1466–1478 *Nucleic Acids Research, 2020, Vol. 48, No. 3* *Published online 27 December 2019*
doi: 10.1093/nar/gkz1195

Structure of HhaI endonuclease with cognate DNA at an atomic resolution of 1.0 Å

John R. Horton[1], Jie Yang[1], Xing Zhang[1], Theresa Petronzio[2], Alexey Fomenkov[2], Geoffrey G. Wilson[2], Richard J. Roberts[2] and Xiaodong Cheng[1,*]

[1]Department of Epigenetics & Molecular Carcinogenesis, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA and [2]New England Biolabs, Inc., Ipswich, MA 01938, USA

Received October 15, 2019; Revised December 08, 2019; Editorial Decision December 08, 2019; Accepted December 11, 2019

ABSTRACT

HhaI, a Type II restriction endonuclease, recognizes the symmetric sequence 5′-GCG↓C-3′ in duplex DNA and cleaves ('↓') to produce fragments with 2-base, 3′-overhangs. We determined the structure of HhaI in complex with cognate DNA at an ultra-high atomic resolution of 1.0 Å. Most restriction enzymes act as dimers with two catalytic sites, and cleave the two strands of duplex DNA simultaneously, in a single binding event. HhaI, in contrast, acts as a monomer with only one catalytic site, and cleaves the DNA strands sequentially, one after the other. HhaI comprises three domains, each consisting of a mixed five-stranded β sheet with a defined function. The first domain contains the catalytic-site; the second contains residues for sequence recognition; and the third contributes to non-specific DNA binding. The active-site belongs to the 'PD-D/EXK' superfamily of nucleases and contains the motif SD-X_{11}-EAK. The first two domains are similar in structure to two other monomeric restriction enzymes, HinP1I (G↓CGC) and MspI (C↓CGG), which produce fragments with 5′-overhangs. The third domain, present only in HhaI, shifts the positions of the recognition residues relative to the catalytic site enabling this enzyme to cleave the recognition sequence at a different position. The structure of M.HhaI, the biological methyltransferase partner of HhaI, was determined earlier. Together, these two structures represent the first natural pair of restriction-modification enzymes to be characterized in atomic detail.

INTRODUCTION

Produced in almost all free-living bacteria and archaea to combat phage-infections, Type II restriction enzymes (REases) bind to specific, 4–8 base pair (bp), sequences in duplex DNA and, in conjunction with divalent metal ions, typically Mg(II), catalyze double-strand (ds) cleavage of the DNA within or close to their recognition sequence (1,2). Hundreds of Type II REases have been discovered, each highly specific for cleaving a different DNA sequence. Their adoption as enzymatic tools for the precise fragmentation, cloning and analysis of DNA has made them the 'workhorses' of modern molecular biology laboratories (3).

To better understand their mechanisms of cleavage and sequence recognition, the structures of almost 40 Type II REases bound to DNA have been solved by X-ray crystallography. Most act as dimers (EcoRI (4), EcoRV (5), BamHI (6), PvuII (7) and BglI (8)) or tetramers (Cfr10I (9), NgoMIV (10), SfiI (11), SgrAI (12), MspJI (13,14)) of identical subunits. Each subunit contains one catalytic site such that when the dimer binds to its target sequence, one catalytic site cleaves one DNA strand and the other site cleaves the other strand. Usually, these two cleavage reactions proceed simultaneously, in a single binding event. Because the dimers are symmetric overall, the sequences they recognize are also symmetric ('palindromic') as, too, are the positions of cleavage (2).

Initially, all REases with symmetric recognition sequences were expected to act as homodimers, but several have since been found to be monomers, including MspI (C↓CGG) (15), HinP1I (G↓CGC) (16,17), MvaI (CC↓WGG; W = A or T) (18) and BcnI (CC↓SGG; S = G or C) (19,20). These monomeric enzymes possess a single catalytic site, and they accomplish dsDNA cleavage sequentially, binding to the sequence first in one orientation and cleaving one strand, and then detaching, re-binding in the opposite orientation, and cleaving the other strand (20). They can bind in either orientation because the sequences they recognize happen to be symmetric. As would be expected, when such enzymes are constrained by mutation to bind in one only orientation, they now cleave only one strand—they 'nick' the DNA, that is—rather than both ((21) and unpublished observations). We report here that HhaI also acts as a monomer, and thus belongs to this second group.

*To whom correspondence should be addressed. Tel: +1 713 834 6274; Email: xcheng5@mdanderson.org

HhaI, produced by the bacterium *Haemophilus haemolyticus*, was among the first restriction enzymes discovered (22). Initial attempts to clone its gene, based on the acquisition of phage-resistance, led instead to the isolation of a different Type II restriction enzyme encoded by the same bacterium: HhaII (G↓ANTC) (23). Subsequent attempts using alternative methods showed that the gene for HhaI lay next to the gene for the antagonistic M.HhaI methyltransferase (MTase)—the means by which *H. haemolyticus* protects its own DNA from HhaI-cleavage—the two enzymes forming a simple, compact, restriction-modification (R-M) system (24,25). M.HhaI recognizes and binds to the same sequence in dsDNA as HhaI, and converts the first cytosine in each strand to 5-methylcytosine (M), forming 5′-GMGC-3′. The methyl groups protrude into the major groove of DNA and create obstructions that prevent HhaI from binding, thereby rendering the DNA insensitive to the presence of this restriction enzyme.

The structure presented here allows us to contrast HhaI and M.HhaI. Both proteins bind as monomers to the same 4-bp sequence in duplex DNA, and catalyze step-wise transformations of the sequence. HhaI breaks a phosphodiester bond in each strand; M.HhaI methylates a cytosine in each strand. The two enzymes, encoded by adjacent genes, share no similarities and act independently, but the net effect of their complementary activities is a primitive, yet remarkable and widespread, form of innate immunity to infectious DNA molecules.

MATERIALS AND METHODS

The genes encoding the HhaI restriction enzyme and the M.HhaI modification enzyme (*hhaIR* and *hhaIM*, respectively) were cloned into *Escherichia coli* as described (25). The sequences of *hhaM* (24) and *hhaIR* and the linkage between them was determined (J. Barsomian and G.G. Wilson, unpublished). Three single-nucleotide (nt) discrepancies in the sequence were uncovered in the course of the present work leading us to re-sequence *hhaIR* and to determine the entire *H. haemolyticus* genome by PacBio SMRT sequencing.

Genome sequencing

Genomic DNA from an overnight L-broth (26) culture of *H. haemolyticus* strain NEB129 (now called *H. parahaemolyticus*) was purified using a modified protocol from the Monarch DNA purification kit (NEB), and sequenced using the Pacific Biosciences (PacBio) RSII sequencing platform. SMRTbell libraries were constructed from a genomic DNA sample sheared to ∼10–20 kb using the G-tubes protocol (Covaris), end repaired and ligated to PacBio hairpin adapters. Incompletely formed SMRTbell templates and linear DNAs were digested with a combination of Exonuclease III and Exonuclease VII (NEB). DNA qualification and quantification were performed using the Qubit fluorimeter (Invitrogen) and 2100 Bio analyzer (Agilent Technology), and sequenced using the Pacific Biosciences (PacBio) RSII sequencing platform.

One 14 kb SMRTbell library was prepared according to modified PacBio sample preparation protocols including additional separation on a BluePippin (Sage Science), sequenced with P6-C4 chemistry with two SMRT cells: one with non-size selected (14 kb) and one with size selected (20 kb) libraries sequenced with 300- and 360-min collection time respectively. A total of 184 659 sequencing reads, with mean sub-read length of 17 358 bp, gave 1.3 Gb of total sequence that was *de novo* assembled using the HGAP_Assembly.3 version 2.3.0 with default quality and read length parameters, and polished using Quiver (27). The polished assembly generated two closed-circular genome elements with 40.3% GC content for the main chromosome (2 090 133 bp) and 44.8% GC content for the plasmid (1733 bp). The assembled sequence was annotated using the NCBI Prokaryotic Genomes Annotation Pipeline (28,29).

Protein purification

Recombinant HhaI restriction endonuclease was purified by chromatography over DEAE Hyper D and Heparin Hyper D (Pall Bioscience), Source 15S, Heparin HP, Source 15Q and Superdex 75 (GE Lifescience) resins. The enzyme was followed by activity and peak fractions of HhaI restriction activity were pooled. After purification, HhaI protein was stored in 50 mM KCl, 10 mM Tris–HCl 7.4, 1 mM dithiothreitol (DTT), 0.1 mM ethylenediaminetetraacetic acid, 50% glycerol and frozen at −80°C.

Cleavage assay

Supercoiled pUC19 plasmid DNA (2.7 kb; 17 HhaI sites, one each of EcoRI and BamHI sites) was used as the template for cleavage assays. A total of a 50 μl reaction (20 nM plasmid DNA and 2 nM HhaI, with a molar ratio of 170:1 cleavage sites to HhaI) were performed at room temperature in NEB buffer 3.1 (50 mM Tris–HCl pH 7.9, 100 mM NaCl, 10 mM $MgCl_2$, 100 μg/ml bovine serum albumin). An aliquot of 5 μl was taken out at indicated time (5 s, 15 s, 30 s, 1 min, 3 min, 7.5 min and 15 min) and mixed with 5 μl of NEB 6×purple loading dye. For EcoRI and BamHI controls, the 50 μl reaction contained 20 nM pUC19 DNA and 2 nM (20 Units) of enzyme (molar ratio of 10:1 cleavage site to enzyme) and was also carried out at room temperature, with BamHI in NEB buffer 3.1 and EcoRI in NEB buffer-EcoRI (100 mM Tris–HCl pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$, 0.025% Triton® X-100). The samples (10 μl each) were visualized by electrophoresis (1% agarose gel).

Crystallography

After thawing, the storage buffer was changed to contain 0.5 mM tris (2-carboxyethyl) phosphine (TCEP) and 10% glycerol (instead of 1 mM DTT and 50% glycerol) by several rounds of concentration and dilution and then diluted to a final protein concentration of 0.7 mM. HhaI was mixed with a 13 or 14 bp double strand oligonucleotide in the presence of $CaCl_2$ at 4°C for 30 min (0.35 mM protein monomer, 0.2 mM dsDNA and 4 mM $CaCl_2$). We note that cleavage by restriction endonucleases generally occurs in the presence of Mg(II) ions, but not in the presence of Ca(II) ions,

which was used here to capture the pre-cleavage complex of HhaI.

An Art Robbins Gryphon Crystallization Robot was used to set up screens using the sitting drop vapor diffusion method at room temperature (~19°C). We observed crystals under many conditions. The best reported X-ray diffraction datasets were collected from crystals that formed in two pH conditions (pH 4.2 and pH 6.0). The low pH conditions consisted of 10% or 35% (v/v) 2-propanol, 0.1 M sodium phosphate dibasic/citric acid pH 4.2, with or without 0.2 M lithium sulfate. The crystals were in $P2_1$ space group, and three structures were determined (PDB IDs: 6UKE, 6UKF and 6UKG) (Supplementary Table S1). The high pH conditions consisted of 10% (v/v) 2-propanol, 0.1 M MES/sodium hydroxide pH 6.0, 0.2 M calcium acetate. The crystals were in either $P4_12_12$ (PDB ID: 6UKH) or $P2_12_12_1$ space groups (PDB ID: 6UKI). Crystals were cryo-protected by soaking in mother liquor supplemented with 20% (v/v) ethylene glycol before plunging into liquid nitrogen. The metal Ca(II) was present in the mixture of protein and DNA, but only observed in the high pH structures.

A dataset for *ab initio* phase determination (PDB ID: 6UKE) was collected from a single crystal grown from low pH condition containing iodinated DNA (synthesized by B. Baker of New England Biolabs using 5-Iodo-Uridine-phosphoramidite purchased from Glen Research). The 4-axis goniometer at our local X-ray facility equipped with Rigaku instrumentation including a MicroMax-003 Microfocus sealed tube X-ray generator, an AFC11 partial-χ and an HyPix-6000HE hybrid photon counting detector, along with the Rigaku Collection Strategy program (version 1.8.1), allowed us to drive the crystal and detector around so as to collect an overly redundant (~16-fold) dataset (with ~86% unique reflections containing Friedel mates measured from different regions many times to minimize noise) (Supplementary Table S1). This dataset was processed with CrysAlis^{Pro} (Rigaku) and were scaled and merged with the AIMLESS suite of the CCP4Interface (30). The resultant dataset for *ab initio* phasing was examined using the PHENIX Xtriage module (Zwart PH, Grosse-Kunstleve RW, Adams PD. Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation. CCP4 Newsletter, 2005 Winter: Contribution 7), which reported a very good anomalous signal to 3.7Å. The PHENIX AutoSol module (31) found the iodine atom positions with a Figure-Of-Merit of 0.46 and gave a density-modified map with an R-factor of 0.33. The initial electron density showed recognizable molecular features of β-sheets and α-helices. Reinserting the iodine positions into AutoSol and utilizing the full resolution of the dataset allowed some sulfur and phosphorous atoms to be located and gave a better map in which protein side chains and DNA could be easily identified. The AutoBuild module of PHENIX (32) was utilized to begin building the model, and manual building of the protein and the DNA duplex was completed with COOT (33), which was also utilized for corrections between PHENIX refinement rounds.

Other datasets were collected at the 22-ID (SER-CAT) beamline at the Advanced Photon Source (Argonne National Laboratory), equipped with a Dectris Eiger 16M detector and MD2 micro diffractometer. Molecular replacement was performed using this *ab initio* determined structure (PDB ID: 6UKE) as the search model with the PHENIX PHASER module (34). Structure quality was analyzed during PHENIX refinements and later validated by the PDB validation server. For the atomic resolution structures (PDB IDs: 6UKF and 6UKG), hydrogen atoms were included and, since these crystals formed below the pKa of the ionizing side chains of aspartate, glutamate, and histidine residues, many of these side chains were also modeled as protonated in the structure. [We note that the pKa value of DNA phosphate groups is extremely low (0–2), and at pH 4.2 they are still deprotonated.] In addition, in the 1.16Å structure (PDB ID: 6UKG), the electron density revealed that the 14 bp oligonucleotide utilized was in both orientations in the crystal, i.e. the two DNA strands appeared interchanged to some extent and were modeled as such. During the refinement at the atomic resolution, the weight for stereochemical restraints were minimized (weight = 0.1) but not eliminated. The accuracy of the distances given reflects the quality of the electron density. The main reason we kept minimal restraints is that some parts of the complex (i.e. protein–DNA interfaces) are better ordered and could be refined without restraints, whereas some other parts (e.g. the two ends of DNA molecules) are less ordered and may be distorted without restraints. Thus, we took a conservative approach without total elimination of all stereochemical restraints.

We note that neither the uncorrected nor correct protein sequence yield any homology to existing structures that would be useful as a molecular replacement model for structure solution as reported by PHYRE2 (35), a web portal for protein modeling, prediction and analysis. Using the NCBI VAST search engine, MspI (PDB ID: 1SA3) and HinP1I (PDB ID: 1YNM) are identified as the two closest structural matches to HhaI. Neither of these have the extra domain (residues ~160–224) present in HhaI.

RESULTS

Structure at resolution of 1.0 Å

We determined the full-length structure of HhaI (258 residues) in complex with a 13 or 14-bp oligonucleotide (oligo) containing the recognition sequence (GCGC) in three space groups at two pH conditions (4.2 and 6.0). The metal Ca(II) was present in the mixture of protein and DNA, but only observed in the high pH structures. In the $P2_1$ space group formed under pH 4.2, there is one molecule of HhaI in complex with DNA in the crystallographic asymmetric unit. The crystal lattice is packed and mediated by protein–protein interactions and protein-DNA interactions at the two ends of the DNA molecule (Figure 1A). The crystal contained less solvent (~40%) and diffracted X-rays to high resolution near or at 1.0 Å (Supplementary Table S1). At pH 6.0, two orthogonal space groups, $P4_12_12$ and $P2_12_12_1$, formed with solvent content of ~60% and diffraction limit of ~2.7–2.8 Å. In both, the DNA oligos coaxially stacked head-to-end forming a pseudo continuous duplex with each HhaI molecule bound in the middle of the DNA (Figure 1B). We will focus on the description of the

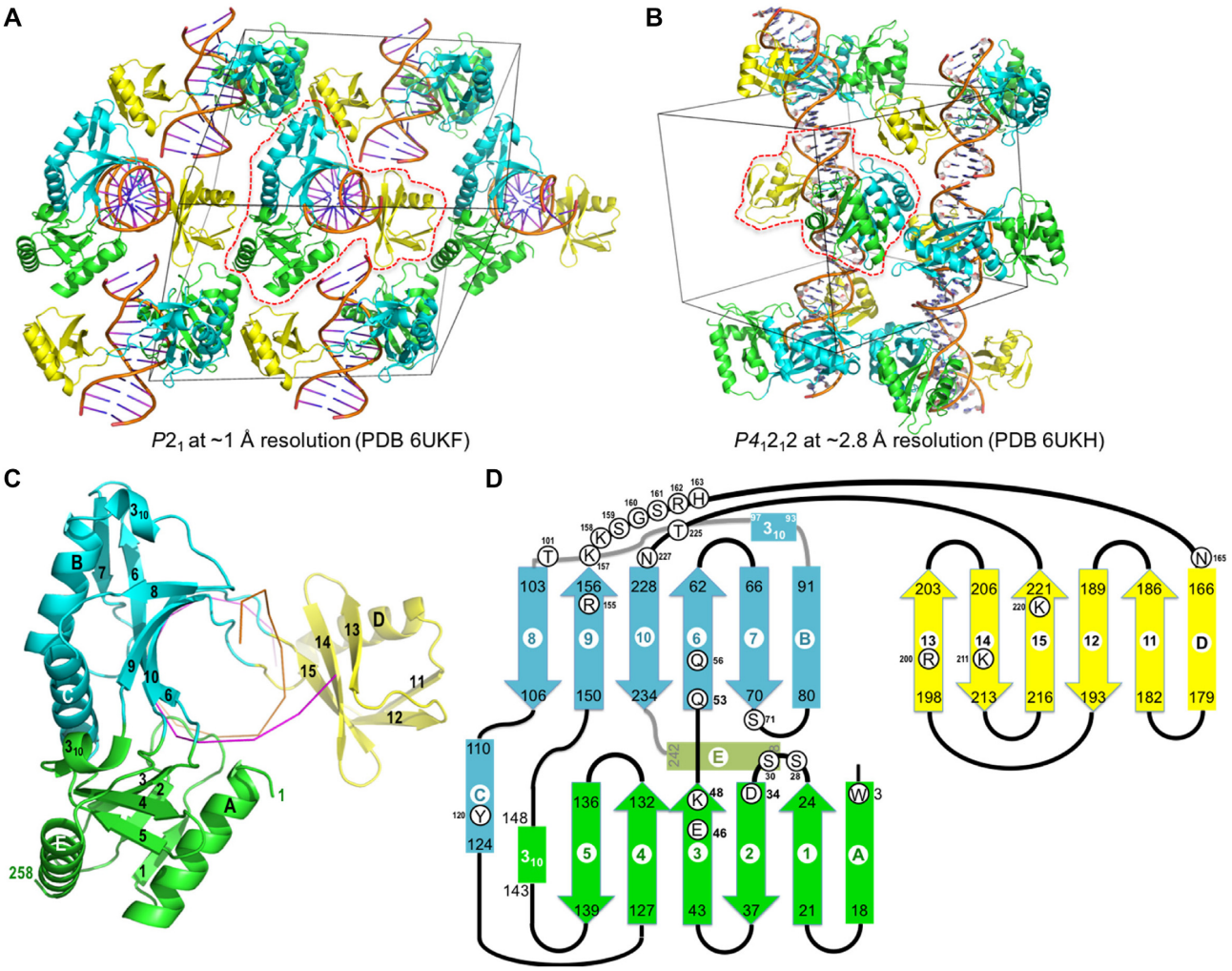

Figure 1. Overall structure of HhaI endonuclease. (**A**) crystal packing in $P2_1$ space group. (**B**) crystal packing in $P4_12_12$ space group. (**C**) The three β-structures of HhaI are colored in green, cyan and yellow. The DNA is depicted as ribbon lines. (**D**) Schematic diagram of HhaI secondary structure. Helices are labeled from A to E and strands are labeled as 1–15.

high-resolution structure in $P2_1$ space group and describe the metal binding in the active-site of a high pH structure.

We initially crystallized HhaI with a 13-bp oligo (5′-CGITGCGCTIGGA-3′) containing two 5-iodo-uridine bases (underlined) outside of the recognition sequence (PDB ID: 6UKE in $P2_1$ space group). The 5-iodo-uridine was used for crystallographic phasing (Supplementary Figure S1). Next, we collected X-ray diffraction data from a second crystal (by means of more than 1.2 millions reflections) using the synchrotron X-ray radiation at a wavelength of 0.8 Å (PDB ID: 6UKF in $P2_1$ space group). The resultant dataset is ∼94% complete for a resolution range of 26.40–0.99 Å, and 68.7% complete for the highest resolution shell of 0.02 Å between 1.01–0.99 Å. The atomic resolution structure, with an overall crystallographic thermal B-factor of 17 Å^2 for the protein component and 27 Å^2 for the DNA component, allowed us to position every atom of the 26 DNA nucleotides, nearly every atom of the 258 protein residues and 342 water molecules.

We also crystallized HhaI in the same $P2_1$ space group with a 14-bp oligo (5′-CTGTTGCGCTTGGA-3′), with GCGC in the middle and 5 flanking base pairs at each end (PDB ID: 6UKG in $P2_1$ space group). However, the last base pair on one end was not observed in the electron density, which effectively resulted in a 13-bp oligo with 4-bp on one side of the GCGC sequence and 5-bp on the other. Because the missing base pair could occur on either end of the oligo, we observed a mixture of the DNA base pairs outside of the symmetric GCGC sequence. Nevertheless, the crystal still diffracted X-rays to a resolution of 1.16 Å (PDB ID: 6UKG; Supplementary Table S1), and the discordance of DNA base pairs outside of the central GCGC did not affect the quality of the protein component. We note that the two high-resolution structures differ by only 0.17 Å in resolution (1.16 Å in 6UKG versus 0.99 Å in 6UKF), the number of unique reflections increased by 53,488, which accounts for approximately one-third of total reflections in the 0.99-Å datasets or more than one-half in the 1.16-Å dataset (Supplementary Table S1). We kept all three structures at $P2_1$ space group because the conformational difference in the active-site loop (see 'Discussion' section).

Table 1. Methylated DNA sequence motifs in *Haemophilus haemolyticus*

Motif[a]	Type	% motif detected	Responsible MTase[b]
G**C**GC	II	100	M.HhaI
G**A**N**T**C	II	100	M.HhaII
G**A**TC	II	100	M.HhaIII
CG**A**N_7**T**AA	II	99.8	(S.HhaORF1590P)
ACCN_5GG**T**	I	100	(S1.HhaORF9470P)
GG**A**N_6RTAY	I	100	(S1.HhaORF6340P +S2.HhaORF6340P)

[a]Locations of the methylated bases are indicated in bold and underlined for 5-methylcytosine and N6-methyladenine. Bold **G** and **T** indicate that the complementary bases are methylated.
[b]Enzymes in parentheses are the current best guesses, but cloning will be necessary to confirm these predictions. Note that the specificity of Type I methylases is defined by these S subunits. In each case, an M subunit is also present, and it is the complex that performs the methylation.

HhaI protein sequence

The DNA sequence of the HhaI gene was determined by manual methods in the early-1990s (NEB accession no. NEBM150; J. Barsomian and G.G. Wilson, unpublished). In the course of the present work, three errors were corrected in the original amino acid sequence for HhaI derived from this nt sequence. They emerged as discrepancies during structural model building, and BLASTP comparisons with otherwise identical 'hypothetical' proteins from *H. parahaemolyticus* in GenBank (WP_005706967), and they were confirmed by *de novo* sequencing of the entire *H. haemolyticus* genome by PacBio SMRT sequencing. Two were point errors: codon 65, originally thought to be AAA (Lys), corrected to GAA, coding for Glu; and codon 161, originally thought to be CCA (Pro), corrected to TCA, coding for Ser. The third was a frame-shift error: an extra A (nt 694) near the 3′-end of our original sequence shortened the C-terminus of the protein by 20 amino acids, from …NA NVIFSISLKNNISLFILNEDRKAFEAAISL* (the correct sequence) to …NANVI*IFNIIKK** (erroneous amino acids in italics).

H. *haemolyticus* genome sequence

One advantage of the PacBio sequencing platform is its ability to detect the epigenetic state of sequenced DNA (36,37). Five DNA MTase recognition motifs were detected by SMRT motif and modification analysis, each containing N6-methyladenine modifications. Four Type I restriction-modification systems were found of which at most two are active. In addition, two known Type II systems, including M.HhaII (20), were clearly detected and one additional system is likely. While PacBio methylome analysis does not reliably detect 5-methylcytosine methylation, the only 5-methylcytosine motif was that for M.HhaI (GCGC), which had been identified previously (24,25). A summary of the results are shown in Table 1 and have been deposited in REBASE (1).

Overall structure of HhaI

HhaI is approximately C-shaped (Figure 1C) and spirals around the DNA helix following the curvature of the major groove. HhaI comprises three mixed, five-stranded, alpha-beta structures (colored in green, cyan and yellow in Figure 1C). The β sheets form the inner surfaces of the structures and face the DNA, while the helices pack against the sheets and form the outer surfaces (Figure 1C). The first sheet (β1–β5) packs against helices αA from the extreme N-terminus, a short 3_{10} helix (residues 143–148), and αE from the extreme C-terminus (Figure 1C and D). Antiparallel strands β2 and β3 contain the conserved catalytic residues D34, E46 and K48 of the $SDX_{11}EAK$ catalytic motif, with D34 located at the beginning of β2 and K48 at the end of β3 (Figure 1D). We note that HhaI has a preserved core architecture of helix αA followed by strands β1–β3, as conserved catalytic residues located in canonical positions in the family of 'PD-(D/E)XK' restriction endonucleases.

The second sheet (β6–β10) packs against helices αB and αC and a short 3_{10} helix (residues 93–97). The third sheet (β11–β15) packs against helix αD and forms a compact domain inserted between strands β9 and β10. The residues for DNA sequence specificity (amino acids 155–161) occur at the end of strand β9 and a continuous string after the strand (Figure 1D). The third sheet (β11–β15) contributes to non-specific DNA binding (see below).

Comparisons with the four characterized monomeric REases (Figure 2) show that the features of the β strands in the first and second sheets are preserved, particularly strands 3, 4 and 5 of the first sheet, and strands 6, 9 and 10 of the second sheet. The third β sheet structure is unique to HhaI; the corresponding region in the other monomeric REases is a loop connecting strands 9 and 10 (Figure 2). Superimposition of HinP1I and HhaI, both which recognize the same 5′-GCGC-3′ but cleave at a different location, either between the 5′ GC dinucleotide (Hinp1I) or the 3′ GC dinucleotide (HhaI), revealed that the third sheet in HhaI might shift the positions of the recognition loop residues relative to the catalytic site enabling this enzyme to recognize sequence 5′ to the cleavage site, whereas HinP1I recognizes the 3′ sequence to the cleavage site (Figure 2F). We also note that several monomeric restriction enzymes have the catalytic domains fused to an additional DNA binding domain: UbaLAI consists of an MvaI-like catalytic domain at the C-terminal and an EcoRII-N-like effector domain (38), DpnI consists of an N-terminal catalytic PD-(D/E)XK domain and a C-terminal winged helix domain (39,40), whereas Sau3aI has two similar domains in the N- and C-terminal halves (41).

DNA structure

We refer to the DNA strand that is cleaved upon binding by HhaI as the 'target strand' of the 5′-G1-C2-G3-C4–3′

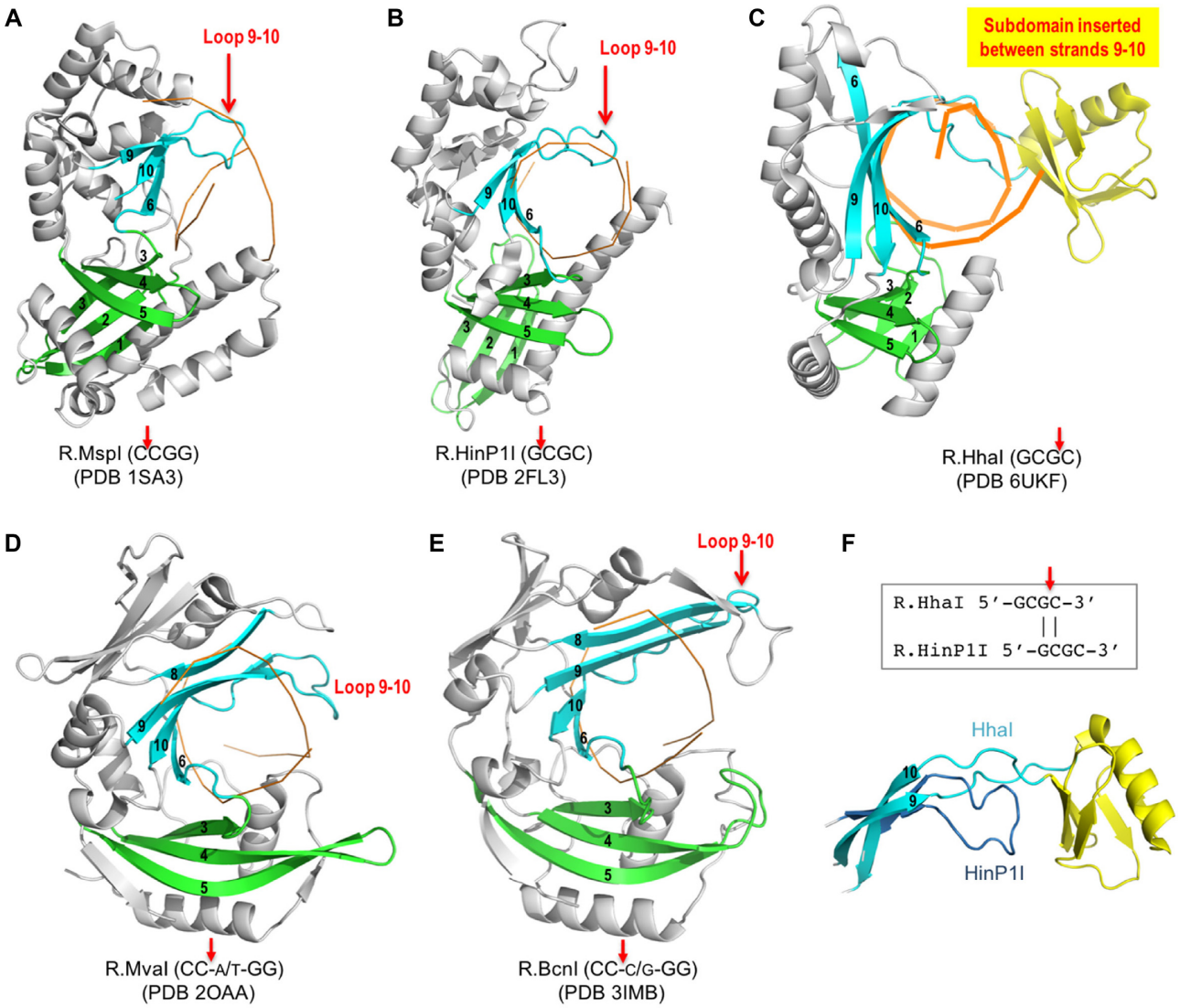

Figure 2. Comparsion of five monomeric endonucleases. (**A**) MspI, (**B**) HinP1I, (**C**) HhaI, (**D**) MvaI and (**E**) BcnI. The conserved β-strands are colored in green for catalysis and cyan for base recognition. (**F**) Superimposition of Hinp1I (β9-loop-β10) and the corresponding strands in HhaI. The box includes the recognition sequence of GCGC aligned by the cleavage site.

recognition sequence, cleavage occurring between guanine G3 and cytosine C4, and to the complementary strand as 5′-G4-C3-G2-C1–3′ (Figure 3A; also see Figure 4A). The DNA is encircled by a single HhaI molecule (Figure 3B and C). Two long loops traverse the major groove, one following strand β9 and the other preceding strand β10. β-structures 2 and 3 (in cyan and yellow) are located on opposite sides of the DNA helix and interconnect by these loops (Figure 3B). Pronounced distortion of the DNA occurs at the G4:C4 end of the recognition site, beyond the target phosphate group. The guanidinium side chain of R155 forms hydrogen bonds (H-bonds) with G4 of the complementary strand, and stacks with the adjacent base outside the recognition sequence (Figure 3D). This interaction is stabilized by Q53, which intercalates into the major groove after C4 of the target strand, causing a kink of approximately 30° (Figure 3E). It is also augmented by the aliphatic side chain of K157, which H-bonds with the preceding base G3 (Figure 3D). The net result is that base-pair G4:C4 has a propeller twist of 35° and base-pair C3:G3 has a buckle of −30° (Figure 3A). Base-pair G2:C2, in contrast, has a perfect in-frame configuration, while base-pair C1:G1 has a minor (−13°) propeller twist (Supplementary Table S2).

Interactions with DNA phosphate backbone

HhaI interacts with the phosphate groups of the two DNA strands asymmetrically. Ten phosphates of the target strand are contacted by all three β structures, and six phosphates of the complement strand are contacted by β structures 2 and 3 (Figure 4A). Due to the low pH (4.2) of crystallization, the imidazole rings of the histidine residues (pKa ∼6) are protonated (= positively charged) and the carboxylate groups of the aspartate and glutamate residues (pKa ∼4.8) are protonated (= uncharged). The low pH leads to some unusual protein–DNA interactions. For example, the

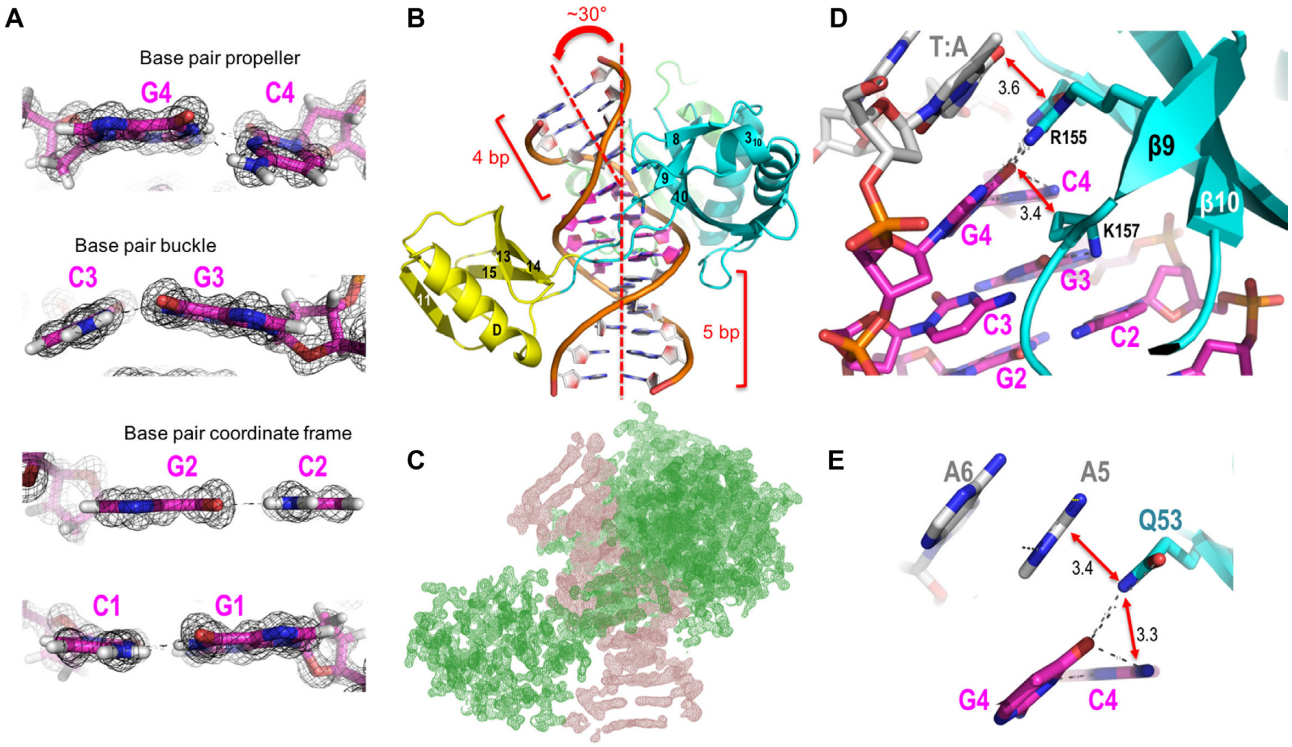

Figure 3. Movement of bases within the four G:C base pairs of the recognition sequence. (**A**) base pair propeller (G4:C4), buckle (C3:G3), in-frame (G2:C2) and a minor propeller twist (C1:G1). (**B**) Two long loops traverse the major groove, one following strand β9 and the other preceding strand β10, between β-structures in cyan and yellow. (**C**) As in the same orientation of panel B, electron densities 2Fo-Fc, contoured at 2σ above the mean, are shown respectively for the entire DNA (orange) and protein (green). (**D**) R155 and K157 approach DNA from major groove. (**E**) Q53 wedges into the cytosine C4 base and the base immediately outside of recognition sequence.

active-site residues D34 and E46, which normally coordinate the catalytic magnesium ion are unable to do so and instead form H-bonds with one of the non-bridging oxygen atoms of the target phosphate group between guanine G3 and cytosine C4 (Figure 4B). The active site K48—also protonated—forms an H-bond with an oxygen of the same phosphate. In addition, uncharged E5 and D205 stack with ends of neighboring DNA molecules, respectively (Supplementary Figure S1). These interactions are not observed in the higher pH condition (6.0), which resulted in a different space group (PDB ID: 6UKH and 6UKI).

In addition to phosphate interactions, some deoxyribose rings also interact with the protein. For example, the deoxyribose of cytosine C4 is co-planar with the indole ring of W3, with the hydrogens of the 2′ and 4′ carbons pointing directly toward the two π-holes (Figure 4C). The deoxyribose of guanine G1 might also form a weak C-H•••O type bond (2.7 Å between the hydrogen and the oxygen or 3.2 Å between the carbon and the oxygen) with the side chain oxygen atom of N74 (Figure 4D).

Base specific Interactions

Supporting the notion that HhaI behaves as a monomer (Supplementary Figure S2A and B), all four base pairs of the recognition sequence engage in major groove H-bonds with a single HhaI molecule (Figure 5A). Although the X-ray diffraction technique does not provide direct information on hydrogen atoms (42), our structures determined at ultra-high atomic resolution (PDB ID: 6UKF at 1.0 Å and PDB ID: 6UKG at 1.16 Å) coupled with the known chemical natures of interacting groups allow us to include hydrogen atoms in the refinement (colored in light gray in Figure 5). The presence of these hydrogen atoms enables us to describe the protein-base interactions in three ways: (i) conventional H-bonds between acceptor atoms and the hydrogens covalently attached to donors; (ii) C-H•••O type bonds (43) that occur frequently in biomolecular interactions but remain generally underappreciated; and (iii) water-mediated interactions. As an internal reference, the 12 Watson–Crick H-bonds between the four G:C base pairs of the recognition sequence display interatomic distances of 2.0–2.3 Å between the hydrogen and the acceptor atom. For example, base-pair G2:C2, which adopts a perfect in-frame configuration, displays H-bond lengths of 2.0–2.1 Å (Figure 5D), and base-pair G4:C4, which has a propeller twist of 35°, displays lengths of 2.0–2.3 Å (Figure 5B). In the following description, we used italics for atoms that are attached to a purine or pyrimidine ring atom but not part of the ring itself (e.g. guanine *N*2 and *O*6; cytosine *N*4). The number refers to the ring atom itself or the ring atom to which these are attached.

Two conventional H-bonds are present at the G4:C4 bp of the recognition sequence where the side chain of R155 donates bidentate H-bonds to the N7 (2.2 Å between N7…H) and *O*6 (2.0 Å between *O*6…H) atoms of guanine G4. Q53 donates an alternative H-bond to the *O*6 atom (2.3 Å between *O*6 and H; Figure 5B). These inter-

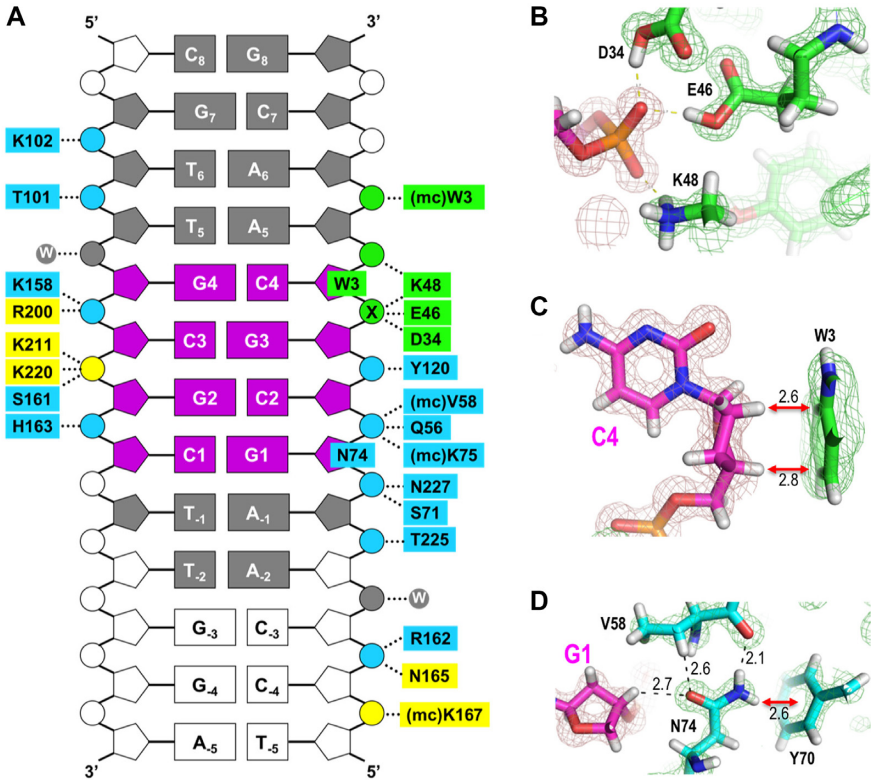

Figure 4. DNA phosphate interactions. (**A**) Schematic of HhaI-DNA backbone interactions: protein residues and their interacting phosphate groups are colored (green, cyan and yellow) accordingly to their locations in the structure (see Figure 1C); mc, main-chain-atom mediated contacts; w, water-mediated hydrogen bonds; X, the target phosphate. The four G:C base pairs of recognition are in magenta; outer sequence is in gray. For the DNA molecule, ten out of 13 bp (shaded) have holes in the electron density in the middle of the ring, including deoxyribose. (**B**) At pH 4.2 (in PDB ID: 6UKG), the active-site residues D34 and E46 are protonated and form H-bonds with one of the non-bridging oxygen atoms of the target phosphate group. (**C**) Contact between cytosine C4 deoxyribose and W3. (**D**) Contact between guanine G1 deoxyribose with N74.

atomic distances are comparable to those of Watson–Crick H-bonds. These Arg–Gua interactions are sufficient to unambiguously discriminate a G:C bp, and are almost identical to the interaction between R109 of the restriction enzyme SfiI and the first guanine of its recognition sequence (11), as well as to those in the tandem array of zinc fingers of human PRDM9 (44).

At the C3:G3 bp, the side chain amino group of K157 donates an H-bond to guanine *O*6 (2.0 Å between *O*6…H), and the main-chain carbonyl of K158 accepts an H-bond from the 4-amino group of cytosine (2.1 Å between *N*4-H…O; Figure 5C). (In addition, a water molecule forms an H-bond with the N7 atom of G3). In principal, these interactions do not discriminate a C:G bp from an A:T bp because a similar pattern of H-bonding can arise with both base pairs (i.e. adenine *N*6 in place of the cytosine *N*4, and thymine *O*4 in place of guanine *O*6). However, if A:T were encountered at this position instead of C:G, the interacting atoms would not be located at the exact same positions and the bulky 5-methyl group of thymine would likely clash with K157 or a water molecule, and thereby prevent binding.

The next two base pairs form both conventional H-bonds and C-H●●●O type bonds. G2:C2 engages in direct contacts in both the major and minor grooves, mainly with main-chain atoms (Figure 5D and E). One major-groove H-bond is present between the main chain nitrogen of S161 and guanine N7 (2.5 Å between N7…H-N). Two C-H●●●O type bonds also appear to be present with the guanine: one with *O*6 (2.4 Å) via the H-Cα atom of G160, and another with C8-H (2.7 Å) via the side chain oxygen atom of S161 (Figure 5D). The former is shorter (2.4 Å) and the latter longer (2.7 Å) than the conventional H-bond (2.5 Å) present at this base. We assume that shape-complementarity along the edge of guanine G2 contributes to base discrimination at this location. In the minor groove, one conventional H-bond and one C-H●●●O type bond are also present between the main-chain carbonyl oxygen of S28 and guanine *N*2 (2.1 Å) and between the side chain Cβ atom of S30 and cytosine *O2* (2.5 Å; Figure 5E).

In contrast to the other base pairs in the recognition sequence, the H-bonding capacity of base pair 1 is fully saturated in the major groove. C1:G1 interacts with S159, G160, and S161 (Figure 5F). The side chain hydroxyl of S159 donates an H-bond to guanine N7 (2.0 Å between N7…H-O); the main chain nitrogen of G160 donates another to guanine *O*6 (2.0 Å between *O*6…H-N); and the main-chain carbonyl of S161 accepts an H-bond from the 4-amino group of cytosine (2.2 Å between *N*4-H…O). In addition, a C-H●●●O type bond is formed between cytosine ring carbon C5-H and the main-chain carbonyl of S161 (2.5 Å). These interactions are sufficient to specify the base pair unambiguously.

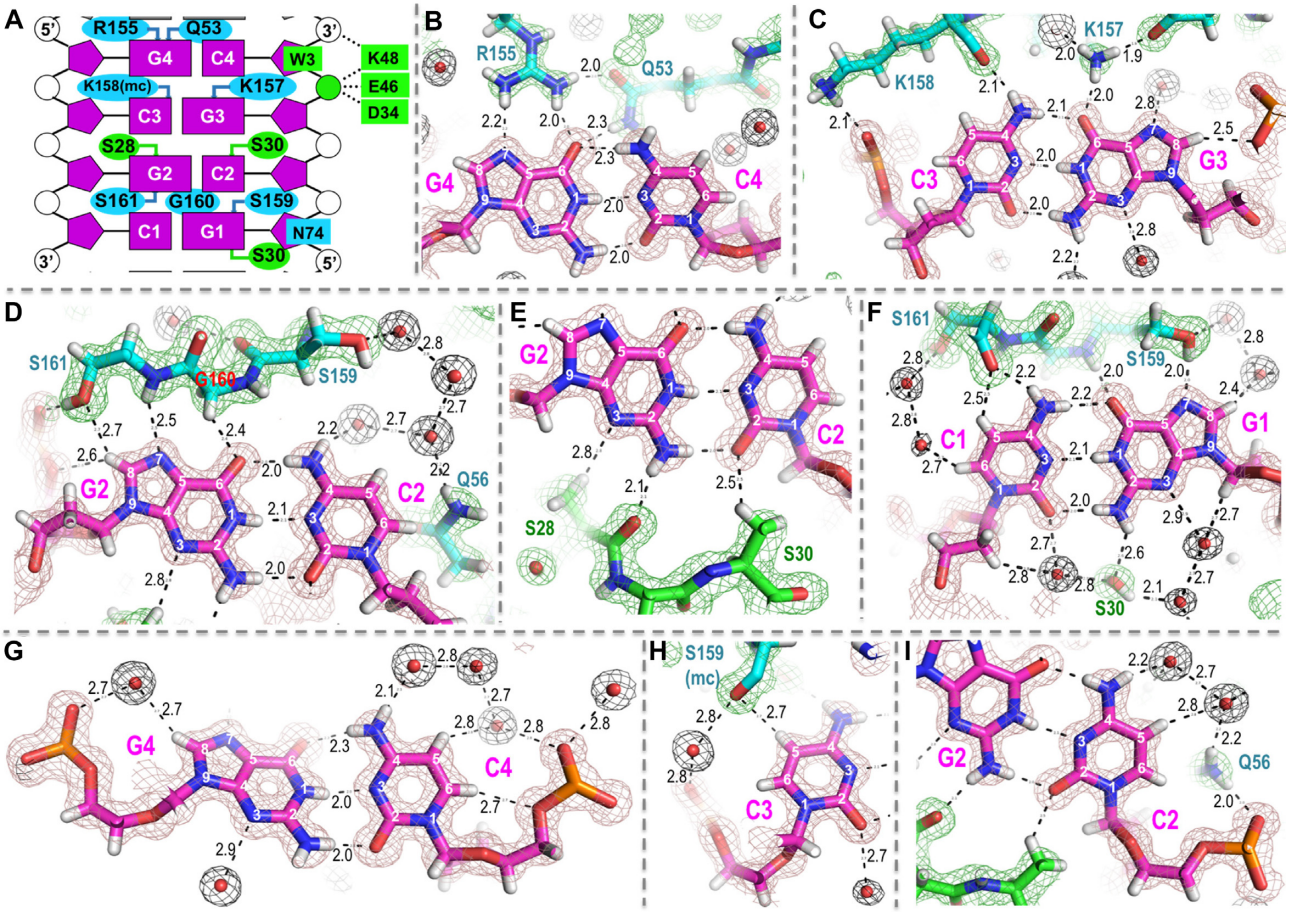

Figure 5. Base specific interaction of four G:C base pairs of recognition sequence. (**A**) Schematic of HhaI-DNA base interactions: residues are colored (green and cyan) according to their locations in the structure (Figure 1C); mc, main-chain-atom mediated contacts. The target phosphate group is colored in green. (**B**) Interactions with the G4:C4 base pair. Interatomic distances are shown in angstroms. Hydrogen atoms, light gray; nitrogen atoms, dark blue; oxygen atoms, red; carbon atoms in DNA, magenta; carbon atoms in protein, green or cyan. Water molecules are shown in red spheres. Electron densities 2Fo-Fc are contoured at 2σ above the mean. (**C**) Interactions with C3:G3 base pair. (**D** and **E**) Interactions with G2:C2 base pair in the major groove (panel D) and minor groove (panel E). (**F**) Interactions with C1:G1 base pair. (**G**–**I**) Examples of water-mediated interactions.

In addition to these direct protein-DNA base interactions, many of the polar atoms of the 8 nt are involved in water-mediated interactions in both DNA grooves. The exposed ring carbon atoms of the nucleotides (i.e. carbon-8 of guanine, and carbon-5 and -6 cytosine) are also shielded by a hydration layer of water molecules, which might form additional C-H•••O type H-bonds (Figure 5G–I) that might contribute to sequence-discrimination. A complex overall network of water molecules interconnects bases (particularly cytosines), phosphate groups and amino acids (Figures 5G–I). In the cases of Cyt2 and Cyt3, S159 and Q56 function as one of the water molecules in a water network surrounding the nonpolar carbon atoms (Figure 5H and I).

Comparison with HinP1I endonuclease and HhaI methyltransferase

We compared base-recognition by HhaI endonuclease to that of HinP1I endonuclease and the M.HhaI methyltransferase, since all three enzymes bind as monomers to the same 4-bp sequence, 5′-GCGC-3′ (Figure 6). Superimposing the active sites of HhaI and HinP1I overlapped the GpC dinucleotides that straddle the target phosphate group (G1-p-C2 in HinP1I and G3-p-C4 in HhaI) (Figure 6A and C) and revealed some similarities in recognition patterns. Recognition of guanine G1 by K96 of HinP1I is equivalent to recognition of G3 by HhaI K157. Likewise, recognition of base pair G2:C2 by K223 and Q93 of HinP1I is equivalent to recognition of G4:C4 by HhaI R155 and Q53 (Figure 6B and D). Interestingly, a similar Arg-Gln pair is also seen in M.HhaI where R240 and Q237 combine to recognize the outermost G1:C1 base pair, next to the target cytosine (Figure 6E and F). Aside from these elements of similarity, the three enzymes use different overall contact strategies, indicating that multiple ways exist to recognize DNA sequences accurately.

The HhaI catalytic site at pH 6.0

As seen previously with PvuII-DNA cocrystals grown at low pH (45), the HhaI-DNA crystals grown in the presence of calcium at pH 4.2 did not contain divalent metal ions at the catalytic site. D34 and E46 appear to be unionized (uncharged) at this pH (Figure 4B), and thus unable

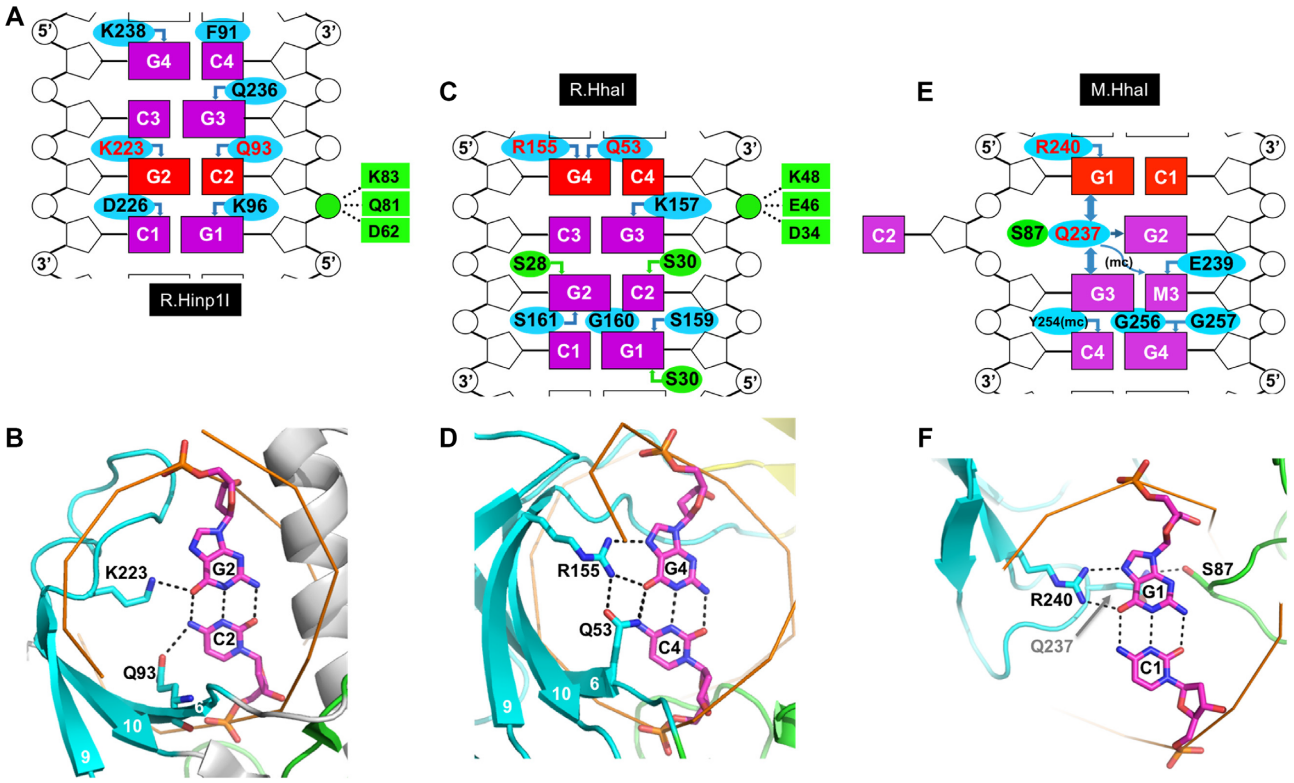

Figure 6. Comparison of three enzymes recognizing GCGC sequence. (**A** and **B**) HinP1I, (**C** and **D**) HhaI and (**E** and **F**) HhaI methyltransferase.

to bind cations. To visualize the catalytic site in the presence of metal ions, we examined two structures at pH 6 (PDB ID: 6UKH in $P4_12_12$ and PDB ID: 6UKI in $P2_12_12_1$). PDB ID: 6UKH contains one protein–DNA complex, whereas PDB ID: 6UKI contains two such complexes. All three complexes contain two Ca(II) ions in the active site (Figure 7A). The two metal ions (M1 and M2) are bridged by carboxylate oxygen atoms of D34 and one of the non-bridging oxygen atoms of the target phosphate group (Figure 7B). The calcium ion bound at M1 has octahedral geometry and is coordinated by six oxygen atoms: one each from the side chain carboxyl oxygens of D34 and E46, the OP1 oxygen of the target phosphate, the backbone carbonyl of A47, and two water molecules (Figure 7C). One of the water molecules (w1) is poised to act as the attacking nucleophilic hydroxide that initiates strand cleavage. It contacts the metal ion at M1, H-bonds to the side-chain amino group of K48 (the conserved general base), and bridges between two neighboring phosphate oxygen atoms (Figure 7D). The second metal ion appears to be coordinated incompletely by only four oxygen atoms: the same OP1 of the target phosphate, the other carboxylate oxygen atoms of D34, the 3′-leaving oxygen of guanine G3 and a water molecule (w3), which in turn H-bonds with side chain of E7 (not shown). In sum, the HhaI active-site geometry in the presence of DNA at pH 6.0 supports a two-metal mechanism of DNA cleavage and is consistent with HhaI cleaving DNA one strand at a time (Supplementary Figure S2C). Under the laboratory conditions, the earliest product (5 s) was a nicked open circle intermediate and a linear product. The amount of nicked intermediate was accumulated before being converted to a linear product, consistent with HhaI cleaving DNA one strand at a time. However, the appearance of the linear product was earlier than one would expect from a completely random nicking reaction, considering that pUC19 DNA contains 17 HhaI sites. This might suggest that the likelihood of strand-hydrolysis at any HhaI site increases greatly once the other strand is already hydrolyzed.

DISCUSSION

We show here that HhaI is a monomer that recognizes its palindromic DNA sequence in an asymmetric manner. The single active site belongs to the 'PD-D/EXK' superfamily of nucleases, contains the motif SD-X_{11}-EAK, and supports the two-metal mechanism of DNA cleavage. The structure at ultra-high atomic resolution of 1.0 Å gave us the opportunity to analyze the protein–DNA interactions in detail. The contacts at the base pairs flanking the target phosphate group result in DNA distortions including bending, propeller twist and base-pair buckle. The four G:C base pairs of the recognition sequence are recognized by traditional H-bonds, C-H•••O type bonds, a water-mediated hydration layer surrounding nonpolar atoms, and by van der Waals interactions, together ensuring high enzymatic specificity. Almost all ring atoms of the four cytosine and four guanine bases of the recognition sequence are in close contact with protein residues, many of which are main-chain atoms of small amino acids (serine, alanine and glycine). This might enhance intimate contact between the enzyme and its substrate.

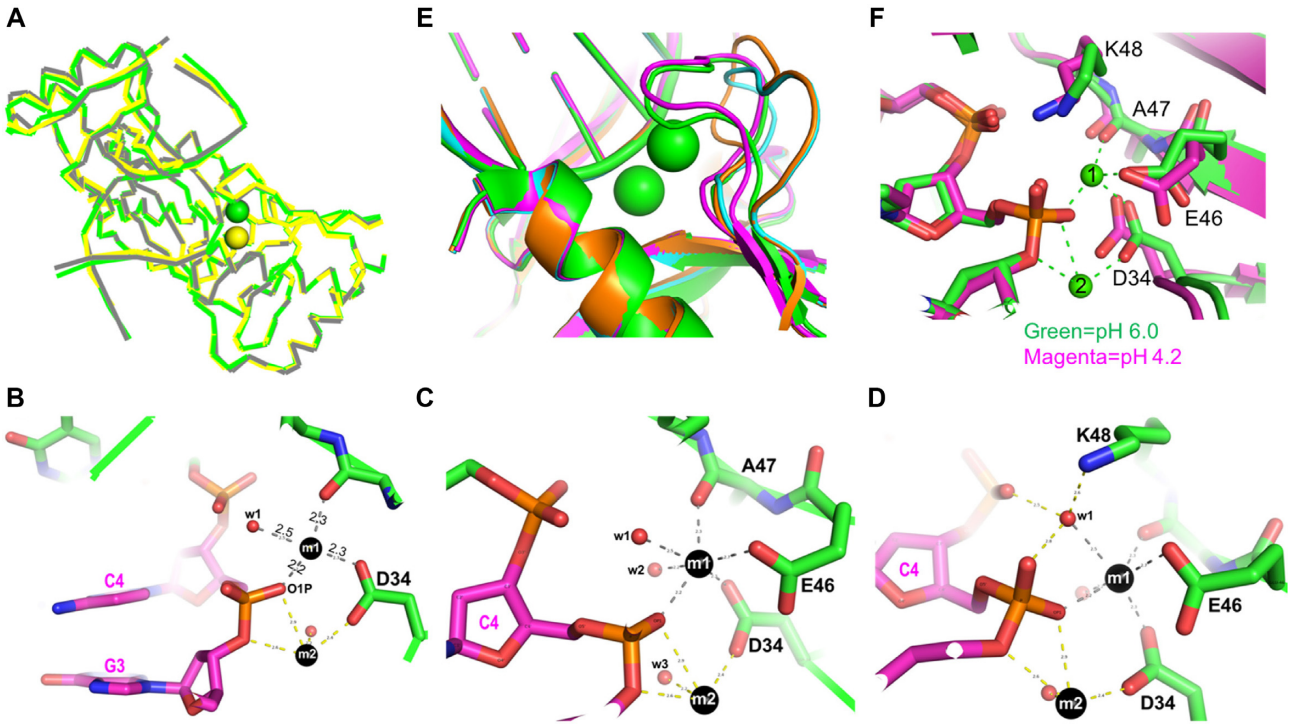

Figure 7. The two-metal binding site in HhaI. (**A**) Three complexes at pH 6.0 (yellow in 6UKH and green and gray in PDB ID: 6UKI). The two Ca(II) ions are shown as spheres. (**B**) The two Ca(II) binding sites (m1 and m2) are bridged by the carboxylate oxygen atoms of D34 and one of the non-bridging oxygen atoms of the target phosphate group. (**C**) The metal at binding site m1 has octahedral coordination, and (**D**) the water at binding site w1 has tetrahedral coordination. (**E**) comparison of the active-site loop at pH 6.0 (green) and three structures at pH 4.2 (6UKG in magenta, 6UKE in orange and 6UKF in cyan). (**F**) Superimposition of active-site residues at pH 6.0 (green) and at pH 4.2 (magenta).

HhaI distorts the substrate DNA upon binding with the overall effect of better aligning the bases with their contact amino acids. This is especially noticeable for guanine G4 which is twisted ∼30° to better H-bond with R155. The deformation is promoted by the side-chains of K157, which intrudes into the helix from below the guanine, and Q53 which intrudes from above cytosine C4, who pairs with G4 (Figure 3D). Comparable deformations are seen with other restriction enzymes such as EcoRI, where A118 from each subunit of EcoRI dimer wedges into the major groove at the center of the sequence GAATTC, and in EcoRV where K38 from each subunit of EcoRV dimer intrudes into the minor groove in the center of the sequence GATATC (4,5).

Interestingly, we noticed a conformational difference in the active-site loop among the three $P2_1$ structures crystallized at pH 4.2 (where divalent metal ions were included in the crystallization, none were present at the catalytic site): PDB ID: 6UKG had the same conformation as the active-site loop at pH 6.0, whereas PDB ID: 6UKE and 6UKF adopted a different and less ordered conformation (Figure 7E and F). The rest of the protein–DNA complexes were closely similar, with pairwise comparison of root-mean-square deviations of <0.5 Å. The observation suggested the enzyme binds the cognate substrate and the active-site loop samples different conformations prior to metal binding for catalysis.

An arginine-glutamine pair cooperates in recognition of the outer G:C base pair next to the catalytic site in both the HhaI endonuclease and its biological methyltransferase partner, M.HhaI. In other respects, the ways in which the two enzymes recognize the sequence differs. HhaI and M.HhaI are interrelated biologically in that, together, they give rise to protective immunity from virus infections. But the reactions they catalyze are entirely unrelated biochemically, and aside from targeting the same substrate sequence, the two enzymes have little else in common: HhaI hydrolyzes a phosphodiester bond in regular, backbone component of DNA, while M.HhaI methylates a cytosine within the otherwise irregular, coding component. The methyl group introduced by M.HhaI at the internal cytosine of GCGC (underlined) would disrupt the interaction with K158 of HhaI and the water network connecting Q56 and S159 (C3 in Figure 5C and C2 in Figure 5D), resulting in protection of methylated DNA from cleavage.

DATA AVAILABILITY

The complete genome sequence of *H. haemolyticus* NEB129 (called *H. parahaemolyticus* 129 by NCBI) is available in GenBank under the accession numbers (CP038817 and CP038818). Original sequence reads have been deposited at NCBI under SRA (SRR8945295 and SRR8945296) and, together with relevant PacBio data, Bioproject (PRJNA530980). The X-ray structure (coordinates) and the source data (structure factor files) of HhaI with bound DNA have been deposited to PDB under accession numbers: 6UKE (5-iodouridine), 6UKF ($P2_1$ at 0.99 Å), 6UKG ($P2_1$ at 1.16 Å), 6UKH ($P4_12_12$),

6UKI ($P2_12_12_1$). The diffraction images for 6UKE, 6UKF and 6UKG have been deposited to Integrated Resource for Reproducibility in Macromolecular Crystallography (https://proteindiffraction.org/) (46). The images are archived with the PDB IDs. The diffraction images were also deposited in SBGrid (doi:10.15785/SBGRID/743 for PDB ID: 6UKF; doi:10.15785/SBGRID/744 for PDB ID: 6UKG; doi:10.15785/SBGRID/745 for PDB ID: 6UKE).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGMENTS

We thank B. Baker of New England Biolabs for synthesizing the oligonucleotides containing 5-iodouridine.
Authors contributions: J.R.H. performed X-ray data collection and structure determination. J.Y. performed crystallization and DNA cleavage experiment under supervision of X.Z. T.P. and A.F. performed genome sequencing. R.J.R. initiated the study. G.G.W. performed data analysis. X.C. R.J.R. and G.G.W. prepared the manuscript.

FUNDING

U.S. National Institutes of Health (NIH) [GM049245–24 to X.C.]; Cancer Prevention and Research Institute of Texas [RR160029 to X.C. who is a CPRIT Scholar in Cancer Research]; New England Biolabs. The open access publication charge for this paper has been waived by Oxford University Press–*NAR* Editorial Board members are entitled to one free paper per year in recognition of their work on behalf of the journal.
Conflict of interest statement. T.P., A.F., G.G.W. and R.J.R. are employees of New England Biolabs, a company that sells DNA research reagents including restriction enzymes.

REFERENCES

1. Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2015) REBASE–a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.*, **43**, D298–D299.
2. Pingoud,A., Wilson,G.G. and Wende,W. (2014) Type II restriction endonucleases–a historical perspective and more. *Nucleic Acids Res.*, **42**, 7489–7527.
3. Roberts,R.J. (2005) How restriction enzymes became the workhorses of molecular biology. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 5905–5908.
4. Kim,Y.C., Grable,J.C., Love,R., Greene,P.J. and Rosenberg,J.M. (1990) Refinement of Eco RI endonuclease crystal structure: a revised protein chain tracing. *Science*, **249**, 1307–1309.
5. Winkler,F.K., Banner,D.W., Oefner,C., Tsernoglou,D., Brown,R.S., Heathman,S.P., Bryan,R.K., Martin,P.D., Petratos,K. and Wilson,K.S. (1993) The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.*, **12**, 1781–1795.
6. Newman,M., Strzelecka,T., Dorner,L.F., Schildkraut,I. and Aggarwal,A.K. (1994) Structure of restriction endonuclease BamHI and its relationship to EcoRI. *Nature*, **368**, 660–664.
7. Cheng,X., Balendiran,K., Schildkraut,I. and Anderson,J.E. (1994) Structure of PvuII endonuclease with cognate DNA. *EMBO J.*, **13**, 3927–3935.
8. Newman,M., Lunnen,K., Wilson,G., Greci,J., Schildkraut,I. and Phillips,S.E. (1998) Crystal structure of restriction endonuclease BglI bound to its interrupted DNA recognition sequence. *EMBO J.*, **17**, 5466–5476.
9. Bozic,D., Grazulis,S., Siksnys,V. and Huber,R. (1996) Crystal structure of Citrobacter freundii restriction endonuclease Cfr10I at 2.15 A resolution. *J. Mol. Biol.*, **255**, 176–186.
10. Deibert,M., Grazulis,S., Sasnauskas,G., Siksnys,V. and Huber,R. (2000) Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. *Nat. Struct. Biol.*, **7**, 792–799.
11. Vanamee,E.S., Viadiu,H., Kucera,R., Dorner,L., Picone,S., Schildkraut,I. and Aggarwal,A.K. (2005) A view of consecutive binding events from structures of tetrameric endonuclease SfiI bound to DNA. *EMBO J.*, **24**, 4198–4208.
12. Lyumkis,D., Talley,H., Stewart,A., Shah,S., Park,C.K., Tama,F., Potter,C.S., Carragher,B. and Horton,N.C. (2013) Allosteric regulation of DNA cleavage and sequence-specificity through run-on oligomerization. *Structure*, **21**, 1848–1858.
13. 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.
14. 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.
15. Xu,Q.S., Roberts,R.J. and Guo,H.C. (2005) Two crystal forms of the restriction enzyme MspI-DNA complex show the same novel structure. *Protein Sci.*, **14**, 2590–2600.
16. Yang,Z., Horton,J.R., Maunus,R., Wilson,G.G., Roberts,R.J. and Cheng,X. (2005) Structure of HinP1I endonuclease reveals a striking similarity to the monomeric restriction enzyme MspI. *Nucleic Acids Res.*, **33**, 1892–1901.
17. Horton,J.R., Zhang,X., Maunus,R., Yang,Z., Wilson,G.G., Roberts,R.J. and Cheng,X. (2006) DNA nicking by HinP1I endonuclease: bending, base flipping and minor groove expansion. *Nucleic Acids Res.*, **34**, 939–948.
18. Kaus-Drobek,M., Czapinska,H., Sokolowska,M., Tamulaitis,G., Szczepanowski,R.H., Urbanke,C., Siksnys,V. and Bochtler,M. (2007) Restriction endonuclease MvaI is a monomer that recognizes its target sequence asymmetrically. *Nucleic Acids Res.*, **35**, 2035–2046.
19. Sokolowska,M., Kaus-Drobek,M., Czapinska,H., Tamulaitis,G., Szczepanowski,R.H., Urbanke,C., Siksnys,V. and Bochtler,M. (2007) Monomeric restriction endonuclease BcnI in the apo form and in an asymmetric complex with target DNA. *J. Mol. Biol.*, **369**, 722–734.
20. Sasnauskas,G., Kostiuk,G., Tamulaitis,G. and Siksnys,V. (2011) Target site cleavage by the monomeric restriction enzyme BcnI requires translocation to a random DNA sequence and a switch in enzyme orientation. *Nucleic Acids Res.*, **39**, 8844–8856.
21. Kostiuk,G., Sasnauskas,G., Tamulaitiene,G. and Siksnys,V. (2011) Degenerate sequence recognition by the monomeric restriction enzyme: single mutation converts BcnI into a strand-specific nicking endonuclease. *Nucleic Acids Res.*, **39**, 3744–3753.
22. Roberts,R.J., Myers,P.A., Morrison,A. and Murray,K. (1976) A specific endonuclease from Haemophilus haemolyticus. *J. Mol. Biol.*, **103**, 199–208.
23. Mann,M.B., Rao,R.N. and Smith,H.O. (1978) Cloning of restriction and modification genes in E. coli: the HhaII system from Haemophilus haemolyticus. *Gene*, **3**, 97–112.
24. Caserta,M., Zacharias,W., Nwankwo,D., Wilson,G.G. and Wells,R.D. (1987) Cloning, sequencing, in vivo promoter mapping, and expression in Escherichia coli of the gene for the HhaI methyltransferase. *J. Biol. Chem.*, **262**, 4770–4777.
25. Barsomian,J.M., Card,C.O. and Wilson,G.G. (1988) Cloning of the HhaI and HinPI restriction-modification systems. *Gene*, **74**, 5–7.
26. Bertani,G. (1951) Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. *J. Bacteriol.*, **62**, 293–300.
27. Chin,C.S., Alexander,D.H., Marks,P., Klammer,A.A., Drake,J., Heiner,C., Clum,A., Copeland,A., Huddleston,J., Eichler,E.E. *et al.* (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods*, **10**, 563–569.
28. Tatusova,T., DiCuccio,M., Badretdin,A., Chetvernin,V., Nawrocki,E.P., Zaslavsky,L., Lomsadze,A., Pruitt,K.D., Borodovsky,M. and Ostell,J. (2016) NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.*, **44**, 6614–6624.
29. Haft,D.H., DiCuccio,M., Badretdin,A., Brover,V., Chetvernin,V., O'Neill,K., Li,W., Chitsaz,F., Derbyshire,M.K., Gonzales,N.R. *et al.*

(2018) RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res.*, **46**, D851–D860.
30. Evans,P.R. and Murshudov,G.N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. D. Biol. Crystallogr.*, **69**, 1204–1214.
31. Terwilliger,T.C., Adams,P.D., Read,R.J., McCoy,A.J., Moriarty,N.W., Grosse-Kunstleve,R.W., Afonine,P.V., Zwart,P.H. and Hung,L.W. (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr. D. Biol. Crystallogr.*, **65**, 582–601.
32. Afonine,P.V., Grosse-Kunstleve,R.W., Echols,N., Headd,J.J., Moriarty,N.W., Mustyakimov,M., Terwilliger,T.C., Urzhumtsev,A., Zwart,P.H. and Adams,P.D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D. Biol. Crystallogr.*, **68**, 352–367.
33. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.*, **66**, 486–501.
34. 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. Crystallogr.*, **40**, 658–674.
35. Kelley,L.A., Mezulis,S., Yates,C.M., Wass,M.N. and Sternberg,M.J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, **10**, 845–858.
36. Flusberg,B.A., Webster,D.R., Lee,J.H., Travers,K.J., Olivares,E.C., Clark,T.A., Korlach,J. and Turner,S.W. (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods*, **7**, 461–465.
37. Clark,T.A., Murray,I.A., Morgan,R.D., Kislyuk,A.O., Spittle,K.E., Boitano,M., Fomenkov,A., Roberts,R.J. and Korlach,J. (2012) Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res.*, **40**, e29.
38. Sasnauskas,G., Tamulaitiene,G., Tamulaitis,G., Calyseva,J., Laime,M., Rimseliene,R., Lubys,A. and Siksnys,V. (2017) UbaLAI is a monomeric Type IIE restriction enzyme. *Nucleic Acids Res.*, **45**, 9583–9594.
39. Siwek,W., Czapinska,H., Bochtler,M., Bujnicki,J.M. and Skowronek,K. (2012) Crystal structure and mechanism of action of the N6-methyladenine-dependent type IIM restriction endonuclease R.DpnI. *Nucleic Acids Res.*, **40**, 7563–7572.
40. Mierzejewska,K., Siwek,W., Czapinska,H., Kaus-Drobek,M., Radlinska,M., Skowronek,K., Bujnicki,J.M., Dadlez,M. and Bochtler,M. (2014) Structural basis of the methylation specificity of R.DpnI. *Nucleic Acids Res.*, **42**, 8745–8754.
41. Friedhoff,P., Lurz,R., Luder,G. and Pingoud,A. (2001) Sau3AI, a monomeric type II restriction endonuclease that dimerizes on the DNA and thereby induces DNA loops. *J. Biol. Chem.*, **276**, 23581–23588.
42. McCusker,L.B. (2017) Electron diffraction and the hydrogen atom. *Science*, **355**, 136.
43. Horowitz,S. and Trievel,R.C. (2012) Carbon-oxygen hydrogen bonding in biological structure and function. *J. Biol. Chem.*, **287**, 41576–41582.
44. Patel,A., Horton,J.R., Wilson,G.G., Zhang,X. and Cheng,X. (2016) Structural basis for human PRDM9 action at recombination hot spots. *Genes Dev.*, **30**, 257–265.
45. Horton,J.R. and Cheng,X. (2000) PvuII endonuclease contains two calcium ions in active sites. *J. Mol. Biol.*, **300**, 1049–1056.
46. Grabowski,M., Langner,K.M., Cymborowski,M., Porebski,P.J., Sroka,P., Zheng,H., Cooper,D.R., Zimmerman,M.D., Elsliger,M.A., Burley,S.K. *et al.* (2016) A public database of macromolecular diffraction experiments. *Acta Crystallogr. D. Struct. Biol.*, **72**, 1181–1193.