



OPEN ACCESS

EDITED BY

Alberto Danielli,
University of Bologna, Italy

REVIEWED BY

Robert Martin Blumenthal,
University of Toledo, United States
Ankan Banerjee,
University of Marburg, Germany

*CORRESPONDENCE

Matthias Bochtler
✉ mbochtler@iimcb.gov.pl
Shuang-yong Xu
✉ xus2007@gmail.com

[†]These authors have contributed equally to this work

RECEIVED 31 August 2023

ACCEPTED 08 February 2024

PUBLISHED 09 April 2024

CITATION

Helbrecht I, Heiter D, Yang W, Vincze T, Hanneman A, Lutz T, Ettwiller L, Bochtler M and Xu S-y (2024) Characterization of winged helix domain fusion endonucleases as N6-methyladenine-dependent type IV restriction systems. *Front. Microbiol.* 15:1286822. doi: 10.3389/fmicb.2024.1286822

COPYRIGHT

© 2024 Helbrecht, Heiter, Yang, Vincze, Hanneman, Lutz, Ettwiller, Bochtler and Xu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Characterization of winged helix domain fusion endonucleases as N6-methyladenine-dependent type IV restriction systems

Igor Helbrecht^{1,2,3†}, Daniel Heiter^{1†}, Weiwei Yang¹, Tamas Vincze¹, Andrew Hanneman¹, Thomas Lutz¹, Laurence Ettwiller¹, Matthias Bochtler^{2,3*} and Shuang-yong Xu^{1*}

¹New England Biolabs, Inc., Ipswich, MA, United States, ²Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland, ³International Institute of Molecular and Cell Biology, Warsaw, Poland

Winged helix (wH) domains, also termed winged helix-turn-helix (wHTH) domains, are widespread in all kingdoms of life and have diverse roles. In the context of DNA binding and DNA modification sensing, some eukaryotic wH domains are known as sensors of non-methylated CpG. In contrast, the prokaryotic wH domains in DpnI and HhiV4I act as sensors of adenine methylation in the 6mA_PT (N6-methyladenine, 6mA, or N6mA) context. DNA-binding modes and interactions with the probed dinucleotide are vastly different in the two cases. Here, we show that the role of the wH domain as a sensor of adenine methylation is widespread in prokaryotes. We present previously uncharacterized examples of PD-(D/E)XK-wH (FcyTI, Psp4BI), PUA-wH-HNH (HtuIII), wH-GIY-YIG (Ahi29725I, Apa233I), and PLD-wH (Aba4572I, CbaI) fusion endonucleases that sense adenine methylation in the Dam⁺ Gm6ATC sequence contexts. Representatives of the wH domain endonuclease fusion families with the exception of the PLD-wH family could be purified, and an *in vitro* preference for adenine methylation in the Dam context could be demonstrated. Like most other modification-dependent restriction endonucleases (MDREs, also called type IV restriction systems), the new fusion endonucleases except those in the PD-(D/E)XK-wH family cleave close to but outside the recognition sequence. Taken together, our data illustrate the widespread combinatorial use of prokaryotic wH domains as adenine methylation readers. Other potential 6mA sensors in modified DNA are also discussed.

KEYWORDS

winged helix fusion endonucleases, HNH endonuclease, GIY-YIG endonuclease, PLD family endonuclease, N6mA-dependent restriction system, genome conflict

Introduction

In most restriction-modification (R-M) scenarios, nucleobase modification serves as a mark of self and provides protection against endonuclease digestion. In some cases, however, phages have learned to exploit this principle by modifying their own DNA, either by incorporation of non-standard nucleoside triphosphates or by post-replicative modifications catalyzed either by host or phage enzymes. Modification-dependent restriction endonucleases (MDREs) specifically target such modified DNA (modified base or backbone). The MDREs

come in two main groups, distinguished by the presence or absence of nucleoside triphosphate (NTP)-consuming motor proteins. The NTP-independent proteins are typically modular, with separate modification sensing and DNA cleavage domains. Because of this architecture, DNA cleavage typically takes place at a distance from the site of modification. For some enzymes, a single site is sufficient, but typically, cleavage is most efficient when it is directed by appropriately spaced modifications, which cooperate to position an endonuclease dimer for a double strand (ds) cut in the DNA.

The catalytic domains present in restriction can be grouped into the almost universally used hydrolases (Orlowski and Bujnicki, 2008) and the very rarely used lyases (Miyazono et al., 2014). The hydrolases, in turn, can be grouped into a surprisingly small set of phylogenetically unrelated enzyme groups. PD-(D/E)XK enzymes are named for characteristic amino acids (aa) built around a central β -sheet, which harbors one or two catalytic Mg^{2+} ions (Pingoud et al., 2005). The metal ions are held in place in part by the D and D or E (abbreviated as D/E) interacting residues, which, together with a K residue, activate a water molecule for direct inline attack on the scissile phosphate (Bujnicki and Rychlewski, 2001; Kosinski et al., 2005). HNH enzymes, also called $\beta\beta\alpha$ -Me enzymes or His-Me finger enzymes (Jablonska et al., 2017; Wu et al., 2020), harbor a single metal cation in their active site. Metal identity requirements are less strict than for PD-(D/E)XK enzymes. Many divalent transition metal ions are acceptable (Pommer et al., 2001). Like PD-(D/E)XK enzymes, the HNH enzymes are believed to catalyze attack on the scissile phosphate by a water molecule. However, water activation is not by a lysine residue but by the first histidine of the HNH motif (Flick et al., 1998; Sokolowska et al., 2009). GIY-YIG enzymes (Dunin-Horkawicz et al., 2006; Kaminska et al., 2008) also bind a single metal cation in the active site. These enzymes activate the water molecule with a tyrosine residue, most likely from the GIY motif (Sokolowska et al., 2011). Finally, there are also completely metal-independent endonuclease domains. They resemble phospholipase D; therefore, the enzymes containing them are known as PLD endonucleases (Grazulis et al., 2005; Chan et al., 2007). The PLD enzymes are believed to catalyze phosphodiester cleavage via a covalent intermediate (Sasnauskas et al., 2010).

The modification sensor domains, like the endonuclease domains, are now understood to be classifiable into only a few groups of phylogenetically unrelated sensors. The largest group of sensors is the PUA (PseudoUridine synthase and Archaeosine transglycosylase) superfamily (Lutz et al., 2019). PUA superfamily sensors comprise SRA (SET and Ring finger Associated) domains with specificity for 5mC, as in MspJI (Cohen-Karni et al., 2011), and related domains (Kazrani et al., 2014; Shao et al., 2014), originally also termed SRA domains, with specificity for 5-hydroxymethylcytosine (5hmC) and glucosyl-5-hydroxymethyl-cytosine (g5hmC), as in the PvuRtsII family of restriction endonucleases (Janosi et al., 1994; Borgaro and Zhu, 2013). The PUA superfamily also comprises EVE (according to the PDB identifier 2eve for a prototypical protein; Bertoni et al., 2009) domains specific for 5mC and 5hmC, as found in VcaM4I (Pastor et al., 2021), and YTH (YT521-B Homologs) domains (YTH-McrB/NTPase fusion) specific for 6-methyladenine (6mA) (Iyer et al., 2006; Hosford et al., 2020; Xu et al., 2020), as well as ASCH (ASC-1 Homology) domains. Bioinformatic analysis has suggested that ASCH domains might be 6mA readers (Iyer et al., 2016), but this prediction has not been confirmed by experimental data so far. Instead, it has been shown that the *E. coli* YqfB, an ASCH domain protein, is able to

hydrolyze various N4-acylated cytosines (4acC) and cytidines (Stanislauskiene et al., 2020). All PUA superfamily domains are engaged in nucleotide flipping (Cerrudo et al., 2014; Pastor et al., 2021). Irrespective of their detailed specificity, they scrutinize the modified base in a dedicated pocket of the enzyme (Roberts and Cheng, 1998; Horton et al., 2014). It has been shown that *E. coli* McrBC endonuclease also recognizes modified cytosines by base flipping (Sukackaite et al., 2012).

Apart from the PUA superfamily, other modification sensor domains may also be involved in restriction, such as the NEco domain in EcoKMcrA with affinity for 5mC and 5hmC (Czapinska et al., 2018). Unlike the PUA superfamily domains, the NEco domain senses 5mC or 5hmC without nucleotide flipping in the context of dsDNA (Slyvka et al., 2019). Finally, a winged helix (wH) domain has been described as a 6mA sensor in DpnI. Like the NEco domain, the wH (winged helix) domain senses nucleobase modifications in the context of dsDNA without flipping (Mierzejewska et al., 2014). However, in contrast to the NEco domain, it has specificity for 6mA rather than 5mC. Also, in contrast to NEco, which recognizes methyl groups of fully methylated CpG in two separate pockets, the wH domain recognizes methyl groups of fully methylated ApT in a single pocket, exploiting their proximity in space. The wH domain in DpnI is unusual in being fused to a nuclease domain, which has a separate sequence (GATC) and modification (6mA) specificity (Siwiek et al., 2012). Therefore, it acts more like an effector domain in type IIE enzymes (Senesac and Allen, 1995; Roberts et al., 2003), except that both the nuclease and sensor/effector domain are specific for methylated rather than non-methylated DNA.

Winged helix (wH) domains are a group of DNA-binding domains that belong to the superfamily of helix-turn-helix (HTH) proteins (Brennan, 1993; Lai et al., 1993; Gajiwala and Burley, 2000). Structurally, canonical winged helix domains consist of an N-terminal α -helices and β -strand, the HTH motif, and a β -hairpin. The “wings” of the motif are the loops connecting the strands of the β -hairpin and immediately downstream of it (Iyer et al., 2016; Figure 1A). Winged helix motifs were first found in transcription factors, but it is now clear that they also have roles in transcription initiation complexes (Teichmann et al., 2012), in the binding of left-handed Z-DNA (Schwartz et al., 2001) or RNA (Tang et al., 2021), or in protein-protein interactions (Wah et al., 1997). In transcription factors, wH domains tend to interact with DNA, just as would be expected for the HTH motif that is embedded within them. In other words, they insert the second helix of the HTH motif, which is the third helix of the wH domain, into the major groove of DNA (Gajiwala and Burley, 2000). However, other DNA-binding modes are also possible in special cases (Gajiwala et al., 2000; Wolberger and Campbell, 2000). A recent example of such alternative binding modes is the complexes of eukaryotic winged helix domains with dsDNA containing non-methylated CpG (Stielow et al., 2021; Becht et al., 2023; Weber et al., 2023). A winged helix motif in a restriction endonuclease (REase) was first noticed in the DNA-binding domain of FokI (Wah et al., 1997), but this particular wH domain does not appear to be involved in interactions with DNA.

The role of the winged helix domain in adenine methylation sensing was first noticed in DpnI. DpnI is a G6mATC-specific endonuclease that cleaves within the recognition sequence and has a strong preference for DNA that is adenine-methylated in both strands (Siwiek et al., 2012). In DpnI, the winged helix domain plays the role

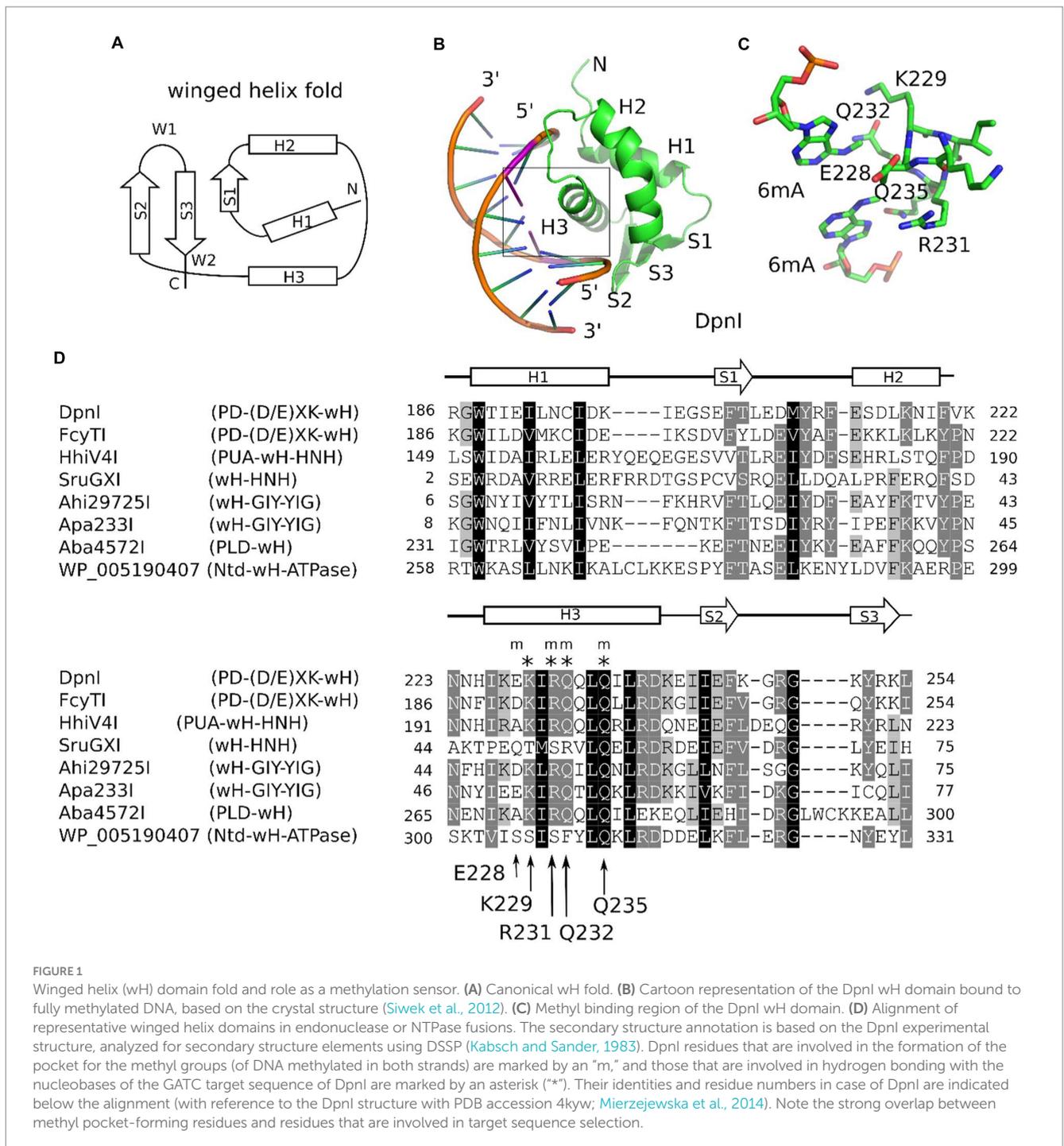


FIGURE 1

Winged helix (wH) domain fold and role as a methylation sensor. **(A)** Canonical wH fold. **(B)** Cartoon representation of the DpnI wH domain bound to fully methylated DNA, based on the crystal structure (Siwek et al., 2012). **(C)** Methyl binding region of the DpnI wH domain. **(D)** Alignment of representative winged helix domains in endonuclease or NTPase fusions. The secondary structure annotation is based on the DpnI experimental structure, analyzed for secondary structure elements using DSSP (Kabsch and Sander, 1983). DpnI residues that are involved in the formation of the pocket for the methyl groups (of DNA methylated in both strands) are marked by an "m," and those that are involved in hydrogen bonding with the nucleobases of the GATC target sequence of DpnI are marked by an asterisk (*). Their identities and residue numbers in case of DpnI are indicated below the alignment (with reference to the DpnI structure with PDB accession 4kyw; Mierzejewska et al., 2014). Note the strong overlap between methyl pocket-forming residues and residues that are involved in target sequence selection.

of an effector domain that senses 6mA separately from and with slightly relaxed sequence specificity compared to the PD-(D/E)XK nuclease domain (Mierzejewska et al., 2014; Figures 1B,C). More recently, a winged helix domain was also implicated in the sensing of adenine methylation, also in the G6mATC context. HHPV4I (also called HhiV4I; Lu et al., 2023) is a three-domain enzyme, with a PUA (SRA)-like domain at the N-terminus, a winged helix domain in the middle, and an HNH endonuclease domain at the C-terminus. The PUA superfamily domain, described as an SRA domain by Lu et al., appears not to be involved in DNA modification sensing (Lu et al., 2023). By contrast, the winged helix domain directed preferential

cleavage of Dam⁺ over Dam⁻ DNA, and it has much higher affinity to Dam⁺ than to Dam⁻ DNA in gel shift experiments. In contrast to DpnI, HHPV4I (HhiV4I) cleaves at a distance from the site of adenine methylation, suggesting that the endonuclease domain is directed by the winged helix domain and does not sense adenine methylation on its own (Lu et al., 2023).

In this study, we show that the winged helix domain is widely used as an adenine methylation sensor in MDREs (Figure 1D). We present additional examples of proteins that share the PD-(D/E)XK-wH architecture with DpnI or the PUA-wH-HNH architecture with HHPV4I (HhiV4I). Additionally, we show that the wH domain is can

also be naturally paired with an HNH domain in the absence of a PUA superfamily domain, with a GIY-YIG endonuclease domain, with a PLD (phospholipase D) nuclease domain, or with an NTPase (GTPase/ATPase) domain. For the PD-(D/E)XK—wH, PUA—wH—HNH, wH—HNH, wH—GIY-YIG, and PLD—wH enzymes, we detect Dam⁺-dependent toxicity in *E. coli* cells, either by tight binding or digestion near the modified sites. For the fusion endonucleases PD-(D/E)XK—wH, PUA—wH—HNH, wH—HNH, and wH—GIY-YIG, but not the PLD—wH enzymes, we find representatives that are active also *in vitro*, and we show that their preferred substrate is fully methylated DNA with one enzyme exception. Unlike DpnI, many of the new wH fusion endonucleases cleave DNA outside the G6mATC recognition sequence and two modified sites in tandem with a short spacer, enhance their cleavage activity.

Materials and methods

Materials

E. coli T7 expression strains C2566 (Dam⁺) and its isogenic Dam-deficient strain ER2948 [constructed and provided by Dr. Lise Raleigh, New England Biolabs (NEB)], expression vector pTXB1, pBR322, phage λ DNA (Dam⁺ or Dam⁻), 2-log (1 kb plus) DNA ladder, chitin beads, restriction enzymes, EcoGII methylase (frequent adenine methylase), Q5 DNA polymerase PCR master mix and cloning kit (Hi-Fi DNA assembly enzyme mix), NEBExpress Ni-NTA magnetic beads, and dZTP (2-aminoadenine triphosphate is abbreviated as base Z in the literature, here we use dZ to denote the 2'-deoxynucleoside) were provided by Michael Kuska (NEB Organic Synthesis Division). Fast-flow Ni-agarose beads were from Qiagen or NEB. The T7 expression vector pET21b with C-terminal 6 \times His tag was originally purchased from Novagen (NdeI-XhoI). 5hmdCTP/dGTP/dATP/dTTP mix was purchased from Zymo Research. FPLC DEAE and Heparin columns (5 mL) were purchased from GE HealthCare or Cytiva.

Endonuclease assays

For restriction of modified DNA, Dam⁺ pBR322 and λ DNA were used. In some cases, we also used M.EcoGII-modified pBR322 (Dam⁻), with all modified adenine bases except in polyA tracks. Restriction buffers used were: NEB buffer 2.1 (medium salt, 50 mM NaCl) or CutSmart buffer (with 50 mM potassium acetate). Endonucleases: we usually perform an enzyme titration in endonuclease activity assays; enzyme concentrations are indicated in each digestion. Buffer compositions in 1 \times NEB restriction buffers: buffer 1.1 (10 mM Bis-Tris-Propane-HCl, 10 mM MgCl₂, 100 μ g/mL BSA or recombinant albumin, pH 7.0 at 25°C); buffer 2.1 (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl₂, 100 μ g/mL BSA or recombinant albumin, pH 7.9 at 25°C); buffer 3.1 (100 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl₂, 100 μ g/mL BSA or recombinant albumin, pH 7.9 at 25°C); CutSmart buffer (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 100 μ g/mL BSA or recombinant albumin, pH 7.9 at 25°C). For restriction digestions in different divalent cations, a medium salt buffer (50 mM NaCl, 20 mM

Tris-HCl, pH 7.5) was supplemented with divalent cations in 0.1, 1, and 10 mM final concentrations.

Synthetic oligos with modified and unmodified GATC sites

Single-stranded DNA oligos were synthesized by the NEB organic synthesis division:

Top strand Gm6ATC (top M+).

5'/56FAM/ACTCATGCAGGCATGCAGG/m6A/

TCGCAGTCAGATTTATGTGTTCATATAGT

ACGTGATTCAAG-3'.

Bottom strand Gm6ATC (bottom M+).

5'-CTTGAATCACGTACTATATGACACATAAATCTGACTGC

G/m6A/TC CTGCATGCCTGCATGAGT-3'.

Top strand GATC (unmodified, top M-).

5'/56FAM/ACTCATGCAGGCATGCAGGATCGCAGTCAGAT

TTATGTGTTCATATAGTACGTGATTCAAG-3'.

Bottom strand GATC (unmodified, bottom M-).

5'-CTTGAATCACGTACTATATGACACATAAATCTGACTG

CGATCCTGCATGCCTGCATGAGT-3'.

Duplex oligos abbreviation:

Fully modified = M+ top/M+ bottom = M+/M+.

Unmodified = M- top/M- bottom = M-/M-.

Hemi-modified = M+ top/M- bottom = M+/M-.

Hemi-modified = M- top/M+ bottom = M-/M+.

Duplex oligos were digested by DpnI (2 U), MboI (5 U), FcyTI (0.1 μ g), Ahi29725I and Apa233I (1 μ g) in NEB buffer 2.1 at 37°C for 1 h. For HhiV4I digests, reactions were carried out in NEB buffer 2.1 supplemented with 1 mM MnCl₂ (this enzyme is a Mn²⁺-dependent REase; see below).

Protein expression and purification

C2566 (Dam⁺) competent cells (cloning grade) were provided by NEB. ER2948 (Dam⁻) competent cells were prepared by a modified rubidium chloride method. *E. coli* cells were cultured at 37°C to mid-log phase, and IPTG was added to the culture at 0.5 mM final concentration for protein production (at 18°C overnight). Cells were lysed by sonication in chitin column buffer or Ni-agarose column buffer. Clarified cell lysates with over-expressed proteins (target protein-intein-CBD) or C-terminal 6 \times His-tagged proteins were loaded onto chitin or Ni-agarose columns, respectively, for affinity purification. The protein purification protocols were used as recommended by the manufacturers. In some cases, the partially purified proteins were further purified by chromatography through DEAE (flow-through to remove nucleic acids at 0.3 M NaCl concentration) and Hi-Trap Heparin (5 mL). BigDye[®] Terminator v3.1 Cycle Sequencing Kit was purchased from Thermo-Fisher (Applied Biosystems). Restriction gene inserts in plasmids were sequenced to verify the correct sequences. Dam⁺ pBR322 DNA fragments after restriction digestion were sequenced to determine the cut sites. DNA sequence edits were carried out using DNASTar or Geneious software packages. BlastP searches in the GenBank and UniProtKB databases were performed using the respective web servers. NCBI Pfam and conserved domains were used to visualize protein domains of REase homologs.

Plasmid preparation and transformation

Plasmid mini-prep kits and competent cells were provided by NEB. Plasmid mini-preparation and bacterial transformation were done according to the manufacturer's recommendation.

Clustering locus-specific annotations analysis

For clustering locus-specific annotations (CLANS) analysis, the homologs of the five groups of wH-containing enzymes (sequences of the reference proteins are listed in [Supplementary material](#)) were obtained by blast (BlastP, default settings in the UniProtKB website) using the UniProtKB reference proteomes and Swiss-Prot database. A total of 1,258 wH fusion endonuclease homolog proteins were analyzed. The resulting homolog protein fasta files were combined and subjected to CLANS analysis using the MPI Bioinformatics Toolkit. The result of the CLANS analysis is visualized with the CLANS Java application ([Frickey and Lupas, 2004](#)). CLANS analyzes all-against-all pairwise sequence similarities to establish relationships within a protein family. In the CLANS network figure, each node (a small colored dot) represents a full-length protein (or wH domain in the lower box), and the line connects two proteins/domains that share sequence similarities. The lengths of the lines represent the degree of sequence similarities, with short lines representing close similarities and vice versa. Each node is the same size, but the size of the color "blob" is related to the number of nodes clustered together.

Phylogenetic analysis

To perform phylogenetic analysis, we first built a hmm (hidden Markov model) profile with the wH domains of the six representative wH-containing enzymes (note: wH domains only, not full-length proteins). The resulting profile was used to search homologs using hmmsearch (HMMER v3.1b2) on the combined fasta file (see CLANS analysis). We extracted the wH domain in each homolog protein and performed multiple alignments with the five representative wH domains using MAFFT (v7.508). The maximum likelihood tree was constructed using RAxML (v8.2.12) with option `-f a` to enable rapid bootstrapping for 100 times. Other options used were `-p 1237`, `-x 1,237`, `-m PROTGAMMAAUTO`. The best-scoring maximum likelihood tree with bootstrap values was visualized with iTOL.¹

Proteome analysis

The four samples (10 µg each) were digested with trypsin using S-Trap micro-columns as recommended by the manufacturer (PROTIFI). LC-MS/MS of the digests was conducted on the Thermo Orbitrap Exploris, with two injections made per sample. Data were searched using Proteome Discoverer v2.5 against a combined FASTA database containing the Uniprotkb_ *E. coli*-B strain proteins and the four winged helix (wH) endonuclease protein sequences (FcyTI, HhiV4I, Ahi29725I, and Apa233I). Results are presented in 8 Excel spreadsheets, one file for each injection (see raw data on the proteome

of the purified proteins). The most confident protein ID is determined by Score Sequest HT (the last column in each table—the higher the score, the more confident the identification). Proteins identified are reported with >1 unique peptide per protein and a 1% false discovery rate.

Results

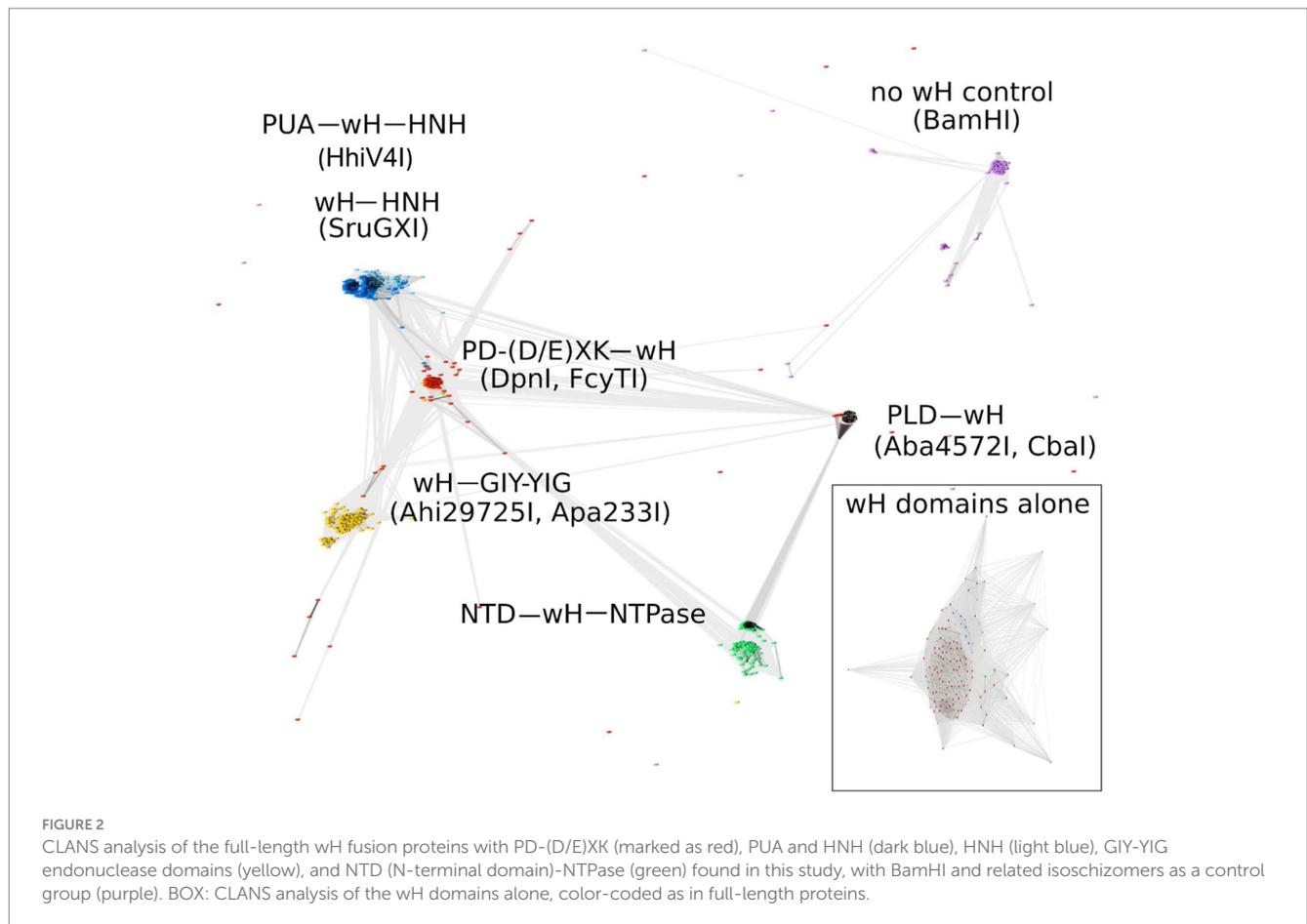
Bioinformatic screen of wH domain endonucleases

Some known wH domains are adenine methylation (6mA)-specific. This notion was originally suggested by the demonstration of 6mA specificity of the DpnI wH domain and further strengthened by the observation of adenine specificity of the wH domain of HHPV4I ([Lu et al., 2023](#)), which was reported when this study was being finalized. In the hope of finding new adenine methylation-specific endonucleases, we searched for fusions of wH domains with endonuclease domains known to play roles in R-M ([Pingoud and Jeltsch, 2001](#)). Apart from additional PD-(D/E)XK—wH and PUA—wH—HNH endonucleases, we also identified fusion proteins with a wH—HNH, wH—GIY-YIG, and PLD—wH architecture. In addition, we found PD-(D/E)XK—wH—NTPase cases that can be considered as fusions of a DpnI-like protein with an McrB-like NTPase (GTPase/ATPase) domain and NTD—wH—NTPase cases, apparently without a nuclease domain and with an N-terminal domain of unknown function. One additional example is the wH—Mrr catalytic domain (PD-QXX)—NTPase fusion endonuclease. For this study, we concentrated on the NTP-independent enzymes with wH and endonuclease domains. To better understand the sequence relationships between the new wH fusion proteins, we carried out a CLANS ([Frickey and Lupas, 2004](#)) analysis. CLANS determines all-against-all pairwise sequence similarities to establish relationships within a protein family. It is not intended to find sequence motifs within protein sequences, which are better detected using other software. In CLANS analysis of the full-length proteins [with BamHI (GGATCC) and related enzymes as a control], the wH domain-containing endonucleases segregated clearly into separate groups, driven by the sequence similarity between endonuclease domains of the same group ([Figure 2](#)). However, when we limited the CLANS analysis to the wH domains alone, the segregation into groups according to the endonuclease domain was not so clear. In particular, wH domains from PD-(D/E)XK and GIY-YIG endonucleases were fully intermingled, possibly suggesting multiple separate fusion events ([Figure 2](#), box). The finding suggests that fusions of the same type may have arisen independently several times in evolution. Nevertheless, much of the diversity of the new fusion proteins is clearly due to divergent evolution. This is also supported by the phylogenetic tree of the wH domains ([Supplementary Figure S1](#)). Note, however, that bootstrap values for most branches of the tree are very low, making the tree very tentative overall.

Rare occurrence of methyltransferase genes in the immediate genetic neighborhood of the new fusion proteins

wH proteins frequently bind DNA, but only some of them are methylation-dependent (e.g., DpnI), whereas others are not (e.g.,

¹ <https://itol.embl.de/>



FokI). If the new wH fusion proteins were modification-specific, they should occur as stand-alone enzymes. Otherwise, they should be associated with a host genome protecting DNA methyltransferase of any type (N6mA, N4mC, C5). Typically, such a methyltransferase would be located in the immediate genomic neighborhood of the endonuclease, so that the entire system could work as a defense island (Makarova et al., 2011). To test for possible association with DNA methyltransferases, we inspected the genomic neighborhoods of over 1,000 wH fusion proteins. In 87% of cases, none of the three genes adjacent upstream or downstream to the wH fusion gene encoded a DNA methyltransferase, suggesting that most of the new wH fusion proteins acted as stand-alone endonucleases, possibly as type IV restriction systems.

Likely specificity of at least some of the wH fusion proteins for 6mA in the GATC context

Adenine methylation in bacteria occurs frequently in the GATC context i.e. the target sequence of the Dam methyltransferase (MTase) (Marinus and Casadesus, 2009), which is widely distributed in bacteria because of its diverse house-keeping roles, including DNA replication (Boye and Lobner-Olesen, 1990) and mismatch repair (Au et al., 1992; Josephs et al., 2015). Hence, it was likely that the putative MDREs with the wH domain might detect adenine methylation in this sequence context. This idea was further supported by the precedent of the wH

domain in DpnI, which is known to be specific for adenine methylation in the Dam context (with some leeway for the outer bases S6mATS, where S is G or C). Inspection of the crystal structure of the DpnI domain in complex with target DNA revealed that the same residues contribute to both the methyl binding pocket and the sequence specificity, suggesting that methyl sensing and detection of the G6mATC target sequence are intricately linked (Mierzejewska et al., 2014). Large-scale analysis of the wH domain fusion proteins showed that the motif for 6mA and GATC recognition (see arrows in Figure 1D), or closely related motifs, were present in approximately 10% of the new fusion proteins. With the exception of SruGX1 as a representative of the wH-HNH endonucleases, we selected for further characterization the fusion proteins that had the motif for 6mA and GATC specificity. Such fusion proteins are very likely to recognize m6A in the GATC context. The reminder is that stand-alone wH fusion proteins are likely to recognize modified DNA (otherwise they would be toxic to the host). However, it is currently unclear whether the modification is m6A, and, if so, whether the sequence context is GATC.

Avoidance of genomic conflict

Inspection of a 10 kb interval around genes encoding the new fusion proteins revealed association with methyltransferases in some cases. PD-(D/E)XK-wH domains co-occurred with predicted C5 methyltransferases in 40 cases. In some cases, they also co-occurred

with a predicted 4mC (N4mC) or 6mA MTase directly adjacent to it. In these cases (e.g., *Bacteroidota* bacterium isolate CP064983.1, *Moraxella ovis* strain CP011158.1), the putative DNA MTase has been inactivated by a frame shift. The PUA—wH—HNH and wH—HNH co-occurred in 47 cases with Eco57I-like MTases (of type IIG R-M-S fusion enzymes). These MTases are predicted to be 6mA MTases, with CTGAAG (site of methylation underlined) as the target sequence. In the case of the wH—GIY-YIG endonucleases, we found four instances of an EcoRI-like adenine MTase nearby. These MTases are expected to methylate GAATTC. Finally, for the PLD—wH endonucleases, we detected 17 cases of proximity to EcoEI-like (GAGN₇ATGC) or EcoR124I-like (GAAN₆RTCG) type I methyltransferases, also causing no conflict. Genetic conflict would not be expected in any of these cases if the new wH fusion proteins had specificity for 6mA in the GATC context.

Next, we looked for possible genetic conflicts on a genome-wide scale, assuming that the new wH fusion proteins were specific for m6A in a GATC context. Three types of such conflict are conceivable. First, a frequent adenine methyltransferase, such as M.EcoGII, may methylate adenine to m6A in GATC, among many other contexts. Second, a Dam-like methyltransferase may specifically modify GATC sequences. Finally, there are also methyltransferases that methylate target sequences that are longer than GATC but include the GATC site in the recognition sequence. We searched for cases of such potential conflict, scanning the entire genome, not just the genomic neighborhoods. Overall, less than 100 cases of potential conflict were identified (Supplementary Tables S1–S3) for over 1,000 wH fusion proteins. Most of the wH fusion proteins in potential conflict are stand-alone enzymes without an associated methyltransferase. Genomic conflict could be avoided if these proteins recognized modified DNA containing a mark other than m6A in the GATC context (i.e., either another methylation type or m6A in a different sequence context). Alternatively, conflict may be avoided or mitigated by tight expression control of the endonuclease or the methyltransferase.

Selection of wH fusion proteins for experimental characterization

Four types of endonuclease domains are noted in wH domain fusions: (1) DpnI-like PD-(E/D)XK endonuclease; (2) HNH endonuclease domain; (3) GIY-YIG endonuclease domain; and (4) PLD family endonuclease domain. An NTD-wH-NTPase fusion is usually paired with another endonuclease subunit, such as McrC-like catalytic subunit, which is not discussed in detail here. We have not studied evolutionary relationships within each endonuclease family since the endonuclease families have been the subject of numerous review articles and research papers (Mehta et al., 2004; Grazulis et al., 2005; Pingoud et al., 2005; Dunin-Horkawicz et al., 2006). For experimental characterization, we chose representatives of the PD-(D/E)XK—wH (FcyTI), PUA—wH—HNH (HhiV4I), wH—HNH (SruGX1), wH—GIY-YIG (Ahi29725I, Apa233), and PLD—wH (Aba4572I, CbaI) fusions for further experimental characterization. In the case of the PUA—wH—HNH architecture, many additional candidate MDREs were tested in *E. coli* cells only. An attempt to purify an NTD—wH—NTPase/McrC-like subunit fusion protein was

unsuccessful. Therefore, we focused this study exclusively on single-chain proteins.

The wH fusion endonucleases exhibit Dam⁺-dependent toxicity in *E. coli* cells

Endonuclease toxicity is a good proxy for restriction in bacterial cells (Heitman and Model, 1990; Fomenkov et al., 1994). If the putative MDREs were specific for Dam-methylated DNA, they should be more toxic to Dam⁺ (C2566) than to Dam⁻ (ER2948) *E. coli* cells. We tested this prediction with our IPTG-inducible expression constructs, both under basal (no IPTG) conditions and under induction conditions (0.5 mM IPTG). Since Dam⁻ competent cells were roughly an order of magnitude less competent than Dam⁺ cells, we avoided direct comparisons of transformation efficiency between Dam⁺ and Dam⁻ cells. Instead, we quantified the reduction in colony counts for transformations with expression plasmids compared to colony counts with an empty vector. With the exception of the Aba4572I expression construct, the plasmids for all other endonuclease-containing clones caused a reduction in colony count by two to three orders of magnitude compared to the empty vector control in Dam⁺ cells under induced conditions (Figure 3). Aba4572I endonuclease may have strong non-specific endonuclease activity since it is toxic in a Dam-deficient strain under IPTG induction. The toxicity appeared to be less severe in Dam⁺ cells. This result is not well understood. If the Aba4572I outlier is disregarded, the experimental results indicate that the wH fusion endonucleases display typical “restriction” on Dam⁺ host but not Dam⁻ cells. Whether this *in vivo* “restriction” was caused by tight binding to modified sites to inhibit replication or endonuclease cleavage of modified DNA remains to be investigated.

PD-(D/E)XK—wH endonucleases

As representatives of the PD-(D/E)XK—wH family, we selected Psp4BI and FcyTI (GenBank accession numbers WP_102090895 and WP_094411979). The two enzymes have 58.4 and 58.7% amino acid (aa) sequence identity to DpnI, respectively. Psp4BI was chosen because the source organism is psychrophilic, suggesting that the enzyme might be susceptible to heat inactivation, which would be desirable for biotechnological applications. The synthetic genes with *E. coli* optimized codons were cloned into pTXB1 in fusion with intein and CBD (chitin-binding domain) and expressed in the Dam-deficient T7 expression strain. The two enzymes were affinity purified on a chitin column and released from the column by DTT-triggered cleavage. The yield of Psp4BI was low due to poor expression of the fusion protein (Psp4BI-intein-CBD) (not shown); partially purified Psp4BI gave rise to a partial digestion pattern that was retained after 4 h at 25–37°C. The low activity could be caused by a low enzyme concentration or inhibition by some impurities in the preparation. By contrast, purified FcyTI was active on Dam⁺ pBR322, pUC19 (HindIII-linearized), and phage λ DNA (Supplementary Figures S2A,B). FcyTI-specific activity was determined as approximately 32,000 U/mg protein in buffer 2.1. FcyTI could be inactivated by heating at 65°C for 30 min, which is a useful enzyme property (Supplementary Figure S3). FcyTI endonuclease was originally found in the genome of *Flavobacterium*

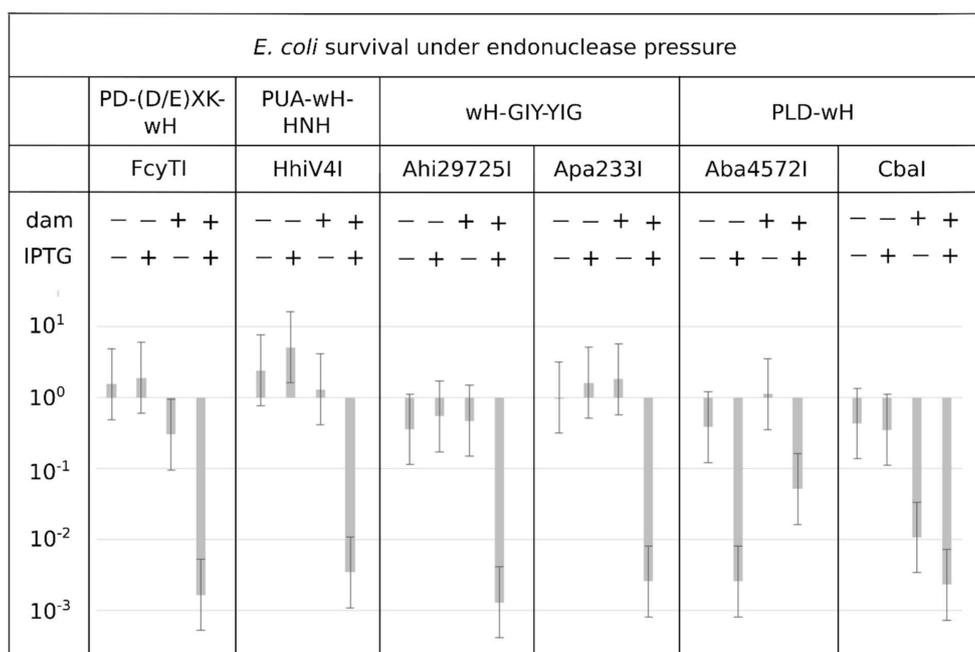


FIGURE 3 Toxicity of selected wH domain fusion endonucleases to a Dam⁺ but not a Dam⁻ host. Expression vectors containing open reading frames for putative MDREs or empty plasmid (50 ng) were transformed into Dam-positive C2566 (+) or Dam-negative ER2948 (-) *E. coli* cells, with (+) or without (-) IPTG induction. The reduction in colony counts for expression plasmid compared to the empty vector, plotted on the ordinate, is a measure of toxicity.

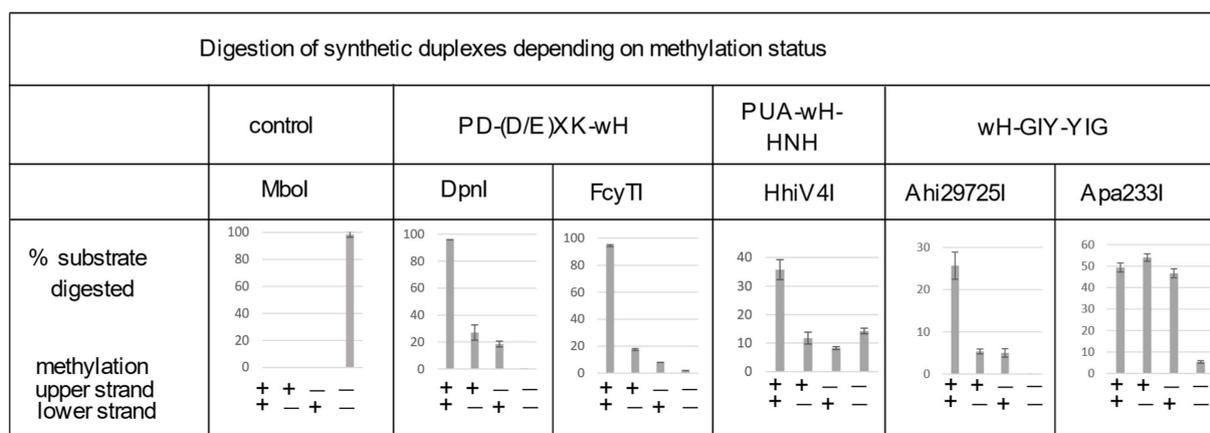


FIGURE 4 Digestion of synthetic DNA oligoduplexes depending on top and bottom strand methylation status in the Gm6ATC sequence context. 5 U of Mbol (GATC) and 2 U of Dpnl (Gm6ATC) were used as controls. FcyTI input was 0.1 μg protein (~3 U) (higher concentrations could obscure the digestion of hemi-modified duplex oligos). Duplex oligos and protein concentration in restriction digests: The duplex oligos concentration is approximately 18 nM (60mer, 21 ng in 30 μL reaction volume). The protein concentration is calculated below in the 30 μL reaction volume: (1) Ahi29725I protein (dimer) MW = 24.62 × 2 = 49.24 kDa, Ahi29725I, 1 μg = ~677 nM. (2) Apa233I protein (dimer) MW = 24.17 × 2 = 48.34 kDa, Apa233I, 1 μg = ~690 nM. (3) HhiV4I protein (dimer) MW = 44.67 × 2 = 89.34 kDa, HhiV4I, 1 μg = ~373 nM. (4) FcyTI protein (dimer) MW = 30.76 kDa × 2 = 61.52 kDa, FcyTI 0.1 μg = 54 nM.

cyanobacterium, which grows at 20–30°C. Due to its better biochemical properties, FcyTI was used for the *in vivo* toxicity study (Figure 3) and for the digestion of modified oligos (see Figure 4). The FcyTI expression plasmid showed over a 1,000-fold reduction in transformation efficiency in the Dam⁺ T7 strain compared to a Dam⁻ host (Figure 3). FcyTI could be over-expressed only in the Dam⁻ T7 expression strain. Run-off sequencing demonstrated that the enzyme was able to cleave within the Gm6A↓TC recognition sequence

(Supplementary Figure S4). The purified 6× His-tagged FcyTI shown in Supplementary Figure S5 was used for the digestion of modified or hemi-modified oligoduplexes.

To compare the activity of the enzyme toward fully methylated, hemi-methylated, and non-methylated DNA, we digested synthetic DNA oligoduplexes and quantified substrate and product amounts after restriction digestion by capillary electrophoresis (CE) (Figure 4). The results showed that FcyTI was most active on fully methylated

DNA but also had partial activity on hemi-methylated DNA, similar to DpnI. No digestion product was detected for non-methylated DNA. MboI was used as a control and digested only unmodified GATC oligos. Fully and hemi-modified substrates were resistant to MboI restriction (Figure 4). The original CE digestion results are shown in raw data (PeakScan analysis of CE peaks). Consistent with the duplex oligos digestion, unmodified pUC19 (HindIII-linearized) or phage DNA were poorly cleaved by FcyTI (Supplementary Figure S2), although weak activity was observed on Dam⁻ pUC19, probably due to the high enzyme concentration.

PUA-wH-HNH fusion endonuclease HhiV4I

6× His-tagged HhiV4I (see Supplementary Figure S5) was subjected to three-step chromatography (Ni-agarose column, DEAE column, and Heparin agarose column). Compared to the recently published paper on the same enzyme (Lu et al., 2023), two additional chromatography steps were used (DEAE and Heparin columns). Unfortunately, the Heparin agarose chromatography step was less efficient for purification than is typical for other DNA-binding proteins because HhiV4I was in the flow-through and did not bind to the Heparin column, as would be expected for a typical nucleic acid-binding protein. As a result, HhiV4I was not purified to homogeneity (Supplementary Figure S5). Mass spectrometry analysis of the contaminations identified, among other proteins, *E. coli* exonuclease (Exonuclease VII, Exo VII) as a minor contaminant (see raw data for the HhiV4I mass spectrometry study). Exo VII cleaves single-stranded DNA (ssDNA) from both the 5′→3′ and 3′→5′ directions. This enzyme is not active on linear or circular dsDNA. The contaminating exonuclease would not likely interfere with major cut site determination, but it may interfere with minor cut site(s) by removing a few nucleotides for the cleaved ends by HhiV4I if the overhang is single-stranded.

The partially purified 6× His-tagged HhiV4I was used for HhiV4I characterization. Consistent with the findings of Lu et al. (2023), we observed that HhiV4I was much more active in the presence of Mn²⁺ ions than in the presence of other divalent metal cations (Figure 5A). HhiV4I showed weak DNA-nicking activity in Mg²⁺ buffer.

In agreement with the toxicity experiments (Figure 3) and the results of Lu et al. (2023), we found that the enzyme had higher activity against Dam⁺ than Dam⁻ pBR322, pUC19, λ DNA, and synthetic duplex oligos. If HhiV4I cleaved at or near Dam⁺ sites, its cleavage products should be of similar size as those of DpnI digestion, and discrete bands (as opposed to a smear on the gel) should be observed. In our experiments, we saw only a partial match of fragment sizes, likely due to incomplete digestion (Figure 6A) (see below for the two-site requirement for efficient cleavage). Dam⁺ phage λ DNA was also only partially digested while Dam⁻ λ DNA was not cut at all (λ DNA was partially methylated by the host Dam methylase during rapid phage replication in *E. coli*, unpublished observation) (Figure 6B). When Dam⁻ λ DNA was methylated *in vitro* by Dam methylase or M.EcoGII, the DNA substrates now became cleavable by HhiV4I (Figure 6C), further demonstrating that GATC methylation is required for restriction. M.EcoGII-modified λ DNA appeared to be a slightly better substrate for HhiV4I restriction than Dam-modified λ DNA, indicating that the wH might not be strictly limited to the detection of 6mA in the GATC context.

We could digest non-methylated DNA with excess HhiV4I, suggesting that the dependence of the enzyme on adenine methylation was not absolute. Most restriction enzymes display star activity at high enzyme, high glycerol concentration, or low salt. This conclusion was confirmed with the digestion of synthetic DNA with a defined adenine methylation status. As expected, HhiV4I was most active on fully methylated DNA but had some activity on hemi- and non-methylated DNA (Figure 4). In agreement with the findings by Lu et al. (2023) we did not detect activity of HhiV4I toward PCR products, which contained 5mC or 5hmC instead of C, in conditions conducive to digestion of m6A containing DNA (Figure 5B). Since the PCR products contain 5mC and 5hmC in many different contexts, this result suggests that the enzyme has no activity against methylated or hydroxymethylated DNA, despite the presence of the PUA (SRA-like) domain. This was surprising because it had been shown previously that PUA superfamily REases VcaM4I, SRA-like domain-containing endonuclease TagI, and PvuRtsII restricted DNA containing modified cytosines (Janosi et al., 1994; Pastor et al., 2021). Possible activity against WT T4 [glucosylated(g)-5hmC] modified DNAs remains to be tested. HhiV4I shows no activity on dZ (modified adenine, 2-aminoadenine, or 2,6-diaminopurine)-modified PCR DNA (Figure 5B).

HhiV4I prefers to cut between two G6mATC sites with optimal spacers of 13–27 bp in Dam⁺ pBR322. Shorter spacers of 8–11 bp or longer spacers >42 bp were cleaved more slowly. Run-off sequencing of Dam⁺ HhiV4I DNA confirmed that the enzyme cleaved in the vicinity of but not within the G6mATC sequence, as previously reported (Supplementary Figure S6).

PUA-wH-HNH and wH-HNH endonucleases

In contrast to the prophage-encoded HhiV4I, most PUA-wH-HNH enzymes (375–496 aa long) and wH-HNH endonucleases (224–283 aa long) are bacterial/archaeal enzymes. For 15 of these enzymes and HhiV4I as a positive control, we attempted expression in the Dam⁻ *E. coli* cells. Moreover, we analyzed the transformation efficiency into Dam⁺ (C2566) and Dam⁻ (ER2948) cells compared to the empty vector. Restriction activity was examined in the presence of IPTG induction to elevate the genome conflict (Table 1). Some restriction genes, such as HhiV4I and SruGX1, had a strong toxic effect, as detected by a 100–1,000-fold reduction in transformation efficiency in the Dam⁺ host. Other ORFs caused an approximately 10-fold reduction in transformation efficiency, consistent with partial restriction (+/-). The transformation of the HhaN23I gene caused the formation of very small colonies in the presence or absence of IPTG, indicating partial restriction. A few ORF constructs showed no difference in transformation efficiency in the Dam⁺ host, presumably as a result of poor expression or lack of activity (e.g., HboP9I). As a control, the pTXB1 empty vector could be readily transferred into C2566 (Dam⁺) or ER2948 (Dam⁻) cells in the presence of IPTG (Table 1; Figure 7).

Selected enzymes that appeared to be promising as Dam⁺-dependent MDREs were partially purified, and their activity was tested on Dam⁺ pBR322 or λ DNA. The partially purified HtuIII enzyme (GenBank accession number NC_013743, PUA-wH-HNH fusion) shows a low nicking activity in Mn²⁺ or Co²⁺ buffer (Supplementary Figure S7). DNA run-off sequencing of the partially nicked pBR322 indicated that the nick occurred upstream of the

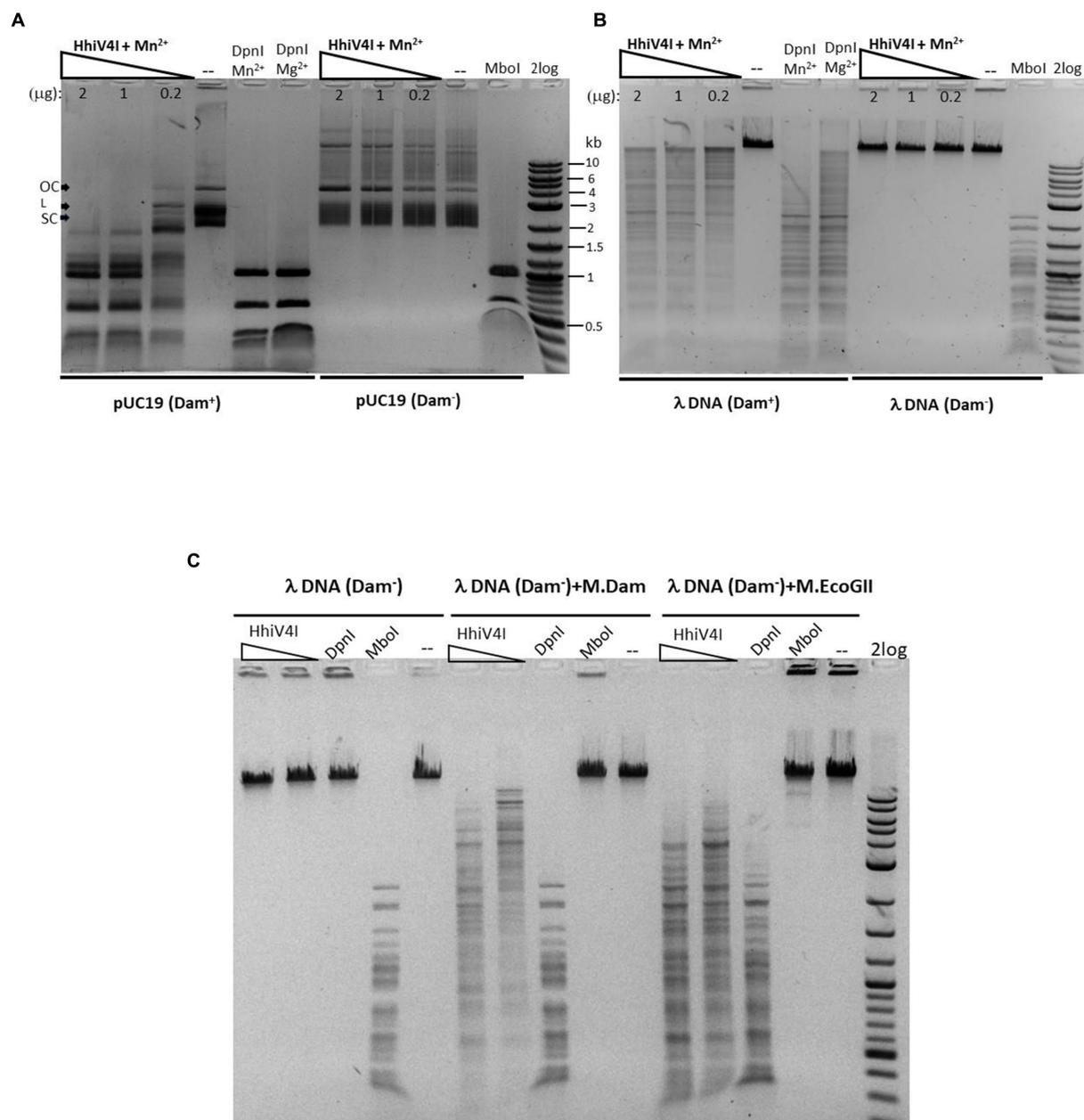


FIGURE 6

Methylation dependence of HhiV4I endonuclease. **(A)** Restriction digests of 1 μg of Dam^+ or Dam^- pUC19 in Mn^{2+} buffer by HhiV4I (447.7, 223.9, and 44.8 nM, respectively, in 50 μL reaction volume). **(B)** Restriction digests of 1 μg of Dam^+ or Dam^- phage λ DNA in Mn^{2+} buffer by HhiV4I (447.7, 223.9, and 44.8 nM, respectively, in 50 μL reaction volume). DpnI (2 U) cleaves Dam^+ DNA only. Mbol (5 U) cleaves unmodified DNA only. **(C)** HhiV4I digestion (447.7 and 44.8 nM) of 1 μg of *in vitro* modified λ DNA (Dam^-) by Dam methylase or frequent adenine methylase M.EcoGII in Mn^{2+} buffer in 50 μL reaction volume. Following methylation reactions, the methylases were inactivated by heating at 65°C for 20 min. The methylated DNA was then diluted for restriction digestion in Mn^{2+} buffer.

(Dam^+). It has a nicking activity in Ca^{2+} (10 mM) buffer. To assess the 6mA-dependent restriction activity, Ahi29725I was also assayed on Dam^- pBR322 in the presence of Mg^{2+} , Mn^{2+} , or Co^{2+} in restriction digests. The results showed that the enzyme digested Dam^- DNA into a smearing pattern in Mn^{2+} buffer without discrete bands, probably as a result of loss of specificity (data not shown). It is known that type II REases and homing endonucleases (HEases) with the GIY-YIG endonuclease domain preferentially use Mg^{2+} divalent cation as a cofactor. The purified Apa233I showed a similar divalent cation

preference as Ahi29725I since it is active in restriction buffers with Mg^{2+} or Mn^{2+} (data not shown).

The purified Ahi29725I enzyme was assayed on Dam^+ and Dam^- λ DNA to test modification dependence (Figure 8). Dam^- λ DNA was also methylated by Dam methylase (M.Dam) or EcoGII frequent adenine methylase (M.EcoGII) in the test tube and used for Ahi29725I digestions. The Ahi29725I endonuclease generated a partial digestion pattern on Dam^+ λ DNA. It showed no cleavage activity on Dam^- λ DNA, indicating restriction activity dependent on Dam modification.

TABLE 1 *In vivo* toxicity of PUA-wH-HNH and wH-HNH endonucleases.

| GenBank accession number (protein) | <i>Halobacteria</i> (archaea) or bacterial strain | Enzyme name | Protein expression level | Toxicity <i>in vivo</i> | Activity <i>in vitro</i> |
|------------------------------------|--|--------------------|--------------------------|-------------------------|--------------------------|
| 1. ARM71120.1 | <i>Haloarcula hispanica</i> pleomorphic virus 4 | HhiV4I (381 aa) | +++ | + | + |
| 2. WP_007980235.1 | <i>Haladaptatus paucihalophilus</i> DX253 | HpaD253I (234 aa) | +++ | + | – |
| 3. WP_009365977.1 | <i>Halogramnum salarium</i> B-1 | Hsa1 (375 aa) | +++ | + | – |
| 4. WP_098725488.1 | <i>Natrinema</i> sp. CBA1119 | NspC1119I (480 aa) | +++ | + | – |
| 5. WP_049951637.1 | <i>Halostagnicola larsenii</i> XH-48 | HlaX48I (378 aa) | +++ | + | – |
| 6. WP_135828390.1 | <i>Halorussus halobios</i> HD8-83 | HhaH883I (224 aa) | +++ | + | – |
| 7. WP_103425724.1 | <i>Salinigranum rubrum</i> GX10 | SruGX1 (232 aa) | +++ | + | ND |
| 8. WP_128477186.1 | <i>Halorussus</i> sp. RC-68 | HspR68I (283 aa) | ++ | + | ND |
| 9. WP_012944509.1 | <i>Haloterrigena turkmenica</i> DSM5511 | HtuIII (496 aa) | ++ | +/- | +/- |
| 10. WP_008893710.1 | <i>Haloterrigena salina</i> JCM13891 | Hsa13891I (496 aa) | +++ | +/- | +/- |
| 11. WP_126662294.1 | <i>Haloterrigena salifodinae</i> ZY19 | HsaZ19I (496 aa) | +++ | +/- | – |
| 12. WP_126597564.1 | <i>Dictyobacter aurantiacus</i> S27 (G ⁺ bacterium) | DauS27I (372 aa) | – | +/- | – |
| 13. WP_009486715.1 | <i>Halobacterium</i> sp. DL1 | HspD11 (230 aa) | – | +/- | – |
| 14. WP_117591244.1 | <i>Haloprofundus halophilus</i> NK23 | HhaN23I (235 aa) | +++ | +/- (small colonies) | ND |
| 15. WP_006089674.1 | <i>Natronorubrum tibetense</i> GA33 | NtiG33I (492 aa) | +++ | – | – |
| 16. WP_159527272.1 | <i>Halobacterium bonnevilliei</i> PCN9 | HboP9I (230 aa) | +++ | – | ND |

Protein expression, *in vivo* toxicity, and *in vitro* activity of PUA-wH-HNH and wH-HNH endonucleases. The *in vivo* toxic effect of the restriction gene was measured by transformation into Dam⁺ and Dam⁻ *E. coli* competent cells under IPTG induction. +, strong restriction (100–1,000-fold reduction in Dam⁺ cells), +/-, mild restriction (~10-fold reduction in Dam⁺ cells or formation of sick small colonies), –, no restriction. Protein expression level: +++, 2–10 mg protein per liter of IPTG-induced cells; ++, 1–1.5 mg/L; –, target protein not detected after chitin column purification or in IPTG-induced cell extract. Proteins with 375–496 aa residues are PUA-wH-HNH fusions; proteins below 300 aa residues are wH-HNH fusions. DauS27I is found in the G⁺ bacterium *Dictyobacter aurantiacus* S27. The other enzymes are found in Archaea (*Halobacteria*). ND, not determined. Dam⁺ pBR322 was used for *in vitro* cleavage assays.

When Dam⁻ λ DNA was methylated *in vitro* by Dam methylase or M.EcoGII, the modified substrates now became cleavable by Ahi29725I (Figure 8). In control digestion, MboI, DpnII, and Sau3AI are able to cleave Dam⁻ λ DNA, but DpnI cannot. Similarly, Ahi29725I and Apa233I endonucleases are also active on Dam⁺ pBR322 and inactive on Dam⁻ pBR322 (Figure 9). However, high enzyme concentrations of Apa233I resulted in non-specific digestion (smearing) of Dam⁻ DNA. The finding was attributed to the non-specific activity on unmodified DNA, since most restriction enzymes display star activity at high enzyme, high glycerol concentration, or low salt.

The Ahi29725I and Apa233I digested pBR322 (Dam⁺) DNA was subjected to run-off sequencing with primers annealing near the

Gm6ATC sites. Cleavages occurred outside the recognition sequence, at a variable distance from the site of methylation (i.e., Ahi29725I cleaves G6mATC at N₁₋₂₃) (Supplementary Figures S9, S10). Ahi29725I and Apa233I have limited, if any, preference for cleavage at NN/RN and NN/GN sites, respectively (Supplementary Figure S11), which would have to be attributed to endonuclease sequence preferences.

To test whether Ahi29725I catalyzed DNA cleavage could be directed by adenine methylation in addition to the G6mATC sequence context, we digested M.EcoGII-modified pBR322 DNA (Dam⁻) to see any enhancement of activity due to frequent adenine methylation. M.EcoGII is capable of methylating all adenines in DNA substrates except in polyA tracks (Murray et al., 2018). Ahi29725I activity was enhanced on M.EcoGII-methylated DNA

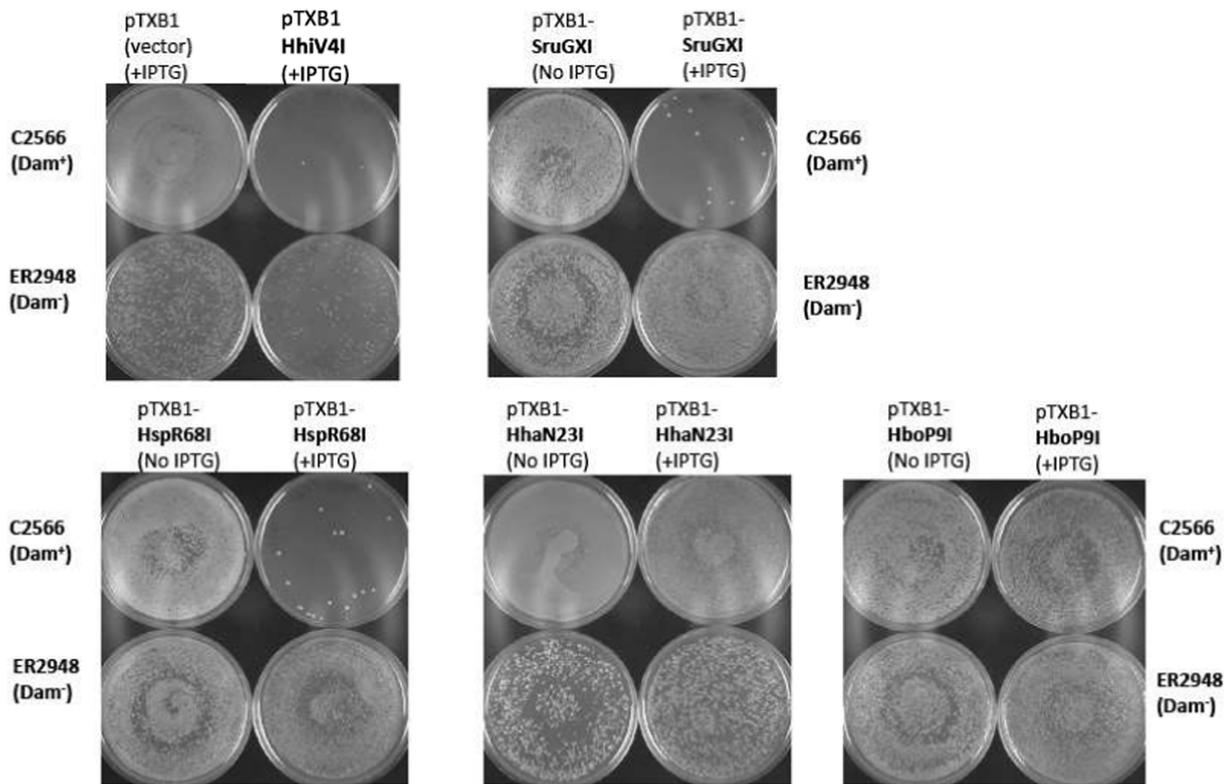


FIGURE 7

In vivo toxicity study: plasmid transfer into C2566 (Dam⁺) and ER2948 (Dam⁻) competent cells by transformation (~50 ng plasmid DNA). Three types of restriction phenotypes were observed: a strong reduction in transformation efficiency due to gene conflict (e.g., HhiV4I, SruGXl, and HspR68I); small colony formation in Dam⁻ hosts presumably due to mild toxicity of the restriction gene (e.g., HhaN23I); and no noticeable change in transformation efficiency (e.g., HboP9I) compared to the empty vector control. The assay was done semi-quantitatively based on visualization of the transformation plates and not quantitatively since the number of colonies was not counted. Toxicity was more apparent with IPTG induction (0.5 mM IPTG in Amp plates). The overall transformation efficiency is lower in the Dam-deficient host.

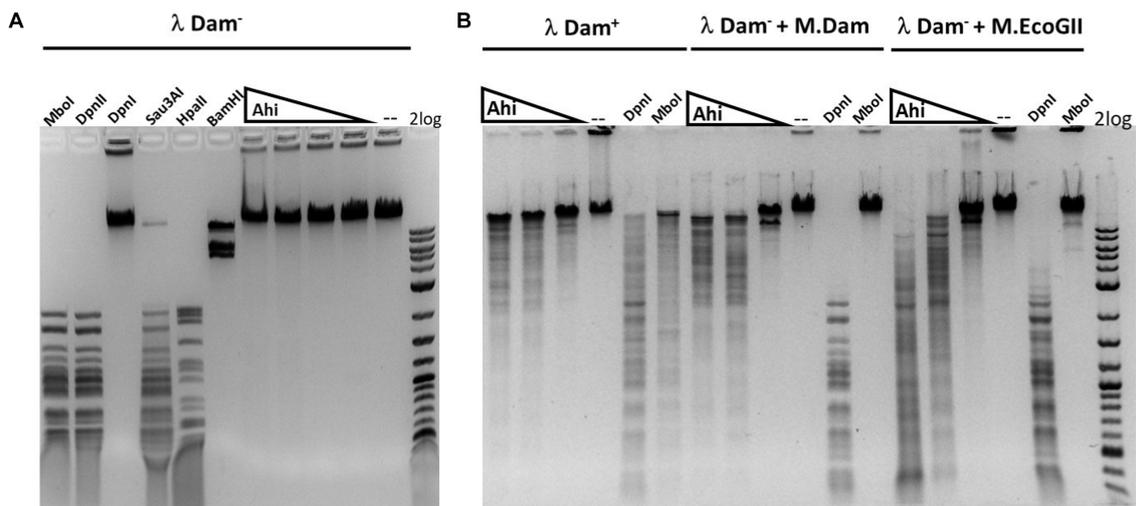


FIGURE 8

Restriction activity assay on Dam⁺, Dam⁻ λ DNA, Dam⁻ λ DNA further modified by Dam methylase (M.Dam) or M.EcoGII. (A) Restriction digestion of Dam⁻ λ DNA with Ahi29725I (Ahi) 10-fold serial dilution: 812.4, 81.2, and 8.1 nM (2, 0.2, and 0.02 μ g) of protein incubated with 0.5 μ g λ DNA in NEB B2.1 in 50 μ L reaction volume at 37°C for 1 h. (B) 10-fold serial dilution of Ahi29725I (Ahi) in digestion of Dam⁺ λ DNA, Dam⁻ λ DNA methylated by Dam methylase or by M.EcoGII. λ DNA was only partially modified by *E. coli* host Dam methylase during rapid phage DNA replication. In control digests, MboI and DpnII cleave unmodified GATC sites only; DpnI cleaves Gm6ATC sites only; Sau3AI cleaves GATC sites regardless of m6A methylation. HpaII (CCGG) and BamHI-HF (GGATCC, not affected by Dam methylation) are additional controls.

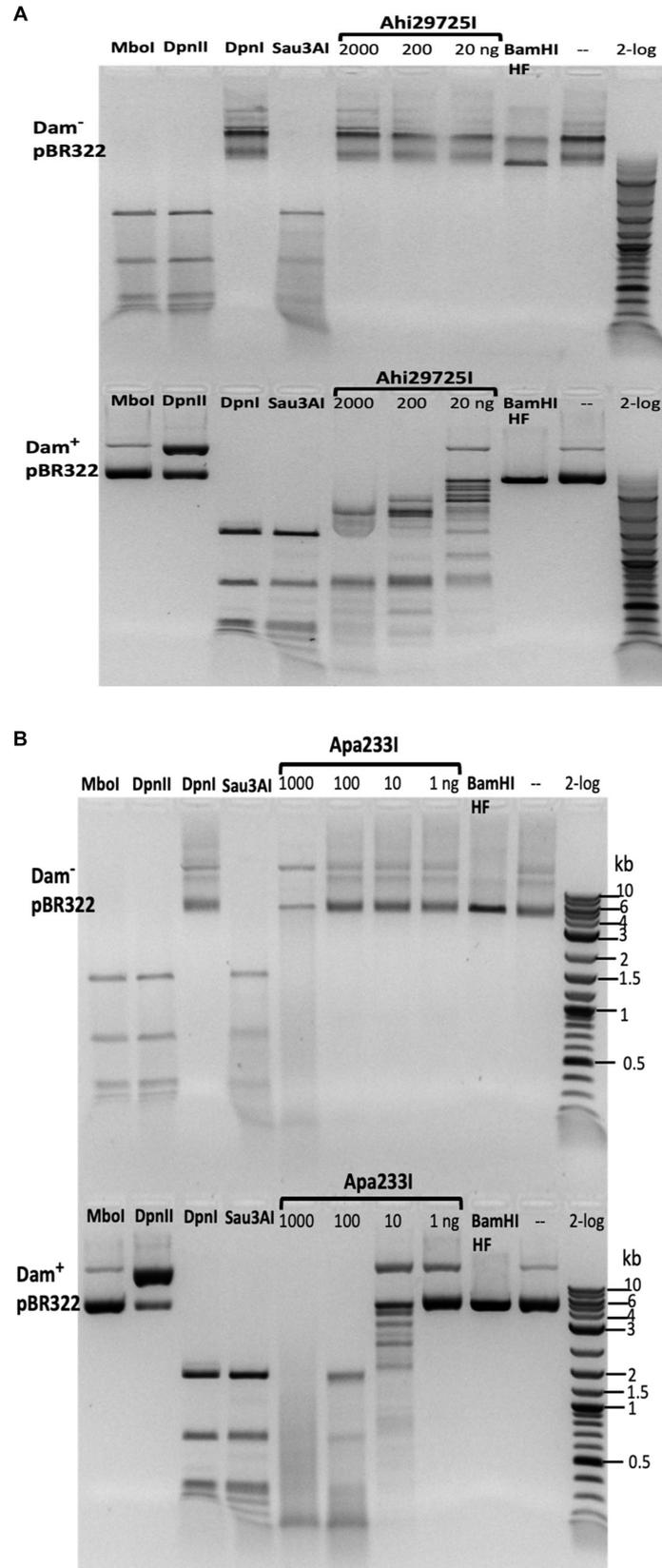


FIGURE 9

In vitro activity of Ahi29725I (A) and Apa233I (B) on Dam⁻ (top) and Dam⁺ (bottom) DNA. Ahi29725I (~812.4, 81.2, and 8.1 nM, respectively) or Apa233I (~413.7, 41.4, 4.1, and 0.4 nM, respectively) digestion of plasmid DNA (0.5 μg) was done in NEB B.2.1 at 37°C for 1 h in a 50 μL reaction volume. In control digests, MboI and DpnII cleave unmodified GATC sites only; DpnI is unable to cut Dam⁻ DNA. Sau3AI cleaves GATC sites regardless of 6m A methylation. BamHI-HF (GGATCC) was used as an additional control. Apa233I showed non-specific endonuclease activity at high enzyme concentrations (~1 μg/0.4 μM). A smeared pattern was detected in both Dam⁺ and Dam⁻ pBR322.

substrate (Supplementary Figure S12). Three large fragments of Dam⁺ DNA were further digested into smaller fragments after M.EcoGII methylation. However, it was not clear whether the enhanced activity was due to 6mA-dependent relaxed sequence recognition (e.g., cleavage near the Cm6ATC star site or Sm6ATS sites, S = G and C). The enhanced activity on M.EcoGII-modified DNA remains to be characterized in future by using defined modified oligos or restriction digestion/NGS sequencing mapping of M.EcoGII-modified λ DNA.

In digestion of methylated duplex oligos with a single G6mATC site, it was noted that Ahi29725I preferentially cleaved fully methylated oligos (M+/M+) over hemi-modified substrates (M+/M- or M-/M+). However, Apa233I endonuclease was able to cut both fully modified and hemi-modified oligos (Figure 4). This discrepancy in methylation dependence between the two enzymes is unexplained.

If the *in vitro* divalent cation requirement of Ahi29725I was relevant in cells, invading 6mA-modified DNA would be digested by Mg²⁺-bound Ahi29725I. By contrast, activation of the non-specific endonuclease activity of Ahi29725I with Mn²⁺, Co²⁺, or Ni²⁺ in the active site could lead to cleavage of both cellular and invading DNA regardless of modifications, triggering cell death and preventing phage release.

PLD—wH endonucleases

We identified 27 predicted PLD—wH fusion endonucleases in bacterial genomes. Two putative restriction genes from *Anaerolineaceae bacterium* (Aba4572I) and *Chloroflexi bacterium* (CbaI) were cloned into the pTXB1 expression vector. However, upon IPTG induction, no over-expressed proteins were detected. In the gene neighborhood analysis, the Aba4572I ORF resides in a genomic region of (1) DNA MTase (predicted specificity CGATCG, amino-MTase), (2) PLD—wH endonuclease, and (3) and (4) hypothetical proteins. If the CGATCG site is methylated to become CGm6ATCG, it would be a substrate for the PLD—wH endonuclease, which could potentially result in self-restriction. The CbaI enzyme is located in a region with (1) Leu-tRNA ligase, (2) restriction endonuclease, (3) PLD—wH endonuclease, (4) hypothetical protein, and (5) dimethyl-menaquinone MTase. Since the transformation of Aba4572I and CbaI was less toxic in Dam⁻ cells in a non-induced condition (see Figure 3), the lack of expression in Dam⁻ cells is surprising and requires further investigation. Expression of two more PLD—wH fusion proteins containing the conserved catalytic residues HxDx(4)K and HxEx(4)K in the PLD endonuclease domain in *E. coli* Dam⁻ cells was not successful due to toxicity. More work is necessary to explain the reasons for the poor expression of PLD—wH fusion endonucleases in *E. coli*.

NTD—wH—NTPase fusions

The N-terminal domain-wH-NTPase fusions are usually paired with another catalytic subunit, such as an McrC (Ross et al., 1989) McrC-like protein with a PD-(D/E)XK endonuclease motif (Pieper and Pingoud, 2002). This arrangement is reminiscent of McrBC, a type IV restriction system acting on modified cytosines (Stewart et al.,

2000). We only made one unsuccessful attempt to purify a putative heterodimeric NTD—wH—NTPase/McrC complex. Therefore, we have not studied the possible activity of these enzymes toward 6mA-containing DNA or their more general activity in the restriction of modified DNA. More work is needed to characterize this group of putative type IV restriction systems with wH-NTPase fusion.

Discussion

wH domain as a sensor of fully methylated ApT in a dsDNA context

The wH domain was first associated with adenine methylation because of its presence in the C-terminal region of *E. coli* and phage T4 Dam methyltransferases (Teichmann et al., 2012), and later because of its presence in the adenine methylation-dependent DpnI restriction endonuclease. A subsequent study on DpnI showed that the wH domain binds dsDNA at fully methylated ApT sites without base flipping. The two methyl groups, which are in close proximity, are bound in a single pocket of the wH domain of DpnI (Mierzejewska et al., 2014). The study on DpnI also showed that the specificity of the domain for the flanking sequence was somewhat relaxed with respect to the Dam Gm6ATC consensus and the Sm6ATS sites (where S stands for G or C) (Siwiek et al., 2012). In this study, we show that the properties of an adenine methylation reader carry over to many fusions with HNH, GIY-YIG, and likely also PLD endonuclease domains. If methylation is seen in the ApT context, there is a preference for fully methylated sites except for Apa233I (see Figures 3, 4). Our study shows that in all tested fusion proteins, the wH domain can operate as an adenine methylation reader for the Gm6ATC context. A study on the wH—GIY-YIG endonucleases further indicates that additional cleavage sites are likely created when DNA is hypermethylated by M.EcoGII (see Supplementary Figure S12). Hence, the wH domains of the wH—GIY-YIG endonucleases also suggest that the Gm6ATC preference may be relaxed, but the star binding sites remain to be characterized (star sites are usually defined as DNA sequences with one base off from the cognate site; if there are two bases off, these sequences are usually called non-cognate sites; Pingoud et al., 2016).

The identification of prokaryotic winged helix domains as sensors of adenine methylation contrasts with the role of some eukaryotic winged helix domains as sensors of non-methylated CpG (Stielow et al., 2021; Becht et al., 2023; Weber et al., 2023). The superposition of the winged helix domains of prokaryotic DpnI (Mierzejewska et al., 2014) and eukaryotic KAT6A (also called histone acetyltransferase KAT6A, lysine acetyltransferase 6A, zinc finger protein 220, and MYST-3) (Weber et al., 2023) shows that the dsDNA molecules are bound to opposite faces of the wH domain (Figure 10), indicating that the two DNA-binding modes have likely evolved independently for needs that are characteristic for prokaryotes (sensing of Dam methylation) and eukaryotes (sensing of the absence of CpG methylation).

Cooperation with endonuclease domains

For most NTP-independent MDREs, there is a clear division of labor between the modification reader and endonuclease domains. The

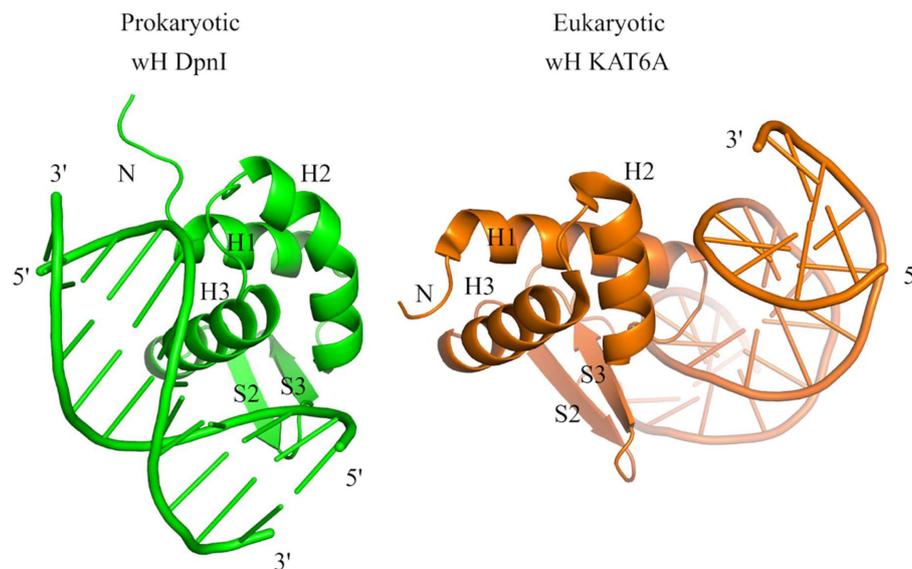


FIGURE 10

Structural comparison of the wH domains of DpnI (G6mATC) with bacterial origin and mammalian KAT6A protein that recognizes unmodified CpG sites. The docking of wH domains on dsDNA shows a major difference in recognition. The wH domain-containing protein KAT6A is a histone lysine acetyltransferase that acetylates lysine residues in histones H3 and H4 (*in vitro*) (Weber et al., 2023).

former recruits the enzyme to modification sites, and the latter cleaves the DNA at a distance from the recognition site, which is likely defined by the length of the linker that connects the two domains. The nuclease domain has generally low or only a very relaxed sequence specificity, and it is likely not modification-specific. How the modification sensor domain keeps the activity of the nuclease domain in check is not well understood. In some cases, it can be shown that the linker has an inhibitory role for the endonuclease that is only relieved once modified DNA is bound to the reader domain and the complex reorganizes structurally (Pastor et al., 2021). The PUA—wH—HNH (HhiV4I) and wH—GIY-YIG (Ahi29725I and Apa233I) that were tested by run-off sequencing are consistent with this expectation. As for the cytosine modification-specific MDREs, cleavage occurred mostly at a distance from the recognition sequence except for the *BisI* family REases (e.g., *Eco15I* and *NhoI*) that cut within the recognition sequence GCNGC with 2–4 modified cytosines (Xu et al., 2016).

Among the wH fusion endonucleases, DpnI and its isochizomers are the exception to the rule that cleavage occurs always outside of and not within the recognition sequence. Mechanically, DpnI DNA cleavage within the recognition sequence is a consequence of the fact that the endonuclease domain has separate sequence and modification specificities (Siwek et al., 2012). In this scenario, the role of the wH domain is similar to the role of the extra specificity domain in type IIE restriction endonucleases (Roberts et al., 2003), except that the target sequence contains a modified base. Type IIE restriction endonucleases are one of the subfamilies that require a pair of target sequences in order to show activity (Senesac and Allen, 1995; Colandene and Topal, 1998). Run-off sequencing shows that *FcyTI* and *Psp4BI* cleave inside the recognition sequence, like DpnI, pointing to the separate sequence and modification specificity of the catalytic PD-(D/E)XK domain. Given the high sequence conservation of the PD-(D/E)XK—wH

endonuclease family, it is likely that this property is general for the entire family.

Sensors/readers for N6-methyladenine in DNA

Despite the many roles of DNA adenine methylation in prokaryotes (and eukaryotic organelles), the repertoire of reader domains for m6A in DNA is still surprisingly limited (Iyer et al., 2016). Perhaps the best-known adenine methylation sensors are the YTH (Liao et al., 2018) domains, which belong to (or are related to) the PUA superfamily domains. The PUA superfamily domains are believed to flip the modified 2'-deoxynucleotide out of duplex DNA (Shao et al., 2014) or to bind a single nucleotide in RNA in the reader pocket (Li et al., 2014; Xu et al., 2014). Therefore, at least when acting in isolation, they can be considered as sensors of a single modified adenine. Consistent with this role, most YTH domains sense adenine methylation in RNA rather than DNA. However, some YTH and ASCH domains in prokaryotes are considered as DNA adenine methylation sensors (Iyer et al., 2016). For the ASCH domains, this remains to be experimentally shown, since currently only a 4mC reader role is experimentally supported (Stanislauskiene et al., 2020). Apart from the YTH and ASCH domains, the HARE-HTH and RAMA (Restriction enzyme and Adenine Methylation Associated) domains have also been suggested to serve as readers of adenine methylation in DNA (Teichmann et al., 2012). The HARE-HTH domains are related to winged helix domains, which would be consistent with a role as adenine methylation sensors. However, they have an extra helix inserted into the HTH motif of the winged helix domain, and recent analysis suggests that they are more likely to sense cytosine modifications (Aravind and Iyer, 2012). Unlike

the HARE-HTH domains, the RAMA domains are unrelated to the wH domain in fold (Yang et al., 2023). For the RAMA domain-containing MPND protein, there is some biochemical evidence for adenine methylation sensing (Kweon et al., 2019). However, a preference for adenine-methylated DNA could not be experimentally confirmed (Yang et al., 2023). We noticed the occurrence of RAMA—Mrr catalytic domain (PD-QXK)—NTPase (three-domain fusion) and GIY-YIG—RAMA (two-domain fusion) in prokaryotes, which might indicate that the RAMA domain is utilized similarly to the wH domain in these fusions. Future studies will be focused on the characterization of YTH-NTPase, YTH-HNH, and RAMA fusion endonucleases as 6mA readers/sensors in type IV restriction systems.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

IH: Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – review & editing. DH: Data curation, Formal analysis, Investigation, Methodology, Writing – review & editing. WY: Data curation, Formal analysis, Investigation, Methodology, Software, Writing – review & editing. TV: Data curation, Investigation, Writing – review & editing. AH: Data curation, Investigation, Writing – review & editing. TL: Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – review & editing. LE: Data curation, Formal analysis, Investigation, Methodology, Software, Writing – review & editing. MB: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing. S-yX: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Part of this

References

- Aravind, L., and Iyer, L. M. (2012). The HARE-HTH and associated domains: novel modules in the coordination of epigenetic DNA and protein modifications. *Cell Cycle* 11, 119–131. doi: 10.4161/cc.11.1.18475
- Au, K. G., Welsh, K., and Modrich, P. (1992). Initiation of methyl-directed mismatch repair. *J. Biol. Chem.* 267, 12142–12148. doi: 10.1016/S0021-9258(19)49816-5
- Becht, D. C., Klein, B. J., Kanai, A., Jang, S. M., Cox, K. L., Zhou, B. R., et al. (2023). MORF and MOZ acetyltransferases target unmethylated CpG islands through the winged helix domain. *Nat. Commun.* 14:697. doi: 10.1038/s41467-023-36368-5
- Bertonati, C., Punta, M., Fischer, M., Yachdav, G., Forouhar, F., Zhou, W., et al. (2009). Structural genomics reveals EVE as a new ASCH/PUA-related domain. *Proteins* 75, 760–773. doi: 10.1002/prot.22287
- Borgaro, J. G., and Zhu, Z. (2013). Characterization of the 5-hydroxymethylcytosine-specific DNA restriction endonucleases. *Nucleic Acids Res.* 41, 4198–4206. doi: 10.1093/nar/gkt102

study was supported by the Polish NAWA (PPI/APM/2018/1/00034/U/001) and NCN grants (2018/30/Q/NZ2/00669). This study was also supported by New England Biolabs, Inc. (NEB). This publication cost was paid for by NEB.

Acknowledgments

We thank Andy Gardner, Tom Evans, Rich Roberts, Jim Ellard, and Salvatore Russello for their continued support. We are grateful to Lise Raleigh for providing the Dam-deficient T7 expression strain ER2948. We thank Honorata Czapinska and Lise Raleigh for their critical reading of the manuscript.

Conflict of interest

WY, DH, TV, AH, LE, and S-yX (retired) are employees of New England Biolabs, Inc., a company providing molecular biology reagents to the research and diagnostic community. IH and TL were short-term visiting scientists at New England Biolabs, Inc.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of *Frontiers*, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2024.1286822/full#supplementary-material>

- Boye, E., and Lobner-Olesen, A. (1990). The role of dam methyltransferase in the control of DNA replication in *E. coli*. *Cell* 62, 981–989. doi: 10.1016/0092-8674(90)90272-G

- Brennan, R. G. (1993). The winged-helix DNA-binding motif: another helix-turn-helix takeoff. *Cell* 74, 773–776. doi: 10.1016/0092-8674(93)90456-Z

- Bujnicki, J. M., and Rychlewski, L. (2001). Grouping together highly diverged PD-(D/E)XK nucleases and identification of novel superfamily members using structure-guided alignment of sequence profiles. *J. Mol. Microbiol. Biotechnol.* 3, 69–72. doi: 10.1016/s0378-1119(01)00405-x

- Cerrudo, C. S., Ghiringhelli, P. D., and Gomez, D. E. (2014). Protein universe containing a PUA RNA-binding domain. *FEBS J.* 281, 74–87. doi: 10.1111/febs.12602

- Chan, S. H., Bao, Y., Ciszak, E., Laget, S., and Xu, S. Y. (2007). Catalytic domain of restriction endonuclease BmrI as a cleavage module for engineering endonucleases

- with novel substrate specificities. *Nucleic Acids Res.* 35, 6238–6248. doi: 10.1093/nar/gkm665
- Cohen-Karni, D., Xu, D., Apone, L., Fomenkov, A., Sun, Z., Davis, P. J., et al. (2011). The MspJI family of modification-dependent restriction endonucleases for epigenetic studies. *Proc. Natl. Acad. Sci. USA* 108, 11040–11045. doi: 10.1073/pnas.1018448108
- Colandene, J. D., and Topal, M. D. (1998). The domain organization of NaeI endonuclease: separation of binding and catalysis. *Proc. Natl. Acad. Sci. USA* 95, 3531–3536. doi: 10.1073/pnas.95.7.3531
- Czapinska, H., Kowalska, M., Zagorskaite, E., Manakova, E., Slyvka, A., Xu, S. Y., et al. (2018). Activity and structure of EcoKMcrA. *Nucleic Acids Res.* 46, 9829–9841. doi: 10.1093/nar/gky731
- Dunin-Horkawicz, S., Feder, M., and Bujnicki, J. M. (2006). Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics* 7:98. doi: 10.1186/1471-2164-7-98
- Flick, K. E., Jurica, M. S., Monnat, R. J. Jr., and Stoddard, B. L. (1998). DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature* 394, 96–101. doi: 10.1038/27952
- Fomenkov, A., Xiao, J. P., Dila, D., Raleigh, E., and Xu, S. Y. (1994). The 'endo-blue method' for direct cloning of restriction endonuclease genes in *E. coli*. *Nucleic Acids Res.* 22, 2399–2403. doi: 10.1093/nar/22.12.2399
- Frickey, T., and Lupas, A. (2004). CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* 20, 3702–3704. doi: 10.1093/bioinformatics/bth444
- Gajiwala, K. S., and Burley, S. K. (2000). Winged helix proteins. *Curr. Opin. Struct. Biol.* 10, 110–116. doi: 10.1016/S0959-440X(99)00057-3
- Gajiwala, K. S., Chen, H., Cornille, F., Roques, B. P., Reith, W., Mach, B., et al. (2000). Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* 403, 916–921. doi: 10.1038/35002634
- Grazulis, S., Manakova, E., Roessle, M., Bochtler, M., Tamulaitiene, G., Huber, R., et al. (2005). Structure of the metal-independent restriction enzyme BfiI reveals fusion of a specific DNA-binding domain with a nonspecific nuclease. *Proc. Natl. Acad. Sci. U. S. A.* 102, 15797–15802. doi: 10.1073/pnas.0507949102
- Heitman, J., and Model, P. (1990). Substrate recognition by the EcoRI endonuclease. *Proteins* 7, 185–197. doi: 10.1002/prot.340070207
- Horton, J. R., Wang, H., Mabuchi, M. Y., Zhang, X., Roberts, R. J., Zheng, Y., et al. (2014). Modification-dependent restriction endonuclease, MspJI, flips 5-methylcytosine out of the DNA helix. *Nucleic Acids Res.* 42, 12092–12101. doi: 10.1093/nar/gku871
- Hosford, C. J., Bui, A. Q., and Chappie, J. S. (2020). The structure of the *Thermococcus gammatolerans* McrB N-terminal domain reveals a new mode of substrate recognition and specificity among McrB homologs. *J. Biol. Chem.* 295, 743–756. doi: 10.1016/S0021-9258(17)49932-7
- Iyer, L. M., Burroughs, A. M., and Aravind, L. (2006). The ASCH superfamily: novel domains with a fold related to the PUA domain and a potential role in RNA metabolism. *Bioinformatics* 22, 257–263. doi: 10.1093/bioinformatics/bti767
- Iyer, L. M., Zhang, D., and Aravind, L. (2016). Adenine methylation in eukaryotes: apprehending the complex evolutionary history and functional potential of an epigenetic modification. *Bioessays* 38, 27–40. doi: 10.1002/bies.201500104
- Jablonska, J., Matelska, D., Steczkiewicz, K., and Ginalski, K. (2017). Systematic classification of the his-me finger superfamily. *Nucleic Acids Res.* 45, 11479–11494. doi: 10.1093/nar/gkx924
- Janosi, L., Yonemitsu, H., Hong, H., and Kaji, A. (1994). Molecular cloning and expression of a novel hydroxymethylcytosine-specific restriction enzyme (PvuRts1I) modulated by glucosylation of DNA. *J. Mol. Biol.* 242, 45–61. doi: 10.1006/jmbi.1994.1556
- Josephs, E. A., Zheng, T., and Marszalek, P. E. (2015). Atomic force microscopy captures the initiation of methyl-directed DNA mismatch repair. *DNA Repair* 35, 71–84. doi: 10.1016/j.dnarep.2015.08.006
- Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637. doi: 10.1002/bip.360221211
- Kaminska, K. H., Kawai, M., Boniecki, M., Kobayashi, I., and Bujnicki, J. M. (2008). Type II restriction endonuclease R.Hpy188I belongs to the GIY-YIG nuclease superfamily, but exhibits an unusual active site. *BMC Struct. Biol.* 8:48. doi: 10.1186/1472-6807-8-48
- Kazrani, A. A., Kowalska, M., Czapinska, H., and Bochtler, M. (2014). Crystal structure of the 5hmC specific endonuclease PvuRts1I. *Nucleic Acids Res.* 42, 5929–5936. doi: 10.1093/nar/gku186
- Kosinski, J., Feder, M., and Bujnicki, J. M. (2005). The PD-(D/E)XK superfamily revisited: identification of new members among proteins involved in DNA metabolism and functional predictions for domains of (hitherto) unknown function. *BMC Bioinformatics* 6:172. doi: 10.1186/1471-2105-6-172
- Kweon, S. M., Chen, Y., Moon, E., Kvederaviciute, K., Klimasauskas, S., and Feldman, D. E. (2019). An adversarial DNA N(6)-Methyladenine-sensor network preserves Polycomb silencing. *Mol. Cell* 74, 1138–1147.e6. doi: 10.1016/j.molcel.2019.03.018
- Lai, E., Clark, K. L., Burley, S. K., and Darnell, J. E. Jr. (1993). Hepatocyte nuclear factor 3/fork head or "winged helix" proteins: a family of transcription factors of diverse biologic function. *Proc. Natl. Acad. Sci. U. S. A.* 90, 10421–10423. doi: 10.1073/pnas.90.22.10421
- Li, F., Zhao, D., Wu, J., and Shi, Y. (2014). Structure of the YTH domain of human YTHDF2 in complex with an m(6)a mononucleotide reveals an aromatic cage for m(6)a recognition. *Cell Res.* 24, 1490–1492. doi: 10.1038/cr.2014.153
- Liao, S., Sun, H., Xu, C., and Domain, Y. T. H. (2018). A family of N(6)-methyladenosine (m(6)a) readers. *Genomics Proteomics Bioinformatics* 16, 99–107. doi: 10.1016/j.gpb.2018.04.002
- Liu, G., Ou, H. Y., Wang, T., Li, L., Tan, H., Zhou, X., et al. (2010). Cleavage of phosphorothioated DNA and methylated DNA by the type IV restriction endonuclease ScoMcrA. *PLoS Genet.* 6:e1001253. doi: 10.1371/journal.pgen.1001253
- Lu, X., Huang, F., Cheng, R., and Zhu, B. (2023). A unique m6A-dependent restriction endonuclease from an archaeal virus. *Microbiol. Spectr.* 11:e0426222. doi: 10.1128/spectrum.04262-22
- Lutz, T., Flodman, K., Copelas, A., Czapinska, H., Mabuchi, M., Fomenkov, A., et al. (2019). A protein architecture guided screen for modification dependent restriction endonucleases. *Nucleic Acids Res.* 47, 9761–9776. doi: 10.1093/nar/gkz755
- Makarova, K. S., Wolf, Y. I., Snir, S., and Koonin, E. V. (2011). Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J. Bacteriol.* 193, 6039–6056. doi: 10.1128/JB.05535-11
- Marinus, M. G., and Casadesu, J. (2009). Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more. *FEMS Microbiol. Rev.* 33, 488–503. doi: 10.1111/j.1574-6976.2008.00159.x
- Mehta, P., Katta, K., and Krishnaswamy, S. (2004). HNH family subclassification leads to identification of commonality in the his-me endonuclease superfamily. *Protein Sci.* 13, 295–300. doi: 10.1110/ps.03115604
- Mierzejewska, K., Siwek, W., Czapinska, H., Kaus-Drobek, M., Radlinska, M., Skowronek, K., et al. (2014). Structural basis of the methylation specificity of R.DpnI. *Nucleic Acids Res.* 42, 8745–8754. doi: 10.1093/nar/gku546
- Miyazono, K., Furuta, Y., Watanabe-Matsui, M., Miyakawa, T., Ito, T., Kobayashi, I., et al. (2014). A sequence-specific DNA glycosylase mediates restriction-modification in *Pyrococcus abyssi*. *Nat. Commun.* 5:3178. doi: 10.1038/ncomms4178
- Murray, I. A., Morgan, R. D., Luyten, Y., Fomenkov, A., Correa, I. R. Jr., Dai, N., et al. (2018). The non-specific adenine DNA methyltransferase M.EcoGII. *Nucleic Acids Res.* 46, 840–848. doi: 10.1093/nar/gkx1191
- Orlowski, J., and Bujnicki, J. M. (2008). Structural and evolutionary classification of type II restriction enzymes based on theoretical and experimental analyses. *Nucleic Acids Res.* 36, 3552–3569. doi: 10.1093/nar/gkn175
- Pastor, M., Czapinska, H., Helbrecht, I., Krakowska, K., Lutz, T., Xu, S. Y., et al. (2021). Crystal structures of the EVE-HNH endonuclease VcaM4I in the presence and absence of DNA. *Nucleic Acids Res.* 49, 1708–1723. doi: 10.1093/nar/gkaa1218
- Pieper, U., and Pingoud, A. (2002). A mutational analysis of the PD...D/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. doi: 10.1021/bi0156862
- Pingoud, A., Fuxreiter, M., Pingoud, V., and Wende, W. (2005). Type II restriction endonucleases: structure and mechanism. *Cell. Mol. Life Sci.* 62, 685–707. doi: 10.1007/s00018-004-4513-1
- Pingoud, A., and Jeltsch, A. (2001). Structure and function of type II restriction endonucleases. *Nucleic Acids Res.* 29, 3705–3727. doi: 10.1093/nar/29.18.3705
- Pingoud, A., Wilson, G. G., and Wende, W. (2016). Type II restriction endonucleases - a historical perspective and more. *Nucleic Acids Res.* 44:8011. doi: 10.1093/nar/gkw513
- Pommer, A. J., Cal, S., Keeble, A. H., Walker, D., Evans, S. J., Kuhlmann, U. C., et al. (2001). Mechanism and cleavage specificity of the H-N-H endonuclease colicin E9. *J. Mol. Biol.* 314, 735–749. doi: 10.1006/jmbi.2001.5189
- Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J., et al. (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.* 31, 1805–1812. doi: 10.1093/nar/gkg274
- Roberts, R. J., and Cheng, X. (1998). Base flipping. *Annu. Rev. Biochem.* 67, 181–198. doi: 10.1146/annurev.biochem.67.1.181
- Ross, T. K., Achberger, E. C., and Braymer, H. D. (1989). Identification of a second polypeptide required for McrB restriction of 5-methylcytosine-containing DNA in *Escherichia coli* K12. *Mol. Gen. Genet.* 216, 402–407. doi: 10.1007/BF00334382
- Sasnauskas, G., Zakrys, L., Zaremba, M., Cosstick, R., Gaynor, J. W., Halford, S. E., et al. (2010). A novel mechanism for the scission of double-stranded DNA: BfiI cuts both 3'-5' and 5'-3' strands by rotating a single active site. *Nucleic Acids Res.* 38, 2399–2410. doi: 10.1093/nar/gkp1194
- Schwartz, T., Behlke, J., Lowenhaupt, K., Heinemann, U., and Rich, A. (2001). Structure of the DLM-1-Z-DNA complex reveals a conserved family of Z-DNA-binding proteins. *Nat. Struct. Biol.* 8, 761–765. doi: 10.1038/nsb0901-761
- Senesac, J. H., and Allen, J. R. (1995). Oligonucleotide activation of the type IIe restriction enzyme NaeI for digestion of refractory sites. *BioTechniques* 19, 990–993.

- Shao, C., Wang, C., and Zang, J. (2014). Structural basis for the substrate selectivity of PvuRtsII, a 5-hydroxymethylcytosine DNA restriction endonuclease. *Acta Crystallogr. D Biol. Crystallogr.* 70, 2477–2486. doi: 10.1107/S139900471401606X
- Siwek, W., Czapinska, H., Bochtler, M., Bujnicki, J. M., and Skowronek, K. (2012). Crystal structure and mechanism of action of the N6-methyladenine-dependent type IIM restriction endonuclease R.DpnI. *Nucleic Acids Res.* 40, 7563–7572. doi: 10.1093/nar/gks428
- Slyvka, A., Zagorskaitė, E., Czapinska, H., Sasnauskas, G., and Bochtler, M. (2019). Crystal structure of the EcoKMcrA N-terminal domain (NEco): recognition of modified cytosine bases without flipping. *Nucleic Acids Res.* 47, 11943–11955. doi: 10.1093/nar/gkz1017
- Sokolowska, M., Czapinska, H., and Bochtler, M. (2009). Crystal structure of the beta beta alpha-me type II restriction endonuclease Hpy99I with target DNA. *Nucleic Acids Res.* 37, 3799–3810. doi: 10.1093/nar/gkp228
- Sokolowska, M., Czapinska, H., and Bochtler, M. (2011). Hpy188I-DNA pre- and post-cleavage complexes—snapshots of the GIY-YIG nuclease mediated catalysis. *Nucleic Acids Res.* 39, 1554–1564. doi: 10.1093/nar/gkq821
- Stanislauskienė, R., Lauryėnas, A., Rutkienė, R., Aucynaite, A., Tauraitė, D., Meskienė, R., et al. (2020). YqfB protein from *Escherichia coli*: an atypical amidohydrolase active towards N(4)-acylcytosine derivatives. *Sci. Rep.* 10:788. doi: 10.1038/s41598-020-57664-w
- Stewart, F. J., Panne, D., Bickle, T. A., and Raleigh, E. A. (2000). Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. *J. Mol. Biol.* 298, 611–622. doi: 10.1006/jmbi.2000.3697
- Stielow, B., Zhou, Y., Cao, Y., Simon, C., Pogoda, H. M., Jiang, J., et al. (2021). The SAM domain-containing protein 1 (SAMD1) acts as a repressive chromatin regulator at unmethylated CpG islands. *Sci. Adv.* 7:eabf2229. doi: 10.1126/sciadv.abf2229
- Sukackaite, R., Grazulis, S., Tamulaitis, G., and Siksnys, V. (2012). The recognition domain of the methyl-specific endonuclease McrBC flips out 5-methylcytosine. *Nucleic Acids Res.* 40, 7552–7562. doi: 10.1093/nar/gks332
- Tang, Q., Rigby, R. E., Young, G. R., Hvidt, A. K., Davis, T., Tan, T. K., et al. (2021). Adenosine-to-inosine editing of endogenous Z-form RNA by the deaminase ADAR1 prevents spontaneous MAVS-dependent type I interferon responses. *Immunity* 54, 1961–1975.e5. doi: 10.1016/j.immuni.2021.08.011
- Teichmann, M., Dumay-Odelot, H., and Fribourg, S. (2012). Structural and functional aspects of winged-helix domains at the core of transcription initiation complexes. *Transcription* 3, 2–7. doi: 10.4161/trns.3.1.18917
- Wah, D. A., Hirsch, J. A., Dorner, L. F., Schildkraut, I., and Aggarwal, A. K. (1997). Structure of the multimodular endonuclease FokI bound to DNA. *Nature* 388, 97–100. doi: 10.1038/40446
- Weber, L. M., Jia, Y., Stielow, B., Gisselbrecht, S. S., Cao, Y., Ren, Y., et al. (2023). The histone acetyltransferase KAT6A is recruited to unmethylated CpG islands via a DNA binding winged helix domain. *Nucleic Acids Res.* 51, 574–594. doi: 10.1093/nar/gkac1188
- Wolberger, C., and Campbell, R. (2000). New perch for the winged helix. *Nat. Struct. Biol.* 7, 261–262. doi: 10.1038/74004
- Wu, C. C., Lin, J. L. J., and Yuan, H. S. (2020). Structures, mechanisms, and functions of his-me finger nucleases. *Trends Biochem. Sci.* 45, 935–946. doi: 10.1016/j.tibs.2020.07.002
- Xu, S. Y., Klein, P., Degtyarev, S., and Roberts, R. J. (2016). Expression and purification of the modification-dependent restriction enzyme BslI and its homologous enzymes. *Sci. Rep.* 6:28579. doi: 10.1038/srep28579
- Xu, D., Shao, J., Song, H., and Wang, J. (2020). The YTH Domain family of N6-Methyladenosine “readers” in the diagnosis and prognosis of colonic adenocarcinoma. *Biomed. Res. Int.* 2020:9502560. doi: 10.1155/2020/9502560
- Xu, C., Wang, X., Liu, K., Roundtree, I. A., Tempel, W., Li, Y., et al. (2014). Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nat. Chem. Biol.* 10, 927–929. doi: 10.1038/nchembio.1654
- Yang, M., Li, X., Tian, Z., Ma, L., Ma, J., Liu, Y., et al. Structures of MPND reveal the molecular recognition of nucleosomes. *Int. J. Mol. Sci.* 24:3368. doi: 10.3390/ijms24043368