Volume 14 Number 9 1986 Nucleic Acids Research

Nucleotide sequence of the *tag* gene from *Escherichia coli*

Anne-Lill Steinum and Erling Seeberg*

Norwegian Defence Research Establishment, Division for Environmental Toxicology, N-2007 Kjeller, Norway

Received 18 March 1986; Accepted 11 April 1986

ABSTRACT

We have determined the complete nucleotide sequence of the tag gene, encoding 3-methyladenine DNA glycosylase I from Escherichia coli. From the nucleotide sequence it is deduced that the tag enzyme consists of 187 amino-acids and has a calculated molecular weight of 21.1 kdaltons. The tag enzyme is unusually rich in cysteine (8 residues) with a cluster of three consecutive cysteines near the C-terminal end. The tag coded DNA glycosylase does not show significant sequence homology to the alkA coded glycosylase in spite of that both of these enzymes catalyze the release of free 3-methyladenine from alkylated DNA.

INTRODUCTION

The tag and alkA genes of Esherichia coli encode two distinct DNA glycosylases which participate in DNA repair of cells exposed to alkylating agents (1-5). Both enzymes catalyze the excision of 3-methyladenine residues from DNA exposed to alkylating mutagens such as methyl-methanesulfonate (MMS) and N-methyl-N'-nitro-N-nitrosoguanidine (MNNG)(6-8). The N3-position of adenine is quantitatively a major site for DNA alkylations (9) and 3-methyladenine residues in DNA are the major cytotoxic lesions in alkylated cells (2,3,10,11).

Both the tag and alkA genes have recently been cloned and the gene products identified radiochemically as proteins of Mr 21.000 and 30.000, respectively (4,5,12). The alkA gene has also been sequenced and the amino-acid composition of the alkA coded glycosylase (TagII) deduced from the nucleotide sequence (13). It appears that TagII contains 282 amino-acids and has a calculated molecular weight of 31.4 kdaltons. We report here the nucleotide sequence of the tag gene and the amino-acid sequence of the tag coded glycosylase (TagI) as deduced from the nucleotide sequence. It appears that TagI has 187 amino-acids and a calculated molecular weight of 21.1 kdaltons. In spite of the similar catalytic properties of TagI and TagII there is no obvious amino-acid sequence homology between the two enzymes.

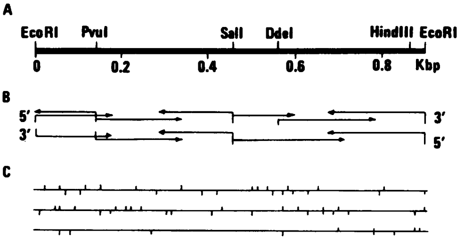

Fig. 1. Restriction map (A), sequencing strategy (B) and distribution of stop codons in different reading frames (C). A. Only restriction enzyme sites for relevant enzymes are shown. The right end of the fragment extending from HindIII to EcoRI originates from the pBR322 vector in which HindIII-site the tag gene was originally cloned. B. Both strands of the fragment were sequenced from the sites indicated using the method of Maxam and Gilbert (15). C. The distribution of stopcodons in the three reading frames on both strands are plotted based on the sequence shown in Fig. 2.

MATERIALS AND METHODS

Bacterial strains and plasmids.

Plasmid pBK202 carries the tag gene cloned in the EcoRI site of pBR322 (12).

```
                                                                            -35
AATTCGGCAATATTATTGTCATTGTATGAAGGATATCGGGCATAGTAGCCCTGTATTAAATATTGACTTTTTCACCGA
EcoRI    *         *         *         *         *         *         *
                                                            1
             -10                          S-D seq.          Met Glu Arg Cys Gly Trp
TGCGTCAAGAAAAGCGGCTGAAATTTTTACGATCGGGTACATAGCGAGGGAAAGT ATG GAA CGT TGC GGC TGG
*         *         100     PvuI          *         *              *           *
            10                                      20
Val Ser Gln Asp Pro Leu Tyr Ile Ala Tyr His Asp Asn Glu Trp Gly Val Pro Glu Thr
GTG AGT CAG GAC CCG CTT TAT ATT GCC TAC CAT GAT AAT GAG TGG GGC GTG CCT GAA ACT
          *             *            *            *             200          *
            30                                      40
Asp Ser Lys Lys Leu Phe Glu Met Ile Cys Leu Glu Gly Gln Gln Ala Gly Leu Ser Trp
GAC AGT AAA AAA CTG TTC GAA ATG ATC TGC CTT GAA GGG CAG CAG GCT GGA TTA TCG TGG
          *             *            *            *             *            *
            50                                      60
Ile Thr Val Leu Lys Lys Arg Glu Asn Tyr Arg Ala Cys Phe His Gln Phe Asp Pro Val
ATC ACC GTC CTC AAA AAA CGC GAA AAC TAT CGC GCC TGC TTT CAT CAG TTC GAT CCG GTG
          *             *           300           *             *            *
            70                                      80
Lys Val Ala Ala Met Gln Glu Glu Asp Val Glu Arg Leu Val Gln Asp Ala Gly Ile Ile
AAG GTC GCA GCA ATG CAG GAA GAG GAT GTC GAA AGA CTG GTA CAG GAC GCC GGG ATT ATC
          *             *            *            *             *            *
            90                                      100
Arg His Arg Gly Lys Ile Gln Ala Ile Ile Gly Asn Ala Arg Ala Tyr Leu Gln Met Glu
CGC CAT CGA GGG AAA ATT CAG GCA ATT ATT GGT AAT GCG CGG GCG TAC CTG CAA ATG GAA
          400           *            *            *             *            *
            110                                     120
Gln Asn Gly Glu Pro Phe Val Asp Phe Val Trp Ser Phe Val Asn His Gln Pro Gln Val
CAG AAC GGC GAA CCG TTT GTC GAC TTT GTC TGG TCG TTT GTA AAT CAT CAG CCA CAG GTG
          *             SalI         *            *             500          *
            130                                     140
Thr Gln Ala Thr Thr Leu Ser Glu Ile Pro Thr Ser Thr Ser Ala Ser Asp Ala Leu Ser
ACA CAA GCC ACA ACG TTG AGC GAA ATT CCC ACA TCT ACG TCC GCC TCC GAC GCC CTA TCT
          *             *            *            *             *          DdeI
            150                                     160
Lys Ala Leu Lys Lys Arg Gly Phe Lys Phe Val Gly Thr Thr Ile Cys Tyr Ser Phe Met
AAG GCA CTG AAA AAA CGT GGT TTT AAG TTT GTC GGC ACC ACA ATC TGT TAC TCC TTT ATG
          *             *           600           *             *            *
            170                                     180
Gln Ala Cys Gly Leu Val Asn Asp His Val Val Gly Cys Cys Cys Tyr Pro Gly Asn Lys
CAG GCA TGT GGG CTG GTG AAT GAT CAT GTG GTT GGC TGC TGT TGC TAT CCG GGA AAT AAA
          *             *            *            *             *            *

Pro End      --------------->  <----------------
CCA TGA TTCGGAAAGCGCAACGTTCAGAACTTCCCGCGATCCTCGAACTGTGGCTGGAAAGTACAACCTGGGGGCAT
         700           *         *         *         *         *         *

CCCTTTATAAAAGCGAATTACTGGCGTGACTGCATCCGCTGGTGCGGGATGCCTATCTTGCCAACGCGCAAAACTGGGT
*         *         *         800       *         *         *         *

                   <--pBR322--------------------------
CTGGGAAGAAGACGGTAAGCTTATCGATGATAAGCTGTCAAACATGAG
  *         *   HindIII          *         *    EcoRI
```

Fig. 2. Nucleotide sequence of the tag gene and flanking regions and deduced amino-acid sequence of 3-methyladenine DNA glycosylase I.

Table 1. Amino-acid composition of the tag gene product.

	tag	ung	alkA	ada
Alanine	14 (7.5)	15 (8.1)	40(14.2)	48(13.6)
Arginine	8 (4.3)	9 (4.8)	20 (7.1)	34 (9.6)
Asparagine	7 (3.7)		6 (2.1)	11 (3.1)
Aspartic	9 (4.8)	}15 (8.1)	13 (4.6)	18 (5.1)
Cysteine	8 (4.3)	1 (0.5)	4 (1.4)	12 (3.4)
Glutamic	12 (6.4)		13 (4.6)	20 (5.7)
Glutamine	13 (7.0)	}25(13.4)	14 (5.0)	21 (5.9)
Glycine	13 (7.0)	14 (7.5)	20 (7.1)	19 (5.4)
Histidine	5 (2.7)	6 (3.2)	4 (1.4)	8 (2.3)
Isoleucine	10 (5.4)	9 (4.8)	12 (4.3)	12 (3.4)
Leucine	11 (5.9)	19(10.2)	33(11.7)	34 (9.6)
Lysine	11 (5.9)	10 (5.4)	8 (2.8)	14 (4.0)
Methionine	5 (2.7)	3 (1.6)	8 (2.8)	6 (1.7)
Phenylalan.	9 (4.8)	7 (3.8)	9 (3.2)	12 (3.4)
Proline	8 (4.3)	12 (6.5)	20 (7.1)	16 (4.5)
Serine	10 (5.4)	8 (4.3)	9 (3.2)	19 (5.4)
Threonine	9 (4.8)	10 (5.4)	13 (4.6)	20 (5.7)
Tryptophane	4 (2.1)	5 (2.7)	9 (3.2)	5 (1.4)
Tyrosine	6 (3.2)	6 (3.2)	12 (4.3)	6 (1.7)
Valine	15 (8.0)	12 (6.5)	15 (5.3)	19 (5.4)

The table includes published values for amino-acid compositions of the ada protein (18), the alkA gene product (TagII, ref 13) and for ung (uracil DNA glycosylase, ref 17). The first two are derived from nucleotide sequence analysis and the latter from amino-acid analysis of purified protein.

Table 2. Codon usage of the tag gene.

Phe	TTT	7	Leu	CTT	2·	Ile	ATT	6	Val	GTT	1
Phe	TTC	2	Leu	CTC	1	Ile	ATC	4	Val	GTC	6
Leu	TTA	1	Leu	CTA	1	Ile	ATA*	0	Val	GTA	2
Leu	TTG	1	Leu	CTG	5	Met	ATG	5	Val	GTG	6
Ser	TCT	2	Pro	CCT*	1	Thr	ACT	1	Ala	GCT	1
Ser	TCC	3	Pro	CCC*	1	Thr	ACC	2	Ala	GCC	6
Ser	TCA	0	Pro	CCA	2	Thr	ACA	4	Ala	GCA	5
Ser	TCG*	2	Pro	CCG	4	Thr	ACG*	2	Ala	GCG	2
Tyr	TAT	3	His	CAT	5	Asn	AAT*	5	Asp	GAT	4
Tyr	TAC	3	His	CAC	0	Asn	AAC	2	Asp	GAC	5
	TAA	0	Gln	CAA*	2	Lys	AAA	8	Glu	GAA	10
	TAG	0	Gln	CAG	11	Lys	AAG	3	Glu	GAG	2
Cys	TGT	3	Arg	CGT	2	Ser	AGT	2	Gly	GGT	2
Cys	TGC	5	Arg	CGC	3	Ser	AGC	1	Gly	GGC	5
	TGA	1	Arg	CGA	1	Arg	AGA	1	Gly	GGA	2
Trp	TGG	4	Arg	CGG	1	Arg	AGG*	0	Gly	GGG	4

Asterix represents rarely used codons as suggested by Konigsberg and Godson (20).

Purification of plasmid DNA.

Plasmid DNA was extracted from transformed cultures of AB1157 by an SDS/NaCl lysis procedure as previously described (14). The high-salt lysate was treated with proteinase K (100 µg/ml), dialyzed, extracted with phenol, dialyzed again, and the DNA was banded in CsCl/Ethidium bromide. Ethidium bromide was extracted with isopropanol and the DNA was finally purified by sedimentation in neutral sucrose followed by dialysis.

Restriction enzyme digests.

The DNA was digested with EcoRI, SalI, PvuI (Boehringer) and DdeI (Toyobo Chem.) according to manufacturers instructions.

DNA sequence analysis.

Sequences were determined according to Maxam and Gilbert (15). The DNA was end-labelled at the 5'-end with polynucleotide kinase and γ-[^{32}P]-ATP and at the 3'-end with terminal transferase with either cordycepin-α-[^{32}P]-ATP or dideoxy-α-[^{32}P]-ATP. Fragments with label at one end only was generated by cutting with a second restriction enzyme following labelling. Fig. 1 shows the sequencing strategy. The entire sequence was determined on both DNA strands. The sequence was analyzed by a microcomputer program for DNA and protein sequence analysis written in Turbo-Pascal (Seeberg, unpublished).

RESULTS

Nucleotide sequence and coding region.

The tag gene is contained within an 895 bp EcoRI fragment cloned in pBR322 (Fig. 2). The 31-bp sequence at the 3'-end (extending from the HindIII-site) originates from the pBR322 vector of the original clone (12). Fig. 1C indicates the distribution of stop codons in the different reading frames of both strands. It appears that the tag gene fragment only has one open reading frame which will allow for the synthesis of a polypeptide larger than 10 kdaltons. The tag gene product has been identified radiochemically as a polypeptide of 21 kdaltons (12). The open reading frame allows for the synthesis of a protein of 21.1 kdaltons starting from a putative start codon at position 134 from the upper EcoRI site. This codon is preceeded by a sequence with a good match to a ribosomal binding site (16) and we conclude that this is the initiation codon of TagI. There is an alternative ATG in frame upstream for the indicated start codon at position 26. However, this can be eliminated as the start codon because synthesis from this site will result in a polypeptide of 25 kd which is considerably larger than the size of the identified protein. Furthermore, there is no Shine-Dalgarno sequence preceeding that site.

Amino-acid composition and codon usage.

The amino-acid composition of the tag product is compiled in Table 1. For comparison are included the published values for ung (17), alkA (TagII, ref. 13) and ada (18). The unmodified protein of the tag gene contains 187 aminoacids with 21 negatively charged and 24 positively charged residues. Five of the positive ones are histidines which aminogroup only represents a weak base at neutral pH. From the composition of the charged residues, one would predict the pI value of TagI to be in between 6.5 and 7.0 which is consistent with that the enzyme only binds weakly to either cation or anion exchange materials (6,7, unpublished data).

The tag product has a very high proportion of cysteines even higher than the ada product which use cysteines as alkyl-acceptor residues for its alkyl-transferase activities (18). The amino-acid composition of the tag product is rather different from alkA, but remarkably similar to ung (Table

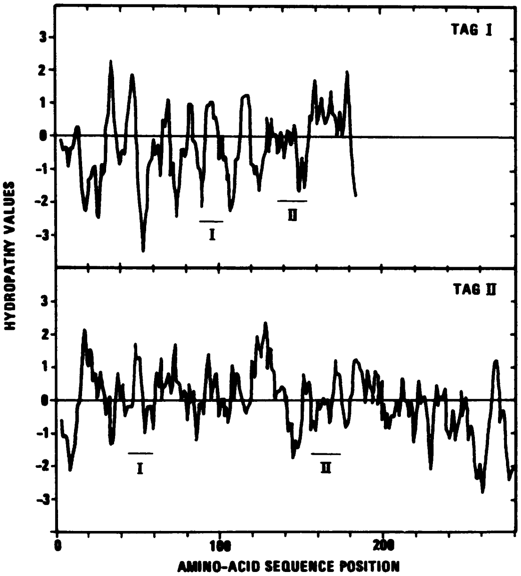

Fig. 3. Hydrophobicity plots for the tag and the alkA coded glycosylases. The hydrophobicity values are calculated as described by Kyte and Doolitle (26). The calculations for the alkA coded glycosylase are based on the sequence of Nakabeppu et al. 1984 (13). Region I (see bars on figure) represents the amino-acid sequence "RGKIQAIIGNARAYL" in tag and "RGVVTAIPDIARHTL" in alkA, while II represents "PISTSASDALSKALKKRG" and "PIPQRLAAADPQALKALG", respectively.

1). The alkA composition is more similar to ada which may relate to both being part of the adaptive response system.

The codon usage of the tag protein (Table 2) deviates somewhat from the codon usage in genes for other non-regulatory proteins in *E. coli* (19,20). The proportion of rarely used codons versus the total number of cognate codons is 21% as compared to 13% for the commonly expressed proteins. Genes with a high proportion of rare codons usually encode products which are present in only a few copies per cell presumably because of limiting amounts of tRNAs for rare codons (20,21). The uvrC gene also has a 21% frequency of rare codons like the tag gene and that protein appears to present only in about 10 copies per cell (21). The ada and alkA genes are even more extreme in this respect with a frequency of rare codons of 26 and 34%, respectively. Again, these are only present in few copies in uninduced cells. Other repair genes like uvrD (22), recA (23), and ssb (24) all have normal codon usage.

Putative promoter and terminator sequences.

Upstream for the initiation codon of tag, we have indicated a sequence with partial homology to the concensus promoter of *E. coli* (Fig. 2, ref. 25). The sequence has 4 out of 6 matches in the -10 region, and 3 out of 6 in the -35 region, and thus is not expected to be a strong promoter. Surprisingly, the plasmid pBK202 overproduces much more TagI enzyme than the original clone pBK201 (12). We think this might be due to the possible construction of a fusion promoter with a -10 region from the cloned DNA ("TATTAT", 10 bases downstream from the upper EcoRI site) and a -35 region from the pBR322 vector ("CTGTCA", 8 bases upstream from the EcoRI site, see end of sequence in Fig. 2 which is repeated in front of the first EcoRI site). This view is supported by the analysis of plasmids with the cloned DNA in opposite orientation which do not lead to such a large overproduction (data not shown).

Downstream from the stop codon, we have identified a region of dyad symmetry which could represent a transcriptional stop signal (25).

DISCUSSION

We have deduced the amino-acid sequence of 3-methyladenine DNA glycosylase I (TagI) from nucleotide sequence analysis of the tag gene. Recently, Nakabeppu et al. (13) reported the sequence of the alkA coded glycosylase. Both of these enzymes catalyze the release of 3-methyladenines from alkylated DNA, and one could expect to find extensive homology between the two enzymes at the amino-acid level. However, a search for such homology between the two glycosylases did not reveal regions with homology greater

A.

```
                                                    alkA - START CODON

        AGCAAAGCGCAGCG TCTGAATAACGTTTATGCTGAAAGCGGATGAATAAGGAGATGCG ATG
         | |||||||| || || |||
A TGA TTCGGAAAGCGCAACGTTCAGAA

tag - STOP CODON
```

B.

```
denV    -Leu-Lys-Arg-Gly-Phe-Asn-Ile-Lys-Asp-Thr-Thr-Val-
         TTA AAA CGT GGT TTT AAT ATC AAG GAT ACT ACA GTC
           | ||| ||| ||| ||| ||   |      |   ||  |||  ||
         AAA AAA CGT GGT TTT AAG TTT GTC GGC ACC ACA ATC
tag     -Lys-Lys-Arg-Gly-Phe-Lys-Phe-Val-Gly-Thr-Thr-Ile-
```

Fig. 4. Sequence homology to other repair genes. A. Sequence homology between the control region of alkA (13) and the putative terminator region of tag. B. Amino-acid and nucleotide sequence homology between tag (amino-acid no. 150-161) and denV (amino-acid no. 79-90 out of 138, ref. 28).

than was found in comparisons with other non-related proteins. Two regions were detected in which a minimum of 7 amino-acids were identical, one in a range of 15, the other in a range of 18 consecutive aminoacids (see Fig. 3). Four homologous triplets were found, no quadruplets, nor any pentapeptides with only one mismatch (data not shown).

Kyte and Doolitle (26) have deviced a method for displaying the profile of the hydrophobic and hydrophilic regions of proteins based on amino-acid sequence. Each amino-acid is given a value for its hydrophilic character (hydropathy values) and the average value calculated in a window of 7 amino-acids. We have applied this method to analyze the hydrophilic characters of TagI and TagII (Fig. 4). It appears that TagI in general has a more hydrophilic character than TagII and also has larger variations between hydrophilic and hydrophobic regions. It is notable that one of the regions with some homology occurs in a hydrophobic region for both proteins, while the other region occurs where the hydropathy profile scatters around zero value. Prediction of protein secondary structures by the Chou and Fasman method (27) indicates that for both proteins a structural change between β-turn and α-helix occurs within region I, while the entire region II for both proteins is in α-helix structure. Further work is needed to see if either of these regions is part of the active sites for the enzymes.

TagII is atypical for DNA glycosylases in having a broad substrate specificity, being capable of removing several different methylated bases from alkylated DNA. TagI is substrate specific and in this respect is similar to the other DNA glycosylases described, for instance the uracil DNA glycosylase (17). It could be that TagI acts in a different fashion than TagII and perhaps is more similar to the other substrate specific DNA glycosylases. This notion is supported by the remarkably similar amino-acid compositions of TagI and the uracil enzyme (Table 1). The sequences of the uracil enzyme are not available so far, however, the nucleotide sequence of the denV gene of bacteriophage T4 encoding a pyrimidine dimer specific DNA glycosylase was recently published (28). A comparison between denV and tag reveals a region of 15 identical consecutive nucleotides which encode 4 identical amino-acids (Fig. 4A). Towards the C-terminal end after a gap of 4 non-homologous amino-acids there are two threonines in both enzymes. It will be of interest to see wether such sequences also will be found in other glycosylases.

In spite of the lack of homology at the protein level between TagI and TagII, there is a nucleotide sequence homology between the control region of TagII and the terminator region of TagI (Fig.4B). Nakabeppu et al (13) have pointed out a sequence "AAAGCGCA" which occurs twice in the control region of alkA and also in the control region of ada (18). This sequence also occurs in the dyad symmetry which we have suggested as a terminator for tag. This could represent a box for the binding of the ada protein (Lindahl, personal communication). If this is the case one might speculate that under adaptive conditions the transcript from the tag gene could be longer and allow for the synthesis of an additional protein which could belong to the tag operon. There is a possibility for the coding of a protein starting at an ATG overlapping with the stop codon of tag which will proceed at least to the end of the fragment we have cloned (i.e. to the HindIII site). We are presently looking further at this possibility by cloning DNA downstream for tag.

One feature of the tag protein is clusters of positively charged amino-acid residues. Corresponding homologous groups are also found in other proteins which interact with DNA like for instance uvrC (21) and lexA (29). It is conceivable that such groups are responsible for "non-specific" binding between the negatively charged DNA and the proteins as an intial step in their action on the DNA.

Near the C-terminal end of tag, there is a group of three consecutive cysteines followed by tyrosine and proline. We have searched protein sequence data bases with a total of 4400 protein entries for such a sequence without finding a match. Three consecutive cysteins occur, but only in structural proteins like for instance keratin (30) and capsid protein B of phage lambda

(31). It could be that this group in tag contributes to maintain protein structure.

Recently, we learned that M. Sekiguchi and collaborators (personal communication) also have determined the nucleotide sequence of the tag gene. The amino-acid sequence derived from their results is identical to the one reported here.

ACKNOWLEDGMENTS

We are grateful to Michelle Ginsburg at Imperial Cancer Research Fund in London for conducting the protein data base search and to Mutsuo Sekiguchi for communicating results prior to publication.

*To whom correspondence should be addressed

REFERENCES

1. Yamamoto, Y., Katsuki, M., Sekiguchi, M. and Otsuji, N. (1978). J. Bacteriol. 135, 144-152
2. Karran, P., Lindahl, T., Øfsteng, I., Evensen, G. and Seeberg, E. (1980). J. Mol. Biol. 140, 101-127
3. Evensen, G. and Seeberg, E. (1982). Nature 296, 773-775
4. Nakabeppu, Y., Kondo, H. and Sekiguchi, M. (1984). J. Biol. Chem. 259, 13723-13729
5. Clarke, N. D., Kvaal, M. and Seeberg, E. (1984). Mol. Gen. Genet. 197, 368-372
6. Riazuddin, S. and Lindahl, T. (1978). Biochemistry 17, 2110-2118
7. Thomas, L., Yang, C.-H. and Goldthwait, D. (1982). Biochemistry 21, 1162-1169
8. Karran, P., Hjelmgren, T. and Lindahl, T. (1982). Nature 296, 770-773
9. Beranek, D.T., Weis, C.C. and Swenson, D.H. (1980). Carcinogenesis 1, 595-606
10. Boiteux, S., Huisman, O. and Laval, J. (1984). EMBO Journal 3, 2569-2573
11. Seeberg, E., Clarke, N.D., Evensen G., Kaasen, I. and Steinum,A.-L. (1986). In "Repair of DNA lesions introduced by N-nitrosocompounds" (eds. Krokan, H. and Myrnes, B.), pp, (in press)
12. Kaasen, I., Evensen, G. and Seeberg, E. (1986). (submitted for publication)
13. Nakabeppu, Y., Miyata, T., Kondo, H., Iwanaga, S. and Sekiguchi, M. (1984). J. Biol. Chem. 259, 13730-13736
14. Seeberg, E. (1978). Proc. Natl. Acad. Sci. USA 75, 2569-2573
15. Maxam, A. and Gilbert, W. (1980). Methods Enzymol. 65, 499-560
16. Shine, J. and Dalgarno, L. (1974). Proc. Natl. Acad. Sci. USA 71, 1342-1346
17. Lindahl, T., Ljungquist, S., Siegert, W., Nyberg, B. and Sperens, B. (1977). J. Biol. Chem. 252, 3286-3294
18. Demple, B., Sedgwick, B., Robins, P., Totty, N., Waterfield, M.D. and Lindahl, T. (1985). Proc. Natl. Acad. Sci. USA 82, 2688-2692
19. Ikemura, T. (1981). J. Mol. Biol. 146, 1-21
20. Konigsberg, W. and Godson, N. (1983). Proc. Natl. Acad. Sci. USA 80, 687-691
21. Sancar, G.B., Sancar, A. and Rupp, W.D. (1984). Nucl. Acid Res. 12, 4593-4608

22. Finch, P.W. and Emmerson, P. (1984). Nucl. Acids Res. 12, 5789-5799
23. Sancar, A., Williams, K.R., Chase, J.W. and Rupp, W.D. (1984). Proc. Natl. Acad. Sci. USA 78, 4274-4278
24. Sancar, A., Stachelek, C., Konigsberg, W. and Rupp, W.D. (1980). Proc. Natl. Acad. Sci. USA 78, 2611-2615
25. Rosenberg, M. and Court, D. (1980). Annu. Rev. Genet. 13, 319-353
26. Kyte, J. and Doolitle, R.F. (1982). J. Mol. Biol. 157, 105-132
27. Chou, P.Y. and Fasman, G. (1978). Adv. Enzymol. 47, 45-148
28. Radany, E.H., Naumovski, L., Love, J., Gutekumat, K., Hall, D. and Friedberg, E.C. (1984). J. Virol. 52, 846-851
29. Markham, B.E., Little, J.W. and Mount, D.W. (1981). Nucl. Acids Res. 9, 4149-4161
30. Elleman, T.C. and Dopheide, T.A. (1972). J. Biol. Chem. 247, 3900-3909
31. Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F. and Petersen, G.B. (1982). J. Mol. Biol. 162, 729-774