12092–12101 *Nucleic Acids Research, 2014, Vol. 42, No. 19* *Published online 27 September 2014*
doi: 10.1093/nar/gku871

Modification-dependent restriction endonuclease, MspJI, flips 5-methylcytosine out of the DNA helix

John R. Horton[1], Hua Wang[2], Megumu Yamada Mabuchi[2], Xing Zhang[1], Richard J. Roberts[2], Yu Zheng[2], Geoffrey G. Wilson[2] and Xiaodong Cheng[1,*]

[1]Department of Biochemistry, Emory University School of Medicine, 1510 Clifton Road, Atlanta, GA 30322, USA and [2]New England Biolabs, 240 County Road, Ipswich, MA 01938, USA

Received August 14, 2014; Revised September 09, 2014; Accepted September 10, 2014

ABSTRACT

MspJI belongs to a family of restriction enzymes that cleave DNA containing 5-methylcytosine (5mC) or 5-hydroxymethylcytosine (5hmC). MspJI is specific for the sequence 5(h)mC-N-N-G or A and cleaves with some variability 9/13 nucleotides downstream. Earlier, we reported the crystal structure of MspJI without DNA and proposed how it might recognize this sequence and catalyze cleavage. Here we report its co-crystal structure with a 27-base pair oligonucleotide containing 5mC. This structure confirms that MspJI acts as a homotetramer and that the modified cytosine is flipped from the DNA helix into an SRA-like-binding pocket. We expected the structure to reveal two DNA molecules bound specifically to the tetramer and engaged with the enzyme's two DNA-cleavage sites. A coincidence of crystal packing precluded this organization, however. We found that each DNA molecule interacted with two adjacent tetramers, binding one specifically and the other non-specifically. The latter interaction, which prevented cleavage-site engagement, also involved base flipping and might represent the sequence-interrogation phase that precedes specific recognition. MspJI is unusual in that DNA molecules are recognized and cleaved by different subunits. Such interchange of function might explain how other complex multimeric restriction enzymes act.

INTRODUCTION

Cytosine in eukaryotic DNA can occur in several chemical forms, including ordinary cytosine (C), 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (1–7). There is much interest in understanding where these forms occur in genomes, how they differ between cell types and change with time and circumstances and how they affect gene expression. Recently discovered 'modification-dependent' restriction enzymes such as MspJI/AspBHI (8,9), and PvuRts1I/AbaSI (10,11), provide a way of answering some of these questions at single-base resolution. These enzymes bind to duplex DNA at sequences containing certain modifications and cleave the DNA a short distance away, generating fragments that can be sequenced and analyzed by bioinformatics (7,12,13). MspJI, for example, recognizes 5mCNNR and 5hmCNNR (R = purine, A or G) and cleaves downstream ∼9 bases away on the modified, 'top', strand, and ∼13 bases away on the complementary, 'bottom', strand, thus: 5(h)mCNNR 9/13 (12,14). At CpG dinucleotide sequences—the principle sites of eukaryotic modification—when the cytosine in both DNA strands is appropriately modified, MspJI can cleave on both sides to produce short fragments of the form 13/9 YN5(h)mCGNR 9/13 (Y = pyrimidine, C or T). These fragments capture some, although not all, of the (hydroxy)methylated sites in the genome (12).

In common with many restriction enzymes (REases) that cut outside of their recognition sequence, cleavage by MspJI is stimulated by the addition of oligonucleotides (oligos) that contain its recognition sequence (12,14). This suggests that cleavage takes place cooperatively, catalyzed by two or more allosterically regulated enzyme molecules acting together. Earlier, we determined the structure of MspJI in the absence of DNA and found that it forms an unusual homotetramer, the subunits of which possess two different conformations (8). Each subunit comprises an N-terminal putative DNA-binding domain and a C-terminal putative DNA-cleavage domain joined by an ∼10-residue linker. In subunits A and B of the tetramer, the binding and cleavage domains are close together ('closed' conformation), and in subunits C and D they are farther apart with respect to each other ('open' conformation).

Based on structural and biochemical observations, we proposed a complex model for MspJI in which the tetramer binds two DNA molecules at once, one subunit recognizing

*To whom correspondence should be addressed. Tel: +1 404 727 8491; Fax: +1 404 727 3746; Email: xcheng@emory.edu
Correspondence may also be addressed to Geoffrey G. Wilson. Tel: +1 978 380 7370; Fax: +1 978 921 1350; Email: wilson@neb.com

each duplex and two other subunits cooperating to cleave it (8). Awkward as this organization might seem, further structural analysis we have now performed, of MspJI bound to DNA, largely confirms it. We report this bound structure here and our interpretation for how MspJI acts. The new structure extends our understanding of restriction enzyme mechanisms and provides a precedent for how cleavage by other REases that act as complexes might also take place (15–18).

MATERIALS AND METHODS

Crystallography

MspJI wild type (WT) was expressed and purified as previously described (8,12). For co-crystallization, final concentrations utilized were ∼12 mg ml^{-1} MspJI (0.24 mM monomer or 0.06 mM tetramer) in 20 mM Tris-HC1 (pH 8.0), 150 mM NaCl, 10% glycerol (v/v), 1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM dithiothreitol (DTT) and 0.13 mM 27-bp oligo, in the presence of 4 mM $CaCl_2$ (to avoid cleavage of the oligo). Crystals were grown at 16°C by the hanging-drop vapor-diffusion method using equal amounts of protein/DNA mixture and well solutions with 10–15% polyethylene glycol MME 2000, 12.5–15% Tacsimate™ (Hampton Research), pH 6.4–6.6. The Tacsimate solution (100% at pH 7.0) was adjusted with HCl to lower the pH to 6.4–6.6. Crystals large enough to acquire with a nylon loop (Hampton) were quickly transferred to the well solution containing 20% ethylene glycol before being flash-frozen directly in liquid nitrogen, stored and later used in X-ray diffraction experiments.

The data set was processed using the program HKL2000 (19). During initial processing, data first appeared to be $P6_122$, but were determined to be $P6_1$ by merohedral twinning and Matthews probability analysis. Molecular replacement with the MspJI tetramer (PDB 4F0Q) (8) as the initial search model, map production and model refinement were conducted using the PHENIX software suite (20). As the twin fraction refinement revealed the data to exhibit near perfect merohedral twinning, we proceeded with map examination and refinement carefully using restraints for secondary structure and NCS (non-crystallographic symmetry) afforded in PHENIX. Restraints were relaxed in later rounds of refinement. Only group thermal B-factor refinement was utilized until the last rounds of refinement, where individual thermal B-factors refinement and TLS (translation/libration/screw) parameters were applied. Maps and models were visualized with COOT (21) as well as conducting manual model manipulation during refinement rounds.

Mutagenesis and activity assays

pNEB206A-His6MspJI, a derivative of plasmid pUC19 containing a codon-optimized *MspJI* gene with an N-terminal 6X-His tag, was used as the template for mutagenesis. Mutants were generated by polymerase chain reaction using oligos synthesized by Integrated DNA Technologies (IDT). All mutants were sequence-verified. For enzyme preparations, 2 l cultures were grown to stationary phase overnight at 37°C without induction, the cloned MspJI gene expressing constitutively from the upstream LacZ promoter. Cells (∼6 g/l) were collected by centrifugation, resuspended in 50 ml of sonication buffer (50 mM Tris, pH 8.0, 50 mM NaCl, 1 mM EDTA, 1 mM DTT, 5% glycerol) and broken by sonication. Cell lysates were clarified by centrifugation and chromatographed using an AKTA FPLC machine (GE Healthcare), first on a 1-ml HisTrap HP column and then on a 1-ml HiTrap Heparin HP column. The clarified sonicate was loaded onto the HisTrap column in sonication buffer. The column was washed with 40-column volumes of 20 mM Na phosphate, pH 7.5, 400 mM NaCl, 20 mM imidazole, 5% glycerol. Bound protein was eluted with a 20–500-mM gradient of imidazole in the same buffer. Fractions containing MspJI were combined and diluted 20-fold into 20 mM Tris, pH 7.5, 20 mM NaCl, 1 mM EDTA, 1 mM DTT, 5% glycerol and loaded onto the HiTrap Heparin column. The column was washed with 25-column volumes of the same buffer and then eluted with a 20–1000-mM gradient of NaCl in the same buffer. Proteins purified in this way were largely homogeneous, as judged by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), with concentrations ranging from ∼0.1 to 1 mg/ml. When necessary, proteins were concentrated using a VivaSpin concentrator.

Activity assays were performed as 4-fold serial titrations in NEBuffer 4 + bovine serum albumin (BSA) (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 100 microg/ml BSA, pH 7.9). Each 30-μl reaction contained 1 μg of supercoiled, Dcm-modified, pBR322 DNA (0.35 pMol substrate; six recognition sites = 2.1 pMol sites). MspJI (0.325 μg or 1.6-pMol tetramer) was added to the first tube in each series, and 10-μl aliquots of this mixture were transferred serially to subsequent tubes. Reactions were incubated at 37°C for 1 h and electrophoresed in 1% agarose gels. Most reactions were performed in duplicate; one set of each, as described above, included no activator. Each tube of the other set contained, in addition, 1.5 pMol of double-stranded (ds) activator oligo containing a fully methylated Dcm/MspJI target site, C5mCWGG (12).

RESULTS

We crystallized MspJI in the presence of a 27-bp asymmetric dsDNA oligo containing the hemimethylated recognition site, **5mCGGG**, close to the 5′ end of the top strand. Previous experiments indicated that this was the shortest oligo MspJI could cleave. The structure was solved by molecular replacement and refined to a resolution of 3.0 Å (Table 1). The crystallographic asymmetric unit contained one MspJI tetramer and one DNA duplex (Figure 1a), even though crystallization was carried out under conditions of two duplexes per tetramer. This discrepancy stemmed from crystal packing. Although each tetramer interacted with two symmetry-related DNA molecules (Figure 1b and c), every DNA molecule also interacted with two symmetry-related tetramers, a consequence of crystallization that reduced the overall stoichiometry to 1:1 (Figure 1d). The proximal portion of each duplex, containing the methylated MspJI recognition sequence (Figure 1e), interacted specifically with the DNA-binding domain of subunit C of one tetramer, and the distal portion interacted

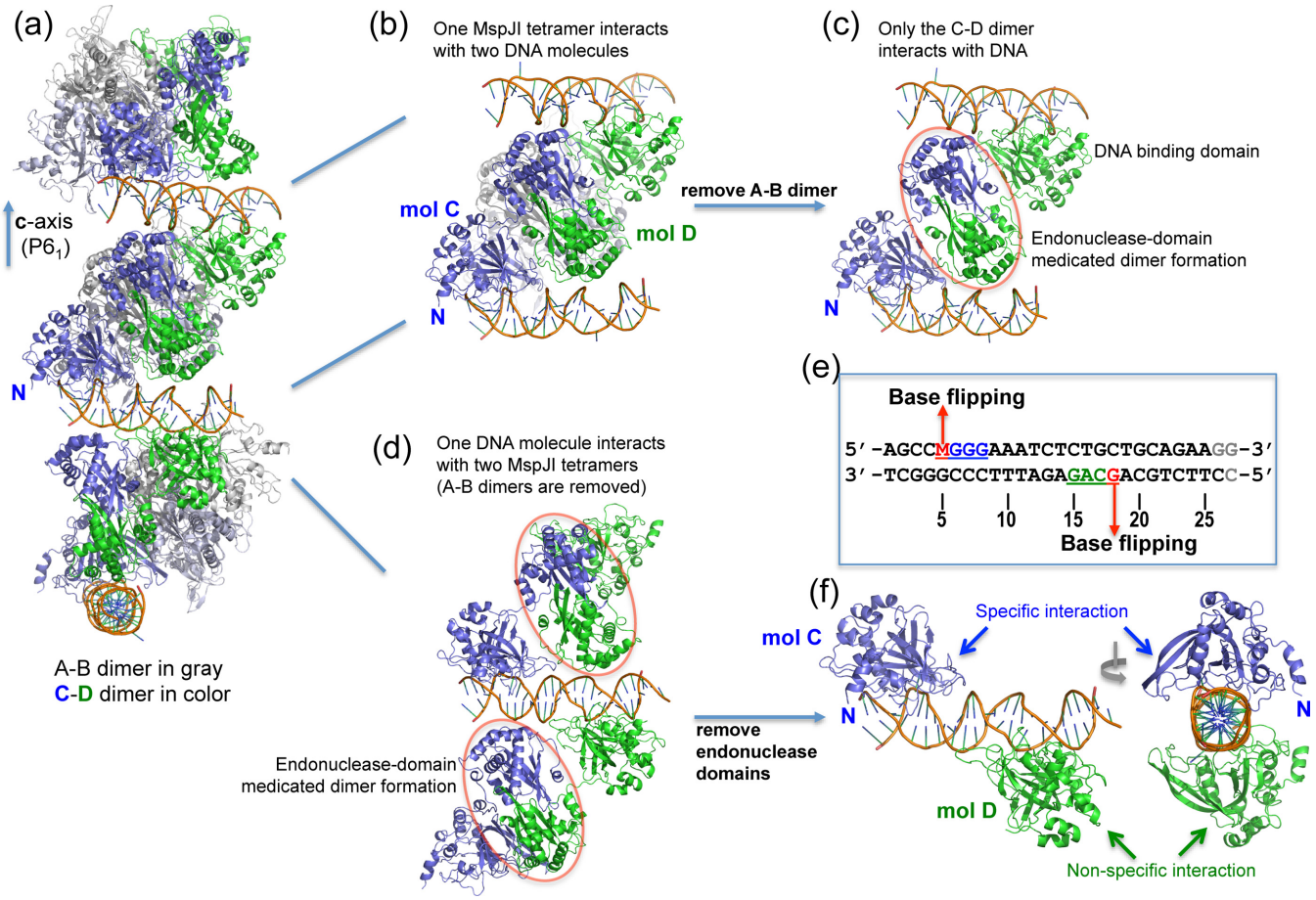

Figure 1. Packing interaction of the MspJI–DNA complex in the crystal lattice. (**a**) Three tetramer–DNA complexes packed together in space group $P6_1$. Each complex is rotated 120° along the crystallographic **c**-axis. (**b**) One tetramer interacts with two symmetry-related DNA molecules. (**c**) Only the C-D dimer of the tetramer interacts with DNA. (**d**) Alternatively, one DNA molecule interacts with two symmetry-related tetramers. (**e**) The 27-bp DNA oligo used for crystallization. (**f**) The subunits C and D of two different tetramers interact with one DNA molecule.

non-specifically with the DNA-binding domain of subunit D of the symmetry-related tetramer (Figure 1f), mediating crystal lattice interactions along the crystallographic **c**-axis (Figure 1a).

The MspJI tetramer

The DNA-bound form of the MspJI homotetramer is similar in overall structure to the unbound form that we described previously (8). Each subunit comprises an N-terminal DNA sequence-recognition domain (residues 1–260) and a C-terminal DNA strand-hydrolysis domain (residues 271–456), with an ∼10-residue connecting linker (amino acids 261–270). The catalytic domains of all four subunits are structurally similar (root-mean-squared deviation = 1.3 Å across 186 Cα atoms), as too for the most part are the recognition domains (rmsd = 1.7 Å across 251 Cα atoms). The spatial relationship between the two domains differs among the subunits, however. In subunits A and B, they are close together ('closed' conformation) and in subunits C and D, they are separated and rotated with respect to each other ('open' conformation) (Figure 2a). The recognition domains of the A:B dimer are held close to the catalytic body of enzyme, while those of the C:D dimer are swung out (Figure 2a). Only the latter, 'open' recognition domains of the C:D dimer interact with DNA in the co-crystal (Figure 2a).

The catalytic sites of the A:B dimer face away from one another, in locations where they cannot cooperate to accomplish cleavage of dsDNA (Figure 2b). Those of the C:D dimer do the same. Pairing of the A:B dimer with the C:D dimer to form the tetramer brings these catalytic sites together in positions suitable for cleavage. The catalytic site of subunit A juxtaposes that of D, and catalytic site of subunit B, that of C (Figure 2b). Together, the four catalytic domains form the central, box-shaped, body of the enzyme comprising two catalytic centers for dsDNA cleavage (Figure 2b). These centers are located on opposite sides of the box, accessible from solution and with geometries appropriate for dsDNA cleavage generating 4-nt 5′-overhangs. The DNA molecule bound specifically by subunit C in the structure is aligned with, and closest to, the catalytic center formed by subunits A and D (Figure 2a and b). Modeling showed that if DNA were bound specifically by subunit D, it would be aligned with, and closest to, the catalytic center formed by subunits B and C, whereas DNA bound by subunits A and B would be misaligned, and far from, either

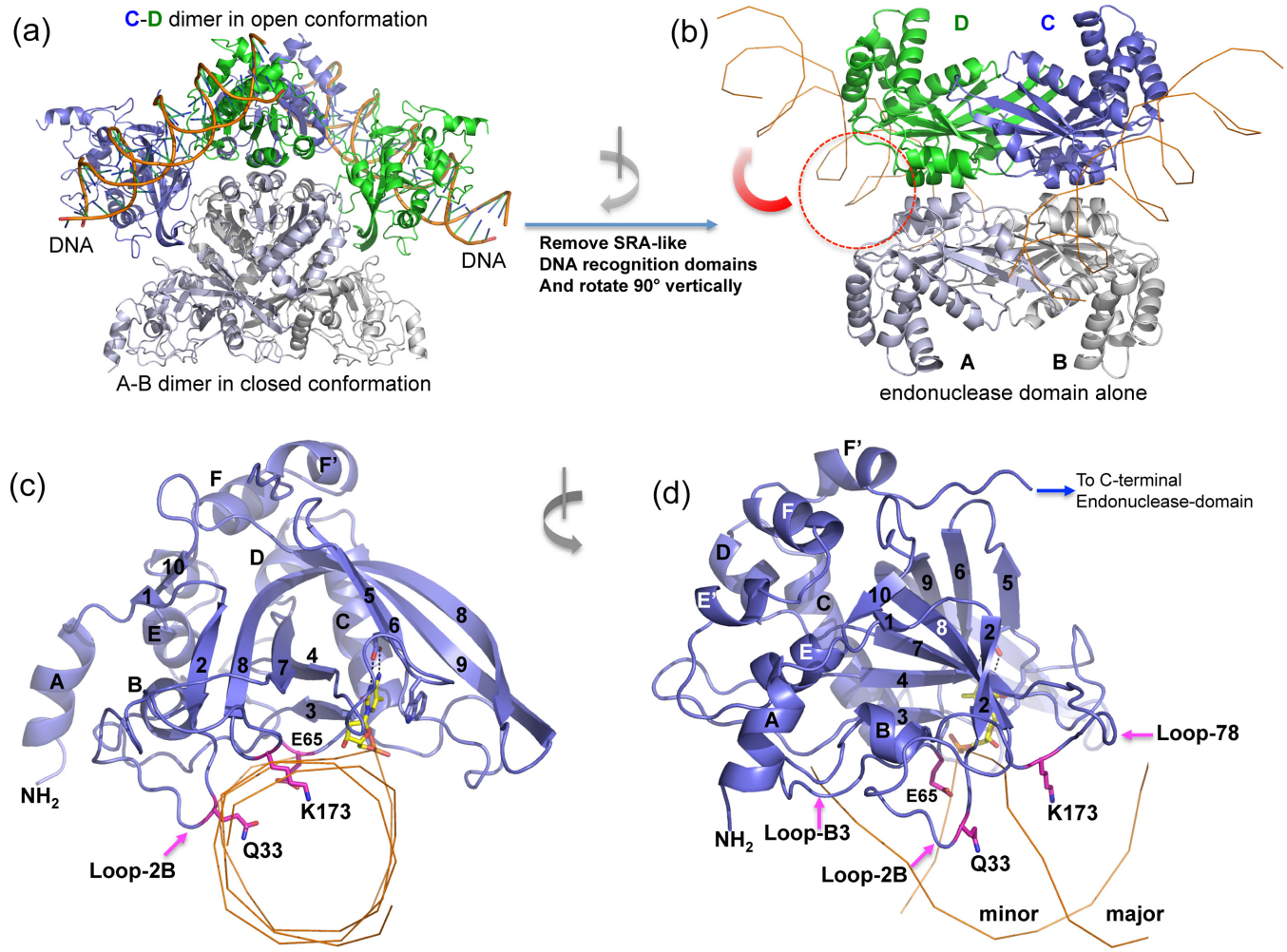

Figure 2. General features of MspJI–DNA interaction. (**a**) One tetramer interacts with two DNA molecules. The A-B dimer is in closed conformation (bottom), while the C-D dimer is in the open conformation (top). The DNA bound specifically by the recognition domain of subunit C (in blue on the left side of the tetramer) is aligned with, and closest to, the catalytic center formed by subunits A (green) and D (see panel (**b**)). (b) Simplified tetramer formation mediated by the four catalytic domains (after removing all four DNA recognition domains). The DNA molecule on the left side of the tetramer is close to the catalytic center formed by subunits A and D (dashed red circle). In our structure, the DNA is moved away, due to its interactions with the adjacent tetramer. (**c, d**) Two views of the recognition domain of subunit C binding DNA specifically. The side chains of Gln33 and Glu65 are located in the minor DNA groove, while Lys173 is in the next major groove.

catalytic center. The MspJI tetramer thus appears to have two centers for dsDNA cleavage (the paired catalytic sites of A+D and the paired catalytic sites of B+C); two DNA-binding sites for productive recognition and cleavage (recognition domains C and D); and two DNA-binding sites for non-productive recognition but, potentially, allosteric activation (recognition domains A and B).

Protein–DNA interactions

The 27-bp oligo with which MspJI co-crystallized interacts directly with only two of the four subunits of the tetramer (Figure 1f). Base pairs 5–8 of the top strand of the oligo correspond to a hemimethylated MspJI recognition sequence, 5mCNNR (Figure 1e). The N-terminal domain of subunit C binds this sequence intimately, centered on the 5mC. We refer to this as the 'specific' protein–DNA interaction, which involves the proximal section of the oligo (subunit C in Figure 1f). Thirteen base pairs downstream from the 5mC, the N-terminal domain of subunit D of the neighboring tetramer binds the same DNA molecule less intimately, centered on a guanine in the complementary strand at base pair 18 (Figure 1e). We refer to this as the 'non-specific' interaction, which involves the distal section of the oligo (subunit D in Figure 1f). We discuss these interactions and their significance below.

MspJI–DNA-specific interactions

The N-terminal recognition domain of each MspJI subunit resembles the eukaryotic SET and RING-associated (SRA) domain (22–24), which binds to hemimethylated 5mCpG dinucleotide sequences. It contains a striking arch-like structure of seven β-strands, central to which is the highly curved, 20-residue β8 strand (amino acids 175–194; Figure 2c and d). Helices pack on the outer surface of the arch. The arch of subunit C surrounds the DNA and forms the DNA-binding surface. The methylated cytosine is

Table 1. Summary of X-ray data collection and refinement statistics

Data collection	MspJI/DNA (27 bp)
Space group	$P6_1$
Cell dimensions	$\alpha = 90, \beta = 90, \gamma = 120$ (°) a = 88, b = 88, c = 512 (Å)
Beamline (SERCAT)	APS 22-ID
Wavelength (Å)	1.0000
Resolution (Å)[a]	35.03–3.06 (3.16–3.06)
R_{merge}[a]	0.117 (0.698)
$I/\sigma I$[a]	10.8 (2.3)
Completeness (%)[a]	99.8 (100.0)
Redundancy[a]	4.7 (4.8)
Observed reflections	201 684
Unique reflections[a]	42 723 (4314)
Refinement	
Resolution (Å)	3.06
No. reflections	42 615 (4297)
R_{work}/R_{free}	0.227/0.258
Twinning fraction (operator)	0.49 (h,-h-k,-l)
No. atoms	
Protein	13 193
DNA	1042
B-factors (Å^2)	
Protein	95.7
DNA	96.9
Rms deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.56

[a]Values in the parentheses correspond to the highest resolution shell.
$R_{merge} = \Sigma|I - <I>|/\Sigma I$, where I is the observed intensity and $<I>$ is the averaged intensity from multiple observations.
$<I/\sigma I>$ = averaged ratio of the intensity (I) to the error of the intensity (σI).
$R_{work} = \Sigma|F_{obs} - F_{cal}|/\Sigma|F_{obs}|$, where F_{obs} and F_{cal} are the observed and calculated structure factors, respectively.
R_{free} was calculated using a randomly chosen subset (5%) of the reflections not used in refinement.

flipped out from the helix into a slot-like pocket in the undersurface of the arch. Phosphate contacts involving both the recognition (modified) strand and the complementary strand span several base pairs flanking the flipped 5mC (Figure 3a and b).

The 5mC-binding pocket

The extra-helical 5mC is bound by a combination of hydrophobic and acidic residues (Figure 3c and d). Trp101 and Tyr114 sandwich the base and contribute to stability via π-stacking (Figure 3c). The polar groups of the cytosine ring that normally engage in Watson–Crick base pairing form one hydrogen bond (H-bond) with Ser90 (main chain NH: Cyt *O*2), two with Asp103 (side chain carboxylate: Cyt N3 and *N*4) and one with Phe115 (main chain carbonyl: Cyt *N*4) (Figure 3d). The latter interactions, with the 4-amino group, make the binding pocket specific for cytosine rather than thymine. The cytosine 5-methyl group is in van der Waals contact with the Cα atom of Asp117 (Figure 3d) and also appears to interact with the main chain carbonyl of Gly106 through a C–H…O type of hydrogen bond (Figure 3d)—a common but underappreciated interaction in bio-molecular recognition (25). The latter interactions might serve to distinguish 5(h)mC, for which the enzyme is specific, from unmodified cytosine.

We assume that the side chain of Asp103 remains protonated (COOH) in the binding pocket, in spite of its customary low pKa of ∼4.1. The same is true for the corresponding residue of the SRA domain (Asp474 of mouse UHRF1; (24)), and the conserved 'motif V' glutamate (ENV) of the 5mC-methyltransferases (26–28), which likewise donate an H-bond to the flipped substrate cytosine. The 5(h)mC-binding pocket of MspJI is not static, but changes shape when the methylated base is present. Comparison of the binding pocket of subunit C with the empty binding pockets of subunits A and B shows that the dimensions of the pocket change as Trp101 rotates onto the base to form one layer of the π-stacked sandwich.

Sequence recognition

Outside of the 5mC-binding pocket, residues from three loops interact with DNA bases from the binding surface of the arch: Gln33 (Loop-2B between strand β2 and helix αB), Glu65 (Loop-B3 between helix αB and strand β3) and Lys173 (Loop-78 between strands β7 and β8) (Figure 2c and d). Gln33 and Glu65 are located in the minor DNA groove, and Lys173 in the major groove (Figure 2d). Glu65 forms a minor groove H-bond with the 2-amino group of the intra-helical, orphan guanine. Gln33 does the same with the guanine 3′ to the flipped 5mC (Figure 3e), and if it were closer, it could also interact with the ring N3 atom of the succeeding guanine. The interaction(s) with Gln33 are unlikely to play a role in specificity determination since any base (N) can occupy these two positions in the MspJI recognition sequence, 5mC**N**N**R**. To assess the importance of Gln33, we changed this residue to alanine (Q33A) to eliminate the H-bond(s), and to asparagine (Q33N) to reduce the length of the side chain without otherwise altering its properties. Both mutants displayed WT specificity and cleavage behavior (data not shown), confirming that Gln33 is enzymatically unimportant.

MspJI has a strong preference for the third base down from the flipped 5mC to be a purine (A or G). Specificity for purine is usually associated with a major groove H-bond to the N7 ring atom, often donated by the electropositive side chain of an amino acid such as lysine (29–31). Lys173 of subunit C is positioned to function in this way in MspJI, but is slightly too far (4.3 Å) from the purine N7 atom to form a strong H-bond. It is possible that this is a crystallization consequence due to non-specific binding of the distal portion of the oligo (see the later section) and that Lys173 is closer in solution. It is also possible that the structure observed represents a post-recognition stage and that the transition from recognition to catalysis involves small conformation adjustments that dissolve the earlier contact at this position. To assess the importance of Lys173, we replaced it with all other 19 amino acids. Most of the resulting proteins proved to be completely inactive, indicating that Lys173 is enzymatically essential. Three mutants with large side chains, in which the Lys173 was replaced by glutamate (K173E), tyrosine (K173Y) and arginine (K173R), displayed slight nicking activity on substrate plasmid DNA, but otherwise were also inactive (data not shown).

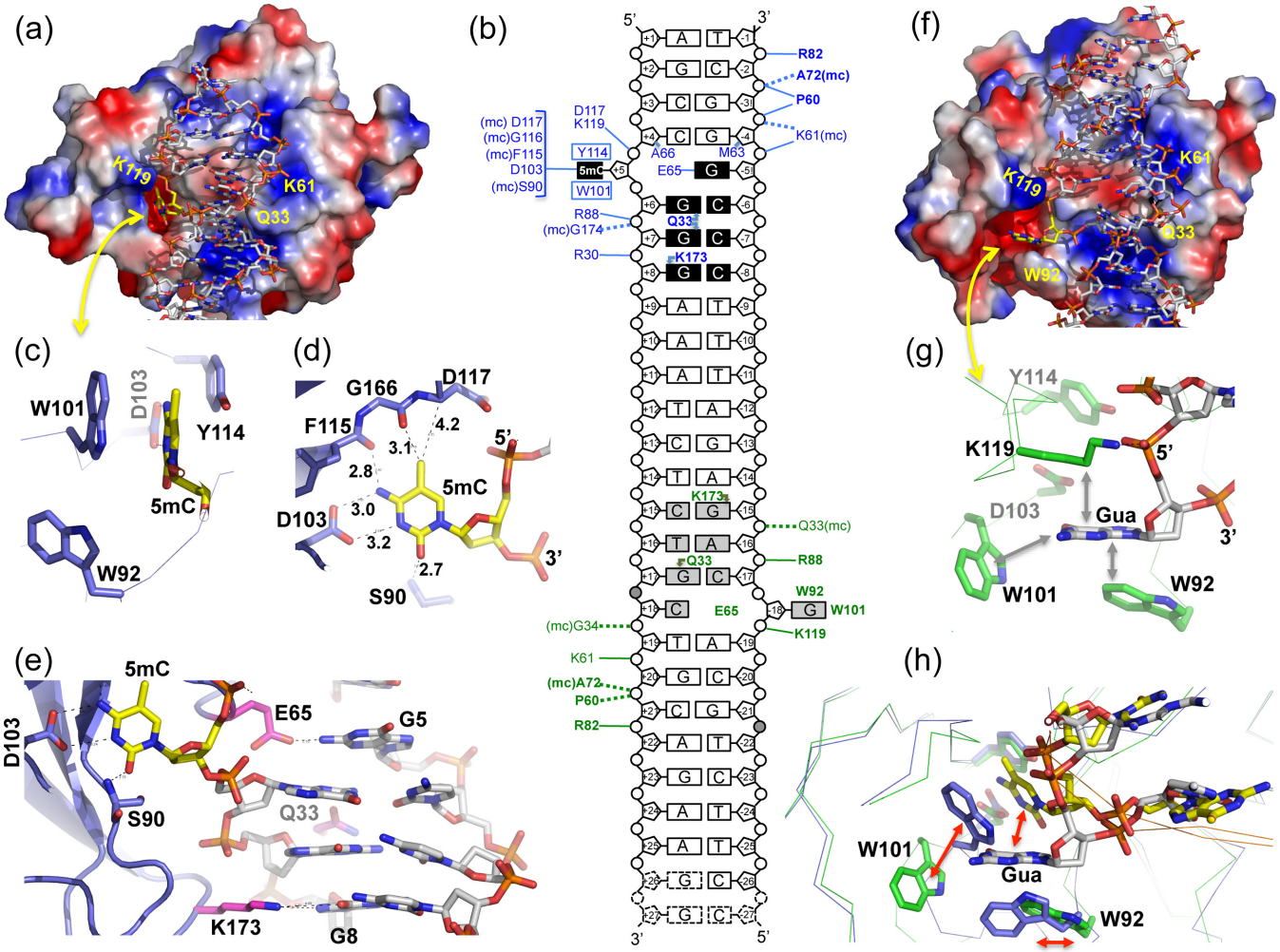

Figure 3. Details of MspJI–DNA interaction. (**a**) The recognition domain of subunit C binds specifically to 5mCNNG. The flipped 5mC is in an acidic slot-like pocket. The surface charge distribution at neutral pH is displayed as blue for positive, red for negative and white for neutral. (**b**) Schematic MspJI–DNA interactions via subunit C (blue) close to the 5′ end of the modified strand (at base-pair position 5) and subunit D of second tetramer (green) at base-pair position 18. (**c, d**) Two views of specific interactions involving the flipped 5mC in the slot-like pocket. Inter-atom distances are shown in Å. (**e**) DNA base interactions involving Gln33, Glu65 and Lys173. (**f**) The recognition domain of subunit D binds non-specifically GNNG sequence. The orientation is similar to that of subunit C–DNA interaction (panel (a)). The flipped Gua is on the edge of an open pocket. (**g**) The flipped Gua is surrounded by side chains of Trp92, Trp101 and Lys119 (which interacts with the 5′ phosphate). (**h**) Superimposition of 5mC-bound subunit C (in blue) and Gua-bound subunit D (in green) indicates that conformational changes (red arrows) occur between specific (yellow DNA) and non-specific interactions (gray DNA).

MspJI–DNA non-specific interactions

The distal portion of the crystallization oligo was found to interact with the binding site of subunit D of the next tetramer in the crystal lattice (Figure 1f). The nucleotide sequence of this portion of the oligo does not contain an MspJI recognition site, and so this interaction is sequence non-specific. Despite the absence of 5mC at this position, a base is nevertheless flipped from the helix, we observe, leaving behind an orphan base in the other DNA strand. This non-specific binding and concomitant base flipping were entirely unexpected.

The flipped base, guanine at position 18 of the complementary strand (Figure 3b) is partially, but not fully, within the 5mC-binding pocket of subunit D (Figure 3f–h). Because the C and D subunits have opposite orientations in the tetramer, and the flipped 5mC and G bases occur in opposite strands, the strand specificity of base flipping is the same at the non-specific site as it is at the specific site. The flipped guanine stacks against Trp92, Trp101 and the aliphatic portion of side chain of Lys119 (Figure 3g). The pocket is more open than in subunit C mainly due to Trp101, the side chain of which is mid-way between the unlatched (subunits A and B) and latched (subunit C) configurations (Figure 3h). Trp92 and Trp101, from the loop between strands β4 and β5, both undergo large conformational changes in conjunction with specific binding of the flipped 5mC (indicated by red arrows in Figure 3h).

The fact that the base is flipped in the non-specific interaction implies an unexpected stability for the flipped conformation. It also suggests that flipping might be an integral part of the sequence-discrimination mechanism of MspJI, the way in which it distinguishes the ‘right’ sequence to bind from all of the other sequences that are ‘wrong’. Most sequence-specific proteins discriminate between bases

while they are intra-helical, through contacts in the major, and to some extent the minor, DNA grooves. Lacking any major groove contacts with which to discriminate 5mC while intra-helical, MspJI might do this instead by flipping every base in turn as it interrogates DNA, and assessing whether or not it fits in the pocket. Those that fit—5mC and 5hmC—are then accepted as 'right', and those that do not fit—everything else—are rejected as 'wrong'. Trp101 might be the key determinant in this regard, latching down on the right bases and capturing them, but not latching down on the others.

A similar interrogation mechanism involving the trapping of sterically correct flipped bases has been proposed for DNA repair enzymes, such as human uracil DNA glycosylase (32). These enzymes distort DNA by bending, interrogate it intra-helically to detect lesions, flip potential substrate nucleotides to varying degrees and reject non-substrates back to DNA helices allowing only the appropriate ones to reach the catalytic site (reviewed in (33,34)). The HhaI methyltransferase has also been found to flip 'wrong' bases into its catalytic site pocket at mismatched target sites within its recognition sequence (35).

In contrast to the large movements of the 5mC binding-pocket residues, those of base-interacting residues, Lys173, Gln33 and Glu65, are small. The third base down from flipped guanine at the non-specific site is also guanine (Gua 15; Figure 3b), as it is too at the specific site. Lys173 in the major groove is farther from the guanine *O*6 atom at this site (4.3 Å, compared to 4.0 Å), and closer to the purine N7 atom (3.1 Å), sufficiently close to form a strong H-bond with the latter. Lys173 is thus better positioned to discriminate A or G (R) from C or T (Y), here, than it is at the proximal site. This might imply that once sequence recognition has occurred, this particular contact is lost due to subsequent conformational changes. Or, it might just indicate that the DNA at the specific site is pulled away in the crystal by its interaction with the neighboring tetramer. The position of Glu65 in the minor groove barely changes between proximal and distal sites. It does not interact with the orphan partner (Cyt 18) of the flipped base, as it does at the proximal site, and this base assumes an unusual perpendicular orientation within the helix. Gln33, also in the minor groove, forms an H-bond with the 2-amino group of Gua 17 located in the opposite strand to the flipped base, rather than in the same strand, as occurs at the proximal site (Figure 3b). In AbaSI, another modification-dependent REase we recently studied, Gln209 interacts with either a G:C or an A:T base pair in the DNA minor groove (11), again indicating that glutamine-mediated minor groove interactions can be non-specific.

MspJI mutagenesis

The properties of a number of point mutants of MspJI (8), and of a related enzyme, AspBHI (9), were reported earlier. Additional mutations, described below, were constructed during this study. Loop-45 (residues 90–103), between strands β4 and β5, includes Ser90 and Trp92 at the beginning of the loop, and Trp101 and Asp103 at the end, which perform critical functions in the 5mC-recognition pocket (Figure 3c and d). This loop adopts a somewhat different conformation according to whether the pocket is occupied (subunit C), partially occupied (subunit D) or empty (subunits A and B). Deletion derivatives were constructed that lacked residues 95 and 96 (mutant Δ95–96) or residues 95 through 98 (Δ95–98) (Figure 4a). Both mutant proteins were almost inactive, confirming the importance of the loop. The residual activity they displayed resulted mainly in nicked substrate plasmid DNA (Figure 4b).

Loop-78 (residues 162–174), between strands β7 and β8, partially occupies the DNA major groove downstream of the 5mC and includes Lys173, the likely determinant of the purine (A or G) specificity of MspJI (Figure 3e). Deletion derivatives were constructed that lacked residues 167 and 168 (Δ167–168), 167 through 170 (Δ167–170) and 167 through 173 (Δ167–173) (Figure 4a). Proteins Δ167–168 and Δ167–173 were completely inactive. Protein Δ167–170 displayed very low activity, and like the loop-56 mutants resulted mainly in nicked substrate DNA (Figure 4b). These results confirm the importance of this loop, too, in enzyme function.

Lys173 is the likely determinant of purine (R) specificity in the MspJI recognition sequence, 5(h)mCNN**R**. As mentioned above, all of the Lys173 mutants we constructed proved to be inactive or nearly-so, indicating that this residue is enzymatically essential. Alone among these, the phenylalanine replacement, K173F, displayed a striking change in behavior. Its endonuclease activity was also much reduced, but rather than generating a fragment pattern, or only nicking DNA, K173F generated a smear at high enzyme concentration, signifying random dsDNA cleavage (Figure 4c). This suggests that Lys173 might couple DNA cleavage to sequence recognition in the WT enzyme, making the one dependent on the proper functioning of the other. In K173F, cleavage and recognition appear to have become uncoupled, such that cleavage can now occur anywhere within the DNA instead of only at recognition sites.

DISCUSSION

The catalytic residues of MspJI comprise Asp334, Gln355, A356 and Lys357 of the $DX_{20}QAK$ motif (8,14), and represent a variation of the common 'PD-(D/E)XK' class of endonuclease sites that require Mg^{2+}-ions for activity (36). To avoid cleavage of the oligo, we co-crystallized MspJI in the presence of Ca^{2+}-ions. This precaution might not have been necessary because the non-specific binding interaction with the adjacent tetramer in the crystal lattice prevents the oligo from engaging with the catalytic centers (Figure 2b). Because MspJI cleaves some distance from its recognition sequence, the distal section of the oligo that must interact with the catalytic residues in order for cleavage to occur was found, instead, to interact with the recognition residues of the neighboring binding domain (Figure 1e). The base pairs involved in the non-specific interaction were the same ones that would otherwise interact with the catalytic sites. This is a coincidence of crystal packing, in our view, rather than a 'bona fide' aspect of the restriction mechanism. The non-specific interaction is nevertheless interesting because it exemplifies, perhaps, the initial phase of MspJI activity in which the enzyme diffuses along the DNA checking for the 'correct' recognition sequence to bind. The co-crystal struc-

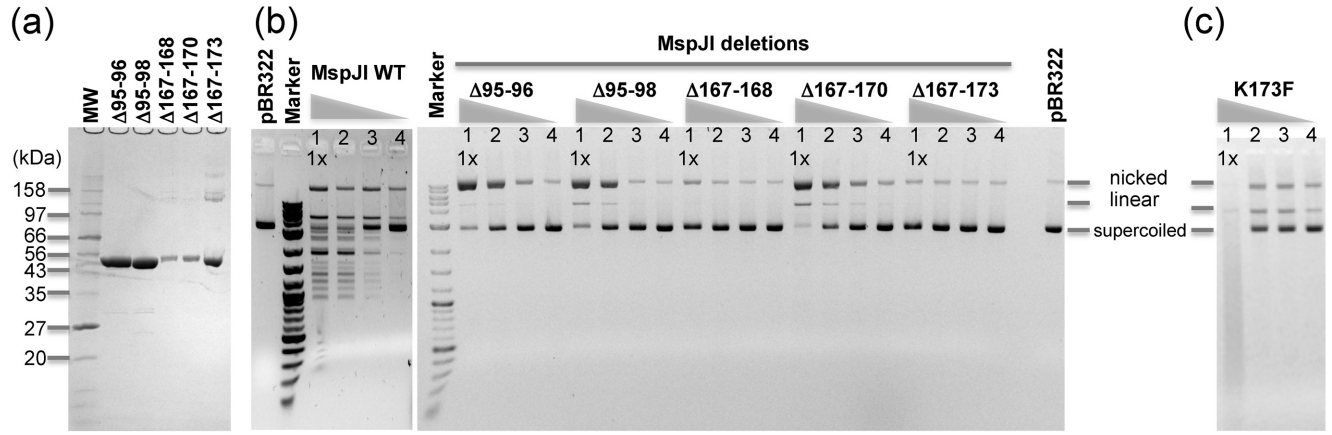

Figure 4. MspJI mutagenesis. (**a**) SDS-PAGE showing the deletion mutants used for activity. (**b**) Activities of MspJI WT and deletion mutants performed as 4-fold serial titrations (lanes 1–4). Approximately equal amount of purified enzymes was added into the starting point of each serial dilution. Based on the cleavage activity showed on the gel, we estimate that the decrease in activity for all the deletion mutants is greater than 64-fold. (**c**) Activity of K173F mutant.

ture fails to show in detail, then, how the catalytic residues of MspJI interact with the DNA. Many PD-(D/E)XK-type catalytic sites have been crystallized in contact with DNA (discussed in (37)), however, and much is known about their variety and modes of action (38,39), and so it is unlikely that MspJI is novel in this regard.

The MspJI tetramer has two identical catalytic centers for double-strand DNA cleavage. These face outward on opposite faces of the box-like body of the enzyme (Figure 2b). One center comprises the paired catalytic domains of subunits A and D, and the other, those of subunits B and C. Each center contains two catalytic sites, one from each constituent subunit. vector alignment search tool (VAST) analysis showed that the catalytic centers are closely similar in structure and geometry to the PD-(D/E)XK catalytic center of the restriction enzyme HindIII (8), which also cleaves dsDNA to produce 4-nt 5′-overhangs. Superimposing the specific HindIII–DNA complex (pdb:2E52) onto our MspJI co-crystal structure indicated where the DNA would be located during cleavage by MspJI.

The distal portion of the oligo bound specifically by subunit C is aligned with, and closest to, the catalytic center formed by subunits A and D (Figure 2b). The catalytic site of subunit D is positioned and oriented to hydrolyze the 'top' (modified) strand of this DNA molecule ∼9 bases away from the MspJI recognition sequence, and that of subunit A is positioned and oriented to hydrolyze the bottom strand ∼13 bases away. Modeling shows that, reciprocally, the top strand of a DNA molecule specifically bound by subunit D would be hydrolyzed by the catalytic site of subunit C, and the bottom strand by the catalytic site of subunit B. In order to engage with the catalytic sites, the DNA must curve. A channel is present on the surface of the protein between the binding sites and the catalytic centers that could accommodate the curved DNA molecules (red circle in Figure 2b), but in our structure the DNA is not curved, and it does not follow this channel, due to its interactions with the adjacent tetramer.

Cleavage by MspJI is strongly affected by substrate concentration. Maximal activity occurs at a molar ratio of approximately four substrate DNA molecules to one MspJI tetramer, and declines markedly when enzyme is present in excess (8). This suggests that MspJI is active only when all four SRA-like DNA recognition domains are specifically bound to recognition sequences. In the co-crystal structure reported here, only the binding domains of the 'open' subunits, C and D, interact with DNA. We consider this a consequence of crystal packing rather than protein structure. Modeling suggests that DNA could interact specifically with the binding domains of the 'closed' A and B subunits in solution, but due to steric conflicts, not in the crystal lattice. DNA modeled into the A and B binding domains is not properly oriented relative to the catalytic centers to undergo cleavage, we find. It seems likely, then, that the recognition domains of subunits A and B do not participate in catalysis, but instead might act indirectly by triggering allosteric activation. Previously, we tested whether non-methylated CNNG oligos would act as activator, instead of 5mCNNG, and no stimulation effect on activity was observed (14).

Comparisons with other restriction enzymes

The individual subunits of MspJI are reminiscent of the Type IIS restriction enzyme, FokI. FokI (specificity: GGATG 9/13) also comprises an N-terminal DNA-recognition domain connected to a C-terminal DNA-cleavage domain through a short linker. FokI cleaves DNA a similar distance away from its recognition sequence as MspJI, and also generates 4-nt 5′-overhangs. Its catalytic domain contains a single PD-(D/E)XK catalytic site, and pairing between two such domains is needed for cleavage. FokI is thought to bind to its recognition sequence as a monomer, and then to cleave following 'transient' dimerization of its catalytic domain with that of another FokI molecule bound to a second site, or less efficiently, free in solution (40,41). Many Type IIS restriction enzymes appear

to act in this way, in fact, and require two recognition sites for cleavage (42).

Two crystal structures of FokI have been reported, one with DNA (43) and one without (44). In the former, the protein is monomeric; the DNA is bound specifically by the recognition domain, but the catalytic domain is far from the DNA, in a 'sequestered' position nestled against the binding domain. This conformation is comparable to the 'closed' forms of the MspJI A and B subunits. In the other FokI structure, without DNA, the catalytic domains of two subunits—both in sequestered conformation—pair to form a catalytic center for dsDNA cleavage. Modeling and electron microscopy (41,45) suggest that at the time of DNA cleavage, one of the FokI monomers adopts an alternative conformation that brings its catalytic domain close to the DNA molecule to which it is bound. Following transient dimerization, this domain then hydrolyzes the bottom DNA strand 13 bases from the recognition sequence, and the other, 'recruited' domain, which remains sequestered, hydrolyzes the top strand 9 bases away (46). The alternative conformation of FokI has not been observed directly, and remains speculative. It might resemble in a general way the alternative, 'open', form of the MspJI C and D subunits.

A number of combined, restriction-and-modification enzymes act as large multimers that bind and cleave several recognition sequences at once. These enzymes, variously referred to as 'Type IIB, IIC or IIG' (reviewed in (37)), cleave outside of their recognition sequences and occur in a variety of oligomeric forms, some comprising separate subunits (15,47), others joined into multi-domain, single-chain proteins (48–50). The structures of their individual restriction, modification and specificity components are fairly well understood, but precisely how these fit together into the active forms of the enzymes is unclear. The structure of the MspJI tetramer sheds some light on this matter. It shows that the domains of otherwise identical subunits can assume different conformations with respect to each other in the assembled enzyme. This might help to explain how enzymes such as TstI compensate for the otherwise puzzling numerical imbalance in their specificity and catalytic domains (18). The MspJI structure also shows that subunits do not necessarily cleave the same DNA molecule they bind, but can cleave those bound by other subunits, instead. This might explain how enzymes such as BcgI cleave apparently upstream of their recognition sequence, while structurally similar enzymes such as MmeI cleave with identical geometry apparently downstream (16,51). What emerges from these considerations is that MspJI, which appears to be a strange restriction enzyme indeed, might not be so unusual after all. As we learn more about the structures of complex restriction enzymes, it might become clear that MspJI shares features with many of them.

ACCESSION NUMBER

Protein Data Bank: the coordinates and structure factors of MspJI–DNA complex have been deposited with accession number 4R28.

ACKNOWLEDGMENTS

The authors thank Don Comb and Jim Ellard for enlightened support for the research presented here; Brenda Baker and John Buswell of the Organic Synthesis group for DNA oligonucleotides; and Devora Cohen-Karni for involvement in the early stage of the project; and together with other members of the restriction enzymes and epigenetics research groups, for numerous thoughtful discussions. X.C. is a Georgia Research Alliance Eminent Scholar.
Author Contributions: J.R.H. performed crystallographic work; H.W. and M.Y.M. performed mutagenesis and activity assays; G.G.W. performed structural analysis and assisted in preparing the manuscript; X.C., X.Z., Y.Z. and R.J.R. organized and designed the scope of the study; and all were involved in analyzing data and preparing the manuscript.

FUNDING

The U.S. National Institutes of Health [GM095209, GM105132 to Y.Z.; GM049245-21 to X.C.]; New England Biolabs; Department of Biochemistry, Emory University School of Medicine, supports the use of Southeast Regional Collaborative Access Team (SER-CAT) beamline at the Advanced Photon Source (APS), Argonne National Laboratory; Use of the APS was supported by the U.S. department of Energy, Office of Science. Funding for open access charge: New England Biolabs.
Conflict of interest statement. None declared.

REFERENCES

1. Kriaucionis,S. and Heintz,N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, **324**, 929–930.
2. Globisch,D., Munzel,M., Muller,M., Michalakis,S., Wagner,M., Koch,S., Bruckl,T., Biel,M. and Carell,T. (2010) Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE*, **5**, e15367.
3. Stroud,H., Feng,S., Morey Kinney,S., Pradhan,S. and Jacobsen,S.E. (2011) 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol.*, **12**, R54.
4. Booth,M.J., Branco,M.R., Ficz,G., Oxley,D., Krueger,F., Reik,W. and Balasubramanian,S. (2012) Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science*, **336**, 934–937.
5. Yu,M., Hon,G.C., Szulwach,K.E., Song,C.X., Zhang,L., Kim,A., Li,X., Dai,Q., Shen,Y., Park,B. *et al.* (2012) Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell*, **149**, 1368–1380.
6. Raiber,E.A., Beraldi,D., Ficz,G., Burgess,H.E., Branco,M.R., Murat,P., Oxley,D., Booth,M.J., Reik,W. and Balasubramanian,S. (2012) Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biol.*, **13**, R69.
7. Sun,Z., Terragni,J., Borgaro,J.G., Liu,Y., Yu,L., Guan,S., Wang,H., Sun,D., Cheng,X., Zhu,Z. *et al.* (2013) High-resolution enzymatic mapping of genomic 5-hydroxymethylcytosine in mouse embryonic stem cells. *Cell Rep.*, **3**, 567–576.
8. Horton,J.R., Mabuchi,M.Y., Cohen-Karni,D., Zhang,X., Griggs,R.M., Samaranayake,M., Roberts,R.J., Zheng,Y. and Cheng,X. (2012) Structure and cleavage activity of the tetrameric MspJI DNA modification-dependent restriction endonuclease. *Nucleic Acids Res.*, **40**, 9763–9773.
9. Horton,J.R., Nugent,R.L., Li,A., Mabuchi,M.Y., Fomenkov,A., Cohen-Karni,D., Griggs,R.M., Zhang,X., Wilson,G.G., Zheng,Y. *et al.* (2014) Structure and mutagenesis of the DNA

modification-dependent restriction endonuclease AspBHI. *Sci. Rep.*, **4**, 4246.
10. 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.
11. 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-hydroxymethylcytosine-specific restriction enzyme, AbaSI, in complex with DNA. *Nucleic Acids Res.*, **42**, 7947–7959.
12. Cohen-Karni,D., Xu,D., Apone,L., Fomenkov,A., Sun,Z., Davis,P.J., Kinney,S.R., Yamada-Mabuchi,M., Xu,S.Y., Davis,T. *et al.* (2011) The MspJI family of modification-dependent restriction endonucleases for epigenetic studies. *Proc. Natl Acad. Sci. U.S.A.*, **108**, 11040–11045.
13. Huang,X., Lu,H., Wang,J.W., Xu,L., Liu,S., Sun,J. and Gao,F. (2013) High-throughput sequencing of methylated cytosine enriched by modification-dependent restriction endonuclease MspJI. *BMC Genet.*, **14**, 56.
14. Zheng,Y., Cohen-Karni,D., Xu,D., Chin,H.G., Wilson,G., Pradhan,S. and Roberts,R.J. (2010) A unique family of Mrr-like modification-dependent restriction endonucleases. *Nucleic Acids Res.*, **38**, 5527–5534.
15. Marshall,J.J., Gowers,D.M. and Halford,S.E. (2007) Restriction endonucleases that bridge and excise two recognition sites from DNA. *J. Mol. Biol.*, **367**, 419–431.
16. Marshall,J.J., Smith,R.M., Ganguly,S. and Halford,S.E. (2011) Concerted action at eight phosphodiester bonds by the BcgI restriction endonuclease. *Nucleic Acids Res.*, **39**, 7630–7640.
17. Smith,R.M., Marshall,J.J., Jacklin,A.J., Retter,S.E., Halford,S.E. and Sobott,F. (2013) Organization of the BcgI restriction-modification protein for the cleavage of eight phosphodiester bonds in DNA. *Nucleic Acids Res.*, **41**, 391–404.
18. Smith,R.M., Pernstich,C. and Halford,S.E. (2014) TstI, a Type II restriction-modification protein with DNA recognition, cleavage and methylation functions in a single polypeptide. *Nucleic Acids Res.*, **42**, 5809–5822.
19. Otwinowski,Z., Borek,D., Majewski,W. and Minor,W. (2003) Multiparametric scaling of diffraction intensities. *Acta Crystallogr. A*, **59**, 228–234.
20. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
21. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
22. Arita,K., Ariyoshi,M., Tochio,H., Nakamura,Y. and Shirakawa,M. (2008) Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature*, **455**, 818–821.
23. Avvakumov,G.V., Walker,J.R., Xue,S., Li,Y., Duan,S., Bronner,C., Arrowsmith,C.H. and Dhe-Paganon,S. (2008) Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature*, **455**, 822–825.
24. Hashimoto,H., Horton,J.R., Zhang,X., Bostick,M., Jacobsen,S.E. and Cheng,X. (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature*, **455**, 826–829.
25. Horowitz,S. and Trievel,R.C. (2012) Carbon-oxygen hydrogen bonding in biological structure and function. *J. Biol. Chem.*, **287**, 41576–41582.
26. Wu,J.C. and Santi,D.V. (1987) Kinetic and catalytic mechanism of HhaI methyltransferase. *J. Biol. Chem.*, **262**, 4778–4786.
27. Baker,D.J., Kan,J.L. and Smith,S.S. (1988) Recognition of structural perturbations in DNA by human DNA(cytosine-5)methyltransferase. *Gene*, **74**, 207–210.
28. O'Gara,M., Klimasauskas,S., Roberts,R.J. and Cheng,X. (1996) Enzymatic C5-cytosine methylation of DNA: mechanistic implications of new crystal structures for HhaI methyltransferase-DNA-AdoHcy complexes. *J. Mol. Biol.*, **261**, 634–645.
29. van der Woerd,M.J., Pelletier,J.J., Xu,S. and Friedman,A.M. (2001) Restriction enzyme BsoBI-DNA complex: a tunnel for recognition of degenerate DNA sequences and potential histidine catalysis. *Structure*, **9**, 133–144.
30. Townson,S.A., Samuelson,J.C., Vanamee,E.S., Edwards,T.A., Escalante,C.R., Xu,S.Y. and Aggarwal,A.K. (2004) Crystal structure of BstYI at 1.85A resolution: a thermophilic restriction endonuclease with overlapping specificities to BamHI and BglII. *J. Mol. Biol.*, **338**, 725–733.
31. Hashimoto,H., Shimizu,T., Imasaki,T., Kato,M., Shichijo,N., Kita,K. and Sato,M. (2005) Crystal structures of type II restriction endonuclease EcoO109I and its complex with cognate DNA. *J. Biol. Chem.*, **280**, 5605–5610.
32. Parker,J.B., Bianchet,M.A., Krosky,D.J., Friedman,J.I., Amzel,L.M. and Stivers,J.T. (2007) Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature*, **449**, 433–437.
33. Parikh,S.S., Putnam,C.D. and Tainer,J.A. (2000) Lessons learned from structural results on uracil-DNA glycosylase. *Mutat. Res.*, **460**, 183–199.
34. David,S.S., O'Shea,V.L. and Kundu,S. (2007) Base-excision repair of oxidative DNA damage. *Nature*, **447**, 941–950.
35. O'Gara,M., Horton,J.R., Roberts,R.J. and Cheng,X. (1998) Structures of HhaI methyltransferase complexed with substrates containing mismatches at the target base. *Nat. Struct. Biol.*, **5**, 872–877.
36. 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.
37. Pingoud,A., Wilson,G.G. and Wende,W. (2014) Type II restriction endonucleases-a historical perspective and more. *Nucleic Acids Res.*, **42**, 7489–7527.
38. Laganeckas,M., Margelevicius,M. and Venclovas,C. (2011) Identification of new homologs of PD-(D/E)XK nucleases by support vector machines trained on data derived from profile-profile alignments. *Nucleic Acids Res.*, **39**, 1187–1196.
39. Steczkiewicz,K., Muszewska,A., Knizewski,L., Rychlewski,L. and Ginalski,K. (2012) Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. *Nucleic Acids Res.*, **40**, 7016–7045.
40. Bitinaite,J., Wah,D.A., Aggarwal,A.K. and Schildkraut,I. (1998) FokI dimerization is required for DNA cleavage. *Proc. Natl Acad. Sci. U.S.A.*, **95**, 10570–10575.
41. Vanamee,E.S., Santagata,S. and Aggarwal,A.K. (2001) FokI requires two specific DNA sites for cleavage. *J. Mol. Biol.*, **309**, 69–78.
42. Bath,A.J., Milsom,S.E., Gormley,N.A. and Halford,S.E. (2002) Many type IIs restriction endonucleases interact with two recognition sites before cleaving DNA. *J. Biol. Chem.*, **277**, 4024–4033.
43. 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.
44. Wah,D.A., Bitinaite,J., Schildkraut,I. and Aggarwal,A.K. (1998) Structure of FokI has implications for DNA cleavage. *Proc. Natl Acad. Sci. U.S.A.*, **95**, 10564–10569.
45. Vanamee,E.S., Berriman,J. and Aggarwal,A.K. (2007) An EM view of the FokI synaptic complex by single particle analysis. *J. Mol. Biol.*, **370**, 207–212.
46. Sanders,K.L., Catto,L.E., Bellamy,S.R. and Halford,S.E. (2009) Targeting individual subunits of the FokI restriction endonuclease to specific DNA strands. *Nucleic Acids Res.*, **37**, 2105–2115.
47. Kong,H., Roemer,S.E., Waite-Rees,P.A., Benner,J.S., Wilson,G.G. and Nwankwo,D.O. (1994) Characterization of BcgI, a new kind of restriction-modification system. *J. Biol. Chem.*, **269**, 683–690.
48. Jurenaite-Urbanaviciene,S., Serksnaite,J., Kriukiene,E., Giedriene,J., Venclovas,C. and Lubys,A. (2007) Generation of DNA cleavage specificities of type II restriction endonucleases by reassortment of target recognition domains. *Proc. Natl Acad. Sci. U.S.A.*, **104**, 10358–10363.
49. Morgan,R.D., Bhatia,T.K., Lovasco,L. and Davis,T.B. (2008) MmeI: a minimal Type II restriction-modification system that only modifies one DNA strand for host protection. *Nucleic Acids Res.*, **36**, 6558–6570.
50. Morgan,R.D., Dwinell,E.A., Bhatia,T.K., Lang,E.M. and Luyten,Y.A. (2009) The MmeI family: type II restriction-modification enzymes that employ single-strand modification for host protection. *Nucleic Acids Res.*, **37**, 5208–5221.
51. 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.