Transcription of the *uvrD* gene of *Escherichia coli* is controlled by the *lexA* repressor and by attenuation

Alan M. Easton and Sidney R. Kushner

Department of Molecular and Population Genetics, University of Georgia, Athens, GA 30602, USA

Received 14 September 1983; Revised and Accepted 14 November 1983

ABSTRACT

The nucleotide sequence of the control region and the presumptive N-terminal portion of the *uvrD* gene of *Escherichia coli* K-12 has been determined. The 1190 base pairs of DNA examined include the likely coding sequence for the first 258 amino acids of the *uvrD* protein. The transcription promoter for the *uvrD* gene was identified upstream of the protein coding region. Synthesis of messenger RNA *in vitro* from this promoter was inhibited by purified *lexA* protein. The *lexA* protein was found to bind downstream from the promoter at a sequence, CTGTATATATACCCAG, which is homologous to other known *lexA* protein binding sites. In the absence of the *lexA* protein, approximately half of the messages initiated *in vitro* at the *uvrD* promoter terminate after about 60 nucleotides at a sequence which resembles a rho-independent terminator. These results indicate that the *uvrD* gene is induced during the SOS response, and that the expression of the gene may also be regulated by transcription attenuation.

INTRODUCTION

The response of *Escherichia coli* cells to agents which damage DNA or block DNA replication is controlled by the SOS regulatory system (1). The levels of expression of over ten genes are increased during SOS induction, a process which depends on the *recA* and *lexA* proteins (1). For some of the induced genes, it has been shown that the *lexA* protein acts by binding to operator sites near the promoters of the genes to repress their transcription. SOS induction appears to involve cleavage of unbound molecules of the *lexA* repressor. The resulting decrease in the *lexA* protein concentration brings about derepression of the SOS inducible genes (1). Genes known to be under this type of regulation include *uvrA* (2), *uvrB* (3), *sulA* (4), *himA* (5), the *umuDC* operon (6), the colicin E1 gene (7), the cloacin gene (8), other *din* (damage inducible) genes (9), as well as the *recA* (10) and *lexA* (11,12) genes themselves.

lexA protein binding sites near the promoters of the *uvrA* (13), *uvrB* (14), *recA* (15,16) and *lexA* (15,16) genes have been identified and sequenced. These sites, called SOS boxes (1), have in common the nucleotides CTG and

CAG separated by ten nucleotides which are mostly A or T residues. Sequences like these have been found near the promoters of the sulA gene (17), the colicin E1 and cloacin genes (8) and the rpsU-dnaG-rpoD operon (18), although there are not yet any reports that the lexA protein binds to these particular sites.

Since the number of genes under the control of lexA is rapidly growing, it was of interest to examine the regulation of additional DNA repair loci. The uvrD gene seemed to particularly warrant analysis, because mutants in this locus exhibit higher levels of spontaneous mutagenesis (19), are deficient in methyl directed DNA mismatch repair (20) and demonstrate increased sensitivity to ultraviolet light (21). With the recent cloning of the uvrD structural gene (22,23,24) it has become possible to examine the DNA sequence of the control region.

In this communication the transcription promoter for the uvrD gene is identified. Evidence from in vitro experiments indicates that the transcription of the uvrD gene is regulated by the lexA protein. In addition, the in vitro transcription experiments have demonstrated a short mRNA species encoded near the beginning of the uvrD gene. The synthesis of this mRNA also did not occur in the presence of the purified lexA protein. Accordingly, it is suggested that the uvrD gene may be controlled by transcription attenuation as well as by the lexA repressor.

MATERIALS AND METHODS

Strains and plasmids and plasmid DNA isolation

The DNA used for determination of the nucleotide sequence was obtained from the plasmid pVMK52 which was constructed by Valerie Maples in this laboratory. pVMK52 was made by insertion of a 2.9 kb SalI fragment which contains the coding sequence of the entire uvrD gene, into the SalI site of the plasmid pBR322 (25). The 2.9 kb SalI fragment was derived from a genomic 2.9 kb PvuII fragment whose termini had been converted in vitro into SalI recognition sites. pVMK52 plasmid DNA was isolated from the *E. coli* K12 strain SK3952 (22). pBR322 and pJL3 (26) plasmid DNA were isolated from the *E. coli* strain SK1590 (27) according to the method of Ish-Horowicz and Burke (28).

DNA sequencing

The nucleotide sequence of DNA was determined using the method of Maxam and Gilbert (29) except that the A+G reactions were accomplished using formic acid (30). Restriction endonucleases used in this procedure were purchased

from Bethesda Research Laboratories. T4 polynucleotide kinase was obtained from PL Biochemicals and bacterial alkaline phosphatase came from Worthington. Electrophoresis of DNA fragments was performed using 5% acrylamide (Fisher), 10% glycerol gels in TBE buffer (89 mM Tris, 2.5 mM EDTA, 89 mM boric acid, pH 8.1). DNA restriction fragments were obtained by electroelution of fragment DNA out of gel slices into dialysis bags containing TA buffer (40 mM Tris, 20 mM sodium acetate, 2 mM EDTA, pH 7.8). The DNA was collected by ethanol precipitation and the pellets were rinsed with 70% ethanol/30% water to remove salt. Separation of individual strands of labeled DNA was performed as outlined (31) except that the DNA fragments were denatured by heating at 90°C for one minute in a solution of 80% formamide (Fisher), 10 mM NaOH, 1 mM EDTA 0.1% xylene cyanol and 0.1% bromophenol blue prior to loading on the gel.

In vitro transcription

Detection of in vitro synthesized RNAs was performed as previously described (32) except that only 2 to 5 microcuries of [α -³²P] UTP (Amersham, >400 Ci/mmol) and only 1 pmole of RNA polymerase were used per reaction. Transcription reactions were terminated by adding to the 20 μ l reaction volumes 8 μ l of TBE buffer containing 2.0% SDS and 8 μ l of formamide containing 0.1% xylene cyanol and 0.1% bromophenol blue. The volumes of the samples were reduced to about 8 μ l in a Speed Vac concentrator (Savant). Transcripts were displayed on 8% polyacrylamide 7 M urea gels which were 0.3 mm thick and 40 cm long. The gels were run at 52°C. E. coli RNA polymerase was obtained from New England Biolabs. Ribonucleotide triphosphates were obtained from P-L Biochemicals. Purified uvrD protein was provided by Valerie Maples.

DNA restriction fragments containing the uvrD gene promoter were isolated from restriction digests of pVMK52 DNA. A 315 base pair EcoRI to Sau3A fragment containing the β -lactamase gene promoter was isolated from the plasmid pBR322 (25). The runoff RNA transcript produced from this fragment is 144 bases long (33). A 283 base pair (SstII to HaeIII fragment containing the recA gene promoter was obtained from the plasmid pJL3 (26) which was provided by John Little. This fragment, containing the recA promoter, serves as a template for a 95 base long runoff RNA transcript (34).

Size standards were single-stranded DNA fragments. These were obtained as follows: pBR322 plasmid DNA was cleaved with HinfI and EcoRI, the 5' ends of the fragments were labeled with ³²PO₄ as in the Maxam and Gilbert procedure (29), and the fragments subsequently were cleaved with HpaII.

Figure 1. (a) Restriction map of the control region and N-terminal coding sequence of the *uvrD* gene. (b) Scheme for determining the nucleotide sequence of the region. The arrows indicate the location and direction of the nucleotide sequencing runs. All sequenced fragments were labeled at their 5' ends. (c) Location of the promoter (referred to as promoter 1 in text), SOS box and protein coding region. The first nucleotide of the protein coding region is designated as +1. The numbers are the scale of nucleotide bases.

Protection of *lexA* binding sites from DNaseI degradation

DNA restriction fragments labeled at one 5' end were partially degraded with DNase I in the presence and absence of purified *lexA* protein. The procedure followed was that of Little *et al.* (15). In our experiments, the concentration of the *lexA* repressor protein was 400 nM. The DNA fragment concentration was 20 nM. Purified *lexA* protein was the generous gift of John Little.

RESULTS

Nucleotide sequence of the N-terminus and control region of the *uvrD* gene

The approximate location of the protein coding region of the *uvrD* gene was determined by Maples and Kushner (22). Their results showed that the coding sequence of the N-terminus of the protein was within a few hundred base pairs of one end of the 2.9 kb *PvuII* restriction fragment which contains the *uvrD* gene.

To find the precise location of the start of the protein coding sequence and of the control region of the *uvrD* gene, the nucleotide sequence of 1190 base pairs of DNA at the end of the *PvuII* fragment was determined. The

```

-416          -400          .          .          .          .          -350          .
TGGCCGCTGGTGGCGATCACCAACGGTAACGCCAGCCGGAGCTGTTGGTTGGGGGATTATTTGAGTTTGTGCTGCG

          -300
CGCTGGCCCGCACGGGCGCTCAAAACGGTCAGCGATATGTACTTTTGGCTGCGGAAAACTCAACGTGCGCATCGGCG

-250          .          .          .          .          -200          .
AGATCTTACATCTTGGGACGATCTCACCACCTGACGTGGGTGGGCAATTCGACGCGAATGCAGCGTTGTTGGATCAGA

          -150          .          .          .          .          -100          .
CCGGAAAATGGCGATCTGATGCAAACTGGGACAGCCGTTTACTGCCGCATCTGGAAATTTCCCGCTTGGCATCTCGAC

CTCGCTGATATAATCAGCAAACTCTATATATACCCACCTTTTGGCGGAGGGCGTTGCCCTTCTCGCCCAACCTATTT

          -50
CTCGCTGATATAATCAGCAAACTCTATATATACCCACCTTTTGGCGGAGGGCGTTGCCCTTCTCGCCCAACCTATTT

-1+1
TACCGGCGCTGCCA ATG GAC GTT TCT TAC CTG CTC GAC AGC CTT AAT GAC AAA CAG CGC GAA
met asp val ser tyr leu leu asp ser leu asn asp lys gln arg glu

+50          .          .          .          .          .          .          +100
GCG GTG GCC GGG CCA CGC AGC AAC CTT CTG GTG CTG GCG GGC GCG GGC AGT GGT AAG ACG
ala val ala ala pro arg ser asn leu leu val leu ala gly ala gly ser gly lys thr

          .          .          .          .          .          .          +150
GCG GTA CTG GTG CAT CGT ATC GCC TGG TTG ATG AGC GTG GAA AAC TGC TCG CCA TAC TCG
arg val leu val his arg ile ala trp leu met ser val glu asn cys ser pro tyr ser

          .          .          .          .          .          .          +200
ATT ATG GCG GTG ACG TTT ACC AAC AAA GCG GCG GCG GAG ATG CGT CAT CGT ATC GGC CAA
ile met ala val thr phe thr asn lys ala ala ala glu met arg his arg ile gly gln

          .          .          .          .          .          .          +250
CTG ATG GGC AGC AGC CAG GGC GGT ATG TGG GTC GGC ACC TTC CAC GGG CTG GCG CAC CGT
leu met gly thr ser gln gly gly met trp val gly thr phe his gly leu ala his arg

          .          .          .          .          .          .          +300
TTG CTG CGT GCG CAC CAT ATG GAC GCC AAT CTG CCG CAG GAT TTC CAG ATC CTC GAC AGT
leu leu arg ala his his met asp ala asn leu pro gln asp phe gln ile leu asp ser

          .          .          .          .          .          .          +350          +400
GAA GAC CAG CTA CGC CTG CTT AAG CGT CTG ATC AAA GCC ATG AAC CTC GAC GAG AAG CAG
glu asp gln leu arg leu leu lys arg leu ile lys ala met asn leu asp glu lys gln

          .          .          .          .          .          .          +450
TGG CCG CCG CGG CAG GCA ATG TGG TAC ATC AAC AGC CAG AAA GAT GAA GGC CTG CGT CCG
trp pro pro arg gln ala met trp tyr ile asn ser gln lys asp glu gly leu arg pro

          .          .          .          .          .          .          +500
CAT CAT ATT CAA AGC TAC GGT AAT CCG GTG GAG CAG ACC TGG CAG AAG GTG TAT CAG GCG
his his ile gln ser tyr gly asn pro val glu gln thr trp gln lys val tyr gln ala

          .          .          .          .          .          .          +550
TAT CAG GAA GCG TGT GAC CGC GCG GGC CTG GTG GAC TTC GCC GAG CTG CTG CTG CGC GCT
tyr gln glu ala cys asp arg ala gly leu val asp phe ala glu leu leu arg ala

          .          .          .          .          .          .          +600
CAC GAG TTG TGG CTT AAC AAG CCG CAT ATC CTG CAA CAC TAC CGC GAA CGT TTT ACC AAT
his glu leu trp leu asn lys pro his ile leu gln his tyr arg glu arg phe thr asn

          .          .          .          .          .          .          +650          +700
ATC CTG GTG GAC GAA TTC CAG GAT ACC AAC AAC ATT CAG TAC GCG TGG ATC CGC CTG CTG
ile leu val asp glu phe gln asp thr asn asn ile gln tyr ala trp ile arg leu leu

          .          .          .          .          .          .          +750
GCG GCG GAC ACC GGC AAA GTG ATG ATC GTC GGT GAT GAC GAC GAC TCA ATC TAC GGC TGG
ala gly asp thr gly lys val met ile val gly asp asp asp gln ser ile tyr gly trp

          .          .          .          .          .          .          +774
CGC GCG
arg gly

```

Figure 2. Sequence of 1190 nucleotides which includes the control region and N-terminal protein coding sequence of the *uvrD* gene. The boxed sequences indicate the -35 region of promoter 1 (at coordinate -110), the -10 region of promoter 1 (at -88), the SOS box (at -74) and the two possible ribosome binding sites (at -14 and -11).

Figure 3. Comparison of the nucleotide sequences of the control regions of the *uvrD* and the *lexA* genes. Stars indicate locations of identical bases in this alignment. The designations -35₁ and -10₁ refer to sequences which define the *uvrD* promoter 1. The location of the second promoter-like sequence is labeled -35₂ and -10₂. SD₁ and SD₂ indicate the two possible ribosome binding sites² for *uvrD* translation initiation. The term SOS refers to the *lexA* protein binding sites. The *lexA* promoter is located at positions -35 and -10.

sequencing scheme and detailed restriction map for this region are shown in Figure 1. Over 80% of the sequence was determined on both strands. The nucleotide sequence is presented in Figure 2.

An examination of the nucleotide sequence revealed only one long "open" reading frame which is the presumptive uvrD coding region. This open reading frame begins with a methionine codon, ATG, at coordinate +1, and continues to the end of the sequence determined. It includes 258 codons and should encode 29 kilodaltons (kd) of the the full size (75 kd) uvrD protein. The codon utilization for this portion of the gene is similar to that of other E. coli genes which are translated at a high level (35). Within the same sequence there are eleven and eight nonsense codons in the other two possible reading frames. It is assumed that this open reading frame corresponds to the coding region of the uvrD gene, since the ATG codon at +1 is the first one downstream from the promoter described below. There is, however, no amino acid sequence data to confirm this assignment.

Identification of the *uvrD* Control Region

The transcription and translation of genetic information in E. coli requires the presence of appropriately spaced -10 and -35 sequences for the

initiation of transcription by RNA polymerase (36) and a ribosome binding site for association of the mRNA to the ribosome (37). Visual inspection of the nucleotide sequence 5' to the ATG which is the presumed start codon indicated the presence of two potential ribosome binding sites, two possible promoters and one presumptive lexA binding site.

Ribosome Binding Sites. The sequence 5' CGGCGGT 3' which begins eleven base pairs upstream (nucleotide -11, Fig. 2 and Fig. 3) from the presumptive ATG initiator codon of the uvrD gene contains in five of the seven positions nucleotides which would base pair with the 3' end of the 16S rRNA (3'AUUCCUCCA5') (37). However, the spacing between this sequence and the ATG initiator codon is only four base pairs, a distance which is rarely found in E. coli genes (38). It is possible that the sequence 5' ACGCGGC 3' starting at coordinate -14 which contains four out of seven bases which are complementary with the 16S rRNA is the actual ribosome binding site. This sequence is separated from the ATG codon by seven base pairs, the average separation among E. coli genes (38). It should also be noted that the protein coding region of uvrD could begin with the GTG codons at either nucleotide -6 or +118 (39).

Promoter Sequences. A sequence which resembles many other known E. coli promoters, which we call promoter 1, begins at the sequence TTGGCA at nucleotide -110 (Fig. 2 and 3) and is similar to that found in many promoters at the so called -35 position (36). Additionally, the sequence which begins at nucleotide -88, TATAAT, is the same as that of the canonical -10 sequence (Pribnow box, 36). The separation between the -35 and -10 sequences is 16 base pairs, a spacing found in many promoters (36). A second possible promoter (designated promoter 2) (Fig. 3) is determined by a -35 sequence, TTGCGC, starting at nucleotide -41 and a -10 sequence, TATTTT, located at nucleotide -21. The functionality of this promoter is suspect, however, because a 14 nucleotide spacer has not been observed with any known E. coli promoter (36).

lexA Binding Sites. Binding sites for the lexA protein have been shown to consist of the sequences CTG ... CAG separated by ten base pairs which are normally rich in A and T (1). A sequence with these characteristics begins at nucleotide -74, nine base pairs downstream from the Pribnow box of promoter 1 (Fig. 3). This possible SOS box, CTGTATATATACCCAG, differs from one of the two known binding sites upstream from the lexA gene, CTGTATATACACCCAG (15,16), at only one position (Fig. 3). A second sequence, CTGATATATACAG, starts at nucleotide -92 and encompasses the Pribnow box of promoter 1. This sequence contains the terminal CTG ... CAG found in SOS boxes but

Figure 4. *In vitro* transcription from the *uvrD* promoter. A. Autoradiogram of denaturing (7M urea) 8% acrylamide gel containing *in vitro* synthesized transcripts. DNA templates containing the *uvrD* promoter were used at a concentration of 20 nM. Lane (1) *RsaI* to *HaeIII* (-295 to +55); Lane (2) *HaeIII* to *HaeIII* (-330 to +55); Lane (3) *RsaI* to *RsaI* (-295 to +113); Lane (4) *RsaI* to *RsaI* with GTP replaced by ITP in the transcription mix; Lane (5) contains single-stranded DNA fragments used as size standards (see Materials and Methods); Lanes (6), (7), and (8) transcripts made from a mixture containing three DNA templates: the *RsaI* to *RsaI* fragment which encodes the *uvrD* promoter at a concentration of 20 nM, a fragment containing the β -lactamase promoter at 5 nM, and a fragment containing the *recA* promoter at 5 nM (see Materials and Methods). The transcripts produced by the β -lactamase

(bla) and recA promoter fragments are indicated. The two RNA species below the 144 base long β -lactamase transcript are also produced from the β -lactamase fragment. The reactions were carried out as described in Materials and Methods except that in lane (7) lexA protein was added to 400 nM ten minutes prior to the addition of RNA polymerase and in lane (8) the lexA protein was added to 400 nM ten minutes after the addition of the RNA polymerase. B. Diagram showing the restriction fragments from the uvrD promoter region used as templates for *in vitro* transcription. The location marked 5' is the start of the transcript. The attenuation site is marked by att and the start of the protein coding sequence by met.

has only a seven nucleotide spacer (Fig. 3).

In vitro Transcription Experiments

Small DNA restriction fragments containing the presumed uvrD promoters were used as templates for *in vitro* transcription with *E. coli* RNA polymerase. When the RsaI fragment containing DNA between nucleotides -295 and +113 was used as a template, transcripts 190 and 58 bases long were produced (Fig. 4A, lane 3). If this same template was cleaved at the HaeIII site at nucleotide +55, the transcripts produced were 136 and 58 bases long (Fig. 4A, lane 1). Templates cleaved at the TaqI site at position +20 produced transcripts 92 and 58 bases long (data not shown). These observations suggest that there is a transcript whose 5' end is at about nucleotide -76 and is read in the direction of the uvrD gene to the ends of each of the restriction fragments used as templates. Since the 5' end of the transcript is several base pairs downstream from the Pribnow box of promoter 1 (Fig. 3) this region must serve as the promoter for the uvrD gene. No runoff transcripts from the second potential promoter (Fig. 3) were observed.

The location of the coding region of the 58 base long RNA was deduced from the following information. This transcript was produced from the RsaI, HaeIII, and TaqI cleaved templates some of which differ from each other at both of their ends (Fig. 4A, lanes 2 and 3). Thus, this RNA is not a runoff transcript from these fragments but it must terminate within them. The 58 base long RNA was not produced from the RsaI template if it had been cleaved with AluI or with CfoI (data not shown), indicating that the transcript is encoded from the region of DNA in which these enzymes cleave (Fig. 4B). Since this region of DNA contains inverted complementary sequences, the RNA may be capable of adopting a secondary structure (Fig. 6) which is involved in the termination event (40). To test this possibility, the transcription reactions were performed with GTP replaced with ITP (see Discussion). This resulted in elimination of production of the 58 base long RNA and in an alteration of the migration of the longer runoff transcript (Fig. 4A, lane 3).

Figure 5. Binding of the *lexA* protein to the *uvrD* promoter region. Panels A and B. Autoradiograms of a denaturing (7M urea) 8% acrylamide gel which contains partially degraded DNA restriction fragments labeled at one of their ends. A fragment labeled at the *TagI* site at coordinate +20 and extending to the *HaeIII* site at position -330 (Panel A) and a fragment labeled at the *Sau3A* site at position -163 and extending to the *HaeIII* site at position +55 (Panel B) were used in these experiments. These fragments were partially degraded with DNase I in the presence of 400 nM *lexA* protein (+*lexA*) and in

the absence of lexA protein (-lexA). Lanes G, A+G, T+C, C contain the chemical degradation products from the fragments treated as in the Maxam and Gilbert procedure (29). The DNase I generated fragments migrate one-half base pair more slowly than do those fragments generated by the chemical cleavage (17). Panel C. The nucleotide sequence of the lexA protected region. The nucleotides protected by the lexA protein are surrounded by brackets. The dotted portion of the brackets indicates ambiguity in the end point of the protected region. The SOS box and -10 region of the uvrD promoter are boxed. The arrows indicate the points of enhanced DNase I cleavage.

As shown in Figure 4A (lanes 6-8) both the 58 base long transcript and the runoff transcript were made in reduced quantities in the presence of purified lexA protein. There was a greater inhibition of transcription when the lexA protein was added prior to the addition of RNA polymerase (Fig. 4A, lanes 7 and 8). Controls for the action of the lexA protein were DNA fragments containing the recA promoter and the β -lactamase promoter. As expected from results reported elsewhere (15,16), transcription from the recA promoter was inhibited by the lexA protein while transcription from the β -lactamase promoter was unaffected.

Experiments designed to determine if uvrD synthesis was autoregulated gave inconclusive results. When purified uvrD protein was added to the transcription reaction after the addition of RNA polymerase, the production of transcripts was reduced whether a uvrD promoter fragment or a β -lactamase promoter fragment was used as a template (data not shown).

Binding of the lexA Protein to the Control Region of the uvrD Gene

The evidence presented in Figure 5 shows that the purified lexA protein binds to the sequence near promoter 1 which is homologous to other known lexA binding sites. The binding of lexA protein protects a region of DNA, which includes the sequence CTGTATATATACCCAG, from degradation by DNase I. There is one position within this sequence on each of the two strands at which the binding of the lexA protein enhances the DNase I cleavage. These points of enhanced cleavage are symmetrically located within the protected region.

The sequence CTGATATAATCAG starting at position -92, contains CTG ... CAG sequences separated by seven base pairs instead of the ten base pairs found in other lexA binding sites. This sequence resembles that of the SOS box 1 in the lexA promoter region, CTGTATATACTCACAG (15,16), (Fig. 3), but was not protected by the lexA protein from degradation by DNase I.

DISCUSSION

The data presented above have permitted the identification of the uvrD promoter and have shown that the lexA protein binds to a sequence downstream from this promoter, repressing transcription of the gene *in vitro*. These observations are consistent with *in vivo* evidence that uvrD expression is increased three to four-fold upon treatment of cells with DNA damaging agents (41,42,43).

In addition, the uvrD regulatory region shows a high degree of homology (>78%) with the similar region of the lexA gene (Fig. 3). It should be pointed out, however, that there are two functional lexA binding sites for the lexA gene (15,16), while the lexA protein only binds to one site in front of the uvrD gene (Fig. 5). The failure of the lexA protein to bind to the CTGATATAATCAG sequence in the uvrD promoter region provides further evidence of the importance of the 10 nucleotide spacer region between the CTG ... CAG sequences.

In a preliminary report from this laboratory (44) it was suggested that promoter 2 was active in producing transcripts. The results presented here, however, do not confirm this observation. Accordingly, we conclude that there is a single promoter (designated promoter 1 in Fig. 3) for the uvrD gene.

Another important finding relates to the apparent existence of transcription attenuation. A 58 base long RNA (Fig. 4) was reproducibly synthesized from small DNA fragments with differing ends that contained the uvrD promoter. This RNA species was not observed when the DNA templates were cleaved within 60 bases of the 5' end of the uvrD transcript. Nor was the RNA seen when lexA protein was added to transcription reactions (Fig. 4). Although other explanations are possible, we hypothesize that this mRNA is initiated at promoter 1 (Fig. 3) and represents a set of messages which terminate at the series of T residues between nucleotides -19 to -15 (Fig. 2, Fig. 3). *In vitro* it appears that approximately one half of the initiation events are terminated prior to reaching the uvrD coding sequence. The length of the RNA predicted from the sequence is 62 to 64 bases while the observed length is only 58 bases. This could result from inaccuracy in estimating the start point of transcription or from discrepancies in migration of RNA and of single-stranded DNA in the gels.

Two possible secondary structures for this RNA are shown in Figure 6. In one case the RNA would contain a GC rich stem and a loop of 12 bases with the stem being followed by five U residues. This structure is very similar

Figure 6. Hypothetical secondary structures of the small RNA produced from the *uvrD* promoter. The ΔG of formation of structure (a) is about -16 kilocalories and that of structure (b) is about -20 kilocalories according to the rules of Tinoco *et al.* (45). The RNA as drawn here is 62 to 64 bases long which is a few nucleotides longer than expected from the size (58 bases) of the transcript as determined from the gels of *in vitro* synthesized RNAs. The nucleotide at which the transcript is initiated is not precisely known.

to that of the RNA for many rho-independent transcription terminators (40). A second possible structure has a longer stem with the loop reduced to three nucleotides (Fig. 6). Since the strength of stem structures is dependent primarily on GC base pairing (45), the replacement of GMP by IMP in the RNA will weaken such structures because an IC base pair has only two hydrogen bonds. The use of ITP in place of GTP in *in vitro* transcription systems has been shown by others to reduce rho-independent transcription termination (46,47). Therefore the failure to see the 58 base RNA in the presence of ITP (Fig. 4A, lane 4) can be taken as an indication that the formation of the GC rich stem shown in Figure 6 is important in the *in vitro* transcription termination event.

Based on these results, we hypothesize that some level of transcription termination occurs *in vivo* and that the presence of this small RNA indicates that *uvrD* gene expression is regulated, in part, by transcription attenuation. A possible means of controlling transcription attenuation at the beginning of the *uvrD* gene is the presence of a ribosome on the incipient message. While there are no commonly used translation initiation or ter-

mination codons within the 58 base RNA, there is within the stem a sequence, GGAGG, which could base pair with the 3' end of the 16S rRNA (Fig. 6). If the ribosome does bind to this sequence, it would prevent the stem formation and thereby permit continued transcription. A control mechanism similar to this has been proposed for the ampC gene of E. coli (48).

With the recent identification of uvrD as the structural gene for DNA helicase II (Maples and Kushner, manuscript in preparation, 49,50), the finding of two types of transcriptional control suggest a complex role for this protein in the cell. While DNA helicase II has clearly been implicated in the repair of UV induced DNA lesions (21), the finding that rep uvrD double mutants cannot be constructed (51, Shepherd, Hamilton and Kushner, unpublished results) suggests some essential role for the protein in DNA replication. Specifically, it has been proposed that both the rep helicase and DNA helicase II are required at the replication fork (52,53). Additionally, Kuhn et al. (53) have shown that DNA helicase II is required in stoichiometric amounts for DNA unwinding to take place. If such a dual role for the protein is correct, the existence of multiple forms of transcription regulation makes more sense.

Accordingly, the following model for the regulation of uvrD synthesis is proposed. In exponentially growing bacteria the demand for DNA helicase II can be accommodated by increased transcription of the gene. However, since rapid growth has not been shown to trigger the SOS response, the increased demand for uvrD protein could be met by relieving transcription attenuation. Since the number of ribosomes increases in rapidly growing cells (54), this would be accomplished by transient binding of excess ribosomes to the GGAGG sequence on the proposed stem of the 58 base RNA (Fig. 6).

When DNA synthesis is stopped either by UV light, DNA crosslinking or the use of temperature sensitive DNA synthesis mutations, already synthesized DNA helicase II protein will presumably remain at the growth fork. New protein synthesis may be required for DNA helicase II to carry out its DNA repair functions. Since the level of ribosomes under conditions of DNA synthesis cessation is probably reduced, transcription attenuation may increase. In order to circumvent this problem, a second level of transcription regulation is invoked. The lexA repressor protein is cleaved permitting full utilization of the promoter by RNA polymerase. The presence of attenuation could explain why only a 3 to 4-fold induction of the uvrD gene has been observed (41,43). Even though this model is highly speculative, it does provide the framework for experiments which can test its validity.

ACKNOWLEDGEMENTS

The authors wish to thank Valerie Maples for many helpful discussions and Kimberly Hearn and Colleen McElfresh for preparation of the manuscript. This work was supported in part by grants from NIH to S.R.K. (GM27797 and GM28760) and A.M.E. (GM08457).

REFERENCES

1. Little, J.W. and Mount, D.W. (1982) *Cell* 29, 11-22.
2. Kenyon, C.J. and Walker, G.C. (1981) *Nature* 289, 808-810.
3. Fogliano, M. and Schendel, P.F. (1981) *Nature* 289, 196-198.
4. Huisman, O. and D'Ari, R. (1981) *Nature* 290, 797-799.
5. Miller, H.I., Kirk, M. and Echols, H. (1981) *Proc. Natl. Acad. Sci. USA* 78, 6754-6758.
6. Bagg, A., Kenyon, C.J. and Walker, G.C. (1981) *Proc. Natl. Acad. Sci. USA* 78, 5749-5753.
7. Hull, R.A. (1975) *J. Bacteriol.* 123, 775-776.
8. van den Elzen, P.J.M., Maat, J., Walters, H.H.B., Veltkamp, E., and Nijkamp, H.J.J. (1982) *Nucl. Acids Res.* 10, 1913-1928.
9. Kenyon, C.J. and Walker, G.C. (1980) *Proc. Natl. Acad. Sci. USA* 77, 2819-2823.
10. McPartland, A., Green, L. and Echols, H. (1980) *Cell* 20, 731-737.
11. Little, J.W. and Harper, J.E. (1979) *Proc. Natl. Acad. Sci. USA* 76, 6147-6151.
12. Brent, R. and Ptashne, M. (1980) *Proc. Natl. Acad. Sci. USA* 77, 1932-1936.
13. Sancar, A., Sancar, G.B., Rupp, W.D., Little, J.W. and Mount, D.W. (1982) *Nature* 298, 97-98.
14. Sancar, G.B., Sancar A., Little, J.W. and Rupp, W.D. (1982) *Cell* 28, 523-530.
15. Little, J.W., Mount, D.W. and Yanisch-Perron, C.R. (1981) *Proc. Natl. Acad. Sci. USA* 78, 4199-4203.
16. Brent, R. and Ptashne, M. (1981) *Proc. Natl. Acad. Sci. USA* 78, 4204-4208.
17. Cole, S.T. (1983) *Mol. Gen. Genet.* 189, 400-404.
18. Lupski, J.R., Smiley, B.L. and Godson, G.N. (1983) *Mol. Gen. Genet.* 189, 48-57.
19. Smirnov, G.B., Filkova, E.V. and Skavronskaya, A.G. (1972) *Mol. Gen. Genet.* 118, 51-56.
20. Lu, A.-L., Clark, S. and Modrich, P. (1983) *Proc. Natl. Acad. Sci. USA* 80, 4639-4643.
21. Ogawa, H., Shimada, K. and Tomizawa, J-I. (1968) *Mol. Gen. Genet.* 101, 227-244.
22. Maples, V.F. and Kushner, S.R. (1982) *Proc. Natl. Acad. Sci. USA* 79, 5616-5620.
23. Oeda, K., Horiuchi, T. and Sekiguchi, M. (1981) *Mol. Gen. Genet.* 184, 191-199.
24. Arthur, H.M., Bramhill, D., Eastlake, P.B. and Emmerson, P.T. (1982) *Gene* 19, 285-295.
25. Bolivar, F., Rodriguez, R.L., Green, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crosa, J.H. and Falkow, S. (1977) *Gene* 2, 95-113.
26. Little, J.W. (1979) *Mol. Gen. Genet.* 177, 13-22.
27. Kushner, S.R. (1978) in Genetic Engineering, eds. Boyer, H.W. and Nicosia, S. (Elsevier/North Holland, Amsterdam), pp. 17-23.
28. Ish-Horowicz, D. and Burke, J.F. (1981) *Nucleic Acids Res.* 9, 2989-2998.

29. Maxam, A.M. and Gilbert, W. (1980) *Methods in Enzymology* 65, 499-560.
30. Krayev, A.S., Kramerov, D.A., Skryabin, K.G., Ryskov, A.P., Bayev, A.A. and Goergiev, G.P. (1980) *Nucl. Acids Res.* 8, 1201-1215.
31. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning, a Laboratory Manual* (Cold Spring Harbor Laboratory), p. 181.
32. Johnson, R.C. and Reznikoff, W.S. (1981) *Nucl. Acids Res.* 9, 1873-1883.
33. Brosius, J., Cate, R.L. and Perlmutter, A.P. (1982) *J. Biol. Chem.* 257, 9205-9210.
34. Sancar, A., Stachelek, C., Konigsberg, W. and Rupp, W.D. (1980). *Proc. Natl. Acad. Sci. USA* 77, 2611-2615.
35. Konigsberg, W. and Godson, G.N. (1983) *Proc. Natl. Acad. Sci. USA* 80, 687-691.
36. Hawley, D.K. and McClure, W.R. (1983) *Nucl. Acids Res.* 11, 2237-2255.
37. Shine, J. and Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* 71, 1342-1346.
38. Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B.S. and Stormo, G. (1981) *Ann. Rev. Microbiol.* 35, 365-403.
39. Kozak, M. (1983) *Microbiol. Rev.* 47, 1-45.
40. Farnham, P.J. and Platt, T. (1981) *Nucl. Acids Res.* 9, 563-577.
41. Siegel, E. (1982) XIII International Congress of Microbiology, P28:1, p. 100.
42. Pang, P.P. and Walker, G.C. (1983) *J. Bacteriol.* 154, 1502-1504.
43. Kumura, K., Oeda, K., Akyama, M., Horiuchi, T. and Sekiguchi (1983) in *Cellular Responses to DNA Damage*, UCLA Symposia on Molecular and Cellular Biology, New Series Vol XI (eds. Friedberg, E.C. and Bridges, B.R.) Alan R. Liss, Inc. New York, in press.
44. Kushner, S.R., Maples, V.F., Easton, A., Farrance, I. and Peramachi, P. (1983) in *Cellular Responses to DNA Damage*, UCLA Symposia on Molecular and Cellular Biology, New Series Vol XI (eds. Friedberg, E.C. and Bridges B.R.) Alan R. Liss, Inc. New York, in press.
45. Tinoco, I., Jun., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M., and Gralla, J. (1973) *Nature New Biology* 246, 40-41.
46. Lee, F. and Yanofsky, C. (1977) *Proc. Natl. Acad. Sci. USA* 74, 4365-4369.
47. Turnbough, C.L. Jr., Hicks, K.L. and Donahue, J.P. (1983) *Proc. Natl. Acad. Sci. USA* 80, 368-372.
48. Jaurin, B., Grundstrom, T., Edlund, T. and Normark, S. (1981) *Nature* 290, 221-225.
49. Oeda, K., Horiuchi, T. and Sekiguchi, M. (1982) *Nature* 298, 98-100.
50. Hickson, I.D., Arthur, H.M., Bramhill, D. and Emmerson, P.T. (1983) *Mol. Gen. Genet.* 190, 265-270.
51. Taucher-Scholz, G., Abdel-Monem, M. and Hoffmann-Berling, H. (1983) in *Cellular Responses to DNA Damage*, UCLA Symposia on Molecular and Cellular Biology, New Series Vol XI (eds. Friedberg, E.C. and Bridges B.R.) Alan R. Liss, Inc. New York, in press.
52. Klinkert, M-Q., Klein, A. and Abdel-Monem, M. (1980) *J. Biol. Chem.* 255, 9746-9752.
53. Kuhn, B., Abdel-Monem, M., Krell, H. and Hoffman-Berling, H. (1979) *J. Biol. Chem.* 254, 11343-11350.
54. Maaloe, O. (1979) in *Biological Regulation and Development* Vol 1. (ed. Goldberger, R.F.) (Plenum, New York) pp. 487-542.