Structure and Function of the *Escherichia coli* RecE Protein, a Member of the RecB Nuclease Domain Family*

Received for publication, September 6, 2001
Published, JBC Papers in Press, October 4, 2001, DOI 10.1074/jbc.M108627200

Hoshing Wan Chang and Douglas A. Julin‡

From the Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742

The RecB subunit of the *Escherichia coli* RecBCD enzyme has both helicase and nuclease activities. The helicase function was localized to an N-terminal domain, whereas the nuclease activity was found in a C-terminal domain. Recent analysis has uncovered a group of proteins that have weak amino acid sequence similarity to the RecB nuclease domain and that are proposed to constitute a family of related proteins (Aravind, L., Walker, D. R., and Koonin, E. V. (1999) *Nucleic Acids Res.* 27, 1223–1242). One is the *E. coli* RecE protein (exonuclease VIII), an ATP-independent exonuclease that degrades the 5′-terminated strand of double-stranded DNA. We have made mutations in several residues of RecE that align with the critical residues of RecB, and we find that the mutations reduce or abolish the nuclease activity of RecE but do not affect the enzyme binding to linear double-stranded DNA. Proteolysis experiments with subtilisin show that a stable 34-kilodalton C-terminal domain that contains these critical residues has nuclease activity, whereas no stable proteolytic fragments accumulate from the N-terminal portion of RecE. These results show that RecE has a nuclease domain and active site that are similar to RecB, despite the very weak sequence similarity between the two proteins. These similarities support the hypothesis that the nuclease domains of the two proteins are evolutionarily related.

The RecBCD enzyme from *Escherichia coli* and other bacteria is a nuclease and a DNA helicase that plays an important role in homologous recombination and recombinational DNA repair (reviewed in Ref. 1). The RecB subunit of this enzyme by itself catalyzes both DNA unwinding (2) and DNA cleavage (3, 4), although at greatly reduced rates compared with RecBCD. These two activities are catalyzed by independent structural domains of RecB (3). The N-terminal three-quarters of the protein has helicase activity, whereas the ~30-kDa C-terminal portion of RecB constitutes a separate domain that exhibits very weak endonuclease activity (3, 5). Despite its very low nuclease activity, the C-terminal domain of RecB is critical for RecBCD enzyme activity, because mutations in this part of RecB abolish the nuclease activity of RecBCD (4, 6).

The N-terminal helicase domain of RecB is clearly related by amino acid sequence similarity to a large group of DNA helicases that includes the Rep, UvrD, and PerA helicases from bacteria as well as enzymes from eukaryotic organisms and viruses (7). On the other hand, the RecB nuclease domain bears no obvious relation to any other nuclease. The only proteins that have significant similarity scores in a simple BLAST data base search that uses the nuclease domain from *E. coli* RecB as the query sequence (residues 930–1180 of RecB) are RecB homologues from other bacteria. This analysis thus sheds no light on the evolutionary origin of the nuclease domain in RecB nor on whether any other enzymes have domains that are related to the RecB nuclease domain.

More sophisticated sequence analysis by Koonin and coworkers (8) using PSI-BLAST and other programs has uncovered a weak similarity between the C-terminal region of RecB and a number of other sequences in the GenBankTM data base. The set of 33 proteins that turned up was dubbed the “RecB nuclease domain family” and is proposed to constitute a group of proteins that are related structurally and evolutionarily. The structures of five of these proteins are shown schematically in *Fig. 1A*, and the amino acid sequence that defines the “RecB family” is given in *Fig. 1B*. Some of the family members are hypothetical and uncharacterized open reading frames from completely sequenced genomes (e.g. the sequence shown from *Archaeoglobus fulgidus*), but a few are proteins that have been studied previously. AddA is a RecB homologue from *Bacillus subtilis* that assembles with AddB to make a two-subunit enzyme (AddAB) that performs essentially the same enzymatic and biological functions as RecBCD (9). The eukaryotic Dna2 protein has both nuclease and helicase activities (10, 11), and the enzyme encoded by the *recE* gene is a dsDNA-specific nuclease known as exonuclease VIII (12).

The *recE* gene is situated within a cryptic prophage called *rac* that is found in the chromosome of some *E. coli* strains (13, 14). The RecE protein and the RecT protein, encoded by a neighboring gene and functionally similar to the *E. coli* RecA protein, can substitute for RecBCD and RecA in some types of homologous recombination (15). The RecE enzyme is a 5′–3′-specific exonuclease that degrades processively one strand of a linear dsDNA substrate to mononucleotide products (16, 17). RecE (866 amino acid residues) is quite a bit larger than the RecB nuclease domain (253 residues), but the sequence that RecE has in common with RecB is found near the C terminus of RecE (Fig. 1A) (8). Previous work using deletion mutants has shown that a large fraction of the N terminus of RecE can be deleted without loss of nuclease activity (18–20).

The overall sequence similarity among the proteins in the putative RecB family is very low, and thus the question arises as to its structural and mechanistic significance. The answer must come from biochemical and structural study of these...

* This work was supported by Grant GM39777 from the National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed. Tel.: 301-405-1821; Fax: 301-314-9121; E-mail: dj13@umail.umd.edu.

---

1 The abbreviations used are: dsDNA, double-stranded DNA; bp, base pair; HisRecE, hexahistidine-tagged RecE protein; MALDI-TOF, matrix-assisted laser desorption/ionization-time of flight; nt, nucleotide residue; PCR, polymerase chain reaction; NTA, nitrilotriacetic acid; ssDNA, single-stranded DNA; DTT, dithiothreitol.
proteins. We have begun by carrying out mutagenesis and proteolysis experiments with the *E. coli* RecE protein. First, we altered by site-directed mutagenesis the residues that are conserved among the RecB family members and determined their importance in the nuclease reaction. Second, we investigated the domain structure of RecE by limited proteolysis. The results confirm the relation between RecE and RecB that was reported by the manufacturers.

**EXPERIMENTAL PROCEDURES**

**Materials**

Restriction enzymes, calf intestinal phosphatase, and T₄ DNA ligase were obtained from Promega or New England Biolabs and were used as recommended by the suppliers. Subtilisin Carlsberg and trypsin were purchased from Sigma, and thrombin was from Novagen, Inc. Turbo *Pfu* DNA polymerase for PCR was obtained from Stratagene. Oligonucleotides for PCR and mutagenesis were obtained from Life Technologies, Inc. The plasmid pRAC31 that contains the *E. coli recE* gene (13) was a gift from Prof. A. J. Clark, University of California, Berkeley. ³²P-Labeled plasmid DNA (pTZ19R-recE<sub>recE</sub>) was isolated from cultures of *E. coli* strains JM109 or XL-1 grown as described previously (22). The parental DNA was purified using the Qiagen Plasmid Maxi kit (Qiagen, Inc.) according to the manufacturer’s protocol.

**Methods**

**Construction of a Plasmid Expressing the His-tagged RecE Protein (pET15b-recE)**—The His-tagged RecE protein (HisRecE) was expressed using the vector pET15b (Novagen, Inc.). The *recE* gene was transferred to the vector as follows. A 2601-bp fragment from the plasmid pRAC31, which encodes the C terminus of RecE protein, was first subcloned to pET15b to produce pET15b-recE. The *recE* gene insert in pET15b-recE was sequenced completely.

**Site-directed Mutagenesis of the RecE Gene**—The site-directed mutagenesis was done using the QuikChange site-directed mutagenesis kit (Stratagene) by one of the following two procedures. 1) The pET15b-recE plasmid was used directly as the template for site-directed mutagenesis for the *Y785F* mutation. 2) A 1165-bp BamHI fragment, which encodes the C terminus of RecE protein, was first subcloned to pTZ19R to make pTZ19R-recE<sub>recE</sub>. This plasmid was then used as the template for the D748A, D759A, K761Q, and Y785N mutations. The different mutated 1165-bp fragments were ligated back into pET15b-recE to express the mutant proteins.

The oligonucleotides used for the mutagenesis are as follows: 5'-GT GGT TGC TCG CCT GGG GCC AAA ATC ATC CCT G and 5'-CGT AAG GAT ATT TTT GGC CGG ACA CCG AC for the D748A mutation; 5'-CAC TGG ATC ATG GCC GTG AAA ACT ACG GCG and 5'-CGG CTG ATG TTT CAC GGC CAT GAT CCA GTG for the D759A mutation; 5'-TCT GAG GAT CCA TCT TTC AGT GAC GGT TAT GAA GCA C and 5'-GGT TGC TTC ATA ACC GTC ACC ACT AQA GAA TGC ATC CTG AAC G for the Y785A mutation; and 5'-CGT GGT CAT GCA TCT AAC ACT GAC GGT TAT GAA GCA C and 5'-GGT TGC TTC ATA ACC GTC ACC ACT AQA GAA TGC ATC CTG AAC G for the Y785N mutation. The mutagenic nucleotides are underlined. The reaction conditions and subsequent manipulations were performed according to the manufacturer’s protocol. Plasmids containing the mutant genes were identified by restriction digestion and DNA sequencing.

**Purification of His-tagged RecE Proteins**—The HisRecE protein was expressed in *E. coli* strain BL21(DE3) transformed with pET15b recE. A 70-mL overnight culture in LB broth containing ampicillin (50 μg/ml) was transferred to 2 liters of the same medium. The cells were grown at 37 °C until the A₆₀₀nm was 0.5. Isopropyl-β-D-thiogalactopyranoside was then added to 1 mM, and growth was continued for 1 h before the cells were harvested. The cell pellet (8 g) was resuspended in 40 ml of native binding buffer (20 mM sodium phosphate, pH 7.8, 0.5 mM NaCl). A protease inhibitor mixture for polyhistidine-tagged proteins (0.4 ml, Sigma, Inc.) and lysozyme (100 μg/ml cell suspension) was added to the cell suspension followed by incubation on ice for 15 min. The mixture was sonicated until it was no longer viscous, and the cell debris was removed by the lysate by centrifugation at 39,000 × g for 20 min at 4 °C. Purity was assessed on SDS-PAGE of the supernatant.
The crude cell extract was applied to a 5-ml Ni²⁺-NTA column (ProBond resin, Invitrogen Corp.) in native binding buffer. The column was then washed with 5 volumes of native wash buffer (20 mM sodium phosphate, pH 6.0, 0.5 M NaCl, 40 mM imidazole), and the HisRecE protein was eluted in a gradient of 60 mM to 1 M imidazole in 20 mM sodium phosphate, pH 6.0, 0.5 M NaCl. The fractions containing HisRecE, based on analysis by SDS-PAGE, were collected and dialyzed against buffer A (20 mM Tris-Cl, pH 7.5, 1 mM DTT, 0.5 mM EDTA) containing 50 mM NaCl. The dialyzed pool was then applied to a 5-ml ssDNA agarose column (Amersham Pharmacia Biotech) in buffer A containing 50 mM NaCl and 60% (v/v) glycerol. The resulting protein solution was stored at -20°C. The HisRecE protein concentration was determined from the absorbance at 280 nm.

The concentration of the RecE proteins is always given as the monomer yield was about 0.45 mg from a 2-liter culture. The mutant HisRecE was purified using the same procedures.

MALDI-TOF Mass Spectrometry—The HisRecE protein was prepared as described above, except that the purified protein was dialyzed against ultrapure water (Amicon) into 10 mM NH₄HCO₃. A sample of HisRecE (18 mg/ml) was diluted 10-fold by mixing with 0.1% trifluoroacetic acid. The diluted protein was mixed with an equal volume (0.3 μl) of sinapinic acid solution (50 mM sodium phosphate, pH 3.0/3.2 v/v of acetone, 0.1% trifluoroacetic acid) and then characterized on a Kratos MALDI 4 TOF mass spectrometer equipped with a 337 nm ultraviolet laser (Kratos Analytical Instruments, UK). The mass spectrometer was externally calibrated using bovine serum albumin. This test was kindly done by Dr. Xudong Yao of Prof. Catherine Fenselau's research group at the University of Maryland.

Exonuclease Assays—The standard exonuclease reaction conditions were 20 mM Tris-Cl, pH 8.0, 10 mM MgCl₂, 1 mM DTT, at 37°C. The DNA substrate was [γ-³²P]labeled plasmid DNA (pTZ19R-recE genomic, or pPeVSm19) linearized by cleavage with HindIII. The RecE protein was diluted 10–50-fold when necessary, in a buffer containing 10 mM Tris-Cl, pH 7.5, 1 mM DTT, and 0.5 mg/ml bovine serum albumin and kept on ice. Nuclease activity was determined by measuring the production of trichloroacetic acid-soluble mononucleotide products (22). The concentration of the RecE proteins is always given as the monomer concentration, although the active form may be a larger oligomer (16) of subtilisin Carlsberg per mg of HisRecE and buffer A containing 50 mM NaCl and 60% (v/v) glycerol. The resulting protein solution was stored at -20°C. The HisRecE protein concentration was determined from the absorbance at 280 nm, using ε₂⁸⁰ 102,840 M⁻¹ cm⁻¹ calculated for HisRecE using the program ProtParam (ca.expasy.org/tools/protparam.html). The typical yield was about 0.45 mg from a 2-liter culture. The mutant HisRecE proteins and the C-terminal segment of RecE (see below) were purified using these same procedures.

DNA Binding Assays—DNA binding assays were done using the two-filter method developed by Wong and Lohman (23) in which DNA-protein complexes were trapped on a nitrocellulose filter (BA85, Schleicher & Schuell), and the unbound DNA is trapped on a DEAE filter (DE81, Whatman) put immediately beneath the nitrocellulose filter, as described. The standard binding mixtures contained 20 mM Tris-HCl, pH 8.0, 0.5 mM DTT, 100 μM MgCl₂, 1 mM EDTA, and [γ-³²P]labeled DNA (51.1 μM nt, equivalent to 12.7 nM ends). The protein and DNA were mixed and incubated for 2 min at room temperature before being applied to the nitrocellulose/DEAE filter sandwich.

The radioactivity bound to each dried filter was determined by liquid scintillation counting. The fraction of DNA bound to the nitrocellulose filter (f) was calculated as shown in Equation 1,

\[
f = \frac{cpm_{NC} - cpm_{ME}}{cpm_{NC} + cpm_{ME}} \quad (\text{Eq. 1})
\]

where cpm_{NC} and cpm_{ME} are the counts bound to the nitrocellulose and DEAE filters, respectively, for a given binding mixture, and \( \alpha \) indicates cpm_{NC}/cpm_{ME} for a mixture having no RecE enzyme. This corrects for the background of DNA that binds nonspecifically to the nitrocellulose filter in the absence of protein (23).

The filter binding data were then analyzed assuming that RecE molecules can bind independently to each end of the DNA with identical dissociation constant \( K_a \). The binding data were fit to Equation 2 (24) using the SigmaPlot program (SPSS, Inc.),

\[
f = 1 - \left( 1 + \left( \frac{[\text{RecE}]}{K_a} \right)^2 \right)^{-1} \quad (\text{Eq. 2})
\]

where [RecE] is the free RecE concentration, and \( \zeta \) is the probability that a DNA molecule with one RecE enzyme bound to one end will not be retained on the filter (the fitted values of \( \zeta \) were very small (10⁻⁴) for most binding experiments).

Limited Proteolysis of the HisRecE Protein and N-terminal Sequencing—The purified HisRecE protein was treated with 6.25 μg of subtilisin Carlsberg per mg of HisRecE in buffer A (20 mM Tris-Cl, pH 7.5, 1 mM DTT, 0.5 mM EDTA) containing 50 mM NaCl and 60% (v/v) glycerol, at room temperature. Samples were removed at the indicated times, quenched with 33 mM (final concentration) phenylmethylsulfonyl fluoride (Sigma), and analyzed by SDS-PAGE. HisRecE protein was also treated with trypsin (0.1 μg of trypsin per mg of HisRecE) under the same conditions, at room temperature for 0–60 min. Samples were quenched and analyzed as above. To prepare the samples for N-terminal sequencing, the digested protein bands in the unstained SDS-polyacrylamide gel were transferred to a polyvinylidene difluoride membrane (Millipore). The N-terminal sequencing was performed by Dr. Brian Martin (National Institutes of Health).

Construction of a Plasmid to Express HisRecE 34kDa—The 34-kDa C-terminal domain of RecE protein (HisRecE 34kDa) was also expressed using the vector pET15b (Novagen, Inc.). The truncated recE gene was transferred to the vector as follows. An 849-bp fragment from the plasmid pET15b-RecE was amplified by PCR. The downstream primer annealed at the 3’-end of the gene (primer recE xho1, see above), and the upstream primer annealed at nt 1692–1712 of the gene (primer recE-34 nde1, 5’-CAG GAA CAT ATG GAA CAT CCG CAC AAT GAG AAT GC). Primer recE-34 nde1 was partially complementary to the recE gene (codons 564–571, underlined) and introduced an NdeI site (bold) and an ATG start codon. The truncated recE gene fragment was inserted into pET15b as described above for the full-length recE gene. The plasmid containing the truncated gene was identified by restriction digestion and DNA sequencing.

RESULTS

Purification and Monomer Molecular Weight of the RecE Protein—A protein of about 120–140 kDa was observed by SDS-PAGE analysis of the lysate of cells transformed with pET15b-recE, the RecE overexpression plasmid, after induction with isopropyl-β-D-thiogalactopyranoside (not shown). This HisRecE protein can be purified in a single step by chromatography on Ni²⁺-NTA resin, and further purification can be achieved by chromatography on ssDNA-agarose. The apparent size of this protein on SDS-PAGE is larger than the calculated molecular weight of HisRecE (98 kDa; see Fig. 2). Native RecE protein was found previously to migrate on SDS-PAGE with an apparent molecular weight of 140 kDa (16), also greater than that predicted from the recE DNA sequence (96 kDa; (25)). The molecular mass of the purified HisRecE was determined by MALDI-TOF mass spectrometry to be 99,274.2 Da (data not shown), in good agreement for a protein of this size with the predicted molecular mass (98,516 Da). This result shows that HisRecE protein is not glycosylated.

Nuclease Activity of HisRecE Protein—The purified HisRecE protein has ATP-independent nuclease activity on linear DNA. The activity is high and is not inhibited by ATP. For example, a 1000-fold dilution of the purified enzyme is still active at 100-fold greater concentrations of ATP.
dsDNA (Fig. 3). The reaction time courses are linear for the first ~10 min, but then the reactions slow down considerably even though most of the substrate DNA has not been digested (Fig. 3). The reaction resumed if more enzyme was added (data not shown), indicating that the enzyme is unstable under the assay conditions, as observed previously with the native RecE protein (16). The amount of acid-soluble DNA product did not exceed about 50% of the total DNA substrate present, even in reactions with high concentrations of active HisRecE (not shown), consistent with degradation of only one strand of the duplex by RecE (16). The specific activity of purified HisRecE, determined from the linear part of the nuclease reaction time courses (0–4 min), was as high as 200 mol of nucleotides produced per mol of enzyme per min but varied somewhat among different enzyme preparations (~50–200 min⁻¹). The purified HisRecE protein retained its activity for at least 3 months when stored at ~20 °C in the storage buffer, suggesting that there was variable loss of activity during the purification procedure itself that accounted for the irregularity among the different preparations.

Mutagenesis of Residues in HisRecE That Are Conserved in the RecB Nuclease Domain Family—We made five mutations (D748A, D759A, K761Q, Y785F, and Y785N) in the residues of the RecE protein that are conserved among the RecB family members (Fig. 1). Asp-748, Asp-759, and Lys-761 correspond to the residues that are important for the RecBCD nuclease activities (4, 6). The mutant proteins (HisRecED748A, HisRecED759A, HisRecEY785F, and HisRecEY785N) were expressed and purified by chromatography on Ni²⁺-NTA and ssDNA-agarose exactly as for the wild-type protein. All tests with the mutant HisRecE proteins (nuclease and DNA binding) were done immediately after the proteins were purified (within 4 days or less), because at least one of the mutants (HisRecEK761Q) gradually lost the ability to bind dsDNA after about 1 month of storage at ~20 °C.

Nuclease Activities of the Mutant Enzymes—Nuclease assays were done on the purified mutant enzymes with linear dsDNA under the same reaction conditions used for the wild-type enzyme. The HisRecED748A, HisRecED759A, and HisRecEY785N mutants had no detectable nuclease activity in these reactions (Fig. 3A). The amount of acid-soluble DNA produced by these mutant enzymes was indistinguishable from the background of soluble radioactivity that was present initially in the reaction mixture (0.5–1.5% of the total radioactivity present), in reactions that were followed for 60 min. Wild-type HisRecE (19.3 nM) solubilized 10–20% of the linear DNA substrate in ~20 min under the same conditions (Fig. 3, A and B).

The nuclease activity of the HisRecEY785F mutant was similar to that of comparable amounts of the wild-type enzyme (Fig. 3B), and thus the mutation appeared to have essentially no effect on the enzyme activity (the variability of the wild-type specific activity does not allow a more rigorous comparison of their activities). The removal of the hydroxyl group of Tyr-785 (Fig. 3) thus did not have much effect on the exonuclease activity of HisRecE. The HisRecEY785N mutant had very low but detectable activity (~1% of the DNA was made acid-soluble in 60 min, after subtracting out the background (Fig. 3B)). Thus the hydroxyl group of Tyr-785 is not necessary for catalysis, but replacing the large aromatic ring of Tyr or Phe with the smaller and more polar Asn side chain may partially disrupt the structure of the enzyme leading to loss of activity. Interestingly, the conserved tyrosine in RecB (Tyr-1114; Fig. 1) is replaced by Tyr-785 in RecE, leading to the specificity of RecE for long DNA substrates.

The Mutant Enzymes Bind to the Ends of Linear dsDNA with Similar Affinity as the Wild-type Enzyme—Protein-DNA binding assays were done to test whether the mutant enzymes that were inactive as nuclease had retained the ability to bind to the DNA substrate. Ca²⁺ was used in the binding buffer instead of Mg²⁺ because the exonuclease activity of the RecE protein requires Mg²⁺ and is inhibited by Ca²⁺ (16). Even though the HisRecED748A, HisRecED759A, and HisRecEK761Q mutant proteins have lost the ability to hydrolyze DNA, they retain essentially the same DNA binding activity as the wild-type protein (Fig. 4). The equilibrium dissociation constants (Kd) were obtained by fitting the binding data to Equation 2 (see “Experimental Procedures”). The Kd values for the wild-type, HisRecED748A, HisRecED759A, and HisRecEK761Q mutant proteins were obtained by fitting the binding data to Equation 2 (see “Experimental Procedures”). The Kd values for the wild-type, HisRecED748A, HisRecED759A, and HisRecEK761Q mutant proteins were obtained by fitting the binding data to Equation 2 (see “Experimental Procedures”). The Kd values for the wild-type, HisRecED748A, HisRecED759A, and HisRecEK761Q mutant proteins were obtained by fitting the binding data to Equation 2 (see “Experimental Procedures”). The Kd values for the wild-type, HisRecED748A, HisRecED759A, and HisRecEK761Q mutant proteins were obtained by fitting the binding data to Equation 2 (see “Experimental Procedures”). The Kd values for the wild-type, HisRecED748A, HisRecED759A, and HisRecEK761Q mutant proteins were obtained by fitting the binding data to Equation 2 (see “Experimental Procedures}).
proteins under these conditions are 182, 154, 85, and 77 nM, respectively, from duplicate determinations. The wild-type enzyme bound specifically to the DNA ends under these conditions, because very little (<1%) of uncut circular plasmid DNA was retained on the filter with 400 nM HisRecE (data not shown).

The Nuclease Active Site of the RecE Protein Resides in a C-terminal Domain of the RecE Protein—Truncated RecE proteins, in which as many as 587 residues are deleted from the N terminus and the recE gene is fused to an upstream gene, retain high levels of nuclease activity (18, 19) and recombination function in vivo (20). We sought to test for the existence of separate structural domains in RecE and further define their limits by proteolysis experiments. HisRecE was treated with either trypsin or the nonspecific protease subtilisin (26), and samples were taken at various times and analyzed by SDS-PAGE. Several large fragments in the range 50–70 kDa were seen after short digestion times with subtilisin (Fig. 5 and data not shown), indicating that there is no single site that is particularly prone to proteolysis. These larger fragments were themselves cleaved by subtilisin, and they disappeared at later times. Two small fragments of about 35 kDa persisted and accumulated after lengthy digestion times (Fig. 5). Trypsin also degraded further, whereas a prominent band of 28 kDa, itself cleaved by subtilisin, and it disappeared at later times. Two small fragments of about 35 kDa persisted and accumulated after lengthy digestion times (Fig. 5). Trypsin also was retained on the filter with 400 nM HisRecE (data not shown).

The two polypeptide fragments of about 35 kDa that persisted after a prolonged (2-h) incubation with subtilisin (Fig. 5) were transferred to a polyvinylidene difluoride membrane and analyzed by automated Edman degradation. The resulting N-terminal sequences (ENDPEEMEGAEH and EHPHNNAG) correspond to residues 554–565 and 564–572 of RecE protein, respectively. The calculated molecular masses of these two polypeptide fragments were 35 and 34 kDa, assuming that they extend to the C terminus of RecE.

The C-terminal 34-kDa Domain of RecE Is an Exonuclease—The 34-kDa fragment binding with residue 564 (HisRecE34kDa) was overexpressed and purified as for the full-length enzyme. Unlike the full-length enzyme, HisRecE34kDa migrates with the expected molecular weight on the SDS gel (Fig. 2). Thus, the 60-kDa N-terminal segment of the RecE protein contains sequences that cause the aberrant migration of RecE on the SDS gel, as suggested previously (19, 25). The truncated HisRecE34kDa has exonuclease activity with linear DNA (Fig. 6). The specific activity of HisRecE34kDa was about the same as we observe for the full-length protein (100–400 mol of nucleotides per mol enzyme per min) and varied somewhat among different HisRecE34kDa enzyme preparations, as described above for the full-length enzyme. The amount of acid-soluble DNA produced by HisRecE34kDa did not exceed more than about 50% of the total DNA substrate (Fig. 6), consistent with degradation of only one strand by HisRecE34kDa.

These results suggest that the RecE protein consists of a loosely structured N-terminal region that is quite susceptible to proteolytic degradation, and a more rigidly structured C-terminal domain of about 34 kDa that contains the exonuclease active site and is relatively resistant to proteolysis. The results from deletion experiments (20) suggest that smaller protein fragments can also be active, but the proteolysis indicates that the −34-kDa fragment may represent an independently folded structural unit within the larger RecE protein.

**DISCUSSION**

**Mutagenesis Experiments Support the Existence of the RecB Nuclease Domain Family**—The results of the mutagenesis experiments in this and previous reports (3, 4, 6) show that the*E. coli* RecE and RecB proteins share an array of amino acid residues that are critical for their nuclease activities in isolatable domains of about the same size. Mutagenesis experiments were reported recently for a third member of the RecB nuclease domain family, the Dna2 protein of*Saccharomyces cerevisiae*(27, 28). The residues in Dna2 that correspond to Asp-748, Asp-759, and Tyr-785 in RecE (see Fig. 1B) were found to be essential for the nuclease activity of Dna2, supporting its inclusion in the RecB family. The domain structure of Dna2 has not been investigated biochemically, and so it is not known whether its nuclease activity can be localized to a smaller module of this rather large protein (170 kDa in*S. cerevisiae*(28)). Together these results support the hypothesis that the three proteins are related to each other through common ancestry (8).
Active Site Residues of the RecE Exonuclease

RecE and RecB Nucleases, Similar Catalytic Sites but Very Different Enzymatic Activities—Although RecE and RecB have the same residues at their active sites, these two nucleases have very different enzymatic properties. First, RecE has much higher specific activity than RecB and especially than the RecB nuclease domain alone. The isolated RecB domain has barely detectable nuclease activity (~0.002 phosphodiester bonds cleaved per h per protein molecule (5)), whereas the full-length RecB protein cut about 0.5 bonds per h (4). Both are strikingly less than RecE (~100–1000 cleavage per min). RecE and RecB also have different substrate specificity and products. RecE is an exonuclease that releases mononucleotides from the 3’-terminated strand of a double-stranded substrate (17), whereas RecB cleaves single-stranded DNA endonucleolytically (4). RecBCD has much greater nuclease activity than RecB (4), but unlike RecE, it degrades dsDNA to short single-stranded oligonucleotide fragments rather than to mononucleotides (29), and it is able to cleave either strand of a dsDNA molecule (30). Mutagenesis experiments indicate that RecB carries the only nuclease active site in RecBCD (6), and so the activity and specificity of the RecB nuclease domain must be altered quite substantially once it has assembled with RecC and RecD to form RecBCD.

What Is the Relation of RecE to Other Bacteriophage Proteins?—We pointed out previously (6) that the residues shown by mutagenesis to be essential for nuclease activity in RecE, RecB, and Dna2 (the Asp…Asp-Xaa-Lys sequence, Fig. 1B) are similar to a motif found in a number of other nucleases including several restriction endonucleases, the 5’–3’-exonuclease of bacteriophage λ, and MutH: Pro-Asp…Asp(Xaa,Glu)-Xaa-Lys (31–33). These latter enzymes have little if any sequence similarity with each other, but several have been found to have similar three-dimensional structures, at least in their active site vicinity (32–34).

A recent analysis by Koonin and colleagues (35) proposes a distant relationship between the RecB family and these other nucleases that have this active site motif, including the λ exonuclease, although there is no readily detectable sequence similarity between the RecB family and these other nucleases. This implies a distant connection between the λ exonuclease and RecE. Indeed, the active site motif of the λ exonuclease, Pro-Asp…Xaa, Glu-Leu-Lys (33), is very close to that of RecE (Pro-Asp…Xaa, Glu-Leu-Lys (33)). The RecE nuclease domain (303 amino acid residues) is also roughly the same size as λ exonuclease (226 residues). These observations thus are the first biochemical evidence of any structural relationship between RecE and the bacteriophage λ exonuclease. The affiliation of the two is not unexpected because both enzymes are 5’–3’-specific exonucleases that degrade one strand of a linear dsDNA substrate to mononucleotide products (17). Moreover, the rac prophage that encodes RecE is a lambda phage (14, 15), and the RecE protein most likely served the same function for the progenitor of that phage as does the λ exonuclease (encoded by the recA gene of bacteriophage λ). On the other hand there was no previous hint of such a connection from either sequence homology searches (8, 15, 19) or antibody inhibition experiments (12).

The function, if any, of the large N-terminal extension of RecE is unknown. There were no stable products from this region in the proteolysis experiments, and it is not required for either the nuclease or for the recombination activity of RecE (20). Perhaps the recE gene once encoded a smaller nuclease corresponding to the C-terminal domain of RecE (and about the same size as the RecB nuclease domain and the λ exonuclease) that became fused to another gene or to extraneous DNA as the rac prophage underwent genetic degradation after its irreversible entrapment in the chromosome. The recE gene is not found in most bacteria, and there are very few close homologues of E. coli RecE in the sequence data bases. Thus the question of whether the N-terminal extension is conserved cannot be addressed. However, an interesting RecE homologue has been identified in a phage-like genetic element from Legionella pneumophila (GenBankTM gene identifier = 13,186,141 (36)). This 32-kDa (280 residues) protein has 29% identity to residues 601–862 of the E. coli RecE protein (BLAST E value = 3 × 10^-21), including residues that align with the RecE/RecB nuclease active site motif. The L. pneumophila protein may be a nuclease that is a close relative of RecE but that includes only the nuclease domain and lacks the N-terminal extension. The protein has not yet been purified and tested for nuclease activity.

Evolutionary Implications—The designation of a set of proteins as a protein family based on detectable amino acid sequence similarity among them implies their evolutionary descent from a common ancestor (37, 38). On this basis, the RecB nuclease domain is an evolutionarily mobile domain or protein “module” as defined in Ref. 39 that is found in proteins from Eubacteria, Archaea, bacteriophage, and Eukaryota. The sequence analysis predicts that the nuclease module has been joined to domains with other functions (helicase) in at least two family members (8), because Dna2 has a helicase domain like that in RecB (Fig. 1A), and the isolated Dna2 protein has been reported to unwind dsDNA (10). That the nuclease domain appears in both a bacterial protein (RecB) and a phage protein (RecE) is presumably an example of the frequent exchange of genetic information that occurs between phages and their bacterial hosts (40, 41).

An alternative explanation for the similarity between RecE and RecB is convergent evolution, whereby a small number of residues suitable for catalyzing DNA cleavage (i.e. Asp-Asp-Lys) arose independently in two otherwise unrelated ancestral proteins. It has been argued that convergent evolution to produce proteins with significantly similar primary sequences (as opposed to similar three-dimensional structures without primary sequence similarity) is very unlikely (42, 43). However, there is a “twilight zone” at low percent sequence identity that is more difficult to judge, and the use of sensitive sequence comparison programs that detect very subtle sequence similarities might group some sequences together in a family that are in fact examples of convergent evolution (42, 43). On the other hand, it is clear that homologous proteins arising from divergent evolution may have no detectable sequence similarity (39, 44, 45), and thus the low overall sequence similarity of RecE to
ReceB cannot be taken to disprove the supposition that they have common ancestry. High resolution structure information for RecE and ReceB (and other family members) could clinch the case for divergent evolution, because true homologues should have similar folds with the active site residues in comparable geometric arrangements on corresponding secondary structural elements (46–49).

Conclusion—Genome sequencing projects are adding tremendous numbers of new open reading frame sequences to the data bases, many of which encode proteins that appear to be unlike any that have been studied before. Sequence comparison methods are critical for predicting possible functions for these hypothetical proteins. At the same time, great effort is ongoing to refine the sequence analysis programs so that they will detect ever more distantly related protein homologues based on subtle sequence similarities (50, 51), such as those that define the ReceB nuclease domain family shown in Fig. 1B. The results reported here on RecE and those with Dna2 cited above show that these analysis methods have correctly identified critical residues in these proteins despite their low sequence similarity. Most other members of the ReceB nuclease domain family are known so far only as hypothetical open reading frames that have not been studied (8). Future work may show whether these novel proteins are in fact nucleases and what their biological functions are.

Acknowledgments—We are grateful to Prof. A. J. Clark, University of California, Berkeley, for the gift of pRAC31, to Dr. Brian Martin, National Institutes of Health, for N-terminal sequencing, and to Dr. Xudong Yao and Prof. Catherine Fenselau of our department for mass spectrometry.

REFERENCES
1. Kowalczykowski, S. C. (2000) Trends Biochem. Sci. 25, 156–165
2. Boehmer, P. E., and Emmerong, P. T. (1992) J. Biol. Chem. 267, 4981–4987
3. Yu, M., Souaya, J., and Julian, D. A. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 981–986
4. Yu, M., Souaya, J., and Julian, D. A. (1998) J. Mol. Biol. 283, 797–808
5. Zhang, X. J., and Julian, D. A. (1999) Nucleic Acids Res. 27, 4200–4207
6. Wang, J., Chen, R., and Julian, D. A. (2000) J. Biol. Chem. 275, 507–513
7. Koonin, E. V. (1992) Trends Biochem. Sci. 17, 465–497
8. Aravind, L., Walker, D. R., and Koonin, E. V. (1999) Nucleic Acids Res. 27, 1233–1242
9. Kooistra, J., Hjilene, B. J., and Venema, G. (1993) Mol. Microbiol. 7, 915–923
10. Budd, M. E., Choe, W. C., and Campbell, J. L. (1995) J. Biol. Chem. 270, 26766–26769
11. Bae, S. H., Choi, E., Lee, K. H., Park, J. S., Lee, S. H., and Seo, Y. S. (1998) J. Biol. Chem. 273, 26880–26890
12. Kushner, S. R., Nagaishi, H., and Clark, A. J. (1974) Proc. Natl. Acad. Sci. U. S. A. 71, 3593–3597
13. Willis, D. K., Satin, L. H., and Clark, A. J. (1985) J. Bacteriol. 162, 1166–1172
14. Campbell, A. M. (1996) in Escherichia coli and Salmonella typhimurium (Neishartd, F. C., ed) Vol. 2, pp. 2041–2046, American Society for Microbiology, Washington, D.C.
15. Kolodner, R., Hall, S. D., and Luisi-DeLuca, C. (1994) Mol. Microbiol. 11, 25–39
16. Joseph, J. W., and Kolodner, R. (1983) J. Biol. Chem. 258, 10411–10417
17. Joseph, J. W., and Kolodner, R. (1983) J. Biol. Chem. 258, 10418–10424
18. Luisi-DeLuca, C., Clark, A. J., and Kolodner, R. D. (1988) J. Biol. Chem. 170, 5797–5805
19. Chu, C. C., Templin, A., and Clark, A. J. (1989) J. Bacteriol. 171, 2101–2109
20. Muyrers, J. P., Zhang, Y., Buchholz, F., and Stewart, A. F. (2000) Genes Des. 14, 1971–1982
21. Korangy, F., and Julian, D. A. (1992) J. Biol. Chem. 267, 1727–1732
22. Korangy, F., and Julian, D. A. (1992) J. Biol. Chem. 267, 1733–1740
23. Wong, I., and Lohman, T. M. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 5428–5432
24. Woodbury, C. P., and von Hippel, P. H. (1983) Biochemistry 22, 4730–4737
25. Clark, A. J., Sharma, V., Brounnewitz, S., Chu, C. C., Sandler, S., Satin, L., Templin, A., Berger, J., and Cohen, A. (1965) J. Bacteriol. 175, 7673–7682
26. Hubbard, S. J. (1998) Biochim. Biophys. Acta 1382, 191–206
27. Lee, K. H., Kim, D. W., Bae, S. H., Kim, J. A., Ryu, G. H., Kwon, Y. N., Kim, K. A., Koo, H. S., and Seo, Y. S. (2000) Nucleic Acids Res. 28, 2873–2881
28. Budd, M. E., Choe, W., and Campbell, J. L. (2000) J. Biol. Chem. 275, 16518–16529
29. Goldman, P. F., and Linn, S. (1972) J. Biol. Chem. 247, 1849–1860
30. Anderson, D. G., Churchill, J. J., and Kowalczykowski, S. C. (1997) Genes Cells 2, 117–128
31. Pingoud, A., and Jeltsch, A. (1997) Eur. J. Biochem. 246, 1–22
32. Ban, C., and Yang, W. (1998) EMBO J. 17, 1528–1534
33. Kovval, R. A., and Matthews, B. W. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 7893–7897
34. Aggarwal, A. K. (1995) Curr. Opin. Struct. Biol. 5, 11–19
35. Aravind, L., Makarova, K. S., and Koonin, E. V. (2000) Nucleic Acids Res. 28, 3417–3432
36. Luneberg, E., Mayer, B., Daryab, N., Kooistra, O., Zahringer, U., Rohde, M., Swanson, J., and Frosh, M. (2001) Mol. Microbiol. 39, 1259–1271
37. Murzin, A. G. (1990) Curr. Opin. Struct. Biol. 6, 396–394
38. Henkoff, S., Greene, E. A., Pietrukovski, S., Bork, P., Attwood, T. K., and Hood, L. (1997) Science 276, 609–614
39. Doolittle, R. F. (1995) Annu. Rev. Biochem. 64, 287–314
40. Hendrix, R. W., Smith, M. C., Burns, R. N., Ford, M. E., and Hafell, G. F. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 2192–2197
41. Moreira, D. (2000) Mol. Microbiol. 35, 1–5
42. Patterson, C. (1988) Mol. Biol. Ecol. 5, 603–625
43. Doolittle, R. F. (1994) Trends Biochem. Sci. 19, 15–18
44. Rost, B. (1997) Fold. Des. 2, 819–824
45. Ponting, C. P., Schults, J., Copley, R. R., Andrade, M. A., and Bork, P. (2000) Adv. Protein Chem. 54, 185–244
46. Orengo, C. A., Jones, D. T., and Thornton, J. M. (1994) Nature 372, 631–634
47. Murzin, A. G. (1998) Curr. Opin. Struct. Biol. 8, 380–387
48. Matsuo, Y., and Bryant, S. H. (1999) Proteins 35, 70–79
49. Wolf, Y. I., Grishin, N. V., and Koonin, E. V. (2000) J. Mol. Biol. 299, 897–905
50. Park, J., Karplus, K., Barrett, C., Hughey, R., Hausler, D., Hubbard, T., and Chothia, C. (1998) J. Mol. Biol. 284, 1201–1210
51. Schaffer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spouge, J. L., Wolf, Y. I., Koonin, E. V., and Altschul, S. F. (2001) Nucleic Acids Res. 29, 2994–3005
Structure and Function of the *Escherichia coli* RecE Protein, a Member of the RecB Nuclease Domain Family

Hoshing Wan Chang and Douglas A. Julin

*J. Biol. Chem. 2001, 276:46004-46010.*

doi: 10.1074/jbc.M108627200 originally published online October 4, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M108627200

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 51 references, 22 of which can be accessed free at 
http://www.jbc.org/content/276/49/46004.full.html#ref-list-1