JOURNAL OF BACTERIOLOGY, Jan. 1997, p. 310–316 Vol. 179, No. 2
0021-9193/97/$04.00+0

nfi, the Gene for Endonuclease V in *Escherichia coli* K-12

GUANGMING GUO, YUNHONG DING, AND BERNARD WEISS*

Department of Pathology, University of Michigan Medical School, Ann Arbor, Michigan 48109-0602

Received 12 September 1996/Accepted 1 November 1996

Endonuclease V is specific for single-stranded DNA or for duplex DNA that contains uracil or that is damaged by a variety of agents (B. Demple and S. Linn, J. Biol. Chem. 257:2848–2855, 1982). Thus, it may be a versatile DNA repair enzyme. The protein was purified to apparent homogeneity, and from its N-terminal sequence, its gene, *nfi*, was identified. *nfi* is immediately downstream of *hemE*, at kb 4208 (90.4 min) on the current chromosomal map of *Escherichia coli* K-12. This region was cloned, and plasmid insertion and deletion mutants were used to study its molecular organization. Although *nfi* is the third of four closely spaced, codirectional genes, it is expressed independently.

Endonuclease V of *Escherichia coli* (11, 12, 14) selectively cleaves untreated single-stranded DNA, duplex DNA containing uracil in place of thymine, and DNA that has been treated with acid, alkali, OsO_4, UV radiation, or 7-bromomethylbenz[*a*]anthracene. It appears to work processively, attacking at multiple lesions on one molecule before moving on to the next. Unlike the members of the glycosylase-endolyase class of repair enzymes that cleave DNA at damaged bases (28), endonuclease V does not appear to release free bases (specifically uracil) from DNA, and it produces 3′-hydroxyl and 5′-phosphoryl end groups. In contrast to the UvrABC system (28), which also recognizes lesions produced by a broad range of agents but consists of a high-molecular-weight protein complex, endonuclease V is a simple polypeptide of about 25 kDa. These properties suggest that endonuclease V may be a highly useful DNA repair enzyme that, because of its simplicity, might be primitive and therefore universal.

Unfortunately, study of this interesting enzyme has been hampered by an absence of mutants and by its scarcity. Homogeneous preparations of endonuclease V have not been produced, and it has been difficult to obtain large amounts of the enzyme because of its low concentration in the cell and its instability during purification. We now report our first steps in overcoming these difficulties. We describe the identification, cloning, and molecular organization of *nfi*, the gene for endonuclease V.

(While this manuscript was in review, we were made aware [18] that *nfi* was also identified as the structural gene for deoxyinosine 3′ endonuclease [41], an enzyme that cleaves DNA near dIMP residues, mismatched bases, urea residues, or abasic sites [39, 40]).

MATERIALS AND METHODS

Microbial strains and methods. The *E. coli* strains, phage, and plasmids used are listed in Table 1. Phage PBS2 and its host, *Bacillus subtilis* SB19, were obtained from A. Price (26). *E. coli* was grown in Luria-Bertani (LB) medium supplemented with antibiotics, as required, for plasmid maintenance (23). Plasmids containing sequences counterclockwise of *hemE* (pGG4, pGG15, pGG6, pGG10, and pGG13) were unstable and did not survive serial propagation even in antibiotic media; therefore, cultures were prepared from colonies of freshly transformed cells. Strains with plasmids carrying the *hemE* gene were photosensitive and were therefore protected from strong light. Phage P1 transduction (32) and transformations with plasmids (10, 16) were performed as described.

Molecular biological methods. Methods not specifically cited are to be found in references 2 and 29. To ligate incompatible DNA ends, they were first blunted by incubation with T4 DNA polymerase and the four deoxynucleoside triphosphates. The blunt ends of a partially deleted plasmid were joined directly, whereas those of separate fragments were joined after adding dTMP and dAMP, respectively, to the 3′ ends of the vector and insert DNAs (22).

DNA substrates. The following were as previously described: tritium labeling (15) and purification (2) of M13 phage and M13 plasmid (RFI) DNAs and the isolation of phage PBS2 DNA (26). Phage λ DNA was purchased from New England BioLabs.

Preparation of cell extracts. Strains bearing plasmid pUC19 or its derivatives were grown at 37°C in LB medium with ampicillin to a density of 5×10^8 ml^{-1}. For the induction of the expression vector pET-29c(+) and its derivative pGG21, the strains were grown at 37°C in LB medium with kanamycin to a density of 2×10^8 ml^{-1}. Isopropyl-1-β-D-thiogalactoside was added to 0.4 mM, and incubation was continued for 3 h. The cells were collected by centrifugation at 4°C, and extracts were prepared by sonication in 50 mM Tris-HCl buffer (pH 7.6) containing 1 mM sodium phosphate and 1 mM dithiothreitol.

Endonuclease V assays. The standard assay for endonuclease V detected breaks in single-stranded circular DNA through an increase in the susceptibility of the DNA to digestion by exonuclease I (Amersham Life Science Corp.). The assay was performed as previously described (11), except that phage M13 DNA was substituted for fd DNA. Where needed, yeast tRNA was added to a final concentration of 34 μg/ml to block interference by endonuclease I (20). One unit of enzyme is defined as the amount that produces 1 nmol of acid-soluble product in 30 min (11).

For the comparison of activities on single- and double-stranded DNA, the reaction mixtures (100 μl) contained 4 nmol of M13 or M13 RFI [^{3}H]DNA (12,000 cpm), 3.4 μg of yeast tRNA, and a portion of a sonicated cell extract (≤320 ng of protein) in the standard reaction solution consisting of 67 mM glycine-NaOH buffer (pH 9.5), 25 mM KCl, and 10 mM $MgCl_2$. After incubation of the tubes at 37°C for 30 min, 75 μl of 0.1 M NaOH was added with immediate mixing to denature the duplex DNA. After 2 min at room temperature, the pH was adjusted to 9.5 with 25 μl of 0.83 M glycine. Two units of *E. coli* exonuclease I was then added, and the mixtures were incubated at 37°C for 30 min. Acid-soluble radioactive material was measured as for the standard endonuclease V assay.

For the comparison of activities on thymine- versus uracil-containing linear DNA, endonucleolytic cleavages were detected by an increase in the ability of the substrates to serve as primers for DNA polymerase I. The reaction mixtures (50 μl) contained 0.5 μg of phage λ or PBS2 DNA, 3.4 μg of yeast tRNA, and sonicated cell extracts (0 to 80 ng of protein) in the standard reaction solution. After 15 min at 37°C, the tubes were incubated at 65°C for 10 min and chilled. The pH was adjusted to 7.5 with 13 μl of 0.1 M HCl, and other additions were made to produce the following concentrations in a final volume of 100 μl: 20 μM (each) dCTP, dTTP, and dGTP; 4 μM dATP; 50 mM Tris-HCl buffer (pH 7.5); 15 mM $MgCl_2$; 1 mM dithiothreitol; and 50 μg of bovine serum albumin per ml. Deoxyadenosine 5′-[α-^{35}S]thiotriphosphate (1 pmol, 1 μCi) and 1 U of *E. coli* DNA polymerase I (Boehringer Mannheim Co.) were then added. After incubation for 15 min at 37°C followed by 10 min at 65°C, 10-μl samples were applied to DEAE-cellulose filters (Whatman DE-81) and dried at room temperature. The filters were rinsed in 0.5 M sodium phosphate buffer (pH 7.0) and subjected to liquid scintillation counting.

Purification of endonuclease V. Late-log-phase cells of strain BW1138 were prepared by the fermentation facility at the University of Alabama at Birmingham. The growth medium contained the following per liter: tryptone, 10 g; yeast

* Corresponding author. Mailing address: Department of Pathology, University of Michigan Medical School, 1301 Catherine Rd., Ann Arbor, MI 48109-0602. Phone: (313) 764-2212. Fax: (313) 763-6476. E-mail: weiss@umich.edu.

TABLE 1. Bacterial strains, plasmids, and phage used

Strain, plasmid, or phage	Description	Source or reference
Strains[a]		
BD2008	Hfr KL16 *ung-151*::Tn*10 thi-1 relA1 spoT1*	13
BL21(DE3)	*E. coli* B; *hsdS gal* [λ *imm*21 *int*::Φ(*lacUV5*-T7 gene *1*)]	33
BW1032	BL21(DE3) *ung-151*::Tn*10*	P1(BD2008) × BL21(DE3)
BW1138	*ung-153*::*kan*[b] *endA thyA gal his thi sup*	Transductant of JC4582
DR1984	*uvrC34 recA1* plus markers of AB1157[c]	30
JC4582	*gal his thi thyA endA sup*	A. J. Clark
NovaBlue	*endA1 hsdR17 glnV44 thi-1 recA1 gyrA96 relA1 lac* (F′::Tn*10 proAB lacI*^{q}ΔM15)	Novagen Corp.
Plasmids		
pET-29c(+)	Expression vector with phage T7 promoter; Kan^r	Novagen Corp.
pGG4	4.8-kb segment of λ533 cloned in pUC19 via an *Eco*RI-*Kpn*I substitution; Amp^r	This study
pGG6	*cat* element[d] *in Bsp*EI site of *hemE* in pGG4; Amp^r Cam^r	This study
pGG10	*cat* element in second *Bsr*BI site of *nfi* in pGG4; *cat* gene codirectional with *nfi*; Amp^r Cam^r	This study
pGG13	As pGG10, but *cat* element in reverse orientation	This study
pGG15	*cat* element in *Bsr*BI site of *yjaD*; Amp^r Cam^r	This study
pGG21	pET-29c(+)::*nfi*; cloned PCR product; Kan^r	This study
pGG41	*Eco*RI-*Bse*RI deletion of pGG4; Amp^r	This study
pGG42	*Eco*RI-*Bsp*EI deletion of pGG4; Amp^r	This study
pGG43	*Nar*I-*Bst*BI deletion of pGG4; Amp^r	This study
pMOB02	*bla*::Tn*9* (*cat*); Amp^s Cam^r	7
pUC19	Cloning vector; Amp^r	38
Phage		
λ533	Recombinant λ phage 7B7 (miniset no. 533) containing 19 kb of *E. coli* DNA including the *nfi* region; *b*189 Δ(*int-c*III) KH54 [Δ(*c*I-*rex*)] *nin5*	17

[a] Unless otherwise stated, bacterial strains are derivatives of *E. coli* K-12 F$^-$ λ$^-$.
[b] *ung-153*::*kan* consists of a *Bam*HI fragment of plasmid pUC4K (36) inserted into a *Bam*HI site in *ung*. The allele was obtained from strain BD2639 provided by Bruce Duncan and transduced several times to separate it from linked auxotrophic markers.
[c] *cat* element, 1.87-kb *Fdi*II fragment of Tn*9*, obtained from pMOB02 and containing the chloramphenicol transacetylase gene.
[d] Markers of AB1157 (see reference 3).

extract, 5 g; NaCl, 5 g; glucose, 10 g; K_2HPO_4, 10 g; KH_2PO_4, 1.85 g; and thymine, 50 mg. The culture was grown at 37°C with aeration, and it was maintained at pH 7.2 by the addition of NaOH. In late log phase (A_{650} = 6.5), the cells were chilled, harvested by centrifugation, and washed and resuspended in an equal volume of 10% sucrose containing 10 mM Tris-HCl buffer (pH 7.0). They were then frozen in liquid N_2 and stored at −80°C.

Unless otherwise stated, subsequent procedures were performed at 0 to 4°C and centrifugations were performed at 25,000 × *g* for 20 min. A frozen suspension containing 200 g of cells was thawed and diluted to 1,200 ml with a solution containing 50 mM Tris-HCl buffer (pH 7.6), 1 mM sodium phosphate, and 1 mM dithiothreitol. Phenylmethylsulfonyl fluoride (5% solution in ethanol) was added to a final concentration of 0.002%, and the cells were disrupted by sonication in 200-ml batches at ≤8°C. Cell debris was removed by centrifugation, yielding fraction I. Solid NaCl was slowly added with stirring to a final concentration of 0.5 M, and 5% polyethyleneimine (adjusted to pH 7.0 with HCl) was added to a concentration of 0.3%. After 20 min, the suspension was centrifuged. Solid $(NH_4)_2SO_4$ was added slowly to the supernatant, with stirring, in the proportion of 36.6 g per 100 ml. After 20 min, the suspension was centrifuged, and the pellet was dissolved in PD (20 mM sodium phosphate buffer [pH 6.8], 1 mM dithiothreitol) and extensively dialyzed against PD containing 0.1 M NaCl to produce fraction II.

In the following procedures, all adsorbents were pretreated according to the manufacturer's directions. Phosphocellulose P-11 was obtained from Whatman. Other adsorbents were from Bio-Rad. In each chromatography step, buffer of the same composition was used to equilibrate the column (8 column volumes), to wash the column (1 volume), and to contain the NaCl gradient for elution (8 volumes). Before each protein sample was applied to a column, the concentrations of its other solutes were adjusted to match those of the equilibrating buffer.

A column (2 by 32 cm) of phosphocellulose P-11 was equilibrated with PD containing 0.1 M NaCl. Fraction II was applied and eluted at 1 ml/min with a gradient of 0.1 to 0.7 M NaCl. Active fractions (at 0.350 to 0.385 M) were pooled (fraction III).

Fraction III was applied to a column (1 by 25.5 cm) of hydroxylapatite (Bio-Gel HT) that had been equilibrated with PD containing 0.1 M NaCl and 1 mM dithiothreitol. Elution was performed at 0.3 ml/min with an NaCl gradient of 0.1 to 0.6 M. Active fractions (at 0.30 to 0.36 M) were pooled (fraction IV).

Fraction IV was applied to a column of the hydrophobic resin Macro-Prep *t*-butyl HIC that had been equilibrated with 1.2 M NaCl in PD. Elution was performed at 0.2 ml/min with a decreasing gradient of NaCl, from 1.2 to 0.1 M. Active fractions (at 0.86 to 0.74 M) were pooled (fraction V).

Fraction V was dialyzed at room temperature overnight against a solution containing 20 mM 2-amino-2-methyl-1-propanol–HCl buffer (pH 9.5), 6 M urea, and 1 mM dithiothreitol. The urea was deionized before use with RG 501-X8 mixed bed resin (Bio-Rad Laboratories). A column (0.5 by 8.2 cm) of Macro-Prep 50 Q anion-exchange resin was equilibrated with the same solution and operated at room temperature. Elution was performed with an NaCl gradient of 0.1 to 0.8 M at a flow rate of 0.1 ml/min. Active fractions (at 0.20 to 0.25 M) were pooled (fraction VI).

Isolation of endonuclease V by SDS-PAGE. Fraction VI was precipitated in cold 10% trichloroacetic acid and redissolved in 0.2 ml of a solution containing 50 mM Tris-HCl buffer (pH 8.8), 10 mM dithiothreitol, 1% sodium dodecyl sulfate (SDS), and 10% glycerol. After incubation at 30°C for 1 h, 20-μl samples were fractionated by discontinuous SDS-polyacrylamide gel electrophoresis (PAGE) in a Tris-glycine buffer system with a 15% separating gel. One lane of the gel was stained with Coomassie blue, a second was used to locate the enzyme activity, and the contents of two others were electroeluted onto a polyvinylidene difluoride membrane (19) for sequence analysis. To locate the enzyme, a lane of the 8-cm-long separating gel was sliced at 1.5-mm intervals. The pieces were ground and then shaken overnight at room temperature in TD (50 mM Tris-HCl buffer [pH 7.8], 1 mM dithiothreitol) containing 0.1 M NaCl and 0.1% Triton X-100. The eluates were dialyzed against TD containing 6 M urea for 1 h at room temperature. SDS was removed with Dowex-1-X2 resin, and the samples were dialyzed at 4°C for 2 h against TD containing 0.1 M NaCl and 10% glycerol.

Cloning of *nfi* by PCR. The *nfi* gene was amplified by thermal cycling with *Taq* DNA polymerase, using ATGGATCTCGCGTCATTAC and CAGTTTACCTG AATTAGGG as primers and *E. coli* K-12 chromosomal DNA as the template. Plasmid pET-29c(+) was cut with *Nde*I and *Bam*HI, blunted with mung bean nuclease, tailed with dTMP (22), ligated to the dAMP-tailed PCR product, and used to transform strain NovaBlue.

Other methods. Protein concentrations were measured with a Coomassie blue reagent (Pierce Chemical Co.). DNA and protein sequences were analyzed with software from the Genetics Computer Group, Madison, Wis. The N-terminal sequence of endonuclease V was determined with an Applied Biosystems Sequencer at the W. M. Keck Foundation, Yale University.

Nucleotide sequence accession number. The nucleotide sequences of *nfi* and its neighboring genes are in the GenBank-EMBL database under accession

TABLE 2. Purification of endonuclease V

Step	Vol (ml)	Total enzyme (10^3 U)	Sp act (10^3 U/mg of protein)	Yield (%)
I. Sonicate	1,090	1,242	0.084	100
II. Ammonium sulfate	149	1,230	0.126	99
III. Phosphocellulose	83	559	4.73	45
IV. Hydroxylapatite	10.6	137	37.6	11
V. *t*-Butyl HIC	1.2	28.7	174	2.3
VI. Macro-Prep 50	0.9	12.4[a]	729–6,200[a,b]	1.0

[a] In order to reflect the recovery of enzyme protein, which was the major object of this purification, these values for fraction VI were corrected for the loss of activity due to treatment with urea. The actual assay values were 20% of these numbers.

[b] Protein concentration was too low to be measured without sacrificing most of the sample. The range for the specific activity is based on an estimate of the sensitivity of the Coomassie blue assay and the amount (4 μg) determined by protein sequencing after electrophoresis.

number U00006, in which *yjaE*, *yjaD*, *yjaF* (*nfi*), and *yjaG* are referred to as ORF_f158, ORF_o257, ORF_o225, and ORF_o196, respectively.

RESULTS

Purification of endonuclease V. Our strategy was to purify the enzyme to homogeneity and determine a partial sequence that would help us to locate its gene. Our first purification scheme was similar to that originally described (11, 14). The purified enzyme preparation contained three proteins detectable by two-dimensional gel electrophoresis, only one of which yielded an unambiguous N-terminal sequence. The sequence matched a gene carried by the recombinant phage λ533 from the genomic library of Kohara et al. (17). Although this gene would be discovered later to encode endonuclease V, cells infected by the phage did not overexpress the activity. Therefore, we assumed that endonuclease V was a minor protein in our purified enzyme preparation, and we devised a new purification scheme (Table 2) to obtain a homogeneous preparation of it in a large enough yield to produce an unambiguous sequence.

The strain used as an enzyme source, BW1138, contained mutations in the genes for endonuclease I (*endA*) and for uracil-DNA glycosylase (*ung*), which otherwise would have interfered with measurements for DNase activity against single-stranded DNA and uracil-containing DNA, respectively. The major obstacle to purification was that the enzyme became denatured and precipitated at NaCl concentrations below 0.1 M, whereas higher salt concentrations interfered with its adsorption to anion exchangers. This problem was bypassed in step VI of the purification scheme by denaturing the enzyme with urea, in the presence of which it would remain soluble at low ionic strength. When transferred to the reaction mixture for the standard assay, the enzyme was found to have either retained or recovered about 20% of the activity it had before the denaturing treatment. We confirmed that the purified enzyme was relatively specific for single-stranded over double-stranded DNA and for uracil-containing over thymine-containing DNA (results not shown).

Fraction VI was analyzed by SDS-PAGE. It contained only one protein band detectable by staining with Coomassie blue. Its mobility was near that of rabbit muscle triosephosphate isomerase (26.6 kDa) and corresponded to that of endonuclease V activity recovered from slices of the gel obtained from a parallel lane. Unstained protein (0.4 μg) was recovered by electroelution and analyzed by partial Edman degradation.

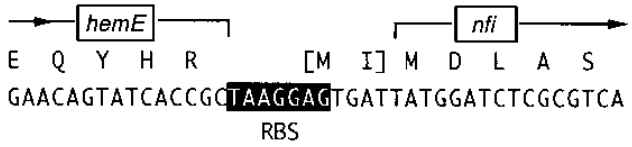

FIG. 1. The *hemE-nfi* intergenic region and neighboring sequences. The amino acids in brackets are translations of the first two codons of the *yjaF* open reading frame and are not found in the Nfi protein. RBS, putative ribosome binding site (highlighted region).

The gene for the endonuclease. The sequence of the first 11 amino acids of the purified protein was determined: MXLA SLRAQQI. It matched the sequence of a theoretical polypeptide starting at the third codon of *yjaF*, an open reading frame immediately downstream of the *hemE* gene (Fig. 1). We have designated the open reading frame starting from the third codon of *yjaF* as *nfi* (a mnemonic for endonuclease V). Its product, as deduced from its sequence (8), should have a molecular weight of 24,672 and a pI of 8.36. This value for the molecular weight is consistent with previous findings that endonuclease V activity sedimented with proteins of roughly 20 kDa and was copurified with a polypeptide that was about 27 kDa by SDS-PAGE (14).

Endonuclease V is the product of *nfi*. We had three pieces of evidence that the DNase we purified was that previously described as endonuclease V. (i) During electrophoresis, the Nfi protein comigrated with a single-stranded-DNA-specific endonuclease activity. (ii) The protein had the same N-terminal sequence as the major protein we had obtained earlier with the standard purification scheme for endonuclease V. (iii) Its molecular weight was consistent with values obtained by others. Additional evidence for its identity was provided by an analysis of the DNase activity of cells that overexpress it. The *nfi* gene was amplified by PCR and cloned in the expression vector pET-29c(+) to produce plasmid pGG21 (Table 1), in which the only cloned gene is *nfi*. The cloned fragment extended from the first codon of *nfi* to 13 nucleotides (nt) beyond its stop codon. The *nfi* gene was now regulated by the vector's phage T7 promoter, and its position downstream of the T7 gene *10* ribosome binding site was the same as that of gene *10* in the phage. The plasmid was introduced into strain BW1032, an *endA ung* host containing an inducible phage T7 RNA polymerase. After induction of the polymerase by isopropyl-1-β-D-thiogalactoside, endonuclease V activity increased to a level 200 times that of the plasmid-free cell, as measured by the standard assay. The induced DNase was tested for its substrate specificity. Its activity on the single-stranded circular DNA of phage M13 was compared to that on the double-stranded circular form, or RFI (Table 3). The standard assay for endonu-

TABLE 3. Endonuclease activity produced by an *nfi* plasmid: specificity for single- versus double-stranded DNA

Substrate	DNase activity in strain bearing plasmid (cpm/ng of protein)[a]:		
	pET-29::*nfi*[b]	pET-29	Δ[c]
M13 DNA	32	0.8	31
M13 RFI DNA	0.7	0.3	0.4

[a] The relative rates of endonucleolytic cleavage were determined by measuring the susceptibility of the product to exonucleolytic degradation (Materials and Methods).

[b] pET-29::*nfi* is pGG21; pET-29 is pET-29c(+). The host strain was BW1032.

[c] Δ, the difference in the results obtained with extracts of the two strains, indicating *nfi*-specified activity.

TABLE 4. Endonuclease activity produced by an *nfi* plasmid: specificity for uracil- versus thymine-containing DNA

Substrate	DNase activity in strain bearing plasmid (cpm/ng of protein)[a]:		
	pET-29::*nfi*	pET-29	Δ[b]
PBS2 DNA	689	13	676
λ DNA	24	4	20

[a] The relative rates of endonucleolytic cleavage were measured by determining the increase in the ability of the substrate to serve as a primer for nick translation by DNA polymerase I (Materials and Methods). The cell extracts were those used for Table 3.

[b] Δ, the difference in the results obtained with extracts of the two strains, indicating *nfi*-specified activity.

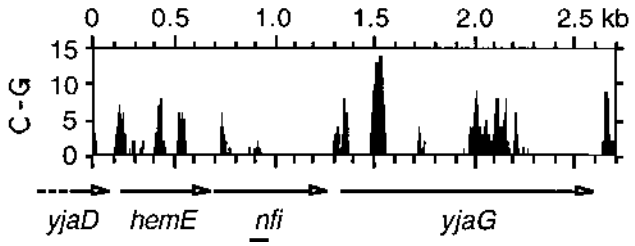

FIG. 3. Possible Rho-dependent transcriptional terminators (regions enriched for cytosine over guanine) in the *yjaD-hemE-nfi-yjaG* gene cluster. The number of cytosine residues minus the number of guanine residues in the preceding 78 nt was plotted against each position on the sense strand (1). The data were generated by the WINDOWS program of the Genetics Computer Group software.

clease V was modified to include an alkaline denaturation step, so that cleaved strands in duplexes would become susceptible to exonuclease I, a single-stranded-DNA-specific enzyme. As measured by this assay, the *nfi* plasmid specified a 40-fold overproduction of an endonuclease active on single-stranded DNA, and this induced enzyme was 78 times more specific for single- than double-stranded DNA (Table 3).

Endonuclease V is also specific for uracil-containing duplexes (11, 12). To test this activity in our plasmid-bearing strains, we used the linear duplex DNA of the *B. subtilis* phage PBS2, which contains uracil instead of thymine (34). For comparison, we used the thymine-containing linear duplex DNA of coliphage λ. The assay employed labeling by nick translation to detect the new 3′-hydroxyl end groups formed by endonucleolytic cleavage. The host strain was deficient in endonuclease I and uracil-DNA glycosylase. The strain bearing the *nfi* plasmid demonstrated a 53-fold overproduction of an endonuclease active on uracil-containing DNA, and this activity was 34-fold more specific for DNA containing uracil than for that containing thymine (Table 4). This combined preference for single-stranded and for uracil-containing DNA is characteristic of endonuclease V and distinguishes it from all other known endonucleases of *E. coli* (21).

Molecular organization of the *nfi* region. The *nfi* gene is preceded by a putative ribosome binding site (Fig. 1) that is complementary to the first 7 nt at the 3′ end of 16S RNA. From the eighth nucleotide of this site to the ATG codon, there is a distance of 4 nt. Although this spacing is rare and much shorter than the optimum of 8 nt, it is still within the range of functionality (27). The products of the first two codons of *yjaF* were not found in purified endonuclease V, which might be explained by the lack of any discernible ribosome binding site further upstream (Fig. 1). Therefore, it is more likely that these codons are never translated than that their amino acids are removed after translation.

The *nfi* gene is located in a cluster of four codirectional genes, *yjaD-hemE-nfi-yjaG* (Fig. 2). YjaD contains the motif of a MutT-like nucleotidohydrolase, HemE is uroporphyrinogen decarboxylase, and YjaG is of unknown function. The spaces between the genes are 40, 12, and 43 nt, respectively. The intergenic regions contained no Rho-independent transcriptional terminators that were detectable by a sensitive algorithm (9). Beyond the end of each gene, however, is a cytosine-rich, guanosine-poor region (Fig. 3). Such regions are characteristic of Rho-dependent terminators (1). However, the consensus sequence for such terminators is unknown, and because of the narrow spacing and orientation of the genes, the cluster might still contain an operon. To examine this possibility, we con-

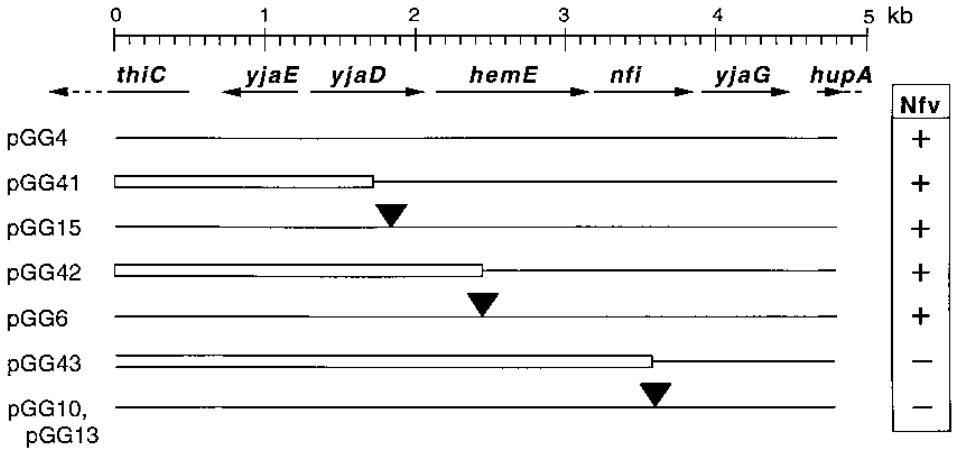

FIG. 2. Transcriptional independence of *nfi* from upstream genes. (Top) A portion of the current map of the *E. coli* K-12 chromosome (6) is shown, corresponding to kb 4204 to 4209 on the physical map or 90.4 min on the linkage map. The designation *nfi* replaces *yjaF*. Arrows indicate the positions and orientations of open reading frames. This 4.8-kb *Eco*RI-*Kpn*I DNA segment was cloned in pUC19 to yield plasmid pGG4, and partial deletions and insertions were produced (Table 1). (Bottom) In the plasmid maps, solid lines represent cloned DNA, open rectangles represent deletions, and triangles represent the sites of inserted *cat* elements. Nfi phenotypes were determined by enzymatic assays of sonicates of strain NovaBlue carrying each of the plasmids. +, overproduction of endonuclease V (4,000 to 7,000 U/mg of protein). −, activity equivalent to that of strain NovaBlue(pUC19) (63 to 81 U/mg of protein).

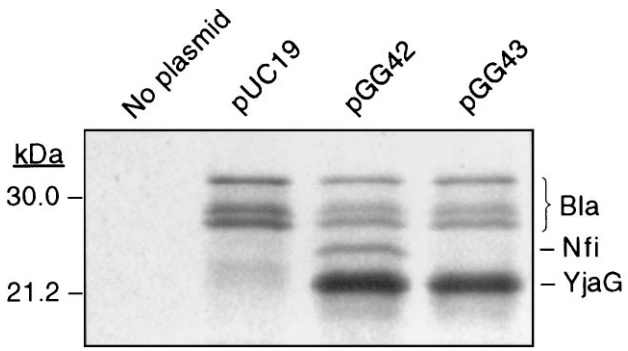

FIG. 4. Independent expression of Nfi and YjaG proteins. The maxicell technique (31) was used to label plasmid-specified proteins in strain DR1984. Plasmids pUC19, pGG42, and pGG43 correspond respectively to vector, pUC19::(*nfi-yjaG*), and pUC19::*yjaG* (Table 1 and Fig. 2). Proteins were labeled in vivo with a mixture of [^{35}S]cysteine and [^{35}S]methionine (ICN Biomedicals Corp.) and analyzed by SDS-PAGE with a 15% separating gel. Markers (bands not shown) were ^{35}S-labeled dCTP deaminase (21.2 kDa) from a maxicell preparation (37) and ^{14}C-labeled carbonic anhydrase (30.0 kDa; Amersham Life Science). Bla, vector-encoded β-lactamase gene products. Two of the Bla bands represent the proprotein (31.8 kDa) and the mature protein (28.9 kDa), respectively. The third is of unknown structure but is probably derived from *bla*, the only open reading frame of the vector that is large enough to encode it. The displayed portion of the autoradiogram contains all visible bands. The predicted fusion products of the truncated *lacZ* and *hupA* genes should have migrated beyond the end of the gel with other proteins of less than 13 kDa.

structed deletion and insertion mutations for the gene cluster on a plasmid.

We first cloned the entire region in the high-copy-number vector plasmid pUC19. The resulting plasmid, pGG4 (Fig. 2), specified a 60-fold overproduction of endonuclease V activity. In this plasmid and its derivatives, the cloned *nfi* region has an orientation opposite to that of the nearby *lac* promoter region of the vector. Therefore, *nfi* transcription must be initiated from a promoter within the cloned region. Multiple deletions and insertions were analyzed to reduce the possibility of a consistent artifact from the fusion of *nfi* to an unknown promoter outside the deletions or within the insertions. When genes upstream of *nfi* in plasmid pGG4 were deleted or interrupted by insertions, endonuclease V activity was unaffected, provided that the *nfi* gene was left intact (Fig. 2). Therefore, *nfi* is expressed independently of *yjaD* and *hemE*, at least under normal growth conditions.

To see if *yjaG*, the gene downstream of *nfi*, is expressed from the *nfi* promoter, we analyzed the proteins specified in vivo by two of the plasmids (Fig. 4). A plasmid carrying both genes (pGG42) produced proteins with the electrophoretic mobilities expected for Nfi (24.7 kDa) and YjaG (22.6 kDa). Plasmid pGG43, which had a deletion extending into *nfi*, no longer produced Nfi but continued to express YjaG. Therefore, *yjaG* is expressed independently of *nfi.*

Plasmid pGG6 contains 490 nt of uninterrupted sequence upstream of *nfi*, and it is Nfi$^+$. Assuming that there is no outward reading promoter in the inserted *cat* element, the *nfi* promoter must be within this 490-bp region which, except for the last 12 bp, is within the *hemE* gene.

Sequence comparisons. Of the proteins of known sequence, the one to which endonuclease V has the greatest similarity is an N^6 adenine-specific DNA methyl transferase encoded by a plasmid of *Pseudomonas aeruginosa* (35). The aligned sequences displayed 28% identity and 41% similarity over a length of 111 amino acids. This region covered 20% the length of the methylase and was distant from its methylase-specific motif. However, this similarity is of questionable significance, because alignment of the sequences necessitated 6 gaps of 1 or 2 amino acids each, and in only one instance were 3 adjacent amino acids paired. Endonuclease V had no significant homology to other proteins in the SWISS-PROT database (4) or with translations, in all possible reading frames, of the current DNA sequence databases. The protein sequence appeared to contain none of the functional domains that have been catalogued in the PROSITE database (5).

Corresponding chromosomal regions of *E. coli* and *Haemophilus influenzae* were compared. *H. influenzae*, which is unable to synthesize protoporphyrin, lacks homologs of *hemE* and *nfi*. However, it possesses homologs of the two outer genes of the cluster, *yjaD* and *yjaG*, which are codirectional and separated by only 39 nt. This finding raises the interesting possibility that during the evolution of these organisms, *nfi* and *hemE* might have been lost or acquired together and may therefore have a physiological relationship (see Discussion).

DISCUSSION

The arrangement of the four genes of the *nfi* cluster suggested at first that they might be in an operon. Although our studies indicated that *nfi* is expressed independently of the others during normal growth, this result does not preclude the possibility that under some stressful conditions, the genes might be expressed coordinately under the influence of a strong inducible promoter. We should now be able to use the radiolabeled products of these genes to identify them in two-dimensional polyacrylamide gels of the total proteins of *E. coli* and thus determine if they belong to any known stimulons.

A curious finding was that a λ*nfi* (λ533) did not significantly overproduce endonuclease V on infection of a host cell. In an earlier unreported experiment, we used a bioassay to screen for *nfi* in the genomic library to which λ533 belonged. We tested for the efficiency of plating of the recombinant λ phages on a *dut ung* host. Growth of the phage in this host, which lacks both dUTPase and uracil-DNA glycosylase, should allow the λ to incorporate large amounts of uracil into its DNA (13). If a recombinant phage overproduces either endonuclease V or uracil-DNA glycosylase, it should self-destruct. λ*ung* phages were detected by this method, but not λ*nfi*. Although λ533 did not appear to overproduce *nfi* by either an enzymatic or a physiological assay, it did have a functional *nfi* gene that was active when subcloned in plasmid vectors (Fig. 2). It is not evident how λ-directed transcription could have interfered with that of *nfi*; the *nfi* gene in the phage is over 15 kb from λ p_L and p_R and is in the same direction as λ p_L. Other possibilities, now under investigation, include a suppressible *nfi* mutation in the phage or an inhibitor produced by the phage vector or the cloned segment. For example, the phages, but not the plasmids, contain an intact *hupA* gene (Fig. 2), the product of which is a subunit of HU protein, a histone-like DNA-binding protein that might interfere either with the expression of *nfi* or with the enzymatic action of its product.

The proximity of *hemE* and *nfi* in *E. coli* and their combined absence from *Haemophilus* does not necessarily mean that the genes are functionally related. However, we cannot ignore this possibility. Why might the two genes be linked? The product of *hemE*, uroporphyrinogen decarboxylase, is needed for the biosynthesis of heme, one of the functions of which is to serve as a prosthetic group for catalase. Therefore, *nfi* and *hemE* might both protect against oxidants that would damage DNA. A second possible connection is based on our knowledge that HemE leads to the biosynthesis of photosensitizing metabolites. *hemH* mutants, which accumulate a late intermediate in

the heme pathway, are killed by visible light, but their photosensitivity is reversed by a *hemE* mutation (25), which affects an earlier step. In this study, we have noted that strains carrying a high-copy-number *hemE* plasmid (e.g., pGG41) are light sensitive. As reviewed in reference 24, this effect may be mediated by the photochemical formation of singlet oxygen or superoxide. These active oxygen species are a potential source of DNA damage that might be repaired by endonuclease V. It should be noted, however, that concomitant overexpression of *nfi* by pGG41 was not sufficient by itself to reverse the photosensitivity caused by the overexpression of *hemE*. Because of these considerations, we shall be prompted to test some properties of *nfi* mutants that we would not have considered otherwise, namely, survival and mutagenesis after exposure to visible light.

It was concurrently found that deoxyinosine 3′ endonuclease is encoded by the same gene as endonuclease V (41). That the two enzymes are the same could not be deduced from their published traits. Deoxyinosine 3′ endonuclease reportedly had no measurable activity on untreated single-stranded DNA or on DNA containing some uracil, and it cleaved to the 3′ side of dIMP, an abasic site, or a base mismatch (39, 40), whereas endonuclease V appeared to cleave 5′ to a lesion (12). However, these disparate results can be reconciled as follows: (i) the activity of deoxyinosine 3′ endonuclease on single-stranded DNA may have been eclipsed by its much greater activity on dIMP-containing DNA, especially at the pH used (7.5 instead of 9.5); (ii) the activity of endonuclease V on uracil-containing DNA is manifest only at high levels of uracil substitution (12); and (iii) determination of the endonuclease V cleavage site using exonuclease and a heterogeneous substrate (12) is apt to be relatively unreliable when compared to a method using DNA sequencing and defined oligonucleotides (39, 40), such as that used for the deoxyinosine 3′ endonuclease. The discovery that the two enzymes are probably the same expands the wide substrate spectrum of endonuclease V and poses a question as to what feature is common to all of its substrates.

ACKNOWLEDGMENTS

We thank Yoke Wah Kow for the communication of his results prior to publication and Susan Wallace for making us aware of his work. We also wish to acknowledge the capable technical assistance of Fred Kung.

This work was supported by Public Health Service grant ES06047 from the National Institute of Environmental Health Sciences. Computer facilities were provided in part by NIH grant MO1 RR00042.

REFERENCES

1. **Alifano, P., F. Rivellini, D. Limauro, C. B. Bruni, and M. S. Carlomagno.** 1991. A consensus motif common to all Rho-dependent prokaryotic transcription terminators. Cell **64:**553–563.
2. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1989. Current protocols in molecular biology. John Wiley & Sons, Inc., New York, N.Y.
3. **Bachmann, B. J.** 1996. Derivations and genotypes of some mutant derivatives of *Escherichia coli* K-12, p. 2460–2488. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. ASM Press, Washington, D.C.
4. **Bairoch, A., and B. Boeckmann.** 1992. The SWISS-PROT protein sequence data bank. Nucleic Acids Res. **20**(Suppl.)**:**2019–2022.
5. **Bairoch, A., and P. Bucher.** 1994. PROSITE: recent developments. Nucleic Acids Res. **22:**3583–3589.
6. **Berlyn, M. K., K. B. Low, and K. E. Rudd.** 1996. Linkage map of *Escherichia coli* K-12, edition 9, p. 1715–1902. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. ASM Press, Washington, D.C.
7. **Bittner, M., and D. Vapnek.** 1981. Versatile cloning vectors derived from the runaway-replication plasmid pKN402. Gene **15:**319–329.
8. **Blattner, F. R., V. Burland, G. Plunkett III, H. J. Sofia, and D. L. Daniels.** 1993. Analysis of the *Escherichia coli* genome. IV. DNA sequence of the region from 89.2 to 92.8 minutes. Nucleic Acids Res. **21:**5408–5417.
9. **Brendel, V., G. H. Hamm, and E. N. Trifonov.** 1986. Terminators of transcription with RNA polymerase from *Escherichia coli*: what they look like and how to find them. J. Biomol. Struct. Dyn. **3:**705–723.
10. **Chung, C. T., S. L. Niemela, and R. H. Miller.** 1989. One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. Proc. Natl. Acad. Sci. USA **86:**2172–2175.
11. **Demple, B., F. T. Gates, and S. Linn.** 1980. Purification and properties of *Escherichia coli* endodeoxyribonuclease V. Methods Enzymol. **65:**224–231.
12. **Demple, B., and S. Linn.** 1982. On the recognition and cleavage mechanism of *Escherichia coli* endodeoxyribonuclease V, a possible DNA repair enzyme. J. Biol. Chem. **257:**2848–2855.
13. **Duncan, B. K.** 1985. Isolation of insertion, deletion, and nonsense mutations of the uracil-DNA glycosylase (*ung*) gene of *Escherichia coli* K-12. J. Bacteriol. **164:**689–695.
14. **Gates, F. T., and S. Linn.** 1977. Endonuclease V of *Escherichia coli*. J. Biol. Chem. **252:**1647–1653.
15. **Goldmark, P. J., and S. Linn.** 1970. An endonuclease activity from *Escherichia coli* absent from certain *rec*$^-$ strains. Proc. Natl. Acad. Sci. USA **67:**434–441.
16. **Inoue, H., H. Nojima, and H. Okayama.** 1990. High efficiency transformation of *Escherichia coli* with plasmids. Gene **96:**23–28.
17. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole E. coli chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50:**495–508.
18. **Kow, Y. W.** Personal communication.
19. **LeFendre, N., and P. T. Matsudaira.** 1989. Purification of proteins and peptides by SDS-PAGE, p. 49–69. *In* P. T. Matsudaira (ed.), A practical guide to protein and peptide purification for microsequencing. Academic Press, New York, N.Y.
20. **Lehman, I. R., G. G. Roussos, and E. A. Pratt.** 1962. The deoxyribonucleases of *Escherichia coli*. II. Purification and properties of a ribonucleic acid-inhibitable endonuclease. J. Biol. Chem. **237:**819–828.
21. **Linn, S.** 1996. The DNases, topoisomerases, and helicases of *Escherichia coli.* *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. ASM Press, Washington, D.C.
22. **Marchuk, D., M. Drumm, A. Saulino, and F. S. Collins.** 1991. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. Nucleic Acids Res. **19:**1154.
23. **Miller, J. H.** 1992. A short course in bacterial genetics. A laboratory manual and handbook for *Escherichia coli* and related bacteria. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
24. **Nakahigashi, K., K. Nishimura, K. Miyamoto, and H. Inokuchi.** 1991. Photosensitivity of a protoporphyrin-accumulating, light sensitive mutant (*visA*) of *Escherichia coli* K-12. Proc. Natl. Acad. Sci. USA **88:**10520–10524.
25. **Nishimura, K., T. Nakayashiki, and H. Inokuchi.** 1993. Cloning and sequencing of the *hemE* gene encoding uroporphyrinogen III decarboxylase (UPD) from *Escherichia coli* K-12. Gene **133:**109–113.
26. **Price, A. R., and S. J. Cook.** 1972. New deoxyribonucleic acid polymerase induced by *Bacillus subtilis* bacteriophage PBS2. J. Virol. **9:**602–610.
27. **Ringquist, S., S. Shinedling, D. Barrick, L. Green, J. Binkley, G. D. Stormo, and L. Gold.** 1992. Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site. Mol. Microbiol. **6:**1219–1229.
28. **Rupp, W. D.** 1996. DNA repair mechanisms, p. 2277–2294. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. ASM Press, Washington, D.C.
29. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
30. **Sancar, A., and W. D. Rupp.** 1979. Cloning of *uvrA*, *lexC*, and *ssb* genes of *Escherichia coli*. Biochem. Biophys. Res. Commun. **90:**123–129.
31. **Sancar, A., R. P. Wharton, S. Seltzer, B. M. Kacinski, N. D. Clarke, and W. D. Rupp.** 1981. Identification of the uvrA gene product. J. Mol. Biol. **148:**45–62.
32. **Sternberg, N. L., and R. Maurer.** 1991. Bacteriophage-mediated generalized transduction in *Escherichia coli* and *Salmonella typhimurium*. Methods Enzymol. **204:**18–24.
33. **Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff.** 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. **185:**60–89.
34. **Takahashi, I., and J. Marmur.** 1963. Replacement of thymidylic acid by deoxyuridylic acid in the deoxyribonucleic acid of a transducing phage for *Bacillus subtilis*. Nature (London) **197:**794–795.

35. **Theriault, G., P. H. Roy, K. A. Howard, J. S. Benner, J. E. Brooks, A. F. Waters, and T. R. Gingeras.** 1985. Nucleotide sequence of the PaeR7 restriction/modification system and partial characterization of its protein products. Nucleic Acids Res. **13:**8441–8461.
36. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13amp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19:**259–268.
37. **Wang, L., and B. Weiss.** 1992. *dcd* (dCTP deaminase) gene of *Escherichia coli*: mapping, cloning, sequencing, and identification as a locus of suppressors of lethal *dut* (dUTPase) mutations. J. Bacteriol. **174:**5647–5653.
38. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.
39. **Yao, M., Z. Hatahet, R. J. Melamede, and Y. W. Kow.** 1994. Purification and characterization of a novel deoxyinosine-specific enzyme, deoxyinosine 3′ endonuclease, from *Escherichia coli*. J. Biol. Chem. **269:**16260–16268.
40. **Yao, M., and Y. W. Kow.** 1994. Strand-specific cleavage of mismatch-containing DNA by deoxyinosine 3′-endonuclease from *Escherichia coli*. J. Biol. Chem. **269:**31390–31396.
41. **Yao, M., and Y. W. Kow.** J. Biol. Chem., in press.