JOURNAL OF BACTERIOLOGY, Oct. 1990, p. 6042–6047
0021-9193/90/106042-06$02.00/0

Vol. 172, No. 10

Sequence and Expression of the *Escherichia coli recR* Locus

THOMAS YEUNG,[1]† DAVID A. MULLIN,[2] KEN-SHIUNG CHEN,[1] ELIZABETH A. CRAIG,[3] JAMES C. A. BARDWELL,[3]‡ AND JAMES R. WALKER[1]*

Department of Microbiology, University of Texas, Austin, Texas 78712[1]; Department of Biology, Tulane University, New Orleans, Louisiana 70118[2]; and Department of Physiological Chemistry, University of Wisconsin Medical School, Madison, Wisconsin 53706[3]

Received 9 March 1990/Accepted 17 July 1990

The *Escherichia coli* RecR protein participates in a recombinational DNA repair process. Its gene is located in a region of chromosome that extends from 502 to 509 kilobases on the physical map and that contains *apt*, *dnaX*, *orf12-recR*, *htpG*, and *adk*. Most, if not all, of these are involved in nucleic acid metabolism. The *orf12-recR* reading frames consist of 935 base pairs and overlap by one nucleotide, with the 3′ A of the *orf12* termination codon forming the 5′ nucleotide of the *recR* initiation codon. The *orf12-recR* promoter was located upstream of *orf12* by sequence analysis, promoter cloning, and S1 nuclease protection analysis. The start point of transcription was determined by primer extension. The transcript 5′ end contained a long, apparently untranslated region of 199 nucleotides. Absence of a detectable promoter specific for *recR* and the overlap of the *orf12* and *recR* reading frames suggest that translation of *recR* is coupled to that of *orf12*. By maxicell analysis, it was determined that both *orf12* and *recR* are translated.

Escherichia coli recR mutants were identified by Mahdi and Lloyd (20) as derivatives of a *recB sbcB sbcC* strain which became recombination deficient and UV sensitive. *recR* mutations reduced recombination after conjugation or transduction in a *recBC sbcBC* background but had little effect in a $recBC^+$ $sbcBC^+$ background (20). *recR* mutations also increased UV sensitivity, but in both *recBC sbcBC* and $recBC^+$ $sbcBC^+$ strains. It was concluded, therefore, that the RecR product participates in a recombinational repair pathway (20). The fact that *recR* mutations decreased recombination proficiency and UV repair when combined as *recB recR* but not as *recF recR* double mutants indicated that *recR* is part of the RecF pathway (20). Although the *recR* mutation had little effect on recombination after conjugation or transduction in an otherwise wild-type strain, it did cause deficiency in plasmid recombination (20).

The *recR* gene was mapped near min 11 (20), clockwise of and near the DNA replication gene *dnaX* (16, 22). We report here that the region between *dnaX* and the nearby *htpG* gene (3) contains two overlapping reading frames which encode proteins of 12 and 22 kilodaltons (kDa). The first reading frame is designated *orf12* in accordance with the proposal of Mahdi and Lloyd (21), who also sequenced this region. (Their report appeared while this manuscript was in preparation.) The second is the *recR* gene, as shown also by Mahdi and Lloyd (21).

The *orf12* and *recR* frames overlap by one nucleotide pair, suggesting translational coupling, and the promoter which expresses both is located upstream of *orf12* within the *dnaX* coding sequence.

MATERIAL AND METHODS

Bacterial strains, plasmids, and bacteriophages. The *E. coli* K-12 strains are listed in Table 1. pBJ1 is a 6.2-kilobase-pair (kb) *Eco*RI fragment carrying *apt-dnaX-orf12-recR-htpG* and a portion of the *adk* region cloned into pBR322 (Fig. 1). pBJ10 is a pBJ1 derivative with a 1.3-kb kanamycin resistance (Kan^r) cassette from pUC4K (31, 33) inserted into the *recR Sal*I site. pTTG20 is a *Nar*I-*Xmn*I fragment carrying most of the *dnaX* gene, the entire *orf12* and *recR* genes, and a portion of *htpG* cloned into the pUC18 (36) *Nar*I and *Sma*I sites. pTTG30 and pTTG40 are pTTG20 derivatives deleted for a portion of *recR* and portions of *orf12* and *recR*, respectively. pTTG30 was constructed by opening pTTG20 with *Sal*I, blunt-ending the *Sal*I site, digestion with *Hpa*I, and religating. pTTG40 was constructed by opening pTTG20 with *Sal*I and *Bst*EII, blunting both sites, and ligating. pKKC11 consists of the 514-base-pair (bp) *Bst*EII-*Ssp*I (positions 2225 to 2738) *recR* promoter fragment cloned into pKK232-8 (6). pTTG1 is a 664-bp *Pst*I-*Sal*I fragment cloned into pUC19 (36) restricted by the same enzymes. TTG2 is an M13mp19 (36) derivative carrying the 2,303-bp *Nar*I-*Hpa*I fragment cloned into its *Acc*I-*Sma*I sites. TTG3 is an M13mp19 derivative with the 747-bp *Pst*I-*Xmn*I fragment in its *Pst*I and *Sma*I sites.

A 311-bp *Bam*HI *tac* fragment was cloned into pKK232-8 to generate pTG1, in which *tac* is fused to the chloramphenicol acetyltransferase (CAT) gene. The *tac* fragment was cut from pMB211 (M. Bröker, personal communication) after first deleting an *Eco*RI *lacZ* fragment.

Recombinant DNA technology. Standard techniques (2) were used for plasmid DNA isolation, restriction, generation of blunt ends by use of T4 DNA polymerase, Bal31 exonuclease digestion, ligation, transformation, gel electrophoresis, and Southern blotting.

DNA sequencing. The chain termination method (26) was used to sequence a set of overlapping fragments from pTTG1 cloned into M13mp18 or mp19 (36). The *Pst*I site (position 3074) and the *Sal*I site (position 3732) were sequenced across by using TTG2 and TTG3, respectively, as templates.

Plasmid recombination. The method of Kolodner et al. (18) was used to measure recombination between two mutant tetracycline resistance alleles carried on one plasmid to generate a tetracycline-resistant wild-type allele.

Transduction. P1 *vir*a was used for transduction by the method of Willetts et al. (34).

* Corresponding author.
† Present address: Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720.
‡ Present address: Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA 02115.

TABLE 1. *E. coli* strains used

Strain	Relevant genotype	Source of derivation
AB1157	*recB*$^+$ *recC*$^+$ *sbcB*$^+$ *sbcC*$^+$	B. J. Bachmann
JC7623	*recB recC sbcB sbcC*	B. J. Bachmann
JCB325	*recB recC sbcB sbcC recR*::Kan^r	This study
NK5992	*recB*$^+$ *recC*$^+$ *argA*::Tn*10*	B. J. Bachmann
TY108	*recB*$^+$ *recC*$^+$ *sbcB*$^+$ *sbcC*$^+$ *recR*::Kan^r	P1 · JCB325 × AB1157
TY203	*recB*$^+$ *recC*$^+$ *argA*::Tn*10* *sbcB sbcC*	P1 · NK5992 × JC7623
TY313	*recB*$^+$ *recC*$^+$ *argA*::Tn*10* *sbcB sbcC recR*::Kan^r	P1 · NK5992 × JCB325
A19	*rna met relA*	B. A. Hardesty
CSR603	*phr-1 recA1 uvrA6*	C. F. Earhart
HB101	Plasmid host	W. Folk (5)
JM103	Plasmid host	J. Messing (33)

Radiolabeling of proteins in vivo. Proteins were labeled with [^{35}S]methionine in the maxicell (29) system with strain CSR603 as the host for the plasmids and with D-cycloserine (100 μg/ml) instead of ampicillin.

Electrophoresis of radiolabeled proteins. For electrophoresis, 15% denaturing, discontinuous polyacrylamide-sodium dodecyl sulfate gels and 5% stacking gels were used (2). Labeled proteins were visualized by exposing gels to Kodak X-Omat XAR-1 film. Molecular weight standards were from Bio-Rad Laboratories.

Enzyme assays. CAT and β-lactamase were assayed in the same extract by the procedures of Seed and Sheen (27) and Lupski et al. (19), respectively. Extracts were prepared as described before (19).

S1 nuclease protection assay. The procedure of Ohta et al. (23) was used for S1 assays. The 650-bp *Bst*EII fragment (positions 2225 to 2875) (Fig. 1) was purified by electroelution and 5′-end labeled. This fragment was denatured and used as probe, or it was cut at the *Hae*II site (position 2354) and the resulting 526-bp, 3′-*Hae*II-*Bst*EII-5′ subfragment (labeled only at the 5′ end) was purified and used. Probe sufficient to provide 200,000 cpm was used.

Primer extension. The method of Inouye et al. (14) was used for primer extension, except that the crude RNA preparation was purified by sedimentation through a 5.7 M CsCl solution (8). Avian myeloblastosis virus reverse transcriptase was from Promega. The primer was the deoxyribonucleotide 5′-CCGACGCAGGGTCTGAA-3′.

Nucleotide sequence accession number. The nucleotide sequence data reported in this paper have been submitted to GenBank under accession no. M37084.

RESULTS

Nucleotide sequence of the *orf*12-*recR* region. The region between the *dnaX* and *htpG* genes (Fig. 1A) was sequenced by the chain termination method (26). The sequence downstream of *dnaX* to the *Pst*I site beginning at nucleotide 3074 was reported by Flower and McHenry (9); the sequence upstream of *htpG* to the *Sal*I site at position 3732 was reported by Bardwell and Craig (3). We determined the sequence between the *Pst*I and *Sal*I sites on both strands and sequenced across both sites on one strand. While this manuscript was in preparation, Mahdi and Lloyd (21) reported the sequence between the *Bst*EII (position 2875) and *Kpn*I (position 4198) sites. The sequences are identical except that we read as CCG the nucleotides beginning at

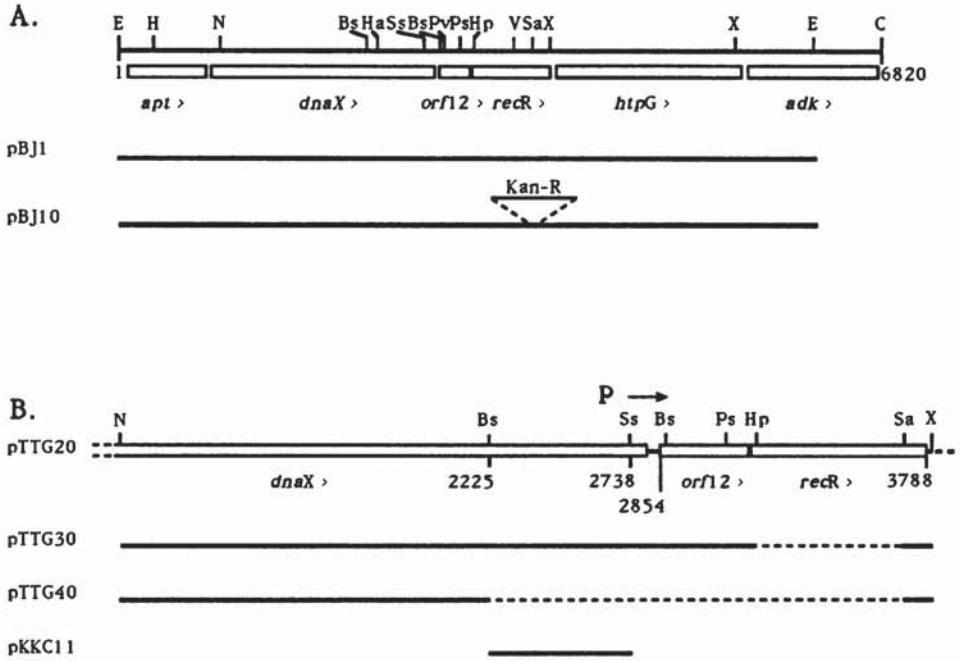

FIG. 1. Structure of the *apt-dnaX-orf*12-*recR-htpG-adk* region. (A) A 6,820-bp fragment with reading frames indicated by open bars and derivative plasmids and phages. The nucleotides are numbered beginning with the first nucleotide of the *Eco*RI site upstream of *apt*. Where differences in sequence were reported in the interval between *htpG* and *adk* (3, 7), the more recent data were used. The overlap between the *orf*12 and *recR* frames is indicated by the darkened area of the bar. E, *Eco*RI; H, *Hin*dIII; N, *Nar*I; Bs, *Bst*EII; Ha, *Hae*II; Ps, *Pst*I; Hp, *Hpa*I; V, *Eco*RV; Sa, *Sal*I; X, *Xmn*I; C, *Cla*I. (B) Plasmids constructed for use in identifying the Orf12 and RecR proteins. These plasmids contain the *dnaX* gene interrupted at the *Nar*I site. Deletions are indicated by dashed lines. The promoter and transcription direction are indicated by the P and the horizontal arrow, respectively.

3274, whereas Mahdi and Lloyd published CGC for that region.

The interval between *dnaX* and *htpG* contains two open reading frames which would be transcribed from left to right (Fig. 1B). Assuming that translation begins at the ATG, the first starts at 2854 and extends for 109 codons to the stop codon at 3183 to 3185, encoding a protein deduced to have a molecular weight of 12,015. This reading frame is designated *orf*12 (21); its 5′ end was first reported by Flower and McHenry (9). The second reading frame, assuming that ATG initiates its translation also, would begin at position 3185 and extend to 3788. This is designated *recR*, based on the observation that disruption of this frame by a kanamycin resistance cassette (below) generated a mutant similar in phenotype to the original *recR* strain of Mahdi and Lloyd (20), the mapping of *recR* to the *dnaX* region of the chromosome (20), and the demonstration (21) that a clone of this region complemented the original *recR* strain. The *recR* product is deduced to be a 201-amino-acid protein of M_r 21,951. The two reading frames overlap by one nucleotide, the A of the *orf*12 terminator serving as the first nucleotide of the *recR* initiator. Both proteins are predicted to be strongly acidic, with isoelectric points of 4.9 and 5.1 for Orf12 and RecR, respectively.

Codon usage indicates that Orf12 should be efficiently expressed, whereas the codon distribution of RecR is more typical of a rarely expressed gene. Strongly expressed genes favor NNC over NNU for phenylalanine, tyrosine, asparagine, and isoleucine, whereas weakly expressed genes favor NNC over NNU for proline, arginine, glycine, and alanine (11). The *orf*12 sequence fits the strongly expressed pattern in five of the eight pairs, two pairs have members used equally, and one pair is typical of weakly expressed genes. On the other hand, the *recR* sequence is more likely to be weakly expressed because four of the pairs are typical of rare proteins, one pair does not discriminate, and three pairs favor the strongly expressed pattern.

Strongly expressed genes also discriminate against codons corresponding to minor isoacceptor tRNAs, whereas weakly expressed genes do not (13, 28). The *orf*12 sequence contained only one rare codon (GGA, Gly) and no rare CUA (Leu), AUA (Ile), UCG (Ser), CCC (Pro), CGA, CGG, AGA, AGG (Arg), or GGG (Gly) codons, consistent with its assignment as a highly expressed gene. RecR, however, depends on translation of three UCG, one CCC, one each CGC and CGG, and five GGG rare codons.

Genes of the *recR* region. Availability of the *orf*12-*recR* sequence allowed completion of a physical map of a 6,820-bp *Eco*RI-*Cla*I fragment which carries the *recR* region and adjacent genes (Fig. 1A). These are *apt* (12), *dnaX* (9, 37), *orf*12-*recR*, *htpG* (3), and *adk* (7). All these genes are transcribed left to right, as presented in Fig. 1. Based on a comparison of restriction sites, this segment of the chromosome is located, in clockwise orientation, between kb 502 and 509 on the Kohara physical map (17) of the whole chromosome.

Insertion mutagenesis of the chromosomal *recR* gene. Site-directed insertion mutagenesis (15, 35) was used to disrupt the chromosomal *recR* allele. The 1.3-kb kanamycin resistance fragment from pUC4K (31, 33) was introduced into the *Sal*I site within the *recR* reading frame on pBJ1, producing pBJ10 (Fig. 1A). pBJ10 was linearized by *Pst*I and used to transform the *recR*$^+$ *recB recC sbcB sbcC* strain JC7623 to kanamycin resistance. One of the recombinants with a *recR*::Kan^r allele (strain JCB325) was chosen for further study. Southern analysis confirmed the *recR* disruption (data not shown).

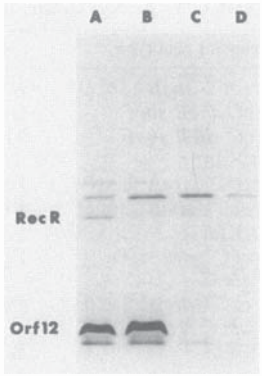

FIG. 2. Products of *orf*12 and *recR* synthesized in maxicells. A 20-μl amount of extract from maxicells containing the indicated plasmids was loaded in each lane. Lanes: A, pTTG20 (*orf*12$^+$ *recR*$^+$); B, pTTG30 (*orf*12$^+$ *recR*); C, pTTG40 (*orf*12 *recR*); D, pUC18 vector.

Strain JCB325 (*recR*::Kan^r *recBC sbcBC*) grew slowly at all temperatures and could not form colonies at 17°C. This cold sensitivity resulted from the combination of *recR*::Kan^r with *recBC sbcBC* rather than from the *recR* mutation directly. The *recR*::Kan^r allele, moved from the *recBC sbcBC* background to generate strain TY108, caused UV sensitivity and plasmid DNA recombination deficiency. Measured by the method of Kolodner et al. (18), plasmid recombination frequency decreased 50-fold in strain TY108 compared with that in the wild-type strain AB1157 (data not shown). This confirms the identity of the reading frame containing the *Sal*I site, rather than *orf*12 (Fig. 1B), as the *recR* gene, because these are the properties associated with *recR* mutants (20).

Protein products of *orf*12 and *recR*. The translation products of *orf*12 and *recR* were labeled by [^{35}S]methionine in vivo in maxicells and in vitro in a coupled transcription-translation system (22, 25) by using plasmids which carried both *orf*12 *recR*, only *orf*12, or neither *orf*12 nor *recR*. Plasmid pTTG20, the insert of which contains *orf*12, *recR*, and their promoter (Fig. 1B) (see below), was deleted for most of *recR* (i.e., between the *Hpa*I and *Sal*I sites) to form pTTG30 and for both *orf*12 and *recR* (i.e., between the *Bst*EII site at position 2225 and the *Sal*I site) to form pTTG40. In vivo, both the 12-kDa Orf12 protein and the RecR product of 26 kDa apparent size were encoded by pTTG20 (Fig. 2). Identification of the apparent 26-kDa protein as the *recR* product was based on the fact that it was not produced from the deleted pTTG30 or pTTG40. The RecR protein had a molecular weight of about 22,000 determined by sequence analysis; the reason for the discrepancy is unknown. In vitro, the *orf*12 product was detectable as a 12-kDa protein encoded by both pTTG20 and pTTG30 (data not shown). Synthesis in vitro of the RecR protein was not observed. Mahdi and Lloyd (21) also identified the *orf*12 and *recR* products as 12- and 22-kDa proteins.

Translation of *recR* was much less efficient than that of *orf*12. Its level was about 5 to 10% of the level of Orf12 (the different content of methionines, nine in *orf*12 and seven in *recR*, did not contribute significantly to the observed difference in intensity of the autoradiogram). The failure to observe RecR synthesis in vitro could be related to inefficient translational coupling in extracts.

***orf*12-*recR* promoter.** Analysis of the nucleotide sequence upstream of *orf*12 identified a potential promoter, TTGAAC-N_{17}-GATAAT, within the 3′ end of the *dnaX* reading frame

(9). To determine whether this region could act as a promoter, the 514-bp *Bst*EII-*Ssp*I fragment (positions 2225 to 2738) (Fig. 1B) containing the predicted promoter was cloned into the promoter-cloning vector pKK232-8 so that it would direct transcription of CAT (6). This plasmid, pKKC11 (Fig. 1B), directed the synthesis of CAT and conferred chloramphenicol resistance on the host strain HB101, whereas the vector pKK232-8 did neither. This confirmed the identity of the promoter, which directed the synthesis of 170 CAT units per unit of plasmid-directed β-lactamase activity. This was about 0.3% of the activity of the induced *tac* promoter. When the *tac* promoter was cloned into the same vector, generating pTC1, and induced with isopropyl-β-D-thiogalactoside, it directed the synthesis of 47,000 CAT units per unit of β-lactamase.

Located 6 bp upstream of the *orf*12 was the sequence AGAGAG, which should serve as a ribosome-binding site (30), as predicted by Flower and McHenry (9).

On the other hand, the *recR* reading frame appears not to have a separate promoter or ribosome-binding site, as judged by sequence analysis. Its translation is probably coupled to that of *orf*12, as proposed also by Mahdi and Lloyd (21), although the presence of a weak promoter specific for *recR* cannot be excluded.

***orf*12-*recR* transcript.** S1 nuclease protection experiments were used to localize, approximately, the 5′ end of the messenger. RNA was extracted from cultures of strains JM103, JM103 carrying the vector pUC18, and JM103 carrying the *orf*12$^+$ *recR*$^+$ plasmid pTTG20 (Fig. 1B). A probe complementary to the messenger was prepared by purifying the *orf*12-*recR* promoter-containing 650-bp *Bst*EII fragment (Fig. 1A). This fragment was 5′-end labeled, denatured, and hybridized to cellular RNA extracts. S1 nuclease-protected fragments were identified by electrophoresis and autoradiography (Fig. 3). The principal protected fragment from the strain carrying *orf*12 and *recR* on a multicopy plasmid was approximately 225 nucleotides in length (Fig. 3, lane E), which confirms the *recR* promoter location predicted from the sequence and indicates that transcription starts at about position 2655, about 10 nucleotides downstream of the −10 sequence (Fig. 4). The same conclusion was reached when the probe was the 526-nucleotide 3′-*Hae*II-*Bst*EII-5′ fragment (Fig. 1A) labeled at the 5′ end (data not shown), which also confirmed that the 650-bp *Bst*EII probe, which was labeled on both ends, was protected by *orf*12-*recR* messenger and not an antisense RNA. Similarly sized *orf*12-*recR* transcripts were present in cells carrying only a haploid copy of the *recR* region (Fig. 3, lanes C and D) but at very low levels. These transcripts were clearly visible when the autoradiogram was overexposed.

Minor transcripts of about 650 nucleotides, which protected the full-length probe, were detected in all extracts and could represent *dnaX* transcripts which extend into the *orf*12-*recR* region or reannealed probe. Minor amounts of fragments of approximately 195, 185, and 145 nucleotides were detected; their identities are unclear.

Transcription start point. The S1 nuclease protection experiment indicated that transcription initiated about 10 nucleotides downstream from the center of the −10 sequence (Fig. 4). To locate the exact start point, primer extension analysis was used. The same RNA preparation used for S1 mapping was further purified by sedimentation through a 5.7 M CsCl solution (8) and used as a substrate for reverse transcription after priming with a 5′-end-labeled synthetic oligonucleotide complementary to nucleotides 2740 to 2756 (Fig. 4). The product was denatured and its size was measured by electrophoresis and autoradiography. The

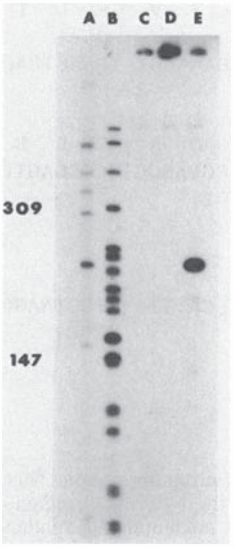

FIG. 3. Identification of the *orf*12-*recR* transcript. RNA extracted from strain JM103 carrying no plasmid (lane C), pUC18 (lane D), or *orf*12$^+$-*recR*$^+$ plasmid pTTG20 (lane E) was hybridized to the 5′-end-labeled 650-bp *Bst*EII fragment (Fig. 1A). S1 nuclease-protected fragments were visualized by electrophoresis and autoradiography. Lanes A and B are 5′-end-labeled *Hin*fI- and *Msp*I-digested pBR322 DNA, respectively. The 147 and 309 refer to the lengths (in nucleotides) of two of the standard fragments.

standards were provided by chain-terminating sequencing reaction products provided by the same primer hybridized to TTG2 DNA, an M13mp19 derivative containing the 2,303-bp *Nar*I-*Hpa*I fragment (Materials and Methods).

The longest RNA detected (number 1 in lane A, Fig. 5) initiated with the G at nucleotide 2655 (Fig. 4). Based on the similarity of the sequence over nucleotides 2620 to 2648 to the consensus *E. coli* promoter, the S1 mapping, and the primer extension, it is proposed that *orf*12-*recR* transcription begins at nucleotide 2655 and that the messenger has a long, untranslated region of 199 nucleotides. Minor amounts of transcripts beginning at positions 2656 and 2657 were also observed.

A transcript with the 5′ end apparently at nucleotide 2675 (number 2 in lane A, Fig. 5) was more abundant than that beginning at position 2655. It is possible that two promoters initiate transcription at positions 2655 and 2675 or that the majority of the transcript initiated at 2655 is processed between 2674 and 2675. However, we favor the interpretation that the reverse transcriptase terminated primer extension synthesis prematurely at position 2675 because the S1 nuclease protection analysis did not detect a major transcript beginning at 2675 and because there is no sequence analysis evidence to indicate the presence of a second promoter.

The significance, if any, of the transcript with a 5′ end at position 2700 (number 3, lane A, Fig. 5) is unknown.

DISCUSSION

The *E. coli orf*12 and *recR* reading frames encode proteins of 12 and 22 kDa, respectively, assuming that they initiate with the ATG codons at positions 2854 and 3183, respectively. These products have been identified after synthesis in vivo in maxicells as proteins which migrate on sodium dodecyl sulfate-polyacrylamide gels at 12 and 26 kDa. We conclude that these reading frames are transcribed from one promoter, located 206 bp upstream of *orf*12, which initiates

```
M  L  K  G   S  T  V   E  L  T   I  V  E  D   D  N  P   A  V  R   T  P  L  E   W  R  Q   A  I  Y
DnaX>              - -35--                    --10--      →
ATGTTAAAAG GTTCAACGGT TGAACTGACT ATCGTTGAAG ATGATAATCC CGCGGTGCGT ACGCCGCTGG AGTGGCGTCA GGCGATATAC
^2601                                                  ^2651

E  E  K  L   A  Q  A   R  E  S   I  I  A  D   N  N  I   Q  T  L   R  R  F  F   D  A  E   L  D  E
GAAGAAAAAC TTGCGCAGGC GCGCGAGTCC ATTATTGCGG ATAATAATAT TCAGACCCTG CGTCGGTTCT TCGATGCGGA GCTGGATGAA
          ^2701                                                ^2751

                                                                          M  F  G   K  G  G
E  S  I  R   P  I  *                                          --SD--      Orf12>
GAAAGTATCC GCCCCATTTG ATCGTAAGCA CAGCTTACGT TCGTCATCCT TAACGTGATT GAGAGAGAAA CCTATGTTTG GTAAAGGCGG
                     ^2801                                                ^2851

 L  G  N  >
TCTGGGTAACC
    ^BstEII
```

FIG. 4. *orf*12-*recR* promoter region. This sequence was taken from Yin et al. (37) and Flower and McHenry (9) and shows only the distal portion of *dnaX* and the N-terminal region of *orf*12. The −35 and −10 represent promoter regions; SD is the ribosome-binding site. The horizontal arrow over nucleotide 2655 indicates the start point for *orf*12-*recR* transcription. The asterisk indicates a stop codon.

transcription starting at nucleotide 2655. The function of the long (199-nucleotide) untranslated leader is, at this time, unknown. The apparent absence of a separate promoter for *recR* and the one-nucleotide overlap between the *orf*12 termination codon and the *recR* initiation codon suggest that RecR translation is coupled to that of Orf12. Of the two models proposed for translational coupling (i.e., translation of the downstream reading frame is facilitated by translating ribosomes opening the secondary structure of the messenger and the "hand-over" model for coupled genes separated by short intercistronic regions [24]), the hand-over model seems more nearly appropriate for Orf12 and RecR. Whatever the mechanism, RecR translation is less efficient by a factor of about 10 to 20 than that of Orf12, at least when labeled in maxicells.

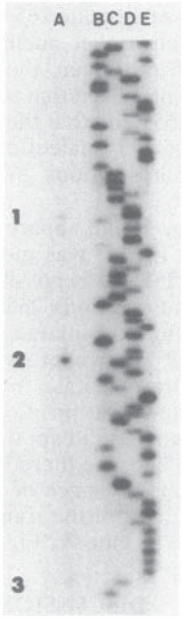

FIG. 5. *orf*12-*recR* transcription start point. RNA extracted for the S1 protection analysis was reverse transcribed with 5′-end-labeled synthetic oligonucleotide complementary to nucleotides 2740 to 2756. The products (lane A) were denatured, electrophoresed, and located by autoradiography. Size markers were provided by chain-terminating sequencing reactions with the same labeled primer and TTG2 (Fig. 1) as the template (lanes B to E, sequencing reactions A, G, C, and T, respectively). Numbers 1, 2, and 3 are explained in the text.

The specific function of the RecR protein is unknown. It is required, however, for plasmid recombination and UV repair. This suggests that it participates in a repair process which depends on *recBC*-independent recombination (21).

It is interesting that *orf*12 and *recR* are located within a group of genes (Fig. 1), most (and possibly all) of which are involved in nucleic acid metabolism. *apt* encodes adenine phosphoribosyltransferase, which catalyzes the synthesis of AMP from adenine and phosphoribosylpyrophosphate (12). *dnaX* encodes the τ and γ subunits of DNA polymerase III (9, 37). *htpG* encodes a heat shock protein which is dispensable but which is required for normal growth rate (3, 4); some heat shock proteins are involved in nucleic acid metabolism (1, 32). Adenylate kinase, an enzyme essential for growth, is the product of *adk* (7). All are transcribed in the same direction, there are short intervals between the open reading frames, and there are overlapping termination (or processing) sites and promoters. Each gene except *recR* seems to have a unique promoter, but in the case of *orf*12 and *recR*, this promoter is located within the *dnaX* reading frame. This location raises the question of whether *dnaX* transcripts also extend through *orf*12-*recR*. Other interesting questions include the function of the untranslated *orf*12-*recR* leader and the significance, if any, of the gene organization in this region.

ACKNOWLEDGMENTS

We thank R. Kolodner for pRDK41, M. Bröker for pMB211, B. J. Bachmann, B. A. Hardesty, C. F. Earhart, W. R. Folk, and J. Messing for bacterial strains, and Peggy Centilli for synthesis of the oligonucleotide.

This work was supported by American Cancer Society grant MV429, in part by Public Health Service grants GM34471 and GM27870 from the National Institutes of Health, and by Welch Foundation grant F949.

ADDENDUM IN PROOF

Alonso, Shirahige, and Ogasawara (personal communication) have found that *Bacillus subtilis* contains two adjacent open reading frames, *orf*107 and *recM*, which are homolo-

gous to *orf*12 and *recR* of *E. coli*. *orf*107 encodes a 107-amino-acid protein that is 39.8% identical to *orf*12 protein; *recM* encodes a 217-amino-acid protein that is 43.2% identical to RecR. Moreover, *orf*107 and *recM* are adjacent to, and transcribed in the same direction as, *dnaX*. There are, however, major differences between these regions in the two organisms. First, the *B. subtilis dnaX* gene has 562 codons, compared with 643 in *E. coli*. Second, the *B. subtilis dnaX-orf*107-*recM* genes seem to form one operon.

LITERATURE CITED

1. **Ang, D., G. N. Chandrasekhar, M. Zylicz, and C. Georgopoulos.** 1986. *Escherichia coli grpE* gene codes for heat shock protein B25.3, essential for both λ DNA replication at all temperatures and host growth at high temperature. J. Bacteriol. **167**:25–29.
2. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. A. Smith, J. G. Seidman, and K. Struhl.** 1987. Current protocols in molecular biology. John Wiley & Sons, Inc., New York.
3. **Bardwell, J. C. A., and E. A. Craig.** 1987. Eukaryotic M_r 83,000 heat shock protein has a homologue in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **84**:5177–5181.
4. **Bardwell, J. C. A., and E. A. Craig.** 1988. Ancient heat shock gene is dispensable. J. Bacteriol. **170**:2977–2983.
5. **Bolivar, F., and K. Backman.** 1979. Plasmids of *Escherichia coli* as cloning vectors. Methods Enzymol. **68**:245–267.
6. **Brosius, J.** 1984. Plasmid vectors for the selection of promoters. Gene **27**:151–160.
7. **Brune, M., R. Schumann, and F. Wittinghofer.** 1985. Cloning and sequencing of the adenylate kinase gene (*adk*) of *Escherichia coli*. Nucleic Acids Res. **13**:7139–7151.
8. **Fisher, S. H., and L. V. Wray, Jr.** 1989. Regulation of glutamine synthetase in *Streptomyces coelicolor*. J. Bacteriol. **171**: 2378–2383.
9. **Flower, A. M., and C. S. McHenry.** 1986. The adjacent *dnaZ* and *dnaX* genes of *Escherichia coli* are contained within one continuous open reading frame. Nucleic Acids Res. **14**:8091–8101.
10. **Granthan, R., C. Gautier, M. Gouy, M. Jacobzone, and R. Mercier.** 1981. Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res. **9**:r43–r74.
11. **Grosjean, H., and W. Fiers.** 1982. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene **18**:199–209.
12. **Hershey, H. V., and M. W. Taylor.** 1986. Nucleotide sequence and deduced amino acid sequence of *Escherichia coli* adenine phosphoribosyltransferase and comparison with other analogous enzymes. Gene **43**:287–293.
13. **Ikemura, T.** 1981. Correlation between the abundance of *E. coli* transfer RNAs and the occurrence of the respective codons in its protein genes. J. Mol. Biol. **146**:1–21.
14. **Inouye, S., A. Nakazawa, and T. Nakazawa.** 1987. Expression of the regulatory gene *xylS* on the TOL plasmid is positively controlled by the *xylR* gene product. Proc. Natl. Acad. Sci. USA **84**:5182–5186.
15. **Jasin, M., and P. Schimmel.** 1984. Deletion of an essential gene in *Escherichia coli* by site-specific recombination with linear DNA fragments. J. Bacteriol. **159**:783–786.
16. **Kodaira, M., S. B. Biswas, and A. Kornberg.** 1983. The *dnaX* gene encodes the DNA polymerase III holoenzyme τ subunit, precursor of the γ subunit, the *dnaZ* gene product. Mol. Gen. Genet. **192**:80–86.
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. **Kolodner, R., R. A. Fishel, and M. Howard.** 1985. Genetic recombination of bacterial plasmid DNA: effect of RecF pathway mutations on plasmid recombination in *Escherichia coli*. J. Bacteriol. **163**:1060–1066.
19. **Lupski, J. R., A. A. Ruiz, and G. N. Godson.** 1984. Promotion, termination, and antitermination in the *rpsU-dnaG-rpoD* macromolecular synthesis operon of *E. coli* K-12. Mol. Gen. Genet. **195**:391–401.
20. **Mahdi, A. A., and R. G. Lloyd.** 1989. Identification of the *recR* locus of *Escherichia coli* K-12 and analysis of its role in recombination and DNA repair. Mol. Gen. Genet. **216**:503–510.
21. **Mahdi, A. A., and R. G. Lloyd.** 1989. The *recR* locus of *Escherichia coli* K-12: molecular cloning, DNA sequencing and identification of the gene product. Nucleic Acids Res. **17**: 6781–6794.
22. **Mullin, D. A., C. L. Woldringh, J. M. Henson, and J. R. Walker.** 1983. Cloning of the *Escherichia coli dnaZX* region and identification of its products. Mol. Gen. Genet. **192**:73–79.
23. **Ohta, N., L.-S. Chen, E. Swanson, and A. Newton.** 1985. Transcriptional regulation of a periodically controlled flagellar gene operon in *Caulobacter crescentus*. J. Mol. Biol. **186**: 107–115.
24. **Oppenheim, D. S., and C. Yanofsky.** 1980. Translational coupling during expression of the tryptophan operon of *Escherichia coli*. Genetics **95**:785–795.
25. **Pratt, I. M., G. I. Boulnois, V. Darby, E. Orr, E. Wahle, and I. B. Holland.** 1981. Identification of gene products programmed by restriction endonuclease DNA fragments using an *Escherichia coli in vitro* system. Nucleic Acids Res. **9**:4459–4474.
26. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
27. **Seed, B., and J. Y. Sheen.** 1988. A simple phase-extraction assay for chloramphenicol acetyltransferase activity. Gene **67**:271–277.
28. **Sharp, P. M., and W.-H. Li.** 1986. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for "rare" codon. Nucleic Acids Res. **14**:7737–7749.
29. **Stoker, N. G., J. M. Pratt, and I. B. Holland.** 1984. In vivo gene expression systems in prokaryotes, p. 171–177. *In* B. D. Hames and S. J. Higgins (ed.), Transcription and translation. IRL Press, Oxford.
30. **Stormo, G. D., T. D. Schneider, and L. Gold.** 1982. Characterization of translational initiation sites in *E. coli*. Nucleic Acids Res. **10**:2971–2996.
31. **Taylor, L. A., and R. E. Rose.** 1988. A correction in the nucleotide sequence of the Tn*903* kanamycin resistance determinant in pUC4K. Nucleic Acids Res. **16**:358.
32. **Tilly, K., N. McKittrick, M. Zylicz, and L. Georgopoulos.** 1983. The *dnaK* protein modulates the heat-shock response of *Escherichia coli*. Cell **34**:641–646.
33. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
34. **Willets, N. S., A. J. Clark, and B. Low.** 1969. Genetic location of certain mutations conferring recombination deficiency in *Escherichia coli*. J. Bacteriol. **97**:244–249.
35. **Winans, S. C., S. J. Elledge, J. H. Kruger, and G. C. Walker.** 1985. Site-directed insertion and deletion mutagenesis with cloned fragments in *Escherichia coli*. J. Bacteriol. **161**: 1219–1221.
36. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.
37. **Yin, K.-C., A. Blinkowa, and J. R. Walker.** 1986. Nucleotide sequence of the *Escherichia coli* replication gene *dnaZX*. Nucleic Acids Res. **14**:6541–6549.