Novel phosphotransferase system genes revealed by bacterial genome analysis – a gene cluster encoding a unique Enzyme I and the proteins of a fructose-like permease system

Jonathan Reizer, Aiala Reizer and Milton H. Saier, Jr

Author for correspondence: Milton H. Saier, Jr. Tel: +1 619 534 4084. Fax: +1 619 534 7108. e-mail: msaier@ucsd.edu

Department of Biology, University of California at San Diego, La Jolla, CA 92037–0116, USA

Previous publications have demonstrated the presence of a cryptic gene encoding a novel Enzyme I of the phosphoenolpyruvate:sugar phosphotransferase system (PTS). Recent Escherichia coli genome sequencing revealed a gene (ptsA) encoding a new Enzyme I homologue in the 89.1–89.3 centisome region. We have analysed this region, and here describe and characterize open reading frames (ORFs) encoding (1) a fused PTS Enzyme I-IAIIBF homologue, (2) a glycerol dehydrogenase, (3) a transaldolase homologue, (4) two PTS IIIBF homologues, (5) a PTS IICF homologue, and (6) homologues of pyruvate formate-lyase and its activating enzyme. Binary comparison scores, multiple alignments and phylogenetic trees establish the families of proteins to which each of the relevant ORFs belong. Identification of the putative products of this gene cluster leads to the proposal that several of the proteins encoded in this region function in anaerobic carbon metabolism.

Keywords: sugar transport, phosphotransferase system, anaerobiosis, glycerol dehydrogenase, pyruvate formate-lyase

INTRODUCTION

Bacteria take up and concomitantly phosphorylate sugars via a phosphoenolpyruvate:sugar phosphotransferase system (Postma et al., 1993). This system transports dozens of sugars via a chain of phosphoryl transfer proteins that consists of two non-sugar-specific components, Enzyme I (I) and HPr, and three or four protein domains of the sugar-specific Enzyme I1 complex (Enzymes or domains IIA, IIB, IIC and sometimes IID) (Saier & Reizer, 1992). The phosphoryl transfer chain of the PTS, regardless of the sugar transported, is as follows:

\[ \text{PEP} \rightarrow \text{I} \rightarrow \text{HPr} \rightarrow \text{IIA} \rightarrow \text{IIB} \rightarrow \text{sugar} \]

Of these proteins or protein domains, only the IIC component, the sugar recognition constituent that transports the sugar, is not phosphorylated.

IIA and IIB are cytoplasmic proteins or peripherally membrane-associated domains of the Enzyme II complex, and IIC is an integral membrane permease. The phylogenies of these protein domains have been characterized in previous publications (Reizer et al., 1991b; Saier et al., 1992; Saier & Reizer, 1992, 1994a). The characterized fructose-specific and fructose-like Enzyme II complexes exhibit IIA and IIB domains that resemble those of the mannitol-specific Enzyme II complex, but the IIC domains are more divergent (see Reizer et al., 1994a).

In 1987 we published results showing that strains of Salmonella typhimurium that were deleted for the pts genes encoding Enzyme I and HPr could be mutated so as to regain PTS function (Chin et al., 1987). A novel Enzyme I was identified and partially characterized, and it proved to have catalytic properties similar to those of the deleted Enzyme I (S. Sutrina, J. Reizer & M. H. Saier, Jr, unpublished results). The gene, designated ptsI, which when mutated gave rise to expression of this novel Enzyme I appears to encode a transcriptional regulatory protein within the GntR family (Reizer et al., 1991a; Titgemeyer et al., 1994).

Recent operon and genome sequence analyses in Escheri-
**METHODS**

**Computer-aided analyses.** All sequence analyses and database searches were performed using the GCG package from the University of Wisconsin (Devereux et al., 1984) and the DNAsystem package (Smith, 1988). Comparison scores (expressed in SD) were calculated using the RDF2 program with 100 or 200 shuffles as indicated in the table legends. A value of 6 SD is suggestive of homology whereas a value of 9 SD establishes homology. Construction of phylogenetic trees and estimation of the relative evolutionary distances among members of a protein family were as described by Reizer & Reizer (1994) using the progressive alignment method of Feng & Doolittle (1990). Mean hydrophyt and mean similarity were calculated using a sliding window of 20 residues. Hydropathy plots were designed according to Kyte & Doolittle (1982). The calculated mean similarity was an arithmetic mean of all pairwise comparisons.

**RESULTS**

**Organization and identification of genes in the 89.1–89.3 centosome region of the *E. coli* genome**

Fig. 1 presents the gene structure of the 89.1–89.3 centosome region of the *E. coli* chromosome, recently sequenced as part of the *E. coli* genome project (Blattner et al., 1993). The figure presents (1) the proposed gene designations, (2) the directions of transcription, (3) the nature of the gene products, (4) the functional assignments of the protein products based on homology to previously characterized proteins, (5) the GC contents of the open reading frames (ORFs), (6) the codon adaptation indices (CAI) of the corresponding ORFs, and (7) information regarding the putative intercistronic regions. This region contains nine clearly identifiable ORFs, three of which probably comprise a leftwardly directed operon. orf f711 (ptsA) codes for a 711 amino acid protein that exhibits sequence similarity to Enzyme I of the PTS in its N-terminal domain and to Enzyme IIAFrU in its C-terminal domain. This protein will hereafter be designated Enzyme IAni (IIA-included). orf f220 (talC) encodes a protein (TalC) that falls into the transaldolase family, while orf f380 (gldA) encodes a glycerol dehydrogenase (GldA) (Truniger & Boos, 1994). In the rightward direction, a short ORF, orf a109, encodes a serine-rich protein lacking significant sequence similarity to all proteins in the current databases. Following a large intercistronic region of 341 bp, orf a359 (frwC) encodes a IICFrW-like protein (IICFrW or FrwC). Fourteen basepairs downstream of orf a359 is the initiation codon (ATG) for orf a106, encoding the first of two IIBFrW-like proteins (here designated IIBFrW or FrwB). Thereafter follow orf a765 (pfLD), encoding a protein homologous to pyruvate formate-lyase (PflD), orf a315 (pfC), encoding a pyruvate-formate-lyase-activating enzyme homologue (PflC), and orf a113, encoding a second IIBFrW-like protein (IIBFrW or FrwD).

**Fig. 1.** Genetic map of the GldA-Enzyme IAni-Frw PTS region in *E. coli*. The map shows the proposed gene assignments for the DNA region that comprises nine ORFs. Designations for the encoded proteins are also provided. The number of basepairs in the eight intergenic regions is given in parentheses. A minus sign denotes an overlapping region. Calculated codon adaptation indices (CAI; Sharp & Li, 1987) and the GC content of each gene are indicated below that gene.
**Novel phosphotransferase system genes**

**Fig. 2.** Phylogenetic tree of the family of proteins that include the bacterial, protozoan and plant pyruvate phosphate dikinases (PPDKs), phosphoenolpyruvate synthase (PPS) of *E. coli* and the Enzyme I proteins and protein domains of the bacterial PTS. Branch length (in arbitrary units) is proportional to phylogenetic distance. Abbreviations used and references to the published protein sequences are as follows: PPS, phosphoenolpyruvate synthase of *E. coli* (Eco) (Niersbach et al., 1992); PPDK, pyruvate phosphate dikinases of *Bacteroides symbiosus* (Bsy) (Pocalyko et al., 1990), *Flaveria trinervia* (Ftr) (Rosche & Westhoff, 1990), *Entamoeba histolytica* (Ehi) (Bruchhaus & Tannich, 1993), and *Zea mays* (Zma) (Matsuoka et al., 1988); Enzyme I, Enzymes I of *Alcaligenes eutrophus* (Aeu) (Pries et al., 1991), *Escherichia coli* (Eco) (Saffen et al., 1987; De Reuse & Danchin, 1988), *Rhodobacter capsulatus* [Rca; fructose specific (Fru); Wu et al., 1990], *Staphylococcus carnosus* (Sca) (Kohlbrecher et al., 1991), *Streptococcus salivarius* (Ssa) (Gagnon et al., 1992), *Streptococcus mutans* (Smu) (Boyd et al., 1994), and *Salmonella typhimurium* (Sty) (Byrne et al., 1988; LiCalsi et al., 1991). The Enzyme I protein domain (Enzyme I*) is encoded by the *ptsA* gene of *E. coli* (Eco) described here. Not presented is the homologous, C-terminal region (110 residues) of the *Bacillus stearothermophilus* pyruvate kinase (Sakai & Ohta, 1993).

**Table 1.** Binary comparisons of the seven transaldolase family (TAF) members

Values in parentheses below the designations of the proteins refer to the numbers of residues in the intact proteins. The FASTA program using the dipeptide identities mode (ktup = 2) (Pearson & Lipman, 1988) was used to assess similarities of the indicated proteins. Values presented in the table which are not in brackets or parentheses represent percentage identities for segments having the numbers of compared residues indicated in parentheses. Comparison scores in standard deviations, using the RDF2 program (Pearson & Lipman, 1988) and 150 shuffles are given in brackets below the values for percentage identity. Abbreviations and references to the published sequences are as follows. The hypothetical protein in the spoOF region of *B. subtilis* [ORFU (Bsu); Trach et al., 1988]; the ORF (ORF220) located downstream of *ptsA* of *E. coli* [ORF220 (Eco); Blattner et al., 1993]; transaldolase of *Saccharomyces cerevisiae* [Tal1 (Sce); Schauf et al., 1990]; transaldolase of *Kluyveromyces lactis* [Tal1 (Kla); Jacoby et al., 1993]; transaldolase of *Homo sapiens* [Tal (Hsa); Banki et al., 1994]; transaldolase of *E. coli* [TalB (Eco); Yura et al., 1992]. The partially sequenced transaldolase A of *E. coli* (G. A. Sprenger, unpublished data cited in SwissProt, accession no. P80218) is abbreviated TalA (Eco).

<table>
<thead>
<tr>
<th></th>
<th>ORFU (Bsu) (186)</th>
<th>Tal1 (Sce) (335)</th>
<th>Tal1 (Kla) (334)</th>
<th>Tal (Hsa) (336)</th>
<th>TalB (Eco) (317)</th>
<th>TalA (Eco) (40; fragment)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tal1 (Sce) (335)</td>
<td>76 (335) [137]</td>
<td>58 (333) [109]</td>
<td>55 (317) [109]</td>
<td>53 (36) [100]</td>
<td>53 (36) [100]</td>
<td></td>
</tr>
<tr>
<td>Tal1 (Kla) (334)</td>
<td>59 (332) [109]</td>
<td>57 (318) [87]</td>
<td>49 (39) [9]</td>
<td>49 (39) [9]</td>
<td>49 (39) [9]</td>
<td></td>
</tr>
<tr>
<td>TalB (Eco) (317)</td>
<td>80 (40) [22]</td>
<td>80 (40) [22]</td>
<td>80 (40) [22]</td>
<td>80 (40) [22]</td>
<td>80 (40) [22]</td>
<td></td>
</tr>
</tbody>
</table>
Fig. 3. Mean hydropathy (a) and similarity (b) plots of the fructose IIC-like family (FCF) members. Mean similarity and hydropathy values were calculated as described in Methods. The mean similarity across the entire alignment in (b) is represented by the dashed line. The signature sequence of the FCF proteins and its location are shown above the mean similarity plot in (b). Residues in brackets indicate alternative possibilities at a single position; X, any residue.

Fig. 4. Evolutionary tree of the fructose IIC-like family (FCF) of proteins or protein domains. Relative phylogenetic distances (in arbitrary units) are provided adjacent to the branches. Abbreviations and references to the published sequences are as follows. Sequenced IIC protein domains of fructose permeases (IICFruf) include those from Escherichia coli (Eco; Prior & Kornberg, 1988), Rhodobacter capsulatus (Rca; Wu & Saier, 1990) and Xanthomonas campestris (Xca; de Crecy-Lagard et al., 1991). The IIC protein domain of the Frv permease of E. coli [IICFruf(Eco); Plunkett et al., 1993; Reizer et al., 1994a] and the IIC protein of the Frv permease of E. coli [IICFruf(Eco); Blattner et al., 1993] are also represented. For a multiple alignment of all fructose-specific and fructose IIC-like proteins or protein domains except IICFruf(Eco), see Fig. 3 in Reizer et al. (1994a).

All of these ORFs were analysed for sequence similarity to proteins in the current databases, and homologous proteins identified are discussed below for the individual proteins. When appropriate, multiple alignments, mean similarity plots, binary comparison tables and phylogenetic trees were derived. Data not presented in this paper will be supplied upon request to M. H. Saier, Jr.

Enzyme IAni

The amino acid sequence of the N-terminal 560 residues of Enzyme IAni (EIAni) exhibits striking sequence similarity to all of the Enzymes I of the PTS throughout most of its length (503–556-residue overlap). It was 34–37% identical to all previously sequenced Enzymes I except that from Alcaligenes eutrophus, to which it was 32% identical. Comparison scores, determined with the RDF2 program (Pearson & Lipman, 1988) gave values ranging between 79 and 121 sd, thereby establishing homology. A multiple alignment of 10 distinct Enzymes I, the E. coli phosphoenolpyruvate synthase and the four sequenced pyruvate phosphate dikinases (from bacteria, protozoa and plants) revealed that almost all fully conserved residues in the Enzyme I family, and all established catalytic residues (the TSH motif that includes the phosphorylated histidine and the C-terminal CG residue pair that includes the conserved cysteine, both believed to be of catalytic significance), are also conserved in EIAni (see Reizer et al., 1993, for a multiple alignment of 11 members of this family). Comparison scores obtained with phosphoenolpyruvate synthase of E. coli (19 sd) and various pyruvate phosphate dikinases (8–15 sd) were also of sufficient magnitude to establish homology. Phylogenetic tree construction surprisingly revealed that EIAni is more similar to the Enzymes I of Gram-positive bacteria than to those of enteric bacteria (Fig. 2). It was most distant from Enzymes I of A. eutrophus and Rhodobacter capsulatus. The IIA domain (147 residues) at the C-terminus of EIAni (here designated IIAniFruf) showed greatest sequence similarity to IIAniFruf of E. coli (28% identity in an amino acid overlap of 144 residues; 20 sd).

Transaldolase homologue (TalC)

Adjacent to the ptsA gene, and downstream from it, is an ORF, orf J220, which we have designated talC (Fig. 1). The encoded protein exhibits 37% identity in 181 residues (28 sd) to a protein encoded by an ORF in Bacillus subtilis (orfU) that maps adjacent to fructose-1,6-bisphosphate aldolase (Trach et al., 1988). As revealed by the statistical analyses recorded in Table 1, OrfU is homologous to transaldolases of bacteria, yeast and man. TalC exhibits only very weak sequence similarity with the latter proteins, but based on the superfamily principle (Doolittle, 1986) all proteins included in Table 1 are members of a single family, here designated the transaldolase family (TAF). A signature sequence was derived for this family and was shown to be specific for the seven known members of the family. The signature sequence is: TTNPS[LIVM][LIVMA] (residues in brackets represent alternative residues at a particular position).
alcohol dehydrogenases (data not presented).

Table 2. Binary statistical comparisons of members of the fructose IIB-like protein or protein domain family (FBF)

<table>
<thead>
<tr>
<th></th>
<th>IIBFrw (Rca)</th>
<th>IIBFrw (Xca)</th>
<th>IIBFrw (Eco)</th>
<th>ORF107 (Mge)</th>
<th>IIBFrw (Eco)</th>
<th>IIBFrw (ECO)</th>
<th>IIBFrw (Rca)</th>
<th>IIBFrw (Xca)</th>
<th>IIBFrw (Eco)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IIBFrw (Eco)</td>
<td>57 (98)</td>
<td>55 (98)</td>
<td>48 (85)</td>
<td>51 (85)</td>
<td>51 (95)</td>
<td>37 (95)</td>
<td>43 (88)</td>
<td>28 (83)</td>
<td>21 (28)</td>
</tr>
<tr>
<td>(100)</td>
<td></td>
<td>[31]</td>
<td>[23]</td>
<td>[24]</td>
<td>[28]</td>
<td>[23]</td>
<td>[25]</td>
<td>[10]</td>
<td>[1]</td>
</tr>
<tr>
<td>IIBFrw (Rca)</td>
<td>54 (97)</td>
<td>52 (86)</td>
<td>48 (85)</td>
<td>44 (93)</td>
<td>42 (95)</td>
<td>42 (71)</td>
<td>47 (85)</td>
<td>24 (86)</td>
<td>26 (43)</td>
</tr>
<tr>
<td>(99)</td>
<td></td>
<td>[35]</td>
<td>[20]</td>
<td>[26]</td>
<td>[28]</td>
<td>[26]</td>
<td>[22]</td>
<td>[13]</td>
<td>[3]</td>
</tr>
<tr>
<td>IIBFrw (Xca)</td>
<td>56 (80)</td>
<td>49 (86)</td>
<td>42 (95)</td>
<td>42 (71)</td>
<td>42 (71)</td>
<td>41 (87)</td>
<td>27 (85)</td>
<td>21 (38)</td>
<td>21 (38)</td>
</tr>
<tr>
<td>(99)</td>
<td></td>
<td>[25]</td>
<td>[25]</td>
<td>[28]</td>
<td>[16]</td>
<td>[21]</td>
<td>[10]</td>
<td>[2]</td>
<td>[2]</td>
</tr>
<tr>
<td>IIBFrw (Eco)</td>
<td>46 (84)</td>
<td>52 (85)</td>
<td>46 (70)</td>
<td>45 (85)</td>
<td>45 (85)</td>
<td>22 (83)</td>
<td>29 (28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(92)</td>
<td></td>
<td>[24]</td>
<td>[30]</td>
<td>[19]</td>
<td>[20]</td>
<td>[6]</td>
<td>[3]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF107 (Mge)</td>
<td>49 (86)</td>
<td>44 (71)</td>
<td>38 (85)</td>
<td>26 (69)</td>
<td>26 (69)</td>
<td>67 (3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(107)</td>
<td></td>
<td>[24]</td>
<td>[17]</td>
<td>[16]</td>
<td>[5]</td>
<td>[-1]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IIBFrw (Eco)</td>
<td>43 (103)</td>
<td>41 (91)</td>
<td>26 (88)</td>
<td>26 (50)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(106)</td>
<td></td>
<td>[28]</td>
<td>[18]</td>
<td>[9]</td>
<td>[3]</td>
<td>[2]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IIBFrw (Eco)</td>
<td>38 (74)</td>
<td>37 (27)</td>
<td>15 (34)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(113)</td>
<td></td>
<td>[13]</td>
<td>[3]</td>
<td>[2]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IIBFrw (Rca)</td>
<td>31 (88)</td>
<td>24 (33)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(90)</td>
<td></td>
<td>[10]</td>
<td>[1]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IIBFrw (Xca)</td>
<td>20 (20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(90)</td>
<td></td>
<td>[0.1]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Glycerol dehydrogenase (GldA)

Downstream from and adjacent to talC is the gldA gene encoding the E. coli glycerol dehydrogenase (Truniger & Boos, 1994). The encoded protein exhibits 50% identity in a 362-residue overlap to the glycerol dehydrogenase of Bacillus stearothermophilus (comparison score of 105 sd) (Mallinder et al., 1992). It also proved to be homologous to propanediol oxidoreductase (lactaldehyde reductase; FucO) of E. coli (21% identity in a 274-residue overlap, comparison score of 11 sd (Chen et al., 1989; Lu & Lin, 1989). Lactaldehyde reductase proved to be homologous to many oxidoreductases including numerous sequenced alcohol dehydrogenases (data not presented).

ORF o109

The putative protein encoded by orf o109 (ORF o109) possesses an N-terminal 70-residue segment that is rich in serine (30%). However, the encoded sequence does not exhibit significant similarity to those of other proteins in the current databases, and it lacks a demonstrable helix-turn-helix motif.

IICFrw (FrwC)

ORF o359 (herein designated IICFrw or FrwC) is a hydrophobic protein of 359 residues. This protein proved to be homologous to the IIC domains of all sequenced Enzymes II specific for fructose as well as to the IIC domain of the recently described fructose-like PTS protein, FrvC (comparison scores ranging from 29 to 49 sd). We refer to this family of proteins as the fructose-specific or fructose-like IIC family (FCF).

The multiple alignment of these five protein sequences revealed many fully conserved residues, including the glutamate that is conserved in most Enzymes II (Wu & Saier, 1990) and is essential for function in the mannitol Enzyme IIC (Jacobsen & Saraceni-Richards, 1993; for the multiple alignment of all fructose-specific IIC domains and fructose IIC-like domains except IICFrw see Fig. 3 in
Reizer et al., 1994a). Several conserved hydrophilic residues in members of this family may be important for catalysis, while the strongly conserved hydrophobic regions may correspond to transmembrane α-helical spanners, similar to those of other PTS permeases (Sugiyama et al., 1991; Buhr & Erni, 1993).

The mean hydropathy and similarity plots for these five proteins are presented in Fig. 3. It can be seen that in general, the putative hydrophobic spanners are better conserved than are the hydrophilic loops. The most conserved region was used to construct a signature sequence for this family of proteins (see Fig. 3), and this sequence, DMGGP[LIVM]NKXA, proved to be specific to the five members of the family when the SwissProt database (version 28) was screened. Six to nine transmembrane spanners were predicted on the basis of the method of Kyte & Doolittle (1982) for the various members of the family.

The phylogenetic tree for these proteins is shown in Fig. 4. IIC\textsuperscript{FrW} (Eco) is apparently distant from all other members, whereas the three biochemically characterized fructose-specific Enzymes IIC are closely related to each other. This fact suggests that IIC\textsuperscript{FrW} and IIC\textsuperscript{FrW'} may possess a function dissimilar from those of the biochemically well-characterized fructose-specific Enzymes IIC.

IIB\textsuperscript{FrW} and IIB\textsuperscript{FrW'} (FrwB and FrwD)

The frw operon contains two ORFs that are homologous to each other as well as to the IIB\textsuperscript{FrW} domains of fructose-specific Enzymes IIC. Correlating with the fact that all biochemically characterized Enzymes IIC specific for fructose contain two adjacent IIB domains, the frw operon contains two IIB\textsuperscript{FrW} domains as separate polypeptide chains encoded by distinct genes.

The binary statistical comparison scores for the ten members of the IIB\textsuperscript{FrW} family (FBF) are presented in Table 2. These scores are sufficient in magnitude to establish homology of IIB\textsuperscript{FrW} and IIB\textsuperscript{FrW'} to all other members of the family except the N-terminal IIB domain (IIB\textsuperscript{FrW'}) of the functionally characterized fructose-specific Enzyme II of \textit{E. coli}. The data in Table 2 reveal that IIB\textsuperscript{FrW} exhibits greater sequence similarity to other...
Novel phosphotransferase system genes

Fig. 6. Evolutionary tree of members of the fructose IIB-like family (FBF). Relative branch lengths, proportional to phylogenetic distance, are shown adjacent to the branches. Abbreviations are as indicated in the legend to Table 2. Construction of the tree was as described by Reizer & Reizer (1994) using the progressive alignment method of Feng & Doolittle (1990).

members of the family than does \text{IIB}^{Frw'} (hence the designation of the latter protein as \text{IIB}^{Frw'}). It is noteworthy that both of the Frw IIB proteins are much more similar to the catalytic IIB domains of the fructose Enzymes II than to the noncatalytic, N-terminal IIB' domains of the same Enzymes II.

The multiple alignment of all but one of these protein sequences is presented in Fig. 5. The noncatalytic N-terminal \text{IIB}^{Frw'} sequence of the \textit{E. coli} fructose Enzyme II was omitted due to its extensive divergence (Wu et al., 1990; see similarity scores in Table 2). Only three residues are conserved in all members of the family (E, G and A at positions 38, 44 and 57, respectively). Several residues are conserved in all but one member of the family (boxed residues), and both proteins, \text{FrwB} and \text{FrwD}, contain the residues proposed to be essential for the phosphoryl transfer reactions involved in sugar uptake. These include the cysteine at position 10 and the histidine at position 16 of the multiple alignment. As is apparent from the alignment, the N-terminal regions of these proteins are more strongly conserved than the C-terminal regions. The phylogenetic tree of the fructose IIB-like family (FBF) is shown in Fig. 6. All catalytically active \text{IIB}^{Frw} domains cluster tightly together on this tree. The \text{IIB}^{Frw}, the \text{IIB}^{Frw'} and the \textit{R. capsulatus} \text{IIB}^{Frw} are about equidistant from this cluster while the \text{IIB}^{Frw'} domain is much more distant. This result suggests that the \text{frwB} gene product may provide the primary function of the Frw Enzyme II complex.

A pyruvate formate-lyase homologue (PflD)

Downstream of \textit{frwB} is \textit{orf} 6765 (\textit{pflD}), a gene that encodes a protein that is homologous throughout most of its length to pyruvate formate-lyase (PflB) of \textit{E. coli} (759 residues). The two proteins exhibit 21\% identity in a 518-residue overlap with a binary comparison score of 44 sd. Other homologues were also identified (Table 3). The glycyl residue that is converted to a free radical by the Pfl-activating enzyme (Wagner et al., 1992) is conserved in PflD. Surrounding this glycyl residue at the C-terminal region of these proteins is a sequence that is fully conserved between these two proteins and is largely conserved in the other homologues listed in Table 3. This sequence is: Y[AP][NQ][IL][TV][IV][RV][SA][GYS][A-V][SA][A-V]XF (amino acid at a position in which the residue is not specified is denoted by X). Although an anaerobic \textit{E. coli} ribonucleotide reductase, NdR, exhibits similar mechanisms of activation and action (Sun et al., 1993), it is not demonstrably homologous to any of the members of the Pfl family listed in Table 3.

A pyruvate-formate-lyase-activating enzyme homologue (PflC)

Downstream of \textit{pflD} is \textit{orf} 6315 (\textit{pflC}), a gene that encodes a homologue of the known pyruvate-formate-lyase-activating enzyme, PflA of \textit{E. coli}, 245 residues in length. These two proteins exhibit 27\% identity in a 197 amino acid overlap with a comparison score of 27 sd. The \textit{E. coli} PflB-activating enzyme is also homologous to a phage T4 ORF (ORF156) (35\% identity in a 71-residue overlap; 13 sd). The results reported in this and the previous section suggest that this gene cluster encodes an enzyme that exhibits mechanisms of activation and action like those of the \textit{E. coli} PflB. In spite of the similarities noted above between PflD and ORF 6315 with PflB and PflA, respectively, the low percentage identities observed between these pairs of homologous proteins do not allow the unequivocal conclusion that the former enzyme is a pyruvate formate-lyase.
Table 3. Binary comparisons of the six pyruvate formate lyase (PFL) family members

Values in parentheses below the designation of a protein refer to the number of residues in that protein. The **FASTA** program using the dipeptide identities mode (ktup = 2) (Pearson & Lipman, 1988) was used to assess similarities of the indicated proteins. Values presented in the table which are not in brackets or parentheses represent percentage identities for segments having the numbers of compared residues indicated in parentheses. Comparison scores in standard deviations using the RDF2 program (Pearson & Lipman, 1988) and 150 shuffles are given in brackets below the values for percentage identity. Abbreviations and references to the published sequences are as follows. The **ORF** (ORF0765) in the Frw PTS region of *E. coli* [ORF0765 (Eco); Blattner et al., 1993], pyruvate formate lyase of *E. coli* [PFB (Eco); Rodel et al., 1988], the partially sequenced pyruvate formate lyase of *Chlamydomonas reinhardtii* [Pf (Cre); submitted to EMBL (accession no. S24997) by F. Dumont], and the hypothetical proteins of *Serratia liquefaciens* [ORF106 (Sli); Givskov et al., 1988], of *E. coli* [ORF127 (Eco); Varshney et al., 1988], and of the phage T4 [ORF120 (T4); Valerie et al., 1986].

<table>
<thead>
<tr>
<th>Protein</th>
<th>Length (residues)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF0765 (Eco)</td>
<td>21 (518)</td>
</tr>
<tr>
<td>Pf (Cre)</td>
<td>35 (192)</td>
</tr>
<tr>
<td>ORF106 (Sli)</td>
<td>62 (200)</td>
</tr>
<tr>
<td>ORF127 (Eco)</td>
<td>68 (60)</td>
</tr>
<tr>
<td>ORF120 (T4)</td>
<td>68 (60)</td>
</tr>
</tbody>
</table>

**DISCUSSION**

In 1987 we demonstrated that *Salmonella typhimurium*, deleted of the *pts* operon which encodes Enzyme I and HPr, could be mutated to allow expression of a novel and presumably cryptic Enzyme I. This novel enzyme could readily substitute for the deleted Enzyme I (Chin et al., 1987). A mutation which allowed expression of the encoding gene was mapped by transduction and shown to be localized between *crr* and *cysM* at 49.5 min on the *Salmonella* chromosome. This putative gene was designated *ptsJ*.

We have recently sequenced a 4.5 kbp region of the *Salmonella* chromosome between *cysM* and *crr* and have identified an ORF which encodes a protein that possesses an identifiable N-terminal helix–turn–helix motif homologous to those of members of the *B. subtilis* gluconate repressor (GntR) family (Titgemeyer et al., 1994; Buck & Guest, 1989; Reizer et al., 1991a). No ORF encoding an Enzyme I-like protein was found in this region. The structural gene for the cryptic Enzyme I therefore had to map elsewhere on the chromosome.

Geese (1989) confirmed the observation of Chin et al. (1987) that a mutation in *S. typhimurium* near the *crr* gene allowed expression of a cryptic Enzyme I. Similarly, Kornberg (1990) reported that mutations in the genetic background of an *E. coli* strain that lacks the *ptsH* and *ptsI* genes restore fructose utilization. In the former work, mutation in addition to that in *ptsJ* was thought to be requisite for this expression, and this gene was claimed to cotransduce with the *fru* operon (Geese, 1989). Extensive sequencing around the *fru* operon (G. M. Church and others, data submitted to GenBank, accession number U00007) has not yet revealed the presence of an Enzyme I-encoding gene in this region.

By contrast, genome sequencing efforts in the 89.2–92.8 region (Blattner et al., 1993) revealed an ORF encoding an Enzyme I-like protein linked to a IIA*^Fru* protein domain at its C-terminus. Analysis of other ORFs in the same gene cluster revealed that other PTS-protein-encoding genes, two encoding IIBFrU-like proteins, and one encoding a IICFrU-like protein, are present. This gene cluster therefore encodes all of the constituents of a fructose-specific PTS (Enzyme I, IIA*^Fru*, IIB*^Fru*, IIB*^Fru* and IIC*^Fru*), except for an HPr-like protein. In this regard it is interesting to note that the *fru* operon of enteric bacteria encodes a protein (DTP) in which an HPr-like domain is fused to a IIA*^Fru* domain (Geese et al., 1989; Wu et al., 1990; Reizer et al., 1994b). This HPr-like domain (termed FPr or fructose-inducible HPr) can fully replace HPr in sugar phosphorylation (Sutrina et al., 1988). Further, many bacteria possess only a fructose-specific PTS (see Romano & Saier, 1992, for a summary of the data) and in some of these, the Enzyme I-like domain (C-terminus) is linked to an HPr domain (middle) and a IIA*^Fru* domain (N-terminus) (Wu et al., 1990; de Crecy-Lagard et al., 1991). It has been postulated that the fructose PTS was the primordial system (Saier, 1977; Saier et al., 1985).

In *E. coli* and other bacteria, several gene products homologous to proteins of the fructose PTS have now
been found. The fru operon of this organism encodes (a) DTP, the three-domain diphosphoryl transfer protein referred to above that includes the IIA\textsuperscript{Fru} and FPr domains, (b) a fructose-1-phosphate kinase, and (c) a fructose-specific Enzyme II having tandemly duplicated IIB domains as well as a IIC domain, all in a single polypeptide chain. *Rhodobacter capsulatus* and *Xanthomonas campestris* possess similar systems but with Enzyme I, HPr and IIA\textsuperscript{Fru} linked within a single polypeptide chain (Wu et al., 1990; de Crecy-Lagard et al., 1991). The recently described frw operon has a gene encoding a IIA\textsuperscript{Fru}-like protein as well as a gene encoding a fused IIB\textsuperscript{Fru}-IIC\textsuperscript{Fru} like protein. This last-mentioned protein has only one IIB domain (see Fig. 7 in Reizer et al., 1994a, for a schematic depiction of these proteins). Finally, the present analyses have revealed a third frw-like gene cluster with four distinct PTS-protein-encoding genes, one for the Enzyme I-IIA\textsuperscript{Fru}-like fusion protein, two encoding IIB\textsuperscript{Fru}-like proteins, and one encoding a IIC\textsuperscript{Fru}-like protein. In terms of its PTS-protein-domain-encoding content, this frw gene cluster therefore resembles the fru operon rather than the frw operon.

The diversity of the fructose-like PTS proteins noted above evidently resulted from domain shuffling and splicing which must have occurred during the evolution of these and other PTS proteins (Saier & Reizer, 1992). Thus, in various systems, IIA, IIB and IIC may be linked together in a single polypeptide chain or they may be found as two or three autonomous polypeptide chains encoded by distinct genes. Further, the various domains can be linked to each other in different but non-random orders (Saier & Reizer, 1992, 1994b).

The Enzyme I-IIA\textsuperscript{Fru} protein reported here is unique to currently recognized PTS proteins in its domain content and order. Interestingly, previously sequenced IIA domains were always found in Enzyme II-containing polypeptide chains as C-terminal domains (Saier & Reizer, 1992, 1994b). By contrast, all previously known general energy-coupling PTS proteins bearing a IIA domain exhibit the IIA domain at their N-termini. Enzyme I\textsuperscript{Ant} is therefore the first general energy-coupling protein of the PTS with a C-terminal IIA domain.

Other enzymes encoded within the frw gene cluster include a glycerol dehydrogenase, a transaldolase homolog, a pyruvate formate-lyase (PFL) homologue and a PFL-activating enzyme homologue. The well-characterized PFL is known to function only under anaerobic conditions (Iuchi & Lin, 1993; Sapers & Suppmann, 1992). Further, several bacteria induce synthesis of PTS proteins only under anaerobic conditions (Pelliccione et al., 1979; S. Klinker & M. H. Saier, Jr, unpublished results). It would therefore be reasonable to propose that proteins encoded in the frw gene cluster function primarily in anaerobic carbon metabolism.

The functions of the proteins encoded within the frw gene cluster are at present unknown. Will the included PTS proteins prove to have significance with respect to sugar uptake, chemotaxis, or regulation? Alternatively, will they prove to be cryptic genes with no identifiable function (Hall et al., 1983; Parker & Hall, 1990)? Is Enzyme I\textsuperscript{Ant} the Enzyme I expressed in ptsJ mutants of *S. typhimurium* (Chin et al., 1987)? Both genetic and biochemical approaches are currently under way to answer these important questions.

**ACKNOWLEDGEMENTS**

We are grateful to Mary Beth Hiller for invaluable assistance provided in the preparation of this manuscript. This work was supported by US Public Health Service grants 5RO1AI 21702 and 2RO1AI 14176 from the National Institute of Allergy and Infectious Diseases.

**REFERENCES**


Buck, D. & Guest, J. R. (1989). Overexpression and site-directed mutagenesis of the succinyl-CoA synthetase of *Escherichia coli* and nucleotide sequence of a gene (g30) that is adjacent to the suc operon. *Biochem J* 260, 737–747.


Doolittle, R. F. (1986). Of Urfs and Orfs: a Primer on How to Analyze


Novel phosphotransferase system genes


Received 12 July 1994; revised 19 October; accepted 11 November 1994.