The structure of the *Thermococcus gammatolerans* McrB N-terminal domain reveals a new mode of substrate recognition and specificity among McrB homologs

Christopher J. Hosford, Anthony Q. Bui, and Joshua S. Chappie

From the Department of Molecular Medicine, Cornell University, Ithaca, New York 14853

Edited by Patrick Sung

McrBC is a two-component, modification-dependent restriction system that cleaves foreign DNA-containing methylated cytosines. Previous crystallographic studies have shown that *Escherichia coli* McrB uses a base-flipping mechanism to recognize these modified substrates with high affinity. The side chains stabilizing both the flipped base and the distorted duplex are poorly conserved among McrB homologs, suggesting that other mechanisms may exist for binding modified DNA. Here we present the structures of the *Thermococcus gammatolerans* McrB DNA-binding domain (TgΔ185) both alone and in complex with a methylated DNA substrate at 1.68 and 2.27 Å resolution, respectively. The structures reveal that TgΔ185 consists of a YT521-B homology (YTH) domain, which is commonly found in eukaryotic proteins that bind methylated RNA and is structurally unrelated to the *E. coli* McrB DNA-binding domain. Structural superposition and co-crystallization further show that TgΔ185 shares a conserved aromatic cage with other YTH domains, which forms the binding pocket for a flipped-out base. Mutational analysis of this aromatic cage supports its role in conferring specificity for the methylated adenines, whereas an extended basic surface present in TgΔ185 facilitates its preferential binding to duplex DNA rather than RNA. Together, these findings establish a new binding mode and specificity among McrB homologs and expand the biological roles of YTH domains.

Modification-dependent restriction systems recognize and cleave modified DNA (1). Some enzymes like Mrr, McrA, Mspl, and McrBC are directed against methylated cytosines (2), whereas others like GmrSD and members of the PvuRts1I family show specificity toward glucosylated nucleic acids (3, 4). Collectively these proteins play a role in establishing the epigenetic landscape of bacterial genomes (5) and are especially important in protecting against predatory bacteriophages, many of which incorporate modified bases into their DNA to evade detection by other defense systems (6).

McrBC is a two-component, motor protein complex that was initially identified in *Escherichia coli* (Ec) genetic screens by its ability to restrict glucosylation-deficient mutants of T4 phage (7). EcMcrB is a 53-kDa protein with an N-terminal domain (pfam: DUF3578) that binds fully or hemi-methylated RMC recognition elements (where R is a purine base and MC is a 4-methyl-, 5-methyl-, or 5-hydroxymethyl-cytosine) (8–13) and a C-terminal extended ATPases associated with various cellular activities (AAA+) domain that binds/hydrolyzes GTP and mediates nucleotide-dependent oligomerization (14). EcMcrB exhibits a low basal GTPase activity (~0.5–1 min⁻¹) that can be stimulated ~30–40-fold via interaction with its partner EcMcrC (15), a 40-kDa protein that contains a C-terminal PD-(D/E)XXK family endonuclease domain and lacks the ability to bind DNA on its own (16). Biochemical studies suggest a model for cleavage in which EcMcrB and EcMcrC assemble at two RMC sites separated by up to 3 kilobases and translocate DNA in a manner that depends on stimulated GTP hydrolysis (17). Collision of these assemblies cleaves both DNA strands near one of the RMC sites (12, 18), suggesting that the complexes remain bound and translocate via DNA looping or twisting (19). These mechanochemical properties are reminiscent of type I and III restriction-modification systems, which bind DNA at non-modified sites separated by up to thousands of base pairs and use ATP hydrolysis to power similar long-range translocation events that trigger cleavage either by collision or stalling (20).

EcMcrB achieves specificity through a base flipping mechanism (13, 21). Modified bases are rotated out of the DNA duplex

2 The abbreviations used are: Ec, *Escherichia coli*; TgΔ185, N-terminal DNA-binding domain of *T. gammatolerans* McrB protein; RMC, methylated binding site where R is a purine and MC is a methylcytosine; m⁵A, 6-methyladenosine; Tg, *T. gammatolerans*; EcΔ155, N-terminal DNA-binding domain of *Escherichia coli* McrB; m⁵C, 5-methylcytosine; SEC, size-exclusion chromatography; rm, nonmethylated; SAD, single-wavelength anomalous diffraction; YTH, YT521-B homology; Hs, human (*Homo sapiens*); ssDNA, single-stranded DNA; AAA+, extended ATPases associated with various cellular activities; SeMet; selenomethionine-labeled; PDB, Protein Data Bank; meDNA, methylated DNA; EMSA, electrophoretic mobility shift assay; IPTG, isopropyl-β-D-galactopyranoside; PMSF, phenylmethylsulfonyl fluoride.
Structure of the T. gammatolerans McrB N-terminal domain

and positioned into a pocket in the N-terminal domain, where they form numerous hydrogen bonds and hydrophobic interactions (Fig. S1). The concomitant insertion of a tyrosine residue (Tyr185) into the resulting gap stabilizes the duplex via base stacking. This strategy, although elegant, cannot simply be extrapolated to other McrB homologs because their N-terminal domains vary significantly in sequence, size, and predicted structural fold across different bacterial and archaeal species (see Fig. 1). In the handful of sequences that show identifiable homology to EcMcrB in this region (e.g. Rhizobium sp. CF907), the tyrosine plug is not conserved, and its mutation to the corresponding residue at that position—either alanine or glutamine—results in loss of DNA binding in vitro (21). These findings imply that McrB homologs have evolved different mechanisms for substrate binding and/or may preferentially target other sequences and modifications. In support of this, we previously showed that the N-terminal domain of Helicobacter pylori LLaJR1, a distant relative of the McrB family, uses a B3 domain to recognize DNA site-specifically (22).

Here we present the crystal structure of the N-terminal DNA-binding domain of Thermococcus gammatolerans McrB (TgΔ185) both alone and in complex with methylated DNA at 1.68 and 2.27 Å, respectively. TgΔ185 is structurally distinct from the EcMcrB DNA-binding domain, adopting a YTH domain fold commonly found in eukaryotic proteins that bind methylated RNA. Filter-binding experiments show that TgΔ185 does not bind RNA and instead preferentially associates with 6-methyladenosine-modified DNA. Structural characterization of the TgΔ185–DNA complex coupled with mutagenesis reveals that TgMcrB uses base flipping and an aromatic cage to recognize the modified base and an extended basic surface to associate with DNA preferentially. Together, these findings highlight a new biological function for YTH domains and underscore the notion that McrBC is a modular nuclease that can be adapted to a broad array of targets.

Results

TgMcrB does not preferentially bind m^6C DNA

To understand the broader species-specific determinants of McrB DNA binding, we identified McrB homologs in the Department of Energy Integrated Microbial Genomes database by BLAST using the E. coli McrB AAA+ domain amino acid sequence as the query. Bona fide McrBC homologs were selected based on the presence of both the McrB consensus motif in the AAA+ domain (15) and a neighboring McrC nuclease gene immediately downstream in the genome (Fig. 1). Of these, we chose the McrB homolog from T. gammatolerans (Tg) and purified its full-length protein (TgMcrB) and isolated N-terminal domain (TgΔ185; Fig. 2A). We reasoned that this homolog would provide new insights into McrB specificity because structural modeling algorithms failed to assign any known fold with high confidence and would be amenable to crystallographic and biochemical studies because Tg is a hyperthermophilic, radiation-tolerant archaea with enhanced thermostability (23).

Specificity for DNA-containing methylated cytosines is a defining feature of EcMcrB (8–13). Because TgΔ185 shares little sequence homology with the EcMcrB DNA-binding domain (EcΔ155; Fig. 2B), we first asked whether it could bind 5-methylcytosine (m^5C) modified DNA substrates (Table S1). Initial characterization by analytical size-exclusion chromatography (SEC) showed that TgΔ185 forms stable complexes with m^5C DNA similar to EcΔ155 (Fig. 2, C and D). To assess these interactions quantitatively, we examined the retention of radiolabeled m^5C and nonmethylated (nm) DNA in the presence of full-length TgMcrB or EcMcrB on alkaline-treated nitrocellulose filter paper (24). Filter binding shows that EcMcrB has a strong preference for m^5C DNA with a calculated binding constant on the order of ~160 nM (Fig. 2E and Table 1). TgMcrB, in contrast, binds both m^5C and nm DNA almost equally but with weaker affinity than EcMcrB (calculated binding constants of ~700 nM) (Fig. 2E and Table 1). These data indicate that TgMcrB is distinct from EcMcrB and displays a different sensitivity to modified DNA.

TgΔ185 adopts a YTH domain fold and preferentially binds m^6A DNA

To understand the molecular basis for the observed specificity differences, we determined the crystal structure of TgΔ185 at 1.68 Å by selenium single-wavelength anomalous diffraction (SAD) phasing (25) (Fig. 3A and Table 2). TgΔ185 is comprised of a six-stranded β-sheet—ordered β6–β1–β3–β4–β5–β2—that is flanked by clusters of α-helices (Fig. 3B). The strands adopt a mainly antiparallel arrangement with only β1 and β3 oriented in a parallel fashion. The extended β4 strand subdivides the sheet and induces a sharp curvature that nearly folds the two opposing segments onto one another. Helical segments insert in loops that flank the β-sheet: α1 and α2 in the β1–β2 loop; α3 and α4 in the β4–β5 loop; and α5 and α6 in the β5–β6 loop. Importantly, the overall topology of the TgΔ185 fold differs from that of EcΔ155 (Fig. 3, C and D).

The DALI alignment algorithm (26) indicates that TgΔ185 shares structural homology with YT521-B homology (YTH) domains (Z score, 7.5–8.5; root-mean-square deviation, 3.0–3.5) (Fig. 3, E and F). YTH domains are conserved RNA-binding modules that specifically recognize 6-methyl-adenosine (m^6A) modifications (27–29). In eukaryotes, m^6A modifications are linked to the regulation of alternative splicing, RNA processing, mRNA degradation, and the circadian clock (30–33). Given the structural similarity to YTH domains and lack of specificity toward m^5C DNA, we tested whether TgΔ185 can associate with m^6A-modified RNA. Filter binding shows that although the human (Hs) YTHDC1 YTH domain specifically associates with m^6A RNA (calculated binding constant of ~400 nM), TgΔ185 shows little affinity for either the methylated or nonmethylated RNA substrates (Fig. 4B and Table 1). We next asked whether TgΔ185 could bind m^6A-modified DNA. Surprisingly, TgΔ185 associates more tightly with m^6A dsDNA, exhibiting a ~5.5-fold increase in affinity compared with m^5C or nonmethylated dsDNA substrates (Fig. 4B and Table 1). This enhancement appears to be driven solely by the modification, because single-stranded DNA oligonucleotides show the same binding profile (Fig. 4C and Table 1). These data indicate that TgΔ185 is a DNA-specific YTH domain that preferentially targets substrates containing m^6A modifications.
An aromatic cage in TgΔ185 confers specificity for m^6^A DNA

Crystallographic studies have shown that YTH domains recognize m^6^A via a conserved “aromatic cage,” wherein two to three aromatic residues provide stabilizing π-stacking and hydrophobic interactions (34–40). Structural superposition with the m^6^A-bound YTH domain from HsYTHDF2 (PDB code 4rdn; Z score, 8.5; root-mean-square deviation, 3.1) identifies Trp^53^, Trp^115^, and Phe^121^ as putative cage residues in TgΔ185, poised to serve as a binding site for modified bases (Fig. S2).

To confirm this hypothesis, we determined the crystal structure of TgΔ185 in complex with DNA (Fig. 5A and Table 2). Although TgΔ185 crystallized with a variety of different modified substrates, suitable diffraction could only be obtained with a 19-mer dsDNA substrate that had single-base pair overhangs and contained two mismatches flanking internal m^5^C modifications in each strand (meDNA; Fig. 5B and Table S1). Incorporation of mismatches did not significantly alter the binding profile of TgMcrB (Fig. 4D and Table 1). Initial maps at 2.64 Å revealed partial, discontinuous DNA density associated with each TgΔ185 monomer and strong peaks for backbone phosphates. Numerous bases throughout the duplex, however, remained poorly resolved. An incomplete model for the TgΔ185–meDNA complex was built and used for molecular replacement into a 2.27 Å resolution isomorphous data set (Table 2). The higher resolution data set yielded vastly improved phases and interpretable electron density for both a flipped-out base and the base pairs within the surrounding DNA duplex (Fig. 5, C–E).

Despite a 19-mer substrate being used for crystallization, the asymmetric unit contains a single TgΔ185 monomer and strong peaks for backbone phosphates. Numerous bases throughout the duplex, however, remained poorly resolved. An incomplete model for the TgΔ185–meDNA complex was built and used for molecular replacement into a 2.27 Å resolution isomorphous data set (Table 2). The higher resolution data set yielded vastly improved phases and interpretable electron density for both a flipped-out base and the base pairs within the surrounding DNA duplex (Fig. 5, C–E).

Figure 1. N-terminal domains of McrB homologs are not conserved. The diagram illustrates phylogenetic analysis of representative McrB homologs. Conserved C-terminal, GTP-specific AAA+ domains are colored light blue. Divergent N-terminal domains are colored differently according to the predicted fold. The protein folds of homologs highlighted in red have been experimentally validated by X-ray crystallography. Department of Energy Integrated Microbial Genomes codes (62) and any applicable PDB codes are as follows: Yersinia pestis sv. Orientalis CO-92 McrB, 637199492; Acinetobacter baumannii D1279779 McrB, 2563734192; Bacillus cereus 03BB102 McrB, 643761446; Thermococcus gammatolerans EJ3 McrB, 644807740; Staphylococcus aureus MRSA252 McrB, 637153557; Lysinibacillus fusiformis SW-B9 McrB, 2598933124; Firmicutes bacterium JGI 000019-P10 McrB, 2519130374; Rhizobium sp. CF097 McrB, 2585392831; E. coli K-12 MG1655 McrB, 646316336; PDB code 35SC; Staphylococcus marinus F1, DSM 3639 McrB, 64019242; PDB code 6N05; Lactococcus lactis 1AA59 LlaI.1, 263206860; L. lactis 1AA59 LlaI.2, 2652068006; L. lactis LlaI.R1, 642916737; and H. pylori LlaI.R1, 637022177, PDB code 6C5D.
TgΔ185 monomers throughout the lattice interact with different DNA sequences. This implies that the resulting electron density attributed to the DNA represents the average distribution of the bases over the length of the duplex rather than a single, defined sequence. A similar scenario has been observed with *Streptomyces coelicolor* IHF, wherein crystallization with a 19-mer DNA substrate yielded an asymmetric unit with eight nucleotides. During refinement, we modeled all possible sequence registers of the substrate and chose the one that yielded the lowest $R_{\text{free}}$ value and the strongest base density. The preferred sequence based on these parameters positions an adenine as the flipped-out base (Fig. 5D). The apo- and DNA-bound TgΔ185 superimpose with an average root-mean-square deviation of 0.549 Å, indicating no significant structural changes occur in the protein upon substrate binding. We do note, however, a significant widening of the major groove (Fig. 5F) that likely arises from both TgΔ185-induced base flipping (Fig. 5D) and the presence of mismatches in the DNA substrate that enhanced crystallization (Fig. 5B).

As predicted, Trp$^{53}$, Trp$^{115}$, and Phe$^{121}$ form an aromatic cage that stabilizes each flipped-out adenine base (Fig. 6A). The organization of this pocket mirrors the stabilization of m$^6$A in the HsYTHDC1 YTH domain–m$^2$A ssRNA complex (PDB code 4r3i; Z score, 7.7; root-mean-square deviation, 3.3) (Fig. 6B). In HsYTHDC1, mutation of either cage tryptophan (W377A or W428A) completely abolishes m$^6$A RNA binding. To assess how the Tg aromatic cage contributes to DNA binding and modified base recognition, we engineered a triple alanine mutant (W53A/W115A/F121A) in full-length TgMcrB and measured how this construct interacts with m$^6$A-modified dsDNA by filter binding (Fig. 6C). W53A/W115A/F121A shows a 5.3-fold reduction in binding relative to WT (Table 1). This finding was corroborated using electrophoretic mobility shift assays (EMSAs) to measure the association of TgMcrB with 5-methylcytosine modified (m$^5$C) and nonmethylated (nm) phage DNA (Fig. 7). We observe a significant gel shift with WT TgMcrB on m$^6$A DNA (Fig. 7A) with higher affinity compared with nonmethylated DNA (Fig. 7B). The W53A/W115A/F121A triple mutant, however, significantly impairs binding to m$^6$A DNA (Fig. 7C versus A) but not to nonmethylated DNA (Fig. 7D versus B). Importantly, these changes reduce binding to a level that is comparable with WT TgMcrB's...
affinity for m^3C or nm DNA (Table 1). This suggests that the cage residues primarily confer specificity for methylated adenines and that other structural features mediate the preferred association with DNA. Although Glu^16 and Asn^20 also form hydrogen bonds to the flipped-out base (Fig. S3A), disruption of these interactions by mutagenesis (E16A/N19A) has no significant effect on m^6A–DNA binding (Fig. S3B and Table 1).

### An expanded basic patch facilitates TgMcrB DNA binding

HsYTHDC1 and TgΔ185 both contain a basic patch surrounding the aromatic cage that interacts with the negatively charged backbone of nucleic acids (Fig. 6, D and E). The area of this interaction surface is dramatically increased in TgΔ185 (Fig. 6E, dashed magenta circle), which facilitates the binding of a duplex rather than a single strand of nucleic acids. Several arginines within these patches contact the bound substrate in each structure (Fig. S3, C–E). In HsYTHDC1, Arg^275 stabilizes the resulting gap caused by base flipping and π-stacks with the G-1 base, whereas Arg^402 engages the phosphate backbone (Fig. S3C). Arg^275 appears to be more critical, because mutation of this side chain to alanine decreases binding affinity over 100-fold (38). In TgΔ185, Arg^78 and Arg^81 engage the major groove near the flipped-out base, whereas Arg^35 and Arg^162 contact the phosphate backbone on opposite strands (Fig. S3, D and E). Mutation of both Arg^78 and Arg^81 to alanine, surprisingly, has no effect on DNA binding via filter binding (Fig. S3B and Table 1) despite a similar spatial orientation that is analogous to Arg^275 in HsYTHDC1 (Fig. S3, C and D). The R78A/R81A double mutant, however, shows reduced affinity for both methylated (Fig. 7E) and nonmethylated (Fig. 7F) 5′ phage DNA via EMSA, suggesting that these side chains may play a role in mediating alternative structural contacts with DNA.

### Discussion

Chemical modifications in nucleic acids serve as important markers that critically control a wide array of cellular processes (42). DNA modifications are central to the epigenetic regulation of gene expression and transcriptional events (43, 44), activation of DNA repair pathways (45, 46), and defense machineries that underlie the ongoing arms race between bacterial hosts and predatory bacteriophage viruses (1, 6). m6A modification of RNA affects stem cell pluripotency, cancer, splicing, circadian rhythm, immunity, sex determination, and viral replication (27–29). Recent structural and biochemical studies have established that eukaryotic YTH domains act as “readers” of this RNA methylation and orchestrate the recruitment of different effector complexes to these sites (47). Here we showed that the N-terminal domain of the archaeal McrB homolog from *T. gammatolerans* (TgΔ185) adopts a YTH domain fold and shows a preference for m^6A modified DNA *in vitro*. This specificity sets it apart from every other YTH domain and opens the potential capabilities for how this fold can be utilized in nature. It remains to be seen whether TgΔ185 is an outlier among the family that simply co-opted this binding module or whether other noneukaryotic YTH domains exist and share a similar propensity for targeting DNA. A more exhaustive bioinformatic analysis coupled with structural and biochemical validation will be necessary to clarify this in the future.

Canonical YTH domains recognize m^6A modifications using base flipping and an aromatic cage that when mutated completely abolishes RNA binding (34, 35, 36, 37, 38, 39, 40). Our structural data indicate that TgΔ185 employs the same general strategy. The contribution of the aromatic cage to the overall substrate binding, however, is less significant than in other YTH domains: cage mutations only impair binding to m^6A DNA by ~5-fold, reducing it to a level that approaches TgMcrB’s intrinsic affinity for nonmethylated DNA (Figs. 6C and 7C and Table 1). We also note that TgΔ185 contains two arginine residues (Arg^78 and Arg^81) that are spatially conserved near the flipped-out base. In other YTH domains, one or more of these residues are important in sequence-specific recognition of the 5′ base immediately upstream of the modified base (35, 36, 38, 39). Interestingly, these residues have little influence on DNA binding via filter binding, but display a drastic decrease in both m6A and nonmethylated 5′ phage DNA binding. We interpret the disparity between the two assays as being a consequence of additional sequence specificity exhibited by TgMcrB that manifests when exposed to the greater sequence diversity present in the pool of lambda fragments. Moreover, unlike the aromatic cage triple mutant that only displays decreased affinity for m^6A DNA, the R78A/R81A double mutant impacts both…

---

Table 1

| Construct                  | DNA or RNA | K_d (mM) | Error |
|----------------------------|------------|----------|-------|
| EcMcrB WT                  | m^3C dsDNA | 0.1605   | 0.02810 |
| EcMcrB WT                  | nmC dsDNA  | ND^a     | ND^a  |
| TgMcrB WT                  | m^3C dsDNA | 0.7070   | 3.5809 |
| TgMcrB WT                  | nmC dsDNA  | 0.7234   | 6.2383 |
| TgMcrB WT                  | m^6A dsDNA | 0.1297   | 3.5809 |
| TgMcrB WT                  | nmA dsDNA  | 0.4906   | 3.0492 |
| TgMcrB WT                  | m^6A RNA 7-mer | ND^a   | ND^a  |
| TgMcrB WT                  | nm RNA 7-mer | ND^a   | ND^a  |
| HsYTHDC1                  | m^6A RNA 7-mer | 0.4102 | 5.918B |
| HsYTHDC1                  | nm RNA 7-mer | ND^a   | ND^a  |
| TgMcrB WT                  | m^6C dsDNA mm | 0.6155 | 4.564G |
| TgMcrB WT                  | nmC dsDNA mm | 0.6478 | 4.1890 |
| TgMcrB WT                  | m^6A dsDNA mm | 0.1005 | 1.904B |
| TgMcrB WT                  | nmA dsDNA mm | 0.4980 | 4.9150 |
| TgMcrB WT                  | m^6C ssDNA (US) | 0.5763 | 7.5342 |
| TgMcrB WT                  | nmC ssDNA (US) | 0.6609 | 6.0508 |
| TgMcrB WT                  | m^6A ssDNA (US) | 0.2019 | 1.8898 |
| TgMcrB WT                  | nmA ssDNA (US) | 0.8485 | 3.8188 |
| TgMcrB W53A/W11SA/F121A    | m^6A dsDNA | 0.6923   | 2.2231 |
| TgMcrB E17A/N19A           | m^6A dsDNA | 0.0779   | 1.5973 |
| TgMcrB Y61A/N82A           | m^6C dsDNA | ND^a     | ND^a  |
| TgMcrB R78A/R81A           | m^6A dsDNA | 0.0994   | 2.4268 |

^a ND, not determined because of incomplete saturation within the data acquisition range.
m^6A and nonmethylated DNA binding. This suggests that these side chains form alternative contacts to DNA that are independent from m^6A recognition. These findings argue that the aromatic cage primarily dictates TgMcrB’s preferred specificity for the m^6A modification and that overall DNA binding is mediated by other structural features. To this end, we observe that TgΔ185 contains an extended basic surface that associates with the second strand of the DNA duplex (Fig. 6, D and E). These subtle structural differences further distinguish TgΔ185 from other YTH domains.

Although aromatic cage mutations reduce DNA binding, the Y61A/N82A double mutant increases TgMcrB’s affinity for m^6A DNA by nearly 7-fold (Fig. S2 and Table 1). Together, these side chains shape the contours of the binding surface and form a wedge into the major groove of the bound DNA (Fig. S3F). We hypothesize that removing these features may relax the structural constraints needed for binding and may increase tolerance for different substrates, similar to how distortions in the substrate caused by mismatched base pairs helped facilitate stable interactions in the crystal lattice.

Bacterial McrBC homologs function as defense systems that restrict foreign bacteriophage DNA (2, 7). Despite sharing conserved AAA+ motor and nuclease machineries, each complex characterized to date exhibits a unique specificity that is determined by the nonconserved N-terminal domain of its associated McrB protein (Fig. 1). Thus, although E. coli McrB targets DNA containing methylated cytosines via a DUF3578 fold (8–13), more distantly related family members like LlaJI, LlaI, and BsuMI recognize DNA site-specifically (48–50) using modules like a B3 domain in some instances (22). Our TgΔ185 structures and biochemical data define a new modality of binding—using a YTH domain to bind m^6A-modified DNA—not previously observed or predicted for any McrB protein. Numerous archaeal viruses have been found that exhibit m^6A genomic methylation and/or carry genes encoding for adenine methyltransferases (51–53). This suggests that archaea like T. gammatolerans have modified the modular McrBC scaffold in response to evolutionary pressures imposed by their viral pathogens much in the same manner as their bacterial counterparts.

Because of its ability to recognize and cleave m^5C DNA, EcMcrBC is commonly used as a diagnostic tool to monitor epigenetic changes underlying mammalian gene expression (54), tissue-specific development (55), and perturbation to normal methylation patterning associated with human diseases like Prader–Willi and Angelman syndromes (56) and Fragile-X mental retardation (57). Members of the PvuRts1I family have been similarly employed to map 5-hydroxymethylcytosine

---

**Figure 3. TgΔ185 adopts a YTH fold that is distinct from the EcΔ155 fold.** A and B, structure (A) and topology (B) of TgΔ185 (yellow). C and D, structure (C) and topology (D) of EcΔ155 (orange). E, topology diagram of HsYTHDF2 YTH domain (light blue). F, structural superposition of TgΔ185 (yellow) and the HsYTHDF2 YTH domain (light blue).
Recent studies have implicated N\(^6\)-methyladenine modification as an important epigenetic marker in mammalian cells (60, 61). Our structural and biochemical results suggest TgMcrBC could be utilized in a similar capacity to track and map dynamic changes in patterns of m\(^6\)A methylation. Further biochemical characterization of the full restriction complex will provide a platform for this application.

### Experimental procedures

**Identification and phylogenetic analysis of McrB homologs**

Putative McrB homologs were initially identified by BLAST using the sequence of the *E. coli* McrB AAA+ domain to search against the Department of Energy Integrated Microbial Genomes database (62). These candidates were only considered if they contained the conserved McrB consensus motif MNXX-DRS and the presence of an adjacent McrC gene that could be confirmed by neighbor analysis. Homologs were then subdivided into groups according to their divergent N-terminal domains. A phylogenetic tree incorporating a representative from each group was generated using the Department of Energy Integrated Microbial Genomes analysis tools. Structural fold prediction for each unique N-terminal domain was carried out using the Phyre 2 protein fold recognition server.

**Cloning, expression, and purification of TgMcrB constructs**

DNA encoding the *T. gammatolerans* EJ3 McrB protein (Department of Energy Integrated Microbial Genomes database code 644807740) was codon-optimized for *E. coli* expression and synthesized commercially by GENEART. DNA encoding full-length TgMcrB was amplified by PCR and cloned into pET21b, introducing a His\(_6\) tag at the C terminus. DNA encoding the N-terminal domain (Tg\(|\Delta\)185, residues 1–185) was amplified by PCR and cloned into pET15bP, a modified pET15b expression and synthesized commercially by Novagen. DNA encoding the N-terminal domain (Tg\(|\Delta\)185, residues 1–185) was amplified by PCR and cloned into pET15bP, a modified pET15b (Novagen) plasmid in which an Hrv3C protease site (LEV-GP) replaces the thrombin site after the N-terminal His\(_6\) tag. Native TgMcrB and Tg\(|\Delta\)185 were transformed into BL21(DE3) cells, grown at 37 °C in Terrific Broth to an OD\(_{600}\) of 1.0, and then induced with 0.3 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG) overnight at 19 °C. All cells were harvested, washed with nickel load buffer (20 mM HEPES, pH 7.5, 500 mM NaCl, 30 mM imidazole, 5% glycerol (v/v), and 5 mM β-mercaptoethanol), and pelleted a second time. Pellets were flash frozen in liquid nitrogen and stored at −80 °C. Selenomethionine-labeled (SeMet) Tg\(|\Delta\)185 was expressed in minimal medium in the absence of auxotrophs as described previously (63).

Thawed Pellets from 500-ml cultures were resuspended in 30-ml of nickel load buffer supplemented with 10 mM phenylmethylsulfonyl fluoride (PMSF), 5 mg of DNase (Roche), 5 mM MgCl\(_2\), and a complete protease inhibitor mixture tablet (Roche). Lysozyme was added to 1 mg/ml, and the mixture was incubated for 15 min rocking at 4 °C. The cells were disrupted by sonication, and the lysate was cleared of debris by centrifugation at 13,000 rpm (19,685 × g) for 30 min at 4 °C.
For native and SeMet TgΔ185, the supernatant was filtered, loaded onto a 5-ml HiTrap chelating column charged with NiSO₄ and then washed with nickel load buffer. TgΔ185 was eluted with an imidazole gradient from 30 mM to 1 M. Pooled fractions were dialyzed overnight at 4 °C into nickel-loading buffer with reduced salt (50 mM NaCl) in the presence of Hrv3C protease to remove the N-terminal His tag. The sample was reapplied to a 5-ml HiTrap chelating column charged with NiSO₄. The flow through was fractionated to collect cleaved TgΔ185, concentrated, and further purified by SEC using a Superdex 75 16/600 pg column.

For full-length TgMcrB, the supernatant from sonication was filtered, heated to 65 °C for 20 min, centrifuged at 4,000 rpm (6,057 × g) for 10 min at 4 °C, and filtered again prior to purification on a 5-ml HiTrap chelating column as described above. Pooled peak fractions were concentrated and purified further by SEC.

All proteins were exchanged into a final buffer of 20 mM HEPES, pH 7.5, 150 mM KCl, 5 mM MgCl₂, and 1 mM DTT during SEC and concentrated to 5–40 mg/ml. SeMet TgΔ185 was purified similarly but was supplemented with 5 mM DTT in the SEC buffer. TgMcrB mutants were generated by QuickChange mutagenesis (Agilent Technologies) and confirmed by sequencing.

Cloning, expression, and purification of EcMcrB constructs

DNA encoding the full-length E. coli McrB protein (UniProt P15005; Department of Energy Integrated Microbial Genomes database code 646316336) was codon-optimized for E. coli expression and synthesized commercially by GENEART. DNA encoding the full-length EcMcrB (residues 1–459) and the N-terminal domain (Ec/H9004 155, residues 1–155) were cloned into pMAL-c2Xp, a modified pMAL-c2X (New England Biolabs) plasmid in which an Hrv3C protease site replaces the Factor Xa site after the N-terminal MBP tag. Both constructs were transformed into BL21(DE3) cells, grown at 37 °C in Terrific Broth to an A₆₀₀ of 1.0, and then induced with 0.3 mM IPTG overnight at 19 °C. All cells were harvested, washed with TGED500 (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 1 mM EDTA, 5% glycerol (v/v), and 1

Figure 4. TgMcrB preferentially binds DNA containing m⁶A modifications. All data points represent average of three independent experiments (means ± S.D.). Binding constants were determined by nonlinear curve fitting using Kaleidagraph (Synergy Software) and defined as the concentration of the protein at which 50% of the labeled DNA substrate is retained. Substrate sequences and calculated Kᵣ values are listed in Table S1 and Table 1, respectively. m⁵C and m⁶A denote 5-methylcytosine and 6-methyladenine modifications, respectively. nmC and nmA denote nonmethylated versions of the same substrates. A, filter-binding analysis of TgMcrB and HsYTHDC1 YTH domain interactions with RNA substrates. B, filter-binding analysis of TgMcrB interactions with dsDNA substrates. C, filter-binding analysis of TgMcrB interaction with different single stranded DNA (ssDNA) substrates. D, filter-binding analysis of TgMcrB with different mismatched dsDNA substrates.
mM DTT), and pelleted a second time. The pellets were flash frozen in liquid nitrogen and stored at −80 °C.

Thawed pellets from 500-ml cultures were resuspended in 30 ml of TGED500 supplemented with 10 mM PMSF, 5 mg of DNase (Roche), 5 mM MgCl₂, and a complete protease inhibitor mixture tablet (Roche). Lysozyme was added to 1 mg/ml, and the mixture was incubated for 15 min of rocking at 4 °C. The cells were disrupted by sonication, and the lysate was cleared of debris by centrifugation at 13,000 rpm (19,685 × g) for 30 min at 4 °C. Each supernatant was filtered, loaded onto 30–40 ml of amylose resin, washed with TGED500, and eluted with TGED500 supplemented with 10 mM D-maltose. Pooled fractions were dialyzed overnight at 4 °C into TGED with reduced salt (TGED50, 50 mM NaCl) in the presence of Hrv3C protease to remove the N-terminal MBP tag. Samples were then applied to a 5-ml HiTrap Q HP ion-exchange column in TGED50 and eluted with a NaCl gradient from 50 to 500 mM. Pooled fractions were concentrated and further purified by SEC using a Superdex 75 10/300 GL column. Both full-length and Ec155 McrB were exchanged into a final buffer of 20 mM HEPES, pH 7.5, 150 mM KCl, 5 mM MgCl₂, and 1 mM DTT during SEC and concentrated to 5–40 mg/ml.

**Cloning, expression, and purification of HsYTHDC1**

DNA encoding the human YTHDC1 YTH domain (residues 344–509) was codon-optimized for E. coli expression and synthesized commercially by Integrated DNA Technologies and cloned into pET15b. The HsYTHDC1 344–509 was transformed into BL21(DE3) cells, grown at 37 °C in Terrific Broth to an A₆₀₀ of 1.0, and then induced with 0.3 mM IPTG overnight at 19 °C. All cells were harvested, washed with nickel load buffer, and pelleted a second time. The pellets were flash frozen in liquid nitrogen and stored at −80 °C. Thawed pellets from 500-ml cultures were resuspended in 30-ml of
nickel load buffer supplemented with 10 mM PMSF, 5 mg of DNase (Roche), 5 mM MgCl₂, and a complete protease inhibitor mixture tablet (Roche). Lysozyme was added to 1 mg/ml, and the mixture was incubated for 15 min rocking at 4 °C. The cells were disrupted by sonication, and the lysate was cleared of debris by centrifugation at 13,000 rpm (19,685 g) for 30 min at 4 °C. The supernatant was filtered, loaded onto a 5-ml HiTrap chelating column charged with NiSO₄, washed with nickel load buffer, and eluted with an imidazole gradient from 30 mM to 1 M. Pooled fractions were concentrated and further purified by SEC using a Superdex 75 10/300 GL column. HsYTHDC1 344–509 was exchanged into a final buffer of 20 mM HEPES, pH 7.5, 150 mM KCl, 5 mM MgCl₂, and 1 mM DTT during SEC and concentrated to 5–40 mg/ml.

Preparation of oligonucleotide substrates

All DNA and RNA substrates for analytical SEC, filter binding, and crystallization were purchased from Integrated DNA Technologies. Lyophilized nonmethylated and HPLC-purified modified single-stranded oligonucleotides were resuspended in to 1 mM in 10 mM Tris-HCl and 1 mM EDTA and stored at −20 °C until needed. Single-stranded oligonucleotides were 5′ end-labeled with [γ-³²P]ATP using polynucleotide kinase (New England Biolabs) and then purified on a P-30 spin column (Bio-Rad) to remove unincorporated label. Duplex substrates were prepared by heating equimolar concentrations of complementary strands (denoted with suffixes “us” and “ls” indicating upper and lower strands) to 95 °C for 15 min followed by cooling to room temperature overnight and then purification on an S-300 spin column (GE Healthcare) to remove single-stranded DNA. Table S1 shows the sequence of each oligonucleotide used in this work.

Analytical size-exclusion chromatography

Samples (50 μl) of 100 μM EcΔ155 or TgΔ185 were mixed with m⁵C dsDNA in a 2:1.2 molar ratio in 20 mM HEPES, pH 7.5, 150 mM KCl, 5 mM MgCl₂, and 1 mM DTT and incubated at
Figure 7. EMSA analysis of predicted TgMcrBΔ185 DNA-binding mutants. Binding was carried out at 25 °C for 30 min in a 16-μl reaction mixture containing 5 ng/μl of digested (BamHI/NdeI) m⁵A methylated (dam⁺) and nonmethylated (dam⁻) λ-phage DNA and increasing concentrations (0–10 μM) of each full-length TgMcrB construct. The gels were stained with SYBR® Green in 1× TAE overnight at 25 °C. Calculated sizes (bp) of the digested DNA products are noted on the left of each gel.
**Structure of the T. gammatolerans McrB N-terminal domain**

room temperature for 10–15 min. Each reaction was fractionated via gel filtration on a Superdex 75 3.2/300 analytical SEC column equilibrated with 20 mM HEPES, pH 7.5, 150 mM KCl, 5 mM MgCl₂, and 1 mM DTT. Fractions containing samples were subjected to 4–20% gradient SDS-PAGE, silver-stained for DNA, and Coomassie-stained for protein.

**Filter-binding assays**

The standard buffer for the DNA-binding assays contained 25 mM MES, pH 6.5, 2.0 mM MgCl₂, 0.1 mM DTT, 0.01 mM DNA, and Coomassie-stained for protein.

**Electrophoretic mobility shift assays**

The standard buffer for the EMSAs contained 10 mM Tris-HCl, pH 8.0, 250 mM NaCl, 1 mM MgCl₂, and 1 mM DTT. Binding was performed with purified full-length TgMcrB (WT or mutants) or HsYTHDC1 YTH domain at 30 °C for 10 min in a 30-μl reaction mixture containing 14.5 nM unlabeled DNA and 0.5 nM labeled DNA. The samples were filtered through KOH-treated nitrocellulose filters (Whatman Protran BA 85, 0.45 μm) using a Hoefer FH225V filtration device for ~1 min. The filters were subsequently analyzed by scintillation counting on a 2910TR digital, liquid scintillation counter (PerkinElmer Life Sciences). All measured values represent the average of at least three independent experiments (means ± S.D.) and were compared with a negative control to determine fraction bound. Binding constants were determined by nonlinear curve fitting using Kaleidagraph (Synergy Software) and defined as the concentration of the protein at which 50% of the labeled DNA substrate is retained. Calculated $K_d$ values are listed in Table 1. Error values were calculated automatically in Kaleidagraph (Synergy Software) and represent the overall percentage deviations of the data from the final curve fit.

**Crystallization, X-ray data collection, and structure determination**

SeMet TgΔ185 was crystallized by sitting-drop vapor diffusion in 0.1 M MES, pH 6.5, 3.2 M (NH₄)₂SO₄ with a drop size of 2 μl and a reservoir volume of 650 μl. Crystals appeared within 10–14 days at 20 °C and were of the space group $P2_12_12_1$ with unit cell dimensions $a = 41.87$ Å, $b = 56.50$ Å, $c = 109.28$ Å, $α = 90.00°$, $β = 90.00°$, and $γ = 90.00°$. The samples were cryoprotected with Parabar 10312 and frozen in liquid nitrogen. An initial 2.64 Å data set (TgMcrB D185 + meDNA 1) was collected at Northeastern Collaborative Access Team 24-ID-E Beamline at the Advanced Photon Source at the selenium edge energy at 12.663 keV (0.9791 Å) (Table 2). The data were integrated and scaled using the Northeastern Collaborative Access Team RAPD pipeline. Heavy atom sites were located using SHELX (65), and phasing, density modification, and initial model building were carried out using the Autobuild routines of the PHENIX package (66). Further model building and refinement was carried out manually in COOT (67) and PHENIX, respectively (66). The final model contained one molecule in the asymmetric unit containing residues 1–175 and was refined to 1.68 Å resolution with $R_{work}/R_{free}$ values of 0.1770/0.1907 (Table 2).

SeMet TgΔ185 was crystallized in complex with a 19-mer methylated DNA substrate (meDNA; Table S1) by sitting-drop vapor diffusion in 0.1 M HEPES, pH 7.5, 20% PEG 3350, and 0.20 M (NH₄)₂SO₄ with a drop size of 2 μl and reservoir volume of 650 μl. meDNA contained a single mC modification in each strand (meDNA upper strand and lower strand oligonucleotides; Table S1) and flanking sequences that produced base pair mismatches in the annealed double stranded duplex that were necessary to obtain diffraction quality crystals. TgΔ185 and meC15 mismatched DNA were mixed at a molar ratio of 2:1.2 and incubated at room temperature for 10–15 min prior to crystallization experiments. Crystals appeared within 10–14 days at 20 °C and were of the space group $P2_12_12_1$ with unit cell dimensions $a = 41.87$ Å, $b = 56.50$ Å, $c = 109.28$ Å, $α = 90.00°$, $β = 90.00°$, and $γ = 90.00°$. The samples were cryoprotected with Parabar 10312 and frozen in liquid nitrogen. An initial 2.64 Å data set (TgMcrB D185 + meDNA 1) was collected at Northeastern Collaborative Access Team 24-ID-E Beamline at the selenium edge energy at 12.663 keV (0.9791 Å) and solved by molecular replacement in PHASER (68) using the unbound TgΔ185 monomer structure determined from selenium SAD phasing as the search model (Table 2). Discontinuous portions of the DNA could be visualized and built; however, the overall model did not improve significantly beyond the initial rounds of refinement. A more complete model was obtained using the diffraction data from a second crystal, TgΔ185 + meDNA 2 (Table 2). This structure was solved by molecular replacement with PHASER (67) using the MR-derived structure from TgΔ185 + meDNA 1 as the search model. The statistics and resulting maps following subsequent rounds of manual model building and refinement continued to improve, ultimately revealing density for the DNA backbone and individual bases. This strategy proved critical, because the density for the DNA remained poorly resolved if the unbound TgΔ185 monomer was instead used as a search model for molecular replacement. The final model of crystal 2 contained one molecule in the asymmetric unit containing residues 3–175 with 6 bp of the DNA substrate and was refined to 2.27 Å resolution with $R_{work}/R_{free}$ of 0.2378/0.2893 (Table 2).

Structural superpositions were carried out in Chimera (69). All structural renderings were generated using PyMOL (Schrodinger), and surface electrostatics were calculated using APBS (70).
References

1. Labrie, S. J., Samson, J. E., and Moineau, S. (2010) Bacteriophage resistance mechanisms. Nat. Rev. Microbiol. 8, 317–327 CrossRef Medline

2. Loenen, W. A., and Raleigh, E. A. (2014) The other face of restriction: modification-dependent enzymes. Nucleic Acids Res. 42, 56–69 CrossRef Medline

3. Bair, C. L., and Black, L. W. (2007) A type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs. J. Mol. Biol. 366, 768–778 CrossRef Medline

4. Borgaro, J. G., and Zhu, Z. (2013) Characterization of the 5-hydroxymethylcytosine-specific DNA restriction endonucleases. Nucleic Acids Res. 41, 4198–4206 CrossRef Medline

5. Ishikawa, K., Fukuda, E., and Kobayashi, I. (2010) Conflicts targeting epigenetic systems and their resolution by cell death: novel concepts for methyl-specific and other restriction systems. DNA Res. 17, 325–342 CrossRef Medline

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

7. Juria, S. E., and Human, M. L. (1952) A nonhereditary, host-induced variation of bacterial viruses. J. Bacteriol. 64, 557–569 Medline

8. Sutherland, E., Coo, L., and Raleigh, E. A. (1992) McrB: a multisubunit GTP-dependent restriction endonuclease. J. Mol. Biol. 225, 327–348 CrossRef Medline

9. Krüger, T., Wild, C., and Noyer-Weidner, M. (1995) McrB: a prokaryotic protein specifically recognizing DNA containing modified cytosine residues. EMBO J. 14, 2661–2669 CrossRef Medline

10. Gast, F. U., Brinkmann, T., Pieper, U., Krüger, T., Noyer-Weidner, M., and Pingoud, A. (1997) The recognition of methylated DNA by the GTP-dependent restriction endonuclease McrBC resides in the N-terminal domain of McrB. Biol. Chem. 378, 975–982 Medline

11. Pieper, U., Schweitzer, T., Groll, D. H., and Pingoud, A. (1999) Definition of the location and function of domains of McrB by deletion mutagenesis. Biol. Chem. 380, 1225–1230 Medline

12. Stewart, F. J., Panne, D., Bickle, T. A., and Raleigh, E. A. (2000) Methylation-specific DNA binding by McrBC, a modification-dependent restriction enzyme. J. Mol. Biol. 298, 611–622 CrossRef Medline

13. Zagorskaite, E., Manakova, E., and Samsasauskas, G. (2018) Recognition of modified cytosine variants by the DNA-binding domain of methyl-directed endonuclease McrBC. FEBS Lett. 592, 3335–3345 CrossRef Medline

14. Panne, D., Müller, S. A., Wirtz, S., Engel, A., and Bickle, T. A. (2001) The McrBC restriction endonuclease assembles into a ring structure in the presence of G nucleotides. EMBO J. 20, 3210–3217 CrossRef Medline

15. Pieper, U., Schweitzer, T., Groll, D. H., Gast, F. U., and Pingoud, A. (1999) The GTP-binding domain of McrB: more than just a variation on a common theme? J. Mol. Biol. 292, 547–556 CrossRef Medline

16. Pieper, U., and Pingoud, A. (2002) A mutational analysis of the PD/EXK motif suggests that McrC harbors the catalytic center for DNA cleavage by the GTP-dependent restriction enzyme McrBC from Escherichia coli. Biochemistry 41, 5236–5244 CrossRef Medline

17. Panne, D., Raleigh, E. A., and Bickle, T. A. (1999) The McrBC endonuclease translocates DNA in a reaction dependent on GTP hydrolisis. J. Mol. Biol. 290, 49–60 CrossRef Medline

18. Pieper, U., Groll, D. H., Wünsch, S., Gast, F. U., Speck, C., Mücke, N., and Pingoud, A. (2002) The GTP-dependent restriction enzyme McrBC from Escherichia coli forms high-molecular mass complexes with DNA and produces a cleavage pattern with a characteristic10-base pair repeat. Biochemistry 41, 5245–5254 CrossRef Medline

19. Bourniquel, A. A., and Bickle, T. A. (2002) Complex restriction enzymes: NTP-driven molecular motors. Biochimie 84, 1047–1059 CrossRef Medline

20. Dryden, D. T., Murray, N. E., and Rao, D. N. (2001) Nucleoside triphosphate-dependent restriction enzymes. Nucleic Acids Res. 29, 3728–3741 CrossRef Medline

21. Sukackaite, R., Grazulis, S., Tamulaitys, G., and Siksnys, V. (2012) The recognition domain of the methyl-specific endonuclease McrBC flips out 5-methylcytosine. Nucleic Acids Res. 40, 7552–7562 CrossRef Medline

22. Hosford, C. J., and Chappie, J. S. (2018) The crystal structure of the Helicobacter pylori LjII.R1 N-terminal domain provides a model for site-specific DNA binding. J. Mol. Biol. 293, 11758–11771 CrossRef Medline

23. Jolivet, E., L’Haridon, S., Corre, E., Forterre, P., and Prieur, D. (2003) Thermococcus gammatolerans sp. Nov., a hyperthermophilic archaeon from a deep-sea hydrothermal vent that resists ionizing radiation. Int. J. Syst. Evol. Microbiol. 53, 847–851 CrossRef Medline

24. Papoulas, O. (2002) Rapid separation of protein-bound DNA from free DNA using nitrocellulose filters. Curr. Protoc. Mol. Biol. 12.8.1–12.8.9 CrossRef Medline

25. Hendrickson, W. A. (2014) Anomalous diffraction in crystallographic phase evaluation. Q. Rev. Biophys. 47, 49–93 CrossRef Medline

26. Holm, L., and Rosenström, P. (2010) DalI server: conservation mapping in 3D. Nucleic Acids Res. 38, W545–W549 CrossRef Medline

27. Zhang, Z., Theler, D., Kaminska, K. H., Miller, M., de la Grange, P., Pudimat, R., Rafalska, I., Heinrich, B., Bujnicki, J. M., Allain, F. H., and Stammb, S. (2010) The YTH domain is a novel RNA binding domain. J. Biol. Chem. 285, 14701–14710 CrossRef Medline

28. Liu, K., Ding, Y., Ye, W., Liu, Y., Yang, J., Liu, J., and Qi, C. (2016) Structural and functional characterization of the proteins responsible for N^4-methyladenosine modification and recognition. Curr. Protein Pept. Sci. 17, 306–318 CrossRef Medline

29. Zhao, Y. L., Liu, Y. H., Wu, R. F., Bi, Z., Yao, Y. X., Liu, Q., Wang, Y. Z., and Wang, X. C. (2019) Understanding m^5A function through uncovering the diversity roles of YTH domain-containing proteins. Mol. Biotechnol. 61, 355–364 CrossRef Medline

30. Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Dixon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Aragiliro, N., Kupiec, M., Sorek, R., and Rechavi, G. (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature 485, 201–206 CrossRef Medline

31. Schwartz, S., Agarwala, S. D., Mumbach, M. R., Jovanovic, M., Mertins, P., Shikshin, A., Tabach, Y., Mikkelsen, T. S., Satija, R., Czech, P., Garr, S. A., Lander, E. S., Fink, G. R., and Regev, A. (2013) High-resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. Cell 155, 1409–1421 CrossRef Medline

32. Fustin, J. M., Doi, M., Yamaguchi, Y., Hida, H., Nishimura, S., Yoshida, M., Isagawa, T., Morikawa, S. M., Kakeya, H., Morita, S., and Okamura, H. (2010) The RNA-methylation-dependent RNA processing controls the speed of the circadian clock. Cell 155, 793–806 CrossRef Medline

33. Wang, X., Lu, Z., Gomez, A., Hon, G. C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., Ren, B., Pan, T., and He, C. (2014) N^4-Methyladenosine-dependent regulation of messenger RNA stability. Nature 505, 117–120 CrossRef Medline
34. Li, F., Zhao, D., Wu, J., and Shi, Y. (2014) Structure of the YTH domain of human YTHDF2 in complex with an mA mononucleotide reveals an aromatic cage for mA recognition. *Cell Res.* 24, 1490–1492. [CrossRef] [Medline]

35. Luo, S., and Tong, L. (2014) Molecular basis for the recognition of methylated adenosines in RNA by the eukaryotic YTH domain. *Proc. Natl. Acad. Sci. U.S.A.* 111, 13834–13839. [CrossRef] [Medline]

36. Theler, D., Dominguez, C., Blatter, M., Boudet, J., and Allain, F. H. (2014) Solution structure of the YTH domain in complex with N⁶-methyladenosine RNA: a reader of methylated RNA. *Nucleic Acids Res.* 42, 13911–13919. [CrossRef]

37. Zhu, T., Roundtree, I. A., Wang, P., Wang, X., Wang, L., Sun, C., Tian, Y., Li, J., He, C., and Xu, Y. (2014) Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N⁶-methyladenosine. *Cell Res.* 24, 1493–1496. [CrossRef]

38. Xu, C., Wang, X., Liu, K., Roundtree, I. A., Tempel, W., Li, Y., Li, Z., He, C., and Min, J. (2014) Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nat. Chem. Biol.* 10, 927–929. [CrossRef]

39. Xu, C., Liu, K., Ahmed, H., Loppnau, P., Schapira, M., and Min, J. (2015) Structural basis for the discriminative recognition of N⁶-methyladenosine RNA by the human YT521-B homology domain family of proteins. *J. Biol. Chem.* 290, 24902–24913. [CrossRef]

40. Wang, C., Zhu, Y., Bao, H., Jiang, Y., Xu, C., Wu, J., and Shi, Y. (2016) A novel RNA binding mode of the YTH domain reveals the mechanism for recognition of determinant of selective removal by Mmi1. *Nucleic Acids Res.* 44, 969–982. [CrossRef]

41. Swierz, J. P., Nianji, T., Gloyd, M., Guarné, A., and Elliot, M. A. (2013) A novel nucleoid-associated protein specific to the actinobacteria. *Nucleic Acids Res.* 41, 4171–4184. [CrossRef]

42. Chen, K., Zhao, B. S., and He, C. (2016) Nucleic acid modifications in RNA by the human YT521-B homology domain family of proteins. *Cell Chem. Biol.* 23, 74–85. [CrossRef]

43. Goldberg, A. D., Allis, C. D., and Bernstein, E. (2007) Epigenetics: a landscape takes shape. *Cell* 128, 635–638. [CrossRef]

44. Adhikari, S., and Curtis, P. D. (2016) DNA methyltransferases and epigenetic regulation in bacteria. *FEMS Microbiol. Rev.* 40, 575–591. [CrossRef]

45. Dalhus, B., Laerdahl, J. K., Backe, P. H., and Bjarås, M. (2009) DNA base repair: recognition and initiation of catalysis. *FEMS Microbiol. Rev.* 33, 1044–1078. [CrossRef]

46. Mullins, E. A., Rodriguez, A. A., Bradley, N. P., and Eichman B. F. (2019) Emerging roles of DNA glycosylases and the base excision repair pathway. *Trends Biochem. Sci.* 44, 765–781. [CrossRef] [Medline]

47. Liao, S., Sun, H., and Xu, C. (2018) YTH domain: a family of N⁶-methyladenosine (mA) readers. *Genomics Proteomics Bioinformatics* 16, 99–107. [CrossRef]

48. Jentsch, S. (1983) Restriction and modification in *Bacillus subtilis: sequence specificities of restriction/modification systems BsuM, BsuE, and BsaF.* *J. Bacteriol.* 156, 880–888. [Medline]

49. Hill, C., Miller, L. A., and Kleinhamer, T. R. (1991) *In vivo* genetic exchange of a functional domain from a type II A methylase between lactococcal plasmid pTR2030 and a virulent bacteriophage. *J. Bacteriol.* 173, 4363–4370. [CrossRef]

50. O’Driscoll, J., Heiter, D. F., Wilson, G. G., Fitzgerald, G. F., Roberts, R., and van Sinderen, D. (2006) A genetic dissection of the Llal restriction casette reveals insights on a novel bacteriophage resistance system. *BMC Microbiol.* 8, 40. [CrossRef]

51. Witte, A., Baranyi, U., Klein, R., Sulzner, M., Luo, C., Wanner, G., Krüger, D. H., and Lubitz, W. (1997) Characterization of *Natronobacterium gregarium* phage phiCh1, a unique archaeal phage containing DNA and RNA. *Mol. Microbiol.* 23, 603–616. [CrossRef]

52. Arnold, H. P., Ziese, U., and Zillig, W. (2000) SNDV, a novel virus of the extremely thermophilic and acidophilic archaeon Sulfolobus. *Virology* 272, 409–416. [CrossRef]