Volume 12 Number 11 1984 Nucleic Acids Research

Sequences of the *E. coli uvrC* gene and protein

Gwendolyn B.Sancar+*, Aziz Sancar+ and W.Dean Rupp

Departments of Therapeutic Radiology and Molecular Biophysics and Biochemistry, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510, USA

Received 17 April 1984; Accepted 8 May 1984

ABSTRACT

We have determined the sequence of a 2400 bp region of *E. coli* chromosomal DNA containing the *uvrC* gene. The coding region of *uvrc* is 2267 bp in length, encodes a polypeptide with a calculated molecular weight of 66,038 daltons, and is preceded by a typical *E. coli* ribosome binding site. By constructing deletion derivatives we have established that a *uvrC* promoter lies within the 113 bp region preceding the translational start of *uvrC*. The codon usage in *uvrC* is strongly biased in favor of codons used infrequently in *E. coli*, which may contribute to the relatively low intracellular concentration of *uvrC* protein.

INTRODUCTION

In *Escherichia coli* removal of pyrimidine dimers from DNA via the excision repair pathway requires intact *uvrA*, *uvrB* and *uvrC* genes (1). During the last few years, we (2-4) and others (5,6) have identified the products of these genes as polypeptides of M_r 114,000 (UVRA), 84,000 (UVRB) and 70,000 (UVRC). *In vivo* evidence indicates that UVRA, UVRB and UVRC are involved in the first step of excision repair, the production of single strand endonucleolytic cuts near the dimer and in the damage containing strand (7,8). Using purified UVR proteins, Sancar and Rupp (9) have confirmed this observation *in vitro* and have demonstrated that the precise locations of the single strand scissions are at the 8th phosphodiester bond 5' and the 4th or 5th phosphodiester bond 3' to the dimer; because the simultaneous presence of all three proteins is required for the reaction to proceed *in vitro* these authors have proposed the name UVRABC excision nuclease to emphasize the concerted action of UVRA, UVRB and UVRC.

Although the *uvr* genes are not linked (1), the functional relationship of the *uvrA*, *uvrB and uvrC* gene products suggests that these genes may share a common regulatory pathway. Evidence for such a pathway has been obtained from *in vivo* experiments the results of which

indicate that all three genes are under the control of the *recA-lexA* regulon and are induced as a part of the coordinated expression of genes following DNA damage (10-13) called the SOS response (14). We have previously shown *in vitro* that *lexA* protein (LEXA) binds to specific regions overlapping the *uvrA and uvrB* promoters and thereby inhibits transcription of the genes (15,16); activation of the proteolytic activity of *recA* protein (RECA), which occurs in response to DNA damage or stalling of the replication fork, leads to cleavage of LEXA by RECA (17) and increased transcription of *uvrA* and *uvrB* (18,19). Whether the mechanism of *recA-lexA* control of *uvrC* expression is similar to that of *uvrA* and *uvrB* is unknown.

The molecular mechanism of action of UVRABC excision nuclease ultimately must be solved by analysis of the specific amino acid to nucleotide contacts made when the enzyme is bound to DNA as well as determination of the contacts sites between the three *uvr* proteins. Meaningful interpretation of such data will require knowledge of the primary sequence of the proteins. As a first step in this process we report here the nucleotide sequence of the *uvrC* gene and the amino acid sequence of *uvrC* protein. In addition to the structural gene, we also report the nucleotide sequence of the 502 bp region preceding the translational start of *uvrC;* our results indicate that regulation of *uvrC* expression may differ from that of the *uvrA* and *uvrB* genes.

MATERIALS AND METHODS

Enzymes and Reagents

Restriction endonucleases, Bal 31 nuclease, bacterial alkaline phosphatase, T4 polynucleotide kinase, T4 DNA ligase, *E. coli* DNA polymerase I-Klenow fragment and terminal deoxynucleotide transferase were obtained from Bethesda Research Laboratories or New England Biolabs and were used according to manufacturers directions. γ-^{32}P-dATP (>7000 Ci/mmole) and 3'-α-^{32}P-cordycepin 5' triphosphate (5000Ci/mmole) were from New England Nuclear as were reagents used in DNA sequencing reactions. Unlabeled nucleotides were from P-L Biochemicals.

Bacterial Strains and Plasmids

E. coli K-12 strain DR1984 (*recA1 uvrC34,* ref. 4) and DH1 (20) were used for propagation of plasmids. The construction of plasmids pDR3003, pDR3024 and pLC13-12 have been previously described (4,21). Plasmid pDR2803 was constructed by cutting pDR3003 with PvuII and PstI followed

by ligation to pBR328 (22,23) which had been treated with the same restriction enzymes; Tet^R Cam^S Amp^S clones were screened by restriction analysis using PvuII and PstI. BglII digestion of pDR2803 followed by ligation to BamHI-cut pBR322 and selection for Amp^R Tet^S colonies yielded plasmids pDR3101 and pDR3102; the orientation of the inserted BglII fragment with respect to the *tet* gene was determined by restriction mapping of HinfI sites in these plasmids. Deletion derivatives of pDR2803 were constructed by incubating 1 μg of DNA with 2 units of Bal31 nuclease for 10 min. at 20°C; after phenol extraction the mix was incubated overnight with T4 DNA ligase, cut with PvuII and used to transform pDR1984 to tetracycline resistance. DNA was prepared from individual clones (24) and the extent of the deletion in each plasmid was determined by restriction mapping with the enzymes BglII, EcoRI, HindIII, PvuII, HinfI, BstNI and BalI.

UV Survivals

All tests for UV survival were performed using DR1984 as host. Cultures were grown overnight in luria broth; antibiotics were added to the growth media according to the plasmid present. Stationary phase cultures at an A_{600} of approximately 4.0 were diluted, plated on luria broth agar and irradiated at 254 nm using a General Electric G8 bulb at a dose rate of $0.02 J/m^2/sec$. Following irradiation, the plates were incubated in the dark at 37°C for 24 h prior to counting survivors.

DNA Sequencing

Plasmid DNA was prepared by CsCl-ethidium bromide equilibrium centrifugation as previously described (16). Purification of restriction fragments was by polyacrylamide gel electrophoresis and electroelution (25). Labeling of 5' ends and of recessed 3' ends using T4 polynucleotide kinase or *E. coli* DNA polymerase I-Klenow fragment, as well as DNA sequence analysis using the A+G, G, T+C and C specific reactions, were performed as described by Maxam and Gilbert (26). Protruding 3' ends were labeled using terminal deoxynucleotide transferase and 3'-α^{32}P-cordycepin triphosphate following the procedure of Roychoudhury *et al* (27). The products of the sequencing reactions were separated by electrophoresis on 20% or 8% polyacrylamide gels as described by Sanger and Coulson (28).

Amino Acid Analysis of *uvrC* Protein

UvrC protein was purified as described by Sancar and Rupp (9). A sample was dialyzed against distilled water and lyophilized; aliquots

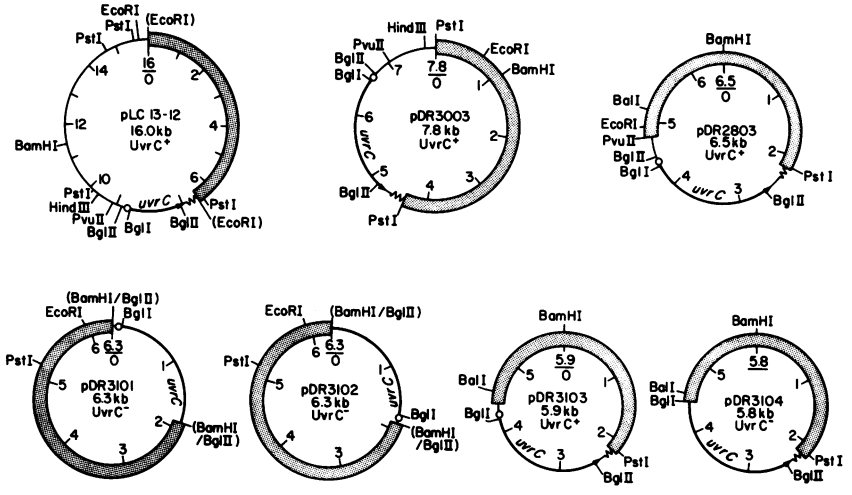

Figure 1. Map of restriction endonuclease recognition sites in various *uvrC* plasmids. Vector sequences are indicated by the filled double line and are ColE1 for pLC13-12, pBR322 for pDR3003, pDR3101, and pDR3102, and pBR328 for pDR2803, pDR3103 and pDR3104. The restriction map of pLC13-12 is from reference 43. The location of *uvrC* and the direction of transcription are indicated by the heavy line and arrow; the *uvrC* promoter defined in this paper is shown as an open circle. BglII sites in the vectors are not shown; in addition, we have not attempted to map any BglII sites upstream from the *uvrC* promoter-proximal PstI site in pLC13-12. (EcoRI), EcoRI site destroyed by poly dA:dT tailing (indicated by the wavy line) during the construction of pLC13-12 (21); (BamHI/BglII), fusion of BamHI and BglII sites destroying both sites.

containing approximately 20 µg of protein were hydrolyzed at 115°C in 6N HCl/0.2% phenol for 16 h, 48 h or 72 h, then analyzed on a Beckman 121M amino acid analyzer. Cysteine was quantitated as cysteic acid and methionine as methionine sulfone following hydrolysis of protein which had been subjected to performic acid oxidation at 0°C for 4h.

RESULTS

Boundaries of *uvrC* and Sequencing Strategy

Sancar *et al.* (4) have previously reported the subcloning of the *uvrC* gene from the Clarke-Carbon plasmid pLC13-12 into pBR322. The resulting plasmid pDR3003 carries a 3.4 kbp DNA fragment, composed of 3.2 kbp of *E. coli* chromosomal DNA followed by approximately 0.2 kbp of poly dA:dT linker plus ColE1 DNA, inserted into the PstI site of the vector (Figure 1). Both pLC13-12 and pDR3003 fully complement the *uvrC* mutation in DR1984 (*recA1 uvrC34*) (Figure 2). Analysis of Tn1000 insertion

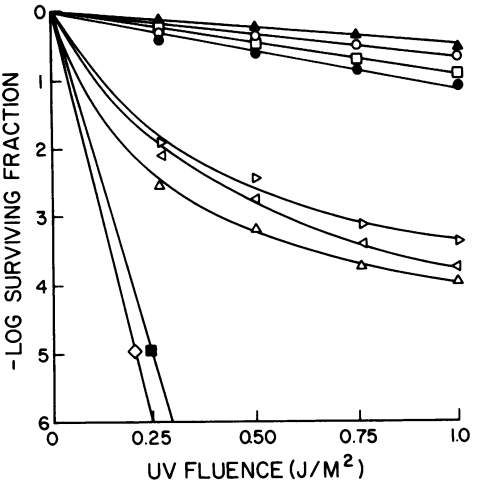

Figure 2. UV survival of DR1984 (*recA1 uvrC34*) and its plasmid-containing derivatives. Stationary phase cultures were diluted, plated, and irradiated as described in Materials and Methods. Symbols: DR1984,◇; DR1984 containing plasmid pBR322,■; pLC13-12,□; pDR3003,●; pDR2803,○; pDR3101,▷; pDR3102,◁; pDR3103,▲; pDR3104,△.

derivatives of pDR3003 in which *uvrC* complementing activity had been lost established that (i) *uvrC* encodes a polypeptide of M_r 70,000 (ii) the direction of transcription is counter-clockwise relative to pBR322, and (iii) the start of the gene is at approximately 7 kbp on the pDR3003 map (4).

To further delineate the 5' boundary of *uvrC* we constructed plasmid pDR2803 by joining the 2.6 kbp PvuII-PstI fragment of pDR3003 to the unique PvuII and PstI sites of the vector pBR328 (22,23) (Figure 2). As is seen in Figure 2, pDR2803 confers the same level of UV resistance as do the parental $UvrC^+$ plasmids; thus in pDR2803 the *uvrC* gene is still intact. In contrast plasmids pDR3101 and pDR3102, which carry only the 1.9 kbp BglII fragment of *uvrC* (Figure 1), confer a greatly diminished level of UV resistance (Figure 2) indicating that one or both of the BglII sites in the chromosomal insert lie within the *uvrC* gene. With these facts in mind, we determined the nucleotide sequence of a portion of the chromosomal insert in pDR2803 beginning at the PvuII site and extending through 150 nucleotides beyond the 3' BglII site. The sequencing strategy employed is shown in Figure 3. Approximately 85% of the sequence was determined on both strands; all internal restriction sites

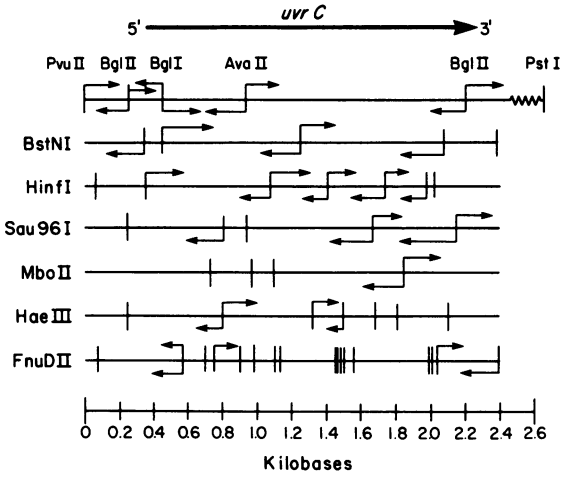

Figure 3. Strategy used in sequencing the *uvrC* gene. The 2.7 kbp PvuII-PstI fragment from pDR2803 containing the *uvrC* gene is shown. The heavy line and arrow indicate the location and direction of transcription of *uvrC*. Other arrows show the direction and extent of sequencing from the indicated restriction endonuclease cutting sites; restriction sites between the 3' FnuDII site and the PstI site have not been mapped. The wavy line represents the poly dA:dT tail added during construction of pLC13-12 (21); note that the PstI site is the junction between the original ColE1 vector of pLC13-12and pBR328 and does not derive from chromosomal DNA 3' to *uvrC* (21, 43, and our unpublished observations).

used for sequencing were crossed by reading from a second restriction site.

Nucleotide Sequence of *uvrC*

The sequence of the 2400 bp region of chromosomal DNA carrying the *uvrC* gene is shown in Figure 4. The *uvrC* coding region begins with an ATG at nucleotides 503-505; beginning 13 bp upstream from this ATG is the sequence AGG which is a good match to a portion of the Shine-Dalgarno consensus sequence for a ribosome binding site (29,30). Translation of *uvrC* terminates at a TGA triplet at nucleotides 2267-2269. To confirm the overall accuracy of the DNA sequence within the coding region we compared the amino acid composition predicted from the DNA sequence with that obtained from amino acid analysis of purified *uvrC* protein; as can be seen in Table 1, the agreement is excellent. We have been unsuccessful in attempts to verify the amino-terminal amino acid sequence using standard stepwise degradation techniques, which leads us to believe that the amino terminus of *uvrC* protein is blocked. The predicted molecular mass of *uvrC* protein is 66,038 daltons, compared to an

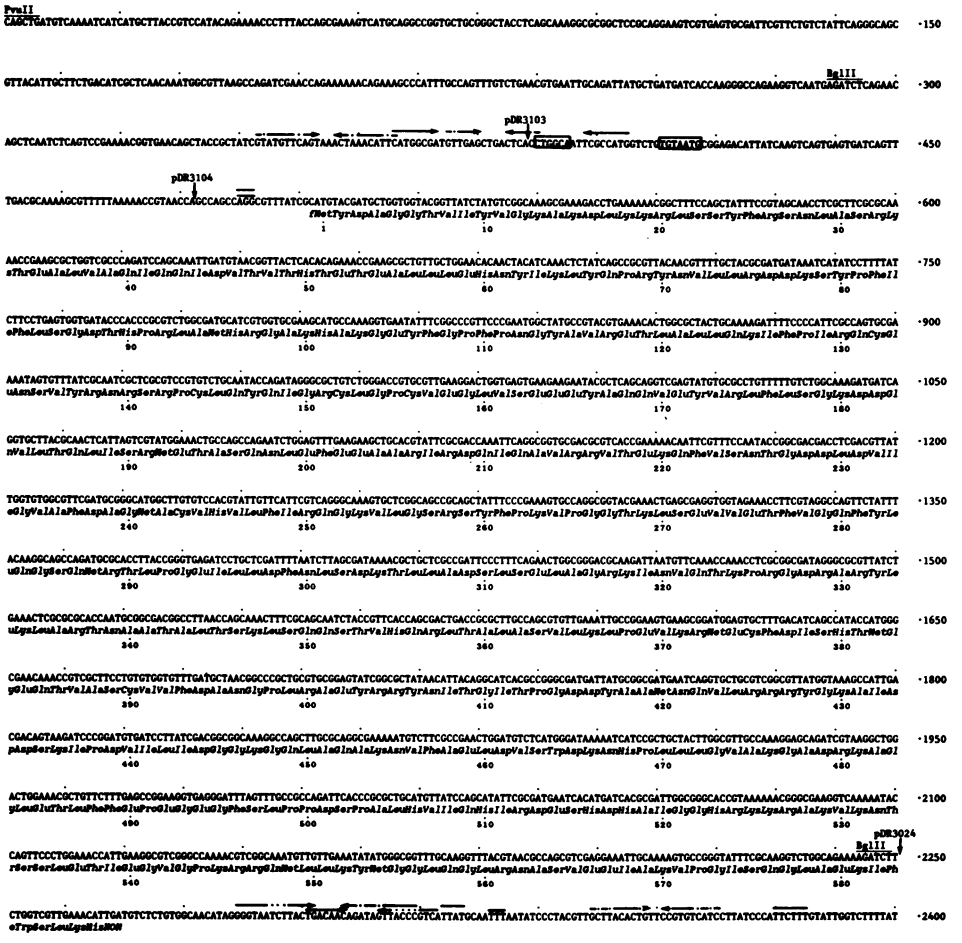

Figure 4. Nucleotide sequence of the *uvrC* gene and the amino acid sequence of *uvrC* protein. The BglII and PvuII sites within the sequence are shown for reference with nucleotide numbering beginning at the first nucleotide in the PvuII recognition site. The proposed -35 and -10 sequences of the *uvrC* promoter described in this paper are enclosed in boxes. Horizontal arrows above the sequence indicate regions of dyad symmetry preceeding and following the *uvrC* gene with mismatches indicated by dots. The Shine-Dalgarno ribosome binding site is doubly overscored and T-rich regions which may be sites of transcription termination are overscored with one line. Vertical arrows indicate deletion break points in pDR3103 and pDR3104 and the Tn1000 insertion site in pDR3024.

observed M_r of 70,000 obtained by SDS-polyacrylamide gel electrophoresis (4).

It was noted above that plasmids carrying only the 1.9 kbp BglII fragment from *uvrC* are incapable of restoring full UV resistance to

Table I. Amino Acid Composition of *uvrC* Protein

Amino Acid	Predicted From DNA Sequence	Found By Amino Acid Analysis†
Cys	7	8.3
Asx	49	49.4
Thr	29	28.6
Ser	36	37.6
Glx	68	72.4
Pro	23	23.7
Gly	46	47.4
Ala	46	46.6
Val	42	41.6
Met	10	8.9
Ile	29	26.4
Leu	63	61.3
Phe	22	22.0
His	15	14.5
Lys	38	37.4
Arg	42	40.9
Trp	2	N.D.

† Based on 586 non-tryptophan residues.

DR1984. Inspection of the DNA sequence of *uvrC* reveals that BglII cuts within the *uvrC* coding sequence at the 3' end of the gene. This prediction was confirmed by sequencing across the *uvrC*: Tn1000 junction in plasmid pDR3024 which carries a Tn1000 insertion previously mapped to the extreme 3' end of *uvrC* (4). As is indicated in Figure 4, the Tn1000 insertion in pDR3024 begins one nucleotide 3' to the BglII recognition site.

In addition to the open reading frame comprising *uvrC*, a second open reading frame beginning an undetermined distance 5' to the PvuII site and ending at TAA at nucleotides 413-415 is present within the 2400 bp sequence. It is not known at present whether this region actually encodes a protein although it is tempting to speculate that this may be the 3' end of the coding region for a 27 kd polypeptide previously mapped 5' to *uvrC* (4,31).

Transcription Signals

Within the 502 bp region preceding the translational start site of *uvrC* there are several regions which display at least partial homology to the -35 (TTGACA) and -10 (TATAATA) consensus sequences known to be important for promoter function in *E. coli* (32-35). To determine which of these regions constitutes the *uvrC* promoter, we constructed deletion derivatives of pDR2803 by cutting the plasmid with PvuII, digesting with the double-strand exonuclease Bal 31, then resealing the plasmids with DNA ligase; these plasmids were then used to transform DR1984 to

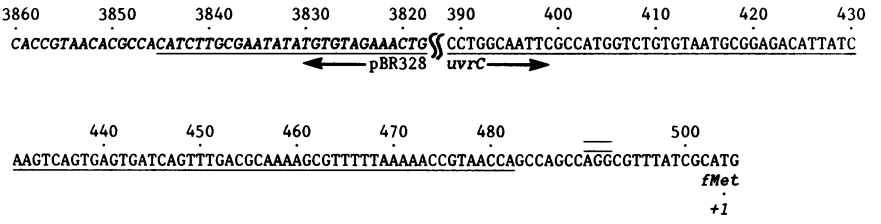

Figure 5. Nucleotide sequence across the deletion break points in pDR3103 and pDR3104. The sequence across the deletion break point in pDR3103 is shown. pBR328 sequences are in italics; numbering of the pBR328 sequence corresponds to the scheme in pBR328 in the absence of the chromosomal insert (22, 23). Numbering of *uvrC* sequences is as in Figure 4. The Shine-Dalgarno ribosome binding site is overlined twice. The underlined sequence is absent in pDR3104. ⟆⟆ , deletion break point in pDR3103.

tetracycline resistance. Twelve such transformants were selected at random and tested for UV sensitivity; DNA was also prepared from these plasmids and subjected to fine-structure restriction mapping to determine the extent of the deletions. Among nine plasmids still capable of complementing the *uvrC* defect in DR1984, three carried deletions encompassing the BglII site near the 5' end of *uvrC*; all plasmids which failed to complement had also lost this BglII site (data not shown). Thus sequences upstream of the 5' BglII site are not required for *uvrC* expression.

Two of the Bal31-generated deletion derivatives, pDR3103 and pDR3104 (Figure 1), proved to be particularly interesting as restriction mapping revealed that the extent of the deletions in these plasmids differed by no more than 150 bp, yet DR1984 carrying pDR3103 was $UvrC^+$ while DR1984 carrying pDR3104 was $UvrC^-$ (Figure 2). The exact break points of the deletions in these plasmids were determined by DNA sequence analysis and are shown in Figures 4 and 5. In both pDR3103 and pDR3104 the *uvrC* ribosome binding site remains intact, leading to the conclusion that the difference in the ability of these plasmids to complement a *uvrC* mutation must lie at the level of transcription. Both the 5' and 3' deletion break points differ in pDR3103 and pDR3104. Inspection of the pBR328 sequence (20) present in pDR3103 but absent in pDR3104 does not reveal any promoter-like regions (Figure 5); therefore a *uvrC* promoter must lie between the 3' break points in pDR3103 and pDR3104, that is between base pairs 389 and 482. The region most likely to comprise the *uvrC* promoter is defined by the sequences CTGGCA (390-395 bp) and TGTAATG (411-417 bp) which display reasonable homology

to the consensus -35 and -10 sequences respectively. In addition, the 15 bp distance separating these two elements is consistent with the 17±2bp spacing seen in most authentic promoters (32-35), and the sequence CAT, which is often observed beginning 5-8 bp downstream from the -10 sequence (32, 35) is seen at base pairs 424-426 in the *uvrC* promoter.

It is interesting to note that the level of UV resistance in the DR1984/pDR3104 strain is greater than that of DR1984 alone or DR1984 containing pBR322. In fact, we have observed that *uvrC* plasmid containing strains show a complete UvrC$^-$ phenotype only when sequences downstream from and including the ribosome binding site are removed (data not shown). The residual UV resistance imparted by pDR3104 probably is the result of a low level of transcription originating in pBR328 coupled with the fact that only about 10 molecules of *uvrC* protein per cell are required for full complementation in a RecA$^-$ strain (5).

Van Sluis *et al*, (12) have recently demonstrated that expression of *uvrC* is regulated by the *recA* and *lexA* proteins. As we have previously demonstrated the presence of functional *lexA* protein binding sites overlapping the promoters of the *uvrA* and *uvrB* genes (15,16), we searched the *uvrC* sequence for such a site; within the region sequenced we did not find any areas displaying close homology to the consensus sequence for the *lexA* protein binding site,

$$\frac{A}{T} - CTGT \frac{A}{T} Y - \frac{A}{T} - Y \frac{A}{T} Y - CAG,$$

where Y = a pyrimidine base (13,15,36-38). However, the region from base pairs 343-405 contains three hyphenated molecular palindromes consisting of 24 bp, 40 bp and 17 bp (Figure 4); two of these palindromes overlap or surround the -35 region of the proposed *uvrC* promoter. Whether these regions are involved in regulation of *uvrC* expression, perhaps by a mechanism indirectly involving *lexA*, remains to be elucidated; it is clear that deletion of these palindromes, as in pDR3103, does not reduce the ability of the cloned gene, when present on a multicopy plasmid, to complement a chromosomal *uvrC* mutation.

As with promoters, transcription termination sites in *E. coli* tend to share common features, namely: (i) a region of hyphenated dyad symmetry including or followed by a G/C rich region of at least 3 contiguous nucleotides and (ii) a run of several consecutive T's at a

Table 2. Codon Usage in *uvrC* Compared to that Seen in Non-Regulatory Proteins in *E. coli*

Amino Acid	*E. coli*† Codon	% Synonym Use	*uvrC* Total Codons	% Synonym Use	Amino Acid	*E. coli*† Codon	% Synonym Use	*uvrC* Total Codons	% Synonym Use
Phe	UUU	44	9	41	Tyr	UAU	41	16	76
	UUC	56	13	59		UAC	59	5	24
Leu	UUA^{I}	6	4	6	His	CAU	39	8	53
	UUG^{I}	8	11	18		CAC	61	7	47
	CUU^{I}	9	9	14					
	CUC^{I}	7	8	13					
	CUA^{I}	2	3	5	Gln	CAA^{R}	27	13	41
	CUG	69	28	44		CAG	73	19	59
Ile	AUU	37	16	55	Asn	AAU^{R}	24	12	67
	AUC	62	12	42		AAC	76	6	33
	AUA^{R}	1	1	3					
Met	AUG	--	10	--	Lys	AAA	77	30	79
						AAG^{I}	23	8	21
Val	GUU	38	11	26					
	GUC	13	11	26	Asp	GAU	51	22	71
	GUA	23	5	12		GAC	49	9	29
	GUG	27	15	36					
Ser	UCU	27	2	6	Glu	GAA	73	27	75
	UCC	26	5	14		GAG^{I}	27	9	25
	UCA^{I}	8	5	14					
	UCG^{R}	11	5	14	Cys	UGU	42	4	57
	AGU^{I}	6	7	19		UGC	58	3	43
	AGC	22	12	33					
Pro	CCU^{R}	9	2	9	Trp	UGG	--	2	--
	CCC^{R}	6	2	9					
	CCA	20	3	13	Arg	CGU	58	17	40
	CCG	65	16	70		CGC	35	16	38
						CGA^{I}	2	2	5
Thr	ACU	24	2	7		CGG^{I}	3	6	14
	ACC	51	15	52		AGA^{I}	1	0	--
	ACA^{I}	6	4	14		AGG^{R}	0.25	1	2
	ACG^{R}	20	8	28					
Ala	GCU	28	8	17	Gly	GGU	48	14	30
	GCC	19	11	24		GGC	41	23	50
	GCA	23	4	9		GGA^{I}	5	6	13
	GCG	30	23	50		GGG^{I}	7	3	7

†Data from reference 40.
^{R}*Codons rarely used in E. coli.*
^{I}Codons infrequently used in *E. coli.*

distance of 20±4 bp from the center of the symmetry element with termination occuring at one of these T's (32). Based upon these criteria three potential transcription termination sites are found in the 134 bp region following the translation termination site of *uvrC*; these sites are indicated in Figure 4.

Codon Usage in *uvrC*

It has been noted that usage of synonymous codons is nonrandom in *E. coli* genes and that the frequency with which a particular codon is used appears to reflect the relative abundance of the cognate tRNA (39). However, in certain genes codon usage deviates from this pattern; as all genes in which such a deviation has been noted encode products which are present in only a few copies per cell (40), it has been proposed that the increased use of rare codons, by decreasing the rate of translation, may be one mechanism involved in the modulation of gene expression. To determine whether codon usage might play a role in the relatively low copy number of *uvrC* protein, we have compared the codon usage in *uvrC* with that of 25 nonregulatory proteins from *E. coli*. As can be seen in Table 2, there is a definite bias in favor of the use of rare or infrequent codons in *uvrC*. Overall the frequency of use of rare codons (defined as the total number of rare codons divided by the total number of cognate codons) is 21% for *uvrC* as compared to 13% for the 25 nonregulatory proteins considered in Table 2; when infrequently used codons are included in this calculation, the frequency of use rises to 35% for *uvrC* but only 20% for nonregulatory proteins.

DISCUSSION

We have determined the sequence of the *uvrC* gene of *E. coli* and have confirmed the overall accuracy of the sequence by comparison of the amino acid composition predicted from the nucleotide sequence with that obtained by amino acid analysis of purified *uvrC* protein. By constructing and sequencing *uvrC* deletion derivatives we have established that a *uvrC* promoter lies within a 113 bp region 5' to the *uvrC* translational start site; based upon homology with the consensus sequence of known *E. coli* promoters, we have proposed that the *uvrC* promoter is located at bp 390-417. This location is consistent, within the limits of the techniques used, with previous reports by Sancar *et al.* (4) and Yoakum and Grossman (5) in which the approximate site of the *uvrC* promoter was determined by restriction mapping. The DNA sequence

predicts that the 3' end of the *uvrC* structural gene lies beyond the promoter distal BglII site in pDR2803. This prediction was confirmed by sequencing across the insertion site of Tn1000 in plasmid pDR3024 which, as a result of the insertion, encodes a *uvrC*:Tn1000 hybrid polypeptide of slightly greater molecular weight than authentic *uvrC* protein (4).

Our results concerning the location of the *uvrC* promoter are at variance with previous reports by Sharma and Moses (6,31) who have concluded on the basis of plasmid-conferred UV resistance and RNA polymerase binding studies that the *uvrC* promoter is at least 0.9 kbp upstream from the structural gene. However, as the plasmids used by these investigators were missing all sequences downstream from the 3' BglII site in *uvrC*, it is difficult to assess the reported data on UV survival; our data indicate that insertion of new sequences beyond the BglII site greatly impairs *uvrC* function (cf. pDR2803, pDR3101, pDR3102, and pDR3024 in Figure 2). In addition, the conditions used by Sharma *et al.* (31) for RNA polymerase binding discriminated against detection of weak promoters as the *uvrC* promoter is expected to be (see below).

Van Sluis *et al.* (12) have recently reported that a *uvrC* promoter exists upstream from the 5' BglII site at nucleotides 231-260 in the sequence shown in Figure 4. Our data do not rule out the presence of this promoter but do indicate that only sequences downstream from bp388 are required for complete complementation when present on a multicopy plasmid. Similarly, the results of S-1 mapping of *uvrC* transcripts by these workers do not rule out a *uvrC* promoter at the site proposed in this report as no data was published using a uniquely labeled probe for RNA transcripts initiating at this promoter.

Both the promoter identified by van Sluis *et al.* (12) and the promoter proposed in this report display limited homology to the consensus -35 and -10 sequences (32-35); the number and location of the mismatches suggests that these promoters will be relatively inefficient in initiating transcription of *uvrC*. This is consistent with the report by Yoakum and Grossman (5) that in a $RecA^-$ strain the number of *uvrC* protein molecules is on the order of 10 per cell. In addition the increased proportion of infrequent and rare codons utilized in the *uvrC* structural gene may contribute to this low level of expression. In this regard *uvrC* appears to be similar to the *E. coli phr* gene which encodes DNA photolyase, the enzyme responsible for photoreactivation of pyrimidine dimers; photolyase is present in 10-20 copies per cell and

rare codon usage in this gene is 25% (41) as compared to 21% for *uvrC* and 13% for "average" *E. coli* nonregulatory proteins (40).

The results of both *in vivo* and *in vitro* experiments indicate that in *E. coli* the simultaneous presence of *uvrA*, *uvrB* and *uvrC* proteins is required for incision near a pyrimidine dimer in DNA (7-9). Given this functional relationship it might be expected that the *uvrA*, *uvrB* and *uvrC* genes share a common regulatory mechanism; indeed it has been shown that *in vivo* all three genes are under the control of the *recA* - *lexA* regulon (10-12). We have demonstrated *in vitro* that in the case of the *uvrA* and *uvrB* genes this control is manifested at the level of transcription with *lexA* protein functioning as a repressor by binding to sequences overlapping the promoters of these genes (15,16). The fact that all UV-inducible cellular genes previously shown *in vitro* to be under *lex A* control contain a *lexA*-repressible operator matching the consensus sequence

$$\underset{T}{\underline{A}} - CTGT \underset{T}{\underline{A}} Y - \underset{T}{\underline{A}} - Y \underset{T}{\underline{A}} Y - CAG$$

where Y = pyrimidine (13, 15, 36-38) makes the absence of this sequence in *uvrC* unexpected. Van Sluis *et al.* (12) have proposed that the sequence GTCTGAACGTGAATTGCAG at bp 233-251 serves as a *lexA* protein binding site regulating transcription from a *uvrC* promoter located at bp 231-260. If this region is an authentic *lexA* binding site then it differs substantially from all previously identified sites in that the spacing between the CTG-CAG palindrome is 11 bp rather than the usual 10 bp spacing and, at the three positions underlined in the sequence above, the sequence of the proposed *lexA* binding site departs from the consensus sequence. Clearly *in vitro* transcription and footprinting experiments using purified *lexA* protein must be performed to verify the existence of such an atypical *lexA* binding site.

The presence of two promoters for *uvrC* raises the question of whether transcription from these promoters is differentially regulated both with respect to one another and as compared with the *uvrA* and *uvrB* genes. Although all three UVR proteins are required for excision repair of pyrimidine dimers and transcription of these genes is induced by UV irradiation (10-12), a recent report by Tang *et al.* (42) indicates that only *uvrC* protein is needed for repair of N-hydroxyamino-fluorene induced damage of DNA; also treatment of cells with nalidixic acid results in induction of the SOS response including increased transcription of *uvrA* and *uvrB* (10,11) but not of *uvrC* (12). Thus it

seems possible that the requirement for *uvrC* in repair of damaged DNA may depend upon the nature of the lesion and in some cases may differ from that for *uvrA* and *uvrB*. *LexA* protein is clearly involved in the regulation of all three genes but whether its participation is direct or indirect in the case of *uvrC* remains to be elucidated.

The amino acid sequence reported here for *uvrC* protein is the first obtained for any of the proteins constituting UVRABC excision nuclease. It has been demonstrated that *uvrC* protein binds to DNA (4); in addition, the fact that *uvrA*, *uvrB*, and *uvrC* proteins must be present simultaneously for incision of dimer-containing DNA suggests that *uvrC* protein interacts with one or both of these proteins. Knowledge of the amino acid sequence of *uvrC* protein should be useful in attempting to define the DNA and protein binding domains of UVRC. In this context it is interesting to note that the results reported here indicate that the 7 carboxy-terminal amino acids of UVRC are important for UVRC function as plasmids encoding proteins in which this region is deleted confer limited UV-resistance. Additional studies aimed at defining the functional domains of UVRC are now in progress.

ACKNOWLEDGEMENTS

We thank Dr. Gary Davis of the Membrane Center, Yale University, for determining the amino acid composition of *uvrC* protein. This work was supported by NIH Grants CA06519, CA09259, GM24178, GM32833 and American Cancer Society grant PDT-80.

+Present address: Department of Biochemistry, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA

*To whom correspondence should be addressed

REFERENCES

1. Howard-Flanders, P. Boyce, R.P. and Theriot, L. (1966) *Genetics* 53, 1119-1136.
2. Sancar, A., Wharton, R.P., Seltzer, S., Kacinski, B.M., Clarke, N.D. and Rupp, W.D. (1981) *J. Mol. Biol.* 148, 45-62.
3. Sancar, A., Clarke, N.D., Griswold, J., Kennedy, W.J. and Rupp, W.D. (1981) *J. Mol. Biol.* 148, 63-76.
4. Sancar A., Kacinski, B.M., Mott, L.D. and Rupp, W.D. (1981) *Proc. Natl.Acad. Sci. USA* 78, 5450-5454.
5. Yoakum, G.H. and Grossman, L. (1981) *Nature* 292, 171-173.
6. Sharma, S., Ohta, A., Dowhan, W. and Moses, R.E. (1981) *Proc. Natl. Acad. Sci. USA* 78, 6033-6037.
7. Seeberg, E., Rupp, W.D. and Strike, P. (1980) *J. Bacteriol* 144, 94-104.
8. Seeberg. E., Nissen-Meyer, J. and Strike, P. (1976) *Nature* 263, 524-526.
9. Sancar, A. and Rupp, W.D. (1983) *Cell* 33, 249-260.

10. Kenyon, C.J. and Walker, G.C. (1981) *Nature* 289, 808-810.
11. Fogliano, M. and Schendel, P. (1981) *Nature* 289, 196-198.
12. van Sluis, C.A., Moolenar, G.F. and Backendorf, C. (1983) *EMBO J.* 2 2313-2318.
13. Little, J.W. and Mount, D.W. (1982) *Cell* 29, 11-22.
14. Radman, M. (1975) in *Molecular Mechanisms for DNA Repair, Part A,* Hanawalt, P. and Setlow, R.B., Eds., pp. 355-367, Plenum Press, New York.
15. Sancar, A., Sancar, G.B., Rupp, W.D., Little, J.W. and Mount, D.W. (1982) *Nature* 298, 96-98.
16. Sancar, G.B., Sancar, A., Little, J.W. and Rupp, W.D. (1982) *Cell* 28, 523-530.
17. Little, J.W., Edmiston, S.H., Pacelli, L.Z. and Mount, D.W. (1980) *Proc. Natl. Acad. Sci. USA* 77, 3225-3229.
18. van den Berg, E.A., Geerse, R. H., Pannekoek, H. and van de Putte, P. (1983) *Nucleic Acids Res.* 11, 4355-4363.
19. Backendorf, C., Brandsma, J. A., Kartasova, T. and van de Putte, P. (1983) *Nucleic Acids Res.* 11, 5795-5810.
20. Hanahan, D. (1983) *J. Mol. Biol.* 166, 557-580.
21. Clarke, L. and Carbon, J. (1976) *Cell* 9, 91-99.
22. Soberon, X., Covarrubias, L., and Bolívar, F. (1980), *Gene* 9, 287-305.
23. Prentki, P., Karch, F., Iida, S., and Meyer, J. (1981) *Gene* 14, 289-299.
24. Birnboim, H. C. and Doly, J. (1979) *Nucleic Acids Res.* 7, 1513-1523.
25. McConell, M.W., Simon, M.N. and Studier, F.W. (1977) *J. Mol. Biol.* 110, 119-146.
26. Maxam, A.M. and Gilbert, W. (1980) in *Methods in Enzymology,* Moldave, K. and Grossman, L., Eds., 65, 499-560, Academic Press, New York.
27. Roychoudhury, R., Jay, E. and Wu, R.(1978) *Nucleic Acids Res.* 3, 863-877.
28. Sanger, F. and Coulson, A.R. (1978) *FEBS Letters* 87, 107-110.
29. Shine, J. and Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* 71, 1342-1346.
30. Steitz, J.A. (1979) in *Ribosomes*, Chambliss, G., Craven, G.R., Davies, J., Kahan, L. and Nomura, M. Eds., pp.349-399, University Park Press, Baltimore.
31. Sharma, S., Dowhan, W. and Moses, R.E. (1982) *Nucleic Acids Res.* 10, 5209-5221.
32. Rosenberg, M. and Court, D. (1979) *Ann. Rev. Genetics* 13, 319-353.
33. Siebenlist, U., Simpson, R.B. and Gilbert, W. (1980) *Cell* 20, 269-281.
34. Aoyama, T., Takanami, M., Ohtsuka, E., Taniyama, Y., Marumoto, R., Sato, H. and Ikehara, M. (1983) *Nucleic Acids Res.* 11, 5855-5864.
35. Hawley, D.K. and McClure, W.R. (1983) *Nucleic Acids Res.* 11, 2237-2255.
36. Easton, A.M. and Kushner, S.R. (1983) *Nucleic Acids Res.* 11, 8625-8640.
37. Beck, E. and Bremer, E. (1980) *Nucleic Acids Res.* 8, 3011-3027.
38. Mizusawa, S., Court, D. and Gottesman, S. (1983) *J. Mol. Biol.* 171, 337-343.
39. Ikemura, T. (1981) *J. Mol. Biol.* 151, 389-409.
40. Konigsberg, W. and Godson, G.N. (1983) *Proc. Natl. Acad. Sci.USA* 80, 687-691.
41. Sancar, G.B., Smith, F.W., Lorence, M.C., Rupert, C.S. and Sancar, A. (1984) *J. Biol. Chem. in press.*
42. Tang, M.-S., Lieberman, M.W. and King, C.M. (1982) *Nature* 299, 646-648.
43. Ohta, A., Waggoner K., Radominska-Pyrek, A. and Dowhan, W. (1981) *J. Bacteriol.* 147, 552-562.