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

Eralp Arikan, Mahmooda S.Kulkarni¹, David C.Thomas² and Aziz Sancar^{2*}

Dicle University School of Medicine, Department of Microbiology, Diyarbakir, Turkey, ¹University of North Carolina School of Medicine, Department of Pathology, Chapel Hill and ²University of North Carolina School of Medicine, Department of Biochemistry 231H, Chapel Hill, NC 27514, USA

Received 15 January 1986; Accepted 17 February 1986

ABSTRACT

The UvrB protein is one of the three subunits of the *E. coli* ABC exci-nuclease. We have reported the sequences of the other two subunits, the UvrA and UvrC proteins. In this paper the sequence of the UvrB protein is presented. The protein sequence was determined from the DNA sequence of the *uvrB* gene and was confirmed by sequencing the NH₂-terminus of the UvrB protein and analyzing its overall amino acid composition. The coding region of *uvrB* is 2019 basepairs, specifying a protein of 672 amino acids and M_r of 76,118. The sequence of the UvrB protein shows a moderate level of homology to that of the UvrC protein and to the ATP binding site of the UvrA protein. During purification of UvrB protein a proteolytic product, UvrB*, is produced in high quantities. We find that UvrB* results from removal of about 40 amino acids from the COOH-terminus of the UvrB protein. The *uvrB* gene has complex regulatory features. On the 5' side, the coding region is preceded by 3 promoters, a DnaA box and an SOS box. On the 3' side the gene is followed by an REP (Repetitive Extragenic Palindrome) sequence which has been implicated in gene regulation by an unknown mechanism.

INTRODUCTION

ABC excinuclease of *Escherichia coli* is a DNA repair enzyme which removes modified bases from DNA in the form of a 12-13 base-long oligonucleotide by incising the damaged strand at the 8th phosphodiester bond 5' and the 4th or 5th phosphodiester bond 3' to the damaged nucleotide (1). The enzyme is made up of three subunits, UvrA (103,866), UvrB (76,118) and UvrC (66,038) proteins. We previously reported the sequences of the regulatory regions of all three genes and the entire sequences of the UvrA and UvrC proteins as predicted from the DNA sequences (2-5). In this communication we present the entire sequence of the *uvrB* gene and the amino acid sequence predicted from the DNA sequence. The amino acid sequence of the UvrB protein was confirmed by determining the amino acid composition as well as the sequences of the NH₂-termini of the purified protein and that of its proteolytic product, UvrB* (6).

The *uvrB* gene is preceded by three promoters as reported previously (3), has a coding region of 2019 base pairs encoding a protein of 672 amino acids

and $M_r = 76,118$. The sequence of the UvrB protein shows several regions of homology to that of the UvrC protein and a region homologous to that of the UvrA protein's presumptive ATP binding site. The coding region of *uvrB* is followed by a 25 bp hyphenated palindrome located 37 bp 3' to the termination codon. The palindrome, which probably acts as a transcriptional termination signal, has the consensus sequence of the REP (repetitive extragenic palindromic) sequences and therefore may act in regulating *uvrB* expression by an additional mechanism (see Ref. 7 for a review) that encompasses at least 25 other *E. coli* genes, including *uvrD*.

MATERIALS AND METHODS

Bacterial Strains and Plasmids. *E. coli* K12 strains DR153 (*recA⁻ uvrB⁻*) and DH1 (*recA⁻ hsdR⁻*) were used to maintain and isolate the *uvrB* plasmid, pUNC20, as the source of the *uvrB* gene in this study. pUNC20 was constructed as follows: pDR1494, which contains a 5.7 kbp chromosomal fragment carrying the *uvrB* gene (8) was first digested with Bgl I, and then with the Bal31 exonuclease to remove approximately 100 bp from the termini of the Bgl I fragments. Following the removal of Bal31 nuclease by phenol extraction Bam HI linkers were ligated to the now blunt-ended Bgl I fragments and the 3.4 kbp fragment carrying the *uvrB* gene was purified by polyacrylamide gel electrophoresis followed by electroelution. This fragment was inserted into the Bam HI site of pBR322 to obtain pUNC20 (*uvrB⁺ bla⁺ tet⁻*). This plasmid complemented the *uvrB* deletion mutation of DR153 (8) and after the initial isolation in that strain it was propagated in DH1 (9). Plasmid DNA from this latter strain was isolated by Sarkosyl lysis and CsCl-ethidium bromide density centrifugation.

DNA Sequencing. The 3.4 kbp Bam HI fragment of pUNC20 was isolated and mapped with the restriction enzymes Bgl II, Bst NI, Hae III, Hinf I, Pvu II and Rsa I. The restriction fragments were terminally labeled at 5' ends using polynucleotide kinase and sequenced by the method of Maxam and Gilbert (10).

Sequence Analysis. The amino acid sequence of the UvrB protein was compared to those of the UvrA and UvrC protein using the ALIGN program from the Protein Identification Resource of the National Biomedical Research Foundation (11). The sequence of the UvrB protein was also compared to the protein sequences in the National Biomedical Research Foundation Protein Sequence Library using the FASTP program of Lipman and Pearson (12).

Protein Chemistry. The UvrB protein was purified as described previously (6). The UvrB* polypeptide was purified from the side fraction of the F2 step of the UvrB purification procedure (6). The details of the purification of

Fig. 1. Restriction Map of the *uvrB* plasmid pUNC20 used in this study. The vector sequence is speckled. The location and direction of transcription of the *uvrB* gene is indicated by the heavy arrow.

UvrB* and of its properties will be published elsewhere (M. Levy and A. Sancar, in preparation). The sequences of the NH_2 -termini of the *UvrB* and *UvrB** proteins were determined by Dr. C. Noyes of the University of North Carolina using a Beckman 890 C Sequencer. The amino acid composition was obtained from analysis of acid hydrolysates on a Beckman 121M amino acid analyzer. Cysteine and methionine were quantitated as cysteic acid and methionine sulfone from performic acid oxidized protein.

RESULTS

Sequencing of the *uvrB* Gene. The *uvrB* gene was previously identified in the Clarke-Carbon *E. coli* gene library (13) and subcloned into pBR322 on a 5.6 kbp-long Pst I fragment (8). Further restriction mapping of this fragment showed that the gene was surrounded on either side by Bgl I sites outside of the *uvrB* gene. This fragment was digested with Bgl I , joined to Bam HI linkers and inserted into pBR322 to obtain pUNC20 (Figure 1). To sequence the gene the Bam HI fragment carrying *uvrB* was purified and mapped with restriction enzymes. Terminally labeled fragments were purified and sequenced by the method of Maxam and Gilbert (10) as outlined in Figure 2. About 90% of the gene was sequenced in both strands and all the restriction sites used for terminal labeling within the gene were crossed by a sequence obtained from a fragment generated by a different restriction enzyme.

Sequence of the *uvrB* Gene. The promoter-operator region of *uvrB* was sequenced and the transcriptional and translational start points were determined

Fig. 2. Sequencing strategy for the *uvrB* gene. In this figure only the cutting sites of the restriction enzymes whose sites were mapped throughout the 3.4 kbp chromosomal fragment are shown. In addition to these enzymes *Hin* P1, *Fnu* 4HI, *Nci* I, *Sau* 3A and *Ava* II were used for generating fragments for terminal labeling and/or for cutting of terminally labeled fragments. The arrows indicate the direction and length of sequences obtained from the corresponding restriction sites.

previously (3). This previously published sequence as well as the rest of the *uvrB* gene sequence obtained in this work is shown in Figure 3. The gene is preceded by three promoters: P1, P2 and P3. Transcription from P3 terminates within the P2 sequence *in vitro* whereas transcriptions from P1 and P2 continue into the coding region. P3 overlaps with a DnaA box (14) and P2 with an SOS box (3). Transcription from P2 is inhibited by LexA protein *in vitro* (3) and *in vivo* (15) while transcription from P1 is not inhibited by LexA protein *in vitro* (3) but may be *in vivo* (15). The translational start point is at +34 (with regard to the P1 transcriptional start point). The open reading frame starting at this point continues for 673 codons terminating at a TAA at bp +2053. The protein encoded by this open reading frame has a calculated $M_r = 76,118$ (excluding the first Met), in reasonable agreement with the M_r reported for the UvrB protein based on its migration in sodium dodecylsulfate polyacrylamide gels (8). The open reading frame is followed, 37 bp 3' to the termination codon, by a 25 bp hyphenated palindrome which presumably acts as a transcriptional terminator (16). It is interesting to note that this potential termination signal matches the consensus sequence for REP (repetitive extragenic palindrome) that has been found outside the coding regions of about 25 other

Fig. 3. Sequences of the *uvrB* gene and protein. The -10 and -35 sequences of the three promoters (P1, P2, P3), the three transcriptional start sites (S1, S2, S3) and the LexA binding site ("SOS box") as well as the DnaA binding site ("DnaA box") are indicated. The ribosome binding site is marked by a double underline and the REP sequence by the two opposing arrows. The potential cleavage site of the protein that results in production of UvrB* is indicated by the broken arrow.

E. coli genes including *uvrD* (7, 17-19). This sequence is presumed to be involved in gene regulation by some novel mechanism (17). As this sequence, with the exception of stable RNA genes, is found only in extragenic regions of

the *E. coli* chromosome, we conclude that the *uvrB* gene constitutes a single-gene transcriptional unit (operon).

Protein Chemistry of UvrB Protein. To confirm the amino acid sequence of the UvrB protein predicted from the DNA sequence we analyzed the UvrB protein purified to homogeneity (1, 6). Amino acid sequencing yielded NH₂-Ser-Lys-Pro-Phe-Lys-Leu-Asx as the first 7 N-terminal amino acids indicating that the N-formylmethionine is removed from the NH₂-terminus of the mature protein. This sequence is in agreement with the amino acid sequence predicted from the nucleotide sequence in Figure 3 and is the same as the sequence we reported previously (3), even though in our previous report there was a discrepancy between the DNA sequence and the protein sequence due to a typographical error in recording the DNA sequence (see 20). Further confirmation of the UvrB protein sequence came from the analysis of the amino acid composition of the purified protein. The result obtained is compared to the predicted composition in Table 1. There is a good agreement between the compositions suggesting that the amino acid sequence shown in Figure 3 is the correct sequence of the UvrB protein. An interesting feature of the UvrB subunit according to the DNA sequence is the absence of Trp in the protein. Accordingly, the protein should have an absorption maximum at 275 nm and not at 280 nm as commonly is the case with an average protein. This was found to be the case as there was good agreement between the experimentally determined relative absorbance values at these two wavelengths and the predicted values (data not shown). Thus, the spectrophotometric analysis lends further support to the UvrB protein sequence predicted from the DNA sequence.

During purification of the UvrB protein from overproducing cells 80-90% of the overproduced protein is cleaved at a specific site immediately following cell lysis to produce a protein (UvrB*) of $M_r \cong 70,000$ (6). The same phenomenon is observed in maxicells as well, indicating that a great level of overproduction is not necessary for this cleavage. However, we have no indication that cleavage occurs at all *in vivo* either in overproducing or wild type cells and therefore we doubt if this cleavage has any physiological significance. Nevertheless the cleavage site might be a hinge point in the polypeptide chain and its localization may be of importance in studying the structure-function relationship of the UvrB protein. We, therefore, determined the NH₂-terminal sequence of UvrB* to localize the cleavage site. The sequence of the first seven N-terminal amino acids were the same as those of intact UvrB protein suggesting that UvrB* resulted from the removal of about 40 amino acids from the COOH-terminus of the protein. It is interesting to note that this cleavage

Table 1. Amino Acid Composition of UvrB Protein

Amino Acid	Predicted From DNA Sequence	Found by Amino Acid Analysis*
Asx	61 (Asp 40 + Asn 21)	59
Thr	34	33
Ser	30	28
Glx	99 (Glu 67 + Gln 32)	92
Pro	34	33
Gly	42	47
Ala	50	52
Val	46	48
Met	19	15
Ile	42	38
Leu	74	74
Tyr	23	23
Phe	23	23
His	10	14
Lys	31	32
Arg	53	61
Cys	1	3
Trp	<u>0</u>	<u>0</u>
TOTAL	672	675

*Based on a molecular weight of 76,118 for the UvrB protein.

results in a great loss of activity with regard to the nucleolytic function of ABC excinuclease.

Codon Usage in UvrB Protein. In *E. coli* it has been found that codons corresponding to rare cognate tRNAs are used rarely in highly expressed genes and more frequently in low copy-usually regulatory-proteins (21, 22). While the suggestion that this type of codon usage may play a role in translational regulation (22,23) has been questioned (see 24), it is generally accepted that there is a correlation between the abundance of a particular protein and the lack of rare codons in its gene. Therefore it was of interest to tabulate the codon usage in the UvrB protein and compare it to those of the other two subunits of ABC excinuclease, the UvrA and UvrC proteins. It has been reported

Table 2. Codon Usage in the UvrA, UvrB and UvrC Proteins

Amino Acid	Codon	E. coli		% Synonym Use		Amino Acid		Codon	E. coli		% Synonym Use	
		UvrA	UvrB	UvrA	UvrB				UvrA	UvrB	UvrA	UvrB
Phe	UUU	44	32	35	41	Ala		GCU	28	9	8	17
Leu	UUC	56	68	65	59			GCC	19	26	18	24
	UUA ^I	6	3	7	6			GCA	23	22	14	9
	UUG ^I	8	7	13	18			GCG	30	43	60	50
	CUU ^I	9	15	13	14	Tyr		UAU	41	52	43	76
	CUC ^I	7	14	11	13			UAC	59	48	57	24
Ile	CUA	2	1	--	5	His		CAU	39	48	50	53
	CUG	69	60	56	44			CAC ^R	61	52	50	47
	AUU	37	50	60	55	Gln		CAA ^R	27	26	22	41
	AUC ^R	62	50	38	42			CAG ^R	73	74	78	59
	AUA ^R	1	--	2	3	Asn		AAU ^R	24	33	29	67
Met	AUG	--	--	--	--			AAC	76	67	71	33
Val	GUU	38	19	22	26	Lys		AAA ^I	77	65	77	79
	GUC	13	18	19	26			AAG ^I	23	35	23	21
	GUA	23	9	11	12	Asp		GAU	51	50	65	71
	GUG	27	54	48	36			GAC	49	50	35	29
	UCU	27	15	10	6	Glu		GAA ^I	73	62	54	75
Ser	UCC ^I	26	26	30	14			GAG ^I	27	38	46	25
	UCA ^I	8	13	13	14	Cys		UGU	42	14	--	57
	UCG ^I	11	19	40	14			UGC	58	86	--	43
	AGU ^I	6	6	7	19	Trp		UGG	--	--	--	--
	AGC ^R	22	21	--	33	Arg		CGU	58	44	45	40
Pro	CCU ^R	9	11	12	9			CGC ^I	35	45	34	38
	CCC ^R	6	2	3	9			CGA ^I	2	5	8	5
	CCA	20	17	23	13			CGG ^I	3	6	9	14
	CCG	65	70	62	70			AGA ^I	1	--	4	--
	ACU	24	2	12	7			AGG ^R	0.25	--	--	2
Thr	ACC ^I	51	55	56	52	Gly		GGU	48	29	26	30
	ACA ^R	6	5	12	14			GGC ^I	41	52	45	50
	ACG ^R	20	38	20	28			GGA ^I	5	4	10	13
								GGG ^I	7	15	19	7

R = rare, I = infrequent codons (superscripts)

that an uninduced *E. coli* cell contains about 25 molecules of UvrA protein (5,25), 150-200 molecules of UvrB protein (8) and 10 molecules of UvrC protein (26). Table 2 shows the codon usage in the three proteins. As we reported previously (4,5) rare codons are used frequently in the UvrC but not in the UvrA protein. The UvrB protein, as expected from its moderately high level does not contain rare codons at high frequency (Table 3). This finding, in turn, supports our earlier conclusion, based on the relative intensities of UvrA and UvrB bands in maxicells that the *uvrB* protein is relatively abundant in an uninduced *E. coli* cell (8). It is of interest to note that the AGC codon for Ser which occurs at a frequency of about 20-30 percent in all reported *E. coli* proteins including UvrA and UvrC, is not utilized in the UvrB protein.

Sequence Homology Between the Uvr Proteins. We compared the sequence of the UvrB protein to those of the UvrA (5) and UvrC (4) proteins. The UvrA and UvrB proteins do not share significant sequence homology. However, the UvrB protein has a region (residues 39-53) that has the consensus Walker "A type" ATP binding sequence G-X₄-GKT-X₆-I (27). Two such sequences have been found in the UvrA protein (5) and in Figure 4 one of these sequences is compared to that found in the UvrB protein. The homology is not more than expected for two ATPases that contain Walker's A type sequence. The interesting fact is that while the UvrA protein is a *bona fide* ATPase the UvrB protein is not, even though it stimulates the ATPase activity of the UvrA protein by a factor of 2.5 in the presence of DNA (6). Whether the UvrB protein has any ATP binding activity remains to be determined.

The UvrB and UvrC proteins contain regions of homology that are scattered throughout both proteins; however, the distribution of the homologous regions is such that the two proteins cannot be aligned. In Figure 4 we present the two regions of both proteins that show the most extensive homology. It is of interest to note that the 14-amino acid stretch that is 12 amino acids from the COOH-terminus of the UvrB protein is homologous (13 out of 14) to a region in the middle of the UvrC protein.

DISCUSSION

The *E. coli* ABC excinuclease of *E. coli* is made up of 3 subunits which probably do not associate with one another in the absence of substrate. When the subunits (UvrA, UvrB and UvrC) are mixed with damaged DNA they assemble sequentially to constitute ABC excinuclease which removes the damaged bases in a 12-13 base-long oligonucleotide. To identify the functional domains of each subunit it is necessary to know the primary structures of the proteins. The

PROTEIN	RESIDUES	SEQUENCE
UvrA	(28-45)	V V T <u>Q</u> L S <u>G S <u>Q</u> <u>Q</u> <u>S</u></u> L A F D T <u>Q</u>
UvrB	(39-53)	<u>G V T</u> - - - <u>G S <u>Q</u> <u>Q</u> <u>Q</u></u> T I A N V <u>Q</u>
UvrB	(318-360)	<u>P G E</u> P P T L F D Y L P A D G L L V D E S H V T I P Q I G C H Y <u>R G D R A R</u> K E T
UvrC	(292-334)	<u>P G E</u> I L L D F N L S D K T L L A D S L S D L A G R K I N V Q T K P <u>R G D R A R</u> Y L K
UvrB	(641-672)	<u>H M Q H A</u> <u>Q N L E F E E A A</u> <u>Q I R D Q</u> L H Q L <u>R</u> E L F I A A S * * *
UvrC	(192-223)	<u>H M T A S</u> <u>Q N L E F E E A A</u> <u>R I R D Q</u> I Q A V <u>R</u> A V T E K Q

Fig. 4. Sequence homologies between the Uvr proteins. The region of the UvrB protein that has homology to Walker's consensus ATP binding site is compared to one of the two potential ATP binding sites in the UvrA protein (5). The circled amino acids are the ones conserved in Walker's consensus sequence. Also shown in the figure are two regions in the UvrB and UvrC proteins that show the most extensive homology. There are other regions of moderate homology whose significance is uncertain. ***, the COOH-terminus of the UvrB protein.

sequences of the UvrA and UvrC proteins were reported previously (4,5) and in this paper we have presented the sequences of the *uvrB* gene and protein.

The *uvrB* gene sequence reported here as well as the promoter-operator sequence reported previously (3) indicate that the *uvrB* gene by itself constitutes a transcriptional unit and is not part of a larger operon. However, as was reported previously (3,28) it is one of the genes of the SOS regulon and is perhaps subject to transcriptional regulation by other regulatory circuits (2). Regarding the regulation of the *uvrB* gene the following points are worthy of consideration. First, the gene is preceded by three promoters, P1, P2 and P3 that start transcription at +1 (33 bp upstream from the initiation codon), -31 and -342 bp, respectively. P1 seems to be transcribed constitutively *in vitro* (3) and partially repressed by the LexA protein *in vivo* (15). P2 overlaps with a SOS box and is inhibited by the LexA repressor strongly both *in vitro* (3) and *in vivo* (15). The transcript of P3 terminates within the SOS box even in the absence of LexA protein *in vitro*, and in the presence of the repressor the transcriptional start from this promoter is inhibited by some unknown mechanism since the LexA protein does not bind to P3 with any measurable affinity. Second, the AATTGTGTCATAATTAA sequence overlapping the Pribnow box of the *uvrA* promoter is homologous to the sequence AATTGTGGGATAA TTAA that overlaps the -35 sequence of P1 and the Pribnow box of P2 of *uvrB* gene suggesting that this sequence may be the recognition site for a protein that regulates these two genes by a mechanism separate from the SOS response (2). Third, a DnaA protein recognition sequence homologous to that found in

oriC overlaps the -35 sequence of P3 (14). That this sequence does function as a DnaA binding site *in vivo* is suggested by the fact that cells containing multicopy plasmids carrying the P3 promoter fail to grow in minimal medium (8) but do grow in a *dnaA*ts background under permissive conditions (29). In addition to these complex regulatory structures found on the 5' side of *uvrB* there is a Repetitive Extragenic Palindrome (REP) located 37 bp downstream from the termination codon. Although the REP sequences are widespread in the *E. coli* chromosome (7,18) their exact function is not known. Nevertheless, it has been found that deletions of REP sequences 3' to the *S. typhimurium hisJ* gene (18) and 3' to the *E. coli gadA* gene (7) decrease the expression of the corresponding upstream genes. We, therefore, propose that the expression of the *uvrB* gene is influenced by the REP sequence as well as the LexA and DnaA proteins.

The coding region of the *uvrB* gene is 2019 basepairs specifying a protein of 672 amino acids and $M_r = 76,118$. The amino acid predicted from the DNA sequence has been confirmed by sequencing the amino terminus of the protein and determining its amino acid composition. These studies also revealed that the initiation codon f-Met is removed from the mature protein as amino acid sequence analysis shows that the UvrB protein starts with a Ser residue, the second amino acid predicted from the DNA sequence. The protein has a somewhat unusual composition in that it lacks Trp, and even though the protein contains about equal numbers of positively and negatively charged (at pH 7.0) residues it is an acidic protein having a $pI \approx 5.0$ (data not shown), perhaps because of differential solvent accessibility of these residues.

The codon usage in the UvrB protein is like that of moderately expressed proteins, in line with our previous reports that there were about 200 molecules of UvrB protein in wild type *E. coli* (8). With regard to codon utilization UvrB is similar to the UvrA protein even though in uninduced cells there are about 10-fold as many UvrB molecules as UvrA molecules. However, DNA damaging treatments that elicit the SOS response produce approximately the same levels of both proteins (30) and this might explain why rare codons are not utilized in UvrA protein. On the other hand rare codons are utilized at high frequency in the UvrC protein whose constitutive level is about 10 molecules per cell (26) and is induced slightly (31) or not at all (32) by SOS treatments.

The UvrB protein sequence obtained in this study was compared to the sequences of the UvrA and UvrC proteins. There was no significant homology between the UvrA and UvrB proteins, but surprisingly the UvrB protein - which has no ATPase activity - was found to contain Walker's consensus "A type

sequence" (27) found in many ATPases including the UvrA protein. At present we have no explanation for the presence of this sequence. The UvrB protein is known to stimulate the ATPase activity of the UvrA protein in the presence of DNA (6). While it is possible that this effect is brought about by a conformational change resulting from binding of the UvrB protein to the UvrA-DNA complex, it is conceivable that the UvrB protein is an ATP binding protein that becomes an ATPase only in the UvrA-UvrB-DNA ternary complex. Further studies are needed to determine whether the UvrB protein binds ATP and whether only the A subunit or both the A and B subunits function as ATPases in the UvrA-UvrB-DNA complex.

Several regions of homology were detected between the UvrB and UvrC proteins. At present it is difficult to speculate on the evolutionary or functional significance of this homology as the distribution of the homologous regions in the two proteins precludes alignment of the two subunits along their full length and the functions of the individual subunits of the ABC excinuclease in the excision reaction are not known at present. It is of interest to note that the region of most extensive homology is that near the C-terminus of the B subunit and in the middle of the C subunit. It is also of importance to point out that the UvrB* protein which is missing this region of homology has lost most but not all of the B activity as measured by the specific incision of UV irradiated DNA (data not shown).

We do not know whether the specific cleavage of the UvrB protein to produce UvrB* has any physiological significance. As we pointed out before we have no evidence that this cleavage occurs *in vivo* (6). Regarding the mechanism of cleavage, we do not think it is self proteolysis as we do not observe the cleavage with the purified UvrB protein. We are more inclined to think that this cleavage is caused by a specific *E. coli* protease. The amino acid sequence of the UvrB protein shown below is homologous to one of the two cleavage sites found in the Ada protein (32, I.A. Teo, personal communication) and cleavage of the UvrB protein at this site would produce a protein approximately the size of UvrB*.

UvrB	Met ₆₂₆ -Ser-Pro-Lys-Ala ₆₃₀
Ada	Met ₁₅₆ -Thr-Pro-Lys-Ala ₁₆₀

Arrows show the cleavage site in the Ada protein and the potential cleavage site in the UvrB protein. Further studies are needed to determine whether this sequence constitutes a "recognition site" for a sequence specific protease.

ACKNOWLEDGEMENT

We thank Dr. C. Noyes (University of North Carolina) who sequenced the N-termini and Dr. K. Williams (Yale University) for obtaining the amino acid composition of UvrB protein. We are grateful to Michael Levy who purified the UvrB* peptide, Drs. I. Husain and B. Van Houten for their help at every phase of this project and Dr. I. A. Teo for communicating unpublished results. This work was supported by NIH Grant GM32833 and NSF Grant PCM51212 and partly by grants from The Miller High Life Foundation and R. J. Reynolds Tobacco Co.

*To whom reprint requests should be sent

REFERENCES

1. Sancar, A. and Rupp, W.D. (1983) *Cell* **33**, 249-260.
2. Sancar, A., Sancar, G.B., Little, J.W., Mount, D.W. and Rupp, W.D. (1982) *Nature* **298**, 96-98.
3. Sancar, G.B., Sancar, A., Little, J.W. and Rupp, W.D. (1982) *Cell* **28**, 523-530.
4. Sancar, G.B., Sancar, A. and Rupp, W.D. (1984) *Nucl. Acids Res.* **12**, 4593-4608.
5. Husain, I., Van Houten, B., Thomas, D.C. and Sancar, A. (1986) *J. Biol. Chem.* *in press*.
6. Thomas, D.C., Levy, M. and Sancar, A. (1985) *J. Biol. Chem.* **260**, 9875-9883.
7. Becerril, B., Valle, F., Merino, E., Riba, L. and Bolivar, F. (1985) *Gene* **37**, 53-62.
8. Sancar, A., Clarke, N., Griswold, J., Kennedy, W. and Rupp, W.D. (1981) *J. Molec. Biol.* **148**, 63-76.
9. Hanahan, D. (1983) *J. Mol. Biol.* **166**, 557-580.
10. Maxam, A.M. and Gilbert, W. (1980). *Meth. Enzymol.* **65**, 499-560.
11. Dayhoff, M.O. (1979) Vol. 5, Supplement 3, pp. 1-8 and 353-358, National Biomedical Research Foundation, Washington, D.C.
12. Lipman, D.J., and Pearson, W.R. (1985) *Science* **227**, 1435-1441.
13. Clarke, L. and Carbon, J. (1977) *Cell* **9**, 91-99.
14. Fuller, R.S., Funnell, B.E. and Kornberg, A. (1984) *Cell* **38**, 889-900.
15. Van den Berg, E.A., Geerse, R.H., Pannekoek, H. and Van de Putte, P. (1983) *Nucleic Acids Res.* **11**, 4355-4363.
16. Rosenberg, M. and Court (1979) *Ann. Rev. Genet.* **13**, 319-353.
17. Higgins, C.F., Ames, G.F.L., Barnes, W.M., Clement, J.M. and Hofnung, M. (1982) *Nature* **298**, 760-762.
18. Gilson, E., Clement, J.M., Brutlag, D. and Hofnung, M. (1984) *EMBO J.* **3**, 1417-1421.
19. Stern, M.J., Ames, G.F.L., Smith, N.H., Robinson, E.C. and Higgins, C.F. (1984) *Cell* **37**, 1015-1026.
20. Friedberg, E.C. (1985) *DNA Repair* (W. H. Freeman and Co., New York, New York).
21. Ikemura, T. (1981) *J. Mol. Biol.* **151**, 389-409.
22. Konigsberg, W. and Godson, G.N. (1983) *Natl. Acad. Sci. USA* **80**, 687-691.
23. Robinson, M., Lilley, R., Little, S., Emtage, J.S., Yarranton, G., Stephens, P., Millican, A., Eaton, M. and Humphreys, G. (1984) *Nucl. Acids Res.* **12**, 6663-6671.
24. Bonekamp, F., Andersen, H.D., Christensen, T. and Jensen, K.F. (1985) *Nucleic Acids Res.* **13**, 4113-4123.
25. Sancar, A., Wharton, R.P., Seltzer, S., Kacinski B., Clarke, N., and Rupp, W.D. (1981) *J. Molec. Biol.* **148**, 45-62.
26. Yoakum, G.H. and Grossman, L. (1981) *Nature* **292**, 171-173.

27. Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982) *EMBO J.* 1, 945-951.
28. Van den Berg, E.A., Zwetsloot, J., Noordermeer, I., Pannekoek, H., Dekker, B., Dijkema, K. and Van Ormondt, H. (1981) *Nucleic Acids Res.* 9, 5623-5643.
29. Van den Berg, E.A., Geerse, R.H., Memlink, J., Bovenberg, R.A.L., Magnée, F.A. and Van de Putte, P. (1985) *Nucleic Acids Res.* 13, 1829-1840.
30. Markham, B.E., Harper, J.E., Mount, D.W., Sancar, G.B., Sancar, A., Rupp, W.D., Kenyon, C.J. and Walker, G.C. (1984) *J. Mol. Biol.* 178, 237-248.
31. Van Sluis, C.A., Moolenaar, G.F., and Backendorf, C. (1983). *EMBO J.* 2, 2313-2318.
32. Forster, J.W. and Strike, P. (1985) *Gene* 35, 71-82.
33. Demple, B., Sedgwick, B., Robins, P., Totty, N., Waterfield, M.D., and Lindahl, T. (1985) *Proc. Natl. Acad. Sci. USA* 82, 2688-2692.