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 Shefield, 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 ColEl–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 [thr supE ΔproAB–lac] F’traD36 proA+ B+ lacZ ΔM15], 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 (Amerham). The L-[35S]methionine-labelled translation products were analysed by two-dimensional PAGE (O’Farrell, 1975) and autoradiography of dried gels: 14C-methylated M, markers (Amersham) were used for calibration.

DNA sequence analysis. The source of DNA was primarily pGS211, and the strategy involved cloning random (MspI or HincPI) 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 $\text{purC}^+$ 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 $\text{purC}$ (Parker, 1984) and $\text{dapA}$ (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 $\text{E. coli}$ chromosome (Kohara et al., 1987), and a series of pUC derivatives was prepared in order to locate the putative $\text{bep}$ 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 $\text{purC}$–$\text{dapA}$ region (Fig. 1). The nucleotide sequence of the 1·85 kb $\text{Bcl}–\text{BamHI}$ segment (Bc1–B2 in Fig. 1) was determined, and subsequently extended to include the adjacent 0·1 kb $\text{BamHI}$–$\text{HindIII}$ fragment in order to complete an open reading-frame that spans the $\text{BamHI}$ site. The sequence is shown in Fig. 2. It overlaps by 96 bp a previously sequenced $\text{dapA}$ 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.

<table>
<thead>
<tr>
<th>Gene–protein index</th>
<th>pLC plasmids*</th>
<th>pSIU103*</th>
<th>pGS211</th>
<th>pGS212</th>
<th>pGS213</th>
<th>Protein identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coordinates</td>
<td>10⁻³ × $\text{M}_r$</td>
<td>43-32 purC</td>
<td>17-30 dapE</td>
<td>25-14 purC</td>
<td>46-7</td>
<td></td>
</tr>
<tr>
<td>Other proteins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99×46</td>
<td>19</td>
<td>.</td>
<td>X</td>
<td>X</td>
<td>.</td>
<td>PurC</td>
</tr>
<tr>
<td>89×51</td>
<td>22</td>
<td>.</td>
<td>X</td>
<td>X</td>
<td>.</td>
<td></td>
</tr>
<tr>
<td>68×52</td>
<td>22</td>
<td>.</td>
<td>X</td>
<td>X</td>
<td>.</td>
<td></td>
</tr>
<tr>
<td>66×72</td>
<td>32</td>
<td>.</td>
<td>X</td>
<td>X</td>
<td>.</td>
<td></td>
</tr>
<tr>
<td>105×79</td>
<td>36</td>
<td>.</td>
<td>X</td>
<td>X</td>
<td>.</td>
<td></td>
</tr>
<tr>
<td>92×64</td>
<td>27</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
</tr>
<tr>
<td>97×66</td>
<td>28</td>
<td>.</td>
<td>X</td>
<td>X</td>
<td>.</td>
<td>DapA</td>
</tr>
<tr>
<td>58×72</td>
<td>32</td>
<td>.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>41×73</td>
<td>32</td>
<td>.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>100×78</td>
<td>35</td>
<td>.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>92×30</td>
<td>15‡</td>
<td>.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>BCP</td>
</tr>
<tr>
<td>Other proteins</td>
<td></td>
<td>.</td>
<td>X</td>
<td>X</td>
<td>.</td>
<td></td>
</tr>
</tbody>
</table>

* Data from Neidhardt et al. (1983) and Parker (1984).
† This protein could have coordinates 41×73 or 58×72.
‡ Comigrates with bacterioferritin ($\text{M}$, 18495; Andrews et al., 1989).
The E. coli linkage map at 53.3 min

---

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 (ΔG = 21.6 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, 20454) and orf2 (156 amino acid residues, M, 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, 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. 2. Nucleotide sequence of a 1985 bp segment in the 53 min region showing translations of four open reading-frames. An adjacent sequence containing the start codon of \textit{dapA} (−274 to −97; Richaud \textit{et al.}, 1986), is included to establish continuity. Potential ribosome-binding sites (RBS, boxed), translational initiation sites (underlined) and stop sites (asterisks) are indicated, as are putative promoters (P), CRP, FNR and FUR sites, and a region of hyphenated dyad symmetry (converging arrows).
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


