Location of the *fpg* gene on the *Escherichia coli* chromosome

Struan H.Coleman and Donald G.Wild
Microbiology Unit, Department of Biochemistry, Oxford University, South Parks Road, Oxford OX1 3QU, UK

Submitted May 15, 1991 GenBank accession no. J01677

The *fpg* gene of *E.coli* specifies a DNA glycosylase that excises a ring-opened form of N7-methylguanine (2-6-diamino-4-hydroxy-5-N-methylformamidopyrimidine [Fapy]) from the chromosome. This gene has been isolated, on plasmids that overproduce the Fapy-DNA glycosylase activity, and sequenced; the protein has been purified to physical homogeneity (1, 2). However, no *fpg* mutants are known and the gene has not been positioned on the genetic map; an attempt (3) to localize *fpg* on the restriction map for *E.coli* (4) gave ambiguous results.

The *rpmB,G* operon codes for ribosomal proteins L28 and L33. There is virtually complete (97.2%) homology between the 3′ end of the published sequence for the *rpmB,G* operon (5) and the 5′ end of the sequence for *fpg* (1). The overlapping homologous sequences (Figure 1) are of 145 nucleotides and include the 3′ end of the coding sequence for L33, the putative promoter sites for transcription of *fpg*, the Shine-Dalgarno sequence and the first 28 nucleotides of the *fpg* coding sequence. The two coding sequences are also separated by an inverted repeat that, in *lac* fusions, blocks transcription of *fpg* from an upstream promoter (2). The four differences in sequence are within this palindrome; these may reflect strain differences and/or early difficulties in reading sequences with a reiterated base.

The *fpg* gene is thus contiguous with the *rpmB,G* operon, at 81.7 minutes on the chromosome (6) and like the latter is transcribed counterclockwise. This location is consistent with the physical map. The BamHI cleavage site immediately 3′ to *rpmB,G* (5) is (using the numbering of ref. 3) at 3882.55 kb and flanked by two BglI sites; these three sites occur within *fpg* in their correct sequence. There are no restriction sites for the other six enzymes used to construct the physical map within the combined 1.7 kb *rpmB,G-fpg* sequence or in this section of the map.

The clockwise order of known genes in this region of the *E.coli* chromosome is now *radC-mutM-pcsA-fpg-rpmG,B-dfp-dut-pyrE-spoT*. Thus *fpg* joins a cluster of genes also involved in DNA transactions; these are separated by *rpmB,G* from other genes concerned with nucleotide metabolism.

REFERENCES

1. Boiteux,S., O'Connor,T.R. and Laval,J. (1987) *EMBO J.* **6**, 3177–3183.
2. O'Connor,T.R., Boiteux,S. and Laval,J. (1989) *Ann. Ist. Super. Sanita* **25**, 27–32.
3. Médigue,C., Bouché,J.P., Hénaut,A. and Danchin,A. (1990) *Mol. Microbiol.* **4**, 169–187.
4. Kohara,Y., Akiyama,K. and Isono,K. (1987) *Cell* **50**, 495–508.
5. Lee,J.S., An,G., Friesen,J.D. and Isono,K. (1981) *Mol. Gen. Genet.* **184**, 218–223.
6. Bachmann,B.J. (1990) *Microbiol. Rev.* **54**, 130–197.

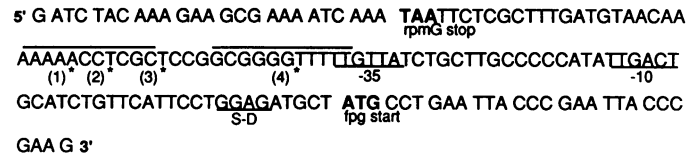

Figure 1. The sequence shown is the 5′ end of the published *fpg* sequence (1). The sequence extending from the 3′ end of *rpmG* (5) differs in the four positions indicated by the asterisks with (1) an A deleted, (2) and (3) a T changed to a C and (4) a T inserted. The palindromic sequence is overlined. The 3′ end of the coding sequence for *rpmG* and the start of that for *fpg* are indicated in bold type. The putative promoter and Shine-Dalgarno sequences for *fpg* are underlined and labelled.