

# Molecular Organization and Nucleotide Sequence of the *recG* Locus of *Escherichia coli* K-12

ROBERT G. LLOYD\* AND GARY J. SHARPLES

Department of Genetics, Queens Medical Centre, University of Nottingham, Nottingham NG7 2UH, United Kingdom

Received 24 May 1991/Accepted 5 September 1991

The nucleotide sequence of the *Escherichia coli* K-12 *recG* gene was determined. *recG* was identified as an open reading frame located between the *spoT* operon and the convergent *gltS* gene. It encodes a polypeptide of 693 amino acids which was identified as a 76-kDa protein by sodium dodecyl sulfate-polyacrylamide gel electrophoresis after it was labeled with [<sup>35</sup>S]methionine in maxicells. The sequence determined revealed no obvious promoter. Synthesis of RecG by plasmids carrying the intact gene varied with the orientation of the insert relative to the vector promoter and with the extent of upstream *spoT* operon sequence included in the construction. It is concluded that *recG* is the fourth and last gene in the *spoT* operon, although a possible promoter for independent transcription of *spoU* and *recG* was identified near the end of the *spoT* gene. The primary sequence of RecG revealed that it is related to proteins that act as helicases and has a well-conserved motif identified with ATP binding.

The *recG* locus is required for normal recombination and DNA repair in *Escherichia coli*. The gene was identified by Storm et al. some 20 years ago (43) but received little or no further attention until quite recently, when Lloyd and Buckman (25) described a mini-*kan* insertion (*recG258*) at the same locus. The insertion, which is located between *pyrE* and *dgo* at min 82.1 on the genetic map, reduces conjugational and transductional recombination in *recBC<sup>+</sup> sbc<sup>+</sup>*, *recBC sbcA*, and *recBC sbcBC* strains, and it also confers sensitivity to mitomycin C, UV light, and ionizing radiation. In these respects, it resembles null alleles of *recA*. However, the defects are much more modest. Thus, while conjugational recombination is abolished by a *recA* null mutation, it is reduced by no more than three- to fivefold by the *recG* insertion (25). Furthermore, *recG* has no effect on plasmid recombination, at least not in a *recBC<sup>+</sup> sbc<sup>+</sup>* genetic background, and is not required for induction of the SOS response (47) to DNA damage.

Our interest in *recG* has been stimulated by what appears to be a functional overlap with the *ruv* genes. Strains carrying the *recG* insertion in combination with a mutation in *ruvA*, *ruvB*, or *ruvC* are very deficient in recombination, much more so than strains carrying these mutations alone. They are also extremely sensitive to UV light (23). This observation means that *recG* probably has a far more critical role in recombination and DNA repair than is indicated by the properties of *recG* single mutants. It also provides a clue to the function of the gene product. The *ruv* genes act late in recombination to convert recombination intermediates into viable products (5, 8). This role is supported by biochemical evidence identifying RuvC as a nuclease that resolves Holliday junctions (9). In this paper, we present the nucleotide sequence for *recG* and show that the gene encodes a polypeptide of 76 kDa. The amino acid sequence predicted for RecG shows a relationship to proteins that are known to act as helicases.

## MATERIALS AND METHODS

**Strains.** The *E. coli* K-12 strains used are described in Table 1.  $\lambda$ 571 is from the Kohara et al. collection (19) and carries *recG<sup>+</sup>* (25). Phage M13K07 was obtained from Pharmacia LKB.

**Media and general methods.** LB broth and agar media have been described elsewhere (26) and were supplemented with 50  $\mu$ g of ampicillin per ml or 40  $\mu$ g of kanamycin per ml to maintain the plasmids used or the *recG258* insertion, respectively. *recG* mutant strains were recognized by their very poor growth on LB agar supplemented with mitomycin C at concentrations of 0.2 or 0.5  $\mu$ g/ml, especially when the inoculated plates were irradiated with 30 J of UV light per m<sup>2</sup> before incubation. Strain AB1157 was used in the plate assays as the control for the *recG* mutant N2731. Sensitivity to UV light was measured by irradiating cells on the surface of LB agar plates as described previously (24). Irradiation was at a dose rate of 1 J/m<sup>2</sup>/s.

**DNA analysis.** DNA manipulations followed standard protocols (36). Restriction enzymes, Klenow polymerase, and T4 DNA ligase were from commercial sources. Plasmid stocks were prepared from the *recA* strain AB2463 by using Qiagen protocols and materials as described by the supplier. Exonuclease III (ExoIII) deletions were generated by using the Erase-A-Base system from Promega.

**Plasmid constructions.** Constructs carrying inserts from the *recG* region are shown in Fig. 1. pBR322 (6), pUC18 (51), and pTZ18/19R (Pharmacia LKB) were used as vectors. pBL125 and pBL130 are derivatives of pBR322 and pUC18, respectively, and have been described previously (25). pBL131 was made by cloning the 1.75-kb *PstI*-*PvuII* fragment from pBL125 into pUC18 digested with *PstI* and *SmaI*. pBL135 contains the 2.6-kb *PstI*-*HindIII* subfragment of the insert in pBL125 cloned into pTZ18R, while pBL136 contains a 7.6-kb *KpnI* fragment from  $\lambda$ 571 inserted into pTZ19R. pBL138 is an *SphI* deletion derivative of pBL136, while pBL140 was made by inverting the *KpnI* insert in pBL136. The deletions of pBL135 shown in Fig. 1 were constructed during the sequencing of *recG*. They were made by cutting the vector multiple cloning site with *XbaI* and *SstI* and then digesting the DNA from the *XbaI* end through the *PstI* site into the insert, using ExoIII. The DNA ends were

\* Corresponding author.

TABLE 1. *E. coli* K-12 strains

Strain	Genotype	Reference
AB1157	F <sup>-</sup> <i>thi-1 his-4 Δ(gpt-proA)62 argE3 thr-1 leuB6 kdgK51 rfbD1(?) ara-14 lacY1 galK2 xyl-5 mtl-1 tsx-33 supE44 rpsL31</i>	4
AB2463	Like AB1157, but <i>recA13</i>	4
AB2480	Like AB1157, but <i>recA13 uvrA6</i>	16
N2731	Like AB1157, but <i>recG258::Tn10 mini-kan<sup>a</sup></i>	25
N2419	F <sup>-</sup> <i>thi-1 metE70 leuB6 proC32 lacI3 lacZ118 ara-14 mtl-1 xyl-5 gyrA supD rpsL109 rpsE</i>	25
N2712	Like N2419, but <i>recJ284::Tn10<sup>a</sup></i>	25
N2971	Like N2419, but <i>recG258</i>	25
N2975	Like N2419, but <i>recG258 recJ284</i>	25
JM101	F <sup>-</sup> 128 <i>proAB<sup>+</sup> lacI<sup>q</sup> ΔM15 traD36/Δ(lac-pro) endA1 hsdR17 gyrA96 supE44</i>	51

<sup>a</sup> After the first listing, the transposon insertions are abbreviated to the allele number.

made blunt with S1 nuclease and ligated. pBL147 was made by deleting the *HpaI-HincII* fragment of the insert in pBL138. pBL149 was made by digesting pBL135 with *BglII* and filling in the recessed 3' ends before religating. pBL144 was made by deleting the *PmaCI-StuI* fragment of the insert from pBL140. The deletion removes all but 260 bp of the DNA between the *EcoRI* and *StuI* sites (Fig. 1).

**Identification of plasmid-encoded proteins.** Plasmids were transformed into strain AB2480, and the proteins made were labeled with L-[<sup>35</sup>S]methionine (Amersham) by using the maxicell method of Sancar et al. (37). Labeled proteins were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and visualized by fluorogra-

phy, as described by West and Emmerson (49). <sup>14</sup>C-methylated molecular weight markers were from Amersham.

**DNA sequencing.** Sequencing was by the dideoxy chain termination method of Sanger et al. (38), using T7 sequencing kits from Pharmacia LKB, single-stranded DNA from pTZ18/19R constructs, and [<sup>35</sup>S]dATP. Template DNA was prepared as directed from the appropriate plasmid-containing derivatives of JM101 after infection with M13KO7 helper phage. pBL135, pBL138, pBL140, pBL144, and a series of deletion derivatives generated by digestion with appropriate restriction enzymes and/or ExoIII were used to provide a series of overlapping sequences. When suitable deletions were not available, oligonucleotide primers (18-mers) based on sequences already determined were used to generate the required overlaps. Compressions in GC-rich regions were resolved by replacing dGTP with 7-deaza-dGTP in the sequencing reactions. The entire sequence was read from both strands of the DNA and was compiled and analyzed with software packages from DNA STAR, Ltd.

**Nucleotide sequence accession number.** The sequence of *recG* and the flanking region has been submitted to EMBL and has been given the accession number X59550.

## RESULTS

**Nucleotide sequencing of the *recG* region.** Previous studies (25) showed that the *PstI-HindIII* DNA fragment cloned in pBL130 (Fig. 1) restores UV resistance to strains carrying *recG258*. The nucleotide sequence of this region and of the adjacent 0.42-kb *SphI-PstI* fragment was determined as described in Materials and Methods. The 3,041-bp sequence obtained is shown in Fig. 2. A search of the GenBank data base (release 65) revealed an overlap (100% match) at the 5'

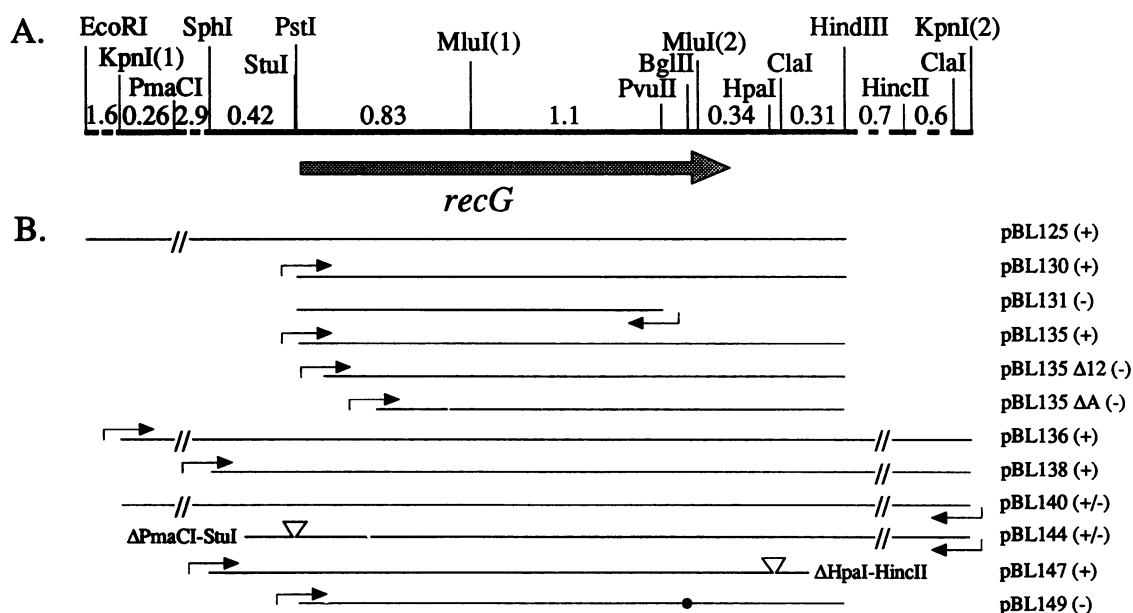


FIG. 1. (A) Restriction map of the *E. coli recG* region and location of the *recG* reading frame (shaded arrow). The sizes shown for the major restriction fragments are in kilobases. (B) Diagram showing the extent of the DNA inserts carried by the plasmid constructs used. The inserts are aligned with respect to the restriction map in panel A. Plasmids are identified on the right; their ability (+) or inability (-) to make strain N2731 resistant to mitomycin C and UV light in plate tests is shown in parenthesis (+/- denotes partial complementation). The orientations of the *lac* promoters in the pUC18 and pTZ18/19R constructs are indicated by arrows. The closed circle in the bottom line denotes a mutated *BglIII* site in pBL149.

*Sph*I  
GCATGCAGATTCTGGCAACCCATCTTTCTGATAACGCTGTCGATTTCGCGCAAATGATTACACTCGCCCGACCTGCATTTTGTATGGGACAGGAGAAAAC 100  
GGGCATCACGCAGGAAGCATTGGCCCTGGCGGATCAGGACATCATCATTCCGATGATCGGCATGGTGCAGTCGCTGAATGTTCCGTTGCCTCAGCCCTC 200  
ATTCTTTACGAAGCCAGCGTCAGCGCAAATGCAGGCATGTACCTGCGTGAAAACAGCATGTTGCCGGAAGCAGAGCAACAACCCCTGTGTTTGAAG 300  
CGCGCTATCCGGTGTGGCGAAAGTCGCAAAACGCAAGGCCTGCCTTATCCCCACGTCATCAGCAAGGCGAGATCGAAGCTGATGCCGACTGGTGGGC 400  
TACTATGCAGGCTGCAGGGTAAAGTGCATGAAAGGTGCGCTGTTAGATGCTGCCCACTCAGTTCCTAACCGGCGTGGCGCAGCACTTAGTAACAAAAC 500  
--->*spoU* M K G R L L D A V P L S S L T G V G A A L S N K L  
TGGCGAAAATCAACCTGCATACCGTGCAGGATCTACTCTTACACCTTCCCTCGGCTACGAAGATCGCACCCATCTCTACCCCATCGGAGAATCTACTGCC 600  
A K I N L H T V Q D L L L H L P L R Y E D R T H L Y P I G E L L P  
GGCGCTTATGCCACGGTGAAGGCGAAGTGTGAACGTCAATATCTCCTTCGGCGGTGGCGGATGATGACCTGCCAGATCAGCGACGGTCCCGGCATC 700  
G V Y A T V E G E V L N C N I S F G G R R M M T C Q I S D G S G I  
CTCACCATGCGCTTTTCAATTCAGCGGGCAATGAAAAATAGCTGGCGGGCGCCGCGTACTGGCTTATGGCGAAGCAAAGCGGTAATATG 800  
L T M R F F N F S A A M K N S L A A G R R V L A Y G E A K R G K Y G  
GTGCGGAGATGATCCACCCGAATACCGCGTGCAGGGCGATCTCAGCACGCCAGAATACAGGAAACGCTCAGCGCGTTTATCCAACAACGGAAGCGCT 900  
A E M I H P E Y R V Q G D L S T P E L Q E T L T P V Y P T T E G V  
AAAGCAGGCCACGCTGCGTAAATTAACCGACCGGCGCTGGATCTGCTGCACACCTGCGCCATTGAAGAACTCCTGCCCGGAACTGTCACAAGGAATG 1000  
ATGACGCTACCGGAAGCGTGGCGCACTTGCACCGCCCGCCACGCGTACTAGCTGGAACCGGGCAGCATCCGGCGCAACGTCGCTGTA 1100  
M T L P E A L R T L H R P P P T L Q L S D L E T G Q H P A Q R R L I  
TTCTGGAAGAACTGCTGGCGCACAACTCAGCATGTTAGCCTTACGTGCCGAGCACAGCGTTTTCATGCCAGCGCTGAGCGCAATGACACGCTGAA 1200  
L E E L L A H N L S M L A L R A G A Q R F H A Q P L S A N D T L K  
AAATAAATCCTCGCCGCTTACCCTTCAAGCCAACGGGCGCACAGGCACGCGTAGTGGCGGAGATCGAGCGGATGAGCGCTGGATGTCGCCGATGATG 1300  
N K L L A A L P F K P T G A Q A R V V A E I E R D M A L D V P M M  
CGTCTGGTCAGGGCGATGAGTTCGGTAAACCGTGGTCCGCGCCCTCGCCGCTGCGTGGCATTGCCACGCAACAGTAGCATTGATGGCAC 1400  
R L V I Q G D V G S G K T L V A A L A L R A I A H G K Q V A L M A P  
CAACCGAAATTAATCGCGAGCAGCACGCAATAACTTCCGCAACTGGTTGCACCGCTCGGTATCGAAGTGGGCTGGCTCGCCGGAAGCAGAAAGTAA 1500  
T E L L A E Q H A N N F R N W F A P L G I E V G W L A G K Q K G K  
AGCAGCGTGGCAGCAGGAAGCCATCGCAGCGGTGAGTGCAGATGATTGTCGGTACACACGCCATCTCCAGGAACAGGTGCGATTAAAGCCCTG 1600  
A R L A Q Q E A I A S G Q V Q M I V G T H A I F Q E Q V Q F N G L  
GCGTGGTGAATTCGACGACAGCATCGTTTGGCGTGCATCAGCTGCGTGGGAGAAAGGCCAGCAGCGCTCCATCCGCACTCAGTTGA 1700  
A L V I I D E Q H R F G V Q R L A L W E K G Q Q Q G F H P H Q L I  
TCATGACCGCCACGCGCATCCCCCGCAGCTGGCAATGACTGCGTATGCCGATCTCGATACCTCGGTGATAGATGAGTGCAGCGAGCGCCAGT 1800  
M T A T P I P R T L A M T A Y A D L D T S V I D E L P P G R T P V  
GACTACGGTGCATTTCTGATACCCGCGTACCACATCATGACCGGTGCACCGCCTGCATAACTGAAGTGCAGGATGAGTGGTGGTGGTGGTGGT 1900  
T T V A I P D T R R T D I I D R V H H A C I T E G R Q A Y W V C T  
TTGATTGAAGAGTCGGAATGCTGGAAGCGCAGGCGCGGAAGTACCTGGGAAGAGTTGAACTGGCGCTACCAGAGTTGAACGTTGGCTGGTACAG 2000  
L I E E S E L L E A Q A A E A T W E E L K L A L P E L N V L H G  
GGCGGATGAAACCTGCCGAGAAACAGGCGGTGATGGCTCGTTTAAACAAGGTGAGCTACACCTGCTGGTTGCCACAACCGTTATTGAAGTCGGCTGA 2100  
R M K P A E K Q A V M A S F K Q G E L H L L V A T T V I E V G V D  
TGTGCTAACCGCAGTCTGATGATTATCGAAAACCGGAGCGTGGGTGCTGGCGCAGTTACACCGCTGCGCGGGCGGTAGTCTGGCGCGTGGCT 2200  
V P N A S L M I I E N P E R L G L A Q L H Q L R G R V G R G A V A  
TCTACTGCGTCTGCTTACAAAACCGCGCTTTCTAAAACCGCGCAATTCGCCCTGCAAGTCTGCGCGACAGTAACGACGGTTTGTGATTGCGCAAA 2300  
S H C V L L Y K T P L S K T A Q I R L Q V L R D S N D G F V I A Q K  
AAGATCTGGAGATTCGCGCCCTGGCGAATGTTAGGCACGCGTACAGCGGTAATGCTGAATTTAAAGTGGCGGATTTACTGCGCATCAGGCGATGAT 2400  
D L E I R G P G E L L G T R Q T G N A E F K V A D L L R D Q A M I  
CCCGAAGTTCAGCGCTGGCAGCCATATTCACGAACGTTACCCACAACAGGCAAAAGCCCTGATAGAAGCTGGATGCCGAGACGGAACGTTACTCG 2500  
P E V Q R L A R H I H E R Y P Q Q A K A L I E R W M P E T E R Y S  
AATGCGTAAAAGCGCAGTGAAGACCGCCATTTACAGTTACCTACCTCTCTGCCGATGCGATTATCACCTACAATTAATAAATATGATATC 2600  
N A  
AATACGAGGCTTGATAAGCGAAGCGTATCAGGCATTTGTCTTTGTTGGCGGGCACTTCTACCGCGCTTCATCGGTTAACCGGCAAAAATCGGCAACA 2700  
HpaI  
gltS <---  
TCAAATACTAATTAATACCAGCATTGACGATATCGATAAAGAAGCGCAGCAGTACCGCACCAAAAAGCCATGTGCGACGGGCAAAAGCGTTC 2800  
EcoRV/ClaI  
AGTATCGCCTGATGTTGGCGATTGCCGTTGCGTTCACCGAGGCAAAACACAGTACCGCAGCAGCACTGCCGATCGTAGTTTTTGGCCATC 2900  
ATGCGCAGGTAACGAAGATGGCATAACAACGCAATGAAGTGGTCTGTACCACAGAAATCGCCAGCATCGGCAGCGCAGCAAGCCAGTCCCACAGTT 3000  
TCAGCCCCATCAACGCCATCGCCAGGAACAACGACAAGCTT 3041  
HindIII

FIG. 2. Nucleotide sequence of the 3,041-bp *Sph*I-*Hind*III fragment of the *recG* region. Relevant restriction sites are labeled above the first base of the sequence recognized. The translation of *recG* is shown beneath the sequence. Stop codons identifying the end of *spoU* and *gltS* are underlined and labeled as such.

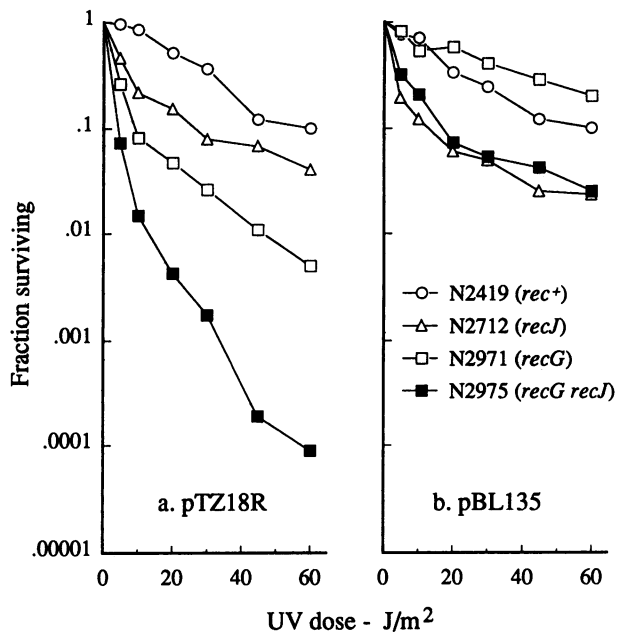


FIG. 3. Effect of pBL135 on survival of UV-irradiated *recG* strains. The strains and plasmids used are identified within the panels.

end with the sequence of the *spo* operon published by Sarubbi et al. (39) which extends to bp 428. The overlap contains the 3' end of *spoU* (50), which terminates with the TAA stop codon at bp 420. A further overlap, this time of 740 bp at the 3' end, with the recently published sequence for *gltS* was found (17). The *gltS* gene is transcribed in the opposite direction to *spoU* and terminates with a TAA stop codon at bp 2680.

**Identification of *recG*.** The DNA sequence between *spoU* and *gltS* contains an open reading frame extending 2,079 nucleotides, from a potential ATG initiation codon at bp 428 to a TAA stop codon at bp 2507. A region of dyad symmetry immediately downstream (bp 2508 to 2536) could provide a transcriptional terminator. A potential terminator for the convergent *gltS* gene is located at bp 2647 to 2668. Translation of the large open reading frame would produce a polypeptide of 693 amino acids with a predicted molecular mass of 76,438 Da. pBL135, which has this open reading frame intact, appears to have a fully functional *recG* gene. It restores UV resistance to a *recG258* single mutant and overcomes completely the more substantial effect seen with the insertion in a *recJ* mutant background (Fig. 3). It also produces a rather poorly expressed protein of 76 kDa (Fig. 4A). The two deletions of pBL135 shown in Fig. 1B ( $\Delta 12$  and  $\Delta A$ ) extend to bp 428 and 807, respectively. As indicated, both have lost the ability to complement *recG* (data not shown). Also, pBL135 $\Delta A$  no longer produces the 76-kDa protein. It produces instead two new proteins migrating in the region of 62 kDa. A protein of about this size would be expected if translation initiated with the ATG at bp 809. Plasmids carrying 3' deletions extending to the *ClaI* site (not shown) or to the *HpaI* site (Fig. 1, pBL147) are able to complement *recG*, whereas a deletion to the *PvuII* site (pBL131) or a frameshift mutation created at the *BglIII* site (pBL149) eliminates this activity. pBL131 also fails to produce the 76-kDa protein (Fig. 4B). We conclude that the reading frame identified is *recG* and that the gene product is a protein of 76 kDa.

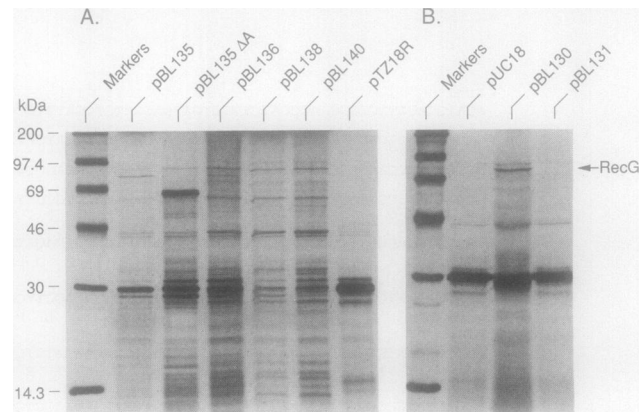


FIG. 4. Fluorogram showing [<sup>35</sup>S]methionine-labeled proteins from AB2480 maxicells separated on a 10% polyacrylamide-SDS gel. The plasmids are identified above each lane. Molecular mass markers are the same in both panels.

**Location of *recG258*.** We made use of the fact that a *recG258* strain is reasonably proficient in recombination to transfer the kanamycin resistance DNA insertion to pBL136 and pBL144. Stocks of these plasmids were prepared from strain N2731(*recG258*), and *recG258* mutant plasmids arising from recombination with the chromosome were identified by transforming strain AB2463 (*recA*) to Ap<sup>r</sup> Km<sup>r</sup>. Four independent *recG258* recombinant plasmids were obtained. Restriction analysis of these plasmids (data not shown) revealed that the insertion is located in the 0.83-kb *PstI-MluI*(1) fragment (Fig. 1).

***recG* is part of the *spo* operon.** Inspection of the sequence immediately upstream of the proposed ATG initiation codon for *recG* reveals no obvious signals for the expression of the gene. The AGG located 9 bp upstream provides a possible ribosome binding site (40), but there are no good matches to the -10 and -35 sequences typical of *E. coli* promoters (14). Sarubbi et al. (39) and Xiao et al. (50) describe two promoters (P1 and P2) that could transcribe through the *spo* operon and lead to the expression of *recG* (Fig. 5). The 5-bp spacing between *spoU* and *recG* is consistent with this view. The plasmid constructs we have examined do not carry these promoters, with the exception of pBL125, which has P2. Therefore, we assume that *recG* is being expressed from vector promoters in those constructs that restore UV resistance to *recG258* strains, or perhaps from secondary promoters in the *spo* coding region. This conclusion is supported by our analysis of the proteins labeled in maxicells from strains carrying different sections of *spo* DNA (Fig. 4).

pBL130 and pBL135 carry identical *PstI-HindIII* inserts,

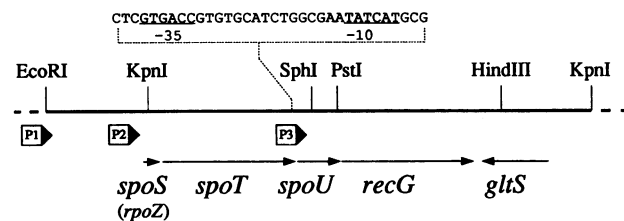


FIG. 5. Organization of the *spoT-recG-gltS* region of the chromosome. P1 and P2 are the promoters defined by Xiao et al. (50). P3 is defined in the text.



FIG. 6. Alignment of RecG with eIF-4A and p68. The alignment is based on the FASTP program (22). Conservative substitutions (identified by colons) indicate evolutionarily related residues. Those residues that score zero in the PAM250 matrix (22) are indicated by a period. Motifs I to VI (identified by asterisks) are as defined by Gorbalenya et al. (12). The consensus (con.) for helices shows both identical and functionally related residues (+, hydrophobic residues [I, L, V, M, F, Y, and W; o, charged or polar residues [S, T, D, E, N, Q, K, and R]) in the region of each motif. The consensus is taken from the lower portion of Fig. 3a by Gorbalenya et al. (12).

and it is clear that both produce the 76-kDa RecG protein. Since they have only a few nucleotides of the upstream *spoU* sequence, it is clear that *recG* must be expressed from the vector *lac* promoter in these constructs. pBL136 and pBL138 carry more of the upstream sequence. Both make the *recG* strain N2731 resistant to UV light and mitomycin C. However, only pBL136 makes enough of the 76-kDa protein to be visible in maxicell extracts (Fig. 4A), despite the fact that *recG* is closer to the vector promoter in pBL138. This suggests that *recG* can be expressed by pBL136 from a promoter located between *KpnI*(1) and *SphI*. Inspection of the sequence reported by Sarubbi et al. (39) reveals a possible candidate (P3 in Fig. 5) 328 bp upstream from the *SphI* site. The proposed -10 region has a good match (five of six nucleotides) to the consensus (14), but the -35 region does not match as well (four of six nucleotides), and the spacing of 16 bp is not optimal. pBL140 carries the same insert as pBL136 but in the reverse orientation with respect to the *lac* promoter. The 76-kDa protein is not seen with this construct, and the plasmid is unable to make N2731 fully resistant to UV light and mitomycin C (Fig. 4A) (data not shown). When the insert was reversed once more, the plasmid was able to confer full resistance. P3 is either a weak promoter or is inhibited in pBL140 by transcription from the *lac* promoter. It is quite noticeable that the truncated RecG proteins encoded by pBL135ΔA are expressed quite strongly. We assume that this is because the deletion has placed the truncated *recG* reading frame at a more favorable distance from the *lac* promoter.

**RecG is related to a family of proteins that act as helices.**  
A search of the NBRF-PIR protein data base (release 26) by the methods of Lipman and Pearson (22) revealed significant homologies between RecG and a family of related RNA and DNA helices (12, 15). An alignment of RecG with the known RNA helices eIF-4A (31) and p68 (10) is shown in Fig. 6. While the similarity is not great, the overall alignment of the seven motifs is quite striking given that eIF-4A and p68 are mammalian proteins. Furthermore, within these motifs, RecG has most of the amino acid residues identified by Gorbalenya et al. (12) as being particularly highly conserved (Fig. 6). RecG also aligns well with the DNA helicase n' (21, 32) over several of these motifs. Motifs I and II, which are presumed to be associated with ATP binding (7, 13, 48), are conserved well in *E. coli* UvrD, UvrB, RecB, and RecQ, all of which are DNA helices, or components of helices, involved in recombination and DNA repair (29, 46).

**DISCUSSION**

The DNA region encoding *recG* has been shown to contain a reading frame for a protein of 693 amino acids. Complementation studies using plasmids with deletions extending into this region from either the 5' end or the 3' end established that the reading frame is *recG*. The gene is flanked on the 5' side by the *spoT* operon and on the 3' side by *gltS*, which is transcribed in the opposite direction. Our analysis of *recG* expression suggests that the gene is tran-

scribed from promoters that serve the *spoT* operon. This would make *recG* the fourth gene in this operon (Fig. 5). *spoS* (*rpoZ*) encodes the omega subunit of RNA polymerase (11), while the *spoT* gene itself is involved with the metabolism of guanosine 3',5'-bispyrophosphate (ppGpp) during the stringent response (20, 39). The function of *spoU* is unknown. Mutations in the *spo* region that disconnect *recG* from the upstream promoters might therefore be expected to have a polar effect on the expression of *recG*. This appears to be the case (18). The *recG162* allele identified by Storm et al. (43) could be an example of such a mutation. Lloyd and Buckman (25) reported that *recG162* has a less severe effect on UV repair than the *recG258* insertion.

The results reported by Sarubbi et al. (39) suggest that the *spoT* operon is not highly expressed. This would be consistent with our observation that very little of the *recG* product is needed to overcome the DNA repair defect in strains carrying *recG258*. The 76-kDa product of *recG* was identified by SDS-PAGE after labeling with <sup>35</sup>S in maxicells. However, while some of the plasmid constructs complement *recG* very well, they do not make enough RecG protein to be visible in labeled maxicell extracts. The low expression of *recG* and its positioning with genes specifying diverse functions is not unusual among recombination genes. A similar situation is seen with *recF* (1, 2), *recJ* (27), and *recO* (30).

The primary structure predicted for RecG suggests that it may have ATPase and helicase activities, which supports the idea that it has a direct role in genetic recombination. Several DNA helicases have been linked with recombination, including UvrD (helicase II) (3, 28, 33), RecBCD enzyme (35, 44, 45), and RecQ (46). Ideas abound as to how these enzymes process DNA during the course of recombination (8, 41, 42). One common theme is that a combination of DNA helicase and nuclease activities provides the single-stranded substrate needed by RecA, as has been suggested for RecQ helicase and RecJ nuclease (46). Another is that such a combination of activities may allow strand exchange to initiate after RecA has formed a paranemic joint (34). The latter possibility is particularly attractive for RecG in view of the functional overlap with the Ruv resolvase (23). However, a test of this possibility will have to await the biochemical characterization of the purified protein.

#### ACKNOWLEDGMENTS

We are grateful to M. Cashel, in whose laboratory the *recG* region was sequenced independently, for exchanging information prior to publication and also for telling us about the polar effects of deletion/insertion mutations in the upstream region. We also thank Carol Buckman for excellent technical assistance and our colleague John Keyte in the Department of Biochemistry, University of Nottingham, for synthesizing the oligonucleotides used.

This work was supported by a grant from the Wellcome Trust and, in part, by grants from the Medical Research Council and the Science and Engineering Research Council.

#### REFERENCES

- Armengod, M. E., M. Garcia-Sogo, and E. Lambies. 1988. Transcriptional organization of the *dnaN* and *recF* genes of *Escherichia coli* K-12. *J. Biol. Chem.* **263**:12109-12114.
- Armengod, M. E., and E. Lambies. 1986. Overlapping arrangement of the *recF* and *dnaN* operons of *Escherichia coli*: positive and negative control sequences. *Gene* **43**:183-196.
- Arthur, H. M., D. Bramhill, P. Eastlake, and P. T. Emmerson. 1982. Cloning of the *uvrD* gene of *E. coli* and identification of the product. *Gene* **19**:285-295.
- Bachmann, B. J. 1987. Derivations and genotypes of some mutant derivatives of *Escherichia coli* K-12, p. 1190-1219. In F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
- Benson, F., S. Collier, and R. G. Lloyd. 1991. Evidence of abortive recombination in *ruv* mutants of *Escherichia coli* K-12. *Mol. Gen. Genet.* **225**:266-272.
- Bolivar, F., R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heynecker, H. W. Boyer, J. H. Cross, and S. Falkow. 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* **2**:95-113.
- Bradley, M. K., T. F. Smith, R. H. Lathrop, D. M. Livingston, and T. A. Webster. 1987. Consensus topology in the ATP binding site of the simian virus 40 and polyomavirus large tumor antigens. *Proc. Natl. Acad. Sci. USA* **84**:4026-4030.
- Clark, A. J., and K. B. Low. 1988. Pathways and systems of homologous recombination in *Escherichia coli*, p. 155-215. In K. B. Low (ed.), *The recombination of genetic material*. Academic Press, Inc., New York.
- Connolly, B., C. Parsons, F. E. Benson, H. J. Dunderdale, G. J. Sharples, R. G. Lloyd, and S. C. West. 1991. Resolution of Holliday junctions in vitro requires *Escherichia coli* *ruvC* gene product. *Proc. Natl. Acad. Sci. USA* **88**:6063-6067.
- Ford, M. J., I. A. Anton, and D. P. Lane. 1988. Nuclear protein with sequence homology to translation initiation factor eIF-4A. *Nature (London)* **332**:736-738.
- Gentry, D. R., and R. R. Burgess. 1989. *rpoZ*, encoding the omega subunit of *Escherichia coli* RNA polymerase, is in the same operon as *spoT*. *J. Bacteriol.* **171**:1271-1277.
- Gorbalenya, A. E., E. V. Koonin, A. Donchenko, and V. M. Blinov. 1989. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res.* **17**:4713-4730.
- Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko, and V. M. Blinov. 1988. A conserved NTP-motif in putative helicases. *Nature (London)* **333**:22-23.
- Harley, C. B., and R. P. Reynolds. 1987. Analyses of *E. coli* promoter sequences. *Nucleic Acids Res.* **15**:2343-2361.
- Hodgman, T. C. 1988. A new superfamily of replicative proteins. *Nature (London)* **333**:22-23.
- Howard-Flanders, P., R. P. Boyce, and L. Theriot. 1966. Three loci in *Escherichia coli* K-12 that control the excision of pyrimidine dimers and certain other mutagen products from DNA. *Genetics* **53**:1119-1136.
- Kalman, M., D. R. Gentry, and M. Cashel. 1991. Characterization of the *Escherichia coli* K12 *gluS* glutamate permease gene. *Mol. Gen. Genet.* **225**:379-386.
- Kalman, M., H. Murphy, and M. Cashel. (National Institutes of Health). 1991. Personal communication.
- 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.
- Laffler, T., and J. Gallant. 1974. *spoT*, a new genetic locus involved in the stringent response in *E. coli*. *Cell* **1**:27-30.
- Lee, E. H., H. Masai, G. C. Allen, and A. Kornberg. 1990. The *priA* gene encoding the primosomal replicative n' protein of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **87**:4620-4624.
- Lipman, D. J., and W. R. Pearson. 1985. Rapid and sensitive protein similarity searches. *Science* **227**:1435-1441.
- Lloyd, R. G. 1991. Conjugational recombination in resolvase-deficient mutants of *Escherichia coli* K-12 depends on *recG*. *J. Bacteriol.* **173**:5414-5418.
- Lloyd, R. G., and S. D. Barbour. 1974. The genetic location of the *sbcA* gene of *Escherichia coli*. *Mol. Gen. Genet.* **134**:157-171.
- Lloyd, R. G., and C. Buckman. 1991. Genetic analysis of the *recG* locus of *Escherichia coli* K-12 and of its role in recombination and DNA repair. *J. Bacteriol.* **173**:1004-1011.
- Lloyd, R. G., K. B. Low, N. G. Godson, and E. A. Birge. 1974. Isolation and characterization of an *Escherichia coli* K-12 mutant with a temperature-sensitive RecA<sup>-</sup> phenotype. *J. Bacteriol.* **120**:407-415.

27. Lovett, S. T., and R. D. Kolodner. 1991. Nucleotide sequence of the *Escherichia coli* *recJ* chromosomal region and construction of RecJ-overexpression plasmids. *J. Bacteriol.* **173**:353–364.
28. Maples, V. F., and S. R. Kushner. 1982. DNA repair in *Escherichia coli*: identification of the *uvrD* gene product. *Proc. Natl. Acad. Sci. USA* **79**:5616–5620.
29. Matson, S. W., and K. A. Kaiser-Rogers. 1990. DNA helicases. *Annu. Rev. Biochem.* **59**:289–329.
30. Morrison, P. T., S. T. Lovett, L. E. Gilson, and R. Kolodner. 1989. Molecular analysis of the *Escherichia coli* *recO* gene. *J. Bacteriol.* **171**:3641–3649.
31. Nielsen, P. J., G. K. McMaster, and H. Trachsel. 1985. Cloning of eukaryotic protein synthesis initiation factor genes: isolation and characterization of cDNA clones encoding factor eIF-4A. *Nucleic Acids Res.* **13**:6867–6880.
32. Nurse, P., R. J. Digate, K. H. Zavitz, and K. J. Marians. 1990. Molecular cloning and DNA sequence analysis of *Escherichia coli* *prfA*, the gene encoding the primosomal protein replication factor Y. *Proc. Natl. Acad. Sci. USA* **87**:4615–4619.
33. Oeda, K., T. Horiuchi, and M. Sekiguchi. 1981. Molecular cloning of the *uvrD* gene of *Escherichia coli* that controls ultraviolet sensitivity and spontaneous mutation frequency. *Mol. Gen. Genet.* **184**:191–199.
34. Radding, C. M. 1991. Helical interactions in homologous pairing and strand exchange driven by RecA protein. *J. Biol. Chem.* **266**:5355–5358.
35. Rosamond, J., K. M. Telander, and S. Linn. 1979. Modulation of the action of the *recBC* enzyme of *Escherichia coli* K12 by  $Ca^{2+}$ . *J. Biol. Chem.* **254**:8646–8652.
36. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
37. Sancar, A., A. M. Hack, and W. D. Rupp. 1979. A simple method for identification of plasmid-coded proteins. *J. Bacteriol.* **137**:692–693.
38. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.
39. Sarubbi, E., K. E. Rudd, H. Xiao, K. Ikehara, M. Kalman, and M. Cashel. 1989. Characterization of the *spoT* gene of *Escherichia coli*. *J. Biol. Chem.* **264**:15074–15082.
40. Shine, J., and L. Dalgarno. 1974. The 3'-terminal sequence of *Escherichia coli* 16S RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA* **71**:1342–1346.
41. Smith, G. R. 1988. Homologous recombination in prokaryotes. *Microbiol. Rev.* **52**:1–28.
42. Stahl, F. W., L. C. Thomason, I. Siddiqi, and M. M. Stahl. 1990. Further tests of a recombination model in which  $\chi$  removes the RecD subunit from the RecBCD enzyme of *Escherichia coli*. *Genetics* **126**:519–533.
43. Storm, P. K., W. P. M. Hoekstra, P. G. De Haan, and C. Verhoef. 1971. Genetic recombination in *Escherichia coli*. IV. Isolation and characterization of recombination-deficient mutants of *Escherichia coli* K12. *Mutat. Res.* **13**:9–17.
44. Taylor, A., and G. R. Smith. 1980. Unwinding and rewinding of DNA by the RecBC enzyme. *Cell* **22**:447–457.
45. Telander-Muskavitch, K. M., and S. Linn. 1982. A unified mechanism for the nuclease and unwinding activities of the *recBC* enzyme of *Escherichia coli*. *J. Biol. Chem.* **257**:2641–2648.
46. Umez, K., K. Nakayama, and H. Nakayama. 1990. *Escherichia coli* RecQ protein is a DNA helicase. *Proc. Natl. Acad. Sci. USA* **87**:5363–5367.
47. Walker, G. C. 1984. Mutagenesis and inducible responses to deoxyribonucleic acid damage in *Escherichia coli*. *Microbiol. Rev.* **48**:60–93.
48. Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay. 1982. Distantly related sequences in the  $\alpha$ - and  $\beta$ -subunits of ATP synthase, myosin, kinase and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**:945–951.
49. West, S. C., and P. T. Emmerson. 1977. Induction of protein synthesis in *Escherichia coli* following UV or  $\gamma$ -irradiation, mitomycin C treatment or *tif* expression. *Mol. Gen. Genet.* **151**:57–67.
50. Xiao, H., M. Kalman, K. Ikehara, S. Zemel, G. Glaser, and M. Cashel. 1991. Residual guanosine 3',5'-bispyrophosphate synthetic activity of *relA* null mutants can be eliminated by *spoT* null mutations. *J. Biol. Chem.* **266**:5980–5990.
51. 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.