

## Cloning, Sequencing, and Mapping of the Bacterioferritin Gene (*bfr*) of *Escherichia coli* K-12

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

*The Krebs Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield S10 2TN, United Kingdom*

Received 24 February 1989/Accepted 13 April 1989

The bacterioferritin (BFR) of *Escherichia coli* K-12 is an iron-storage hemoprotein, previously identified as cytochrome *b*<sub>1</sub>. The bacterioferritin gene (*bfr*) has been cloned, sequenced, and located in the *E. coli* linkage map. Initially a gene fusion encoding a BFR- $\lambda$  hybrid protein ( $M_r$  21,000) was detected by immunoscreening a  $\lambda$  gene bank containing *Sau*3A restriction fragments of *E. coli* DNA. The *bfr* gene was mapped to 73 min (the *str-spc* region) in the physical map of the *E. coli* chromosome by probing Southern blots of restriction digests of *E. coli* DNA with a fragment of the *bfr* gene. The intact *bfr* gene was then subcloned from the corresponding  $\lambda$  phage from the gene library of Kohara et al. (Y. Kohara, K. Akiyama, and K. Isono, *Cell* 50:495–508, 1987). The *bfr* gene comprises 474 base pairs and 158 amino acid codons (including the start codon), and it encodes a polypeptide having essentially the same size ( $M_r$  18,495) and N-terminal sequence as the purified protein. A potential promoter sequence was detected in the 5' noncoding region, but it was not associated with an "iron box" sequence (i.e., a binding site for the iron-dependent Fur repressor protein). BFR was amplified to 14% of the total protein in a *bfr* plasmid-containing strain. An additional unidentified gene (*gen-64*), encoding a relatively basic 64-residue polypeptide and having the same polarity as *bfr*, was detected upstream of the *bfr* gene.

Iron is essential for the growth of animals, plants, and most microorganisms. However, in aerobic organisms, iron poses problems of cytotoxicity and bioavailability. This is because it reacts with oxygen to generate toxic free radicals and because it is relatively insoluble in its oxidized state. Many bacteria overcome the latter problem by elaborating specific Fe<sup>3+</sup> uptake mechanisms involving extracellular chelators or siderophores (26) and by depositing iron stores. These iron stores allow growth in iron-deficient medium (21), and their deposition may also provide a means of iron detoxification. Part of the stored iron is associated with bacterioferritin (BFR) (4), a protein that has been detected in at least seven bacterial species (3, 7, 14, 23, 24, 36, 44).

BFR is similar in many respects to ferritin, the well-characterized iron storage molecule of animals (9, 38) and plants (17, 30). Electron microscopy and X-ray diffraction data indicate that the BFR of *Escherichia coli* has a quaternary structure resembling that of horse spleen ferritin (30a, 30b, 44). It comprises an iron core (diameter, approximately 6 nm) surrounded by a roughly spherical protein shell (external diameter, approximately 12 nm) which is composed of 24 structurally identical subunits arranged in 4-3-2 symmetry (30b). The subunit molecular weight of the BFR from *E. coli* has been estimated as 15,000 (44) or 18,000 (31), and a molecular weight of approximately 500,000 has been obtained for the 24-mer holoprotein (43). These values are similar to those reported for plant and animal ferritins (17, 30, 38).

The polynuclear iron core of BFR may contain from 0 to approximately 3,000 iron atoms per molecule in the form of structurally amorphous oxyhydroxide-ferric phosphate (19, 20). However, reconstitution of the BFR iron core in vitro, without phosphate, produces a more highly ordered (ferrihydrite) core structure as found for the iron cores of mammalian ferritins (20).

Unlike animal and plant ferritins, BFR is a hemoferritin containing protoheme IX: one heme per two subunits in the *Azotobacter vinelandii* and *E. coli* proteins (36; J. Yariv, personal communication) and one heme per five subunits in the *Pseudomonas aeruginosa* protein (24). The BFRs of *E. coli* and *A. vinelandii* are known to be identical to the proteins previously designated cytochromes *b*<sub>1</sub> and *b*<sub>557.5</sub>, respectively (31, 36, 43). Antiserum raised against BFR from *Azotobacter chroococcum* shows no reactivity with horse spleen ferritin, but it exhibits partial reactivity with phytoferritin (7), and there is also good reactivity between anti-*E. coli* BFR and the *A. vinelandii* protein (Yariv, personal communication). There is no similarity between the peptide sequences of the BFR of *E. coli* (39) and the sequences of mammalian ferritins, despite their structural and functional relationships.

This paper describes the molecular cloning, nucleotide sequence, and map position of the bacterioferritin gene (*bfr*) of *E. coli*, and it also provides the first primary structure of a bacterioferritin from any source. In a recent short communication this primary structure was compared with previous N-terminal sequence data and amino acid compositions for BFR, and structural similarities with mammalian ferritins were also predicted (1).

### MATERIALS AND METHODS

**Bacterial strains, plasmids, and bacteriophage.** Strains of *E. coli* K-12 were used as follows: JM101 [*thi supE*  $\Delta$ (*proAB-lac*) F' *traD36 proA*<sup>+</sup>*B*<sup>+</sup> *lacI*<sup>q</sup>Z $\Delta$ M15], as host for M13 and pUC derivatives (22), and C600, as host for recombinant  $\lambda$  phages. The phagemid vector pUC119 (40) was used for subcloning, and the replicative forms of M13mp18 and M13mp19 were used for subcloning and preparing templates for DNA sequencing (42). A library containing fragments of *E. coli* C600 DNA obtained by partial digestion with *Sau*3A, inserted at the *Bam*HI site of the  $\lambda$ PE11 replacement vector (2), was kindly provided by P. T. Emmerson (Newcastle

\* Corresponding author.

University, Newcastle-upon-Tyne, United Kingdom). A copy of the  $\lambda$ -*E. coli* W3110 gene bank of Kohara et al. (15), containing 476 recombinant  $\lambda$  phages with overlapping inserts covering almost all of the *E. coli* chromosome, was obtained from D. Buck (Schering, Saffron Walden, Essex, United Kingdom). Cultures were grown with shaking at 37°C in L-broth plus ampicillin (50  $\mu$ g ml<sup>-1</sup>) where appropriate. Bacteriophage and plasmid DNAs were isolated and manipulated as described previously (11, 18).

**Immunochemical detection of BFR and  $\lambda$  *bfr*.** Antiserum raised in goat against BFR purified from *E. coli* K-12 strain 3300 was kindly supplied by J. Yariv (Weizmann Institute, Rehovot, Israel). Immunochemical detection was carried out at room temperature as follows. Nitrocellulose filters (Western [protein] blots or plaque-lifts) were blocked with 1% BSA/TBS (bovine serum albumin [BSA], 1% [wt/vol], in TBS: 50 mM Tris hydrochloride, pH 7.5, 150 mM NaCl) for  $\geq$ 3 h and then washed twice for 5 min with Tween 20 (0.1% [vol/vol] in TBS), incubated for 1 h with a 1:400 dilution of the primary antiserum in 0.25% BSA/TBS, and washed as before. They were then incubated sequentially with washing at each stage, first for 1 h with a 1:500 dilution of biotinylated anti-sheep/goat antibody in 0.25% BSA/TBS, then for  $\geq$ 30 min with a 1:1,000 dilution of streptavidin-biotinylated horseradish peroxidase complex in 0.25% BSA/TBS, and finally for 1 to 5 min with the substrate solution (25 mg of 3, 3'-diaminobenzidine tetrahydrochloride [Sigma], 30 mg of NiCl, and 5  $\mu$ l of 30% [vol/vol] H<sub>2</sub>O<sub>2</sub> in 100 ml of TBS) until color development was satisfactory, before rinsing thoroughly with tap water and air drying.

To detect  $\lambda$  phages expressing the *bfr* gene, the  $\lambda$ PE11 gene bank was plated at 10<sup>3</sup> to 10<sup>4</sup> plaques per plate, and the plaques were transferred to nitrocellulose filters (13) and screened by immunochemical detection as described above.

**PAGE and Western blotting.** Proteins were fractionated by polyacrylamide gel electrophoresis (PAGE) (16) either under denaturing conditions (0.1% sodium dodecyl sulfate [SDS] in all buffers and 15% acrylamide) for whole cells or under nondenaturing conditions (no SDS and 6.5% acrylamide) for soluble extracts. Cell pellets corresponding to 50  $\mu$ l of 16-h L-broth cultures were suspended in 40  $\mu$ l of loading buffer (0.1 M Tris hydrochloride, pH 6.8; glycerol, 10% [vol/vol]; SDS, 2.3% [wt/vol];  $\beta$ -mercaptoethanol, 5% [vol/vol]) and heated at 100°C for 10 min before SDS-PAGE. For nondenaturing gel electrophoresis, the cell pellets were disrupted by ultrasonic treatment (100 W) for 2 min at 0°C (10-s bursts with 30-s intervals for cooling), and soluble extracts were obtained by centrifugation (15,000  $\times$  *g* for 30 min). For Western blotting, samples in SDS-polyacrylamide gels were transferred to nitrocellulose filters by using a BioRad Trans-blot Electrophoretic Transfer Cell according to the manufacturer's instructions. Subsequent immunochemical detection was as described above.

**Southern blotting and plaque hybridization.** Restriction fragments were transferred from agarose gels to Genebind 45 nylon or nitrocellulose filters as described by Southern (32). Plaques representing each phage in the  $\lambda$ -*E. coli* gene library of Kohara et al. (15) were obtained by applying samples of lysates (10<sup>2</sup> to 10<sup>3</sup> PFU in 1  $\mu$ l) to seeded lawns and incubating at 37°C for 16 h. They were then transferred to nitrocellulose filters and screened after fixing (18). In each case, hybridization probes were prepared by random-primed DNA polymerase I (Klenow)-dependent incorporation of digoxigenin-11-dUTP into the complementary strands of single-stranded template DNAs prepared from M13mp18-*bfr*DX (a derivative of M13mp18 containing a 367-base-pair

[bp] *Dra*I-*Xho*II fragment of the *bfr* gene cloned into the *Sma*I and *Bam*HI sites) and from M13mp18 (as control). The prehybridization, hybridization (65°C), and washing conditions were those of Maniatis et al. (18), and the procedure for immunodetection was according to the manufacturer's instructions (Boehringer Mannheim Biochemicals). Molecular weight markers were prepared from *Hind*III and *Hind*III-plus-*Eco*RI digests of  $\lambda$ DNA labeled with digoxigenin-11-dUTP by T4 DNA polymerase-dependent replacement synthesis (18).

**DNA sequence analysis.** The sequencing strategy involved cloning specific segments of pGS277 and pGS280 obtained with different combinations of *Dra*I, *Eco*RI, *Hind*III, *Pst*I, and *Xho*II into the corresponding sites of M13mp18 and M13mp19 (see Fig. 4). Single-stranded M13 templates were sequenced by the chain-termination method using two "universal" primers, [ $\alpha$ -<sup>35</sup>S]thio-dATP, buffer gradient gels, and Sequenase (5, 28, 37). The amount of sequence obtained with some of the clones was increased by using specific oligonucleotide primers designed on the basis of the initial sequences. Oligonucleotides were made with an Applied Biosystems 381A DNA Synthesizer. Nucleotide sequences were compiled and analyzed with the aid of Staden computer programs (34, 35).

**Materials.** Restriction endonucleases, T4 DNA ligase, and T4 DNA polymerase were purchased from Bethesda Research Laboratories. Sequenase was purchased from Cambridge BioScience, [ $\alpha$ -<sup>35</sup>S]thio-dATP was from New England Nuclear Corp., and the immunochemicals (biotinylated anti-sheep/goat antibody and streptavidin-biotinylated horseradish peroxidase complex) were from Amersham International. The Non-Radioactive Labelling and Detection Kit was obtained from Boehringer Mannheim Biochemicals; the M13mp18 and -19 replicative-form DNAs and Genebind 45 nylon membranes were from Pharmacia. Nitrocellulose membranes (0.45- $\mu$ m pore size) were from Schleicher and Schuell. Protein molecular weight standards (12,300 to 78,000) were from British Drug Houses.

## RESULTS

The initial strategy used for cloning the *bfr* gene of *E. coli* was based on information in the gene-protein index of Neidhardt et al. (25), which indicated that BFR is expressed by plasmids pLC25-14 and pLC46-7 containing inserts from the 53- and 83-min regions of the linkage map, respectively. These plasmids were both found to contain overlapping segments of bacterial DNA from the 53-min region, and they expressed a polypeptide that comigrated with BFR in two-dimensional electrophoretic gels. However, it was clear from the nucleotide sequence of the corresponding gene that its product differed from BFR in both N-terminal sequence and amino acid composition (1, 39). A second strategy, using oligonucleotide probes based on segments of the partial BFR sequence published by Tsugita and Yariv (39), also failed, but it is now apparent that the partial sequence differs considerably from a more recent N-terminal sequence (1) and the gene-derived sequence. The third and successful strategy involved the immunochemical screening procedure described below.

**Cloning of a *bfr*- $\lambda$  hybrid gene.** Immunoscreening of a  $\lambda$ PE11-*E. coli* gene bank with anti-BFR antiserum yielded five immunopositive clones: four with identical inserts ( $\lambda$ G207 through -210) and one ( $\lambda$ G206) having a unique and apparently unrelated insert (Fig. 1). Derivatives of pUC119,

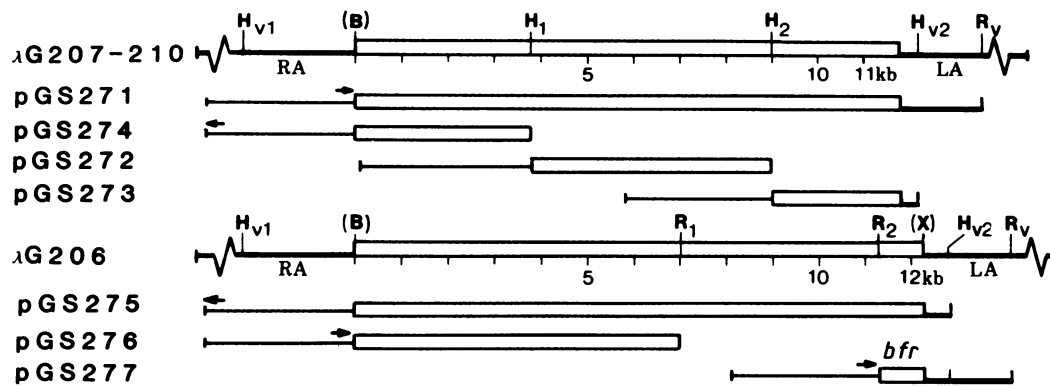


FIG. 1. Restriction maps showing the segments of DNA cloned in the immunopositive  $\lambda$  phages ( $\lambda$ G206 to -210) and in derivatives of pUC119 (pGS271 to -277). Open bars represent bacterial DNA, and vector DNA is denoted by thick ( $\lambda$ ) or thin (pUC119) lines; LA and RA indicate the respective left and right arms of the  $\lambda$  vector. The deduced location of the *bfr* gene is indicated, and where known the polarity of the *lac* promoter in pUC vectors is shown by an arrow. Relevant restriction sites are abbreviated as follows: *Bam*HI, B; *Eco*RI, R; *Hind*III, H; and *Xho*II, X. The *v* subscripts indicate vector sites. The brackets around the *Bam*HI and *Xho*II sites denote that they are hybrid sites formed between vector and insert DNA which may not exist in the bacterial DNA; the *Xho*II site is not unique.

containing the inserts cloned in  $\lambda$ G206 and  $\lambda$ G210 and various subfragments thereof, were constructed and designated pGS271 to -277 (Fig. 1). Cultures of JM101 transformants containing each of the plasmids were analyzed by Western blotting after SDS-PAGE. Each contained an immunoreactive band ( $M_r$  18,500) corresponding to the BFR subunit of the host (Fig. 2A). An additional immunoreactive band ( $M_r$  21,000) was observed with the pGS275 and pGS277 transformants, and this was particularly well expressed with pGS277. All of the strains exhibited a prominent band of  $M_r$  23,000, but subsequent tests showed that its appearance was not dependent upon the presence of the anti-BFR serum. This band reacted directly with the streptavidin-biotin complex and may therefore represent the biotin carboxyl carrier protein ( $M_r$  22,500) of the acetyl coenzyme A carboxylase complex, the sole biotin-containing protein found in *E. coli* (41). In parallel SDS-PAGE analyses of total cell protein, a polypeptide corresponding to the unique immunoreactive

band ( $M_r$  21,000) appeared as a strongly staining component of pGS277 transformants, but it was not as strongly amplified with pGS275 (Fig. 2B). Soluble extracts were also examined, using non-denaturing conditions to prevent dissociation of the BFR 24-mer ( $M_r$  approximately 500,000), and a protein having the corresponding mobility was detected in the pGS277-transformed strain, but not in the host (Fig. 3). Two proteins of much lower mobility which could correspond to larger BFR-like complexes were also detected in the pGS277-transformed strains. These observations, together with the proximity of the active segment of bacterial DNA to the flanking segment of  $\lambda$  DNA, suggested that a BFR- $\lambda$  fusion protein was being expressed from a *bfr*- $\lambda$  hybrid gene in  $\lambda$ G206, pGS275, and pGS277.

**Partial sequence of the *bfr* gene.** The plasmids pGS275 and pGS277 share a common 1.4-kilobase (kb) *Eco*RI-*Hind*III fragment ( $R_2$  to  $H_{v2}$  in Fig. 1) of which approximately 1 kb is bacterial DNA ( $R_2$  to X) and 0.4 kb is from the  $\lambda$  vector (X

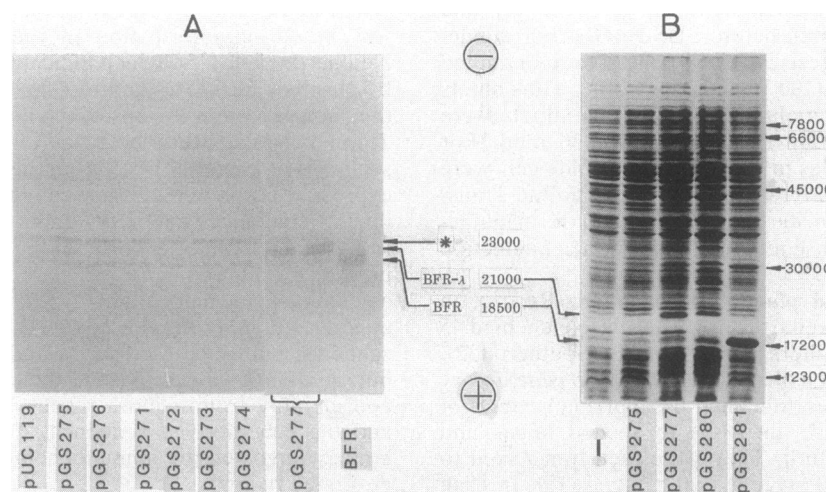


FIG. 2. SDS-PAGE analysis and Western blotting of total cell proteins of JM101 transformants containing different plasmids. Total cell protein was fractionated by SDS-PAGE and either (A) blotted and immunostained or (B) stained directly with Coomassie blue. The positions and estimated sizes of BFR and a putative BFR- $\lambda$  fusion product are indicated. The asterisk denotes a nonspecific staining component (see text), and the lane containing untransformed JM101 is marked with a dash. Sizes were estimated with standards in the range  $M_r$  12,300 to 78,000 as shown.

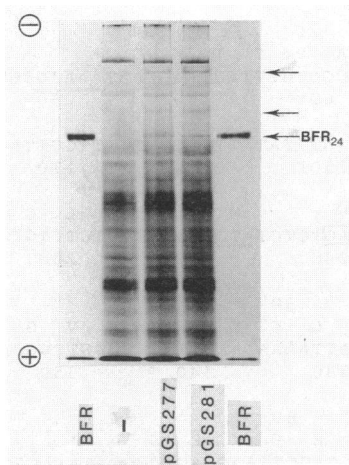


FIG. 3. Native PAGE of purified BFR and soluble extracts of JM101 transformants containing pGS277, pGS281, or no plasmid (-). Proteins were stained with Coomassie blue. The position of the BFR 24-mer (BFR<sub>24</sub>, *M<sub>r</sub>* approximately 500,000) and putative higher BFR oligomers (arrowed) are indicated.

to H<sub>v2</sub>). A conveniently situated *DraI* site was found in the bacterial fragment, and the sequence of the entire bacterial segment (971 bp) was derived according to the strategy outlined in the legend of Fig. 4 and in Materials and

Methods, above. The sequence (Fig. 5) contains an open reading frame of 154 bacterial codons extending up to the *XhoI* site. It was predicted that the reading frame should continue into the adjacent segment of the  $\lambda$  genome for a further 18  $\lambda$ -derived codons before reaching a stop codon (29).

The amino acid sequence translated from the open reading frame matches the 87-residue N-terminal sequence obtained from purified BFR (1), and this confirmed the identity of the cloned DNA. The 154 bacterial codons of the hybrid gene specify a polypeptide having an amino acid composition and a molecular weight (*M<sub>r</sub>* 18,025) that closely resemble those of BFR. Indeed, it was concluded that almost all of the *bfr* gene had been cloned. The sequence predicted for the  $\lambda$ -encoded C-terminal segment of the fusion product was RLPFTSCAVCLQDSMRSR, and, added to the N-terminal BFR segment, it yields a 172-residue polypeptide of *M<sub>r</sub>* 20,077. This is only slightly smaller than the value (21,000) estimated for the *bfr*- $\lambda$  hybrid gene product by SDS-PAGE (Fig. 2).

To obtain the entire *bfr* gene, its position was located in the physical map of the *E. coli* chromosome by hybridization and the gene was then cloned from the corresponding phage in the  $\lambda$ -*E. coli* gene bank of Kohara et al. (15).

**Location of the *bfr* gene in the physical map of the *E. coli* chromosome.** Identical Southern blots containing seven restriction digests of *E. coli* W3110 DNA were probed with M13mp18 and with a derivative (M13mp18-*bfr*DX) contain-

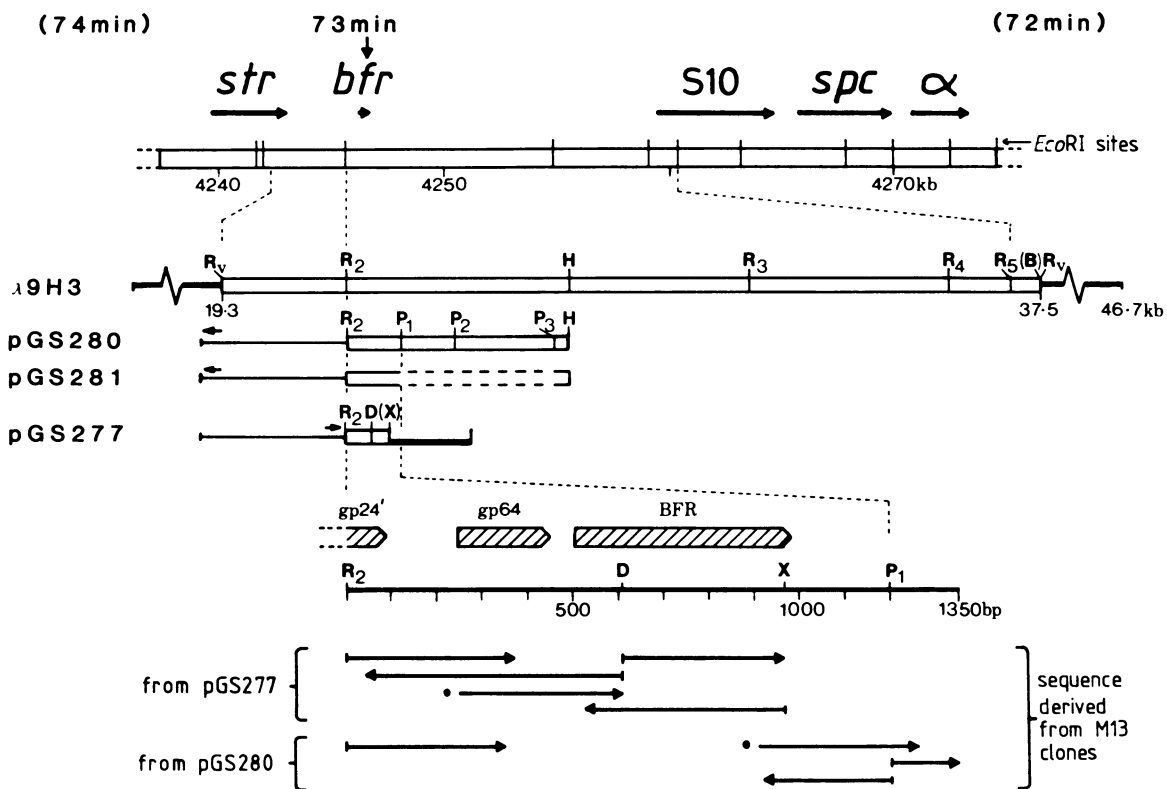


FIG. 4. Restriction map of the *str-spc* region of the *E. coli* chromosome and sequencing strategy for the *bfr* gene. The positions and extents of sequences obtained initially from M13 subclones of pGS277, and subsequently from pGS280, are indicated by arrows; the filled circles identify sequences obtained with specific primers. The coding regions are denoted by hatched arrows, and the restriction sites are abbreviated as for Fig. 1, except for *DraI* (D) and *PstI* (P). The relative positions and polarities of the *str*, S10, *spc*, and  $\alpha$  operons at 73 min in the *E. coli* physical map (15, 27) and the segment of DNA cloned in  $\lambda$ 9H3 are indicated. Vector DNA is denoted by thick ( $\lambda$ ) or thin (pUC119) lines, and the positions and polarities of the *lac* promoters in the pUC119 derivatives are indicated by the small arrows.

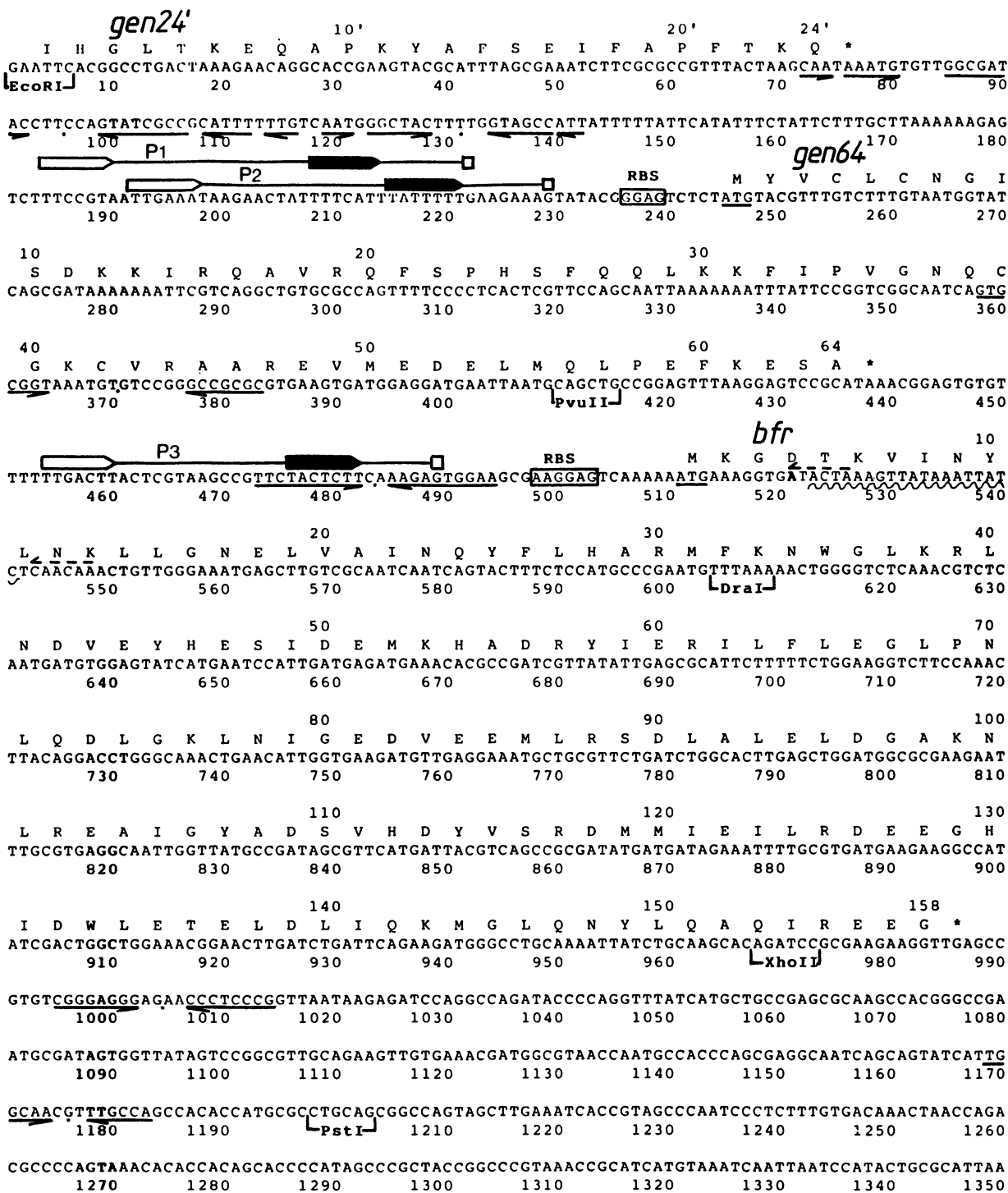


FIG. 5. Nucleotide sequence of the *bfr* region and primary structures of three open reading frames. The primary structures of the *bfr* gene product and those of two unidentified coding regions (*gen-24'* and *gen-64*) are shown above the nucleotide sequence. Relevant restriction sites, potential ribosome binding sites (boxed), translational initiation sites (underlined) and stop sites (asterisks), putative promoters (P, with the corresponding  $-35$ ,  $-10$ , and  $+1$  sites), significant regions of hyphenated dyad symmetry (converging arrows), a putative iron box (wavy underline), and a putative antisense promoter (dashed arrows) are indicated.

ing the 367-bp *DraI-XhoII* fragment which includes 122 *bfr* codons (Fig. 5). The probes were labeled by a random priming procedure using a nonradioactive but immunochemically detectable precursor, and DNA standards labeled with the same precursor were used for estimating the size of

hybridizing fragments. The fragments hybridizing with the M13mp18 probe closely matched the pattern predicted for the physical map of the *lac* region (15). The pattern observed with the M13mp18-*bfrDX* probe contained the same *lac* fragments plus an additional set of fragments corresponding

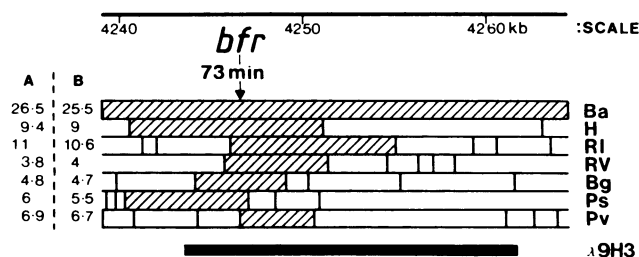


FIG. 6. Physical map of the 73-min region of the *E. coli* chromosome according to Kohara et al. (15). The sizes (in kilobases) of fragments hybridizing specifically with the *bfr* probe M13mp18-*bfr*DX (A) and the corresponding fragments (B) in the physical map (hatched regions) are indicated, as is the position deduced for the *bfr* gene. The scale at the top indicates the nucleotide coordinates. The relative position of the bacterial insert cloned in  $\lambda$ 9H3 is denoted by the solid bar. Restriction sites are abbreviated as follows: *Bam*HI, Ba; *Hind*III, H; *Eco*RI, RI; *Eco*RV, RV; *Bgl*I, Bg; *Pst*I, Ps; and *Pvu*II, Pv.

to a unique region of the *E. coli* physical map at 73 min (Fig. 6).

The inferred position of the *bfr* gene was confirmed by probing a "mini-set" of 476  $\lambda$  clones containing the entire *E. coli* chromosome (15). Plaques from only one clone,  $\lambda$ 9H3, hybridized uniquely with the M13mp18-*bfr*DX probe. This phage contains the region identified by Southern blotting (Fig. 6), and the results confirmed that the *bfr* gene is located at 73 min in the *E. coli* linkage map. A Southern blot analysis of the four BFR-immunopositive phages ( $\lambda$ G207 through -210) failed to show any hybridization with the M13mp18-*bfr*DX probe, and the reason for their initial selection remains obscure.

**Cloning of the intact *bfr* gene.** The *bfr* gene of  $\lambda$ 9H3 was subcloned into pUC119 on a 4.9-kb *Eco*RI-*Hind*III fragment ( $R_2$  to H) to generate pGS280 (Fig. 4). A 3-kb fragment of bacterial DNA ( $P_1$  to  $P_3$ ) was removed by *Pst*I-mediated deletion and religation to generate pGS281 (Fig. 4). An examination of the total cell proteins of JM101 derivatives transformed with pGS280 and pGS281, or infected with M13mp18 carrying the *bfr* region ( $R_2$  to  $P_1$ ), indicated that an immunostaining band corresponding to BFR is very considerably amplified in Western blots of SDS-PAGE-fractionated samples (data not shown). Quantitative estimates from Coomassie Blue-stained SDS-PAGE tracks of the pGS281 transformants indicated that BFR represents 14 to 15% of the total cell protein (Fig. 2B) and the soluble protein (not shown). Native PAGE analysis of soluble extracts of the pGS281 transformant also showed that the BFR 24-mer (and possibly higher oligomers) is amplified relative to the host, where the corresponding band(s) is not visible (Fig. 3). These results confirm that an intact *bfr* gene has been cloned in the 1.55-kb insert of pGS281.

**Complete nucleotide sequence of the *bfr* gene and an unidentified gene (*gen-64*).** The partial sequence of the *bfr* gene obtained with pGS277 was overlapped and extended using clones derived from pGS280 (Fig. 4). The complete sequence is contained in the 1,350-bp segment shown in Fig. 5. The sequence is fully overlapped, and 86% (positions 40 to 1203) is derived from both strands. It can be seen that the *bfr* coding region was completed by just four additional codons. The *bfr* structural gene thus encodes a polypeptide of 158 amino acid residues and  $M_r$  18,495. The amino acid composition corresponds to that reported for the BFR protein (1), and the  $M_r$  matches the value observed in the present

work and reported previously by Smith et al. (31). In addition to the *bfr* gene, two other potential coding regions having the same polarity as *bfr* were detected by using the FRAMESCAN program (35). One (designated *gen-64*) extends from position 245 for 192 bp and encodes an unidentified product (gp64) of 64 amino acid residues and  $M_r$  7,363 (Fig. 5). The other potential coding region (*gen-24'*) extends into the sequenced region for just 24 codons, and it presumably specifies the C terminus of a larger polypeptide (gp24').

## DISCUSSION

The results describe the first cloning and sequencing of a *bfr* gene from any source and the first primary structure of a BFR or hemoferritin. The work also illustrates a novel use of the *E. coli* physical map for defining the chromosomal location of a gene. The amino acid sequence of BFR matches the 87-residue N-terminal sequence derived from the purified protein at all but one of the 70 unambiguously assigned positions: residue 53 was previously identified as methionine, not lysine (1). There are five histidine and seven methionine residues which could function as heme ligands, and secondary-structure predictions suggest that the BFR molecule could consist of a bundle of four antiparallel  $\alpha$ -helices resembling mammalian ferritins and several bacterial cytochromes (1). The *bfr* gene is highly expressed in pGS281 transformants which, according to quantitative SDS-PAGE, represent an excellent source of BFR. The high amplification of BFR apparent in denatured samples (14 to 15%) contrasts with estimates of approximately 1% obtained under nondenaturing conditions. It is possible that the assembly of most of the BFR is incomplete or that it forms aggregates or membrane-bound complexes that are excluded from native gels. The hybrid BFR- $\lambda$  protein expressed from pGS277 appears to assemble into a 24-mer of comparable mobility to BFR, despite the presence of 14 extra C-terminal residues. The C terminus must be located in such a position that it can be extended without gross disruption of the structure, possibly by protruding into the central cavity of BFR.

The *bfr* gene is preceded by a putative ribosome binding site and a good potential promoter ( $P_3$ ) in the relatively short (70-bp) intergenic region (Fig. 5). The *bfr* promoter region contains a partially palindromic sequence which may function as an operator because it overlaps the proposed -10 sequence and the transcriptional start site (Fig. 5). There is also a G+C-rich region of hyphenated dyad symmetry ( $\Delta G = -96 \text{ kJ mol}^{-1}$ ) immediately downstream of the *bfr* coding region (Fig. 5). This could function as a transcriptional terminator although it lacks the run of T(U) nucleotides which is typically associated with *rho*-independent terminators. The codon preferences suggest that *bfr* is a weakly or moderately expressed gene because it exhibits a relatively high proportion of modulatory codons (3.2%) and a low proportion of optimal energy codons (40%) in the diagnostic set (10).

The expression of mammalian ferritins is induced by iron (45), but this has not been investigated for BFRs. The expression of a number of *E. coli* genes concerned with iron uptake is known to be modulated by iron, and in many cases this is mediated by the regulatory protein Fur (12). A Fur- $\text{Fe}^{2+}$  complex is thought to bind to a 19-bp recognition site known as the iron box in the operator regions of the regulated genes (8). An iron box resembling the consensus sequence (GATAATGATAATCATTATC) was not detected in the sequences upstream of the *bfr* and *gen-64* coding

regions. This suggests that *bfr* expression may not be directly modulated by Fur. However, there is a relatively poor iron box having identities at 12 of 19 positions in the *bfr* coding region, positions 523 to 541 (Fig. 5). Furthermore, this sequence is associated with a potential promoter sequence of opposite polarity to the *bfr* gene (Fig. 5). It is therefore conceivable that this promoter expresses an antisense mRNA that could base-pair with the *bfr* transcript, and thus repression of the antisense promoter by the Fur-Fe<sup>2+</sup> complex would lead indirectly to derepression of the *bfr* gene by iron. No binding site for the cyclic AMP receptor protein (CRP) (6) or the anaerobic gene regulator (FNR) (33), was detected.

The unidentified gene, *gen-64*, encodes a product of average hydropathy and relatively high content of basic amino acids. The *gen-64* coding region is preceded by a putative ribosome binding site, and its codon usage is typical of a moderately expressed gene. There are two potential promoter sequences upstream of *gen-64* and no obvious transcriptional terminator, which raises the possibility that *gen-64* and *bfr* are cotranscribed (Fig. 5). The function of gp64 remains unknown, and computer searches of several data bases revealed no significant homologies with gp64 (or BFR). The C-terminal end of the second unidentified gene (*gen-24'*) gives no clues about its function, but it is immediately followed by a region that would generate a transcript of potentially complex secondary structure and by an unusually A+T-rich region (80% A+T for the 143 to 233 segment) (Fig. 5).

It has recently been reported that *E. coli* contains two novel iron proteins (B. F. Matzanke, G. I. Müller, E. Bill, G. Winkelmann, and A. X. Trautwein, Abstr. Sites Actifs Métalliques en Biologie et Analoges Synthétiques II, 1988, p. 35). However, the precise functions of these proteins in bacterial ferrometabolism and their relationship to BFR have yet to be established. Indeed, relatively little is known about the intracellular metabolism of iron in bacteria compared with our understanding of its uptake. The cloning, mapping, and overexpression of the *bfr* gene will facilitate various studies including the isolation and physiological characterization of *bfr* mutants, which should help to redress this imbalance.

#### ACKNOWLEDGMENTS

We thank J. Yarov (Weizmann Institute, Rehovot, Israel) for kindly providing BFR antiserum and purified BFR, and D. Buck (Schering, Saffron Walden, United Kingdom) and Y. Kohara (Department of Molecular Biology, Nagoya University, Nagoya, Japan) for supplying the  $\lambda$ -*E. coli* gene library.

This work was supported by a grant from the Wellcome Trust.

#### LITERATURE CITED

- Andrews, S. C., J. M. A. Smith, J. R. Guest, and P. M. Harrison. 1989. Amino acid sequence of the bacterioferritin (cytochrome *b*<sub>1</sub>) of *Escherichia coli*. *Biochim. Biophys. Res. Commun.* **158**:489-496.
- Arthur, H. M., D. Bramhill, P. B. Eastlake, and P. T. Emmerston. 1982. Cloning of the *uvrD* gene of *E. coli* and identification of the product. *Gene* **19**:285-295.
- Bartsch, R. G., T. Katuno, T. Horio, and M. D. Kamen. 1971. Preparation and properties of *Rhodospirillum rubrum* cytochromes *c*<sub>2</sub>, *cc'*, and *b*<sub>557.5</sub>, and flavin mononucleotide protein. *J. Biol. Chem.* **246**:4489-4496.
- Bauminger, E. R., S. G. Cohen, D. P. E. Dixon, A. Levy, S. Ofer, and J. Yarov. 1980. Mössbauer spectroscopy of *Escherichia coli* and its iron-storage protein. *Biochim. Biophys. Acta* **623**:237-242.
- Biggin, M. D., T. J. Gibson, and G. F. Hong. 1983. Buffer gradient gels and <sup>35</sup>S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA* **80**:3963-3965.
- Busby, S. 1986. Catabolite repression and related processes. *Symp. Soc. Gen. Microbiol.* **39**:51-77.
- Chen, M., and R. R. Crichton. 1982. Purification and characterization of a bacterioferritin from *Azotobacter chroococcum*. *Biochim. Biophys. Acta* **707**:1-6.
- de Lorenzo, V., W. Sechan, M. Herrero, and J. B. Neilands. 1987. Operator sequences of the aerobactin operon of plasmid ColV-K30 binding the ferric uptake regulation (*fur*) repressor. *J. Bacteriol.* **169**:2624-2630.
- Ford, G. C., P. M. Harrison, D. W. Rice, J. M. A. Smith, A. Treffry, J. L. White, and J. Yarov. 1984. Ferritin: design and formation of an iron-storage molecule. *Philos. Trans. R. Soc. London Ser. B* **304**:551-565.
- Grosjean, H., and W. Friers. 1982. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene* **18**:199-209.
- Guest, J. R., R. E. Roberts, and P. E. Stephens. 1983. Hybrid plasmids containing the pyruvate dehydrogenase complex genes and gene-DNA relationships in the 2 to 3 minute region of the *Escherichia coli* chromosome. *J. Gen. Microbiol.* **129**:671-680.
- Hantke, K. 1981. Regulation of ferric iron transport in *E. coli*. Isolation of a constitutive mutant. *Mol. Gen. Genet.* **182**:288-292.
- Huynh, T. V., R. A. Young, and R. W. Davis. 1985. Constructing and screening cDNA libraries in  $\lambda$ gt10 and  $\lambda$ gt11, p. 49-78. *In* D. M. Glover (ed.), *DNA cloning*, vol. 1: A practical approach. IRL Press, Oxford.
- Inoue, Y., and H. Kubo. 1965. The metabolism of *Streptomyces griseus*. X. A study of *Streptomyces griseus* cytochrome *b*. *Biochim. Biophys. Acta* **110**:57-65.
- Kohara, Y., K. Akiyama, and K. Isono. 1987. The 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.
- Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227**:680-685.
- Laulhere, J. P., A. M. Lescurie, and J. F. Briat. 1988. Purification and characterization of ferritins from maize, pea, and soya bean seeds. *J. Biol. Chem.* **263**:10289-10294.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Mann, S., J. V. Bannister, and R. J. P. Williams. 1986. Structure and composition of ferritin cores isolated from human spleen, limpet (*Patella vulgata*) hemolymph and bacterial (*Pseudomonas aeruginosa*) cells. *J. Mol. Biol.* **188**:225-232.
- Mann, S., J. M. Williams, A. Treffry, and P. M. Harrison. 1987. Reconstituted and native iron-cores of bacterioferritin and ferritin. *J. Mol. Biol.* **198**:405-416.
- McIntosh, M. A., and C. F. Earhart. 1977. Coordinate regulation by iron of the synthesis of phenolate compounds and three outer membrane proteins in *Escherichia coli*. *J. Bacteriol.* **131**:331-339.
- Messing, J. 1983. New M13 vectors for cloning. *Methods Enzymol.* **101**:20-78.
- Meyer, T. E., and M. A. Cusanovich. 1985. Soluble cytochrome composition of the purple phototrophic bacterium *Rhodospseudomonas sphaeroides* ATCC 17023. *Biochim. Biophys. Acta* **807**:308-319.
- Moore, G. R., S. Mann, and J. V. Bannister. 1986. Isolation and properties of the complex nonheme-iron-containing cytochrome *b*<sub>557</sub> (bacterioferritin) from *Pseudomonas aeruginosa*. *J. Inorg. Biochem.* **28**:329-336.
- Neidhardt, F. C., V. Vaughn, T. A. Phillips, and P. L. Bloch. 1983. Gene-protein index of *Escherichia coli* K-12. *Microbiol. Rev.* **47**:231-284.
- Neilands, J. B. 1983. Siderophores, p. 137-166. *In* E. C. Theil, G. L. Eichhorn, and L. G. Marzill (eds.), *Iron binding proteins*

- without cofactors or sulfur clusters, vol. 5. Elsevier Science Publishing, Inc., New York.
27. Post, L. E., and M. Nomura. 1980. DNA sequences from the *str* operon of *Escherichia coli*. *J. Biol. Chem.* **255**:4660–4666.
  28. Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe. 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* **143**:161–178.
  29. Sanger, F., A. R. Coulson, G. F. Hong, D. F. Hill, and G. B. Petersen. 1982. Nucleotide sequence of bacteriophage  $\lambda$  DNA. *J. Mol. Biol.* **162**:729–773.
  30. Sczekman, S. R., and J. G. Joshi. 1987. Isolation and characterization of ferritin from soybeans (*Glycine max*). *J. Biol. Chem.* **262**:13780–13788.
  - 30a. Smith, J. M. A., G. C. Ford, and P. M. Harrison. 1988. Very-low-resolution structure of a bacterial ferritin. *Biochem. Soc. Trans.* **16**:836–838.
  - 30b. Smith, J. M. A., G. C. Ford, P. M. Harrison, J. Yariv, and A. J. Kalb. 1989. Molecular size and symmetry of the bacterioferritin of *Escherichia coli*. X-ray crystallography characterization of four crystal forms. *J. Mol. Biol.* **205**:465–467.
  31. Smith, J. M. A., A. V. Quirk, R. W. H. Plank, F. M. Diffin, G. C. Ford, and P. M. Harrison. 1988. The identity of *Escherichia coli* bacterioferritin and cytochrome  $b_1$ . *Biochem. J.* **255**:737–740.
  32. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**:503–517.
  33. Spiro, S., and J. R. Guest. 1987. Regulation and over-expression of the *fmr* gene of *Escherichia coli*. *J. Gen. Microbiol.* **133**:3279–3288.
  34. Staden, R. 1982. An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. *Nucleic Acids Res.* **10**:2951–2961.
  35. Staden, R. 1984. Graphic methods to determine the function of nucleic acid sequences. A summary of ANALYSEQ options. *Nucleic Acids Res.* **12**:521–538.
  36. Stiefel, E. I., and G. D. Watt. 1979. *Azotobacter* cytochrome  $b_{557.5}$  is a bacterioferritin. *Nature (London)* **279**:81–83.
  37. Tabor, S., and C. C. Richardson. 1978. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA* **84**:4767–4771.
  38. Theil, E. C. 1987. Ferritin: structure, gene regulation and cellular function in animals, plants, and microorganisms. *Annu. Rev. Biochem.* **56**:282–315.
  39. Tsugita, A., and J. Yariv. 1985. Preliminary results for the primary structure of bacterioferritin of *Escherichia coli*. *Biochem. J.* **231**:209–212.
  40. Viera, J., and J. Messing. 1987. Production of single-stranded plasmid DNA. *Methods Enzymol.* **153**:3–34.
  41. Wood, H. G., and R. E. Barden. 1977. Biotin enzymes. *Annu. Rev. Biochem.* **46**:385–413.
  42. Yanisch-Peron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13 mp18 and pUC19 vectors. *Gene* **33**:103–119.
  43. Yariv, J. 1983. The identity of bacterioferritin and cytochrome  $b_1$ . *Biochem. J.* **211**:527.
  44. Yariv, J., A. J. Kalb, R. Sperling, E. R. Bauminger, S. G. Cohen, and S. Ofer. 1981. The composition and the structure of bacterioferritin of *Escherichia coli*. *Biochem. J.* **197**:171–175.
  45. Zahring, J., B. S. Baliga, and H. N. Munro. 1976. Novel mechanism for translational control in regulation of ferritin synthesis by iron. *Proc. Natl. Acad. Sci. USA* **73**:857–861.