Sequence of the Gene Encoding Flavocytochrome *c* from *Shewanella putrefaciens*: A Tetraheme Flavoenzyme That Is a Soluble Fumarate Reductase Related to the Membrane-Bound Enzymes from Other Bacteria[†,⊥]

Sara L. Pealing,[‡] Ann C. Black,[‡] Forbes D. C. Manson,[‡] F. Bruce Ward,[‡] Stephen K. Chapman,[§] and Graeme A. Reid*[,‡]

Edinburgh Centre for Molecular Recognition, Institute of Cell and Molecular Biology, and Department of Chemistry, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JR, Scotland, U.K.

Received June 16, 1992; Revised Manuscript Received September 18, 1992

ABSTRACT: Flavocytochrome *c* from the Gram-negative, food-spoiling bacterium *Shewanella putrefaciens* is a soluble, periplasmic fumarate reductase. We have isolated the gene encoding flavocytochrome *c* and determined the complete DNA sequence. The predicted amino acid sequence indicates that flavocytochrome *c* is synthesized with an N-terminal secretory signal sequence of 25 amino acid residues. The mature protein contains 571 amino acid residues and consists of an N-terminal cytochrome domain, of about 117 residues, with four heme attachment sites typical of *c*-type cytochromes and a C-terminal flavoprotein domain of about 454 residues that is clearly related to the flavoprotein subunits of fumarate reductases and succinate dehydrogenases from bacterial and other sources. A second reading frame that may be cotranscribed with the flavocytochrome *c* gene exhibits some similarity with the 13-kDa membrane anchor subunit of *Escherichia coli* fumarate reductase. The sequence of the flavoprotein domain demonstrates an even closer relationship with the product of the yeast *OSM*1 gene, mutations in which result in sensitivity to high osmolarity. These findings are discussed in relation to the function of flavocytochrome *c*.

Shewanella putrefaciens is a Gram-negative, facultative aerobe that is commonly associated with food spoilage. Under anaerobic conditions several *c*-type cytochromes are synthesized, the most abundant of which is flavocytochrome *c* (Morris et al., 1990). The protein has been purified and shown to catalyze fumarate reduction using artificial electron donors such as methyl viologen (Black, 1991). The physiological reductant is unknown, but evidence indicates that flavocytochrome *c* participates in electron transfer from formate to fumarate and possibly also to trimethylamine oxide (TMAO).

Fumarate is a terminal electron acceptor in the anaerobic respiratory pathways of many facultative aerobes. The fumarate reductase of *Escherichia coli* and other bacteria is a membrane-bound enzyme with four subunits (Cole et al., 1982) encoded by the frd operon. The largest of these, encoded by frdA, is a 69-kDa flavoprotein which provides the active site for fumarate reduction (Cole, 1982). The frdB gene product is a 27-kDa iron–sulfur protein. These two subunits form a membrane-extrinsic catalytic complex that is associated with two smaller polypeptides which serve as membrane anchors (Lemire et al., 1982) and are required for electron transfer from menaquinol to substrate (Cecchini et al., 1986; Weiner et al., 1986; Westenberg et al., 1990). The enzyme is structurally and functionally similar to the respiratory succinate dehydrogenase, and the interconversion of succinate and fumarate is readily reversible in both enzymes. In contrast, the soluble fumarate reductase (flavocytochrome *c*) from *S. putrefaciens* is essentially unidirectional (Black, 1991). Soluble fumarate reductases have been identified in yeast (Rossi et al., 1964; Muratsubaki & Katsume, 1982; Muratsubaki & Katsume, 1985) and the sulfate-reducing bacterium *Desulfovibrio multispirans* (He et al., 1986); these enzymes are also essentially unidirectional. The membrane-bound fumarate reductases and succinate dehydrogenases contain covalently bound flavin [8α-(N^3-histidyl)FAD],[1] (Weiner & Dickie, 1979), whereas the soluble fumarate reductases contain noncovalently bound FAD. The biochemical evidence raises the possibility that *S. putrefaciens* flavocytochrome *c* is more closely related to the soluble fumarate reductases than to the membrane-bound enzymes. We have begun to examine the relationships between different fumarate reductases at the structural level by determining the sequence of the gene encoding flavocytochrome *c*. Our findings demonstrate a clear evolutionary relationship between the soluble and membrane-bound fumarate reductases, with residues identified at the active site of *E. coli* fumarate reductase (Schröder et al., 1991) conserved both in other membrane-bound enzymes and in flavocytochrome *c*.

MATERIALS AND METHODS

Strains and Vectors. Shewanella putrefaciens NCMB400 was used as a source of DNA and protein. *E. coli* strains MM294 and TG1 were used as hosts for recombinant plasmids. The expression vector pEX3 (Stanley & Luzio, 1984) was used to construct a library of *S. putrefaciens* genomic DNA fragments. Subclones were constructed in pTZ18R and pTZ19R (Rokeach et al., 1988). The plasmid pCI857 directs

† This work was supported by research grants from the Molecular Recognition Initiative and the Transition Metals in Biology (TRAMPS) Initiative of the Science and Engineering Research Council, which also provided studentships for S.L.P. and A.C.B.

⊥ The sequence in this paper has been submitted to Genbank under accession number L04283.

* To whom correspondence should be addressed.

‡ Edinburgh Centre for Molecular Recognition, Institute of Cell and Molecular Biology.

§ Department of Chemistry.

[1] Abbreviations: FAD, flavin adenine dinucleotide; Fcc, flavocytochrome *c*; Ecf, Pvf, and Wsf, the flavoprotein subunits of fumarate reductase from *Escherichia coli*, *Proteus vulgaris*, and *Wolinella succinogenes*, respectively; Ecs, Scs, and Bss, the flavoprotein subunits of succinate dehydrogenase from *E. coli*, *Saccharomyces cerevisiae*, and *Bacillus subtilis*, respectively.

0006-2960/92/0431-12132$03.00/0

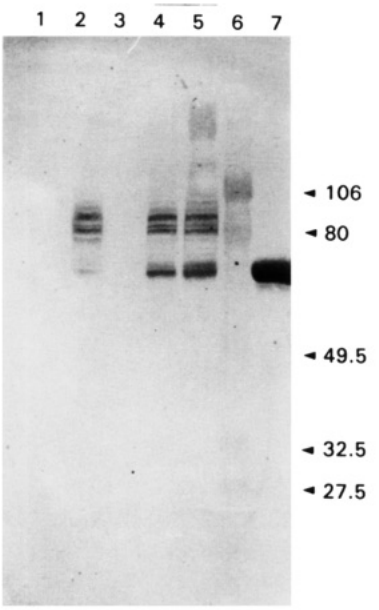

FIGURE 1: Immunochemical detection of flavocytochrome *c* after gel electrophoresis. Total cell protein (100 μg) from *S. putrefaciens* grown aerobically (lane 1), anaerobically (lane 2), aerobically with the addition of fumarate (lane 3), anaerobically with fumarate (lane 4), or anaerobically with fumarate plus lactate (lane 5) was subjected to SDS–PAGE alongside molecular mass markers (lane 6) and purified flavocytochrome *c* (lane 7). After electrophoresis, the gel was subjected to Western blotting, and flavocytochrome *c* was detected with a sheep antiserum. The positions and molecular masses, in kilodaltons, of the markers (which do not show up well in the photograph) are indicated at the right.

expression of the temperature-sensitive λCI repressor and was used to repress unwanted synthesis of proteins encoded by recombinant pEX3 (Remaut et al., 1983).

DNA Isolation. Chromosomal DNA was isolated from a 100-mL culture of stationary-phase, aerobically grown *S. putrefaciens* essentially as described for the isolation of *E. coli* DNA by Sambrook et al. (1989). Plasmid DNA and single-stranded plasmid DNA were isolated from *E. coli* transformants as described (Sambrook et al., 1989; Vieira & Messing, 1987). M13K07 was used as helper phage for single-stranded DNA production.

Construction of S. putrefaciens DNA Libraries. A library of *S. putrefaciens* chromosomal DNA fragments was constructed in the expression vector pEX3. Plasmid DNA was cut with *Bam*HI and treated with calf intestinal alkaline phosphatase. *S. putrefaciens* DNA was partially digested with *Sau*3AI and fractionated by centrifugation on 10–40% (w/v) sucrose density gradients. Fragments of 0.5–4 kb were ligated with vector DNA and used to transform *E. coli* MM294 containing pCI857, selecting for resistance to ampicillin and kanamycin. A total of 2×10^4 independent transformants were obtained.

A second library was constructed in pTZ19R that had been cut with *Bam*HI. *S. putrefaciens* DNA was cut to completion with *Bgl*II, and fragments were ligated with the vector without size fractionation.

DNA Library Screening. The expression library was screened using a sheep antiserum directed against purified flavocytochrome *c*. Twenty plates containing ampicillin and kanamycin to maintain selection for both pEX3 recombinants and pCI857 were each inoculated with 2×10^3 transformants. After growth at 30 °C for 20 h, colonies were replica plated onto nylon filters (Amersham Hybond-N). The replica filters were incubated at 42 °C for 2 h to allow expression of recombinant proteins and were then treated as in Sambrook et al. (1989) for immunochemical screening. After incubation with anti-flavocytochrome *c* antibodies, the filters were thoroughly washed and developed with horseradish peroxidase-conjugated donkey anti-sheep IgG antibodies. After further washing, enzyme activity was visualized using *o*-dianisidine as substrate.

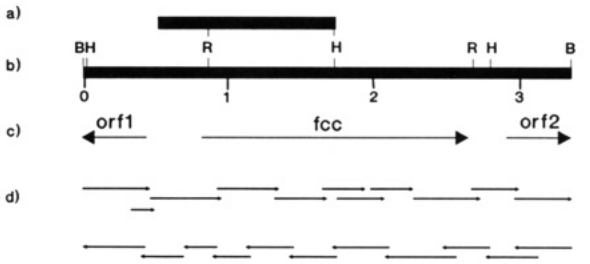

FIGURE 2: Physical map of the cloned flavocytochrome *c* gene. The solid bars in (a) and (b) represent the cloned fragments in pACB1 and pSP1, respectively. Cleavage sites for the restriction endonucleases *Eco*RI (R), *Bgl*II (B), and *Hin*dIII (H) are indicated, and the distance from the left border is shown in kilobases. (c) The major reading frames and their orientations are indicated by the arrows: fcc, flavocytochrome *c*; orf1, upstream reading frame; orf2, downstream reading frame. (d) The arrows indicate the direction and length of sequences read from DNA sequencing gels.

The *Bgl*II fragment library was screened by hybridization with the ^{32}P-labeled 1.4-kb insert from pACB1 (see Results and Figure 2a). The procedure for colony hybridization and fragment labeling by the random-priming method are described in Sambrook et al. (1989).

DNA Sequence Determination. The DNA sequence of the entire cloned 3.3-kb *Bgl*II fragment was determined on both strands using the dideoxy chain termination method (Sanger et al., 1977) with the Sequenase (U.S. Biochemical Corp.) T7 polymerase. Individual clones and subclones were sequenced using either the universal reverse sequencing primer or primers designed specifically to extend the sequence already determined.

Sequence Analysis. DNA sequence information was analyzed using the UWGCG (University of Wisconsin Genetics Computer Group) programs (Devereux et al., 1984) mounted on the SEQNET VAX at Daresbury and on ERCVAX at Edinburgh. Exhaustive database searches were performed on the Edinburgh distributed array processor using the program prosrch (Coulson et al., 1987; Collins et al., 1988).

Amino Acid Sequencing. The N-terminal sequence of purified flavocytochrome *c* was determined with a Waters 743 automated, gas-phase sequenator. PTH-amino acid derivatives were detected in an on-line analyzer by reverse-phase HPLC.

Western Blotting. Proteins were separated by SDS–polyacrylamide gel electrophoresis and electrophoretically transferred to a nylon membrane (Amersham Hybond-N) as described previously (Haid & Suissa, 1984). Flavocytochrome *c* was detected using antibodies as described above for DNA library screening.

RESULTS

Isolation of Flavocytochrome c Coding Sequences from an Expression Library. An antiserum raised in sheep against purified flavocytochrome *c* was shown to precipitate fumarate reductase activity and flavocytochrome *c* protein (Black, 1991; A. C. Black, S. L. Pealing, S. K. Chapman, F. B. Ward, and G. A. Reid, unpublished experiments). Western blotting demonstrated that these antibodies specifically recognize flavocytochrome *c* in cell extracts from *S. putrefaciens*. The

```
                                                                                
agatctgcccaattgtttattaacgccttaccctgcaaggtataagctttagcgtgggttgcatatataggatcagattgcgctaactta  90
   d  a  w  n  n  i  l  a  k  g  q  l  t  y  a  k  a  h  t  a  y  i  p  d  s  q  a  l  k
                                                                                
cgatccatgttcaccgggctaaccacagcaatcacatcgatatcgctgtcaggatcatgaacatcttttaccaattgttgaaaactcttt 180
 r  d  m  n  v  p  s  v  v  a  i  v  d  i  d  s  d  p  d  h  v  d  k  v  l  q  q  f  s  k
                                                                                
cgacgtacaataatgcgcttattggtcagctggccaattctatccatcaactcgatattaaagccttgatcaacgccgttattccgccac 270
 r  r  v  i  i  r  k  n  t  l  q  g  i  r  d  m  l  e  i  n  f  g  q  d  v  g  n  n  r  w
                                                                                
tctaatggcgctgtttttgaatgtacgccaaagacaatgctgtcttgcgcgagcacaggcatagtgatgatgataagtaataacaataga 360
 e  l  p  a  t  k  s  h  v  g  f  v  i  s  d  q  a  l  v  p  m  t  i  i  i  l  l  l  l  l
                                                                                
gttttcatgccgcgatgttacccatttttttccacttgaaactgtgcagcattacccgaattagcaaatcattatgtacgttttcttatga 450
 t  k  m
                                                                                
tgttaactcgatgtgatctatctccacataagctatgcaaaattgcatatacctctttataactatgcatgtttgcatagcgcatttttа 540
                                                                                
ggcttgtgattgagttgtcatttctgccctctattctggttatctggttgttgattgcgccccttaggcaaagattacattgagcgaaa 630
                                                                                
ataatagctgatcaactcgatgctaatttggaccgtatttttaaagtgaagtgaagtcagtttcacttcagcctaactaaaacaataaga 720
                                                                                
ggcttctatagttaatgaataaaccgtcattacgtatagaaggggagcaaaaatgaaaaagatgaatcttgcagtctgtattgctacat 810
                                                  M  K  K  M  N  L  A  V  C  I  A  T  L
                                                                                
taatgggcacagcaggcctaatgggcactgctgttgcggctgataacttagctgaattccatgtacaaaaccaagaatgtgatagctgcc 900
  M  G  T  A  G  L  M  G  T  A  V  A  A  D  N  L  A  E  F  H  V  Q  N  Q  E  C  D  S  C  H  18
                                                                                
atacaccagatggtgaactgtcaaacgacagcttaacctatgaaaatacccaatgcgtatcttgccatggcacacttgctgaagtagctg 990
  T  P  D  G  E  L  S  N  D  S  L  T  Y  E  N  T  Q  C  V  S  C  H  G  T  L  A  E  V  A  E  48
                                                                                
aaaccacaaaacatgaacattataatgctcatgcttctcatttccctggcgaagtagcttgtacctcatgccacagcgcacacgaaaaat 1080
  T  T  K  H  E  H  Y  N  A  H  A  S  H  F  P  G  E  V  A  C  T  S  C  H  S  A  H  E  K  S  78
                                                                                
cgatggtgtattgtgactcttgccacagcttcgatttcaacatgccttatgctaaaaaatggctacgtgacgagccgactattgctgaat 1170
  M  V  Y  C  D  S  C  H  S  F  D  F  N  M  P  Y  A  K  K  W  L  R  D  E  P  T  I  A  E  L  108
                                                                                
tggccaaagacaaatcagaacgtcaggctgctcttgctagcgcacctcacgatactgttgacgtagtggttgtcggttctggcggcgcag 1260
  A  K  D  K  S  E  R  Q  A  A  L  A  S  A  P  H  D  T  V  D  V  V  V  V  G  S  G  G  A  G  138
                                                                                
gtttctcagcggcaatatcagcaacagacagtggtgctaaagtcattcttattgaaaaagagcctgttattggtggtaatgctaagttag 1350
  F  S  A  A  I  S  A  T  D  S  G  A  K  V  I  L  I  E  K  E  P  V  I  G  G  N  A  K  L  A  168
                                                                                
ctgcgggtggcatgaacgctgcttggactgatcaacaaaaagccaaaaaaattactgacagcccagagttaatgttcgaagacaccatga 1440
  A  G  G  M  N  A  A  W  T  D  Q  Q  K  A  K  K  I  T  D  S  P  E  L  M  F  E  D  T  M  K  198
                                                                                
aaggtggccaaaacataaatgaccctgcattagttaaagtattaagctcacactctaaagactctgttgattggatgaccgctatgggtg 1530
  G  G  Q  N  I  N  D  P  A  L  V  K  V  L  S  S  H  S  K  D  S  V  D  W  M  T  A  M  G  A  228
                                                                                
ccgatttaactgatgttggcatgatgggtggcgcatctgttaatcgtgcgcatcgtccaaccggtggtgcgggtgttggtgctcatgttg 1620
  D  L  T  D  V  G  M  M  G  G  A  S  V  N  R  A  H  R  P  T  G  G  A  G  V  G  A  H  V  V  258
                                                                                
ttcaagtactttatgataatgcagtgaaacgcaatatcgacttacgcatgaacactcgcggcattgaagtgcttaaagatgataaaggca 1710
  Q  V  L  Y  D  N  A  V  K  R  N  I  D  L  R  M  N  T  R  G  I  E  V  L  K  D  D  K  G  T  288
                                                                                
ctgttaaaggtattctggttaaggggatgtacaaaggttactactgggtgaaagccgatgcggtaatcttagcaacgggtggtttcgcta 1800
  V  K  G  I  L  V  K  G  M  Y  K  G  Y  Y  W  V  K  A  D  A  V  I  L  A  T  G  G  F  A  K  318
                                                                                
aaaataacgagcgtgtcgcaaagcttgatccttcactaaaaggctttatctctactaaccaacctggtgcagtaggtgatggactggatg 1890
  N  N  E  R  V  A  K  L  D  P  S  L  K  G  F  I  S  T  N  Q  P  G  A  V  G  D  G  L  D  V  348
                                                                                
tagctgaaaatgcgggtggcgcattgaaagacatgcagtatatccaagctcacccaacactatctgttaaaggtggcgtaatggtcactg 1980
  A  E  N  A  G  G  A  L  K  D  M  Q  Y  I  Q  A  H  P  T  L  S  V  K  G  G  V  M  V  T  E  378
```

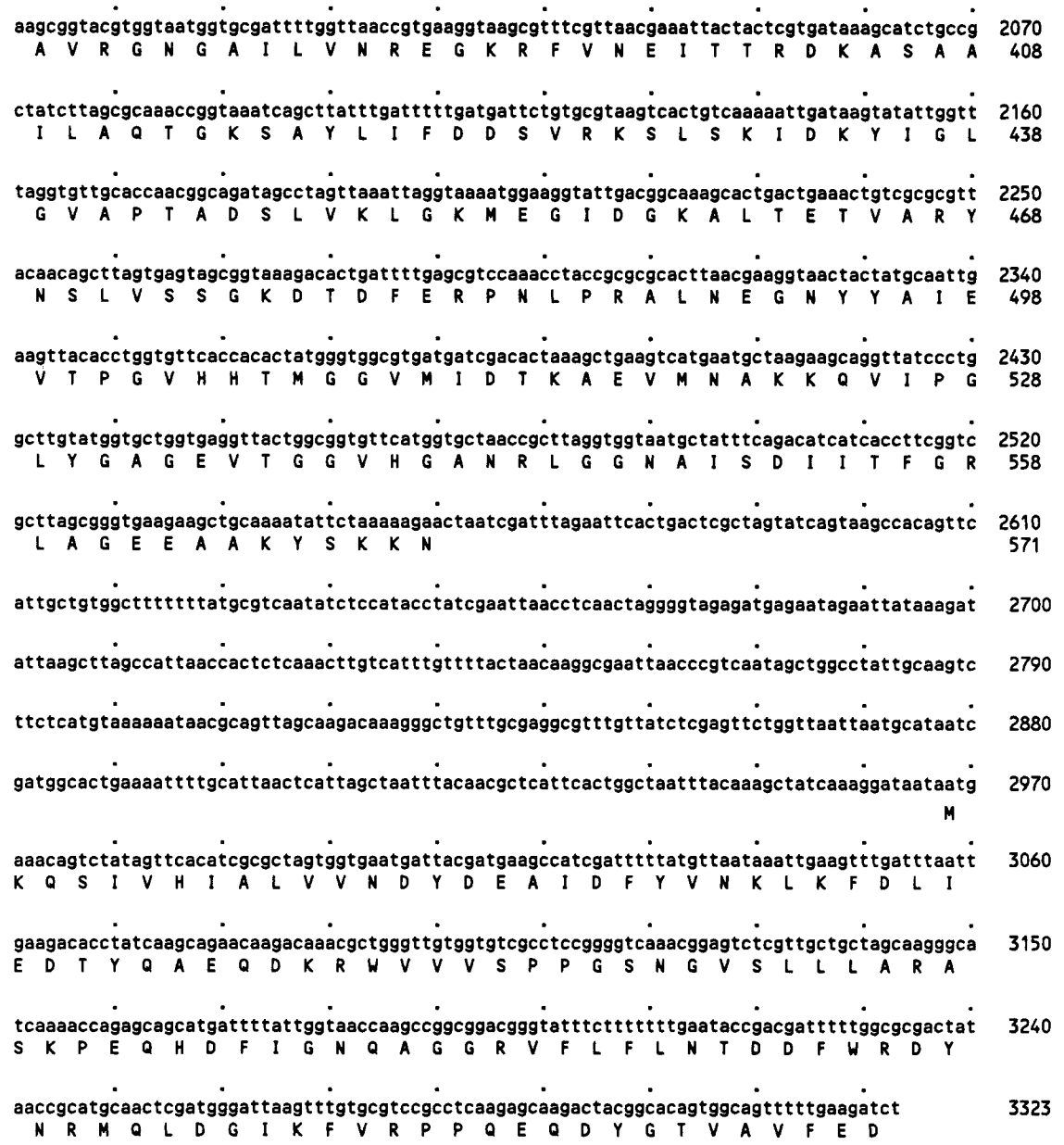

FIGURE 3: DNA sequence of the entire 3.3-kb *Bgl*II fragment containing the flavocytochrome *c* coding region indicated along with the translation throughout the reading frame. Two further reading frames are indicated, one of which is found upstream from flavocytochrome *c* and read from the other DNA strand (indicated in lowercase letters). The other has the same polarity as the flavocytochrome *c* coding sequence and extends to the end of the cloned fragment. The nucleotide sequence is numbered from the start of the cloned *Bgl*II fragment, and the flavocytochrome *c* polypeptide sequence is numbered from the start of the mature protein. The N-terminal sequence of the flavocytochrome *c* polypeptide is underlined. The translation of the flavocytochrome *c* coding sequence is shown from the second of five in-frame ATG codons preceding the mature N-terminus (see text for explanation).

reaction with bands of lower mobility than purified flavocytochrome *c* was only detected when flavocytochrome *c* itself was expressed and may be due to aggregation with other cellular proteins. Expression of flavocytochrome *c* is clearly induced by anaerobiosis and by the presence of fumarate in the growth medium (Figure 1). No cross-reaction was observed with *E. coli* fumarate reductase (not shown). This antiserum was subsequently used to isolate the flavocytochrome *c* coding sequence from an *S. putrefaciens* library constructed in the plasmid pEX3 (Stanley & Luzio, 1984). This vector is designed to direct the expression of a fusion protein consisting of the bulk of *E. coli* β-galactosidase with the insert-encoded peptide at its C-terminus. A single clone was identified that reacted with anti-flavocytochrome *c* antibodies through two rounds of screening. Proteins isolated from this transformant were subjected to Western blotting. A single major band was found with an apparent molecular mass of 45 kDa (not shown), much smaller than β-galactosidase and thus, apparently, not a fusion protein. The recombinant plasmid, pACB1, was isolated from this transformant and mapped by restriction enzyme analysis (Figure 2a). The 1.4-kb insert from pACB1 was subcloned into pTZ19R and pTZ18R, and its DNA sequence was determined. No extended open reading frame was found within over 300 base pairs from the cloning site, but a long reading frame extended, with the same polarity as that of the β-galactosidase coding sequence, for 1.05 kb to the end of the insert. This finding explains the absence of a fusion protein in pACB1 transformants, and it can be assumed that translation initiates within the cloned fragment.

Isolation of the Complete Flavocytochrome c Gene. DNA sequence analysis (described below) indicated that pACB1 contains only part of the flavocytochrome *c* coding sequence. Southern blot analysis of chromosomal DNA indicated that the entire coding sequence was contained on a 3.3-kb *Bgl*II fragment. To isolate the complete gene, we constructed a

library of *Bgl*II fragments of *S. putrefaciens* DNA in pTZ19R. The library was screened with a ^{32}P-labeled probe from the insert of pACB1. Positive clones were purified and shown to contain a recombinant plasmid with an insert of the expected size and with the restriction sites expected according to the map of the partial clone in pACB1. One recombinant plasmid, pSP1, was subjected to restriction mapping (Figure 2b), and the DNA sequence of the entire 3323-bp insert was determined on both strands using synthetic oligonucleotide primers (Figure 2d).

Analysis of the Flavocytochrome c Coding Sequence. DNA sequencing showed the existence of a long open reading frame encoding flavocytochrome *c* (Figures 2c and 3). The N-terminal amino acid sequence of the purified protein was determined as AQNLAEFHVQNQEXD, and this sequence is found close to the start of the reading frame. The unidentified residue 14 is expected to be one of the cysteines that is covalently linked to a heme group and would therefore not be released upon peptide bond cleavage during the fourteenth cycle of Edman degradation. The *c*-type cytochromes contain a consensus sequence for heme attachment of CxxCH. The two cysteines are linked through thioether bonds to vinyl groups of protoheme IX, and the histidine provides a ligand to the heme iron. Four CxxCH motifs are found in the N-terminal region of flavocytochrome *c* in what appears to comprise a domain of just over 100 amino acid residues.

The N-terminal alanine codon is preceded by five in-frame methionine codons, one of which is presumably the site for initiation of translation. We have predicted that one of these in particular, 25 codons upstream of the Ala1 codon, is the authentic initiation site because this would give rise to a typical bacterial secretory signal sequence (von Heijne, 1986a) with two positively charged residues at the extreme N-terminus followed by a hydrophobic, uncharged region and a signal peptidase cleavage site (von Heijne, 1986b). This predicted signal sequence is 25 residues in length and presumably directs the localization of flavocytochrome *c* to the periplasm where the mature enzyme is found (Black, 1991). None of the other potential initiation codons would result in a recognizable signal sequence.

The cytochrome domain is followed by a sequence (residues 129–143) typical of a wide range of FAD-containing enzymes and believed to interact with the adenine ring of FAD (Wood et al., 1984). This sequence is found at the N-terminus in many flavoproteins, and in the case of flavocytochrome *c* it marks the N-terminal end of a flavoprotein domain of over 450 residues. This domain exhibits extensive sequence similarity with the flavoprotein subunits of fumarate reductases and succinate dehydrogenases (Figure 4) and is clearly a member of the same family of proteins. FAD is covalently bound to a histidine residue in bovine heart mitochondrial succinate dehydrogenase, and the sequence of the flavopeptide (Kenney et al., 1972) can readily be aligned with the sequences of bacterial succinate dehydrogenase and fumarate reductase flavoproteins (Wood et al., 1984). In the membrane-bound enzymes with covalently-bound FAD, the modified histidine is found in a well-conserved sequence environment some 20 residues down from the adenine binding motif. No histidine is found near the corresponding positions in flavocytochrome *c* (Figure 4). Similarly, no histidine is found at the equivalent position in lipoamide dehydrogenase, mercuric reductase, and other enzymes with noncovalently bound FAD, nor in Osm1p (see below).

Extensive searching of protein sequence databases revealed a close relationship with the product (Osm1p) of the *Saccharomyces cerevisiae OSM*1 gene (Melnick & Sherman, 1990). Mutations in the *OSM*1 gene result in sensitivity to high osmolarity in the external medium, but the biochemical function of the protein product has not been identified. Melnick and Sherman (1990) suggested on the basis of sequence similarity that this polypeptide is a member of the G-protein family. This now seems very unlikely since Osm1p is much more similar to flavocytochrome *c* than to G-proteins, the relationship with the latter being restricted to the nucleotide-binding region that is also related to the FAD-binding region in flavoproteins. The published sequence indicates that the *OSM*1 gene encodes a protein of 301 amino acid residues. We have found that one of the other reading frames extends the relationship with flavocytochrome *c* for a further 150 residues, and we assume that this is accounted for by a sequence error resulting in an apparent frameshift. For the purposes of sequence comparison, we have assumed that a single C has been omitted from codon 281 but this remains to be experimentally verified. This proposed correction would restore the sequence QxHPTG that is well-conserved throughout the frd family, with the histidine in this sequence having been identified at the active site in *E. coli* frdA (Schröder et al., 1991).

The multiple alignment program PILEUP generates a dendrogram representing the relatedness of proteins in a family. We have used this program to compare the sequences of flavocytochrome *c*, *OSM*1 protein, and the flavoprotein subunits of fumarate reductases and succinate dehydrogenases, and the derived dendrogram (Figure 5) demonstrates that flavocytochrome *c* is more closely related to OSM1 (31% identity) than to any of the membrane-bound flavoproteins, though the relationship with these latter proteins is very clear. For example, flavocytochrome *c* and *E. coli* frdA are 26% identical.

The similarity of the flavin domain of flavocytochrome *c* to other flavoproteins indicates that it will fold similarly in three dimensions and thus independently of the N-terminal cytochrome domain. Similarly, the membrane-bound fumarate reductase and succinate dehydrogenase flavoproteins extend about 200 residues beyond the position equivalent to the C-terminus of flavocytochrome *c*. We would therefore expect this to fold as a separate domain in these enzymes. The domain structure of these different proteins is shown schematically in Figure 6.

In addition to the flavocytochrome *c* coding sequence, two further reading frames are found within the cloned *Bgl*II fragment (Figure 2). One of these has the opposite polarity compared with flavocytochrome *c* and extends for 123 codons to one end of the cloned DNA and presumably beyond. We have been unable to identify homologues of the presumed product of this sequence by database searching. Downstream from the flavocytochrome *c* coding sequence is a reading frame of 118 codons that extends to the end of the cloned fragment. The predicted protein sequence exhibits some similarity with the 13-kDa subunit of *E. coli* fumarate reductase (Figure 7) with 22% identity over the available sequence. The similarity with the 13-kDa subunit of *Proteus vulgaris* fumarate reductase and with *E. coli* succinate dehydrogenase is less obvious, but this possible relationship is particularly interesting because of the proximity of this coding sequence in the *S. putrefaciens* genome to the flavocytochrome *c* coding sequence. We have not attempted to analyze transcription of the cloned region in *S. putrefaciens*. The upstream sequence must be

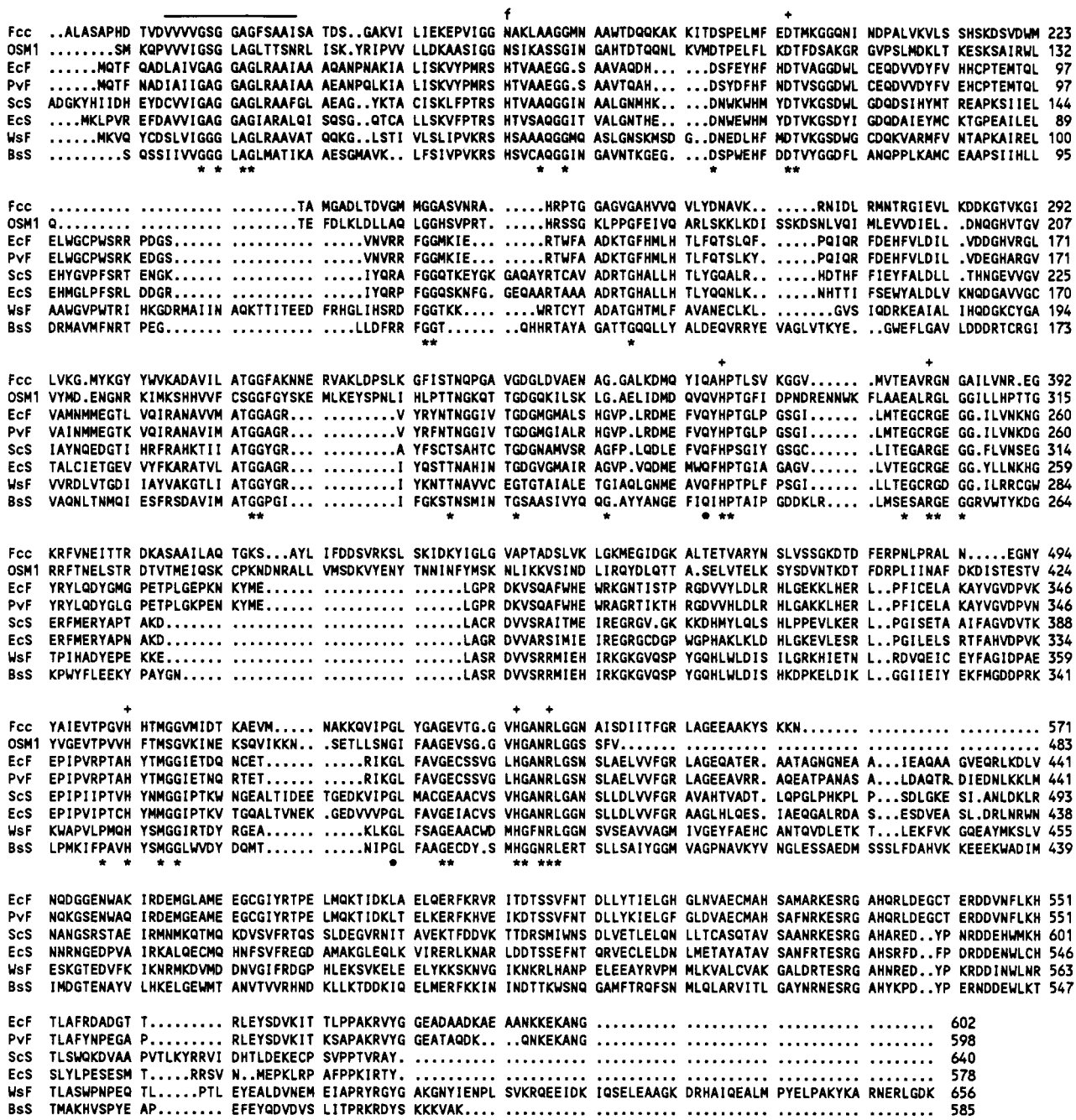

FIGURE 4: Sequence alignment of the fumarate reductase family. The sequences are labeled as follows: Fcc, flavocytochrome *c*; OSM1, the yeast Osm1 protein (Melnick & Sherman, 1990); EcF, *E. coli* frdA (Cole, 1982); PvF, *P. vulgaris* frdA (Cole, 1987), ScS, *S. cerevisiae* sdhA (Chapman et al., 1992; Robinson & Lemire, 1992); EcS, *E. coli* sdhA (Wood et al., 1984); WsF *Wolinella succinogenes* frdA (Lauterbach et al., 1990); and BsS, *Bacillus subtilis* sdhA (Phillips et al., 1987). The sequences were aligned using the program PILEUP in the University of Wisconsin Genetics Computer Group (UWGCG) package. This program first compares all possible pairs of sequences and groups these in clusters according to their pairwise sequence similarity. The alignment begins with the most similar sequences which are aligned within the cluster, and these are then extended by alignment with further sequences and the other clusters. The FAD-binding region is overlined, and the position of the histidine through which FAD is covalently attached to some of these proteins is indicated above by the letter f. Residues conserved in all 8 sequences are indicated below the alignment with an asterisk (*), and those residues implicated in the active site are marked with a plus (+) above the aligned sequences. The sequence of the OSM1 protein is identical to that translated from the published DNA sequence up to His280 but then continues in the +1 reading frame. The signal sequence and cytochrome domain of flavocytochrome *c*, the extreme N-terminus of the OSM1 protein, and the mitochondrial targeting sequence from yeast succinate dehydrogenase are omitted from the alignment.

transcribed separately from flavocytochrome *c* since the two reading frames have opposite polarities. The downstream sequence is separated from the flavocytochrome *c* coding sequence by 406 nucleotides, and the two may be cotranscribed.

DISCUSSION

We have determined the sequence of the DNA encoding flavocytochrome *c* from *Shewanella putrefaciens*. The protein is apparently synthesized with an N-terminal signal sequence that leads to its localization to the periplasmic space. The mature protein is predicted to contain two distinct domains, an N-terminal tetraheme cytochrome domain and a larger C-terminal flavoprotein domain which is the site of fumarate reduction. Flavocytochrome *c* has been shown biochemically to contain at least four *c*-type heme groups, and the presence of four heme-binding consensus sequences is consistent with a stoichiometry of four heme groups per polypeptide chain. As with other *c*-type cytochromes, one of the axial ligands is

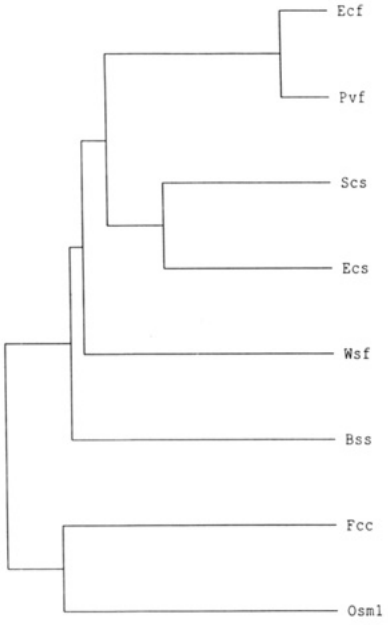

FIGURE 5: Fumarate reductase family tree. The dendrogram shows the output of the UWGCG program PILEUP, indicating clusters on the basis of sequence similarity. The protein names are as in Figure 4. The more closely related sequences are joined by branches lower down in the figure, for example frdA from *E. coli* and *P. vulgaris*. The diagram shows that flavocytochrome *c* and Osm1p are more closely related to one another than they are to the other members of the family.

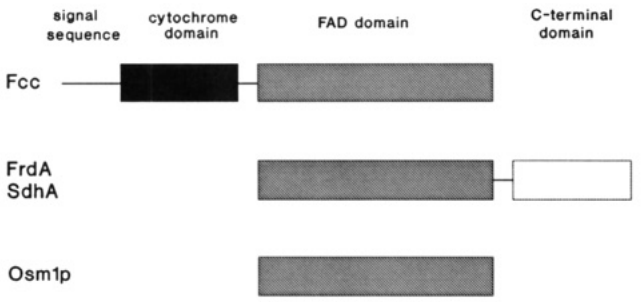

FIGURE 6: Domain structure of flavocytochrome *c* and its relatives. Flavocytochrome *c*, Osm1p, and the flavoprotein subunits of membrane-bound fumarate reductases and succinate dehydrogenases share extensive sequence similarity over a region of about 450 amino acid residues. This is inferred to form a structurally and functionally distinct domain in each protein. Flavocytochrome *c* contains an N-terminal cytochrome domain not found in other members of the family, and the membrane-bound proteins contain an additional C-terminal domain of unknown function.

```
ORF2  MKQSIVHIAL VVNDYDEAID FYVNKLKFDL IEDTYQAEQD KRWVVVSPPG
FrdD  .......... MINPNPKRSD ...EPVFWGL FGAG.....G MWSAIIAPVM
Cons  ---------- --N------D ---------L ---------- -------P--

ORF2  SNGVSLLLAR ASKP...... EQHDFIGNQA GGRVFLFLNT D.DFWRDYNR
FrdD  ILLVGILLPL GLFPGDALSY ERVLAFAQSF IGRVFLFLMI VLPLWCGLHR
Cons  ---V--LL-- ---P------ E--------- -GRVFLFL-- ----W----R

ORF2  MQ..LDGIKF VRPPQEQDYG TVAVFED... .......
FrdD  MHHAMHDLKI HVPAGKWVFY GLAAILTVVT LIGVVTI
Cons  M-------K- --P------- --A------- -------
```

FIGURE 7: Alignment of the predicted protein sequence 3′ to the flavocytochrome *c* coding sequence (ORF2) with the 13-kDa subunit of *E. coli* fumarate reductase (FrdD). The sequences were aligned using the UWGCG program GAP. The consensus (Cons) where both sequences are identical is also shown.

provided by a histidine residue immediately adjacent to the heme attachment site. Spectroscopic studies, particularly electron paramagnetic resonance, indicate that flavocytochrome *c* contains two very different pairs of hemes. Similarly redox potentiometry indicates that two hemes have a midpoint potential of −204 mV, whereas the other pair titrates at −320 mV (Morris, 1987). The differences between the two pairs of hemes could be due to different protein environments or to differences in the coordination of the heme iron. Magnetic circular dichroism indicates that the sixth ligand in each case is most likely also histidine and is clearly not methionine as in some *c*-type cytochromes (our unpublished results). The cytochrome domain contains 10 histidine residues in all, more than the eight required for bishistidine ligation of the four hemes. Two lysine residues are found within the cytochrome domain and it is possible, but perhaps less likely, that two of the hemes are His–Lys coordinated.

The association of four heme groups with a relatively small protein is not unprecedented. Cytochrome c_3 from *Desulfovibrio* contains four *c*-type hemes in a polypeptide of a little over 100 amino acid residues (Ambler et al., 1971). Apart from the similarity of the heme attachment sites, there is no obvious relationship between the sequences of the cytochrome domain of flavocytochrome *c* and cytochrome c_3. As in flavocytochrome *c* (Morris, 1987; A. C. Black, S. L. Pealing, S. K. Chapman, F. B. Ward, and G. A. Reid, unpublished experiments), the heme groups of cytochrome c_3 have very low midpoint redox potentials (Pettigrew & Moore, 1987) but the latter is involved in sulfur metabolism and thus not apparently functionally related to the former. In neither case is it clear why four heme redox centers should be required for two electron transfer reactions—reduction of flavin in the case of flavocytochrome *c* and sulfur reduction in cytochrome c_3. The cytochrome domain of flavocytochrome *c* shows no sequence similarity to other known tetraheme cytochromes, including the cytochromes *c* found in the photosynthetic reaction centers of *Rhodopseudomonas viridis* (Weyer et al., 1987), nor to that found in association with nitrite reductase (cytochrome cd_1) from *Pseudomonas stutzeri* (Jungst et al., 1991), nor to that found in *Chromatium vinosum* (Dolata et al., 1992), encoded adjacent to a flavocytochrome *c* gene.

Flavocytochromes *c* have been isolated from several other bacterial sources (Cusanovich et al., 1991). Flavocytochrome *c* from *Pseudomonas putida* is a *p*-cresol methylhydroxylase composed of $\alpha_2\beta_2$ tetramers with covalently-bound FAD in the larger α subunits and a single heme in the β subunit (Mathews et al., 1991). The prosthetic groups of flavocytochrome *c* from *Chromatium vinosum* are also contained in separate subunits, but in this case the heme subunit contains two covalently-bound heme groups (Cusanovich et al., 1991; Dolata et al., 1992). The flavocytochrome *c* from *Chlorobium thiosulfatophilum* is, like the *Chromatium* enzyme, involved in sulfur metabolism but contains a monoheme cytochrome subunit. Sequence analysis reveals no significant similarity of any of the heme or flavoprotein subunits of these other flavocytochromes *c* to the *Shewanella* fumarate reductase.

The flavoprotein domain of flavocytochrome *c* is structurally and functionally related to the flavoprotein subunits of several fumarate reductases and succinate dehydrogenases. The enzymes from this family that have thus far been sequenced, other than flavocytochrome *c*, are all membrane-bound and consist of four distinct subunits. Membrane binding is apparently mediated by the two small subunits (Lemire et al., 1982), and overexpression of the *E. coli* frd operon can lead to apparent saturation of membrane sites with the excess enzyme accumulating in novel tubular lipid–protein structures (Weiner et al., 1984; Elmes et al., 1986). Flavocytochrome *c* behaves as a soluble, periplasmic protein, but the finding of a sequence encoding an apparent relative of one of the small membrane anchor subunits from *E. coli* fumarate reductase immediately adjacent to the flavocytochrome *c* coding sequence is intriguing. The *S. putrefaciens* sequence is less hydrophobic

than the *E. coli* polypeptide, and it may not function in precisely the same way. The 13-kDa and 15-kDa subunits of *E. coli* fumarate reductase, in addition to their membrane-binding function, are required for electron transfer from the physiological reductant menaquinol since mutations in the frdC and frdD genes lead to a loss of quinol oxidation (Cecchini et al., 1986; Weiner et al., 1986; Westenberg et al., 1990).

By comparing the sequence of flavocytochrome *c* with those of its membrane-bound relatives, we have inferred that the latter proteins consist of a large flavoprotein domain and a smaller C-terminal domain (Figure 6). The function of this C-terminal domain is unknown, but its absence from the single-subunit, soluble fumarate reductase, flavocytochrome *c*, clearly indicates that it is not required for fumarate reduction per se but may be involved in intersubunit interactions or perhaps in interactions with electron donors. The product of the *E. coli* nadB gene is an aspartate oxidase required for the synthesis of quinolinate, a precursor for nicotinamide nucleotide biosynthesis (Flachmann et al., 1988; Seifert et al., 1990). This protein is closely related to the fumarate reductases, with all the active site residues identified to date being conserved. Interestingly, the similarity extends also to the C-terminal domain. Aspartate oxidase is a component of a soluble quinolinate synthase complex, along with the nadA gene product; therefore its interactions with other polypeptides are very different from those of the membrane-bound fumarate reductases.

The sequence similarity between flavocytochrome *c* and the *S. cerevisiae OSM*1 gene product was unexpected and may give some clues to the function of Osm1p in yeast. Mutations in the *OSM*1 gene result in osmotic sensitivity, but this is not necessarily a direct effect of the lesion in Osm1p. Osmoregulation is a complex process, and mutations in several genes affecting a wide range of metabolic processes result in an osmosensitive phenotype in yeast. Our sequence comparisons indicate a probable error in the published *OSM*1 DNA sequence such that the reading frame is considerably longer than originally predicted. We have inferred a specific correction on the grounds that extensive sequence similarity is found both before and after the assumed frame-shift, and the precise modification restores a well-conserved sequence around the active site histidine. The similarity with flavocytochrome *c* extends along the full length of Osm1p except for the putative N-terminal signal sequence (Melnick & Sherman, 1990). The degree of similarity of Osm1p with flavocytochrome *c* and other members of the fumarate reductase family leads us to suggest that Osm1p may be a fumarate reductase and may correspond to a soluble enzyme characterized previously (Muratsubaki & Katsume, 1985). This enzyme has a molecular mass of 58 kDa, as estimated by SDS–PAGE (Muratsubaki & Katsume, 1982), compared with 53 kDa calculated for the corrected sequence of Osm1p. The fumarate reductase activity in yeast is cytosolic, but Osm1p has been predicted to enter the secretory pathway because of a putative signal sequence at the N-terminus (Melnick & Sherman, 1990). This sequence does not, however, conform well to the criteria noted by von Heijne (1985) for signal sequences, and it seems likely that it is not a functional targeting sequence. Osm1p is therefore likely to be a cytosolic protein.

Arg248 of *E. coli* frdA has been shown by site-directed mutagenesis to be required for efficient fumarate reductase activity (Schröder et al., 1991). This residue is conserved in all known members of the family (Figure 4) and has been predicted to be involved in substrate binding (Kotlyar & Vinogradov, 1984; Schröder et al., 1991). Its conservation in flavocytochrome *c* and Osm1p, which were not previously known to be associated with this family of proteins, reinforces the importance of this residue. Schröder et al. (1991) suggested that a second arginine might be involved in binding substrate. The arginine guanidinium group not only is particularly well suited to electrostatic interaction with a substrate carboxylate but also is capable of hydrogen bonding, thereby orientating substrate within the active site. We have found that, apart from Arg381 in flavocytochrome *c* (Arg248 in *E. coli* frdA), only one other arginine residue is conserved throughout the family and this is found at position 544 in flavocytochrome *c* and 390 in frdA (Figure 4).

Similarly, a histidine residue has been implicated in succinate dehydrogenase activity and proposed to act as a proton donor/acceptor (Vinogradov, 1986). His232 of *E. coli* fumarate reductase has been substituted by serine with the result that enzyme activity is reduced by 75% in the direction of fumarate reduction but the rate of succinate dehydrogenation is lowered to 1.6% compared with that of the wild-type enzyme (Schröder et al., 1991). This histidine is apparently modified by the inhibitor, Rose Bengal, since the His232Ser mutant enzyme was insensitive to Rose Bengal. The data are consistent with a role for this residue as an active site base, though the retention of 25% of the wild-type fumarate reductase activity would indicate that an alternative proton donor/acceptor can functionally replace His232. This histidine is conserved in all members of the family and is most likely to act as a proton donor to fumarate.

The finding of an apparent relative of one of the smaller subunits of *E. coli* fumarate reductase associated genetically with flavocytochrome *c* raises intriguing questions relating to the interaction of flavocytochrome *c* with other proteins. The *E. coli* fumarate reductase is reduced by menaquinol, but the nature of the reductant of flavocytochrome *c* has yet to be determined. Flavocytochrome *c* is isolated as a soluble, monomeric, periplasmic protein. Its interaction with other proteins may, however, depend upon ionic conditions such that dissociation is favored when the protein is isolated. The possible existence of other genetically linked proteins that might also interact with flavocytochrome *c* cannot be confirmed without extending the sequence determination.

ACKNOWLEDGMENT

We are grateful to Bryan Dunbar for assistance with the protein sequence determination, to OSWEL Oligonucleotide Synthesis Unit for the synthesis of sequencing primers, to the SEQNET computing facility for access to software and databases, to Bernard Lemire for helpful comments, and to Andrew Coulson for help with sequence similarity searches using the distributed array processor.

REFERENCES

Ambler, R. P., Bruschi, M., & LeGall, J. (1971) *FEBS Lett. 18*, 347–350.

Black, A. C. (1991) Ph.D. Thesis, University of Edinburgh.

Cecchini, G., Thompson, C. R., Ackrell, B. A. C., Westenberg, D. J., Dean, N., & Gunsalus, R. P. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8898–8902.

Chapman, K. B., Solomon, S. D., & Boeke, J. D. (1992) *Gene 118*, 131–136.

Cole, S. T. (1982) *Eur. J. Biochem. 122*, 479–484.

Cole, S. T. (1987) *Eur. J. Biochem. 167*, 481–488.

Cole, S. T., Grundström, T., Jaurin, B., Robinson, J. J., & Weiner, J. H. (1982) *Eur. J. Biochem. 126*, 211–216.

Collins, J. F., Coulson, A. F. W., & Lyall, A. (1988) *Comput. Appl. Biosci. 4*, 67–71.
Coulson, A. F. W., Collins, J. F., & Lyall, A. (1987) *Comput. J. 30*, 420–424.
Cusanovich, M. A., Meyer, T. E., & Bartsch, R. G. (1991) In *Chemistry and Biochemistry of Flavoenzymes* (Muller, F., Ed.) Vol. II, pp 377–393, CRC Press, Boca Raton, FL.
Devereux, J., Haeberli, P., & Smithies, O. (1984) *Nucleic Acids Res. 21*, 387–395.
Dolata, M. M., van Beeumen, J. J., Ambler, R. P., Meyer, T. E., & Cusanovich, M. A. (1992) *Biochemistry* (in press).
Elmes, M. L., Scraba, D. G., & Weiner, J. H. (1986) *J. Gen. Microbiol. 132*, 1429–1439.
Flachmann, R., Kunz, N., Seifert, J., Gütlich, M., Wientjes, F.-J., Läufer, A., & Gassen, H. G. (1988) *Eur. J. Biochem. 175*, 221–228.
Haid, A., & Suissa, M. (1984) *Methods Enzymol. 96*, 192–205.
He, S. H., DerVartanian, D. V., & LeGall, J. (1986) *Biochem. Biophys. Res. Commun. 135*, 1000–1007.
Jungst, A., Wakabayashi, S., Matsubara, H., & Zumft, W. G. (1991) *FEBS Lett. 279*, 205–209.
Kenney, W. C., Walker, W. H., & Singer, T. P. (1972) *J. Biol. Chem. 247*, 4510–4513.
Kotlyar, A. B., & Vinogradov, A. D. (1984) *Biochem. Int. 8*, 545–552.
Lauterbach, F., Körtner, C., Albracht, S. P. J., Unden, G., & Kröger, A. (1990) *Arch. Microbiol. 154*, 386–393.
Lemire, B. D., Robinson, J. J., & Weiner, J. H. (1982) *J. Bacteriol. 152*, 1126–1131.
Mathews, F. S., Chen, Z.-W., Bellamy, H. D., & McIntire, W. S. (1991) *Biochemistry 30*, 238–247.
Melnick, L., & Sherman, F. (1990) *Gene 87*, 157–166.
Morris, C. J. (1987) Ph.D. Thesis, University of Edinburgh.
Morris, C. J., Gibson, D. M., & Ward, F. B. (1990) *FEMS Microbiol. Lett. 69*, 259–262.
Muratsubaki, H., & Katsume, T. (1982) *Agric. Biol. Chem. 46*, 2909–2917.
Muratsubaki, H., & Katsume, T. (1985) *J. Biochem. 97*, 1201–1209.
Pettigrew, G. W., & Moore, G. R. (1987) *Cytochromes c, Biological Aspects*, Springer Verlag, Berlin.
Phillips, M. K., Hederstedt, L., Hasnain, S., Rutberg, L., & Guest, J. R. (1987) *J. Bacteriol. 169*, 864–873.
Remaut, E., Tsao, H., & Fiers, W. (1983) *Gene 22*, 103–113.
Robinson, K. M., & Lemire, B. D. (1992) *J. Biol. Chem. 267*, 10101–10107.
Rokeach, L. A., Haselby, J. A., & Hoch, S. O. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4832–4836.
Rossi, C., Hauber, J., & Singer, T. P. (1964) *Nature 204*, 167–170.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning. A laboratory manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Schröder, I., Gunsalus, R. P., Ackrell, B. A. C., Cochran, B., & Cecchini, G. (1991) *J. Biol. Chem. 266*, 13572–13579.
Seifert, J., Kunz, N., Flachmann, R., Läufer, A., Jany, K.-D., & Gassen, H.-G. (1990) *Biol. Chem. Hoppe-Seyler 371*, 239–248.
Stanley, K. K., & Luzio, J. P. (1984) *EMBO J. 3*, 1429–1434.
Vieira, J., & Messing, J. (1987) *Methods Enzymol. 153*, 3–11.
Vinogradov, A. D. (1986) *Biokhimiya 51*, 1944–1973.
von Heijne, G. (1985) *J. Mol. Biol. 184*, 99–105.
von Heijne, G. (1986a) *J. Mol. Biol. 189*, 239–242.
von Heijne, G. (1986b) *Nucleic Acids Res. 14*, 4683–4690.
Weiner, J. H., & Dickie, P. (1979) *J. Biol. Chem. 154*, 8590–8593.
Weiner, J. H., Lemire, B. D., Elmes, M. L., Bradley, D. G., & Scraba, D. G. (1984) *J. Bacteriol. 158*, 590–596.
Weiner, J. H., Cammack, R., Cole, S. T., Condon, C., Honoré, U., Lemire, B. D., & Shaw, G. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 2056–2060.
Westenberg, D. J., Gunsalus, R. P., Ackrell, B. A. C., & Cecchini, G. (1990) *J. Biol. Chem. 265*, 19560–19567.
Weyer, K. A., Lottspeich, F., Grünberg, H., Lang, F., Oesterhelt, D., & Michel, H. (1987) *EMBO J. 6*, 2197–2202.
Wood, D., Darlison, M. G., Wilde, R. J., & Guest, J. R. (1984) *Biochem. J. 222*, 519–534.