The *in vitro* activity of Mom on short DNA was tested in Activity Buffer 2 (AB2) in 100 μL reaction volume. The mixture contained 17 μg of 11 bp or 3 bp DNA, 0.6 μM Mom protein (monomer), 1.6 μM tRNA, 0.09 μM GlyRS (hetero-tetramer), 10 mM glycine, and 2 mM ATP. The reaction was incubated at 37 °C for 5 hours followed by nucleoside analysis[100]. For Mom assays including WT AlaRS or C666A AlaRS, 7 μM of enzyme and 14 μM of tRNA were included[34].

***In vitro* coupled Mom activity assay with restriction challenge**

Activity assays with radiolabeled substrates included 2 pmoles of 34-mer dsDNA and other reaction components (in AB2 and assay conditions used for short DNA, see previous section) with an overnight incubation. After incubation, purified DNA was treated by AasI FastDigest (Thermo), and the digestion products were resolved on 20% Urea-PAGE and visualized by autoradiography using Amersham Typhoon Biomolecular Imager. The activity of Mom with and its variants was also tested on 3.6 µg of substrate plasmid (pET28a-empty in activity assay conditions for short DNA). Purified plasmid DNA was subjected to restriction digestion as above.

Single-nucleotide extension of tRNA using CCA2

rA from ATP (NEB) or 2'-dA from 2'-dATP (NEB) was added to the 3' end of $tRNA^{Gly/GCC}$-CC using purified WT CCA2 from *S. pombe*[47,101–103] (see also Supplementary Methods and Supplementary Fig. 14) at 25 °C under single turnover conditions at a ~300 µg scale in activity buffer containing 50 mM Tris (pH 8), 100 mM KCl, 1 mM TCEP, 10 mM magnesium acetate, 1 µM BSA, and 1-2 mM (d)NTP. tRNA was cleaned up using spin columns, treated with apyrase to deplete nucleotide triphosphate carried over during purification, cleaned up using spin columns, and extension was verified by intact mass spectrometry as above.

Charging, labeling, and quantification of tRNA aminoacylation

tRNA (~7 µM) was aminoacylated (charged) for 20-30 minutes at 37 °C in 100 mM HEPES pH 7.6, 2 mM ATP, 20 mM $MgCl_2$, 1 mM DTT, 7 µM aaRS, Pyrophosphatase, Inorganic (NEB, 5 units), and 2 mM amino acid in a total volume of 50 µL as described[48–51]. Reactions were quenched with 1/10 volume of 3 M sodium acetate (pH 5.2), purified by ethanol precipitation, and resuspended in 40 µL of 5 mM sodium acetate (pH 5.2). The Qubit™ Broad Range Assay Kit (Invitrogen™) was used to quantify the tRNA recovered from the aminoacylation reaction, and samples were stored at -20 °C until use. Roughly 150 pmol (~3.6 µg, ~2.5 µM) of tRNA recovered from the charging reaction was labeled with EZ-Link™ Sulfo-NHS-LC-Biotin (Thermo Scientific™) that was freshly prepared in RNase-free water (15 mM) in a total volume of 60 µL on ice. Labeling reactions were cleaned up with a Micro-Bio Spin 6 column pre-equilibrated with 5 mM sodium acetate (pH 5.2). About 8 pmol of recovered tRNA was incubated with 8 µL of streptavidin (1 mg/mL) at room temperature for 20 min. tRNA (200 ng total) was mixed 1:1 with acidic loading dye (100 mM sodium acetate [pH 5.2], 7.5 mM urea, 0.05 bromophenol blue) without heating and resolved by urea-PAGE using 6% Novex™ TBE-Urea Gels, Invitrogen™ at 200 V for 30 min. Gels stained with SYBR™ Gold were visualized using an Amersham Typhoon Biomolecular Imager, and ImageLab (Bio-Rad) was used to quantify band intensities of uncharged and charged (shifted) bands. Percent charging was calculated by (shifted band)/(shifted band + unshifted band). Generally, 20-60% charging was observed.

Mom crystallization and structure determination

Purified Δ10 Mom S124A protein was applied to a Superdex 200 Increase 10/300 GL column equilibrated with Buffer 9 (20 mM HEPES-KOH, pH 7.5, 550 mM NaCl, 50 mM KCl, 5% glycerol, 10 mM 2-ME). The best fractions were combined and concentrated to ~23 mg/mL using an Amicon 10 MWCO centrifugal filter unit. Mom crystals were grown by mixing 2 µL of the protein with 2 µL of the condition H4 of the PACT premier crystal screen (Molecular Dimensions) (0.2 M Potassium Thiocyanate, 0.1 M Bis–Tris propane, pH 8.5, 20% PEG3350) using the vapor diffusion method in hanging drops. For cryo-protection, crystals were soaked in mother liquor containing 25% glycerol and flash-cooled in liquid nitrogen. Mom crystal diffraction data was collected on PETRAIII P11 beamline at the DESY synchrotron. Crystals had a space group $P\ 2_1$

$2_1\ 2_1$ and diffracted up to 2.03Å at using X-ray wavelength of 1.0332Å. The diffraction data was automatically processed by the XDSAPP program at the synchrotron [104]. The structure of Mom was solved by molecular replacement using the Phaser program and the Alphafold 2 predicted model [71,73]. The programs Phenix and Coot were used for the refinement and manual building [72,105]. The final model had following Ramachandran statistics (%): favored 98.3, allowed 1.7, outliers 0.00. The model contained 1.6% rotamer outliers and had a MolProbity clashscore of 2. The data collection and refinement statistics are presented in Supplementary Table 5. Representative 2mFo-DFc and composite omit electron density maps supporting data quality are depicted in Supplementary Fig. 47. The final model coordinates and the structure factors were deposited in Protein Data Bank under accession code: 8BV8.

3D structural analysis and graphic representation

Structure manipulation and image rendering was performed using UCSF ChimeraX program[106]. The MatchMaker tool in ChimeraX was used for structure superposition. Secondary structure assessment (Fig. 3b and 3c) was guided by the PDBsum server[75]. Vector Alignment Search Tool (VAST) was used to search for the structural similarity between Mom and the available GNAT structures on the medium redundancy setting[107]. GNATs with productively bound CoA and acyl-CoA were used for superposition in Fig. 3d and Fig. 3e and were selected from the top 164 hits (the cut-off was manually determined). A subset (11) of retrieved GNATs with productively bound substrates or products were used for superposition in Fig. 4f. See Supplementary Table 6 and Supplementary Table 7.

Metagenome mining and selection of homologs to screen for activity *in vivo*

A search for homologs of Mom in metavirome databases was completed as described[32,33,92,93]. Briefly, using Mu Mom as a query sequence, Geneious Prime software (v 2020.1.1) was used to run a Basic Local Alignment Search Tool (BLAST) search of the Nucleotide collection (nr/nt). 500 sequences only containing Mom domains were used to build an HMM profile with the hmmbuild program from the HMMER software package[108]. The hmm profile for Mom was used to search two metavirome sequence databases: the Joint Genome Institute's Integrated Microbial Genomes Viral Resource version 2 (IMG/VR v2)[109] and the Global Oceanic Viromes 2.0 (GOV2.0)[110]. Open reading frames in the recovered contigs (>8,000) were predicted and annotated using Prokka [111] within JGI's KBase Knowledge Discovery Environment[112]. A more thorough annotation process using the Pfam database was completed using the hmmscan program[113].

All-against-all BLAST comparisons of protein sequences including Mu Mom and >8000 predicted homologs were run using EFI-EST (https://efi.igb.illinois.edu/efi-est/) (Enzyme Function Initiative – Enzyme Similarity Tool)[114] with an alignment score threshold of 35[115]. To facilitate visualization, the final network generated was 95% representative, collapsing sequences of 95% identity into representative nodes (3799 nodes, 589,294 edges, representative of 9,647 sequences). The similarity network was visualized using Cytoscape[116]. The alignment threshold and representative node percentage reported were selected based on empirical analysis. The SSN for Mom and homologs was used to guide initial screening using the *in vivo* activity assay. The sequences for Mom Homolog A and Mom Homolog B were selected at random. The structures of Mom Homolog A and Mom Homolog B were predicted using ColabFold[117] and superimposed with Chain A of Mom using PyMOL[118].

Data availability

The final model of the structure solved in this work was deposited in Protein Data Bank under accession code 8BV8. ORCA output files for all energy minimizations of educts, products, and the spirocyclic intermediate, as well as ORCA output files for the saddle points/transition states, and the corresponding PDB files (translated from the ORCA files using Avogadro) are made available in a Zip drive. A sequencing read archive (SRA) containing BAM files produced by PacBio SMRT sequencing platform and extended to include kinetic data for the modification motif analysis in this work can be obtained at NCBI Bioproject PRJNA1406289. Metagenome information for Mom Homolog A can be found at https://img.jgi.doe.gov/cgi-bin/m/main.cgi?section=TaxonDetail&page=taxonDetail&taxon_oid=3300002484. Metagenome information for Mom Homolog B can be found at https://img.jgi.doe.gov/cgi-bin/m/main.cgi?section=TaxonDetail&page=taxonDetail&taxon_oid=2872672955. All other data supporting the findings of this work are available within the Article, Supplementary Information, or from the corresponding author(s) upon request.

References

1. Isaev, A. B., Musharova, O. S. & Severinov, K. V. Microbial Arsenal of Antiviral Defenses – Part I. *Biochem Mosc* **86**, 319–337 (2021).

2. Isaev, A. B., Musharova, O. S. & Severinov, K. V. Microbial Arsenal of Antiviral Defenses. Part II. *Biochem Mosc* **86**, 449–470 (2021).

3. Hall, R. H. N6-(δ2-Isopentenyl)adenosine: Chemical Reactions, Biosynthesis, Metabolism, and Significance to the Structure and Function of †RNA. *Prog Nucleic Acid Re* **10**, 57–86 (1970).

4. Warren, R. A. J. Modified Bases in Bacteriophage DNAs. *Annu Rev Microbiol* **34**, 137–158 (1980).

5. H.Gommers-Ampt, J. & Borst, P. Hypermodified bases in DNA. *Faseb J* **9**, 1034–1042 (1995).

6. Weigele, P. & Raleigh, E. A. Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses. *Chem Rev* **116**, 12655–12687 (2016).

7. Parker, M. J., Lee, Y.-J., Weigele, P. R. & Saleh, L. Comprehensive Natural Products III. 465–488 (2020) doi:10.1016/b978-0-12-409547-2.14838-3.

8. Hutinet, G., Lee, Y.-J., Crécy-Lagard, V. de & Weigele, P. R. Hypermodified DNA in Viruses of E. coli and Salmonella. *EcoSal Plus* **9**, eESP-0028-2019 (2021).

9. Flodman, K. *et al.* Type II Restriction of Bacteriophage DNA With 5hmdU-Derived Base Modifications. *Front Microbiol* **10**, 584 (2019).

10. Flodman, K., Corrêa, I. R., Dai, N., Weigele, P. & Xu, S. In vitro Type II Restriction of Bacteriophage DNA With Modified Pyrimidines. *Front Microbiol* **11**, 604618 (2020).

11. Hall, R. H. *The Modified Nucleosides in Nucleic Acids*. (Columbia University Press, 1971).

12. Galburt, E. A., Pelletier, J., Wilson, G. & Stoddard, B. L. Structure of a tRNA Repair Enzyme and Molecular Biology Workhorse T4 Polynucleotide Kinase. *Structure* **10**, 1249–1260 (2002).

13. Shuman, S. DNA Ligases: Progress and Prospects*. *J Biol Chem* **284**, 17365–17369 (2009).

14. Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).

15. *Phage Mu*. vol. 28 (Cold Spring Harbor Laboratory, 1988).

16. Allet, B. & Bukhari, A. I. Analysis of bacteriophage Mu and λ-Mu hybrid DNAs by specific endonucleases. *J Mol Biol* **92**, 529–540 (1975).

17. Toussaint, A. The DNA modification function of temperate phage Mu-1. *Virology* **70**, 17–27 (1976).

18. Hattman, S. Unusual Modification of Bacteriophage Mu DNA. *J Virol* **32**, 468–475 (1979).

19. Khatoon, H. & Bukhari, A. I. Bacteriophage Mu-induced modification of DNA is dependent upon a host function. *J Bacteriol* **136**, 423–428 (1978).

20. Hattman, S. Specificity of the bacteriophage Mu mom+ -controlled DNA modification. *J Virol* **34**, 277–279 (1980).

21. Swinton, D. *et al.* Purification and characterization of the unusual deoxynucleoside, a-N-(9-fi-D-2'-deoxyribofuranosylpurin-6-yl)glycinamide, specified. *Proceedings of the National Academy of Science* (1983).

22. Toussaint, A. My life with Mu. *Bacteriophage* **5**, e1034336 (2015).

23. Karambelkar, S. *et al.* Emergence of a novel immune-evasion strategy from an ancestral protein fold in bacteriophage Mu. *Nucleic Acids Res* **48**, 5294–5305 (2020).

24. Newton, R. P., Kingston, E. E. & Overton, A. Identification of novel nucleotides found in the red seaweed Porphyra umbilicalis. *Rapid Commun. Mass Spectrom.* **9**, 305–311 (1995).

25. Strzelecka, D., Chmielinski, S., Bednarek, S., Jemielity, J. & Kowalska, J. Analysis of mononucleotides by tandem mass spectrometry: investigation of fragmentation pathways for phosphate- and ribose-modified nucleotide analogues. *Sci Rep-uk* **7**, 8931 (2017).

26. Kahmann, R., Seiler, A., Wulczyn, F. G. & Pfaff, E. The mom gene of bacteriophage Mu: a unique regulatory scheme to control a lethal function. *Gene* **39**, 61–70 (1985).

27. Krüger, D. H. & Bickle, T. A. Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiol Rev* **47**, 345–360 (1983).

28. Kaminska, K. H. & Bujnicki, J. M. Bacteriophage Mu Mom protein responsible for DNA modification is a new member of the acyltransferase superfamily. *Cell Cycle* **7**, 120–121 (2008).

29. Iyer, L. M., Tahiliani, M., Rao, A. & Aravind, L. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* **8**, 1698–1710 (2009).

30. Iyer, L. M., Zhang, D. & Aravind, L. Adenine methylation in eukaryotes: Apprehending the complex evolutionary history and functional potential of an epigenetic modification. *Bioessays* **38**, 27–40 (2016).

31. Iyer, L. M., Anantharaman, V., Krishnan, A., Burroughs, A. M. & Aravind, L. Jumbo Phages: A Comparative Genomic Overview of Core Functions and Adaptions for Biological Conflicts. *Viruses* **13**, 63 (2021).

32. Lee, Y.-J. *et al.* Identification and biosynthesis of thymidine hypermodifications in the genomic DNA of widespread bacterial viruses. *P Natl Acad Sci Usa* **115**, E3116–E3125 (2018).

33. Lee, Y.-J. *et al.* Pathways of thymidine hypermodification. *Nucleic Acids Res* **50**, 3001–3017 (2021).

34. Shimizu, Y. & Ueda, T. Cell-Free Protein Production, Methods and Protocols. *Methods Mol Biology* **607**, 11–21 (2009).

35. Tuckey, C., Asahara, H., Zhou, Y. & Chong, S. Protein Synthesis Using a Reconstituted Cell-Free System. *Curr Protoc Mol Biology* **108**, 16.31.1-16.31.22 (2014).

36. Burckhardt, R. M. & Escalante-Semerena, J. C. Small-Molecule Acetylation by GCN5-Related N -Acetyltransferases in Bacteria. *Microbiol Mol Biol R* **84**, 10.1128/mmbr.00090-19 (2020).

37. Favrot, L., Blanchard, J. S. & Vergnolle, O. Bacterial GCN5-Related N-Acetyltransferases: From Resistance to Regulation. *Biochemistry-us* **55**, 989–1002 (2016).

38. Baumgartner, J. T. *et al.* Gcn5-Related N-Acetyltransferases (GNATs) With a Catalytic Serine Residue Can Play Ping-Pong Too. *Frontiers Mol Biosci* **8**, 646046 (2021).

39. Ostrem, D. L. & Berg, P. Glycyl transfer ribonucleic acid synthetase from Escherichia coli: purification, properties, and substrate binding. *Biochemistry* **13**, 1338–48 (1974).

40. Toth, M. J. & Schimmel, P. Deletions in the large (beta) subunit of a hetero-oligomeric aminoacyl-tRNA synthetase. *J. Biol. Chem.* **265**, 1000–1004 (1990).

41. Valencia-Sánchez, M. I. *et al.* Structural Insights into the Polyphyletic Origins of Glycyl tRNA Synthetases. *J. Biol. Chem.* **291**, 14430–46 (2016).

42. Gomez, M. A. R. & Ibba, M. Aminoacyl-tRNA Synthetases. *RNA* **26**, rna.071720.119 (2020).

43. Ju, Y. *et al.* X-shaped structure of bacterial heterotetrameric tRNA synthetase suggests cryptic prokaryote functions and a rationale for synthetase classifications. *Nucleic Acids Res.* **49**, 10106–10119 (2021).

44. Grünberg, S. *et al.* Enhanced expression and purification of nucleotide-specific ribonucleases MC1 and Cusativin. *Protein Expr. Purif.* **190**, 105987 (2022).

45. Wolf, E. J. *et al.* Human RNase 4 improves mRNA sequence characterization by LC–MS/MS. *Nucleic Acids Res.* **50**, e106–e106 (2022).

46. Wolf, E. J., Dai, N., Chan, S.-H. & Corrêa, I. R. Selective Characterization of mRNA 5′ End-Capping by DNA Probe-Directed Enrichment with Site-Specific Endoribonucleases. *ACS Pharmacol. Transl. Sci.* **6**, 1692–1702 (2023).

47. Sprinzl, M. & Cramer, F. Site of aminoacylation of tRNAs from Escherichia coli with respect to the 2'- or 3'-hydroxyl group of the terminal adenosine. *Proc. Natl. Acad. Sci.* **72**, 3049–3053 (1975).

48. Asahara, H. & Uhlenbeck, O. C. The tRNA Specificity of Thermus thermophilus EF-Tu. *Proc National Acad Sci* **99**, 3499–3504 (2002).

49. Walker, S. E. & Fredrick, K. Preparation and evaluation of acylated tRNAs. *Methods* **44**, 81–86 (2008).

50. Gamper, H. & Hou, Y.-M. A Label-Free Assay for Aminoacylation of tRNA. *Genes-basel* **11**, 1173 (2020).

51. Xu, X. *et al.* MenT nucleotidyltransferase toxins extend tRNA acceptor stems and can be inhibited by asymmetrical antitoxin binding. *Nat. Commun.* **14**, 4644 (2023).

52. Chan, P. P. & Lowe, T. M. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res* **37**, D93–D97 (2009).

53. Chan, P. P. & Lowe, T. M. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res* **44**, D184–D189 (2016).

54. Leiva, L. E. *et al.* Modulation of Escherichia coli Translation by the Specific Inactivation of tRNAGly Under Oxidative Stress. *Frontiers Genetics* **11**, 856 (2020).

55. STEWART, T. S., ROBERTS, R. J. & STROMINGER, J. L. Novel Species of tRNA. *Nature* **230**, 36–38 (1971).

56. ROBERTS, R. J. Structures of Two Glycyl-tRNAs from Staphylococcus epidermidis. *Nat New Biology* **237**, 44–45 (1972).

57. Beebe, K., Pouplana, L. R. de & Schimmel, P. Elucidation of tRNA-dependent editing by a class II tRNA synthetase and significance for cell viability. *EMBO J.* **22**, 668–675 (2003).

58. Pasman, Z. *et al.* Substrate Specificity and Catalysis by the Editing Active Site of Alanyl-tRNA Synthetase from Escherichia coli. *Biochemistry* **50**, 1474–1482 (2011).

59. Kavoor, A., Kelly, P. & Ibba, M. Escherichia coli alanyl-tRNA synthetase maintains proofreading activity and translational accuracy under oxidative stress. *J. Biol. Chem.* **298**, 101601 (2022).

60. Fonvielle, M. *et al.* Aminoacyl-tRNA recognition by the FemX Wv transferase for bacterial cell wall synthesis. *Nucleic Acids Res.* **37**, 1589–1601 (2009).

61. Villet, R. *et al.* Idiosyncratic features in tRNAs participating in bacterial cell wall synthesis. *Nucleic Acids Res* **35**, 6870–6883 (2007).

62. Bikmetov, D. *et al.* GNAT toxins evolve toward narrow tRNA target specificities. *Nucleic Acids Res.* **50**, 5807–5817 (2022).

63. Vetting, M. W. *et al.* Structure and functions of the GNAT superfamily of acetyltransferases. *Arch Biochem Biophys* **433**, 212–226 (2005).

64. Ud-Din, A. I. M. S., Tikhomirova, A. & Roujeinikova, A. Structure and Functional Diversity of GCN5-Related N-Acetyltransferases (GNAT). *Int. J. Mol. Sci.* **17**, 1018 (2016).

65. Bochtler, M. & Fernandes, H. DNA adenine methylation in eukaryotes: Enzymatic mark or a form of DNA damage? *Bioessays* **43**, 2000243 (2021).

66. Fu, D., Calvo, J. A. & Samson, L. D. Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat Rev Cancer* **12**, 104–120 (2012).

67. Xu, G.-L. & Bochtler, M. Reversal of nucleobase methylation by dioxygenases. *Nat Chem Biol* **16**, 1160–1169 (2020).

68. Macon, J. B. & Wolfenden, R. 1-Methyladenosine. Dimroth rearrangement and reversible reduction. *Biochemistry-us* **7**, 3453–3458 (1968).

69. Katritzky, A. R., Ramsden, C. A., Joule, J. A. & Zhdankin, V. V. Handbook of Heterocyclic Chemistry (Third Edition). *Part 3 React Heterocycles* 473–604 (2010) doi:10.1016/b978-0-08-095843-9.00010-0.

70. Neese, F., Wennmohs, F., Becker, U. & Riplinger, C. The ORCA quantum chemistry program package. *J. Chem. Phys.* **152**, 224108 (2020).

71. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674 (2007).

72. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr Sect D* **66**, 486–501 (2010).

73. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

74. Holm, L. Dali server: structural unification of protein families. *Nucleic Acids Res* **50**, W210–W215 (2022).

75. Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S. & Thornton, J. M. PDBsum: Structural summaries of PDB entries. *Protein Sci Publ Protein Soc* **27**, 129–134 (2018).

76. Cheverton, A. M. *et al.* A Salmonella Toxin Promotes Persister Formation through Acetylation of tRNA. *Mol Cell* **63**, 86–96 (2016).

77. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J Mol Biol* **372**, 774–797 (2007).

78. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).

79. Yashiro, Y., Zhang, C., Sakaguchi, Y., Suzuki, T. & Tomita, K. Molecular basis of glycyl-tRNAGly acetylation by TacT from Salmonella Typhimurium. *Cell Reports* **37**, 110130 (2021).

80. Hall, R. H. Isolation of N 6 -(Aminoacyl)adenosine from Yeast Ribonucleic Acid *. *Biochemistry-us* **3**, 769–773 (1964).

81. Chheda, G. B. & Hall, R. H. Aminoacyl Nucleosides. III. A Novel Rearangement: Conversion of N6-(α-Aminoacyl) adenines into N-(6-Purinyl)amino Acids*. *Biochemistry* (1966).

82. Chheda, G. B., Hall, R. H. & Tanna, P. M. Aminoacyl Nucleosides. V. The Mechanism of the N6-(α-Aminoacyl)adenines into N-(6-purinyl)amino Acids. *The Journal of Organic Chemistry* (1969).

83. Masurier, N. *et al.* Synthesis of Peptide-Adenine Conjugates as a New Tool for Monitoring Protease Activity. *Eur J Org Chem* **2019**, 176–183 (2018).

84. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715 (1993).

85. Moutiez, M., Belin, P. & Gondry, M. Aminoacyl-tRNA-Utilizing Enzymes in Natural Product Biosynthesis. *Chem Rev* **117**, 5578–5618 (2017).

86. Hattman, S. Unusual transcriptional and translational regulation of the bacteriophage Mu mom operon. *Pharmacol Therapeut* **84**, 367–388 (1999).

87. Chimnaronk, S. *et al.* RNA helicase module in an acetyltransferase that modifies a specific tRNA anticodon. *Embo J* **28**, 1362–1373 (2009).

88. Huang, B., Johansson, M. J. O. & Byström, A. S. An early step in wobble uridine tRNA modification requires the Elongator complex. *Rna* **11**, 424–436 (2005).

89. Dauden, M. I. *et al.* Architecture of the yeast Elongator complex. *Embo Rep* **18**, 264–279 (2017).

90. Lee, Y.-J. & Weigele, P. R. Detection of Modified Bases in Bacteriophage Genomic DNA. *Methods Mol Biology Clifton N J* **2198**, 53–66 (2020).

91. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).

92. Parker, M. J., Weigele, P. R. & Saleh, L. Insights into the Biochemistry, Evolution, and Biotechnological Applications of the Ten-Eleven Translocation (TET) Enzymes. *Biochemistry-us* **58**, 450–467 (2019).

93. Burke, E. J. *et al.* Phage-encoded ten-eleven translocation dioxygenase (TET) is active in C5-cytosine hypermodification in DNA. *Proc National Acad Sci* **118**, e2026742118 (2021).

94. Cavaluzzi, M. J. & Borer, P. N. Revised UV extinction coefficients for nucleoside-5′-monophosphates and unpaired DNA and RNA. *Nucleic Acids Res* **32**, e13–e13 (2004).

95. Dunn, D. B. & Hall, R. H. Handbook of Biochemistry and Molecular Biology. 269–358 (2010) doi:10.1201/b10501-47.

96. Francois, C. J. L., Jang, Y. H., Cagin, T., Goddard, W. A. & Sowers, L. C. Conformation and Proton Configuration of Pyrimidine Deoxynucleoside Oxidation Damage Products in Water. *Chem Res Toxicol* **13**, 462–470 (2000).

97. Milo, R., Jorgensen, P., Moran, U., Weber, G. & Springer, M. BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Res* **38**, D750–D753 (2010).

98. NATHAN, T., JENNIFER, O., ESTA, S., L, M. L. & V, R. S. RNase Inhibitors. (2024).

99. Edwards, A. L., Garst, A. D. & Batey, R. T. Determining structures of RNA aptamers and riboswitches by X-ray crystallography. *Methods Mol Biology Clifton N J* **535**, 135–63 (2009).

100. Quinlivan, E. P. & Gregory, J. F. DNA digestion to deoxyribonucleoside: A simplified one-step procedure. *Anal Biochem* **373**, 383–385 (2008).

101. SPRINZL, M., STERNBACH, H., HAAR, F. V. D. & CRAMER, F. Enzymatic Incorporation of ATP and CTP Analogues into the 3′ End of tRNA. *Eur. J. Biochem.* **81**, 579–589 (1977).

102. Katoh, T. & Suga, H. Flexizyme-catalyzed synthesis of 3′-aminoacyl-NH-tRNAs. *Nucleic Acids Res* **47**, e54–e54 (2019).

103. Kent, A. D. *et al.* Thioesters Support Efficient Protein Biosynthesis by the Ribosome. *ACS Cent. Sci.* **11**, 404–412 (2025).

104. Krug, M., Weiss, M. S., Heinemann, U. & Mueller, U. XDSAPP : a graphical user interface for the convenient processing of diffraction data using XDS. *J Appl Crystallogr* **45**, 568–572 (2012).

105. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr Sect D* **75**, 861–877 (2019).

106. Pettersen, E. F. *et al.* UCSF ChimeraX : Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70–82 (2020).

107. Gibrat, J.-F., Madej, T. & Bryant, S. H. Surprising similarities in structure comparison. *Curr Opin Struc Biol* **6**, 377–385 (1996).

108. Eddy, S. R. Accelerated Profile HMM Searches. *Plos Comput Biol* **7**, e1002195 (2011).

109. Paez-Espino, D. *et al.* IMG/VR v.2.0: an integrated data management and analysis system for cultivated and environmental viral genomes. *Nucleic Acids Res* **47**, D678–D686 (2019).

110. Gregory, A. C. *et al.* Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell* **177**, 1109-1123.e14 (2019).

111. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

112. Arkin, A. P. *et al.* KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol* **36**, 566–569 (2018).

113. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res* **47**, D427–D432 (2019).

114. Gerlt, J. A. *et al.* Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochimica Et Biophysica Acta Bba - Proteins Proteom* **1854**, 1019–1037 (2015).

115. O'Toole, K. H., Imperiali, B. & Allen, K. N. Glycoconjugate pathway connections revealed by sequence similarity network analysis of the monotopic phosphoglycosyl transferases. *Proc National Acad Sci* **118**, e2018289118 (2021).

116. Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* **13**, 2498–2504 (2003).

117. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679–682 (2022).

118. LLC, S. & DeLano, W. *PyMOL*. (2020).

119. Francklyn, C. S. & Minajigi, A. tRNA as an active chemical scaffold for diverse chemical transformations. *Febs Lett* **584**, 366–375 (2010).

120. Ulrich, E. C. & Donk, W. A. van der. Cameo appearances of aminoacyl-tRNA in natural product biosynthesis. *Curr Opin Chem Biol* **35**, 29–36 (2016).

121. Su, Z., Wilson, B., Kumar, P. & Dutta, A. Noncanonical Roles of tRNAs: tRNA Fragments and Beyond. *Annu Rev Genet* **54**, 1–23 (2020).

122. Maruyama, C. & Hamano, Y. tRNA-dependent amide bond–forming enzymes in peptide natural product biosynthesis. *Curr. Opin. Chem. Biol.* **59**, 164–171 (2020).

123. Hentchel, K. L. & Escalante-Semerena, J. C. Acylation of Biomolecules in Prokaryotes: a Widespread Strategy for the Control of Biological Function and Metabolic Stress. *Microbiol Mol Biol R* **79**, 321–346 (2015).

124. Yashiro, Y., Sakaguchi, Y., Suzuki, T. & Tomita, K. Mechanism of aminoacyl-tRNA acetylation by an aminoacyl-tRNA acetyltransferase AtaT from enterohemorrhagic E. coli. *Nat Commun* **11**, 5438 (2020).

125. Skiba, M. A. *et al*. Repurposing the GNAT Fold in the Initiation of Polyketide Biosynthesis. *Structure* **28**, 63-74.e4 (2020).

Acknowledgments We thank Johanna Hakanpää and Guillaume Pompidor for access to the PETRA III P11 beamline at DESY, and Dr. Honorata Czapińska, Dr. Elzbieta Nowak and Michal Pastor for data collection. We gratefully acknowledge Poland's high-performance Infrastructure PLGrid [ACK Cyfronet AGH] for providing computer facilities and support within computational grant no [plgmomylation and plgmomylation2]. The authors are grateful for purified WT GlyRS and WT AlaRS from Corinna Tuckey, Dr. Ying Zhou, Dr. Bradley J. Landgraf and Dr. Emily Chen, T7 RNA polymerase from Professor Janusz Bujnicki's lab, a plasmid encoding $tRNA^{Gly/GCC}$ from Dr. Haruichi Asahara, and plasmid pRY from Dr. Robert M. Yarrington. We thank Dr. Nan Dai, Dr. Eric J. Wolf, Dr. Edwin E. Escobar, and the Protein Research Mass Spec Facility at NEB for their expertise using intact mass spectrometry to characterize tRNA. We also thank Dr. Andrew P. Sikkema for sharing his methods and expertise regarding WT CCA2 as well as Dr. Esta Tamanaha and James Elliot for sharing immobilized aptamers during purification of CCA2. Critical readings of our manuscript were provided by Dr. Ivan R. Corrêa Jr., Dr. Richard J. Roberts, Dr. Timothy R. Blower, Dr. Gregory J. S. Lohman, and colleagues at NEB/IIMCB. Tashe José provided key feedback on scientific illustrations. We are thankful to Dr. Shweta Karambelkar for sharing her unpublished thesis from the Indian Institute of Science, Bangalore. R.M.B.S., Y.J.L., C.G., E.A.S., S.R.L., and P.R.W. are grateful for the generous support from the Comb Family and New England Biolabs, Inc. without whom this work would not have been possible. M.B. and A.S. are grateful for funding from the Polish National Science Centre (NCN, 2018/30/Q/NZ2/00669) and the Foundation for Polish Science (FNP, POIR.04.04.00-00-5D81/17-00 to M.B. and START 79.2020 to A.S.). Some figures were made using BioRender. We dedicate this work to the memory of Professor Stanley Hattman.

Author contributions P.R.W. and M.B. conceptualized the project. R.M.B.S., A.S., and Y.J.L. developed the methodologies used in this study. R.M.B.S., A.S., Y.J.L., C.G., S.R.L., E.A.R., K.S., M.S.K, M.B., and P.R.W. conducted experiments and completed data analysis. R.M.B.S. and A.S. administered the project and managed data visualization. M.B. and P.R.W supervised the project and secured funding. R.M.B.S, P.R.W., Y.-J.L., A.S., and M.B. wrote the manuscript with support from all the other authors.

Competing interests R.M.B.S, Y.-J.L., C.G., E.A.S., S.R.L., M.S.K, and P.R.W. are employees of New England Biolabs, a manufacturer and vendor of molecular biology reagents. This affiliation does not affect the authors' impartiality, adherence to journal standards and policies, or availability of data. The remaining authors declare no competing interests.

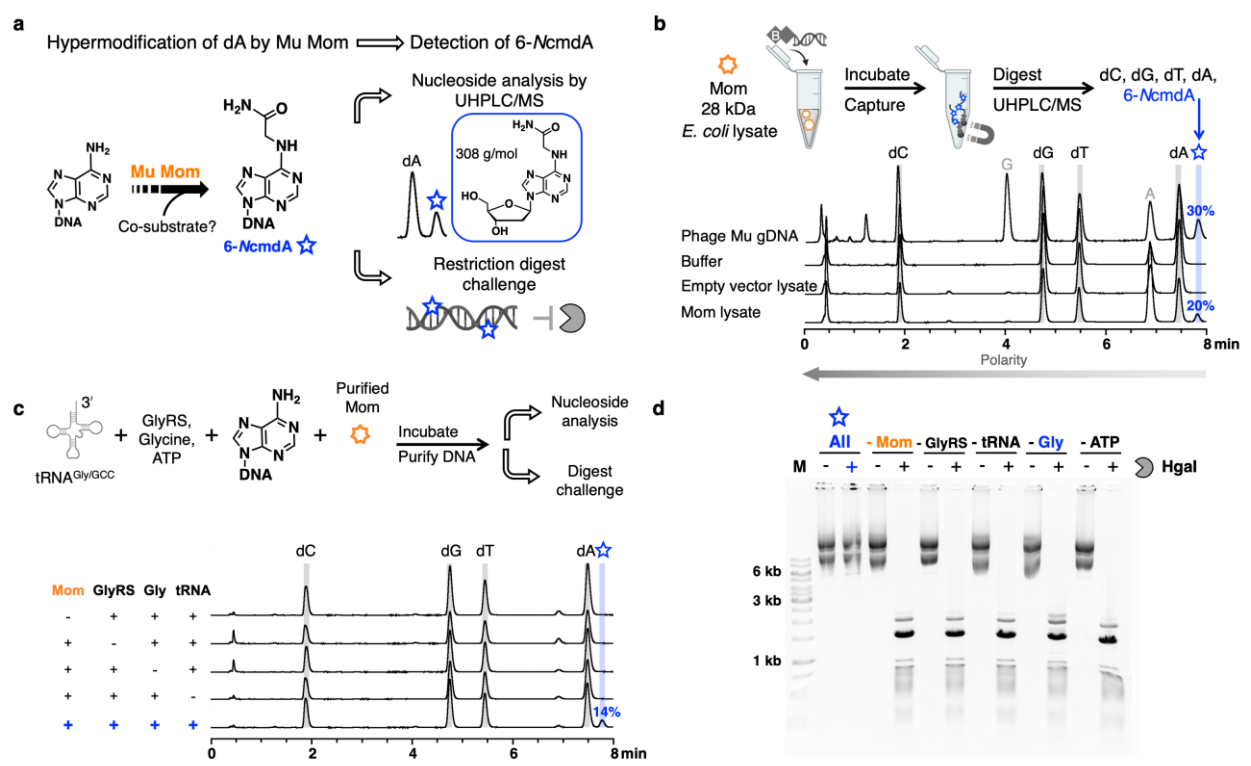

Fig. 1 | Mom hypermodifies the *N*6 position of dA post-replicatively through tRNA-dependent transfer of an activated glycine. **a**, Detecting Mom-catalyzed hypermodification. Following nucleoside digestion of dsDNA hypermodified by Mom, 6-*N*cmdA can be resolved and detected by UHPLC/MS. Alternatively, hypermodified dsDNA can be visualized following challenge with restriction endonucleases. **b**, Hypermodification of dA by Mom is post-replicative. In an *ex vivo* assay, clarified *E. coli* lysates (20 µg total protein) containing heterologously expressed Mom were incubated with biotinylated dsDNA substrates (250-1000 ng, see Methods). Nucleoside analysis shows that 6-*N*cmdA is installed by Mom post-replicatively on dsDNA (bottom trace). **c**, *In vitro* reconstitution of Mom activity demonstrated by nucleoside analysis of hypermodified DNA substrates. 6-*N*cmdA is detected only when purified Mom (2 µM, orange), dsDNA (1000 ng), $tRNA^{Gly/GCC}$, glycine, and ATP are included in a coupled activity assay (bottom trace). **d**, *In vitro* reconstitution of Mom activity demonstrated by restriction challenge of hypermodified DNA substrates. Protection from HgaI digestion was only observed when purified Mom (0.6 µM), substrate plasmid (3.6 µg), $tRNA^{Gly/GCC}$, glycine, and ATP are included in a coupled activity assay (see lanes under All). Elements of this figure use icons from BioRender https://BioRender.com/j9mbjgb.

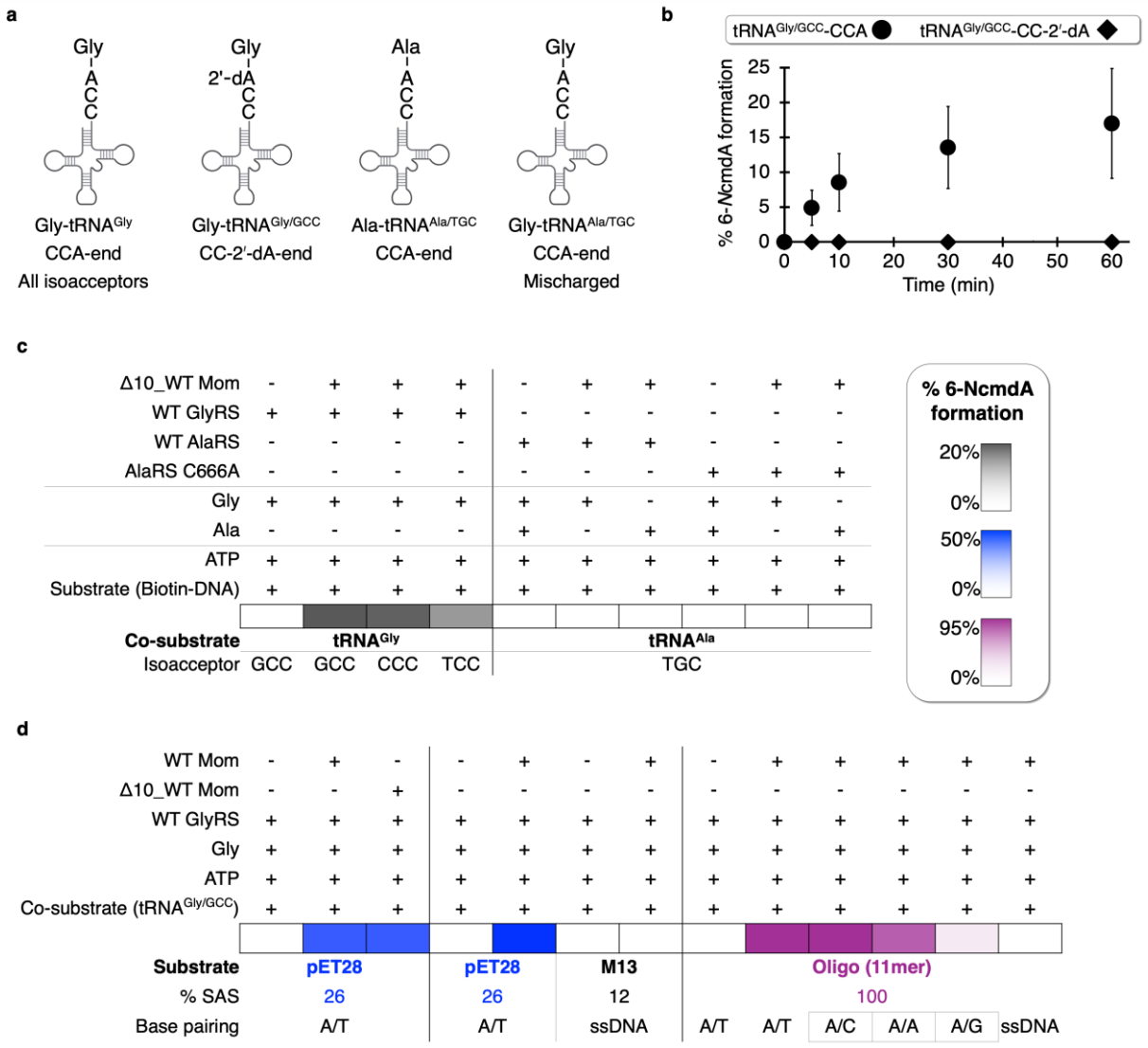

Fig. 2 | Mom is selective for Gly-tRNA$^{Gly/GCC}$ as a co-substrate and dsDNA as a substrate. **a**, tRNA species used to probe co-substrate specificities. Full-length tRNA$^{Gly/GCC}$-CCA was generated either by IVT or by a combination of IVT of tRNA$^{Gly/GCC}$-CC and add addition of rA from ATP to the 3' end by CCA2 (see Supplementary Fig. 9 and Supplementary Fig. 14) . tRNA$^{Gly/GCC}$-CC-2'-dA was generated by IVT and CCA2 addition of 2'-dA from 2'-dATP. Aminoacylation of tRNA was monitored using a gel-based derivatization method developed by Gamper and Hou (see Supplementary Fig. 17). Charging with glycine was achieved *in situ* using WT GlyRS. Full-length tRNA$^{Ala/TGC}$ was generated by IVT and charged with alanine *in situ* using WT AlaRS or mischarged with glycine using C666A AlaRS. **b**, Time courses of Mom assays including tRNA$^{Gly/GCC}$-CCA or tRNA$^{Gly/GCC}$-CC-2'-dA. All uncharged tRNAs were verified by intact mass prior to Mom assays. Three independent time courses were run as above with each tRNA species, and 6-*N*cmdA formation on pUC19 plasmid substrates was monitored by UHPLC. Averages and standard deviations for the activity measured at each time point are shown. Since

about 25% of all dA are in an SAS sequence context for pUC19, full momylation corresponds to about 25% 6-*N*cmdA formation. No 6-*N*cmdA formation was detected where $tRNA^{Gly/GCC}$-CC-2'-dA was included at any time point taken, up to 24 hours (see Supplementary Fig. 18 and Supplementary Fig. 19). Since charging of $tRNA^{Gly/GCC}$-CC-2'-dA and $tRNA^{Gly/GCC}$-CCA by GlyRS both proceed through a glycyladenylate intermediate, these data confirm that Gly-$tRNA^{Gly/GCC}$-CCA is the co-substrate used by Mom to hypermodify dA. **c**, Co-substrate preference of Mom. Each $tRNA^{Gly}$ isoacceptor from *E. coli* supported momylation. $tRNA^{Ala/TGC}$ charged with alanine or mischarged with glycine did not support momylation for Δ10_WT Mom. **d**, Substrate preference of Mom. Momlylation can be detected on dsDNA using long, linear (biotinylated), plasmid (pUC19 or pET28a), or short, linear (oligomers) substrates. No activity for Mom can be detected on long (M13 circular) or short (oligomer) substrates. Duplex DNA containing mismatches are poorer substrates, with lower activity observed on purine:purine mismatches. For the conversion for the A/A MM dsDNA substrate is given per strand. Elements of this figure use icons from BioRender https://BioRender.com/j9mbjgb.

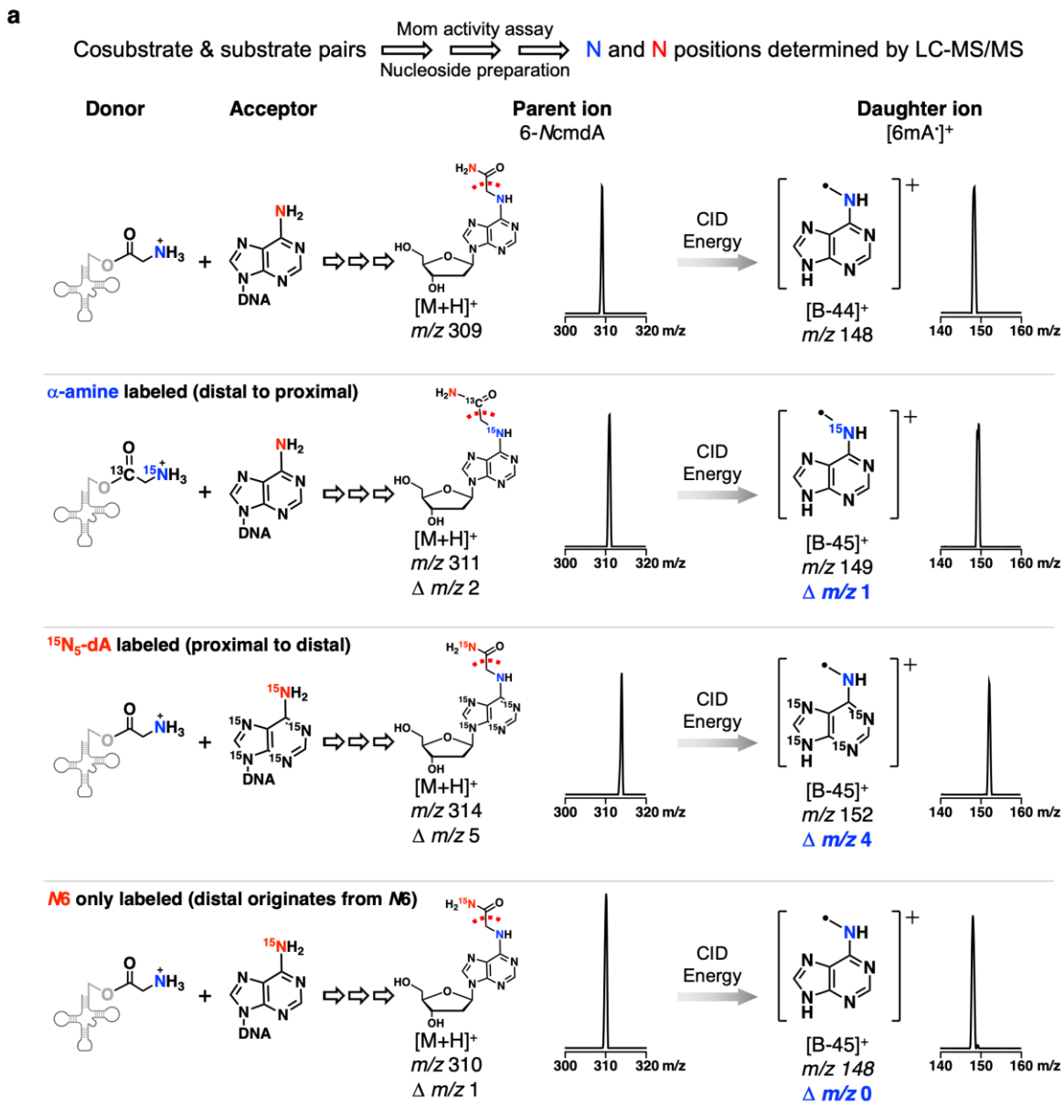
a
Cosubstrate & substrate pairs
Mom activity assay
Nucleoside preparation
N and N positions determined by LC-MS/MS
Donor
Acceptor
Parent ion
6-NcmdA
Daughter ion
[6mA•]+
CID Energy
[M+H]+
m/z 309
[B-44]+
m/z 148
α-amine labeled (distal to proximal)
[M+H]+
m/z 311
Δ m/z 2
[B-45]+
m/z 149
Δ m/z 1
15N5-dA labeled (proximal to distal)
[M+H]+
m/z 314
Δ m/z 5
[B-45]+
m/z 152
Δ m/z 4
N6 only labeled (distal originates from N6)
[M+H]+
m/z 310
Δ m/z 1
[B-45]+
m/z 148
Δ m/z 0
300 310 320 m/z
140 150 160 m/z
b
Aminoacylation
Mom
tRNA
Rearrangement
Spirocyclic intermediate
6-NcmdA
DNA

Fig. 3. | Mom installs 6-*N*cmdA through transfer and rearrangement of the glycyl group. a, On-base rearrangement detected through fragmentation of 6-*N*cmdA and site-specific isotopes. Combinations of unlabeled and isotopically-labeled co-substrate and substrate pairs (left) were used in activity assays to track the positions of the α amine nitrogen from glycine (blue) and the *N*6 nitrogen from dA (red) in the mature hypermodification (center). One daughter ion of 6-*N*cmdA, $[6mA^{\bullet}]^+$ (*m/z* 148, top), fragments at the exocyclic C-C bond, which separates the two exocyclic nitrogens (top right). In the second reaction, $[6m\text{-}^{15}N\text{-}A^{\bullet}]^+$ is 1 Dalton heavier in mass (*m/z* 149), indicating that the ^{15}N-labeled α amine nitrogen originating from glycine at the distal position (relative to the purine ring) has rearranged to the *N*6 position (proximal). In a complementary experiment where the proximal nitrogen originating from dA is labeled (third reaction), $[6m\text{-}^{15}N_4\text{-}A^{\bullet}]^+$ (*m/z* 152) is 4 Daltons heavier (rather than 5), demonstrating that an on-base rearrangement occurs following transfer of the glycyl group. The fourth activity assay in the series included a DNA substrate labeled with ^{15}N only at the *N*6 position of dA. The $[6mA^{\bullet}]^+$ daughter ion produced is equal in mass (*m/z* 148, top) to the daughter ion in the top reaction where unlabeled donor and acceptor were used. The fragmentation pattern here confirms that the nitrogen at the *N*6 position of dA makes the nucleophilic attack on glycine. **b**, Proposed mechanism for formation of 6-*N*cmdA. Rearrangement of exocyclic groups covalently-bound to dA at *N*6, resulting in a mixture of products, have been reported in abiotic reactions. In a biological context, we propose that following aminoacylation of *N*6 by Mom, an analogous rearrangement could occur through a five-membered ring intermediate where 6-*N*cmdA is the favored product (bracketed steps). Quantum chemical calculations indicate protonation states are important during the rearrangement. Elements of this figure use icons from BioRender https://BioRender.com/j9mbjgb.

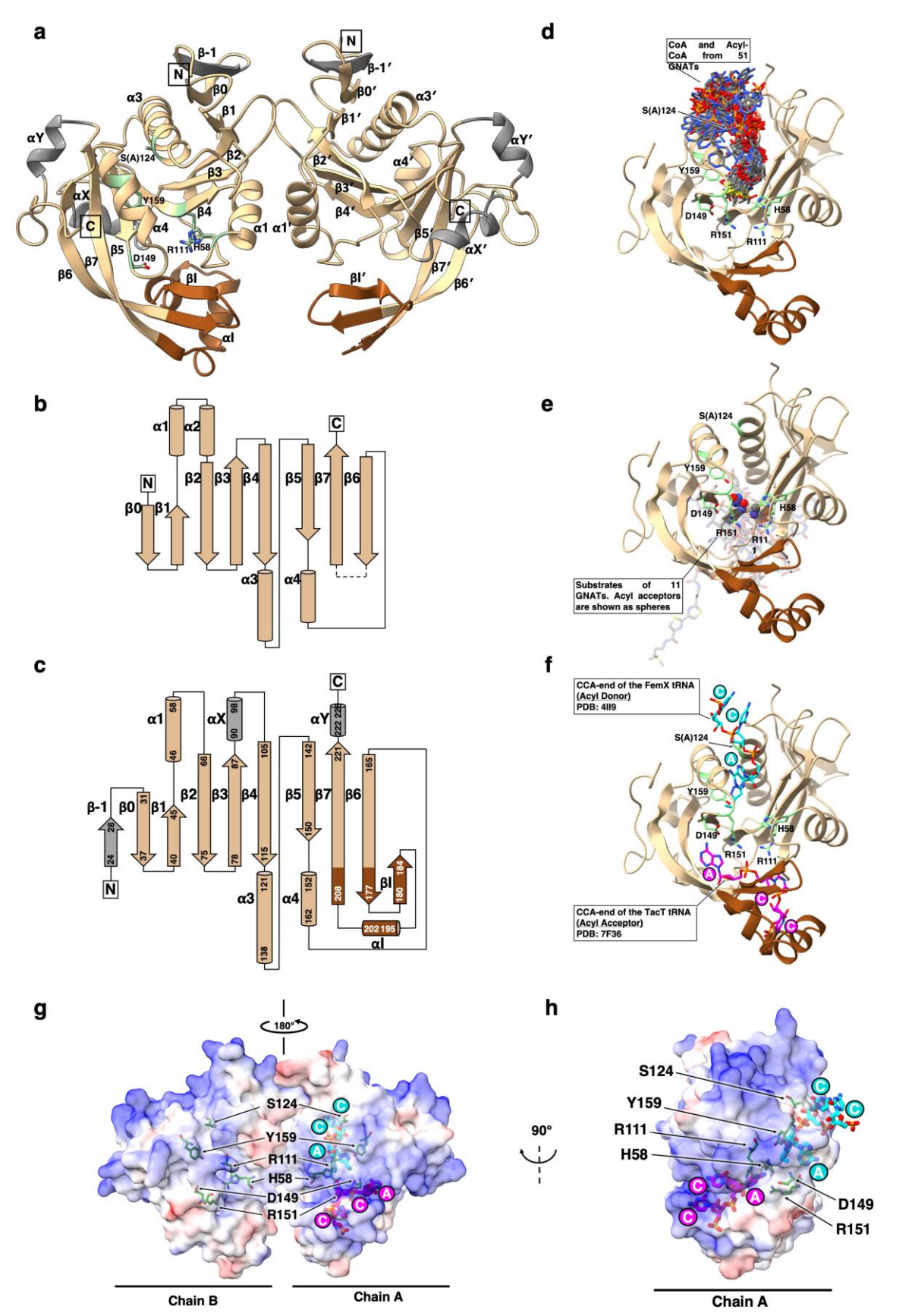
a
N
β-1
β0
β1
α3
αY
S(A)124
β2
β3
αX
Y159
β4
C
α4
α1
β5
R111
H58
D149
β7
β6
βI
αI
N
β-1′
β0′
β1′
α3′
αY′
β2′
α4′
β3′
β4′
C
β5′
αX′
β7′
β6′
βI′
b
α1
α2
N
β0
β1
β2
β3
β4
C
β5
β7
β6
α3
α4
c
C
α1
αX
αY
β-1
β0
β1
β2
β3
β4
β5
β7
β6
N
α3
α4
βI
αI
d
CoA and Acyl-CoA from 51 GNATs
S(A)124
Y159
D149
H58
R151
R111
e
S(A)124
Y159
D149
H58
R151
R111
Substrates of 11 GNATs. Acyl acceptors are shown as spheres
f
CCA-end of the FemX tRNA (Acyl Donor) PDB: 4II9
S(A)124
Y159
D149
H58
R151
R111
CCA-end of the TacT tRNA (Acyl Acceptor) PDB: 7F36
g
180°
S124
Y159
R111
H58
D149
R151
Chain B
Chain A
90°
h
S124
Y159
R111
H58
D149
R151
Chain A

Fig. 4. | Crystal structure of S124A Mom. a, Ribbon representation of the Mom dimer. Featured are the conserved GNAT core (pale brown), the insertion unique to Mom (brown), residues important for activity (sticks, pale green), and additional regions of low sequence conservation in the GNAT superfamily (gray). **b**, General GNAT topology diagram. The line that connects β-6 and β-7 is dashed, because β-7 can be contributed by another protomer in some GNATs. **c**, Mom topology diagram. **d**, Superposition of Mom with 51 diverse GNAT:CoA structures. A Mom protomer (Chain A) was superposed with 51 diverse GNATs in complex with CoA or acyl-CoA (see Methods and Supplementary Table 6). Only the overlaid CoA or acyl-CoA molecules of these GNATs are shown. **e**, Out of 51 superposed GNAT structures 11 contained the substrates or the products of the acylation (shown as half-transparent sticks). The acyl acceptor atoms are shown as spheres colored by heteroatom. The list of substrates is available in the Supplementary Table 7. **f**, Superposition of Mom with GNAT:tRNA structures. Chain A of Mom was superposed with $FemX_{Wv}$ (PBD: 4II9, Ala-$tRNA^{Ala}$ co-substrate) and TacT (PBD: 7F36, Gly-$tRNA^{Gly}$ substrate). Only the 3´-CCA ends of tRNAs are shown (cyan and magenta, respectively). Mom could utilize a FemX-like strategy for binding tRNA and a TacT-like strategy for binding DNA, aided by the positively charged cleft. **g, h,** Electrostatic representation of the Mom dimer and monomer surfaces. Residues important for catalysis are labeled. CCA-ends of the superposed structures of FemX and TacT (colored same as in 3f) are shown to demonstrate the local positive charge of the potential tRNA/DNA binding surfaces of Mom. In the chain B, the missing part of the insertion (residues 187-206) was completed with the one from chain A for better representation.

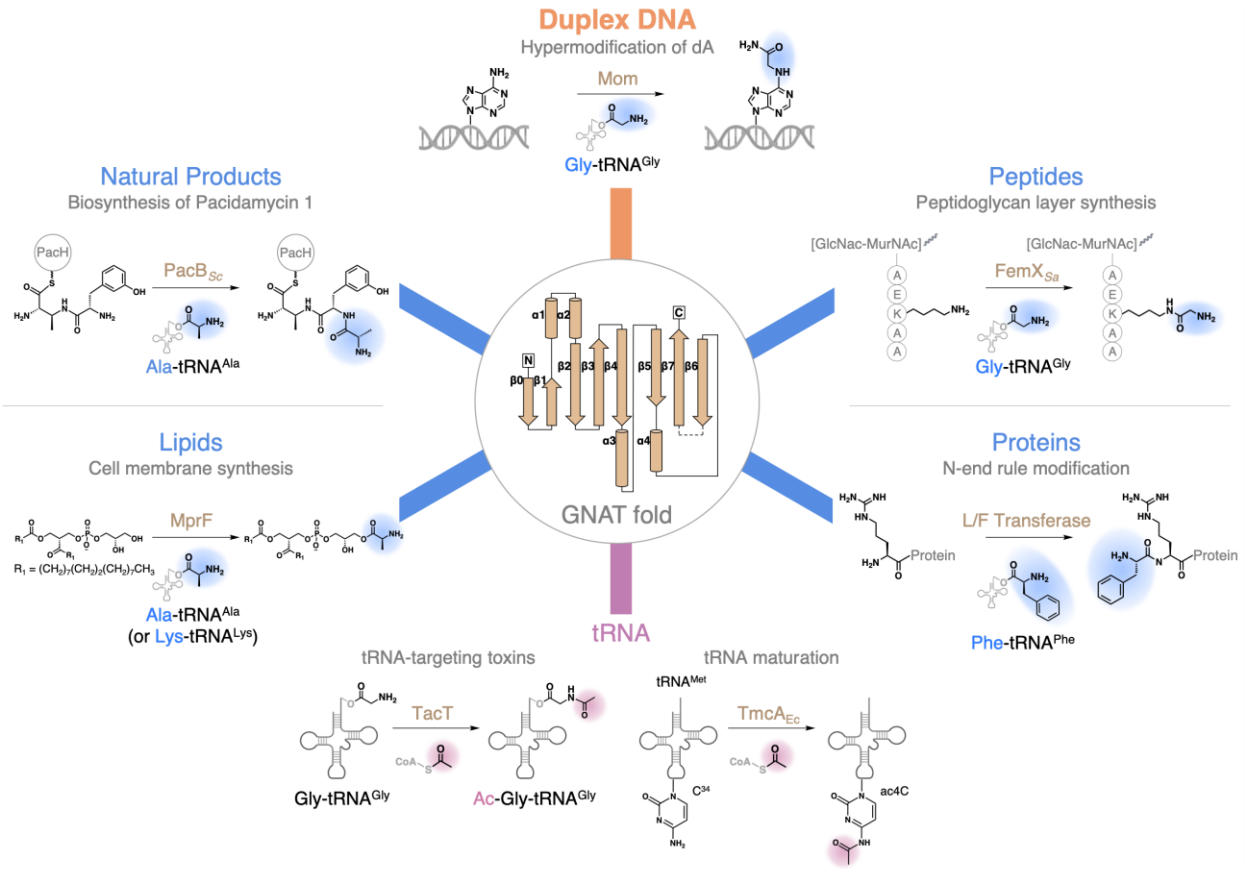

Fig. 5. | Co-substrate, substrate, and functional diversity of the GNAT superfamily. Noncanonical roles for aminoacylated tRNA outside of ribosome-dependent protein synthesis are found in all domains of life and continue to emerge in diverse cellular processes[55,56,119–122]. Analogous to its canonical role, the aminoacyl group at the 3′ end of a charged tRNA is utilized as an amino acid donor to modify diverse target substrates. Highlighted here are selected examples from microbes where tRNA has been found either as the co-substrate used by GNAT superfamily enzymes to aminoacylate target substrates (blue panels) or as the substrate that is the target of modification (mauve panel)[36,37,63,79,87,123–125]. Our work expands the properties of the GNAT superfamily (orange panel), where charged tRNA is the donor and DNA is the target. Charged tRNA has been found to be the aminoacyl donor in natural product biosynthesis (shown for PacB from *Streptomyces coeruleorubidus*), cell membrane synthesis (shown for MprF, multiple peptide resistance factor), peptides (shown for FemX, factors essential for methicillin resistance, from *Staphylococcal aureus*), and proteins (shown for L/F transferase, Leucyl/phenylalanyl-tRNA-protein transferase). To date, tRNA is known to be the substrate for GNAT enzymes in a few systems. The α-amine at the charged 3′ end can be acetylated by Type II Toxins (ex. AtaT, TacT, etc.), or the wobble base in the anticodon region can be acetylated by TmcA (example for $tRNA^{Met}$ cytidine acetyltransferase from *E. coli* shown) as part of maturation for $tRNA^{Met}$. These examples are not comprehensive, but instead emphasize the broad range of cellular processes where GNATs and tRNA are involved. The factors that govern co-substrate and substrate specificity for this superfamily are not fully understood. For example, the co-substrate and substrate of Mom could not be predicted *de novo* from the structure alone or through alignments of Mom with the most

closely related structural homologs found by the DALI server. Continued advances in computational and experimental approaches may give insights on what features have been repurposed or acquired over protein evolution and are driving function. Elements of this figure use icons from BioRender https://BioRender.com/j9mbjgb.

Editor Summary:
Phages install chemically complex DNA hypermodifications as anti-defense strategies. Here, the authors uncover transfer and rearrangement of glycine on dA driven by Mom, a tRNA-dependent DNA aminoacyltransferase encoded by phage Mu.
Peer Review Information:
Nature Communications thanks anonymous reviewers for their contribution to the peer review of this work. [A peer review file is available.]