JOURNAL OF BACTERIOLOGY, Jan. 1991, p. 353–364
0021-9193/91/010353-12$02.00/0

Vol. 173, No. 1

Nucleotide Sequence of the *Escherichia coli recJ* Chromosomal Region and Construction of RecJ-Overexpression Plasmids

SUSAN T. LOVETT† AND RICHARD D. KOLODNER*

Division of Cellular and Molecular Biology, Dana-Farber Cancer Institute, 44 Binney Street, and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115*

Received 25 June 1990/Accepted 13 October 1990

The nucleotide sequence of the *recJ* gene of *Escherichia coli* K-12 and two upstream coding regions was determined. Three regions were identified within these two upstream genes that exhibited weak to moderate promoter activity in fusions to the *galK* gene and are candidates for the *recJ* promoter. *recJ* appeared to be poorly translated: the *recJ* nucleotide sequence revealed a suboptimal initiation codon GUG, no discernible ribosome-binding consensus sequence, and relatively nonbiased synonymous codon usage. Comparison of the sequence of this region of the chromosome with DNA data bases identified the gene immediately downstream of *recJ* as *prfB*, which encodes translational release factor 2 and has been mapped near *recJ* at 62 min. No significant homology between *recJ* and other previously sequenced regions of DNA was detected. However, protein sequence comparisons with a gene upstream of *recJ*, denoted *xprB*, revealed significant homology with several site-specific recombination proteins. Its genetic function is presently unknown. Knowledge of the nucleotide sequence of *recJ* allowed the construction of a plasmid from which overexpression of RecJ protein could be induced. Supporting the notion that translation of *recJ* is limiting, a strong T7 bacteriophage promoter upstream of *recJ* did not, by itself, allow high-level expression of RecJ protein. The addition of a ribosome-binding sequence fused to the initiator GTG of *recJ* in this construction was necessary to promote expression of high levels of RecJ protein.

recJ mutations were initially identified by their effect on RecBC-independent conjugational recombination pathways (the RecF and RecE pathways). These two pathways share a strong dependence on *recA*, *recF*, *recO*, and *recR* as well as *recJ* (14, 16, 18, 22, 26, 28, 30, 32, 34, 53). It appears that *recJ* is required for many types of recombinational events, although the stringency of the requirement for *recJ*, as for other recombination genes in *Escherichia coli*, appears to vary with the type of recombinational event monitored and the other recombination gene products which are available. Loss of *recJ* function can be tolerated for some types of recombination events (e.g., UV repair and transductional and conjugational recombination) if functional RecBCD enzyme is available (28, 31). On the other hand, *recJ* mutations cause a severe reduction in recombination between alleles carried on plasmids, even in *recBCD*$^+$ strains (22).

The genetics of recombination in *E. coli* point to a complex interaction of many genes (reviewed in references 33 and 51). However, the biochemical function of only a few of these genes is known. Because *recJ* is not a highly expressed gene (29), the sequence of the *recJ* gene was determined and overproduction of *recJ* was engineered with this information. This work led to the purification of the RecJ protein to homogeneity and its identification as a single-strand DNA-specific exonuclease (30).

MATERIALS AND METHODS

Strains and plasmids. *E. coli* AB1157 [*argE3* Δ(*gpt-proA*)*62 his-4 leuB6 thr-1 thi-1 ara-14 lacY1 galK2 mtl-1 xyl-5 kdgK51 supE44 rpsL31 tsx-1*] (4) was used as the host for expression vectors. Complementation of *recJ* was assayed by transformation of plasmids into strain RDK1656 (*recB21 recC22 sbcA23 recJ284*::Tn*10* derivative of AB1157 [32]). Quantitative matings employed *E. coli* Hfr JC158 (Hfr PO1 *serA6 rel-1 thi-1 lacI22*) (8). pGP1-2 (54) and pT7-5 (a derivative of pT7-1 [54] in which the *bla* gene was inverted relative to the T7 promoter) were provided by S. Tabor (Harvard Medical School, Boston, Mass.) and used in overexpression experiments. M13 cloning vectors M13mp18 and M13mp19 were from laboratory stocks and propagated in strain JM101 (38). pKO1 and pKO500 (vectors for identifying promoter regions), pKL200 (a derivative of pKO1 carrying the *lac* promoter), and a *galK* mutant derivative of *E. coli* C600 were provided by K. McKenney (National Institutes of Health, Bethesda, Md.) (37). Plasmids constructed for this study are described briefly in Table 1 and in more detail in the appropriate sections below.

Recombination and UV tests for complementation. Procedures for cell growth, transformation, quantitative Hfr matings, and UV survival assays have been described previously (31, 32). In conjugational recombination tests, selection was for leucine prototrophy with contraselection for serine prototrophy and streptomycin resistance (at 100 μg/ml) with 56/2 minimal medium (58) after a 1-h mating. Plasmid transformants of RDK1656 were grown in L broth containing 50 μg of ampicillin per ml and 20 μg of tetracycline per ml to ensure maintenance of both the plasmid and the chromosomal *recJ*::Tn*10* mutation.

Enzymes. Restriction endonucleases were purchased from New England BioLabs, Inc. (Beverly, Mass.) and used as specified by the manufacturer. Partial restriction enzyme digestion was performed by the addition of ethidium bromide ranging from 50 to 500 μg/ml in the reaction to achieve optimal production of the desired partial product (44). T4 DNA ligase was purified by an unpublished method of R.

* Corresponding author.
† Present address: Rosenstiel Center, Brandeis University, Waltham, MA 02254.

TABLE 1. Plasmids and oligonucleotides

Name	Description[a]	Source or reference
Starting plasmids		
pJC763	*Hin*dIII-*Sal*I *recJ* fragment (nucleotides 1–3772) inserted in pBR322	29
pJC765	*Eco*RI-*Sal*I *recJ* fragment (nucleotides 1697–3772) inserted in pBR322	29
Promoter assay plasmids		
pKO1	Promoterless *galK* vector	37
pKO500	Promoterless *galK* vector	K. McKenney
pKL200	p*lac galK*	K. McKenney
pRDK117	*Eco*RI-*Nru*I pJC765 fragment (nucleotides 1697–1949) + *Eco*RI-*Sma*I fragment of pKO1	This study
pRDK168	*Hin*dIII-*Eco*RI pJC763 fragment (nucleotides 1–1702) + *Hin*dIII-*Eco*RI fragment of pKO500	This study
pRDK169	*Nru*I fragment of pJC763 (nucleotides 199–369) + *Sma*I fragment of pKO500	This study
pRDK170	*Nru*I deletion of pRDK168 (deletion of nucleotides 1–369)	This study
pRDK171	*Sac*II deletion of pRDK168 (deletion of nucleotides 232–992)	This study
pRDK172	*Sac*II inversion of pRDK168 (inversion of nucleotides 232–992)	This study
pRDK173	*Nru*I-*Eco*RI fragment of pRDK168 (nucleotides 370–1702) + *Sma*I-*Eco*RI fragment of pKO500	This study
recJ overexpression plasmids		
pT7-5	T7 promoter vector, derived from pBR322, Ap^r	S. Tabor
pGP1-2	T7 RNA polymerase gene under p_L, cI857 control, derived from pACYC177, Km^r	54
pRDK110	*Eco*RI-*Sal*I pJC765 fragment (nucleotides 1697–3772) + *Eco*RI-*Sal*I fragment of pT7-5	This study
pRDK112	Oligonucleotides 1 and 1′ + *Dde*I-*Eco*RI fragment of pRDK110 (retains nucleotides 1903–3772, oligonucleotide reconstructs coding sequence to nucleotide 1884 and adds ribosome-binding sequence)	This study
pRDK113	*Taq*I deletion of pJC763 (retains nucleotides 1778–3772)	This study
pRDK114	Oligonucleotides 2 and 2′ + *Eco*RI-*Cla*I fragment of pRDK113 (retains nucleotides 1778–3772, oligonucleotide reconstructs sequence back to nucleotide 1744 and adds ribosome-binding sequence)	This study
pRDK115	*Eco*RI-*Sal*I fragment of pRDK114 + *Eco*RI-*Sal*I fragment of pT7-5	This study
Oligonucleotides		
1	AATTCAGATCTAGAAAGGAGGTACAAGCATGCTTCCCTGGCAACAAC	
1′	TCAGTTGTTGCCAGGGAAGCATGCTTGACCTCCTTTCTAGATCGT	
2	AATTCAGATCTAGAAAGGAGGTACAAGCGTGAAACAACAGATACACCTT CGTCGCCGTGAAGT	
2′	CGACTTCACGGCGACGAAGGTGTATCTGTTGTTTCACGCTTGTACCTCCTT TCTAGATCTG	

[a] The numbers given in parentheses refer to the nucleotides of *recJ* sequence as illustrated in Fig. 1. Ap^r, Ampicillin resistance; Km^r, kanamycin resistance.

Kolodner. T4 polynucleotide kinase was purified as described previously (43). Reaction conditions for recombinant plasmid construction were as described previously (11, 19, 39).

DNA purification and oligonucleotides. Plasmid DNA was purified by the alkaline-sodium dodecyl sulfate (SDS) method (6). This was followed by two cycles of equilibrium centrifugation in CsCl-ethidium bromide density gradients. Restriction fragments of DNA were purified by electrophoresis through agarose gels run in Tris-acetate–EDTA buffer and then by glass bead extraction (55). Oligonucleotides were synthesized by the Dana-Farber Cancer Institute Molecular Biology Facility with an Applied Biosystems (Foster City, Calif.) 380A DNA synthesizer using β-cyanoethyl phosphoramidites. Those oligonucleotides used in the construction of overexpression plasmids were further purified by gel electrophoresis and chromatography through Sep-Pac C18 columns (Waters Associates, Milford, Mass.) as previously described (27). Purified complementary oligonucleotides were annealed by incubation for 5 min at 65°C in TE buffer (10 mM Tris hydrochloride, 1 mM EDTA) containing 100 mM NaCl and cooled slowly before use in ligation reactions.

DNA sequence analysis. The DNA sequence of M13 clones was determined by the dideoxynucleotide-chain termination sequencing method with [^{35}S]dATP (Amersham Corp., Arlington Heights, Ill.) and sequencing kits obtained from either New England BioLabs or U.S. Biochemicals (Cleveland, Ohio). The sequencing strategy employed either different restriction fragments cloned into M13mp18 or M13mp19 or nested deletion derivatives of selected fragments constructed by the method of Dale et al. (10) with a kit from IBI (New Haven, Conn.). In some cases, specific oligonucleotide primers were synthesized and used to sequence selected regions of DNA. Computer analysis of DNA sequence data was performed by the DNA Inspector II program (Textco, West Lebanon, N.H.) or the Intelligenetics package available through Bionet. Homology searches to PIR/NBRL release 17.0, SWISS-PROT release 8.0, and GenBank release 56 data bases were performed by using the FASTA program (45).

Promoter assays. Various fragments were inserted into promoter assay vector pKO1 or pKO500 with T4 DNA ligase and transformed into *E. coli* C600 *galK* followed by selection for Ap^r on MacConkey-galactose plates. These plasmids and their derivations are described in Table 1 and Fig. 2. Assays of galactokinase activity in toluenized cells were performed as described previously (37) except that cells were grown in 56/2 minimal medium containing 0.2% fructose, 0.5% Casamino Acids, 1 μg of thiamine per ml, and 50 μg of

ampicillin per ml. Values for promoter-positive plasmids and pKO1 and pKO500 vectors are the averages of two or more determinations.

Construction of RecJ-overexpressing plasmids. The plasmids used to analyze the expression of RecJ protein and the oligonucleotide linkers used in their construction are described in Table 1, and restriction maps of these plasmids are presented in Fig. 5. The derivations of these plasmids were as follows. pRDK110 was constructed by the ligation of the 2.07-kb *Eco*RI-to-*Sal*I fragment of pJC765 (29) between the same sites of pT7-5. pRDK112 was constructed by first partially digesting pRDK110 with *Dde*I, followed by purification of the full-length linear fragments by preparative agarose gel electrophoresis. This DNA was further digested to completion with *Eco*RI and cyclized with DNA ligase in the presence of annealed oligonucleotides 1 and 1′. pRDK115 was constructed in several steps. First, pJC763 was partially digested with *Taq*I. Those fragments in the size range of 5.4 to 5.8 kb were purified by electrophoresis through an agarose gel, cleaved with *Cla*I, and cyclized with DNA ligase. The resulting plasmid, pRDK113, was analyzed for the presence of a reconstructed *Cla*I site and appropriate *Nru*I, *Sal*I, and *Cla*I sites. pRDK113 was digested with *Eco*RI and *Cla*I and cyclized in the presence of annealed oligonucleotide linkers 2 and 2′, producing plasmid pRDK114. pRDK114 was digested with *Eco*RI and *Sal*I, and the 2-kb *recJ*-containing fragment was inserted into pT7-5, giving pRDK115.

Induction and labeling of proteins expressed from pT7-5-derived plasmids. The strains used for induction experiments contained plasmid pGP1-2 (carrying Km^r and the T7 RNA polymerase gene under control of heat-sensitive lambda *c*I857 repressor [54]) in addition to one of the Ap^r plasmids pT7-5, pRDK110, pRDK112, and pRDK115. Strains were grown in 2YT medium (38) containing 50 mM potassium phosphate (pH 7.0), 0.2% glycerol, 30 μg of kanamycin per ml, and 100 μg of ampicillin per ml at 30°C. After reaching an optical density of 0.5 at 590 nm, the cells were collected by centrifugation, washed in 56/2 buffer, and resuspended in 56/2 minimal medium containing 0.2% glucose, 0.001% thiamine, and 50 μg each of the amino acids arginine, histidine, proline, threonine, and leucine per ml. The growth temperature was then shifted to 42°C for 30 min. Rifampin (Sigma Chemical Co., St. Louis, Mo.), which inhibits *E. coli* but not the T7 RNA polymerase, was then added to 200 μg/ml, and the cell suspensions were incubated for an additional 15 min at 42°C. After additional incubation at 30°C for 30 min, 10 μCi of [^{35}S]methionine (Amersham) per ml (specific activity, 800 Ci/mmol) was added and incubation was continued for 5 min. The cells were collected by centrifugation and resuspended in sample buffer, and the proteins present were analyzed by SDS-polyacrylamide gel electrophoresis (PAGE) (36). The gels were stained with Coomassie blue R250, destained, dried, and exposed to Kodak XAR5 film.

Viability measurements of cultures that had been similarly induced were performed by determining the number of CFU before and after a 45-min incubation at 42°C. Induction was performed as described above but without the addition of rifampin. Viable cell counts were determined by dilution in 56/2 buffer and plating on L-kanamycin-ampicillin medium with subsequent incubation at 30°C.

Nucleotide sequence accession number. The DNA sequence reported in this communication has been submitted to GenBank and is listed under accession number M54884.

RESULTS

Sequence of *recJ*. The nucleotide sequence of a 3,772-bp *Hin*dIII-to-*Sal*I fragment obtained from pJC765 was determined and is shown in Fig. 1. A restriction map of the *recJ* region is shown in Fig. 2. The *recJ* region contained an open reading frame from nucleotides 1735 to 3475, potentially encoding a protein of 63,000 molecular weight. There was no discernible consensus ribosome-binding site (Shine-Dalgarno sequence [50]) upstream of the potential initiation codons (GTG at nucleotide 1744 or ATG at nucleotide 1885) in this reading frame. However, complementation and protein labeling experiments using *recJ* expression plasmids (described below) indicated that translation initiates at the first GTG at nucleotide 1744 rather than at the first ATG at position 1885. In addition, as is characteristic of poorly expressed genes, the *recJ* reading frame exhibited a relatively low bias of synonymous codon usage (23), having a "codon adaptation index" (49) value of 0.314.

The termination codon TAG is found at position 3475. An inverted repeat, reminiscent of a rho-independent transcription terminator (47), is located at position 3496 immediately downstream of the TAG. Contained within this hairpin is a box A sequence, CGCTCTTA. Box A sites have been implicated in the binding of NusA protein, perhaps causing stalling of the transcriptional complex (15). Because NusA promotes both transcriptional termination and antitermination, the consequence of this sequence at this site is unclear.

Using DNA sequence homology searches, the region downstream of *recJ* was identified as the gene *prfB*, encoding translational release factor 2 (9). The region from nucleotide 3267 to the *Sal*I site of our sequence is identical to that reported by Craigen et al. (9). The translational initiation site for *prfB* is at nucleotide 3568, as indicated in Fig. 1. The *prfB* gene has been mapped to 62 min, near *recJ* (20). This allows the placement of the *recJ* gene on the *serA*-proximal side of *prfB*, given the published orientation of *prfB* relative to cotransducible markers *lysA* and *serA* (20).

Sequence upstream of *recJ*. Two open reading frames are present upstream of *recJ*. The first upstream open reading frame in this region occurs from nucleotide positions 1004 to 1736. It potentially encodes a protein of about 26,000 Da if translation is initiated at the first ATG located at nucleotide 1028. A second upstream open reading frame was identified from nucleotide positions 98 to 1001. It potentially encodes as much as a 34,000-Da protein, but because of a lack of a detectable Shine-Dalgarno sequence, the location of the translation initiation site is unclear. The first and second reading frames correspond, respectively, to the previously identified *xprA* and *xprB* genes coding for proteins of unknown function identified by maxicell labeling of *recJ* plasmids (29).

Promoter analysis. To locate potential promoters, various portions of the *recJ* upstream region were fused to *galK* promoter assay plasmids (37). These vectors lack a promoter for *galK*, and the level of galactokinase expression observed is dependent on the strength of the promoters placed upstream. Figure 2 illustrates the extent of *recJ* upstream sequence carried by each construction, its GalK phenotype as determined by plating on MacConkey-galactose medium, and the amount of galactokinase expressed. Fusion of the *Hin*dIII-to-*Eco*RI fragment (nucleotides 1 to 1702) promotes expression of the *galK* gene, indicating that a promoter does lie somewhere in the region upstream of *recJ*. The level of galactokinase promoted by this region is roughly 25% of that promoted by the *E. coli lac* promoter on a similar plasmid

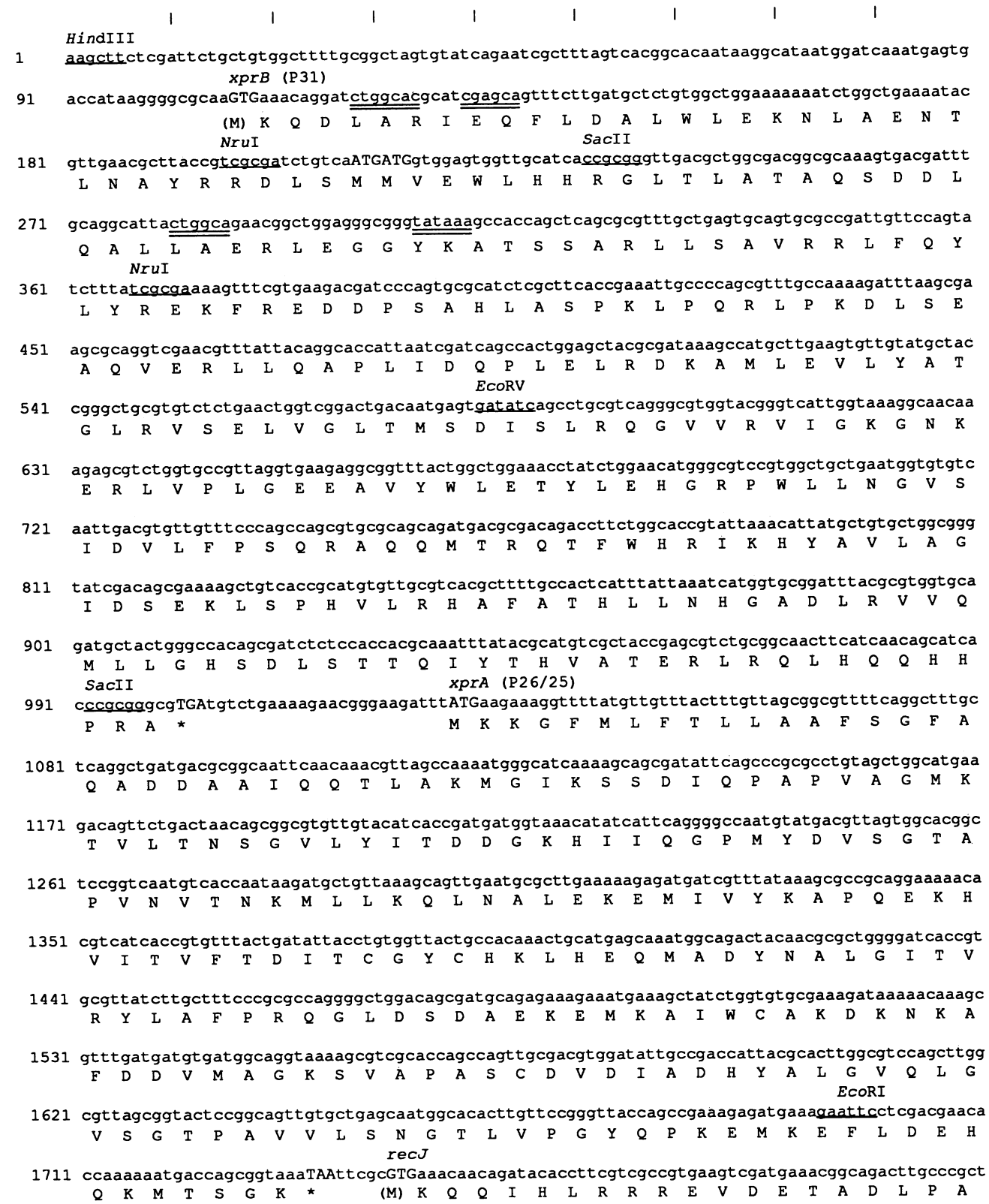


FIG. 1. Nucleotide sequence of the 3.8-kb *Hin*dIII-*Sal*I fragment containing *recJ*. Illustrated below the nucleotide sequence is the predicted amino acid sequence for two reading frames upstream of *recJ* (*xprB* and *xprA*) and *recJ*. Some potential start and stop codons for these reading frames and the known start codon for the downstream *prfB* gene are indicated in capital letters. Two potential promoter sequences beginning at nucleotides 119 and 305 are double underlined. Underlined are pertinent restriction sites and a box A-containing palindrome (a potential terminator sequence).

```
1801 gctgaattgcctcccttgctgcgccgtttatacgccagccggggagtacgcagtgcgcaagaactggaacgcagtgttaaaggtATGctg
     Q  E  L  P  P  L  L  R  R  L  Y  A  S  R  G  V  R  S  A  Q  E  L  E  R  S  V  K  G  M  L
                                                                  NruI
1891 ccctggcagcaactgagcggcgtcgaaaaggccgttgagatcctttacaacgctttTCGCGAaggaacgcggattattgtggtcggtgat
     P  W  Q  Q  L  S  G  V  E  K  A  V  E  I  L  Y  N  A  F  R  E  G  T  R  I  I  V  V  G  D

1981 ttcgacgccgacggcgcgaccagcacggctctaagcgtgctggcgatgcgctcgcttggttgcagcaatatcgactacctggtaccaaac
     F  D  A  D  G  A  T  S  T  A  L  S  V  L  A  M  R  S  L  G  C  S  N  I  D  Y  L  V  P  N

2071 cgtttcgaagacggttacggcttaagcccggaagtggtcgatcaggcccatgcccgtggcgcgcagttaattgtcacggtggataacggt
     R  F  E  D  G  Y  G  L  S  P  E  V  V  D  Q  A  H  A  R  G  A  Q  L  I  V  T  V  D  N  G

2161 atttcctcccatgcgggggttgagcacgctcgctcgttgggcatcccggttattgttaccgatcaccatttgccaggcgacacattaccc
     I  S  S  H  A  G  V  E  H  A  R  S  L  G  I  P  V  I  V  T  D  H  H  L  P  G  D  T  L  P

2251 gcagcggaagcgatcattaaccctaacttgcgcgactgtaatttcccgtcgaaatcactggcaggcgtgggtgtggcgttttatctgatg
     A  A  E  A  I  I  N  P  N  L  R  D  C  N  F  P  S  K  S  L  A  G  V  G  V  A  F  Y  L  M

2341 ctggcgctgcgcacctttttgcgcgatcagggctggtttgatgagcgtaacatcgcaattcctaacctggcagaactgctggatctggtc
     L  A  L  R  T  F  L  R  D  Q  G  W  F  D  E  R  N  I  A  I  P  N  L  A  E  L  L  D  L  V

2431 gcgctggggacagtggcggacgtcgtgccgctggacgctaataatcgcattctgacctggcaggggatgagtcgcatccgagccggaaag
     A  L  G  T  V  A  D  V  V  P  L  D  A  N  N  R  I  L  T  W  Q  G  M  S  R  I  R  A  G  K

2521 tgccgtccggggattaaagcgctgcttgaagtggcaaaccgtgatgcacaaaaactcgccgccagcgatttaggttttgcgctggggcca
     C  R  P  G  I  K  A  L  L  E  V  A  N  R  D  A  Q  K  L  A  A  S  D  L  G  F  A  L  G  P

2611 cgtctcaatgctgccggacgactggacgatatgtccgtcggtgtggcgctgttgttgtgcgacaacatcggcgaagcgcgcgtgctggca
     R  L  N  A  A  G  R  L  D  D  M  S  V  G  V  A  L  L  L  C  D  N  I  G  E  A  R  V  L  A

2701 aatgaactcgatgcgctaaaccagacgcgaaaagagatcgaacaaggaatgcaaattgaagccctgaccctgtgcgagaaactggagcgc
     N  E  L  D  A  L  N  Q  T  R  K  E  I  E  Q  G  M  Q  I  E  A  L  T  L  C  E  K  L  E  R

2791 agccgtgacacgctacccggcgggctggcaatgtatcaccccgaatggcatcagggcgttgtcggtattctggcttcgcgcatcaaagag
     S  R  D  T  L  P  G  G  L  A  M  Y  H  P  E  W  H  Q  G  V  V  G  I  L  A  S  R  I  K  E

2881 cgttttcaccgtccggttatcgcgtttgcgccagcaggtgacggtacgctgaaaggttccggtcgctccattcaggggctgcatatgcgt
     R  F  H  R  P  V  I  A  F  A  P  A  G  D  G  T  L  K  G  S  G  R  S  I  Q  G  L  H  M  R

2971 gatgcgctggagcgattagacacactctaccctggcatgatgctgaagtttggcggtcatgcgatggcggcgggtttgtcgctggaagag
     D  A  L  E  R  L  D  T  L  Y  P  G  M  M  L  K  F  G  G  H  A  M  A  A  G  L  S  L  E  E

3061 gataaattcaaactctttcaacaacggtttggcgaactggttactgagtggctggacccttcgctattgcaaggcgaagtggtatcagac
     D  K  F  K  L  F  Q  Q  R  F  G  E  L  V  T  E  W  L  D  P  S  L  L  Q  G  E  V  V  S  D

3151 ggtccgttaagcccggccgaaatgaccatggaagtggcgcagctgctgcgcgatgctggcccgtgggggcagatgttcccggagccgctg
     G  P  L  S  P  A  E  M  T  M  E  V  A  Q  L  L  R  D  A  G  P  W  G  Q  M  F  P  E  P  L

3241 tttgacggtcatttccgtctgctgcaacagcggctggtgggcgaacgtcatttgaaggtgatggtcgaaccggtcggcggcggtccactg
     F  D  G  H  F  R  L  L  Q  Q  R  L  V  G  E  R  H  L  K  V  M  V  E  P  V  G  G  G  P  L

3331 ctggatggtattgcttttaatgtcgataccgccctctggccggataacggcgtgcgcgaagtgcaactggcttataagctcgatatcaac
     L  D  G  I  A  F  N  V  D  T  A  L  W  P  D  N  G  V  R  E  V  Q  L  A  Y  K  L  D  I  N

3421 gagtttcgcggcaaccgcagcctgcaaattatcatcgacaatatctggccaatttAGcgtcatcttctctataaaaaaagagcgtggattg
     E  F  R  G  N  R  S  L  Q  I  I  I  D  N  I  W  P  I  *
               BoxA                                          prfB
3511 ggtacaatccCGCTCTTAtcaccgcattttgactagctcaataaaagaaatcagaccATGtttgaaattaatccggtaaataatcgcatt

3601 caggacctcacggaacgctccgacgttcttaggggtatctttgactacgacgccaagaaagagcgtctggaagaagtaaacgccgagct

3691 ggaacagccggatgtctggaacgaacccgaacgcgcacaggcgctgggtaaagagcgttcctccctcgaagccgttgtcgac
```

FIG. 1—*Continued.*

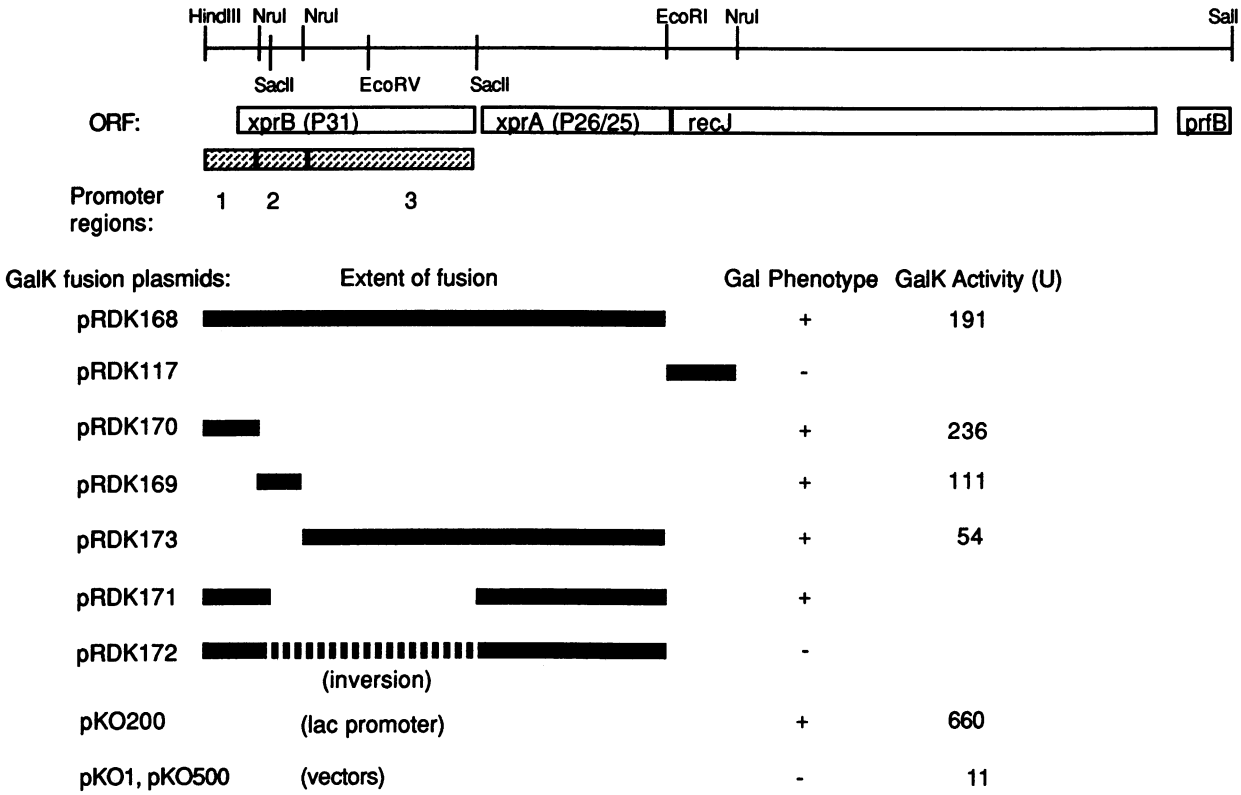


FIG. 2. Structure of *galK* promoter fusion plasmids. Diagrammed is the *recJ* chromosomal region and the extent of DNA fused to various *galK* promoter assay vectors. Also indicated is the GalK phenotype as determined on MacConkey-galactose medium and the amount of galactokinase activity assayed for each construction. Values for promoterless vectors and for a *lac* promoter construction, pKL200, are given for comparison.

(pKL200), which gave approximately 660 U of galactokinase activity.

The region between the *Hin*dIII and *Nru*I sites located at positions 1 and 199, respectively, exhibited approximately the same promoter activity as the larger *Hin*dIII-to-*Eco*RI fragment. This promoter could theoretically direct transcription of all three genes. Although no strong homology to the σ^{70} consensus promoter sequence was found within this region, a potential promoter utilizing the alternate sigma factor σ^{54} (24) was found from positions 119 to 135 (Fig. 1 and 3). Figure 3 shows that this sequence is as homologous to the σ^{54} consensus promoter (24) as is the putative *rcsB* promoter, a gene which is involved in capsular polysaccharide synthesis and known to require *rpoN* (σ^{54}) for expression (52).

A small *Nru*I fragment located at positions 199 to 369 exhibited a lower level of promoter activity. The best candidate for such a promoter is the sequence CTGGCA-N_{18}-TATAAA present at positions 281 to 310. The promoter-scoring algorithm of Mulligan et al. (40) predicts that this sequence would be a poor promoter (score of 39) if active at all. This promoter could direct the transcription of both the *xprA* and *recJ* genes as well as a portion of *xprB*. Such a putative transcript could encode a shortened version of the 31,000-Da XprB protein since translational initiation at the GTG at nucleotide position 343 would allow a 22,000-Da protein to be expressed from *xprB*.

```
RecJ putative promoter ctggcacgcatcgagca
                       :::::::    :  :::
RpoN consensus         ctggcacnnnnncttgca
                       ::::  ::    :  :::
RcsB putative promoter ctggtacccggcaagca
```

FIG. 3. Identification of a potential *rpoN*-dependent promoter. Aligned are a potential *recJ* promoter sequence at nucleotides 119 to 135, the *rpoN*-dependent promoter consensus, and the *rcsB* putative *rpoN*-dependent promoter (52).

A third region which exhibited even poorer promoter activity was located in the region between the *Nru*I site at 370 and the *Sac*II site at 995. The position of this promoter region was deduced as follows. The fragment located between the *Nru*I site at 370 and the *Eco*RI site at 1702 was found to contain promoter activity. The 1.7-kb *Hin*dIII-to-*Eco*RI fragment with an inversion of the *Sac*II fragment located between positions 232 and 992 had no promoter activity (data not shown). These results suggest that there is no promoter between the second *Sac*II site (position 992) and the *Eco*RI site (position 1702) and that the inverted *Sac*II fragment must terminate transcription from promoters upstream of the first *Sac*II site. This indicates that the *Sac*II fragment contains a promoter which must be positioned downstream from the *Nru*I site at 370. This latter weak promoter activity region did not contain any sequence which was strongly homologous to consensus promoter sequences.

No promoter was found between the *Eco*RI site at position 1702 and the *Nru*I site at position 1947 just after the start of

```
XprB     HVLRHATATHLLNHG-ADLRVVQMLLGHSD-LSTTQIYT

FimB     HMLRHSCGFALANMG-IDTRLIQDYLGHRN-IRHTVWYT
FimE     HMLRHACGYELAERG-ADTRLIQDYLGHRN-IRHTVRYT
GeneD    HTFRHSYAMHMLYAG-IPLKVLQSLMGHKS-ISSTEVYT
TnpA     HMLRHTHATQLIREG-WDVAFVQKRLGHAHVQTTLNTYV
TnpB     HAFRHTVGTRMINNG-MPQHIVQKFLGHES-PEMTSRYA
TnpI     HQLRHFFCTNAIEKG-FSIHEVANQAGHSN-IHTTLLYT
Int P2   HALRHSFATHFMING-GSIITLQRILGHTR-IEQTMVYA
 186     HVLRHTFASHFMMNG-GNILVLQRVLGHTD-IKMTMRYA
 P22     HDLRHTWASWLVQAG-VPISVLQEMGGWES-IEMVRRYA
 P1      HSARVGAARDMARAG-VSIPEIMQAGGWTN-VNIVMNYI
 lambda  HELRSLSA-RLYEKQ-ISDKFAQHLLGHKS-DTMASQYR
 phi80   HDMRRTIATNLSELG-CPPHVIEKLLGHQM-VGVMAHYN
 P4      HGFRTMARGALGESGLWSDDAIERQLSHSERNNVRAAYI
Flp      HIGRHLMTSFLSMKGLTELTNVVGNWSDKRASAVRTTYT
```

FIG. 4. Homology of predicted XprB protein sequence and C-terminal regions of integrase-related proteins. Shown are amino acids 244 to 280 of XprB (assuming initiation at the first GUG) aligned with the C-terminal region of integrase-related proteins. This figure is modeled after that of Dorman and Higgins (12) showing C-terminal regions of FimB and FimE (21), gene *D* of F (25), TnpA and TnpB of Tn*554* (41), TnpI of Tn*4430* (35), Int proteins of bacteriophages P2, 186, P22, P1, λ, φ80, and P4, and yeast FLP protein (3). Underlined in the XprB sequence are invariant amino acids H, R, and Y.

recJ translation, which is in agreement with published S1 transcript mapping of this region (20). No consensus LexA-binding sequence was found within the 3,772-base *Hin*dIII-*Eco*RI fragment, suggesting that expression of the *recJ* region is not directly regulated by *lexA* (56).

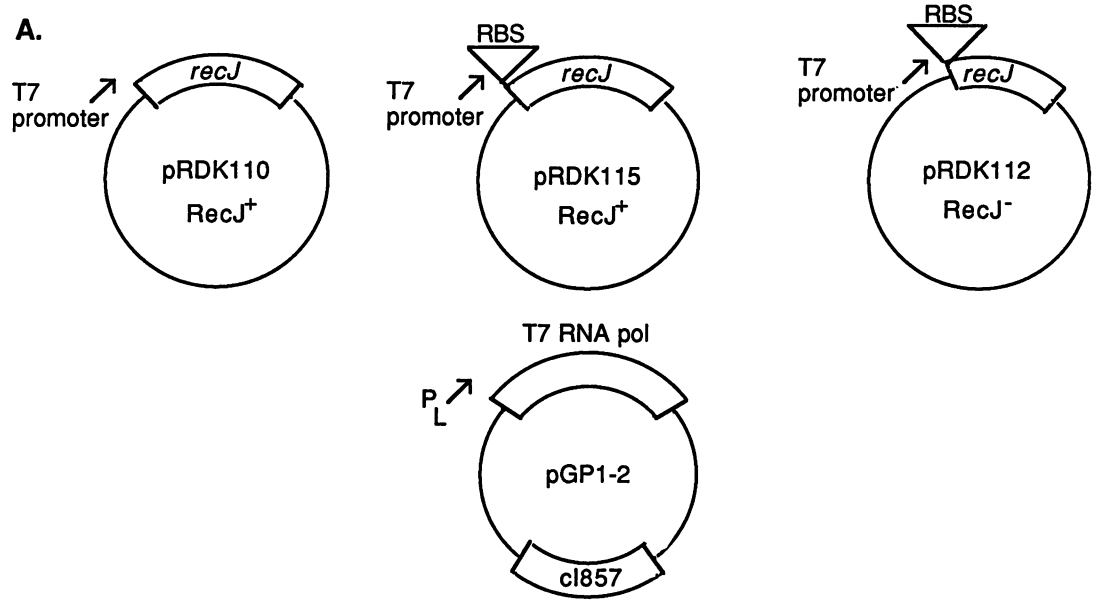
Protein sequence. No striking protein homology was seen in comparisons of the hypothetical RecJ protein sequence with the SWISS-PROT and PIR data bases with the FASTA algorithm. The protein sequence does not contain a type-A nucleotide-binding consensus (57) nor a zinc-finger binding motif (5), which are found in some recombination and repair proteins. The protein sequence predicted for XprA also failed to yield any striking similarities among data base proteins. However, it contained a motif at its amino terminus that was similar to the motif associated with signal peptide cleavage during protein export (42): Met Lys Lys, followed by 14 uncharged amino acids, followed by Ala Gln Ala. In addition, a region near the carboxyl terminus of XprA was relatively hydrophobic, containing a run of 28 uncharged amino acids.

The XprB protein sequence showed significant homology with the conserved C-terminal regions in many proteins that catalyze site-specific recombination reactions (Fig. 4), including the invariant arginine, histidine, and tyrosine residues. Within this group of proteins, the strongest overall homology of XprB protein was seen with the phase-variation regulatory proteins of *E. coli* (*fimB* and *fimE*). These proteins are thought to mediate site-specific inversion of fimbrial structural genes. The *fimB* and *fimE* genes are themselves strongly homologous, with 48% identical amino acids (21), and of these 93 amino acids, 38 are also shared by XprB. No clues as to an origin or potential function for XprB were found from the analysis of surrounding DNA sequences. No homologies to bacteriophage or transposon sequences were found which would suggest that this gene was derived from a cryptic phage or transposon. Furthermore, no restriction site bimorphisms in clones carrying these regions were observed which might indicate the presence of an inverting DNA segment like that associated with the *fim* genes (1).

Engineering overexpression of *recJ*. Three different *recJ* expression plasmids were constructed as described in Materials and Methods and briefly here. Maps of these plasmids along with relevant DNA sequence information are presented in Fig. 5. Plasmid pRDK110 contains the *Eco*RI-to-*Sal*I fragment of *recJ* cloned into the T7 promoter expression vector, pT7-5. pRDK112 is derived from pRDK110, with the region 5′ to the first ATG of the *recJ* coding sequence replaced with a sequence containing a Shine-Dalgarno ribosome-binding sequence and several restriction endonuclease cleavage sites. pRDK115 is similar to pRDK112 except that the ribosome-binding site was placed upstream of the first GTG in the *recJ* coding region.

Figure 6 shows total Coomassie blue-stained protein and labeled protein from extracts prepared before and after induction of expression strains carrying one of the three *recJ* plasmids, pRDK110, -112, and -115, or control plasmid pT7-5. With the pRDK110 plasmid carrying the native *recJ* gene, a 60,000-Da protein was specifically labeled in the presence of rifampin after heat induction. However, very little of this protein accumulated, as judged from the Coomassie blue-stained protein levels. pRDK115, which contained the GTG–ribosome-binding site fusion, promoted labeling of the same 60,000-Da protein along with substantial accumulation of this protein. The size of this protein is somewhat smaller than that predicted by the amino acid sequence (63,000 Da) and somewhat larger than that reported previously from maxicell labeling experiments (55,000 Da [29]). Direct comparison of the protein encoded by pRDK115 with the RecJ protein labeled in maxicell experiments performed exactly as previously described (29) showed that these two proteins comigrated during analysis by SDS-PAGE (data not shown). pRDK112 induced a slightly smaller 55,000-Da peptide, a size consistent with an N-terminal deletion of the normal *recJ* coding sequence. This truncated peptide accumulated, although less than that seen for the native RecJ protein expressed by pRDK115.

When the same strains were induced as above but without the addition of rifampin and returned to 30°C growth, only the pRDK115 strain showed a loss of viability. The plating efficiency of the pRDK115-carrying strain after 45 min at



B.

```
                rbs
pRDK115:   AAAGGAGGTACAAGCG GTG ...
                CGGTAAATAATTCGC GTG AAA CAA CAG ATA CAC CTT CGT CGC CGT GAA GTC
                GAT GAA ACG GCA GAC TTG CCC GCT GAA TTG CCT CCC TTG CTG CGC CGT
recJ+:          TTA TAC GCC AGC CGG GGA GTA CGC AGT GCG CAA GAA CTG GAA CGC
                AGT GTT AAA GGT ATG CTG CCC TGG CAG CAA
pRDK112:   AAAGGAGGTACAAGC ATG CTT CCC TGG CAA CAA ...
                rbs
```

FIG. 5. Construction of RecJ-overproducing plasmids. (A) Schematic illustration of different *recJ* derivatives of pT7-5: pRDK110, pRDK115, and pRDK112. Also shown is pGP1-2 (54). (B) Sequence changes introduced during the construction of RecJ expression plasmids. The sequence labeled *recJ*$^+$ is the wild-type *recJ* sequence between nucleotide positions 1726 and 1902 from Fig. 1. Shown in boldface are the sequences introduced to provide the ribosome-binding site (rbs, RBS) fusions to the first GTG codon (plasmid pRDK115) or the first ATG codon (plasmid pRDK112) of the *recJ* open reading frame. Note that additional changes have been made downstream of the ATG in pRDK112 to destabilize the potential palindrome marked by half arrows.

42°C was reduced to 0.3% of that with 30°C incubation. The cell viability observed when similar experiments were performed with pRDK110-, pRDK112-, and pT7-5-carrying strains was 50, 70, and 50%, respectively. These results indicate that overexpression of functional RecJ protein is lethal to the cell and may explain why RecJ is normally poorly expressed in the cell.

Finally, the overexpression plasmids were tested for ability to complement a *recJ* mutation. These plasmids were introduced into a *recB recC sbcA* genetic background in which *recJ* mutations produce a severe reduction in conjugational recombination frequencies and UV survival (28). Unlike the experiments described above, the plasmid carrying the T7 RNA polymerase gene was not introduced into these strains. Any complementation observed is therefore due to expression of *recJ* from other promoters present on the vector. Table 2 summarizes the results of these complementation assays. pRDK110, which contains the intact *Eco*RI-to-*Sal*I fragment, substantially increased the UV resistance and recombination proficiency of the *recJ* mutant strain. pRDK115, the GTG fusion plasmid, also complemented the *recJ* mutation, and the level of complementation observed was slightly higher than observed with pRDK110. pRDK112, the ATG fusion plasmid, and vector pT7-5 showed no detectable complementation. This confirmed that the first GTG corresponds to the natural initiation site of translation. In support of this, plasmid pRDK113 (an intermediate to pRDK115), which deletes *recJ* sequence to position 1701, including the proposed GTG start, but leaves intact other potential translational start sites (such as TTG), also failed to complement (data not shown). The full complementation of *recJ* mutations observed with pRDK115 also provides evidence that no detrimental changes to the *recJ* sequence were made during the construction of pRDK115.

DISCUSSION

The nucleotide sequence of the *recJ* region revealed a 578-amino-acid open reading frame in the region previously determined by complementation analysis to encode *recJ* (29). Downstream of *recJ* is the *prfB* gene, which encodes translation release factor RF2 and has been genetically mapped to the 62-min region (20). This gene is probably not

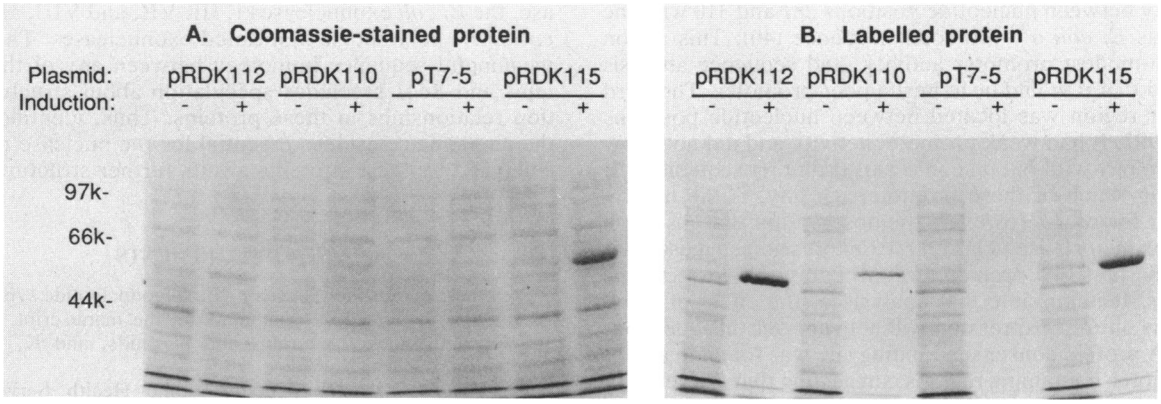


FIG. 6. Expression of RecJ protein by expression plasmids. (A) Coomassie blue-stained protein gel of [^{35}S]methionine-labeled protein expressed from pRDK112 (Δ*recJ* ribosome-binding site–ATG fusion in pT7-5), pRDK110 (*recJ*$^+$ in pT7-5), pT7-5 (T7 promoter vector), and pRDK115 (*recJ*$^+$ ribosome-binding site–GTG fusion in pT7-5), with and without temperature induction or rifampin treatment. (B) Autoradiograph of the same gel. Numbers on left show molecular weights (10^3).

coexpressed with *recJ*, because the promoter for *prfB* lies immediately upstream, near or after the C-terminal end of *recJ* (20). Two open reading frames potentially encoding 235- and 299-amino-acid proteins were found immediately upstream of the *recJ* gene. These two open reading frames appear to correspond to the previously described *xprA* and *xprB* genes (29). The orientation and close proximity of these two genes to *recJ* and the polar effects that insertions in these genes have on RecJ protein expression observed in maxicell labeling experiments (29) suggest that, at least in some situations, these three genes are coexpressed.

The predicted protein sequences of the two upstream genes that may be coexpressed with *recJ* provide some clues to their function. The N-terminal amino acid sequence predicted for the product of the *xprA* gene immediately upstream of *recJ* contained a signal peptide cleavage motif, suggesting that the protein encoded by this gene is exported to the periplasm or outer membrane. In addition, a hydrophobic C terminus suggests that the final product is membrane associated. This agrees with the observation that in maxicell labeling experiments (29), the *xprA* gene encoded two peptides of 26,000 and 25,000 Da, with the latter protein apparently being derived from the former. No significant homology of the *xprA* protein to known proteins was found in data base searches. The 31,000-Da *xprB* protein found further upstream of *recJ* has homology to a class of site-specific recombination proteins, typified by lambda bacteriophage *int* protein (3). The members of this "integrase"-related group include (in addition to bacteriophage *int* proteins) *fimE* and *fimB*. These are related proteins of *E. coli* involved in the site-specific inversion reaction controlling expression of fimbriae structural genes (12, 21). Other members of the group are protein D of F plasmid, which acts as a site-specific resolvase (25), the *tnpI* resolvase of the Tn*4430 Bacillus thuringiensis* (35), and the *tnpA* and *tnpB* proteins of *Staphylococcus aureus* transposon Tn*554* (41). The *xprB* and *fim* genes may be evolutionarily related since they show strong homology: *xprB* and *fimB* have 24% identity over 200 amino acids; *xprB* and *fimE* share 25% identity over 198 amino acids. It has yet to be determined whether *xprB* protein is involved in promoting any type of site-specific recombination event. The relationship of *xprA* and *xprB* to previously identified genes remains unknown. The *ops* gene (59) controlling exopolysaccharide production has been mapped near the 62.0-min region. However, it remains unclear whether it is identical to *xprA* or *xprB* or why they may be coexpressed with *recJ*. Note that *lysS*, which also maps at 62 min (13), is located on the other side of *prfB* from *recJ* (48).

TABLE 2. Complementing ability of RecJ overexpression plasmids[a]

Relevant mutant chromosomal genotype	Plasmid[b]	Conjugational recombination frequency (% Leu$^+$ [Ser$^+$ Sm^r])	% UV survival
recB recC sbcA recJ	pT7-5	0.021	0.002
recB recC sbcA recJ	pRDK110	6.2	11
recB recC sbcA recJ	pRDK115	10	34
recB recC sbcA recJ	pRDK112	0.016	0.033
recB recC sbcA		15	20

[a] UV and recombination phenotypes of a *recJ* mutant strain, RDK1656, carrying various expression plasmids and *recJ*$^+$ control strain JC8679 are illustrated. Conjugational recombination frequencies were determined in matings with JC158 for 1 h with a donor-to-recipient ratio of 0.1. The reported recombination frequency is the number of Leu$^+$ (Ser$^+$ Sm^r) transconjugant cells per donor. UV survival was determined from the titer of viable cells obtained after a 20-J/m^2 UV dose relative to that with no UV exposure.

[b] Plasmids are illustrated in Fig. 5.

No promoter was found immediately upstream of *recJ*. Three regions located within the *xprA/xprB* region upstream of *recJ* were found to exhibit weak to moderate promoter activity when fused to a promoterless *galK* gene. The first and strongest promoter region was furthest upstream of *recJ* (which could theoretically transcribe all three genes) and contained homology to the consensus σ^{54} promoter between nucleotide positions 119 and 135 (24). This class of alternate promoter uses the σ^{54} factor, originally discovered for nitrogen-regulated genes, and is required for the transcription of diverse groups of genes in many species of bacteria (2, 24). Initiation at σ^{54}-dependent promoters appears to require additional positive regulatory proteins whose activity is modulated by environmental sensor proteins (24). Whether the promoter found here requires *rpoN* (the structural gene for σ^{54}) and what genes might act as positive effectors remain to be determined. The second promoter region had some

homology between nucleotide positions 281 and 310 with the consensus *E. coli* σ^{70} promoter sequence (40). This region had only modest promoter activity, and sequence analysis predicted that it would be at best a poor promoter. The third promoter region was located between nucleotide positions 370 and 997. It had weak promoter activity and did not show any homology with *E. coli* consensus promoter sequences. It is unclear which of these promoters, if any, is the natural promoter for *recJ*. However, the observation that insertion mutations in *xprB* are polar on *recJ* expression suggests that one of the first two promoters is probably the correct *recJ* promoter. It is possible that analysis of the effect of *rpoN* mutations on *recJ* expression will help answer this question. No LexA protein consensus-binding site was found in any of the promoter-containing regions, suggesting that the genes in the *recJ* region are not *lexA* regulated.

The provision of a strong bacteriophage T7 promoter upstream did not, by itself, lead to high expression of RecJ protein. The *recJ* gene lacked a ribosome-binding consensus sequence, had a suboptimal GTG initiation codon, and exhibited low synonymous codon bias characteristic of poorly expressed *E. coli* genes (17, 23, 46, 49). The fusion of a consensus ribosome-binding sequence to the first GTG of the *recJ* open reading frame, in constructions that additionally contain the strong T7 promoter upstream of *recJ*, resulted in high expression of RecJ protein. This suggests that initiation of translation is limiting for *recJ* expression. Our data indicate that the first GTG, rather than the first ATG, is the natural translation initiation site of *recJ*. Fusion of a ribosome-binding sequence to the first ATG resulted in accumulation of a protein that was detectably smaller than wild-type RecJ, and this construct failed to complement *recJ* mutations. In contrast, the fusion of a ribosome-binding site to the first GTG resulted in accumulation of a protein of a size indistinguishable from wild type, and this construction exhibited full complementation of *recJ*. In the absence of optimal translation initiation signals, translation of *recJ* may be more efficient when coupled to translation of the immediately upstream *xprA* gene, allowing the ribosome to reinitiate at *recJ*. It may be for this reason that all insertion and deletion mutations in *xprA* result in reduced levels of RecJ protein expressed in maxicell labeling experiments (29).

In general, the nucleotide sequence of the *recJ* region revealed elements consistent with a theme of rather poor natural expression of the *recJ* gene. In this respect, *recJ* is similar to the recombination genes *recF* (7) and *recO* (39) with which it shares many phenotypic effects. The poor natural expression of *recJ* suggests that very little RecJ protein is required for recombination and UV repair. This is deduced from the observation that a *recJ* plasmid lacking known promoter sequences to transcribe *recJ* (pRDK115) exhibited full complementation of a *recJ* insertion mutation on the chromosome. In addition, high expression of *recJ* may be deleterious. In our RecJ protein-overproducing plasmid strain, a 45-min period of RecJ protein overexpression was sufficient to kill 99.7% of the cells in the population, whereas overexpression of a truncated form of RecJ protein was without effect. Similar lethality caused by the expression of the RecF protein has been observed (17a).

The plasmid construction which overexpresses RecJ protein led to the purification of RecJ protein and its identification as a single-stranded DNA-specific exonuclease. We found no significant gross homology of RecJ to other proteins in data base searches, which included several exonucleases, such as lambda exonuclease, T7 gene *6* exonuclease, the *E. coli* exonucleases I, III, VII, and VIII, and the *E. coli* DNA polymerase-associated exonucleases. The lack of meaningful sequence homology between any of these proteins and RecJ precludes speculation about structure-function relationships in these proteins. Thus, identification of the amino acid residues essential for the nuclease reactions catalyzed by these proteins awaits further structural analysis.

ACKNOWLEDGMENTS

We thank P. Morrison for help with oligonucleotide synthesis, C. Luisi-DeLuca for helpful comments on the manuscript, S. Tabor and K. McKenney for strains and plasmids, and K. Rudd for information regarding *lysS*.

This work was supported by Public Health Service grant GM26017 to R.K. from the National Institutes of Health and an American Cancer Society, Massachusetts Division, postdoctoral fellowship to S.L.

REFERENCES

1. **Abraham, J. M., C. S. Freitag, J. R. Clements, and B. I. Eisenstein.** 1985. An invertible element of DNA controls phase variation of type 1 fimbriae of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **82**:5724–5727.
2. **Albright, L. M., E. Huala, and F. M. Ausubel.** 1989. Prokaryotic signal transduction mediated by sensor and regulator protein pairs. Annu. Rev. Genet. **23**:311–336.
3. **Argos, P., A. Landy, K. Abremski, J. B. Egan, E. Haggard-Ljungquist, R. H. Hoess, M. L. Kahn, B. Kalionis, S. V. L. Narayana, L. S. Pierson III, N. Sternberg, and J. M. Leong.** 1986. The integrase family of site-specific recombinases: regional similarities and global diversity. EMBO J. **5**:433–440.
4. **Bachmann, B. J.** 1972. Pedigrees of some mutant strains of *Escherichia coli*. Bacteriol. Rev. **36**:525–557.
5. **Berg, J. M.** 1986. Potential metal-binding domains in nucleic acid binding proteins. Science **232**:485–487.
6. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
7. **Blanar, M. A., S. J. Sandler, M.-E. Armengod, L. W. Ream, and A. J. Clark.** 1984. Molecular analysis of the *recF* gene of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **81**:4622–4626.
8. **Clark, A. J.** 1963. Genetic analysis of a "double-male" strain of *Escherichia coli* K-12. Genetics **48**:105–120.
9. **Craigen, W. J., R. G. Cook, W. P. Tate, and C. T. Caskey.** 1985. Bacterial peptide chain release factors: conserved primary structure and possible frameshift regulation of release factor 2. Proc. Natl. Acad. Sci. USA **82**:3616–3620.
10. **Dale, R., B. A. McClure, and J. P. Houchins.** 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18S rRNA. Plasmid **13**:31–40.
11. **Doherty, M. J., P. T. Morrison, and R. Kolodner.** 1983. Genetic recombination of bacterial plasmid DNA: physical and genetic analysis of the products of plasmid recombination in *Escherichia coli*. J. Mol. Biol. **167**:539–560.
12. **Dorman, C. J., and C. F. Higgins.** 1987. Fimbrial phase variation in *Escherichia coli*: dependence on integration host factor and homologies with other site-specific recombinases. J. Bacteriol. **169**:3840–3843.
13. **Emmerich, R., and I. N. Hirshfield.** 1987. Mapping of the constitutive lysyl-tRNA synthetase gene of *Escherichia coli* K-12. J. Bacteriol. **169**:5311–5313.
14. **Fishel, R. A., A. A. James, and R. Kolodner.** 1981. RecA-independent general genetic recombination of plasmids. Nature (London) **294**:184–186.
15. **Friedman, D. I., and E. R. Olson.** 1983. Evidence that a nucleotide sequence, "box A," is involved in the action of NusA protein. Cell **34**:143–149.
16. **Gillen, J., D. K. Willis, and A. J. Clark.** 1981. Genetic analysis

of the RecE pathway of genetic recombination in *Escherichia coli* K-12. J. Bacteriol. **145**:521–532.
17. **Gren, E. J.** 1984. Recognition of messenger RNA during translational initiation in *Escherichia coli*. Biochimie **66**:1–29.
17a. **Griffin, T. J., IV, and R. D. Kolodner.** 1990. Purification and preliminary characterization of the *Escherichia coli* K-12 RecF protein. J. Bacteriol. **172**:6291–6299.
18. **Horii, Z. I., and A. J. Clark.** 1973. Genetic analysis of the RecF pathway of genetic recombination in *Escherichia coli* K12: isolation and characterization of mutants. J. Mol. Biol. **80**:327–344.
19. **James, A. A., P. T. Morrison, and R. Kolodner.** 1982. Genetic recombination of bacterial plasmid DNA: analysis of the effect of recombination-deficient mutations on plasmid recombination. J. Mol. Biol. **160**:411–430.
20. **Kawakami, K., Y. H. Jonsson, G. R. Bjork, H. Ikeda, and Y. Nakayama.** 1988. Chromosomal location and structure of the operon encoding peptide-chain-release factor 2 of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **85**:5620–5624.
21. **Klemm, P.** 1986. Two regulatory *fim* genes, *fimB* and *fimE*, control the phase variation of type 1 fimbriae in *Escherichia coli*. EMBO J. **5**:1389–1393.
22. **Kolodner, R. D., 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.
23. **Konigsberg, W., and G. N. Godson.** 1983. Evidence for rare codons in the *dnaG* gene and other regulatory genes in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **80**:687–691.
24. **Kustu, S., E. Santero, J. Keener, D. Popham, and D. Weiss.** 1989. Expression of σ^{54} (*ntrA*)-dependent genes is probably united by a common mechanism. Microbiol. Rev. **53**:367–376.
25. **Lane, D., R. de Feyter, M. Kennedy, S. H. Phua, and D. Semon.** 1988. D protein of mini-F acts as a repressor of transcription and as a site-specific resolvase. Nucleic Acids Res. **14**:9713–9728.
26. **Lloyd, R. G., C. Buckman, and F. E. Benson.** 1987. Genetic analysis of conjugational recombination in *Escherichia coli* K12 strains deficient in RecBCD enzyme. J. Gen. Microbiol. **133**:2531–2538.
27. **Lo, K.-M., S. S. Jones, N. R. Hackett, and H. G. Khorana.** 1984. Specific amino acid substitutions in bacterioopsin: replacement of a restriction fragment in the structural gene by synthetic DNA fragments containing altered codons. Proc. Natl. Acad. Sci. USA **81**:2285–2289.
28. **Lovett, S. T., and A. J. Clark.** 1984. Genetic analysis of the *recJ* gene of *Escherichia coli* K-12. J. Bacteriol. **157**:190–197.
29. **Lovett, S. T., and A. J. Clark.** 1985. Cloning of the *Escherichia coli recJ* chromosomal region and identification of its encoded proteins. J. Bacteriol. **162**:280–285.
30. **Lovett, S. T., and R. D. Kolodner.** 1989. Identification and purification of a single-strand-specific exonuclease encoded by the *recJ* gene of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **86**:2627–2631.
31. **Lovett, S. T., C. A. Luisi-DeLuca, and R. D. Kolodner.** 1988. The genetic dependence of recombination in *recD* mutants of *Escherichia coli*. Genetics **120**:37–45.
32. **Luisi-DeLuca, C., S. T. Lovett, and R. D. Kolodner.** 1989. Genetic and physical analysis of plasmid recombination in *recB recC sbcB* and *recB recC sbcA Escherichia coli* K-12 mutants. Genetics **122**:269–278.
33. **Mahajan, S. K.** 1988. Pathways of homologous recombination in *Escherichia coli*, p. 87–140. *In* R. Kucherlapati and G. R. Smith (ed.), Genetic recombination. American Society for Microbiology, Washington, D.C.
34. **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 repair. Mol. Gen. Genet. **216**:503–510.
35. **Mahillon, J., and D. Lereclus.** 1988. Structural and functional analysis of Tn*4430*: identification of an integrase-like protein involved in the cointegrate-resolution process. EMBO J. **7**:1515–1526.
36. **Manservigi, R., P. G. Spear, and A. Buchan.** 1977. Cell fusion induced by herpes simplex virus is promoted and suppressed by different viral glycoproteins. Proc. Natl. Acad. Sci. USA **74**:3913–3917.
37. **McKenney, K., H. Shimatake, D. Court, U. Schmeissner, and M. Rosenberg.** 1981. A system to study promoter and terminator signals recognized by *Escherichia coli* RNA polymerase, p. 383–415. *In* J. Chirikjian and T. Papas (ed.), Gene amplification and analysis, vol. 2. Elsevier/North Holland Publishing Co., New York.
38. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:2–78.
39. **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.
40. **Mulligan, R. E., D. K. Hawley, R. Entriken, and W. R. McClure.** 1984. *Escherichia coli* promoter sequences predict in vitro RNA polymerase selectivity. Nucleic Acids Res. **12**:789–800.
41. **Murphy, E., L. Huwyler, and M. do Carmo de Freire Bastos.** 1985. Transposon Tn*554*: complete nucleotide sequence and isolation of transposition-defective and antibiotic-sensitive mutants. EMBO J. **4**:3357–3365.
42. **Oliver, D. B.** 1987. Periplasm and protein secretion, p. 56–69. *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, vol. 1. American Society for Microbiology, Washington, D.C.
43. **Panet, A., J. H. van de Sande, P. C. Loewen, H. G. Khorana, A. J. Raae, J. R. Lillehaug, and K. Kleppe.** 1973. Physical characterization and simultaneous purification of bacteriophage T4 induced polynucleotide kinase, polynucleotide ligase and deoxyribonucleic acid polymerase. Biochemistry **12**:5045–5050.
44. **Parker, R. C., R. M. Watson, and J. Vinograd.** 1977. Mapping of closed circular DNAs by cleavage with restriction endonucleases and calibration by agarose gel electrophoresis. Proc. Natl. Acad. Sci. USA **74**:851–855.
45. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85**:2444–2448.
46. **Reddy, P., A. Peterkofsky, and K. McKenney.** 1985. Translational efficiency of the *Escherichia coli* adenylate cyclase gene: mutating the UUG initiation codon to GUG or AUG results in increased gene expression. Proc. Natl. Acad. Sci. USA **82**:5656–5660.
47. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–352.
48. **Rudd, K.** Unpublished data.
49. **Sharp, P. M., and W.-H. Li.** 1987. The codon adaption index—a measure of directional synonymous codon usage bias and its potential applications. Nucleic Acids Res. **15**:1281–1295.
50. **Shine, J., and L. Dalgarno.** 1974. The 3′ terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
51. **Smith, G. R.** 1988. Homologous recombination in prokaryotes. Microbiol. Rev. **52**:1–28.
52. **Stout, V., and S. Gottesman.** 1990. RcsB and RcsC: a two-component regulator of capsule synthesis in *Escherichia coli*. J. Bacteriol. **172**:659–669.
53. **Symington, L. S., P. T. Morrison, and R. Kolodner.** 1985. Intramolecular recombination of linear DNA catalyzed by the *Escherichia coli* RecE recombination system. J. Mol. Biol. **186**:515–525.
54. **Tabor, S., and C. C. Richardson.** 1985. A bacteriophage T7 RNA polymerase promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA **82**:1074–1078.
55. **Vogelstein, B, and D. Gillespie.** 1979. Preparative and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. USA **76**:615–619.
56. **Walker, G. E.** 1984. Mutagenesis and inducible responses to deoxyribonucleic acid damage in *Escherichia coli*. Microbiol.

Rev. **48**:60–93.

57. **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, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1**:945–951.
58. **Willetts, 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.
59. **Zinkewich-Peotti, K., and J. M. Fraser.** 1988. New locus for exopolysaccharide overproduction in *Escherichia coli* K-12. J. Bacteriol. **170**:1405–1407.