



Crystal structure of a novel domain of the motor subunit of the Type I restriction enzyme EcoR124 involved in complex assembly and DNA binding

Received for publication, May 16, 2018, and in revised form, July 19, 2018. Published, Papers in Press, July 27, 2018, DOI 10.1074/jbc.RA118.003978

Pavel Grinkevich^{‡§}, Dhiraj Sinha[‡], Iuliia Iermak^{‡¶}, Alena Guzanova^{||}, Marie Weiserova^{||}, Jost Ludwig[‡], Jeroen R. Mesters^{**}, and Rüdiger H. Etrich^{‡††1}

From the [‡]Center for Nanobiology and Structural Biology, Institute of Microbiology, Academy of Sciences of the Czech Republic, Zamek 136, 373 33 Nove Hradky, Czech Republic, the [§]Faculty of Sciences, University of South Bohemia in Ceske Budejovice, Branišovská 1760, 370 05 České Budějovice, Czech Republic, the [¶]Department of Structural Cell Biology, Molecular Mechanisms of DNA Repair, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany, the ^{||}Institute of Microbiology, Academy of Sciences of the Czech Republic, Vídeňská 1083, 142 20 Praha 4, Czech Republic, the ^{**}Institute of Biochemistry, University of Lübeck, Ratzeburger Allee 160, Lübeck, Germany, and the ^{††}College of Biomedical Sciences, Larkin University, Miami, Florida 33169

Edited by Joseph M. Jez

Although EcoR124 is one of the better-studied Type I restriction-modification enzymes, it still presents many challenges to detailed analyses because of its structural and functional complexity and missing structural information. In all available structures of its motor subunit HsdR, responsible for DNA translocation and cleavage, a large part of the HsdR C terminus remains unresolved. The crystal structure of the C terminus of HsdR, obtained with a crystallization chaperone in the form of pHluorin fusion and refined to 2.45 Å, revealed that this part of the protein forms an independent domain with its own hydrophobic core and displays a unique α -helical fold. The full-length HsdR model, based on the WT structure and the C-terminal domain determined here, disclosed a proposed DNA-binding groove lined by positively charged residues. *In vivo* and *in vitro* assays with a C-terminal deletion mutant of HsdR supported the idea that this domain is involved in complex assembly and DNA binding. Conserved residues identified through sequence analysis of the C-terminal domain may play a key role in protein–protein and protein–DNA interactions. We conclude that the motor subunit of EcoR124 comprises five structural and functional domains, with the fifth, the C-terminal domain, revealing a unique fold characterized by four conserved motifs in the IC subfamily of Type I restriction-modification systems. In summary, the structural and biochemical results reported here support a model in which the C-terminal domain of the motor subunit HsdR of the endonuclease EcoR124 is involved in complex assembly and DNA binding.

This work was supported by the Czech Science Foundation (P207/12/2323) and by a Memorandum of Agreement between the Institute of Microbiology, Czech Academy of Sciences, and the College of Biomedical Sciences, Larkin University. Access to instruments and other facilities was supported by the Czech Research Infrastructure for Systems Biology C4SYS (LM2015055). The authors declare that they have no conflicts of interest with the contents of this article.

This article contains Model S1.

The atomic coordinates and structure factors (codes 5J3N and 6H2J) have been deposited in the Protein Data Bank (<http://www.pdb.org/>).

¹ To whom correspondence should be addressed. Tel./Fax: 420-386-361-297; E-mail: etrich@nh.cas.cz.

Restriction (R)² enzymes, or restriction endonucleases, are a diverse group of site-specific DNA cutters, acting in tandem with methyltransferases (MTases) capable of DNA modification (M) through nucleotide methylation. Tens of thousands of putative restriction endonucleases and MTases have been identified throughout bacterial and archaeal kingdoms, and thousands of restriction-modification (R-M) proteins have been biochemically or genetically described (1). The staggering diversity of R-M enzymes is attributed to their role as a prokaryotic “immune system” (2); these proteins are able to recognize specific sequences in DNA, read out their methylation state, and act accordingly, either by cutting unmethylated DNA (“foreign DNA,” *e.g.* bacteriophages) or by methylating the daughter DNA strand generated during the preceding replication event, thereby protecting the cell’s own DNA from degradation. R-M enzymes have also been implicated in horizontal gene transfer (3), differential gene expression through epigenetic modification (4, 5), and driving divergent evolution (6, 7). The practical impact of R-M enzymes on genetic engineering and DNA fingerprinting can hardly be overstated (8).

R-M enzymes are historically divided into four types (I–IV) based on their subunit composition and biochemical, genetic, and other properties (1, 9). EcoR124 is one of the best-described Type I R-M enzymes (10) along with EcoAI (11), EcoBI (12), and EcoKI (13). Two key features of the Type I enzymes are their multimeric composition (one specificity (HsdS, S) and two modification subunits (HsdM, M) form the trimeric MTase (M₂S) that can be supplemented by two restriction subunits (HsdR, R) to complete the pentameric holoenzyme (R₂M₂S)) and their ability to translocate DNA before cleaving it at a distant random site (14). Both structural and functional complexity are implied by these two characteristics and have indeed become apparent in decades of research. Type I enzymes split into five families, A–E, based on complementation tests, anti-

² The abbreviations used are: R, restriction; MTase, methyltransferase; M, modification; EOP, efficiency of plating; EMSA, electrophoretic mobility shift assay.

Novel domain of the motor subunit of Type I R-M system

body cross-reactivity, and sequence similarity (8). EcoR124 belongs to the Type IC family, along with EcoPrnI, EcoDXXI, and NgoAV (8).

To study DNA recognition, methylation, translocation, and cleavage, attempts have been made to obtain structures of the individual Type I R-M subunits and their assemblies using X-ray crystallography and EM techniques. These attempts have been met with varying success: a low-resolution structure (~ 18 Å) of MTase.EcoKI (PDB codes 2y7c and 2y7h (15)) and the 3.2-Å DNA-bound MTase from *Thermoanaerobacter tengcongensis* were solved (PDB code 5YBB (16)) as well as unbound HsdM (PDB codes 2AR0, 3KHK, 2LKD, and 3UFB (17)), HsdS (PDB codes 1YF2 (18), 1YDX (19), and 3OKG (20)), and HsdR subunits (PDB codes 2W00 (21) and 3H1T (22)) from several species of microorganisms. However, the structural information available to date is incomplete and fragmented; structures of the subunits often originate from different R-M systems or even species, and some are only partially resolved.

The first structure of HsdR was obtained from EcoR124 in 2009 (PDB code 2W00 (21)), along with four single-point mutants in later years (PDB codes 4BE7, 4B4B, and 4BEC (23) and 4XJX). HsdR was described as having four domains (endonuclease, two RecA-like helicases, and one helical domain); however, a large stretch of the C terminus (residues 893–1038) was not resolved and said to be disordered (21). It was initially suggested that these amino acids add to the fourth helical domain. As demonstrated for a related Type I R-M enzyme, EcoKI, in partial proteolysis experiments, cleaving off 155 terminal residues of HsdR produced a fragment stable in the presence of ATP but unable to bind to MTase (24). *De novo* modeling of HsdR from EcoR124 supported the notion of the C terminus actively participating in complex formation but revealed little about its structural organization except for a relatively short coiled-coil fragment and an overall disordered nature (25). Another known crystal structure of HsdR from *Vibrio vulnificus* also lacks ~ 220 C-terminal residues and folds into three domains (endonuclease and two RecA-like helicases), providing no further insight into the 3D structure of the C terminus. A structure of the full EcoR124I complex was derived using EM and small-angle scattering (neutron and X-ray) (26). However, the ring-like shape together with a resolution of only 21 Å unfortunately does not allow orientation of HsdR in the EM map unambiguously. Additionally, the HsdM subunit used for fitting into the EM map was a low-resolution model based on a distant homolog with missing parts modeled *de novo*. Although this model appropriately describes the open and closed conformations of the full pentameric complex, it is not suitable to explore the HsdM–HsdR interface. In the absence of a crystal structure, it has become apparent that atomic-level structural information could shed first light on the way HsdR subunits attach to the MTase to form a holoenzyme capable of DNA translocation and cleavage.

Bioinformatic analysis of HsdR is hindered because of a low sequence identity of ~ 23 – 27% for the classified HsdR subunits between families of Type I R-M enzymes (8). Within the Type IC family, full sequences are only available for EcoR124 and NgoAV from *Neisseria gonorrhoeae* (27) that display $\sim 75\%$ sequence identity. An alignment between members of different

families was nevertheless achieved in endonuclease and helicase I and II domains because of the presence of clearly defined motifs (25, 21). The helical and C-terminal domains exhibit an even lower sequence identity than the rest of the protein, thereby prohibiting a direct identification of conserved regions through multiple pairwise alignment. So far, no motifs or folds have been identified for these two domains, although they have been said to belong to an α -helical class (21).

The obtained crystal structure of the C terminus of HsdR from EcoR124 in the form of a pHluorin fusion protein reveals that the C terminus stands as a separate domain featuring a helical bundle with its own hydrophobic core and a unique fold. The full-length HsdR model based on a re-refinement of the original WT crystal structure proposes that the C-terminal domain, attached by a short linker region, points outward from the four-domain arrangement and thereby clearly breaks the planar nature of the HsdR subunit of EcoR124. Sequence analysis using putative HsdR sequence data allowed a tentative assignment of conserved amino acid sequences in the C-terminal domain. Combined with the results of biochemical and physiological experiments, the model augments the hypothesis of the involvement of the C-terminal domain in complex assembly and DNA binding.

Results and discussion

Structure determination and overall architecture of the C-terminal domain

The vector design, purification, crystallization, and X-ray data analysis for pHluorin-HsdR887 fusion protein were reported previously (28). Initial molecular replacement revealed two pHluorin molecules in the asymmetric unit. Using the Buccaneer pipeline in CCP4, 74% of all amino acids were built in eight fragments. After combining the fragments and manual model building, the C-terminal domain was resolved for one of the two molecules in the asymmetric unit. A pore of ~ 42 -Å diameter with almost no electron density delineates the missing residues from the second molecule in the asymmetric unit (Fig. 1). In total, 382 residues were built in one molecule and 228 in the other, along with 198 solvent molecules; the structure was refined to a final resolution of 2.45 Å (see Table 1 for refinement statistics).

The following description applies to molecule A of the asymmetric unit, as it contains the resolved HsdR domain. Residues 8–244 correspond to the pHluorin sequence and residues 247–398 to the C terminus of HsdR (residues 887–1038 in “HsdR coordinates”). Electron density was missing for residues 263–267 and therefore modeled as a loop in the YASARA software. Residues with missing or poor electron density for the side chains were modeled to $C\beta$. Two histidines from the His₆ tag, 6 and 7, are also present in molecule A, along with a short linker of Glu²⁴⁵-Phe²⁴⁶ between the pHluorin and HsdR sequences. Linker design is a trial and error process (29), and, as a first possibility, a rigid linker of two amino acids was chosen to reduce conformational heterogeneity and promote interactions between the C-terminal domain and pHluorin. This first trial proved successful.

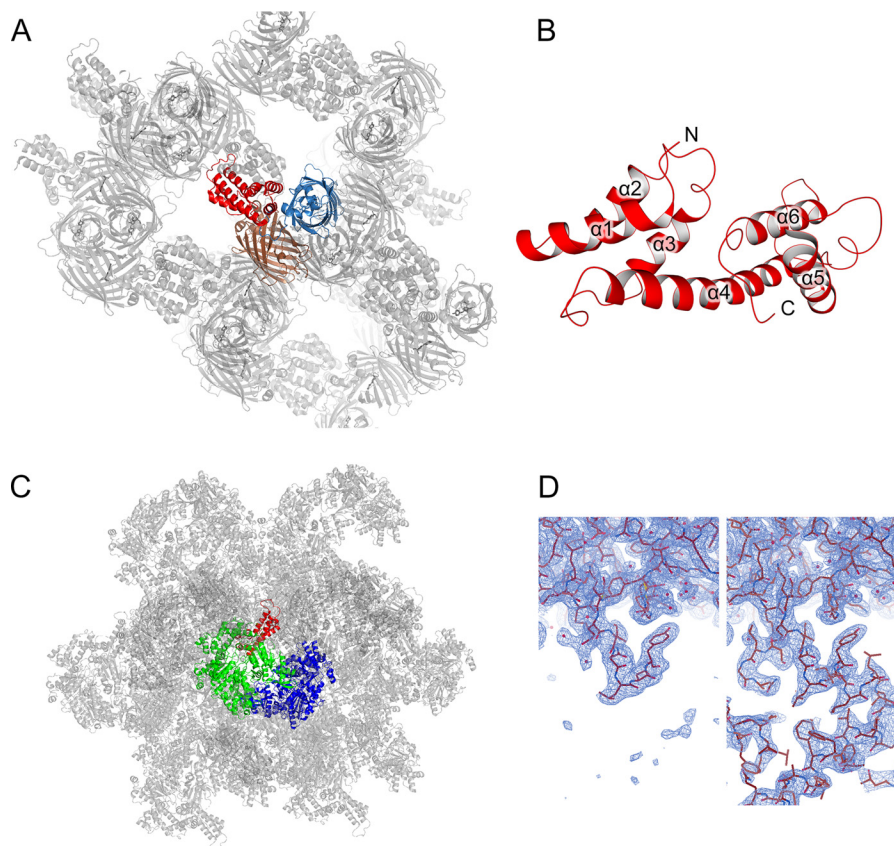


Figure 1. Features of the C-terminal domain. A, crystal packing of the pHLuorin-HsdR887 fusion protein with two pHLuorin molecules (molecule A, brown; molecule B, blue) and the C-terminal domain of HsdR belonging to molecule A (red). Symmetric molecules are displayed in gray. B, overall structure of the C-terminal domain of HsdR consisting of α -helices 1–6 numbered from the N to the C terminus. C, crystal packing of the WT HsdR based on re-refinement of PDB entry 2W00 (molecule A, green; molecule B, blue; C-terminal domain, red). D, sample electron density of WT HsdR before (left panel) and after (right panel) re-refinement of the full structure with the C-terminal domain. The figure was prepared in PyMOL (A and C), YASARA (B), and Coot (D).

Table 1

Data collection and refinement statistics for pHLuorin-HsdR887 and re-refinement statistics for the extended HsdR structure

	pHLuorin-HsdR887 (PDB code 5JN3)	Extended HsdR with C-terminus (PDB code 6H2J)
Data collection		
Space group	C222 ₁	P2 ₁ 2 ₁ 2 ₁
<i>a</i> , <i>b</i> , <i>c</i> (Å), α (°)	83.39, 176.51, 126.02, 90.00	123.91, 129.92, 161.01, 90.00
Resolution range (Å)	20.00–2.45 (2.59–2.45)	19.97–2.60
Unique reflections ^a	34,663 (5303)	
Multiplicity ^a	4.68 (4.63)	
Completeness (%) ^a	99.2 (97.4)	
<i>I</i> / δ ^a	11.70 (2.74)	
<i>R</i> _{merge} (%) ^a	10.1 (51.0)	
Overall B-factor (Å ²)	41.27	
Refinement		
Protein atoms/residues	4810/610	14,917/1832 (13,740/1694)
Solvent molecules	198	383 (582)
<i>R</i> _{work} / <i>R</i> _{free} (%) ^{b,c}	19.2/25.4	20.27/26.38 (22.1/26.5)
RMSD ^b		
Bond length (Å) ^b	0.01	0.009 (0.014)
Bond angles (°) ^b	1.73	1.008 (1.418)
Ramachandran plot (%)^b		
Core region ^b	98.2	94.14 (91.7)
Allowed region ^b	1.8	4.7 (6.15)
Outliers ^b	0	1.16 (2.15)

^a Values in parentheses are for the highest resolution shell.

^b Values in parentheses are those of the original HsdR structure (PDB code 2W00).

^c *R*_{free} was calculated using 5% of the data excluded from refinement.

Despite the prevailing assumption that the C terminus is a part of the helical domain, it actually forms a separate domain characterized by a unique fold and its own hydrophobic core. The domain is stable on its own in solution, as demonstrated in

molecular dynamics simulations (data not shown). The C-terminal domain consists of six α -helices arranged in two bundles around the hydrophobic core (Fig. 1B), shaped similar to a tube with dimensions of 50 by 26 Å. According to secondary struc-

Novel domain of the motor subunit of Type I R-M system

ture analysis in STRIDE (30), of 152 residues in the C-terminal domain, 60.5% form α -helices, 6.0% 3_{10} -helices, 11.2% turns, 19% coils, and 3.3% are unresolved. The longest α -helix is the fourth one (27 residues), whereas the shortest is the third one (nine residues), which is preceded by three residues forming a 3_{10} -helix. All α -helices are arranged in an antiparallel manner, except $\alpha 5$ and $\alpha 6$. They are linked by a 27-residue-long coil–turn– 3_{10} -helix–turn–coil–turn region, bringing helix $\alpha 6$ to an $\sim 45^\circ$ angle in relation to $\alpha 5$ and almost perfectly antiparallel to $\alpha 4$. The N and C termini of the domain are ~ 27 Å apart. Six contacts between the C-terminal domain and the crystallization chaperone pFluorin are present (Asn²⁸⁴-Phe²⁴⁶ (linker), Arg²⁸⁵-Tyr⁴⁵, Asn³³³-Lys⁴⁷, Arg³⁴⁹-Thr⁴⁴, Thr³⁵¹-Asp⁸², and Lys³⁷⁶-Thr²³⁶).

A fold analysis by the Dali server (31), which compares newly solved structures against structures in the Protein Data Bank to compute the amount of structural similarity as a so-called Z-score, was unable to assign the C terminus to a specific fold family. Of 30 structures with a Z-score of 4.0 or higher reported by Dali, the highest percentage of aligned residues is 71% (PDB code 5W1H, chain A) and the lowest root mean square deviation is 2.4 (PDB code 2DOB, chain A). Most of the structures with a Z-score ≥ 4.0 have less than 50% aligned residues corresponding at best to a three-helix bundle. Hence, the C-terminal domain represents a novel six-helix bundle.

Modeling of full-length HsdR and electrostatic surface potential analysis

The extended HsdR crystal structure was obtained by combining the previous WT HsdR coordinates and structure factors PDB code 2W00 (21) with the structure of the C-terminal domain determined here. Using Phaser from the CCP4 software bundle, the main features of the secondary structure of the C-terminal domain were fitted into the electron density map, even though many side chains and some regions were not fully resolved after refinement (Fig. 1, C and D, and Table 1). High mobility of the C-terminal domain that formerly prevented structure determination correlates well with increased B-factors of amino acids most distal from the core of the enzyme. The resulting structure represents the first experimentally determined full-length structure of HsdR deposited in the Protein Data Bank (PDB code 6H2J). Nevertheless, to be able to calculate the electrostatic surface potential and determine hydrogen-bond interactions, a full-length HsdR model was generated that includes the missing loops, side chains, and hydrogens. To attain more side chain information, the C-terminal domain from the pFluorin-HsdR887 fusion structure was aligned in YASARA using the Mustang software (32), and atom coordinates were replaced by higher-resolution information from the pFluorin-HsdR887 crystal structure. A root mean square deviation of 0.884 Å over 119 aligned residues of the C-terminal domain was observed during the alignment. The connecting region between the helical domain and the C-terminal domain (residues 887–902) was modified manually to fit the electron density map of the WT structure. The resulting model represents the first full-length HsdR subunit of a Type I restriction enzyme (Fig. 2). The full-length model was used for all further analysis (see Model S1).

The four previously identified domains of HsdR are the endonuclease domain (residues 13–260), RecA-like helicase domains 1 and 2 (residues 261–461 and 470–731), and the helical domain (residues 732–892) (21). The C-terminal fifth domain connects to the helical domain via a short linker (residues 886–888). Consequently, the helical domain should be reassigned to include residues 732–886 only, and residues 887–1038 should be assigned to the C-terminal domain. The most striking feature of the C-terminal domain is the way it is arranged with respect to the rest of the protein; although the four previously identified domains are arranged to form a square-planar body, the fifth domain protrudes outward (Fig. 2).

Calculations of the common molecular surface between the C-terminal and other domains of HsdR show a common surface of 39.67 Å² for the helicase 2 domain and 142.86 Å² for the helical domain. The hydrophobicity of the common surface between the C-terminal and helicase 2 domain is 57.5% and 57.9% for the helical domain. The endonuclease and helicase 1 domains do not share a common molecular surface with the C-terminal domain.

Potential interactions with the helicase 2 domain are as follows: Glu⁸⁹⁹ and Lys⁵⁷² with a 3.3-Å distance between C δ of glutamate and C ϵ of lysine; Gln⁹⁰³ and Thr⁵⁷⁰ with a 3.6-Å distance between the nitrogen of the amino group of glutamine and the backbone oxygen of threonine; and Tyr⁸⁹² and Pro⁵³³ with a 3.5-Å distance between the oxygen of tyrosine and the C β atom of proline. The helical domain has the following potential interactions with the C-terminal domain: a hydrophobic interaction between Leu⁹²² and Leu⁸⁸³ (3.5 Å between C $\delta 1$); Ser⁹²⁰ and Gln⁸⁴⁴ with a distance of 3.4 Å between the backbone oxygen of serine and the amino group nitrogen of glutamine and a hydrogen bond of 2.1 Å between the backbone oxygen of Leu⁹²² and the H $\delta 2$ of Asn⁸⁵¹. Another potential interaction zone for the helical domain is in the 858–862 loop that comes as close as 3.5 Å (measured between C δ of Tyr⁹⁹¹ and the backbone oxygen of Arg⁸⁵⁹) to the potentially flexible loop–coil region connecting helices 5 and 6 in the C-terminal domain. The potential interactions and common molecular surfaces of the C-terminal domain are shown in Fig. 3.

Electrostatic surface potential calculations were carried out for the full structure (Fig. 2). A prominent groove between the endonuclease and the C-terminal domain (and, partially, the helical domain) provides a prevalently positively charged protein surface that extends the DNA-binding cleft between the two helicase domains (Fig. 2, A and C). This groove could potentially bind to DNA, bringing the dsDNA close to the catalytic residues Asp¹⁵¹, Glu¹⁶⁵, and Lys¹⁶⁷ in the endonuclease domain.

Bacteriophage restriction assay with the C terminus deletion mutant of HsdR

A standard bacteriophage restriction test was used to compare the C terminus deletion mutant with the WT HsdR in its restriction activity and MTase binding (33). The pTrcR124 plasmid (34) carrying a gene for either WT or mutated HsdR was introduced into a restriction-deficient (JM109(DE3)/pACMS r–m+) or restriction-proficient host (JM109(DE3)/

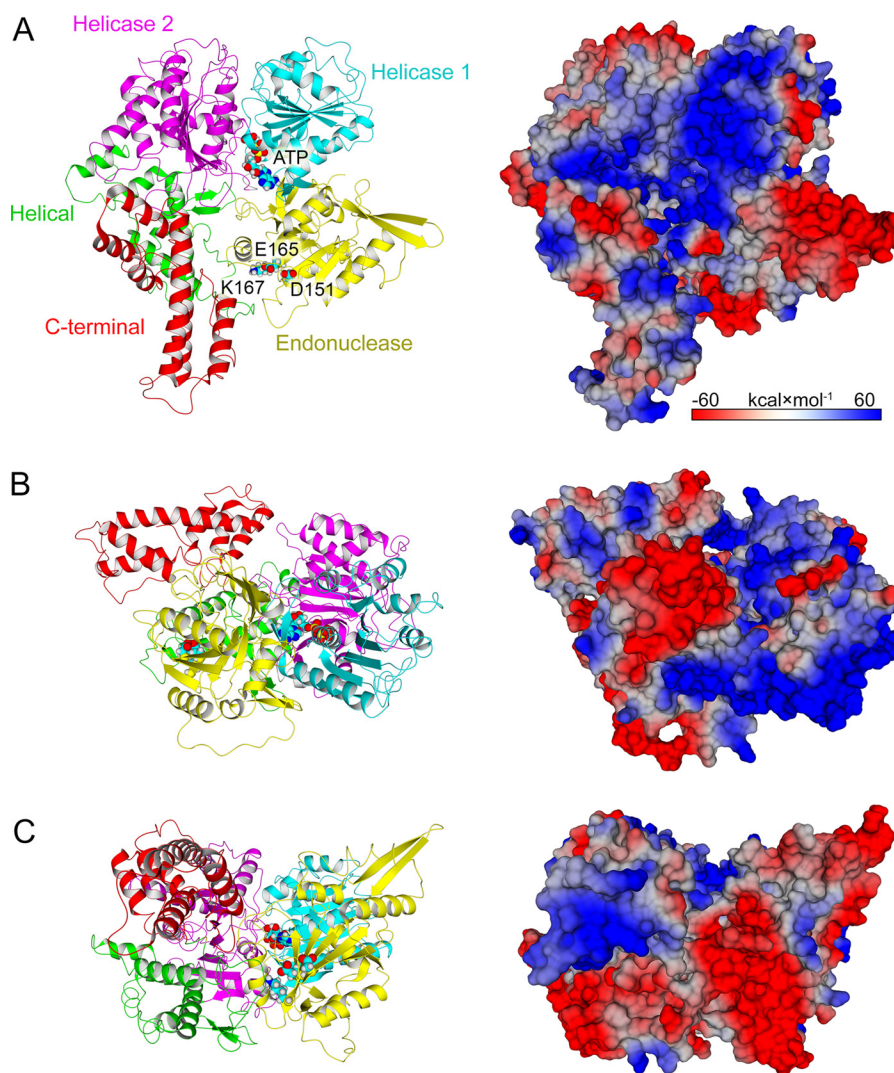


Figure 2. A–C, full-length model of the HsdR subunit “front” (A) and “side” (B and C) views of the full-length HsdR model obtained by combining four previously known domains and the C-terminal domain. Domains are color-coded to match the first published HsdR structure (21); the C-terminal domain is shown in red. The catalytic triad (Asp¹⁵¹, Glu¹⁶⁵, and Lys¹⁶⁷) and ATP are marked in A. The corresponding electrostatic potential surfaces are shown on the right (blue, positive; red, negative). Positively charged surface areas in the C-terminal domain extend the proposed DNA path. The figure was prepared in YASARA.

pKF650 r+m+) as described under “Experimental procedures.” Positive and negative complementation tests described previously (35) were employed to distinguish between DNA cutting deficiency and inability to assemble with MTase, respectively. The results in Table 2 represent relative efficiency of plating (EOP) values, calculated as the number of plaques relative to that of a nonrestricting strain (r–m+), JM109(DE3).

The C terminus deletion mutant HsdR shows a restriction-deficient phenotype in the r–m+ strain with an EOP of ~0.7 (values from 0.1 to 1 correspond to a nonrestricting strain), whereas the WT protein complements the r–m+ strain, showing an ~100-fold decrease in phage growth (EOP, 0.0084). Such a difference in DNA cutting ability can be explained in at least two different ways: inability of the mutant to function properly as a DNA cutter or defective interaction with MTase. The latter appears to be true, according to the results of a negative complementation experiment using the r+m+ strain, where the mutant and WT HsdRs compete for MTase binding. Similar values of EOP for the C-terminal deletion mutant and full-

length HsdR in this test (0.0041 and 0.0075, respectively) show no significant competition between the subunits, leaving restriction proficiency almost intact for the strain expressing both mutant and WT HsdR.

Electrophoretic mobility shift assay with C terminus deletion HsdR mutant

Further indication of the involvement of the C-terminal domain in binding to MTase comes from a direct observation of complex formation of purified proteins in an electrophoretic mobility shift assay (EMSA). In this experiment, ³²P-labeled duplex oligonucleotides were mixed with MTase, and the WT or C terminus deletion mutant HsdR was added in increasing concentrations. The position of the labeled DNA duplex, visualized on a native polyacrylamide gel, is shifted according to the physical properties of the protein complexes bound to it, allowing to track the interactions between MTase and the HsdR subunit. Fig. 4 shows the formation of distinct S₁M₂R₁ and S₁M₂R₂ for the WT HsdR as the ratio of HsdR to MTase increases. At

Novel domain of the motor subunit of Type I R-M system

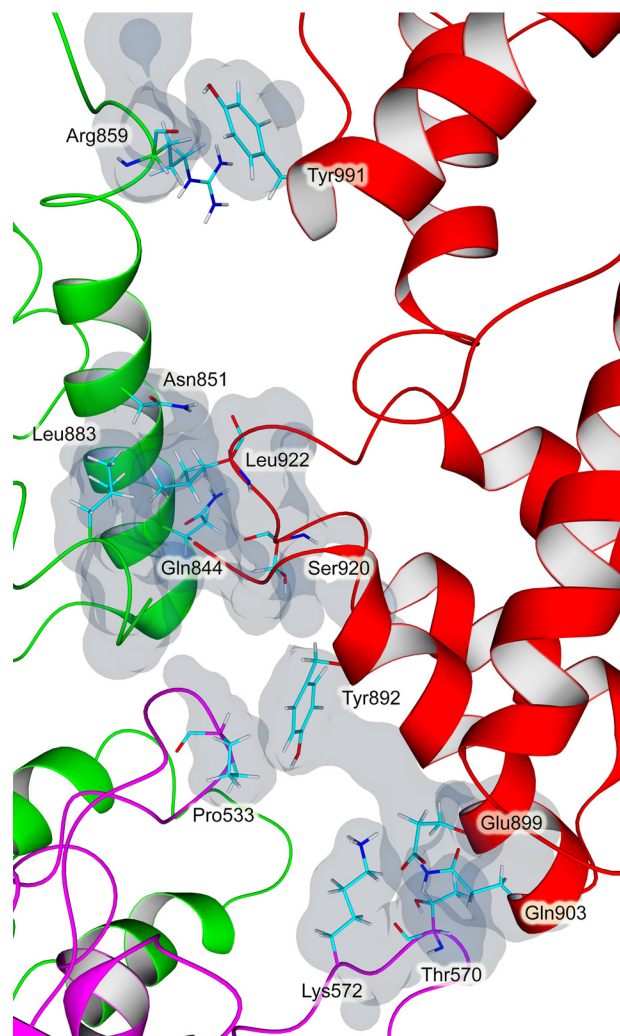


Figure 3. Interdomain interactions of the C-terminal domain. Shown is part of the HsdR subunit, showcasing the interface between the helixase 2 (magenta) and helical (green) and C-terminal domain (red). The molecular surface is drawn in transparent gray for each domain for atoms within the 5-Å distance from the interacting domain. The figure was prepared in YASARA.

Table 2

Average efficiency of plating of the wildtype and C-terminal deletion mutant HsdR relative to a non-restricting strain, JM109(DE3) r-m-

HsdR	r- Host ^a	r+ Host ^b
Wildtype	0.0084 ± 0.0083 ^c	0.0041 ± 0.0038
C-terminal deletion mutant	0.71 ± 0.27	0.0075 ± 0.0023

^a Positive complementation in the r-m+ *E. coli* strain JM109(DE3) (pACMS).

^b Negative complementation the r+m+ *E. coli* strain JM109(DE3) (pKF650).

^c The standard deviation derived from at least three independent experiments.

the same time, only the S_1M_2 complex forms for the mutant HsdR, showing that no interaction occurs up to a 3× molar excess of HsdR over MTase.

Identification of conserved regions in the C-terminal domain

Because the direct multiple alignment between the characterized HsdRs of the Type I families is impossible because of low sequence identity of the helical and C-terminal domains, a different approach using putative enzymes was implemented here. All available putative HsdR sequences were downloaded from REBASE, and redundant sequences were removed as described

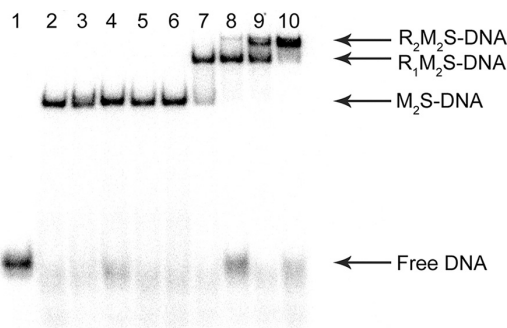


Figure 4. 887 deletion mutant and WT HsdR binding to DNA in EMSA. The 10 nm oligonucleotide labeled with [γ^{32} P]ATP is in lane 1 (the corresponding band is marked *free DNA*), the oligonucleotide mixed with 40 nM MTase is in lane 2 (the band is marked M_2S -DNA), and either the C-terminal deletion mutant (lanes 3–6) or wild-type HsdR (lanes 7–10) was added to the DNA-MTase mixture at concentrations of 20, 40, 80, and 120 nM. Complexes with either one (R_1M_2S -DNA) or two (R_2M_2S -DNA) HsdR subunits are indicated with arrows where appropriate. Gel borders are outlined in gray.

under “Experimental procedures.” The final trimmed dataset contained 8851 sequences, among which ~92% of sequences were 20–40% identical to R.EcoR124, whereas other percentiles are much less populated (Fig. 5). This dataset served to identify conserved motifs in the C terminus using Clustal Omega for generating multiple alignments (Fig. 6). The identified motifs were then cross-checked in all available sequence data based on pairwise alignment to R.EcoR124 to evaluate their conservation (Fig. 5). The main motifs identified in sequences with a sequence identity as low as 40–45% with EcoR124 were as follows: $^{887}EXNXDYIL^{894}$, $^{925}RKKX-XLXXXFI^{935}$, $^{996}G-^{1004}PXXS^{1007}$, and $^{1016}KKXXXXXXK^{1023}$ close to the C terminus. These motifs seem to have very specific functions in the C-terminal domain. Motif $^{887}EXNXDYIL^{894}$ is part of the N-terminal “hinge” region connecting the helical and C-terminal domain (Fig. 7A), with Tyr⁸⁹² contacting Pro⁵³³ of the helixase 2 domain (with a distance of 2.7 Å between C ζ of Tyr⁸⁹² and H β of Pro⁵³³). Motif $^{925}RKKXLXXXFI^{935}$ is flanking the $\alpha 3$ helix (Fig. 7B), with Arg⁹²⁵ and Lys⁹²⁷ forming a nearly helical structure just before the first real helix–turn and Phe⁹³⁴ and Ile⁹³⁵ being in the C-terminal turn of the helix, contacting helices 4 and 2, respectively. The conserved Leu⁹³⁰ is facing and contacting helix 4. Motif $^{996}G-^{1004}PXXS^{1007}$ helix (Fig. 7C) is part of the 24-amino acid-long loop connecting helices 5 and 6, with Pro¹⁰⁰⁴ and Ser¹⁰⁰⁷ forming a local structure placed in a part of the loop that passes through a cleft formed by helices 4 and 6. In the fourth motif, $^{1016}KKXXXXXXK^{1023}$ (Fig. 7D), Lys¹⁰¹⁶ is contacting Glu⁹⁷² (H-bond donor-acceptor distance, 2.76 Å) in the last turn of helix 4, with Glu⁹⁷² being conserved throughout the analyzed sequences. Lys¹⁰²³ has a very similar interaction, as it is contacting Glu⁹⁷¹ (H-bond donor-acceptor distance, 2.76 Å) in the last turn of helix 4 neighboring the conserved Glu⁹⁷². However, Glu⁹⁷¹ is less conserved, being either aspartic acid or glutamic acid in sequences around 40–50% identity but not providing a negative charge in sequences with higher identity (70–75%). Lys¹⁰¹⁷ is oriented away from the protein, pointing into bulk water, and is thus potentially a candidate for contacting another subunit or DNA.

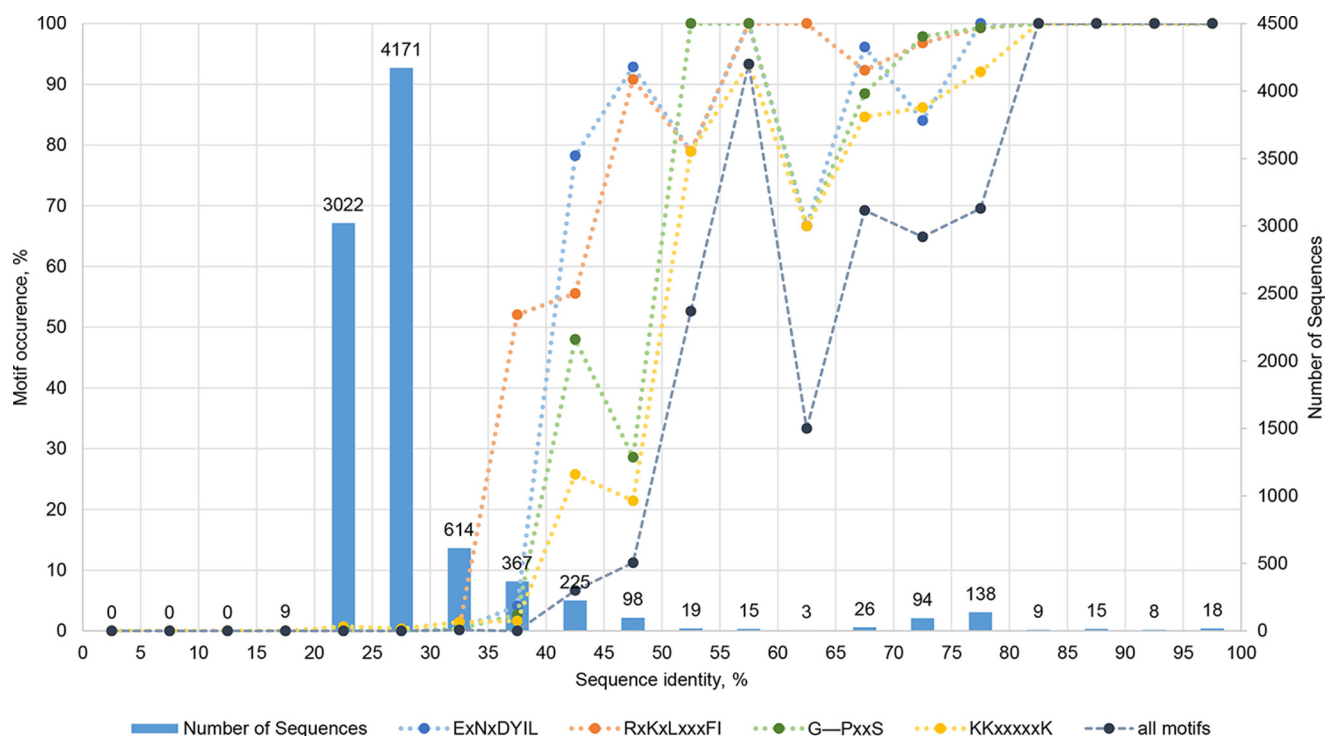


Figure 5. Occurrence of conserved regions in the C-terminal domain of putative HsdR sequences broken by sequence identity to R.EcoR124. Most of the sequences (~92%) share 20–40% sequence identity with the HsdR of EcoR124, whereas ~3% fall into 65–100%, representing the Type IC family. Four conserved regions were identified in sequences as remote as 40–50%.

The multiple alignment in Fig. 6 shows a sample of sequences of putative HsdRs based on alignment scores and phylogenetic representativeness. So far, 37 Type I enzymes were characterized biochemically, and only for ten is the HsdR sequence is available (EcoKI, EcoAI, StySBLI, NgoAV, KpnBI, KpnAI, CjeFII, CjeFIV, and the N-terminal fragments of EcoPrRI and StySKI). All four motifs were found only in NgoAV (IC family). The other IC member, EcoPrRI, could not be analyzed because only a 313-amino-acid N-fragment was sequenced. Other sequences shown in the alignment are putative HsdRs, one representing a similar high sequence identity as NgoAV, whereas others are representing lower identities in the range of 44–55%. Among all putative HsdRs, 3.5% of the sequences were more than 65% identical, probably representing subfamily IC (Fig. 5). This distribution has its maximum at 70–80%, with 232 sequences of 308 sequences being in that range. Nevertheless, even in sequences with a sequence identity as low as 40%, all four motifs are conserved in ~7% of all sequences. Here, motif ⁸⁸⁷EXNXDYIL⁸⁹⁴ shows the highest conservation, being present in 78% of all sequences with a sequence identity of 40–45%. Motifs ⁹²⁵RXKXLXXXFI⁹³⁵ and ⁹⁹⁶G—¹⁰⁰⁴PXXS¹⁰⁰⁷ are still present in ~50% of all sequences, with a sequence identity of 40–45%. The least conservation is observed for motif ¹⁰¹⁶KKXXXXXK¹⁰²³, appearing in sequences with a sequence identity above 50%. The general threshold for being assigned to the same family is a sequence identity of at least 30%. As approximately one thousand sequences in the data set show a sequence identity of 30–40% despite not having any of the identified motifs, the novel six-helix bundle observed in the C-terminal domain is probably common to subfamily IC and

several other closely related subfamilies of Type I R-M systems.

Conclusions

Despite a few structures of HsdR from the EcoR124 R-M system that have already been solved, the absence or poor quality of the electron density around the C-terminal domain barred previous efforts to obtain a robust atomic-level crystallographic structure. We propose that the consistently poor quality of the electron density in this region is due to its high flexibility and, hence, irregularities in crystal packing. The fusion with ratiometric pHluorin (a pH-sensitive variant of GFP) was a measure to induce additional contacts to make the C-terminal domain less flexible within the crystal packing context. The fusion partner also helped to overcome the absence of expression when trying to express the C-terminal domain separately (data not shown). The gene of the fusion protein, on the other hand, was easy to express, and the overproduced protein was easy to purify and crystallize (28).

Additionally, the fusion approach allowed us to obtain phases for the C-terminal domain without carrying out experimental phasing. Combining crystallographic data from the previously obtained WT structure allowed us to create the first full model of the HsdR subunit for a Type I RM enzyme. The whole procedure further validates the crystallization chaperone approach (29).

So far, the motor subunit has been described to be comprised of four structural and functional domains (61) forming a planar array, and the C-terminal region of HsdR had been thought to enlarge the helical domain, making the square-planar

Novel domain of the motor subunit of Type I R-M system

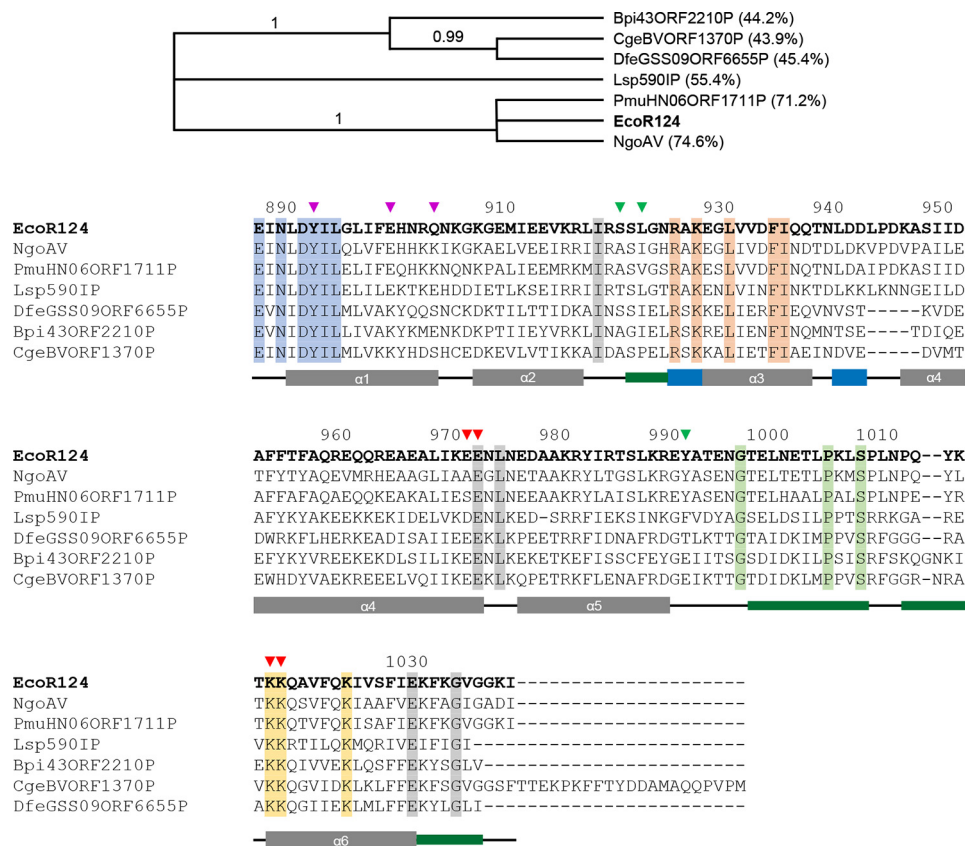


Figure 6. Multiple sequence alignment of the C-terminal domain of putative and described HsdR subunits. The alignment was performed for two biochemically described Type IC HsdR subunits (EcoR124 and NgoAV) and five putative proteins with a sequence identity of ~44–71% compared with R.EcoR124. The sequences were retrieved from REBASE (1). The cladogram was produced using <http://www.phylogeny.fr/> (please note that the JBC is not responsible for the long-term archiving and maintenance of this site or any other third party-hosted site) (59) and PhyML (60); the alignment was performed in Clustal Omega (58). Only the C-terminal domain (amino acids 887–1038) is shown, although the alignments were performed with full-length subunits. The numbers on top of the sequence denote the EcoR124 numbering. Conserved residues are highlighted (⁸⁸⁷EXNXYDYL⁸⁹⁴ motif, blue; ⁹²⁵RKXXLXXXF⁹³⁵, orange; ⁹⁹⁶G₁₀₀₄PXXS¹⁰⁰⁷, green; ¹⁰¹⁶KKXXXXXK¹⁰²³, yellow; others, gray). The residues involved in inter- and intradomain contacts are marked with triangles colored according to the residue's contact partner (the domain coloring scheme is the same as in Fig. 2). The secondary structure elements are given according to the EcoR124 sequence (gray bars, α -helix numbered as in Fig. 1B; blue bars, 3_1 -helix; green bars, turn).

arrangement more symmetrical. The crystal structure refined to 2.45-Å resolution reported here reveals not only that this part of the protein has its own hydrophobic core and forms an independent C-terminal domain that has a unique α -helical fold but also breaks the symmetry and planarity of the motor subunit and reshapes it by protruding outward. The C-terminal domain is probably common to subfamily IC and several other closely related subfamilies of Type I restriction-modification systems.

The results of both *in vivo* and *in vitro* assays demonstrate that the C-terminal domain is an essential component in HsdR's ability to bind to MTase. Because one or two HsdRs are recruited by MTase bound to DNA after recognition of the methylation state of adenines within the recognition sequence, both protein-protein and DNA-protein interactions are expected to happen. In agreement with this, electrostatic surface potential calculations with full-length HsdR revealed a positively charged patch of residues in the C-terminal domain that extend the DNA-binding cleft between the two helicase domains. We suppose that these residues may facilitate the correct DNA positioning at the active site of the endonuclease domain. The conserved motif ¹⁰¹⁶KKXXXXXK¹⁰²³ would be a good candidate for participating in binding to the

methyltransferase. Hence, the motor subunit is comprised of five structural and functional domains, and the fifth, the C-terminal domain, reveals a novel fold characterized by four conserved motifs in subfamily IC of Type I restriction-modification systems, essential for proper complex assembly and probably involved in DNA binding.

Experimental procedures

Fusion protein production and crystallization

A gene coding for the pHluorin-HsdR887 fusion protein, consisting of a His₆ tagged GFP variant, ratiometric pHluorin (36), and amino acids 887–1038 of the HsdR subunit corresponding to C-terminal domain, was cloned and expressed, and the overproduced protein was purified and crystallized as described previously (28).

Briefly, initial crystallization trials were done using a Gryphon crystallization robot (Art Robbins Instruments) employing the crystallization screens Morpheus (Molecular Dimensions) and PEG/Ion (Hampton Research) in MRC 2-well crystallization plates (Hampton Research). For further optimization, protein and precipitant concentrations were varied, and additives from Additive Screen (Hampton Research) were added in 24-well Comb

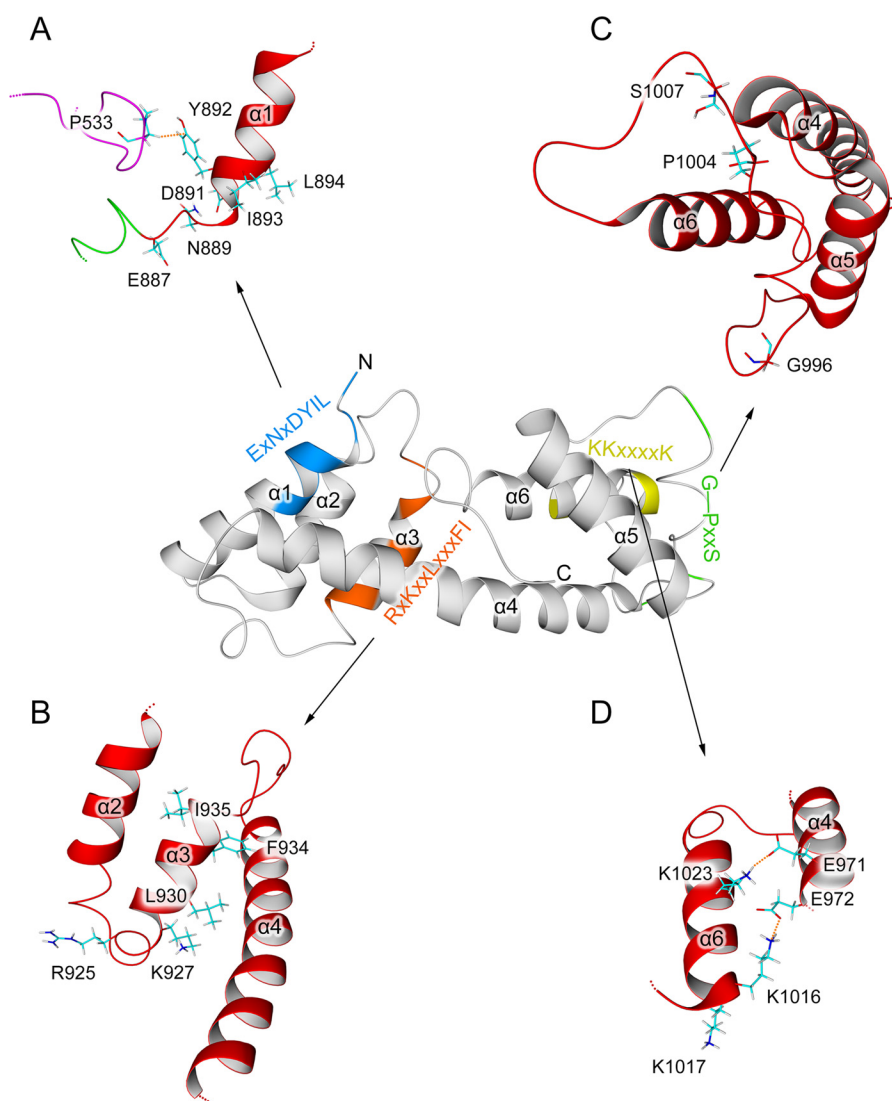


Figure 7. Motifs in the C-terminal domain of HsdR. A–D, conserved motifs mapped to the C-terminal domain (*center*) and their close-ups: ⁸⁸⁷EXNXDYIL⁸⁹⁴ (A), ⁹²⁵RKXXLXXXFI⁹³⁵ (B), ⁹⁹⁶G—¹⁰⁰⁴PXXS¹⁰⁰⁷ (C), and ¹⁰¹⁶KKXXXXK¹⁰²³ (D). Residues constituting the motifs and appropriate contact partners are labeled. The α -helices are numbered as in Fig. 1B. The domain coloring scheme is the same as in Fig. 2. Contacts are shown with dotted orange lines. The figure was prepared in YASARA.

iClover crystallization plates (Jena Bioscience). The sitting drop vapor diffusion technique was used for all trials (37).

Data processing and structure determination

Data collection was done at the PETRA III X-ray radiation source (Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany) on beamline P13 (38) operated by EMBL Hamburg, equipped with a PILATUS 6 M-F detector (DECTRIS Ltd.). Prior to flash-cooling in liquid nitrogen, 1–3 μ l of 50% (w/v) PEG 3350 was added to the drop with crystals for cryoprotection. XDS software (39) was used for integrating data images. The space group was determined with POINTLESS from the CCP4 software package (40), and the Matthews coefficient (41) was calculated using MATTHEWS_COEF from CCP4.

Initial phases were obtained by molecular replacement with Molrep (42) using a GFP structure (PDB code 1W7S (43)) as a template. Missing parts corresponding to the C-terminal domain of HsdR were traced by Buccaneer (44). The model was refined by applying cycles of restrained and rigid body refinement

in Refmac5 (45) and manual model building in Coot (46). The final structure was checked with the PDB validation server and uploaded to the PDB after minor corrections (PDB code 5J3N).

Refinement, modeling, and surface analysis of full-length HsdR

WT HsdR coordinates and structure factors (PDB code 2W00 (21)) were used to position the C-terminal domain with respect to the rest of the protein. In PDB code 2W00, the C-terminal domain is not resolved; however, the missing \sim 150 C-terminal residues of each monomer were confirmed to be present in dissolved crystals by gel electrophoresis, and the electron density reveals several short disordered regions, allowing positioning of the C-terminal domain relative to the rest of the structure. To do so, the C-terminal domain and the WT HsdR structure of PDB code 2W00 were used as two independent ensembles in Phaser (47). The obtained model was then trimmed in Coot, and a few cycles of restrained and TLS (Translation/Libration/Screw) refinement and model building

Novel domain of the motor subunit of Type I R-M system

were performed with Refmac5 (45) and phenix.refine (48). The extended WT structure with the C-terminal domain was deposited in the PDB database (PDB code 6H2J).

The full-length model was obtained by adding hydrogens and modeling the missing four segments (residues 142–147, 585–590, 859–869, and 903–908) using a standard loop remodeling approach in YASARA (49, 32), followed by steepest descent energy minimization. The missing residues 182–189 were modeled as described previously (50).

Electrostatic surface potentials were calculated in YASARA with the particle mesh Ewald method using a 1-nm cutoff, a maximum electrostatic surface potential of 60 kcal mol⁻¹, and a probe radius of 1.4 Å. The molecular surfaces common between the C-terminal and other domains were calculated as the difference of the sum of molecular surfaces of each domain pair and the combined molecular surface of the domain pair. All hydrophobicity calculations and surface area calculations were performed using the standard approach described in the YASARA tutorial (49).

HsdR and MTase production

A deletion in the *HsdR* gene coding for the C-terminal amino acids 887–1038 was introduced into the pTrcR124 plasmid commonly used for HsdR overexpression (34) by means of one-step PCR-based mutagenesis (51). The forward and reverse primers were as follows: 5'-GCTGAAGTCTCAGTAGC-CCAATTCGTGTTTTTC-3' and 5'-ACGAATTGGGCTACTGAGACTTCAGCAAATCG-3'. PCR products were digested with DpnI to degrade the methylated template and used to transform *Escherichia coli* DH5 α cells. DNA was extracted from bacterial cultures using the Zyppy plasmid miniprep kit (Zymo Research). The presence of the deletion and absence of undesired additional mutations was verified by DNA sequencing. Both WT and mutant HsdRs were purified as described previously (34). MTase was produced with a two-step purification procedure in a similar manner as before (52).

Bacteriophage restriction assay

The *in vivo* restriction activity for the WT and C terminus deletion mutant of HsdR was assessed by analyzing the relative EOP in a bacteriophage λ_0 plaque assay. The *E. coli* JM109 (DE3) strain (53) lacking *RecA* and genes of restriction enzymes was transformed with pACMS carrying the *HsdS* and *HsdM* genes coding for MTase or pKF650 carrying the *HsdS*, *HsdM*, and *HsdR* genes (54). In a positive complementation test, the JM109(DE3)/pACMS strain was further transformed with pTrcR124 (34) carrying either the WT or C terminus deletion mutant *HsdR*. Similarly, the JM109(DE3)/pKF650 strain was transformed with pTrcR124 for a negative complementation test. The virulent bacteriophage λ (55) was cultivated, and its titer was determined as described previously (33). 0.5 ml of fresh overnight culture was mixed with 3 ml of soft agar medium (1% Tryptone, 0.5% yeast extract, 0.5% NaCl, 0.6% agar) and spread on the plates with the corresponding antibiotics. Phage dilutions (10² to 10⁶ pfu/ml) were then dropped on the plates, followed by overnight incubation at 37 °C. The aver-

age EOP was calculated relative to a nonrestricting JM109(DE3) strain from at least three independent experiments.

EMSA

Binding of the EcoR124 complex to DNA was assessed *in vitro* following the retardation in electrophoretic mobility of a radioactively labeled 30-mer duplex oligonucleotide (5'-CGT-GCAGAATTCGAGGTCGACGGATCCGGG-3'; the recognition site is shown in bold). Equimolar amounts of complementary oligonucleotides were annealed, labeled with [γ ³²P] ATP by T4 polynucleotide kinase (New England Biolabs), and purified with the QIAquick Nucleotide Removal Kit (Qiagen). Binding reactions were performed for 10 min at room temperature in a total volume of 10 μ l in 50 mM Tris-HCl (pH 8.0), 50 mM NaCl, 10 mM MgCl₂, 1 mM DTT, and 10% (v/v) glycerol with 10 nM DNA substrate and varying concentrations of MTase and the WT or mutant HsdR. DNA and protein complexes were separated on a 6% nondenaturing polyacrylamide gel run in TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH 8.3) at 4 °C and 7 V/cm field strength. After being vacuum-dried for 30 min at 80 °C (583 Gel Dryer, Bio-Rad), the gel was visualized on a Molecular Imager FX (Bio-Rad).

Sequence analysis

The putative Type I HsdR sequences for analysis were downloaded from REBASE (1); 18,073 sequences were retrieved on February 24, 2018 (see supporting information). Because many sequences represent the results of whole-genome sequencing of closely related strains of microorganisms, the redundancy of the dataset was reduced by removing nearly identical sequences from the same species (more than 98% sequence identity) and fragments (less than 700 amino acids). The dataset was trimmed using R (56) with the Biostrings package (57). Multiple sequence alignment was performed in Clustal Omega v1.2.4 (58).

Author contributions—P. G. and R. H. E. conceptualization; P. G., D. S., I. I., A. G., and J. L. investigation; P. G. and R. H. E. writing-original draft; P. G., D. S., I. I., A. G., M. W., J. L., J. R. M., and R. H. E. writing-review and editing; M. W., J. L., J. R. M., and R. H. E. supervision; R. H. E. funding acquisition; R. H. E. project administration.

Acknowledgments—Data collection was carried out at the PETRA III light source at DESY, a member of the Helmholtz Association (HGF). We thank Dr. Johanna Kallio for assistance with using beamline P13.

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 [CrossRef Medline](#)
2. Murray, N. E. (2002) Immigration control of DNA in bacteria: self versus non-self. *Microbiology* **148**, 3–20 [CrossRef Medline](#)
3. Oliveira, P. H., Touchon, M., and Rocha, E. P. (2014) The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.* **42**, 10618–10631 [CrossRef Medline](#)
4. Manso, A. S., Chai, M. H., Atack, J. M., Furi, L., De Ste Croix, M., Haigh, R., Trappetti, C., Ogunniyi, A. D., Shewell, L. K., Boitano, M., Clark, T. A., Korlach, J., Blades, M., Mirkes, E., Gorban, A. N., *et al.* (2014) A random

- six-phase switch regulates pneumococcal virulence via global epigenetic changes. *Nat. Commun.* **5**, 5055 [CrossRef Medline](#)
5. Furuta, Y., Namba-Fukuyo, H., Shibata, T. F., Nishiyama, T., Shigenobu, S., Suzuki, Y., Sugano, S., Hasebe, M., and Kobayashi, I. (2014) Methylome diversification through changes in DNA methyltransferase sequence specificity. *PLoS Genet.* **10**(4), e1004272 [CrossRef Medline](#)
 6. Budroni, S., Siena, E., Dunning Hotopp, J. C., Seib, K. L., Serruto, D., Nofroni, C., Comanducci, M., Riley, D. R., Daugherty, S. C., Angiuoli, S. V., Covacci, A., Pizza, M., Rappuoli, R., Moxon, E. R., Tettelin, H., and Medini, D. (2011) *Neisseria meningitidis* is structured in clades associated with restriction modification systems that modulate homologous recombination. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 4494–4499 [CrossRef Medline](#)
 7. Erwin, A. L., Sandstedt, S. A., Bonthuis, P. J., Geelhood, J. L., Nelson, K. L., Unrath, W. C., Diggie, M. A., Theodore, M. J., Pleatman, C. R., Mothershed, E. A., Sacchi, C. T., Mayer, L. W., Gilsdorf, J. R., and Smith, A. L. (2008) Analysis of genetic relatedness of *Haemophilus influenzae* isolates by multilocus sequence typing. *J. Bacteriol.* **190**, 1473–1483 [CrossRef Medline](#)
 8. Loenen, W. A., Dryden, D. T., Raleigh, E. A., Wilson, G. G., and Murray, N. E. (2014) Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Res.* **42**, 3–19 [CrossRef Medline](#)
 9. Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J., Blumenthal, R. M., Degtyarev, S. K. H., Dryden, D. T., Dybvig, K., Firman, K., Gromova, E. S., Gumpert, R. I., Halford, S. E., Hattman, S., et al. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.* **31**, 1805–1812 [CrossRef Medline](#)
 10. Youell, J., and Firman, K. (2008) EcoR124I: from plasmid-encoded restriction-modification system to nanodevice. *Microbiol. Mol. Biol. Rev.* **72**, 365–377, table of contents [CrossRef Medline](#)
 11. Suri, B., Shepherd, J. C., and Bickle, T. A. (1984) The EcoA restriction and modification system of *Escherichia coli* 15T⁻: enzyme structure and DNA recognition sequence. *EMBO J.* **3**, 575–579 [Medline](#)
 12. Sommer, R., and Schaller, H. (1979) Nucleotide sequence of the recognition site of the B-specific restriction modification system in *E. coli*. *Mol. Gen. Genet.* **168**, 331–335 [Medline](#)
 13. Bertani, G., and Weigle, J. J. (1953) Host controlled variation in bacterial viruses. *J. Bacteriol.* **65**, 113–121 [Medline](#)
 14. Horiuchi, K., and Zinder, N. D. (1972) Cleavage of bacteriophage ϕ 1 DNA by the restriction enzyme of *Escherichia coli* B. *Proc. Natl. Acad. Sci. U.S.A.* **69**, 3220–3224 [CrossRef Medline](#)
 15. Kennaway, C. K., Obarska-Kosinska, A., White, J. H., Tuszyńska, I., Cooper, L. P., Bujnicki, J. M., Trinick, J., and Dryden, D. T. (2009) The structure of M. EcoKI Type I DNA methyltransferase with a DNA mimic antirestriction protein. *Nucleic Acids Res.* **37**, 762–770 [CrossRef Medline](#)
 16. Liu, Y. P., Tang, Q., Zhang, J. Z., Tian, L. F., Gao, P., and Yan, X. X. (2017) Structural basis underlying complex assembly and conformational transition of the type I R-M system. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11151–11156 [CrossRef Medline](#)
 17. Park, S. Y., Lee, H. J., Song, J. M., Sun, J., Hwang, H. J., Nishi, K., and Kim, J. S. (2012) Structural characterization of a modification subunit of a putative type I restriction enzyme from *Vibrio vulnificus* YJ016. *Acta Crystallogr. D* **68**, 1570–1577 [CrossRef Medline](#)
 18. Kim, J. S., DeGiovanni, A., Jancarik, J., Adams, P. D., Yokota, H., Kim, R., and Kim, S. H. (2005) Crystal structure of DNA sequence specificity subunit of a type I restriction-modification enzyme and its functional implications. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 3248–3253 [CrossRef Medline](#)
 19. Calisto, B. M., Pich, O. Q., Piñol, J., Fita, I., Querol, E., and Carpena, X. (2005) Crystal structure of a putative type I restriction-modification S subunit from *Mycoplasma genitalium*. *J. Mol. Biol.* **351**, 749–762 [CrossRef Medline](#)
 20. Gao, P., Tang, Q., An, X., Yan, X., and Liang, D. (2011) Structure of HsdS subunit from *Thermoanaerobacter tengcongensis* sheds lights on mechanism of dynamic opening and closing of type I methyltransferase. *PLoS ONE* **6**, e17346–e17346 [CrossRef Medline](#)
 21. Lapkouski, M., Panjikar, S., Janscak, P., Smatanova, I. K., Carey, J., Ettrich, R., and Csefalvay, E. (2009) Structure of the motor subunit of type I restriction-modification complex EcoR124I. *Nat. Struct. Mol. Biol.* **16**, 94–95 [CrossRef Medline](#)
 22. Uyen, N. T., Park, S. Y., Choi, J. W., Lee, H. J., Nishi, K., and Kim, J. S. (2009) The fragment structure of a putative HsdR subunit of a type I restriction enzyme from *Vibrio vulnificus* YJ016: implications for DNA restriction and translocation activity. *Nucleic Acids Res.* **37**, 6960–6969 [CrossRef Medline](#)
 23. Csefalvay, E., Lapkouski, M., Guzanova, A., Csefalvay, L., Baikova, T., Shevelev, I., Bialevich, V., Shamayeva, K., Janscak, P., Kuta Smatanova, I., Panjikar, S., Carey, J., Weiserova, M., and Ettrich, R. (2015) Functional coupling of duplex translocation to DNA cleavage in a type I restriction enzyme. *PLoS ONE* **10**, e0128700 [CrossRef Medline](#)
 24. Davies, G. P., Martin, I., Sturrock, S. S., Cronshaw, A., Murray, N. E., and Dryden, D. T. (1999) On the structure and operation of type I DNA restriction enzymes. *J. Mol. Biol.* **290**, 565–579 [CrossRef Medline](#)
 25. Obarska-Kosinska, A., Taylor, J. E., Callow, P., Orłowski, J., Bujnicki, J. M., and Kneale, G. G. (2008) HsdR subunit of the type I restriction-modification enzyme EcoR124I: biophysical characterisation and structural modelling. *J. Mol. Biol.* **376**, 438–452 [CrossRef Medline](#)
 26. Kennaway, C. K., Taylor, J. E., Song, C. F., Potrzebowski, W., Nicholson, W., White, J. H., Swiderska, A., Obarska-Kosinska, A., Callow, P., Cooper, L. P., Roberts, G. A., Artero, J. B., Bujnicki, J. M., Trinick, J., Kneale, G. G., and Dryden, D. T. (2012) Structure and operation of the DNA-translocating type I DNA restriction enzymes. *Genes Dev.* **26**, 92–104 [CrossRef Medline](#)
 27. Piekarczyk, A., Kłyz, A., Kwiatek, A., and Stein, D. C. (2001) Analysis of type I restriction modification systems in the Neisseriaceae: genetic organization and properties of the gene products. *Mol. Microbiol.* **41**, 1199–1210 [Medline](#)
 28. Grinkevich, P., Iermak, I., Luedtke, N. A., Mesters, J. R., Ettrich, R., and Ludwig, J. (2016) pHluorin-assisted expression, purification, crystallization and X-ray diffraction data analysis of the C-terminal domain of the HsdR subunit of the *Escherichia coli* type I restriction-modification system EcoR124I. *Acta Crystallogr. F Struct. Biol. Commun.* **72**, 672–676 [CrossRef Medline](#)
 29. Kobe, B., Ve, T., and Williams, S. J. (2015) Fusion-protein-assisted protein crystallization. *Acta Crystallogr. F Struct. Biol. Commun.* **71**, 861–869 [CrossRef Medline](#)
 30. Frishman, D., and Argos, P. (1995) Knowledge-based protein secondary structure assignment. *Proteins* **23**, 566–579 [CrossRef Medline](#)
 31. Holm, L., and Laakso, L. M. (2016) Dali server update. *Nucleic Acids Res.* **44**, W351–W355 [CrossRef Medline](#)
 32. Konagurthu, A. S., Whisstock, J. C., Stuckey, P. J., and Lesk, A. M. (2006) MUSTANG: a multiple structural alignment algorithm. *Proteins* **64**, 559–574 [CrossRef Medline](#)
 33. Hubacek, J., and Glover, S. W. (1970) Complementation analysis of temperature-sensitive host specificity mutations in *Escherichia coli*. *J. Mol. Biol.* **50**, 111–127 [CrossRef Medline](#)
 34. Janscak, P., Abadjieva, A., and Firman, K. (1996) The type I restriction endonuclease R. EcoR124I: over-production and biochemical properties. *J. Mol. Biol.* **257**, 977–991 [CrossRef Medline](#)
 35. Sisáková, E., Weiserová, M., Dekker, C., Seidel, R., and Szczelkun, M. D. (2008) The interrelationship of helicase and nuclease domains during DNA translocation by the molecular motor EcoR124I. *J. Mol. Biol.* **384**, 1273–1286 [CrossRef Medline](#)
 36. Miesenböck, G., De Angelis, D. A., and Rothman, J. E. (1998) Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins. *Nature* **394**, 192–195 [CrossRef Medline](#)
 37. McPherson, A. (1982) *Preparation and Analysis of Protein Crystals*, pp. 19–49, Wiley, New York
 38. Cianci, M., Bourenkov, G., Pompidor, G., Karpics, I., Kallio, J., Bento, I., Roessle, M., Cipriani, F., Fiedler, S., and Schneider, T. R. (2017) P13, the EMBL macromolecular crystallography beamline at the low-emittance PETRA III ring for high- and low-energy phasing with variable beam focusing. *J. Synchrotron Radiat.* **24**, 323–332 [CrossRef Medline](#)
 39. Kabsch, W. (2010) XDS. *Acta Crystallogr. D* **66**, 125–132 [CrossRef Medline](#)

Novel domain of the motor subunit of Type I R-M system

40. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 [CrossRef Medline](#)
41. Matthews, B. W. (1968) Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497 [CrossRef Medline](#)
42. Vagin, A., and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022–1025 [CrossRef](#)
43. van Thor, J. J., Georgiev, G. Y., Towrie, M., and Sage, J. T. (2005) Ultrafast and low barrier motions in the photoreactions of the green fluorescent protein. *J. Biol. Chem.* **280**, 33652–33659 [CrossRef Medline](#)
44. Cowtan, K. (2008) Fitting molecular fragments into electron density. *Acta Crystallogr. D* **64**, 83–89 [CrossRef Medline](#)
45. Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 [CrossRef Medline](#)
46. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 [CrossRef Medline](#)
47. 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 [CrossRef Medline](#)
48. 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 [CrossRef Medline](#)
49. Krieger, E., Koraimann, G., and Vriend, G. (2002) Increasing the precision of comparative models with YASARA NOVA: a self-parameterizing force field. *Proteins* **47**, 393–402 [CrossRef Medline](#)
50. Sinha, D., Bialevich, V., Shamayeva, K., Guzanova, A., Sisakova, A., Csefalvay, E., Reha, D., Krejci, L., Carey, J., Weiserova, M., and Ettrich, R. (2018) A residue of motif III positions the helicase domains of motor subunit HsdR in restriction-modification enzyme EcoR124I. *J. Mol. Model.* **24**, 176 [CrossRef Medline](#)
51. Qi, D., and Scholthof, K. B. (2008) A one-step PCR-based method for rapid and efficient site-directed fragment deletion, insertion, and substitution mutagenesis. *J. Virol. Methods* **149**, 85–90 [CrossRef Medline](#)
52. Taylor, I., Patel, J., Firman, K., and Kneale, G. (1992) Purification and biochemical characterisation of the EcoR124 type I modification methylase. *Nucleic Acids Res.* **20**, 179–186 [CrossRef Medline](#)
53. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors. *Gene* **33**, 103–119 [CrossRef Medline](#)
54. Patel, J., Taylor, I., Dutta, C. F., Kneale, G., and Firman, K. (1992) High-level expression of the cloned genes encoding the subunits of and intact DNA methyltransferase, M. EcoR124. *Gene* **112**, 21–27 [CrossRef Medline](#)
55. Jacob, F., and Wollman, E. L. (1955) Etude génétique d'un bactériophage tempéré d'*Escherichia coli*: III: effet du rayonnement ultraviolet sur la recombinaison génétique. *Ann. Inst. Pasteur* **88**, 724–749
56. R Development Core Team (2017) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
57. Pagès, H., Aboyou, P., Gentleman, R., and DebRoy, S. (2017) *Biostrings: Efficient Manipulation of Biological Strings*, Version 2.48.0
58. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 [Medline](#)
59. Dereeper, A., Audic, S., Claverie, J. M., and Blanc, G. (2010) BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evol. Biol.* **10**, 8 [CrossRef Medline](#)
60. Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 [CrossRef Medline](#)
61. Dryden, D. T., Murray, N. E., and Rao, D. N. (2001) Nucleoside triphosphate-dependent restriction enzymes. *Nucleic Acids Res.* **29**, 3728–3741 [CrossRef Medline](#)

Crystal structure of a novel domain of the motor subunit of the Type I restriction enzyme EcoR124 involved in complex assembly and DNA binding

Pavel Grinkevich, Dhiraj Sinha, Iuliia Iermak, Alena Guzanova, Marie Weiserova, Jost Ludwig, Jeroen R. Mesters and Rüdiger H. Etrich

J. Biol. Chem. 2018, 293:15043-15054.

doi: 10.1074/jbc.RA118.003978 originally published online July 27, 2018

Access the most updated version of this article at doi: [10.1074/jbc.RA118.003978](https://doi.org/10.1074/jbc.RA118.003978)

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

This article cites 58 references, 10 of which can be accessed free at <http://www.jbc.org/content/293/39/15043.full.html#ref-list-1>