Volume 14 Number 5 1986 Nucleic Acids Research

Multiple control elements for the *uvrC* gene unit of *Escherichia coli*

Surendra Sharma, Thomas F.Stark[1], Wanda G.Beattie[1] and Robb E.Moses[1*]

Department of Pathology, Section of Pathobiology, University of Texas Health Science Center, M.D.Anderson Hospital and Tumor Institute and [1]Department of Cell Biology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Received 3 December 1985; Accepted 14 January 1986

ABSTRACT

We have sequenced the control region of the uvrC protein including two open reading frames (ORF) encoding polypeptides of 28 kd and 23 kd molecular weight. The uvrC gene is preceded by five promoters. The P1, P2a and P2b promoter sequences are 5' to the 28 kd and the 23 kd proteins respectively. The P3 and P4 promoters are located within the structural gene for the 23 kd protein. The P3 promoter is required for adequate in vivo expression. There are three putative lexA protein binding sites, detected at the 3' end of the 28 kd protein (lexA1), within the coding sequences for the 23 kd protein (lexA2) and within the P3 promoter (lexA3). Promoter P2 is responsible for transcription of the uvrC gene, producing transcripts of 2.8 and 1.6 kb. The upstream region including the 28 kd protein is required for enhanced expression under non-induced conditions. These results show that the uvrC gene is controlled by multiple promoters and is transcribed as part of a multigene unit.

INTRODUCTION

The gene products of the uvrA, uvrB and uvrC genes control nucleotide excision repair of DNA damage (1). These genes have been cloned and code for polypeptides of molecular weights 114,000 (uvrA) (2), 84,000 (uvrB) (3) and 66,000 (uvrC) (4-6). The proteins bind to and excise in a sequential manner the DNA damage caused by UV (7). It has been shown that uvrA and uvrB belong to the "SOS" regulatory system (8-10).

The "SOS" regulatory system includes genes which are induced as a result of exposure to agents which damage DNA or block DNA replication. For example, DNA excision repair and recombination genes (uvrA, uvrB, uvrD, lexA, recA), a cell division gene (SfiA), mutagenesis genes (umuD), and other damage-induced genes (din) are under control of the "SOS" regulatory system (11). A common denominator found in "SOS" genes is the presence of an operator sequence, CTG-N10-CAG, overlapping the promoter region. The lexA protein binds to this operator site and prevents RNA polymerase from binding to the promoter. The uvrA and uvrB genes have been shown to contain lexA protein binding sites

overlapping their promoters (2,3,12). However, the uvrC gene seems to be regulated in a different manner. For example, the uvrC gene can not be induced by nalidixic acid and DNA damaging agents promote only minor and delayed onset of induction (13). These differences between the regulation of the uvrC gene and that of the uvrA and uvrB genes are surprising, since, based upon their functional relationship, these genes might be expected to follow a common regulatory mechanism.

We reported previously that the uvrC gene contains multiple promoter-like sites, determined by RNA polymerase binding experiments. There exist three putative RNA polymerase binding sites located approximately 0.4 kb (P3), 0.9 kb (P2) and 2.3 kb (P1) upstream from the structural gene, P1 showing strong binding and P2 and P3 showing weaker binding (14). Sancar et al. (15) have indicated the existence of an additional proximal promoter adjacent to the structural gene. We would designate this P4 in relation to the other promoters. Complementation testing led us to conclude that the distal promoters are required for optimal uvrC expression (14). In addition to the multiple RNA polymerase binding sites, the "maxicell" technique detected a previously unrecognized polypeptide, with a nominal molecular weight of 27 kd, 5' to the uvrC gene (16). The existence of multiple promoters with at least one 5'-encoded protein along with unknown regulatory mechanisms prompted further investigation of uvrC expression.

In this report, we present the nucleotide sequence of the control region of the uvrC gene and evidence that the uvrC gene is regulated by a complex network requiring the presence of several promoters for optimal expression in vivo. Transcription of the gene occurs as part of a multigene message with mRNA for a newly identified 23 kd protein, and is positively affected by expression of a newly identified 28 kd protein. The natural COOH terminus of the protein is required for optimal activity.

MATERIALS AND METHODS

Materials

Restriction enzymes, T4 DNA ligase and RNase T1 were purchased from Boehringer Mannheim Biochemicals, Collaborative Research, Inc., and Bethesda Research Laboratories. Growth media were purchased from Difco. Agarose and acrylamide were purchased from Bio-Rad. Ampicillin, tetracycline, and ethidium bromide were pruchased from Sigma. ^{32}P-labeled nucleotides and Pall membrane were purchased from ICN. ^{35}S-methionine was purchased from Amersham.

Bacterial Strains

AB1157 wild-type, and AB1884 uvrC$^-$ (1) were used as host for uvrC complementing plasmids. CSR603 uvrA$^-$, recA1, phr^{-1} (17) was used for maxicell analysis. JM103 (18) was used for growth of M13 cloning vectors.

Maxicell Analysis

Maxicell analysis was preformed as described (19) except that the cells were treated with cycloserine three times at 1 hr, 5 hrs and 21 hrs after UV-irradiation. The samples were fractionated by 10% SDS/polyacrylamide gel electrophoresis (20). The gels were treated with Autofluor as described by National Diagnostics, and autoradiography was performed.

Quantitative UV Survival

Cells were grown in L-broth (LB) containing appropriate antibiotic at 37°C to an O.D. 600 = 0.5. Cells were harvested, resuspended in an equal volume of M9 salts, and irradiated with a GE Germicidal lamp at a distance of 50 cm. Aliquots were taken at appropriate times, dilutions made and cells plated on LB plates. Plates were incubated at 37°C overnight and surviving colonies were counted.

Construction of uvrC Complementing Plasmids

All plasmids were grown and selected in the strain AB1884. Plasmids were screened for appropriate drug resistances, UV resistance and by restriction analysis. The plasmid pUVC1234 was constructed by ligating the 7.1 kb BamHI-EcoRI fragment of pPG1-L (21) into BamHI-EcoRI digested pBR322. The plasmid pUVC111 was constructed by ligating the 3.6 kb PstI fragment from pPG1-L into the PstI site of pBR322. The plasmid pUVC234 was made by inserting the 4.1 kb HindIII fragment of pUVC111 into the HindIII site of pBR322. Deletion of the 1.1 kb HincII fragment of pUVC234 was accomplished by partial HincII digestion and religation, resulting in the plasmid pUVC34. Partial BglII digestion of pUVC234 and pUVC1234 followed by complete BamHI digestion and religation resulted in the plasmids pUVC154 and pUVC4.

Plasmid DNA Isolation

Small amounts of DNA for restriction analysis were isolated as described (22). Large scale plasmid preparation was done using the cleared lysate procedure as described (BRL). The DNA was loaded onto a NACS-37 column and eluted with a 0.5 M-0.8 M NaCl gradient. Aliquots of different fractions were visualized on a 1% agarose gel and fractions containing only plasmid DNA were pooled.

Northern Analysis

Total RNA from AB1884 with appropriate plasmid was isolated by the

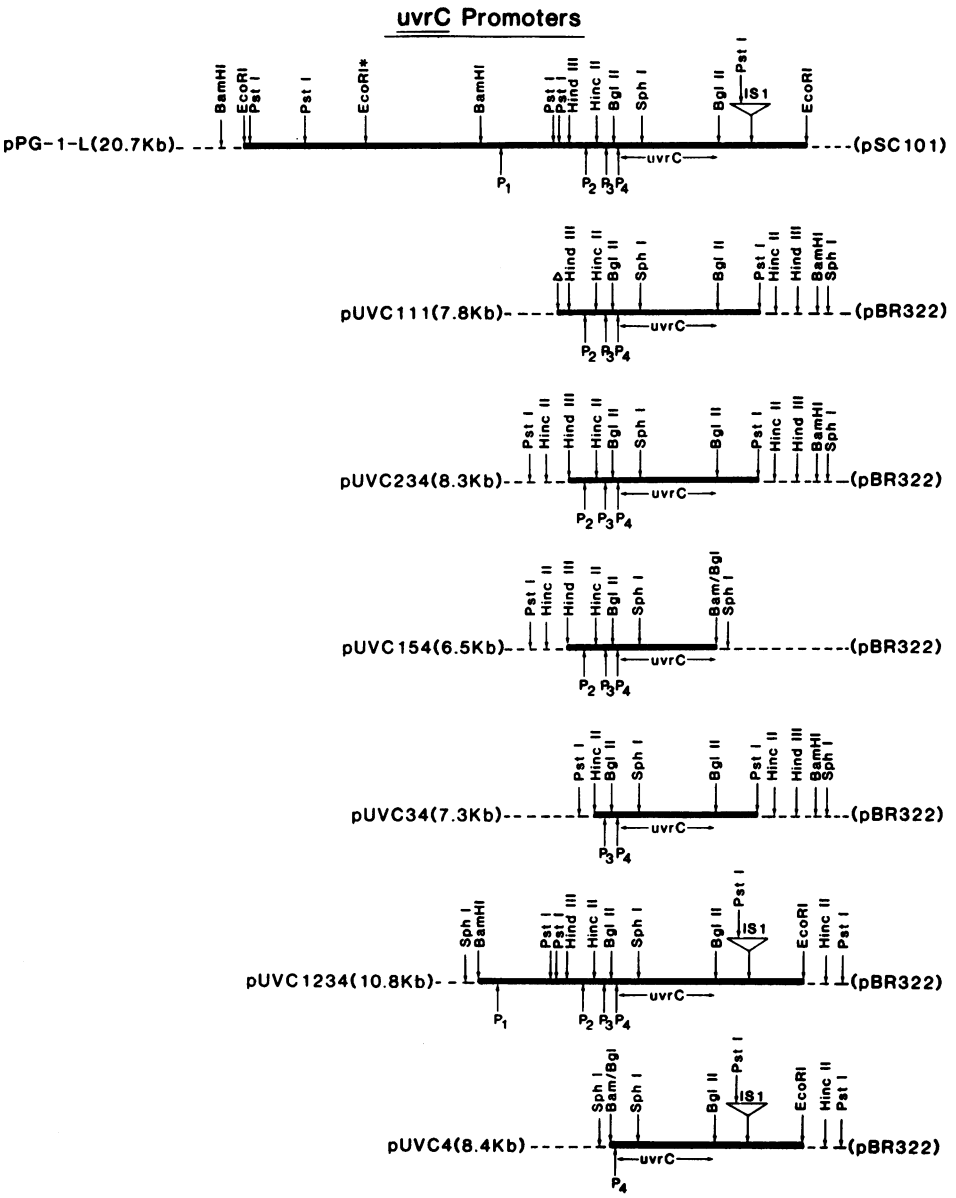


Figure 1. Map of uvrC Complementing Plasmids

procedure of van den Elzen et al. (23) with the exception that the RNA was precipitated in 3 M Na acetate, pH 6.0, two times before use. Twenty-µg samples of RNA were treated with glyoxal and fractionated on a 1% agarose gel

as described (24). Duplicate samples were cut from the gel and stained with ethidium bromide to visualize the 16S and 23S ribosomal RNAs. Unstained lanes were transferred to two Pall membranes by bidirectional transfer (25). Filters were hybidized with nick-translated probes and washed as described (26). The dried filters were exposed to Kodak XAR-5 film at -80°C for 1-2 days.

RESULTS

We have previously described a uvrC plasmid pUV7 (4) which contains the uvrC structural gene and approximately 2.6 kb of upstream sequence, but lacks 22 nucleotides at the uvrC 3' end. A similar plasmid, pUVC1234, containing the intact 3' end of the uvrC gene with the same 5' region has been constructed (Fig. 1). Plasmids pUV7 and pUVC1234 contain multiple promoters in the upstream DNA.

Biological Consequences of 5'- and 3'-Deletions in the uvrC Gene Unit

Existence of multiple promoters upstream from the uvrC gene raises the question of whether these control elements exert any influence on uvrC gene expression. To quantitate biological efficiency of the uvrC complementation in the presence or absence of these control elements, we used plasmids containing combinations of upstream sequences to confer UV-resistance to uvrC- cells. Figure 2A shows the quantitative UV survival of AB1884 (uvrC34) when transformed with pUVC1234, pUVC234, pUVC34, pUVC4 or pBR322. A wild-type strain, AB1157, transformed with pBR322 was also included in the UV test. Deletion of sequences upstream from the P4 promoter greatly diminishes the uvrC complementation, even on the high copy number plasmid pUVC4. These results agree with the finding of very weak affinity of promoter P4 for RNA polymerase, and also with maxicell analysis which shows very little uvrC synthesis with the plasmid pUVC4 (see below). Promoter P3, which may be "SOS"-inducible, is important, since plasmid pUVC34 is quite proficient in complementing the uvrC defect. It is notable that the plasmid pUVC1234 gives "super" complementation, well over wild-type levels.

To delineate the effect of alterations in the 3' region on uvrC complementation (Fig. 2B), we compared plasmid pUVC234, which contains the intact 3' end, with the plasmids pUVC154, pUV18, and pUV201, which were constructed by cutting at the 3' Bgl II site, creating a 22 bp deletion (4,16). In plasmids pUV18 and pVU201, the uvrC gene is preceded by P4 and the promoter for the tetracycline resistance gene. The orientation of the uvrC gene in pUV201 with respect to pBR322 is opposite in pUVC154 and pUV18. The published sequence data of the uvrC structural gene (15) and pBR322 (27) predict the 3'-nucleotide sequence of these plasmids. In both cases the 3' deletion results

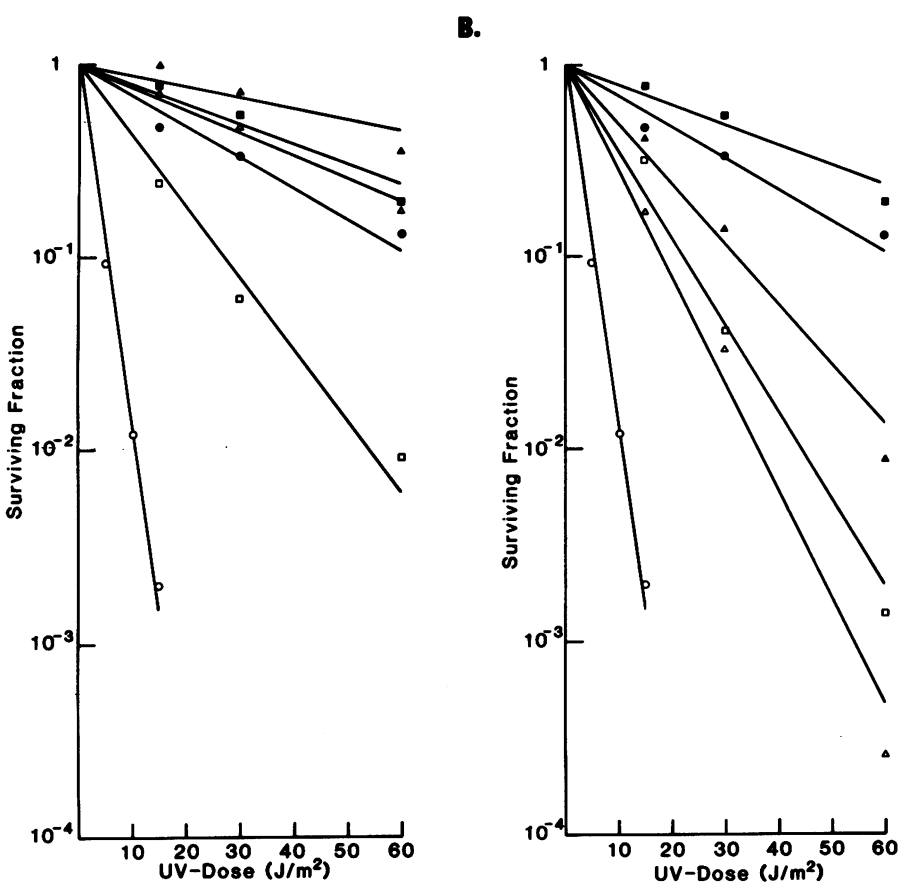


Figure 2. Complementation of uvrC plasmids.
AB1884 (uvrC34) and AB1157 (wild-type) transformed with various plasmids were tested for UV survival as described in Materials and Methods. A, the effect of the 5' region on UV survival. The symbols represent: ●, AB1157/pBR322; ○, AB1884/pBR322; ▲, AB1884/pUVC1234; ■, AB1884/pUVC234; Δ, AB1884/pUVC34; [], AB1884/pUVC4. B, the effect of 3' deletions on UV survival. The symbols represent: ●, AB1157/pBR322; ○, AB1884/pBR322; ■, AB1884/pUVC234; [], Δ, AB1884/pUVC154; ▲, AB1884/pUV201; Δ, AB1884/pUV18.

in longer encoded reading frames, (Fig. 3). The first stop codon reached in the predicted sequence of pUV201 is an amber stop codon, and the strain used in this study, AB1884, contains an amber suppressor. These plasmids promote synthesis of uvrC proteins which show altered mobility with maxicell analysis (data not shown).

When the UV survival curves of pUVC234 and pUVC154, which have the same

PEPTIDE TERMINATIONS

Natural Gene

```
                Bgl II
                  + 3940                3955
GGT CTG GCA GAA AAG ATC TTC TGG TCG TTG AAA CAT TGA
Gly Leu Ala Glu Lys Ile Phe Trp Ser Leu Lys His  *
```

65.7 kd

pUV201

```
              Bam/Bgl
                + 3940                3955                3970                3985                4000                4015
GGT CTG GCA GAA AAG ATC CAC AGG ACG GGT GTG GTC GCC ATG ATC GCG TAG TCG ATA GTG GCT CCA AGT AGC GAA GCG AGC AGG ACT GGG CGG
Gly Leu Ala Glu Lys Ile His Arg Thr Gly Val Val Ala Met Ile Ala  *  Ser Ile Val Ala Pro Ser Ser Glu Ala Ser Arg Thr Gly Arg

          4030                4045                4060                4075                4090
CGG CCA AAG CGG TCG GAC AGT GCT CCG AGA ACG GGT GCG CAT AGA AAT TGC ATC AAC GCA TAT AGC GCT AGC AGC ACG CCA TAG TGA
Arg Pro Lys Arg Ser Asp Ser Ala Pro Arg Thr Gly Ala His Arg Asn Cys Ile Asn Ala Tyr Ser Ala Ser Ser Thr Pro  *  *
```

65.9 kd or 70.5 kd

pUV18

```
              Bam/Bgl
                + 3940                3955                3970                3985                4000                4015
GGT CTG GCA GAA AAG ATC CTC TAC GCC GGA CGC ATC GTG GCC GGC ATC ACC GGC GCC ACA GGT GCG GTT GCT GGC GCC TAT ATC GCC GAC ATC
Gly Leu Ala Glu Lys Ile Leu Tyr Ala Gly Arg Ile Val Ala Gly Ile Thr Gly Ala Thr Gly Ala Val Ala Gly Ala Tyr Ile Ala Asp Ile

          4030                4045                4060                4075                4090                4105
ACC GAT GGG GAA GAT CGG GCT CGC CAC TTC GGG CTC ATG AGC GCT TGT TTC GGC GTG GGT ATG GTG GCA GGC CCG TGG CCG GGG GAC TGT TGG
Thr Asp Gly Glu Asp Arg Ala Arg His Phe Gly Leu Met Ser Ala Cys Phe Gly Val Gly Met Val Ala Gly Pro Trp Pro Gly Asp Cys Trp

        4120                4135                4150                4165                4180
GCG CCA TCT CCT TGC ATG CAC CAT TCC TTG CGG CGG CGG TGC TCA ACG GCC TCA ACC TAC TAC TGG GCT GCT TCC TAA
Ala Pro Ser Pro Cys Met His His Ser Leu Arg Arg Arg Cys Ser Thr Ala Ser Thr Tyr Tyr Trp Ala Ala Ser  *
```

73.3 kd

Figure 3. Predicted uvrC Protein Terminations
Using the published sequence data for the uvrC gene and pBR322, the termination of the uvrC gene in different plasmids is shown. The plasmid pUVC154 has the same predicted termination as the plasmid pUV18. The molecular weights of the resultant proteins are shown.

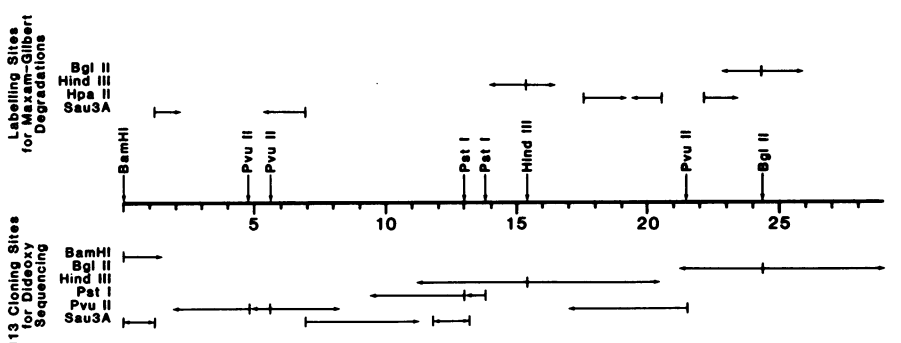
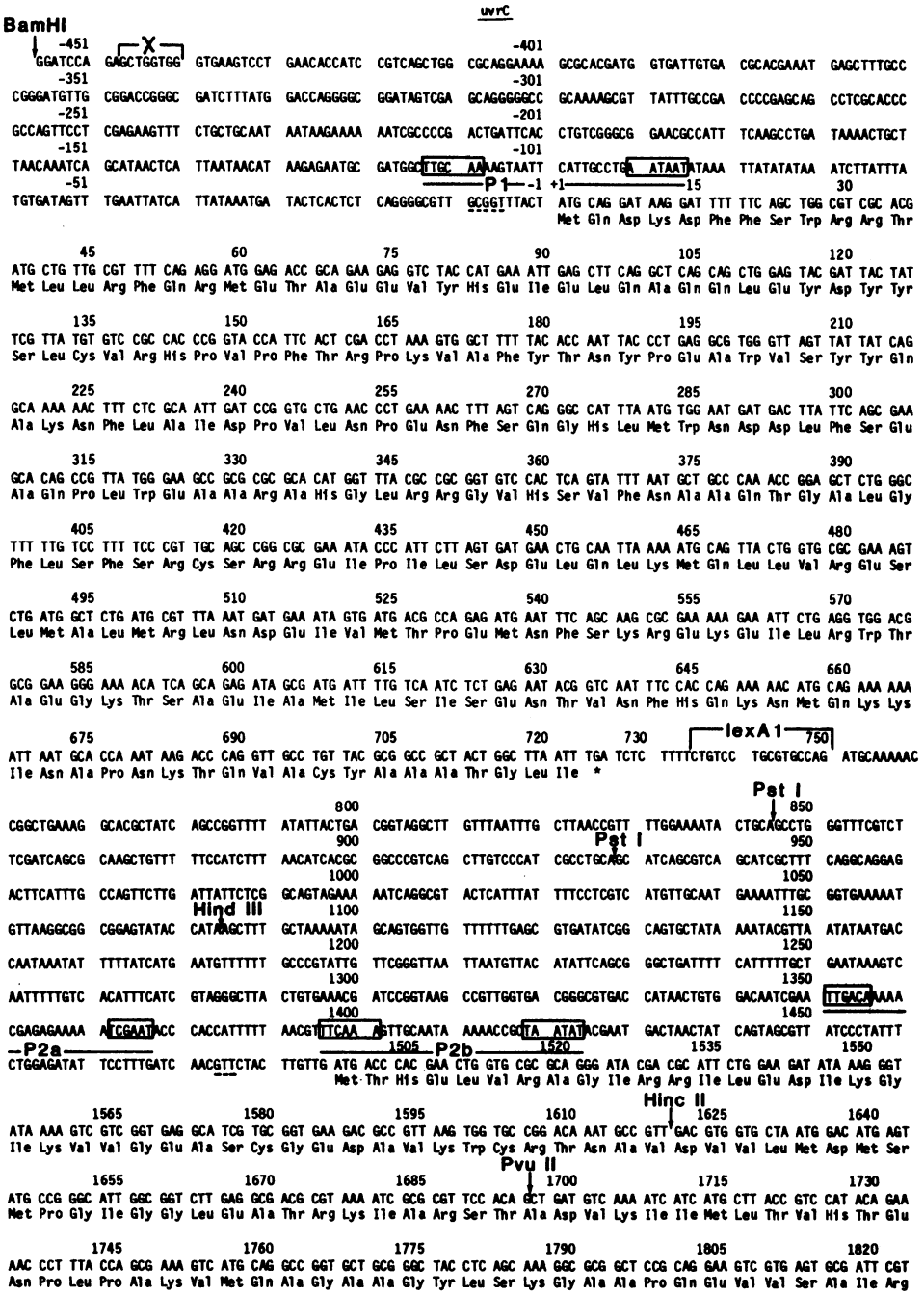


Figure 4. Sequencing Strategy.
A partial restriction map of the uvrC regulatory region is shown by the heavy line. Arrows above the line indicate sequences determined by the Maxam and Gilbert chemical degradation method of Maxam and Gilbert (28). Arrows below the line indicate sequences determined by the M13 dideoxy method of Sanger (29). Right-pointing arrows indicate sequences read from the coding strand, and left-pointing arrows indicate sequences read from the non-coding strand.


BamHI
uvrC
X
P1
P2a
P2b
lexA1
Pst I
Pst I
Hind III
Hinc II
Pvu II

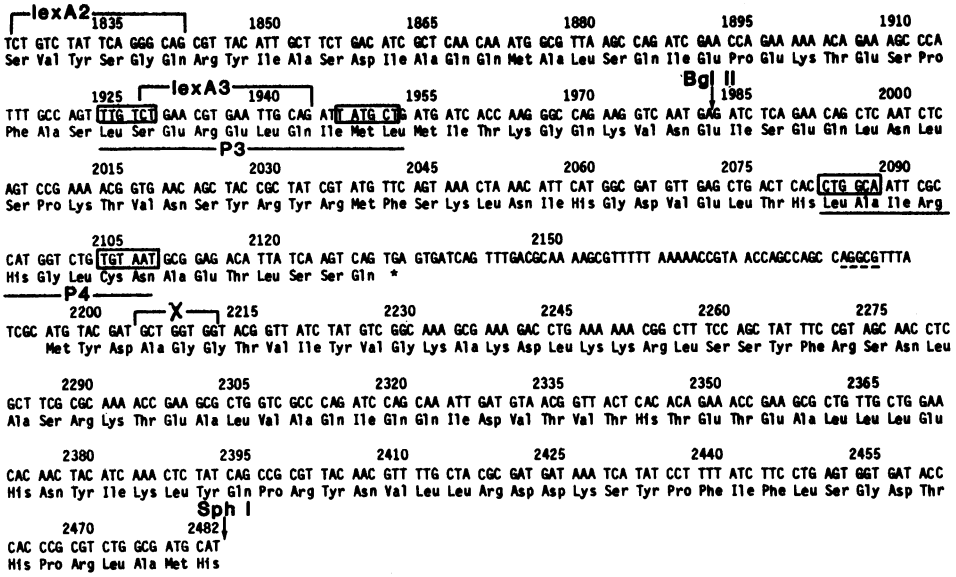


Figure 5. Sequence of the uvrC Regulatory Region
Numbering of the sequence starts with the first nucleotide of open reading frame 1 (ORF1). ORF1 starts at base 1 and ends at base 723. ORF2 starts at base 1497 and ends at base 2129. Promoter sequences are underlined with the -10 and -35 regions boxed. χ sites and lexA protein binding sites are bracketed above the sequence. Shine-Delgarno sequences are shown as dashes under the corresponding bases.

promoter regions, are compared, it is apparent that pUVC154 gives approximately 10% of the survival of pUVC234, thus suggesting an important role for the COOH end of the uvrC protein. When the curves of pUV18 and pUV201, which both contain the same promoter region, are compared, pUV201 is more proficient in complementing the uvrC defect. It appears that the COOH terminus of uvrC cannot be extended and retain full biological activity. These results indicate that the presence of distal-5' end and intact 3'-end sequences are important for optimal activity of the uvrC gene product.

Nucleotide Sequence of the uvrC Control Region

A detailed restriction map of the 2933 bp BamHI-SphI fragment was constructed and used to generate DNA fragments for sequencing using either the M13 Sanger method (28) or the Maxam-Gilbert method (29). The strategy for sequencing this fragment is shown in Fig. 4.

The nucleotide sequence of the Bam HI-Sph I fragment containing the uvrC control region and its N-terminus is shown in Fig. 5. The putative 5' end (N-

Table 1A. uvrC Promoters

	-35		-10
Consensus	TTGACA	N=17±2	TATAAT
P1	TTGCAA	N=17	AATAAT
P2a	TTGACA	N=15	TCGAAT
P2b	TTCAAA	N=17	TAATAT
P3	TTGTCA	N=17	TATGCT
P4	CTGGCA	N=15	TGTAAT

Table 1B. lexA Binding Sites

Consensus	CTG	N=10	CAG
lexA1	CTG	N=10	CAG
lexA2	CTG	N=11	CAG
lexA3	CTG	N=11	CAG

terminus) of the uvrC structural gene was approximately localized by subcloning of the uvrC gene unit and by monitoring the uvrC complementing ability of resultant plasmids (4). To identify the location of the 5' end in the sequence, we looked for an open reading frame downstream from the Bgl II site. The ATG initiator codon at 2195 bp position is followed by such an open reading frame. This agrees with the published nucleotide sequence of the uvrC structural gene (15). A potential Shine-Delgarno (SD) sequence AGGCG is located 13 bp upstream from the initiator ATG codon.

Promoter Regions

The uvrC gene is preceded by three promoters (P1, P2 and P3), identified by heparin-resistant RNA polymerase DNA complex formation (14). P2 and P3 were identified proximal to the 5' end of the uvrC gene while the P1 promoter is separated from the uvrC structural gene by an interposed DNA region of approximately 1.6 kb. In order to identify putative promoter regions in the nucleotide sequence, a search for segments of DNA containing homology to the -35 and -10 regions found in E. coli promoters was performed. Regions containing homology to the consensus sequence (30) are listed in Table 1A and are represented in Fig. 6. The locations for promoters P1, P2 and P3 are consistent with our previous findings. The location of promoter P4 agrees with

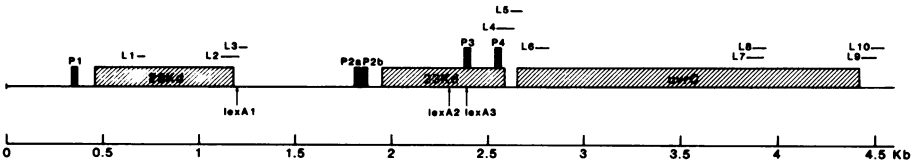


Figure 6. Control Elements of the uvrC Regulatory Region
Positioning of promoters, structural genes and lexA protein binding sites are as determined by sequence analysis. Putative loop structures are marked L1-L10. The thermodynamic stabilities of loop structures are (in kcal): L1, -20.8; L2, -21.4; L3, -25.8; L4, -20.4; L5, -26.8; L6, -24.4; L7, -21.6; L8, -21.0; L9, -21.8; L10, -20.4.

the published observation of this promoter (15). We noted two potential promoters in the region appropriate for P2, and will call these promoters P2a and P2b. The P1 promoter shows strong binding to RNA polymerase whereas

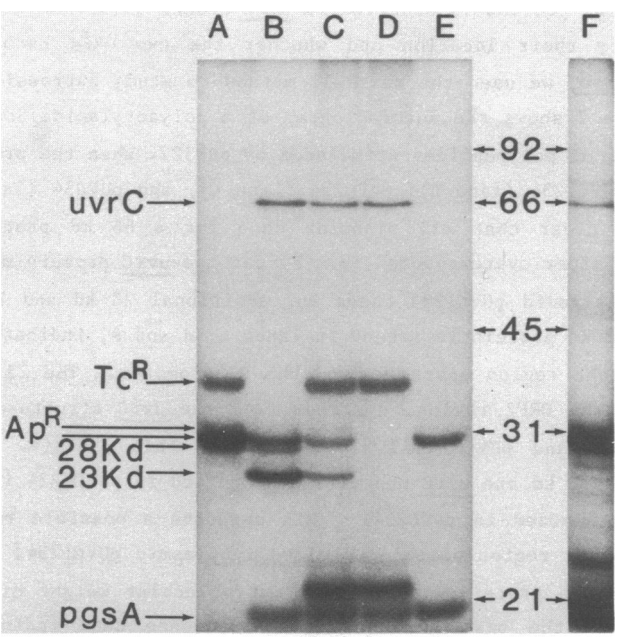


Figure 7. ^{35}S-Methionine Labeled Proteins Encoded by uvrC Complementing Plasmids
Autoradiogram of 10% polyacrylamide gel. Lane: A, pBR322; B, pUVC1234; C, pUVC234; D, pUVC34; E, pUVC4; F, pUVC4 (overexposed). Standard molecular weight markers: phosphorylase b, 92,000; bovine serum albumin, 66,000; ovalbumin, 45,000; carbonic anhydrase, 31,000; soybean trypsin inhibitor, 21,000.

promoters P2 and P3 exhibit weaker binding, and binding to the P4 promoter is not detected (14). These results could be explained on the basis of an AT rich region surrounding these promoters and the extent of homology to the consensus -35 and -10 sequences. For example, promoter P1 is bounded by a very high AT content (Fig. 5), a characteristic enabling easy access by DNA binding proteins, and shows good homology to the consensus -35 and -10 sequences. On the other hand, promoter P4 lacks AT-rich boundaries and does not show a perfect match to the -35 sequence requirements of a model E. coli promoter.

Evidence That the uvrC Upstream Region Encodes Two Polypeptides

A search for transcriptional units indicated that the uvrC upstream region contains two open reading frames (ORFs). ORF1 (Fig. 5) is capable of encoding a 28 kd protein and is preceded by the promoter P1. A 23 kd protein structural gene lies downstream from P2 (ORF2). Interestingly, it contains promoters P3 and P4 within its coding region (Fig. 6). These ORFs can account for "27 kd" proteins previously observed in uvrC clones (4,5).

To verify their location and whether the two ORFs encoded any polypeptides in vivo, we used the maxicell method to study expression of polypeptides. Figure 7 shows the autoradiogram of a polyacrylamide/SDS gel. Lane A shows labeling of polypeptides stimulated by pBR322. When the proteins encoded by plasmids pUVC1234 (lane B), pUVC234 (lane C), and pUVC34 (lane D) are compared, it is clear that all plasmids code for a 66 kd protein, the uvrC protein. Only after overexposure (lane F) can the uvrC protein made from pUVC4 be seen. The plasmid pUVC1234 codes for additional 28 kd and 23 kd polypeptides. The 28 kd protein is absent in lanes c, d and e, indicating that it is coded for by the region upstream from the P2 promoter. The 23 kd protein is coded for by the ORF2 region, upstream from the uvrC structural gene, since plasmids pUVC34 and pUVC4 fail to demonstrate this protein. The relative amounts of the 23 kd and uvrC proteins synthesized in pUVC1234 is much greater than that synthesized in pUVC234. This suggests a possible regulatory role for the distal 5' region of DNA contained in plasmid pUVC1234. Lanes B, C, D, and E also show a protein with an apparent molecular weight of 19 kd. This band represents the pgsA gene product. Sequence data indicate that this protein has a molecular weight of 24 kd, but it has been shown that with the SDS-PAGE system used here, the protein migrates with an apparent molecular weight of 19 kd (31). The 28 kd and 23 kd proteins represent previously unidentified entities. The E. coli genetic map is sparse in this region and suitable candidate genes have not been identified.

Codon Usage of the 23 kd and the 28 kd Proteins

It has become apparent that some codons are used much less frequently in coding reading frames than in noncoding reading frames in E. coli. Konigsberg and Godson (32) have suggested that regulatory proteins show a higher frequency of such rare codons (9.29%), than do nonregulatory proteins (4.24%). Other investigators have suggested that rare codons are used by E. coli as a way of regulating the rate of translation, therefore maintaining a low level of the given protein in the cell. Analysis of the codon usage in the 23 kd and 28 kd proteins shows that the 23 kd protein has normal rare codon usage (6.2%), while the 28 kd protein has increased rare codon usage (10.4%). It has previously been shown that the uvrC gene exhibits increased rare codon usage (15). These results suggest the the cell maintains only small quantities of the 28 kd and uvrC proteins.

lexA Binding Sites

To determine whether the uvrC gene belongs to the "SOS" cascade, we searched the uvrC upstream sequence for lexA protein binding sites. As shown in Table 1B, there exist three such sites (lexA1, lexA2 and lexA3). The lexA1 site is located just 3' to the structural gene of the 28 kd protein, and the lexA2 and lexA3 sites are within the structural gene for the 23 kd protein (lexA2) and within the P3 promoter (lexA3). It has been reported that a lexA protein binding site overlaps the P3 promoter (13), agreeing with the lexA3 site reported here. The spacing between CTG-CAG in the case of lexA3 differs by one nucleotide from the consensus SOS sequence. This, along with the presence of a second lexA binding site (lexA2) only 100 bases upstream, suggests unusual induction of this promoter.

Transcription Stop Signals

Hairpin loop structures which could act as transcription terminators or attenuators were searched for. The loops which have a thermodynamic stability of less than -20 kcal are shown in Fig 6. L2 overlaps the 3' end of the 28 kd coding region and encompases lexA1. L3 has all the features of a ρ-independent terminator and is in the proper position to terminate transcripts of the 28 kd protein. L3 overlaps L2 by one base. Several loops are present between P4 and the 3' end of the uvrC gene. These loops could act as attenuators.

Other Sequences

Although no genetic evidence implicates the uvrC gene locus in recombinational events, a search was made for χ sites, an octameric sequence, GCTGGTGG, known as a recombination-promoting sequence (33). A perfect match for this octamer was found at the beginning of the uvrC coding region (position 2103)

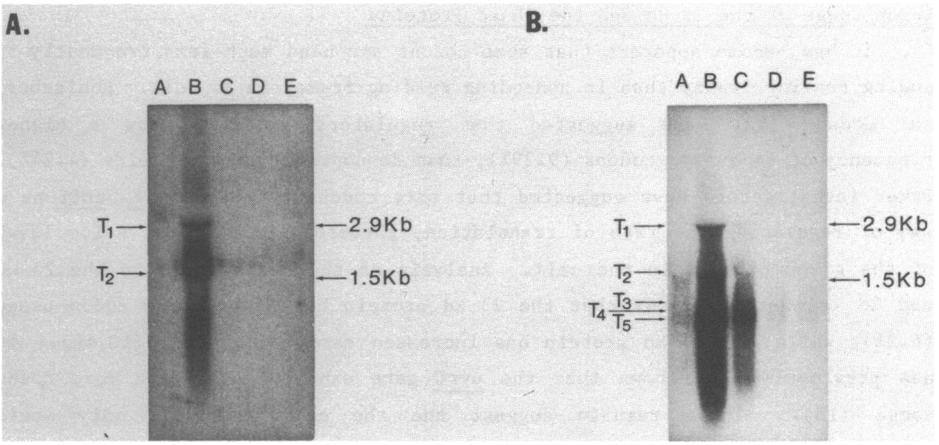


Figure 8. uvrC RNA Hybridization
Total from AB1884 transformed with uvrC complementing plasmids was isolated and hybridized to nick-translated DNA as described in Materials and Methods. Lanes: A, pBR322; B, pUVC1234; C, pUVC234; D, pUVC34; E, pUVC4. A, the RNA was hybridized to the SphI-BglII fragment encompassed by the uvrC structural gene (Fig. 3). B, RNA hybridized to the HindIII-PvuII fragment overlapping P2a and P2b (Figure 3).

and at an upstream location (position -453). A search for the CAP binding site (TGTGA) revealed 3 such sites, at bases -388, -60 and 1292.

In Vivo Transcription Studies

To determine which promoters are involved in uvrC gene transcription under non-induced conditions, in vivo transcription of the uvrC gene was studied. It was also important to determine whether upstream sequences including the 28 kd and 23 kd protein function as regulatory signals. To identify the mRNA species from the uvrC gene unit, we carried out RNA blot hybridization experiments. Total RNA from AB1884 transformed with pUVC1234, pUVC234, pUVC34, pUVC4, or pBR322, was fractionated by electrophoresis, blotted to Pall membranes, and hybridized to a ^{32}P-labeled Sph I-BglII fragment (Fig. 1). This fragment is contained within the uvrC structural gene, and should detect any uvrC transcripts.

An autoradiogram obtained after hybridization is shown in Fig. 8A. Two major transcripts of approximately 2.8 kb and 1.6 kb were detected in lane B, representing plasmid pUVC1234. These bands are much weaker in lane C, which contained RNA from AB1884 transformed with pUVC234. Lanes which contained RNA from AB1884 transformed with pUVC34, pUVC4, or pBR322 show very low signals. However, definite signals are present in such cells. This argues for mRNA

production in this uvrC⁻ mutant and detection of single copy gene expression. Because the 2.8 kb band is present in RNA made from the cells containing pUVC234, this band is not coming from P1. The 1.6 kb band is not big enough to represent the entire uvrC gene, or come from P1. Together, these results indicate that the transcripts are not coming from P1.

To determine at which promoters these messages originate, we repeated the above experiment using a Hind III-Pvu II fragment as a probe. This probe overlaps P2 and should differentiate between transcripts originating from P2 and those originating from P3 or P4. As can be seen in Fig. 8B, this probe hybridizes to both the 2.8 kb and the 1.6 kb bands along with several bands in the 0.8-1.0 kb range. It is also apparent that the majority of the signal is coming from the smaller (0.8-1.0 kb) bands. Since these small bands are not large enough to come from P1, we assume that these are terminated transcripts from P2. The hairpin loops L4, L5, and L6 are the proper distance from P2 to account for bands of this size. Once again, the plasmid pUVC1234 gives a stronger signal than the other plasmids, suggesting a positive regulatory effect from the region of DNA 5' to the HindIII site, including the structural gene for the 28 kd protein.

Based on the size of the transcripts detected by the two probes and the intensity of the signals from the different plasmids, we conclude that all of the transcripts are originating from P2, terminating either in or at the end of the uvrC structural gene, and that the region of DNA 5' to the HindIII site has a positive regulatory affect on P2.

DISCUSSION

Analysis of the uvrC gene unit shows it to be part of a complex regulatory system. Expression of many prokaryotic genes either in a polycistronic operon or in an independent gene unit is controlled by differential translational efficiencies, physiologically-modulated multiple promoters or by attenuated transcriptional stop signals. The experiments described here suggest that all of these regulatory mechanisms may be involved in the control of the uvrC gene.

Identification of promoters P3 and P4 within the 23 kd protein suggests regulatory mechanisms. Transcription from the P3 and P4 promoters may face constraints due to active transcription from P2. A similar phenomenon has been shown to be true for the trp and mms operons (34,35). Such "promoter occlusion" may totally inactivate the already weak promoter P4 when the upstream promoters P2 and P3 are active.

The results with RNA hybridization suggest that the multiple hairpin loops may be acting as attenuators, thereby creating a method for the cell to differentially control two genes with the same promoter (P2). It seems possible that the cell might keep relatively low constitutive levels of the uvrC protein, while maintaining greater levels of the 23 kd protein. When the cell is induced by DNA damage the DNA in this region may undergo a conformational change allowing unimpeded transcription through the uvrC gene. The presence of the overlapping loops L4 and L5 at the start of the uvrC coding region is interesting in this respect. Under normal conditions the more thermodynamically stable L5 may act as a terminator. Increased ribosomal binding to the mRNA during induction may favor formation of L4, thus destroying the terminator L5, leading to increased transcription through the uvrC gene. A similar model has been proposed for regulation of the uvrD gene (36,37). The arrangement of the P3 and P4 promoters within the 23 kd structural gene might allow increased production of mRNA for the uvrC gene, while transcription of the 23 kd mRNA is constant. Such stepped production of the uvrC message would presumably occur by loss of inhibition.

It is also apparent from the maxicell and RNA hybridization data that the region of DNA 5' to the HindIII site is responsible for enhanced transcription from P2. Since it appears that no transcript from P1 is made through this area, it seems that the 28 kd protein has a positive regulatory affect on P2.

It has been shown by Van Sluis et al. (13) that P3 is SOS-inducible, but that the induction is delayed and rather modest as compared to other SOS-induced genes. This result is not in conflict with our findings. It is possible that under induced conditions P3 is activated. This activation could be facilitated by reduced transcription from P2. If in fact transcription from P2 is positively controlled by the 28 kd protein, reduction in levels of the 28 kd protein would create such an effect. The positioning of lexA1 just 3' to the 28 kd protein is interesting in this respect. The activation of P3 and the inactivation of attenuators in the uvrC gene could combine to produce a rather large induction of the uvrC protein. Other investigators do not find SOS induction of P3 (38).

Understanding the biological roles of the 28 kd and 23 kd polypeptides may be necessary to understand the expression of the uvrC gene. The nucleotide sequence of these polypeptides does not seem to show homology to any known regulatory proteins or the uvrC stuctural gene. However, our results suggest that the 28 kd protein may be a prerequisite for enhanced transcription from the P2 promoter.

The presence of multiple promoters, lexA protein binding sites, and putative translational attenuators, along with an upstream encoded regulatory protein, suggests a complex control mechanism for two genes encoded in a single transcript. Such a mechanism could account for the coordinate or independent control of the 23 kd and uvrC proteins, depending on the cellular environment. Sequence data for the pgsA gene and the glyW gene show that the pgsA stuctural gene starts just 56 bp from the uvrC termination and that the glyW gene starts 51 bp 3' to the pgsA gene. The promoter for the pgsA gene is embedded in the uvrC structural gene and the glyW promoter is embedded in the pgsA structural gene (31,39). This suggests that these genes could also be influenced by similar mechanisms.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health.

*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. (1981b) J. Mol. Biol. 148, 45-62.
3. Sancar, A., Clarke, N.D., Griswold, J., Kennedy, W.J. and Rupp, W.D. (1981a) J. Mol. Biol. 148, 63-76.
4. Sharma, S., Ohta, A., Dowhan, W. and Moses, R.E. (1981) Proc. Natl. Acad. Sci. USA. 78, 6033-6037.
5. Sancar, A., Kacinski, B.M., Mott, D.L. and Rupp, W.D. (1981c) Proc. Natl. Acad. Sci. USA. 78, 5450-5454.
6. Yoakum, G.H. and Grossman, L. (1981) Nature 292, 171-173.
7. Sancar, A. and Rupp, W.D. (1983) Cell 33, 249-260.
8. Kenyon, K.J. and Walker, G.C. (1981) Nature 289, 808-810.
9. Fogliano, M. and Schendel, P.F. (1981) Nature 289, 196-198.
10. Little, J.W. and Mount, D.W. (1982) Cell 29, 11-22.
11. Walker, G.C. (1984) Microbiological Rev. 48, 60-93.
12. van den Berg, E., Zwetsloot, J., Noordemeer, I., Pannekoek, H., Dekker, B., Dijkema, R. and van Ormondt, H. (1981) Nucl. Acids Res. 9, 5623-5643.
13. van Sluis, C.A., Moolenaar, G.F. and Backendorf, C. (1983) EMBO 2, 2313-2318.
14. Sharma, S., Stark, T. and Moses, R.E. (1984) Nucl. Acids Res. 12, 5341-5353.
15. Sancar, G.B., Sancar, A. and Rupp, W.D. (1984) Nucl. Acids Res. 12, 4593-4608.
16. Sharma, S., Dowhan, W. and Moses, R.E. (1982) Nucl. Acids Res. 10, 5209-5221.
17. Sancar, A. and Rupert, C.S. (1978) Nature 272, 471-472.
18. Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucl. Acids Res. 9, 309.
19. Sancar, A., Hack, A.M. and Rupp, W.D. (1979) J. Bacteriol. 137, 692-693.
20. Laemmli, U.K. (1970) Nature 227, 680-685.

21. Ohta, A., Waggoner, K., Radominsk-Pyrek, A. and Dowhan, W. (1981) J. Bacteriol. 147, 552-556.
22. Birnboim, H.C. and Doly, J. (1979) Nucl Acids Res. 7, 1513-1523.
23. van den Elzen, P.J.M., Konings, R.N.H., Veltkamp, E. and Nijkamp, H.J.J. (1980) J. Bacteriol. 144, 579-591.
24. Thomas, P. (1980) Proc. Natl. Acad. Sci. USA. 77, 5201-5205.
25. Smith, G.E. and Summers, M.D. (1980) Anal. Biochem. 109, 123-129.
26. Bergsma, D.J., Chang, K.S. and Schwartz, R.J. (1985) Mol. Cell Biol. 5, 1151-1162.
27. Sutcliffe, J.G. (1978) Cold Spring Harbor Symp. Quant. Biol. 43, 77-90.
28. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
29. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 1858-1865.
30. Hawley, D.K. and McClure, W.R. (1983) Nucl. Acids Res. 11, 2237-2255.
31. Gopalakrishnan, A.S., Chen, Y.C., Temkin, M. and Dowhan, W. (1985) J. Biol. Chem., in press.
32. Konigsberg, W. and Godson, N.G. (1983) Proc. Natl. Acad. Sci. USA. 80, 687-691.
33. Smith, G.R., Kunes, S.M., Schultz, D.W., Tayler, A. and Triman K.L. (1981) Cell 24, 429-436.
34. Horowitz, H. and Platt, T. (1982) J. Mol. Biol. 156, 257-267.
35. Lupski, J.R., Ruiz, A.A. and Godson, G.N. (1984) Mol. Gen. Genet. 195, 391-401.
36. Easton, A.M. and Kushner S.M. (1983) Nucl. Acids Res. 11, 8625-8639.
37. Finch, P. and Emmerson, P.T. (1983) Gene 25, 317-323.
38. Forster, J. and Strike, P. (1985) Gene 35, 71-82.
39. Tucker, S.D. and Murgola, E.J. (1985) Submitted.