The nucleotide sequence of a 2224 bp region of the *Escherichia coli* chromosome that carries the LexA regulated recN gene has been determined. A region of 1701 nucleotides encoding a polypeptide of 567 amino acids with a predicted molecular weight of 63,599 was identified as the most probable sequence for the recN structural gene. The proposed initiation codon is preceded by a reasonable Shine-Dalgarno sequence and a promoter region containing two 16 bp sequences, separated by 6 bp, that match the consensus sequence (SOS box) for binding LexA protein. DNA fragments containing this putative promoter region are shown to bind LexA in vitro and to have LexA-regulated promoter activity in vivo. The amino acid sequence of RecN predicted from the DNA contains a region that is homologous to highly conserved sequences found in several DNA repair enzymes and other proteins that bind ATP. A sequence of 9 amino acids was found to be homologous to a region of the RecA protein of *E. coli* postulated to have a role in DNA/nucleotide binding.

**INTRODUCTION**

The recN gene of *Escherichia coli* is a component of the SOS network of genes regulated by LexA protein (1). The gene was first identified in a recBC sbcBC background as a Mudl(Aplac) insertion which led to a reduction in the efficiency of conjugational recombination and in which expression of lacZ was found to be increased following exposure to the DNA crosslinking agent, mitomycin C (2). Subsequent studies identified recN point mutations (3) and Tn5 insertions (4). Analysis of recN::lacZ fusion strains showed that recN is regulated directly by LexA and that its expression can be induced by ultraviolet light, gamma-radiation, and nalidixic acid, in addition to mitomycin C (5). These studies also revealed a correlation in certain strains between the efficiency of recombination and the uninduced level of recN expression. Thus, expression was found to be much reduced in recombination deficient recBC mutants, but was greatly elevated in recBC sbcBC strains where the combination of sbcB (6,7) and sbcC (8) mutations restores recombination to the level of rec+ strains. Furthermore, a lexA(Ind−) mutation, which prevents induction of SOS genes and severely curtails recombination in recBC sbcBC strains (9), reduced expression of recN to the extent that it was no longer detectable. These results,
together with the fact that recN single mutants are highly sensitive to gamma-irradiation and appear unable to repair DNA double-strand breaks (3), suggest that recN specifies an inducible activity involved in recombinational repair of damaged DNA.

A further understanding of recN and of its involvement in DNA repair and recombination will require a knowledge of the recN coding sequence, purification of the gene product, and a study of its properties in vitro. As a first step in this study, we described previously the molecular cloning of recN and the identification of its product as a protein of approximately 60 kDa (10). Here, we report the complete nucleotide sequence of recN and identify promoter sequences involved in its regulation by LexA.

METHODS

Strains and plasmids

The source of recN was pSP100 which carries the gene on a 5.6 kb HindIII fragment of chromosomal DNA (Fig. 1) cloned into the low copy-number vector pHSG415 (10). pUC18 and pUC19 (11) were used to subclone DNA fragments from pSP100 for further analysis. pACYC184 (Tc\textsuperscript{r}, Cm\textsuperscript{r}) was described by Chang and Cohen (12). pKK232-8 (Ap\textsuperscript{r}) carries a promoterless chloramphenicol acetyltransferase (CAT) gene and was used to select DNA fragments containing active promoters (13). pSP103 is a Cm\textsuperscript{r} recombinant of pKK232-8 carrying the 0.9 kb PstI fragment from pSP100. The 0.9 kb PstI fragment was first cloned into the polylinker of pUC18 from which it was then removed by cleavage at the flanking BamHI and HindIII sites and inserted into pKK232-8 cut with the same two enzymes. This construction placed the CAT gene downstream of the promoter and N-terminal coding region of recN (see Fig. 2). Strains JM101 and JM109 were used to grow the phage cloning vectors M13mp18 and M13mp19 and their recombinants and to detect recombinants of pUC18 and pUC19 (11). AB1157 (14) is a standard rec\textsuperscript{+} lex\textsuperscript{+} strain. AB2463 (recA13), DM49 (lexA3), and DM1187 (lexA51 sulA11) are closely related derivatives of AB1157 (14,15).

Enzymes and proteins

Restriction endonucleases, T4 DNA ligase, and the Klenow fragment of DNA polymerase I were purchased from commercial sources and used as recommended by the suppliers. LexA protein, purified essentially as described by Schnarr et al. (16), was a gift from P. Strike.

DNA sequence analysis

Previous studies (10) suggested that the recN gene is encoded within the 2.1 kb BglIII-AvaI fragment of the DNA insert cloned in pSP100 (Fig. 1). Appropriate restriction fragments from this region were inserted into M13mp18 and M13mp19 and
Figure 1.
Genetic location and partial restriction map of the *E. coli* chromosomal region encoding recN. The arrow shows the orientation of recN within the 5.6 kb HindIII DNA fragment cloned into pSP100.

sequenced by the dideoxy chain termination method of Sanger *et al.* (17) using a synthetic 17-mer universal primer and [$\alpha^{35}$S] dATP (Amersham). More than 90% of the sequence was determined from both strands of the DNA.

**Assay of LexA—DNA binding by gel retardation**

A 565 bp Sau3A fragment (nucleotides 109–674, Fig. 2) containing the promoter region for recN cloned at the BamHI site of the polylinker in M13mpl8 was excised from RF DNA by cutting at the flanking EcoRI and HindIII sites. The EcoRI end was partially filled in using Klenow polymerase in the presence of [$\alpha^{35}$S] dATP prior to digestion with HindIII. The appropriate end-labeled fragment was isolated by electroelution from polyacrylamide gels. Binding of LexA was measured by the gel retardation assay of Fried and Crothers (18). Approximately $1\mu$g of labeled DNA was mixed with LexA protein in 10mM Tris pH 7.4, 1mM EDTA, 50mM KCl, 100μg/ml BSA, and incubated for 20 min at 21°C. As a control, LexA was reacted with a 530 bp SalI-HindIII fragment from the *tet* gene of pACYC184 labeled in a similar fashion at the HindIII end. Reaction mixtures were diluted with 0.2 vol loading dye (0.25% bromophenol blue, 40% glycerol, 10mM Tris, 1mM EDTA, pH 7.4), before loading onto 5% polyacrylamide gels (4.94% acrylamide, 0.06% bisacrylamide, 10mM Tris, 1 mM EDTA, pH 7.4). Electrophoresis was at 10V/cm with buffer recirculation.

**Assay of chloramphenicol resistance and CAT activity**

Strains carrying pSP103 were grown in LB broth medium containing 20 μg/ml ampicillin to a density of about 2 x $10^8$ cells/ml. Samples of various dilutions were spread on LB agar plates containing 20 μg/ml ampicillin and various concentrations of chloramphenicol ranging from 0–100 μg/ml. Colonies were scored after incubation for 24 hours at 37°C and resistance to chloramphenicol was taken as the highest concentration that caused no significant reduction in colony number relative to the control. For measures of CAT activity, cells grown in a similar fashion were pelleted by centrifugation, resuspended in the same volume of 50 mM Tris-HCl, 0.03 mM DTT, pH 7.8, and sonicated on ice. Debris was removed by
Figure 2.
Nucleotide sequence of the recN coding region and flanking sequences. Restriction sites for the 565 bp Sau3A fragment and the 0.9 kb and 1.4 kb PstI fragments referred to in the text are underlined. The proposed coding sequence for recN begins with ATG at bp 391. Nucleotides forming part of a putative ribosome binding site are asterisked. Sequences in the region upstream of the recN coding region that are homologous with LexA binding sites are overlined.

5044
centrifugation and CAT activity in the supernatant fraction was measured by the
spectrophotometric assay described by Shaw (19). SOS induction was achieved by
growing the cells in the presence of 0.1 μg/ml mitomycin C (MC) for 2 hours before
making extracts.

RESULTS AND DISCUSSION
Identification of the DNA sequence encoding recN

Figure 2 shows the sequence of the 2224 bp region of the E. coli chromosome
corresponding to the adjacent 0.9 kb and 1.4 kb PstI fragments in pSP100 (Fig. 1) and
is numbered from the left end of the 0.9 kb fragment. The first 268 nucleotides
reveal a possible coding region terminated by a TAA codon and followed by a short
stretch of GC rich region and a 17 bp dyad symmetry between bp 294 and 310. This is
followed by an open reading frame beginning with an ATG codon at bp 391 and extending
1701 nucleotides to a TAA termination codon at bp 2092. We believe this to be the
recN structural gene. A possible ribosome binding site (20) is provided by a region
containing the sequence AGGA which appears 7 nucleotides upstream of the ATG. The
putative recN coding region would direct the synthesis of a polypeptide of 567 amino
acids with a predicted molecular weight of 63,599, which is in reasonable agreement
with the estimates of 60,000-62,000 obtained for RecN by SDS-PAGE (10,21). However,
there is another ATG codon in the same reading frame at bp 460. Initiation at this
point would produce a smaller polypeptide of 544 amino acids with a molecular weight
of 61,012. Until the amino-terminal sequence of RecN is determined experimentally
we cannot exclude the possibility that initiation begins at this second ATG.

Codon usage and amino acid composition of RecN

Analysis of codon usage within recN, based on initiation at bp 391, revealed
that the rare codons (22), ATA (Ile), TCG (Ser), CCT (Pro), CCC(Pro), ACG (Thr), CAA
(Gln), AAT (Asn), AGG (Arg), occur at a frequency of 8.1% in the recN coding frame,
and at 11.8% and 9.5% in the non-coding frames. The frequency of CAA (Gln) is
exceptionally high in the coding frame (3.88% of all codons, and 47.8% of rare
codons) which masks the otherwise low frequency (4.1%) of rare codons. The
predicted sequence of 567 amino acids would give a protein with an approximate net
charge at pH 7.0 of -16. This value is in good agreement with the slightly acidic
isoelectric point (pl = 5.8) determined by Finch et al. (21). The hydrophilicity
plot (not shown) of the polypeptide sequence follows the general pattern of globular
proteins. These analyses are indicative of a moderately expressed cytoplasmic
protein which would be consistent with the fact that RecN is normally produced in
fairly low amounts but can be induced quite rapidly in response to DNA damage (5,21).
The primary structure of RecN may determine a nucleotide binding fold

Since RecN appears to participate in recombinational repair of damaged DNA, it would be expected to have properties associated with DNA/nucleotide processing. A consensus sequence of amino acids thought to be concerned with nucleotide binding has been found in many proteins and is conserved in several repair enzymes with known ATPase activity (23, 24; I. Anton, personal communication). The sequence of RecN between residues 27 and 43 shows very good homology to the consensus sequence (Table 1), suggesting perhaps that the protein may have an adenine nucleotide binding site.

The protein sequence between residues 355 and 363 shows significant homology with residues 124-132 (α-3 helix) of RecA (25). 7 of the 9 residues, QALEIARAL in RecN and QALEICDAL in RecA, are identical. Nucleotide sequences in these regions are also highly conserved, with one stretch of 17 nucleotides (bp 1451-1467 in recN) being identical. The significance of this homology is not yet clear, but Blanar et al. (25) suggested that in RecA, the α-3 helix together with the ATP-binding site (α- 

### Table 1

<table>
<thead>
<tr>
<th>Protein</th>
<th>Sequence</th>
<th>Residues</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>UvrA</td>
<td>DKLIVVTCGLVLSGGKS GSGKSSLAFDTL YAEGQ</td>
<td>24 - 50</td>
<td>31</td>
</tr>
<tr>
<td>UvrB</td>
<td>GLFTCICTGVSGSGKSTSLLIDDTLPPIAQ</td>
<td>633 - 659</td>
<td>31</td>
</tr>
<tr>
<td>UvrD</td>
<td>LASGTVGSGKTFTIANVLADLQRI</td>
<td>32 - 58</td>
<td>32</td>
</tr>
<tr>
<td>RecA</td>
<td>GRIVEIYCGPESGKTTLTDQAALAIQ</td>
<td>59 - 85</td>
<td>33</td>
</tr>
<tr>
<td>RecB</td>
<td>QGERLIESASAAGKTFTIAALYLRLLLeal</td>
<td>16 - 42</td>
<td>23</td>
</tr>
<tr>
<td>RecD</td>
<td>RRSISGCTGTGKTTVAKLHLSLGAALI</td>
<td>164 - 190</td>
<td>34</td>
</tr>
<tr>
<td>RecN</td>
<td>SGMTVIYGCTGAGKSIAIDALGCLG</td>
<td>22 - 48</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Strain</th>
<th>Plasmid</th>
<th>Cm (ug/ml)</th>
<th>CAT activity (units/mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>- MC</td>
<td>+ MC</td>
</tr>
<tr>
<td>AB1157 (recA)</td>
<td>none</td>
<td>2.0</td>
<td>ND</td>
</tr>
<tr>
<td>AB1157 (recA)</td>
<td>pSP103</td>
<td>33.5</td>
<td>5.7</td>
</tr>
<tr>
<td>AB2463 (recA)</td>
<td>pSP103</td>
<td>7.0</td>
<td>1.4</td>
</tr>
<tr>
<td>DM49 (lexA(Ind-))</td>
<td>pSP103</td>
<td>5.0</td>
<td>2.5</td>
</tr>
<tr>
<td>DM1187 (lexA(Def))</td>
<td>pSP103</td>
<td>70.0</td>
<td>94.9</td>
</tr>
</tbody>
</table>

pKK232-8 confers no resistance to chloramphenicol beyond that found in the plasmid free strain AB1157. ND = not determined.
Figure 3.
Titration of the 566 bp Sau3A fragment containing the recN promoter (lanes a-h) and the 530 bp tet fragment of pACYC184 (lanes i-o) with LexA. LexA monomer concentrations are 0 (h,o), $10^{-10}$M (g,n), $5.0 \times 10^{-10}$M (f), $10^{-9}$M (e,m), $0.5 \times 10^{-8}$M (d,l), $10^{-8}$M (c,k), $0.5 \times 10^{-7}$M (b,j), and $10^{-7}$M (a,i). The concentration of the DNA fragments was $0.5 - 1.0 \times 10^{-8}$M. Reaction conditions and electrophoresis were performed as described in the Methods section. The arrow indicates the position of nonspecific DNA/protein complexes.

helix) form the DNA binding core of the protein. Sequences homologous to the QALEIARAL sequence in RecN were also found in organelle ATPases, the Cro protein of λ and in pyrophosphate transferases involved in nucleotide biosynthesis.

Analysis of the recN promoter

Previous studies have shown that expression of recN is regulated by the product of lexA (5). Examination of the nucleotide sequence upstream from the proposed translational start site reveals a possible promoter with a −10 region at bp 344 and a −35 region at bp 322. The putative promoter overlaps two 16 bp regions (positions 326-341 and 348-363) that match the consensus for sequences (SOS boxes) known to bind LexA protein (26,27,28). The presence of these SOS boxes in the promoter region suggests that recN is indeed regulated at the transcriptional level by LexA. To investigate this possibility we transformed pSP103 into AB1157 and its related lexA and recA derivatives and examined the Ap" transformants for their resistance to chloramphenicol and CAT activity. pSP103 contains the putative promoter region within a 0.9 kb Pst I fragment cloned upstream of an otherwise promoterless CAT gene (see Methods). The results summarised in Table 2 show that CAT expression is activated in pSP103 and is regulated by LexA, from which we conclude that the 0.9 kb PstI fragment does contain the recN promoter.
To demonstrate that LexA binds specifically to a DNA fragment containing the recN promoter we made use of the gel retardation assay described by Fried and Crothers (18). The results shown in Figure 3 reveal binding of LexA to a 565 bp Sau3A fragment containing the putative promoter (see Methods). Binding was observed at molar ratios of protein/DNA less than 10\(^{-2}\). At higher protein concentrations, binding becomes a two-phase process and results in a stepwise increase in the molecular weight of the protein/DNA complex until binding becomes non-specific. Under identical conditions, no specific complex was formed between LexA and a 530 bp DNA fragment from the tet gene of pACYC184. We have no direct evidence to suggest that the two SOS boxes have different affinities for LexA, though DNase I protection experiments have established that LexA does bind to both (data not shown). A comparison the SOS boxes in the recN promoter with similar sequences in the promoters of other SOS genes (Table 3) revealed that the 16 nucleotides in the first SOS box are identical to those in the promoter of clel, the structural gene for ColEl (29). LexA protein is known to have a particularly high affinity for the ColEl promoter (1,30). A similar affinity for the recN promoter might account for the tight regulation of recN expression observed in vivo (5).

ACKNOWLEDGEMENTS

This paper is dedicated to the memory of our colleague Andrew J. Smith who died recently in a road traffic accident.

We would like to thank P. Strike and Darerca Owen for the generous gift of purified LexA protein, I. Anton for communicating his unpublished analysis of ATP-binding protein sequences, M. Ginsburg for assistance with the computer analyses, and the Imperial Cancer Research Fund Laboratories for the use of their computing facilities and databases. We would also like to thank Anne Davies for her assistance with some of the experiments. This work was supported by a grant to R.G.L from the Science and Engineering Research Council.
REFERENCES


* Present address: Imperial Cancer Research Fund, Clare Hall Laboratories, South Mimms, Herts., UK