Volume 14 Number 11 1986 Nucleic Acids Research

Complete nucleotide sequence of the *Escherichia coli recC* gene and of the *thyA–recC* intergenic region

Paul W.Finch, Rosemary E.Wilson, Kate Brown, Ian D.Hickson[1], Alan E.Tomkinson* and Peter T.Emmerson

Departments of Biochemistry and [1]Clinical Oncology, The University, Newcastle upon Tyne NE1 4LP, UK

Received 26 March 1986; Revised and Accepted 12 May 1986

ABSTRACT

The nucleotide sequence of a 6,000 bp region of the E. coli chromosome that includes the 3' end of the coding region for the thyA gene and the entire recC gene has been determined. The proposed coding region for the RecC protein is 3369 nucleotides long, which would encode a polypeptide consisting of 1122 amino acids with a calculated molecular mass of 129 kDa. Mung bean nuclease mapping of a recC specific transcript produced in vivo indicates that transcription of recC is initiated 80 bp upstream of the translational start point. A weak promoter sequence situated 5' to the transcription initiation point has been identified. In the 1953 bp thyA-recC intergenic region there are three open reading frames that would code for polypeptides of molecular mass 30 kDa, 13.5 kDa and 12 kDa, respectively. Although the first and third of these open reading frames are preceded by possible ribosome binding sites, no obvious promoter sequences could be identified. Moreover, transcripts for these reading frames could not be detected.

INTRODUCTION

Mutations in the recB and recC genes of E. coli lead to a deficiency in genetic recombination and sensitivity to ultra-violet and ionising radiation (1). After irradiation, these mutants degrade their DNA less than wild type cells (1), presumably because they are deficient in an enzyme known as exonuclease V or the RecBC DNase (2-4). Preparations of this enzyme exhibit a wide variety of catalytic activities including ATP-dependent exonuclease, ATP-stimulated endonuclease, ATP-dependent helicase and DNA-dependent ATPase activities (5).

Homologous recombination in E. coli mediated by the RecBC pathway is locally stimulated by the presence of Chi sites (6). Chi, originally identified as mutations in bacteriophage lambda (7), is determined by a unique 8 bp sequence, 5'-GCTGGTGG-3' (8). A nuclease activity of the RecBC enzyme is required for Chi-dependent cleavage of one DNA strand, that containing the Chi sequence (9).

Phenotypically, recB and recC mutants cannot be distinguished on the

basis of recombination proficiency, radiation sensitivity or post-irradiation degradation of DNA, but were assigned to different genes on the basis of cotransduction frequencies with the neighbouring *thyA* and *argA* genes (1), and complementation analysis with F' merodiploids (10). This assignment was confirmed when the products of cloned *recB* and *recC* genes were identified as non-identical proteins of molecular mass 135 kDa and 125 kDa, respectively (11-13). It has recently been shown that *ptr*, the structural gene for protease III which encodes a protein of molecular mass 110 kDa, lies between the *recC* and *recB* loci, and that all three genes are independently transcribed (13).

It has been estimated that there are only 10 molecules of the RecBC DNase per cell (14). Analysis of the amount of β-galactosidase synthesised from Mud(*bla lac*) fusions to the *recB* and *recC* promoters, indicate that this low intracellular level can be attributed to an extremely low level of transcription of the *recB* and *recC* genes and that this level is not increased following damage to cellular DNA (15,16). This suggests that unlike several other gene products involved in DNA repair and genetic recombination (17,18), the RecBC DNase is not induced as part of the SOS response. This finding is in agreement with the observation that mutations in *lexA*, which prevent induction of the SOS response following damage to DNA, do not significantly alter the levels of conjugal recombination in *recBC*$^+$ strains (19,20).

As part of a study of the structure and control of the *recB* and *recC* genes we have determined their nucleotide sequences. We report here the sequence of the entire *recC* gene and that of a 2541 bp region preceding the translational start of *recC* which includes the 3' end of the *thyA* coding region. The precise transcriptional start point of *recC* has been located by mung bean nuclease mapping of RNA transcripts produced *in vivo*.

METHODS

Enzymes and Biochemicals

Restriction endonucleases were purchased from NBL Enzymes, New England Biolabs, and BCL: DNA polymerase I (Klenow fragment), T4 polynucleotide kinase, T4 DNA polymerase and mung bean nuclease from Pharmacia: calf intestinal phosphatase and T4 DNA ligase from BCL. Deoxynucleoside and dideoxynucleoside triphosphates were from Sigma. Radiochemicals were purchased from Amersham.

Bacterial Strains and Plasmids

The source of the *recC* gene was the plasmid pPE37 (15) which carries *recC*

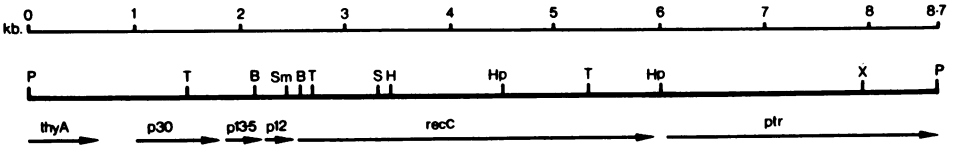


Figure 1 Restriction map of the 8.7 kb PstI fragment from pPE37. The positions of thyA, recC, ptr and the unidentified reading frames are shown. The direction of transcription of thyA (34), recC (13,15,38) and ptr (13) have been previously reported. Restriction endonuclease cleavage sites are abbreviated as follows: P, PstI: T, Tth111I: B, BstXI: S, SalI: Sm, SmaI: H, HindIII: Hp, HpaI: X, XhoI.

on a 8.7 kb PstI fragment (Figure 1). JM105 was used as host for the phage cloning vectors M13mp18 and mp19 (21) and their recombinants. The source of RecC protein was the strain M5219 (22) harbouring the plasmid pPE233, in which transcription of the recC gene is under control of the phage lambda leftward promoter (15). Total cellular RNA for hybridisation experiments was isolated from cells harbouring a recombinant plasmid containing recC subcloned into the runaway replication plasmid pSY343 (23). Thermal induction increases the plasmid copy number and therefore the number of transcripts initiated from recC promoters in vivo. This plasmid was constructed in this laboratory by A. Storey and will be described elsewhere.

DNA Sequence analysis

DNA sequence analysis was performed by the dideoxy chain termination method (24) using single-stranded DNA from clones of M13mp18 and mp19, a synthetic 17 base universal primer (Pharmacia) and [α^{35}S] dATP as radiolabel. The nucleotide sequence was determined by electrophoresis through 0.4 mm polyacrylamide buffer gradient gels (25) followed by exposure to Fuji RX X-ray film.

The sequence was built up by determining the sequences of random clones of the 8.7 kb PstI fragment of pPE37 generated by shearing the DNA randomly into fragments by sonication (26). After sonication, the fragments were incubated with T4 DNA polymerase and dNTPs to repair sheared ends (26) then size fractionated on a 1.5% agarose gel. Fragments of between 300 and 600 nucleotides in length were isolated from the gel by electroelution (27) and inserted into alkaline phosphatase treated, SmaI cleaved M13mp18 replicative form I DNA. The final sequence was demonstrated to be correct by sequencing specific DNA restriction fragments, subcloned into mp18 or mp19.

Computer programs developed by Queen and Korn (28) and Staden (29) were used to assemble and analyse the sequence.

Mapping of in vivo recC transcripts

RNA was isolated from cells by repeated extraction with phenol at 60°C (30). DNA restriction fragments were isolated from agarose gels by electroelution (27) and the 5' terminus labelled with [γ^{32}P] ATP using polynucleotide kinase. Formation of DNA-RNA hybrids was carried out essentially as described by Berk and Sharp (31). ^{32}P-labelled DNA restriction fragments and RNA (approximately 250 µg) were ethanol precipitated, dried and dissolved in 50 µl hybridisation buffer (0.4 M NaCl, 0.04 M PIPES, pH 6.4, 0.001M EDTA in 80% formamide). The solution was incubated at 85°C for 15 minutes to denature duplex DNA and then at 56°C for 3 hours to allow hybrid formation. The products of hybridisation were diluted ten-fold into ice-cold 30 mM Na acetate (pH 4.6), 50 mM NaCl, 1 mM $ZnCl_2$ and incubated for 30 minutes at 37°C with 50 units of mung bean nuclease. After phenol extraction and isopropanol precipitation the protected DNA fragments were analysed on 1.0% alkaline agarose gels (27). Transcripts were accurately mapped by electrophoresis through 6.0% polyacrylamide, 8M urea sequencing gels (32) in parallel with Maxam and Gilbert sequencing ladders (33) of the original restriction fragment.

Isolation of RecC protein and NH_2-terminal sequence determination

The RecC protein was purified to homogeneity as previously described (15). The NH_2 terminal amino acid sequence determination was kindly performed by Prof. Fothergill of the Department of Biochemistry, University of Aberdeen.

RESULTS

Nucleotide sequence

The sequence of a 6,000 bp region of the E. coli chromosome, carrying the 3' end of the thyA coding sequence and the entire recC gene is shown in Figure 2. The first 757 nucleotides, which includes the carboxyl-terminal sequence (bp 1 to bp 607) and putative transcription termination sequence of the thyA gene, has been previously reported (34). Our sequence of this region agrees exactly with that reported.

The recC coding sequence begins with an ATG initiation codon at bp 2542 and extends for 3369 nucleotides until a termination codon, TGA, at bp 5910. This would direct the synthesis of a protein 1122 amino acids long with a molecular mass of 128,860. The sequence of the NH_2-terminus of the RecC protein was determined to be NH_2-Met-Leu-Arg-Val-Tyr-His-Ser-Asn-Arg-Leu-Asp-Val-, which corresponds exactly with the sequence predicted from the

nucleotide sequence. A ribosome binding site, AGGAG (35) precedes the ATG initiation codon by five nucleotides.

An examination of the sequence of the thyA-recC intergenic region reveals three open reading frames. The first begins at an ATG at bp 1014 and continues for 804 nucleotides until a TGA termination codon at bp 1817. This region would code for a putative polypeptide 267 amino acids in length with a molecular mass of 29,666. A ribosome binding site, GGAGG, beginning at bp 1002 precedes the ATG codon.

A 363 nucleotide open reading frame extends from an ATG at bp 1856 to a TGA stop codon at bp 2221 which would direct the synthesis of a polypeptide of molecular mass 13,476. However, there is no obvious ribosome binding site preceding this region.

The third open reading frame directly precedes the recC structural gene and overlaps the second unidentified open reading frame by 15 bp. It begins with an ATG codon at bp 2206 and continues for 324 nucleotides until a TAG codon at bp 2529. This reading frame is preceded by a ribosome binding site, GAGG, beginning at bp 2197 and would encode a polypeptide of molecular mass 12,005.

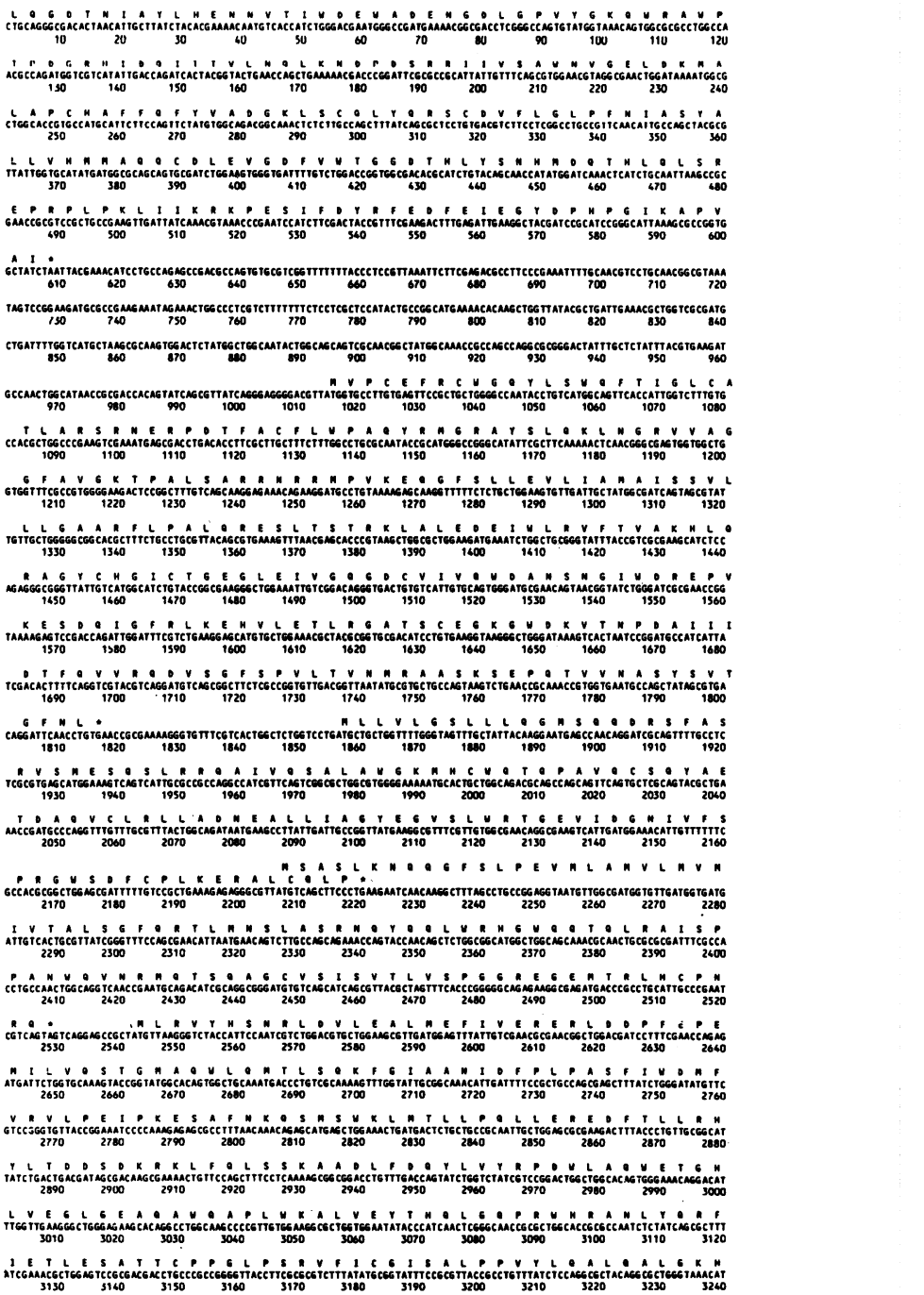
There are no long open reading frames coded for by the opposite DNA strand in the thyA-recC intergenic region, or in the recC coding region.

Localisation of the recC promoter

There are no sequences preceding the recC gene or the open reading frames in the thyA-recC intergenic region, that strongly fit the canonical E. coli promoter -10 (TATAAT) and -35 (TTGaca) sequences (36). Therefore, S1 nuclease mapping (31) was used to locate the transcriptional start point of the recC gene and also to determine whether the other putative coding sequences are transcribed. In these experiments mung bean nuclease was substituted for S1 nuclease because it has been reported that it gives single bands in transcript mapping experiments whereas S1 nuclease will often give multiple bands (37).

A PstI-Tth111I DNA fragment (bp 1 to bp 1503), ^{32}P end-labelled at Tth111I site located within the 804 nucleotide open reading frame, is not protected against digestion by mung bean nuclease (Figure 3, lane 2).

However, a ^{32}P end-labelled Tth111I fragment (bp 1503 to bp 2685), which overlaps the start of the recC coding sequence by 143 nucleotides, when hybridised with total cellular RNA is protected against digestion from mung bean nuclease for approximately 250 nucleotides (Figure 3, lane 4). This is consistent with the S1 mapping results of Sasaki et al. (38) which indicated



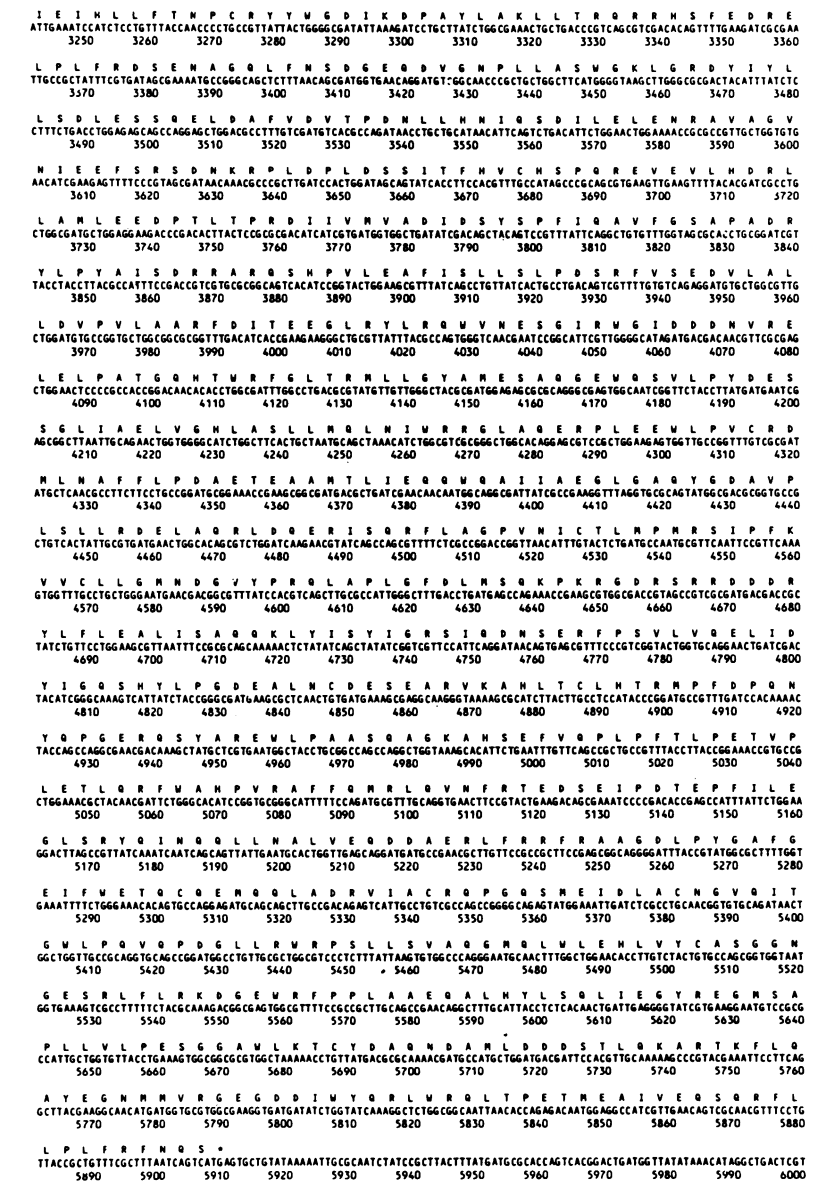


Figure 2 Sequence of the recC gene and 5′ flanking region. The nucleotides are numbered from the PstI (CTGCAG) cleavage site. The recC gene and deduced amino acid sequence begins at position 2542. The carboxyl-terminal sequence of thyA extends from bp 1 to bp 607. The predicted amino acid sequences of three unidentified open reading frames (bps 1014 to 1817, 1856 to 2221 and 2206 to 2529) are also given.

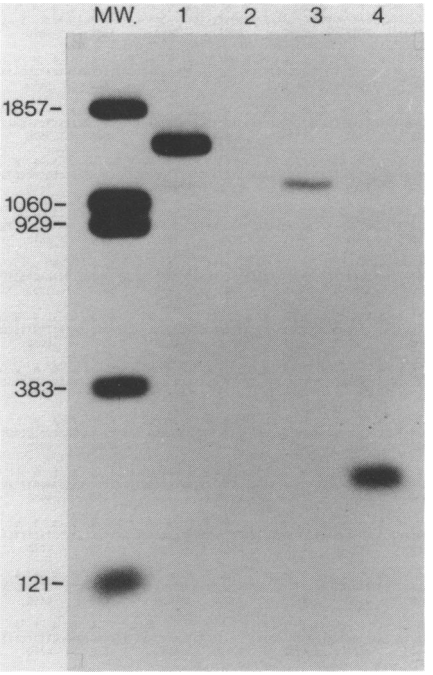


Figure 3 Electrophoretic analysis of mung bean nuclease digested hybrids formed between total cellular RNA and ^{32}P 5' end-labelled DNA restriction fragments from pPE37. Two fragments were used, a 1503 bp PstI-Tth111I fragment (lane 1) which overlaps the 804 nucleotide open reading frame by 389 nucleotides, and a 1182 bp Tth111I fragment (lane 3) which overlaps the start of the recC coding region by 143 nucleotides. Labelled fragments were hybridised to RNA and digested with mung bean nuclease (lanes 2 and 4). Molecular weight markers were 5' ^{32}P end-labelled BstNI fragments of pBR322.

that transcription of recC is initiated approximately 1 kb upstream of the HindIII site at position 3454.

To map accurately the 5' end of the recC specific transcript, a 399 bp ^{32}P end-labelled BstXI fragment (bp 2162 to bp 2561) was used as a hybridisation probe. Mung bean nuclease resistant DNA fragments were electrophoresed in parallel with Maxam and Gilbert sequencing ladders of the original fragment such that the 5' end of the transcript could be read directly from the sequencing gel (Figure 4). The cleavage fragments represent a set of 5' end-labelled fragments, up to but not including the base in the sequence which is eliminated by the cleavage reaction (33). A correction of one base is therefore required to arrive at the exact 5' terminus of the RNA transcript (39). Transcription appears to be initiated at a T residue,

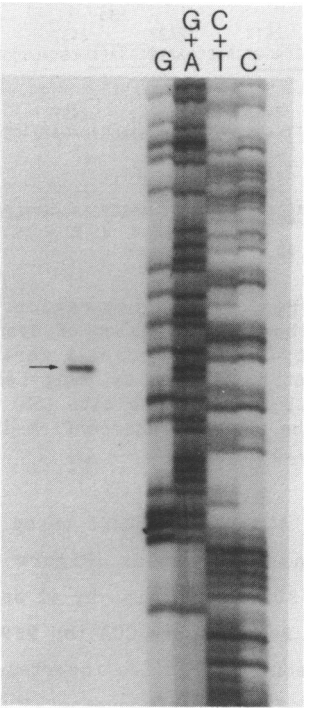


Figure 4 Mapping of 5'-termini of recC specific RNAs. Autoradiogram of mung bean nuclease resistant hybrids formed between a 399 bp BstXI DNA fragment and total cellular RNA. The protected fragment (indicated by an arrow) is shown callibrated against Maxam and Gilbert sequencing ladders of the same fragment.

position 2463 in the sequence. This is somewhat unusual for E. coli in which transcription is usually initiated at a purine residue (36). Close examination of the autoradiograph revealed faint bands corresponding to positions T2462, A2464, T2468, A2469 and at three T residues at positions 2471 to 2473 in the sequence. It seems unlikely that all of these represent transcriptional start points and instead may reflect intracellular RNA degradation occuring during thermal induction of the runaway replication plasmid. The sequences CAGCAT (bp 2451 to bp 2456) and CAGACA (bp 2428 to bp 2433) could possibly act as the -10 and -35 sequences of the recC promoter (Figure 5). Although these sequences display only limited homology to the canonical promoter sequences (36) no other sequences with a better correlation could be identified.

Further analysis of this region reveals a region of dyad symmetry

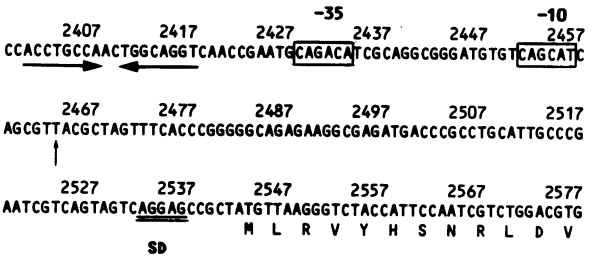


Figure 5 Sequence of the recC regulatory region. The putative -10 and -35 sequences are enclosed by boxes. The region of dyad symmetry preceding the -35 sequence is depicted by facing arrows. A vertical arrow indicates the transcriptional start point determined by mung bean nuclease mapping of in vivo RNA. The predicted ribosome binding site (SD) is indicated. The first 12 amino acids predicted by the sequence and confirmed by $-NH_2$ terminal sequence analysis of the RecC protein are shown.

consisting of the sequence ACCTGCCA (bp 2400 to bp 2407) followed 3 bp later by TGGCAGGT in a 8 bp inverted repeat (Figure 5). Preceding the coding sequence of the putative 30 kDa protein by 42 bp there is a 6 bp inverted repeat consisting of the sequence ATGCCA (bp 959 to bp 964) followed 3 bp later by TGGCAT. A comparison of these two inverted repeats reveals that they both contain the identical 12 bp core sequence, TGCCAACTGGCA.

Codon usage and amino acid composition.

In E. coli there is a strong preference in efficiently-expressed genes for codons that correspond to most abundant tRNA species (40) and there is a

Table 1. Codon usage in the E. coli recC gene.

F	TTT	33.	S	TCT	4.	Y	TAT	23.	C	TGT	6.
F	TTC	17.	S	TCC	11.	Y	TAC	12.	C	TGC	8.
L	TTA	18.	S	TCA	10.	*	TAA	0.	*	TGA	1.
L	TTG	20.	S	TCG	6.	*	TAG	0.	W	TGG	27.
L	CTT	14.	P	CCT	10.	H	CAT	16.	R	CGT	38.
L	CTC	15.	P	CCC	5.	H	CAC	6.	R	CGC	26.
L	CTA	12.	P	CCA	12.	Q	CAA	25.	R	CGA	6.
L	CTG	75.	P	CCG	40.	Q	CAG	52.	R	CGG	8.
I	ATT	24.	T	ACT	7.	N	AAT	7.	S	AGT	11.
I	ATC	23.	T	ACC	17.	N	AAC	23.	S	AGC	27.
I	ATA	3.	T	ACA	5.	K	AAA	18.	R	AGA	1.
M	ATG	29.	T	ACG	8.	K	AAG	5.	R	AGG	3.
V	GTT	15.	A	GCT	11.	D	GAT	41.	G	GGT	18.
V	GTC	11.	A	GCC	23.	D	GAC	33.	G	GGC	25.
V	GTA	3.	A	GCA	15.	E	GAA	60.	G	GGA	8.
V	GTG	24.	A	GCG	36.	E	GAG	24.	G	GGG	12.

bias against the use of certain other codons in the reading frame. In particular, the codons ATA (Ile), TCG (Ser), CAA (Gln), AAT (Asn), CCT and CCC (Pro), ACG (Thr) and AGG (Arg) occur approximately three times more frequently in the noncoding frames than in the coding frame (11% and 10% versus 4%) (41). However, in certain genes which encode products present in only a few copies per cell the codon usage deviates from the usual pattern, and rare codons are found with an equal frequency in all three reading frames. It has been proposed that by increasing the frequency of rare codons in the reading frame, the rate of translation is decreased. This may be one mechanism involved in the regulation of gene expression.

It has been estimated that there are 10 copies of the RecBC DNase per cell and it is possible that codon usage might play a role in maintaining this low level. The codon usage for recC is given in Table 1. The rare codons occur as 6.0% of the total codons in the coding frame and 11.0% and 9.7% in the noncoding frames. It seems unlikely therefore that expression of recC is regulated at the level of translation.

In the other potential coding sequences the rare codons occur at a frequency of 8.6, 6.6 and 8.4%, respectively in the coding frame and at 8.2, 8.3 and 4.6%, and 6.4, 6.6 and 9.7%, respectively in the noncoding frames. This distribution does not conform with the usual pattern (41) and may indicate that these sequences are noncoding.

From the predicted amino acid sequence, the RecC protein contains 23.4% charged amino acids. This consists of 14.1% acidic residues (Asp_{74}, Glu_{84}) and 9.3% basic residues (Arg_{82}, Lys_{23}) and represents a net charge of -53. The isoelectric point of purified RecC protein has been estimated to be approximately 5.5 (16,42) in good agreement with the predicted value.

Sequence homolgy between recC and recB

A standard matrix comparison (28) between the recC nucleotide sequence and that of recB (manuscript in preparation) revealed no significant homology between the two sequences. A similar result was obtained on comparing the predicted amino acid sequences of RecC and RecB. This is consistent with the results of Dykstra et al (13) who found that there was no hybridisation between clones of recC and recB.

DISCUSSION

We have determined the sequence of a 6,000 bp region of the E. coli chromosome which includes the 3' end of the thyA coding region, the entire

recC gene and three unidentified open reading frames in the thyA-recC intergenic region.

The 1122 amino acids encoded by recC would give an unmodified molecular weight of 128,860, which is in good agreement with the values estimated for the molecular weight of the RecC protein from SDS-PAGE (11-13). The translational start of recC has been confirmed by NH_2-terminal sequence analysis of the RecC protein.

We have located the precise transcriptional start point of recC from mung bean nuclease experiments. It is clear from an inspection of the sequence 5' to the transcription initiation point that the sequence of the proposed recC promoter differs significantly from the consensus sequence for E. coli promoters (36). Mulligan et al. (43) have described a method for quantitating the strength of E. coli promoters based on an analysis of sequence alone. Putative promoter sequences are compared with the consensus derived from 112 known promoters (36) to obtain a homology score. The value assigned to each base is determined by the frequency at which it appears in that position in the 112 promoters. The overall weight also includes a score for spacing between the -10 and -35 regions. Using this method, the homology score for the putative recC promoter is calculated as 43.8%, slighty below the arbitrary score of 45% set by Mulligan et al. (43) as a lower limit for an effective promoter in vitro. This may account for the low level of transcription of recC observed in vivo.

There is an 8 bp inverted repeat situated 5' to the recC transcriptional start point and a 6 bp inverted repeat situated upstream of the 804 bp coding sequence for the putative 30 kDa protein which include an identical 12 bp core sequence. This core sequence may act as a common operator site which would suggest that expression of these two coding sequences is coordinately controlled. However, we have been unable to detect a transcript for the 804 bp open reading frame which indicates that if this sequence is expressed, then it is transcribed at a very low level. This might be achieved by a negative control mechanism, such as the binding of a repressor to this potential operator site. Similarly, the binding of a repressor to the possible recC operator could also account for the low level of transcription of recC seen under normal growth conditions. Alternatively, this sequence might be recognised by a positive control element (an activator) which consequently leads to enhanced expression of recC under the appropriate conditions. Activator binding sites are usually located immediately upstream or within the -35 promoter sequence (44). Furthermore, the promoters of

positively controlled genes usually differ significantly from that of the consensus (44), indicating that they are not properly recognised by RNA polymerase alone. Most of the promoters listed by Mulligan et al. (43) which have poor homology scores (<45%) require activators for maximal expression. The location of the putative gene for the 30 kDa protein, close to and upstream of recC, raises the possibility that it may be involved itself in the regulation of expression of recC.

All UV-inducible cellular genes previously shown in vitro to be under lexA control contain a LexA protein binding site matching the consensus sequence (A/T)-CTGT(A/T)Y-(A/T)-Y(A/T)Y-CAG, where Y is a pyrimidine (17). Such a sequence is not present in the recC regulatory region which is consistent with the observation that expression of recC is not induced in the SOS response (15).

There is a region in the deduced amino acid sequence of the putative 30 kDa protein, R-V-G----GKT (residues 59 to 70) which matches the consensus found in adenine nucleotide binding proteins (45, 46). No such region could be identified in the deduced amino sequence of the RecC protein. This is consistent with biochemical data which indicates that the ATPase activity of the RecBC enzyme resides in the RecB subunit since purified RecB protein but not RecC protein has this activity (15,16). A consensus adenine nucleotide binding sequence is found in the deduced amino acid sequence of both the RecB protein and of a protein encoded immediately downstream of recB which appears to be required for maximal activity of the RecBC enzyme (manuscript in preparation).

ACKNOWLEDGEMENTS

We thank Professor J. Fothergill for kindly performing the NH_2-terminal amino acid analysis on the RecC protein.

This work was supported by the Medical Research Council.

*Present address: Department of Biochemistry, University of California, Berkeley, CA 94720, USA

REFERENCES

1. Emmerson, P.T. (1968) Genetics 60, 19-30.
2. Buttin, G. and Wright, M. (1968) Cold Spring Harbour Symp. Quant Biol. 33, 259-264.
3. Oishi, M. (1969) Proc. Natl. Acad. Sci. USA 64, 1292-1299.
4. Goldmark, P.J. and Linn, S. (1970) Proc. Natl. Acad. Sci. USA 67, 434-441.

5. Muskavitch, K.M.T. and Linn, S. (1981) in The Enzymes (Boyer, P.D., ed) Vol. 14, pp233-250, Academic Press, New York.
6. Gillen, J.R. and Clark, A.J. (1974) in Mechanisms In Recombination (Grell, R.F., ed) p123, Plenum Press, New York.
7. Lam, S.T., Stahl, M.M., McMilin, K.D. and Stahl, F.W. (1974) Genetics 77, 425-433.
8. Smith, G.R., Kunes, S.M., Schultz, D.W., Taylor, A. and Triman, K.L. (1981) Cell 24, 429-436.
9. Ponticelli, A.S., Schultz, D.W. Taylor, A.F. and Smith, G.R. (1985) Cell 41, 145-151.
10. Willets, N.S. and Mount, D.W. (1969) J. Bacteriol. 100, 923-934.
11. Hickson, I.D. and Emmerson, P.T. (1981) Nature 294, 578-580.
12. Umeno, M., Sasaki, M., Anai, M. and Takagi, Y. (1983) Biochem. Biophys. Res. Commun. 116, 1144-1150.
13. Dykstra, C.C., Prasher, D. and Kushner, S.R. (1984) J. Bacteriol. 157, 21-27.
14. Taylor, A. and Smith, G.R. (1980) Cell 22, 447-457.
15. Hickson, I.D., Atkinson, K.E., Hutton, L., Tomkinson, A.E. and Emmerson, P.T. (1984) Nucl. Acids Res. 12, 3807-3819.
16. Hickson, I.D., Robson, C.R., Atkinson, K.E., Hutton, L. and Emmerson, P.T. (1985) J. Biol. Chem. 260, 1224-1229.
17. Little, J.W. and Mount, D.W. (1982) Cell 29, 11-22.
18. Walker, G.C. (1984) Microbiol Rev. 48, 60-93.
19. Lloyd, R.G. and Thomas, A. (1983) Mol. Gen. Genet. 190, 156-161.
20. Lovett, S.T. and Clark, A.J. (1983) J. Bacteriol. 153, 1471-1478.
21. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) 33, 103-119.
22. Remaut, E., Stanssens, P. and Fiers, W. (1981) Gene 15, 81-83.
23. Yasuda, S. and Takagi, T. (1983) J. Bacteriol. 154, 1153-1161.
24. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
25. Biggin, M.D., Gibson, T.J. and Hong, C.F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963-3965.
26. Bankier, A.T. and Barrell, B.G. (1983) Nucleic Acid Biochemistry, (Flavell, R.A., ed) Vol. B5, pp1-34, Elsevier Scientific Publishers (Ireland) Ltd.
27. McDonnel, M.W., Simon, M.N. and Studier, F.W. (1977) J. Mol. Biol. 110, 119-146.
28. Queen, C. and Korn, L.J. (1984) Nucl. Acids Res. 12, 581-599.
29. Staden, R. (1984) Nucl. Acids Res. 12, 551-567.
30. van den Elzen, P.J.M., Konings, R.N.H., Veltkamp, E. and Nijkamp, H.J.H. (1980) J. Bacteriol. 144, 579-591.
31. Berk, A.J. and Sharp, P.A. (1977) Cell 12, 721-732.
32. Sanger, F. and Coulson, A.R. (1978) FEBS Lett. 87, 107-110.
33. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-559.
34. Belfort, M., Maley, G., Pedersen-Lane, J. and Maley, F. (1983) Proc. Natl. Acad. Sci USA 80, 4914-4918.
35. Steitz, J.A. (1979) In Ribosomes (Chambliss, G., Craven, G.R., Davies, J., Kahan, L. and Nomura, M., eds) pp349-399, University Park Press, Baltimore.
36. Hawley, D.K. and McClure, W.R. (1983) Nucl. Acids Res. 11, 2237-2255.
37. Green, M.R. and Roeder, R.G. (1980) Cell 22, 231-242.
38. Sasaki, M., Fujiyoshi, T., Shimada, K. and Takagi, Y. (1982) Biochem. Biophys. Res. Commun. 109, 414-422.
39. Hentschel, C., Irminger, J-C., Bucher, P. and Binsteil, M.L. (1980) Nature 285, 147-151.
40. Ikemura, T. (1980) J. Mol. Biol. 151, 389-409.

41. Konigsberg, W. and Godson, G.N. (1983) Proc. Natl. Acad. Sci. USA 80, 687-691.
42. Umeno, M., Anai, M., Sasaki, M. and Takagi, Y. (1985) J. Biochem. 98, 681-685.
43. Mulligan, E.M., Hawley, D.K., Entriken, R. and McClure, W.R. (1984) Nucl. Acids Res. 12, 789-800.
44. Raibaud, O. and Schwartz, M. (1984) Ann. Rev. Genet. 18, 173-206.
45. Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982) EMBO J. 1, 645-951.
46. Finch, P.W. and Emmerson, P.T. (1984) Nucl. Acids Res. 12, 5789-5799.