The crystal structure of the *Helicobacter pylori* LlaJI.R1 N-terminal domain provides a model for site-specific DNA binding

Received for publication, January 16, 2018, and in revised form, June 4, 2018 Published, Papers in Press, June 12, 2018, DOI 10.1074/jbc.RA118.001888

Christopher J. Hosford and Joshua S. Chappie

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

Edited by Patrick Sung

Restriction modification systems consist of an endonuclease that cleaves foreign DNA site-specifically and an associated methyltransferase that protects the corresponding target site in the host genome. Modification-dependent restriction systems, in contrast, specifically recognize and cleave methylated and/or glucosylated DNA. The LlaJI restriction system contains two 5-methylcytosine (5mC) methyltransferases (LlaJI.M1 and LlaJI.M2) and two restriction proteins (LlaJI.R1 and LlaJI.R2). LlaJI.R1 and LlaJI.R2 are homologs of McrB and McrC, respectively, which in *Escherichia coli* function together as a modification-dependent restriction complex specific for 5mC-containing DNA. *Lactococcus lactis* LlaJI.R1 binds DNA site-specifically, suggesting that the LlaJI system uses a different mode of substrate recognition. Here we present the structure of the N-terminal DNA-binding domain of *Helicobacter pylori* LlaJI.R1 at 1.97 Å resolution, which adopts a B3 domain fold. Structural comparison to B3 domains in plant transcription factors and other restriction enzymes identifies key recognition motifs responsible for site-specific DNA binding. Moreover, biochemistry and structural modeling provide a rationale for how LlaJI.R1 may bind a target site that differs from the 5-bp sequence recognized by other LlaJI homologs and identify residues critical for this recognition activity. These findings underscore the inherent structural plasticity of B3 domains, allowing recognition of a variety of substrates using the same structural core.

Classical restriction modification (RM) systems are ubiquitous in bacteria and act as a requisite layer of defense against predatory bacteriophage viruses (1). These systems consist of a restriction endonuclease and a methyltransferase, which provide the dual function of cleaving exogenous DNA site-specifically, whereas protecting the host genome via methylation of the corresponding recognition sequence (2). Three variants of RM systems, type I, type II, and type III, have been identified and differ in their structural composition and mechanism of restriction. Type I systems are multifunctional complexes containing separate restriction, methylation, and DNA-sequence recognition subunits. These machines require Mg2+ and ATP, catalyze both restriction and methylation, and cut DNA nonspecifically far from their recognition sites (3). Type II systems are the simplest, generally existing as dimers that carry out the recognition and restriction activities. These enzymes do not require ATP and have a separate, associated methyltransferase (4). Homodimeric type II restriction enzymes recognize DNA sequences that are symmetric, whereas those that are heterodimeric can bind asymmetric sequences (5). Type III systems contain separate modification (Mod) and restriction (Res) subunits that form homodimeric Mod2 and heterotetrameric Res2Mod2 complexes and catalyze both restriction and methylation in a Mg2+ and ATP-dependent manner (6). They differ from type I systems, however, in that they require two inversely oriented recognition sites that can vary in their spatial separation (7).

Modification-dependent restriction systems (MDRSs), colloquially referred to as type IV systems, recognize and cleave modified DNA (8). McrA and McrBC are prototypical MDRSs that target DNA containing 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) (9–14). McrA is a small, dimeric protein that recognizes the symmetrically methylated sequence Y5mCRG (15). The McrB and McrC proteins together form a conservred, two-component restriction complex capable of long-range DNA translocation similar to type I and type III enzymes. *Escherichia coli* (Ec) McrB contains an N-terminal DNA-binding domain (12) and a C-terminal AAA+ motor domain that hydrolyzes GTP and mediates nucleotide-dependent oligomerization into heptameric rings (16). McrB’s basal GTPase activity is stimulated via interaction with its partner endonuclease McrC (13), which cannot bind DNA on its own and in vitro only associates with the McrB oligomer (17). Biochemical studies suggest a model for DNA cleavage in which McrB and McrC assemble at two distant R3C sites (where R is a purine, and 3C is a methylcytosine) and translocate in a manner that requires stimulated GTP hydrolysis (10, 18). Collision...
of McrBC complexes triggers cleavage of both DNA strands close to one of the R^3C sites (14, 19). Other MDRS families display a variable spectrum of specificity for different modifications. These include MspI, which recognizes 5mC and 5hmC (20), the PvuRtsI family, whose members show unique individual specificities for 5hmC and/or 5-glucosylhydroxymethylcytosine (5ghmC) (21), and GmrSD, which recognizes 5ghmC (22). Structural studies of McrB, MspI, PvuRtsI, and AbaSI suggest type IV systems employ a generalized base-flipping mechanism for recognition of the modified DNA (23–27).

The LlaJI restriction cassette was first identified in Lactococcus lactis on the naturally occurring plasmid pNP40 and shown to confer resistance against common lactococcal phages (28). It consists of an operon encoding two 5mC-methyltransferases, LlaJI.M1 and LlaJI.M2, and two restriction proteins, LlaJI.R1 and LlaJI.R2, both of which are absolutely required for restriction activity in vivo (29). The M1 and M2 methyltransferase activities modulate expression of LlaJI operon in vivo (30). Although formally classified as a type II R/M system (REBASE enzyme number 10100, New England Biolabs), LlaJI.R1 and LlaJI.R2 share domain homology with McrB and McrC, respectively. R1 contains sequence motifs that identify its C-terminal portion as a GTP-specific AAA+ domain and R2 contains a conserved C-terminal PD-(D/E)XK endonuclease domain. These features suggest LlaJI enzymes function more like McrB than other type II systems.

Unlike McrB, however, L. lactis LlaJI.R1 binds DNA site-specifically, recognizing the asymmetric 5′-GACGC-3′ sequence in one strand and 5′-GCGTC-3′ in the other strand (29). Other LlaJI homologs have been identified in Helicobacter pylori, Streptococcus pyogenes, Bacillus cereus, and Clostridium cellulovorans (29, 31). Of these, C. cellulovorans LlaJI has also been shown to target the same asymmetric, 5-bp sequence (31). How LlaJI proteins recognize DNA site-specifically is unknown. Here we present the structure of the N-terminal DNA-binding domain of H. pylori LlaJI.R1 (HpR1Δ136) at 1.97-Å resolution, which adopts a B3 domain fold. Structural comparison to B3 domain-containing plant transcription factors and restriction endonucleases identifies the key recognition motifs responsible for site-specific DNA binding. Additional evidence from biochemistry and structural modeling argues that HpLlaJI.R1 binds a target site that differs from the 5-bp sequence recognized by the L. lactis and C. cellulovorans LlaJI homologs. Mutagenesis further identifies residues Arg-17 and Arg-60 as critical determinants of HpLlaJI.R1 DNA binding. Together, these findings underscore the inherent structural plasticity previously noted for B3 domains, which confers specificity to different sequences via the same structural core.

Results

Structure of H. pylori LlaJI.R1 N-terminal domain

Although previous studies show LlaJI.R1 binds DNA site-specifically (29, 31), the molecular means through which this is achieved remains unknown. Numerous attempts to purify either the full-length L. lactis LlaJI.R1 or its isolated N-terminal DNA-binding domain for structural studies were unsuccessful. Bioinformatics identifies various other species harboring the LlaJI operon, including H. pylori (Hp) (29). Computational analyses of these homologs by fold matching and structural prediction algorithms failed to identify a reliable template for modeling DNA interactions. To understand how the LlaJI.R1 binds DNA site-specifically, we therefore crystallized the N-terminal domain of HpLlaJI.R1 (HpR1Δ136) and determined its structure at 1.97 Å by selenium SAD phasing (32) (Fig. 1).

HpR1Δ136 crystallizes in the space group P1 with four molecules (A–D) in the asymmetric unit organized as two homodimers packed end to end, with molecules A and B and molecules C and D pairing together (Fig. 1A). These dimers superimpose with an r.m.s. deviation of 0.589 Å. Each HpR1Δ136 monomer consists of a core six-stranded β sheet that folds into a pseudo-β barrel flanked on four separate edges by α helices (α1–α4) (Fig. 1, B and C). An additional β-strand (β7) inserts at the dimer interface and breaks the symmetry, adopting an antiparallel configuration with β1^B/β1^D and a parallel configuration with β1^A/β1^C (Fig. 1, B and C). Clear connectivity between β7 and α4 can be traced in molecule B (Fig. 1D). Structural superposition of the two asymmetric dimers suggests β7 is connected in the same manner in molecule D despite the lack of density for the α4–β7 loop (Fig. 1E). We observe no density for the corresponding β7 strands in either molecule A or molecule C.

β7 residues Leu-127 and Phe-129 interact with a hydrophobic cluster sandwiched between β1 and the amphipathic α2 helix in each monomer (Fig. 2A). Ile-24, His-27, and Phe-28 in α2 and Val-115, Leu-116, and Leu-119 in α4 provide additional stabilizing contacts across the dimer interface (Fig. 2, A and B). β7 insertion helps space these elements and prevent steric clash that otherwise would occur. A total interaction surface of 800 Å^2 is shared between the monomers. Size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) indicates HpR1Δ136 dimerizes in solution (Fig. 2C), suggesting the observed molecular organization in the crystal lattice is not simply a packing artifact. Deletion of β7 renders HpR1Δ136 insoluble. Point mutations at the dimer interface in α2 (I24N, H27E, F28E) and α4 (L116E and L119E) provide additional stabilizing contacts across the dimer interface (Fig. 2, A and B). α4 (L116E and L119E) display similar solubility phenotypes and could not be purified. Only the V115N mutant retained solubility and, like WT, forms stable dimers in solution when analyzed by SEC-MALS (Fig. 3A). These data support the notion that HpR1Δ136 dimerization is required for stability and biologically relevant.

HpR1Δ136 adopts a B3 domain-fold

The DALI alignment algorithm (33) identifies several structural homologs of HpR1Δ136. These include the DNA-binding domain of the Arabidopsis thaliana (At) auxin-dependent transcription factor ARF1 (PDB code 4dx, Z-score = 8.6, r.m.s. deviation = 3.0), the C-terminal fragment of the BfiI restriction endonuclease (PDB code 3zi5, Z-score = 9.7, r.m.s. deviation = 2.1), and the N-terminal fragment of the EcoRII restriction endonuclease (PDB code 3huf, Z-score = 8.1, r.m.s. deviation = 2.7). Each is comprised of a B3 domain (34–38). B3 domains share a common pseudo-barrel architecture (SCOP number 101935) and act as recognition modules that bind DNA site-specifically (39). Structural superposition confirms HpR1Δ136 similarly adopts a B3 domain-fold (Fig. 4A, Fig. 51). Impor-
tantly, this fold is structurally distinct from the analogous region in EcMcrB, which preferentially binds DNA containing methylated cytosines.

B3 domains contain two critical regions that confer DNA target site–specificity. These recognition motifs, termed the N-arm and C-arm, reside on opposite edges of the pseudo-barrel core and form a wrench-like structure that contacts the major groove (36, 37, 40). In this arrangement, the N-arm specifically associates with the 5′-half of the target site and the C-arm engages the 3′-half. Comparison to the DNA-bound AtARF1 structure identifies these key features within the Hpr1Δ136 model: the N-arm encompasses the β1–β2 loop and the α1 helix and the C-arm localizes to the β3–β4 loop (Figs. 1C and 4B). Both proteins display a comparable electrostatic surface, with an extensive basic patch positioned between the N- and C-arms and coincident with the DNA-binding face of AtARF1 (Fig. 4C). This mode of substrate binding is conserved among DNA-bound B3 domain structures and is consistent with other Hpr1Δ136 structural superpositions (Fig. S1). An exception is the B3 domain of NgoAVII, whose orientation on DNA is inverted such that the N-arm associates with the 3′-half of the target site and the C-arm with the 5′-half (40).

The BfiI contains other unique motifs (termed the N- and C-loops) that provide additional phosphate backbone and minor groove interactions (37). These are shortened in EcoRII and absent in all previously characterized plant B3 domains (36, 38, 41–43). Hpr1Δ136 similarly lacks these segments, suggesting it either evolved from a more simplified common ancestor or lost these segments over time due to a lack of selective pressure.

The putative DNA-binding surface of each Hpr1Δ136 monomer faces away from the dimer interface, suggesting that
Hpr1Δ136 has the capacity to bind two DNA target sites simultaneously. The asymmetric orientation of the β7 strand, however, positions Glu-131 close to one of the binding sites and alters its surface charge potential in a manner that makes it less basic (Fig. 4D). This intrinsic difference would allow one monomer to bind more efficiently and could bias the arrangement of HpLlaJI on DNA.

All previously characterized B3 domains exist as monomers (36–44). To understand what hinders dimerization in these contexts, we superimposed the coordinates of other B3 domains onto our HpR1Δ136 dimer and examined the orientation of secondary structure features relative to the dimer interface (Fig. 5). Although EcoRII and BfiI have structurally equivalent -strands that partially align with 7, they also contain helical segments that sterically prevent two monomers from coming together (Fig. 5, A and B). The α1 helices of VRN1 and UbaLAI would similarly clash and block dimerization (Fig. 5, C and D). AtARF1, NgoAVII, RAV1, and At1g16640.1, in contrast, lack a corresponding 7 strand (Fig. 5, E–H), suggesting monomers cannot be properly spaced to avoid collision. The stabilizing hydrophobic interactions provided by 7 would also be absent. These observations highlight the importance of secondary structure features in modulating the oligomeric state of B3 domains and will be useful for predicting interactions in other uncharacterized proteins that contain this conserved fold.

**Hpr1Δ136 structure provides model for site-specific binding**

Previous biochemical and genetic studies indicate that *L. lactis* and *C. cellulovorans* LlaJI target the 5-bp sequence 5’-GACGC-3’ (29, 31). Hpr1Δ136 shows weak affinity for DNA containing this sequence (Ll) when assessed by filter binding (Fig. 6A, blue). Scrambling the putative binding sequence (Llscr) has no effect on the affinity (Fig. 6A, green), suggesting this represents the basal level for nonspecific DNA binding by Hpr1Δ136. EcMcrB, in contrast, preferentially binds 5mC-containing DNA (5mC) but does not bind a nonmethylated version of the same substrate (nm) under the same assay conditions (Fig. 6A, black versus red). EcMcrB similarly does not bind either the Ll or Llscr substrates (Fig. 6A, yellow and orange), underscoring how its binding depends on the presence of methylated cytosines.

Unexpectedly, Hpr1Δ136 showed enhanced affinity for the *E. coli*-specific 5mC and nm substrates (Fig. 6A, light blue and purple) relative to Ll and Llscr substrates. We attribute this to subtle sequence differences as the binding is independent of methylation status. The 5mC and nm substrates likely contain sequence fragments that more closely mimic the preferred recognition site of Hpr1LlaJI, which is distinct from both EcMcrB and other LlaJI homologs.

Despite a common fold, B3 domains exhibit divergent sequence preferences. Previous structural and biochemical data show that the C-arm length can influence the length of the recognized target site (Fig. 6B). A longer C-arm is excluded from the major groove (Fig. 6C), decreasing the overall binding footprint and biasing recognition toward a 5-bp site (36, 37). A shorter C-arm affords greater access to the DNA bases, which in some instances increases the number of specific contacts and extends the target site to six bases (37, 38). The amino acid
Figure 3. SEC-MALS of HpR1Δ136 mutants. V115N is located at the dimer interface (see Fig. 2B), whereas H14A, R17A, P59A, and R60A are putative binding site mutations based on structural homology (see Fig. 6, D and E). UV trace (black) and calculated molecular weight based on light scattering (blue) are shown. Dashed red lines denote the predicted molecular weight of an HpR1Δ136 monomer and dimer.
composition of the N- and C-arms ultimately dictates specificity, however, and thus some B3 domains with shorter C-arms still bind 5-bp sites (40, 44). Structural superposition reveals a shorter C-arm in HpR1/H9004 (Fig. 6C).

In the absence of a DNA-bound complex and without explicit knowledge of the HpLlaJI target site, we used the AtARF1–DNA structure as a proxy to identify side chains that might contribute to specificity. The N-arm residue His-136 and C-arm residues Arg-181, Pro-184, and Arg-186 are critical for AtARF1 DNA binding (Fig. 6D). Structural modeling reveals similar residues in HpR1/H9004 (Fig. 6D), with His-14 and Arg-17 in the N-arm and Pro-59 and Arg-60 in the C-arm poised to provide base-specific contacts. Interestingly, Arg-17 is spatially oriented like Arg-181 in AtARF1, hinting that it would contact the 3’-half of the target site despite being localized in the N-arm.

To confirm the significance of our structural modeling, we mutated the predicted binding residues in HpR1/H9004 and assessed how each substitution affects interaction with the E. coli-specific nm DNA substrate via filter binding (Fig. 6F). H14A (red), R17A (orange), and R60A (light blue) mutations show a marked decrease in affinity for nm DNA versus wildtype (WT, purple), whereas P59A (green) shows less of an effect. The R17A/R60A double mutant (brown) completely abolishes binding. This finding was corroborated using electrophoretic mobility shift assays (EMSAs) to measure the association of HpR1/H9004 with digested, nonmethylated λ-phage DNA (Fig. 7). We observed a significant gel shift with WT HpR1/H9004. H14A and P59A show similar shifts in this assay, whereas the individual R17A and R60A substitutions produce a moderate reduction in binding. The R17A/R60A double mutant, however, significantly impairs binding (Fig. 7), similar to its effects on the nm DNA substrate in the filter-binding assay (Fig. 6F). All of these mutants form stable dimers in solution (Fig. 3B), arguing that their effects are not due to global structural perturbations. Together these data implicate
Arg-17 and Arg-60 as critical determinants of HpR1Δ136 DNA binding.

B3 domains contain conserved residues that associate with “clamp” phosphates at the 5’ ends of each strand in the target site (37). Arg-17 and Lys-126 form these interactions in AtARF1 (Arg-81 and Lys-23 in EcoRII; Arg-272 and Lys-340 in BfiI; Lys-27 and Lys-82 in UbaLA1; Lys-212 and Lys-275 in R.NgoAVII). In HpR1Δ136, Arg-6 is poised to act on one strand, whereas Lys-50 could perform a similar function on the opposing strand. Lys-50 is positioned away from the modeled DNA backbone in the apo state and may be reoriented upon target recognition. Conformational rearrangements in the BfiI and R.NgoAVII B3 domains have previously been observed upon DNA binding (35, 37, 40).

Discussion

Here we described the structure of the HpLlaJL.R1 DNA-binding domain and demonstrated that it adopts a B3 domain fold. B3 domains are prevalent among bacterial restriction endonucleases and plant transcription factors, where they function as site-specific DNA-binding modules (37, 39, 45). Previous crystallographic studies revealed that the N- and
Figure 6. Structural modeling of HpR1Δ136 substrate recognition. A, filter binding analysis of HpR1Δ136 (Hp) and full-length EcMcrB (Ec) interactions with different DNA substrates. Substrate abbreviations are as follows: 5mC, methylated EcMcrB-specific substrate; nm, nonmethylated EcMcrB-specific substrate; Ll, site-specific substrate containing the *L. lactis* LlaJI.R1 5′-GACGC-3′ target site sequence; Llscr, substrate with the *L. lactis* LlaJI.R1 target site sequence scrambled as a control. Sequences for each substrate can be found under “Experimental procedures.” Binding was performed at 30 °C for 10 min in a 30-μl reaction mixture containing 14.5 nM unlabeled DNA and 0.5 nM labeled DNA. Samples were filtered through KOH-treated nitrocellulose and binding was assessed by scintillation counting. B, relationship between C-arm length and target site length in previously determined B3 domain–DNA complexes. C, orientation of C-arm loops relative to DNA in various B3 domain homologs. DNA from the AtARF1 complex (PDB code 4ldx) is shown. C-arm coloring is labeled below along with corresponding PDB codes. D, key residues in AtARF1 DNA binding. E, residues predicted to be important for HpR1Δ136 DNA binding based on structural comparison. AtARF1 DNA modeled as in D. F, filter binding analysis of HpR1Δ136 mutants. Point mutations of predicted binding residues identified in D (H14A, red; R17A, orange; P59A, green; R60A, light blue; R17A/R60A, brown) were assessed for binding to the nm DNA substrate. Filter binding was carried out as described in A. The WT curve (purple) is the same as shown in A (Hp + nm).
C-arms determine the specificity of each individual B3 domain and confer structural plasticity to the conserved core scaffold (36–38,40,43). Our structural data and modeling identifies the N- and C-arms in HpR1Δ136 along with key residues that likely form direct contacts with the DNA backbone, clamp phosphates, and specific bases. HpR1Δ136 has weak affinity for DNA containing the asymmetric 5-bp site that other LlaJI homologs target (29,31) and a surprisingly stronger affinity for the EcMcrB-specific DNA substrates, regardless of their methylation status (Fig. 6A). We note that HpR1Δ136 contains a shorter C-arm and thus could potentially bind a 6-bp site. Although further studies are required to pinpoint the target site of HpLlaJI.R1, our findings offer a general model for site-specific binding and provide a structural explanation for why LlaJI homologs do not target modifications despite sharing a similar domain organization with McrBC.

Importantly, we identify the N-arm Arg-17 and C-arm Arg-60 residues as critical determinants of DNA binding and specificity in HpR1Δ136. Individual point mutations at these positions display moderate defects, whereas a combined double mutant completely abolishes DNA binding in all assays tested (Figs. 6F and 7). The H14A and P59A mutations show varying effects depending on the specific substrate used and the sensitivity of the assay. Although His-14 and Pro-59 may also impart specific binding interactions, their contribution is likely context dependent.

HpLlaJI.R1 is unique in that its isolated B3 domain dimerizes, both in solution (Fig. 2C) and in crystallo (Fig. 1). β7 is absolutely essential for HpR1Δ136 dimerization and structural stability, as a truncation of this motif renders the protein insoluble. Our structure shows that direct dimerization of other B3 domains is hindered by either (i) the intrinsic lack of a structur-
ally equivalent $\beta$ strand or (ii) the presence of additional helical motifs at the N or C terminus that sterically clash with $\beta$7 or $\alpha$4 at the dimer interface (Fig. 5). Dimerization of other B domain-containing proteins instead occurs through additional structural elements. For instance, AtARF1 monomers associate through a separate dimerization domain, which facilitates cooperative binding of the B domains to two anti-parallel 5′-TGTCTC-3′ sites on opposing strands (38). BfiI, EcoRII, and R.NgoAVII also dimerize but through their respective nuclease domains (34, 35, 40). These observations will help in classifying uncharacterized B domains and predicting their architectural organization.

Our structural data show that the $\beta$7 strand from one HpR1Δ136 monomer is asymmetrically stabilized at the dimer interface, whereas the corresponding region in the other monomer remains disordered. The orientation of this strand dictates the electrostatic landscape on the dimer surface, making the DNA-binding site in one monomer more basic than the other. Although we cannot completely rule out that this is an artifact of crystallization, an analogous phenomenon was noted in the rotavirus A nonstructural protein 3 (NSP3) homodimer (46). There the asymmetric stabilization of a helix from one monomer creates a single positively charged binding site that ultimately leads to a stoichiometry of 2:1 NSP3:viral mRNA (46). We speculate that HpLlaJI.R1 may bind DNA with a similar 2:1 stoichiometry, but that only one B3 domain will directly contact the target site.

Asymmetric binding could have important implications for the assembly of a cleavage-competent LlaJI restriction complex. Like McrB, LlaJI.R1 contains a conserved GTP-specific AAA+ domain at its C terminus (29). EcMcrB forms heptameric rings in the presence of GTP and this oligomerization is critical for recruiting its partner endonuclease McrC (16), which cannot bind DNA on its own and preferentially associates with the assembled AAA+ domain (17). Although the exact organization of McrBC on DNA has yet to be elucidated, biochemical and structural studies have shown the EcMcrB N-terminal domain binds a single methylated cytosine via base flipping (23). The intrinsic asymmetry of the McrBC complex therefore imposes constraints on how the individual subunits interact with the R$^M$C site and suggests that some monomers are directly engaged whereas others are not. The asymmetric HpR1Δ136 dimer could reflect a similar structural constraint in the LlaJI system wherein the alternative positioning of the $\beta$7 strand dictates which monomers bind the target sequence. Further structural characterization of both systems will be necessary to parse out how substrate binding, GTP-dependent assembly, and nuclease recruitment are coordinated in each case. Although LlaJI and McrBC differ in their specificity and targeting, we predict the general molecular mechanisms governing the function of LlaJI and McrBC will be conserved.

**Experimental procedures**

**Cloning, expression, and purification of HpLlaJI.R1 constructs**

DNA encoding the *H. pylori* LlaJI.R1 protein (DOE IMG/M ID 637022177) was codon optimized for *E. coli* expression and synthesized commercially by Bio Basic Inc. DNA encoding the N-terminal domain (HpR1Δ136; residues 1–136) was amplified by PCR and cloned into pET21b, introducing a His$_{6}$ tag at the C terminus. Selenomethionine-labeled (SeMet) HpR1Δ136 was expressed in minimal media using methionine auxotrophs (T7 Express Crystal Competent *E. coli*, New England Biolabs) according to manufacturer protocols. Native HpR1Δ136 was transformed into BL21(DE3) cells, grown at 37°C in Terrific Broth to an A$_{600}$ of 1.0, and then induced with 0.3 mM isopropyl-1-thio-$\beta$-D-galactopyranoside 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 $\beta$-mercaptoethanol), and pelleted a second time. Pellets were typically 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 phenylmethanesulfonyl fluoride, 5 mg of DNase (Roche Applied Science), 5 mM MgCl$_{2}$, and a Roche 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. 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 HpR1Δ136, the supernatant was filtered, loaded onto a 5-ml HiTrap chelating column charged with NiSO$_{4}$, and then washed with nickel load buffer. HpR1Δ136 was eluted with an imidazole gradient from 30 mM to 1 M. Pooled fractions were dialyzed overnight at 4°C into SP loading buffer (20 mM HEPES, pH 7.5, 50 mM NaCl, 1 mM EDTA, 5% glycerol (v/v), and 5 mM DTT). The sample was applied to a 5-ml HiTrap SP HP column equilibrated with SP loading buffer and then washed with SP loading buffer. HpR1Δ136 was eluted with a NaCl gradient from 50 mM to 1 M. Pooled fractions were concentrated and further purified by SEC using a Superdex 200 10/300 column. All proteins were exchanged into a final buffer of 20 mM HEPES, pH 7.5, 150 mM KCl, 5 mM MgCl$_{2}$, and 1 mM DTT (5 mM for SeMet labeled) during SEC and concentrated to 5–40 mg/ml. Concentrations of purified proteins were determined by SDS-PAGE and densitometry was compared against BSA standards. All amino acid substitutions were introduced into HpR1Δ136 in pET21b by QuikChange PCR and mutant proteins were purified as described for WT.

**Cloning, expression, and purification of EcMcrB**

DNA encoding the *E. coli* McrB protein (Uniprot P15005) was codon optimized for *E. coli* expression and synthesized commercially by GENEART. DNA encoding the full-length protein (EcMcrB FL) was amplified by PCR and cloned into c2xP, a modified pMAL c2x vector with an HRV3C protease site replacing the Factor Xa site directly upstream of the mcrB gene. Native EcMcrB FL was expressed as N-terminal maltose-binding protein fusion in BL21(DE3) cells. Transformed cells were grown at 37°C in Terrific Broth to an A$_{600}$ of 0.8–1.0, and then induced with 0.3 mM isopropyl-1-thio-$\beta$-D-galactopyranoside overnight at 19°C. All cells were harvested, washed with amylose loading buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5% glycerol (v/v), 1 mM EDTA, and 1 mM DTT), and pelleted a second time. Pellets were typically flash frozen in liquid nitrogen and stored at −80°C.
Thawed pellets from 500-ml cultures were resuspended in 30 ml of amylose loading buffer supplemented with 10 mM phenylmethylsulfonyl fluoride, 5 mg of DNase (Roche), 5 mM MgCl₂, and a Roche 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. 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 40 ml of packed amylose resin (New England Biolabs), and then washed with amylose loading buffer.

EcMcrB FL was eluted with amylose loading buffer supplemented with 10 mM maltose. Hrv3C protease was added to the eluate, and amylose loading buffer was added to pooled fractions and dialyzed overnight at 4 °C into Q loading buffer (20 mM Tris–HCl, pH 8.0, 50 mM NaCl, 1 mM EDTA, 5% glycerol (v/v), and 1 mM DTT). The sample was applied to a 5-ml HiTrap Q HP column equilibrated with Q loading buffer and then washed with Q loading buffer. EcMcrB FL was eluted with a NaCl gradient from 50 mM to 1 M. Pooled fractions were concentrated and further purified by SEC using a Superdex 75 10/30pg column. 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.

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

SeMet HpR1Δ136 was crystallized by sitting drop vapor diffusion in 0.1 M 1:2:2 mL-malic acid:MES:Tris base (MMT), pH 6.5, 25% PEG 1,500 (v/v) by mixing 1 μl of protein with 1 μl of the condition with a final drop size of 2 μl and reservoir volume of 65 μl. Crystals appeared within 2–8 days at 20 °C. Samples were cryoprotected with Parabar 10312 and frozen in liquid nitrogen. Single-wavelength anomalous diffraction (SAD) data of two crystals were collected remotely on the tuneable NE-CAT 24-ID-C beamline at the Advanced Photon Source at the selenium edge energy at 12.663 keV (Table 1). Crystal 1 was of the space group P1 with unit cell dimensions \( a = 37.47 \) Å, \( b = 44.39 \) Å, \( c = 84.78 \) Å and \( \alpha = 98.04^\circ \), \( \beta = 94.37^\circ \), \( \gamma = 98.52^\circ \) and showed strong anomalous signal. Crystal 2 was of the space group P1 with unit cell dimensions \( a = 37.53 \) Å, \( b = 43.77 \) Å, \( c = 85.09 \) Å and \( \alpha = 97.87^\circ \), \( \beta = 93.86^\circ \), \( \gamma = 97.77^\circ \). Both crystals were prepared in the same condition but exhibited mosaicities of 0.20871° and 0.11033°, respectively. Data were integrated and scaled using XDS (47) and AIMLESS (48) via the NE-CAT RAPD pipeline. Se-SAD phasing with the data from crystal 1 yielded an initial model that was incomplete and contained a few regions of ambiguity. Heavy atom sites were located using SHELX (49) and phasing, density modification, and initial model building was carried out using the Autobuild routines of the PHENIX package (50). Further cycles of model building and refinement were carried out manually in COOT (51) and PHENIX, respectively (50), but failed to improve significantly the density and refinement statistics. A more complete model was obtained using the diffraction data from crystal 2. The structure was solved by molecular replacement with PHASER (52) using the SAD-derived structure from crystal 1 as the search model. This vastly improved the resulting maps and statistics following subsequent rounds of manual model building and refinement. The final model of crystal 2 was refined to 1.97-Å resolution with \( R_{work}/R_{free} = 0.1927/0.2233 \) (Table 1) and contained four molecules in the asymmetric unit: molecule A, residues 1–121; molecule B, residues 1–131; molecule C, residues 1–121; molecule D, residues 1–131. Threonine 57 exists as a Ramachandran outlier with relatively weak density in the β3–β4 loop of molecules B and D, respectively, and could not be refined further. All structural models were rendered with PyMOL (Schrödinger, Inc.) and surface electrostatics were calculated with APBS (53).

**SEC-MALS**

Purified HpR1Δ136 at 4 mg/ml was subjected to size-exclusion chromatography using a Superdex 200 10/300 column (GE Healthcare) equilibrated in SEC buffer (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**

The following DNA oligonucleotides for filter binding were synthesized commercially by Integrated DNA Technologies (IDT): 5mC, 5'-CCGGGTAAGA(5mC)CGGTAGCGAGC-C CCCG; 5mC, 5'-CCGGGCTCGCTA(5mC)CGGTCTTACCCCG; 5mC, 5'-CCGGGTAAGACCGGTAGCGAGCCCG; 5mC, 5'-CCGGGCTCGCTACGGTCTTACCGCG; Ll_us, 5'-CAGATCTGAGGCTAGAG; Ll_us, 5'-AAGCT-
CTAGCGTCGATGCTG; Llscr_us, 5'-CGATGTCGAGCT-AGAGCTG; Llscr_ls, 5'-AAGCTACGTCGAGATGCTG.

Lyophilized single-stranded oligonucleotides were resuspended 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 ssDNA. Four duplex DNA substrates were prepared: a methylated EcMcrB-specific substrate, 5mC (5mC_us and 5mC_ls), a nonmethylated EcMcrB-specific substrate, nm (nm_us and nm_ls), a site-specific substrate containing the L. lactis LlaJI.R1-binding site LI (LI_us and LI_ls), and a substrate with the L. lactis LlaJI.R1-binding site sequence scrambled as a control Llscr (Llscr_us and Llscr_ls).

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 EDTA, and 40 μg/ml BSA. Binding was performed with purified HpR1Δ136 (WT or mutants) or EcMcrB FL at 30 °C for 10 min in a 30-μl reaction mixture containing 14.5 nM unlabelled DNA and 0.5 nM labeled DNA. Samples were filtered through KOH-treated nitrocellulose filters (Whatman Protran BA 85, 0.45 μm) using a Hoefer FH225V filtration device for ~1 min. 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 two independent experiments and were compared with a negative control to determine fraction bound.

EMSA

The standard buffer for the EMASAs contained 10 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM MgCl2, and 1 mM DTT. Binding was performed with purified HpR1Δ136 (WT or mutants) at 25 °C for 30 min in a 16-μl reaction mixture containing 10 ng/μl of λ-phage DNA (New England Biolabs) digested with BamHI and NdeI (New England Biolabs) and purified via a NucleoSpin® Gel and PCR Clean-Up kit (Machery-Nagel). Following incubation, samples were analyzed by 0.7% agarose gel in 1× TAE at 4 °C and 80 V for 90 min. All gels were stained with SYBR® Green in 1× TAE overnight at 25 °C (Thermo-Fisher Scientific) and visualized using a Bio-Rad Gel Doc™ EZ imager system.

Author contributions—C. J. H. and J. S. C. conceptualization; C. J. H. and J. S. C. data curation; C. J. H. software; C. J. H. and J. S. C. formal analysis; C. J. H. and J. S. C. validation; C. J. H. and J. S. C. investigation; C. J. H. and J. S. C. visualization; C. J. H. and J. S. C. methodology; C. J. H. and J. S. C. writing—original draft; C. J. H. and J. S. C. writing—review and editing; J. S. C. resources; J. S. C. supervision; J. S. C. funding acquisition; J. S. C. project administration.

Acknowledgments—We thank April Lee for assistance with initial crystallization experiments, the Northeastern Collaborative Access Team (NE-CAT) beamline staff at the Advanced Photon Source (APS) for assistance with remote X-ray data collection, Dr. Holger Sondermann for critical reading of the manuscript and use of his SEC-MALS instrument, Dr. John O’Donnell for assistance with SEC-MALS data analysis, and Dr. Eric Alani and Dr. Carol Manhart for guidance and assistance with filter binding experiments. This work is based upon research conducted at the NE-CAT beamlines (24-ID-C and 24-ID-E) under the General User Proposals GUP-51113 and GLIP-41829 (PI: J. S. C). NE-CAT is funded by National Institutes of Health program project Grant P41 GM103403. The Pilatus 6M detector on 24-ID-C beam line is funded by National Institutes of Health-ORIP HEI Grant S10 RR029205. This research used resources of the Advanced Photon Source, a U.S. Dept. of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract DE-AC02-06CH11357.

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., Dryden, D. T., Raleigh, E. A., Wilson, G. G., and Murray, N. E. (2014) Highlights of the DNA cutters: a short history of the restriction enzymes. Nucleic Acids Res. 42, 3–19 CrossRef Medline
3. Loenen, W. A., Dryden, D. T., Raleigh, E. A., and Wilson, G. G. (2014) Type I restriction enzymes and their relatives. Nucleic Acids Res. 42, 20–44 CrossRef Medline
4. Pingoud, A., Fuxreiter, M., Pingoud, V., and Wende, W. (2005) Type II restriction endonucleases: structure and mechanism. Cell Mol. Life Sci. 62, 685–707 CrossRef Medline
5. Pingoud, A., Wilson, G. G., and Wende, W. (2014) Type II restriction endonucleases: a historical perspective and more. Nucleic Acids Res. 42, 7489–7527 CrossRef Medline
6. Raghavendra, N. K., Bheema, S., and Rao, D. N. (2012) Mechanistic insights into type III restriction enzymes. Front. Biosci. 17, 1094–1107 CrossRef
7. Mseis, A., Bickle, T. A., Krüger, D. H., and Schroeder, C. (1992) Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. Nature 355, 467–469 CrossRef Medline
8. Loenen, W. A., and Raleigh, E. A. (2014) The other face of restriction: modification-dependent enzymes. Nucleic Acids Res. 42, 56–69 CrossRef Medline
9. Raleigh, E. A., and Wilson, G. (1986) Escherichia coli K-12 restricts DNA containing 5-methylcytosine. Proc. Natl. Acad. Sci. U.S.A. 83, 9070–9074 CrossRef Medline
10. Sutherland, E., Coe, L., and Raleigh, E. A. (1992) McrBC: a multisubunit GTP-dependent restriction endonuclease. J. Mol. Biol. 225, 327–34 CrossRef Medline
11. 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 Medline
12. Gast, F. U., Brinkmann, T., Pieper, U., Krüger, T., 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
13. Pieper, U., Schweitzer, T., Groll, D. H., and Pingoud, A. (1999) Defining the location and function of domains of McrB by deletion mutagenesis. Bioc. Chem. 380, 1225–1230 Medline
14. Stewart, F. J., Panne, D., Bickle, T. A., and Raleigh, E. A. (2000) Methyl-directed restriction enzymes. N. E. (2014) Highlights of the DNA cutters: a short history of the restriction enzymes. Nucleic Acids Res. 42, 3–19 CrossRef Medline
15. Raleigh, E. A., and Wilson, G. (1986) The other face of restriction: modification-dependent enzymes. Nucleic Acids Res. 42, 56–69 CrossRef Medline
16. Schirmer, T., Wild, C., and Noyer-Weidner, M. (1995) McrB: a prokaryotic protein specifically recognizing DNA containing modified cytosine residues. EMBO J. 14, 2661–2669 Medline
17. 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
18. Pieper, U., Schweitzer, T., Groll, D. H., and Pingoud, A. (1999) Defining the location and function of domains of McrB by deletion mutagenesis. Bioc. Chem. 380, 1225–1230 Medline
19. Stewart, F. J., Panne, D., Bickle, T. A., and Raleigh, E. A. (2000) Methyl-directed DNA binding by McrBC, a modification-dependent restriction enzyme. J. Mol. Biol. 298, 611–622 CrossRef Medline
20. Mulligan, E. A., Hatchwell, E., McCorkle, S. R., and Dunn, J. J. (2010) Differential binding of Escherichia coli McrA protein to DNA sequences...
that contain the dinucleotide m5CpG. Nucleic Acids Res. 38, 1997–2005

16. 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

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

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

19. 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 characteristic 10-base pair repeat. Biochemistry 41, 5245–5254

20. Cohen-Karni, D., Xu, S. Y., Apone, L., Fomenkov, A., Sun, Z., Davis, P. J., Kinney, S. R., Yamada-Mabuchi, M., Xu, S. Y., Davis, T., Pradhan, S., Roberts, R. J., and Zheng, Y. (2011) The MspJI family of modification-dependent restriction endonucleases for epigenetic studies. Proc. Natl. Acad. Sci. U.S.A. 108, 11040–11045

21. Cortes, D., and Zhu, Z. (2013) Characterization of the 5-hydroxymethyl-cytosine-specific DNA restriction endonucleases. Nucleic Acids Res. 41, 4198–4206

22. Bair, C. L., Rifat, D., and Black, L. W. (2007) Exclusion of glucosyl-hydroxymethyl-cytosine DNA containing bacteriophages. J. Mol. Biol. 366, 779–789

23. Sukackaite, R., Grauzulis, 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

24. Horton, J. R., Wang, H., Mabuchi, M. Y., Zhang, X., Roberts, R. J., Zheng, Y., Wilson, G. G., and Cheng, X. (2014) Modification-dependent restriction endonuclease. MspJI, flips 5-methylcytosine out of the DNA helix. Nucleic Acids Res. 42, 12092–12101

25. Kazrani, A. A., Kowalska, M., Czapinska, H., and Bochtler, M. (2014) Crystal structure of the ShmC specific endonuclease PvuRII. Nucleic Acids Res. 42, 5929–5936

26. Horton, J. R., Borgaro, J. G., Griggs, R. M., Quimby, A., Guan, S., Zhang, X., Wilson, G. G., Zheng, Y., Zhu, Z., and Cheng, X. (2014) Structure of 5-hydroxymethyl-cytosine-specific restriction enzyme, AbaSI, in complex with DNA. Nucleic Acids Res. 42, 7947–7959

27. Shao, C., Wang, C., and Zang, J. (2014) Structural basis for the substrate selectivity of PvuRII, a 5-hydroxymethyl-cytosine DNA restriction endonuclease. Acta Crystallogr. Sect. D Biol. Crystallogr. 70, 2477–2486

28. O’Driscoll, J., Glynn, F., Cahalane, O., O’Connell-Motherway, M., Fitzgerald, G. F., and Van Sinderen, D. (2004) Lactococcal plasmid pNP40 encodes a novel, temperature-sensitive restriction-modification system. Appl. Environ. Microbiol. 70, 5546–5556

29. O’Driscoll, J., Heiter, D. F., Wilson, G. G., Fitzgerald, G. F., Roberts, R., and Van Sinderen, D. (2006) A genetic dissection of the LlaJI restriction cassette reveals insights on a novel bacteriophage resistance system. BMC Microbiol. 6, 40–52

30. O’Driscoll, J., Fitzgerald, G. F., and van Sinderen, D. (2005) A dichotomous epigenetic mechanism governs expression of the LlaJI restriction/modification system. Mol. Microbiol. 57, 1532–1544

31. Yang, X., Xu, M., and Yang, S. T. (2016) Restriction modification system analysis and development of in vivo methylation for the transformation of Clostridium cellulosovorans. Appl. Microbiol. Biotechnol. 100, 2289–2299
Structure of H. pylori LlaJI.R1 N-terminal domain

McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 CrossRef Medline

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

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 CrossRef

Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., et al. (2018) Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112–128 CrossRef Medline