The Genes Encoding Formamidopyrimidine and MutY DNA Glycosylases in *Escherichia coli* Are Transcribed as Part of Complex Operons

CHRISTINE M. GIFFORD AND SUSAN S. WALLACE*

Department of Microbiology and Molecular Genetics, The Markey Center for Molecular Genetics, The University of Vermont, Burlington, Vermont 05405-0068

Received 15 March 1999/Accepted 30 April 1999

* Corresponding author. Mailing address: Department of Microbiology and Molecular Genetics, The Markey Center for Molecular Genetics, The University of Vermont, Stafford Hall, Burlington, VT 05405-0068. Phone: (802) 656-2164. Fax: (802) 656-8749. E-mail: swallace@zoo.uvm.edu.

Oxidative stress has been linked to cancer, aging, and a number of neurodegenerative diseases (for reviews, see references 2 and 12). Cells experience oxidative stress due to reactive oxygen species that are produced during normal oxidative metabolism. Reactive oxygen species cause a variety of damage in cells, including damage to proteins, lipids, and DNA. The last includes single-strand breaks, sites of base loss, and damage to the purine and pyrimidine bases. Oxidative base lesions that block DNA polymerases can be lethal to cells, and lesions that pair with an incorrect base are potentially mutagenic. Cells have evolved an efficient and accurate mechanism, base excision repair, to remove these lesions from DNA. Base excision repair is highly conserved from bacteria to humans.

*Escherichia coli* formamidopyrimidine (Fpg) DNA glycosylase and MutY DNA glycosylase are base excision repair proteins that work together to protect cells from the mutagenic effects of the commonly oxidized guanine product 7,8-dihydro-8-oxoguanine. The genes encoding these proteins, *fpg* and *mutY*, are both cotranscribed as part of complex operons. *fpg* is the terminal gene in an operon with the gene order radC, rpmB, rpmG, and *fpg*. This operon has transcription initiation sites upstream of *radC*, in the *radC* coding region, and immediately upstream of *fpg*. There is a strong attenuator in the *rpmC*-*fpg* intergenic region, and three transcription termination sites downstream of *fpg*. There is an additional site, in the *radC-rpmB* intergenic region, that corresponds either to a transcription initiation site or to an RNase E or RNase III cleavage site. *mutY* is the first gene in an operon with the gene order *mutY*, *yggX*, *mltC*, and *nupG*. This operon has transcription initiation sites upstream of *mutY*, in the *mutY* coding region, and immediately upstream of *nupG*. There also appear to be attenuators in the *yggX-mltC* and *mltC-nupG* intergenic regions. The order of genes in these operons has been conserved or partially conserved only in other closely related gram-negative bacteria, although it is not known whether the genes are cotranscribed in these other organisms.

nuclease VIII (*nei*), identified on the basis of its specificity for oxidized pyrimidines, also removes 8-oxoG. The addition of the *nei* mutation to the *fpg* *mutY* double mutant results in an additional threefold synergy and a further increase in GC-ta transversions (3a). MutT is an 8-oxodGTPase that breaks down 8-oxodGTP to 8-oxodGMP in the nucleotide pool, preventing its incorporation into DNA (25). Strains deficient in MutT have elevated rates of AT-to-CG transversions (45). Fpg, MutY, and MutT are often referred to as the GO repair system, an error avoidance pathway devoted to preventing mutations resulting from the presence of 8-oxoG in DNA (GO is a synonym for 8-oxoG) (26–28).

There have been few studies on the regulation of the oxidative DNA glycosylases in *E. coli*. It has been shown that cells exhibit increased Fpg activity when shifted from anaerobic to aerobic growth conditions and when exposed to the superoxide-generating compound paraquat (20, 21). This response still occurs in mutants defective in SoxR and SoxS, demonstrating that *fpg* is not part of the SoxRS regulon (21). It has recently been shown that, under anaerobic growth conditions, Fpg activity increases in strains deficient in the global regulators Fur, FNR, and ArcA (21). Possible Fur, FNR, and ArcA binding sites have been identified in the *fpg* promoter region, suggesting that these proteins may play a regulatory role in *fpg* regulation. In addition, it has been reported that the transcription of *fpg* is decreased in the presence of thioredoxin and glutathione, and it has been speculated that glutathione-mediated disassembly of iron-sulfur centers in key regulatory proteins may govern the transcriptional regulation of *fpg* (17).

There have been no reported studies on MutY regulation. Although our major goals are to further delineate the regulatory controls for Fpg and to define any regulatory controls for MutY, we had to first determine whether the *fpg* and *mutY* genes in *E. coli* are transcribed alone or with neighboring genes that have the same orientation. If *fpg* and *mutY* are part of operons, the genes with which they are transcribed may play a...
role in or provide information about their possible regulation. In this paper, we show that fpg is cotranscribed as the terminal gene in a four-gene operon and that mutY is cotranscribed as the first gene in a four-gene operon. Furthermore, both of these are complex operons containing multiple promoters and terminators. The order of genes in the fpg operon is radC, rpmB, rpmG, and fpg. These genes are arranged in a counter-clockwise orientation on the genome. RadC has been suggested to play a role in growth medium-dependent, recA-dependent repair of DNA single-strand breaks after X-irradiation and postreplication repair after UV irradiation (15). rpmB and rpmG encode the ribosomal proteins (r-proteins) L28 and L33, respectively (22). The order of genes in the mutY operon is mutY, yggX, mlcC, and mutN. These genes are arranged in a clockwise orientation on the genome. yggX encodes a protein of unknown function; mlcC encodes membrane-bound lytic transglycosylase C, which has been shown to have peptidoglycan hydrolase activity (11); and mutN encodes a high-affinity nucleoside transport protein (42).

MATERIALS AND METHODS

Bacterial strains. E. coli GC4668 [DE8194 Lac F+ lacI(Δ74)/rpsL179(smR)], N3433 [lacZΔ(48) hsr- relA spoT1 thi-1], and N3431 [N3433 Δ307(Fr)] were obtained from the Yale University E. coli Genetic Stock Center. Strain IBPC637 (N3431, Δ305 radB51::Tn10) was kindly supplied by Philippe Regnier, Institut de Biologie-Physico-Chimique, Paris, France.

Growth conditions and RNA isolation. E. coli cultures (1 ml) were grown overnight in Luria-Bertani broth with shaking at 250 rpm. The overnight cultures were diluted 1:100 in fresh Luria-Bertani broth and were grown until they reached the desired optical density at 600 nm (OD600). GC4668 was grown at 37°C in T1 medium present 10 μg of streptomycin per ml until an OD600 of 0.7 to 0.8 was reached. N3433 and IBPC637 were grown at 30°C until an OD600 of 0.4 to 0.5 was reached, and then they were shifted to 45°C for 30 min. IBPC637 was grown in the presence of 10 μg of tetracycline per ml. Total RNA was isolated with a Qiagen RNeasy kit according to manufacturer’s recommendations. After elution from the RNasey column, the RNA was treated with DNase, extracted twice with acid-phenol, and ethanol precipitated once with chloroform-isooctanol. The RNA was precipitated with ammonium acetate and ethanol, washed in 75% ethanol, and resuspended in RNA-free water. The integrity of the 16S and 23S rRNAs was checked on a 1% agarose gel.

Reverse transcription (RT)-PCR. RNA was reverse transcribed with Gibco BRL USERSCRIPTS reverse transcriptase. Total RNA (30 μg) and primer (15 pmol) were denatured in water for 10 min at 80°C and then quickly chilled on ice. Gibco BRL 1× first-strand buffer, 10 mM dithiothreitol, and 500 μM each dNTP were added, and the reaction mixture was incubated at 47°C for 2 min. SUPERSCRIPT II (1 μl; 200 U) was added, and the reaction mixture was incubated at 47°C for 1 h. The nucleic acids were precipitated with ammonium acetate and ethanol, washed with 75% ethanol, and resuspended in 5 μl of formamide gel loading buffer. A sequencing ladder was generated from control DNA M13mp18 with a U.S. Biochemicals Sequenase version 2.0 sequencing kit. The sequencing ladder and the 5-μl primer extension reaction mixture were run on a 6% polyacrylamide gel. The gel was dried and exposed to film.

Homology search. The amino acid sequence for the product of each gene in the fpg and mutY operons was used in a BLAST search at the Unfinished NCBI Microbial Genome BLAST Website (29a). On this web page, the “All” option was chosen to search both the finished and the unfinished genomes. Each sequence used was blastn. The amino acid sequences for the products of the genes in the fpg operon were obtained from GenBank accession no. AE000441, those for mutY and yggX were obtained from accession no. AE000578, and those for mlcC and mutN were obtained from accession AE000379. A Blast searching of the following unfinished genomes: Salmonella typhimurium and Yersinia pestis, was performed at the Sanger Centre through projects funded by Beowulf Genomics (the BLAST servers for these organisms at the Sanger Centre site [25a] were also used. Contig 770 from the S. typhimurium database contains the gene radC and was used to splice together overlapping sequences from this site and GenBank accession no. U23405, which contains S. typhimurium rpmB, rpmG, and fpg; Actinobacillus actinomycetemcomitans; and the Actinobacillus Genome Sequencing Project supported by a USPHS/NIH grant from the National Institute of Dental Research; Pseudomonas aeruginosa, sequenced at the University of Washington Genome Center and PathoGenesis Corporation and jointly funded by the Cystic Fibrosis Foundation and PathoGenesis Corporation; and Vibrio cholerae (preliminary sequence data were obtained from the BLAST server at the Institute for Genomic Research [19a]; sequencing of V. cholerae was accomplished with support from NIAID). Sequences from all organisms were aligned with the subject numbers provided by the BLAST searches. In some cases, when genes were located on different contigs, it was possible to align them because of overlapping sequences.

RESULTS

fpg is part of an operon containing four genes. The fpg transcript was initially mapped by RT-PCR to determine if it is transcribed as a single or with other genes as a larger operon. The order of the genes that could potentially be transcribed together is radC, rpmB, rpmG, and fpg. rpmB and rpmG have previously been proposed to be transcribed together, and there is a stem-loop structure, including six T residues beginning 15 bases after the termination codon for rpmG; this structure was shown by an S1 nuclease digestion assay to be an RNA polymerase transcription termination site (22). There are 216 bp between the radC stop codon and the rpmB start codon, 20 bp between the rpmB stop codon and the rpmG start codon, and 97 bp between the rpmG stop codon and the fpg start codon. The gene 5′ to radC, dlp, and the gene 3′ to fpg, kdtB, are in opposite orientations, precluding them from being on the same transcript with fpg. E. coli total RNA was reverse transcribed with the primer fpg1, which anneals 357 bp 3′ to the fpg start codon, and PCR was performed to determine the approximate size of the cDNA. Primer fpg1a was the downstream primer and primers fpg7, fpg9, and fpg10 were the upstream primers used in the PCR; Fig. 1A shows approximate annealing locations of PCR primers. Positive PCR controls were run with the three primer sets and genomic DNA as a template (Fig. 1B, lanes 1, 4, and 7; sizes are indicated in the figure). PCR with primer sets fpg7-fpg1a and fpg9-fpg1a and with cDNA as a template yielded a product (Fig. 1B, lanes 3 and 6), indicating that the transcript containing fpg also contained rpmG, rpmB, and radC. PCR with fpg10-
fpg1a did not yield a product (Fig. 1B, lane 9), indicating that the upstream transcription initiation site for the transcript was within 148 bp of the proposed start codon for radC. The negative control lanes (the PCR template was an RT reaction lacking reverse transcriptase) were empty, indicating that the RNA used for RT was not contaminated with DNA (Fig. 1B, lanes 2, 5, and 8). To confirm that the PCR product obtained with primer set fpg9-fpg1a was specific for the fpg transcript, the band was cut out, gel eluted, and used as a template in PCRs with the same primer sets. The results were the same: primer sets fpg7-fpg1a and fpg9-fpg1a yielded a product, whereas primer set fpg10-fpg1a did not (Fig. 1C, lanes 2, 4, and 6).

The fpg operon contains multiple promoters and terminators. RPAs were performed to confirm the RT-PCR results and to map the locations of promoters, terminators, and transcript processing sites. Several operons containing r-protein genes undergo transcript processing by endonucleases RNase III (rnc) and/or RNase E (rne) (1, 7, 34, 35, 43). For this reason, the fpg RPAs were performed with RNA from wild-type E. coli (N3433) and from rnc mce E. coli (IBPC637). Three reactions were set up for each RNA probe, one with probe plus E. coli RNA and two with probe plus yeast RNA. After overnight hybridization, the E. coli RNA-probe reaction mixture was digested with RNase A and RNase T1. One of the yeast RNA-probe reaction mixtures was also digested with RNase A and RNase T1 to ensure complete digestion of the probe at the RNase concentrations used (data not shown). The other yeast RNA-probe reaction mixture was mock digested; this procedure resulted in a full-length probe when the reaction mixture was run on a gel (only part of the reaction mixture was loaded so the signal would not overpower the other lanes). RNA probes 1, 2, and 3 overlap and cover the length of the transcript from 148 bp 5' to the proposed start codon for radC to 90 bp 3' to the fpg start codon (Fig. 2A). The RPA with probe 1 resulted in a 635-nucleotide protected fragment (Fig. 2B, lanes 2 and 3) as predicted with a semilog plot of the distance traveled by the RNA markers. The RPAs with wild-type RNA (Fig. 2B, lane 2) and rnc mce RNA (Fig. 2B, lane 3) looked the same, except for a more pronounced ladder of bands below the 635-nucleotide band in the double mutant. Digestion occurred from the 3' end of the probe, since the RPA with probe 6 (Fig. 3A) resulted in a full-length product (Fig. 3B, lane 2; this assay was performed with wild-type RNA only). The 635-nucleotide product corresponds to a transcription initiation site about 8 bp upstream from the radC start codon proposed in the GenBank accession no. AE000441 sequence. To more precisely map the site, primer extension analysis was performed with
primer fpg14 (Fig. 4A). The primer was extended to a product of 123 nucleotides (Fig. 4B, lane 1), which corresponds to an initiation site \( \approx 12 \) bp from the proposed start codon for radC.

The RPA with probe 2 and wild-type RNA resulted in a faint full-length product, a 483-nucleotide product, and a 318-nucleotide product (Fig. 2C, lane 2). The full-length product corresponds to transcript readthrough from radC into rpmB and was more abundant in rnc rne cells (Fig. 2C, lane 3). The 483-nucleotide product was present with equal intensities in both wild-type and double-mutant RNAs, and the 318-nucleotide product was present only in wild-type RNA. To determine if the 483-nucleotide and 318-nucleotide protected products were the result of 5' or 3' probe digestion, an RPA was performed with probe 7 (Fig. 3A). If the products were the result of 3' probe digestion, the 483-nucleotide product would appear as full length and the 318-nucleotide product would remain the same size. This was the case when probe 7 was used (Fig. 3B, lane 4; this assay was performed with wild-type RNA only). The 483-nucleotide product obtained with probe 2 indicates that there is a transcription initiation site about 70 bp 5' to the radC stop codon. This site was mapped by primer extension analysis with primer fpg17 (Fig. 4A). The primer was extended to a product of 126 nucleotides (Fig. 4B, lane 2), corresponding to a transcription initiation site \( 104 \) bp 5' to the radC stop codon. The 318-nucleotide product seen with wild-type RNA maps to a site about 118 bp 5' to the rpmB start codon in the radC-rpmB intergenic region. Primer extension analysis was performed with primer fpg5 (Fig. 4A). The primer was extended to a product of 214 nucleotides (Fig. 4C), placing the transcription initiation or cleavage site 138 bp 5' to the rpmB start codon.

The RPA with probe 3 resulted in a full-length product, a 273-nucleotide product, and a 109-nucleotide product (Fig. 2D, lanes 2 and 3). The full-length product was due to transcript readthrough into fpg and was present in a greater abundance in the double mutant. The 273-nucleotide product was present in equal abundances in wild-type and rnc rne cells; the size of the product, if it resulted from 5' probe digestion, would correspond to the location of the previously described terminator downstream of rpmG. Probe 4 was used to confirm this result, and the RPA resulted in a full-length product, a 109-nucleotide product, and a 94-nucleotide product of equal intensities in wild-type and double-mutant cells (Fig. 2D, lanes 5 and 6). The 94-nucleotide product corresponds to the pro-
The thin lines indicate the products obtained from RPAs with these probes along with the sizes of the products. The sizes and exact annealing locations of the probes are as follows: probe 6, 480 nucleotides, anneals from 625 bp 3' to the radC start codon; probe 7, 425 nucleotides, anneals from 201 bp 3' to the rpmB stop codon to 8 bp 5' to the radC stop codon; probe 8, 820 nucleotides, anneals from 494 bp 3' to the fpg stop codon to 326 bp 5' to the fpg stop codon. (B and C) The probes used in the RPAs are indicated, and lane M contains RNA size markers. Lanes 1 in panel B and C and lane 3 in panel B show the full-length probe (probe hybridized with yeast RNA and mock digested; only part of the reaction mixture was loaded so that the signal would not overpower the other lanes). Lanes 2 in panels B and C and lane 4 in panel B show RPA results obtained with wild-type RNA. Numbers beside panels are in base pairs.

FIG. 3. Mapping of the fpg transcript by RPAs. (A) The thick lines designated 6 to 8 indicate the approximate annealing locations of the probes used in the RPAs. The thin lines indicate the products obtained along with the sizes of the products. The sizes and exact annealing locations of the probes are as follows: probe 6, 480 nucleotides, anneals from 625 bp 3' to the radC start codon; probe 7, 425 nucleotides, anneals from 201 bp 3' to the rpmB stop codon to 8 bp 5' to the radC stop codon; probe 8, 820 nucleotides, anneals from 494 bp 3' to the fpg stop codon to 326 bp 5' to the fpg stop codon. (B and C) The probes used in the RPAs are indicated, and lane M contains RNA size markers. Lanes 1 in panel B and C and lane 3 in panel B show the full-length probe (probe hybridized with yeast RNA and mock digested; only part of the reaction mixture was loaded so that the signal would not overpower the other lanes). Lanes 2 in panels B and C and lane 4 in panel B show RPA results obtained with wild-type RNA. Numbers beside panels are in base pairs.

posed terminator, which appears to act as an attenuator, since there is transcript readthrough into fpg. The 109-nucleotide product present with both probes was much more abundant in mc me RNA. The size of the product did not shift with two probes, indicating that the product resulted from 3' probe digestion. This result was confirmed by primer extension analysis with primer fpg4 (Fig. 4A). The primer was extended to an 89-nucleotide product (Fig. 4B, lane 3), placing the transcription initiation site −23 bp from the fpg start codon.

Probe 5 was designed to map the terminator(s) of the operon. The RPA with probe 5 resulted in products of 93, 177, and 360 nucleotides as well as several other, less prominent bands (Fig. 2E, lanes 2 and 3). Since kdtB, the gene 3' to fpg, begins 39 bp from the fpg stop codon and is in the opposite orientation, probe 5 cannot anneal to kdtB RNA. For this reason, it was assumed that all probe digestion took place from the 5' end. The patterns of protected products from wild-type cells (Fig. 2E, lane 2) and mc me cells (Fig. 2E, lane 3) were similar, except for a faint full-length product and a more abundant 93-nucleotide product in the mc me cells. The 93-nucleotide protected product corresponds to a previously predicted region of secondary structure between fpg and kdtB (4) that may act as a bidirectional transcription terminator. This product maps to a site 19 bp 3' to the fpg stop codon. The 177- and 360-nucleotide products correspond to termination sites 103 bp and 286 bp 3' to the fpg stop codon, respectively. Both of these termination sites are within the kdtB coding region. The termination sites were confirmed with probe 8 (Fig. 3A); this assay resulted in protected products of 355, 434, and 644 nucleotides and several less prominent protected products (Fig. 3C, lane 2; this assay was performed with wild-type RNA only). These products correspond to termination sites 29, 108, and 318 bp 3' to the fpg stop codon and, taken together, are within 10 to 30 bp of the sites mapped with probe 5.

These results indicate that fpg is cotranscribed as the terminale gene in an operon with the gene order radC, rpmB, rpmG, and fpg. There are transcription initiation sites upstream of radC, in the radC coding region, and immediately upstream of fpg (Fig. 5A). In addition, a 5' transcript end detected in the radC-rpmB intergenic region corresponds to either a promoter or an RNase III or RNase E cleavage site. There is a strong attenuator in the rpmG-fpg intergenic region, and there are three apparent terminators downstream of fpg.

mutY is part of an operon containing four genes. The mutY transcript was initially mapped by RT-PCR to determine if it is transcribed alone or with surrounding genes in the same orientation. The order of genes that could potentially be transcribed together is mutY, yggX, mltC, and nupG. There are 27 bp between the mutY stop codon and the yggX start codon, 61 bp between the yggX stop codon and the mltC start codon, and 153 bp between the mltC stop codon and the nupG start codon. The gene 3' to mutY, yggH, and the gene 3' to nupG, speC, are
in opposite orientations on the genome, precluding them from being transcribed with mutY. E. coli total RNA was reverse transcribed with primer mutY11, which anneals 465 bp 3' to the proposed start codon for nupG, PCR was performed to determine the approximate size of the cDNA (Fig. 6A). PCR with primers mutY8-mutY5 and mutY12-mutY16 resulted in a product (Fig. 6B, lanes 3 and 6), indicating that mutY, yggX, mltC, and nupG were present on the same transcript. PCR was also performed with primers mutY12-mutY5, which resulted in a product, and mutY3-mutY5, which did not (Fig. 6C, lanes 3 and 6); these results confirm that the four genes were present on the same transcript and that the upstream transcription initiation site for the operon was within 271 bp of the mutY start codon. Positive PCR controls were run with the primers sets and genomic DNA as a template (Fig. 6B, lanes 1 and 4, and Fig. 6C, lanes 1 and 4). The negative control lanes were empty, indicating that the RNA used for RT was not contaminated with DNA (Fig. 6B, lanes 2 and 5, and Fig. 6C, lanes 2 and 5).

The mutY operon contains multiple promoters and terminators. RPAs were performed to confirm the RT-PCR results and to map the locations of promoters and terminators. The assays were performed as described above, except that only RNA from wild-type E. coli (GC4468) was used. Probe 1 (Fig. 7A) was used to map the 5'-most promoter(s) for the operon. The RPA with probe 1 resulted in a faint 243-nucleotide product, a 167-nucleotide product, and a 151-nucleotide product (Fig. 7B, lane 2). The 151-nucleotide product was not consistently seen in other experiments (data not shown) and was not pursued further. The 243-nucleotide product corresponds to a possible minor transcription initiation site about 102 bp 5' to the mutY start codon, and the 167-nucleotide product corresponds to a possible transcription initiation site about 26 bp 5' to the mutY start codon. Attempts to map the minor transcription initiation site by primer extension analysis were inconclu-
The site corresponding to the 167-nucleotide product was mapped by primer extension analysis with primer mutY14 (Fig. 8A). The primer was extended to a product of 106 bp (Fig. 8B), corresponding to a transcription initiation site 25 bp 5' to the mutY start codon.

The RPA with probe 2 resulted in a full-length product and a 237-nucleotide product (Fig. 7B, lane 4). To determine whether the 237-nucleotide product was the result of 5' or 3' probe digestion, an RPA was performed with probe 3. This assay resulted in a full-length product and a 238-nucleotide product (Fig. 7C, lane 2), showing that digestion had occurred from the 3' end of the probe and that there is a possible transcription initiation site about 70 bp 5' to the mutY stop codon. This site was mapped by primer extension analysis with primer mutY16 (Fig. 8A). The primer was extended to a 145-nucleotide product (Fig. 8C), which corresponds to a transcription initiation site 72 bp 5' to the mutY stop codon.

The RPA with probe 4 yielded a full-length product and products of 208, 177, 163, 149, and 143 nucleotides (Fig. 7D, lane 2). The portion of the transcript that this probe anneals to has several regions that are AU rich; this characteristic can result in transient strand separation of the transcript and probe hybrid, leaving it susceptible to cleavage by RNase A, which cleaves 3' to cytosine and uridine residues. To determine whether any of the protected products were due to RNase A cleavage of separated strands, an assay was performed with probe 4 and RNase T1 digestion only. RNase T1 cleaves 3' to guanosine residues, a characteristic which should eliminate non-specific cleavage due to strand separation. After RNase T1 digestion, the 163-, 149-, and 143-nucleotide products disappeared (Fig. 7D, lane 3). The sizes of these products indicate that they were due to the same RNase A cleavage events that occurred with probe 4. The 374-nucleotide product shifted to a 375-nucleotide product after RNase T1 digestion alone. The sizes of the products obtained with probe 5 showed that 5' digestion had occurred with both probes and indicate that there are two termination sites in the intergenic region between yggX and mltC. The second termination site appears to be minor. The 177-nucleotide product obtained by RNase A and RNase T1 digestion of probe 4 and the 374-nucleotide product obtained by RNase A and RNase T1 digestion of probe 5 map to a C residue 26 bp 3' to the yggX stop codon and an A residue 28 bp 3' to the stop codon, respectively. It is not possible to determine with certainty, from these experiments, which residue is the terminating residue, since the size predictions are not exact. A possible stem-loop structure immediately preceding this site may act as an attenuator, although it would appear to be a weak attenuator, since there is substantial transcript readthrough. The minor termination site, corresponding to the 208-nucleotide product obtained by RNase A and RNase T1 digestion of probe 4, and the 408-nucleotide product obtained by RNase A and RNase T1 digestion of probe 5 map to a G residue 3 bp 5' to the mltC start codon and to the A residue of the mltC start codon, respectively.

The RPA with probe 6 resulted in a full-length product and products of 242, 199, 172, and 144 nucleotides (Fig. 7E, lane 2). The 144-nucleotide product disappeared after RNase T1 digestion (data not shown), indicating that it was due to cleavage of separated hybrid strands. A nupG transcription initiation site was previously mapped by S1 nuclease digestion to an A residue 15 bp 5' to the predicted start codon (29). 3' digestion of probe 6 back to this transcription initiation site would result in a product of 172 nucleotides, which is what was seen. To confirm that this product resulted from 3' probe digestion and to determine whether the other products were due to 5' or 3' probe digestion, an RPA was performed with probe 7. This assay resulted in 261-, 242-, and 199-nucleotide products (Fig. 7E, lane 4). The 172-nucleotide product seen with probe 6 was no longer present, confirming that it was due to 3' probe digestion. The 242- and 199-nucleotide products were present.
with both probes, indicating that they were due to 5’ probe digestion and that they correspond to termination sites. The 199-nucleotide product maps to a termination site 34 bp 3’ to the \textit{mltC} stop codon, and the 242-nucleotide product maps to a termination site 76 bp 3’ to the \textit{mltC} stop codon. Both of these sites are immediately preceded by possible stem-loop structures that may act as terminators. The lack of a full-length product and the presence of a 261-nucleotide product not seen with probe 6 suggest that RNase “nibbling” at the 5’ end of the probe-RNA hybrid may have resulted in the 261-nucleotide product.

To confirm the \textit{nupG} transcription initiation site, primer extension analysis was performed with primer mutY5 (Fig. 8A). The primer was extended to a product of 177 nucleotides (Fig. 8D), which corresponds to a transcription initiation site at the previously mapped A residue 15 bp 5’ to the \textit{nupG} start codon. Starting 7 bp downstream of the \textit{nupG} stop codon, there is a structure with the features of a rho-independent terminator (C+G-rich stem-loop structure followed by a stretch of T residues) (42), but this structure was not mapped in the present study.

These results indicate that \textit{mutY} is cotranscribed as the first gene in an operon with the gene order \textit{mutY}, \textit{yggX}, \textit{mltC}, and \textit{nupG}. There are transcription initiation sites upstream of \textit{mutY}, in the \textit{mutY} coding region, and immediately upstream of \textit{nupG} (Fig. 5B). The operon also appears to have attenuators in the \textit{yggX-mltC} and \textit{mltC-nupG} intergenic regions.

The genes in the operons are arranged in a similar manner in other organisms. BLAST searches were performed with the finished and unfinished genome databases at the NCBI Microbial Genome BLAST Website and with unfinished genome databases at other sites (see Materials and Methods) by use of the \textit{E. coli} amino acid sequences for the products of the genes in both operons. The results showed that in \textit{S. typhimurium} and \textit{Y. pestis}, the genes of the \textit{E. coli fpg} operon are arranged in the same order and with similar spacing (Fig. 9A). In \textit{V. cholerae} and \textit{Haemophilus influenzae}, \textit{radC}, \textit{rpmB}, and \textit{rpmG} are arranged in the same order and with similar spacing, but \textit{fpg} is separated from them by 722 and 4,281 bp, respectively. \textit{A. actinomycetemcomitans} may have the genes arranged in the same order, but it was not possible to confirm this notion. Parts of \textit{radC} and \textit{rpmB} were shown to be on the same contig with similar spacing, and the 3’ end of \textit{rpmG} and all of \textit{fpg} were shown to be on the same contig with similar spacing. In \textit{P.}}
aeruginosa, radC is separated from rpmB and rpmG by 2,530 nucleotides, and the gene for fpg appears to be located elsewhere on the genome.

The genes of the E. coli mutY operon are arranged in the same order and with similar spacing in S. typhimurium (Fig. 9B). In Y. pestis, V. cholerae, and H. influenzae, the first three genes are arranged in the same order and with similar spacing, but these organisms do not appear to have a nupG homolog. In A. actinomycetemcomitans and P. aeruginosa, the first two genes are arranged in the same order and with similar spacing, but neither of these organisms appears to have a mltC or nupG homolog.

**DISCUSSION**

fpg is transcribed as the terminal gene in the radC-rpmB-rpmG-fpg operon, and mutY is transcribed as the first gene in the mutY-yggX-mltC-nupG operon (Fig. 5). The genes in these operons have no apparent relationship to the function of Fpg or MutY as DNA glycosylases, with the possible exception of radC, the first gene in the fpg operon. A mutation in radC (radC102) sensitizes cells to UV irradiation and X-irradiation, and cells with this mutation are 60% deficient in recombination ability (15). It has been shown that radC102 mutants are deficient in recA-dependent repair of X-ray-induced single-strand breaks and that this deficiency correlates with their X-ray survival deficiency; the mutants also have a small deficiency in repair of X-ray-induced double-strand breaks (16). radC was cloned as a 297-bp open reading frame (ORF) based, in part, on partial restoration of resistance to a radC102 strain carrying a plasmid containing the ORF (14). The present study shows that the mapped size of the radC transcript is more in accordance with an ORF of 675 bp, as assigned in GenBank accession no. U000441, which places the ATG start site 375 bp 5' to the originally predicted start site.

rpmB and rpmG encode ribosomal proteins L28 and L33, respectively (22), which are 2 of the 52 r-proteins made by E. coli. In E. coli, many of the r-protein genes are arranged in operons, and many of these operons include non-r-protein genes. Three r-protein operons, in particular, are similar to the fpg operon in that they have multiple promoters, undergo transcript processing by RNase III and/or RNase E, and have an attenuator between the r-protein gene(s) and the non-r-protein gene(s). The rplK-rplA-rplJ-rplL-rpoB-rpoC operon encodes r-proteins L11, L1, L10, and L7/12 and RNA polymerase subunits β and β' (6). This operon has promoters upstream of rplK, in the rplA-rplJ intergenic region, and there may be a minor promoter in the rplL-rpoB intergenic region (6, 23, 44).
The attenuator in the *rplL-rpoB* intergenic region terminates 80% of transcripts, and nonattenuated transcripts are cleaved by RNase III (1, 10). The *rpsU-dnaG-rpoD* operon encodes r-protein S21, DNA primase, and sigma-70 (7). This operon has two promoters upstream of *rpsU* and a heat shock promoter in the *dnaG* coding region (7, 38). The attenuator between *rpsU* and *dnaG* terminates 80 to 90% of transcripts, and nonattenuated transcripts are cleaved between *dnaG* and *rpoD* by RNase E, after which the *dnaG* message is rapidly degraded (7, 24, 43). The *rpsO-pnp* operon encodes r-protein S15 and polynucleotide phosphorylase (37). This operon has promoters upstream of *rpsO* and upstream of *pnp*. There is an attenuator between the two genes that terminates 50% of transcripts, and nonattenuated transcripts are processed by both RNase III and RNase E (34, 35, 37). The attenuators in these operons all resemble rho-independent terminators. The obvious difference between these operons and the *fpg* operon is that there is a non–r-protein gene, *radC*, preceding the r-protein genes in the *fpg* operon.

There are at least three promoters in the *fpg* operon (Fig. 5A). The promoter upstream of *radC* appears to be a weak one, since very little RNA message originating from it was detected by RPAs in either the wild-type strain or the *mc me* strain (Fig. 2B, lanes 2 and 3). At least some of the transcript originating from this promoter extends through *fpg*, since it was possible to detect a transcript containing all four genes by RT-PCR. The promoter located in the 3' end of the *radC* coding region, upstream of *rpmB* and *rpmG*, was expressed at equal strengths in wild-type and *mc me* strains (Fig. 2C, lanes 2 and 3; 483-nucleotide product). There is a possible 10 hexamer (TAAACT) 8 bp upstream of the transcription initiation site and separated by 18 bp from a possible 35 hexamer (TTGTGC). The transcription initiation site upstream of *fpg* that was identified in this study is 1 bp 5' to a transcription initiation site mapped in a previous study (5). This initiation site is 2 bp downstream from a proposed -10 hexamer (TTG ACT) separated by 16 bp from a proposed -35 hexamer (TT GTTA). The transcript originating from this promoter was
present in an approximately 30-fold-larger amount in _mc rne_ cells than in wild-type cells (Fig. 2D, lanes 2, 3, 5, and 6; 109-nucleotide product), as determined by phosphorimager analysis. At this point it is unclear what is causing this difference. It is possible that a negative regulator that normally represses transcription from the _fpg_ promoter is not active in _mc rne_ cells. Three repressors of Fpg activity, FNR, Fur, and ArcA, and their possible binding sites in the _fpg_ promoter region have been identified (21), although there are no reports that the transcripts for these proteins require RNase III and/or RNase E processing to be properly expressed. In addition, these regulators appear to repress Fpg activity under anaerobic conditions, so it is unlikely that they would have such a dramatic effect in aerobic growth. It is also possible that RNase III and/or RNase E play a role in the rapid degradation of the _fpg_ transcript in wild-type cells.

The attenuator in the _rpmG-fpg_ intergenic region is located very close to the predicted −35 hexamer for _fpg_ and resembles...
a rho-independent terminator. One possible secondary structure has a 6-bp G+C-rich stem (with one mismatch) followed by a run of six T residues. The last two T residues in this run are the first two in the predicted TTGTTA −35 hexamer. A second possible secondary structure has up to a 15-bp stem by a run of six T residues. The last two T residues in this run

C-rich stem (with one mismatch) followed

1

ture has a 6-bp G

a rho-independent terminator. One possible secondary structure has a 6-bp G

between

creased by rho, even though it appears to be a rho-independent terminator, and decreased by NusA (33). It has also been shown in vitro that NusA decreases termination frequency at the attenuator between rpsU and dnaG in the rpsU-dnaG-rpoD operon (31). Studies are under way to determine if conditions that reduce the termination frequency at the attenuator upstream of fpg exist.

An additional site that corresponds to either a promoter or an RNase III site was detected. This 5′ transcript end (Fig. 2C, lane 2; 318-nucleotide product) in the radC-rpmB intergenic region was detected in wild-type cells but not in rnc me cells and was also detected by primer extension analysis in wild-type cells. This site was previously predicted to be a promoter, and a possible −10 hexamer (TATA CT) and −35 hexamer (TTAGGC) were identified (22). If this site is a promoter, it is completely down-regulated in rnc me cells. Perhaps there is a transcriptional activator which requires the presence of RNase III and/or RNase E for activity. Or this site may be, as it appears, a cleavage site. RNase E cleaves single-stranded RNA, and possible consensus sequences have been proposed. The motif ACAG(A/U)AUUG was predicted based on sequence similarities between RNase E cleavage sites in 9S RNA and RNA I (40). Other investigators, using site-directed mutagenesis of the phage T4 RNase E cleavage site, proposed (A/G)AU(A/U) as the consensus sequence (13). The sequence surrounding the possible cleavage site in the fpg operon is GCCACCUUUG. This sequence is 50% homologous to the first predicted RNase E consensus sequence and 60% homologous to the second. RNase III cleaves double-stranded RNA. A preliminary examination of the region surrounding the possible cleavage site with Genetics Computer Group programs Stemloop and Foldrna did not reveal a secondary structure resembling an RNase III cleavage site that would allow the cleaved residue to be in double-stranded form. No 3′ transcript end was detected with probe 2 or 7 that would correspond to a cleavage event (Fig. 2C, lane 2, and Fig. 3B, lane 4). If the RNA is cleaved at this site by RNase E or RNase III, the upstream RNA must be rapidly degraded. This could be part of the reason that radC appears to be poorly transcribed in wild-type cells. There was not an appreciable increase in the radC 635-nucleotide product in rnc me cells compared to wild-type cells, but the laddering below this band was more intense (Fig. 2B, lanes 2 and 3). It is unclear why this laddering occurred, but it was present with different probe and RNA preparations. A cleavage site between radC and rpmB would explain the increase in the full-length transcript seen in rnc me cells with probe 2 (Fig. 2C, lanes 2 and 3). If there is a cleavage site, though, less transcript should originate from the promoter in the 3′ end of radC because this would also be processed, and this was not the case (Fig. 2C; 483-nucleotide product). Further experiments are needed to determine whether the 5′ transcript end corresponds to a promoter or a cleavage site. Three termination sites were mapped downstream of the fpg operon. The first one corresponds to a predicted stem-loop structure in the fpg-kdltB intergenic region, and the other two correspond to sites in the kdltB coding region (Fig. 2E and 3C).

The mutY-yggX-mltC-nupG operon consists of genes that have no apparent relationship to each other, although the function of the protein encoded by yggX is unknown. There is a possible minor transcription initiation site 102 bp 5′ to the mutY start codon (Fig. 5B). A second transcription initiation site, 25 bp 5′ to the mutY start codon, corresponds to a previously predicted promoter with a proposed −10 hexamer (TG CAAT) and a proposed −35 hexamer (TTTACA) separated by 17 bp (41). A third transcription initiation site, 72 bp 5′ to the mutY stop codon, is 9 bp downstream from a possible −10 hexamer (TATAAC) and a possible −35 hexamer (TTGATG) separated by 16 bp. The last transcription initiation site in the operon is 15 bp 5′ to the nupG start codon and was previously mapped by an S1 nuclease digestion assay (29). There are termination sites in the yggX-mltC and mltC-nupG intergenic regions. Not all transcripts were terminated at these sites, so there must be secondary structures which act as attenuators. These experiments were performed only with wild-type RNA, so we cannot rule out the possibility that the apparent termination sites correspond to RNase III and/or RNase E cleavage sites.

The arrangement of promoters and terminators in the operon appears to allow different combinations of genes to be transcribed together. From the two 5′-most promoters, mutY can be cotranscribed with yggX, with yggX and mltC, or with all three downstream genes. From the promoter in the 3′ end of mutY, yggX can be transcribed alone, with mltC, or with both downstream genes. nupG can be transcribed from its own independent promoter, and the intensity of the RPA product corresponding to this promoter (Fig. 7E, lane 2; 172-nucleotide product) indicates that it is the most active promoter in the operon. The amount of nontagged transcript reading from downstream of nupG into nupG is minor, and it is possible that nupG is predominantly regulated independently of the other genes. It has been shown that the nupG promoter is activated 50-fold by the cyclic AMP-cyclic AMP receptor protein complex and repressed fourfold by CytR and DeoR and that it has binding sites for these proteins (29, 32).

There have been no reports on the regulation of MutY, YggX, or MltC. Since MutY helps repair mismatches resulting from the presence of 8-oxoG in DNA, it is possible that it is up-regulated under conditions that generate the 8-oxoG lesion. MltC is a potentially lytic protein, and it would seem important for cells to control the amount present. The glycan strand of peptidoglycan is made up of N-acetylmuramic acid and N-acetylglucosamine joined together in an alternating sequence by (β-1-4) glycosidic bonds. In order for cells to grow and divide, the glycan strand must be cleaved to allow the insertion of new materials. Lytic transglycosylases can cleave the (β-1-4) bond between N-acetylmuramic acid and N-acetylglucosamine and transfer the muramyl bond to the carbon-6 hydroxyl group of the same N-acetylmuramic acid, forming (1-6) anhydromuramic acid (19). MltC has been shown to be a membrane-bound lytic glycosylase with the same activity (11). E. coli has three other proteins possessing this activity, one soluble (SI70)
and two membrane bound (MltA and MltB). These proteins have been shown to act as exomuramidases in that they start at the glucosamine end and cleave the glycan strand in a processive manner (for a review, see reference 18). It is not known whether each of these proteins plays a specific role in the cell or whether the redundancy serves as a backup mechanism. Since these proteins can cause cell lysis, it would seem necessary to strictly regulate them, but the soluble lytic transglycosylase can be overexpressed 30-fold without harming cells. It was shown that Ski-dependent autolysis is suppressed by RelA-mediated stringent control, possibly by keeping the enzyme in an inactive form or in an unfavorable topological position (3). It has not been shown whether the enzyme activities of the membrane-bound lytic transglycosylases are controlled in a similar manner or whether any of them is regulated at the transcriptional level. Since mlTC does not have an independent promotor, it is possible that the frequency of transcript termination between yggX and mlTC plays a role in regulating MltC.

Searches of the genomes of other organisms revealed that the order of genes in both operons is conserved or partially conserved in several closely related gram-negative bacteria, as shown in Fig. 9. Of course, conservation of gene order does not mean that the transcriptional organization has been conserved, and there are no reports on how these genes are transcribed in the organisms identified. It is tempting to speculate that the conserved gene order implies a functional or regulatory significance, but it may just mean that the organisms are so closely related evolutionarily that the genes have not had time to rearrange. We have not seen any obvious homologies to consensus sequence binding sites for LexA, SoxS, OxyR, or KatF in the promoter regions of the fpg or mutY operon. Studies are currently under way to examine the regulation of each of these operons.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grant R37 CA33657 awarded by the National Cancer Institute. Christine M. Gifford was supported by Environmental Pathology training grant T32 CA33657 awarded by the National Cancer Institute. Christine M. Gifford was supported by Environmental Pathology training grant T32 CA33657 awarded by the National Cancer Institute.

REFERENCES

32. Regnier, P., and C. Portier. 1986. Initiation, attenuation and RNAse III processing of transcripts from the Escherichia coli operon encoding ribo-