Molecular organization of *sbcC*, a gene that affects genetic recombination and the viability of DNA palindromes in *Escherichia coli* K-12

Isam S.Naom, Stuart J.Morton, David R.F.Leach[1] and Robert G.Lloyd

Department of Genetics, University of Nottingham, Queens Medical Centre, Nottingham NG7 2UH and [1]Department of Molecular Biology, University of Edinburgh, King's Buildings, Edinburgh EH9 3JR, UK

ABSTRACT

The *sbcC* gene product of *Escherichia coli* interferes with the growth of a λ*red gam* phage carrying a long palindrome in its DNA. This phenotype was used to identify recombinant plasmids harbouring the wild-type gene and to isolate *sbcC* mutant derivatives carrying Tn*1000* insertions. Analysis of these plasmids located *sbcC* between *proC* and *phoR* at a slightly different position from that reported before (Lloyd, R. G. and Buckman, C. 1985, J. Bacteriol. **164**, 836−844). Nucleotide sequencing revealed that the gene spans a DNA segment of 3.3 kb that encodes a poorly expressed protein of 118 kDa and which lies downstream of a gene of unknown function that encodes a polypeptide of 45 kDa. The amino acid sequence of SbcC contains a nucleotide binding fold similar to that in RecB and other recombination proteins.

INTRODUCTION

The *sbcC* gene of *Escherichia coli* has been linked with genetic recombination, DNA repair and palindrome viability in this organism. The locus was first identified by mutations that arise in *recB recC sbcB* strains and which act in conjunction with the *sbcB* mutation to overcome the deficiencies in DNA repair, recombination and cell viability normally associated with the *recBC* mutations (1). These *sbcC* mutations were discovered in several of the commonly used '*recBC sbcB*' strains where they had presumably arisen spontaneously because of the improvement in viability. More recently, Chalker et al. (2) discovered that the *sbcC* gene interferes with the growth and stability of a λ*red gam* phage carrying a long palindrome in its DNA. Leach and Stahl (3) had earlier reported that phages of this type were unable to form plaques on *rec*$^+$ *sbc*$^+$ hosts, but did so with a reasonable efficiency on what they then thought to be a *recBC sbcB* strain. Chalker et al. demonstrated that the latter effect was in fact due to the *sbcC* mutation also present in the *recBC sbcB* strain used since they found that the palindrome phage grew equally well on an *sbcC* single mutant.

The inviability and instability of long DNA palindromes in *sbcC*$^+$ strains suggests that the *sbcC* product normally interacts with these sequences in some way that reduces the viability of the carrier DNA. How SbcC protein achieves this effect is uncertain. One possibility is that a palindrome is extruded into a cruciform and that SbcC cleaves this structure in much the same way as a Holliday junction intermediate in genetic recombination is resolved into the product molecules (3). Alternatively, SbcC protein may interfere with the replication of DNA by binding to hairpins formed by the palindrome in the single-stranded template. The available evidence seems to favour an effect on replication since palindromes survive well enough in *sbcC*$^+$ cells as long as they are not replicated (4,5).

However, a nuclease activity cannot be excluded, especially if replication helps to extrude palindromes.

As a first step in the molecular analysis of *sbcC* and of its product, we describe the cloning and nucleotide sequencing of the gene and flanking regions. We show that *sbcC* is part of an operon that encodes proteins of 45 kDa and 118 kDa.

METHODS

Bacterial strains, λphages and plasmids

Escherichia coli K-12 strains and λ phages are listed in Table 1. The genotype *pal571* in λDRL116 represents a 571 base pair (bp) palindrome with a 15 bp non-palindromic centre (2). Phages M13mp18 and M13mp19 (7) were used for DNA sequencing. The plasmid constructs used are shown in Figure 1. pJP71 and pJP77 are derivatives of pACYC184 and have been described elsewhere (9), as has pSB43 (10). pBL118 and pBL121 are derivatives of pBR322 (1). The construction of pIN509, pIN510, pGTI27 and pGTI28 is described in the Results section. pGTI27 and pGTI28 are derivatives of pHSG415, a low copy number vector that is temperature sensitive for replication (11). Strains carrying these plasmids or their derivatives were grown at 30°C. pSM118 was made by digestion of pGTI27 with *Xho*I and religating the large fragment. To make pSM122, the small *Hind*III-*Eco*RI junction fragment of pGTI27 was first cloned into pUC18 (7) cut with the same enzymes. The *Eco*RI end of the insert was then removed by digestion with *Eco*RI and *Alu*I, inserted into pUC18 digested with *Eco*RI and *Hinc*II and then excised again with *Eco*RI and *Hind*III. The resulting 0.7 kb fragment was then used to replace the *Hind*III-*Eco*RI junction fragment in pGTI27 (using a partial *Eco*RI digestion). pSM123 was constucted by the same strategy except that *Hae*III was used instead of *Alu*I. To construct pSM124, the larger part of the insert from pSM122 was removed by digestion with *Hind*III and partial digestion with *Pst*I (so as to avoid cutting the *Pst*I site introduced into this construct with the *Hinc*II-*Hind*III section of the multiple cloning site of pUC18), and cloned into *Hind*III-*Pst*I digested pHSG415.

Table 1. *E.coli* K-12 strains.

Strain	Relevant genotype	Other markers[a]	Source or derivation
AB1157	*rec+ sbc+*	a	6
AB2463	*recA13*	a	6
AB2480	*recA13 uvrA6*	a	P. Howard-Flanders
JC7623	*recB21 recC22 sbcB15 sbcC201*	a	6
AC301	*thyA+ recD1009*	*lac ara rpsL*	A. Chaudhury
N1116	*thyA*	a	1
N2667	*rec+ sbcC201 thyA*	a, *pro+*	N1116×N2373 to Pro+
N2679	*rec+ sbcC201*	a, *pro+*	P1.AC301×N2667 to Thy+
NH4104	F42 *lac+*	see ref. 1	K. B. Low
JM109	F128 *proAB+ lacI^qZΔM15 traD36 / Δ(lac-pro) recA1 endA1*	*gyrA96 hsdR17 supE44*	7
N2373	Hfr (PO2A) *sbcC201*	*metB1 relA1*	8
λDRL112	Δ*spi6 cI857*		2
λDRL116	Δ*spi6 cI857 pal571*		2

[a] F⁻ *thi-1 hisG4 Δ(gpt−proA)62 argE3 thr-1 leuB6 kdgK51 rfbD1 ara-14 lacY1 galK2 xyl-5 mtl-1 tsx-33 supE44 rpsL31 rac⁻*

Media and general methods
LB broth and 56/2 salts media have been cited elsewhere (1) along with methods for measuring growth in liquid culture and for strain construction by conjugation or by transduction with P1*vir*. Media contained 20 μg/ml tetracycline (Tc), 25 μg/ml chloramphenicol (Cm), 40 μg/ml kanamycin (Km), 50 μg/ml ampicillin, or 100μg/ml streptomycin, as required for the selection of antibiotic resistant cells. λ stocks were grown on strain JC7623 using the media and protocols of Silhavy et al. (12). To assay for *sbcC* activity, λDRL116 (λ*pal*) and the control, λDRL112, were titrated on lawns of the strains to be tested. On average, the efficiency of plating of λ*pal* on *sbcC*$^+$ strains was reduced by a factor of about 10^4 relative to that on *sbcC* mutants.

Enzymes and DNA analysis
Restriction endonucleases, T4 DNA ligase and Klenow polymerase were obtained from commercial sources and used as directed by the suppliers. Plasmid stocks were prepared from *recA* strain AB2463 by the rapid alkaline lysis method (13). Procedures for analysis of restricted DNA by agarose gel electrophoresis, purification of DNA fragments by electroelution, DNA ligation and transformation followed recipes and protocols described by Maniatis et al. (14).

DNA sequencing
Overlapping DNA fragments spanning *sbcC* (Fig. 1) were cloned into pUC18 or pUC19 and then directed into M13mp18 and M13mp19 using a combination of *Eco*RI and *Hind*III digests. The DNA inserts in these phages were sequenced by the dideoxy chain termination method of Sanger et al. (15), using kits from Pharmacia-LKB containing phage T7 DNA polymerase, deoxynucleoside and dideoxynucleoside triphosphates, and M13 universal primer (17mer). Compressions in GC rich regions were resolved using sequencing reactions containing 7-deaza dGTP or 7-deaza dITP instead of dGTP. The sequence determined was read from both strands of the DNA and was compiled and analysed using microcomputer software packages from DNASTAR Ltd.

Identification of plasmid encoded proteins
Plasmids were transformed into strain AB2480 and the proteins made were labelled with [^{35}S]methionine (Amersham) using the maxicell method of Sancar et al. (16). Labelled proteins were separated by SDS-PAGE as described by West and Emmerson (17). The [^{14}C]methylated proteins mixture used to provide molecular weight markers was obtained from Amersham International and contained myosin (200,000), phosphorylase-b (92,500), bovine serum albumin (69,000), ovalbumin (46,000), carbonic anhydrase (30,000) and lysozyme (14,300).

RESULTS
Molecular cloning of sbcC
Lloyd and Buckman (1) reported that *recBC sbcBC* strains carrying the recombinant plasmids pJP71 or pJP77 grow very slowly, are sensitive to mitomycin C and deficient in recombination. They concluded that these plasmids carried *sbcC*$^+$, as was expected from the fact that they carried the *proC-phoR* interval of the chromosome (Fig. 1). Analysis of Tn*1000* insertions in pJP71 and pJP77 that alleviated the adverse effect of these plasmids on growth suggested that *sbcC* might be located within the 4.3 kb *Bam*HI(1) to *Eco*RI(1) region, which was then cloned into pBR322. However, testing for the presence of *sbcC* in these constructs (pBL118 and pBL121) by measuring their effect on the growth of a *recBC sbcBC* strain proved impracticable (Lloyd, R.G. and Buckman, C., unpublished)

because of their instability, a common property of multi-copy plasmids in strains of this type (18).

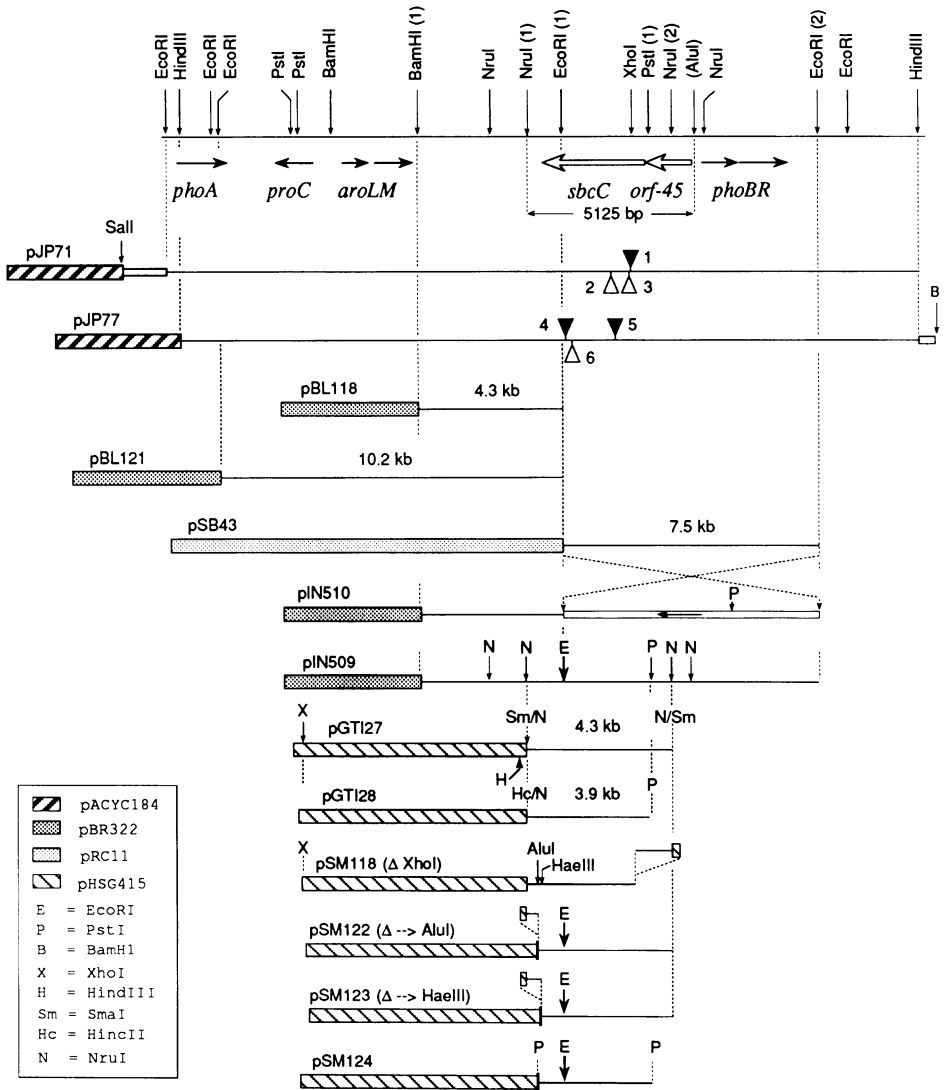The discovery that *sbcC* single mutants allow growth of a λ phage carrying a long

Figure 1. Restriction map of the *phoA-phoR* interval of the *E. coli* chromosome and structure of recombinant plasmids. Vector sequences are identified by the shaded boxes. With the exception of the inversion of the 7.5 *Eco*RI(1) to *Eco*RI(2) fragment (arrowed open box) in pIN510, chromosomal inserts are indicated by a solid line and are aligned with respect to the *Eco*RI(1) site within *sbcC*. Open boxes in pJP71 and pJP77 are non-contiguous chromosomal sequences (9). The sites of insertion of Tn*1000* in these two plasmids are indicated by the open (gamma-delta insertion) and closed (delta-gamma insertion) triangles. The section of the pUC18 multiple cloning site introduced into pSM122, pSM123 and pSM124 is indicated by a solid vertical bar.

palindrome in its DNA, whereas *sbcC*[+] strains do not (2), enabled us to re-examine these plasmids for *sbcC* activity in a *recBC*[+] *sbcB*[+] genetic background, which avoided the problem of plasmid instability. The results (Table 2) revealed that pJP71 prevents λDRL116 from plating with a high efficiency on an *sbcC* mutant. Very similar results were obtained with pJP77 (data not shown). These observations confirm that both plasmids do indeed carry *sbcC*[+]. However, it is clear that pBL118 and pBL121 do not. We therefore re-examined the putative *sbcC*::Tn*1000* insertion derivatives of pJP71 and pJP77. All reduced the plating of λ*pal* on strain N2679 by a factor of 10^4 or more (data not shown). We concluded that the locus defined by these insertions cannot be *sbcC*.

To try and locate *sbcC* we examined the effect of pSB43 on the plating efficiency of λ*pal* on N2679. This plasmid carries the 7.5 kb *Eco*RI(1) to *Eco*RI(2) fragment of DNA that contains the section of DNA between *proC* and *phoR* missing in pBL121 (Fig. 1). Again, it was clear that pSB43 did not carry an intact *sbcC* gene since the plating efficiency was high (data not shown). It seemed likely therefore that *sbcC* extends across the *Eco*RI(1) site common to pBL121 and pSB43. This possibility was confirmed when the 7.5 kb *Eco*RI fragment was cloned into the *Eco*RI site of pBL118 to reconstruct this region of the chromosome. The resulting construct, pIN509 (Fig. 1), severely reduces the plating of λ*pal* on N2679, whereas pIN510, which carries the same 7.5 kb fragment inserted into pBL118 in the reverse orientation does not (Table 2).

To identify the extent of the DNA encoding *sbcC* we isolated a number of pJP71 and pJP77 derivatives in which *sbcC* was inactivated by Tn*1000* insertion. The plasmids were transformed separately into the F42*lac* strain, NH4104, and then mobilized to the *sbcC* mutant N2679 by conjugation to generate Tn*1000* insertion mutants. Cm^r(*rpsL*) transconjugants were selected and tested for their sensitivity to λDRL116. Sensitive isolates were assumed to carry Tn*1000* insertions inactivating *sbcC*. The locations of the insertions in six independent isolates, three from pJP71 and three from pJP77, were determined by

Table 2. Effect of *sbcC* plasmids on plaque formation by λDRL116.

Strain	Plasmid	λ Plaque forming units per ml[a]		Ratio
		DRL116	DRL112	DRL116/DRL112
AB1157	pACYC184	2.0×10^4	4.5×10^8	0.000044
(*sbcC*[+])	pJP71	5.1×10^4	8.0×10^8	0.000064
	pBR322	1.0×10^4	7.1×10^8	0.00013
	pIN509	4.9×10^4	6.9×10^8	0.000071
	pHSG415	2.5×10^4	3.5×10^8	0.000071
N2679	pACYC184	5.7×10^8	5.5×10^8	1.04
(*sbcC201*)	pJP71	5.0×10^4	4.1×10^8	0.00012
	pBR322	5.3×10^8	5.5×10^8	0.96
	pBL118	4.0×10^8	4.5×10^8	0.88
	pBL121	2.6×10^8	4.2×10^8	0.62
	pIN509	5.9×10^4	5.3×10^8	0.00011
	pIN510	4.5×10^8	4.5×10^8	1.00
	pHSG415	2.3×10^8	3.1×10^8	0.74
	pGTI27	2.5×10^4	3.0×10^8	0.000083
	pGTI28	3.1×10^4	3.0×10^8	0.0001
	pSM118	2.6×10^8	4.3×10^8	0.60

[a] The phage stocks used contained approximately 5×10^8 p.f.u. per ml as determined by titration on the *recB recC sbcB sbcC* strain JC7623.

restriction analysis. The insertions were found to span a 1.8 kb section of DNA within the 2.7 kb *Eco*RI(1) to *Pst*I(1) fragment (Fig. 1).

Restriction mapping of pIN509 revealed that the 2.7 kb *Eco*RI(1) to *Pst*I(1) fragment is flanked in the chromosome by *Nru*I sites (Fig. 1). We attempted to clone the 4.3 kb *Nru*I(1) to *Nru*I(2) fragment and the 3.9 kb *Nru*I(1) to *Pst*I(1) sub-section into a variety of standard multi-copy plasmid cloning vectors, without success. However, both fragments were introduced without difficulty into the low copy-number vector, pHSG415. Both constructs, pGTI27 and pGTI28 (Fig. 1), are *sbcC*$^+$ as defined by their effect on the plating of λDRL116 (Table 2). To further delineate *sbcC* a series of deletion derivatives of pGTI27 (pSM118, pSM122, pSM123 and pSM124, Fig. 1) were analysed for *sbcC* activity. Of these, only pSM122 and pSM124 retained the ability to suppress the plating of λDRL116 (Table 2 and data not shown). These results indicate that the DNA essential for *sbcC* activity extends about 3.3 kb from the 0.1 kb *Alu*I-*Hae*III fragment to the 0.6 kb *Xho*I to *Pst*I(1) fragment, a region that could encode a protein of up to 120 kDa.

Identification of the sbcC product

Examination of maxicell extracts revealed that pJP71 does indeed encode a large protein of about 120 kDa (Fig. 2, lane b), as does pJP77 (not shown). In both cases, the 120 kDa band is weak in relation to other products encoded by these plasmids. None of the six *sbcC*::Tn*1000* insertion mutants encoded this protein. The result obtained with insertion
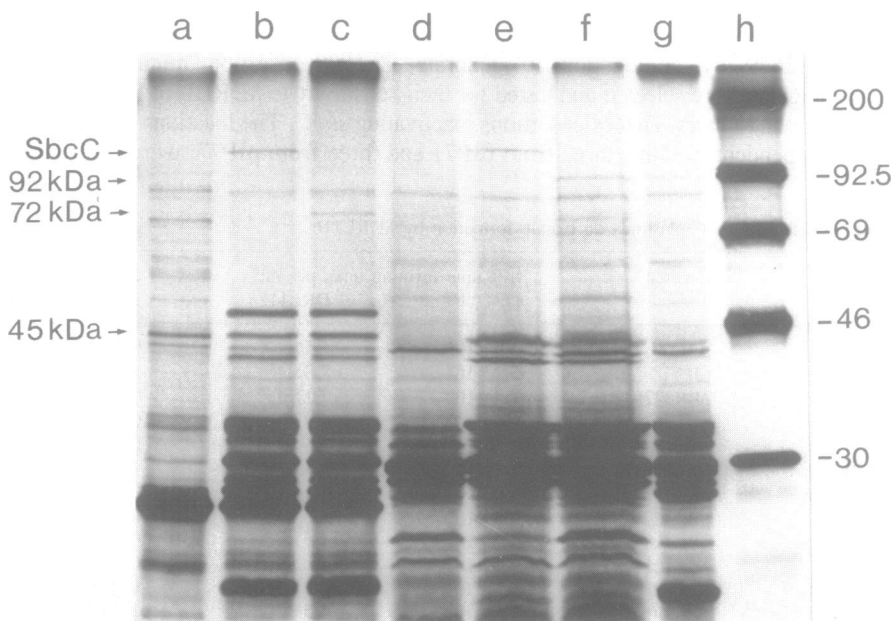

Figure 2. Fluorogram showing [^{35}S]-labelled, plasmid-encoded proteins from maxicell extracts separated on 10.0% polyacrylamide-sodium dodecyl sulphate gels. The plasmids present were a, pACYC184; b, pJP71; c, pJP71 *sbcC*::Tn*1000*-3; d, pBR322; e, pIN509; f, pIN510; g, pBL118. Molecular weight markers (lane h) are identified in kDa. The arrows to the left mark the positions of the possible truncated products of *sbcC* (SbcC protein) detectable in lanes c and f, and the 45 kDa protein encoded by the 7.5 kb *Eco*RI(1) to *Eco*RI(2) fragment (see text).

3 in pJP71 is shown in lane c. The same 120 kDa protein is also encoded by the $sbcC^+$ plasmid, pIN509 (lane e), but not by pIN510 (lane f), pBL118 (lane g) or pBL121 (not shown). It is also detectable in maxicells of pGTI27 and pGTI28, but only just (data not shown), which is not surprising in view of the fact that these are low copy-number constructs. Since this appears to be the only protein encoded by all the $sbcC^+$ constructs that is missing from $sbcC$ insertion or deletion plasmids, we conclude that it must be the product of $sbcC$. The molecular weight of 120 kDa agrees with the predicted size of the gene.

The maxicell extracts of pJP71 $sbcC$::Tn1000 insertion number 3 (Fig. 2, lane c) and pIN510 (lane f) show new protein bands of approximately 72 kDa and 92 kDa. These could well be the truncated products of the interrupted $sbcC$ genes in these constructs. If that is indeed what they are, it would indicate that $sbcC$ is transcribed in a counter-clockwise direction with respect to the genetic map.

Certain other protein bands visible in the various maxicell extracts could be attributed to $phoA$ (49.7 kDa), $proC$ (28 kDa), $aroL$ (17 kDa), $aroM$ (26.5 kDa), $phoB$ (30 kDa) and $phoR$ (49 kDa) on the basis of their mobilities and the plasmids that encode them. Others were located to the 4.3 kb BamHI(1) to EcoRI(1) fragment carried by pBL118, which is genetically poorly defined. Not all of these proteins are visible in Figure 2 since proteins of low molecular weight were run off the end so as to improve the resolution in the high molecular weight range. Two proteins, of about 40 kDa and 45 kDa respectively, are encoded by both pIN509 and pIN510 but not by pBL118 (Fig. 2, and data not shown), which means that they must be encoded within the 7.5 kb EcoRI(1) to EcoRI(2) fragment that includes $phoB$, $phoR$ and most of $sbcC$ (Fig. 1).

Nucleotide sequencing of sbcC

Figure 3 shows the nucleotide sequence of the 5,125 bp section of DNA spanning the $sbcC$ region of the chromosome (Fig. 1) from NruI (1) to an AluI site beyond NruI(2). The sequence is presented with the AluI site at the 5' end. The first 199 bp are complementary (with a 100% match) to the 5' end of the sequence presented by Makino et al. (19) that contains the region immediately upstream of $phoB$ (Fig. 1). Analysis of the sequence determined also revealed a NruI site (TCGCGA) at bp 2442 between the EcoRI(1) and XhoI sites which had not been detected during restriction mapping of this region (Fig. 1). We found that *dam* methylation (20) of the GATC sequence that overlaps the NruI site by 2 bp was responsible for this anomaly. Cleavage at this NruI site was readily detected when the DNA was extracted from a *dam* mutant (data not shown).

Our initial sequencing concentrated on the 3.9 kb NruI(1) to PstI(1) fragment shown to encode $sbcC$. Examination of the sequence obtained revealed that this region contains a large open reading frame that begins with an ATG at bp 1309 and extends 3,144 nucleotides to a TAA termination codon at bp 4453. Translation of this open reading frame would give a polypeptide of 1,048 amino acids with a predicted molecular weight of 118,730 Daltons, which is in good agreement with the estimate of 120 kDa obtained for SbcC protein from its migration on denaturing gels. Since both the beginning and end of this reading frame agree exactly with the limits of the DNA needed for $sbcC$ activity set by the cloning and deletion analysis, we conclude that it must be the $sbcC$ gene.

Examination of the 53 bp region extending from the PstI(1) site (bp 1256) to the proposed start codon for $sbcC$ failed to reveal any sequences that could direct the transcription and translation of the gene. We therefore extended the sequencing beyond PstI(1) until an overlap was generated with the sequence determined previously for the $phoB$ region (19). This

```
AluI                                                                                    -10                   **
AGCTCGTCGCACATATCTTCAGGTTATTGATTTCCGTGGCGCAGAAAAAAGCAAATGGCACATCTGTTTGGGTATAATCGCGCCCATGCTTTTTCGCCAG  100

        orf-45 -->
* GGAACCGTTATGCGCATCCTTCACACCTCAGACTGGCATCTCGGCCAGAACTTCTACAGTAAAAGCCGCGAAGCTGAACATCAGGCTTTTCTTGACTGGC  200
          m  r  i  l  h  t  s  d  w  h  l  g  q  n  f  y  s  k  s  r  e  a  e  h  q  a  f  l  d  w  l

TGCTGGAGACAGCACAAACCCATCAGGTGGATGCGATTATTGTTGCCGGTGATGTTTTCGATACCGGCTCGCCGCCCAGTTACGCCCGCACGTTATACAA  300
  l  e  t  a  q  t  h  q  v  d  a  i  i  v  a  g  d  v  f  d  t  g  s  p  p  s  y  a  r  t  l  y  n

CCGTTTTGTTGTCAATTTACAGCAAACTGGCTGTCATCTGGTGGTACTGGCAGGAAACCATGACTCGGTCGCCACGCTGAATGAATCGCGCGATATCATG  400
  r  f  v  v  n  l  q  q  t  g  c  h  l  v  v  l  a  g  n  h  d  s  v  a  t  l  n  e  s  r  d  i  m
                                                                     NruI(2)
GCGTTCCTCAATACTACCGTGGTCGCCAGCGCCGGACATGCGCCGCAAATCTTGCCTCGTCGCGACGGGACGCCAGGCGCAGTGCTGTGCCCCATTCCGT  500
  a  f  l  n  t  t  v  v  a  s  a  g  h  a  p  q  i  l  p  r  r  d  g  t  p  g  a  v  l  c  p  i  p  f

TTTTACGTCCGCGTGACATTATTACCAGCCAGGCGGGCTTAACGGTATTGAAAAACAGCAGCATTTACTGGCAGCGATTACCGATTATTACCAACAACA  600
  l  r  p  r  d  i  i  t  s  q  a  g  l  n  g  i  e  k  q  q  h  l  l  a  a  i  t  d  y  y  q  q  h

CTATGCCGATGCCTGCAAACTGCGCGGCGATCAGCCTCTGCCCATCATCGCCACGGGACATTTAACGACCGTGGGGGCCAGTAAAAGTGACGCCGTGCGT  700
  y  a  d  a  c  k  l  r  g  d  q  p  l  p  i  i  a  t  g  h  l  t  t  v  g  a  s  k  s  d  a  v  r

GACATTTATATTGGCACGCTGGACGCGTTTCCGGCACAAACCTTTCCACCAGCCGACTACATCGCGCTCGGGCATATTCACCGCGCACAGATTATTGGCG  800
  d  i  y  i  g  t  l  d  a  f  p  a  q  n  f  p  p  a  d  y  i  a  l  g  h  i  h  r  a  q  i  i  g

GCATGGAACATGTTCGCTATTGCGGCTCCCCCATTCCACTGAGTTTTGATGAATGCGGTAAGAGTAAATATGTCCATCTGGTGACATTTTCAAACGGCAA  900
  m  e  h  v  r  y  c  g  s  p  i  p  l  s  f  d  e  c  g  k  s  k  y  v  h  l  v  t  f  s  n  g  k

ATTAGAGAGCGTGGAAAACCTGAACGTACCGGTAACGCAACCCATGGCAGTGCTGAAAGGCGATCTGGCGTCGATTACCGCACAGCTGGAACAGTGGCGC  1000
  l  e  s  v  e  n  l  n  v  p  v  t  q  p  m  a  v  l  k  g  d  l  a  s  i  t  a  q  l  e  q  w  r

GATGTATCGCAGGAGCCACCTGTCTGGCTGGATATCGAAATCACTACTGATGAGTATCTGCATGATATTCAGCGCAAAATCCAGGCATTAACCGAATCAT  1100
  d  v  s  q  e  p  p  v  w  l  d  i  e  i  t  t  d  e  y  l  h  d  i  q  r  k  i  q  a  l  t  e  s  l

TGCCTGTCGAAGTATTGCTGGTACGTCGGAGTCGTGAACAGCGCGAGCGTGTGTTAGCCAGCCAACAGCGTGAAACCCTCAGCGAACTCAGCGTCGAAGA  1200
  p  v  e  v  l  l  v  r  r  s  r  e  q  r  e  r  v  l  a  s  q  q  r  e  t  l  s  e  l  s  v  e  e
                                                               PstI(1)
GGTGTTCAATCGCCGTCTGGCACTGGAAGAACTGGATGAATCGCAGCAGCAACGTCTGCAGCATCTTTTCACCACGACGTTGCATACCCTCGCCGGAGAA  1300
  v  f  n  r  r  l  a  l  e  e  l  d  e  s  q  q  q  r  l  q  h  l  f  t  t  t  l  h  t  l  a  g  e
            sbcC -->
CACGAAGCATGAAAATTCTCAGCCTGCGCCTGAAAAACCTGAACTCATTAAAAGGCGAATGGAAGATTGATTTCACCCGCGAGCCGTTCGCCAGCAACGG  1400
  h  e  a  .       m  k  i  l  s  l  r  l  k  n  l  n  s  l  k  g  e  w  k  i  d  f  t  r  e  p  f  a  s  n  g

GCTGTTTGCTATTACCGGCCCAACAGGTGCGGGGAAAACCACCCTGCTGGACGCCCATTTGTCTGGCGCTGTATCACGAAACTCCGCGTCTCTCTAACGTT  1500
  l  f  a  i  t  g  p  t  g  a  g  k  t  t  l  l  d  a  i  c  l  a  l  y  h  e  t  p  r  l  s  n  v

TCACAATCGCAAAATGATCTCATGACCCGCGATACCGCCGAATGTCTGGCGGAGGTGGAGTTTGAAGTGAAAGGTGAAGCGTACCGTGCATTCTGGAGCC  1600
  s  q  s  q  n  d  l  m  t  r  d  t  a  e  c  l  a  e  v  e  f  e  v  k  g  e  a  y  r  a  f  w  s  q

AGAATCGGGCGCGTAACCAACCCGACGGTAATTTGCAGGTGCCACGCGTAGAGCTGGCGCGCTGCGCCGACGGCAAAATTCTCGCCGACAAAGTGAAAGA  1700
  n  r  a  r  n  q  p  d  g  n  l  q  v  p  r  v  e  l  a  r  c  a  d  g  k  i  l  a  d  k  v  k  d

TAAGCTGGAACTGACAGCGACGTTAACCGGGCTGGATTACGGGCGCTTCACCCGTTCGATGCTGCTTTCGCAGGGGCAATTTGCTGCCTTCCTGAATGCC  1800
  k  l  e  l  t  a  t  l  t  g  l  d  y  g  r  f  t  r  s  m  l  l  s  q  g  q  f  a  a  f  l  n  a
                 XhoI
AAACCCAAAGAACGCGCGGAATTGCTCGAGGAGTTAACCGGCACTGAAATCTACGGGCAAATCTCGGCGATGGTTTTTGAGCAGCACAAATCGGCCCGCA  1900
  k  p  k  e  r  a  e  l  l  e  e  l  t  g  t  e  i  y  g  q  i  s  a  m  v  f  e  q  h  k  s  a  r  t

CAGAGCTGGAGAAGCTGCAAGCGCAGGCCAGCGGCGTCACGTTGCTCACGCCGGAACAAGTGCAATCGCTGACAGCGAGTTTGCAGGTACTTACTGACGA  2000
  e  l  e  k  l  q  a  q  a  s  g  v  t  l  l  t  p  e  q  v  q  s  l  t  a  s  l  q  v  l  t  d  e

AGAAAAACAGTTAATTACCGCGCAGCAGCAAGAACAACAATCGCTAAACTGGTTAACGCGTCAGGACGAATTGCAGCAAGAAGCCAGCCGCCGTCAGCAG  2100
  e  k  q  l  i  t  a  q  q  q  e  q  q  s  l  n  w  l  t  r  q  d  e  l  q  q  e  a  s  r  r  q  q

GCCTTGCAACAGGCGTTAGCCGAAGAAGAAAAAAGCGCAACCTCAACTGGCGGCGCTTAGTCTGGCACAACCGGCACGAAATCTTCGTCCACACTGGGAAC  2200
  a  l  q  q  a  l  a  e  e  e  k  a  q  p  q  a  a  l  s  l  a  q  p  a  r  n  l  r  p  h  w  e  r

GCATCGCAGAACACAGCGCGGCGCTGGCGCATATTCGCCAGCAGATTGAAGAAGTAAATACTCGCTTACAGAGCAATGGCGCTTCGCGCGAGCATTCG  2300
  i  a  e  h  s  a  a  l  a  h  i  r  q  q  i  e  e  v  n  t  r  l  q  s  t  m  a  l  r  a  s  i  r

CCACCACGCGGCGAAGCAGTCAGCAGAATTACAGCAGCAGCAACAAAGCCTGAATACCTGGTTACAGGAACACGACCGCTTCCGTCAGTGGAACAACGAA  2400
  h  h  a  a  k  q  s  a  e  l  q  q  q  q  q  s  l  n  t  w  l  q  e  h  d  r  f  r  q  w  n  n  e

CCGGCGGGTTGGCGTGCGCAGTTCTCCCAACAAACCAGCGATCGGCAGCATCTGCGGCAATGGCAGCAACAGTTAACCCATGCTGAGCAAAAACTTAATG  2500
  p  a  g  w  r  a  q  f  s  q  q  t  s  d  r  e  h  l  r  q  w  q  q  q  l  t  h  a  e  q  k  l  n  a

CGCTTGCGGCGATCACGTTGACGTTAACCGCCGATGAAGTTGCTACCGCCCTGGCGCAACATGCTGAGCAACGCCCACTGCGTCAGCACCTGGTCGCGCT  2600
  l  a  a  i  t  l  t  l  t  a  d  e  v  a  t  a  l  a  q  h  a  e  q  r  p  l  r  q  h  l  v  a  l

GCATGGACAGATTGTTCCCCAACAAAAACGTCTGGCGCAGTTACAGGTCGCTATCCAGAATGTCACGCAAGAACAGACGCAACGTAACGCCGCACTTAAC  2700
  h  g  q  i  v  p  q  q  k  r  l  a  q  l  q  v  a  i  q  n  v  t  q  e  q  t  q  r  n  a  l  n

GAAATGCGCCAGCGTTATAAAGAAAAGACGCAGCAACTTGCCGATGTGAAAACCATTTGCGAGCAGGAAGCGCGCATCAAAAACGCTGGAAGCTCAACGTG  2800
  e  m  r  q  r  y  k  e  k  t  q  q  l  a  d  v  k  t  i  c  e  q  e  a  r  i  k  t  l  e  a  q  r  a

CACAGTTACAGGCGGGTCAGCCTTGCCCCACTTTTGTGGTTCCACCAGCCACCCGGCGGTCGAGGCGTATCAGGCGCTGGAGCCTGGCGTTAATCAGTCTCG  2900
  q  l  q  a  g  q  p  c  p  l  c  g  s  t  s  h  p  a  v  e  a  y  q  a  l  e  p  g  v  n  q  s  r
```

```
ATTACTGGCGCTGGAAAACGAAGTTAAAAAGCTCGGTGAAGAAGGTGCGACGCTACGTGGGCAACTGGACGCCATAACAAAGCAGCTTCAGCGTGATGAA 3000
  l  l  a  l  e  n  e  v  k  k  l  g  e  e  g  a  t  l  r  g  q  l  d  a  i  t  k  q  l  q  r  d  e
AACGAAGCGCAAAGCCTCCGACAAGATGAGCAAGCACTTACTCAACAATGGCAAGCCGTCACGGCCAGCCTCAATATCACCTTGCAGCCACTGGACGATA 3100
  n  e  a  q  s  l  r  q  d  e  q  a  l  t  q  q  w  q  a  v  t  a  s  l  n  i  t  l  q  p  l  d  d  i
TTCAACCGTGGCTGGATGCACAAGATGAGCACGAACGCCAGCTGCGGTTACTCAGCCAACGGCATGAATTACAAGGGCAGATTGCCGCGCATAATCAGCA 3200
  q  p  w  l  d  a  q  d  e  h  e  r  q  l  r  l  l  s  q  r  h  e  l  q  g  q  i  a  a  h  n  q  q
AATTATCCAGTATCAACAGCAAATTGAACAACGCCAGCAACTACTTTTAACGACATTGACGGGTTATGCACTGACATTGCCACAGGAAGATGAAGAAGAG 3300
  i  i  q  y  q  q  q  i  e  q  r  q  q  l  l  l  t  t  l  t  g  y  a  l  t  l  p  q  e  d  e  e
AGCTGGTTGGCGACACGTCAGCAAGAAGCGCAGAGCTGGCAGCAACGCCAGAACGAATTAACCGCGCTGCAAAACCGTATTCAGCAGCTGACGCCGGATTC 3400
  s  w  l  a  t  r  q  q  e  a  q  s  w  q  q  r  q  n  e  l  t  a  l  q  n  r  i  q  q  l  t  p  i  l
TGGAAACGTTGCCGCAAAGTGATGAACTCCCGCACTGCGAAGAAACTGTGTGGTATTGGAAAACTGGCGGCAGGTACATGAACAATGTCTCGCATTACACAG 3500
    e  t  l  p  q  s  d  e  l  p  h  c  e  e  t  v  v  l  e  n  w  r  q  v  h  e  q  c  l  a  l  h  s
CCAGCAGCAGACGTTACAGCAACAGGATGTTCTGGCGGCGCAAAGTCTGCAAAAAGCCCAGGCGCAGTTTGACACCGCGCTACAGGCCAGCGTCTTTGAC 3600
  q  q  q  t  l  q  q  q  d  v  l  a  a  q  s  l  q  k  a  q  a  q  f  d  t  a  l  q  a  s  v  f  d
GATCAGCAGGCGTTCCTTGCGGCGCTAATGGATGAACAAACACTAACGCAGCTGGAACAGCTCAAGCAGAATCTGGAAAAACCAGCGCCGTCAGGCGCAA 3700
  d  q  q  a  f  l  a  a  l  m  d  e  q  t  l  t  q  l  e  q  l  k  q  n  l  e  n  q  r  r  q  a  q  t
CTCTGGTCACTCAGACAGCAGAAACGCTGGCACAGCATCAACAACACCGACCTGACGACGGGTTGGCTCTCACTGTGACGGTGGAGCAGATTCAGCAAGA 3800
   l  v  t  q  t  a  e  t  l  a  q  h  q  q  h  r  p  d  d  g  l  a  l  t  v  t  v  e  q  i  q  q  e
GTTAGCGCAAACTCACCAAAAGTTGCGTGAAAACACCACGAGTCAAGGCGAGATTCGCCAGCAGCTGAAGCAGGATGCAGATAACCGTCAGCAACAACAA 3900
  l  a  q  t  h  q  k  l  r  e  n  t  t  s  q  g  e  i  r  q  q  l  k  q  d  a  d  n  r  q  q  q  q
                                                                    EcoRI
ACCTTAATGCAGCAAATTGCTCAAATGACGCAGCAGGTTGAGGACTGGGGATATCTGAATTCGCTAATAGGTTCCAAAGAGGGCGATAAATTCCGCAAGT 4000
  t  l  m  q  q  i  a  q  m  t  q  q  v  e  d  w  g  y  l  n  s  l  i  g  s  k  e  g  d  k  f  r  k  f
TTGCCCAGGGGCTGACGCTGGATAATTTAGTCCATCTCGCTAATCAGCAACTTACCCGGCTGCACGGGCGCTATCTGTTACAGCGCAAAGCCAGCGAGGC 4100
   a  q  g  l  t  l  d  n  l  v  h  l  a  n  q  q  l  t  r  l  h  g  r  y  l  l  q  r  k  a  s  e  a
GCTGGAAGTCGAGGTTGTTGATACCTGGCAGGCAGATGCGGTACGCGATACCCGTACCCTTTCCGGCGGCGAAAGTTTCCTCGTTAGTCTGGCGCTGGCG 4200
   l  e  v  e  v  v  d  t  w  q  a  d  a  v  r  d  t  r  t  l  s  g  g  e  s  f  l  v  s  l  a  l  a
CTGGCGCTTTCGGATCTGGTCAGCCATAAAACACGTATTGACTCGCTGTTCCTTGATGAAGGTTTTGGCACGCTGGATAGCGAAACGCTGGATACCGCCC 4300
   l  a  l  s  d  l  v  s  h  k  t  r  i  d  s  l  f  l  d  e  g  f  g  t  l  d  s  e  t  l  d  t  a  l
TTGATGCGCTGGATGCCCTGAACGCCAGTGGCAAAACCATCGGTGTGATTAGCCACGTAGAAGCGATGAAAGAGCGTATTCCGGTGCAGATCAAAGTGAA 4400
   d  a  l  d  a  l  n  a  s  g  k  t  i  g  v  i  s  h  v  e  a  m  k  e  r  i  p  v  q  i  k  v  k
     HaeIII
AAAGATCAACGGCCTGGGCTACAGCAAACTGGAAAGTACGTTTGCAGTGAAATAACTATTCAGCAGGATAATGAATACAGAGGGGCGAATTATCTCTTGG 4500
  k  i  n  g  l  g  y  s  k  l  e  s  t  f  a  v  k  .
    AluI
CCTTGCTGGTCGTTATCCTGCAAGCTATCACTTTATTGGCTACGGTGATTGGTAGCCGTTCTGGTGGTTGTGATGGTGGTATGAAAAAAGTCATTTTATC 4600
TTTGGCTCTGGGCACGTTTGGTTTGGGGATGGCCGAATTTGGCATTATGGGCGTGCTCACGGAGCTGGCGCATAACGTAGGGAATTTCGATTCCTGCCGCC 4700
GGGCATATGATCTCGTATTATGCACTGGGGGTGGTGGTCGGTGCGCCAATCATCGCACTCTTTTCCAGCCGCTACTCACTCAAACATATCTTGTTGTTTC 4800
TGGTGGCGTTGTGCGTCATTGGCAACGCCATGTTCACGCTCTCTTCGTCGTTACCTGATGCTCGCCATTGGTCGGCTGGTATCCGGCTTTCCGCATGGCGC 4900
ATTTTTTGGCGTCGGAGCGATCTGTGTTATCAAAAATTATCAAACCGGAAAAGTCACCGCCGCCGTGGCGGGGATGGTTTCCGGGATGACATGCGCCAAT 5000
TTGCTGGGCATTCCGCTGGGAACGTATTTAAGTCAGGAATTTAGCTGGCGTTACACCTTTTTATTGATCGCTGTTTTTAATATTGCGGTGATGGCATCGG 5100
TCTATTTTTGGGTGCCAGATATTCG                                                                         5125
```

Figure 3. Nucleotide sequence of *sbcC* and flanking regions. The orientation of the 5125 bp sequence is from the *Alu*I site to *Nru*I(1) as shown in Figure 1. The relevant restriction enzyme sites are listed above the first nucleotide of the recognition sequence. The putative promoter −10 region for the *orf-45 sbcC* operon is overlined and labelled as such, while the proposed ribosome binding sequence is marked by asterisks. Stop codons for the *orf-45* and *sbcC* reading frames are identified with a dot.

upstream region contains another long open reading frame beginning with an ATG at bp 110 and extending 1200 nucleotides to a TGA stop codon at bp 1310. The start codon is preceded by the sequence AGG which could provide a ribosome binding site (21), and a little further upstream by TATAAT, which matches exactly the consensus sequence for a −10 promoter region (22). The TTT and CCCA flanking this sequence also match the extended consensus for −10 regions proposed by O'Neill (23). Therefore, while there is no good match to promoter −35 regions, we believe that this reading frame is expressed. Translation of this reading frame would give a polypeptide of 400 amino acids with a predicted molecular weight of 44,717 Daltons. We believe therefore that this reading frame is the structural gene for the 45 kDa protein seen in Figure 2 to be encoded by the 7.5

Table 3. Alignement of putative nucleotide binding sequence of SbcC with similar sequences in *E. coli* recombination proteins.

Protein	Residues		Sequence								Reference
SbcC	30–51	NGLFA	I	T	G	PT	G	A	GKT	TLDAIC	
RecA	59–80	GRIVE	I	Y	G	PE	S	S	GKT	TLTLQVI	31
RecB	16–37	QGERL	I	E	A	SA	G	T	GKT	FTIAALY	32
RecD	164–185	RRISV	I	S	G	GP	G	T	GKT	TTVAKLL	33
RecN	22–43	SGMTV	I	T	G	ET	G	A	GKS	IAIDALG	34
RuvB	55–76	LDHLL	I	F	G	PP	G	L	GKT	TLANIVA	35

kb *Eco*RI(1) to *Eco*RI(2) fragment in pJP71, PIN509 and pIN510. We shall refer to this gene as *orf-45*.

The last codon (GCA) for *orf-45* and the ATG start codon for *sbcC* overlap by 1 bp (Fig. 3). In the absence of any promoter sequences for independent transcription of *sbcC*, it seems most likely that these two genes form a single operon transcribed from the putative promoter region identified upstream of *orf-45*. Furthermore, since there is no obvious site for ribosome binding at the appropriate distance upstream of *sbcC*, we assume that expression of *sbcC* depends also on the ribosomes translating *orf-45* being relocated from the TGA stop codon for this reading frame to the ATG start codon for *sbcC*. Since SbcC protein is produced in significantly lower amounts than the 45 kDa protein ((Fig. 2), we assume that this relocation cannot be very efficient.

Analysis of the sequence at the 3′ end of the *sbcC* coding region revealed no regions of dyad symmetry that could act as transcriptional terminators. We were surprised to discover that a sequence of 379 nucleotides extending across the 3′ end of *sbcC* from bp 4303 to bp 4681 matches almost perfectly (98.9%) a sequence of 380 bp thought by Kosiba and Schleif (24) to come from the *araFG* region at 44.8 min on the genetic map, but which they conceded might come from elsewhere. The latter appears to be true since our sequencing strategy established that the 379 bp region is in our case contiguous with the flanking sequences. Kosiba and Schleif identified the sequence because of its ability to act as an arabinose inducible promoter and showed transcription beginning some 30 bp downstream of the TAA stop codon we identify as the end of *sbcC*. Examination of the sequence reveals a possible −10 region (bp 4467−4472) and an open reading frame extending to the end of the sequence determined, though it is not clear where translation would start. We assume therefore that the region downstream of *sbcC* defines the beginning of an arabinose inducible operon. Whether this operon has anything to do with arabinose metabolism is another matter.

Further searches of the GenBank (release 59) DNA sequence database revealed no other significant similarities with the 5125 bp sequence determined. We did find two Chi sequences (5′GCTGGTGG3′, 25) in the reverse complement (underscored in Fig. 3), which is consistent with the average frequency of this site (one for every 5 to 10 kb of DNA) in the *E. coli* chromosome (26).

Codon usage and amino acid composition
The maxicell analyses presented in Figure 2 suggested that *sbcC* is poorly expressed. A low level of expression would certainly be consistent with the apparent lack of any bias against the use of rare codons (27,28). The codons ATA (Ile), TCG (Ser), CCT and CCC (Pro), ACG (Thr), CAA (Gln), AAT (Asn), AGG (Arg) occur with a frequency of 12.4% in the *sbcC* reading frame, compared with frequencies of 12.4% and 8.4% in the non-

coding frames. A similar situation is seen with *orf-45* which has 9.3% of these rare codons in the reading frame compared with 11.8% and 7.8% in the non-coding frames.

The amino acid composition suggests that SbcC, like several other *E. coli* proteins involved in recombination, contains a nucleotide-binding fold (29,30) represented by the sequence GXXGXGKT (Table 3). Chou−Fasman predictions (37) for the secondary structure of SbcC suggest that the amino acids immediately before the fold could form a β-sheet, while those immediately after could form a strong α-helix, which matches exactly the structure suggested for an ATP binding site by Bradley et al. (30). Otherwise, the composition of SbcC appeared not unlike that of a typical globular protein. A search of the NBRF-PIR (release 20) protein database by the methods of Lipman and Pearson (38) revealed some of the highest optimal alignment scores with the products of the *E. coli* *recB* (32), *recN* (34), and *uvrD* (39) genes, and phage T4 gene *46* (36), though the initial scores (38) were not very remarkable (data not shown). We assume that at least some of this similarity is related to the homology at the putative nucleotide binding sites (Table 3; 39).

The 45 kDa product of *orf-45* showed no particularly striking features except for several long stretches of amino acids predicted to form β-sheets. A search of the NBRF-PIR protein database in this case produced the highest optimal alignment score, with another T4 exonuclease, the product of gene *47* (36), with RecC (40) coming a close second. Again, the similarities are not high, but may be significant (see Discussion) in view of the similarities found between SbcC and RecB.

DISCUSSION

We have shown that the *sbcC* gene spans a 3.3 kb region of DNA between *aroM* and *phoB* that encodes a poorly expressed protein of 118 kDa. It lies immediately downstream of a previously unknown gene (*orf-45*) that encodes a protein of 45 kDa. These two genes appear to form a single operon transcribed in a counterclockwise direction.

A previous attempt at the molecular analysis of *sbcC* was frustrated by the instability of multi-copy plasmids in a *recBC sbcBC* genetic background (1). We had the advantage of a simple and very reliable test for *sbcC* activity in *recBC*+ *sbcB*+ strains. Nevertheless, the present studies were not without their difficulties. The 4.3 kb *Nru*I(1) to *Nru*I(2) and 3.9 kb *Nru*I(1) to *Pst*I(1) fragments cloned in the low copy-number constructs pGTI27 and pGTI28, respectively, express *sbcC* despite the absence of the normal upstream sequences. Yet, neither fragment could be inserted stably into pBR322, pACYC184, or pUC18. This failure cannot be a simple effect of plasmid copy-number since *sbcC* is clearly present in the multi-copy constructs pJP71, pJP77 and pIN509. We assume therefore that other sequences present either in the 4.3 kb *Bam*HI(1) to *Eco*RI(1) region common to pJP71, pJP77 and pIN509, or in the vicinity of the *phoBR*, help to stabilize *sbcC* in these constructs. Whether or not this has anything to do with the other locus identified within the 4.3 kb *Bam*HI(1) to *Eco*RI(1) fragment by Lloyd and Buckman (1) is not clear.

The *sbcC* locus was first discovered in connection with the suppression of *recBC* mutations (1). We were intrigued therefore to discover some similarity between the products of *sbcC* and *recB* on the one hand and the products of *orf-45* and *recC* on the other. In neither case is the similarity great, but given that *recC* is located very close to and upstream of *recB* (32,40), they raise the possibility that the two sets of genes may have evolved from common ancestors and that their products have retained some of the same activities. The fact that both SbcC and RecB are large proteins and have similar nucleotide binding folds

lends further support to this possibility, as does the fact that both genetic regions are expressed rather poorly (32,33,40,41).

The idea of common ancestry is consistent with the fact that *sbcC* at 9 min on the genetic map (1) is located approximately 180° removed from the *recBC* genes at 61 min since there is some evidence to suggest that the *E. coli* chromosome has evolved by two genome duplications (42). The major differences between the two regions is that *recC* is separated from *recB* by *ptr* (32,40). The *recC* gene is also much larger than *orf-45*. Since the 45 kDa *orf-45* product aligned best with the central portion of RecC, it may be that the *orf-45-sbcC* region suffered a deletion.

S. K. Kulkarni and F. W. Stahl (Genetics, in press) have independently suggested a relationship between *sbcC* and *recB*. They have found evidence which indicates that the *sbcC* product interacts with the *gam* product of phage λ. Since Gam protein is known to interact with RecBCD enzyme, they suggest that SbcC may have a nuclease activity that is functionally related to one of the nuclease activities of RecBCD enzyme (43). Our results are consistent with this view but do not rule out the alternative possibility that the relationships detected are due to convergence. The latter would account for the fact that similarities of equal or greater magnitude were detected with phage T4 exonucleases.

ACKNOWLEDGEMENTS

REFERENCES

1. Lloyd, R.G. and Buckman, C. (1985) J. Bacteriol. **164**, 836–844.
2. Chalker, A.F., Leach, D.R.F. and Lloyd, R.G. (1988) Gene **71**, 201–205.
3. Leach, D.R.F. and Stahl, F.W. (1983) Nature **305**, 448–451.
4. Shurvinton, C.E., Stahl, M.M. and Stahl, F.W. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 1624–1628.
5. Lindsey, J.C. and Leach, D.R.F (1989) J. Molec. Biol. **206**, 779–782.
6. Bachmann, B.J. (1987), In: Neidhardt, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M. and Umbarger, H.E. (Eds), *Escherichia coli* and *Salmonella typhimurium* cellular and molecular biology. American Soc. for Microbiol., Washington DC, pp 1190–1219.
7. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene **33**, 103–119.
8. Mahdi, A.A. and Lloyd, R.G. (1989) Mol. Gen. Genet. **216**, 503–510.
9. Tommassen, J., Hiemstra, P., Overduin, P. and Lugtenberg, B. (1984) Molec. Gen. Genet. **195**, 190–194.
10. Boidol, W., Simonis, M., Topert, M. and Siewert, G. (1982) Molec. Gen. Genet. **185**, 510–512
11. Hashimoto-Gotoh, T., Franklin, F.C.H., Nordheim, A. and Timmis, K.N. (1981) Gene **16**, 227–235.
12. Silhavy, T.J., Berman, M.L. and Enquist, L.W. (1984) Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
13. Isch-Horowicz, D. and Burke J.F. (1981) Nucleic Acids Res. **9**, 2989–2998.
14. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
15. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467.
16. Sancar, A., Hack, A.M and Rupp, W.D. (1979) J. Bacteriol. **137**, 692–693.
17. West, S.C. and Emmerson, P.T. (1977) Mol. Gen. Genet. **151**, 57–67.
18. Cohen, A. and Clark, A.J. (1986) J. Bacteriol. **167**, 327–335.
19. Makino, K., Shinagawa, H., Amemura, M. and Nakata, A. (1986) J. Molec. Biol. **190**,37–44.

20. Marinus, M.G. (1987), In: Neidhardt, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M. and Umbarger, H.E. (Eds), *Escherichia coli* and *Salmonella typhimurium* cellular and molecular biology. American Soc. for Microbiol., Washington DC, pp 697–702.
21. Shine, J., and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA **71**, 1342–1346.
22. Harley, C.B., and Reynolds, R.P. (1987) Nucl. Acids Res. **15**, 2343–2361.
23. O'Neill, M.C. (1989) J. Biol. Chem. **264**, 5522–5530.
24. Kosiba, K.E. and Schleif, R. (1982) J. Molec. Biol. **156**, 53–66.
25. Smith, G.R., Kunes, S.M., Schultz, D.W., Taylor, A.F. and Triman, K.L. (1981) Cell **24**, 429–436.
26. Malone, R.E., Chattoraj, D.K., Faulds, D.H., Stahl, M.H. and Stahl, F.W. (1978) J. Molec. Biol. **121**, 473–491.
27. Konigsberg, W., and Godson, G.N. (1983) Proc. Natl. Acad. Sci. USA **80**, 687–691.
28. Sharp, P.M. and Li, W. (1987) Nucl. Acids Res. **15**, 1281–1295
29. Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J., (1982) EMBO J. **1**, 645–51.
30. Bradley, M.K., Smith, T.F., Lathrop, R.H., Livingston, D.M. and Webster, T.A. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 4026–4030.
31. Sancar, A., Stachelek, C., Konigsberg, W., and Rupp, W.D. (1980) Proc Natl. Acad. Sci USA **77**, 2611–2615.
32. Finch, P.W., Storey, A., Chapman, K.E., Brown, K., Hickson, I.D. and Emmerson, P.T. (1986) Nucl. Acids Res. **14**, 8573–8582.
33. Finch, P.W., Storey, A., Brown, K., Hickson, I.D. and Emmerson, P.T. (1986) Nucl. Acids Res. **14**, 8583–8594.
34. Rostas, K., Morton, S.J., Picksley, S.M. and Lloyd, R.G. (1987) Nucl. Acids Res. **13**, 5041–5049.
35. Benson, F.E., Illing, G.T., Sharples, G.J. and Lloyd, R.G. (1988) Nucl. Acids Res. **16**, 1541–1549.
36. Gram, H. and Ruger, W. (1985) EMBO J. **4**, 257–264.
37. Chou, P.Y., and Fassman, G.D. (1978) Adv. Enzymol. **47**, 45–148.
38. Lipman, D.J. and Pearson, W.R. (1985) Science **227**, 1435–1441.
39. Finch, P.W. and Emmerson, P.T. (1984) Nucl. Acids. Res. **12**, 5789–5799.
40. Finch, P.W., Wilson, R.E., Brown, K., Hickson, I.D., Tompkinson, A.E. and Emmerson, P.T. (1986) Nucl. Acids Res. **14**, 4437–4451.
41. Hickson, I.D., Robson, C.N., Atkinson, K.E., Hutton, L.W. and Emmerson, P.T. (1985) J. Biol. Chem. **260**, 1224–1229.
42. Riley, M. and Anilionis, A. (1978) Ann. Rev. Microbiol. **32**, 519–560.
43. Smith, G.R. (1988) Microbiol. Rev. **52**, 1–28.