

---

**Nucleotide sequencing of the *ruv* region of *Escherichia coli* K-12 reveals a LexA regulated operon encoding two genes**

---

Fiona E. Benson, Graham T. Illing, Gary J. Sharples and Robert G. Lloyd

---

Department of Genetics, Queen's Medical Centre, University of Nottingham, Nottingham NG7 2UH, UK

---

Received December 10, 1987; Revised and Accepted January 21, 1988

---

**ABSTRACT**

The nucleotide sequence of a 2505 bp region of the *Escherichia coli* chromosome containing the LexA regulated *ruv* gene has been determined. A sequence of 1631 bp encoding two non-overlapping open reading frames that constitute a single operon and which specify polypeptides with predicted molecular weights of 22172 daltons and 37177 daltons respectively, was identified as the most probable sequence for *ruv*. Each of the two open reading frames, designated *ruvA* and *ruvB*, is preceded by a reasonable Shine-Dalgarno sequence. Two 16 bp sequences (SOS boxes) that match the consensus sequence for binding LexA protein are located 5' to *ruvA* in a region that provides a possible single promoter for expression of both *ruvA* and *ruvB*, with the second SOS box overlapping the putative -35 region. A possible transcriptional terminator is located 137 bp downstream of *ruvB*. The amino acid sequence predicted for *RuvB* contains a region that matches a highly conserved sequence found in several DNA repair and recombination proteins that bind ATP.

**INTRODUCTION**

The *ruv* locus of *Escherichia coli* was originally defined by mutations that increase sensitivity to mitomycin C, ultraviolet light, and ionising radiation, and cause a defect in cell division that results in an unusual degree of filamentous growth (1). Later studies identified further point mutations, phage  $\text{Mud}(\text{Ap}^{\text{r}}\text{lac})$  fusions, and *Th10* insertions, that confer a similar phenotype (2,3). Investigation of  $\beta$ -galactosidase synthesis in the *ruv::lac* fusion strains demonstrated that expression of *ruv* is induced by damage to DNA and is regulated by LexA as part of the SOS response (3). Furthermore, *ruv* derivatives of *recBC sbcBC* and *recBC sbcA* were found to show a significant reduction in the ability to foster recombinants in conjugational crosses, although *ruv* single mutants appeared reasonably proficient in recombination, (4,5). From these additional properties, it was suggested that *ruv*, along with other LexA regulated recombination genes such as *recA*, *recN*, and *recQ* (6,7,8), is involved in a recombinational process for repair of damaged DNA (4,5).

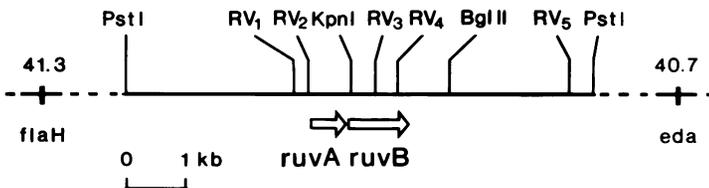
Analysis of recombinant plasmids that carry *ruv*<sup>+</sup> succeeded in locating the gene to a 7.7 kb *Pst*I fragment of the chromosome (2,9). The cloned DNA was shown to encode three polypeptides of approximately 27, 33, and 41 kilodaltons, of which only

the 41 kd protein was required for ruv<sup>+</sup> activity (9). Further studies (unpublished) indicated that the genes for the 27 kd and 41 kd proteins were adjacent and that a Tn10 insertion (ruv-59::Tn10), which on the basis of the phenotype conferred was thought to be within ruv (2), was in fact located in the gene for the 27 kd protein. Given that a Tn10 insertion can exert a polar effect, these observations suggested that the genes for the 27 kd and 41 kd proteins may comprise an operon. To gain further insight into the molecular organisation of ruv, we determined the nucleotide sequence of the DNA encoding the 27 kd and 41 kd proteins. The 2505 bp sequence presented in this paper reveals that the genes specifying these proteins do indeed form part of an operon regulated by LexA.

METHODS

Strains and plasmids

The original source of ruv was pPVA101, which carries a 10.4 kb HindIII fragment of DNA cloned into the low copy number vector pHSG415 (2). Fig. 1 shows the location and transcriptional orientation of ruv within the 7.7 kb PstI sub-fragment of the 10.4 kb insert in pPVA101 (2,9). pFB803 carries a 590 bp Sau3AI DNA fragment spanning EcoRV sites 1 and 2 cloned into the BamHI site of pKK232-8, which carries a promoterless chloramphenicol acetyl transferase (CAT) gene (10). pFB512 and pGS700 are derivatives of pUC18 (11) carrying respectively the BamHI-KpnI and KpnI-BglII fragments of pPVA101 that span the ruv region (Fig. 1). Neither plasmid encodes the intact 41 kd product of ruv since the gene is interrupted at the KpnI site. pFB512 encodes the 27 kd and 33 kd proteins (unpublished work). Since the ruv region could not be cloned intact into high copy number plasmids (9), pFB512 and pGS700 were used to provide more convenient sources of DNA for further subcloning of this region.



**Figure 1**

Partial restriction map of the 7.7 kb PstI fragment of the E. coli chromosome encoding ruv based on analysis of pPVA101 and its derivatives (9) and the DNA sequence (this work). EcoRV is abbreviated to RV. The flanking markers eda and flaH, and their map co-ordinates in minutes, are included to show the orientation of ruv relative to the chromosome. The DNA insert in pPVA101 contains a BamHI site approximately 1.0 kb to the left of the PstI fragment (9). This site is absent from the chromosome (unpublished work). The arrows indicate the approximate location and direction of transcription of the ruvA and ruvB genes.

Recombinant plasmids and the phage cloning vectors M13mp18 and M13mp19 and their derivatives were maintained in strain JM101 (11).

#### Enzymes and biochemicals

Restriction enzymes, T4 DNA ligase and DNA Polymerase I Klenow fragment were from commercial sources and were used as recommended by the suppliers. Deoxynucleoside and dideoxynucleoside triphosphates were obtained from Sigma. Radiolabelled nucleotides and M13 universal sequencing primer (17mer) were obtained from Amersham.

#### DNA sequence analysis

Previous studies (9) located ruv to a 2.2 kb section of DNA spanning the KpnI site (Fig. 1). Overlapping restriction fragments from EcoRV site 1 to the BglII site were therefore inserted into the multiple cloning site of M13mp18 and M13mp19 and sequenced by the dideoxy chain termination technique of Sanger et al. (12), using a synthetic universal primer and [ $\alpha$   $^{35}$ S] dATP. A primer of 25 nucleotides, kindly provided by New Brunswick Scientific (UK) Ltd., was used to determine the sequence immediately 3' of EcoRV site 4. Sequencing reactions containing dITP instead of dGTP were used to confirm the sequence of GC rich regions (13). The entire sequence was determined for both strands of the DNA and was compiled and analysed using commercial computer software packages.

### RESULTS AND DISCUSSION

#### Nucleotide sequence of the region encoding ruv.

Figure 2 shows the nucleotide sequence of the 2505 bp region of the chromosome between EcoRV site 1 and the BglII site (Fig. 1). Analysis of the sequence revealed two major open reading frames. The first, beginning with a GTG codon at bp 317, extends 609 nucleotides to a TGA termination codon at bp 926. A possible ribosome binding site (14) is provided by the sequence AGGAGC located 11 bp upstream of the GTG initiation codon. Translation of this open reading frame would give a polypeptide of 203 amino acids with a predicted molecular weight 22172 daltons. We believe this to be the structural gene for the 27 kd protein (9). This conclusion is supported by the fact that Tn1000 insertions in pFB512 that abolish synthesis of the 27 kd protein have been mapped to the interval between EcoRV site 1 and the KpnI site (unpublished work). The molecular weight of the protein predicted by the DNA sequence is within the limits of the estimate of 27 kd obtained by SDS-PAGE (9).

The second open reading frame begins with an ATG codon at bp 937 and extends 1008 nucleotides to a TAA termination codon at bp 1944. The sequence ATGAGG, which appears 12 bp upstream of the ATG initiation codon and overlaps the last nucleotide of the previous reading frame, provides a possible ribosome binding site. We believe

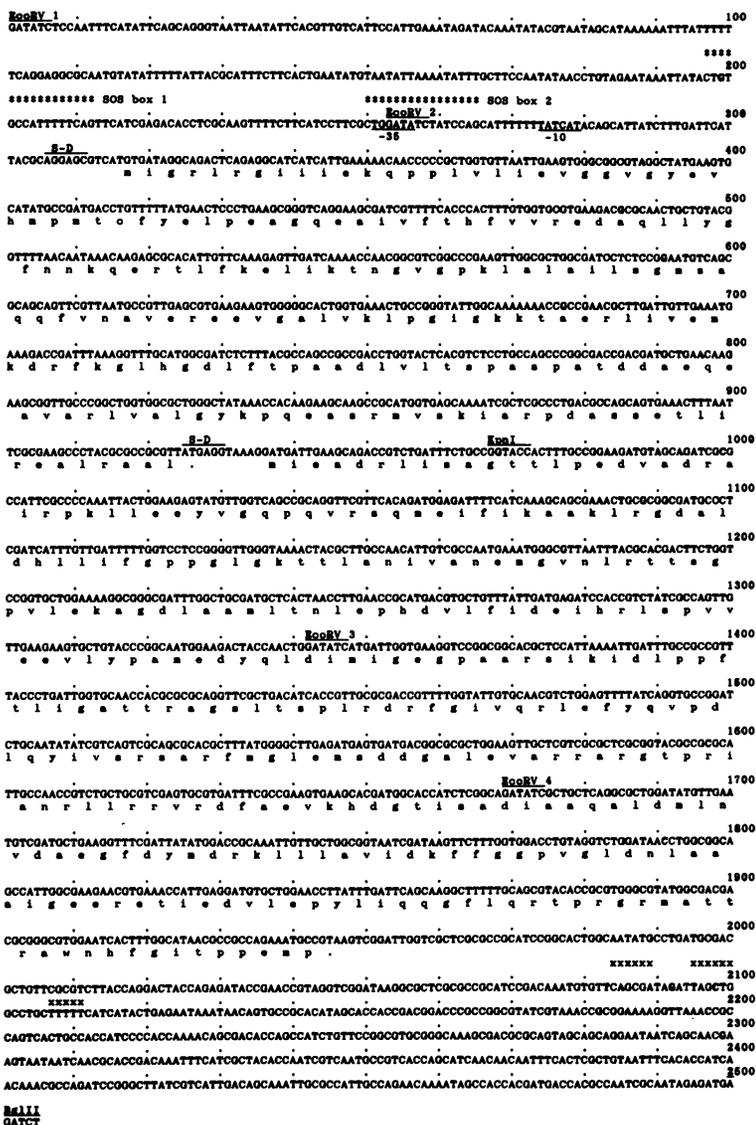


Figure 2

Nucleotide sequence of the 2505 bp chromosomal *ruv* region from EcoRV site 1 to the BglII site. The EcoRV, KpnI, and BglII sites referred to in the text are overlined, and the sequence is numbered from EcoRV site 1. LexA boxes are identified by asterisks above the sequence. Putative promoter -35 and -10 regions are underlined and labelled as such. The proposed transcriptional terminator is indicated by a string of Xs above the nucleotides. Putative Shine-Dalgarno (S-D) sequences are overlined. The predicted amino acid sequence for each of the two major open reading frames is aligned below the first base of each codon.

**Table 1**

Alignment of SOS boxes in the promoters of ruv and other LexA regulated genes. The symmetrical consensus sequences are boxed.

Gene	Sequence		Reference
<u>ruv</u> 1	CTG	TGCCATTTTT CAG	
2	CTG	GATATCTATC CAG	
<u>lexA</u> 1	CTG	TATATACTCA CAG	28,29
2	CTG	TATATACACC CAG	30
<u>recA</u>	CTG	TATGAGCATA CAG	28,31
<u>recN</u> 1	CTG	TATATAAAAC CAG	32
2	CTG	TACACAATAA CAG	
<u>recQ</u>	CTG	TTTTTATTT CAG	33
<u>uvrA</u>	CTG	TATATTCATT CAG	34
<u>uvrB</u>	CTG	TTTTTTTATC CAG	35
<u>uvrD</u>	CTG	TATATATACC CAG	36
<u>umuD</u>	CTG	TATATAAAAA CAG	37,38
<u>mucAB</u>	CTG	TATAAATAAA CAG	37
<u>impAB</u>	CTG	TATATACATA CAG	39
<u>sulA</u>	CTG	TACATCCATA CAG	40
consensus	CTG	tatatatata CAG	27

this second open reading frame to be the structural gene for ruv since it spans the KpnI site. Previous studies (9) established that ruv is inactivated by interruption of the sequence at this restriction site. Furthermore, translation of this open reading frame would give a polypeptide of 336 amino acids with a predicted molecular weight of 37177 daltons, which is in reasonably good agreement with the estimate of 41 kd for Ruv obtained by SDS-PAGE (9).

The only other major open reading frame begins with a GTG initiation codon at bp 2254. However, its significance is unclear as it extends beyond the BglII end of the sequence and analysis of the region downstream of ruv failed to reveal sequences for transcriptional and translation initiation.

#### Identification of a putative promoter for ruv

Sequences similar to those commonly found in E. coli promoter regions (15) are located upstream of the first open reading frame, with a possible -10 region provided by TATCAT, beginning at bp 274 and a -35 region by TGGATA beginning at bp 251 (Fig. 2). This region also contains two sequences that provide possible sites for binding LexA protein (16). The first, designated LexA box 1 in Fig. 2, extends from bp 197 to bp 212, while the second, designated LexA box 2, extends from bp 250 to bp 265 and overlaps the proposed -35 region. Table 1 shows that these LexA boxes match the consensus sequence for LexA binding established from a comparison of similar sequences found in the promoter regions of other SOS regulated genes. Since expression of ruv is known to be regulated by LexA (3), it seems likely that the

region immediately upstream of the first open reading frame is the promoter for expression of ruv. This conclusion is supported by the fact that expression of the CAT gene in pFB803, which carries a 590 bp Sau3AI DNA fragment extending from the left of EcoRV site 1 to the Sau3AI site beginning at bp 451, is regulated in a LexA-dependent fashion (data not shown). Furthermore, the gel retardation assay of Fried and Crothers (17) revealed specific binding of LexA protein to this 590 bp Sau3AI fragment and to the two sub-fragments of this region generated by cleavage at the TaqI site beginning at bp 217 (data not shown). The latter observation indicated that both LexA boxes bind LexA protein independently, as would be expected from the fact that they are separated by 37 bp. The affinity of LexA for the 590 bp fragment was approximately one order of magnitude lower than for a DNA fragment containing the recN promoter, which also contains two LexA boxes (18). This lower affinity for the DNA fragment containing the putative ruv promoter is in agreement with the observation that ruv is not so tightly regulated as recN (3,19).

Analysis of the sequence upstream of ruv failed to reveal any other region that could provide promoter activity. A sequence beginning at bp 2084, 137 bp downstream of the ruv stop codon, shows the characteristics of a rho factor-independent terminator (20). Initiation of transcription from the promoter suggested above and its extension to this putative terminator would imply that ruv is part of a LexA regulated operon that encodes the 22 kd polypeptide in addition to the 37 kd Ruv protein needed to correct the DNA repair deficiency of ruv mutants. An operon structure of this nature is supported by the fact that  $\beta$ -galactosidase synthesis in ruv-61::Mud(Ap<sup>r</sup>lac) insertion strains (2,3) is subject to LexA regulation (3) and that this insertion is located in the KpnI-BglII region in the correct orientation for expression of lac from the suggested promoter (unpublished work). It would also explain how a Tn10 insertion (ruv-59::Tn10) previously thought to be within ruv (2) but subsequently located between the Sau3AI sites beginning at bp 451 and bp 541 (unpublished work) is able to confer a typical ruv mutant phenotype. Presumably, any transcription initiated at the LexA regulated promoter would terminate at the Tn10 insertion and prevent expression of the otherwise intact ruv gene.

On the basis of these studies, we propose that the DNA sequence encoding the 22 kd and 37 kd proteins should be referred to as the ruv operon. Furthermore, we suggest that the gene for the 22 kd protein be designated ruvA and that the ruv gene be renamed as ruvB. During the course of these studies, we learnt that the ruv operon had been sequenced independently by Shinagawa, H., Makino, K., Amemura, M., Kimura, S., and Nakata, A. (H. Shinagawa, personal communication). The sequence of the ruv operon presented here is identical to that found by Shinagawa et al.

Table 2

Alignment of the putative ATP binding sequence of RuvB with similar sequences found in other *E. coli* proteins involved in DNA metabolism. Identical or similar sequences are boxed.

Protein	Residues	Sequence										Reference
UvrA	24-45	DKLIV	V	T	G	LS	G	S	GKS	SLAFDT	L	41
	633-654	GLFTC	I	T	G	VS	G	S	GKS	TLINDT	L	
UvrB	32-53	LAHQI	L	L	G	VT	G	S	GKT	FTIANV	I	42,43
UvrD	22-43	RSNLL	V	L	A	GA	G	S	GKT	RVLVHR	I	
RecB	16-37	QGERL	I	E	A	SA	G	T	GKT	FTIAAL	Y	45
DnaB	223-244	SDLII	V	A	A	RP	S	M	GKT	TFAMNL	V	46
Rho	172-192	GORGL	I	V	A	PP	K	A	GKT	MLIQNI	A	47
RecN	22-48	SGMTV	I	T	G	ET	G	A	GKS	LAIDAL	G	32
RecD	164-185	RRLSV	I	S	G	GP	G	T	GKT	TTVAKL	L	48
RecA	59-80	GRIVE	I	Y	G	PE	S	S	GKT	TLTLQV	I	25
RuvB	55-77	LDHLL	I	F	G	PP	G	L	GKT	TLANIV	A	

#### Codon usage and amino acid composition of *ruvA* and *ruvB*.

Initiation of RuvA synthesis with a GTG codon suggests that the level of expression of *ruvA* may be controlled at both translational and transcriptional levels. GTG serves as the initiation codon in 8% of the translation initiation sites characterised in *E. coli* so far (21), and is approximately 66% as efficient as the more common ATG codon (22). Initiation of RuvB with an ATG codon possibly allows for RuvB to be produced in greater amounts than RuvA. In *ruvA*, the initiation codon is immediately followed by the rare codon ATA (isoleucine), which on the basis of 288 translational start sites characterised has been predicted to occur only 3 times as a second codon per 1000 genes, but which is one of the most efficient second codons in terms of its effects on translation initiation (23).

Further possibilities for translational control relate to the frequency with which rare codons are used (24). The rare codons ATA(Ile), TCG(Ser), CCT(Pro), CCC(Pro), ACG(Thr), CAA(Gln), AAT(Asn), AGG(Arg) occur with a frequency of 7.4% in *ruvA* and 8.8% and 13% in the two non-coding frames. Rare codons occur with a frequency of 6.8% in *ruvB* and 9.3% and 11% in the non-coding frames. These values are consistent with expression of the *ruv* genes being controlled primarily at a transcriptional level.

The secondary structure of RuvA and RuvB predicted from Chou-Fasman analysis (50) of the amino acid sequence suggests both proteins have conformations typical of moderately expressed globular proteins. Analysis of the amino acid composition of RuvB revealed a region between residues 55 and 77 (Table 2) that is similar to highly conserved sequences found among proteins known to bind adenine nucleotides (25,26).

## Nucleic Acids Research

---

This raises the possibility that RuvB, like several other proteins involved in DNA repair and recombination, requires ATP as a co-factor.

### ACKNOWLEDGEMENTS

We would like to thank Hideo Shinagawa for communicating results prior to publication, Paul Attfield for his clone of the ruv region, and Peter Strike and Darerca Owen for the generous gift of purified LexA protein. This work was supported by a grant to R.G.L. from the Science and Engineering Research Council.

### References

1. Otsuji,N, Iyehara,H, and Hideshima,Y. (1974) *J. Bacteriol.* **117**, 337-344.
2. Shurvinton,C.E., Lloyd,R.G., Benson,F.E., and Attfield,P.V. (1984) *Mol. Gen. Genet.* **194**, 322-329.
3. Shurvinton,C.E., and Lloyd,R.G. (1982) *Mol. Gen. Genet.* **185**, 352-355.
4. Lloyd,R.G., Benson,F.E., and Shurvinton,C.E. (1984) *Mol. Gen. Genet.* **194**, 303-309.
5. Lloyd,R.G., Buckman,C. and Benson,F.E. (1987) *J. Gen. Micro.* **133**, 2531-2538.
6. Lloyd,R.G., Picksley,S.M., and Prescott,C. (1983) *Mol. Gen. Genet.* **195**, 162-167.
7. Walker,G.C. (1985) *Ann. Rev. Biochem.* **54**, 425-457.
8. Irino,N., Nakayama,K., and Nakayama,H. (1986) *Mol. Gen. Genet.* **205**, 298-304.
9. Attfield,P.V., Benson,F.E., and Lloyd,R.G. (1985) *J. Bacteriol.* **164**, 276-281.
10. Brosius,J. (1984) *Gene* **27**, 151-160.
11. Yanisch-Perron,C., Vieira,J., and Messing,J. (1985) *Gene* **33**, 103-119.
12. Sanger,F., Nicklen,S., and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
13. Mills,D.R., and Kramer F.R. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 2232-2235.
14. Shine,J., and Dalgarno,L. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 1342-1346.
15. Harley,C.B., and Reynolds,R.P. (1987) *Nucl. Acids Res.* **15**, 2343-2361.
16. Walker,G.C. (1984) *Microbiol. Rev.* **48**, 60-93.
17. Fried,M., and Crothers,D.M. (1981) *Nucl. Acids Res.* **9**, 6505-6525
18. Rostas,K., Morton,S.J., Picksley,S.M., and Lloyd,R.G. (1987) *Nucl. Acids Res.* **15**, 5041-5049
19. Picksley,S.M., Lloyd,R.G., and Buckman,C. (1984) *Cold Spring Harbor Symp. Quant. Biol.* vol. XLIX, pp.469-474. Cold Spring Harbor, New York.
20. Brendel,V., Hamm,G.H., and Trifonov,E.N. (1986) *J. Biomolecular Structure and Dynamics* **3**, 705-723.
21. Gren,E.J. (1984) *Biochimie* **66**, 1-29.
22. Reddy P., Peterkofsky A., and McKenney,K. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 5656-5660.
23. Looman,A.C., Bolaender,J., Comstock,L.J., Eaton,D., Jhurani,P., DeBoer,H.A., and Van Knippenberg,P.H. (1987) *EMBO J.* **6**, 2489-2492.
24. Konigsberg,W., and Godson,G.N. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 687-691.
25. Walker,J.E., Saraste,M., Runswick,M.J., and Gay,N.J., (1982) *EMBO J.* **1**, 645-651.
26. Fry,D.C., Kuby,S.A., and Mildvan,A.S., (1980) *Proc. Natl. Acad. Sci. USA* **83**, 907-911.
27. Wertman,K.F., and Mount,D.W. (1985) *J. Bacteriol.* **163**, 376-381.
28. Horii,T., Ogawa,T., and Ogawa,H. (1981) *Cell* **23**, 689-697.

29. Little, J.W., Mount, D.W., and Yanisch-Perron, C.R. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 4199-4203.
30. Brent, R., and Ptashne, M. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 4204-4208.
31. Sancar, A., Stachelek, C., Konigsberg, W., and Rupp, W.D. (1980) *Proc Natl. Acad. Sci USA* **77**, 2611-2615.
32. Rostas, K., Morton, S.J., Picksley, S.M., and Lloyd, R.G. (1987) *Nucl. Acids Res.* **13**, 5041-5049.
33. Irino, N., Nakayama, K., and Nakayama, H. (1986) *Mol. Gen. Genet.* **205**, 298-304.
34. Sancar, A., Sancar, G.B., Rupp, W.D., Little, J.W., and Mount, D.W. (1982) *Nature* **289**, 96-98.
35. Sancar, G.B., Sancar, A., Little, J.W., and Rupp, W.D., (1982) *Cell* **28**, 523-530.
36. Finch, P., and Emmerson, P.T. (1983) *Gene* **25**, 317-323.
37. Perry, K.L., Elledge, S.J., Mitchell, B.B., Marsh, L., and Walker, G.C. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4331-4335.
38. Kitagawa, Y., Akaboshi, E., Shinagawa, H., Horii, T., Ogawa, H., and Kato, T. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4336-4340.
39. Strike, P., and Lodwick, D. (1987) *J. Cell Sci. Suppl.* **6**, 303-321.
40. Cole, S.A. (1983) *Mol. Gen. Genet.* **189**, 400-404.
41. Husain, I., Van Houten, B., Thomas, D.C., and Sancar, A. (1986) *J. Biol. Chem.* **261**, 4895-4901.
42. Arikan, E., Kulkarni, M.S., Thomas, D.C., and Sancar, A. (1986) *Nucl. Acids Res.* **14**, 2637-2650.
43. Backendorf, C., Spaink, H., Barbeiro, A.P., and Van de Putte, P. (1986) *Nucl. Acids Res.* **14**, 2877-2890.
44. Finch, P.W. and Emmerson, P.T. (1984) *Nucl. Acids Res.* **12**, 5789-5799.
45. Finch, P.W., Storey, A., Chapman, K.E., Brown, K., Hickson, I.D., and Emmerson, P.T. (1986) *Nucl. Acids Res.* **14**, 8583-8594.
46. Nakayama, N., Arai, N., Bond, M.W., Kaziro, Y., and Arai, K. (1984) *J. Biol. Chem* **259**, 97-101.
47. Pinkham, J.L., and Platt, T. (1983) *Nucl. Acids Res.* **11**, 3531-3545.
48. Finch, P.W., Storey, A., Brown, K., Hickson, I.D., and Emmerson, P.T. (1986) **14**, 8583-8594.
49. Lupski, J.R., Ruiz, A.A., and Godson, G.N. (1984) *Mol. Gen. Genet.* **195**, 394-401
50. Chou, P.Y., and Fassman, G.D. (1978) *Adv. Enzymol.* **47**, 45-148.