Structure of the motor subunit of type I restriction-modification complex EcoR124I

Mikalai Lapkouski1,2, Santosh Panjikar3, Pavel Janscak4, Ivana Kuta Smatanova1,2, Jannette Carey5, Rüdiger Ettrich1,2 & Eva Csefalvay1,2

Type I restriction-modification enzymes act as conventional adenine methylases on hemimethylated DNAs, but unmethylated recognition targets induce them to translocate thousands of base pairs before cleaving distant sites nonspecifically. The first crystal structure of a type I motor subunit responsible for translocation and cleavage suggests how the pentameric translocating complex is assembled and provides a structural framework for translocation of duplex DNA by RecA-like ATPase motors.

Type I restriction-modification enzymes are molecular machines that recognize specific, typically asymmetric, short DNA sequences. Three subunits either act together as a typical methyltransferase or recruit two endonuclease motor subunits that translocate duplex DNA through the stationary enzyme, driven by helicase-like motors that consume about 60 ATP per base pair5–7. We report the crystal structure of motor subunit HsdR of the Escherichia coli pR124 plasmid-borne type I restriction-modification enzyme EcoR124I, revealing structural constraints on assembly and the mechanism of the translocation and cleavage complex.

EcoR124I HsdR with selenomethionine labeling was expressed in E. coli, and its structure was solved in complex with Mg2+–ATP at 2.6-Å resolution (Supplementary Table 1 and Supplementary Methods online). HsdR presents four globular domains in a square-planar arrangement, generating prominent grooves between adjacent domain pairs (Fig. 1a and Supplementary Fig. 1 online). Similarity (SSM, http://www.ebi.ac.uk) identifies its endonuclease domain (residues 13–260) and RecA-like helicase domains 1 and 2 (261–461 and 470–731); a helical domain (732–892) has no apparent structural relatives, and residues 893–1038 were confirmed present but disordered.

The endonuclease domain presents the typical zβz core of type II restriction enzymes. Asp151, Glu165 and Lys167 clustered opposite the helical domain are proposed to be involved in catalyzing DNA cleavage. Lys220 on z is 3.1 Å from N3 on the exposed edge of ATP bound at the helicase domains (Fig. 1b), potentially coupling endonuclease and helicase functions. Adenine stacks on Arg273 and is surrounded by other residues of motif Q. A variant DEAD box (408-DECHR-412) makes the closest approach to ATP, 4.85 Å from Glu409 O1 to ATP O1e. Unlike helicases that present a bridging water molecule7, the position of the Glu409 carboxylate is altered by interaction with Lys313, allowing a small change in the Glu409 Cβ–Cγ dihedral angle to bring O1γ to within 3.4 Å of ATP O1e.

A uniformly positive surface groove with a clear match to the size and shape of duplex DNA (Fig. 2a,b) proceeds from a canonical helicase cleft in a continuous path down the ‘front’ of the motor subunit between the helical and endonuclease domains, where the cleavage site is recessed slightly from the surface (Fig. 2c). The continuity of exonuclease III footprints of EcoKI–DNA complexes6–9 indicates that DNA follows the protein surface without looping; footprinting of EcoR124I used DNAs too short to accommodate HsdR8. DNA following the surface groove would be deflected 80° from linear, a degree found in other protein-DNA complexes. The helicase cleft superimposes with cocrystal structures of

1Department of Structure and Function of Proteins, Institute of Systems Biology and Ecology, Academy of Sciences of the Czech Republic, and 2Institute of Physical Biology, University of South Bohemia in Ceske Budejovice, Zamek 136, CZ-373 33 Nove Hrady, Czech Republic. 3EMBL Hamburg Outstation, Oo DESY, Notkestrasse 85, D22603 Hamburg, Germany. 4Institute of Molecular Cancer Research, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. 5Chemistry Department, Princeton University, Princeton, New Jersey 08544-1009, USA. Correspondence should be addressed to R.E. (ettrich@greentech.cz) or J.C. (jcarey@princeton.edu).

Received 7 July; accepted 29 October; published online 14 December 2008, doi:10.1038/nsmb.1523
Rad54–dsDNA\(^1\) and helicases complexed with single-stranded nucleic acids (helicase–ss)\(^1\) and would similarly contact the strand defined in those cocrystals as the 5′-to-3′ strand, which would also contact helix 29 in domain 2 via motif IVa; the 5′-to-3′ strand would contact domain 1 via motif IIa, as in Rad54–dsDNA. Superposition of the HsdR endonuclease core with type II restriction enzymes\(^6\) predicts that the major groove of the 5′-to-3′ strand faces the active site.

Although HsdR belongs to helicase superfamily 2 with RecA-like sequence motifs common to all helicases\(^1\), EcoR124I tracks mainly on EcoKI–DNA based on biochemical and genetic evidence. The axial helical domain near methyltransferase as in a cartoon model for Rad54–dsDNA\(^1\) and helicases complexed with single-stranded nucleic acids (helicase–ss)\(^1\) and would similarly contact the single strand of the helicase–ss cocrystal structure, suggesting that transport by Rad54 might use a helicase-like, ATP-driven screw mechanism\(^10\), but the extremely rotated state of crystalline Rad54 helicase domain 2 raised doubt about its similarity to helicases\(^11\). Transport direction in the helicases is known to occur from the point of strand separation toward the helicase cleft\(^12\), consistent with fluorescence resonance energy transfer (FRET) analysis of Rad54 translocation stages\(^13\). The path of dsDNA down the front of the HsdR motor subunit supports a common mechanism for helicases and translocases because it predicts transport toward the endonuclease active site, that is, in the same direction as proposed for Rad54.

Given the transport direction, DNA translocation toward the stationary enzyme\(^1,2,3\) requires that motor subunits flank methyltransferase with their helicase domains facing outward, bringing the HsdR helical domain near methyltransferase as in a cartoon model\(^14\) for EcoKI–DNA based on biochemical and genetic evidence. The axial orientation should align the HsdR surface groove with DNA emerging from methyltransferase, leading to the pentamer model shown in Supplementary Figure 2 online. Atomic force microscopy (AFM) images of EcoR124I initiation complexes show shortening of the DNA contour length by \(\sim 8\) nm with one motor subunit and by \(\sim 11\) nm with two\(^16\), representing contributions from structural alterations and two-dimensional projection. The observed shortening would be in agreement with a pentamer in which HsdR and HsdM abut with little or no overlap, bringing the two DNA bends to \(\sim 87\) bp apart and thus \(\sim 90^\circ\) out of the helical phase on average B-form DNA.

DNA exits the bottom of the motor subunit where clustered positive residues (Fig. 2d) could favor maintenance of the duplex and confine accumulating negative supercoils to the extruding loop\(^3,4\). Supercoil extrusion is likely to occur between the motor subunit and methyltransferase, where DNA topology can couple methylation status to translocation and cleavage\(^4\).

**Accession codes.** Protein Data Bank: Coordinates for EcoR124I HsdR have been deposited with accession code 2W00.

**Note:** Supplementary information is available on the Nature Structural & Molecular Biology website.

**ACKNOWLEDGMENTS**

We gratefully acknowledge support from the Ministry of Education, Youth and Sports of the Czech Republic (MSM600766808, LC06010), Academy of Sciences of the Czech Republic (AVOZ6070520), Grant Agency of the Czech Republic (203/08/0114 to R.E.), the European Molecular Biology Organization (to M.L.), the Swiss National Science Foundation (to P.J.), and joint Czech and US National Science Foundation International Research Cooperation (INT03-09049 to J.C.). We thank the European Molecular Biology Laboratory for access to the X12 beamline at the DORIS storage ring, DESY, Hamburg. J.C. and R.E. thank F. Hughson, T. Lohman, B. Matthews, L. Raleigh and S. Weller for insightful discussion during preparation of the manuscript.

Published online at http://www.nature.com/nsmb/

Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

---

Supplementary Information

Structure of the motor subunit of type I restriction-modification complex EcoR124I

Mikalai Lapkouski¹, Santosh Panjikar², Pavel Janscak³, Ivana Kuta Smatanova¹, Jannette Carey⁴*, Rüdiger Ettrich¹*, Eva Csefalvay¹

¹Department of Structure and Function of Proteins, Institute of Systems Biology and Ecology, Academy of Sciences of the Czech Republic; and Institute of Physical Biology, University of South Bohemia in Ceske Budejovice, Zamek 136, CZ-373 33 Nove Hrady, Czech Republic
²EMBL Hamburg Outstation, c/o DESY, Notkestrasse 85, D22603 Hamburg, Germany
³Institute of Molecular Cancer Research, University of Zürich, Wintherthurerstrasse 190, CH-8057 Zürich, Switzerland
⁴Chemistry Department, Princeton University, Princeton, New Jersey 08544-1009, USA

*corresponding authors
R. Ettrich, ettrich@greentech.cz FAX 420386361279
J. Carey, jcarey@princeton.edu FAX 1-609-258-6746
Supplemental Figure 1. Structure-based multiple alignment. Genomic sequences of HsdR subunits from EcoR124I, EcoAI (UniProt accession number Q07736), EcoKI (UniProt accession number P08956), StySBLI (REBASE Enz. Num. 2890), and KpnBI (REBASE Enz. Num. 2600) representing each of the sub-family types. Residue numbers above and secondary structures below are those of EcoR124I HsdR (bars, helices; arrows, strands). Unresolved residues are in italics. Inferred contacts are indicated by symbols above in domain color code: filled circles, interdomain contacts; triangles, Mg-ATP contacts coded to atomic colors of Figure 1. The extremely low homology among sub-families makes sequence alignments nearly impossible even when aided by the structure. Thus alignment took account of additional constraints besides sequence identity and similarity scores: motif sequences were aligned as closely as possible; secondary structure elements were not interrupted; secondary and tertiary elements of the three-dimensional core of the endonuclease and helicase domains identified by structural
superposition were maintained; the approximate tertiary symmetry of the endonuclease and helical domains was preserved. **a. Endonuclease, helicase 1, and helicase 2 domains.** Grey shaded residues are conserved in canonical ATPase motifs whose extent and numbering in the EcoR124I HsdR sequence are indicated by bars with Roman numerals below. Helicase interdomain linker helix 22 is also grey. **b. Helical domain.**
Supplemental Figure 2. Pentamer/DNA complex. DNA as black backbone ribbons; extruded loops omitted for clarity. Proteins in semi-transparent surface outline: HsdR in domain colors; HsdS brown; HsdMh (helical) and HsdMc (catalytic) domains gold; dashes indicate uncertain relative positions of the domains (17).
**Supplementary Table 1** Data collection and refinement statistics

<table>
<thead>
<tr>
<th></th>
<th>R.EcoR124I</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P2₁2₁2₁</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
</tr>
<tr>
<td>(a), (b), (c) (Å)</td>
<td>123.91, 129.92, 161.01</td>
</tr>
<tr>
<td>(\alpha), (\beta), (\gamma) (°)</td>
<td>90</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>20.0-2.60 (2.64-2.60)</td>
</tr>
<tr>
<td>(R_{merge})</td>
<td>13% (49.7%)</td>
</tr>
<tr>
<td>(I / \sigma I)</td>
<td>17.9 (5.8)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>9.4 (9.1)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>2.6</td>
</tr>
<tr>
<td>No. reflections</td>
<td>758388</td>
</tr>
<tr>
<td>(R_{work} / R_{free})</td>
<td>22.1 / 26.5</td>
</tr>
<tr>
<td>No. atoms</td>
<td>14386</td>
</tr>
<tr>
<td>Protein</td>
<td>13740</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>62/2</td>
</tr>
<tr>
<td>Water</td>
<td>582</td>
</tr>
<tr>
<td>(B)-factors</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>40.91</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>26.95/12.5</td>
</tr>
<tr>
<td>Water</td>
<td>43.77</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.014</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.4</td>
</tr>
</tbody>
</table>

* Values in parentheses are for highest-resolution shell.
Supplementary methods

Protein production, crystallization, and diffraction

Plasmid pTrcR124\textsuperscript{1} expressing HsdR under lac repressor control in methionine-auxotrophic *E. coli* B834 (DE3) (Novagen) was used to produce the full-length EcoR124I HsdR, residues 1-1038 (UniProt Q304R3), in M9 minimal medium containing L-selenomethionine (50 µg ml\textsuperscript{-1}; Sigma) using procedures reported for unlabeled protein\textsuperscript{2}. Purified HsdR at ~15mg/mL in 20 mM potassium phosphate, pH 7.5, 100mM NaCl, 5mM disodium ATP was stored at 250 K. Crystals grew in sitting drops (1 µl protein, 2 µl reservoir solution 0.2M Li\textsubscript{2}SO\textsubscript{4}, 8% PEG20000, 8% PEG550 monomethyl ether, 1.5mM dithiothreitol) to ~200x100x50 µm in one week at 277K. Crystals cryoprotected in 18% 2-methyl-2,4-pentanediol were flash-cooled in liquid nitrogen for data collection at 100K at the experimentally-determined selenium absorption edge on tunable beamline X12 at EMBL Hamburg Outstation.

Diffraction Analysis

460 frames of data were collected at 12.671 keV (0.9784 Å) on an MAR Mosaic 225mm CCD area detector. Images were processed using *DENZO* and scaled using *SCALEPACK*\textsuperscript{3}, in space group *P2\textsubscript{1}2\textsubscript{1}2\textsubscript{1}2\textsubscript{1}* to 2.6 Å. Heavy atoms sites determination, phase calculation using SAD, solvent flattening, and partial model building were performed automatically by *SHELXD*\textsuperscript{4}, *SHARP*\textsuperscript{5}, *DM*\textsuperscript{6}, and *ARP/wARP*\textsuperscript{7}, respectively, within the software pipeline AutoRickshaw\textsuperscript{8}. Further manual model building was performed using *Coot*\textsuperscript{9}. Initial rounds of refinement were carried out applying restrained noncrystallographic symmetry using the CNS program package\textsuperscript{10}. ATP molecules were manually built using COOT in the difference map \( (F_o - F_c) \). An anomalous difference
Fourier peak near the ATP was interpreted as Mg$^{2+}$ on the basis of its coordination number and geometry. Parameter and topology files for ATP were generated using PRODRG$^{11}$ and used in CNS refinement. Finally, 10 cycles of restrained and TLS$^{12}$ refinement were performed using REFMAC5$^{13}$ with starting atomic residual isotropic $B$ factors 20 Å$^2$. TLS groups were calculated by analyzing the spatial distribution of individual atomic thermal parameters using TLSMD web server$^{14}$. The final $R_{\text{cryst}}$ and $R_{\text{free}}$ were 22.1% and 26.5% respectively. The structure was analyzed using PROCHECK$^{15}$. The structure has two monomers in the asymmetric unit. Hydrodynamic methods and small-angle neutron scattering indicate HsdR is monomeric in solution$^{16,17}$. Approximately 150 C-terminal residues of each monomer confirmed to be present in dissolved crystals by gel electrophoresis (data not shown) are not resolved in the structure; several short disordered regions are also present. One ATP molecule and one Mg$^{2+}$ ion are associated with each HsdR monomer. The presence of unhydrolysed ATP is consistent with the weak ATPase activity of the isolated subunit$^{16}$.

**Loop Modeling**

The seven-residue loop Ala863 to Ser869 was modeled using the ModLoop routine for automated modeling of loops in protein structures$^{18}$. The structurally-resolved anchoring atoms of the protein backbone are separated by ~18.0 Å. Given that one residue contributes ~3 Å in backbone length, the number of degrees of freedom for a loop of seven residues is significantly constrained and the backbone positions of individual residues in the loop are predicted with high confidence.
Supplementary References