A molecular analysis of the 53·3 minute region of the *Escherichia coli* linkage map

SIMON C. ANDREWS, PAULINE M. HARRISON and JOHN R. GUEST*

Krebs Institute for Biomolecular Research, Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield S10 2TN, UK

(Received 13 August 1990; revised 10 October 1990; accepted 25 October 1990)

The coding characteristics of four plasmids expressing a protein (BCP) which comigrates with bacterioferritin were examined and the nucleotide sequence of a common 1985 bp segment from the 53 min region of the *Escherichia coli* linkage map was determined. Three open reading-frames (*orf1*, *orf2* and *orf3*) were detected, and *orf2* (*bcp*, 156 amino acid codons) appeared to encode the bacterioferritin comigratory protein, BCP. The translation product of *orf3* (205 amino acid codons) resembled the iron-sulphur protein component (DMS B subunit) of the anaerobic dimethylsulphoxide reductase complex of *E. coli*.

Introduction

The 'Gene-Protein Index of *Escherichia coli* K-12' is an invaluable source of information correlating the two-dimensional gel coordinates of 800 discrete proteins with the linkage map locations of the corresponding genes (Neidhardt *et al.*, 1983). The proteins expressed from 275 plasmids of the Clarke-Carbon ColE1-*E. coli* gene library (Clarke & Carbon, 1976) were analysed, and the linkage map positions were identified for 130 plasmids.

One of the standard proteins used to calibrate the two-dimensional gel map was bacterioferritin (BFR), an iron-storage haemoprotein (Yariv *et al.*, 1981). Two plasmids carrying inserts assigned to different regions of the linkage map, pLC25-14 (53 min) and pLC46-7 (83 min), were shown to express proteins that comigrate with bacterioferritin (BFR) and to have peptide maps like BFR (Neidhardt *et al.*, 1983). During attempts to clone the bacterioferritin gene (*bfr*) it was realized that both plasmids contain inserts from the 53 min region and that they encode a BFR comigratory protein (BCP) rather than BFR: the *bfr* gene has since been located at 73 min (Andrews *et al.*, 1989). These studies, including the

nucleotide sequence of a 1985 bp segment containing the putative *bcp* gene, are reported here.

Methods

Bacterial strains, plasmids and bacteriophages. The strains of *E. coli* K12 were: JM101 [*thi supE* Δ (*proAB-lac*) *F'* *traD36 proA⁺B⁺ lacI^qZ* Δ M15], the host for M13 and pUC derivatives (Messing, 1983); and a JA200 derivative containing pLC25-14 (Clarke & Carbon, 1976), kindly supplied by Dr B. J. Bachmann. Phagemid vectors pUC118 and pUC119 (Vieira & Messing, 1987), were used for subcloning specific subfragments of pLC25-14 to generate the series of derivatives shown in Fig. 1: pGS211, pGS212, pGS213, pGS339, pGS340 and pGS370. M13mp18 and M13mp19 were used for subcloning and preparing templates for DNA sequencing. Bacteriophage and plasmid DNA was isolated and manipulated according to Sambrook *et al.* (1989).

Expression of plasmid DNA in vitro and analysis of translation products. Plasmid DNA was expressed in the presence of L-[³⁵S]methionine, using a DNA-directed *in vitro* transcription-translation kit (Amersham). The L-[³⁵S]methionine-labelled translation products were analysed by two-dimensional PAGE (O'Farrell, 1975) and autoradiography of dried gels: ¹⁴C-methylated *M_r* markers (Amersham) were used for calibration.

DNA sequence analysis. The source of DNA was primarily pGS211, and the strategy involved cloning random (*Msp*I or *Hinc*PI) and specific fragments, into M13mp18 and M13mp19. Single-stranded M13 templates were sequenced using either Sequenase or the Klenow fragment of DNA polymerase I (Sanger *et al.*, 1980; Tabor & Richardson, 1987; Biggin *et al.*, 1983) and 'universal' or specific oligonucleotide primers. Plasmid DNA (pGS339, pGS340 and pGS370) was sequenced directly using the double-stranded sequencing method (Zhang *et al.*, 1988). Nucleotide sequences were compiled and analysed with the aid of computer programs (Staden, 1982, 1984). Sequence databases were searched using the VAX computer facility at the Daresbury Laboratory.

Abbreviations: BFR, bacterioferritin; BCP, bacterioferritin comigratory protein.

The nucleotide sequence reported in this paper has been submitted to GenBank and given the accession number M37689.

Results and Discussion

Location of the orf encoding the bacterioferritin comigratory protein

The two plasmids thought to encode bacterioferritin, pLC25-14 (53 min) and pLC46-7 (83 min), each express a protein now designated the bacterioferritin comigratory protein (BCP) and a set of five additional proteins that give the same pattern of mobilities in two-dimensional PAGE (Neidhardt *et al.*, 1983). Two other plasmids bearing inserts from the 53 min region, pLC43-32 and pLC17-30, express the same set of five proteins, and additionally, pLC17-30 and pLC25-14 each express a further comigrating set of five proteins (Table 1). This analysis clearly indicates that all four pLC plasmids possess overlapping segments of DNA from the 53 min region, although pLC46-7 could still contain some additional DNA from the 83 min region. A *purC*⁺ plasmid (pSIU103) containing part of the 53 min region also expresses five proteins with mobilities that match some of the pLC-encoded proteins (Parker, 1984). Two of these proteins have been identified as products of the *purC* (Parker, 1984) and *dapA* (Richaud *et al.*, 1986) genes, and a third migrated like BCP (Table 1). It is important to note that the vertical arrangement of the groups of proteins boxed in Table 1 reflects the linkage map of the corresponding genes.

A restriction map constructed for pLC25-14 was entirely consistent with that of the 2595.5 to 2616.0 kb (53 min) region of the complete physical map of the *E. coli* chromosome (Kohara *et al.*, 1987), and a series of pUC derivatives was prepared in order to locate the putative *bcp* gene (Fig. 1). A two-dimensional electrophoretic analysis of the proteins expressed from these plasmids confirmed that the BFR co-migratory protein detected by Neidhardt *et al.* (1983) is encoded by the 4.1 kb subfragment, E-B₂ in Fig. 1 (Table 1).

Nucleotide sequence of a 1985 bp segment in the 53 min region

If BCP is uniquely encoded, rather than being a product of overlapping genes, it should be located immediately clockwise of the *purC*-*dapA* region (Fig. 1). The nucleotide sequence of the 1.85 kb *Bcl*I-*Bam*HI segment (Bc₁-B₂ in Fig. 1) was determined, and subsequently extended to include the adjacent 0.1 kb *Bam*HI-*Hind*III fragment in order to complete an open reading-frame that spans the *Bam*HI site. The sequence is shown in Fig. 2. It overlaps by 96 bp a previously sequenced *dapA* region (Richaud *et al.*, 1986) and the common segment is identical except for a C→G substitution at position 28.

Table 1. Comparison of proteins expressed by plasmids carrying inserts from the 53 min region

The proteins observed by two-dimensional PAGE analysis (X) are listed according to their expression by different plasmids. The linkage relationships of the corresponding genes are reflected in the vertical alignment of the boxes which enclose products in each group.

Gene-protein index		pLC plasmids*				pSIU103*		Protein identity
Coordinates	10 ⁻³ × M _r	43-32 <i>purC</i>	17-30 <i>dapE</i>	25-14 <i>purC</i>	46-7	pGS211 pGS212 <i>purC</i> , <i>dapA</i>	pGS213 <i>dapA</i>	
Other proteins		.	.	5X	.	.	.	
99 × 46	19	.	X	X	.	.	.	
89 × 51	22	.	X	X	.	.	.	
68 × 52	22	.	X	X	.	.	.	
66 × 72	32	.	X	X	.	.	.	
105 × 79	36	.	X	X	.	.	.	
92 × 64	27	X	X	X	X	X	.	PurC
97 × 66	28	X	X	X	X	X	X	
58 × 72	32	X	X	X	X	X†	X†	DapA
41 × 73	32	X	X	X	X	X	X	
100 × 78	35	X	X	X	X	X	X	
92 × 30	15‡	.	.	X	X	X	X	BCP
Other proteins		.	.	.	25X	.	.	

* Data from Neidhardt *et al.* (1983) and Parker (1984).

† This protein could have coordinates 41 × 73 or 58 × 72.

‡ Comigrates with bacterioferritin (M_r 18495; Andrews *et al.*, 1989).

Fig. 1. Restriction map for the segment of bacterial DNA in pLC25-14 and the structures of several pUC derivative plasmids. The positions and polarities of previously characterized and unidentified genes, *purC*, *dapA* and *urf* (Parker, 1984; Richaud *et al.*, 1986), are indicated, as are those of *orf1*, 2 (*bcp*), and 3. Plasmid pSIU103 is a pBR322 derivative that contains the same insert as pGS211 and pGS212 (Parker, 1984). Open bars represent bacterial DNA and single lines represent vector DNA. The polarities of the *lac* promoters in pUC vectors are indicated by arrows. Relevant restriction sites are abbreviated as follows: *Bam*HI, B; *Bcl*I, Bc; *Eco*RI, E; *Hind*III, H; *Pst*I, P.

The sequence contains three potential coding regions identified by FRAMESCAN (Staden, 1984): they have codon preferences consistent with weak (*orf1*), moderate (*orf2*) and weak-moderate (*orf3*) expression (Figs 1 and 2). The stop codon of *orf3* is overlapped by another *orf*, *orf4*, which extends to the end of the sequenced region (Fig. 2) and suggests that *orf3* and *orf4* are translationally coupled. A computer search revealed putative promoters, binding sites for CRP (Busby, 1986), FNR (Spiro & Guest, 1990), and FUR (Griggs & Konisky, 1989), and a terminator for *orf2* ($\Delta G -21.6 \text{ kJ mol}^{-1}$ for the stem-loop). These are indicated in Fig. 2, but they have not been confirmed experimentally.

Features of the amino acid sequences

The translation products of *orf1* (184 amino acid residues, M_r 20454) and *orf2* (156 amino acid residues, M_r 17634) bear no significant similarity to BFR or any sequences in available databases. The *orf2* product most closely matches BCP with respect to size and degree of expression, but the identity of BCP remains obscure. The *orf3* product, ORF3 (205 amino acid residues, M_r 22154), resembles the iron-sulphur protein subunit (DMS B) of the anaerobic dimethylsulphoxide reductase of *E. coli* (Bilous *et al.*, 1988) and, to a

lesser extent, other proteins containing ferredoxin-like cysteine clusters. The ORF3 and DMS B amino acid sequences are optimally aligned in Fig. 3: 28% of the 167 equivalenced residues are identical and the similarity rises to 49% when conservative substitutions are included. Both sequences contain 16 cysteine residues arranged in four clusters, as shown in Fig. 3. The cysteine clusters closely resemble iron-sulphur centres of the [3Fe-4S] and [4Fe-4S] type (Beinert, 1990). Clusters I and III share a high level of sequence similarity, as do clusters II and IV, and this raises the possibility that ORF3 has evolved by duplication of a gene encoding a ferredoxin-like ancestral protein with two cysteine clusters.

The ORF4 polypeptide has a high proportion of hydrophobic amino acids amongst the first 34 residues. It is therefore possible that ORF3 functions as an electron-transfer subunit of a multicomponent enzyme complex in which ORF4 may be a membrane-intrinsic subunit anchoring the catalytic subunits to the cytoplasmic membrane, as in dimethylsulphoxide reductase, fumarate reductase and succinate dehydrogenase (Bilous *et al.*, 1988; Darlison & Guest, 1984). The relevant region of the linkage map was formerly thought to encode an NADH-nitrite reductase activity, but the corresponding locus (*nirF*) has now been removed because the defect appears to be in anaerobic glucose metabolism rather than specifically in nitrite reduction (Bachmann, 1990; MacDonald *et al.*, 1985).

Fig. 3. Alignment of the amino acid sequences of ORF3 and DMS B. Asterisks signify identical residues and colons conserved substitutions scoring ≥ 0.1 in the MDM₇₈ mutation data matrix (Schwartz & Dayhoff, 1978). Hyphens denote insertions used to optimize the alignment. The cysteine clusters are boxed and labelled I to IV.

We thank B. J. Bachmann for providing the pLC plasmids and R. E. Roberts for technical assistance. This work was supported by a grant from the Wellcome Trust.

References

- ANDREWS, S. C., HARRISON, P. M. & GUEST, P. M. (1989). Cloning, sequencing, and mapping of the bacterioferritin gene (*bfr*) of *Escherichia coli* K-12. *Journal of Bacteriology* **171**, 3940-3947.
- BACHMANN, B. J. (1990). Linkage map of *Escherichia coli* K-12, edition 8. *Microbiological Reviews* **54**, 130-197.
- BEINERT, H. (1990). Recent developments in the field of iron-sulfur proteins. *The FASEB Journal* **4**, 2483-2491.
- BIGGIN, M. D., GIBSON, T. J. & HONG, G. F. (1983). Buffer gradient gels and ³⁵S label as an aid to rapid DNA sequence determination. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 3963-3965.
- BLOUS, P. T., COLE, S. T., ANDERSON, W. F. & WEINER, J. H. (1988). Nucleotide sequence of the *dmsABC* operon encoding the anaerobic dimethylsulphoxide reductase of *Escherichia coli*. *Molecular Microbiology* **2**, 785-795.
- BUSBY, S. (1986). Catabolite repression and related processes. *Symposia of the Society for General Microbiology* **39**, 51-77.
- CLARKE, L. & CARBON, J. (1976). A colony bank containing synthetic ColE1 hybrid plasmids representative of the entire *E. coli* genome. *Cell* **9**, 91-99.
- DARLISON, M. G. & GUEST, J. R. (1984). Nucleotide sequence encoding the iron-sulphur protein subunit of the succinate dehydrogenase of *Escherichia coli*. *Biochemical Journal* **223**, 507-517.
- GRIGGS, D. W. & KONISKY, J. (1989). Mechanism for iron-regulated transcription of the *Escherichia coli* *cir* gene: metal-dependent binding of Fur protein to the promoters. *Journal of Bacteriology* **171**, 1048-1054.
- KOHARA, Y., AKIYAMA, K. & ISONO, K. (1987). Physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* **50**, 495-508.
- MACDONALD, H., POPE, N. R. & COLE, J. A. (1985). Isolation, characterization and complementation analysis of *nirB* mutants of *Escherichia coli* deficient only in NADH-dependent nitrite reductase activity. *Journal of General Microbiology* **131**, 2771-2782.
- MESSING, J. (1983). New M13 vectors for cloning. *Methods in Enzymology* **101**, 20-78.
- NEIDHARDT, F. C., VAUGHN, V., PHILLIPS, T. A. & BLOCH, P. L. (1983). Gene-protein index of *Escherichia coli* K-12. *Microbiological Reviews* **47**, 231-284.
- O'FARRELL, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. *Journal of Biological Chemistry* **250**, 4007-4021.
- PARKER, J. (1984). Identification of the *purC* gene product of *Escherichia coli*. *Journal of Bacteriology* **157**, 712-717.
- RICHAUD, F., RICHAUD, C., RATET, P. & PATTE, J.-C. (1986). Chromosomal location and nucleotide sequence of the *Escherichia coli* *dapA* gene. *Journal of Bacteriology* **166**, 297-300.
- SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. (1989). *Molecular Cloning. A Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- SANGER, F., COULSON, A. R., BARRELL, B. G., SMITH, A. J. H. & ROE, B. A. (1980). Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *Journal of Molecular Biology* **143**, 161-178.
- SCHWARTZ, R. M. & DAYHOFF, M. O. (1978). Matrices for detecting distant relationships. In *Atlas of Protein Sequence and Structure*, vol. 5, supplement 3, pp. 353-358. Edited by M. O. Dayhoff. Washington, DC: National Biomedical Research Foundation.
- SPIRO, S. & GUEST, J. R. (1990). FNR and its role in oxygen-regulated gene expression in *Escherichia coli*. *FEMS Microbiology Reviews* **75**, 399-428.
- STADEN, R. (1982). An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. *Nucleic Acids Research* **10**, 2951-2961.
- STADEN, R. (1984). Graphic methods to determine the function of nucleic acid sequences. A summary of ANALYSEQ options. *Nucleic Acids Research* **12**, 521-538.
- TABOR, S. & RICHARDSON, C. C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 4767-4771.

- VIEIRA, J. & MESSING, J. (1987). Production of single-stranded plasmid DNA. *Methods in Enzymology* **153**, 3–34.
- YARIV, J., KALB, A. J., SPERLING, R., BAUMINGER, E. R., COHEN, S. G. & OFER, S. (1981). The composition and structure of bacterioferritin of *Escherichia coli*. *Biochemical Journal* **197**, 171–175.
- ZHANG, H., SCHOLL, R., BROWSE, J. & SOMERVILLE, C. (1988). Double stranded DNA sequencing as a choice for DNA sequencing. *Nucleic Acids Research* **16**, 1220.