Molecular analysis of the recF gene of Escherichia coli
(nucleotide sequence/genetic recombination/DNA repair/DNA synthesis/maxicells)

MICHAEL A. BLANAR, STEVEN J. SANDLER, MARIA-EUGENIA ARMENGOD*, LLOYD W. REAM†,
AND ALVIN J. CLARK

Department of Molecular Biology, University of California, Berkeley, CA 94720

Communicated by Franklin W. Stahl, March 13, 1984

ABSTRACT We analyzed the nucleotide sequence of a 1.325-kilobase region of wild-type Escherichia coli containing a functional recF gene and six Tn3 mutations that inactivate recF. The analysis shows a potentially translatable reading frame of 1071 nucleotides, which is interrupted by all six insertions. A protein of 40.5 kilodaltons would result from translation of the open reading frame, and a radioactive band of protein of an apparent molecular weight of ~40 kilodaltons was seen by the maxicell method using a recF+ plasmid. Putative truncated peptides were seen when two recF::Tn3 mutant plasmids were used. Differential expression of dnaN and recF from a common promoter was noted. recF332::Tn3 was transferred to the chromosome where, in hemizygous condition, it produced UV sensitivity indistinguishable from that produced by two presumed recF point mutations.

The recF gene of Escherichia coli lies within a cluster of genes involved in DNA metabolism, the gene order being gyrB recF dnaN (1), gyrB has been shown to code for one of the subunits of DNA gyrase (2), dnaN determines the β subunit of DNA polymerase III holoenzyme (3), and dnaA encodes a 52-kilodalton (kDa) protein (4, 5) involved with the initiation of replication of the Escherichia coli chromosome (6).

The phenotype of a recF mutant indicates that recF also is involved in DNA metabolism and possibly with replication. UV irradiation of recF143 mutants produces a wide variety of effects (e.g., refs. 7–9) economically interpretable by the hypothesis that induction of the lexA regulon is partially inhibited (10). Such induction is thought to depend in part on formation of single-strand gaps because of incomplete replication of UV-damaged DNA and in part on association of recA, lexA, and cl gene products with these gaps (11). Therefore, it is plausible that the recF gene product is involved in some way with formation of these gaps or association of single-stranded DNA with proteins. recF143 strains also behave abnormally when exposed to treatments by other DNA-damaging agents (e.g., psoralen and visible light (12), thymine starvation (unpublished results), nalidixic acid (7, 13), and coumermycin (14)), leading one to the hypothesis that recF+ genotype is required for normal SOS inducibility (15).

Recently, evidence has appeared that recF143 substantially inhibits a variety of DNA metabolism events in unirradiated cells: plasmid recombination (16, 17) and one type of mismatch correction (ref. 18; unpublished results). recF143 also substantially reduces conjugal recombination (7, 14, 19) and λ phage–prophage recombination (20) in special mutant backgrounds. These effects are plausibly explained by hypothesizing that recF encodes an endonuclease (21) or a protein involved in DNA synthesis. Cloning of the recF region has been described (22), and in this paper we identify the recF gene product.

MATERIALS AND METHODS

Bacterial Strains. All bacterial strains used are derivatives of E. coli K-12. Gene symbols are those used by Bachmann (23). JC11033 was produced by transforming the Hfr strain JC158 (24) with the plasmid pML2.

Plasmids. pJC605 is a 7.9-kilobase (kb) plasmid derived by cloning a 3.6-kb fragment carrying recF+, called cJC1, with pBR322 (22). pMAB4 is a 3.6-kb plasmid derived by cloning the 1.3-kb EcoRI–Pvu II recF+ fragment of cJC1 into the 2.3-kb EcoRI–Pvu II bla + fragment of pBR322. pJC655 is a 7.7-kb de(bla)300 ampicillin-sensitive (Amp+) deletion mutant derivative of pJC605 (22). Tn3 mutant derivatives of pJC655 (12.7 kb) were isolated by induction with the conjugative plasmid pJC753. pJC753 is a Tn3 insertion mutant of R1 drd-19-K1, a plasmid that has lost Tn3 and Tn4 by deletion and was produced upon transposition of Tn3 from pJC752, a pSC101::Tn3 derivative (unpublished results). pML2 is a 13.5-kb plasmid containing a 6.7-kb EcoRI apH+ fragment cloned into ColEl1 (25). Construction of pSJS9, pSJS40, pSJS76, and pSJS77 using the temperature-dependent runaway-copy vector pBEU28 (26) will be published elsewhere.

Media, Materials, and Mutagenesis. Media, materials, and some methods have been described or cited by Ream and Clark (22). To detect transposon mutagenesis, JC13225, a strain carrying the target plasmid pJC655 and the conjugative plasmid pJC753 was used as donor, and JC9239 (recF143) was used as recipient in a 1-hr conjugation at 37°C. Colonies of Amp+ TetR (Amp- and tetracycline-resistant) [Met+ Trp+ Ile+ Val+] transconjugants appeared at frequencies of 3 × 10−4 and 1 × 10−3 per donor in two experiments while the conjugative plasmid was transmitted at 3 × 10−3 and 1 × 10−3, respectively. Six UV+ transconjugants were found among 159 tested, indicating that the plasmid-borne recF gene was mutant. Hydroxyurea (H3NOH) mutagenesis of plasmid DNA was performed in vitro as described by Humphreys et al. (27). Plasmid DNA (5 µg) was incubated with 1 M H3NOH (in 1 mM EDTA) at 72°C for 30 min in one experiment and at 60°C for 15 min in another.

Sequence Determination. Sequencing was done essentially by the procedure of Maxam and Gilbert (28). 3' termini were labeled with [γ-32P]ATP (Amersham; >5000 Ci/mmol; 1 Ci = 37 GBq) and polynucleotide kinase (P-L Biochemicals). 3' ends were labeled with the large fragment of DNA polymerase.

Abbreviations: Amp, ampicillin; Tet, tetracycline; Str, streptomycin; Kan, kanamycin; Sup, superscripts R, S, +, and −, resistant/resistance, sensitive/sensitivity, independent/independence, and dependent/independence; kb, kilobase(s).

*Present address: Instituto de Investigaciones Cientificas, Amadeo de Saboya 4, Valencia-10, Spain.
†Present address: Department of Biology, Indiana University, Bloomington, IN 47405.
ase (Klenow; Boehringer Mannheim) and α-32P-labeled deoxyribonucleoside triphosphates (Amersham; >3000 Ci/mmole). Maxicells. Labeling with [35S]methionine (Amersham; ≈1200 Ci/mmole) was performed essentially as described by Sancar et al. (29). Labeling was done at 42°C for 1 hr.

RESULTS

Tn3 and Hydroxylamine-Induced Mutations. In order to ascertain the limits of the recF gene, we generated and characterized six insertion mutations in the cloned recF gene. All the insertions resulted in the inability of the plasmid-borne recF gene to confer UV sensitivity to an otherwise UV-sensitive recF143 chromosomal mutant strain. Each of the resultant plasmids complemented the temperature-sensitive mutation dnaN608 (data not shown), indicating that none of these Tn3 insertions eliminated dnaN function. Restriction enzyme mapping with BamHI and HincII (30) was used to locate roughly and orient the transposon in each mutant insertion (data not shown). The six insertions occurred within a region of about 950 base pairs. Exact positions as determined by nucleotide sequencing are indicated in Fig. 1. As delimited by these Tn3 insertion mutations, the recF gene and any other DNA fragment necessary for expression must extend from at least position 41 to position 983 in Fig. 1.

In addition to the Tn3 insertion mutations, we also generated a nonconditional, H2NOH-induced recF mutation, recF289. We transformed a sup+ recF143 strain, JC11810, with H2NOH-treated DNA of pJC605 and selected for TetR. The resultant transformants were screened for sensitivity to 3.5 μg of nitrofurantoin per ml at 30°C and 42°C. Of 4500 transformants screened, 7 highly sensitive transformants were identified, and 1 of these, nitrofurantoin-sensitive at both temperatures, was found to contain a plasmid whose recF mutation could not be suppressed by various nonsense suppressors. This plasmid, pJC610, carries recF289 and complements dnaN608 (data not shown), demonstrating that it is dnaN.

Translocation of Plasmid-Carried recF Mutation. The plasmid-borne recF mutations were all detected in heterozygous condition with the chromosomal mutation recF143. To produce cells hemizygous for these mutations, we did the following: (i) transformed a recF+ dnaA508 (42°C) strain with the appropriate plasmid; (ii) selected and purified two spontaneous 42°C mutant strains of the recF::Tn3 region; (iii) incubated an inoculum from each clone at 30°C in the absence of antibiotics for 24 hr, followed by subculturing at 30°C for an additional 24 hr before screening individual clones for 42°C and for UV sensitivity, to detect plasmid excision and either a hemizygous or homozygous recF mutation; and (iv) used the resultant isolates as donors in P1 transduction of the nearby tnaA300::Tn10 marker, selecting TetR 42°C transductants and screening for UV sensitivity, to isolate those that inherited the recipient’s dnaA+ and the donor’s recF mutant allele. In steps iii and iv, we were able to isolate respectively dnaA508 and dnaA+ strains that were hemizygous for recF289 and recF332:recF143. Translocation was verified for the latter mutation by the Southern transfer method (31) using EcoRI digests of chromosomal DNA and pJC605 DNA as a labeled probe (data not shown). Strains hemizygous for recF289 and recF332:recF143 were found to be as UV-sensitive as the hemizygous recF143 strain, JC11803 (data not shown). When crossed with donor JC11033, each produced essentially the same frequency of Thr+ Leu+ [StrR (streptomycin-resistant) Ser+] and KanR (kanamycin-resistant) [StrR Ser+] transconjugants as JC11803, both with chromosomal and plasmid pML2 markers, respectively (data not shown).

As the recF::Tn3 mutations had been detected by their failure to complement recF143, we tested their ability to complement recF289. Strains carrying plasmids with the Tn3 insertions at the extremes of the recF region and with recF289 on the chromosome were UV-sensitive, whereas those with recF143 on the chromosome were UV-resistant.

DNA Sequence Analysis of the recF Gene. The nucleotide sequence of a 1.325-kb Msp I–EcoRI fragment, which has been shown to contain a completely functional recF gene (ref. 22; unpublished results) is presented in Fig. 1. It includes two long open reading frames. The shorter consists of 210 nucleotides and begins at the Msp I sequence. It encodes the carboxyl-terminal 70 amino acids of dnaN protein (32) and ends at a TAA that overlaps, by one nucleotide, the first ATG of the longer open reading frame. Beginning at this ATG, the latter consists of 1071 nucleotides, which could encode a 357-amino-acid protein of 40,519 daltons. The sequence A-A-T-G-A-G is centered 9 nucleotides prior to the A of the first ATG. Five of the 6 nucleotides of this sequence (underlined) are complementary with the 3' end of E. coli 16S RNA. Since the average ribosome-binding site has a 4.8-nucleotide complementarity centered 9.8 nucleotides from the A of an initiating ATG (33, 34), we hypothesize that the A-A-T-G-A-G sequence is a suitable ribosome-binding sequence. Thus, the longer open reading frame is potentially translatable and may be recF.

Maxicell Analysis of the recF Protein. We transformed pSSJ9, pSJS40, pSJS76, and pSJS77 into an appropriate UV-sensitive recF143 recipient mutant strain and screened the KanR transformants for recF332 complementation. The strains harboring pSSJ9 and pSJS40 yielded UV-sensitive colonies at 30°C, whereas those containing pSJS76 and pSJS77 remained UV-resistant (data not shown).

Using [35S]methionine-labeled extracts from maxicells carrying these four plasmids, we obtained autoradiograms shown in Fig. 2b. Temperature induction of pSSJ9 (recF332) for 1 hr results in the expression of four proteins with molecular masses of 59, 47, 43, and 29 kDa. Identities of these proteins have been confirmed by deletion analysis, and this evidence will be presented elsewhere. The 59-kDa protein is the result of a fusion of 93 codons of λ phage exo (35) with 446 codons of dnaA (36). The 47-kDa protein has not been identified, but it seems to be vector-encoded. The 43- and 29-kDa proteins are the dnaN and λ phage bet gene products, respectively (35). λ phage N protein was not observed in these experiments. It is of interest to note that all genes downstream of the temperature-induced λPl and Pr promoters produce substantial amounts of proteins, but that the recF gene product is not observed. We were successful in identifying the recF gene product only upon construction of deletion derivatives of pSSJ9.

From the DNA sequence analysis, we expected pSJS40 (recF332) to encode proteins of 7.2 and 43 kDa in addition to the 40.5-kDa recF protein. The smaller protein would result from a fusion of the 55 amino-terminal codons of the N phage N (35) to the 10 carboxyl-terminal codons of dnaN and would not be seen under the conditions we used. The larger protein would result from translation initiating at an ATG in the λ phage N gene, 23 codons upstream of the recF initiating codon. As shown in Fig. 2b, we observed two proteins of approximately the expected weights upon incubation at 42°C to initiate transcription from λP and Pr. With the Tn3 mutant plasmid pJS76, we saw neither the 42- nor 40-kDa bands. Instead a new band appeared at 20.5 kDa. This is the size expected of the product of fusion between recF and the codons of Tn3 in the appropriate reading frame. A larger peptide (24 kDa) with the additional 23 amino acids is probably obscured by the heavily labeled cl and aphA gene products. With plasmid pSJS77, we saw proteins of apparent molecular masses 36 kDa and 40 kDa as expected from the recF and upstream initiation codons, respectively.
4624 Biochemistry: Blanar et al.


-212

CGG GAA CAG GAA GAA GGA GAG ATC TCC GAC
CGG GAA CAG GAA GGA GAG ATC TCC GAC

Terminus of proximal to the 1103. The recF gene begins at position 1103. Restriction endonuclease recognition sequences are underlined. A, Positions of insertion of the six Tn3 mutations. Tn3 elements were found in both orientations. In recF334 and recF336, Tn3 is oriented so that the bla gene is proximal to the gyrB side of recF, whereas bla is distal to the gyrB side in the insertion mutations recF332, recF333, recF335, and recF337. Description of the sequencing strategy is available upon request.

**Fig. 1.** DNA sequence of the recF gene and adjacent regions. The recF gene and the deduced amino acid sequence begin at position +1 and extend to the TAA termination codon at position 1074. The carboxyl-terminal sequence of dnaN extends from position −212 to +1, and the amino terminus of the gyrB gene begins at position 1103. Restriction endonuclease recognition sequences are underlined. A, Positions of insertion of the six Tn3 mutations. Tn3 elements were found in both orientations. In recF334 and recF336, Tn3 is oriented so that the bla gene is proximal to the gyrB side of recF, whereas bla is distal to the gyrB side in the insertion mutations recF332, recF333, recF335, and recF337. Description of the sequencing strategy is available upon request.
DISCUSSION

Almost all of the results described are consistent with the hypothesis that most of an open reading frame between dnaN and gyrB is recF and that the recF gene product is a 40.5-kDa protein. The exception is our observance of RecF- phenotype at 30°C (i.e., UV8) without detecting recF gene product by maxicell analysis at this temperature. At present, we think there is too little recF gene product produced at 30°C to be detectable by the maxicell method. We were unable to determine the phenotype at 42°C because the plasmid vector we used is lethal at that temperature.

An analysis of the amino acid composition shows that the presumptive recF gene product contains 30.3% charged amino acids and has an isoelectric point of ~7.6. A comparison of the primary structure of the recF gene product with various DNA binding proteins has not revealed any significant similarity.

Observations have indicated that recF and dnaN gene products are produced in substantially different amounts from a common promoter. In an earlier paper, Ream and Clark (22) were easily able to identify the dnaN product in minicells but were unable to identify a recF gene product. Other reports (4, 37) have identified only the dnaA and dnaN gene products from DNA of this region, although a complete recF gene was present in these studies.

There are three reasons that could explain this differential expression. First, it is possible that transcription terminates early in the recF sequence. A region of dyad symmetry begins at position 21 in Fig. 1 and ends at position 55 with a stem length of 13 nucleotide pairs and a loop distance of 9 nucleotides. This sequence may be a transcriptional pause sequence and a potential p-dependent terminator (38). Since UV irradiation suppresses polarity effects in the gal operon by reducing p activity (39, 40), the amount of recF gene product might increase after irradiation without having a lexA gene product binding site nearby (unpublished results). Expression of a putative pBEU28 gene downstream from recF and detectable by its 47-kDa protein product would not be affected by such a terminator perhaps because of the intervening gyrB promoter at the carboxyl-terminal end of recF (41). Second, it is possible that a recF transcript is cleaved by RNase action. A second region of dyad symmetry extends from position 305 through 338 (Fig. 1). The potential stem of an RNA structure from this region contains the sequence 5'-A-A-G-G-U-C, which is found in all symmetrical sequences of T7 early mRNAs cleaved by RNase III (42). Each of these two possibilities would result in truncated mRNAs, incapable of encoding a complete recF protein. Third, it is possible that translation of recF mRNA may be slow. The recF reading frame has a high percentage (43.3%) of nonoptimal codons (43), whereas the adjacent 70 codons of the carboxyl terminus of dnaN contain only 22% nonoptimal codons. All of these mechanisms may operate to prevent toxic effects due to overproduction of recF gene product.
An interesting feature of the presented sequence is the one-base overlap of recF and dnaN. Overlaps of one nucleotide are seen at the junctions of other genes in E. coli (44, 45) and have been hypothesized (45, 46) to couple translation of adjacent reading frames. This feature, with the neutralization of the limiting factors we have described, should allow for the overproduction and further characterization of the recF gene product.

We are very grateful to the following people: M. Gellert and H. Ohmori, who sent us sequence information on the recF and dnaN regions, respectively, prior to publication; Steven Polevoi, who performed the last step in construction of pSJ76 and pSJ77; M. Cherry, C. Gautier, W. Gish, M. Gouy, and H. Martinez for their invaluable assistance with the computer analyses; J. Nowak, for allowing us to use his construct, pJC753; and L. Margossian, who determined the UV sensitivity of recF289 and recF332: Tn3 hemizygous strains and performed the complementation experiments. M.-E.A. was supported by a fellowship from the Fundacion Juan March. This work was supported by Grant NP-237 from the American Cancer Society and by Grant AI 05731 from the National Institute of Allergy and Infectious Diseases.