Ann. Ist. Super. Sanità
Vol. 25, N. 1 (1989), pp. 27-32

REPAIR OF IMIDAZOLE RING-OPENED PURINES IN DNA: OVERPRODUCTION OF THE FORMAMIDOPYRIMIDINE-DNA GLYCOSYLASE OF *ESCHERICHIA COLI* USING PLASMIDS CONTAINING THE fpg+ GENE

T. R. O'CONNOR, S. BOITEUX and J. LAVAL

Institut Gustave Roussy, Villejuif Cedex, France

Summary. - *The formamidopyrimidine-DNA glycosylase (Fapy-DNA glycosylase) of* Escherichia coli (E. coli) *was overexpressed by cloning the fpg+ gene on a multicopy plasmid and placing this gene under the control of the lac promoter. The lac promoter contributed significantly to the overall expression of the fpg gene only after the deletion of an inverted repeat sequence located immediately upstream from the fpg promoter. The biological purpose of the inverted repeat sequence may be associated with the termination of an adjacent gene transcribed in the same direction as the fpg gene in* E. coli. *Cells harboring the fpg gene under the control of the lac promoter were able to produce the Fapy-DNA glycosylase as at least 17% of the total soluble proteins. Such strains allow the preparation of milligram quantities of pure protein for use in the study of its catalytic properties and three dimensional crystal structure.*

Riassunto (Riparazione nel DNA delle purine con un anello imidazolico aperto: alta produzione della formamidopirimidina-DNA glicosilasi di *E. coli* usando plasmidi contenenti il gene fpg+). - *E' stata osservata alta espressione dell'enzima formamidopirimidina-DNA glicosilasi (Fapy-DNA glicosilasi) di* Escherichia coli (E. coli) *clonando il gene fpg+ in un plasmide multicopie sotto il controllo del promotore lac. Il promotore lac ha contribuito in modo significativo all'espressione del gene fpg solo dopo delezione di una sequenza invertita ripetuta e localizzata strettamente a monte del promotore di fpg. Il significato biologico della sequenza invertita ripetuta potrebbe essere associato con la terminazione di un gene adiacente trascritto nella stessa direzione del gene fpg in* E. coli. *Le cellule che contengono il gene fpg sotto il controllo del promotore lac producono la Fapy-DNA glicosilasi in quantità equivalenti ad almeno il 17% delle proteine solubili totali. Questi ceppi rendono possibile la preparazione di milligrammi di proteina pura permettendo quindi lo studio delle sue proprietà catalitiche e della struttura cristallina tridimensionale.*

Introduction

N7-methylguanine is the major product in DNA treated by chemical carcinogens such as N-methylnitrosourea or N-methyl-N'-nitro-N-nitroso-guanidine [1]. This lesion does not seem to be harmful to the cell as it does not interfere with DNA synthesis *in vitro* [2-4]. However, N7 alkylation favors the cleavage of the glycosidic bond or the cleavage of the imidazole ring, which yields 2-6 diamino-4-hydroxy-5-N-methylformamidopyrimidine (Fapy) [5]. Several observations suggest that the imidazole ring-opened form of N7-methylguanine might play a significant role in processes leading to mutagenesis and/or cell death by alkylating agents. *In vitro* DNA synthesis experiments show that Fapy residues inhibit DNA synthesis by *E. coli* DNA polymerase I [2]. Furthermore, the termination pattern of the *in vitro* DNA synthesis showed that *E. coli* DNA polymerase I stops one base before the Fapy residues [4]. Since DNA synthesis terminates one base before N3-methyladenine lesions [3], and this lesion is a major cell killing lesion after treatment with alkylating agents [6], by analogy, Fapy residues would be expected to be involved in cell death. The occurence of a DNA glycosylase which excises the ring-opened form of N7-methylguanine either in *E. coli* [7-9] or in mammalian cells [10] implies that this enzyme may play a role in the overall maintenance of genetic material. Study of this enzyme, however, has been limited as a result of its low expression level in *E. coli* wild type strains. Therefore, we cloned the Fapy-DNA glycosylase gene (fpg) of *E. coli* on the pBR322 plasmid [9]. The fpg+ gene was sequenced and found to code for a protein corresponding to the appropriate size (30.2 kDa) and amino acid composition of the Fapy-DNA glycosylase [9]. The fpg+ gene was further subcloned in the pUC18 and pUC19 plasmids to yield pFPG50 and 60, respectively. Despite the 60-fold overproducer character of the cells harboring the pFPG60 plasmid, the Fapy-DNA glycosylase did not exceed 1% of the total soluble proteins. This relatively low expression of the gene was

mainly due to the inefficiency of the transcription from the lac promoter of the pFPG60 plasmid. In this paper we report experiments which show that this low expression level of the cloned Fapy-DNA glycosylase in plasmids using the lac promoter is associated with a palindromic sequence which is located upstream from the fpg promoter. Deletion of this palindromic sequence allowed the level of Fapy-DNA glycosylase to increase to 17% of the total soluble cellular proteins.

Materials and methods

DNA. - Plasmid DNA used was isolated and treated as recommended by Maniatis *et al.* [11].

Enzymes. - All commercial enzymes (restriction endonucleases, T4 DNA ligase, and calf intestine phosphatase) used were purchased from Boehringer-Mannheim and used as recommended.

DNA sequencing. - DNA sequencing was performed by generating clones of the appropriate plasmid in M13 and sequencing the single stranded phage as previously described [9].

Preparation of [^{3}H]-poly(dG-dC)$_n$ containing Fapy residues. - The poly(dG-dC)$_n$ containing [^{3}H]-Fapy residues was prepared as previously described [12].

Bacterial strains. - HB101 and JM105 were used from laboratory stocks.

Preparation of crude lysates. - 20 ml of LB broth, containing 50 µg/ml ampicilin, was inoculated with 0.2 ml of overnight cultures and grown at 37 °C until the $O.D._{600nm}^{1\,cm}$ = 2.0. The cultures were centrifuged, and the pellet resuspended in 0.5 ml of lysis buffer (300 mM Tris-HCl pH 8.0, 5 mM Na_2EDTA). The cell suspension was lysed and centrifuged as previously described [9] to yield a crude lysate.

Enzymatic assay for Fapy-DNA glycosylase activity. - The assay for Fapy-DNA glycosylase activity has been previously described [9]. Briefly, [^{3}H]-Fapy-poly(dG-dC)$_n$(2000 cpm/assay) in 70 mM Hepes-KOH, pH 7.6, 100 mM KCl, 2 mM Na_2EDTA, 5% glycerol was incubated with limited amounts of crude lysate for 10 min at 37 °C. Following the incubation, an equal volume of carrier DNA and BSA solution was added and the ethanol soluble radioactivity determined. Dilution of the crude lysate in the reaction buffer was performed when the assay was saturated. Total soluble protein was determined and SDS polyacrylamide gel electrophoresis was performed on the crude lysates as previously described [9].

Overproduction of the Fapy-DNA glycosylase using IPTG. - JM105 cells hosting either the pFPG210 or 220 plasmid were grown at 37 °C until $O.D._{600nm}^{1\,cm}$ = 1.0. At this stage, the cultures were made 1 mM in IPTG, and grown for an additional 17 h at 37 °C.

Densitometry. - Densitometry was performed by directly scanning the protein gel using a Joyce-Loebel scanning densitometer. Peak areas were determined by weight.

Results

Plasmid construction

To overproduce the Fapy-DNA glycosylase, the fpg$^+$ gene of *E. coli* was cloned into a multicopy plasmid. After subcloning, a 1.7 kb (*SalI*/*Eco*RI) DNA fragment of the pFPG40 plasmid was isolated which contained 1.4 kb of *E. coli* DNA carrying the fpg$^+$ gene and two flanking regions of pBR322 DNA [9]. This 1.7 kb fragment was cloned into the pUC18 and 19 plasmids yielding pFPG50 and 60 (Fig. 1) [9]. The nucleotide sequence of the fpg$^+$ gene was determined and shown to consist of an 807 base pair open-reading frame [9]. The cloned Fapy-DNA glycosylase was purified to homogeneity. From this purified protein, the sequence of the 25 first amino acids at the NH_2-terminus was determined to be identical to the amino acid sequence predicted from the nucleotide sequence (Boiteux *et al.*, unpublished data). The last observation demonstrates that we have cloned the structural gene for the Fapy-DNA glycosylase.

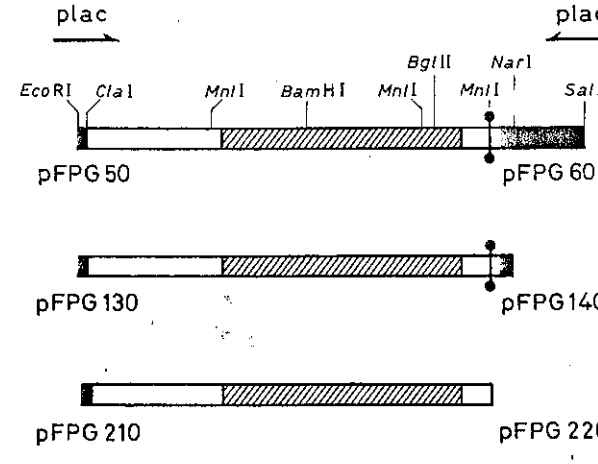

Fig. 1. - Structure of various plasmids containing the fpg$^+$ gene of *E. coli*. The left side of the figure lists the plasmids which have the fpg gene transcribed in the opposite direction from that of the lac promoter, whereas the right side of the figure lists the plasmids which have the fpg gene transcribed in the same direction as the lac promoter. Several of the important restriction sites are also indicated. The coding region for the fpg gene of *E. coli* is indicated by the cross-hatched area (▨). The black region (■) depicts DNA from pBR322 and the white region (□) indicates *E. coli* DNA. The inverted repeat structure just before the coding region is shown as a cruciform structure. Although this cruciform structure exists as isolated from the cell on the plasmid, we do not wish to imply that the DNA adopts only this structure.

The fpg+ gene was further subcloned by the creation of plasmids pFPG130 and 140 (Fig. 1) which were constructed by isolating the *EcoRI/SalI* fragment of pFPG60, restricting with *NarI*, and cloning the fragment into the *Eco*RI/*Acc*I site of pUC18 and 19.

The final two subclones shown in Fig. 1, pFPG210 and 220, were constructed by isolating the 290 bp *Bgl*II/*Hind*III fragment of pFPG60, restricting with *MnlI* and then cloning the fragment of 190 bp into the *Bgl*II/*Hind*II cut pFPG50 or 60.

Overproducer character of plasmids containing the fpg gene

For each of the plasmids in Fig. 1, the amount of the Fapy-DNA glycosylase produced in HB101 was estimated by the glycosylase activity assay and the intensity of the 31 kDa band visualized by SDS-electrophoresis of crude lysates. The results of the assays for each of the plasmids are shown in Figs 2 and 3. Fig. 2 shows that cells hosting pFPG50 and 60 exhibited approximately 60- and 70-fold more enzyme activity than the HB101 hosting only pUC19. In concordance with the increase in glycosylase activity, the SDS polyacrylamide gel of the soluble crude lysate proteins in Fig. 3 shows a new band at 31 kDa, which is the same molecular weight as the Fapy-DNA glycosylase. Sequencing showed that transcription from the lac promoter and the fpg promoter are in the same direction in the pFPG60 plasmid. Surprisingly, we do not observe a significant effect of the orientation on the fpg gene expression when the gene in pFPG60 is in the same orientation as the lac promoter. Previously we also observed that only a modest increase (~2-fold) in Fapy-DNA glycosylase production occurs when JM105 cells hosting the pFPG60 plasmid are incubated with IPTG. These two

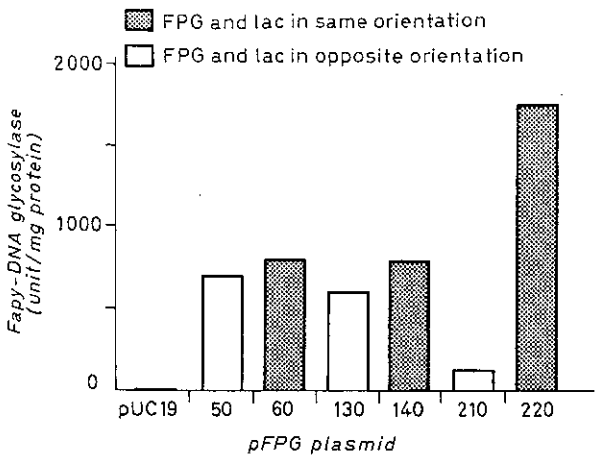

Fig. 2. - Production of the Fapy-DNA glycosylase in HB101 cells hosting different plasmids by the glycosylase assay. The production is indicated as Fapy-DNA glycosylase units/mg of protein in the crude lysate. The plasmid constructions are indicated in Fig. 1 and further details are given in the "Materials and methods". One unit of Fapy-DNA glycosylase enzyme released 1 pmol of Fapy in 5 min at 37 °C.

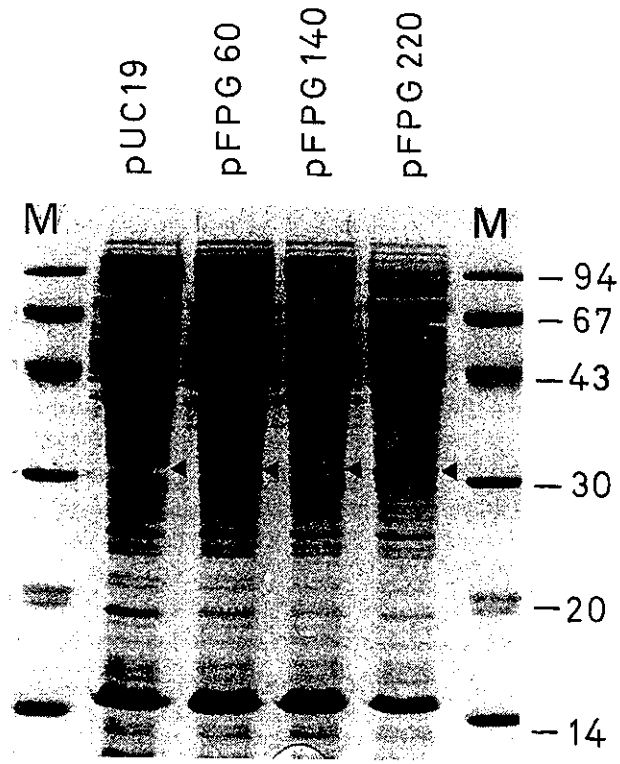

Fig. 3. - Production of the Fapy-DNA glycosylase in HB101 cells hosting different plasmids by SDS polyacrylamide gel electrophoresis of soluble crude protein lysates. Lanes 1 and 6 are control lanes with molecular weight markers. The plasmids are indicated at the top of the figure. The location of the Fapy-DNA glycosylase is shown by an arrow. The dark band at the bottom of the gel corresponds to lysozyme.

observations suggest that transcription from the lac promoter does not contribute significantly to the expression of the fpg+ gene.

Since the lac promoter is located upstream from the putative fpg promoter in pFPG60, we hypothesized that some sequence in the ~350 bp separating the two promoters was responsible for the attenuation of the expression (Fig. 4). A first deletion was performed to eliminate the possibility that a sequence of pBR322 DNA between *Nar*I and *Sal*I (positions 413 and 650, respectively, from pBR322) caused the lower expression of the fpg gene (Fig. 1). The deletion of the *NarI/SalI* fragment to form pFPG130 and 140, does not significantly alter the fpg gene expression compared to that observed using the parental pFPG50 and 60 as shown in Figs 2 and 3. Therefore, these two observations suggest that the lower expression levels are most likely not associated with pBR322 DNA sequences.

Sequence analysis of the 100 bp of *E. coli* DNA immediately upstream from the fpg coding region showed that there is a palindromic sequence just before the putative -35 fpg promoter sequence as shown in Fig. 4. The two plasmids which lack half of the palindromic sequence, pFPG210 and 220, eliminate the formation of possible secondary structures in DNA or RNA which involve the inverted repeat sequence (Figs 1 and 4). Analysis of the glycosylase activity in Fig. 2 shows

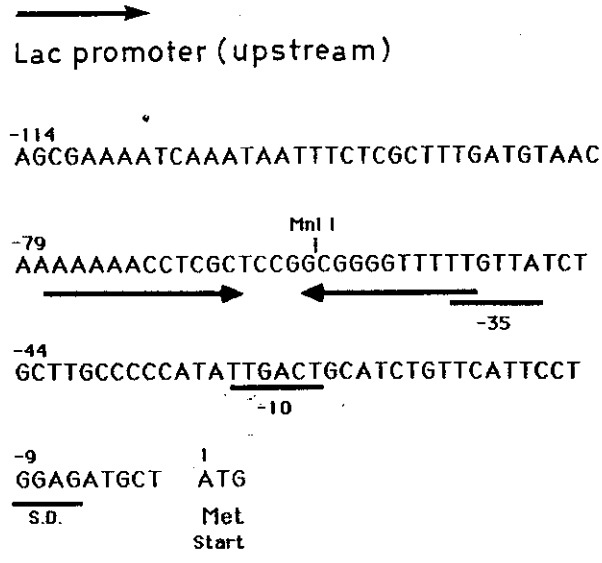

Fig. 4. - DNA sequence of the fpg promoter region of the fpg gene from *E. coli* in the pFPG60 plasmid. The sequence of pBR322 consisting of approximately 275 bp and the lac promoter region just before the pBR322 sequence are not shown. The putative promoter (-35 and -10) regions, the Shine-Dalgarno (S.D.) region, the *MnlI* cleavage site, and the direction of transcription of the lac promoter are indicated. The arrows designate the inverted repeat sequence. The Methionine (Met) codon for the start of the fpg gene is also shown.

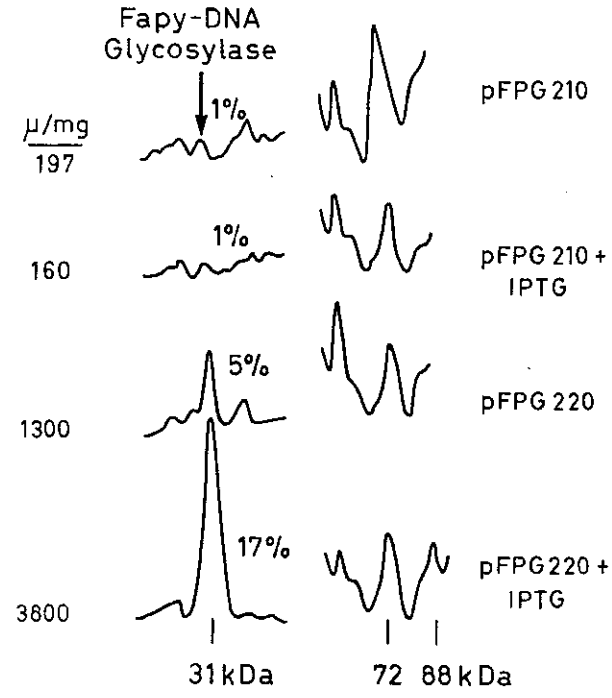

Fig. 5. - Densitometric scans of the polyacrylamide gels for the overproduction of the Fapy-DNA glycosylase following treatment with IPTG in JM105. The portions of the scans which indicate the overproduction of the Fapy-DNA glycosylase are labeled. Another region which maintains a relatively constant amount of protein is shown for comparison. The percentage of Fapy-DNA glycosylase compared to the total soluble cellular proteins in the crude lysates as estimated from densitometry and the number of Fapy-DNA glycosylase units per milligram of total protein are also indicated in the figure. The molecular weight range of a region with constant levels of protein bands are indicated in kilodaltons.

that the activity is 140-fold greater than the background in cells hosting the pFPG220 and only 10-fold above background in cells containing pFPG210. In fact, removing the sequence with the inverted repeat increases the expression of the fpg gene in cells hosting pFPG220 to such an extent that the Fapy-DNA glycosylase is one of the major soluble cellular proteins as suggested from Fig. 3. These results suggest that the inverted repeat sequence was responsible for the attenuation of the fpg gene expression from the lac promoter. In addition, the expression of the fpg$^+$ gene without the lac promoter is higher when the palindromic sequence is present as indicated by a comparison of the Fapy-DNA glycosylase level observed for the pFPG50 and 210 (Fig. 2). The fpg promoter in pFPG210 does not appear to be deleted as suggested by the nucleotide sequence and the fact cells harboring pFPG210 still overproduce the Fapy-DNA glycosylase 10-fold.

Overexpression of the Fapy-DNA glycosylase

Since the presence of the pFPG220 plasmid in HB101 cells has a deleterious effect on cell growth, we transformed JM105 cells harboring a lac I^q mutation with pFPG220. In order to further stimulate the production of the Fapy-DNA glycosylase, we added IPTG, a non-metabolizable inducer of the lac promoter, to cells in log phase growth. These cells expressed the Fapy-DNA glycosylase at very high levels as demonstrated by the densitometric scans of the polyacrylamide gels as shown in Fig. 5. Without IPTG the Fapy-DNA glycosylase is overproduced as indicated by the band at 31 kDa, but when IPTG is added, the Fapy-DNA glycosylase is 17% of the total soluble proteins.

Discussion

In this paper we have shown that a naturally occurring, inverted repeat DNA sequence of *E. coli* located upstream from the fpg promoter has a regulatory effect on the expression of the fpg gene when it is cloned in the pUC18 and 19 systems as i) a strong attenuator of the expression when the lac promoter is used for transcription and ii) a possible stimulator of the fpg promoter. If there is another gene transcribed in the same direction as the fpg gene, this palindromic sequence may act as a transcription terminator of the previous gene's RNA synthesis. This may be a reasonable explanation if the preceding gene is transcribed at a high level. Alternatively, the effect of the inverted repeat sequence on the gene expression may be the result of the formation of a cruciform, which is a non-B DNA secondary structure [13, 14]. In fact, the pFPG60 plasmid as isolated from the cell has an S1 hypersensitive site which maps to the

palindrome (T. O'Connor, unpublished data), which is one of the criteria for the existence of a cruciform structure [13, 14]. Moreover, we hypothesize that the formation of a cruciform structure could inhibit transcription while under control of the lac promoter and stimulate transcription from the native fpg promoter. The stimulation of transcription may result from the formation of an entry point for RNA polymerase. However, the assessment of the precise role of the palindromic sequence on the fpg promoter in this plasmid await the mapping of the transcription start for the fpg gene and analysis of the *in vitro* transcripts from the different plasmids.

In conclusion, JM105 cells hosting the pFPG220 plasmid showed an increase in Fapy-DNA glycosylase content following treatment with IPTG to such an extent that this protein becomes the major soluble cellular protein. This overproduction of the Fapy-DNA glycosylase should be important in the isolation of milligram quantities of pure protein for the study of catalytic properties and three dimensional structure.

Acknowledgements

We are very grateful to Patricia Auffret Van der Kemp, Claudine Lagravere, and Florence Lopès for their excellent technical assistance. This work was supported by the Centre National de la Recherche Scientifique (CNRS) LA147 and the Institut National de la Santé et de la Recherche Medicale (INSERM) U140, and grants from the Association pour la Recherche sur le Cancer (ARC). T.R. O'Connor was the recipient of international fellowships from INSERM, ARC, and the Ligue Nationale contre le Cancer.

Review submitted on invitation by the Editorial Board of the *Annali*. Accepted for publication: October 1988.

REFERENCES

1. SINGER, B. & GRUNBERGER, D. 1983. *Molecular biology of mutagens and carcinogens.* Plenum Press, New York. pp. 55-78.

2. BOITEUX, S. & LAVAL, J. 1983. Imidazole ring-opened 7-methylguanine: an inhibitor of DNA synthesis. *Biochem. Biophys. Res. Commun.* **110**: 552-558.

3. LARSON, K., SHAM, J., SHENKAR, R. & STRAUSS, B. 1985 Methylation-induced blocks to *in vitro* DNA replication. *Mutat. Res.* **150**: 77-84.

4. O'CONNOR, T.R., BOITEUX, S. & LAVAL, J. 1988. Formamido ring-opened pyrimidines are a block to DNA synthesis. *J. Cell. Biochem.* **12A**(Suppl.): 270.

5. HAINES, J.A., REESE, C.B. & LORD TODD, J. 1962. The methylation of guanosine and related compounds with diazomethane. *J. Chem. Soc.* 5281-5288.

6. KARRAN, P., HJELMGREN, T. & LINDAHL, T. 1982. Induction of a DNA-glycosylase for N-methylated purines is part of the adaptive response to alkylating agents. *Nature* **296**: 770-773.

7. CHETSANGA, C.J. & LINDAHL, T. 1979. Release of 7-methylguanine whose imidazole rings have been opened from damaged DNA by a DNA glycosylase from *Escherichia coli*. *Nucleic Acids Res.* **6**: 3673-3683.

8. CHETSANGA, C.J., LOZON, M., MAKAROFF, C. & SAVAGE, L. 1981. Purification and characterization of *E. coli* formamidopyrimidine-DNA glycosylase that excises damaged 7-methylguanine from DNA. *Biochemistry* **20**: 5201-5207.

9. BOITEUX, S., O'CONNOR, T.R. & LAVAL, J. 1987. Formamidopyrimidine-DNA glycosylase of *Escherichia coli*: cloning and sequencing of the fpg structural gene and overproduction of the protein. *EMBO J.* **6**: 3177-3183.

10. MARGISON, G.P. & PEGG, A.E. 1981. Enzymatic release of 7-methylguanine from methylated DNA by rodent liver extracts. *Proc. Natl. Acad. Sci. USA* **78**: 861-865.

11. MANIATIS, T., FRITSCH, E.F. & SAMBROOK, J. 1982. *Molecular Cloning.* CSH Press, Cold Spring Harbor, New York.

12. BOITEUX, S., BELLENEY, J., ROQUES, B.P. & LAVAL, J. 1984. Two rotameric forms of ring-open 7-methylguanine are present in alkylated polynucleotides. *Nucleic Acids Res.* **6**: 3673-3683.

13. PANAYATATOS, N. & WELLS, R.D. 1981. Cruciform structures in supercoiled DNA. *Nature* **289**: 466-470.

14. LILLEY, D.M.J. 1980. The inverted repeat as a recognisable structural feature in supercoiled DNA molecules. *Proc. Natl. Acad. Sci. USA* **77**: 6468-6472.