© 1992 Oxford University Press

Nucleic Acids Research, Vol. 20, No. 4 839–845

Sequence and complementation analysis of *recF* genes from *Escherichia coli*, *Salmonella typhimurium*, *Pseudomonas putida* and *Bacillus subtilis*: evidence for an essential phosphate binding loop

Steven J.Sandler, Bryce Chackerian*, John T.Li+ and Alvin J.Clark
Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA

Received October 21, 1991; Revised and Accepted January 13, 1992 EMBL accession nos X62504, X62505

ABSTRACT

We have compared the *recF* genes from *Escherichia coli* K-12, *Salmonella typhimurium*, *Pseudomonas putida*, and *Bacillus subtilis* at the DNA and amino acid sequence levels. To do this we determined the complete nucleotide sequence of the *recF* gene from *Salmonella typhimurium* and we completed the nucleotide sequence of *recF* gene from *Pseudomonas putida* begun by Fujita *et al.* (1). We found that the RecF proteins encoded by these two genes contain respectively 92% and 38% amino acid identity with the *E. coli* RecF protein. Additionally, we have found that the *S. typhimurium* and *P. putida recF* genes will complement an *E. coli recF* mutant, but the *recF* gene from *Bacillus subtilis* [showing about 20% identity with *E. coli* (2)] will not. Amino acid sequence alignment of the four proteins identified four highly conserved regions. Two of these regions are part of a putative phosphate binding loop. In one region (position 36), we changed the lysine codon (which is essential for ATPase, GTPase and kinase activity in other proteins having this phosphate binding loop) to an arginine codon. We then tested this mutation (*recF4101*) on a multicopy plasmid for its ability to complement a *recF* chromosomal mutation and on the *E. coli* chromosome for its effect on senstivity to UV irradiation. The strain with *recF4101* on its chromosome is as sensitive as a null *recF* mutant strain. The strain with the plasmid-borne mutant allele is however more UV resistant than the null mutant strain. We conclude that lysine-36 and possibly a phosphate binding loop is essential for full *recF* activity. Lastly we made two chimeric *recF* genes by exchanging the amino terminal 48 amino acids of the *S. typhimurium* and *E. coli recF* genes. Both chimerias could complement *E. coli* chromosomal *recF* mutations.

INTRODUCTION

The *recF* gene of *E. coli* K-12 was first detected as an UV-sensitive/recombination deficient mutant in a *recB21 recC22 sbcB15 sbcC101* background (3). This was the first gene mapped of ten which are needed for the RecF Pathway of recombination [(4, 5) and S. Lovett, unpublished results]. This pathway of recombination has been associated with DNA repair and called recombinational repair, daughter strand gap repair or post replication repair. Single mutants in most RecF pathway genes (i.e.*recF,O,N,Q,R, ruvA,B*) but not *recJ* (5) are sensitive to UV irradiation. *recF* single mutants are conjugally recombination proficient (3), deficient in the ability to do plasmid recombination (6), sensitive to UV irradiation (and other DNA damaging agents), deficient in a *mutLHS* independent pathway of mismatch repair (7), have an attenuated ability to induce the SOS (8) and adaptive responses (9) and are UV-immutable for ssDNA phages (10). Even with these diverse phenotypes, the *recF* gene has never been identified in a screen for any phenotype other than recombination deficiency.

Due to the pleiotropic effects of *recF* mutations, we hypothesize that *recF*, like *recA* (11) has dual functions in the cell. One function is enzymatic, that for recombinational repair; the other is regulatory, that for induction of the SOS Response. This hypothesis is supported by indirect suppressor studies of two different *recA* mutants, *recA441* (12) and *recA803* (13, 14). Both mutant genes partially suppress the UV sensitive defect of a *recF* mutant. *recA441* is thought to compensate for *recF*'s role in induction of the SOS Response by being better able than wildtype RecA protein to compete with single-stranded DNA binding protein for ssDNA in the cell (15). *recA803* is thought to compensate for *recF*'s role in recombinational repair by being better able than wildtype RecA protein to utilize ssDNA in the cell for recombination (13, 16). Recently Griffin and Kolodner (17) and independently by Madiraju and Clark (18) purified RecF protein to homogeneity. Both studies showed it to be a single-stranded DNA binding protein.

Present addresses: *University of Washington, School of Medicine, Department of Microbiology, Seattle, WA 98195 and +Tulane University, School of Medicine, New Orleans, LA 70112, USA

The *recF* gene is located at 83 minutes on the *E. coli* chromosome in a group of genes encoding enzymes which metabolize DNA (19). The order is *dnaA, dnaN, recF* and *gyrB*. The above genes are transcribed in the same direction and have their own promoters which can be differentially regulated (20, 21, 22, 23). Armengod *et al.* (23) have shown that *dnaN* and *recF* may function as an operon since mutations in *dnaN*, which do not affect the *recF* promoter, decrease expression of *recF-galK* transcriptional fusions. *recF* and *dnaN* overlap in their start and stop codons respectively and are thought to be translationally coupled (24). Recently Sandler and Clark (25) have shown that *recF* is subject to translational control which inhibits over-expression 100 fold.

Mutants in a *recF*-like gene have been isolated in only one other organism: *Bacillus subtilis* (26). These mutations cause increased sensitivity to UV irradiation, decreased recombination frequency (26) and attenuated induction of the SOS Response (27). DNA sequence analysis of the *dnaA* to *gyrB* region of the *B. subtilis* chromosome shows that the sizes and arrangement of the open reading frames are similar to the sizes and arrangement of the *dnaA, dnaN, recF* and *gyrB* genes from *E. coli* (Figure 1) (2, 28, 29). The open reading frame in the *B. subtilis* region, corresponding in position to that position of the *E. coli recF* gene, codes for a 37 kDa protein which has 20% identity with the *E. coli recF* protein. Recently, the start codon for the *B. subtilis* RecF protein was correctly assigned to be TTG at nucleotide 1640 in the sequence of Moriya (2) to yield a slightly longer open reading frame (27) than orginally reported (2). Maxicell anaylsis shows that the *B. subtilis recF* gene produces a 38 kDa protein (30). Several other genes similar to recombination genes of *E. coli* have also been identified in *B. subtilis* (26).

Much work has been done in characterizing the *dnaA* genes (and operons) from several organisms at the genetic, DNA, and protein levels (29, 31). Comparison showed that the DnaA proteins could be divided into two highly conserved domains and a highly variable region. In order to gain more information on the structure of the *recF* proteins and which parts of the protein may be important for function, we have taken a similar appproach to that used with *dnaA* and isolated and compared *recF* genes from several organisms.

MATERIALS AND METHODS

All bacterial strains used in this work were derivatives of *E. coli* K-12. AB1157 has the following partial genotype: *thr-1 leu-6 thi-1 lacY1 ara-14 xyl-5 mtl-1 proA2 argE3 rpsL31 tsx-33* and *supE44*. JC15359 is AB1157 with *recF349* (32). *recF349* is a deletion mutation which removes nucleotides 60 to 138 of the *recF* gene (24). JC10937 is AB1157 with *recF143 supE*$^+$ λ^+ lysogen. HC120 contains *dnaN59* which is a temperature conditional allele (33).

Plasmid Constructions

The plasmids used in this work are listed in Table 1. Unless otherwise stated, all plasmid constructions were made by restricting DNA with appropriate enzymes, fractionating the DNA in a 1% low melting agarose gel by electrophoresis, excising appropriate regions of the gel, mixing the fragments and treating them with ligase. The DNA mixtures where then used to transform competent cells.

A 3.8 kb DNA fragment possibly carrying the *recF* gene from *Salmonella typhimurium* (*Sty recF*) was available as plasmid pFF48 (28). To make an isogenic *E. coli recF* counterpart, we first restricted both pSJS9 (24) and pUC18 (34) with *Eco* RI. Then fragments were isolated and treated with ligase. The mixture was used to transform HC120. Temperature resistant Amp^R transformants were selected and characterized. One was saved with the same fragment orientation as pFF48 and called pSJS521. To mutate the putative *Sty recF* we restricted pFF48 with *Nco* I and then treated the digest with DNA Polymerase I fragment (Klenow Fragment) in the presence of dNTPs to blunt the ends. The mixture was then treated with DNA ligase and used to transform JC15359 to ampicillin resistance (Amp^R). This resulting plasmid was called pSJS520 and carries *Sty recF4102*.

The chimeric DNA constructions pSJS525 and pSJS526 were made by exchanging *Nco* I to *Pst* I fragments between pFF48 and pSJS521. This was done by restricting both pSJS521 and pFF48 with *Nco* I and *Pst* I. Fragments were isolated and treated with ligase. The mixture was used to transform JC15359 to Amp^R. One clone named pSJS525 contains the *S. typhimurium Pst* I to *Nco* I fragment which includes part of *dnaA, dnaN*, and the first 48 codons of *recF* fused to the remaining 309 codons of the *E. coli recF* gene contained on the *Nco* I to *Pst* I fragment. The complementary chimera, pSJS526, contains *dnaA, dnaN* and first 48 codons of *recF* sequences from *E. coli* and the remainder of the *recF* gene from *S. typhimurium*.

pSJS626 was constructed by restricting λgt10 *Eco* RI 5.2 (1) and pUC120 with *Sph* I and *Sal* I. A 1.6 kb fragment containing the *recF* gene from *P. putida* and the vector fragment from pUC120 (Viera, 1987 # 17) were isolated as described above. The DNAs were mixed, treated with ligase and used to transform

Table I. Plasmid List

Plasmid	Description	Ref.
pFF48	pUC18 with 3.8 Kb *Eco* RI fragment with *dnaA* to *recF* region from *Salmonella typhimurium*	(28)
pRG1	pACYC derivative containing the *lacI*^{q} gene	(16)
pSJS9	derivative with *E. coli recF* gene on a 3.8 kb DNA fragment	(24)
pSJS159	A pUC118 derivative containing the *recF* gene downstream of the lambda *cIAt2* gene and P_L promoter	(25)
pSJS520	pFF48 with *recF4102*	This work
pSJS521	pUC18 with 3.8 kb *Eco* RI fragment with *dnaA* to *recF* region from *E. coli*	This work
pSJS525	pUC18 with 3.8 kb *Eco* RI fragment with *dnaA* to *recF* region containing *recF* chimera *E. coli-S. typhimurium*	This work
pSJS526	pUC18 with 3.8 kb *Eco* RI fragment with *dnaA* to *recF* region containing *recF* chimera *S. typhimurium-E. coli*	This work
pSJS626	pUC120 clone with a 1.6 kb region containing the *recF* gene from *P. putida*.	This work
pSJS630	pUC118 clone with a 1.4 kb region containing the *recF* gene from *P. putida*.	This work
pSJS641	pSJS630 with *recF4103*	This work
pSJS684	Like pSJS521 but using pUC118 instead of pUC18	This work
pSJS703	pSJS684 but with *recF4101*	This work
pSJS711	pBR322 derivative containing the 3.8 kb *Eco* RI fragment *dnaA* to *recF* with *recF4101*	This work
pSM2003	pBR322 derivative with 4 kb. Sal I fragment containing the *B. subtilis dnaA* to *recF* region.	(2)
pUC18	high multi-copy number vector containing *bla*	(34)
pUC118	like pUC18 but with M13 *ori*	(34)
pUC120	high multi-copy number vector containing *bla* and M13 *ori* an additional *Nco* I site in the Multipurpose Cloning Site.	(34)

a derivative of AB1157. Amp^R transformants were selected and analyzed by restriction enzyme cleavage pattern. One clone receiving the *P. putida recF* gene (*Ppu recF*) was saved and called pSJS626. This clone contains the *recF* gene in the orientation so that it is not transcribed by the *lac* promoter contained on pUC120. To make a clone in which *Ppu recF* would be transcribed by the *lac* promoter, we used pUC118 (Viera, 1987 # 17) as the vector. This was done by restricting pSJS626 with *Pst* I and then treating the digest with Klenow fragment in the presence of dNTPs to blunt the ends. The DNA was further restricted with *Sal* I. pUC118 was first restricted with *Eco* RI, similarly treated to blunt the ends and then was further cleaved with *Sal* I. The appropriate fragments from the two digested plasmids were isolated, mixed, treated with ligase and used to transform AB1157 containing pRG1. This plasmid contains a copy of the *lacI^q* gene. This was done to ensure that the *lac* promoter on pUC120 which would be oriented towards the *recF* gene would be repressed. Amp^R transformants were selected and screened for the proper restriction enzyme cleavage pattern. One clone was found and called pSJS630. pSJS641 contains an internal deletion of the *P. putida recF* gene. It was made from pSJS630 by restricting with *Sst* II and treating with ligase. The resulting plasmid, pSJS641, carries *Ppu recF4103*.

pSJS684 was constructed in a similar manner to pSJS521 except that we used pUC118 instead of pUC18. Site directed-mutagenesis was performed using the oligonucleotide prSJS56 5′ GTGCCAACGGCAGCGGCCGTACCAGCGTGCTG 3′ (an *Eag* I site is underlined) and pSJS684 as the template plasmid as previously described (25). The resulting plasmid has a mutant *recF* gene with a single change in the predicted amino acid sequence from a lysine to arginine at position 36. The presence of the *Eag* I site was used to confirm the sequence of this mutation. This new allele of *recF* is called *recF4101*. The *Eco* RI fragment containing this mutation was then recloned with vector pBR322 (pSJS711) and then transferred to the chromosome using the method of integrative suppression (24). Essentially this entails allowing a plasmid carrying *recF4101* first to integrate with the chromosome of a *dnaA508 tnaA300*::Tn*10* strain and then to excise. The original strain was heterozygous *recF4101/recF$^+$* and was UV^R. UV^S homozygotes were detected by replica plating. One UV^S strain was used as a donor to cotransduce *recF4101* with *tnaA300*::Tn*10* into AB1157. *dnaA$^+$* (heat resistant) transductants both UV^S (JC18146) and UV^R (JC18148) were obtained from this cross. Since *recF4101* also introduces an *Eag* I site , we were able to verify by Southern Blot analysis (data not shown) that this mutation was causing the UV^S phenotype since the UV^S clone and not the UV^R clone had inherited an *Eag* I site.

UV sensitivity tests

Quantitative UV survival curves were done by growing the cells in 10 ml Luria Broth plus 50 mgm/ml ampillicin at 37°C to mid log phase. The cells were then collected by centrifugation, washed once in 56/2 buffer (35), resuspended in 5 mls of 56/2 and then irradiated for the desired length of time at 1 joule/sec/m^2. Samples were taken for viable cell counts.

DNA sequencing, alignments of proteins, and computer analysis

Clones, primers, and sequencing strategy are available upon request to SJS. Sequence alignments were done using the computer program package by Smith and Smith (36). The alignment and parsimony analysis was done by Temple Smith. The accession numbers for the *S. typhimurium* and *P. putida recF* genes in the EMBL Data Library are X62505 and X62504 respectively.

RESULTS

Of the four bacterial *recF* genes which will be compared in this paper, three originate from *E. coli*, *S. typhimurium*, and *P. putida* which are members of the purple bacteria group (37). The fourth *recF* gene comes from *B. subtilis*, a more distantly related eubacterium and member of the gram positive group (37). Since we knew the arrangement of the genes surrounding *recF* in *E. coli* and *B. subtilis*, the two most distantly related organisms, we reasoned that other closely related organisms may share the same arrangement. Therefore we obtained clones of *P. putida* and *S. typhimurium* containing the *dnaA* and *dnaN* genes and looked downstream for the *recF* gene using two criteria. The first was whether or not the clone could complement *E. coli recF* mutants. Since this criterion assumes that these foreign genes are correctly expressed and make the same interactions with other *E. coli* proteins that its homologue makes, we also applied a sequence criterion. We sequenced the region of DNA where the *recF* gene should be located and looked for an open reading frame which could encode a protein which would be similar to *recF*.

Identification of clones which may contain *recF* homologues in *S. typhimurium* and *P. putida*

In both cases, clones from *S.typhimurium* and *P. putida* which encoded the *dnaA* and at least part of the *dnaN* genes had been previously identified and sequenced. In the case of *S. typhimurium*, a clone called pFF48 contained a 3.8 kb *Eco* RI fragment known to encode part of the *dnaA* gene and the complete *dnaN* gene (28). Since a similarly sized fragment from *E. coli* encodes similar portions of the two genes and the entire *recF* gene (38), we hypothesized that it may encode the *S. typhimurium recF* gene. In the case of *P. putida*, previous work had revealed an open reading frame encoding the N-terminus of a putative protein which was highly similar to the *recF* protein of *E. coli*

Subunit for DNA gyrase	Recombination and DNA Repair		Processivity Factor for Pol III		Initiation of DNA Replication	
gyrB	*recF*		*dnaN*		*dnaA*	
Name of Species	gene (bp)	protein (aa)	gene (bp)	protein (aa)	gene (bp)	protein (aa)
E.coli	1074	357	1101	366	1403	467
S. typhimurium	1068	355	nd	nd	1403	467
P. putida	1098	365	1104	367	1521	506
B. subtilis	1113	370	1133	378	1337	446

Figure 1. General diagrams of the *recF* region of the chromosomes of *E. coli*, *P. putida*, *B. subtilis* and *S. typhimurium*. Columns are headed by the name of the gene. The order of the columns reflects the genetic map order. Genes whose names are surrounded by solid lines are described are the table. Dashed lines indicate flanking genes not described. Above the column headings is listed the process and/or enzymes in which the gene product functions. No attempt is made to indicate the untranslated space between or the translational overlaps of the *recF*, *dnaN*, and *dnaA* genes. Sizes of the genes have been determined in this work or have been taken from the literature. nd is not determined.

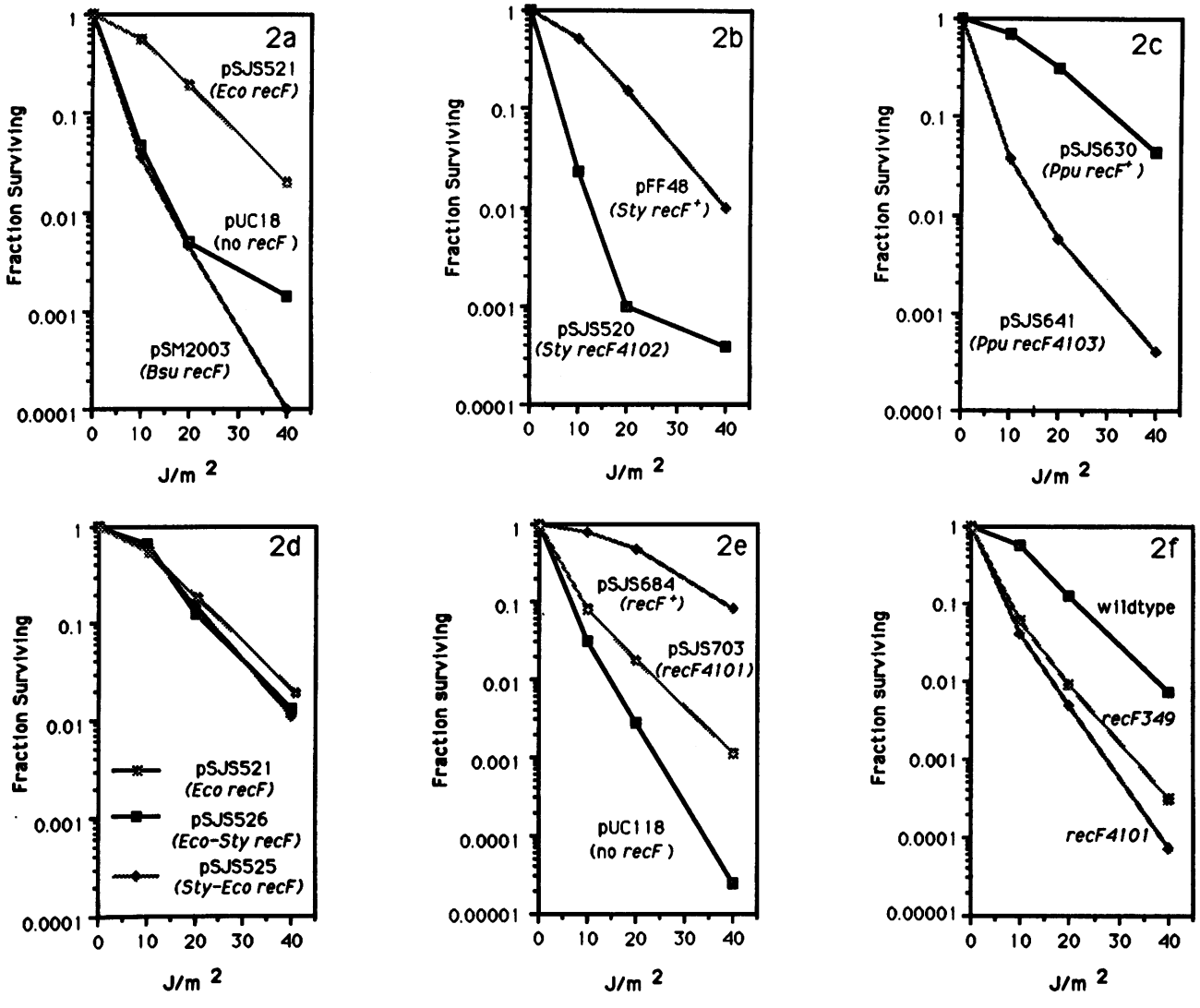

Figure 2. UV survival curves of JC15359 containing the indicated plasmids. 2a shows plasmids pUC18 (no *recF*), pSJS521 (*Eco recF*$^+$) and pSM2003 (*Bsu recF*$^+$). 2b shows pFF48 (*Sty recF*$^+$) and pSJS520 (*Sty recF4102*). 2c shows pSJS630 (*Ppu recF*$^+$) and pSJS641 (*Ppu recF4103*). 2d shows pSJS525 (*Eco-Sty recF*) and pSJS526 (*Sty-Eco recF*). 2e shows pSJS684 (*recF*$^+$), pSJS703 (*recF4101*) and pUC118 (no *recF* gene). 2f shows JC18148 (*recF*$^+$), JC18146 (*recF4101*), and JC15359 (*recF349*). Note that the y axis in 2a–d are different from y axis in 2e–f. The experiments were done as descibed in the materials and methods.

K-12 (1). We therefore obtained a clone of this entire region (l gt10 *Eco* RI 5.2) and subcloned a 1.4 *Sph* I-*Sal* I fragment with vector pUC120 which we predicted would contain the entire *recF* coding sequence to facilitate further analysis. This plasmid was called pSJS630.

Complemetation of *E. coli recF349*

To ascertain whether or not these clones contained a *recF*-like gene, we tested their ability to complement an *E. coli recF* mutant. pFF48 and pSJS630 were used to transform *recF349* strain JC15359. These transformants were tested for the sensitivity to UV irradiation. The results are shown in Fig 2b,c. Both pSJS630 and pFF48 impart the UV resistance normally seen for a clone containing the *E. coli recF* gene (pSJS521); hence these clones are likely to carry *recF* counterparts. Since it was not known if the *B. subtilis recF* gene would complement an *E. coli recF* mutation, we performed a similar test with pSM2003, a clone carrying the *recF* gene of *B. subtilis* (2). This clone failed to complement the *E. coli recF* chromosomal mutation (Fig. 2a).

DNA sequence analysis and identification of *S. typhimurium* and *P. putida recF* genes

Since two clones complemented *recF349*, we decided to sequence the regions of DNA which were most likely to contain the *recF* genes from those organisms. Figure 3 shows the DNA sequences of two long open reading frames found in these two clones. The open reading frame for the *S. typhimurium* clone is 355 codons and the open reading frame for the *P. putida* clone is 367 codons (Figure 1). These sizes are very similar to the sizes of the *E. coli* K-12 and *B. subtilis recF* genes (Figure 1). In order to verify that it is these open reading frames which are responsible for the complementing ability of these clones, mutations which disrupt these open reading frames were introduced. A frameshift mutation, *recF4102* (carried on pSJS520), in the beginning of the *S. typhimurium* gene is predicted to produce a nonsense polypeptide of 64 amino acids instead of the wildtype protein. A deletion mutation in the *P. putida* gene, *recF4103* (carried on pSJS641), deletes nucleotides 27 through 878. In both cases, these mutations eliminate the complementing activity of these clones

Figure 3. DNA sequences of the putative *recF* genes from *S. typhimurium* [A] and *P. putida* [B]. The sequence begins with the first ATG start codon in the longest open reading frame. Flanking sequences are omitted to simplify the presentation. Full lines contain 70 bps.

as measured by UV sensitivity (Figure 2b,c). Hence we identify these two open reading frames as the *recF* genes from *P. putida* and *S. typhimurium*.

Amino Acid sequence analysis of the four *recF* proteins

The amino acid sequences for the four proteins from *E. coli* , *S. typhimurium*, *P. putida* and *B. subtilis* and their alignment according to the method of Smith and Smith (36) are shown in Figure 4A. Features of the alignment include two major gaps of 3 amino acids or more in the middle of the genes of the three purple bacteria and none in that of the gram positive bacterium. These gaps can be used to divide the proteins into N-terminal and C-terminal portions for functional and structural analysis. The alignment shows that 15% of the amino acids are identical in all four proteins. Pairwise comparisons are shown in Figure 4B. The most concentrated region of identities between the four proteins occurs in the C-terminal portion where there are 5 consecutive identities: SELDD (aa 306−310). Three other regions of high identities are: GxNxxGKT [(aa 30−37) first part of the phosphate binding loop {also identified in (39)}—see below], GxPxRRxFLD (aa 127−136), and LxQRNxxL (aa 158−165).

Nucleotide Binding Motif

It has been shown through x-ray crystallographic studies, primary sequence alignment, and mutational analysis of proteins with ATPase, kinase and GTPase activities that the amino acid sequence GxxGxGKT constitutes a loop binding the phosphate residue of a nucleotide di- or triphosphate (40, 41, 42). For short we will call this a phosphate binding loop. An aspartate residue elsewhere in each protein participates in phosphate binding by binding a divalent cation (e.g. Mg^{+2}). In particular, recombination and DNA repair genes such as *recA*, *recB*, *uvrA*, *uvrB*, and *uvrD* from *E. coli* and *RAD3* from *Saccharomyces*

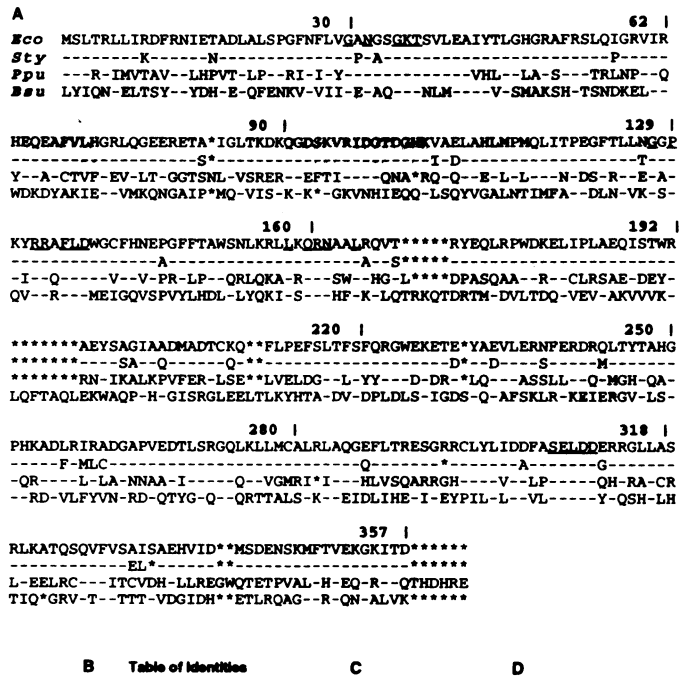

Figure 4. A. Alignment of the four RecF proteins. Numbering is for the amino acid of the *Eco* RecF protein. Dashes indicate an identical amino acid to *E. coli* in that position. Asterisks indicate a space put in by the computer to make this alignment. Underlined amino acids sequences show absolutely conserved residues in the regions of highest identity. B. A table of the percent identitical residues in pairwise combinations of the RecF proteins aligned as in Figure 4A. C and D. Dendrograms which indicate two different ways the four *recF* genes could be evolutionarily related.

cerevisiae encode proteins with this motif (43). We see the sequence for such a loop in all four *RecF* proteins. In the following three cases mutations in the highly conserved lysine {K} residue to arginine {R} makes a protein which is inactive for an ATPase activity and shows the mutant phenotype *in vivo* : *RAD3* (43), *uvrB* (44), and *recA* (R. Devoret and and S. Kowalczykowski, personal communication). To test whether or not this putative phosphate binding loop was essential for *recF* function. We changed lysine to arginine at position 36. This new *recF* allele is called *recF4101* and is contained on the plasmid pSJS703. Figure 2e shows that it can partially complement the UV sensitivity of a *recF* chromosomal null mutant (*recF349*). This partial complemetation could be due to two reasons: (1) *recF4101* inactivates only one of two or more activities of *recF* needed for complete UV resistance or (2) RecF4101 protein is weakly active and mutiple copies of the gene produce enough protein for a little repair. To discriminate between these two hypotheses, *recF4101* was transferred to the chromosome. The results (Fig. 2f) show that *recF4101* confers slightly more sensitivity than the *recF* chromosomal null mutation *recF349*. Hence we conclude that *recF4101* makes a largely inactive protein, which when encoded on multicopy number plasmid can contribute to UV resistance. We also conclude that lysine-36 and possibly a phosphate binding loop is important for *recF* function.

Parsimony Analysis

Figure 4B shows a table which lists the percent identical amino acids between pairwise combinations of the four *recF* proteins given the alignment in Figure 4a. From these data it is easy to see that the two *recF* genes from *E. coli* and *S. typhimurium* share a more recent ancestor than the *B. subtilis* and *P. putida genes*. It is less clear however whether the gene from *P. putida* shares a more recent ancestor with *B. subtilis* (Fig. 4C) or with *S. typhimurium* and *E. coli* (Fig. 4D). To discriminate between these two possibilities, a parsimony analysis of the third codon position variation in the DNA sequences of the four genes using the protein pattern induced alignments (36) and the PAUP software (45) was done. It was found that the diagram in 4D best matches this analysis (T. Smith, personal communication). Hence the three purple bacterial *recF* genes share a common ancestor which is not shared by the gram positive bacterial *recF* gene.

Chimeric *recF* genes

Since the two RecF proteins from *S. typhimurium* and *E. coli* were so similar, we wanted to test whether or not a few differences could be tolerated in the absence of the others. To do this, we made chimeric genes by exchanging codons 1–48. There are four codon differences [at positions 9 (R to K), 16 (T to N), 31 (A to P), and 33 (G to A)]. These represent two conservative differences (positions 9 and 33) and two non-conservative differences (positions 16 and 31). Two of these differences are also in the region of the putative nucleotide binding fold (positions 31 and 33). The results of such an exchange are shown in Fig.2. Both pSJS525 (*Eco* 1–48 *Sty* 49–355) and pSJS526 (*Sty* 1–48 *Eco* 49–357) are fully able to complement the UV sensitivity defect of a *recF349* chromosomal mutant. Hence if these substitutions do alter the activity of RecF protein, the alteration is too small to have a phenotypic effect detectable by this measure.

DISCUSSION

In this work we have identified and sequenced the *recF* gene from *Salmonella typhimurium* and completed the sequence of the *recF* gene from *Pseudomonas putida*. Although the carboxy terminal end of the *recF* gene from *P. putida* had been published (1), only one strand was sequenced. Hence we were able to correct the partial sequence and complete it for the middle of the gene by sequencing both strands of the DNA. We have shown that the *E. coli, S. typhimurium,* and *P. putida recF* genes will complement the UV sensitivity of a *E. coli recF* chromosomal mutation. Then by using the previously published sequences of the *B. subtilis* and *E. coli recF* gene sequences in addition to the new data, we aligned the protein sequences of the four proteins to look for regions which may define structural motifs that are important for *recF* function. We identified a possible phosphate binding loop in the N-terminal portion of the protein and a very highly conserved group of amino acids in the carboxy terminal portion of the protein (SELDD). Finally we probed for functional relationships between portions of the *recF* gene by making chimeric proteins between *E. coli* and *S. typhimurim recF* genes.

Several points have come from this study. First the fact that there are *recF* genes in three other organisms beside *E. coli* is significant. The *recA* gene has been found in 45 different strains of bacteria (from four different kingdoms) (11). The finding of *recF* genes in multiple organisms implies that at least the RecF pathway of recombinational repair may be equally widespread. Hence the genetic studies of the pathways of homologous recombination in *E. coli* may take on a broader meaning. Since there has been much work linking the roles of *recA* and *recF* in recombinational repair and induction of the SOS response in *E. coli* (13, 14), it would also be interesting to sequence and study the pairs of *recA* and *recF* genes from these organisms.

The complementation studies have shown that the *recF* genes from the purple bacteria will complement the *E. coli recF* chromosomal mutation for UV sensitivity, but the *recF* gene from gram positive bacteria, *B. subtilis*, will not. Expression of the *B. subtilis recF* gene product has been demonstrated in *E. coli* maxicells (30). If one assumes that the *Bsu* RecF protein has the same activities as *Eco* RecF protein, there may be several reasons why *Bsu recF*$^+$ does not complement *Eco recF349* : insufficient expression, poor protein stability, or an inability to interact with other *E. coli* proteins. With respect to the last possibility, we note that *Pseudomonas putida* protein with only 38% identitical amino acids as *E. coli* appears to be able to make the proper interactions. This may imply either that there is great conservation in the areas of the protein that do interact with *E. coli* proteins or that *Ppu* RecF protein may have eliminated the need for the interactions. The parsimony analysis showed good agreement with the complementation analysis in that the three genes which share the most recent common ancestor (*E. coli, S. typhimurium, and P. putida*) will subsitute for the *E. coli recF* gene.

The *recF4101* mutation, a lysine to arginine change at position 36 in the putative phosphate binding loop eliminates *recF* function when tested as a single copy gene on the chromosome and may produce a partially active protein as deduced by the partial complementation when carried on a multicopy plasmid. Since a phosphate binding loop is often found in proteins with ATPase, GTPase and kinase actvities (40, 41, 42), it is possible that *recF* has one of these types of activities. Griffin and Kolodner (17) showed that the RecF protein *in vitro* binds single stranded DNA *in vitro* in the absence of ATP or GTP but fails to act as a DNA dependent ATPase. RecF protein might however be a cryptic ATPase or nucleotide binding protein. For example the cryptic ATPase acitivity of UvrB protein is only active in the presence of UvrA protein (44). Thus if RecF protein associates with other proteins, then it might be active as an ATPase (or GTPase or kinase) in the presence of these other proteins. RecO protein and RecR proteins have been identified as possible associated proteins (18). Alternatively RecF protein might require a nucleotide cofactor to react with its native substrate.

ACKNOWLEDGEMENTS

We would like to thank Temple Smith of the MBCRR at the Dana-Farber Cancer Institute/Harvard School of Public Health for doing the protein alignment and the parsimony analysis with his computer programs. We would also like to thank Russell Maurer, Naotake Ogasawara and Masaki Q. Fujita for sending plasmids and phages and Juan Alonzo for his unpublished results. This research was supported by grant AI05371 from the National Institutes of Health.

REFERENCES

1. Fujita, M.Q., Yoshikawa, H. and Ogasawara, N. (1989) *Mol. Gen. Genet.*, **215**, 381–387.
2. Moriya, S., Ogasawara, N. and Yoshikawa, H. (1985) *Nucleic Acids Res.*, **13**, 2251–2265.
3. Horii, Z.-I. and Clark, A.J. (1973) *J. Mol. Biol.*, **80**, 327–344.

4. Clark, A.J. and Low, K.B. (1988) In Low, K.B. (ed.), Pathways and Systems of Homologous Recombination in *Escherichia coli*. Academic Press, Inc., San Diego, pp. 155–215.
5. Mahajan, S.K. (1989) In Kucherlapati, R. and Smith, G.R. (ed.), Pathways of Homolgous recombination in *Escherichia coli*. American Society for Microbiology, Washington, D.C., pp. 87–140.
6. Cohen, A. and Laban, A. (1983) *Mol. Gen. Genet.*, **189**, 471–474.
7. Fishel, R. and Kolodner, R. (1989) *J. Bacteriol.*, **171**, 3046–3052.
8. Thoms, B. and Wackernagel, W. (1987) *J. Bacteriol.*, **169**, 1731–1736.
9. Volkert, M.R. (1989) *J. Bacteriol.*, **171**, 99–103.
10. Ciesla, Z., O'Brian, P. and Clark, A.J. (1987) *Mol. Gen. Genet.*, **207**, 1–8.
11. Roca, A.I. and Cox, M.M. (1990) *Critical Reviews in Biochem. and Mol. Biol.*, **25**, 415–456.
12. Volkert, M.R., Margossian, L.J. and Clark, A.J. (1984) *J. Bacteriol.*, **160**, 702–705.
13. Madiraju, M.V.V.S., Templin, A. and Clark, A.J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 6592–6569.
14. Volkert, M.R. and Hartke, M.A. (1984) *J. Bacteriol.*, **169**, 498–506.
15. Lavery, P.E. and Kowalczykowski, S.C. (1990) *J. Biol. Chem.*, **265**, 4004–4010.
16. Madiraju, M.V.V.S., Lavery, P.E., Kowalczykowski, S.C. and Clark, A.J. (1991) *Biochemistry* (submitted).
17. Griffin, T.J. and Kolodner, R.D. (1990) *J. Bacteriol.*, **172**, 6291–6299.
18. Madiraju, M.V.V.S. and Clark, A.J. (1991) *Nucleic Acids Res.*, **19**, 6295–6300.
19. Ream, L.W., Margossian, L., Clark, A.J., Hansen, F.G. and Meyenburg, K.v. (1980) *Mol. Gen. Genet.*, **180**, 115–121.
20. Wang, Q. and Kaguni, J.M. (1987) *Mol. Gen: Genet.*, **209**, 518–525.
21. Menzel, R. and Gellert, M. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 4185–4189.
22. Armengod, M.-E. and Lambies, E. (1986) *Gene*, **43**, 183–196.
23. Armengod, M.-E., Garcia-Sogo, M. and Lambies, E. (1988) *J. Biol. Chem.*, **263**, 12109–12114.
24. Blanar, M.A., Sandler, S.J., Armengod, M.-E., Ream, L.W. and Clark, A.J. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4622–4626.
25. Sandler, S.J. and Clark, A.J. (1990) *Gene*, **86**, 35–43.
26. Alonso, J.C., Tailor, R.H. and Luder, G. (1988) *J. Bacteriol.*, **170**, 3001–3007.
27. Alonzo, J. (1991) *Mol. Gen. Genet.* (in press).
28. Engstrom, J., Wong, A. and Maurer, R. (1986) *Genetics*, **113**, 499–515.
29. Ogasawara, N., Moriya, S., Meyenberg, K.v., Hansen, F.G. and Yoshikawa, H. (1985) *EMBO J.*, **4**, 3345–3350.
30. Ogasawara, N., Moriya, S. and Yoshikawa, H. (1985) *Nuclec. Acids Res.*, **13**, 2267–2279.
31. Skovgaard, O. and Hansen, F.G. (1987) *J. Bacteriol.*, **169**, 3976–3981.
32. Blanar, M.A. (1985) Ph. D. Thesis, University of California, Berkeley.
33. Sakakibara, Y. and Mizuami, T. (1980) *Mol. Gen. Genet.*, **178**, 541–553.
34. Vieira, J. and Messing, J. (1987) In Wu, R. and Grossman, L. (ed.), Production of single-stranded plasmid DNA. Academic Press, Vol.153, pp. 1–11.
35. Willetts, N.S., Clark, A.J. and Low, B. (1969) *J. Bacteriol.*, **97**, 244–249.
36. Smith, R.F. and Smith, T.F. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 118–122.
37. Woese, C.R. (1987) *Microbiol. Rev.*, **51**, 221–271.
38. Ream, L.W. and Clark, A.J. (1983) *Plasmid*, **10**, 101–110.
39. Gorbalenya, A.E. and Koonin, E.V. (1990) *J. Mol. Biol.*, **213**,
40. Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982) *EMBO J.*, **1**, 945–951.
41. Saraste, M., Sibbald, P.R. and Wittinghofer, A. (1990) *TIBS*, **15**, 430–434.
42. Mimura, C.S., Holbrook, S.R. and Ames, G.F.-L. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 84–88.
43. Sung, P., Higgins, D., Prakash, L. and Prakash, S. (1988) *EMBO J.*, **7**, 3263–3269.
44. Seeley, T.W. and Grossman, L. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6577–6581.
45. Swofford, D.L. (1985) PAUP User's Guide, Illinois Natural History Survey.