JOURNAL OF BACTERIOLOGY, July 1989, p. 3641–3649
0021-9193/89/073641-09$02.00/0

Vol. 171, No. 7

Molecular Analysis of the *Escherichia coli recO* Gene

PAUL T. MORRISON, SUSAN T. LOVETT, LYNN E. GILSON, AND RICHARD 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 3 January 1989/Accepted 1 April 1989

The plasmid pLC7-47, which contains *lep*, *rnc*, and *era*, was found to complement the UV-sensitive and recombination-deficient phenotypes caused by the *recO1504*::Tn*5* mutation. Southern blotting analysis demonstrated that pLC7-47 contained a segment of *Escherichia coli* DNA that covered the region of the *E. coli* chromosome containing the *recO1504*::Tn*5* mutation. A combination of deletion mapping and insertional mutagenesis localized the *recO*-complementing region to an approximately 1-kilobase region of a 1.6-kilobase *Bam*HI fragment. The DNA sequence of the 1.6-kilobase *Bam*HI fragment was determined and contained part of *era* and a 726-base-pair *recO* open reading frame. The *recO* open reading frame contained three possible translation start codons and could potentially encode a polypeptide of M_r 26,000. Computer analysis indicated that the putative RecO protein had suboptimal codon usage and did not show significant homology with previously identified proteins whose sequences were present in protein data bases. A combination of primary sequence analysis and secondary structure predictions suggested that *recO* contains a mononucleotide-binding fold.

Genetic analysis of recombination and repair in *Escherichia coli* has identified 11 genes, *recA*, *recB*, *recC*, *recD*, *recE*, *recF*, *recJ*, *recN*, *recO*, *recQ*, and *ruv*, whose gene products are required for recombination in *E. coli* (2, 4, 12, 16, 20, 24, 26, 29, 31, 40, 52). Mutations in genes whose products are involved in other aspects of DNA metabolism are also known to affect recombination (46). Historically, *E. coli* recombination genes have been classified as acting in the RecBC, RecE, or RecF recombination pathway depending on whether they were required for conjugal recombination in the wild-type, *recB recC sbcA*, or *recB recC sbcB* strain, respectively (10). The relationship between recombination genes and recombination pathways has not proven to be this simple. Extensive genetic analysis has demonstrated that the gene products required for a particular recombination event depend on the structure of the recombination substrate studied and the genetic background of the cell (10, 18, 20, 27, 30, 32, 34, 46; C. A. Luisi-Deluca, S. T. Lovett, and R. D. Kolodner, Genetics, in press). In some cases, a particular recombination pathway even fails to recognize a particular recombination substrate. To understand this apparent diversity of recombination events, it will be necessary to obtain a greater understanding of the biochemical activities of recombination proteins and the mechanism of individual recombination events.

The most recently identified *E. coli* recombination gene is the *recO* gene (26). The *recO* gene product has a broad involvement in different recombination and repair events in *E. coli*. In wild-type *E. coli* strains, the *recO* gene product is required for recombination of circular plasmids, the repair of UV damage to DNA, and the formation of a portion of the transcribable recombination intermediates that are formed during conjugal recombination (26, 30). In *recB recC sbcB* mutant strains, the *recO* mutation behaves like a typical RecF pathway mutation, decreasing the frequency of both conjugal recombination, circular plasmid recombination, linear dimer plasmid recombination, and the repair of DNA damage caused by UV irradiation or mitomycin C (26; Luisi-Deluca et al., in press). In *recB recC sbcA* mutant strains, the *recO* mutation, like many other RecF pathway mutations, decreases conjugal recombination and the repair of UV-induced DNA damage (Luisi-Deluca et al., in press). The *recO* mutation also decreases the frequency of recombination of circular plasmid substrates in *recB recC sbcA* strains but has no effect on the recombination of linear dimer plasmid substrates (Luisi-Deluca et al., in press). Exactly how *recO* plays such a broad role in recombination and in different "recombination pathways" is unclear at present. In this communication we describe experiments on the molecular characterization of the *recO* gene which were performed as a step in the process of clarifying the role that *recO* plays in recombination.

MATERIALS AND METHODS

Strains, plasmids, and phages. Strains AB1157 (*rec*$^+$) (3), RDK1541 (*recO1504*::Tn*5*) (26), JC8679 (*recB21 recC22 sbcA23*) (20), and RDK1658 (*recB21 recC22 sbcA23 recO1504*::Tn*5*) (Luisi-Deluca et al., in press) have been described previously and were used for *recO* plasmid complementation assays. These strains also carry the following mutations: *argE3* Δ(*gpt-proA*)*62 thi-1 his-4 thr-1 leuB6 ara-14 lacY1 xyl-5 mtl-1 galK2 kgdK51 supE44 rpsL31 tsx-1*. JC158 (Hfr PO1 *serA6 rel-1 thi-1 lacI22*) (9) was used as the donor in conjugational recombination tests. MG1063 (F$^+$ *recA56*) (22) and JC2926 (*recA13* derivative of AB1157 described above) (12) were used for γδ mutagenesis. pBR322 and M13 cloning vectors mp18 and mp19 were from laboratory collections. The bacteriophage T7 promoter plasmid expression vector pT7-5 was from Stanley Tabor (Harvard Medical School, Boston, Mass.) and is almost identical to pT7-1 except that the polarity of the *amp* gene relative to the T7 promoter is inverted (47). pLC7-47 (11) was obtained from Barbara Bachmann of the *E. coli* Genetic Stock Center. The *recO* plasmids constructed in this study are illustrated in Fig. 1 or described in the legend to Fig. 4.

Media. Strains were grown routinely in LB medium (52). Plate minimal medium consisted of 56/2 salts (52) with 0.2% glucose, 0.5 μg of thiamine per ml, and 50 μg of amino acid

* Corresponding author.

supplements per ml. Streptomycin, ampicillin, and kanamycin were added to 100, 100, and 30 μg/ml, respectively. Selections with colicin E1 were carried out as previously described (11).

Recombination and UV tests. Published procedures for the quantitation of conjugational recombination assays and UV survival were used (34). Plasmid-containing strains were grown in L-ampicillin broth; those with additional *recO*::Tn*5* mutations were grown in L-ampicillin-kanamycin broth to insure maintenance of plasmid and the chromosomal *recO* mutation. One-hour matings were performed with Hfr JC158 at a 1:10 donor-to-recipient ratio, with subsequent selection for Leu^+ Ser^+ Sm^r. UV survival was determined by plating on LB plates. Plate assays for conjugational recombination and UV survival (12) were employed for determining the complementation ability of γδ insertion mutants of plasmid pRDK124.

Insertional mutagenesis and transformation procedures. Insertional mutagenesis with γδ was carried out essentially as described by Guyer (22). Briefly, pRDK124 was transformed into MG1063 and then transferred to JC2926 by mating and selecting Ap^r Sm^r transconjugants. Plasmid DNA was isolated from a number of individual transconjugants, and the position and orientation of individual γδ insertions were determined by restriction mapping with a *Bam*HI single digest to locate the site of insertion and a *Eco*RI-*Sal*I double digest to determine the orientation of the transposon. Plasmids of interest were then transformed into RDK1658 for complementation tests. Transformations were performed essentially as described previously (51).

Enzymes and reaction conditions. Restriction endonucleases were obtained from New England BioLabs (Beverly, Mass.) and used according to the instructions supplied by the manufacturer. *E. coli* DNA polymerase I and the Klenow fragment of *E. coli* DNA polymerase I were also from New England BioLabs. Homogeneous T4 DNA ligase was purified by an unpublished method of R. Kolodner. Ligase reactions were carried out at 12.5°C in buffer containing 40 mM Tris hydrochloride (pH 7.8), 8 mM $MgCl_2$, 5 mM 2-mercaptoethanol, 0.067 mM ATP, and 0.1 to 1 PP_i exchange unit of ligase per ml. DNA fragments were joined to each other in two-step ligation reactions; the first reaction was at high DNA concentration (50 to 100 μg/ml) to favor end-to-end joining, and the second reaction was at low DNA concentration (0.1 of the DNA concentration where $I = J$ for the desired construct [15]) to favor cyclization.

DNA purification and analysis. Plasmid DNA was prepared by the method of Holmes and Quigley (23) or by the alkaline lysis method (5). For plasmid constructions, the DNA was further purified as follows. DNA was treated with RNase A at 200 μg/ml in buffer containing 10 mM Tris hydrochloride (pH 8.0), 1 mM EDTA, and 100 mM NaCl. After 15 min at 23°C the DNA was extracted once with phenol and chromatographed on a Pasteur pipette column of agarose A5m (200–400 mesh; Bio-Rad Laboratories, Richmond, Calif.) run in 10 mM Tris hydrochloride (pH 8.0)–1 mM EDTA–100 mM NaCl. The DNA-containing fractions were pooled, precipitated with ethanol, and suspended in 10 mM Tris hydrochloride (pH 8.0)–1 mM EDTA. Minipreparations of single-stranded or double-stranded M13 DNA were prepared exactly as described previously (37). M13 replicative-form DNA was from New England BioLabs or was purified as previously described (37). DNA samples were analyzed by electrophoresis through gels containing 0.8% agarose run in Tris acetate-EDTA buffer containing 0.5 μg of ethidium bromide per ml (18). Southern blot analysis was carried out as described previously (8). Hybridization probes labeled with [α-^{32}P]dATP (New England Nuclear Corp., Boston, Mass.) were prepared by random priming (17) of *recO*-M13 templates with random primers purchased from Boehringer Mannheim Biochemicals (Indianapolis, Ind.).

DNA sequence analysis. DNA sequence analysis was performed by the dideoxynucleotide-chain termination sequencing method with [^{35}S]dATP (New England Nuclear) and sequencing kits obtained from either New England BioLabs or US Biochemicals (Cleveland, Ohio). Two different sequencing strategies were used. In some cases nested deletions were made by the method of Dale et al. (13) by using a commercially available kit (IBI Inc., New Haven, Conn.), and then the deletions were sequenced by using the universal sequencing primer. In other cases, oligonucleotide primers that were complementary to regions of known sequence were synthesized and used to sequence adjacent regions. Oligonucleotides were synthesized with an Applied Biosystems 380A DNA synthesizer and β-cyanoethyl phosphoramadites. When compressions were observed, their sequence was resolved by substituting dITP for dGTP as well as by sequencing from multiple primer sites. The accuracy of the sequence was confirmed by verifying the existence of restriction endonuclease cleavage sites predicted from the DNA sequence and by comparison with regions that had been sequenced by others.

Computer analysis. Computer analysis of DNA sequence data was carried out by using either the DNA Inspector II program (Textco, West Lebanon, N.H.) or computer programs available on BIONET or at the Dana-Farber Cancer Institute Molecular Biology Computer Research Resource. Homology searches against the protein databases PIR/NBRF release 17.0 and SWIS-PROT release 8.0 (12/1988) were performed by using the FASTA program (41) and a Ktup value of 2. Signature sequence searches were performed by using the QUEST program and KEYBANK release 3.0. Secondary structure predictions were made using the PRSTRC adaptation (44) of the method of Chou and Fasman (7) and the following values: α-helix former, 1.12 (threshold, 1.08; cutoff, 1.00); β-strand former, 1.25 (cutoff, 0.97); β-turn minimum value, 0.20; omega loop threshold, 1.10.

RESULTS

RecO plasmid derivatives. Previous genetic mapping studies located *recO* at 55.4 min on the standard *E. coli* genetic map and suggested that it might be no more than 0.1 min from *lep* (26). Because *lep* was known to be encoded on pLC7-47 (14), this plasmid was tested for its ability to complement the UV-sensitive phenotype of *recO* single mutants and was found to complement this defect (data not shown). To better locate *recO*, a series of deletion and subcloning experiments was carried out with pLC7-47 (Fig. 1). Representative results of complementation experiments are present in Fig. 2 and Table 1 and are discussed in more detail below. An 8.4-kilobase (kb) *Hin*dIII fragment, when subcloned into pBR322 to yield pRDK123, complemented *recO*. By deleting different restriction fragments from pRDK123, the *recO* complementing region was located to a 3.2-kb region containing a 1.6-kb *Bam*HI fragment and a 1.6-kb *Bam*HI-to-*Hin*dIII fragment. When the 1.6-kb *Bam*HI fragment was subcloned into pBR322 to yield pRDK125, it complemented *recO* mutations.

Complementation of *recO1504*::Tn*5*. Initial studies with *recO* mutations demonstrated that they cause a modest

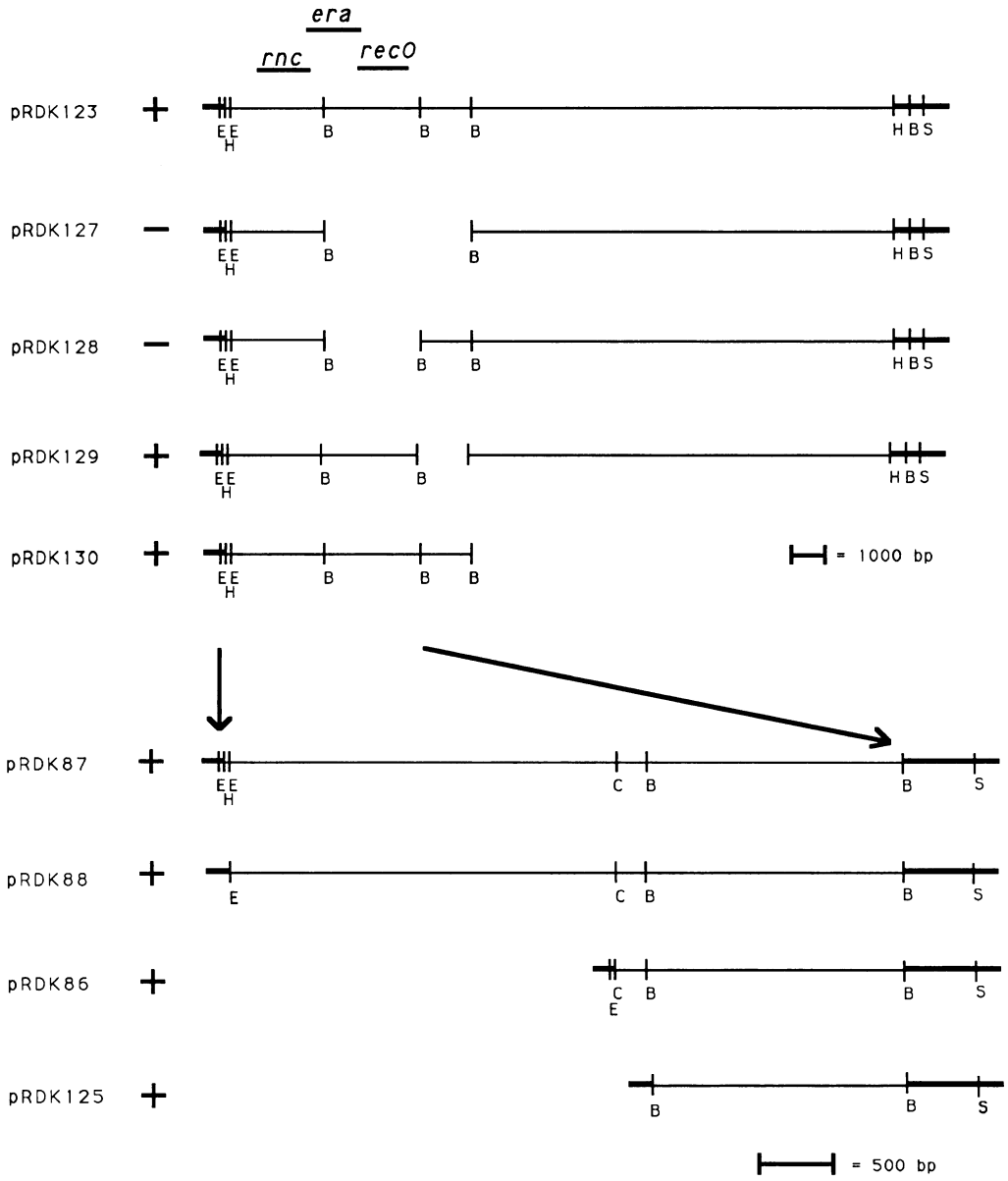

FIG. 1. Maps of plasmids constructed for *recO* complementation tests. All plasmids contain pBR322 oriented so that the *amp* gene is to the left of the left *Hin*dIII site of pRDK123 and the *tet* gene is to the right of the right *Hin*dIII site. Linear maps of the *E. coli* DNA inserts are shown along with some of the flanking pBR322 DNA indicated by the bold lines. Note that the region of pBR322 DNA between the *Eco*RI and *Hin*dIII sites contains a *Cla*I site that is not indicated. All plasmids shown were derived from pRDK123 by deletion of the indicated fragments. pRDK87, pRDK127, pRDK129, and pRDK130 were derived from pRDK123 by deletion of the indicated *Bam*HI fragments followed by cyclization under sticky-end ligation conditions. Similarly, pRDK88 and pRDK86 were derived from pRDK87 by deletion of the indicated *Eco*RI and *Cla*I fragments. pRDK125 was derived from pRDK86 by deleting the indicated *Bam*HI-to-*Cla*I fragment, filling in the cohesive ends, and cyclizing by blunt end ligation so that the *Bam*HI site was regenerated. The top of the figure shows the relative positions of *rnc*, *era*, and *recO*. Plasmids are drawn to the scale indicated below each group of plasmids. + and − refer to the results of complementation tests measuring the effect of each plasmid on the UV sensitivity of RDK1541 and RDK1658. Complementation, when it was observed, occurred in both strains, and typical results of complementation tests are shown in Fig. 2. Restriction endonuclease cleavage sites: H, *Hin*dIII; R, *Eco*RI; B, *Bam*HI; C, *Cla*I.

UV-sensitive phenotype when present in *E. coli* AB1157 and cause a much more severe UV-sensitive phenotype as well as a defect in conjugal recombination when present in combination with *recB*, *recC*, and *sbcB* mutations (26). The more extreme phenotype of *recO recB recC sbcB* strains would make complementation studies easier to interpret; however, such studies might be complicated by the known instability of some, but not all, plasmids in strains containing *recB*, *recC*, and *sbcB* mutations (49). Therefore, the effect of *recO1504* on UV sensitivity and conjugal recombination was tested in a *recB recC sbcA* strain, where plasmids are more stable (18; Luisi-Deluca et al., in press), and the *recO1504*::Tn*5* mutation causes a large defect in the repair of UV damage and conjugal recombination (30; Luisi-Deluca et al., in press).

During the deletion and subcloning studies described

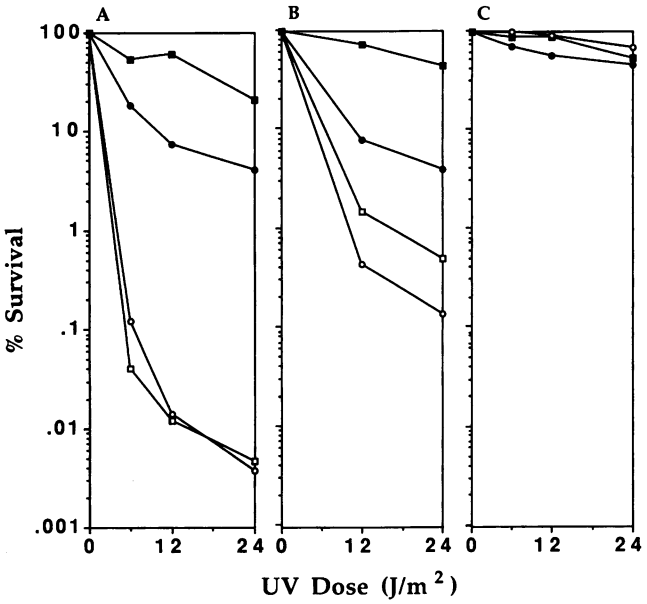

FIG. 2. Complementation of the *recO* UV-sensitive phenotype by *recO* containing plasmids. (A) *recB recC sbcA recO* strains: ■, JC8679; ●, RDK1658(pRDK86); ○, RDK1658(pBR322); □, RDK1658. (B) *recO* strains: ■, AB1157; ●, RDK1541(pRDK86); ○, RDK1541(pBR322); □, RDK1541. (C) Wild-type strains: ■, AB1157; ●, AB1157(pRDK86); ○, AB1157(pBR322).

above, the ability of all plasmids to complement the defects caused by *recO1504* in both the wild type and a *recB recC sbcA* strain was tested. The results of complementation tests with pRDK86, which contains the 1.6-kb *Bam*HI fragment and a 200-base-pair (bp) *Cla*I-to-*Bam*HI fragment, are presented in Fig. 2 and Table 1. The results obtained with other *recO*-complementing plasmids were similar (data not shown). pRDK86 showed almost complete complementation of the UV-sensitive defect (Fig. 2A) caused by the *recO* mutation in the *recB recC sbcA* strain and showed substantial but not complete complementation of the UV-sensitive phenotype of the *recO* single-mutant strain (Fig. 2B), whereas pBR322 showed no complementation in any *recO* strain. pRDK86 also showed substantial complementation of the conjugal recombination defect in the *recO recB recC sbcA* strain, whereas pBR322 had no effect (Table 1).

Incomplete complementation has been observed in studies with other cloned recombination and repair genes (33, 35, 39). Two possible explanations for the observed incomplete complementation of *recO* mutant strains by the *recO* plasmids are as follows: (i) the presence of the *recO* gene on a multicopy plasmid in wild-type strains causes a UV-sensitive phenotype, as has been observed for *recN* (43) and *uvrD* (35), and the presence of the *recO* gene on a multicopy plasmid in *recB recC sbcA* strains causes a recombination defect; or (ii) *recO* is not expressed well enough from pRDK86 to complement these defects. To investigate the first possibility, the effect of pRDK86 and pBR322 on the repair of UV damage in *E. coli* AB1157, the parental strain, was investigated (Fig. 2C). pRDK86 slightly increased the UV sensitivity of *E. coli* AB1157 but not enough to explain the observed incomplete complementation. Similarly, pRDK86 did not cause a significant recombination defect when it was present in a *recB recC sbcA* strain (data not shown). Supporting the second possibility, the DNA sequence analysis described below indicates that *recO* should be expressed poorly, and maxicell analysis with pRDK86 (unpublished observations) has failed to detect the RecO protein. Loss of the plasmid from some cells in the population due to instability may also explain the observed incomplete complementation results. Additional information about the maintenance and expression of *recO* plasmids and the function of *recO* is clearly required before it will be possible to explain the incomplete complementation of the *recO* mutation by *recO*-containing plasmids.

TABLE 1. Complementation of the conjugal recombination deficiency caused by *recO1504*::Tn*5* mutations

Relevant genetic background[a]	Plasmid	Recombination frequency[b] (%)
recB21 recC22 sbcA23		62
recB21 recC22 sbcA23 recO1504::Tn*5*		0.027
recB21 recC22 sbcA23 recO1504::Tn*5*	pBR322	0.027
recB21 recC22 sbcA23 recO1504::Tn*5*	pRDK86	1.5

[a] The *recB21 recC22 sbcA23* strain used was JC8679, and the *recB21 recC22 sbcA23 recO1504* strain used was RDK1658.

[b] This recombination frequency is the number of Leu$^+$ (Sm^r Ser$^+$) transconjugants divided by the number of Hfr donor cells. The frequency of F′ inheritance was similar (86 to 100% of JC8679) with all strains tested.

Mapping the chromosomal location of *recO1504*::Tn*5*. The results described above identified a 1.6-kb segment of *E. coli* chromosomal DNA that is able to complement a *recO* mutation but do not prove that this segment of DNA corresponds to the chromosomal *recO* locus. The position of the chromosomal *recO1504*::Tn*5* mutation was determined by Southern blotting with a hybridization probe derived from M13mp19 containing the cloned 1.6-kb *recO*-complementing region (Fig. 3). Analysis of a *Hin*dIII digest of *E. coli* AB1157 (wild type) DNA, the parental DNA, detected a single 8.4-kb *Hin*dIII fragment corresponding in size to the *recO*-complementing *Hin*dIII fragment present in pLC7-47. Similar analysis of chromosomal DNA isolated from the *recO1504*::Tn*5* mutant detected, instead, 3.7- and 6.3-kb fragments. Because each IS*50* element of Tn*5* contains a *Hin*dIII site 1.05 kb in from the outside end (25), the most likely explanation of this result is that *recO1504*::Tn*5* is caused by the insertion of Tn*5* at a site 2.6 kb from one end of the 8.4-kb *Hin*dIII fragment. Analysis of a *Bam*HI digest of *E. coli* AB1157 (wild type) DNA, the parental DNA, detected a single 1.6-kb *Bam*HI fragment corresponding in size to the *recO*-complementing *Bam*HI fragment present in pRDK88 and other *recO*-complementing plasmids. Similar analysis of chromosomal DNA isolated from the *recO1504*::Tn*5* mutant detected an approximately 4-kb band that appeared to be a doublet on some underexposed autoradiograms (data not shown). Because Tn*5* contains single *Bam*HI sites 3.0 and 2.8 kb from its left and right ends, respectively (25), the most likely explanation for this result is that *recO1504*::Tn*5* is caused by the insertion of Tn*5* at a site approximately in the middle of the 1.6-kb *Bam*HI fragment. The location of the 1.6-kb *Bam*HI fragment within the 8.4-kb *Hin*dIII fragment is known (Fig. 1) and places the two *Bam*HI sites 1.65 and 3.25 kb from the *Hin*dIII closest to the *rnc* gene. Thus the middle of the *Bam*HI fragment is 2.6 kb from the closest *Hin*dIII site; this is consistent with the site of *recO1504*::Tn*5* being 2.6 kb from the *Hin*dIII site closest to the *rnc* gene. This analysis places the *recO1504*::Tn*5* insertion approximately 300 bp internal from the amino terminus of the *recO* ORF discussed below (see Fig. 4 and 6). Identical conclusions were obtained when nick-translated

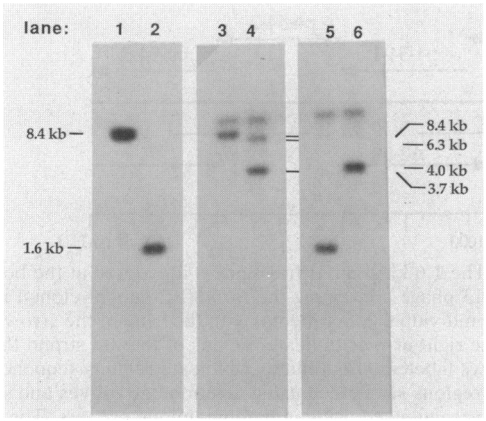

FIG. 3. Southern blot mapping of the chromosomal *recO1504*::Tn*5* mutation. Lanes: 1, pRDK123 DNA digested with *Hin*dIII; 2, pRDK123 DNA digested with *Bam*HI; 3, AB1157 chromosomal DNA digested with *Hin*dIII; 4, RDK1541 (*recO*) chromosomal DNA digested with *Hin*dIII; 5, AB1157 chromosomal DNA digested with *Bam*HI; 6, RDK1541 (*recO*) chromosomal DNA digested with *Bam*HI. The hybridization probe used was derived from M13mp19 containing the cloned 1.6-kb *recO*-complementing *Bam*HI fragment from pRDK125 (Fig. 1). The sizes indicated were determined relative to standards produced by digesting λ DNA with *Hin*dIII or pRDK123 DNA with either *Hin*dIII or *Bam*HI and run in parallel lanes. The hybridization to the most slowly migrating DNA observed in lanes 3 through 6 appears to represent hybridization to partially digested DNA that was visible on the ethidium bromide-stained gel and probably also represents hybridization to labeled *lacI* and *lacZ* sequences present on M13mp19. Identical results were obtained when nick-translated pRDK88 was used as the hybridization probe in similar Southern blotting experiments (data not shown).

pRDK88 was used as the hybridization probe in similar Southern blotting experiments (data not shown).

Insertional mutagenesis of the *recO*-complementing region. To better define the *recO*-complementing region, a number of γδ insertion mutations in pRDK124 were isolated and mapped. A total of 32 independent insertions were located by restriction mapping and tested for their ability to complement the UV-sensitive defect of a *recO recB recC sbcA* strain. The location and orientation of 14 of these insertion mutations and their effect on complementation are summarized in Fig. 4. All of the insertions that inactivated complementation mapped within an 800-bp region of the 1.6-kb *Bam*HI *recO*-complementing fragment, which contains the mapped site of the *recO1504*::Tn*5* insertion mutation. The insertions that mapped upstream of *recO* did not appear to be polar on *recO*. The reason for this is unclear; however, several possible explanations for this are that only a small amount of *recO* expression is necessary for complementation, the complementation assay is not sensitive enough to detect a polar effect as has been seen in studies with *recJ* (33), or there is a promoter close to the start of the *recO* open reading frame (ORF).

Sequence analysis of the *recO*-complementing region. To analyze the structure of the *recO*-complementing region, the 1.6-kb *Bam*HI *recO*-complementing restriction fragment was cloned in both orientations into M13 sequencing vectors and sequenced by the dideoxynucleotide sequencing method. The sequencing strategy is outlined in Fig. 5, and the DNA sequence is presented in Fig. 6. A portion of this sequence corresponding to the *era* gene has been previously reported (1, 36). Five differences in the sequence of the *era* region were observed: omission of a G at position 359, addition of a G at position 365, substitution of a G for an A at position 434, substitution of a G for a C at position 723, and addition of a G at position 837. These regions fall within compressions that can obscure some sequence information and give rise to sequencing errors. We have compared our sequence from positions 721 to 1496 with the sequence determined by Takiff et al. (48) in their studies on the *rnc* operon and found 100% agreement between the two sequences.

Our sequence contained a portion of an ORF that corresponded to the previously described *era* gene (1, 36). This ORF overlapped by 35 nucleotides a different ORF, which started at position 771 and ended at position 1497, terminating with two TGA stop codons. The 726-bp ORF covered the entire region where the *recO* insertion mutations mapped identifying it as the *recO* ORF. The *recO* ORF had three potential start codons: an ATG at position 771, an ATG at position 831, and a GTG at position 864. The first ATG had a marginal ribosome-binding site (21) preceding it, and it is possible that the 35-base overlap with the *era* ORF could prevent translation from initiating from this ATG. The second ATG and the GTG both had reasonable ribosome-binding sites preceding them, although the distance relative to the start codons was not optimal (21) (Fig. 6). The GTG codon was within a 12-bp palindrome (containing two mispairs) whose formation on the mRNA level could conceivably interfere with the ability of the GTG codon to serve as an initiation codon. At present, it is unclear which is the correct initiation codon. The second ATG seems to be the most likely candidate, but it will require additional analysis to resolve this point. Because of this uncertainty, all of the

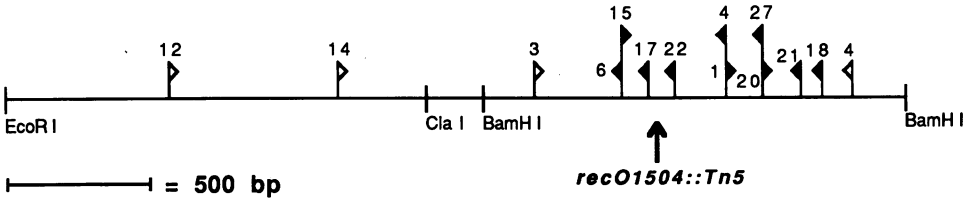

FIG. 4. γδ insertional mutagenesis of pRDK124. pRDK124 contains the *Eco*RI-to-*Sal*I fragment of pRDK88 inserted between the *Eco*RI and *Sal*I polylinker sites of pT7-5; this fragment contains the *Eco*RI-to-*Bam*HI segment of *E. coli* DNA present in pRDK88 (Fig. 1) and the 276-bp *Bam*HI-to-*Sal*I fragment of pBR322, which is located to the right of the *Bam*HI site in pRDK88 (Fig. 1). The map of the *recO*-containing portion of the insert contained in pRDK124 is presented in the same orientation as in Fig. 1. The pT7-5 portion is not shown, but a relevant map can be found in reference 45. The positions of the γδ insertions are indicated by the flags. The plasmids containing the insertions marked by the open flags complemented the UV-sensitive phenotype of RDK1658 (*recB recC sbcA recO*), and the plasmids containing the insertions marked by the closed flags failed to complement. Left-facing flags indicate the orientation of γδ having the largest γδ *Eco*RI junction fragment on the left side of the γδ insertion, and right-facing flags indicate the opposite orientation. The position of *recO1504*::Tn*5* is indicated by the arrow.

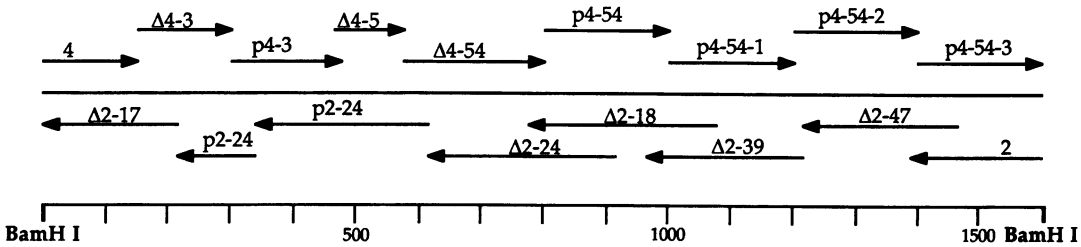

FIG. 5. Strategy for sequencing the 1.6-kb *Bam*HI fragment containing *recO*. The 1.6-kb *Bam*HI fragment is illustrated at the bottom. It was sequenced by the dideoxy sequencing technique by utilizing recombinant M13 phage containing the *Bam*HI fragment cloned in either M13mp18 and M13mp19. The arrows indicate the region of sequence determined in individual experiments, with the head of the arrow located at the 3′ end of the sequence determined and pointing in the 5′-to-3′ direction. The right arrows indicate regions of the top strand that were determined, and the left arrows indicate the sequence of the bottom strand. Arrows labeled with a number indicate regions sequenced with standard M13 sequencing primers, arrows labeled with a Δ and a number indicate regions sequenced using deletion derivatives and standard M13 sequencing primers, and arrows labeled with a p and a number indicate regions sequenced using individual unique primers. The primers used are listed in the 5′-to-3′ direction as follows: p4-3, GCGCGAAGGCAAGC; p4-54, GTGGAGCGAAACCAG; p4-54-1, AAGT TCAAAACGCTGCG; p4-54-2, TGCGCCGCTTTGAACT; p4-54-3, TAAACGCACGGGAATT; p2-24, TTCGAACATTTCCTG; p2-24-1, ACGCACGATTGCCGCAATAG.

analysis of the putative RecO protein described below uses the first ATG as the initiation codon.

Analysis of a putative RecO protein. An extensive computer analysis of the structure of the protein encoded by the *recO* ORF has been performed to gain possible insight into the function of *recO*. Because in the absence of confirming biochemical data such studies only provide correlative information, the results will only be summarized here. Analysis of the structure of *recO* by the methods of Chou and Fasman (7) suggests that *recO* encodes a globular protein lacking a membrane-spanning segment and containing a relatively hydrophilic carboxy terminus. Analysis of the codon usage in *recO* revealed a codon adaptation index (45) of 0.307, suggesting that *recO* might be poorly translated. Searching the RecO protein sequence against the protein data bases available on Bionet did not reveal any major homology with known proteins. Searching *recO* against all of the KEYBANK data available on Bionet as of December 1988 with the QUEST program did not reveal any significant matches.

recO did not contain a match with the nucleotide-binding fold signature sequence GxxxxGKT (50). However, starting at nucleotide position 952 (amino acid position 61), *recO* contained the amino acid sequence GgrgeVKTlrsaeaV, which is an excellent match with nucleotide-binding fold segments 1 and 2, except for the G-to-V change (19, 50). In vitro mutagenesis experiments with *E. coli* F1 ATPase have shown that the G (of GKT) is not absolutely required for ATPase activity (42). Structural predictions demonstrated that starting with amino acid 50 (nucleotide position 919) this region of *recO* matched the primary sequence-secondary structure descriptor for nucleotide-binding folds: NH_2-(β strand)-GXXXX-(GKα-helix)-(0 to 11 amino acids)-(β strand) (Fig. 7) (6). Starting at nucleotide position 1282 (amino acid position 171), *recO* contained the amino acid sequence GfiasVVID, which is predicted to be a segment of hydrophobic β strand terminating in a negative charge (Fig. 7). This appears to be an excellent match with nucleotide-binding fold segments 3, virtually all of which contain an N-terminal G and a C-terminal D (19, 50). In addition, there is an R that is highly conserved at amino acid position 167 (50). This latter β strand has been implicated in Mg binding (19, 28). The structural predictions also indicate that *recO* contains at least three other segments that could have β-strand-forming potential and could contribute to the formation of a stable β-pleated sheet structure. This analysis suggests, but does not prove, that RecO protein might interact with a nucleotide cofactor or even hydrolyze nucleoside triphosphates.

DISCUSSION

The *E. coli recO* gene was identified by the isolation of a Tn*5* insertion mutation linked to *tyrA*, *nadB*, and *purI*. This mutation in *recO* caused a variety of defects in genetic recombination and DNA repair. The general properties of the *recO* mutation were typical of RecF pathway genes. The present study has identified a 726-bp ORF that appears to encode the *recO* gene. The evidence for this includes the following. (i) The original chromosomal *recO1504*::Tn*5* mutation maps in the *recO* ORF. (ii) Plasmids containing a portion of the upstream *era* ORF, the complete *recO* ORF, and no other downstream sequences complement all of the defects caused by the *recO1504*::Tn*5* mutation; this complementation is consistent with the *recO1504*::Tn*5* mutation being an insertion in *recO* rather than being polar on *recO*. (iii) Insertional mutagenesis experiments demonstrated that the *recO*-complementing region maps to within the *recO* ORF and does not require other sequences. Takiff et al. (48) identified a 726-bp ORF that maps downstream of *era*. Because this ORF and the *recO* ORF have the same sequence and map in the same place, the data presented here indicate that *recO* could be a member of the *rnc* operon, although additional evidence will be required to prove that *recO* is expressed as a member of this operon (48). *rnc* encodes RNase III, and *era* is an essential gene encoding a protein of unknown function in *E. coli* (1, 36, 48). The significance of the relationship between *recO* and the *rnc* operon is unclear at present.

The *recO* ORF could, in principle, encode a protein of approximately M_r 26,000. However, at present, the exact start codon for this putative RecO protein has not been resolved. The possible 35-base overlap between *recO* and the upstream translated *era* sequences combined with the *recO* codon usage suggest that expression of the RecO protein might be inefficient. Expression studies in which the 1.6-kb *Bam*HI fragment containing *recO* and the upstream *era* sequences was inserted downstream of the T7 promoter present in pT7-5 and expressed under control of T7 RNA polymerase identified a weakly expressed, plasmid insert-encoded protein of M_r 26,000 (data not shown). Although this protein may not be the RecO protein, this result is

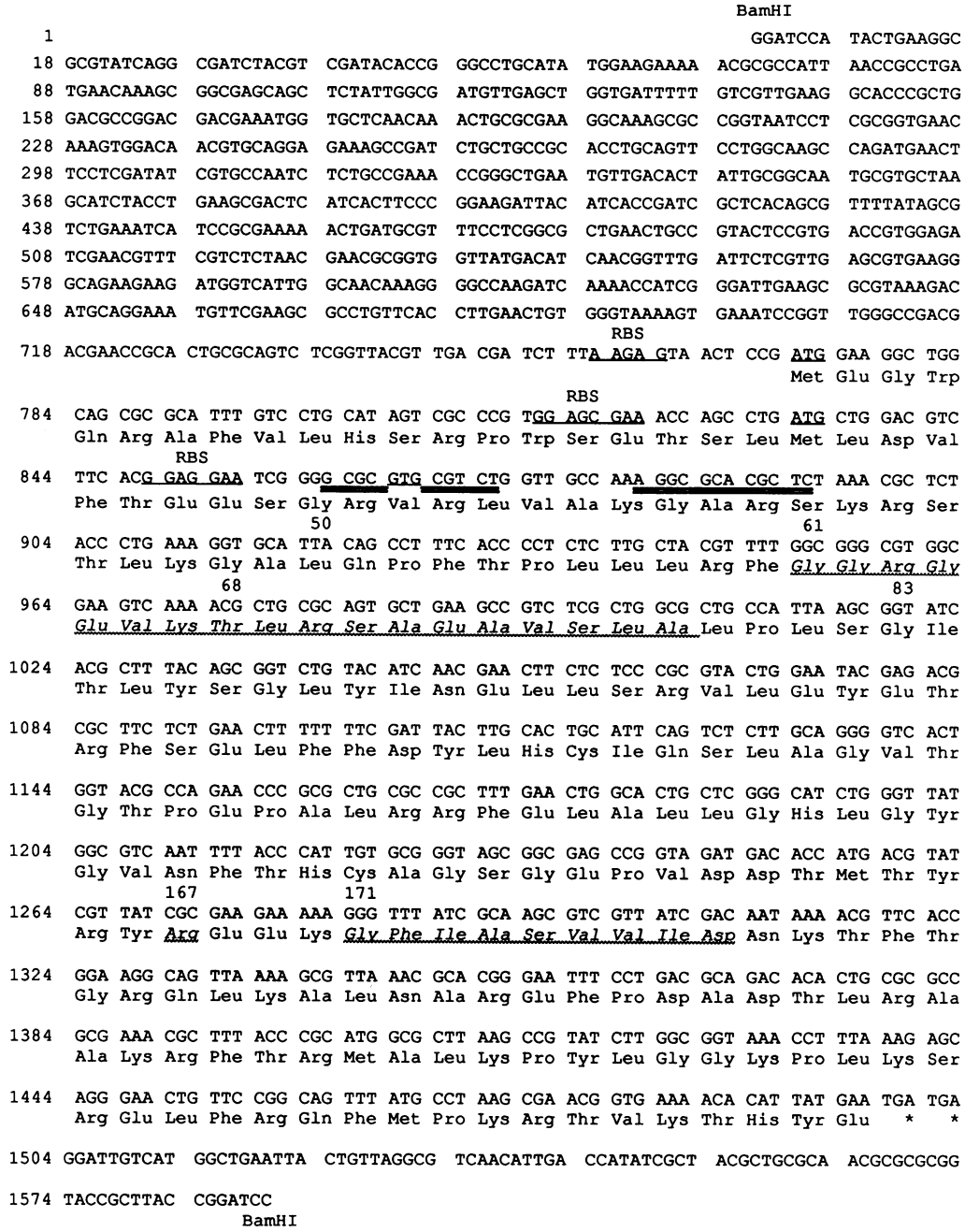

FIG. 6. DNA sequence of the 1.6-kb *Bam*HI fragment containing *recO*. The DNA sequence of the 1.6-kb *Bam*HI fragment is illustrated in the 5′-to-3′ direction along with the translated *recO* ORF starting with the first possible initiation codon. The three potential ATG and GTG start codons and their adjacent putative ribosome-binding sites (RBS) are underlined with a single underline. The 12-bp palindrome covering the putative GTG start is, except for the GTG, indicated by double underline. The regions covering nucleotides 952 to 1005 and 1282 to 1308 are indicated by italics and wavy underlines and correspond to the regions 1 plus 2 and region 3 of the putative mononucleotide-binding fold. The numbers above individual amino acids refer to the amino acid positions discussed in the text and in Fig. 7. The two TGA stop codons are indicated by asterisks (*).

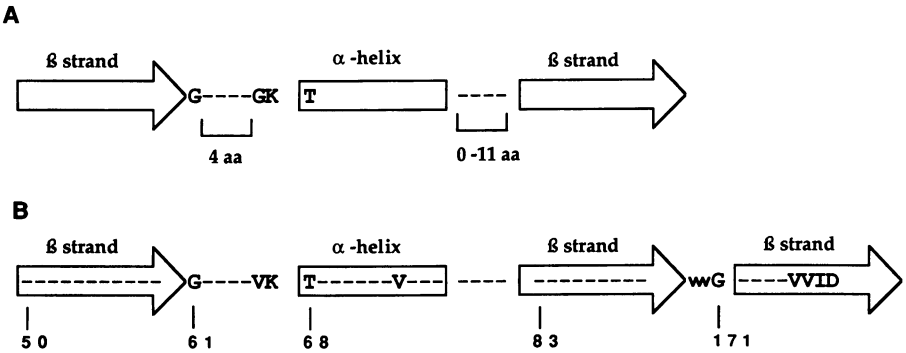

FIG. 7. Comparison of the mononucleotide-binding fold primary sequence-secondary structure descriptor with the predicted structure of *recO*. (A) Primary sequence-secondary structure descriptor described by Bradley et al. (6). (B) Secondary structure prediction of amino acid positions 50 to 90 aligned with the descriptor. The structural prediction is also indicated for amino acid positions 171 to 179, which match with mononucleotide-binding fold segment 3 as indicated in the text (19, 50). Individual amino acids in panel B are indicated by the standard one-letter code or by a "–" to designate a fixed number of amino acids. The numbers refer to the amino acid positions in the *recO* ORF of Fig. 6 starting from the first putative initiation codon. Arrowheads in β strands indicate the carboxy-terminal ends. The output of the computer analysis leading to this structural prediction is available on request.

consistent with poor expression of the *recO* gene product, given the strength of the T7 promoter. Computer analysis of the putative RecO protein did not identify a previously characterized protein having significant homology with *recO*. This analysis has suggested that the RecO protein might contain a nucleotide-binding fold. However, it will require extensive biochemical analysis to verify this and to identify a biochemical function for *recO*. Our efforts are now directed at overproducing and purifying the RecO protein to begin identifying its biochemical activities.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant GM26017 from the National Institutes of Health and by American Cancer Society faculty research award FRA-271 to R.K. S.T.L. was supported by a postdoctoral fellowship from the American Cancer Society, Inc., Massachusetts Division.

We thank M. Bradley for helpful discussions about mononucleotide-binding folds, C. A. Luisi-Deluca for helpful comments about the manuscript, and H. Takiff, S.-M. Chen, and D. Court for communicating their unpublished results on the *rnc* operon.

LITERATURE CITED

1. **Ahnn, J., P. E. March, H. E. Takiff, and M. Inouye.** 1986. A GTP-binding protein of *Escherichia coli* has homology to yeast RAS proteins. Proc. Natl. Acad. Sci. USA **83**:8849–8853.
2. **Amundsen, S. K., A. F. Taylor, A. M. Chaudhury, and G. R. Smith.** 1986. *recD*: the gene for an essential third subunit of exonuclease V. Proc. Natl. Acad. Sci. USA **83**:5558–5562.
3. **Bachmann, B. J.** 1972. Pedigrees of some mutant strains of *Escherichia coli*. Bacteriol. Rev. **36**:525–557.
4. **Biek, D. P., and S. N. Cohen.** 1986. Identification and characterization of *recD*, a gene affecting plasmid maintenance and recombination in *Escherichia coli*. J. Bacteriol. **167**:594–603.
5. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
6. **Bradley, M. K., T. F. Smith, R. H. Lathrop, D. M. Livingston, and T. A. Webster.** 1987. Consensus topology in the ATP binding site of the simian virus 40 and polyomavirus large tumor antigens. Proc. Natl. Acad. Sci. USA **84**:4026–4030.
7. **Chou, P. Y., and G. S. Fasman.** 1978. Emperical predictions of protein conformation. Annu. Rev. Biochem. **47**:251–276.
8. **Church, G. M., and W. Gilbert.** 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA **81**:1991–1995.
9. **Clark, A. J.** 1963. Genetic analysis of a "double male" strain of *Escherichia coli* K-12. Genetics **48**:105–120.
10. **Clark, A. J.** 1973. Recombination-deficient mutants of *E. coli* and other bacteria. Annu. Rev. Genet. **7**:67–86.
11. **Clarke, L., and J. Carbon.** 1976. A colony bank containing synthetic ColEI hybrid plasmids representative of the entire *E. coli* genome. Cell **9**:91–99.
12. **Clark, A. J., and A. D. Margulies.** 1965. Isolation and characterization of recombination-deficient mutants of *Escherichia coli* K12. Proc. Natl. Acad. Sci. USA **53**:451–459.
13. **Dale, R. M. K., 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 rDNA. Plasmid **13**:31–40.
14. **Date, T., and W. Wickner.** 1981. Isolation of the *Escherichia coli* leader peptidase and effects of leader peptidase overproduction *in vivo*. Proc. Natl. Acad. Sci. USA **78**:6106–6110.
15. **Dugaiczyk, A., H. Boyer, and H. Goodman.** 1975. Ligation of *Eco*RI endonuclease-generated DNA fragments into linear and circular structures. J. Mol. Biol. **96**:171–184.
16. **Emerson, P. T., and P. Howard-Flanders.** 1967. Cotransduction with *thy* of a gene required for genetic recombination in *Escherichia coli*. J. Bacteriol. **93**:1729–1731.
17. **Feinberg, A. P., and B. Vogelstein.** 1983. A new technique for radiolabeling DNA restriction fragments to high specific activity. Anal. Biochem. **132**:6–13.
18. **Fishel, R. A., A. A. James, and R. Kolodner.** 1981. *recA*-independent general genetic recombination of plasmids. Nature (London) **294**:184–186.
19. **Fry, D. C., S. A. Kuby, and A. S. Mildvan.** 1986. ATP-binding site of adenylate kinase: mechanistic implications of its homology with *ras*-encoded p21, F1-ATPase and other nucleotide binding proteins. Proc. Natl. Acad. Sci. USA **83**:907–911.
20. **Gillen, R., 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.
21. **Gold, L., and G. Stormo.** 1987. Translational initiation, p. 1302–1306. *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. 2. American Society for Microbiology, Washington, D.C.
22. **Guyer, M. S.** 1978. The γδ sequence of F is an insertion sequence. J. Mol. Biol. **126**:347–365.
23. **Holmes, D. S., and M. Quigley.** 1981. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. **141**: 193–196.
24. **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.

25. **Jorgensen, R. A., S. J. Rothstein, and W. S. Reznikoff.** 1979. A restriction enzyme cleavage map of Tn*5* and location of a gene encoding neomycin resistance. Mol. Gen. Genet. **177:**65–72.
26. **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.
27. **Laban, A., and A. Cohen.** 1981. Interplasmidic and intraplasmidic recombination in *Escherichia coli* K-12. Mol. Gen. Genet. **184:**200–207.
28. **laCour, T. F. M., J. Nyborg, S. Thirup, and B. F. C. Clark.** 1985. Structural details of the binding of guanosine diphosphates to elongation factor Tu from *E. coli* as studied by X-ray crystallography. EMBO J. **4:**2385–2388.
29. **Lloyd, R. G., F. E. Benson, and C. E. Shurvinton.** 1984. Effect of *ruv* mutations on recombination and DNA repair in *Escherichia coli*. Mol. Gen. Genet. **194:**303–309.
30. **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.
31. **Lloyd, R. G., S. M. Picksley, and C. Prescott.** 1983. Inducible expression of a gene specific to the RecF pathway for recombination in *Escherichia coli* K-12. Mol. Gen. Genet. **190:**162–167.
32. **Lloyd, R. G., M. C. Porton, and C. Buckman.** 1988. Effect of *recF*, *recJ*, *recN*, *recO* and *ruv* mutations on ultraviolet survival and genetic recombination in a *recD* strain of *Escherichia coli* K12. Mol. Gen. Genet. **212:**317–324.
33. **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.
34. **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.
35. **Maples, V. F., and S. R. Kushner.** 1982. DNA repair in *Escherichia coli*: identification of the *uvrD* gene product. Proc. Natl. Acad. Sci. USA **79:**5616–5620.
36. **March, P. E., and M. Inouye.** 1985. Characterization of the *lep* operon of *Escherichia coli*. J. Biol. Chem. **260:**7206–7213.
37. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101:**20–77.
38. **Nakayama, H., K. Nakayama, R. Nakayama, N. Irino, Y. Nakayama, and P. C. Hanawalt.** 1984. Isolation and genetic characterization of a thymineless death-resistant mutant of *Escherichia coli* K12: identification of a new mutation (*recQ1*) that blocks the RecF recombination pathway. Mol. Gen. Genet. **195:**474–480.
39. **Nakayama, K., N. Irino, and H. Nakayama.** 1985. The *recQ* gene of *Escherichia coli* K12: molecular cloning and isolation of insertion mutants. Mol. Gen. Genet. **200:**266–271.
40. **Otsuji, N., H. Iyehara, and U. Hideshima.** 1974. Isolation and characterization of an *Escherichia coli ruv* mutant which forms nonseptate filaments after low doses of ultraviolet light irradiation. J. Bacteriol. **117:**337–344.
41. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:**2444–2448.
42. **Personage, D., S. Wilke-Mounts, and A. E. Senior.** 1987. Directed mutagenesis of the β-subunit of F1-ATPase from *Escherichia coli*. J. Biol. Chem. **262:**8022–8026.
43. **Picksley, S. M., S. J. Morton, and R. G. Lloyd.** 1985. The *recN* gene of *Escherichia coli* K12: molecular analysis and identification of the gene product. Mol. Gen. Genet. **201:**301–307.
44. **Ralph, W., T. A. Webster, and T. F. Smith.** 1987. A modified Chow Fasman protein structure algorithm. Comput. Appl. Biosci. **3:**211–216.
45. **Sharp, P. M., and W.-H. Li.** 1987. The codon adaptation index—a measure of directional synonomous codon usage bias, and its potential applications. Nucleic Acids Res. **15:**1281–1295.
46. **Smith, G. R.** 1988. Homologous recombination in procaryotes. Microbiol. Rev. **52:**1–28.
47. **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.
48. **Takiff, H. E., S.-M. Chen, and D. L. Court.** 1989. Genetic analysis of the *rnc* operon of *Escherichia coli*. J. Bacteriol. **171:**2581–2590.
49. **Vapnek, D., N. Alton, C. Bassett, and S. R. Kushner.** 1976. Amplification in *Escherichia coli* of enzymes involved in genetic recombination: construction of hybrid plasmids carrying the structural gene for exonuclease I. Proc. Natl. Acad. Sci. USA **73:**3492–3496.
50. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay.** 1982. Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinase and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1:**945–951.
51. **Wensink, P. C., D. J. Finnegan, P. J. Donelson, and D. S. Hogness.** 1974. A system for mapping DNA sequences in the chromosomes of *Drosophila melanogaster*. Cell **3:**315–325.
52. **Willetts, N. S., A. J. Clark, and K. B. Low.** 1969. Genetic location of certain mutations conferring recombination deficiency in *Escherichia coli*. J. Bacteriol. **97:**244–249.